



# Cloudera Impala



## **Important Notice**

(c) 2010-2015 Cloudera, Inc. All rights reserved.

Cloudera, the Cloudera logo, Cloudera Impala, and any other product or service names or slogans contained in this document are trademarks of Cloudera and its suppliers or licensors, and may not be copied, imitated or used, in whole or in part, without the prior written permission of Cloudera or the applicable trademark holder.

Hadoop and the Hadoop elephant logo are trademarks of the Apache Software Foundation. All other trademarks, registered trademarks, product names and company names or logos mentioned in this document are the property of their respective owners. Reference to any products, services, processes or other information, by trade name, trademark, manufacturer, supplier or otherwise does not constitute or imply endorsement, sponsorship or recommendation thereof by us.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Cloudera.

Cloudera may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Cloudera, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property. For information about patents covering Cloudera products, see <http://tiny.cloudera.com/patents>.

The information in this document is subject to change without notice. Cloudera shall not be liable for any damages resulting from technical errors or omissions which may be present in this document, or from use of this document.

**Cloudera, Inc.**  
**1001 Page Mill Road Bldg 2**  
**Palo Alto, CA 94304**  
**[info@cloudera.com](mailto:info@cloudera.com)**  
**US: 1-888-789-1488**  
**Intl: 1-650-362-0488**  
**[www.cloudera.com](http://www.cloudera.com)**

## **Release Information**

Version: Impala 2.2.x / CDH 5.4.x  
Date: July 14, 2015

# Table of Contents

<b>Introducing Cloudera Impala.....</b>	<b>16</b>
Impala Benefits.....	16
How Cloudera Impala Works with CDH.....	16
Primary Impala Features.....	17
<b>Impala Concepts and Architecture.....</b>	<b>18</b>
Components of the Impala Server.....	18
<i>The Impala Daemon</i> .....	18
<i>The Impala Statestore</i> .....	18
<i>The Impala Catalog Service</i> .....	19
Programming Impala Applications.....	19
<i>Overview of the Impala SQL Dialect</i> .....	19
<i>Overview of Impala Programming Interfaces</i> .....	20
How Impala Fits Into the Hadoop Ecosystem.....	21
<i>How Impala Works with Hive</i> .....	21
<i>Overview of Impala Metadata and the Metastore</i> .....	21
<i>How Impala Uses HDFS</i> .....	22
<i>How Impala Uses HBase</i> .....	22
<b>Planning for Impala Deployment.....</b>	<b>23</b>
Cloudera Impala Requirements.....	23
<i>Product Compatibility Matrix</i> .....	23
<i>Supported Operating Systems</i> .....	23
<i>Hive Metastore and Related Configuration</i> .....	23
<i>Java Dependencies</i> .....	24
<i>Networking Configuration Requirements</i> .....	24
<i>Hardware Requirements</i> .....	24
<i>User Account Requirements</i> .....	24
Cluster Sizing Guidelines for Impala.....	25
Guidelines for Designing Impala Schemas.....	27
<b>Installing Impala.....</b>	<b>30</b>
What is Included in an Impala Installation.....	30
Impala Installation Procedure for CDH 4 Users.....	30
Installing Impala with Cloudera Manager.....	31

Installing Impala without Cloudera Manager.....	31
---	----

## **Managing Impala.....34**

Post-Installation Configuration for Impala.....	34
Configuring Impala to Work with ODBC.....	37
Configuring Impala to Work with JDBC.....	39
<i>Configuring the JDBC Port.....</i>	39
<i>Choosing the JDBC Driver.....</i>	39
<i>Enabling Impala JDBC Support on Client Systems.....</i>	40
<i>Establishing JDBC Connections.....</i>	41

## **Upgrading Impala.....43**

Upgrading Impala through Cloudera Manager - Parcels.....	43
Upgrading Impala through Cloudera Manager - Packages.....	44
Upgrading Impala without Cloudera Manager.....	45

## **Starting Impala.....47**

Starting Impala through Cloudera Manager.....	47
Starting Impala from the Command Line.....	47
Modifying Impala Startup Options.....	48
<i>Configuring Impala Startup Options through Cloudera Manager.....</i>	48
<i>Configuring Impala Startup Options through the Command Line.....</i>	48
<i>Checking the Values of Impala Configuration Options.....</i>	51
<i>Startup Options for impalad Daemon.....</i>	51
<i>Startup Options for statestored Daemon.....</i>	51
<i>Startup Options for catalogd Daemon.....</i>	51

## **Impala Tutorials.....52**

Tutorials for Getting Started.....	52
<i>Explore a New Impala Instance.....</i>	52
<i>Load CSV Data from Local Files.....</i>	57
<i>Point an Impala Table at Existing Data Files.....</i>	59
<i>Describe the Impala Table.....</i>	60
<i>Query the Impala Table.....</i>	60
<i>Data Loading and Querying Examples.....</i>	61
Advanced Tutorials.....	63
<i>Attaching an External Partitioned Table to an HDFS Directory Structure.....</i>	63
<i>Switching Back and Forth Between Impala and Hive.....</i>	65
<i>Cross Joins and Cartesian Products with the CROSS JOIN Operator.....</i>	66

<b>Impala Administration.....</b>	<b>69</b>
Admission Control and Query Queuing.....	69
<i>Overview of Impala Admission Control.....</i>	70
<i>How Impala Admission Control Relates to YARN.....</i>	70
<i>How Impala Schedules and Enforces Limits on Concurrent Queries.....</i>	71
<i>How Admission Control works with Impala Clients (JDBC, ODBC, HiveServer2).....</i>	71
<i>Configuring Admission Control.....</i>	72
<i>Guidelines for Using Admission Control.....</i>	77
Integrated Resource Management with YARN.....	78
<i>The Llama Daemon.....</i>	78
<i>Controlling Resource Estimation Behavior.....</i>	78
<i>Checking Resource Estimates and Actual Usage.....</i>	79
<i>How Resource Limits Are Enforced.....</i>	79
<i>Enabling Resource Management for Impala.....</i>	79
<i>Limitations of Resource Management for Impala.....</i>	80
Setting Timeout Periods for Daemons, Queries, and Sessions.....	81
Using Impala through a Proxy for High Availability.....	82
<i>Overview of Proxy Usage and Load Balancing for Impala.....</i>	82
<i>Special Proxy Considerations for Clusters Using Kerberos.....</i>	83
<i>Example of Configuring HAProxy Load Balancer for Impala.....</i>	84
Managing Disk Space for Impala Data.....	85
<b>Impala Security.....</b>	<b>87</b>
Security Guidelines for Impala.....	87
Securing Impala Data and Log Files.....	88
Installation Considerations for Impala Security.....	89
Securing the Hive Metastore Database.....	89
Securing the Impala Web User Interface.....	89
Configuring SSL for Impala.....	90
<i>Using Cloudera Manager.....</i>	90
<i>Using the Command Line.....</i>	90
Enabling Sentry Authorization for Impala.....	91
<i>The Sentry Privilege Model.....</i>	91
<i>Starting the impalad Daemon with Sentry Authorization Enabled.....</i>	92
<i>Using Impala with the Sentry Service (CDH 5.1 or higher only).....</i>	93
<i>Using Impala with the Sentry Policy File.....</i>	93
<i>Setting Up Schema Objects for a Secure Impala Deployment.....</i>	98
<i>Privilege Model and Object Hierarchy.....</i>	99
<i>Debugging Failed Sentry Authorization Requests.....</i>	101
<i>Configuring Per-User Access for Hue.....</i>	102
<i>Managing Sentry for Impala through Cloudera Manager.....</i>	102

<i>The DEFAULT Database in a Secure Deployment</i> .....	103
<b>Impala Authentication.....</b>	<b>103</b>
<i>Enabling Kerberos Authentication for Impala.....</i>	<i>103</i>
<i>Enabling LDAP Authentication for Impala.....</i>	<i>106</i>
<i>Using Multiple Authentication Methods with Impala.....</i>	<i>108</i>
<b>Auditing Impala Operations.....</b>	<b>108</b>
<i>Durability and Performance Considerations for Impala Auditing.....</i>	<i>108</i>
<i>Format of the Audit Log Files.....</i>	<i>109</i>
<i>Which Operations Are Audited.....</i>	<i>109</i>
<i>Reviewing the Audit Logs.....</i>	<i>109</i>
<b>Viewing Lineage Information for Impala Data.....</b>	<b>110</b>

## **Impala SQL Language Reference.....111**

<b>Comments.....</b>	<b>111</b>
<b>Data Types.....</b>	<b>111</b>
<i>BIGINT Data Type.....</i>	<i>112</i>
<i>BOOLEAN Data Type.....</i>	<i>113</i>
<i>CHAR Data Type (CDH 5.2 or higher only).....</i>	<i>114</i>
<i>DECIMAL Data Type (CDH 5.1 or higher only).....</i>	<i>116</i>
<i>DOUBLE Data Type.....</i>	<i>125</i>
<i>FLOAT Data Type.....</i>	<i>125</i>
<i>INT Data Type.....</i>	<i>126</i>
<i>REAL Data Type.....</i>	<i>127</i>
<i>SMALLINT Data Type.....</i>	<i>128</i>
<i>STRING Data Type.....</i>	<i>128</i>
<i>TIMESTAMP Data Type.....</i>	<i>130</i>
<i>TINYINT Data Type.....</i>	<i>135</i>
<i>VARCHAR Data Type (CDH 5.2 or higher only).....</i>	<i>136</i>
<b>Literals.....</b>	<b>138</b>
<i>Numeric Literals.....</i>	<i>139</i>
<i>String Literals.....</i>	<i>140</i>
<i>Boolean Literals.....</i>	<i>141</i>
<i>Timestamp Literals.....</i>	<i>141</i>
<i>NULL.....</i>	<i>141</i>
<b>SQL Operators.....</b>	<b>142</b>
<i>Arithmetic Operators.....</i>	<i>142</i>
<i>BETWEEN Operator.....</i>	<i>143</i>
<i>Comparison Operators.....</i>	<i>144</i>
<i>EXISTS Operator.....</i>	<i>144</i>
<i>IN Operator.....</i>	<i>147</i>
<i>IS NULL Operator.....</i>	<i>149</i>
<i>LIKE Operator.....</i>	<i>149</i>
<i>Logical Operators.....</i>	<i>150</i>

<i>REGEXP Operator</i> .....	152
<i>RLIKE Operator</i> .....	153
Schema Objects and Object Names.....	154
<i>Aliases</i> .....	154
<i>Databases</i> .....	155
<i>Functions</i> .....	155
<i>Identifiers</i> .....	155
<i>Tables</i> .....	156
<i>Views</i> .....	158
SQL Statements.....	162
<i>DDL Statements</i> .....	162
<i>DML Statements</i> .....	163
<i>ALTER TABLE Statement</i> .....	164
<i>ALTER VIEW Statement</i> .....	168
<i>COMPUTE STATS Statement</i> .....	170
<i>CREATE DATABASE Statement</i> .....	175
<i>CREATE FUNCTION Statement</i> .....	177
<i>CREATE ROLE Statement (CDH 5.2 or higher only)</i> .....	179
<i>CREATE TABLE Statement</i> .....	180
<i>CREATE VIEW Statement</i> .....	186
<i>DESCRIBE Statement</i> .....	188
<i>DROP DATABASE Statement</i> .....	192
<i>DROP FUNCTION Statement</i> .....	192
<i>DROP ROLE Statement (CDH 5.2 or higher only)</i> .....	193
<i>DROP STATS Statement</i> .....	193
<i>DROP TABLE Statement</i> .....	197
<i>DROP VIEW Statement</i> .....	198
<i>EXPLAIN Statement</i> .....	199
<i>GRANT Statement (CDH 5.2 or higher only)</i> .....	201
<i>INSERT Statement</i> .....	202
<i>INVALIDATE METADATA Statement</i> .....	210
<i>LOAD DATA Statement</i> .....	212
<i>REFRESH Statement</i> .....	215
<i>REVOKE Statement (CDH 5.2 or higher only)</i> .....	217
<i>SELECT Statement</i> .....	218
<i>SET Statement</i> .....	241
<i>SHOW Statement</i> .....	242
<i>USE Statement</i> .....	256
Built-in Functions.....	257
<i>Impala Mathematical Functions</i> .....	259
<i>Impala Type Conversion Functions</i> .....	267
<i>Impala Date and Time Functions</i> .....	268
<i>Impala Conditional Functions</i> .....	275
<i>Impala String Functions</i> .....	279

<i>Impala Miscellaneous Functions</i> .....	284
<i>Impala Aggregate Functions</i> .....	284
<i>Impala Analytic Functions</i> .....	296
<i>Impala User-Defined Functions (UDFs)</i> .....	308
SQL Differences Between Impala and Hive.....	323
<i>HiveQL Features not Available in Impala</i> .....	323
<i>Semantic Differences Between Impala and HiveQL Features</i> .....	324
Porting SQL from Other Database Systems to Impala.....	326
<i>Porting DDL and DML Statements</i> .....	326
<i>Porting Data Types from Other Database Systems</i> .....	326
<i>SQL Statements to Remove or Adapt</i> .....	328
<i>SQL Constructs to Doublecheck</i> .....	330
<i>Next Porting Steps after Verifying Syntax and Semantics</i> .....	331

## Using the Impala Shell (`impala-shell` Command).....**332**

<code>impala-shell</code> Configuration Options.....	332
<i>Summary of impala-shell Configuration Options</i> .....	332
<i>impala-shell Configuration File</i> .....	334
Connecting to impalad through <code>impala-shell</code> .....	335
Running Commands in <code>impala-shell</code> .....	336
<code>impala-shell</code> Command Reference.....	337
Query Options for the SET Statement.....	339
<i>ABORT_ON_DEFAULT_LIMIT_EXCEEDED Query Option</i> .....	339
<i>ABORT_ON_ERROR Query Option</i> .....	339
<i>ALLOW_UNSUPPORTED_FORMATS Query Option</i> .....	340
<i>APPX_COUNT_DISTINCT Query Option</i> .....	340
<i>BATCH_SIZE Query Option</i> .....	340
<i>COMPRESSION_CODEC Query Option</i> .....	341
<i>DEBUG_ACTION Query Option</i> .....	341
<i>DEFAULT_ORDER_BY_LIMIT Query Option</i> .....	341
<i>DISABLE_CODEGEN Query Option</i> .....	342
<i>DISABLE_UNSAFE_SPILLS Query Option</i> .....	342
<i>EXEC_SINGLE_NODE_ROWS_THRESHOLD Query Option</i> .....	342
<i>EXPLAIN_LEVEL Query Option</i> .....	343
<i>HBASE_CACHE_BLOCKS Query Option</i> .....	348
<i>HBASE_CACHING Query Option</i> .....	348
<i>MAX_ERRORS Query Option</i> .....	349
<i>MAX_IO_BUFFERS Query Option</i> .....	349
<i>MAX_SCAN_RANGE_LENGTH Query Option</i> .....	349
<i>MEM_LIMIT Query Option</i> .....	349
<i>NUM_NODES Query Option</i> .....	350
<i>NUM_SCANNER_THREADS Query Option</i> .....	350
<i>PARQUET_COMPRESSION_CODEC Query Option</i> .....	350

<i>PARQUET_FILE_SIZE</i> Query Option.....	350																																																												
<i>QUERY_TIMEOUT_S</i> Query Option.....	351																																																												
<i>REQUEST_POOL</i> Query Option.....	352																																																												
<i>RESERVATION_REQUEST_TIMEOUT</i> Query Option (CDH 5 only).....	352																																																												
<i>SUPPORT_START_OVER</i> Query Option.....	352																																																												
<i>SYNC_DDL</i> Query Option.....	352																																																												
<i>V_CPU_CORES</i> Query Option (CDH 5 only).....	353																																																												
<b>Tuning Impala for Performance.....</b>	<b>354</b>																																																												
Impala Performance Guidelines and Best Practices.....	354																																																												
Performance Considerations for Join Queries.....	357	<i>How Joins Are Processed when Statistics Are Unavailable</i> .....	358	<i>Overriding Join Reordering with STRAIGHT_JOIN</i> .....	358	<i>Examples of Join Order Optimization</i> .....	358	How Impala Uses Statistics for Query Optimization.....	363	<i>Overview of Table Statistics</i> .....	363	<i>Overview of Column Statistics</i> .....	364	<i>Overview of Incremental Statistics</i> .....	364	<i>Detecting Missing Statistics</i> .....	365	<i>Keeping Statistics Up to Date</i> .....	367	<i>Setting Statistics Manually through ALTER TABLE</i> .....	368	<i>Examples of Using Table and Column Statistics with Impala</i> .....	368	Benchmarking Impala Queries.....	371	Controlling Impala Resource Usage.....	372	Using HDFS Caching with Impala (CDH 5.1 or higher only).....	372	<i>Overview of HDFS Caching for Impala</i> .....	372	<i>Setting Up HDFS Caching for Impala</i> .....	372	<i>Enabling HDFS Caching for Impala Tables and Partitions</i> .....	373	<i>Loading and Removing Data with HDFS Caching Enabled</i> .....	374	<i>Administration for HDFS Caching with Impala</i> .....	375	<i>Performance Considerations for HDFS Caching with Impala</i> .....	376	Testing Impala Performance.....	377	Understanding Impala Query Performance - EXPLAIN Plans and Query Profiles.....	378	<i>Using the EXPLAIN Plan for Performance Tuning</i> .....	379	<i>Using the SUMMARY Report for Performance Tuning</i> .....	379	<i>Using the Query Profile for Performance Tuning</i> .....	380	Detecting and Correcting HDFS Block Skew Conditions.....	386	<b>Scalability Considerations for Impala.....</b>	<b>388</b>	Scalability Considerations for the Impala Statestore.....	388	SQL Operations that Spill to Disk.....	389	Limits on Query Size and Complexity.....	392
<i>How Joins Are Processed when Statistics Are Unavailable</i> .....	358																																																												
<i>Overriding Join Reordering with STRAIGHT_JOIN</i> .....	358																																																												
<i>Examples of Join Order Optimization</i> .....	358																																																												
How Impala Uses Statistics for Query Optimization.....	363																																																												
<i>Overview of Table Statistics</i> .....	363																																																												
<i>Overview of Column Statistics</i> .....	364																																																												
<i>Overview of Incremental Statistics</i> .....	364																																																												
<i>Detecting Missing Statistics</i> .....	365																																																												
<i>Keeping Statistics Up to Date</i> .....	367																																																												
<i>Setting Statistics Manually through ALTER TABLE</i> .....	368																																																												
<i>Examples of Using Table and Column Statistics with Impala</i> .....	368																																																												
Benchmarking Impala Queries.....	371																																																												
Controlling Impala Resource Usage.....	372																																																												
Using HDFS Caching with Impala (CDH 5.1 or higher only).....	372	<i>Overview of HDFS Caching for Impala</i> .....	372	<i>Setting Up HDFS Caching for Impala</i> .....	372	<i>Enabling HDFS Caching for Impala Tables and Partitions</i> .....	373	<i>Loading and Removing Data with HDFS Caching Enabled</i> .....	374	<i>Administration for HDFS Caching with Impala</i> .....	375	<i>Performance Considerations for HDFS Caching with Impala</i> .....	376	Testing Impala Performance.....	377	Understanding Impala Query Performance - EXPLAIN Plans and Query Profiles.....	378	<i>Using the EXPLAIN Plan for Performance Tuning</i> .....	379	<i>Using the SUMMARY Report for Performance Tuning</i> .....	379	<i>Using the Query Profile for Performance Tuning</i> .....	380	Detecting and Correcting HDFS Block Skew Conditions.....	386	<b>Scalability Considerations for Impala.....</b>	<b>388</b>	Scalability Considerations for the Impala Statestore.....	388	SQL Operations that Spill to Disk.....	389	Limits on Query Size and Complexity.....	392																												
<i>Overview of HDFS Caching for Impala</i> .....	372																																																												
<i>Setting Up HDFS Caching for Impala</i> .....	372																																																												
<i>Enabling HDFS Caching for Impala Tables and Partitions</i> .....	373																																																												
<i>Loading and Removing Data with HDFS Caching Enabled</i> .....	374																																																												
<i>Administration for HDFS Caching with Impala</i> .....	375																																																												
<i>Performance Considerations for HDFS Caching with Impala</i> .....	376																																																												
Testing Impala Performance.....	377																																																												
Understanding Impala Query Performance - EXPLAIN Plans and Query Profiles.....	378	<i>Using the EXPLAIN Plan for Performance Tuning</i> .....	379	<i>Using the SUMMARY Report for Performance Tuning</i> .....	379	<i>Using the Query Profile for Performance Tuning</i> .....	380	Detecting and Correcting HDFS Block Skew Conditions.....	386	<b>Scalability Considerations for Impala.....</b>	<b>388</b>	Scalability Considerations for the Impala Statestore.....	388	SQL Operations that Spill to Disk.....	389	Limits on Query Size and Complexity.....	392																																												
<i>Using the EXPLAIN Plan for Performance Tuning</i> .....	379																																																												
<i>Using the SUMMARY Report for Performance Tuning</i> .....	379																																																												
<i>Using the Query Profile for Performance Tuning</i> .....	380																																																												
Detecting and Correcting HDFS Block Skew Conditions.....	386																																																												
<b>Scalability Considerations for Impala.....</b>	<b>388</b>																																																												
Scalability Considerations for the Impala Statestore.....	388																																																												
SQL Operations that Spill to Disk.....	389																																																												
Limits on Query Size and Complexity.....	392																																																												

Scalability Considerations for Impala I/O.....	392
Scalability Considerations for Table Layout.....	393
<b>Partitioning for Impala Tables.....</b>	<b>394</b>
When to Use Partitioned Tables.....	394
SQL Statements for Partitioned Tables.....	394
Static and Dynamic Partitioning Clauses.....	395
Permissions for Partition Subdirectories.....	395
Partition Pruning for Queries.....	396
<i>Checking if Partition Pruning Happens for a Query.....</i>	396
<i>What SQL Constructs Work with Partition Pruning.....</i>	396
Partition Key Columns.....	397
Setting Different File Formats for Partitions.....	398
Managing Partitions.....	398
<b>How Impala Works with Hadoop File Formats.....</b>	<b>399</b>
Choosing the File Format for a Table.....	400
Using Text Data Files with Impala Tables.....	400
<i>Query Performance for Impala Text Tables.....</i>	401
<i>Creating Text Tables.....</i>	401
<i>Data Files for Text Tables.....</i>	402
<i>Loading Data into Impala Text Tables.....</i>	403
<i>Using LZO-Compressed Text Files.....</i>	404
<i>Using gzip, bzip2, or Snappy-Compressed Text Files.....</i>	406
Using the Parquet File Format with Impala Tables.....	407
<i>Creating Parquet Tables in Impala.....</i>	408
<i>Loading Data into Parquet Tables.....</i>	409
<i>Query Performance for Impala Parquet Tables.....</i>	410
<i>Snappy and GZip Compression for Parquet Data Files.....</i>	411
<i>Exchanging Parquet Data Files with Other Hadoop Components.....</i>	414
<i>How Parquet Data Files Are Organized.....</i>	414
<i>Compacting Data Files for Parquet Tables.....</i>	416
<i>Schema Evolution for Parquet Tables.....</i>	416
<i>Data Type Considerations for Parquet Tables.....</i>	417
Using the Avro File Format with Impala Tables.....	417
<i>Creating Avro Tables.....</i>	418
<i>Using a Hive-Created Avro Table in Impala.....</i>	419
<i>Specifying the Avro Schema through JSON.....</i>	419
<i>Loading Data into an Avro Table.....</i>	420
<i>Enabling Compression for Avro Tables.....</i>	420
<i>How Impala Handles Avro Schema Evolution.....</i>	420
<i>Data Type Considerations for Avro Tables.....</i>	421

Using the RCFile File Format with Impala Tables.....	422
<i>Creating RCFile Tables and Loading Data.....</i>	422
<i>Enabling Compression for RCFile Tables.....</i>	423
Using the SequenceFile File Format with Impala Tables.....	423
<i>Creating SequenceFile Tables and Loading Data.....</i>	424
<i>Enabling Compression for SequenceFile Tables.....</i>	424
<b>Using Impala to Query HBase Tables.....</b>	<b>426</b>
Overview of Using HBase with Impala.....	426
Configuring HBase for Use with Impala.....	426
Supported Data Types for HBase Columns.....	427
Performance Considerations for the Impala-HBase Integration.....	427
Use Cases for Querying HBase through Impala.....	432
Loading Data into an HBase Table.....	432
Limitations and Restrictions of the Impala and HBase Integration.....	432
Examples of Querying HBase Tables from Impala.....	433
<b>Using Impala to Query the Amazon S3 Filesystem (Unsupported Preview)...</b>	<b>436</b>
Specifying Impala Credentials to Access Data in S3.....	436
Loading Data into S3 for Impala Queries.....	436
Creating Impala Databases, Tables, and Partitions for Data Stored on S3.....	437
Internal and External Tables Located on S3.....	438
Running and Tuning Impala Queries for Data Stored on S3.....	440
<i>Understanding and Tuning Impala Query Performance for S3 Data.....</i>	440
Restrictions on Impala Support for S3.....	441
<b>Using Impala with Isilon Storage.....</b>	<b>442</b>
<b>Using Impala Logging.....</b>	<b>443</b>
Locations and Names of Impala Log Files.....	443
Managing Impala Logs through Cloudera Manager or Manually.....	444
Rotating Impala Logs.....	444
Reviewing Impala Logs.....	444
Understanding Impala Log Contents.....	445
Setting Logging Levels.....	445
Redacting Sensitive Information from Impala Log Files.....	446
<b>Troubleshooting Impala.....</b>	<b>447</b>
Troubleshooting Impala SQL Syntax Issues.....	447

Impala Troubleshooting Quick Reference.....	447
Impala Web User Interface for Debugging.....	449
<i>Debug Web UI for impalad.</i> .....	449
<b>Ports Used by Impala.....</b>	<b>451</b>
<b>Impala Reserved Words.....</b>	<b>453</b>
<b>Cloudera Impala Frequently Asked Questions.....</b>	<b>456</b>
Trying Impala.....	456
Impala System Requirements.....	457
Supported and Unsupported Functionality In Impala.....	458
How do I?.....	459
Impala Performance.....	460
Impala Use Cases.....	462
Questions about Impala And Hive.....	463
Impala Availability.....	464
Impala Internals.....	465
SQL.....	467
Partitioned Tables.....	468
HBase.....	469
<b>Cloudera Impala Release Notes.....</b>	<b>470</b>
Cloudera Impala Release Notes.....	470
New Features in Impala.....	470
<i>New Features in Impala Version 2.2.3 / CDH 5.4.3.</i> .....	470
<i>New Features in Impala for CDH 5.4.x.</i> .....	470
<i>New Features in Impala Version 2.2.0 / CDH 5.4.0.</i> .....	470
<i>New Features in Impala Version 2.1.4 / CDH 5.3.4.</i> .....	472
<i>New Features in Impala Version 2.1.3 / CDH 5.3.3.</i> .....	472
<i>New Features in Impala Version 2.1.2 / CDH 5.3.2.</i> .....	472
<i>New Features in Impala Version 2.1.1 / CDH 5.3.1.</i> .....	473
<i>New Features in Impala Version 2.1.0 / CDH 5.3.0.</i> .....	473
<i>New Features in Impala Version 2.0.5 / CDH 5.2.6.</i> .....	473
<i>New Features in Impala Version 2.0.4 / CDH 5.2.5.</i> .....	473
<i>New Features in Impala Version 2.0.3 / CDH 5.2.4.</i> .....	473
<i>New Features in Impala Version 2.0.2 / CDH 5.2.3.</i> .....	473
<i>New Features in Impala Version 2.0.1 / CDH 5.2.1.</i> .....	474
<i>New Features in Impala Version 2.0.0 / CDH 5.2.0.</i> .....	474
<i>New Features in Impala Version 1.4.4 / CDH 5.1.5.</i> .....	476

<i>New Features in Impala Version 1.4.3 / CDH 5.1.4.....</i>	476
<i>New Features in Impala Version 1.4.2 / CDH 5.1.3.....</i>	476
<i>New Features in Impala Version 1.4.1 / CDH 5.1.2.....</i>	476
<i>New Features in Impala Version 1.4.0 / CDH 5.1.0.....</i>	476
<i>New Features in Impala Version 1.3.3 / CDH 5.0.5.....</i>	478
<i>New Features in Impala Version 1.3.2 / CDH 5.0.4.....</i>	478
<i>New Features in Impala Version 1.3.1 / CDH 5.0.3.....</i>	478
<i>New Features in Impala Version 1.3.0 / CDH 5.0.0.....</i>	478
<i>New Features in Impala Version 1.2.4.....</i>	479
<i>New Features in Impala Version 1.2.3.....</i>	480
<i>New Features in Impala Version 1.2.2.....</i>	480
<i>New Features in Impala Version 1.2.1.....</i>	481
<i>New Features in Impala Version 1.2.0 (Beta).....</i>	483
<i>New Features in Impala Version 1.1.1.....</i>	484
<i>New Features in Impala Version 1.1.....</i>	484
<i>New Features in Impala Version 1.0.1.....</i>	485
<i>New Features in Impala Version 1.0.....</i>	486
<i>New Features in Version 0.7 of the Cloudera Impala Beta Release.....</i>	486
<i>New Features in Version 0.6 of the Cloudera Impala Beta Release.....</i>	486
<i>New Features in Version 0.5 of the Cloudera Impala Beta Release.....</i>	486
<i>New Features in Version 0.4 of the Cloudera Impala Beta Release.....</i>	486
<i>New Features in Version 0.3 of the Cloudera Impala Beta Release.....</i>	487
<i>New Features in Version 0.2 of the Cloudera Impala Beta Release.....</i>	487
<i>Incompatible Changes in Impala.....</i>	487
<i>Incompatible Changes Introduced in Impala for CDH 5.4.x.....</i>	487
<i>Incompatible Changes Introduced in Impala 2.2.0 / CDH 5.4.0.....</i>	487
<i>Incompatible Changes Introduced in Cloudera Impala 2.1.4 / CDH 5.3.4.....</i>	488
<i>Incompatible Changes Introduced in Impala 2.1.3 / CDH 5.3.3.....</i>	488
<i>Incompatible Changes Introduced in Impala 2.1.2 / CDH 5.3.2.....</i>	488
<i>Incompatible Changes Introduced in Impala 2.1.1 / CDH 5.3.1.....</i>	488
<i>Incompatible Changes Introduced in Impala 2.1.0 / CDH 5.3.0.....</i>	488
<i>Incompatible Changes Introduced in Cloudera Impala 2.0.5 / CDH 5.2.6.....</i>	488
<i>Incompatible Changes Introduced in Impala 2.0.4 / CDH 5.2.5.....</i>	488
<i>Incompatible Changes Introduced in Impala 2.0.3 / CDH 5.2.4.....</i>	489
<i>Incompatible Changes Introduced in Impala 2.0.2 / CDH 5.2.3.....</i>	489
<i>Incompatible Changes Introduced in Impala 2.0.1 / CDH 5.2.1.....</i>	489
<i>Incompatible Changes Introduced in Impala 2.0.0 / CDH 5.2.0.....</i>	489
<i>Incompatible Changes Introduced in Impala 1.4.4 / CDH 5.1.5.....</i>	491
<i>Incompatible Changes Introduced in Impala 1.4.3 / CDH 5.1.4.....</i>	491
<i>Incompatible Changes Introduced in Impala 1.4.2 / CDH 5.1.3.....</i>	491
<i>Incompatible Changes Introduced in Impala 1.4.1 / CDH 5.1.2.....</i>	491
<i>Incompatible Changes Introduced in Impala 1.4.0 / CDH 5.1.0.....</i>	491
<i>Incompatible Changes Introduced in Impala 1.3.3 / CDH 5.0.5.....</i>	492
<i>Incompatible Changes Introduced in Impala 1.3.2 / CDH 5.0.4.....</i>	492
<i>Incompatible Changes Introduced in Impala 1.3.1 / CDH 5.0.3.....</i>	492

<i>Incompatible Changes Introduced in Impala 1.3.0 / CDH 5.0.0.....</i>	492
<i>Incompatible Changes Introduced in Impala 1.2.4.....</i>	493
<i>Incompatible Changes Introduced in Impala 1.2.3.....</i>	493
<i>Incompatible Changes Introduced in Impala 1.2.2.....</i>	493
<i>Incompatible Changes Introduced in Impala 1.2.1.....</i>	494
<i>Incompatible Changes Introduced in Impala 1.2.0 (Beta).....</i>	494
<i>Incompatible Changes Introduced in Impala 1.1.1.....</i>	495
<i>Incompatible Change Introduced in Cloudera Impala 1.1.....</i>	495
<i>Incompatible Changes Introduced in Impala 1.0.....</i>	495
<i>Incompatible Change Introduced in Version 0.7 of the Cloudera Impala Beta Release.....</i>	495
<i>Incompatible Change Introduced in Version 0.6 of the Cloudera Impala Beta Release.....</i>	495
<i>Incompatible Change Introduced in Version 0.4 of the Cloudera Impala Beta Release.....</i>	496
<i>Incompatible Change Introduced in Version 0.3 of the Cloudera Impala Beta Release.....</i>	496
<b>Known Issues and Workarounds in Impala.....</b>	496
<i>Known Issues in the Current Production Release (Impala 2.2.x / CDH 5.4.x).....</i>	496
<b>Fixed Issues in Impala.....</b>	500
<i>Issues Fixed in Impala for CDH 5.4.3.....</i>	500
<i>Issues Fixed in Impala for CDH 5.4.1.....</i>	501
<i>Issues Fixed in the 2.2.0 Release / CDH 5.4.0.....</i>	501
<i>Issues Fixed in the 2.1.4 Release / CDH 5.3.4.....</i>	503
<i>Issues Fixed in the 2.1.3 Release / CDH 5.3.3.....</i>	504
<i>Issues Fixed in the 2.1.2 Release / CDH 5.3.2.....</i>	505
<i>Issues Fixed in the 2.1.1 Release / CDH 5.3.1.....</i>	506
<i>Issues Fixed in the 2.1.0 Release / CDH 5.3.0.....</i>	506
<i>Issues Fixed in the 2.0.5 Release / CDH 5.2.6.....</i>	507
<i>Issues Fixed in the 2.0.4 Release / CDH 5.2.5.....</i>	507
<i>Issues Fixed in the 2.0.3 Release / CDH 5.2.4.....</i>	508
<i>Issues Fixed in the 2.0.2 Release / CDH 5.2.3.....</i>	508
<i>Issues Fixed in the 2.0.1 Release / CDH 5.2.1.....</i>	509
<i>Issues Fixed in the 2.0.0 Release / CDH 5.2.0.....</i>	509
<i>Issues Fixed in the 1.4.4 Release / CDH 5.1.5.....</i>	511
<i>Issues Fixed in the 1.4.3 Release / CDH 5.1.4.....</i>	512
<i>Issues Fixed in the 1.4.2 Release / CDH 5.1.3.....</i>	512
<i>Issues Fixed in the 1.4.1 Release / CDH 5.1.2.....</i>	512
<i>Issues Fixed in the 1.4.0 Release / CDH 5.1.0.....</i>	513
<i>Issues Fixed in the 1.3.3 Release / CDH 5.0.5.....</i>	515
<i>Issues Fixed in the 1.3.2 Release / CDH 5.0.4.....</i>	515
<i>Issues Fixed in the 1.3.1 Release / CDH 5.0.3.....</i>	515
<i>Issues Fixed in the 1.3.0 Release / CDH 5.0.0.....</i>	517
<i>Issues Fixed in the 1.2.4 Release.....</i>	519
<i>Issues Fixed in the 1.2.3 Release.....</i>	521
<i>Issues Fixed in the 1.2.2 Release.....</i>	521
<i>Issues Fixed in the 1.2.1 Release.....</i>	522
<i>Issues Fixed in the 1.2.0 Beta Release.....</i>	523
<i>Issues Fixed in the 1.1.1 Release.....</i>	523

<i>Issues Fixed in the 1.1.0 Release.....</i>	524
<i>Issues Fixed in the 1.0.1 Release.....</i>	525
<i>Issues Fixed in the 1.0 GA Release.....</i>	526
<i>Issues Fixed in Version 0.7 of the Beta Release.....</i>	529
<i>Issues Fixed in Version 0.6 of the Beta Release.....</i>	530
<i>Issues Fixed in Version 0.5 of the Beta Release.....</i>	531
<i>Issues Fixed in Version 0.4 of the Beta Release.....</i>	532
<i>Issues Fixed in Version 0.3 of the Beta Release.....</i>	533
<i>Issues Fixed in Version 0.2 of the Beta Release.....</i>	534

# Introducing Cloudera Impala

Cloudera Impala provides fast, interactive SQL queries directly on your Apache Hadoop data stored in HDFS, HBase, or the Amazon Simple Storage Service (S3). In addition to using the same unified storage platform, Impala also uses the same metadata, SQL syntax (Hive SQL), ODBC driver, and user interface (Cloudera Impala query UI in Hue) as Apache Hive. This provides a familiar and unified platform for real-time or batch-oriented queries.

Cloudera Impala is an addition to tools available for querying big data. Impala does not replace the batch processing frameworks built on MapReduce such as Hive. Hive and other frameworks built on MapReduce are best suited for long running batch jobs, such as those involving batch processing of Extract, Transform, and Load (ETL) type jobs.

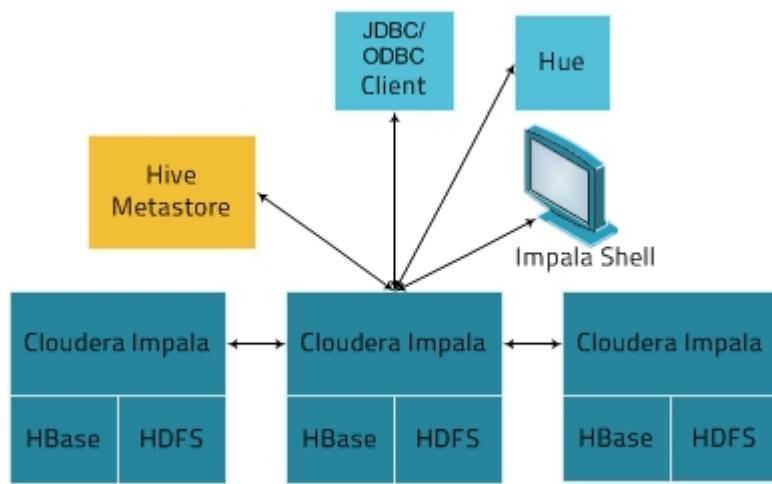
## Impala Benefits

Impala provides:

- Familiar SQL interface that data scientists and analysts already know
- Ability to interactively query data on big data in Apache Hadoop
- Distributed queries in a cluster environment, for convenient scaling and to make use of cost-effective commodity hardware
- Ability to share data files between different components with no copy or export/import step; for example, to write with Pig, transform with Hive and query with Impala
- Single system for big data processing and analytics, so customers can avoid costly modeling and ETL just for analytics

## How Cloudera Impala Works with CDH

The following graphic illustrates how Impala is positioned in the broader Cloudera environment:



The Impala solution is composed of the following components:

- Clients - Entities including Hue, ODBC clients, JDBC clients, and the Impala Shell can all interact with Impala. These interfaces are typically used to issue queries or complete administrative tasks such as connecting to Impala.
- Hive Metastore - Stores information about the data available to Impala. For example, the metastore lets Impala know what databases are available and what the structure of those databases is. As you create, drop,

and alter schema objects, load data into tables, and so on through Impala SQL statements, the relevant metadata changes are automatically broadcast to all Impala nodes by the dedicated catalog service introduced in Impala 1.2.

- Cloudera Impala - This process, which runs on DataNodes, coordinates and executes queries. Each instance of Impala can receive, plan, and coordinate queries from Impala clients. Queries are distributed among Impala nodes, and these nodes then act as workers, executing parallel query fragments.
- HBase and HDFS - Storage for data to be queried.

Queries executed using Impala are handled as follows:

1. User applications send SQL queries to Impala through ODBC or JDBC, which provide standardized querying interfaces. The user application may connect to any `impalad` in the cluster. This `impalad` becomes the coordinator for the query.
2. Impala parses the query and analyzes it to determine what tasks need to be performed by `impalad` instances across the cluster. Execution is planned for optimal efficiency.
3. Services such as HDFS and HBase are accessed by local `impalad` instances to provide data.
4. Each `impalad` returns data to the coordinating `impalad`, which sends these results to the client.

## Primary Impala Features

Impala provides support for:

- Most common SQL-92 features of Hive Query Language (HiveQL) including [SELECT](#), [joins](#), and aggregate functions.
- HDFS, HBase, and Amazon Simple Storage System (S3) storage, including:
  - [HDFS file formats](#): Text file, SequenceFile, RCFile, Avro file, and Parquet.
  - Compression codecs: Snappy, GZIP, Deflate, BZIP.
- Common data access interfaces including:
  - [JDBC driver](#).
  - [ODBC driver](#).
  - Hue Beeswax and the Cloudera Impala Query UI.
- Impala [command-line interface](#).
- [Kerberos authentication](#).

# Impala Concepts and Architecture

The following sections provide background information to help you become productive using Cloudera Impala and its features. Where appropriate, the explanations include context to help understand how aspects of Impala relate to other technologies you might already be familiar with, such as relational database management systems and data warehouses, or other Hadoop components such as Hive, HDFS, and HBase.

## Components of the Impala Server

The Impala server is a distributed, massively parallel processing (MPP) database engine. It consists of different daemon processes that run on specific hosts within your CDH cluster.

### The Impala Daemon

The core Impala component is a daemon process that runs on each node of the cluster, physically represented by the `impalad` process. It reads and writes to data files; accepts queries transmitted from the `impala-shell` command, Hue, JDBC, or ODBC; parallelizes the queries and distributes work to other nodes in the Impala cluster; and transmits intermediate query results back to the central coordinator node.

You can submit a query to the Impala daemon running on any node, and that node serves as the *coordinator node* for that query. The other nodes transmit partial results back to the coordinator, which constructs the final result set for a query. When running experiments with functionality through the `impala-shell` command, you might always connect to the same Impala daemon for convenience. For clusters running production workloads, you might load-balance between the nodes by submitting each query to a different Impala daemon in round-robin style, using the JDBC or ODBC interfaces.

The Impala daemons are in constant communication with the *statestore*, to confirm which nodes are healthy and can accept new work.

They also receive broadcast messages from the `catalogd` daemon (introduced in Impala 1.2) whenever any Impala node in the cluster creates, alters, or drops any type of object, or when an `INSERT` or `LOAD DATA` statement is processed through Impala. This background communication minimizes the need for `REFRESH` or `INVALIDATE METADATA` statements that were needed to coordinate metadata across nodes prior to Impala 1.2.

**Related information:** [Modifying Impala Startup Options](#) on page 48, [Starting Impala](#) on page 47, [Setting the Idle Query and Idle Session Timeouts for impalad](#) on page 81, [Ports Used by Impala](#) on page 451, [Using Impala through a Proxy for High Availability](#) on page 82

### The Impala Statestore

The Impala component known as the *statestore* checks on the health of Impala daemons on all the nodes in a cluster, and continuously relays its findings to each of those daemons. It is physically represented by a daemon process named `statestored`; you only need such a process on one node in the cluster. If an Impala node goes offline due to hardware failure, network error, software issue, or other reason, the statestore informs all the other nodes so that future queries can avoid making requests to the unreachable node.

Because the statestore's purpose is to help when things go wrong, it is not critical to the normal operation of an Impala cluster. If the statestore is not running or becomes unreachable, the other nodes continue running and distributing work among themselves as usual; the cluster just becomes less robust if other nodes fail while the statestore is offline. When the statestore comes back online, it re-establishes communication with the other nodes and resumes its monitoring function.

**Related information:**

[Scalability Considerations for the Impala Statestore](#) on page 388, [Modifying Impala Startup Options](#) on page 48, [Starting Impala](#) on page 47, [Increasing the Statestore Timeout](#) on page 81, [Ports Used by Impala](#) on page 451

## The Impala Catalog Service

The Impala component known as the ***catalog service*** relays the metadata changes from Impala SQL statements to all the nodes in a cluster. It is physically represented by a daemon process named `catalogd`; you only need such a process on one node in the cluster. Because the requests are passed through the statestore daemon, it makes sense to run the `statestored` and `catalogd` services on the same node.

This new component in Impala 1.2 reduces the need for the `REFRESH` and `INVALIDATE_METADATA` statements. Formerly, if you issued `CREATE DATABASE`, `DROP DATABASE`, `CREATE TABLE`, `ALTER TABLE`, or `DROP TABLE` statements on one Impala node, you needed to issue `INVALIDATE_METADATA` on any other node before running a query there, so that it would pick up the changes to schema objects. Likewise, if you issued `INSERT` statements on one node, you needed to issue `REFRESH table_name` on any other node before running a query there, so that it would recognize the newly added data files. The catalog service removes the need to issue `REFRESH` and `INVALIDATE_METADATA` statements when the metadata changes are performed by statement issued through Impala; when you create a table, load data, and so on through Hive, you still need to issue `REFRESH` or `INVALIDATE_METADATA` on an Impala node before executing a query there.

This feature, new in Impala 1.2, touches a number of aspects of Impala:

- See [Installing Impala](#) on page 30, [Upgrading Impala](#) on page 43 and [Starting Impala](#) on page 47, for usage information for the `catalogd` daemon.
- The `REFRESH` and `INVALIDATE_METADATA` statements are no longer needed when the `CREATE TABLE`, `INSERT`, or other table-changing or data-changing operation is performed through Impala. These statements are still needed if such operations are done through Hive or by manipulating data files directly in HDFS, but in those cases the statements only need to be issued on one Impala node rather than on all nodes. See [REFRESH Statement](#) on page 215 and [INVALIDATE\\_METADATA Statement](#) on page 210 for the latest usage information for those statements.

By default, the metadata loading and caching on startup happens asynchronously, so Impala can begin accepting requests promptly. To enable the original behavior, where Impala waited until all metadata was loaded before accepting any requests, set the `catalogd` configuration option `--load_catalog_in_background=false`.

▪ **Note:**

In Impala 1.2.4 and higher, you can specify a table name with `INVALIDATE_METADATA` after the table is created in Hive, allowing you to make individual tables visible to Impala without doing a full reload of the catalog metadata. Impala 1.2.4 also includes other changes to make the metadata broadcast mechanism faster and more responsive, especially during Impala startup. See [New Features in Impala Version 1.2.4](#) on page 479 for details.

**Related information:** [Modifying Impala Startup Options](#) on page 48, [Starting Impala](#) on page 47, [Ports Used by Impala](#) on page 451

## Programming Impala Applications

The core development language with Impala is SQL. You can also use Java or other languages to interact with Impala through the standard JDBC and ODBC interfaces used by many business intelligence tools. For specialized kinds of analysis, you can supplement the SQL built-in functions by writing [user-defined functions \(UDFs\)](#) in C++ or Java.

### Overview of the Impala SQL Dialect

The Impala SQL dialect is highly compatible with the SQL syntax used in the Apache Hive component (HiveQL). As such, it is familiar to users who are already familiar with running SQL queries on the Hadoop infrastructure. Currently, Impala SQL supports a subset of HiveQL statements, data types, and built-in functions. Impala also

## Impala Concepts and Architecture

includes additional built-in functions for common industry features, to simplify porting SQL from non-Hadoop systems.

For users coming to Impala from traditional database or data warehousing backgrounds, the following aspects of the SQL dialect might seem familiar:

- The `SELECT` statement includes familiar clauses such as `WHERE`, `GROUP BY`, `ORDER BY`, and `WITH`. You will find familiar notions such as [joins](#), [built-in functions](#) for processing strings, numbers, and dates, [aggregate functions](#), [subqueries](#), and [comparison operators](#) such as `IN()` and `BETWEEN`. The `SELECT` statement is the place where SQL standards compliance is most important.
- From the data warehousing world, you will recognize the notion of [partitioned tables](#). One or more columns serve as partition keys, and the data is physically arranged so that queries that refer to the partition key columns in the `WHERE` clause can skip partitions that do not match the filter conditions. For example, if you have 10 years worth of data and use a clause such as `WHERE year = 2015`, `WHERE year > 2010`, or `WHERE year IN (2014, 2015)`, Impala skips all the data for non-matching years, greatly reducing the amount of I/O for the query.
- In Impala 1.2 and higher, [UDFs](#) let you perform custom comparisons and transformation logic during `SELECT` and `INSERT...SELECT` statements.

For users coming to Impala from traditional database or data warehousing backgrounds, the following aspects of the SQL dialect might require some learning and practice for you to become proficient in the Hadoop environment:

- Impala SQL is focused on queries and includes relatively little DML. There is no `UPDATE` or `DELETE` statement. Stale data is typically discarded (by `DROP TABLE` or `ALTER TABLE ... DROP PARTITION` statements) or replaced (by `INSERT OVERWRITE` statements).
- All data creation is done by `INSERT` statements, which typically insert data in bulk by querying from other tables. There are two variations, `INSERT INTO` which appends to the existing data, and `INSERT OVERWRITE` which replaces the entire contents of a table or partition (similar to `TRUNCATE TABLE` followed by a new `INSERT`). There is no `INSERT ... VALUES` syntax to insert a single row.
- You often construct Impala table definitions and data files in some other environment, and then attach Impala so that it can run real-time queries. The same data files and table metadata are shared with other components of the Hadoop ecosystem. In particular, Impala can access tables created by Hive or data inserted by Hive, and Hive can access tables and data produced by Impala. Many other Hadoop components can write files in formats such as Parquet and Avro, that can then be queried by Impala.
- Because Hadoop and Impala are focused on data warehouse-style operations on large data sets, Impala SQL includes some idioms that you might find in the import utilities for traditional database systems. For example, you can create a table that reads comma-separated or tab-separated text files, specifying the separator in the `CREATE TABLE` statement. You can create [external tables](#) that read existing data files but do not move or transform them.
- Because Impala reads large quantities of data that might not be perfectly tidy and predictable, it does not impose length constraints on string data types. For example, you can define a database column as `STRING` with unlimited length, rather than `CHAR(1)` or `VARCHAR(64)`. (Although in Impala 2.0 and later, you can also use length-constrained `CHAR` and `VARCHAR` types.)

**Related information:** [Impala SQL Language Reference](#) on page 111, especially [SQL Statements](#) on page 162 and [Built-in Functions](#) on page 257

## Overview of Impala Programming Interfaces

You can connect and submit requests to the Impala daemons through:

- The [impala-shell](#) interactive command interpreter.
- The [Hue](#) web-based user interface.

- [JDBC](#).
- [ODBC](#).

With these options, you can use Impala in heterogeneous environments, with JDBC or ODBC applications running on non-Linux platforms. You can also use Impala in combination with various Business Intelligence tools that use the JDBC and ODBC interfaces.

Each `impalad` daemon process, running on separate nodes in a cluster, listens to [several ports](#) for incoming requests. Requests from `impala-shell` and Hue are routed to the `impalad` daemons through the same port. The `impalad` daemons listen on separate ports for JDBC and ODBC requests.

## How Impala Fits Into the Hadoop Ecosystem

Impala makes use of many familiar components within the Hadoop ecosystem. Impala can interchange data with other Hadoop components, as both a consumer and a producer, so it can fit in flexible ways into your ETL and ELT pipelines.

### How Impala Works with Hive

A major Impala goal is to make SQL-on-Hadoop operations fast and efficient enough to appeal to new categories of users and open up Hadoop to new types of use cases. Where practical, it makes use of existing Apache Hive infrastructure that many Hadoop users already have in place to perform long-running, batch-oriented SQL queries.

In particular, Impala keeps its table definitions in a traditional MySQL or PostgreSQL database known as the **metastore**, the same database where Hive keeps this type of data. Thus, Impala can access tables defined or loaded by Hive, as long as all columns use Impala-supported data types, file formats, and compression codecs.

The initial focus on query features and performance means that Impala can read more types of data with the `SELECT` statement than it can write with the `INSERT` statement. To query data using the Avro, RCFile, or SequenceFile [file formats](#), you load the data using Hive.

The Impala query optimizer can also make use of [table statistics](#) and [column statistics](#). Originally, you gathered this information with the `ANALYZE TABLE` statement in Hive; in Impala 1.2.2 and higher, use the Impala [COMPUTE STATS](#) statement instead. `COMPUTE STATS` requires less setup, is more reliable, and does not require switching back and forth between `impala-shell` and the Hive shell.

## Overview of Impala Metadata and the Metastore

As discussed in [How Impala Works with Hive](#) on page 21, Impala maintains information about table definitions in a central database known as the **metastore**. Impala also tracks other metadata for the low-level characteristics of data files:

- The physical locations of blocks within HDFS.

For tables with a large volume of data and/or many partitions, retrieving all the metadata for a table can be time-consuming, taking minutes in some cases. Thus, each Impala node caches all of this metadata to reuse for future queries against the same table.

If the table definition or the data in the table is updated, all other Impala daemons in the cluster must receive the latest metadata, replacing the obsolete cached metadata, before issuing a query against that table. In Impala 1.2 and higher, the metadata update is automatic, coordinated through the `catalogd` daemon, for all DDL and DML statements issued through Impala. See [The Impala Catalog Service](#) on page 19 for details.

For DDL and DML issued through Hive, or changes made manually to files in HDFS, you still use the `REFRESH` statement (when new data files are added to existing tables) or the `INVALIDATE METADATA` statement (for entirely new tables, or after dropping a table, performing an HDFS rebalance operation, or deleting data files). Issuing `INVALIDATE METADATA` by itself retrieves metadata for all the tables tracked by the metastore. If you know that only specific tables have been changed outside of Impala, you can issue `REFRESH table_name` for each affected table to only retrieve the latest metadata for those tables.

### How Impala Uses HDFS

Impala uses the distributed filesystem HDFS as its primary data storage medium. Impala relies on the redundancy provided by HDFS to guard against hardware or network outages on individual nodes. Impala table data is physically represented as data files in HDFS, using familiar HDFS file formats and compression codecs. When data files are present in the directory for a new table, Impala reads them all, regardless of file name. New data is added in files with names controlled by Impala.

### How Impala Uses HBase

HBase is an alternative to HDFS as a storage medium for Impala data. It is a database storage system built on top of HDFS, without built-in SQL support. Many Hadoop users already have it configured and store large (often sparse) data sets in it. By defining tables in Impala and mapping them to equivalent tables in HBase, you can query the contents of the HBase tables through Impala, and even perform join queries including both Impala and HBase tables. See [Using Impala to Query HBase Tables](#) on page 426 for details.

# Planning for Impala Deployment

Before you set up Impala in production, do some planning to make sure that your hardware setup has sufficient capacity, that your cluster topology is optimal for Impala queries, and that your schema design and ETL processes follow the best practices for Impala.

## Cloudera Impala Requirements

To perform as expected, Impala depends on the availability of the software, hardware, and configurations described in the following sections.

### Product Compatibility Matrix

The ultimate source of truth about compatibility between various versions of CDH, Cloudera Manager, and various CDH components is the online [Product Compatibility Matrix](#).

For Impala, see the [Impala compatibility matrix page](#).

### Supported Operating Systems

The relevant supported operating systems and versions for Impala are the same as for the corresponding CDH 4 and CDH 5 platforms. For details, see the *Supported Operating Systems* page for [CDH 4](#) or [CDH 5](#).

### Hive Metastore and Related Configuration

Impala can interoperate with data stored in Hive, and uses the same infrastructure as Hive for tracking metadata about schema objects such as tables and columns. The following components are prerequisites for Impala:

- MySQL or PostgreSQL, to act as a metastore database for both Impala and Hive.

▪ **Note:**

Installing and configuring a Hive metastore is an Impala requirement. Impala does not work without the metastore database. For the process of installing and configuring the metastore, see [Installing Impala](#) on page 30.

Always configure a **Hive metastore service** rather than connecting directly to the metastore database. The Hive metastore service is required to interoperate between possibly different levels of metastore APIs used by CDH and Impala, and avoids known issues with connecting directly to the metastore database. The Hive metastore service is set up for you by default if you install through Cloudera Manager 4.5 or higher.

A summary of the metastore installation process is as follows:

- Install a MySQL or PostgreSQL database. Start the database if it is not started after installation.
- Download the [MySQL connector](#) or the [PostgreSQL connector](#) and place it in the /usr/share/java/ directory.
- Use the appropriate command line tool for your database to create the metastore database.
- Use the appropriate command line tool for your database to grant privileges for the metastore database to the `hive` user.
- Modify `hive-site.xml` to include information matching your particular database: its URL, user name, and password. You will copy the `hive-site.xml` file to the Impala Configuration Directory later in the Impala installation process.

## Planning for Impala Deployment

- **Optional:** Hive. Although only the Hive metastore database is required for Impala to function, you might install Hive on some client machines to create and load data into tables that use certain file formats. See [How Impala Works with Hadoop File Formats](#) on page 399 for details. Hive does not need to be installed on the same data nodes as Impala; it just needs access to the same metastore database.

## Java Dependencies

Although Impala is primarily written in C++, it does use Java to communicate with various Hadoop components:

- The officially supported JVM for Impala is the Oracle JVM. Other JVMs might cause issues, typically resulting in a failure at `impalad` startup. In particular, the JamVM used by default on certain levels of Ubuntu systems can cause `impalad` to fail to start.
- Internally, the `impalad` daemon relies on the `JAVA_HOME` environment variable to locate the system Java libraries. Make sure the `impalad` service is not run from an environment with an incorrect setting for this variable.
- All Java dependencies are packaged in the `impala-dependencies.jar` file, which is located at `/usr/lib/impala/lib/`. These map to everything that is built under `fe/target/dependency`.

## Networking Configuration Requirements

As part of ensuring best performance, Impala attempts to complete tasks on local data, as opposed to using network connections to work with remote data. To support this goal, Impala matches the **hostname** provided to each Impala daemon with the **IP address** of each DataNode by resolving the hostname flag to an IP address. For Impala to work with local data, use a single IP interface for the DataNode and the Impala daemon on each machine. Ensure that the Impala daemon's hostname flag resolves to the IP address of the DataNode. For single-homed machines, this is usually automatic, but for multi-homed machines, ensure that the Impala daemon's hostname resolves to the correct interface. Impala tries to detect the correct hostname at start-up, and prints the derived hostname at the start of the log in a message of the form:

```
Using hostname: impala-daemon-1.cloudera.com
```

In the majority of cases, this automatic detection works correctly. If you need to explicitly set the hostname, do so by setting the `--hostname` flag.

## Hardware Requirements

During join operations, portions of data from each joined table are loaded into memory. Data sets can be very large, so ensure your hardware has sufficient memory to accommodate the joins you anticipate completing.

While requirements vary according to data set size, the following is generally recommended:

- CPU - Impala version 2.2 and higher uses the SSSE3 instruction set, which is included in newer processors.
  - **Note:** This required level of processor is the same as in Impala version 1.x. The Impala 2.0 and 2.1 releases had a stricter requirement for the SSE4.1 instruction set, which has now been relaxed.
- Memory - 128 GB or more recommended, ideally 256 GB or more. If the intermediate results during query processing on a particular node exceed the amount of memory available to Impala on that node, the query writes temporary work data to disk, which can lead to long query times. Note that because the work is parallelized, and intermediate results for aggregate queries are typically smaller than the original data, Impala can query and join tables that are much larger than the memory available on an individual node.
- Storage - DataNodes with 12 or more disks each. I/O speeds are often the limiting factor for disk performance with Impala. Ensure that you have sufficient disk space to store the data Impala will be querying.

## User Account Requirements

Impala creates and uses a user and group named `impala`. Do not delete this account or group and do not modify the account's or group's permissions and rights. Ensure no existing systems obstruct the functioning of these

accounts and groups. For example, if you have scripts that delete user accounts not in a white-list, add these accounts to the list of permitted accounts.

For the resource management feature to work (in combination with CDH 5 and the YARN and Llama components), the `impala` user must be a member of the `hdfs` group. This setup is performed automatically during a new install, but not when upgrading from earlier Impala releases to Impala 1.2. If you are upgrading a node to CDH 5 that already had Impala 1.1 or 1.0 installed, manually add the `impala` user to the `hdfs` group.

For correct file deletion during `DROP TABLE` operations, Impala must be able to move files to the HDFS trashcan. You might need to create an HDFS directory `/user/impala`, writeable by the `impala` user, so that the trashcan can be created. Otherwise, data files might remain behind after a `DROP TABLE` statement.

Impala should not run as root. Best Impala performance is achieved using direct reads, but root is not permitted to use direct reads. Therefore, running Impala as root negatively affects performance.

By default, any user can connect to Impala and access all the associated databases and tables. You can enable authorization and authentication based on the Linux OS user who connects to the Impala server, and the associated groups for that user. [Impala Security](#) on page 87 for details. These security features do not change the underlying file permission requirements; the `impala` user still needs to be able to access the data files.

## Cluster Sizing Guidelines for Impala

This document provides a very rough guideline to estimate the size of a cluster needed for a specific customer application. You can use this information when planning how much and what type of hardware to acquire for a new cluster, or when adding Impala workloads to an existing cluster.

- **Note:** Before making purchase or deployment decisions, consult your Cloudera representative to verify the conclusions about hardware requirements based on your data volume and workload.

Always use hosts with identical specifications and capacities for all the nodes in the cluster. Currently, Impala divides the work evenly between cluster nodes, regardless of their exact hardware configuration. Because work can be distributed in different ways for different queries, if some hosts are overloaded compared to others in terms of CPU, memory, I/O, or network, you might experience inconsistent performance and overall slowness.

For analytic workloads with star/snowflake schemas, and using consistent hardware for all nodes (64 GB RAM, 12 2 TB hard drives, 2x E5-2630L 12 cores total, 10 GB network), the following table estimates the number of data nodes needed in the cluster based on data size and the number of concurrent queries, for workloads similar to TPC-DS benchmark queries:

**Table 1: Cluster size estimation based on the number of concurrent queries and data size with a 20 second average query response time**

Data Size	1 query	10 queries	100 queries	1000 queries	2000 queries
250 GB	2	2	5	35	70
500 GB	2	2	10	70	135
1 TB	2	2	15	135	270
15 TB	2	20	200	N/A	N/A
30 TB	4	40	400	N/A	N/A
60 TB	8	80	800	N/A	N/A

### Factors Affecting Scalability

A typical analytic workload (TPC-DS style queries) using recommended hardware is usually CPU-bound. Each node can process roughly 1.6 GB/sec. Both CPU-bound and disk-bound workloads can scale almost linearly.

## Planning for Impala Deployment

with cluster size. However, for some workloads, the scalability might be bounded by the network, or even by memory.

If the workload is already network bound (on a 10 GB network), increasing the cluster size won't reduce the network load; in fact, a larger cluster could increase network traffic because some queries involve "broadcast" operations to all data nodes. Therefore, boosting the cluster size does not improve query throughput in a network-constrained environment.

Let's look at a memory-bound workload. A workload is memory-bound if Impala cannot run any additional concurrent queries because all memory allocated has already been consumed, but neither CPU, disk, nor network is saturated yet. This can happen because currently Impala uses only a single core per node to process join and aggregation queries. For a node with 128 GB of RAM, if a join node takes 50 GB, the system cannot run more than 2 such queries at the same time.

Therefore, at most 2 cores are used. Throughput can still scale almost linearly even for a memory-bound workload. It's just that the CPU will not be saturated. Per-node throughput will be lower than 1.6 GB/sec. Consider increasing the memory per node.

As long as the workload is not network- or memory-bound, we can use the 1.6 GB/second per node as the throughput estimate.

### A More Precise Approach

A more precise sizing estimate would require not only queries per minute (QPM), but also an average data size scanned per query (D). With the proper partitioning strategy, D is usually a fraction of the total data size. The following equation can be used as a rough guide to estimate the number of nodes (N) needed:

$$\text{Eq 1: } N > \text{QPM} * D / 100 \text{ GB}$$

Here is an example. Suppose, on average, a query scans 50 GB of data and the average response time is required to be 15 seconds or less when there are 100 concurrent queries. The QPM is  $100/15*60 = 400$ . We can estimate the number of node using our equation above.

$$\begin{aligned} N &> \text{QPM} * D / 100\text{GB} \\ N &> 400 * 50\text{GB} / 100\text{GB} \\ N &> 200 \end{aligned}$$

Because this figure is a rough estimate, the corresponding number of nodes could be between 100 and 500.

Depending on the complexity of the query, the processing rate of query might change. If the query has more joins, aggregation functions, or CPU-intensive functions such as string processing or complex UDFs, the process rate will be lower than 1.6 GB/second per node. On the other hand, if the query only does scan and filtering on numbers, the processing rate can be higher.

### Estimating Memory Requirements

Impala can handle joins between multiple large tables. Make sure that statistics are collected for all the joined tables, using the [COMPUTE STATS](#) statement. However, joining big tables does consume more memory. Follow the steps below to calculate the minimum memory requirement.

Suppose you are running the following join:

```
select a.* , b.col_1, b.col_2, ... b.col_n
from a, b
where a.key = b.key
and b.col_1 in (1,2,4...)
and b.col_4 in (....);
```

And suppose table B is smaller than table A (but still a large table).

The memory requirement for the query is the right-hand table (`B`), after decompression, filtering (`b.col_n in ...`) and after projection (only using certain columns) must be less than the total memory of the entire cluster.

```
Cluster Total Memory Requirement = Size of the smaller table *
selectivity factor from the predicate *
projection factor * compression ratio
```

In this case, assume that table `B` is 100 TB in Parquet format with 200 columns. The predicate on `B(b.col_1 in ... and b.col_4 in ...)` will select only 10% of the rows from `B` and for projection, we are only projecting 5 columns out of 200 columns. Usually, Snappy compression gives us 3 times compression, so we estimate a 3x compression factor.

```
Cluster Total Memory Requirement = Size of the smaller table *
selectivity factor from the predicate *
projection factor * compression ratio
= 100TB * 10% * 5/200 * 3
= 0.75TB
= 750GB
```

So, if you have a 10-node cluster, each node has 128 GB of RAM and you give 80% to Impala, then you have 1 TB of usable memory for Impala, which is more than 750GB. Therefore, your cluster can handle join queries of this magnitude.

## Guidelines for Designing Impala Schemas

The guidelines in this topic help you to construct an optimized and scalable schema, one that integrates well with your existing data management processes. Use these guidelines as a checklist when doing any proof-of-concept work, porting exercise, or before deploying to production.

If you are adapting an existing database or Hive schema for use with Impala, read the guidelines in this section and then see [Porting SQL from Other Database Systems to Impala](#) on page 326 for specific porting and compatibility tips.

### Prefer binary file formats over text-based formats.

To save space and improve memory usage and query performance, use binary file formats for any large or intensively queried tables. Parquet file format is the most efficient for data warehouse-style analytic queries. Avro is the other binary file format that Impala supports, that you might already have as part of a Hadoop ETL pipeline.

Although Impala can create and query tables with the RCFile and SequenceFile file formats, such tables are relatively bulky due to the text-based nature of those formats, and are not optimized for data warehouse-style queries due to their row-oriented layout. Impala does not support `INSERT` operations for tables with these file formats.

#### Guidelines:

- For an efficient and scalable format for large, performance-critical tables, use the Parquet file format.
- To deliver intermediate data during the ETL process, in a format that can also be used by other Hadoop components, Avro is a reasonable choice.
- For convenient import of raw data, use a text table instead of RCFile or SequenceFile, and convert to Parquet in a later stage of the ETL process.

### Use Snappy compression where practical.

Snappy compression involves low CPU overhead to decompress, while still providing substantial space savings. In cases where you have a choice of compression codecs, such as with the Parquet and Avro file formats, use Snappy compression unless you find a compelling reason to use a different codec.

## Planning for Impala Deployment

### Prefer numeric types over strings.

If you have numeric values that you could treat as either strings or numbers (such as `YEAR`, `MONTH`, and `DAY` for partition key columns), define them as the smallest applicable integer types. For example, `YEAR` can be `SMALLINT`, `MONTH` and `DAY` can be `TINYINT`. Although you might not see any difference in the way partitioned tables or text files are laid out on disk, using numeric types will save space in binary formats such as Parquet, and in memory when doing queries, particularly resource-intensive queries such as joins.

### Partition, but don't over-partition.

Partitioning is an important aspect of performance tuning for Impala. Follow the procedures in [Partitioning for Impala Tables](#) on page 394 to set up partitioning for your biggest, most intensively queried tables.

If you are moving to Impala from a traditional database system, or just getting started in the Big Data field, you might not have enough data volume to take advantage of Impala parallel queries with your existing partitioning scheme. For example, if you have only a few tens of megabytes of data per day, partitioning by `YEAR`, `MONTH`, and `DAY` columns might be too granular. Most of your cluster might be sitting idle during queries that target a single day, or each node might have very little work to do. Consider reducing the number of partition key columns so that each partition directory contains several gigabytes worth of data.

For example, consider a Parquet table where each data file is 1 HDFS block, with a maximum block size of 1 GB. (In Impala 2.0 and later, the default Parquet block size is reduced to 256 MB. For this exercise, let's assume you have bumped the size back up to 1 GB by setting the query option `PARQUET_FILE_SIZE=1g`.) If you have a 10-node cluster, you need 10 data files (up to 10 GB) to give each node some work to do for a query. But each core on each machine can process a separate data block in parallel. With 16-core machines on a 10-node cluster, a query could process up to 160 GB fully in parallel. If there are only a few data files per partition, not only are most cluster nodes sitting idle during queries, so are most cores on those machines.

You can reduce the Parquet block size to as low as 128 MB or 64 MB to increase the number of files per partition and improve parallelism. But also consider reducing the level of partitioning so that analytic queries have enough data to work with.

### Always compute stats after loading data.

Impala makes extensive use of statistics about data in the overall table and in each column, to help plan resource-intensive operations such as join queries and inserting into partitioned Parquet tables. Because this information is only available after data is loaded, run the `COMPUTE STATS` statement on a table after loading or replacing data in a table or partition.

Having accurate statistics can make the difference between a successful operation, or one that fails due to an out-of-memory error or a timeout. When you encounter performance or capacity issues, always use the `SHOW STATS` statement to check if the statistics are present and up-to-date for all tables in the query.

When doing a join query, Impala consults the statistics for each joined table to determine their relative sizes and to estimate the number of rows produced in each join stage. When doing an `INSERT` into a Parquet table, Impala consults the statistics for the source table to determine how to distribute the work of constructing the data files for each partition.

See [COMPUTE STATS Statement](#) on page 170 for the syntax of the `COMPUTE STATS` statement, and [How Impala Uses Statistics for Query Optimization](#) on page 363 for all the performance considerations for table and column statistics.

### Verify sensible execution plans with EXPLAIN and SUMMARY.

Before executing a resource-intensive query, use the `EXPLAIN` statement to get an overview of how Impala intends to parallelize the query and distribute the work. If you see that the query plan is inefficient, you can take tuning steps such as changing file formats, using partitioned tables, running the `COMPUTE STATS` statement, or adding query hints. For information about all of these techniques, see [Tuning Impala for Performance](#) on page 354.

After you run a query, you can see performance-related information about how it actually ran by issuing the `SUMMARY` command in `impala-shell`. Prior to Impala 1.4, you would use the `PROFILE` command, but its highly technical output was only useful for the most experienced users. `SUMMARY`, new in Impala 1.4, summarizes the most useful information for all stages of execution, for all nodes rather than splitting out figures for each node.

# Installing Impala

Cloudera Impala is an open-source add-on to the Cloudera Enterprise Core that returns rapid responses to queries.

■ **Note:**

Under CDH 5, Impala is included as part of the CDH installation and no separate steps are needed. Therefore, the instruction steps in this section apply to CDH 4 only.

## What is Included in an Impala Installation

Impala is made up of a set of components that can be installed on multiple nodes throughout your cluster. The key installation step for performance is to install the `impalad` daemon (which does most of the query processing work) on *all* data nodes in the cluster.

The Impala package installs these binaries:

- `impalad` - The Impala daemon. Plans and executes queries against HDFS, HBase, and Amazon S3 data. [Run one impalad process](#) on each node in the cluster that has a data node.
- `statestored` - Name service that tracks location and status of all `impalad` instances in the cluster. [Run one instance of this daemon](#) on a node in your cluster. Most production deployments run this daemon on the namenode.
- `catalogd` - Metadata coordination service that broadcasts changes from Impala DDL and DML statements to all affected Impala nodes, so that new tables, newly loaded data, and so on are immediately visible to queries submitted through any Impala node. (Prior to Impala 1.2, you had to run the `REFRESH` or `INVALIDATE METADATA` statement on each node to synchronize changed metadata. Now those statements are only required if you perform the DDL or DML through an external mechanism such as Hive or by uploading data to the Amazon S3 filesystem.) [Run one instance of this daemon](#) on a node in your cluster, preferably on the same host as the `statestored` daemon.
- `impala-shell` - [Command-line interface](#) for issuing queries to the Impala daemon. You install this on one or more hosts anywhere on your network, not necessarily data nodes or even within the same cluster as Impala. It can connect remotely to any instance of the Impala daemon.

Before doing the installation, ensure that you have all necessary prerequisites. See [Cloudera Impala Requirements](#) on page 23 for details.

## Impala Installation Procedure for CDH 4 Users

You can install Impala under CDH 4 in one of two ways:

- Using the Cloudera Manager installer. This is the recommended technique for doing a reliable and verified Impala installation. Cloudera Manager 4.8 or higher can automatically install, configure, manage, and monitor Impala 1.2.1 and higher. The latest Cloudera Manager is always preferable, because newer Cloudera Manager releases have configuration settings for the most recent Impala features.
- Using a manual process for systems not managed by Cloudera Manager. You must do additional verification steps in this case, to check that Impala can interact with other Hadoop components correctly, and that your cluster is configured for efficient Impala execution.

## Installing Impala with Cloudera Manager

Before installing Impala through the Cloudera Manager interface, make sure all applicable nodes have the appropriate hardware configuration and levels of operating system and CDH. See [Cloudera Impala Requirements](#) on page 23 for details.

- **Note:** To install the latest Impala under CDH 4, upgrade Cloudera Manager to 4.8 or higher. Cloudera Manager 4.8 is the first release that can manage the Impala catalog service introduced in Impala 1.2. Cloudera Manager 4.8 requires this service to be present, so if you upgrade to Cloudera Manager 4.8, also upgrade Impala to the most recent version at the same time.

For information on installing Impala in a Cloudera Manager-managed environment, see [Installing Impala](#) or the *Cloudera Manager Installation Guide* for [Cloudera Manager 4](#).

Managing your Impala installation through Cloudera Manager has a number of advantages. For example, when you make configuration changes to CDH components using Cloudera Manager, it automatically applies changes to the copies of configuration files, such as `hive-site.xml`, that Impala keeps under `/etc/impala/conf`. It also sets up the Hive Metastore service that is required for Impala running under CDH 4.1.

In some cases, depending on the level of Impala, CDH, and Cloudera Manager, you might need to add particular component configuration details in some of the free-form option fields on the Impala configuration pages within Cloudera Manager. In Cloudera Manager 4, these fields are labelled **Safety Valve**; in Cloudera Manager 5, they are called **Advanced Configuration Snippet**.

## Installing Impala without Cloudera Manager

Before installing Impala manually, make sure all applicable nodes have the appropriate hardware configuration, levels of operating system and CDH, and any other software prerequisites. See [Cloudera Impala Requirements](#) on page 23 for details.

You can install Impala across many hosts or on one host:

- Installing Impala across multiple machines creates a distributed configuration. For best performance, install Impala on **all** DataNodes.
- Installing Impala on a single machine produces a pseudo-distributed cluster.

### To install Impala on a host:

1. Install CDH as described in the Installation section of the [CDH 4 Installation Guide](#) or the [CDH 5 Installation Guide](#).
2. Install the Hive metastore somewhere in your cluster, as described in the Hive Installation topic in the [CDH 4 Installation Guide](#) or the [CDH 5 Installation Guide](#). As part of this process, you configure the Hive metastore to use an external database as a metastore. Impala uses this same database for its own table metadata. You can choose either a MySQL or PostgreSQL database as the metastore. The process for configuring each type of database is described in the CDH Installation Guide).

Cloudera recommends setting up a Hive metastore service rather than connecting directly to the metastore database; this configuration is required when running Impala under CDH 4.1. Make sure the `/etc/impala/conf/hive-site.xml` file contains the following setting, substituting the appropriate host name for `metastore_server_host`:

```
<property>
<name>hive.metastore.uris</name>
<value>thrift://metastore_server_host:9083</value>
</property>
<property>
<name>hive.metastore.client.socket.timeout</name>
<value>3600</value>
```

## Installing Impala

```
<description>MetaStore Client socket timeout in seconds</description>
</property>
```

3. (Optional) If you installed the full Hive component on any host, you can verify that the metastore is configured properly by starting the Hive console and querying for the list of available tables. Once you confirm that the console starts, exit the console to continue the installation:

```
$ hive
Hive history file=/tmp/root/hive_job_log_root_201207272011_678722950.txt
hive> show tables;
table1
table2
hive> quit;
$
```

4. Confirm that your package management command is aware of the Impala repository settings, as described in [Cloudera Impala Requirements](#) on page 23. (For CDH 4, this is a different repository than for CDH.) You might need to download a repo or list file into a system directory underneath /etc.
5. Use **one** of the following sets of commands to install the Impala package:

### For RHEL, Oracle Linux, or CentOS systems:

```
$ sudo yum install impala          # Binaries for daemons
$ sudo yum install impala-server    # Service start/stop script
$ sudo yum install impala-state-store # Service start/stop script
$ sudo yum install impala-catalog    # Service start/stop script
```

### For SUSE systems:

```
$ sudo zypper install impala          # Binaries for daemons
$ sudo zypper install impala-server    # Service start/stop script
$ sudo zypper install impala-state-store # Service start/stop script
$ sudo zypper install impala-catalog    # Service start/stop script
```

### For Debian or Ubuntu systems:

```
$ sudo apt-get install impala          # Binaries for daemons
$ sudo apt-get install impala-server    # Service start/stop script
$ sudo apt-get install impala-state-store # Service start/stop script
$ sudo apt-get install impala-catalog    # Service start/stop script
```

▪ **Note:** Cloudera recommends that you not install Impala on any HDFS NameNode. Installing Impala on NameNodes provides no additional data locality, and executing queries with such a configuration might cause memory contention and negatively impact the HDFS NameNode.

6. Copy the client `hive-site.xml`, `core-site.xml`, `hdfs-site.xml`, and `hbase-site.xml` configuration files to the Impala configuration directory, which defaults to `/etc/impala/conf`. Create this directory if it does not already exist.
7. Use **one** of the following commands to install `impala-shell` on the machines from which you want to issue queries. You can install `impala-shell` on any supported machine that can connect to DataNodes that are running `impalad`.

### For RHEL/CentOS systems:

```
$ sudo yum install impala-shell
```

### For SUSE systems:

```
$ sudo zypper install impala-shell
```

**For Debian/Ubuntu systems:**

```
$ sudo apt-get install impala-shell
```

8. Complete any required or recommended configuration, as described in [Post-Installation Configuration for Impala](#) on page 34. Some of these configuration changes are mandatory. (They are applied automatically when you install using Cloudera Manager.)

Once installation and configuration are complete, see [Starting Impala](#) on page 47 for how to activate the software on the appropriate nodes in your cluster.

If this is your first time setting up and using Impala in this cluster, run through some of the exercises in [Impala Tutorials](#) on page 52 to verify that you can do basic operations such as creating tables and querying them.

# Managing Impala

This section explains how to configure Impala to accept connections from applications that use popular programming APIs:

- [Post-Installation Configuration for Impala](#) on page 34
- [Configuring Impala to Work with ODBC](#) on page 37
- [Configuring Impala to Work with JDBC](#) on page 39

This type of configuration is especially useful when using Impala in combination with Business Intelligence tools, which use these standard interfaces to query different kinds of database and Big Data systems.

You can also configure these other aspects of Impala:

- [Impala Security](#) on page 87
- [Modifying Impala Startup Options](#) on page 48

## Post-Installation Configuration for Impala

This section describes the mandatory and recommended configuration settings for Cloudera Impala. If Impala is installed using Cloudera Manager, some of these configurations are completed automatically; you must still configure short-circuit reads manually. If you installed Impala without Cloudera Manager, or if you want to customize your environment, consider making the changes described in this topic.

In some cases, depending on the level of Impala, CDH, and Cloudera Manager, you might need to add particular component configuration details in one of the free-form fields on the Impala configuration pages within Cloudera Manager. In Cloudera Manager 4, these fields are labelled **Safety Valve**; in Cloudera Manager 5, they are called **Advanced Configuration Snippet**.

- You must enable short-circuit reads, whether or not Impala was installed through Cloudera Manager. This setting goes in the Impala configuration settings, not the Hadoop-wide settings.
- If you installed Impala in an environment that is not managed by Cloudera Manager, you must enable block location tracking, and you can optionally enable native checksumming for optimal performance.
- If you deployed Impala using Cloudera Manager see [Testing Impala Performance](#) on page 377 to confirm proper configuration.

### Mandatory: Short-Circuit Reads

Enabling short-circuit reads allows Impala to read local data directly from the file system. This removes the need to communicate through the DataNodes, improving performance. This setting also minimizes the number of additional copies of data. Short-circuit reads requires `libhadoop.so` (the Hadoop Native Library) to be accessible to both the server and the client. `libhadoop.so` is not available if you have installed from a tarball. You must install from an `.rpm`, `.deb`, or `parcel` to use short-circuit local reads.

- **Note:** If you use Cloudera Manager, you can enable short-circuit reads through a checkbox in the user interface and that setting takes effect for Impala as well.

Cloudera strongly recommends using Impala with CDH 4.2 or higher, ideally the latest 4.x release. Impala does support short-circuit reads with CDH 4.1, but for best performance, upgrade to CDH 4.3 or higher. The process of configuring short-circuit reads varies according to which version of CDH you are using. Choose the procedure that is appropriate for your environment.

#### To configure DataNodes for short-circuit reads with CDH 4.2 or higher:

1. Copy the client `core-site.xml` and `hdfs-site.xml` configuration files from the Hadoop configuration directory to the Impala configuration directory. The default Impala configuration location is `/etc/impala/conf`.

- On all Impala nodes, configure the following properties in Impala's copy of `hdfs-site.xml` as shown:

```
<property>
  <name>dfs.client.read.shortcircuit</name>
  <value>true</value>
</property>

<property>
  <name>dfs.domain.socket.path</name>
  <value>/var/run/hdfs-sockets/dn</value>
</property>

<property>
  <name>dfs.client.file-block-storage-locations.timeout.millis</name>
  <value>10000</value>
</property>
```

- If `/var/run/hadoop-hdfs/` is group-writable, make sure its group is `root`.

- Note:** If you are also going to enable block location tracking, you can skip copying configuration files and restarting DataNodes and go straight to [Optional: Block Location Tracking](#). Configuring short-circuit reads and block location tracking require the same process of copying files and restarting services, so you can complete that process once when you have completed all configuration changes. Whether you copy files and restart services now or during configuring block location tracking, short-circuit reads are not enabled until you complete those final steps.

- After applying these changes, restart all DataNodes.

#### To configure DataNodes for short-circuit reads with CDH 4.1:

- Note:** Cloudera strongly recommends using Impala with CDH 4.2 or higher, ideally the latest 4.x release. Impala does support short-circuit reads with CDH 4.1, but for best performance, upgrade to CDH 4.3 or higher. The process of configuring short-circuit reads varies according to which version of CDH you are using. Choose the procedure that is appropriate for your environment.

- Enable short-circuit reads by adding settings to the Impala `core-site.xml` file.

- If you installed Impala using Cloudera Manager, short-circuit reads should be properly configured, but you can review the configuration by checking the contents of the `core-site.xml` file, which is installed at `/etc/impala/conf` by default.
- If you installed using packages, instead of using Cloudera Manager, create the `core-site.xml` file. This can be easily done by copying the `core-site.xml` client configuration file from another machine that is running Hadoop services. This file must be copied to the Impala configuration directory. The Impala configuration directory is set by the `IMPALA_CONF_DIR` environment variable and is by default `/etc/impala/conf`. To confirm the Impala configuration directory, check the `IMPALA_CONF_DIR` environment variable value.

- Note:** If the Impala configuration directory does not exist, create it and then add the `core-site.xml` file.

Add the following to the `core-site.xml` file:

```
<property>
  <name>dfs.client.read.shortcircuit</name>
  <value>true</value>
</property>
```

## Managing Impala

- **Note:** For an installation managed by Cloudera Manager, specify these settings in the Impala dialogs, in the options field for HDFS. In Cloudera Manager 4, these fields are labelled **Safety Valve**; in Cloudera Manager 5, they are called **Advanced Configuration Snippet**.

2. For each DataNode, enable access by adding the following to the `hdfs-site.xml` file:

```
<property>
    <name>dfs.client.use.legacy.blockreader.local</name>
    <value>true</value>
</property>

<property>
    <name>dfs.datanode.data.dir.perm</name>
    <value>750</value>
</property>

<property>
    <name>dfs.block.local-path-access.user</name>
    <value>impala</value>
</property>

<property>
    <name>dfs.client.file-block-storage-locations.timeout.millis</name>
    <value>10000</value>
</property>
```

- **Note:** In the preceding example, the `dfs.block.local-path-access.user` is the user running the `impalad` process. By default, that account is `impala`.

3. Use `usermod` to add users requiring local block access to the appropriate HDFS group. For example, if you assigned `impala` to the `dfs.block.local-path-access.user` property, you would add `impala` to the `hadoop` HDFS group:

```
$ usermod -a -G hadoop impala
```

- **Note:** The default HDFS group is `hadoop`, but it is possible to have an environment configured to use an alternate group. To find the configured HDFS group name using the Cloudera Manager Admin Console:

1. Go to the HDFS service.
2. Click the **Configuration** tab.
3. Click **Scope > HDFS service name (Service-Wide)**.
4. Click **Category > Advanced**.
5. The **Shared Hadoop Group Name** property contains the group name.

- **Note:** If you are going to enable block location tracking, you can skip copying configuration files and restarting DataNodes and go straight to [Mandatory: Block Location Tracking](#) on page 37. Configuring short-circuit reads and block location tracking require the same process of copying files and restarting services, so you can complete that process once when you have completed all configuration changes. Whether you copy files and restart services now or during configuring block location tracking, short-circuit reads are not enabled until you complete those final steps.

4. Copy the client `core-site.xml` and `hdfs-site.xml` configuration files from the Hadoop configuration directory to the Impala configuration directory. The default Impala configuration location is `/etc/impala/conf`.
5. After applying these changes, restart all DataNodes.

## Mandatory: Block Location Tracking

Enabling block location metadata allows Impala to know which disk data blocks are located on, allowing better utilization of the underlying disks. Impala will not start unless this setting is enabled.

### To enable block location tracking:

- For each DataNode, adding the following to the `hdfs-site.xml` file:

```
<property>
  <name>dfs.datanode.hdfs-blocks-metadata.enabled</name>
  <value>true</value>
</property>
```

- Copy the client `core-site.xml` and `hdfs-site.xml` configuration files from the Hadoop configuration directory to the Impala configuration directory. The default Impala configuration location is `/etc/impala/conf`.
- After applying these changes, restart all DataNodes.

## Optional: Native Checksumming

Enabling native checksumming causes Impala to use an optimized native library for computing checksums, if that library is available.

### To enable native checksumming:

If you installed CDH from packages, the native checksumming library is installed and setup correctly. In such a case, no additional steps are required. Conversely, if you installed by other means, such as with tarballs, native checksumming may not be available due to missing shared objects. Finding the message "Unable to load native-hadoop library for your platform... using builtin-java classes where applicable" in the Impala logs indicates native checksumming may be unavailable. To enable native checksumming, you must build and install `libhadoop.so` (the [Hadoop Native Library](#)).

## Configuring Impala to Work with ODBC

Third-party products can be designed to integrate with Impala using ODBC. For the best experience, ensure any third-party product you intend to use is supported. Verifying support includes checking that the versions of Impala, ODBC, the operating system, and the third-party product have all been approved for use together. Before configuring your systems to use ODBC, download a connector.

- Note:** You may need to sign in and accept license agreements before accessing the pages required for downloading ODBC connectors.

Versions 2.5 and 2.0 of the Cloudera ODBC Connector, currently certified for some but not all BI applications, use the HiveServer2 protocol, corresponding to Impala port 21050. Impala supports Kerberos authentication with all the supported versions of the driver, and requires ODBC 2.05.13 for Impala or higher for LDAP username/password authentication.

Version 1.x of the Cloudera ODBC Connector uses the original HiveServer1 protocol, corresponding to Impala port 21000.

See the [downloads page](#) for the versions of these drivers for different products, and the [documentation page](#) for installation instructions.

- Important:** If you are using the Cloudera Connector for Tableau, to connect Impala to your Kerberos-secured CDH clusters, contact your Tableau account representative for an updated Tableau Data-connection Customization (TDC) file. The updated TDC file will override the Tableau connection settings to set specific parameters on the connection string that are required for a secure connection.

## Managing Impala

- **Note:** If your JDBC or ODBC application connects to Impala through a load balancer such as haproxy, be cautious about reusing the connections. If the load balancer has set up connection timeout values, either check the connection frequently so that it never sits idle longer than the load balancer timeout value, or check the connection validity before using it and create a new one if the connection has been closed.

To illustrate the outline of the setup process, here is a transcript of a session to set up all required drivers and a business intelligence application that uses the ODBC driver, under Mac OS X. Each .dmg file runs a GUI-based installer, first for the [underlying IODBC driver](#) needed for non-Windows systems, then for the Cloudera ODBC Connector, and finally for the BI tool itself.

```
$ ls -1
Cloudera-ODBC-Driver-for-Impala-Install-Guide.pdf
BI_Tool_Installer.dmg
iodbc-sdk-3.52.7-macosx-10.5.dmg
ClouderaImpalaODBC.dmg
$ open iodbc-sdk-3.52.7-macosx-10.dmg
Install the IODBC driver using its installer
$ open ClouderaImpalaODBC.dmg
Install the Cloudera ODBC Connector using its installer
$ installer_dir=$(pwd)
$ cd /opt/cloudera/impalaodbc
$ ls -1
Cloudera ODBC Driver for Impala Install Guide.pdf
Readme.txt
Setup
lib
ErrorMessages
Release Notes.txt
Tools
$ cd Setup
$ ls
odbc.ini odbcinst.ini
$ cp odbc.ini ~/.odbc.ini
$ vi ~/.odbc.ini
$ cat ~/.odbc.ini
[ODBC]
# Specify any global ODBC configuration here such as ODBC tracing.

[ODBC Data Sources]
Sample Cloudera Impala DSN=Cloudera ODBC Driver for Impala

[Sample Cloudera Impala DSN]

# Description: DSN Description.
# This key is not necessary and is only to give a description of the data source.
Description=Cloudera ODBC Driver for Impala DSN

# Driver: The location where the ODBC driver is installed to.
Driver=/opt/cloudera/impalaodbc/lib/universal/libclouderaimpalaodbc.dylib

# The DriverUnicodeEncoding setting is only used for SimbaDM
# When set to 1, SimbaDM runs in UTF-16 mode.
# When set to 2, SimbaDM runs in UTF-8 mode.
#DriverUnicodeEncoding=2

# Values for HOST, PORT, KrbFQDN, and KrbServiceName should be set here.
# They can also be specified on the connection string.
HOST=hostname.sample.example.com
PORT=21050
Schema=default

# The authentication mechanism.
# 0 - No authentication (NOSASL)
# 1 - Kerberos authentication (SASL)
# 2 - Username authentication (SASL)
# 3 - Username/password authentication (SASL)
# 4 - Username/password authentication with SSL (SASL)
# 5 - No authentication with SSL (NOSASL)
# 6 - Username/password authentication (NOSASL)
```

```

AuthMech=0

# Kerberos related settings.
KrbFQDN=
KrbRealm=
KrbServiceName=

# Username/password authentication with SSL settings.
UID=
PWD
CAIssuedCertNamesMismatch=1
TrustedCerts=/opt/cloudera/impalaodbc/lib/universal/cacerts.pem

# Specify the proxy user ID to use.
#DelegationUID=

# General settings
TSaslTransportBufSize=1000
RowsFetchedPerBlock=10000
SocketTimeout=0
StringColumnLength=32767
UseNativeQuery=0
$ pwd
/opt/cloudera/impalaodbc/Setup
$ cd $installer_dir
$ open BI_Tool_Installer.dmg
Install the BI tool using its installer
$ ls /Applications | grep BI_Tool
BI_Tool.app
$ open -a BI_Tool.app
In the BI tool, connect to a data source using port 21050

```

## Configuring Impala to Work with JDBC

Impala supports the standard JDBC interface, allowing access from commercial Business Intelligence tools and custom software written in Java or other programming languages. The JDBC driver allows you to access Impala from a Java program that you write, or a Business Intelligence or similar tool that uses JDBC to communicate with various database products.

Setting up a JDBC connection to Impala involves the following steps:

- Verifying the communication port where the Impala daemons in your cluster are listening for incoming JDBC requests.
- Installing the JDBC driver on every system that runs the JDBC-enabled application.
- Specifying a connection string for the JDBC application to access one of the servers running the impalad daemon, with the appropriate security settings.

### Configuring the JDBC Port

The default port used by JDBC 2.0 and later (as well as ODBC 2.x) is 21050. Impala server accepts JDBC connections through this same port 21050 by default. Make sure this port is available for communication with other hosts on your network, for example, that it is not blocked by firewall software. If your JDBC client software connects to a different port, specify that alternative port number with the `--hs2_port` option when starting impalad. See [Starting Impala](#) on page 47 for details about Impala startup options. See [Ports Used by Impala](#) on page 451 for information about all ports used for communication between Impala and clients or between Impala components.

### Choosing the JDBC Driver

In Impala 2.0 and later, you have the choice between the Cloudera JDBC Connector and the Hive 0.13 JDBC driver. Cloudera recommends using the Cloudera JDBC Connector where practical.

## Managing Impala

If you are already using JDBC applications with an earlier Impala release, you must update your JDBC driver to one of these choices, because the Hive 0.12 driver that was formerly the only choice is not compatible with Impala 2.0 and later.

Both the Cloudera JDBC 2.5 Connector and the Hive JDBC driver provide a substantial speed increase for JDBC applications with Impala 2.0 and higher, for queries that return large result sets.

### Enabling Impala JDBC Support on Client Systems

#### Using the Cloudera JDBC Connector (recommended)

You download and install the Cloudera JDBC 2.5 connector on any Linux, Windows, or Mac system where you intend to run JDBC-enabled applications. From the [Cloudera Connectors download page](#), you choose the appropriate protocol (JDBC or ODBC) and target product (Impala or Hive). The ease of downloading and installing on non-CDH systems makes this connector a convenient choice for organizations with heterogeneous environments.

#### Using the Hive JDBC Driver

You install the Hive JDBC driver (`hive-jdbc` package) through the Linux package manager, on hosts within the CDH cluster. The driver consists of several Java JAR files. The same driver can be used by Impala and Hive.

To get the JAR files, install the Hive JDBC driver on each CDH-enabled host in the cluster that will run JDBC applications. Follow the instructions for [CDH 5](#) or [CDH 4](#).

- **Note:** The latest JDBC driver, corresponding to Hive 0.13, provides substantial performance improvements for Impala queries that return large result sets. Impala 2.0 and later are compatible with the Hive 0.13 driver. If you already have an older JDBC driver installed, and are running Impala 2.0 or higher, consider upgrading to the latest Hive JDBC driver for best performance with JDBC applications.

If you are using JDBC-enabled applications on hosts outside the CDH cluster, you cannot use the CDH install procedure on the non-CDH hosts. Install the JDBC driver on at least one CDH host using the preceding procedure. Then download the JAR files to each client machine that will use JDBC with Impala:

```
commons-logging-X.X.X.jar  
hadoop-common.jar  
hive-common-X.XX.X-cdhX.X.X.jar  
hive-jdbc-X.XX.X-cdhX.X.X.jar  
hive-metastore-X.XX.X-cdhX.X.X.jar  
hive-service-X.XX.X-cdhX.X.X.jar  
httpclient-X.X.X.jar  
httpcore-X.X.X.jar  
libfb303-X.X.X.jar  
libthrift-X.X.X.jar  
log4j-X.X.XXX.jar  
slf4j-api-X.X.X.jar  
slf4j-logXjXX-X.X.X.jar
```

#### To enable JDBC support for Impala on the system where you run the JDBC application:

1. Download the JAR files listed above to each client machine.

- **Note:** For Maven users, see [this sample github page](#) for an example of the dependencies you could add to a `pom` file instead of downloading the individual JARs.

2. Store the JAR files in a location of your choosing, ideally a directory already referenced in your `CLASSPATH` setting. For example:
  - On Linux, you might use a location such as `/opt/jars/`.

- On Windows, you might use a subdirectory underneath C:\Program Files.
3. To successfully load the Impala JDBC driver, client programs must be able to locate the associated JAR files. This often means setting the CLASSPATH for the client process to include the JARs. Consult the documentation for your JDBC client for more details on how to install new JDBC drivers, but some examples of how to set CLASSPATH variables include:
- On Linux, if you extracted the JARs to /opt/jars/, you might issue the following command to prepend the JAR files path to an existing classpath:
- ```
export CLASSPATH=/opt/jars/*.jar:$CLASSPATH
```
- On Windows, use the **System Properties** control panel item to modify the **Environment Variables** for your system. Modify the environment variables to include the path to which you extracted the files.

**Note:** If the existing CLASSPATH on your client machine refers to some older version of the Hive JARs, ensure that the new JARs are the first ones listed. Either put the new JAR files earlier in the listings, or delete the other references to Hive JAR files.

## Establishing JDBC Connections

The JDBC driver class depends on which driver you select.

**Note:** If your JDBC or ODBC application connects to Impala through a load balancer such as haproxy, be cautious about reusing the connections. If the load balancer has set up connection timeout values, either check the connection frequently so that it never sits idle longer than the load balancer timeout value, or check the connection validity before using it and create a new one if the connection has been closed.

### Using the Cloudera JDBC Connector (recommended)

Depending on the level of the JDBC API your application is targeting, you can use the following fully-qualified class names (FQCNs):

- com.cloudera.impala.jdbc41.Driver
- com.cloudera.impala.jdbc41.DataSource
- com.cloudera.impala.jdbc4.Driver
- com.cloudera.impala.jdbc4.DataSource
- com.cloudera.impala.jdbc3.Driver
- com.cloudera.impala.jdbc3.DataSource

The connection string has the following format:

```
jdbc:impala://Host:Port[ /Schema];Property1=Value;Property2=Value;...
```

The port value is typically 21050 for Impala.

For full details about the classes and the connection string (especially the property values available for the connection string), download the appropriate driver documentation for your platform from [the Impala JDBC Connector download page](#).

### Using the Hive JDBC Driver

For example, with the Hive JDBC driver, the class name is org.apache.hive.jdbc.HiveDriver. Once you have configured Impala to work with JDBC, you can establish connections between the two. To do so for a cluster that

## Managing Impala

does not use Kerberos authentication, use a connection string of the form `jdbc:hive2://host:port/;auth=noSasl`. For example, you might use:

```
jdbc:hive2://myhost.example.com:21050/;auth=noSasl
```

To connect to an instance of Impala that requires Kerberos authentication, use a connection string of the form `jdbc:hive2://host:port/;principal=principal_name`. The principal must be the same user principal you used when starting Impala. For example, you might use:

```
jdbc:hive2://myhost.example.com:21050/;principal=impala/myhost.example.com@H2.EXAMPLE.COM
```

To connect to an instance of Impala that requires LDAP authentication, use a connection string of the form `jdbc:hive2://host:port/db_name;user=ldap_userid;password=ldap_password`. For example, you might use:

```
jdbc:hive2://myhost.example.com:21050/test_db;user=fred;password=xyz123
```

## Upgrading Impala

Upgrading Cloudera Impala involves stopping Impala services, using your operating system's package management tool to upgrade Impala to the latest version, and then restarting Impala services.

- **Note:**

- Each version of CDH 5 has an associated version of Impala. When you upgrade from CDH 4 to CDH 5, you get whichever version of Impala comes with the associated level of CDH. Depending on the version of Impala you were running on CDH 4, this could install a lower level of Impala on CDH 5. For example, if you upgrade to CDH 5.0 from CDH 4 plus Impala 1.4, the CDH 5.0 installation comes with Impala 1.3. Always check the associated level of Impala before upgrading to a specific version of CDH 5. Where practical, upgrade from CDH 4 to the latest CDH 5, which also has the latest Impala.
- When you upgrade Impala, also upgrade Cloudera Manager if necessary:
  - Users running Impala on CDH 5 must upgrade to Cloudera Manager 5.0.0 or higher.
  - Users running Impala on CDH 4 must upgrade to Cloudera Manager 4.8 or higher. Cloudera Manager 4.8 includes management support for the Impala catalog service, and is the minimum Cloudera Manager version you can use.
  - Cloudera Manager is continually updated with configuration settings for features introduced in the latest Impala releases.
- If you are upgrading from CDH 5 beta to CDH 5.0 production, make sure you are using the appropriate CDH 5 repositories shown on the [CDH version and packaging](#) page, then follow the procedures throughout the rest of this section.
- Every time you upgrade to a new major or minor Impala release, see [Incompatible Changes in Impala](#) on page 487 in the *Release Notes* for any changes needed in your source code, startup scripts, and so on.
- Also check [Known Issues and Workarounds in Impala](#) on page 496 in the *Release Notes* for any issues or limitations that require workarounds.
- For the resource management feature to work (in combination with CDH 5 and the YARN and Llama components), the `impala` user must be a member of the `hdfs` group. This setup is performed automatically during a new install, but not when upgrading from earlier Impala releases to Impala 1.2. If you are upgrading a node to CDH 5 that already had Impala 1.1 or 1.0 installed, manually add the `impala` user to the `hdfs` group.

## Upgrading Impala through Cloudera Manager - Parcels

Parcels are an alternative binary distribution format available in Cloudera Manager 4.5 and higher.

- **Important:** In CDH 5, there is not a separate Impala parcel; Impala is part of the main CDH 5 parcel. Each level of CDH 5 has a corresponding version of Impala, and you upgrade Impala by upgrading CDH. See the [CDH 5 upgrade instructions](#) and choose the instructions for parcels. The remainder of this section only covers parcel upgrades for Impala under CDH 4.

To upgrade Impala for CDH 4 in a Cloudera Managed environment, using parcels:

## Upgrading Impala

1. If you originally installed using packages and now are switching to parcels, remove all the Impala-related packages first. You can check which packages are installed using one of the following commands, depending on your operating system:

```
rpm -qa          # RHEL, Oracle Linux, CentOS, Debian  
dpkg --get-selections # Debian
```

and then remove the packages using one of the following commands:

```
sudo yum remove pkg_names    # RHEL, Oracle Linux, CentOS  
sudo zypper remove pkg_names # SLES  
sudo apt-get purge pkg_names # Ubuntu, Debian
```

2. Connect to the Cloudera Manager Admin Console.
3. Go to the **Hosts > Parcels** tab. You should see a parcel with a newer version of Impala that you can upgrade to.
4. Click **Download**, then **Distribute**. (The button changes as each step completes.)
5. Click **Activate**.
6. When prompted, click **Restart** to restart the Impala service.

## Upgrading Impala through Cloudera Manager - Packages

To upgrade Impala in a Cloudera Managed environment, using packages:

1. Connect to the Cloudera Manager Admin Console.
2. In the **Services** tab, click the **Impala** service.
3. Click **Actions** and click **Stop**.
4. Use **one** of the following sets of commands to update Impala on each Impala node in your cluster:

**For RHEL, Oracle Linux, or CentOS systems:**

```
$ sudo yum update impala  
$ sudo yum update hadoop-lzo-cdh4 # Optional; if this package is already installed.
```

**For SUSE systems:**

```
$ sudo zypper update impala  
$ sudo zypper update hadoop-lzo-cdh4 # Optional; if this package is already installed
```

**For Debian or Ubuntu systems:**

```
$ sudo apt-get install impala  
$ sudo apt-get install hadoop-lzo-cdh4 # Optional; if this package is already installed
```

5. Use **one** of the following sets of commands to update Impala shell on each node on which it is installed:

**For RHEL, Oracle Linux, or CentOS systems:**

```
$ sudo yum update impala-shell
```

**For SUSE systems:**

```
$ sudo zypper update impala-shell
```

**For Debian or Ubuntu systems:**

```
$ sudo apt-get install impala-shell
```

6. Connect to the Cloudera Manager Admin Console.
7. In the **Services** tab, click the Impala service.
8. Click **Actions** and click **Start**.

## Upgrading Impala without Cloudera Manager

To upgrade Impala on a cluster not managed by Cloudera Manager, run these Linux commands on the appropriate hosts in your cluster:

1. Stop Impala services.
  - a. Stop `impalad` on each Impala node in your cluster:
 

```
$ sudo service impala-server stop
```
  - b. Stop any instances of the state store in your cluster:
 

```
$ sudo service impala-state-store stop
```
  - c. Stop any instances of the catalog service in your cluster:
 

```
$ sudo service impala-catalog stop
```
2. Check if there are new recommended or required configuration settings to put into place in the configuration files, typically under `/etc/impala/conf`. See [Post-Installation Configuration for Impala](#) on page 34 for settings related to performance and scalability.
3. Use **one** of the following sets of commands to update Impala on each Impala node in your cluster:

**For RHEL, Oracle Linux, or CentOS systems:**

```
$ sudo yum update impala-server
$ sudo yum update hadoop-lzo-cdh4 # Optional; if this package is already installed
$ sudo yum update impala-catalog # New in Impala 1.2; do yum install when upgrading from 1.1.
```

**For SUSE systems:**

```
$ sudo zypper update impala-server
$ sudo zypper update hadoop-lzo-cdh4 # Optional; if this package is already installed
$ sudo zypper update impala-catalog # New in Impala 1.2; do zypper install when upgrading from 1.1.
```

**For Debian or Ubuntu systems:**

```
$ sudo apt-get install impala-server
$ sudo apt-get install hadoop-lzo-cdh4 # Optional; if this package is already installed
$ sudo apt-get install impala-catalog # New in Impala 1.2.
```

4. Use **one** of the following sets of commands to update Impala shell on each node on which it is installed:

**For RHEL, Oracle Linux, or CentOS systems:**

```
$ sudo yum update impala-shell
```

## Upgrading Impala

### For SUSE systems:

```
$ sudo zypper update impala-shell
```

### For Debian or Ubuntu systems:

```
$ sudo apt-get install impala-shell
```

5. Depending on which release of Impala you are upgrading from, you might find that the symbolic links /etc/impala/conf and /usr/lib/impala/sbin are missing. If so, see [Known Issues in the Current Production Release \(Impala 2.2.x / CDH 5.4.x\)](#) on page 496 for the procedure to work around this problem.
6. Restart Impala services:
  - a. Restart the Impala state store service on the desired nodes in your cluster. Expect to see a process named statestored if the service started successfully.

```
$ sudo service impala-state-store start  
$ ps ax | grep [s]tatestored  
 6819 ?          S1      0:07 /usr/lib/impala/sbin/statestored  
 -log_dir=/var/log/impala -state_store_port=24000
```

Restart the state store service *before* the Impala server service to avoid "Not connected" errors when you run `impala-shell`.

5. Depending on which release of Impala you are upgrading from, you might find that the symbolic links /etc/impala/conf and /usr/lib/impala/sbin are missing. If so, see [Known Issues in the Current Production Release \(Impala 2.2.x / CDH 5.4.x\)](#) on page 496 for the procedure to work around this problem.
- b. Restart the Impala catalog service on whichever host it runs on in your cluster. Expect to see a process named catalogd if the service started successfully.

```
$ sudo service impala-catalog restart  
$ ps ax | grep [c]atalogd  
 6068 ?          S1      4:06 /usr/lib/impala/sbin/catalogd
```

- c. Restart the Impala daemon service on each node in your cluster. Expect to see a process named impalad if the service started successfully.

```
$ sudo service impala-server start  
$ ps ax | grep [i]mpalad  
 7936 ?          S1      0:12 /usr/lib/impala/sbin/impalad -log_dir=/var/log/impala  
 -state_store_port=24000 -use_statestore  
 -state_store_host=127.0.0.1 -be_port=22000
```

#### Note:

If the services did not start successfully (even though the `sudo service` command might display [OK]), check for errors in the Impala log file, typically in `/var/log/impala`.

# Starting Impala

To begin using Cloudera Impala:

1. Set any necessary configuration options for the Impala services. See [Modifying Impala Startup Options](#) on page 48 for details.
2. Start one instance of the Impala statestore. The statestore helps Impala to distribute work efficiently, and to continue running in the event of availability problems for other Impala nodes. If the statestore becomes unavailable, Impala continues to function.
3. Start one instance of the Impala catalog service.
4. Start the main Impala service on one or more DataNodes, ideally on all DataNodes to maximize local processing and avoid network traffic due to remote reads.

Once Impala is running, you can conduct interactive experiments using the instructions in [Impala Tutorials](#) on page 52 and try [Using the Impala Shell \(impala-shell Command\)](#) on page 332.

## Starting Impala through Cloudera Manager

If you installed Impala with Cloudera Manager, use Cloudera Manager to start and stop services. The Cloudera Manager GUI is a convenient way to check that all services are running, to set configuration options using form fields in a browser, and to spot potential issues such as low disk space before they become serious. Cloudera Manager automatically starts all the Impala-related services as a group, in the correct order. See [the Cloudera Manager Documentation](#) for details.

■ **Note:**

Currently, Impala UDFs and UDAs are not persisted in the metastore database. Information about these functions is held in the memory of the catalogd daemon. You must reload them by running the CREATE FUNCTION statements again each time you restart the catalogd daemon.

## Starting Impala from the Command Line

To start the Impala state store and Impala from the command line or a script, you can either use the service command or you can start the daemons directly through the impalad, statestored, and catalogd executables.

Start the Impala statestore and then start impalad instances. You can modify the values the service initialization scripts use when starting the statestore and Impala by editing /etc/default/impala.

Start the statestore service using a command similar to the following:

```
$ sudo service impala-state-store start
```

Start the catalog service using a command similar to the following:

```
$ sudo service impala-catalog start
```

Start the Impala service on each data node using a command similar to the following:

```
$ sudo service impala-server start
```

## Starting Impala

- **Note:**

Currently, Impala UDFs and UDAs are not persisted in the metastore database. Information about these functions is held in the memory of the `catalogd` daemon. You must reload them by running the `CREATE FUNCTION` statements again each time you restart the `catalogd` daemon.

If any of the services fail to start, review:

- [Reviewing Impala Logs](#) on page 444
- [Troubleshooting Impala](#) on page 447

## Modifying Impala Startup Options

The configuration options for the Impala-related daemons let you choose which hosts and ports to use for the services that run on a single host, specify directories for logging, control resource usage and security, and specify other aspects of the Impala software.

### Configuring Impala Startup Options through Cloudera Manager

If you manage your cluster through Cloudera Manager, configure the settings for all the Impala-related daemons by navigating to this page: **Clusters > Services > Impala > Configuration > View and Edit**. See the Cloudera Manager documentation for instructions about how to configure Impala through Cloudera Manager: [Impala information for Cloudera Manager 5](#), [Impala information for Cloudera Manager 4](#).

If the Cloudera Manager interface does not yet have a form field for a newly added option, or if you need to use special options for debugging and troubleshooting, the **Advanced** option page for each daemon includes one or more fields where you can enter option names directly. In Cloudera Manager 4, these fields are labelled **Safety Valve**; in Cloudera Manager 5, they are called **Advanced Configuration Snippet**. There is also a free-form field for query options, on the top-level **Impala Daemon** options page.

### Configuring Impala Startup Options through the Command Line

When you run Impala in a non-Cloudera Manager environment, the Impala server, statestore, and catalog services start up using values provided in a defaults file, `/etc/default/impala`.

This file includes information about many resources used by Impala. Most of the defaults included in this file should be effective in most cases. For example, typically you would not change the definition of the `CLASSPATH` variable, but you would always set the address used by the statestore server. Some of the content you might modify includes:

```
IMPALA_STATE_STORE_HOST=127.0.0.1
IMPALA_STATE_STORE_PORT=24000
IMPALA_BACKEND_PORT=22000
IMPALA_LOG_DIR=/var/log/impala
IMPALA_CATALOG_SERVICE_HOST=...
IMPALA_STATE_STORE_HOST=...

export IMPALA_STATE_STORE_ARGS=${IMPALA_STATE_STORE_ARGS:- \
    -log_dir=${IMPALA_LOG_DIR} -state_store_port=${IMPALA_STATE_STORE_PORT} }
IMPALA_SERVER_ARGS=" \
    -log_dir=${IMPALA_LOG_DIR} \
    -catalog_service_host=${IMPALA_CATALOG_SERVICE_HOST} \
    -state_store_port=${IMPALA_STATE_STORE_PORT} \
    -use_statestore \
    -state_store_host=${IMPALA_STATE_STORE_HOST} \
    -be_port=${IMPALA_BACKEND_PORT}"
export ENABLE_CORE_DUMPS=${ENABLE_COREDUMPS:-false}
```

To use alternate values, edit the defaults file, then restart all the Impala-related services so that the changes take effect. Restart the Impala server using the following commands:

```
$ sudo service impala-server restart
Stopping Impala Server: [ OK ]
Starting Impala Server: [ OK ]
```

Restart the Impala statestore using the following commands:

```
$ sudo service impala-state-store restart
Stopping Impala State Store Server: [ OK ]
Starting Impala State Store Server: [ OK ]
```

Restart the Impala catalog service using the following commands:

```
$ sudo service impala-catalog restart
Stopping Impala Catalog Server: [ OK ]
Starting Impala Catalog Server: [ OK ]
```

Some common settings to change include:

- Statestore address. Cloudera recommends the statestore be on a separate host not running the `impalad` daemon. In that recommended configuration, the `impalad` daemon cannot refer to the statestore server using the loopback address. If the statestore is hosted on a machine with an IP address of 192.168.0.27, change:

```
IMPALA_STATE_STORE_HOST=127.0.0.1
```

to:

```
IMPALA_STATE_STORE_HOST=192.168.0.27
```

- Catalog server address (including both the hostname and the port number). Update the value of the `IMPALA_CATALOG_SERVICE_HOST` variable. Cloudera recommends the catalog server be on the same host as the statestore. In that recommended configuration, the `impalad` daemon cannot refer to the catalog server using the loopback address. If the catalog service is hosted on a machine with an IP address of 192.168.0.27, add the following line:

```
IMPALA_CATALOG_SERVICE_HOST=192.168.0.27:26000
```

The `/etc/default/impala` defaults file currently does not define an `IMPALA_CATALOG_ARGS` environment variable, but if you add one it will be recognized by the service startup/shutdown script. Add a definition for this variable to `/etc/default/impala` and add the option `-catalog_service_host=hostname`. If the port is different than the default 26000, also add the option `-catalog_service_port=port`.

- Memory limits. You can limit the amount of memory available to Impala. For example, to allow Impala to use no more than 70% of system memory, change:

```
export IMPALA_SERVER_ARGS=${IMPALA_SERVER_ARGS:- \
-log_dir=${IMPALA_LOG_DIR} \
-state_store_port=${IMPALA_STATE_STORE_PORT} \
-use_statestore -state_store_host=${IMPALA_STATE_STORE_HOST} \
-be_port=${IMPALA_BACKEND_PORT}}
```

to:

```
export IMPALA_SERVER_ARGS=${IMPALA_SERVER_ARGS:- \
-log_dir=${IMPALA_LOG_DIR} -state_store_port=${IMPALA_STATE_STORE_PORT} \
-use_statestore -state_store_host=${IMPALA_STATE_STORE_HOST} \
-be_port=${IMPALA_BACKEND_PORT} -mem_limit=70%}
```

## Starting Impala

You can specify the memory limit using absolute notation such as 500m or 2G, or as a percentage of physical memory such as 60%.

- **Note:** Queries that exceed the specified memory limit are aborted. Percentage limits are based on the physical memory of the machine and do not consider cgroups.

- Core dump enablement. To enable core dumps, change:

```
export ENABLE_CORE_DUMPS=${ENABLE_COREDUMPS:-false}
```

to:

```
export ENABLE_CORE_DUMPS=${ENABLE_COREDUMPS:-true}
```

- **Note:** The location of core dump files may vary according to your operating system configuration. Other security settings may prevent Impala from writing core dumps even when this option is enabled.

- Authorization using the open source Sentry plugin. Specify the `-server_name` and `-authorization_policy_file` options as part of the `IMPALA_SERVER_ARGS` and `IMPALA_STATE_STORE_ARGS` settings to enable the core Impala support for authentication. See [Starting the impalad Daemon with Sentry Authorization Enabled](#) on page 92 for details.
- Auditing for successful or blocked Impala queries, another aspect of security. Specify the `-audit_event_log_dir=directory_path` option and optionally the `-max_audit_event_log_file_size=number_of_queries` and `-abort_on_failed_audit_event` options as part of the `IMPALA_SERVER_ARGS` settings, for each Impala node, to enable and customize auditing. See [Auditing Impala Operations](#) on page 108 for details.
- Password protection for the Impala web UI, which listens on port 25000 by default. This feature involves adding some or all of the `--webserver_password_file`, `--webserver_authentication_domain`, and `--webserver_certificate_file` options to the `IMPALA_SERVER_ARGS` and `IMPALA_STATE_STORE_ARGS` settings. See [Security Guidelines for Impala](#) on page 87 for details.
- Another setting you might add to `IMPALA_SERVER_ARGS` is:

```
-default_query_options='option=value;option=value;...'
```

These options control the behavior of queries performed by this `impalad` instance. The option values you specify here override the default values for [Impala query options](#), as shown by the `SET` statement in `impala-shell`.

- Options for resource management, in conjunction with the YARN and Llama components. These options include `-enable_rm`, `-llama_host`, `-llama_port`, `-llama_callback_port`, and `-cgroup_hierarchy_path`. Additional options to help fine-tune the resource estimates are `--rm_always_use_defaults`, `--rm_default_memory=size`, and `--rm_default_cpu_cores`. For details about these options, see [impalad Startup Options for Resource Management](#) on page 80. See [Integrated Resource Management with YARN](#) on page 78 for information about resource management in general, and [The Llama Daemon](#) on page 78 for information about the Llama daemon.
- During troubleshooting, Cloudera Support might direct you to change other values, particularly for `IMPALA_SERVER_ARGS`, to work around issues or gather debugging information.

The following startup options for `impalad` enable resource management and customize its parameters for your cluster configuration:

- `-enable_rm`: Whether to enable resource management or not, either `true` or `false`. The default is `false`. None of the other resource management options have any effect unless `-enable_rm` is turned on.

- `-llama_host`: Hostname or IP address of the Llama service that Impala should connect to. The default is 127.0.0.1.
- `-llama_port`: Port of the Llama service that Impala should connect to. The default is 15000.
- `-llama_callback_port`: Port that Impala should start its Llama callback service on. Llama reports when resources are granted or preempted through that service.
- `-cgroup_hierarchy_path`: Path where YARN and Llama will create cgroups for granted resources. Impala assumes that the cgroup for an allocated container is created in the path '`cgroup_hierarchy_path + container_id`'.
- `-rm_always_use_defaults`: If this Boolean option is enabled, Impala ignores computed estimates and always obtains the default memory and CPU allocation from Llama at the start of the query. These default estimates are approximately 2 CPUs and 4 GB of memory, possibly varying slightly depending on cluster size, workload, and so on. Cloudera recommends enabling `-rm_always_use_defaults` whenever resource management is used, and relying on these default values (that is, leaving out the two following options).
- `-rm_default_memory=size`: Optionally sets the default estimate for memory usage for each query. You can use suffixes such as M and G for megabytes and gigabytes, the same as with the [MEM\\_LIMIT](#) query option. Only has an effect when `-rm_always_use_defaults` is also enabled.
- `-rm_default_cpu_cores`: Optionally sets the default estimate for number of virtual CPU cores for each query. Only has an effect when `-rm_always_use_defaults` is also enabled.

▪ **Note:**

These startup options for the `impalad` daemon are different from the command-line options for the `impala-shell` command. For the `impala-shell` options, see [impala-shell Configuration Options](#) on page 332.

## Checking the Values of Impala Configuration Options

You can check the current runtime value of all these settings through the Impala web interface, available by default at `http://impala_hostname:25000/varz` for the `impalad` daemon, `http://impala_hostname:25010/varz` for the `statestored` daemon, or `http://impala_hostname:25020/varz` for the `catalogd` daemon. In the Cloudera Manager interface, you can see the link to the appropriate **`service_name` Web UI** page when you look at the status page for a specific daemon on a specific host.

## Startup Options for `impalad` Daemon

The `impalad` daemon implements the main Impala service, which performs query processing and reads and writes the data files.

## Startup Options for `statestored` Daemon

The `statestored` daemon implements the Impala statestore service, which monitors the availability of Impala services across the cluster, and handles situations such as nodes becoming unavailable or becoming available again.

## Startup Options for `catalogd` Daemon

The `catalogd` daemon implements the Impala catalog service, which broadcasts metadata changes to all the Impala nodes when Impala creates a table, inserts data, or performs other kinds of DDL and DML operations.

By default, the metadata loading and caching on startup happens asynchronously, so Impala can begin accepting requests promptly. To enable the original behavior, where Impala waited until all metadata was loaded before accepting any requests, set the `catalogd` configuration option `--load_catalog_in_background=false`.

# Impala Tutorials

This section includes tutorial scenarios that demonstrate how to begin using Impala once the software is installed. It focuses on techniques for loading data, because once you have some data in tables and can query that data, you can quickly progress to more advanced Impala features.

■ **Note:**

Where practical, the tutorials take you from “ground zero” to having the desired Impala tables and data. In some cases, you might need to download additional files from outside sources, set up additional software components, modify commands or scripts to fit your own configuration, or substitute your own sample data.

Before trying these tutorial lessons, install Impala using one of these procedures:

- If you already have a CDH environment set up and just need to add Impala to it, follow the installation process described in [Installing Impala](#) on page 30. Make sure to also install the Hive metastore service if you do not already have Hive configured.
- To set up Impala and all its prerequisites at once, in a minimal configuration that you can use for small-scale experiments, set up the Cloudera QuickStart VM, which includes CDH and Impala on CentOS 6.3 (64-bit). Use this single-node VM to try out basic SQL functionality, not anything related to performance and scalability. For more information, see [the Cloudera QuickStart VM](#).

## Tutorials for Getting Started

These tutorials demonstrate the basics of using Impala. They are intended for first-time users, and for trying out Impala on any new cluster to make sure the major components are working correctly.

### Explore a New Impala Instance

This tutorial demonstrates techniques for finding your way around the tables and databases of an unfamiliar (possibly empty) Impala instance.

When you connect to an Impala instance for the first time, you use the `SHOW DATABASES` and `SHOW TABLES` statements to view the most common types of objects. Also, call the `version()` function to confirm which version of Impala you are running; the version number is important when consulting documentation and dealing with support issues.

A completely empty Impala instance contains no tables, but still has two databases:

- `default`, where new tables are created when you do not specify any other database.
- `_impala_builtins`, a system database used to hold all the built-in functions.

The following example shows how to see the available databases, and the tables in each. If the list of databases or tables is long, you can use wildcard notation to locate specific databases or tables based on their names.

```
$ impala-shell -i localhost --quiet
Starting Impala Shell without Kerberos authentication
Welcome to the Impala shell. Press TAB twice to see a list of available commands.

Copyright (c) 2012 Cloudera, Inc. All rights reserved.

(Shell build version: Impala Shell v...
[localhost:21000] > select version();
+-----+
| version() |
+-----+
```

```

| impalad version ...
| Built on ...
+-----+
[localhost:21000] > show databases;
+-----+
| name |
+-----+
| _impala_builtins |
| ctas |
| d1 |
| d2 |
| d3 |
| default |
| explain_plans |
| external_table |
| file_formats |
| tpc |
+-----+
[localhost:21000] > select current_database();
+-----+
| current_database() |
+-----+
| default |
+-----+
[localhost:21000] > show tables;
+-----+
| name |
+-----+
| ex_t |
| t1 |
+-----+
[localhost:21000] > show tables in d3;
[localhost:21000] > show tables in tpc;
+-----+
| name |
+-----+
| city |
| customer |
| customer_address |
| customer_demographics |
| household_demographics |
| item |
| promotion |
| store |
| store2 |
| store_sales |
| ticket_view |
| time_dim |
| tpc_tables |
+-----+
[localhost:21000] > show tables in tpc like 'customer*';
+-----+
| name |
+-----+
| customer |
| customer_address |
| customer_demographics |
+-----+

```

Once you know what tables and databases are available, you descend into a database with the `USE` statement. To understand the structure of each table, you use the `DESCRIBE` command. Once inside a database, you can issue statements such as `INSERT` and `SELECT` that operate on particular tables.

The following example explores a database named `TPC` whose name we learned in the previous example. It shows how to filter the table names within a database based on a search string, examine the columns of a table, and run queries to examine the characteristics of the table data. For example, for an unfamiliar table you might want to know the number of rows, the number of different values for a column, and other properties such as

## Impala Tutorials

whether the column contains any NULL values. When sampling the actual data values from a table, use a LIMIT clause to avoid excessive output if the table contains more rows or distinct values than you expect.

```
[localhost:21000] > use tpc;
[localhost:21000] > show tables like '*view*';
+-----+
| name |
+-----+
| ticket_view |
+-----+
[localhost:21000] > describe city;
+-----+-----+-----+
| name | type | comment |
+-----+-----+-----+
id	int
name	string
countrycode	string
district	string
population	int
+-----+-----+-----+	
[localhost:21000] > select count(*) from city;	
+-----+	
count(*)	
+-----+	
0	
+-----+	
[localhost:21000] > desc customer;	
+-----+-----+-----+	
name	type
+-----+-----+-----+	
c_customer_sk	int
c_customer_id	string
c_current_cdemo_sk	int
c_current_hdemo_sk	int
c_current_addr_sk	int
c_first_shipto_date_sk	int
c_first_sales_date_sk	int
c_salutation	string
c_first_name	string
c_last_name	string
c_preferred_cust_flag	string
c_birth_day	int
c_birth_month	int
c_birth_year	int
c_birth_country	string
c_login	string
c_email_address	string
c_last_review_date	string
+-----+-----+-----+	
[localhost:21000] > select count(*) from customer;	
+-----+	
count(*)	
+-----+	
100000	
+-----+	
[localhost:21000] > select count(distinct c_birth_month) from customer;	
+-----+	
count(distinct c_birth_month)	
+-----+	
12	
+-----+	
[localhost:21000] > select count(*) from customer where c_email_address is null;	
+-----+	
count(*)	
+-----+	
0	
+-----+	
[localhost:21000] > select distinct c_salutation from customer limit 10;	
+-----+	
c_salutation	
+-----+	
Mr.	
Ms.	
```

|      |
|------|
| Dr.  |
| Miss |
| Sir  |
| Mrs. |

When you graduate from read-only exploration, you use statements such as `CREATE DATABASE` and `CREATE TABLE` to set up your own database objects.

The following example demonstrates creating a new database holding a new table. Although the last example ended inside the `TPC` database, the new `EXPERIMENTS` database is not nested inside `TPC`; all databases are arranged in a single top-level list.

```
[localhost:21000] > create database experiments;
[localhost:21000] > show databases;
+-----+
| name |
+-----+
| _impala_builtin |
| ctas |
| d1 |
| d2 |
| d3 |
| default |
| experiments |
| explain_plans |
| external_table |
| file_formats |
| tpc |
+-----+
[localhost:21000] > show databases like 'exp*';
+-----+
| name |
+-----+
| experiments |
| explain_plans |
+-----+
```

The following example creates a new table, `t1`. To illustrate a common mistake, it creates this table inside the wrong database, the `TPC` database where the previous example ended. The `ALTER TABLE` statement lets you move the table to the intended database, `EXPERIMENTS`, as part of a rename operation. The `USE` statement is always needed to switch to a new database, and the `current_database()` function confirms which database the session is in, to avoid these kinds of mistakes.

```
[localhost:21000] > create table t1 (x int);
[localhost:21000] > show tables;
+-----+
| name |
+-----+
| city |
| customer |
| customer_address |
| customer_demographics |
| household_demographics |
| item |
| promotion |
| store |
| store2 |
| store_sales |
| t1 |
| ticket_view |
| time_dim |
| tpc_tables |
+-----+
[localhost:21000] > select current_database();
+-----+
| current_database() |
+-----+
```

```
+-----+  
| tpc |  
+-----+  
[localhost:21000] > alter table t1 rename to experiments.t1;  
[localhost:21000] > use experiments;  
[localhost:21000] > show tables;  
+---+  
| name |  
+---+  
| t1 |  
+---+  
[localhost:21000] > select current_database();  
+-----+  
| current_database() |  
+-----+  
| experiments |  
+-----+
```

For your initial experiments with tables, you can use ones with just a few columns and a few rows, and text-format data files.

- **Note:** As you graduate to more realistic scenarios, you will use more elaborate tables with many columns, features such as partitioning, and file formats such as Parquet. When dealing with realistic data volumes, you will bring in data using `LOAD DATA` or `INSERT ... SELECT` statements to operate on millions or billions of rows at once.

The following example sets up a couple of simple tables with a few rows, and performs queries involving sorting, aggregate functions and joins.

```
[localhost:21000] > insert into t1 values (1), (3), (2), (4);  
[localhost:21000] > select x from t1 order by x desc;  
+---+  
| x |  
+---+  
| 4 |  
| 3 |  
| 2 |  
| 1 |  
+---+  
[localhost:21000] > select min(x), max(x), sum(x), avg(x) from t1;  
+-----+-----+-----+-----+  
| min(x) | max(x) | sum(x) | avg(x) |  
+-----+-----+-----+-----+  
| 1      | 4      | 10     | 2.5    |  
+-----+-----+-----+-----+  
  
[localhost:21000] > create table t2 (id int, word string);  
[localhost:21000] > insert into t2 values (1, "one"), (3, "three"), (5, 'five');  
[localhost:21000] > select word from t1 join t2 on (t1.x = t2.id);  
+---+  
| word |  
+---+  
| one |  
| three |  
+---+
```

After completing this tutorial, you should now know:

- How to tell which version of Impala is running on your system.
- How to find the names of databases in an Impala instance, either displaying the full list or searching for specific names.
- How to find the names of tables in an Impala database, either displaying the full list or searching for specific names.
- How to switch between databases and check which database you are currently in.
- How to learn the column names and types of a table.
- How to create databases and tables, insert small amounts of test data, and run simple queries.

## Load CSV Data from Local Files

This scenario illustrates how to create some very small tables, suitable for first-time users to experiment with Impala SQL features. TAB1 and TAB2 are loaded with data from files in HDFS. A subset of data is copied from TAB1 into TAB3.

Populate HDFS with the data you want to query. To begin this process, create one or more new subdirectories underneath your user directory in HDFS. The data for each table resides in a separate subdirectory. Substitute your own user name for `cloudera` where appropriate. This example uses the `-p` option with the `mkdir` operation to create any necessary parent directories if they do not already exist.

```
$ whoami
cloudera
$ hdfs dfs -ls /user
Found 3 items
drwxr-xr-x  - cloudera cloudera          0 2013-04-22 18:54 /user/cloudera
drwxrwx---  - mapred   mapred            0 2013-03-15 20:11 /user/history
drwxr-xr-x  - hue      supergroup        0 2013-03-15 20:10 /user/hive

$ hdfs dfs -mkdir -p /user/cloudera/sample_data/tab1 /user/cloudera/sample_data/tab2
```

Here is some sample data, for two tables named TAB1 and TAB2.

Copy the following content to .csv files in your local filesystem:

`tab1.csv`:

```
1,true,123.123,2012-10-24 08:55:00
2,false,1243.5,2012-10-25 13:40:00
3,false,24453.325,2008-08-22 09:33:21.123
4,false,243423.325,2007-05-12 22:32:21.33454
5,true,243.325,1953-04-22 09:11:33
```

`tab2.csv`:

```
1,true,12789.123
2,false,1243.5
3,false,24453.325
4,false,2423.3254
5,true,243.325
60,false,243565423.325
70,true,243.325
80,false,243423.325
90,true,243.325
```

Put each .csv file into a separate HDFS directory using commands like the following, which use paths available in the Impala Demo VM:

```
$ hdfs dfs -put tab1.csv /user/cloudera/sample_data/tab1
$ hdfs dfs -ls /user/cloudera/sample_data/tab1
Found 1 items
-rw-r--r--  1 cloudera cloudera    192 2013-04-02 20:08
/usr/cloudera/sample_data/tab1/tab1.csv

$ hdfs dfs -put tab2.csv /user/cloudera/sample_data/tab2
$ hdfs dfs -ls /user/cloudera/sample_data/tab2
Found 1 items
-rw-r--r--  1 cloudera cloudera    158 2013-04-02 20:09
/usr/cloudera/sample_data/tab2/tab2.csv
```

The name of each data file is not significant. In fact, when Impala examines the contents of the data directory for the first time, it considers all files in the directory to make up the data of the table, regardless of how many files there are or what the files are named.

## Impala Tutorials

To understand what paths are available within your own HDFS filesystem and what the permissions are for the various directories and files, issue `hdfs dfs -ls /` and work your way down the tree doing `-ls` operations for the various directories.

Use the `impala-shell` command to create tables, either interactively or through a SQL script.

The following example shows creating three tables. For each table, the example shows creating columns with various attributes such as Boolean or integer types. The example also includes commands that provide information about how the data is formatted, such as rows terminating with commas, which makes sense in the case of importing data from a `.csv` file. Where we already have `.csv` files containing data in the HDFS directory tree, we specify the location of the directory containing the appropriate `.csv` file. Impala considers all the data from all the files in that directory to represent the data for the table.

```
DROP TABLE IF EXISTS tab1;
-- The EXTERNAL clause means the data is located outside the central location
-- for Impala data files and is preserved when the associated Impala table is dropped.
-- We expect the data to already exist in the directory specified by the LOCATION
-- clause.
CREATE EXTERNAL TABLE tab1
(
    id INT,
    col_1 BOOLEAN,
    col_2 DOUBLE,
    col_3 TIMESTAMP
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION '/user/cloudera/sample_data/tab1';

DROP TABLE IF EXISTS tab2;
-- TAB2 is an external table, similar to TAB1.
CREATE EXTERNAL TABLE tab2
(
    id INT,
    col_1 BOOLEAN,
    col_2 DOUBLE
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
LOCATION '/user/cloudera/sample_data/tab2';

DROP TABLE IF EXISTS tab3;
-- Leaving out the EXTERNAL clause means the data will be managed
-- in the central Impala data directory tree. Rather than reading
-- existing data files when the table is created, we load the
-- data after creating the table.
CREATE TABLE tab3
(
    id INT,
    col_1 BOOLEAN,
    col_2 DOUBLE,
    month INT,
    day INT
)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',';
```

- **Note:** Getting through these `CREATE TABLE` statements successfully is an important validation step to confirm everything is configured correctly with the Hive metastore and HDFS permissions. If you receive any errors during the `CREATE TABLE` statements:

- Make sure you followed the installation instructions closely, in [Installing Impala](#) on page 30.
- Make sure the `hive.metastore.warehouse.dir` property points to a directory that Impala can write to. The ownership should be `hive:hive`, and the `impala` user should also be a member of the `hive` group.
- If the value of `hive.metastore.warehouse.dir` is different in the Cloudera Manager dialogs and in the Hive shell, you might need to [designate the hosts running impalad with the "gateway" role for Hive](#), and [deploy the client configuration files](#) to those hosts.

## Point an Impala Table at Existing Data Files

A convenient way to set up data for Impala to access is to use an external table, where the data already exists in a set of HDFS files and you just point the Impala table at the directory containing those files. For example, you might run in `impala-shell` a `*.sql` file with contents similar to the following, to create an Impala table that accesses an existing data file used by Hive.

The following examples set up 2 tables, referencing the paths and sample data supplied with the Cloudera QuickStart VM. For historical reasons, the data physically resides in an HDFS directory tree under `/user/hive`, although this particular data is entirely managed by Impala rather than Hive. When we create an external table, we specify the directory containing one or more data files, and Impala queries the combined content of all the files inside that directory. Here is how we examine the directories and files within the HDFS filesystem:

```
$ cd ~/cloudera/datasets
$ ./tpcds-setup.sh
... Downloads and unzips the kit, builds the data and loads it into HDFS ...
$ hdfs dfs -ls /user/hive/tpcds/customer
Found 1 items
-rw-r--r-- 1 cloudera supergroup 13209372 2013-03-22 18:09
/usr/hive/tpcds/customer/customer.dat
$ hdfs dfs -cat /user/hive/tpcds/customer/customer.dat | more
1|AAAAAAAAABAAAAAAA|980124|7135|32946|2452238|2452208|Mr.|Javier|Lewis|Y|9|12|1936|CHILE||Javier.Lewis@VFAxlnZEvOx.org|2452508|
2|AAAAAAAAACAAAAAA|819667|1461|31655|2452318|2452288|Dr.|Amy|Moses|Y|9|4|1966|TOGO||Amy.Moses@Ovk9KjHH.com|2452318|
3|AAAAAAAADAAAAAA|1473522|6247|48572|2449130|2449100|Miss|Latisha|Hamilton|N|18|9|1979|NIUE||Latisha.Hamilton@V.com|2452313|
4|AAAAAAAEEAAAAAA|1703214|3986|39558|2450030|2450000|Dr.|Michael|White|N|7|6|1983|MEXICO||Michael.White@i.org|2452361|
5|AAAAAAAFAAAAAAA|953372|4470|36368|2449438|2449408|Sir|Robert|Moran|N|8|5|1956|FIJI||Robert.Moran@Hh.edu|2452469|
...
...
```

Here is a SQL script to set up Impala tables pointing to some of these data files in HDFS. (The script in the VM sets up tables like this through Hive; ignore those tables for purposes of this demonstration.) Save the following as `customer_setup.sql`:

```
-- 
-- store_sales fact table and surrounding dimension tables only
--
create database tpcds;
use tpcds;

drop table if exists customer;
create external table customer
(
    c_customer_sk          int,
    c_customer_id           string,
    c_current_cdemo_sk     int,
    c_current_hdemo_sk     int,
    c_current_addr_sk      int,
    c_first_shipto_date_sk int,
    c_first_sales_date_sk  int,
    c_salutation            string,
    c_first_name             string,
    c_last_name              string,
    c_preferred_cust_flag   string,
    c_birth_day              int,
    c_birth_month             int,
    c_birth_year              int,
    c_birth_country            string,
    c_login                  string,
    c_email_address            string,
    c_last_review_date        string
)
row format delimited fields terminated by '|'
location '/user/hive/tpcds/customer';

drop table if exists customer_address;
```

```
create external table customer_address
(
    ca_address_sk          int,
    ca_address_id           string,
    ca_street_number        string,
    ca_street_name          string,
    ca_street_type          string,
    ca_suite_number         string,
    ca_city                 string,
    ca_county               string,
    ca_state                string,
    ca_zip                  string,
    ca_country              string,
    ca_gmt_offset            float,
    ca_location_type         string
)
row format delimited fields terminated by '|'
location '/user/hive/tpcds/customer_address';
```

We would run this script with a command such as:

```
impala-shell -i localhost -f customer_setup.sql
```

## Describe the Impala Table

Now that you have updated the database metadata that Impala caches, you can confirm that the expected tables are accessible by Impala and examine the attributes of one of the tables. We created these tables in the database named `default`. If the tables were in a database other than the `default`, we would issue a command `use db_name` to switch to that database before examining or querying its tables. We could also qualify the name of a table by prepending the database name, for example `default.customer` and `default.customer_name`.

```
[impala-host:21000] > show databases
Query finished, fetching results ...
default
Returned 1 row(s) in 0.00s
[impala-host:21000] > show tables
Query finished, fetching results ...
customer
customer_address
Returned 2 row(s) in 0.00s
[impala-host:21000] > describe customer_address
+-----+-----+-----+
| name      | type   | comment |
+-----+-----+-----+
| ca_address_sk  int    |
| ca_address_id   string |
| ca_street_number string |
| ca_street_name  string |
| ca_street_type   string |
| ca_suite_number string |
| ca_city          string |
| ca_county         string |
| ca_state          string |
| ca_zip            string |
| ca_country         string |
| ca_gmt_offset     float  |
| ca_location_type  string |
+-----+-----+-----+
Returned 13 row(s) in 0.01
```

## Query the Impala Table

You can query data contained in the tables. Impala coordinates the query execution across a single node or multiple nodes depending on your configuration, without the overhead of running MapReduce jobs to perform the intermediate processing.

There are a variety of ways to execute queries on Impala:

- Using the `impala-shell` command in interactive mode:

```
$ impala-shell -i impala-host
Connected to localhost:21000
[impala-host:21000] > select count(*) from customer_address;
50000
Returned 1 row(s) in 0.37s
```

- Passing a set of commands contained in a file:

```
$ impala-shell -i impala-host -f myquery.sql
Connected to localhost:21000
50000
Returned 1 row(s) in 0.19s
```

- Passing a single command to the `impala-shell` command. The query is executed, the results are returned, and the shell exits. Make sure to quote the command, preferably with single quotation marks to avoid shell expansion of characters such as \*.

```
$ impala-shell -i impala-host -q 'select count(*) from customer_address'
Connected to localhost:21000
50000
Returned 1 row(s) in 0.29s
```

## Data Loading and Querying Examples

This section describes how to create some sample tables and load data into them. These tables can then be queried using the Impala shell.

### Loading Data

Loading data involves:

- Establishing a data set. The example below uses `.csv` files.
- Creating tables to which to load data.
- Loading the data into the tables you created.

### Sample Queries

To run these sample queries, create a SQL query file `query.sql`, copy and paste each query into the query file, and then run the query file using the shell. For example, to run `query.sql` on `impala-host`, you might use the command:

```
impala-shell.sh -i impala-host -f query.sql
```

The examples and results below assume you have loaded the sample data into the tables as described above.

#### Example: Examining Contents of Tables

Let's start by verifying that the tables do contain the data we expect. Because Impala often deals with tables containing millions or billions of rows, when examining tables of unknown size, include the `LIMIT` clause to avoid huge amounts of unnecessary output, as in the final query. (If your interactive query starts displaying an unexpected volume of data, press `Ctrl-C` in `impala-shell` to cancel the query.)

```
SELECT * FROM tab1;
SELECT * FROM tab2;
SELECT * FROM tab2 LIMIT 5;
```

# Impala Tutorials

Results:

| id | col_1 | col_2      | col_3                         |
|----|-------|------------|-------------------------------|
| 1  | true  | 123.123    | 2012-10-24 08:55:00           |
| 2  | false | 1243.5     | 2012-10-25 13:40:00           |
| 3  | false | 24453.325  | 2008-08-22 09:33:21.123000000 |
| 4  | false | 243423.325 | 2007-05-12 22:32:21.334540000 |
| 5  | true  | 243.325    | 1953-04-22 09:11:33           |

| id | col_1 | col_2         |
|----|-------|---------------|
| 1  | true  | 12789.123     |
| 2  | false | 1243.5        |
| 3  | false | 24453.325     |
| 4  | false | 2423.3254     |
| 5  | true  | 243.325       |
| 60 | false | 243565423.325 |
| 70 | true  | 243.325       |
| 80 | false | 243423.325    |
| 90 | true  | 243.325       |

| id | col_1 | col_2     |
|----|-------|-----------|
| 1  | true  | 12789.123 |
| 2  | false | 1243.5    |
| 3  | false | 24453.325 |
| 4  | false | 2423.3254 |
| 5  | true  | 243.325   |

Example: Aggregate and Join

```
SELECT tab1.col_1, MAX(tab2.col_2), MIN(tab2.col_2)
FROM tab2 JOIN tab1 USING (id)
GROUP BY col_1 ORDER BY 1 LIMIT 5;
```

Results:

| col_1 | max(tab2.col_2) | min(tab2.col_2) |
|-------|-----------------|-----------------|
| false | 24453.325       | 1243.5          |
| true  | 12789.123       | 243.325         |

Example: Subquery, Aggregate and Joins

```
SELECT tab2.*
FROM tab2,
(SELECT tab1.col_1, MAX(tab2.col_2) AS max_col2
 FROM tab2, tab1
 WHERE tab1.id = tab2.id
 GROUP BY col_1) subquery1
WHERE subquery1.max_col2 = tab2.col_2;
```

Results:

| id | col_1 | col_2     |
|----|-------|-----------|
| 1  | true  | 12789.123 |
| 3  | false | 24453.325 |

### Example: INSERT Query

```
INSERT OVERWRITE TABLE tab3
SELECT id, col_1, col_2, MONTH(col_3), DAYOFMONTH(col_3)
FROM tab1 WHERE YEAR(col_3) = 2012;
```

Query TAB3 to check the result:

```
SELECT * FROM tab3;
```

Results:

| id | col_1 | col_2   | month | day |
|----|-------|---------|-------|-----|
| 1  | true  | 123.123 | 10    | 24  |
| 2  | false | 1243.5  | 10    | 25  |

## Advanced Tutorials

These tutorials walk you through advanced scenarios or specialized features.

### Attaching an External Partitioned Table to an HDFS Directory Structure

This tutorial shows how you might set up a directory tree in HDFS, put data files into the lowest-level subdirectories, and then use an Impala external table to query the data files from their original locations.

The tutorial uses a table with web log data, with separate subdirectories for the year, month, day, and host. For simplicity, we use a tiny amount of CSV data, loading the same data into each partition.

First, we make an Impala partitioned table for CSV data, and look at the underlying HDFS directory structure to understand the directory structure to re-create elsewhere in HDFS. The columns `field1`, `field2`, and `field3` correspond to the contents of the CSV data files. The `year`, `month`, `day`, and `host` columns are all represented as subdirectories within the table structure, and are not part of the CSV files. We use `STRING` for each of these columns so that we can produce consistent subdirectory names, with leading zeros for a consistent length.

```
create database external_partitions;
use external_partitions;
create table logs (field1 string, field2 string, field3 string)
partitioned by (year string, month string, day string, host string)
row format delimited fields terminated by ',';
insert into logs partition (year="2013", month="07", day="28", host="host1") values
("foo","foo","foo");
insert into logs partition (year="2013", month="07", day="28", host="host2") values
("foo","foo","foo");
insert into logs partition (year="2013", month="07", day="29", host="host1") values
("foo","foo","foo");
insert into logs partition (year="2013", month="07", day="29", host="host2") values
("foo","foo","foo");
insert into logs partition (year="2013", month="08", day="01", host="host1") values
("foo","foo","foo");
```

Back in the Linux shell, we examine the HDFS directory structure. (Your Impala data directory might be in a different location; for historical reasons, it is sometimes under the HDFS path `/user/hive/warehouse`.) We use the `hdfs dfs -ls` command to examine the nested subdirectories corresponding to each partitioning column, with separate subdirectories at each level (with `=` in their names) representing the different values for

## Impala Tutorials

each partitioning column. When we get to the lowest level of subdirectory, we use the `hdfs dfs -cat` command to examine the data file and see CSV-formatted data produced by the `INSERT` statement in Impala.

```
$ hdfs dfs -ls /user/impala/warehouse/external_partitions.db
Found 1 items
drwxrwxrwt - impala hive 0 2013-08-07 12:24
/user/impala/warehouse/external_partitions.db/logs
$ hdfs dfs -ls /user/impala/warehouse/external_partitions.db/logs
Found 1 items
drwxr-xr-x - impala hive 0 2013-08-07 12:24
/user/impala/warehouse/external_partitions.db/logs/year=2013
$ hdfs dfs -ls /user/impala/warehouse/external_partitions.db/logs/year=2013
Found 2 items
drwxr-xr-x - impala hive 0 2013-08-07 12:23
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07
drwxr-xr-x - impala hive 0 2013-08-07 12:24
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=08
$ hdfs dfs -ls /user/impala/warehouse/external_partitions.db/logs/year=2013/month=07
Found 2 items
drwxr-xr-x - impala hive 0 2013-08-07 12:22
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=28
drwxr-xr-x - impala hive 0 2013-08-07 12:23
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=29
$ hdfs dfs -ls
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=28
Found 2 items
drwxr-xr-x - impala hive 0 2013-08-07 12:21
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=28/host=host1
drwxr-xr-x - impala hive 0 2013-08-07 12:22
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=28/host=host2
$ hdfs dfs -ls
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=28/host=host1
Found 1 items
-rw-r--r-- 3 impala hive 12 2013-08-07 12:21
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=28/host=host1/3981726974111751120--8907184999369517436_822630111_data.0
$ hdfs dfs -cat
/user/impala/warehouse/external_partitions.db/logs/year=2013/month=07/day=28/\host=host1/3981726974111751120--8 907184999369517436_822630111_data.0
foo,foo,foo
```

Still in the Linux shell, we use `hdfs dfs -mkdir` to create several data directories outside the HDFS directory tree that Impala controls (`/user/impala/warehouse` in this example, maybe different in your case). Depending on your configuration, you might need to log in as a user with permission to write into this HDFS directory tree; for example, the commands shown here were run while logged in as the `hdfs` user.

```
$ hdfs dfs -mkdir -p /user/impala/data/logs/year=2013/month=07/day=28/host=host1
$ hdfs dfs -mkdir -p /user/impala/data/logs/year=2013/month=07/day=28/host=host2
$ hdfs dfs -mkdir -p /user/impala/data/logs/year=2013/month=07/day=28/host=host1
$ hdfs dfs -mkdir -p /user/impala/data/logs/year=2013/month=07/day=29/host=host1
$ hdfs dfs -mkdir -p /user/impala/data/logs/year=2013/month=08/day=01/host=host1
```

We make a tiny CSV file, with values different than in the `INSERT` statements used earlier, and put a copy within each subdirectory that we will use as an Impala partition.

```
$ cat >dummy_log_data
bar,baz,bletch
$ hdfs dfs -mkdir -p
/user/impala/data/external_partitions/year=2013/month=08/day=01/host=host1
$ hdfs dfs -mkdir -p
/user/impala/data/external_partitions/year=2013/month=07/day=28/host=host1
$ hdfs dfs -mkdir -p
/user/impala/data/external_partitions/year=2013/month=07/day=28/host=host2
$ hdfs dfs -mkdir -p
/user/impala/data/external_partitions/year=2013/month=07/day=29/host=host1
$ hdfs dfs -put dummy_log_data
/user/impala/data/logs/year=2013/month=07/day=28/host=host1
$ hdfs dfs -put dummy_log_data
/user/impala/data/logs/year=2013/month=07/day=28/host=host2
$ hdfs dfs -put dummy_log_data
```

```
/user/impala/data/logs/year=2013/month=07/day=29/host=host1
$ hdfs dfs -put dummy_log_data
/user/impala/data/logs/year=2013/month=08/day=01/host=host1
```

Back in the `impala-shell` interpreter, we move the original Impala-managed table aside, and create a new `external` table with a `LOCATION` clause pointing to the directory under which we have set up all the partition subdirectories and data files.

```
use external_partitions;
alter table logs rename to logs_original;
create external table logs (field1 string, field2 string, field3 string)
partitioned by (year string, month string, day string, host string)
row format delimited fields terminated by ','
location '/user/impala/data/logs';
```

Because partition subdirectories and data files come and go during the data lifecycle, you must identify each of the partitions through an `ALTER TABLE` statement before Impala recognizes the data files they contain.

```
alter table logs add partition (year="2013",month="07",day="28",host="host1")
alter table log_type add partition (year="2013",month="07",day="28",host="host2");
alter table log_type add partition (year="2013",month="07",day="29",host="host1");
alter table log_type add partition (year="2013",month="08",day="01",host="host1");
```

We issue a `REFRESH` statement for the table, always a safe practice when data files have been manually added, removed, or changed. Then the data is ready to be queried. The `SELECT *` statement illustrates that the data from our trivial CSV file was recognized in each of the partitions where we copied it. Although in this case there are only a few rows, we include a `LIMIT` clause on this test query just in case there is more data than we expect.

```
refresh log_type;
select * from log_type limit 100;
+-----+-----+-----+-----+-----+-----+
| field1 | field2 | field3 | year | month | day | host |
+-----+-----+-----+-----+-----+-----+
bar	baz	bletch	2013	07	28	host1
bar	baz	bletch	2013	08	01	host1
bar	baz	bletch	2013	07	29	host1
bar	baz	bletch	2013	07	28	host2
+-----+-----+-----+-----+-----+-----+
```

## Switching Back and Forth Between Impala and Hive

Sometimes, you might find it convenient to switch to the Hive shell to perform some data loading or transformation operation, particularly on file formats such as RCFile, SequenceFile, and Avro that Impala currently can query but not write to.

Whenever you create, drop, or alter a table or other kind of object through Hive, the next time you switch back to the `impala-shell` interpreter, issue a one-time `INVALIDATE METADATA` statement so that Impala recognizes the new or changed object.

Whenever you load, insert, or change data in an existing table through Hive (or even through manual HDFS operations such as the `hdfs` command), the next time you switch back to the `impala-shell` interpreter, issue a one-time `REFRESH table_name` statement so that Impala recognizes the new or changed data.

For examples showing how this process works for the `REFRESH` statement, look at the examples of creating RCFile and SequenceFile tables in Impala, loading data through Hive, and then querying the data through Impala. See [Using the RCFile File Format with Impala Tables](#) on page 422 and [Using the SequenceFile File Format with Impala Tables](#) on page 423 for those examples.

For examples showing how this process works for the `INVALIDATE METADATA` statement, look at the example of creating and loading an Avro table in Hive, and then querying the data through Impala. See [Using the Avro File Format with Impala Tables](#) on page 417 for that example.

- **Note:**

Originally, Impala did not support UDFs, but this feature is available in Impala starting in Impala 1.2. Some `INSERT ... SELECT` transformations that you originally did through Hive can now be done through Impala. See [Impala User-Defined Functions \(UDFs\)](#) on page 308 for details.

Prior to Impala 1.2, the `REFRESH` and `INVALIDATE METADATA` statements needed to be issued on each Impala node to which you connected and issued queries. In Impala 1.2 and higher, when you issue either of those statements on any Impala node, the results are broadcast to all the Impala nodes in the cluster, making it truly a one-step operation after each round of DDL or ETL operations in Hive.

### Cross Joins and Cartesian Products with the CROSS JOIN Operator

Originally, Impala restricted join queries so that they had to include at least one equality comparison between the columns of the tables on each side of the join operator. With the huge tables typically processed by Impala, any miscoded query that produced a full Cartesian product as a result set could consume a huge amount of cluster resources.

In Impala 1.2.2 and higher, this restriction is lifted when you use the `CROSS JOIN` operator in the query. You still cannot remove all `WHERE` clauses from a query like `SELECT * FROM t1 JOIN t2` to produce all combinations of rows from both tables. But you can use the `CROSS JOIN` operator to explicitly request such a Cartesian product. Typically, this operation is applicable for smaller tables, where the result set still fits within the memory of a single Impala node.

The following example sets up data for use in a series of comic books where characters battle each other. At first, we use an equijoin query, which only allows characters from the same time period and the same planet to meet.

```
[localhost:21000] > create table heroes (name string, era string, planet string);
[localhost:21000] > create table villains (name string, era string, planet string);
[localhost:21000] > insert into heroes values
    > ('Tesla','20th century','Earth'),
    > ('Pythagoras','Antiquity','Earth'),
    > ('Zopzar','Far Future','Mars');
Inserted 3 rows in 2.28s
[localhost:21000] > insert into villains values
    > ('Caligula','Antiquity','Earth'),
    > ('John Dillinger','20th century','Earth'),
    > ('Xibulor','Far Future','Venus');
Inserted 3 rows in 1.93s
[localhost:21000] > select concat(heroes.name, ' vs. ',villains.name) as battle
    > from heroes join villains
        > where heroes.era = villains.era and heroes.planet = villains.planet;
+-----+
| battle           |
+-----+
| Tesla vs. John Dillinger |
| Pythagoras vs. Caligula  |
+-----+
Returned 2 row(s) in 0.47s
```

Readers demanded more action, so we added elements of time travel and space travel so that any hero could face any villain. Prior to Impala 1.2.2, this type of query was impossible because all joins had to reference matching values between the two tables:

```
[localhost:21000] > -- Cartesian product not possible in Impala 1.1.
    > select concat(heroes.name, ' vs. ',villains.name) as battle from
heroes join villains;
ERROR: NotImplementedException: Join between 'heroes' and 'villains' requires at least
one conjunctive equality predicate between the two tables
```

With Impala 1.2.2, we rewrite the query slightly to use `CROSS JOIN` rather than `JOIN`, and now the result set includes all combinations:

```
[localhost:21000] > -- Cartesian product available in Impala 1.2.2 with the CROSS JOIN syntax.
      > select concat(heroes.name, ' vs. ',villains.name) as battle from
heroes cross join villains;
+-----+
| battle
+-----+
| Tesla vs. Caligula
| Tesla vs. John Dillinger
| Tesla vs. Xibulor
| Pythagoras vs. Caligula
| Pythagoras vs. John Dillinger
| Pythagoras vs. Xibulor
| Zopzar vs. Caligula
| Zopzar vs. John Dillinger
| Zopzar vs. Xibulor
+-----+
Returned 9 row(s) in 0.33s
```

The full combination of rows from both tables is known as the Cartesian product. This type of result set is often used for creating grid data structures. You can also filter the result set by including `WHERE` clauses that do not explicitly compare columns between the two tables. The following example shows how you might produce a list of combinations of year and quarter for use in a chart, and then a shorter list with only selected quarters.

```
[localhost:21000] > create table x_axis (x int);
[localhost:21000] > create table y_axis (y int);
[localhost:21000] > insert into x_axis values (1),(2),(3),(4);
Inserted 4 rows in 2.14s
[localhost:21000] > insert into y_axis values (2010),(2011),(2012),(2013),(2014);
Inserted 5 rows in 1.32s
[localhost:21000] > select y as year, x as quarter from x_axis cross join y_axis;
+-----+
| year | quarter |
+-----+
| 2010 | 1
| 2011 | 1
| 2012 | 1
| 2013 | 1
| 2014 | 1
| 2010 | 2
| 2011 | 2
| 2012 | 2
| 2013 | 2
| 2014 | 2
| 2010 | 3
| 2011 | 3
| 2012 | 3
| 2013 | 3
| 2014 | 3
| 2010 | 4
| 2011 | 4
| 2012 | 4
| 2013 | 4
| 2014 | 4
+-----+
Returned 20 row(s) in 0.38s
[localhost:21000] > select y as year, x as quarter from x_axis cross join y_axis where
x in (1,3);
+-----+
| year | quarter |
+-----+
| 2010 | 1
| 2011 | 1
| 2012 | 1
| 2013 | 1
| 2014 | 1
| 2010 | 3
| 2011 | 3
+-----+
```

## Impala Tutorials

|                             |   |
|-----------------------------|---|
| 2012                        | 3 |
| 2013                        | 3 |
| 2014                        | 3 |
| Returned 10 row(s) in 0.39s |   |

# Impala Administration

As an administrator, you monitor Impala's use of resources and take action when necessary to keep Impala running smoothly and avoid conflicts with other Hadoop components running on the same cluster. When you detect that an issue has happened or could happen in the future, you reconfigure Impala or other components such as HDFS or even the hardware of the cluster itself to resolve or avoid problems.

## Related tasks:

As an administrator, you can expect to perform installation, upgrade, and configuration tasks for Impala on all machines in a cluster. See [Installing Impala](#) on page 30, [Upgrading Impala](#) on page 43, and [Managing Impala](#) on page 34 for details.

For security tasks typically performed by administrators, see [Impala Security](#) on page 87.

Administrators also decide how to allocate cluster resources so that all Hadoop components can run smoothly together. For Impala, this task primarily involves:

- Deciding how many Impala queries can run concurrently and with how much memory, through the admission control feature. See [Admission Control and Query Queuing](#) on page 69 for details.
- Dividing cluster resources such as memory between Impala and other components, using YARN for overall resource management, and Llama to mediate resource requests from Impala to YARN. See [Integrated Resource Management with YARN](#) on page 78 for details.

## Admission Control and Query Queuing

Admission control is an Impala feature that imposes limits on concurrent SQL queries, to avoid resource usage spikes and out-of-memory conditions on busy CDH clusters. It is a form of "throttling". New queries are accepted and executed until certain conditions are met, such as too many queries or too much total memory used across the cluster. When one of these thresholds is reached, incoming queries wait to begin execution. These queries are queued and are admitted (that is, begin executing) when the resources become available.

In addition to the threshold values for currently executing queries, you can place limits on the maximum number of queries that are queued (waiting) and a limit on the amount of time they might wait before returning with an error. These queue settings let you ensure that queries do not wait indefinitely, so that you can detect and correct "starvation" scenarios.

Enable this feature if your cluster is underutilized at some times and overutilized at others. Overutilization is indicated by performance bottlenecks and queries being cancelled due to out-of-memory conditions, when those same queries are successful and perform well during times with less concurrent load. Admission control works as a safeguard to avoid out-of-memory conditions during heavy concurrent usage.

- **Important:**

- Cloudera strongly recommends you upgrade to CDH 5 or higher to use admission control. In CDH 4, admission control will only work if you *don't* have Hue deployed; unclosed Hue queries will accumulate and exceed the queue size limit. On CDH 4, to use admission control, you must explicitly enable it by specifying `--disable_admission_control=false` in the `impalad` command-line options.
- Use the `COMPUTE STATS` statement for large tables involved in join queries, and follow other steps from [Tuning Impala for Performance](#) on page 354 to tune your queries. Although `COMPUTE STATS` is an important statement to help optimize query performance, it is especially important when admission control is enabled:
  - When queries complete quickly and are tuned for optimal memory usage, there is less chance of performance or capacity problems during times of heavy load.
  - The admission control feature also relies on the statistics produced by the `COMPUTE STATS` statement to generate accurate estimates of memory usage for complex queries. If the estimates are inaccurate due to missing statistics, Impala might hold back queries unnecessarily even though there is sufficient memory to run them, or might allow queries to run that end up exceeding the memory limit and being cancelled.

### Overview of Impala Admission Control

On a busy CDH cluster, you might find there is an optimal number of Impala queries that run concurrently. Because Impala queries are typically I/O-intensive, you might not find any throughput benefit in running more concurrent queries when the I/O capacity is fully utilized. Because Impala by default cancels queries that exceed the specified memory limit, running multiple large-scale queries at once can result in having to re-run some queries that are cancelled.

The admission control feature lets you set a cluster-wide upper limit on the number of concurrent Impala queries and on the memory used by those queries. Any additional queries are queued until the earlier ones finish, rather than being cancelled or running slowly and causing contention. As other queries finish, the queued queries are allowed to proceed.

For details on the internal workings of admission control, see [How Impala Schedules and Enforces Limits on Concurrent Queries](#) on page 71.

### How Impala Admission Control Relates to YARN

The admission control feature is similar in some ways to the YARN resource management framework, and they can be used separately or together. This section describes some similarities and differences, to help you decide when to use one, the other, or both together.

Admission control is a lightweight, decentralized system that is suitable for workloads consisting primarily of Impala queries and other SQL statements. It sets "soft" limits that smooth out Impala memory usage during times of heavy load, rather than taking an all-or-nothing approach that cancels jobs that are too resource-intensive.

Because the admission control system is not aware of other Hadoop workloads such as MapReduce jobs, you might use YARN with static service pools on heterogeneous CDH 5 clusters where resources are shared between Impala and other Hadoop components. Devote a percentage of cluster resources to Impala, allocate another percentage for MapReduce and other batch-style workloads; let admission control handle the concurrency and memory usage for the Impala work within the cluster, and let YARN manage the remainder of work within the cluster.

You could also try out the combination of YARN, Impala, and Llama, where YARN manages all cluster resources and Impala queries request resources from YARN by using the Llama component as an intermediary. YARN is a more centralized, general-purpose service, with somewhat higher latency than admission control due to the requirement to pass requests back and forth through the YARN and Llama components.

The Impala admission control feature uses the same mechanism as the YARN resource manager to map users to pools and authenticate them. Although the YARN resource manager is only available with CDH 5 and higher, internally Impala includes the necessary infrastructure to work consistently on both CDH 4 and CDH 5. You do not need to run the YARN and Llama components for admission control to operate.

In Cloudera Manager, the controls for Impala resource management change slightly depending on whether the Llama role is enabled, which brings Impala under the control of YARN. When you use Impala without the Llama role, you can specify three properties (memory limit, query queue size, and queue timeout) for the admission control feature. When the Llama role is enabled, you can specify query queue size and queue timeout, but the memory limit is enforced by YARN and not settable through resource pools.

For full details about using Impala with YARN, see [Integrated Resource Management with YARN](#) on page 78.

## How Impala Schedules and Enforces Limits on Concurrent Queries

The admission control system is decentralized, embedded in each Impala daemon and communicating through the statestore mechanism. Although the limits you set for memory usage and number of concurrent queries apply cluster-wide, each Impala daemon makes its own decisions about whether to allow each query to run immediately or to queue it for a less-busy time. These decisions are fast, meaning the admission control mechanism is low-overhead, but might be imprecise during times of heavy load. There could be times when the query queue contained more queries than the specified limit, or when the estimated of memory usage for a query is not exact and the overall memory usage exceeds the specified limit. Thus, you typically err on the high side for the size of the queue, because there is not a big penalty for having a large number of queued queries; and you typically err on the low side for the memory limit, to leave some headroom for queries to use more memory than expected, without being cancelled as a result.

At any time, the set of queued queries could include queries submitted through multiple different Impala daemon hosts. All the queries submitted through a particular host will be executed in order, so a `CREATE TABLE` followed by an `INSERT` on the same table would succeed. Queries submitted through different hosts are not guaranteed to be executed in the order they were received. Therefore, if you are using load-balancing or other round-robin scheduling where different statements are submitted through different hosts, set up all table structures ahead of time so that the statements controlled by the queuing system are primarily queries, where order is not significant. Or, if a sequence of statements needs to happen in strict order (such as an `INSERT` followed by a `SELECT`), submit all those statements through a single session, while connected to the same Impala daemon host.

The limit on the number of concurrent queries is a "soft" one. To achieve high throughput, Impala makes quick decisions at the host level about which queued queries to dispatch. Therefore, Impala might slightly exceed the limit from time to time.

To avoid a large backlog of queued requests, you can also set an upper limit on the size of the queue for queries that are delayed. When the number of queued queries exceeds this limit, further queries are cancelled rather than being queued. You can also configure a timeout period, after which queued queries are cancelled, to avoid indefinite waits. If a cluster reaches this state where queries are cancelled due to too many concurrent requests or long waits for query execution to begin, that is a signal for an administrator to take action, either by provisioning more resources, scheduling work on the cluster to smooth out the load, or by doing [Impala performance tuning](#) to enable higher throughput.

## How Admission Control works with Impala Clients (JDBC, ODBC, HiveServer2)

Most aspects of admission control work transparently with client interfaces such as JDBC and ODBC:

- If a SQL statement is put into a queue rather than running immediately, the API call blocks until the statement is dequeued and begins execution. At that point, the client program can request to fetch results, which might also block until results become available.
- If a SQL statement is cancelled because it has been queued for too long or because it exceeded the memory limit during execution, the error is returned to the client program with a descriptive error message.

If you want to submit queries to different resource pools through the `REQUEST_POOL` query option, as described in [REQUEST\\_POOL Query Option](#) on page 352, In Impala 2.0 and higher you can change that query option through

## Impala Administration

a SQL SET statement that you submit from the client application, in the same session. Prior to Impala 2.0, that option was only settable for a session through the `impala-shell` SET command, or cluster-wide through an `impalad` startup option.

Admission control has the following limitations or special behavior when used with JDBC or ODBC applications:

- The `MEM_LIMIT` query option, sometimes useful to work around problems caused by inaccurate memory estimates for complicated queries, is only settable through the `impala-shell` interpreter and cannot be used directly through JDBC or ODBC applications.
- Admission control does not use the other resource-related query options, `RESERVATION_REQUEST_TIMEOUT` or `V_CPU_CORES`. Those query options only apply to the YARN resource management framework.

## Configuring Admission Control

The configuration options for admission control range from the simple (a single resource pool with a single set of options) to the complex (multiple resource pools with different options, each pool handling queries for a different set of users and groups). You can configure the settings through the Cloudera Manager user interface, or on a system without Cloudera Manager by editing configuration files or through startup options to the `impalad` daemon.

### Admission Control Options

The following Impala configuration options let you adjust the settings of the admission control feature. When supplying the options on the command line, prepend the option name with `--`.

#### `default_pool_max_queued`

**Purpose:** Maximum number of requests allowed to be queued before rejecting requests. Because this limit applies cluster-wide, but each Impala node makes independent decisions to run queries immediately or queue them, it is a soft limit; the overall number of queued queries might be slightly higher during times of heavy load. A negative value or 0 indicates requests are always rejected once the maximum concurrent requests are executing. Ignored if `fair_scheduler_config_path` and `llama_site_path` are set.

**Type:** `int64`

**Default:** 200

#### `default_pool_max_requests`

**Purpose:** Maximum number of concurrent outstanding requests allowed to run before incoming requests are queued. Because this limit applies cluster-wide, but each Impala node makes independent decisions to run queries immediately or queue them, it is a soft limit; the overall number of concurrent queries might be slightly higher during times of heavy load. A negative value indicates no limit. Ignored if `fair_scheduler_config_path` and `llama_site_path` are set.

**Type:** `int64`

**Default:** 200

#### `default_pool_mem_limit`

**Purpose:** Maximum amount of memory (across the entire cluster) that all outstanding requests in this pool can use before new requests to this pool are queued. Specified in bytes, megabytes, or gigabytes by a number followed by the suffix `b` (optional), `m`, or `g`, either uppercase or lowercase. You can specify floating-point values for megabytes and gigabytes, to represent fractional numbers such as `1.5`. You can also specify it as a percentage of the physical memory by specifying the suffix `%`. 0 or no setting indicates no limit. Defaults to bytes if no unit is given. Because this limit applies cluster-wide, but each Impala node makes independent decisions to run queries immediately or queue them, it is a soft limit; the overall memory used by concurrent queries might be slightly higher during times of heavy load. Ignored if `fair_scheduler_config_path` and `llama_site_path` are set.

- **Note:** Impala relies on the statistics produced by the `COMPUTE STATS` statement to estimate memory usage for each query. See [COMPUTE STATS Statement](#) on page 170 for guidelines about how and when to use this statement.

**Type:** string

**Default:** "" (empty string, meaning unlimited)

**disable\_admission\_control**

**Purpose:** Turns off the admission control feature entirely, regardless of other configuration option settings.

**Type:** Boolean

**Default:** true

**disable\_pool\_max\_requests**

**Purpose:** Disables all per-pool limits on the maximum number of running requests.

**Type:** Boolean

**Default:** false

**disable\_pool\_mem\_limits**

**Purpose:** Disables all per-pool mem limits.

**Type:** Boolean

**Default:** false

**fair\_scheduler\_allocation\_path**

**Purpose:** Path to the fair scheduler allocation file (`fair-scheduler.xml`).

**Type:** string

**Default:** "" (empty string)

**Usage notes:** Admission control only uses a small subset of the settings that can go in this file, as described below. For details about all the Fair Scheduler configuration settings, see the [Apache wiki](#).

**llama\_site\_path**

**Purpose:** Path to the Llama configuration file (`llama-site.xml`). If set, `fair_scheduler_allocation_path` must also be set.

**Type:** string

**Default:** "" (empty string)

**Usage notes:** Admission control only uses a small subset of the settings that can go in this file, as described below. For details about all the Llama configuration settings, see the [documentation on Github](#).

**queue\_wait\_timeout\_ms**

**Purpose:** Maximum amount of time (in milliseconds) that a request waits to be admitted before timing out.

**Type:** int64

**Default:** 60000

## Configuring Admission Control Using Cloudera Manager

In Cloudera Manager, you can configure pools to manage queued Impala queries, and the options for the limit on number of concurrent queries and how to handle queries that exceed the limit. For details, see [Managing Resources with Cloudera Manager](#).

## Impala Administration

See [Examples of Admission Control Configurations](#) on page 74 for a sample setup for admission control under Cloudera Manager.

### Configuring Admission Control Using the Command Line

If you do not use Cloudera Manager, you use a combination of startup options for the Impala daemon, and optionally editing or manually constructing the configuration files `fair-scheduler.xml` and `llama-site.xml`.

For a straightforward configuration using a single resource pool named `default`, you can specify configuration options on the command line and skip the `fair-scheduler.xml` and `llama-site.xml` configuration files.

For an advanced configuration with multiple resource pools using different settings, set up the `fair-scheduler.xml` and `llama-site.xml` configuration files manually. Provide the paths to each one using the Impala daemon command-line options, `--fair_scheduler_allocation_path` and `--llama_site_path` respectively.

The Impala admission control feature only uses the Fair Scheduler configuration settings to determine how to map users and groups to different resource pools. For example, you might set up different resource pools with separate memory limits, and maximum number of concurrent and queued queries, for different categories of users within your organization. For details about all the Fair Scheduler configuration settings, see the [Apache wiki](#).

The Impala admission control feature only uses a small subset of possible settings from the `llama-site.xml` configuration file:

```
llama.am.throttling.maximum.placed.reservations.queue_name  
llama.am.throttling.maximum.queued.reservations.queue_name
```

For details about all the Llama configuration settings, see [Llama Default Configuration](#).

See [Example Admission Control Configurations Using Configuration Files](#) on page 76 for sample configuration files for admission control using multiple resource pools, without Cloudera Manager.

### Examples of Admission Control Configurations

#### Example Admission Control Configurations Using Cloudera Manager

For full instructions about configuring dynamic resource pools through Cloudera Manager, see [Dynamic Resource Pools](#). The following examples demonstrate some important points related to the Impala admission control feature.

The following figure shows a sample of the Dynamic Resource Pools page in Cloudera Manager, accessed through the **Clusters > Cluster name > Resource Management > Dynamic Resource Pools** menu choice and then the **Configuration** tab. Numbers from all the resource pools are combined into the topmost `root` pool. The `default` pool is for users who are not assigned any other pool by the user-to-pool mapping settings. The `development` and `production` pools show how you can set different limits for different classes of users, for total memory, number of concurrent queries, and number of queries that can be queued.

**Dynamic Resource Pools for Cluster 1**

**Status** **Configuration**

**Resource Pools** **Scheduling Rules** **Placement Rules** **User Limits** **Other Settings**

**Applications** can run in a pool based on the user, the group of the submitting user, as well as specific pools and the default pool. Allocate resources across pools using weights, minimum, and maximum limits. Configuration sets allow switching on different weight and limit settings activated by user-defined schedules.

Pools can be nested, each level of which can support a different scheduler, such as FIFO or fair scheduler. Each pool can be configured to allow only a certain set of users and groups to access the pool.

|             |          | YARN                       |                     |                     |                   | Impala        |                        |                       |               |  |  |
|-------------|----------|----------------------------|---------------------|---------------------|-------------------|---------------|------------------------|-----------------------|---------------|--|--|
| Name        | Weight % | Virtual Cores<br>Min / Max | Memory<br>Min / Max | Max<br>Running Apps | Scheduling Policy | Max<br>Memory | Max<br>Running Queries | Max<br>Queued Queries |               |  |  |
| root        | 1 100.0% | - / -                      | - / -               | -                   | DRF               | -             | -                      | -                     | <b>Edit</b> ▾ |  |  |
| default     | 1 33.3%  | - / -                      | - / -               | -                   | DRF               | 50000MB       | 10                     | 50                    | <b>Edit</b> ▾ |  |  |
| development | 1 33.3%  | - / -                      | - / -               | -                   | DRF               | 200000MB      | 50                     | 100                   | <b>Edit</b> ▾ |  |  |
| production  | 1 33.3%  | - / -                      | - / -               | -                   | DRF               | 1000000MB     | 100                    | 200                   | <b>Edit</b> ▾ |  |  |

Figure 1: Sample Settings for Cloudera Manager Dynamic Resource Pools Page

The following figure shows a sample of the Placement Rules page in Cloudera Manager, accessed through the **Clusters > Cluster name > Resource Management > Dynamic Resource Pools** menu choice and then the **Configuration > Placement Rules** tabs. The settings demonstrate a reasonable configuration of a pool named **default** to service all requests where the specified resource pool does not exist, is not explicitly set, or the user or group is not authorized for the specified pool.

**Dynamic Resource Pools for Cluster 1**

**Status** **Configuration**

**Resource Pools** **Scheduling Rules** **Placement Rules** **User Limits** **Other Settings**

**Applications** can run in a pool based on the user, the group of the submitting user, as well as specific pools and the default pool. Configure how an application will determine in which pool it will run.

**Basic**  
 **Advanced**

Specify the order in which rules are evaluated to determine in which pool an application will run.

If a rule is always satisfied, subsequent rules are not evaluated and appear disabled. If a rule has a condition that is not satisfied, subsequent rules are evaluated. When none of the rules apply, the application is rejected.

|                                                      |   |                                |
|------------------------------------------------------|---|--------------------------------|
| specified pool only if the pool exists. ▾            | - | +                              |
| default pool; create the pool if it doesn't exist. ▾ | - | This rule is always satisfied. |

**Subsequent rules are not evaluated.**

**Save**

Figure 2: Sample Settings for Cloudera Manager Placement Rules Page

### Example Admission Control Configurations Using Configuration Files

For clusters not managed by Cloudera Manager, here are sample `fair-scheduler.xml` and `llama-site.xml` files that define resource pools equivalent to the ones in the preceding Cloudera Manager dialog. These sample files are stripped down: in a real deployment they might contain other settings for use with various aspects of the YARN and Llama components. The settings shown here are the significant ones for the Impala admission control feature.

#### `fair-scheduler.xml`:

Although Impala does not use the `vcores` value, you must still specify it to satisfy YARN requirements for the file contents.

Each `<aclSubmitApps>` tag (other than the one for `root`) contains a comma-separated list of users, then a space, then a comma-separated list of groups; these are the users and groups allowed to submit Impala statements to the corresponding resource pool.

If you leave the `<aclSubmitApps>` element empty for a pool, nobody can submit directly to that pool; child pools can specify their own `<aclSubmitApps>` values to authorize users and groups to submit to those pools.

```
<allocations>
  <queue name="root">
    <aclSubmitApps> </aclSubmitApps>
  <queue name="default">
    <maxResources>50000 mb, 0 vcores</maxResources>
    <aclSubmitApps>*</aclSubmitApps>
  </queue>
  <queue name="development">
    <maxResources>200000 mb, 0 vcores</maxResources>
    <aclSubmitApps>user1,user2 dev,ops,admin</aclSubmitApps>
  </queue>
  <queue name="production">
    <maxResources>1000000 mb, 0 vcores</maxResources>
    <aclSubmitApps> ops,admin</aclSubmitApps>
  </queue>
</queue>
<queuePlacementPolicy>
  <rule name="specified" create="false"/>
  <rule name="default" />
</queuePlacementPolicy>
</allocations>
```

#### `llama-site.xml`:

```
<?xml version="1.0" encoding="UTF-8"?>
<configuration>
  <property>
    <name>llama.am.throttling.maximum.placed.reservations.root.default</name>
    <value>10</value>
  </property>
  <property>
    <name>llama.am.throttling.maximum.queued.reservations.root.default</name>
    <value>50</value>
  </property>
  <property>
    <name>llama.am.throttling.maximum.placed.reservations.root.development</name>
    <value>50</value>
  </property>
  <property>
    <name>llama.am.throttling.maximum.queued.reservations.root.development</name>
    <value>100</value>
  </property>
  <property>
    <name>llama.am.throttling.maximum.placed.reservations.root.production</name>
    <value>100</value>
  </property>
  <property>
    <name>llama.am.throttling.maximum.queued.reservations.root.production</name>
    <value>200</value>
  </property>
</configuration>
```

```
</property>
</configuration>
```

## Guidelines for Using Admission Control

To see how admission control works for particular queries, examine the profile output for the query. This information is available through the `PROFILE` statement in `impala-shell` immediately after running a query in the shell, on the `queries` page of the Impala debug web UI, or in the Impala log file (basic information at log level 1, more detailed information at log level 2). The profile output contains details about the admission decision, such as whether the query was queued or not and which resource pool it was assigned to. It also includes the estimated and actual memory usage for the query, so you can fine-tune the configuration for the memory limits of the resource pools.

Where practical, use Cloudera Manager to configure the admission control parameters. The Cloudera Manager GUI is much simpler than editing the configuration files directly. In Cloudera Manager 4, the admission control settings are not available directly, but you can use the `impalad` safety valve field to configure appropriate startup options.

Remember that the limits imposed by admission control are “soft” limits. Although the limits you specify for number of concurrent queries and amount of memory apply cluster-wide, the decentralized nature of this mechanism means that each Impala node makes its own decisions about whether to allow queries to run immediately or to queue them. These decisions rely on information passed back and forth between nodes by the statestore service. If a sudden surge in requests causes more queries than anticipated to run concurrently, then as a fallback, the overall Impala memory limit and the Linux cgroups mechanism serve as hard limits to prevent overallocation of memory, by cancelling queries if necessary.

If you have trouble getting a query to run because its estimated memory usage is too high, you can override the estimate by setting the `MEM_LIMIT` query option in `impala-shell`, then issuing the query through the shell in the same session. The `MEM_LIMIT` value is treated as the estimated amount of memory, overriding the estimate that Impala would generate based on table and column statistics. This value is used only for making admission control decisions, and is not pre-allocated by the query.

In `impala-shell`, you can also specify which resource pool to direct queries to by setting the `REQUEST_POOL` query option. (This option was named `YARN_POOL` during the CDH 5 beta period.)

The statements affected by the admission control feature are primarily queries, but also include statements that write data such as `INSERT` and `CREATE TABLE AS SELECT`. Most write operations in Impala are not resource-intensive, but inserting into a Parquet table can require substantial memory due to buffering 1 GB of data before writing out each Parquet data block. See [Loading Data into Parquet Tables](#) on page 409 for instructions about inserting data efficiently into Parquet tables.

Although admission control does not scrutinize memory usage for other kinds of DDL statements, if a query is queued due to a limit on concurrent queries or memory usage, subsequent statements in the same session are also queued so that they are processed in the correct order:

```
-- This query could be queued to avoid out-of-memory at times of heavy load.
select * from huge_table join enormous_table using (id);
-- If so, this subsequent statement in the same session is also queued
-- until the previous statement completes.
drop table huge_table;
```

If you set up different resource pools for different users and groups, consider reusing any classifications and hierarchy you developed for use with Sentry security. See [Enabling Sentry Authorization for Impala](#) on page 91 for details.

For details about all the Fair Scheduler configuration settings, see [Fair Scheduler Configuration](#), in particular the tags such as `<queue>` and `<aclSubmitApps>` to map users and groups to particular resource pools (queues).

## Integrated Resource Management with YARN

You can limit the CPU and memory resources used by Impala, to manage and prioritize workloads on clusters that run jobs from many Hadoop components. (Currently, there is no limit or throttling on the I/O for Impala queries.) In CDH 5, Impala can use the underlying Apache Hadoop YARN resource management framework, which allocates the required resources for each Impala query. Impala estimates the resources required by the query on each host of the cluster, and requests the resources from YARN.

Requests from Impala to YARN go through an intermediary service called Llama. When the resource requests are granted, Impala starts the query and places all relevant execution threads into the cgroup containers and sets up the memory limit on each host. If sufficient resources are not available, the Impala query waits until other jobs complete and the resources are freed. During query processing, as the need for additional resources arises, Llama can "expand" already-requested resources, to avoid over-allocating at the start of the query.

After a query is finished, Llama caches the resources (for example, leaving memory allocated) in case they are needed for subsequent Impala queries. This caching mechanism avoids the latency involved in making a whole new set of resource requests for each query. If the resources are needed by YARN for other types of jobs, Llama returns them.

While the delays to wait for resources might make individual queries seem less responsive on a heavily loaded cluster, the resource management feature makes the overall performance of the cluster smoother and more predictable, without sudden spikes in utilization due to memory paging, CPUs pegged at 100%, and so on.

- **Note:**

When using YARN with Impala, Cloudera recommends using the static partitioning technique (through a static service pool) rather than the combination of YARN and Llama. YARN is a central, synchronous scheduler and thus introduces higher latency and variance which is better suited for batch processing than for interactive workloads like Impala (especially with higher concurrency). Currently, YARN allocates memory throughout the query, making it hard to reason about out-of-memory and timeout conditions.

### The Llama Daemon

Llama is a system that mediates resource management between Cloudera Impala and Hadoop YARN. Llama enables Impala to reserve, use, and release resource allocations in a Hadoop cluster. Llama is only required if resource management is enabled in Impala.

By default, YARN allocates resources bit-by-bit as needed by MapReduce jobs. Impala needs all resources available at the same time, so that intermediate results can be exchanged between cluster nodes, and queries do not stall partway through waiting for new resources to be allocated. Llama is the intermediary process that ensures all requested resources are available before each Impala query actually begins.

For management through Cloudera Manager, see [The Impala Llama ApplicationMaster](#).

### Controlling Resource Estimation Behavior

By default, Impala consults the table statistics and column statistics for each table in a query, and uses those figures to construct estimates of needed resources for each query. See [COMPUTE STATS Statement](#) on page 170 for the statement to collect table and column statistics for a table.

To avoid problems with inaccurate or missing statistics, which can lead to inaccurate estimates of resource consumption, Impala allows you to set default estimates for CPU and memory resource consumption. As a query grows to require more resources, Impala will request more resources from Llama (this is called "expanding" a query reservation). When the query is complete, those resources are returned to YARN as normal. To enable this feature, use the command-line option `-rm_always_use_defaults` when starting `impalad`, and optionally `-rm_default_memory=size` and `-rm_default_cpu_cores`. Cloudera recommends always running with `-rm_always_use_defaults` enabled when using resource management, because if the query needs more

resources than the default values, the resource requests are expanded dynamically as the query runs. See [impalad Startup Options for Resource Management](#) on page 80 for details about each option.

## Checking Resource Estimates and Actual Usage

To make resource usage easier to verify, the output of the `EXPLAIN` SQL statement now includes information about estimated memory usage, whether table and column statistics are available for each table, and the number of virtual cores that a query will use. You can get this information through the `EXPLAIN` statement without actually running the query. The extra information requires setting the query option `EXPLAIN_LEVEL=verbose`; see [EXPLAIN Statement](#) on page 199 for details. The same extended information is shown at the start of the output from the `PROFILE` statement in `impala-shell`. The detailed profile information is only available after running the query. You can take appropriate actions (gathering statistics, adjusting query options) if you find that queries fail or run with suboptimal performance when resource management is enabled.

## How Resource Limits Are Enforced

- CPU limits are enforced by the Linux cgroups mechanism. YARN grants resources in the form of containers that correspond to cgroups on the respective machines.
- Memory is enforced by Impala's query memory limits. Once a reservation request has been granted, Impala sets the query memory limit according to the granted amount of memory before executing the query.

## Enabling Resource Management for Impala

To enable resource management for Impala, first you [set up the YARN and Llama services for your CDH cluster](#). Then you [add startup options and customize resource management settings](#) for the Impala services.

### Required CDH Setup for Resource Management with Impala

YARN is the general-purpose service that manages resources for many Hadoop components within a CDH cluster. Llama is a specialized service that acts as an intermediary between Impala and YARN, translating Impala resource requests to YARN and coordinating with Impala so that queries only begin executing when all needed resources have been granted by YARN.

For information about setting up the YARN and Llama services, see the instructions for [Cloudera Manager](#).

## Using Impala with a Llama High Availability Configuration

Impala can take advantage of the Llama high availability (HA) feature, with additional Llama servers that step in if the primary one becomes unavailable. (Only one Llama server at a time services all resource requests.) Before using this feature from Impala, read the background information about Llama HA, its main features, and how to set it up.

### Command-line method for systems without Cloudera Manager:

Setting up the Impala side in a Llama HA configuration involves setting the `impalad` configuration options `-llama_addresses` (mandatory) and optionally `-llama_max_request_attempts`, `-llama_registration_timeout_secs`, and `-llama_registration_wait_secs`. See the next section [impalad Startup Options for Resource Management](#) on page 80 for usage instructions for those options.

The `impalad` daemon on the coordinator host registers with the Llama server for each query, receiving a handle that is used for subsequent resource requests. If a Llama server becomes unavailable, all running Impala queries are cancelled. Subsequent queries register with the next specified Llama server. This registration only happens when a query or similar request causes an `impalad` to request resources through Llama. Therefore, when a Llama server becomes unavailable, that fact might not be reported immediately in the Impala status information such as the `metrics` page in the debug web UI.

**Cloudera Manager method:** See [Llama High Availability](#).

### impalad Startup Options for Resource Management

The following startup options for `impalad` enable resource management and customize its parameters for your cluster configuration:

- `-enable_rm`: Whether to enable resource management or not, either `true` or `false`. The default is `false`. None of the other resource management options have any effect unless `-enable_rm` is turned on.
- `-llama_host`: Hostname or IP address of the Llama service that Impala should connect to. The default is `127.0.0.1`.
- `-llama_port`: Port of the Llama service that Impala should connect to. The default is `15000`.
- `-llama_callback_port`: Port that Impala should start its Llama callback service on. Llama reports when resources are granted or preempted through that service.
- `-cgroup_hierarchy_path`: Path where YARN and Llama will create cgroups for granted resources. Impala assumes that the cgroup for an allocated container is created in the path '`cgroup_hierarchy_path + container_id`'.
- `-rm_always_use_defaults`: If this Boolean option is enabled, Impala ignores computed estimates and always obtains the default memory and CPU allocation from Llama at the start of the query. These default estimates are approximately 2 CPUs and 4 GB of memory, possibly varying slightly depending on cluster size, workload, and so on. Cloudera recommends enabling `-rm_always_use_defaults` whenever resource management is used, and relying on these default values (that is, leaving out the two following options).
- `-rm_default_memory=size`: Optionally sets the default estimate for memory usage for each query. You can use suffixes such as M and G for megabytes and gigabytes, the same as with the [MEM\\_LIMIT](#) query option. Only has an effect when `-rm_always_use_defaults` is also enabled.
- `-rm_default_cpu_cores`: Optionally sets the default estimate for number of virtual CPU cores for each query. Only has an effect when `-rm_always_use_defaults` is also enabled.

The following options fine-tune the interaction of Impala with Llama when Llama high availability (HA) is enabled. The `-llama_addresses` option is only applicable in a Llama HA environment. `-llama_max_request_attempts`, `-llama_registration_timeout_secs`, and `-llama_registration_wait_secs` work whether or not Llama HA is enabled, but are most useful in combination when Llama is set up for high availability.

- `-llama_addresses`: Comma-separated list of `hostname:port` items, specifying all the members of the Llama availability group. Defaults to "`127.0.0.1:15000`".
- `-llama_max_request_attempts`: Maximum number of times a request to reserve, expand, or release resources is retried until the request is cancelled. Attempts are only counted after Impala is registered with Llama. That is, a request survives at most `llama_max_request_attempts - 1` re-registrations. Defaults to 5.
- `-llama_registration_timeout_secs`: Maximum number of seconds that Impala will attempt to register or re-register with Llama. If registration is unsuccessful, Impala cancels the action with an error, which could result in an `impalad` startup failure or a cancelled query. A setting of -1 means try indefinitely. Defaults to 30.
- `-llama_registration_wait_secs`: Number of seconds to wait between attempts during Llama registration. Defaults to 3.

### impala-shell Query Options for Resource Management

Before issuing SQL statements through the `impala-shell` interpreter, you can use the `SET` command to configure the following parameters related to resource management:

- [EXPLAIN\\_LEVEL Query Option](#) on page 343
- [MEM\\_LIMIT Query Option](#) on page 349
- [RESERVATION\\_REQUEST\\_TIMEOUT Query Option \(CDH 5 only\)](#) on page 352
- [V\\_CPU\\_CORES Query Option \(CDH 5 only\)](#) on page 353

### Limitations of Resource Management for Impala

Currently, Impala in CDH 5 has the following limitations for resource management of Impala queries:

- Table statistics are required, and column statistics are highly valuable, for Impala to produce accurate estimates of how much memory to request from YARN. See [Overview of Table Statistics](#) on page 363 and [Overview of Column Statistics](#) on page 364 for instructions on gathering both kinds of statistics, and [EXPLAIN Statement](#) on page 199 for the extended EXPLAIN output where you can check that statistics are available for a specific table and set of columns.
- If the Impala estimate of required memory is lower than is actually required for a query, Impala dynamically expands the amount of requested memory. Queries might still be cancelled if the reservation expansion fails, for example if there are insufficient remaining resources for that pool, or the expansion request takes long enough that it exceeds the query timeout interval, or because of YARN preemption. You can see the actual memory usage after a failed query by issuing a PROFILE command in impala-shell. Specify a larger memory figure with the MEM\_LIMIT query option and re-try the query.

The MEM\_LIMIT query option, and the other resource-related query options, are settable through the ODBC or JDBC interfaces in Impala 2.0 and higher. This is a former limitation that is now lifted.

## Setting Timeout Periods for Daemons, Queries, and Sessions

Depending on how busy your CDH cluster is, you might increase or decrease various timeout values. Increase timeouts if Impala is cancelling operations prematurely, when the system is responding slower than usual but the operations are still successful if given extra time. Decrease timeouts if operations are idle or hanging for long periods, and the idle or hung operations are consuming resources and reducing concurrency.

### Increasing the Statestore Timeout

If you have an extensive Impala schema, for example with hundreds of databases, tens of thousands of tables, and so on, you might encounter timeout errors during startup as the Impala catalog service broadcasts metadata to all the Impala nodes using the statestore service. To avoid such timeout errors on startup, increase the statestore timeout value from its default of 10 seconds. Specify the timeout value using the `-statestore_subscriber_timeout_seconds` option for the statestore service, using the configuration instructions in [Modifying Impala Startup Options](#) on page 48. The symptom of this problem is messages in the impalad log such as:

```
Connection with state-store lost
Trying to re-register with state-store
```

See [Scalability Considerations for the Impala Statestore](#) on page 388 for more details about statestore operation and settings on clusters with a large number of Impala-related objects such as tables and partitions.

### Setting the Idle Query and Idle Session Timeouts for impalad

To keep long-running queries or idle sessions from tying up cluster resources, you can set timeout intervals for both individual queries, and entire sessions.

#### ▪ Note:

The timeout clock for queries and sessions only starts ticking when the query or session is idle. For queries, this means the query has results ready but is waiting for a client to fetch the data. A query can run for an arbitrary time without triggering a timeout, because the query is computing results rather than sitting idle waiting for the results to be fetched. The timeout period is intended to prevent unclosed queries from consuming resources and taking up slots in the admission count of running queries, potentially preventing other queries from starting.

For sessions, this means that no query has been submitted for some period of time.

Specify the following startup options for the impalad daemon:

## Impala Administration

- The `--idle_query_timeout` option specifies the time in seconds after which an idle query is cancelled. This could be a query whose results were all fetched but was never closed, or one whose results were partially fetched and then the client program stopped requesting further results. This condition is most likely to occur in a client program using the JDBC or ODBC interfaces, rather than in the interactive `impala-shell` interpreter. Once the query is cancelled, the client program cannot retrieve any further results.
- The `--idle_session_timeout` option specifies the time in seconds after which an idle session is expired. A session is idle when no activity is occurring for any of the queries in that session, and the session has not started any new queries. Once a session is expired, you cannot issue any new query requests to it. The session remains open, but the only operation you can perform is to close it. The default value of 0 means that sessions never expire.

For instructions on changing `impalad` startup options, see [Modifying Impala Startup Options](#) on page 48.

You can reduce the idle query timeout by using the `QUERY_TIMEOUT_S` query option. Any value specified for the `--idle_query_timeout` startup option serves as an upper limit for the `QUERY_TIMEOUT_S` query option. See [QUERY\\_TIMEOUT\\_S Query Option](#) on page 351 for details.

## Using Impala through a Proxy for High Availability

For most clusters that have multiple users and production availability requirements, you might set up a proxy server to relay requests to and from Impala.

Currently, the Impala statestore mechanism does not include such proxying and load-balancing features. Set up a software package of your choice to perform these functions.

### Overview of Proxy Usage and Load Balancing for Impala

Using a load-balancing proxy server for Impala has the following advantages:

- Applications connect to a single well-known host and port, rather than keeping track of the hosts where the `impalad` daemon is running.
- If any host running the `impalad` daemon becomes unavailable, application connection requests still succeed because you always connect to the proxy server rather than a specific host running the `impalad` daemon.
- The coordinator node for each Impala query potentially requires more memory and CPU cycles than the other nodes that process the query. The proxy server can issue queries using round-robin scheduling, so that each connection uses a different coordinator node. This load-balancing technique lets the Impala nodes share this additional work, rather than concentrating it on a single machine.

The following setup steps are a general outline that apply to any load-balancing proxy software.

1. Download the load-balancing proxy software. It should only need to be installed and configured on a single host.
2. Configure the software (typically by editing a configuration file). Set up a port that the load balancer will listen on to relay Impala requests back and forth.
3. Specify the host and port settings for each Impala node. These are the hosts that the load balancer will choose from when relaying each Impala query. See [Ports Used by Impala](#) on page 451 for when to use port 21000, 21050, or another value depending on what type of connections you are load balancing.
4. Run the load-balancing proxy server, pointing it at the configuration file that you set up.
5. On systems managed by Cloudera Manager, on the page **Impala > Configuration > Impala Daemon Default Group**, specify a value for the **Impala Daemons Load Balancer** field. Specify the address of the load balancer in `host:port` format. This setting lets Cloudera Manager route all appropriate Impala-related operations through the proxy server.
6. For any scripts, jobs, or configuration settings for applications that formerly connected to a specific datanode to run Impala SQL statements, change the connection information (such as the `-i` option in `impala-shell`) to point to the load balancer instead.

## Special Proxy Considerations for Clusters Using Kerberos

In a cluster using Kerberos, applications check host credentials to verify that the host they are connecting to is the same one that is actually processing the request, to prevent man-in-the-middle attacks. To clarify that the load-balancing proxy server is legitimate, perform these extra Kerberos setup steps:

1. This section assumes you are starting with a Kerberos-enabled cluster. See [Enabling Kerberos Authentication for Impala](#) on page 103 for instructions for setting up Impala with Kerberos. See the *CDH Security Guide* for general steps to set up Kerberos: [CDH 5 instructions](#) or [CDH 4 instructions](#).
2. Choose the host you will use for the proxy server. Based on the Kerberos setup procedure, it should already have an entry `impala/proxy_host@realm` in its keytab. If not, go back over the initial Kerberos configuration steps to the keytab on each host running the `impalad` daemon.
3. Copy the keytab file from the proxy host to all other hosts in the cluster that run the `impalad` daemon. (For optimal performance, `impalad` should be running on all DataNodes in the cluster.) Put the keytab file in a secure location on each of these other hosts.
4. On systems not managed by Cloudera Manager, add an entry `impala/actual_hostname@realm` to the keytab on each host running the `impalad` daemon.
5. For each `impalad` node, merge the existing keytab with the proxy's keytab using `ktutil`, producing a new keytab file. For example:

```
$ ktutil
ktutil: read_kt proxy.keytab
ktutil: read_kt impala.keytab
ktutil: write_kt proxy_impala.keytab
ktutil: quit
```

- **Note:** On systems managed by Cloudera Manager 5.1.0 and later, the keytab merging happens automatically. To verify that Cloudera Manager has merged the keytabs, run the command:

```
klist -k keytabfile
```

which lists the credentials for both `principal` and `be_principal` on all nodes.

6. Make sure that the `impala` user has permission to read this merged keytab file.
7. Change some configuration settings for each host in the cluster that participates in the load balancing. Follow the appropriate steps depending on whether you use Cloudera Manager or not:
  - In the `impalad` option definition, or the Cloudera Manager safety valve (Cloudera Manager 4) or advanced configuration snippet (Cloudera Manager 5), add:
 

```
--principal=impala/proxy_host@realm
--be_principal=impala/actual_host@realm
--keytab_file=path_to_merged_keytab
```
  - **Note:** Every host has a different `--be_principal` because the actual host name is different on each host.
8. Restart Impala to make the changes take effect. Follow the appropriate steps depending on whether you use Cloudera Manager or not:
  - On a cluster managed by Cloudera Manager, restart the Impala service.
  - On a cluster not managed by Cloudera Manager, restart the `impalad` daemons on all hosts in the cluster, as well as the `statestored` and `catalogd` daemons.

### Example of Configuring HAProxy Load Balancer for Impala

If you are not already using a load-balancing proxy, you can experiment with [HAProxy](#), a free, open source load balancer. This example shows how you might install and configure that load balancer on a Red Hat Enterprise Linux system.

- Install the load balancer: `yum install haproxy`
- Set up the configuration file: `/etc/haproxy/haproxy.cfg`. See the following section for a sample configuration file.
- Run the load balancer (on a single host, preferably one not running `impalad`):

```
/usr/sbin/haproxy -f /etc/haproxy/haproxy.cfg
```

- In `impala-shell`, JDBC applications, or ODBC applications, connect to the listener port of the proxy host, rather than port 25000 on a host actually running `impalad`. The sample configuration file sets haproxy to listen on port 25003, therefore you would send all requests to `haproxy_host:25003`.

This is the sample `haproxy.cfg` used in this example:

```
global
    # To have these messages end up in /var/log/haproxy.log you will
    # need to:
    #
    # 1) configure syslog to accept network log events. This is done
    #    by adding the '-r' option to the SYSLOGD_OPTIONS in
    #    /etc/sysconfig/syslog
    #
    # 2) configure local2 events to go to the /var/log/haproxy.log
    #    file. A line like the following can be added to
    #    /etc/sysconfig/syslog
    #
    #      local2.*          /var/log/haproxy.log
    #
    log          127.0.0.1 local0
    log          127.0.0.1 local1 notice
    chroot      /var/lib/haproxy
    pidfile    /var/run/haproxy.pid
    maxconn    4000
    user        haproxy
    group       haproxy
    daemon

    # turn on stats unix socket
    #stats socket /var/lib/haproxy/stats

#-----
# common defaults that all the 'listen' and 'backend' sections will
# use if not designated in their block
#
# You might need to adjust timing values to prevent timeouts.
#-----
defaults
    mode          http
    log           global
    option        httplog
    option        dontlognull
    option        http-server-close
    option        forwardfor   except 127.0.0.0/8
    option        redispatch
    retries       3
    maxconn     3000
    contimeout  5000
    clitimeout  50000
    srvtimeout  50000

#
# This sets up the admin page for HA Proxy at port 25002.
```

```

#
listen stats :25002
  balance
  mode http
  stats enable
  stats auth username:password

# This is the setup for Impala. Impala client connect to load_balancer_host:25003.
# HAProxy will balance connections among the list of servers listed below.
# The list of Impalad is listening at port 21000 for beeswax (impala-shell) or original
# ODBC driver.
# For JDBC or ODBC version 2.x driver, use port 21050 instead of 21000.
listen impala :25003
  mode tcp
  option tcplog
  balance leastconn

server symbolic_name_1 impala-host-1.example.com:21000
server symbolic_name_2 impala-host-2.example.com:21000
server symbolic_name_3 impala-host-3.example.com:21000
server symbolic_name_4 impala-host-4.example.com:21000

```

- **Note:** If your JDBC or ODBC application connects to Impala through a load balancer such as `haproxy`, be cautious about reusing the connections. If the load balancer has set up connection timeout values, either check the connection frequently so that it never sits idle longer than the load balancer timeout value, or check the connection validity before using it and create a new one if the connection has been closed.

## Managing Disk Space for Impala Data

Although Impala typically works with many large files in an HDFS storage system with plenty of capacity, there are times when you might perform some file cleanup to reclaim space, or advise developers on techniques to minimize space consumption and file duplication.

- Use compact binary file formats where practical. Numeric and time-based data in particular can be stored in more compact form in binary data files. Depending on the file format, various compression and encoding features can reduce file size even further. You can specify the `STORED AS` clause as part of the `CREATE TABLE` statement, or `ALTER TABLE` with the `SET FILEFORMAT` clause for an existing table or partition within a partitioned table. See [How Impala Works with Hadoop File Formats](#) on page 399 for details about file formats, especially [Using the Parquet File Format with Impala Tables](#) on page 407. See [CREATE TABLE Statement](#) on page 180 and [ALTER TABLE Statement](#) on page 164 for syntax details.
- You manage underlying data files differently depending on whether the corresponding Impala table is defined as an [internal](#) or [external](#) table:
  - Use the `DESCRIBE FORMATTED` statement to check if a particular table is internal (managed by Impala) or external, and to see the physical location of the data files in HDFS. See [DESCRIBE Statement](#) on page 188 for details.
  - For Impala-managed ("internal") tables, use `DROP TABLE` statements to remove data files. See [DROP TABLE Statement](#) on page 197 for details.
  - For tables not managed by Impala ("external" tables), use appropriate HDFS-related commands such as `hadoop fs`, `hdfs dfs`, or `distcp`, to create, move, copy, or delete files within HDFS directories that are accessible by the `impala` user. Issue a `REFRESH table_name` statement after adding or removing any files from the data directory of an external table. See [REFRESH Statement](#) on page 215 for details.
  - Use external tables to reference HDFS data files in their original location. With this technique, you avoid copying the files, and you can map more than one Impala table to the same set of data files. When you drop the Impala table, the data files are left undisturbed. See [External Tables](#) on page 157 for details.

- Use the `LOAD DATA` statement to move HDFS files into the data directory for an Impala table from inside Impala, without the need to specify the HDFS path of the destination directory. This technique works for both internal and external tables. See [LOAD DATA Statement](#) on page 212 for details.
- Make sure that the HDFS trashcan is configured correctly. When you remove files from HDFS, the space might not be reclaimed for use by other files until sometime later, when the trashcan is emptied. See [DROP TABLE Statement](#) on page 197 and the FAQ entry [Why is space not freed up when I issue DROP TABLE?](#) on page 468 for details. See [User Account Requirements](#) on page 24 for permissions needed for the HDFS trashcan to operate correctly.
- Drop all tables in a database before dropping the database itself. See [DROP DATABASE Statement](#) on page 192 for details.
- Clean up temporary files after failed `INSERT` statements. If an `INSERT` statement encounters an error, and you see a directory named `.impala_insert_staging` or `_impala_insert_staging` left behind in the data directory for the table, it might contain temporary data files taking up space in HDFS. You might be able to salvage these data files, for example if they are complete but could not be moved into place due to a permission error. Or, you might delete those files through commands such as `hadoop fs` or `hdfs dfs`, to reclaim space before re-trying the `INSERT`. Issue `DESCRIBE FORMATTED table_name` to see the HDFS path where you can check for temporary files.
- By default, intermediate files used during large sort, join, aggregation, or analytic function operations are stored in the directory `/tmp/impala-scratch`. These files are removed when the operation finishes. (Multiple concurrent queries can perform operations that use the “spill to disk” technique, without any name conflicts for these temporary files.) You can specify a different location by starting the `impalad` daemon with the `--scratch_dirs="path_to_directory"` configuration option or the equivalent configuration option in the Cloudera Manager user interface. You can specify a single directory, or a comma-separated list of directories. The scratch directories must be on the local filesystem, not in HDFS. You might specify different directory paths for different hosts, depending on the capacity and speed of the available storage devices. Impala will not start if it cannot create or read and write files in the “scratch” directory. If there is less than 1 GB free on the filesystem where that directory resides, Impala still runs, but writes a warning message to its log.
- If you use the Amazon Simple Storage Service (S3) as a place to offload data to reduce the volume of local storage, Impala 2.2.0 and higher can query the data directly from S3. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details.

# Impala Security

Impala includes a fine-grained authorization framework for Hadoop, based on the Sentry open source project. Sentry authorization was added in Impala 1.1.0. Together with the Kerberos authentication framework, Sentry takes Hadoop security to a new level needed for the requirements of highly regulated industries such as healthcare, financial services, and government. Impala also includes an auditing capability; Impala generates the audit data, the Cloudera Navigator product consolidates the audit data from all nodes in the cluster, and Cloudera Manager lets you filter, visualize, and produce reports. The auditing feature was added in Impala 1.1.1.

The security features of Cloudera Impala have several objectives. At the most basic level, security prevents accidents or mistakes that could disrupt application processing, delete or corrupt data, or reveal data to unauthorized users. More advanced security features and practices can harden the system against malicious users trying to gain unauthorized access or perform other disallowed operations. The auditing feature provides a way to confirm that no unauthorized access occurred, and detect whether any such attempts were made. This is a critical set of features for production deployments in large organizations that handle important or sensitive data. It sets the stage for multi-tenancy, where multiple applications run concurrently and are prevented from interfering with each other.

The material in this section presumes that you are already familiar with administering secure Linux systems. That is, you should know the general security practices for Linux and Hadoop, and their associated commands and configuration files. For example, you should know how to create Linux users and groups, manage Linux group membership, set Linux and HDFS file permissions and ownership, and designate the default permissions and ownership for new files. You should be familiar with the configuration of the nodes in your Hadoop cluster, and know how to apply configuration changes or run a set of commands across all the nodes.

The security features are divided into these broad categories:

## **authorization**

Which users are allowed to access which resources, and what operations are they allowed to perform? Impala relies on the open source Sentry project for authorization. By default (when authorization is not enabled), Impala does all read and write operations with the privileges of the `impala` user, which is suitable for a development/test environment but not for a secure production environment. When authorization is enabled, Impala uses the OS user ID of the user who runs `impala-shell` or other client program, and associates various privileges with each user. See [Enabling Sentry Authorization for Impala](#) on page 91 for details about setting up and managing authorization.

## **authentication**

How does Impala verify the identity of the user to confirm that they really are allowed to exercise the privileges assigned to that user? Impala relies on the Kerberos subsystem for authentication. See [Enabling Kerberos Authentication for Impala](#) on page 103 for details about setting up and managing authentication.

## **auditing**

What operations were attempted, and did they succeed or not? This feature provides a way to look back and diagnose whether attempts were made to perform unauthorized operations. You use this information to track down suspicious activity, and to see where changes are needed in authorization policies. The audit data produced by this feature is collected by the Cloudera Manager product and then presented in a user-friendly form by the Cloudera Manager product. See [Auditing Impala Operations](#) on page 108 for details about setting up and managing auditing.

# Security Guidelines for Impala

The following are the major steps to harden a cluster running Impala against accidents and mistakes, or malicious attackers trying to access sensitive data:

- Secure the `root` account. The `root` user can tamper with the `impalad` daemon, read and write the data files in HDFS, log into other user accounts, and access other system services that are beyond the control of Impala.

## Impala Security

- Restrict membership in the `sudoers` list (in the `/etc/sudoers` file). The users who can run the `sudo` command can do many of the same things as the `root` user.
- Ensure the Hadoop ownership and permissions for Impala data files are restricted.
- Ensure the Hadoop ownership and permissions for Impala log files are restricted.
- Ensure that the Impala web UI (available by default on port 25000 on each Impala node) is password-protected. See [Impala Web User Interface for Debugging](#) on page 449 for details.
- Create a policy file that specifies which Impala privileges are available to users in particular Hadoop groups (which by default map to Linux OS groups). Create the associated Linux groups using the `groupadd` command if necessary.
- The Impala authorization feature makes use of the HDFS file ownership and permissions mechanism; for background information, see the [CDH HDFS Permissions Guide](#). Set up users and assign them to groups at the OS level, corresponding to the different categories of users with different access levels for various databases, tables, and HDFS locations (URIs). Create the associated Linux users using the `useradd` command if necessary, and add them to the appropriate groups with the `usermod` command.
- Design your databases, tables, and views with database and table structure to allow policy rules to specify simple, consistent rules. For example, if all tables related to an application are inside a single database, you can assign privileges for that database and use the `*` wildcard for the table name. If you are creating views with different privileges than the underlying base tables, you might put the views in a separate database so that you can use the `*` wildcard for the database containing the base tables, while specifying the precise names of the individual views. (For specifying table or database names, you either specify the exact name or `*` to mean all the databases on a server, or all the tables and views in a database.)
- Enable authorization by running the `impalad` daemons with the `-server_name` and `-authorization_policy_file` options on all nodes. (The authorization feature does not apply to the `statestored` daemon, which has no access to schema objects or data files.)
- Set up authentication using Kerberos, to make sure users really are who they say they are.

## Securing Impala Data and Log Files

One aspect of security is to protect files from unauthorized access at the filesystem level. For example, if you store sensitive data in HDFS, you specify permissions on the associated files and directories in HDFS to restrict read and write permissions to the appropriate users and groups.

If you issue queries containing sensitive values in the `WHERE` clause, such as financial account numbers, those values are stored in Impala log files in the Linux filesystem and you must secure those files also. For the locations of Impala log files, see [Using Impala Logging](#) on page 443.

All Impala read and write operations are performed under the filesystem privileges of the `impala` user. The `impala` user must be able to read all directories and data files that you query, and write into all the directories and data files for `INSERT` and `LOAD DATA` statements. At a minimum, make sure the `impala` user is in the `hive` group so that it can access files and directories shared between Impala and Hive. See [User Account Requirements](#) on page 24 for more details.

Setting file permissions is necessary for Impala to function correctly, but is not an effective security practice by itself:

- The way to ensure that only authorized users can submit requests for databases and tables they are allowed to access is to set up Sentry authorization, as explained in [Enabling Sentry Authorization for Impala](#) on page 91. With authorization enabled, the checking of the user ID and group is done by Impala, and unauthorized access is blocked by Impala itself. The actual low-level read and write requests are still done by the `impala` user, so you must have appropriate file and directory permissions for that user ID.

- You must also set up Kerberos authentication, as described in [Enabling Kerberos Authentication for Impala](#) on page 103, so that users can only connect from trusted hosts. With Kerberos enabled, if someone connects a new host to the network and creates user IDs that match your privileged IDs, they will be blocked from connecting to Impala at all from that host.

## Installation Considerations for Impala Security

Impala 1.1 comes set up with all the software and settings needed to enable security when you run the `impalad` daemon with the new security-related options (`-server_name` and `-authorization_policy_file`). You do not need to change any environment variables or install any additional JAR files. In a cluster managed by Cloudera Manager, you do not need to change any settings in Cloudera Manager.

## Securing the Hive Metastore Database

It is important to secure the Hive metastore, so that users cannot access the names or other information about databases and tables through the Hive client or by querying the metastore database. Do this by turning on Hive metastore security, using the instructions in the [CDH 5 Security Guide](#) or the [CDH 4 Security Guide](#) for securing different Hive components:

- Secure the Hive Metastore.
- In addition, allow access to the metastore only from the HiveServer2 server, and then disable local access to the HiveServer2 server.

## Securing the Impala Web User Interface

The instructions in this section presume you are familiar with the [.htpasswd mechanism](#) commonly used to password-protect pages on web servers.

Password-protect the Impala web UI that listens on port 25000 by default. Set up a `.htpasswd` file in the `$IMPALA_HOME` directory, or start both the `impalad` and `statestored` daemons with the `--webserver_password_file` option to specify a different location (including the filename).

This file should only be readable by the Impala process and machine administrators, because it contains (hashed) versions of passwords. The username / password pairs are not derived from Unix usernames, Kerberos users, or any other system. The `domain` field in the password file must match the domain supplied to Impala by the new command-line option `--webserver_authentication_domain`. The default is `mydomain.com`.

Impala also supports using HTTPS for secure web traffic. To do so, set `--webserver_certificate_file` to refer to a valid `.pem` SSL certificate file. Impala will automatically start using HTTPS once the SSL certificate has been read and validated. A `.pem` file is basically a private key, followed by a signed SSL certificate; make sure to concatenate both parts when constructing the `.pem` file.

If Impala cannot find or parse the `.pem` file, it prints an error message and quits.

**Note:**

If the private key is encrypted using a passphrase, Impala will ask for that passphrase on startup, which is not useful for a large cluster. In that case, remove the passphrase and make the `.pem` file readable only by Impala and administrators.

When you turn on SSL for the Impala web UI, the associated URLs change from `http://` prefixes to `https://`. Adjust any bookmarks or application code that refers to those URLs.

# Configuring SSL for Impala

Impala supports SSL network encryption, between Impala and client programs, and between the Impala-related daemons running on different nodes in the cluster. This feature is important when you also use other features such as Kerberos authentication or Sentry authorization, where credentials are being transmitted back and forth.

- **Important:**
  - You can use either Cloudera Manager or the following command-line instructions to complete this configuration.
  - This information applies specifically to CDH 5.4.x. If you use an earlier version of CDH, see the documentation for that version located at [Cloudera Documentation](#).

## Using Cloudera Manager

To configure Impala to listen for Beeswax and HiveServer2 requests on SSL-secured ports:

1. Open the Cloudera Manager Admin Console and navigate to the **Impala** service.
2. Click the **Configuration** tab.
3. Select **Scope > Impala (Service-Wide)**.
4. Select **Category > Security**.
5. Edit the following SSL properties:

**Table 2: Impala SSL Properties**

Property	Description
<b>Enable TLS/SSL for Impala Client Services</b>	Encrypt communication between clients (like ODBC, JDBC, and the Impala shell) and the Impala daemon using Transport Layer Security (TLS) (formerly known as Secure Socket Layer (SSL)).
<b>SSL/TLS Certificate for Clients</b>	Local path to the X509 certificate that will identify the Impala daemon to clients during SSL/TLS connections. This file must be in PEM format.
<b>SSL/TLS Private Key for Clients</b>	Local path to the private key that matches the certificate specified in the Certificate for Clients. This file must be in PEM format.

6. Click **Save Changes** to commit the changes.
7. Restart the Impala service.

For information on configuring SSL communication with the `impala-shell` interpreter, see [Configuring SSL Communication for the Impala Shell](#) on page 91.

## Using the Command Line

To enable SSL for Impala network communication, add both of the following flags to the `impalad` startup options:

- `--ssl_server_certificate`: the full path to the server certificate, on the local filesystem.
- `--ssl_private_key`: the full path to the server private key, on the local filesystem.

If either of these flags are set, both must be set. In that case, Impala starts listening for Beeswax and HiveServer2 requests on SSL-secured ports only. (The port numbers stay the same; see [Ports Used by Impala](#) on page 451 for details.)

## Configuring SSL Communication for the Impala Shell

Typically, a client program has corresponding configuration properties in Cloudera Manager to verify that it is connecting to the right server. For example, with SSL enabled for Impala, you use the following options when starting the `impala-shell` interpreter:

- `--ssl`: enables SSL for `impala-shell`.
- `--ca_cert`: the local pathname pointing to the third-party CA certificate, or to a copy of the server certificate for self-signed server certificates.

If `--ca_cert` is not set, `impala-shell` enables SSL, but does not validate the server certificate. This is useful for connecting to a known-good Impala that is only running over SSL, when a copy of the certificate is not available (such as when debugging customer installations).

## Enabling Sentry Authorization for Impala

Authorization determines which users are allowed to access which resources, and what operations they are allowed to perform. In Impala 1.1 and higher, you use the Sentry open source project for authorization. Sentry adds a fine-grained authorization framework for Hadoop. By default (when authorization is not enabled), Impala does all read and write operations with the privileges of the `impala` user, which is suitable for a development/test environment but not for a secure production environment. When authorization is enabled, Impala uses the OS user ID of the user who runs `impala-shell` or other client program, and associates various privileges with each user.

- **Note:** Sentry is typically used in conjunction with Kerberos authentication, which defines which hosts are allowed to connect to each server. Using the combination of Sentry and Kerberos prevents malicious users from being able to connect by creating a named account on an untrusted machine. See [Enabling Kerberos Authentication for Impala](#) on page 103 for details about Kerberos authentication.

## The Sentry Privilege Model

Privileges can be granted on different objects in the schema. Any privilege that can be granted is associated with a level in the object hierarchy. If a privilege is granted on a container object in the hierarchy, the child object automatically inherits it. This is the same privilege model as Hive and other database systems such as MySQL.

The object hierarchy covers Server, URI, Database, and Table. (The Table privileges apply to views as well; anywhere you specify a table name, you can specify a view name instead.) Currently, you cannot assign privileges at the partition or column level. The way you implement column-level or partition-level privileges is to create a view that queries just the relevant columns or partitions, and assign privileges to the view rather than the underlying table or tables.

A restricted set of privileges determines what you can do with each object:

### **SELECT** privilege

Lets you read data from a table or view, for example with the `SELECT` statement, the `INSERT...SELECT` syntax, or `CREATE TABLE...LIKE`. Also required to issue the `DESCRIBE` statement or the `EXPLAIN` statement for a query against a particular table. Only objects for which a user has this privilege are shown in the output for `SHOW DATABASES` and `SHOW TABLES` statements. The `REFRESH` statement and `INVALIDATE METADATA` statements only access metadata for tables for which the user has this privilege.

### **INSERT** privilege

Lets you write data to a table. Applies to the `INSERT` and `LOAD DATA` statements.

### **ALL** privilege

Lets you create or modify the object. Required to run DDL statements such as `CREATE TABLE`, `ALTER TABLE`, or `DROP TABLE` for a table, `CREATE DATABASE` or `DROP DATABASE` for a database, or `CREATE VIEW`, `ALTER VIEW`, or `DROP VIEW` for a view. Also required for the `URI` of the "location" parameter for the `CREATE EXTERNAL TABLE` and `LOAD DATA` statements.

## Impala Security

Privileges can be specified for a table or view before that object actually exists. If you do not have sufficient privilege to perform an operation, the error message does not disclose if the object exists or not.

Originally, privileges were encoded in a policy file, stored in HDFS. This mode of operation is still an option, but the emphasis of privilege management is moving towards being SQL-based. Although currently Impala does not have GRANT or REVOKE statements, Impala can make use of privileges assigned through GRANT and REVOKE statements done through Hive. The mode of operation with GRANT and REVOKE statements instead of the policy file requires that a special Sentry service be enabled; this service stores, retrieves, and manipulates privilege information stored inside the metastore database.

### Starting the impalad Daemon with Sentry Authorization Enabled

To run the `impalad` daemon with authorization enabled, you add one or more options to the `IMPALA_SERVER_ARGS` declaration in the `/etc/default/impala` configuration file:

- The `-server_name` option turns on Sentry authorization for Impala. The authorization rules refer to a symbolic server name, and you specify the name to use as the argument to the `-server_name` option.
- If you specify just `-server_name`, Impala uses the Sentry service for authorization, relying on the results of GRANT and REVOKE statements issued through Hive. (This mode of operation is available in Impala 1.4.0 and higher.) Prior to Impala 1.4.0, or if you want to continue storing privilege rules in the policy file, also specify the `-authorization_policy_file` option as in the following item.
- Specifying the `-authorization_policy_file` option in addition to `-server_name` makes Impala read privilege information from a policy file, rather than from the metastore database. The argument to the `-authorization_policy_file` option specifies the HDFS path to the policy file that defines the privileges on different schema objects.

For example, you might adapt your `/etc/default/impala` configuration to contain lines like the following. To use the Sentry service rather than the policy file:

```
IMPALA_SERVER_ARGS=" \
-server_name=server1 \
..."
```

Or to use the policy file, as in releases prior to Impala 1.4:

```
IMPALA_SERVER_ARGS=" \
-authorization_policy_file=/user/hive/warehouse/auth-policy.ini \
-server_name=server1 \
..."
```

The preceding examples set up a symbolic name of `server1` to refer to the current instance of Impala. This symbolic name is used in the following ways:

- In an environment managed by Cloudera Manager, the server name is specified through **Impala (Service-Wide) > Category > Advanced > Sentry Service** and **Hive > Service-Wide > Advanced > Sentry Service**. The values must be the same for both, so that Impala and Hive can share the privilege rules. Restart the Impala and Hive services after setting or changing this value.
- In an environment not managed by Cloudera Manager, you specify this value for the `sentry.hive.server` property in the `sentry-site.xml` configuration file for Hive, as well as in the `-server_name` option for `impalad`.

If the `impalad` daemon is not already running, start it as described in [Starting Impala](#) on page 47. If it is already running, restart it with the command `sudo /etc/init.d/impala-server restart`. Run the appropriate commands on all the nodes where `impalad` normally runs.

- If you use the mode of operation using the policy file, the rules in the `[roles]` section of the policy file refer to this same `server1` name. For example, the following rule sets up a role `report_generator` that lets

users with that role query any table in a database named `reporting_db` on a node where the `impalad` daemon was started up with the `-server_name=server1` option:

```
[roles]
report_generator = server=server1->db=reporting_db->table=*->action=SELECT
```

When `impalad` is started with one or both of the `-server_name=server1` and `-authorization_policy_file` options, Impala authorization is enabled. If Impala detects any errors or inconsistencies in the authorization settings or the policy file, the daemon refuses to start.

## Using Impala with the Sentry Service (CDH 5.1 or higher only)

When you use the Sentry service rather than the policy file, you set up privileges through `GRANT` and `REVOKE` statement in either Impala or Hive, then both components use those same privileges automatically. (Impala added the `GRANT` and `REVOKE` statements in Impala 2.0.0 / CDH 5.2.0.)

Hive already had `GRANT` and `REVOKE` statements prior to CDH 5.1, but those statements were not production-ready. CDH 5.1 is the first release where those statements use the Sentry framework and are considered GA level. If you used the Hive `GRANT` and `REVOKE` statements prior to CDH 5.1, you must set up these privileges with the CDH 5.1 versions of `GRANT` and `REVOKE` to take advantage of Sentry authorization.

For information about using the updated Hive `GRANT` and `REVOKE` statements, see [Sentry service](#) topic in the *CDH 5 Security Guide*.

## Using Impala with the Sentry Policy File

The policy file is a file that you put in a designated location in HDFS, and is read during the startup of the `impalad` daemon when you specify both the `-server_name` and `-authorization_policy_file` startup options. It controls which objects (databases, tables, and HDFS directory paths) can be accessed by the user who connects to `impalad`, and what operations that user can perform on the objects.

- **Note:** This mode of operation works on both CDH 4 and CDH 5, but in CDH 5 the emphasis is shifting towards managing privileges through SQL statements, as described in [Using Impala with the Sentry Service \(CDH 5.1 or higher only\)](#) on page 93. If you are still using policy files, plan to migrate to the new approach some time in the future.

The location of the policy file is listed in the `auth-site.xml` configuration file. To minimize overhead, the security information from this file is cached by each `impalad` daemon and refreshed automatically, with a default interval of 5 minutes. After making a substantial change to security policies, restart all Impala daemons to pick up the changes immediately.

### Policy File Location and Format

The policy file uses the familiar `.ini` format, divided into the major sections `[groups]` and `[roles]`. There is also an optional `[databases]` section, which allows you to specify a specific policy file for a particular database, as explained in [Using Multiple Policy Files for Different Databases](#) on page 98. Another optional section, `[users]`, allows you to override the OS-level mapping of users to groups; that is an advanced technique primarily for testing and debugging, and is beyond the scope of this document.

In the `[groups]` section, you define various categories of users and select which roles are associated with each category. The group and user names correspond to Linux groups and users on the server where the `impalad` daemon runs.

The group and user names in the `[groups]` section correspond to Linux groups and users on the server where the `impalad` daemon runs. When you access Impala through the `impalad` interpreter, for purposes of authorization, the user is the logged-in Linux user and the groups are the Linux groups that user is a member of. When you access Impala through the ODBC or JDBC interfaces, the user and password specified through the connection string are used as login credentials for the Linux server, and authorization is based on that user name and the associated Linux group membership.

## Impala Security

In the [roles] section, you specify a set of roles. For each role, you specify precisely the set of privileges available. That is, which objects users with that role can access, and what operations they can perform on those objects. This is the lowest-level category of security information; the other sections in the policy file map the privileges to higher-level divisions of groups and users. In the [groups] section, you specify which roles are associated with which groups. The group and user names correspond to Linux groups and users on the server where the impalad daemon runs. The privileges are specified using patterns like:

```
server=server_name->db=database_name->table=table_name->action=SELECT  
server=server_name->db=database_name->table=table_name->action=CREATE  
server=server_name->db=database_name->table=table_name->action=ALL
```

For the `server_name` value, substitute the same symbolic name you specify with the `impalad -server_name` option. You can use \* wildcard characters at each level of the privilege specification to allow access to all such objects. For example:

```
server=impala-host.example.com->db=default->table=t1->action=SELECT  
server=impala-host.example.com->db=*->table=*->action=CREATE  
server=impala-host.example.com->db=*->table=audit_log->action=SELECT  
server=impala-host.example.com->db=default->table=t1->action=*
```

When authorization is enabled, Impala uses the policy file as a *whitelist*, representing every privilege available to any user on any object. That is, only operations specified for the appropriate combination of object, role, group, and user are allowed; all other operations are not allowed. If a group or role is defined multiple times in the policy file, the last definition takes precedence.

To understand the notion of whitelisting, set up a minimal policy file that does not provide any privileges for any object. When you connect to an Impala node where this policy file is in effect, you get no results for `SHOW DATABASES`, and an error when you issue any `SHOW TABLES`, `USE database_name`, `DESCRIBE table_name`, `SELECT`, and/or other statements that expect to access databases or tables, even if the corresponding databases and tables exist.

The contents of the policy file are cached, to avoid a performance penalty for each query. The policy file is re-checked by each impalad node every 5 minutes. When you make a non-time-sensitive change such as adding new privileges or new users, you can let the change take effect automatically a few minutes later. If you remove or reduce privileges, and want the change to take effect immediately, restart the impalad daemon on all nodes, again specifying the `-server_name` and `-authorization_policy_file` options so that the rules from the updated policy file are applied.

## Examples of Policy File Rules for Security Scenarios

The following examples show rules that might go in the policy file to deal with various authorization-related scenarios. For illustration purposes, this section shows several very small policy files with only a few rules each. In your environment, typically you would define many roles to cover all the scenarios involving your own databases, tables, and applications, and a smaller number of groups, whose members are given the privileges from one or more roles.

### A User with No Privileges

If a user has no privileges at all, that user cannot access any schema objects in the system. The error messages do not disclose the names or existence of objects that the user is not authorized to read.

This is the experience you want a user to have if they somehow log into a system where they are not an authorized Impala user. In a real deployment with a filled-in policy file, a user might have no privileges because they are not a member of any of the relevant groups mentioned in the policy file.

### Examples of Privileges for Administrative Users

When an administrative user has broad access to tables or databases, the associated rules in the [roles] section typically use wildcards and/or inheritance. For example, in the following sample policy file, `db=*` refers to all databases and `db=*->table=*` refers to all tables in all databases.

Omitting the rightmost portion of a rule means that the privileges apply to all the objects that could be specified there. For example, in the following sample policy file, the `all_databases` role has all privileges for all tables in all databases, while the `one_database` role has all privileges for all tables in one specific database. The `all_databases` role does not grant privileges on URIs, so a group with that role could not issue a `CREATE TABLE` statement with a `LOCATION` clause. The `entire_server` role has all privileges on both databases and URIs within the server.

```
[groups]
supergroup = all_databases

[roles]
read_all_tables = server=server1->db=*->table=*->action=SELECT
all_tables = server=server1->db=*->table=*
all_databases = server=server1->db=*
one_database = server=server1->db=test_db
entire_server = server=server1
```

### A User with Privileges for Specific Databases and Tables

If a user has privileges for specific tables in specific databases, the user can access those things but nothing else. They can see the tables and their parent databases in the output of `SHOW TABLES` and `SHOW DATABASES`, `USE` the appropriate databases, and perform the relevant actions (`SELECT` and/or `INSERT`) based on the table privileges. To actually create a table requires the `ALL` privilege at the database level, so you might define separate roles for the user that sets up a schema and other users or applications that perform day-to-day operations on the tables.

The following sample policy file shows some of the syntax that is appropriate as the policy file grows, such as the `#` comment syntax, `\` continuation syntax, and comma separation for roles assigned to groups or privileges assigned to roles.

```
[groups]
cloudera = training_sysadmin, instructor
visitor = student

[roles]
training_sysadmin = server=server1->db=training, \
server=server1->db=instructor_private, \
server=server1->db=lesson_development
instructor = server=server1->db=training->table=*->action=*, \
server=server1->db=instructor_private->table=*->action=*, \
server=server1->db=lesson_development->table=lesson*
# This particular course is all about queries, so the students can SELECT but not INSERT
# or CREATE/DROP.
student = server=server1->db=training->table=lesson_*->action=SELECT
```

### Privileges for Working with External Data Files

When data is being inserted through the `LOAD DATA` statement, or is referenced from an HDFS location outside the normal Impala database directories, the user also needs appropriate permissions on the URIs corresponding to those HDFS locations.

In this sample policy file:

- The `external_table` role lets us insert into and query the Impala table, `external_table.sample`.
- The `staging_dir` role lets us specify the HDFS path `/user/cloudera/external_data` with the `LOAD DATA` statement. Remember, when Impala queries or loads data files, it operates on all the files in that directory, not just a single file, so any Impala `LOCATION` parameters refer to a directory rather than an individual file.
- We included the IP address and port of the Hadoop name node in the HDFS URI of the `staging_dir` rule. We found those details in `/etc/hadoop/conf/core-site.xml`, under the `fs.default.name` element. That is what we use in any roles that specify URIs (that is, the locations of directories in HDFS).
- We start this example after the table `external_table.sample` is already created. In the policy file for the example, we have already taken away the `external_table_admin` role from the `cloudera` group, and replaced it with the lesser-privileged `external_table` role.

## Impala Security

- We assign privileges to a subdirectory underneath /user/cloudera in HDFS, because such privileges also apply to any subdirectories underneath. If we had assigned privileges to the parent directory /user/cloudera, it would be too likely to mess up other files by specifying a wrong location by mistake.
- The cloudera under the [groups] section refers to the cloudera group. (In the demo VM used for this example, there is a cloudera user that is a member of a cloudera group.)

Policy file:

```
[groups]
cloudera = external_table, staging_dir

[roles]
external_table_admin = server=server1->db=external_table
external_table = server=server1->db=external_table->table=sample->action=*
staging_dir =
server=server1->uri=hdfs://127.0.0.1:8020/user/cloudera/external_data->action=*
```

impala-shell session:

```
[localhost:21000] > use external_table;
Query: use external_table
[localhost:21000] > show tables;
Query: show tables
Query finished, fetching results ...
+-----+
| name |
+-----+
| sample |
+-----+
Returned 1 row(s) in 0.02s

[localhost:21000] > select * from sample;
Query: select * from sample
Query finished, fetching results ...
+---+
| x |
+---+
| 1 |
| 5 |
| 150 |
+---+
Returned 3 row(s) in 1.04s

[localhost:21000] > load data inpath '/user/cloudera/external_data' into table sample;
Query: load data inpath '/user/cloudera/external_data' into table sample
Query finished, fetching results ...
+-----+
| summary |
+-----+
| Loaded 1 file(s). Total files in destination location: 2 |
+-----+
Returned 1 row(s) in 0.26s
[localhost:21000] > select * from sample;
Query: select * from sample
Query finished, fetching results ...
+-----+
| x |
+-----+
| 2 |
| 4 |
| 6 |
| 8 |
| 64738 |
| 49152 |
| 1 |
| 5 |
| 150 |
+-----+
Returned 9 row(s) in 0.22s

[localhost:21000] > load data inpath '/user/cloudera/unauthorized_data' into table
```

```
sample;
Query: load data inpath '/user/cloudera/unauthorized_data' into table sample
ERROR: AuthorizationException: User 'cloudera' does not have privileges to access:
hdfs://127.0.0.1:8020/user/cloudera/unauthorized_data
```

### Controlling Access at the Column Level through Views

If a user has `SELECT` privilege for a view, they can query the view, even if they do not have any privileges on the underlying table. To see the details about the underlying table through `EXPLAIN` or `DESCRIBE FORMATTED` statements on the view, the user must also have `SELECT` privilege for the underlying table.

- **Important:**

The types of data that are considered sensitive and confidential differ depending on the jurisdiction the type of industry, or both. For fine-grained access controls, set up appropriate privileges based on all applicable laws and regulations.

Be careful using the `ALTER VIEW` statement to point an existing view at a different base table or a new set of columns that includes sensitive or restricted data. Make sure that any users who have `SELECT` privilege on the view do not gain access to any additional information they are not authorized to see.

The following example shows how a system administrator could set up a table containing some columns with sensitive information, then create a view that only exposes the non-confidential columns.

```
[localhost:21000] > create table sensitive_info
> (
>   name string,
>   address string,
>   credit_card string,
>   taxpayer_id string
> );
[localhost:21000] > create view name_address_view as select name, address from
sensitive_info;
```

Then the following policy file specifies read-only privilege for that view, without authorizing access to the underlying table:

```
[groups]
cloudera = view_only_privs

[roles]
view_only_privs = server=server1->db=reports->table=name_address_view->action=SELECT
```

Thus, a user with the `view_only_privs` role could access through Impala queries the basic information but not the sensitive information, even if both kinds of information were part of the same data file.

You might define other views to allow users from different groups to query different sets of columns.

### Separating Administrator Responsibility from Read and Write Privileges

Remember that to create a database requires full privilege on that database, while day-to-day operations on tables within that database can be performed with lower levels of privilege on specific table. Thus, you might set up separate roles for each database or application: an administrative one that could create or drop the database, and a user-level one that can access only the relevant tables.

For example, this policy file divides responsibilities between users in 3 different groups:

- Members of the `supergroup` group have the `training_sysadmin` role and so can set up a database named `training`.
- Members of the `cloudera` group have the `instructor` role and so can create, insert into, and query any tables in the `training` database, but cannot create or drop the database itself.

## Impala Security

- Members of the visitor group have the student role and so can query those tables in the training database.

```
[groups]
supergroup = training_sysadmin
cloudera = instructor
visitor = student

[roles]
training_sysadmin = server=server1->db=training
instructor = server=server1->db=training->table=*->action=*
student = server=server1->db=training->table=*->action=SELECT
```

## Using Multiple Policy Files for Different Databases

For an Impala cluster with many databases being accessed by many users and applications, it might be cumbersome to update the security policy file for each privilege change or each new database, table, or view. You can allow security to be managed separately for individual databases, by setting up a separate policy file for each database:

- Add the optional [databases] section to the main policy file.
- Add entries in the [databases] section for each database that has its own policy file.
- For each listed database, specify the HDFS path of the appropriate policy file.

For example:

```
[databases]
# Defines the location of the per-DB policy files for the 'customers' and 'sales'
databases.
customers = hdfs://ha-nn-uri/etc/access/customers.ini
sales = hdfs://ha-nn-uri/etc/access/sales.ini
```

To enable URIs in per-DB policy files, add the following string in the Cloudera Manager field **Impala Service Environment Advanced Configuration Snippet (Safety Valve)**:

```
JAVA_TOOL_OPTIONS="-Dsentry.allow.uri.db.policyfile=true"
```

- **Important:** Enabling URIs in per-DB policy files introduces a security risk by allowing the owner of the db-level policy file to grant himself/herself load privileges to anything the `impala` user has read permissions for in HDFS (including data in other databases controlled by different db-level policy files).

## Setting Up Schema Objects for a Secure Impala Deployment

Remember that in your role definitions, you specify privileges at the level of individual databases and tables, or all databases or all tables within a database. To simplify the structure of these rules, plan ahead of time how to name your schema objects so that data with different authorization requirements is divided into separate databases.

If you are adding security on top of an existing Impala deployment, remember that you can rename tables or even move them between databases using the `ALTER TABLE` statement. In Impala, creating new databases is a relatively inexpensive operation, basically just creating a new directory in HDFS.

You can also plan the security scheme and set up the policy file before the actual schema objects named in the policy file exist. Because the authorization capability is based on whitelisting, a user can only create a new database or table if the required privilege is already in the policy file: either by listing the exact name of the object being created, or a \* wildcard to match all the applicable objects within the appropriate container.

## Privilege Model and Object Hierarchy

Privileges can be granted on different objects in the schema. Any privilege that can be granted is associated with a level in the object hierarchy. If a privilege is granted on a container object in the hierarchy, the child object automatically inherits it. This is the same privilege model as Hive and other database systems such as MySQL.

The kinds of objects in the schema hierarchy are:

```
Server
URI
Database
Table
```

The server name is specified by the `-server_name` option when `impalad` starts. Specify the same name for all `impalad` nodes in the cluster.

URIs represent the HDFS paths you specify as part of statements such as `CREATE EXTERNAL TABLE` and `LOAD DATA`. Typically, you specify what look like UNIX paths, but these locations can also be prefixed with `hdfs://` to make clear that they are really URIs. To set privileges for a URI, specify the name of a directory, and the privilege applies to all the files in that directory and any directories underneath it.

There are not separate privileges for individual table partitions or columns. To specify read privileges at this level, you create a view that queries specific columns and/or partitions from a base table, and give `SELECT` privilege on the view but not the underlying table. See [Views](#) on page 158 for details about views in Impala.

URIs must start with either `hdfs://` or `file://`. If a URI starts with anything else, it will cause an exception and the policy file will be invalid. When defining URIs for HDFS, you must also specify the NameNode. For example:

```
data_read = server=server1->uri=file:///path/to/dir, \
server=server1->uri=hdfs://namenode:port/path/to/dir
```

- **Warning:**

Because the NameNode host and port must be specified, Cloudera strongly recommends you use High Availability (HA). This ensures that the URI will remain constant even if the namenode changes.

```
data_read = server=server1->uri=file:///path/to/dir, \
server=server1->uri=hdfs://ha-nn-uri/path/to/dir
```

**Table 3: Valid privilege types and objects they apply to**

Privilege	Object
INSERT	DB, TABLE
SELECT	DB, TABLE
ALL	SERVER, TABLE, DB, URI

- **Note:**

Although this document refers to the `ALL` privilege, currently if you use the policy file mode, you do not use the actual keyword `ALL` in the policy file. When you code role entries in the policy file:

- To specify the `ALL` privilege for a server, use a role like `server=server_name`.
- To specify the `ALL` privilege for a database, use a role like `server=server_name->db=database_name`.
- To specify the `ALL` privilege for a table, use a role like `server=server_name->db=database_name->table=table_name->action=*`.

## Impala Security

Operation	Scope	Privileges	URI	Others
EXPLAIN	TABLE	SELECT		
LOAD DATA	TABLE	INSERT	URI	
CREATE DATABASE	SERVER	ALL		
DROP DATABASE	DATABASE	ALL		
CREATE TABLE	DATABASE	ALL		
DROP TABLE	TABLE	ALL		
DESCRIBE TABLE	TABLE	SELECT/INSERT		
ALTER TABLE .. ADD COLUMNS	TABLE	ALL		
ALTER TABLE .. REPLACE COLUMNS	TABLE	ALL		
ALTER TABLE .. CHANGE column	TABLE	ALL		
ALTER TABLE .. RENAME	TABLE	ALL		
ALTER TABLE .. SET TBLPROPERTIES	TABLE	ALL		
ALTER TABLE .. SET FILEFORMAT	TABLE	ALL		
ALTER TABLE .. SET LOCATION	TABLE	ALL	URI	
ALTER TABLE .. ADD PARTITION	TABLE	ALL		
ALTER TABLE .. ADD PARTITION location	TABLE	ALL	URI	
ALTER TABLE .. DROP PARTITION	TABLE	ALL		
ALTER TABLE .. PARTITION SET FILEFORMAT	TABLE	ALL		
ALTER TABLE .. SET SERDEPROPERTIES	TABLE	ALL		
CREATE VIEW	DATABASE; SELECT on TABLE	ALL		SELECT on TABLE
DROP VIEW	VIEW/TABLE	ALL		
ALTER VIEW	You need ALL privilege on the named view and the parent database, plus SELECT privilege for any tables or views referenced by the view query. Once the view is created or altered by a	ALL, SELECT		

Operation	Scope	Privileges	URI	Others
	high-privileged system administrator, it can be queried by a lower-privileged user who does not have full query privileges for the base tables. (This is how you implement column-level security.)			
ALTER TABLE .. SET LOCATION	TABLE	ALL	URI	
CREATE EXTERNAL TABLE	Database (ALL), URI (SELECT)	ALL, SELECT		
SELECT	TABLE	SELECT		
USE <dbName>	Any			
CREATE FUNCTION	SERVER	ALL		
DROP FUNCTION	SERVER	ALL		
REFRESH <table name>	TABLE	SELECT/INSERT		
INVALIDATE METADATA	SERVER	ALL		
INVALIDATE METADATA <table name>	TABLE	SELECT/INSERT		
COMPUTE STATS	TABLE	ALL		
SHOW TABLE STATS, SHOW PARTITIONS	TABLE	SELECT/INSERT		
SHOW COLUMN STATS	TABLE	SELECT/INSERT		
SHOW FUNCTIONS	DATABASE	SELECT		
SHOW TABLES		No special privileges needed to issue the statement, but only shows objects you are authorized for		
SHOW DATABASES, SHOW SCHEMAS		No special privileges needed to issue the statement, but only shows objects you are authorized for		

## Debugging Failed Sentry Authorization Requests

Sentry logs all facts that lead up to authorization decisions at the debug level. If you do not understand why Sentry is denying access, the best way to debug is to temporarily turn on debug logging:

## Impala Security

- In Cloudera Manager, add `log4j.logger.org.apache.sentry=DEBUG` to the logging settings for your service through the corresponding **Logging Safety Valve** field for the Impala, Hive Server 2, or Solr Server services.
- On systems not managed by Cloudera Manager, add `log4j.logger.org.apache.sentry=DEBUG` to the `log4j.properties` file on each host in the cluster, in the appropriate configuration directory for each service.

Specifically, look for exceptions and messages such as:

```
FilePermission server..., RequestPermission server...., result [true|false]
```

which indicate each evaluation Sentry makes. The `FilePermission` is from the policy file, while `RequestPermission` is the privilege required for the query. A `RequestPermission` will iterate over all appropriate `FilePermission` settings until a match is found. If no matching privilege is found, Sentry returns `false` indicating "Access Denied".

## Configuring Per-User Access for Hue

When users connect to Impala directly through the `impala-shell` interpreter, the Impala authorization feature determines what actions they can take and what data they can see. When users submit Impala queries through a separate application, such as Hue, typically all requests are treated as coming from the same user. In Impala 1.2 and higher, authorization is extended by a new feature that allows applications to pass along credentials for the users that connect to them (known as "delegation"), and issue Impala queries with the privileges for those users. Currently, the delegation feature is available only for Impala queries submitted through the Hue interface; for example, Impala cannot issue queries using the privileges of the HDFS user.

A new startup option for `impalad`, `--authorized_proxy_user_config`, is available in Impala 1.2 and higher. When you specify this option, users whose names you specify (such as `hue`) can delegate the execution of a query to another user. The query runs with the privileges of the delegated user, not the original user such as `hue`. The name of the delegated user is passed using the HiveServer2 configuration property `impala.doas.user`.

You can specify a list of users that the application user can delegate to, or `*` to allow a superuser to delegate to any other user. For example:

```
impalad --authorized_proxy_user_config 'hue=user1,user2;admin=*' ...
```

- **Note:** Make sure to use single quotes or escape characters to ensure that any `*` characters do not undergo wildcard expansion when specified in command-line arguments.

See [Modifying Impala Startup Options](#) on page 48 for details about adding or changing `impalad` startup options. See [this Cloudera blog post](#) for background information about the delegation capability in HiveServer2.

## Managing Sentry for Impala through Cloudera Manager

To enable the Sentry service for Impala and Hive, set **Hive/Impala > Service-Wide > Sentry Service** parameter to the Sentry service. Then restart Impala and Hive. Simply adding Sentry service as a dependency and restarting enables Impala and Hive to use the Sentry service.

To set the server name to use when granting server level privileges, set the **Hive > Service-Wide > Advanced > Server Name for Sentry Authorization** parameter. When using Sentry with the Hive Metastore, you can specify the list of users that are allowed to bypass Sentry Authorization in Hive Metastore using **Hive > Service-Wide > Security > Bypass Sentry Authorization Users**. These are usually service users that already ensure all activity has been authorized.

- **Note:** The **Hive/Impala > Service-Wide > Policy File Based Sentry** tab contains parameters only relevant to configuring Sentry using policy files. In particular, make sure that **Enable Sentry Authorization using Policy Files** parameter is unchecked when using the Sentry service. Cloudera Manager throws a validation error if you attempt to configure the Sentry service and policy file at the same time.

## The DEFAULT Database in a Secure Deployment

Because of the extra emphasis on granular access controls in a secure deployment, you should move any important or sensitive information out of the `DEFAULT` database into a named database whose privileges are specified in the policy file. Sometimes you might need to give privileges on the `DEFAULT` database for administrative reasons; for example, as a place you can reliably specify with a `USE` statement when preparing to drop a database.

## Impala Authentication

Authentication is the mechanism to ensure that only specified hosts and users can connect to Impala. It also verifies that when clients connect to Impala, they are connected to a legitimate server. This feature prevents spoofing such as *impersonation* (setting up a phony client system with the same account and group names as a legitimate user) and *man-in-the-middle attacks* (intercepting application requests before they reach Impala and eavesdropping on sensitive information in the requests or the results).

Impala supports authentication using either Kerberos or LDAP.

Impala currently does not support application data wire encryption.

- **Note:** Regardless of the authentication mechanism used, Impala always creates HDFS directories and data files owned by the same user (typically `impala`). To implement user-level access to different databases, tables, columns, partitions, and so on, use the Sentry authorization feature, as explained in [Enabling Sentry Authorization for Impala](#) on page 91.

Once you are finished setting up authentication, move on to authorization, which involves specifying what databases, tables, HDFS directories, and so on can be accessed by particular users when they connect through Impala. See [Enabling Sentry Authorization for Impala](#) on page 91 for details.

## Enabling Kerberos Authentication for Impala

Impala supports Kerberos authentication. For more information on enabling Kerberos authentication, see the topic on Configuring Hadoop Security in the [CDH 4 Security Guide](#) or the [CDH 5 Security Guide](#).

Impala currently does not support application data wire encryption.

When using Impala in a managed environment, Cloudera Manager automatically completes Kerberos configuration. In an unmanaged environment, create a Kerberos principal for each host running `impalad` or `statestored`. Cloudera recommends using a consistent format, such as `impala/_HOST@Your-Realm`, but you can use any three-part Kerberos server principal.

In Impala 2.0 and later, `user()` returns the full Kerberos principal string, such as `user@example.com`, in a Kerberized environment.

- **Note:** Regardless of the authentication mechanism used, Impala always creates HDFS directories and data files owned by the same user (typically `impala`). To implement user-level access to different databases, tables, columns, partitions, and so on, use the Sentry authorization feature, as explained in [Enabling Sentry Authorization for Impala](#) on page 91.

An alternative form of authentication you can use is LDAP, described in [Enabling LDAP Authentication for Impala](#) on page 106.

### Requirements for Using Impala with Kerberos

On version 5 of Red Hat Enterprise Linux and comparable distributions, some additional setup is needed for the `impala-shell` interpreter to connect to a Kerberos-enabled Impala cluster:

```
sudo yum install python-devel openssl-devel python-pip  
sudo pip-python install ssl
```

- **Important:**

If you plan to use Impala in your cluster, you must configure your KDC to allow tickets to be renewed, and you must configure `krb5.conf` to request renewable tickets. Typically, you can do this by adding the `max_renewable_life` setting to your realm in `kdc.conf`, and by adding the `renew_lifetime` parameter to the `libdefaults` section of `krb5.conf`. For more information about renewable tickets, see the [Kerberos documentation](#).

Currently, you cannot use the resource management feature in CDH 5 on a cluster that has Kerberos authentication enabled.

Start all `impalad` and `statestored` daemons with the `--principal` and `--keytab-file` flags set to the principal and full path name of the `keytab` file containing the credentials for the principal.

Impala supports the Cloudera ODBC driver and the Kerberos interface provided. To use Kerberos through the ODBC driver, the host type must be set depending on the level of the ODBD driver:

- `SecImpala` for the ODBC 1.0 driver.
- `SecBeeswax` for the ODBC 1.2 driver.
- Blank for the ODBC 2.0 driver or higher, when connecting to a secure cluster.
- `HS2NoSasl` for the ODBC 2.0 driver or higher, when connecting to a non-secure cluster.

To enable Kerberos in the Impala shell, start the `impala-shell` command using the `-k` flag.

To enable Impala to work with Kerberos security on your Hadoop cluster, make sure you perform the installation and configuration steps in [Authentication in the CDH 5 Security Guide](#) or the topic on Configuring Hadoop Security in the [CDH4 Security Guide](#). Also note that when Kerberos security is enabled in Impala, a web browser that supports Kerberos HTTP SPNEGO is required to access the Impala web console (for example, Firefox, Internet Explorer, or Chrome).

If the NameNode, Secondary NameNode, DataNode, JobTracker, TaskTrackers, ResourceManager, NodeManagers, HttpFS, Oozie, Impala, or Impala statestore services are configured to use Kerberos HTTP SPNEGO authentication, and two or more of these services are running on the same host, then all of the running services must use the same HTTP principal and keytab file used for their HTTP endpoints.

### Configuring Impala to Support Kerberos Security

Enabling Kerberos authentication for Impala involves steps that can be summarized as follows:

- Creating service principals for Impala and the HTTP service. Principal names take the form:  
`serviceName/fully.qualified.domain.name@KERBEROS.REALM`
- Creating, merging, and distributing key tab files for these principals.
- Editing `/etc/default/impala` (in cluster not managed by Cloudera Manager), or editing the **Security** settings in the Cloudera Manager interface, to accommodate Kerberos authentication.

#### Enabling Kerberos for Impala

1. Create an Impala service principal, specifying the name of the OS user that the Impala daemons run under, the fully qualified domain name of each node running `impalad`, and the realm name. For example:

```
$ kadmin  
kadmin: addprinc -requires_pcreauth -randkey  
impala/impala_host.example.com@TEST.EXAMPLE.COM
```

2. Create an HTTP service principal. For example:

```
kadmin: addprinc -randkey HTTP/impala_host.example.com@TEST.EXAMPLE.COM
```

**Note:** The HTTP component of the service principal must be uppercase as shown in the preceding example.

3. Create keytab files with both principals. For example:

```
kadmin: xst -k impala.keytab impala/impala_host.example.com
kadmin: xst -k http.keytab HTTP/impala_host.example.com
kadmin: quit
```

4. Use ktutil to read the contents of the two keytab files and then write those contents to a new file. For example:

```
$ ktutil
ktutil: rkt impala.keytab
ktutil: rkt http.keytab
ktutil: wkt impala-http.keytab
ktutil: quit
```

5. (Optional) Test that credentials in the merged keytab file are valid, and that the "renew until" date is in the future. For example:

```
$ klist -e -k -t impala-http.keytab
```

6. Copy the impala-http.keytab file to the Impala configuration directory. Change the permissions to be only read for the file owner and change the file owner to the `impala` user. By default, the Impala user and group are both named `impala`. For example:

```
$ cp impala-http.keytab /etc/impala/conf
$ cd /etc/impala/conf
$ chmod 400 impala-http.keytab
$ chown impala:impala impala-http.keytab
```

7. Add Kerberos options to the Impala defaults file, `/etc/default/impala`. Add the options for both the `impalad` and `statestored` daemons, using the `IMPALA_SERVER_ARGS` and `IMPALA_STATE_STORE_ARGS` variables. For example, you might add:

```
-kerberos_reinit_interval=60
-principal=impala_1/impala_host.example.com@TEST.EXAMPLE.COM
-keytab_file=/var/run/cloudera-scm-agent/process/3212-impala-IMPALAD/impala.keytab
```

For more information on changing the Impala defaults specified in `/etc/default/impala`, see [Modifying Impala Startup Options](#).

**Note:** Restart `impalad` and `statestored` for these configuration changes to take effect.

## Enabling Kerberos for Impala with a Proxy Server

A common configuration for Impala with High Availability is to use a proxy server to submit requests to the actual `impalad` daemons on different hosts in the cluster. This configuration avoids connection problems in case of machine failure, because the proxy server can route new requests through one of the remaining hosts in the cluster. This configuration also helps with load balancing, because the additional overhead of being the "coordinator node" for each query is spread across multiple hosts.

## Impala Security

Although you can set up a proxy server with or without Kerberos authentication, typically users set up a secure Kerberized configuration. For information about setting up a proxy server for Impala, including Kerberos-specific steps, see [Using Impala through a Proxy for High Availability](#) on page 82.

### Using a Web Browser to Access a URL Protected by Kerberos HTTP SPNEGO

Your web browser must support Kerberos HTTP SPNEGO. For example, Chrome, Firefox, or Internet Explorer.

**To configure Firefox to access a URL protected by Kerberos HTTP SPNEGO:**

1. Open the advanced settings Firefox configuration page by loading the `about:config` page.
2. Use the **Filter** text box to find `network.negotiate-auth.trusted-uris`.
3. Double-click the `network.negotiate-auth.trusted-uris` preference and enter the hostname or the domain of the web server that is protected by Kerberos HTTP SPNEGO. Separate multiple domains and hostnames with a comma.
4. Click **OK**.

### Enabling LDAP Authentication for Impala

Authentication is the process of allowing only specified named users to access the server (in this case, the Impala server). This feature is crucial for any production deployment, to prevent misuse, tampering, or excessive load on the server. Impala users LDAP for authentication, verifying the credentials of each user who connects through `impala-shell`, Hue, a Business Intelligence tool, JDBC or ODBC application, and so on.

- **Note:** Regardless of the authentication mechanism used, Impala always creates HDFS directories and data files owned by the same user (typically `impala`). To implement user-level access to different databases, tables, columns, partitions, and so on, use the Sentry authorization feature, as explained in [Enabling Sentry Authorization for Impala](#) on page 91.

An alternative form of authentication you can use is Kerberos, described in [Enabling Kerberos Authentication for Impala](#) on page 103.

### Requirements for Using Impala with LDAP

Authentication against LDAP servers is available in Impala 1.2.2 and higher. Impala 1.4.0 adds support for secure LDAP authentication through SSL and TLS.

The Impala LDAP support lets you use Impala with systems such as Active Directory that use LDAP behind the scenes.

### Client-Server Considerations for LDAP

Only client->Impala connections can be authenticated by LDAP.

You must Kerberos authentication mechanism for connections between internal Impala components, such as between the `impalad`, `statesstored`, and `catalogd` daemons. See [Enabling Kerberos Authentication for Impala](#) on page 103 for how to set up Kerberos for Impala.

### Server-Side LDAP Setup

These requirements apply on the server side when configuring and starting Impala:

To enable LDAP authentication, set the following startup options for `impalad`:

- `--enable_ldap_auth` enables LDAP-based authentication between the client and Impala.
- `--ldap_uri` sets the URI of the LDAP server to use. Typically, the URI is prefixed with `ldap://`. In Impala 1.4.0 and higher, you can specify secure SSL-based LDAP transport by using the prefix `ldaps://`. The URI can optionally specify the port, for example: `ldap://ldap_server.cloudera.com:389` or `ldaps://ldap_server.cloudera.com:636`. (389 and 636 are the default ports for non-SSL and SSL LDAP connections, respectively.)

- For `ldaps://` connections secured by SSL, `--ldap_ca_certificate="/path/to/certificate/pem"` specifies the location of the certificate in standard `.PEM` format. Store this certificate on the local filesystem, in a location that only the `impala` user and other trusted users can read.

## Support for Custom Bind Strings

When Impala connects to LDAP it issues a bind call to the LDAP server to authenticate as the connected user. Impala clients, including the Impala shell, provide the short name of the user to Impala. This is necessary so that Impala can use Sentry for role-based access, which uses short names.

However, LDAP servers often require more complex, structured usernames for authentication. Impala supports three ways of transforming the short name (for example, '`henry`') to a more complicated string. If necessary, specify one of the following configuration options when starting the `impalad` daemon on each data node:

- `--ldap_domain`: Replaces the username with a string `username@ldap_domain`.
- `--ldap_baseDN`: Replaces the username with a "distinguished name" (DN) of the form: `uid=userid,ldap_baseDN`. (This is equivalent to a Hive option).
- `--ldap_bind_pattern`: This is the most general option, and replaces the username with the string `ldap_bind_pattern` where all instances of the string `#UID` are replaced with `userid`. For example, an `ldap_bind_pattern` of "`user=#UID,OU=foo,CN=bar`" with a username of `henry` will construct a bind name of "`user=henry,OU=foo,CN=bar`".

For clusters not managed by Cloudera Manager, specify the option on the `impalad` command line. For clusters managed by Cloudera Manager 5.4.0 and higher, search for the configuration field names `ldap_domain`, `ldap_basedn`, or `ldap_bind_pattern`, fill in and save the appropriate field values, and restart the Impala service. Prior to Cloudera Manager 5.4.0, these values were filled in using the **Impala Daemon Command Line Argument Advanced Configuration Snippet (Safety Valve)** field.

These options are mutually exclusive; Impala does not start if more than one of these options is specified.

## Secure LDAP Connections

To avoid sending credentials over the wire in cleartext, you must configure a secure connection between both the client and Impala, and between Impala and the LDAP server. The secure connection could use SSL or TLS.

### Secure LDAP connections through SSL:

For SSL-enabled LDAP connections, specify a prefix of `ldaps://` instead of `ldap://`. Also, the default port for SSL-enabled LDAP connections is 636 instead of 389.

### Secure LDAP connections through TLS:

[TLS](#), the successor to the SSL protocol, is supported by most modern LDAP servers. Unlike SSL connections, TLS connections can be made on the same server port as non-TLS connections. To secure all connections using TLS, specify the following flags as startup options to the `impalad` daemon:

- `--ldap_tls` tells Impala to start a TLS connection to the LDAP server, and to fail authentication if it cannot be done.
- `--ldap_ca_certificate="/path/to/certificate/pem"` specifies the location of the certificate in standard `.PEM` format. Store this certificate on the local filesystem, in a location that only the `impala` user and other trusted users can read.

## LDAP Authentication for `impala-shell` Interpreter

To connect to Impala using LDAP authentication, you specify command-line options to the `impala-shell` command interpreter and enter the password when prompted:

- `-l` enables LDAP authentication.
- `-u` sets the user. Per Active Directory, the user is the short user name, not the full LDAP distinguished name. If your LDAP settings include a search base, use the `--ldap_bind_pattern` on the `impalad` daemon to translate the short user name from `impala-shell` automatically to the fully qualified name.
- `impala-shell` automatically prompts for the password.

## Impala Security

For the full list of available `impala-shell` options, see [impala-shell Configuration Options](#) on page 332.

**LDAP authentication for JDBC applications:** See [Configuring Impala to Work with JDBC](#) on page 39 for the format to use with the JDBC connection string for servers using LDAP authentication.

### LDAP Restrictions for Impala

The LDAP support is preliminary. It currently has only been tested against Active Directory.

## Using Multiple Authentication Methods with Impala

Impala 2.0 and later automatically handles both Kerberos and LDAP authentication. Each `impalad` daemon can accept both Kerberos and LDAP requests through the same port. No special actions need to be taken if some users authenticate through Kerberos and some through LDAP.

Prior to Impala 2.0, you had to configure each `impalad` to listen on a specific port depending on the kind of authentication, then configure your network load balancer to forward each kind of request to a data node that was set up with the appropriate authentication type. Once the initial request was made using either Kerberos or LDAP authentication, Impala automatically handled the process of coordinating the work across multiple nodes and transmitting intermediate results back to the coordinator node.

## Auditing Impala Operations

To monitor how Impala data is being used within your organization, ensure that your Impala authorization and authentication policies are effective, and detect attempts at intrusion or unauthorized access to Impala data, you can use the auditing feature in Impala 1.2.1 and higher:

- Enable auditing by including the option `-audit_event_log_dir=directory_path` in your `impalad` startup options for a cluster not managed by Cloudera Manager, or [configuring Impala Daemon logging in Cloudera Manager](#). The log directory must be a local directory on the server, not an HDFS directory.
- Decide how many queries will be represented in each log files. By default, Impala starts a new log file every 5000 queries. To specify a different number, include the option `-max_audit_event_log_file_size=number_of_queries` in the `impalad` startup options.
- Configure the Cloudera Navigator product to collect and consolidate the audit logs from all the hosts in the cluster.
- Use Cloudera Navigator or Cloudera Manager to filter, visualize, and produce reports based on the audit data. (The Impala auditing feature works with Cloudera Manager 4.7 to 5.1 and Cloudera Navigator 2.1 and higher.) Check the audit data to ensure that all activity is authorized and/or detect attempts at unauthorized access.

## Durability and Performance Considerations for Impala Auditing

The auditing feature only imposes performance overhead while auditing is enabled.

Because any Impala host can process a query, enable auditing on all hosts where the `impalad` daemon runs. Each host stores its own log files, in a directory in the local filesystem. The log data is periodically flushed to disk (through an `fsync()` system call) to avoid loss of audit data in case of a crash.

The runtime overhead of auditing applies to whichever host serves as the coordinator for the query, that is, the host you connect to when you issue the query. This might be the same host for all queries, or different applications or users might connect to and issue queries through different hosts.

To avoid excessive I/O overhead on busy coordinator hosts, Impala syncs the audit log data (using the `fsync()` system call) periodically rather than after every query. Currently, the `fsync()` calls are issued at a fixed interval, every 5 seconds.

By default, Impala avoids losing any audit log data in the case of an error during a logging operation (such as a disk full error), by immediately shutting down `impalad` on the host where the auditing problem occurred. You can override this setting by specifying the option `-abort_on_failed_audit_event=false` in the `impalad` startup options.

## Format of the Audit Log Files

The audit log files represent the query information in JSON format, one query per line. Typically, rather than looking at the log files themselves, you use the Cloudera Navigator product to consolidate the log data from all Impala hosts and filter and visualize the results in useful ways. (If you do examine the raw log data, you might run the files through a JSON pretty-printer first.)

All the information about schema objects accessed by the query is encoded in a single nested record on the same line. For example, the audit log for an `INSERT ... SELECT` statement records that a select operation occurs on the source table and an insert operation occurs on the destination table. The audit log for a query against a view records the base table accessed by the view, or multiple base tables in the case of a view that includes a join query. Every Impala operation that corresponds to a SQL statement is recorded in the audit logs, whether the operation succeeds or fails. Impala records more information for a successful operation than for a failed one, because an unauthorized query is stopped immediately, before all the query planning is completed.

The information logged for each query includes:

- Client session state:
  - Session ID
  - User name
  - Network address of the client connection
- SQL statement details:
  - Query ID
  - Statement Type - DML, DDL, and so on
  - SQL statement text
  - Execution start time, in local time
  - Execution Status - Details on any errors that were encountered
  - Target Catalog Objects:
    - Object Type - Table, View, or Database
    - Fully qualified object name
    - Privilege - How the object is being used (`SELECT`, `INSERT`, `CREATE`, and so on)

## Which Operations Are Audited

The kinds of SQL queries represented in the audit log are:

- Queries that are prevented due to lack of authorization.
- Queries that Impala can analyze and parse to determine that they are authorized. The audit data is recorded immediately after Impala finishes its analysis, before the query is actually executed.

The audit log does not contain entries for queries that could not be parsed and analyzed. For example, a query that fails due to a syntax error is not recorded in the audit log. The audit log also does not contain queries that fail due to a reference to a table that does not exist, if you would be authorized to access the table if it did exist.

Certain statements in the `impala-shell` interpreter, such as `CONNECT`, `SUMMARY`, `PROFILE`, `SET`, and `QUIT`, do not correspond to actual SQL queries, and these statements are not reflected in the audit log.

## Reviewing the Audit Logs

You typically do not review the audit logs in raw form. The Cloudera Manager Agent periodically transfers the log information into a back-end database where it can be examined in consolidated form. See the [Cloudera Navigator documentation](#) for details .

# Viewing Lineage Information for Impala Data

**Lineage** is a feature in the Cloudera Navigator data management component that helps you track where data originated, and how data propagates through the system through SQL statements such as `SELECT`, `INSERT`, and `CREATE TABLE AS SELECT`. Impala is covered by the Cloudera Navigator lineage features in CDH 5.4.0 and higher.

This type of tracking is important in high-security configurations, especially in highly regulated industries such as healthcare, pharmaceuticals, financial services and intelligence. For such kinds of sensitive data, it is important to know all the places in the system that contain that data or other data derived from it; to verify who has accessed that data; and to be able to doublecheck that the data used to make a decision was processed correctly and not tampered with.

You interact with this feature through **lineage diagrams** showing relationships between tables and columns. For instructions about interpreting lineage diagrams, see [Lineage Diagrams](#).

## Column Lineage

**Column lineage** tracks information in fine detail, at the level of particular columns rather than entire tables.

For example, if you have a table with information derived from web logs, you might copy that data into other tables as part of the ETL process. The ETL operations might involve transformations through expressions and function calls, and rearranging the columns into more or fewer tables (**normalizing** or **denormalizing** the data). Then for reporting, you might issue queries against multiple tables and views. In this example, column lineage helps you determine that data that entered the system as `RAW_LOGS.FIELD1` was then turned into `WEBSITE_REPORTS.IP_ADDRESS` through an `INSERT ... SELECT` statement. Or, conversely, you could start with a reporting query against a view, and trace the origin of the data in a field such as `TOP_10_VISITORS.USER_ID` back to the underlying table and even further back to the point where the data was first loaded into Impala.

When you have tables where you need to track or control access to sensitive information at the column level, see [Enabling Sentry Authorization for Impala](#) on page 91 for how to implement column-level security. You set up authorization using the Sentry framework, create views that refer to specific sets of columns, and then assign authorization privileges to those views rather than the underlying tables.

## Lineage Data for Impala

The lineage feature is enabled by default. When lineage logging is enabled, the serialized column lineage graph is computed for each query and stored in a specialized log file in JSON format.

Impala records queries in the lineage log if they complete successfully, or fail due to authorization errors. For write operations such as `INSERT` and `CREATE TABLE AS SELECT`, the statement is recorded in the lineage log only if it successfully completes. Therefore, the lineage feature tracks data that was accessed by successful queries, or that was attempted to be accessed by unsuccessful queries that were blocked due to authorization failure. These kinds of queries represent data that really was accessed, or where the attempted access could represent malicious activity.

Impala does not record in the lineage log queries that fail due to syntax errors or that fail or are cancelled before they reach the stage of requesting rows from the result set.

To enable or disable this feature on a system not managed by Cloudera Manager, set or remove the `-lineage_event_log_dir` configuration option for the `impalad` daemon. For information about turning the lineage feature on and off through Cloudera Manager, see [Impala Lineage Properties](#).

# Impala SQL Language Reference

Cloudera Impala uses SQL as its query language. To protect user investment in skills development and query design, Impala provides a high degree of compatibility with the Hive Query Language (HiveQL):

- Because Impala uses the same metadata store as Hive to record information about table structure and properties, Impala can access tables defined through the native Impala `CREATE TABLE` command, or tables created using the Hive data definition language (DDL).
- Impala supports data manipulation (DML) statements similar to the DML component of HiveQL.
- Impala provides many [built-in functions](#) with the same names and parameter types as their HiveQL equivalents.

Impala supports most of the same [statements and clauses](#) as HiveQL, including, but not limited to `JOIN`, `AGGREGATE`, `DISTINCT`, `UNION ALL`, `ORDER BY`, `LIMIT` and (uncorrelated) subquery in the `FROM` clause. Impala also supports `INSERT INTO` and `INSERT OVERWRITE`.

Impala supports data types with the same names and semantics as the equivalent Hive data types: `string`, `TINYINT`, `SMALLINT`, `INT`, `BIGINT`, `FLOAT`, `DOUBLE`, `BOOLEAN`, `STRING`, `TIMESTAMP`.

For full details about Impala SQL syntax and semantics, see [SQL Statements](#) on page 162.

Most HiveQL `SELECT` and `INSERT` statements run unmodified with Impala. For information about Hive syntax not available in Impala, see [SQL Differences Between Impala and Hive](#) on page 323.

For a list of the built-in functions available in Impala queries, see [Built-in Functions](#) on page 257.

## Comments

Impala supports the familiar styles of SQL comments:

- All text from a `--` sequence to the end of the line is considered a comment and ignored. This type of comment can occur on a single line by itself, or after all or part of a statement.
- All text from a `/*` sequence to the next `*/` sequence is considered a comment and ignored. This type of comment can stretch over multiple lines. This type of comment can occur on one or more lines by itself, in the middle of a statement, or before or after a statement.

For example:

```
-- This line is a comment about a table.
create table ...;

/*
This is a multi-line comment about a query.
*/
select ...;

select * from t /* This is an embedded comment about a query. */ where ...;

select * from t -- This is a trailing comment within a multi-line command.
where ...;
```

## Data Types

Impala supports a set of data types that you can use for table columns, expression values, and function arguments and return values.

- **Note:** Currently, Impala supports only scalar types, not composite or nested types. Accessing a table containing any columns with unsupported types causes an error.

For the notation to write literals of each of these data types, see [Literals](#) on page 138.

See [SQL Differences Between Impala and Hive](#) on page 323 for differences between Impala and Hive data types.

## BIGINT Data Type

An 8-byte integer data type used in `CREATE TABLE` and `ALTER TABLE` statements.

### Syntax:

In the column definition of a `CREATE TABLE` statement:

```
column_name BIGINT
```

**Range:** -9223372036854775808 .. 9223372036854775807. There is no `UNSIGNED` subtype.

**Conversions:** Impala automatically converts to a floating-point type (`FLOAT` or `DOUBLE`) automatically. Use `CAST()` to convert to `TINYINT`, `SMALLINT`, `INT`, `STRING`, or `TIMESTAMP`. Casting an integer or floating-point value `N` to `TIMESTAMP` produces a value that is `N` seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting `-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting `TIMESTAMP` represents a date and time in the local time zone.

### Examples:

```
CREATE TABLE t1 (x BIGINT);
SELECT CAST(1000 AS BIGINT);
```

### Usage notes:

`BIGINT` is a convenient type to use for column declarations because you can use any kind of integer values in `INSERT` statements and they are promoted to `BIGINT` where necessary. However, `BIGINT` also requires the most bytes of any integer type on disk and in memory, meaning your queries are not as efficient and scalable as possible if you overuse this type. Therefore, prefer to use the smallest integer type with sufficient range to hold all input values, and `CAST()` when necessary to the appropriate type.

For a convenient and automated way to check the bounds of the `BIGINT` type, call the functions `MIN_BIGINT()` and `MAX_BIGINT()`.

If an integer value is too large to be represented as a `BIGINT`, use a `DECIMAL` instead with sufficient digits of precision.

**NULL considerations:** Casting any non-numeric value to this type produces a `NULL` value.

**Partitioning:** Prefer to use this type for a partition key column. Impala can process the numeric type more efficiently than a `STRING` representation of the value.

**HBase considerations:** This data type is fully compatible with HBase tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Internal details:** Represented in memory as an 8-byte value.

**Added in:** Available in all versions of Impala.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

### Related information:

[Numeric Literals](#) on page 139, [TINYINT Data Type](#) on page 135, [SMALLINT Data Type](#) on page 128, [INT Data Type](#) on page 126, [BIGINT Data Type](#) on page 112, [DECIMAL Data Type \(CDH 5.1 or higher only\)](#) on page 116, [Impala Mathematical Functions](#) on page 259

## BOOLEAN Data Type

A data type used in CREATE TABLE and ALTER TABLE statements, representing a single true/false choice.

### Syntax:

In the column definition of a CREATE TABLE statement:

```
column_name BOOLEAN
```

**Range:** TRUE or FALSE. Do not use quotation marks around the TRUE and FALSE literal values. You can write the literal values in uppercase, lowercase, or mixed case. The values queried from a table are always returned in lowercase, true or false.

**Conversions:** Impala does not automatically convert any other type to BOOLEAN. All conversions must use an explicit call to the CAST( ) function.

You can use CAST( ) to convert TINYINT, SMALLINT, INT, BIGINT, FLOAT, or DOUBLE BOOLEAN: a value of 0 represents false, and any non-zero value is converted to true.

You can cast DECIMAL values to BOOLEAN, with the same treatment of zero and non-zero values as the other numeric types. You cannot cast a BOOLEAN to a DECIMAL.

You cannot cast a STRING value to BOOLEAN, although you can cast a BOOLEAN value to STRING, returning '1' for true values and '0' for false values.

Although you can cast a TIMESTAMP to a BOOLEAN or a BOOLEAN to a TIMESTAMP, the results are unlikely to be useful. Any non-zero TIMESTAMP (that is, any value other than 1970-01-01 00:00:00) becomes TRUE when converted to BOOLEAN, while 1970-01-01 00:00:00 becomes FALSE. A value of FALSE becomes 1970-01-01 00:00:00 when converted to BOOLEAN, and TRUE becomes one second past this epoch date, that is, 1970-01-01 00:00:01.

**NULL considerations:** An expression of this type produces a NULL value if any argument of the expression is NULL.

### Partitioning:

Do not use a BOOLEAN column as a partition key. Although you can create such a table, subsequent operations produce errors:

```
[localhost:21000] > create table truth_table (assertion string) partitioned by (truth boolean);
[localhost:21000] > insert into truth_table values ('Pigs can fly',false);
ERROR: AnalysisException: INSERT into table with BOOLEAN partition column (truth) is not supported: partitioning.truth_table
```

### Examples:

```
SELECT 1 < 2;
SELECT 2 = 5;
SELECT 100 < NULL, 100 > NULL;
CREATE TABLE assertions (claim STRING, really BOOLEAN);
INSERT INTO assertions VALUES
  ("1 is less than 2", 1 < 2),
  ("2 is the same as 5", 2 = 5),
  ("Grass is green", true),
  ("The moon is made of green cheese", false);
SELECT claim FROM assertions WHERE really = TRUE;
```

**HBase considerations:** This data type is fully compatible with HBase tables.

**Parquet considerations:** This type is fully compatible with Parquet tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

**Related information:** [Boolean Literals](#) on page 141, [SQL Operators](#) on page 142, [Impala Conditional Functions](#) on page 275

## CHAR Data Type (CDH 5.2 or higher only)

A fixed-length character type, padded with trailing spaces if necessary to achieve the specified length. If values are longer than the specified length, Impala truncates any trailing characters.

### Syntax:

In the column definition of a `CREATE TABLE` statement:

```
column_name CHAR(length)
```

The maximum length you can specify is 255.

### Semantics of trailing spaces:

- When you store a `CHAR` value shorter than the specified length in a table, queries return the value padded with trailing spaces if necessary; the resulting value has the same length as specified in the column definition.
- If you store a `CHAR` value containing trailing spaces in a table, those trailing spaces are not stored in the data file. When the value is retrieved by a query, the result could have a different number of trailing spaces. That is, the value includes however many spaces are needed to pad it to the specified length of the column.
- If you compare two `CHAR` values that differ only in the number of trailing spaces, those values are considered identical.

**Partitioning:** This type can be used for partition key columns. Because of the efficiency advantage of numeric values over character-based values, if the partition key is a string representation of a number, prefer to use an integer type with sufficient range (`INT`, `BIGINT`, and so on) where practical.

**HBase considerations:** This data type cannot be used with HBase tables.

### Parquet considerations:

- This type can be read from and written to Parquet files.
- There is no requirement for a particular level of Parquet.
- Parquet files generated by Impala and containing this type can be freely interchanged with other components such as Hive and MapReduce.
- Any trailing spaces, whether implicitly or explicitly specified, are not written to the Parquet data files.
- Parquet data files might contain values that are longer than allowed by the `CHAR(n)` length limit. Impala ignores any extra trailing characters when it processes those values during a query.

### Text table considerations:

Text data files might contain values that are longer than allowed for a particular `CHAR(n)` column. Any extra trailing characters are ignored when Impala processes those values during a query. Text data files can also contain values that are shorter than the defined length limit, and Impala pads them with trailing spaces up to the specified length. Any text data files produced by Impala `INSERT` statements do not include any trailing blanks for `CHAR` columns.

### Compatibility:

This type is available using Impala 2.0 or higher under CDH 4, or with Impala on CDH 5.2 or higher. There are no compatibility issues with other components when exchanging data files or running Impala on CDH 4.

Some other database systems make the length specification optional. For Impala, the length is required.

**Internal details:** Represented in memory as a byte array with the same size as the length specification. Values that are shorter than the specified length are padded on the right with trailing spaces.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

**UDF considerations:** This type cannot be used for the argument or return type of a user-defined function (UDF) or user-defined aggregate function (UDA).

### Examples:

These examples show how trailing spaces are not considered significant when comparing or processing `CHAR` values. `CAST()` truncates any longer string to fit within the defined length. If a `CHAR` value is shorter than the specified length, it is padded on the right with spaces until it matches the specified length. Therefore, `LENGTH()` represents the length including any trailing spaces, and `CONCAT()` also treats the column value as if it has trailing spaces.

```
select cast('x' as char(4)) = cast('x      ' as char(4)) as "unpadded equal to padded";
+-----+
| unpadded equal to padded |
+-----+
| true                         |
+-----+


create table char_length(c char(3));
insert into char_length values (cast('1' as char(3))), (cast('12' as char(3))),
(cast('123' as char(3))), (cast('123456' as char(3)));
select concat("[",c,"]") as c, length(c) from char_length;
+-----+
| c      | length(c) |
+-----+
| [1    ] | 3          |
| [12   ] | 3          |
| [123]  | 3          |
| [123]  | 3          |
+-----+
```

This example shows a case where data values are known to have a specific length, where `CHAR` is a logical data type to use.

```
create table addresses
(id bigint,
 street_name string,
 state_abbreviation char(2),
 country_abbreviation char(2));
```

The following example shows how values written by Impala do not physically include the trailing spaces. It creates a table using text format, with `CHAR` values much shorter than the declared length, and then prints the resulting data file to show that the delimited values are not separated by spaces. The same behavior applies to binary-format Parquet data files.

```
create table char_in_text (a char(20), b char(30), c char(40))
row format delimited fields terminated by ',';

insert into char_in_text values (cast('foo' as char(20)), cast('bar' as char(30)),
cast('baz' as char(40))), (cast('hello' as char(20)), cast('goodbye' as char(30)),
cast('aloha' as char(40)));

-- Running this Linux command inside impala-shell using the ! shortcut.
!hdfs dfs -cat
'hdfs://127.0.0.1:8020/user/hive/warehouse/impala_doc_testing.db/char_in_text/*.*';
foo,bar,baz
hello,goodbye,aloha
```

## Impala SQL Language Reference

The following example further illustrates the treatment of spaces. It replaces the contents of the previous table with some values including leading spaces, trailing spaces, or both. Any leading spaces are preserved within the data file, but trailing spaces are discarded. Then when the values are retrieved by a query, the leading spaces are retrieved verbatim while any necessary trailing spaces are supplied by Impala.

```
insert overwrite char_in_text values (cast('trailing   ' as char(20)), cast('  leading
and trailing   ' as char(30)), cast('    leading' as char(40)));
!hdfs dfs -cat
'hdfs://127.0.0.1:8020/user/hive/warehouse/impala_doc_testing.db/char_in_text/*.*';
trailing,  leading and trailing,  leading

select concat('[' ,a, ',' ]') as a, concat('[' ,b, ',' ]') as b, concat('[' ,c, ',' ]') as c from
char_in_text;
+-----+-----+-----+
| a      | b      | c      |
+-----+-----+-----+
| [trailing] | [leading and trailing] | [leading]
+-----+-----+-----+
```

### Restrictions:

Because the blank-padding behavior requires allocating the maximum length for each value in memory, for scalability reasons avoid declaring CHAR columns that are much longer than typical values in that column.

All data in CHAR and VARCHAR columns must be in a character encoding that is compatible with UTF-8. If you have binary data from another database system (that is, a BLOB type), use a STRING column to hold it.

When an expression compares a CHAR with a STRING or VARCHAR, the CHAR value is implicitly converted to STRING first, with trailing spaces preserved.

```
select cast("foo  " as char(5)) = 'foo' as "char equal to string";
+-----+
| char equal to string |
+-----+
| false                |
+-----+
```

This behavior differs from other popular database systems. To get the expected result of TRUE, cast the expressions on both sides to CHAR values of the appropriate length:

```
select cast("foo  " as char(5)) = cast('foo' as char(3)) as "char equal to string";
+-----+
| char equal to string |
+-----+
| true                 |
+-----+
```

This behavior is subject to change in future releases.

### Related information:

[STRING Data Type](#) on page 128, [VARCHAR Data Type \(CDH 5.2 or higher only\)](#) on page 136, [String Literals](#) on page 140, [Impala String Functions](#) on page 279

## DECIMAL Data Type (CDH 5.1 or higher only)

A numeric data type with fixed scale and precision, used in CREATE TABLE and ALTER TABLE statements. Suitable for financial and other arithmetic calculations where the imprecise representation and rounding behavior of FLOAT and DOUBLE make those types impractical.

### Syntax:

In the column definition of a `CREATE TABLE` statement:

```
column_name DECIMAL[ (precision[,scale]) ]
```

`DECIMAL` with no precision or scale values is equivalent to `DECIMAL(9,0)`.

#### Precision and Scale:

*precision* represents the total number of digits that can be represented by the column, regardless of the location of the decimal point. This value must be between 1 and 38. For example, representing integer values up to 9999, and floating-point values up to 99.99, both require a precision of 4. You can also represent corresponding negative values, without any change in the precision. For example, the range -9999 to 9999 still only requires a precision of 4.

*scale* represents the number of fractional digits. This value must be less than or equal to *precision*. A scale of 0 produces integral values, with no fractional part. If precision and scale are equal, all the digits come after the decimal point, making all the values between 0 and 0.999... or 0 and -0.999...

When *precision* and *scale* are omitted, a `DECIMAL` value is treated as `DECIMAL(9,0)`, that is, an integer value ranging from -999,999,999 to 999,999,999. This is the largest `DECIMAL` value that can still be represented in 4 bytes. If precision is specified but scale is omitted, Impala uses a value of zero for the scale.

Both *precision* and *scale* must be specified as integer literals, not any other kind of constant expressions.

To check the precision or scale for arbitrary values, you can call the [precision\(\) and scale\(\) built-in functions](#). For example, you might use these values to figure out how many characters are required for various fields in a report, or to understand the rounding characteristics of a formula as applied to a particular `DECIMAL` column.

#### Range:

The maximum precision value is 38. Thus, the largest integral value is represented by `DECIMAL(38,0)` (999... with 9 repeated 38 times). The most precise fractional value (between 0 and 1, or 0 and -1) is represented by `DECIMAL(38,38)`, with 38 digits to the right of the decimal point. The value closest to 0 would be .0000...1 (37 zeros and the final 1). The value closest to 1 would be .999... (9 repeated 38 times).

For a given precision and scale, the range of `DECIMAL` values is the same in the positive and negative directions. For example, `DECIMAL(4,2)` can represent from -99.99 to 99.99. This is different from other integral numeric types where the positive and negative bounds differ slightly.

When you use `DECIMAL` values in arithmetic expressions, the precision and scale of the result value are determined as follows:

- For addition and subtraction, the precision and scale are based on the maximum possible result, that is, if all the digits of the input values were 9s and the absolute values were added together.
- For multiplication, the precision is the sum of the precisions of the input values. The scale is the sum of the scales of the input values.
- For division, Impala sets the precision and scale to values large enough to represent the whole and fractional parts of the result.
- For `UNION`, the scale is the larger of the scales of the input values, and the precision is increased if necessary to accommodate any additional fractional digits. If the same input value has the largest precision and the largest scale, the result value has the same precision and scale. If one value has a larger precision but smaller scale, the scale of the result value is increased. For example, `DECIMAL(20,2) UNION DECIMAL(8,6)` produces a result of type `DECIMAL(24,6)`. The extra 4 fractional digits of scale (6-2) are accommodated by extending the precision by the same amount (20+4).
- To doublecheck, you can always call the `PRECISION()` and `SCALE()` functions on the results of an arithmetic expression to see the relevant values, or use a `CREATE TABLE AS SELECT` statement to define a column based on the return type of the expression.

#### Compatibility:

## Impala SQL Language Reference

- Using the `DECIMAL` type is only supported under CDH 5.1.0 and higher.
- Use the `DECIMAL` data type in Impala for applications where you used the `NUMBER` data type in Oracle. The Impala `DECIMAL` type does not support the Oracle idioms of \* for scale or negative values for precision.

### Conversions and casting:

Casting an integer or floating-point value `N` to `TIMESTAMP` produces a value that is `N` seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting `-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting `TIMESTAMP` represents a date and time in the local time zone.

Impala automatically converts between `DECIMAL` and other numeric types where possible. A `DECIMAL` with zero scale is converted to or from the smallest appropriate integral type. A `DECIMAL` with a fractional part is automatically converted to or from the smallest appropriate floating-point type. If the destination type does not have sufficient precision or scale to hold all possible values of the source type, Impala raises an error and does not convert the value.

For example, these statements show how expressions of `DECIMAL` and other types are reconciled to the same type in the context of `UNION` queries and `INSERT` statements:

```
[localhost:21000] > select cast(1 as int) as x union select cast(1.5 as decimal(9,4)) as x;
+-----+
| x   |
+-----+
| 1.5000
| 1.0000
+-----+
[localhost:21000] > create table int_vs_decimal as select cast(1 as int) as x union
select cast(1.5 as decimal(9,4)) as x;
+-----+
| summary      |
+-----+
| Inserted 2 row(s) |
+-----+
[localhost:21000] > desc int_vs_decimal;
+-----+-----+-----+
| name | type        | comment |
+-----+-----+-----+
| x   | decimal(14,4) |         |
+-----+-----+-----+
```

To avoid potential conversion errors, you can use `CAST()` to convert `DECIMAL` values to `FLOAT`, `TINYINT`, `SMALLINT`, `INT`, `BIGINT`, `STRING`, `TIMESTAMP`, or `BOOLEAN`. You can use exponential notation in `DECIMAL` literals or when casting from `STRING`, for example `1.0e6` to represent one million.

If you cast a value with more fractional digits than the scale of the destination type, any extra fractional digits are truncated (not rounded). Casting a value to a target type with not enough precision produces a result of `NULL` and displays a runtime warning.

```
[localhost:21000] > select cast(1.239 as decimal(3,2));
+-----+
| cast(1.239 as decimal(3,2)) |
+-----+
| 1.23                         |
+-----+
[localhost:21000] > select cast(1234 as decimal(3));
+-----+
| cast(1234 as decimal(3,0)) |
+-----+
| NULL                         |
+-----+
WARNINGS: Expression overflowed, returning NULL
```

When you specify integer literals, for example in `INSERT ... VALUES` statements or arithmetic expressions, those numbers are interpreted as the smallest applicable integer type. You must use `CAST()` calls for some combinations of integer literals and `DECIMAL` precision. For example, `INT` has a maximum value that is 10 digits

long, TINYINT has a maximum value that is 3 digits long, and so on. If you specify a value such as 123456 to go into a DECIMAL column, Impala checks if the column has enough precision to represent the largest value of that integer type, and raises an error if not. Therefore, use an expression like `CAST(123456 TO DECIMAL(9, 0))` for DECIMAL columns with precision 9 or less, `CAST(50 TO DECIMAL(2, 0))` for DECIMAL columns with precision 2 or less, and so on. For DECIMAL columns with precision 10 or greater, Impala automatically interprets the value as the correct DECIMAL type; however, because `DECIMAL(10)` requires 8 bytes of storage while `DECIMAL(9)` requires only 4 bytes, only use precision of 10 or higher when actually needed.

```
[localhost:21000] > create table decimals_9_0 (x decimal);
[localhost:21000] > insert into decimals_9_0 values (1), (2), (4), (8), (16), (1024),
(32768), (65536), (1000000);
ERROR: AnalysisException: Possible loss of precision for target table
'decimal_testing.decimals_9_0'.
Expression '1' (type: INT) would need to be cast to DECIMAL(9,0) for column 'x'
[localhost:21000] > insert into decimals_9_0 values (cast(1 as decimal)), (cast(2 as
decimal)), (cast(4 as decimal)), (cast(8 as decimal)), (cast(16 as decimal)), (cast(1024
as decimal)), (cast(32768 as decimal)), (cast(65536 as decimal)), (cast(1000000 as
decimal));

[localhost:21000] > create table decimals_10_0 (x decimal(10,0));
[localhost:21000] > insert into decimals_10_0 values (1), (2), (4), (8), (16), (1024),
(32768), (65536), (1000000);
[localhost:21000] >
```

Be aware that in memory and for binary file formats such as Parquet or Avro, `DECIMAL(10)` or higher consumes 8 bytes while `DECIMAL(9)` (the default for `DECIMAL`) or lower consumes 4 bytes. Therefore, to conserve space in large tables, use the smallest-precision `DECIMAL` type that is appropriate and `CAST()` literal values where necessary, rather than declaring `DECIMAL` columns with high precision for convenience.

To represent a very large or precise DECIMAL value as a literal, for example one that contains more digits than can be represented by a BIGINT literal, use a quoted string or a floating-point value for the number, and CAST( ) to the desired DECIMAL type:

- The result of an aggregate function such as `MAX()`, `SUM()`, or `AVG()` on `DECIMAL` values is promoted to a scale of 38, with the same precision as the underlying column. Thus, the result can represent the largest possible value at that particular precision.
  - `STRING` columns, literals, or expressions can be converted to `DECIMAL` as long as the overall number of digits and digits to the right of the decimal point fit within the specified precision and scale for the declared `DECIMAL` type. By default, a `DECIMAL` value with no specified scale or precision can hold a maximum of 9 digits of an integer value. If there are more digits in the string value than are allowed by the `DECIMAL` scale and precision, the result is `NULL`.

The following examples demonstrate how STRING values with integer and fractional parts are represented when converted to DECIMAL. If the scale is 0, the number is treated as an integer value with a maximum of *precision* digits. If the precision is greater than 0, the scale must be increased to account for the digits both to the left and right of the decimal point. As the precision increases, output values are printed with additional trailing zeros after the decimal point if needed. Any trailing zeros after the decimal point in the STRING value must fit within the number of digits specified by the precision.

```
[localhost:21000] > select cast('100' as decimal); -- Small integer value fits  
within 9 digits of scale.  
+-----+  
| cast('100' as decimal(9,0)) |  
+-----+  
| 100 |  
+-----+[localhost:21000] > select cast('100' as decimal(3,0)); -- Small integer value  
fits within 3 digits of scale.
```

```
+-----+
| cast('100' as decimal(3,0)) |
+-----+
| 100
+-----+
[localhost:21000] > select cast('100' as decimal(2,0)); -- 2 digits of scale is
not enough!
+-----+
| cast('100' as decimal(2,0)) |
+-----+
| NULL
+-----+
[localhost:21000] > select cast('100' as decimal(3,1)); -- (3,1) = 2 digits left
of the decimal point, 1 to the right. Not enough.
+-----+
| cast('100' as decimal(3,1)) |
+-----+
| NULL
+-----+
[localhost:21000] > select cast('100' as decimal(4,1)); -- 4 digits total, 1 to
the right of the decimal point.
+-----+
| cast('100' as decimal(4,1)) |
+-----+
| 100.0
+-----+
[localhost:21000] > select cast('98.6' as decimal(3,1)); -- (3,1) can hold a 3
digit number with 1 fractional digit.
+-----+
| cast('98.6' as decimal(3,1)) |
+-----+
| 98.6
+-----+
[localhost:21000] > select cast('98.6' as decimal(15,1)); -- Larger scale allows
bigger numbers but still only 1 fractional digit.
+-----+
| cast('98.6' as decimal(15,1)) |
+-----+
| 98.6
+-----+
[localhost:21000] > select cast('98.6' as decimal(15,5)); -- Larger precision
allows more fractional digits, outputs trailing zeros.
+-----+
| cast('98.6' as decimal(15,5)) |
+-----+
| 98.60000
+-----+
[localhost:21000] > select cast('98.60000' as decimal(15,1)); -- Trailing zeros
in the string must fit within 'scale' digits (1 in this case).
+-----+
| cast('98.60000' as decimal(15,1)) |
+-----+
| NULL
+-----+
```

- Most built-in arithmetic functions such as `SIN()` and `COS()` continue to accept only `DOUBLE` values because they are so commonly used in scientific context for calculations of IEEE 954-compliant values. The built-in functions that accept and return `DECIMAL` are:

- `ABS()`
- `CEIL()`
- `COALESCE()`
- `FLOOR()`
- `FNV_HASH()`
- `GREATEST()`
- `IF()`
- `ISNULL()`
- `LEAST()`

- NEGATIVE()
- NULLIF()
- POSITIVE()
- PRECISION()
- ROUND()
- SCALE()
- TRUNCATE()
- ZEROIFNULL()

See [Built-in Functions](#) on page 257 for details.

- BIGINT, INT, SMALLINT, and TINYINT values can all be cast to DECIMAL. The number of digits to the left of the decimal point in the DECIMAL type must be sufficient to hold the largest value of the corresponding integer type. Note that integer literals are treated as the smallest appropriate integer type, meaning there is sometimes a range of values that require one more digit of DECIMAL scale than you might expect. For integer values, the precision of the DECIMAL type can be zero; if the precision is greater than zero, remember to increase the scale value by an equivalent amount to hold the required number of digits to the left of the decimal point.

The following examples show how different integer types are converted to DECIMAL.

```
[localhost:21000] > select cast(1 as decimal(1,0));
+-----+
| cast(1 as decimal(1,0)) |
+-----+
| 1
+-----+
[localhost:21000] > select cast(9 as decimal(1,0));
+-----+
| cast(9 as decimal(1,0)) |
+-----+
| 9
+-----+
[localhost:21000] > select cast(10 as decimal(1,0));
+-----+
| cast(10 as decimal(1,0)) |
+-----+
| 10
+-----+
[localhost:21000] > select cast(10 as decimal(1,1));
+-----+
| cast(10 as decimal(1,1)) |
+-----+
| 10.0
+-----+
[localhost:21000] > select cast(100 as decimal(1,1));
+-----+
| cast(100 as decimal(1,1)) |
+-----+
| 100.0
+-----+
[localhost:21000] > select cast(1000 as decimal(1,1));
+-----+
| cast(1000 as decimal(1,1)) |
+-----+
| 1000.0
+-----+
```

- When a DECIMAL value is converted to any of the integer types, any fractional part is truncated (that is, rounded towards zero):

```
[localhost:21000] > create table num_dec_days (x decimal(4,1));
[localhost:21000] > insert into num_dec_days values (1), (2), (cast(4.5 as decimal(4,1)));
[localhost:21000] > insert into num_dec_days values (cast(0.1 as decimal(4,1))), (cast(.9 as decimal(4,1))), (cast(9.1 as decimal(4,1))), (cast(9.9 as decimal(4,1)));
```

```
[localhost:21000] > select cast(x as int) from num_dec_days;
+-----+
| cast(x as int) |
+-----+
| 1
| 2
| 4
| 0
| 0
| 9
| 9
+-----+
```

- You cannot directly cast `TIMESTAMP` or `BOOLEAN` values to or from `DECIMAL` values. You can turn a `DECIMAL` value into a time-related representation using a two-step process, by converting it to an integer value and then using that result in a call to a date and time function such as `from_unixtime()`.

```
[localhost:21000] > select from_unixtime(cast(cast(1000.0 as decimal) as bigint));
+-----+
| from_unixtime(cast(cast(1000.0 as decimal(9,0)) as bigint)) |
+-----+
| 1970-01-01 00:16:40
+-----+
[localhost:21000] > select now() + interval cast(x as int) days from num_dec_days;
-- x is a DECIMAL column.

[localhost:21000] > create table num_dec_days (x decimal(4,1));
[localhost:21000] > insert into num_dec_days values (1), (2), (cast(4.5 as decimal(4,1)));
[localhost:21000] > select now() + interval cast(x as int) days from num_dec_days;
-- The 4.5 value is truncated to 4 and becomes '4 days'.
+-----+
| now() + interval cast(x as int) days |
+-----+
| 2014-05-13 23:11:55.163284000
| 2014-05-14 23:11:55.163284000
| 2014-05-16 23:11:55.163284000
+-----+
```

- Because values in `INSERT` statements are checked rigorously for type compatibility, be prepared to use `CAST()` function calls around literals, column references, or other expressions that you are inserting into a `DECIMAL` column.

**NULL considerations:** Casting any non-numeric value to this type produces a `NULL` value.

### DECIMAL differences from integer and floating-point types:

With the `DECIMAL` type, you are concerned with the number of overall digits of a number rather than powers of 2 (as in `TINYINT`, `SMALLINT`, and so on). Therefore, the limits with integral values of `DECIMAL` types fall around 99, 999, 9999, and so on rather than 32767, 65535,  $2^{32} - 1$ , and so on. For fractional values, you do not need to account for imprecise representation of the fractional part according to the IEEE-954 standard (as in `FLOAT` and `DOUBLE`). Therefore, when you insert a fractional value into a `DECIMAL` column, you can compare, sum, query, `GROUP BY`, and so on that column and get back the original values rather than some "close but not identical" value.

`FLOAT` and `DOUBLE` can cause problems or unexpected behavior due to inability to precisely represent certain fractional values, for example dollar and cents values for currency. You might find output values slightly different than you inserted, equality tests that do not match precisely, or unexpected values for `GROUP BY` columns. `DECIMAL` can help reduce unexpected behavior and rounding errors, at the expense of some performance overhead for assignments and comparisons.

### Literals and expressions:

- When you use an integer literal such as 1 or 999 in a SQL statement, depending on the context, Impala will treat it as either the smallest appropriate `DECIMAL` type, or the smallest integer type (`TINYINT`, `SMALLINT`,

`INT`, or `BIGINT`). To minimize memory usage, Impala prefers to treat the literal as the smallest appropriate integer type.

- When you use a floating-point literal such as `1.1` or `999.44` in a SQL statement, depending on the context, Impala will treat it as either the smallest appropriate `DECIMAL` type, or the smallest floating-point type (`FLOAT` or `DOUBLE`). To avoid loss of accuracy, Impala prefers to treat the literal as a `DECIMAL`.

#### Storage considerations:

- Only the precision determines the storage size for `DECIMAL` values; the scale setting has no effect on the storage size.
- Text, RCFFile, and SequenceFile tables all use ASCII-based formats. In these text-based file formats, leading zeros are not stored, but trailing zeros are stored. In these tables, each `DECIMAL` value takes up as many bytes as there are digits in the value, plus an extra byte if the decimal point is present and an extra byte for negative values. Once the values are loaded into memory, they are represented in 4, 8, or 16 bytes as described in the following list items. The on-disk representation varies depending on the file format of the table.
- Parquet and Avro tables use binary formats. In these tables, Impala stores each value in as few bytes as possible depending on the precision specified for the `DECIMAL` column.
  - In memory, `DECIMAL` values with precision of 9 or less are stored in 4 bytes.
  - In memory, `DECIMAL` values with precision of 10 through 18 are stored in 8 bytes.
  - In memory, `DECIMAL` values with precision greater than 18 are stored in 16 bytes.

#### File format considerations:

- The `DECIMAL` data type can be stored in any of the file formats supported by Impala, as described in [How Impala Works with Hadoop File Formats](#) on page 399. Impala only writes to tables that use the Parquet and text formats, so those formats are the focus for file format compatibility.
- Impala can query Avro, RCFFile, or SequenceFile tables containing `DECIMAL` columns, created by other Hadoop components, on CDH 5 only.
- You can use `DECIMAL` columns in Impala tables that are mapped to HBase tables. Impala can query and insert into such tables.
- Text, RCFFile, and SequenceFile tables all use ASCII-based formats. In these tables, each `DECIMAL` value takes up as many bytes as there are digits in the value, plus an extra byte if the decimal point is present. The binary format of Parquet or Avro files offers more compact storage for `DECIMAL` columns.
- Parquet and Avro tables use binary formats. In these tables, Impala stores each value in 4, 8, or 16 bytes depending on the precision specified for the `DECIMAL` column.
- Parquet files containing `DECIMAL` columns are not expected to be readable under CDH 4. See the **Compatibility** section for details.

**UDF considerations:** When writing a C++ UDF, use the `DecimalVal` data type defined in `/usr/include/impala_udf/udf.h`.

#### Partitioning:

You can use a `DECIMAL` column as a partition key. Doing so provides a better match between the partition key values and the HDFS directory names than using a `DOUBLE` or `FLOAT` partitioning column:

#### Schema evolution considerations:

- For text-based formats (text, RCFFile, and SequenceFile tables), you can issue an `ALTER TABLE ... REPLACE COLUMNS` statement to change the precision and scale of an existing `DECIMAL` column. As long as the values in the column fit within the new precision and scale, they are returned correctly by a query. Any values that do not fit within the new precision and scale are returned as `NULL`, and Impala reports the conversion error. Leading zeros do not count against the precision value, but trailing zeros after the decimal point do.

```
[localhost:21000] > create table text_decimals (x string);
[localhost:21000] > insert into text_decimals values ("1"), ("2"), ("99.99"),
("1.234"), ("000001"), ("1.000000000");
[localhost:21000] > select * from text_decimals;
```

```
+---+  
| x |  
+---+  
| 1 |  
| 2 |  
| 99.99 |  
| 1.234 |  
| 000001 |  
| 1.0000000000 |  
+---+  
[localhost:21000] > alter table text_decimals replace columns (x decimal(4,2));  
[localhost:21000] > select * from text_decimals;  
+---+  
| x |  
+---+  
| 1.00 |  
| 2.00 |  
| 99.99 |  
| NULL |  
| 1.00 |  
| NULL |  
+---+  
ERRORS:  
Backend 0:Error converting column: 0 TO DECIMAL(4, 2) (Data is: 1.234)  
file:  
hdfs://127.0.0.1:8020/user/hive/warehouse/decimal_testing.db/text_decimals/634d4bd3aa0  
e8420-b4b13bab7f1be787_56794587_data.0  
record: 1.234  
Error converting column: 0 TO DECIMAL(4, 2) (Data is: 1.0000000000)  
file:  
hdfs://127.0.0.1:8020/user/hive/warehouse/decimal_testing.db/text_decimals/cd40dc68e20  
c565a-cc4bd86c724c96ba_311873428_data.0  
record: 1.0000000000
```

- For binary formats (Parquet and Avro tables), although an `ALTER TABLE ... REPLACE COLUMNS` statement that changes the precision or scale of a `DECIMAL` column succeeds, any subsequent attempt to query the changed column results in a fatal error. (The other columns can still be queried successfully.) This is because the metadata about the columns is stored in the data files themselves, and `ALTER TABLE` does not actually make any updates to the data files. If the metadata in the data files disagrees with the metadata in the metastore database, Impala cancels the query.

### Examples:

```
CREATE TABLE t1 (x DECIMAL, y DECIMAL(5,2), z DECIMAL(25,0));  
INSERT INTO t1 VALUES (5, 99.44, 123456), (300, 6.7, 999999999);  
SELECT x+y, ROUND(y,1), z/98.6 FROM t1;  
SELECT CAST(1000.5 AS DECIMAL);
```

### Restrictions:

Currently, the `COMPUTE STATS` statement under CDH 4 does not store any statistics for `DECIMAL` columns. When Impala runs under CDH 5, which has better support for `DECIMAL` in the metastore database, `COMPUTE STATS` does collect statistics for `DECIMAL` columns and Impala uses the statistics to optimize query performance.

**HBase considerations:** This data type is fully compatible with HBase tables.

**Parquet considerations:** This type is fully compatible with Parquet tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

### Related information:

[Numeric Literals](#) on page 139, [TINYINT Data Type](#) on page 135, [SMALLINT Data Type](#) on page 128, [INT Data Type](#) on page 126, [BIGINT Data Type](#) on page 112, [DECIMAL Data Type \(CDH 5.1 or higher only\)](#) on page 116, [Impala Mathematical Functions](#) on page 259 (especially `PRECISION()` and `SCALE()`)

## DOUBLE Data Type

An 8-byte (double precision) floating-point data type used in CREATE TABLE and ALTER TABLE statements.

### Syntax:

In the column definition of a CREATE TABLE statement:

```
column_name DOUBLE
```

**Range:** 4.94065645841246544e-324d .. 1.79769313486231570e+308, positive or negative

**Conversions:** Impala does not automatically convert DOUBLE to any other type. You can use CAST( ) to convert DOUBLE values to FLOAT, TINYINT, SMALLINT, INT, BIGINT, STRING, TIMESTAMP, or BOOLEAN. You can use exponential notation in DOUBLE literals or when casting from STRING, for example 1.0e6 to represent one million. Casting an integer or floating-point value N to TIMESTAMP produces a value that is N seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting -use\_local\_tz\_for\_unix\_timestamp\_conversions=true is in effect, the resulting TIMESTAMP represents a date and time in the local time zone.

### Usage notes:

The data type REAL is an alias for DOUBLE.

### Examples:

```
CREATE TABLE t1 (x DOUBLE);
SELECT CAST(1000.5 AS DOUBLE);
```

**Partitioning:** Because fractional values of this type are not always represented precisely, when this type is used for a partition key column, the underlying HDFS directories might not be named exactly as you expect. Prefer to partition on a DECIMAL column instead.

**HBase considerations:** This data type is fully compatible with HBase tables.

**Parquet considerations:** This type is fully compatible with Parquet tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Internal details:** Represented in memory as an 8-byte value.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the COMPUTE STATS statement.

### Restrictions:

Due to the way arithmetic on FLOAT and DOUBLE columns uses high-performance hardware instructions, and distributed queries can perform these operations in different order for each query, results can vary slightly for aggregate function calls such as SUM( ) and AVG( ) for FLOAT and DOUBLE columns, particularly on large data sets where millions or billions of values are summed or averaged. For perfect consistency and repeatability, use the DECIMAL data type for such operations instead of FLOAT or DOUBLE.

### Related information:

[Numeric Literals](#) on page 139, [Impala Mathematical Functions](#) on page 259, [FLOAT Data Type](#) on page 125

## FLOAT Data Type

A 4-byte (single precision) floating-point data type used in CREATE TABLE and ALTER TABLE statements.

### Syntax:

## Impala SQL Language Reference

In the column definition of a `CREATE TABLE` statement:

```
column_name FLOAT
```

**Range:** 1.40129846432481707e-45 .. 3.40282346638528860e+38, positive or negative

**Conversions:** Impala automatically converts `FLOAT` to more precise `DOUBLE` values, but not the other way around. You can use `CAST()` to convert `FLOAT` values to `TINYINT`, `SMALLINT`, `INT`, `BIGINT`, `STRING`, `TIMESTAMP`, or `BOOLEAN`. You can use exponential notation in `FLOAT` literals or when casting from `STRING`, for example `1.0e6` to represent one million. Casting an integer or floating-point value `N` to `TIMESTAMP` produces a value that is `N` seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting `-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting `TIMESTAMP` represents a date and time in the local time zone.

**Examples:**

```
CREATE TABLE t1 (x FLOAT);
SELECT CAST(1000.5 AS FLOAT);
```

**Partitioning:** Because fractional values of this type are not always represented precisely, when this type is used for a partition key column, the underlying HDFS directories might not be named exactly as you expect. Prefer to partition on a `DECIMAL` column instead.

**HBase considerations:** This data type is fully compatible with HBase tables.

**Parquet considerations:** This type is fully compatible with Parquet tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Internal details:** Represented in memory as a 4-byte value.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

**Restrictions:**

Due to the way arithmetic on `FLOAT` and `DOUBLE` columns uses high-performance hardware instructions, and distributed queries can perform these operations in different order for each query, results can vary slightly for aggregate function calls such as `SUM()` and `AVG()` for `FLOAT` and `DOUBLE` columns, particularly on large data sets where millions or billions of values are summed or averaged. For perfect consistency and repeatability, use the `DECIMAL` data type for such operations instead of `FLOAT` or `DOUBLE`.

**Related information:**

[Numeric Literals](#) on page 139, [Impala Mathematical Functions](#) on page 259, [DOUBLE Data Type](#) on page 125

## INT Data Type

A 4-byte integer data type used in `CREATE TABLE` and `ALTER TABLE` statements.

**Syntax:**

In the column definition of a `CREATE TABLE` statement:

```
column_name INT
```

**Range:** -2147483648 .. 2147483647. There is no `UNSIGNED` subtype.

**Conversions:** Impala automatically converts to a larger integer type (`BIGINT`) or a floating-point type (`FLOAT` or `DOUBLE`) automatically. Use `CAST()` to convert to `TINYINT`, `SMALLINT`, `STRING`, or `TIMESTAMP`. Casting an integer or floating-point value `N` to `TIMESTAMP` produces a value that is `N` seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting

`-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting `TIMESTAMP` represents a date and time in the local time zone.

#### Usage notes:

The data type `INTEGER` is an alias for `INT`.

For a convenient and automated way to check the bounds of the `INT` type, call the functions `MIN_INT()` and `MAX_INT()`.

If an integer value is too large to be represented as a `INT`, use a `BIGINT` instead.

**NULL considerations:** Casting any non-numeric value to this type produces a `NULL` value.

#### Examples:

```
CREATE TABLE t1 (x INT);
SELECT CAST(1000 AS INT);
```

**Partitioning:** Prefer to use this type for a partition key column. Impala can process the numeric type more efficiently than a `STRING` representation of the value.

**HBase considerations:** This data type is fully compatible with HBase tables.

#### Parquet considerations:

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Internal details:** Represented in memory as a 4-byte value.

**Added in:** Available in all versions of Impala.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

#### Related information:

[Numeric Literals](#) on page 139, [TINYINT Data Type](#) on page 135, [SMALLINT Data Type](#) on page 128, [INT Data Type](#) on page 126, [BIGINT Data Type](#) on page 112, [DECIMAL Data Type \(CDH 5.1 or higher only\)](#) on page 116, [Impala Mathematical Functions](#) on page 259

## REAL Data Type

An alias for the `DOUBLE` data type. See [DOUBLE Data Type](#) on page 125 for details.

#### Examples:

These examples show how you can use the type names `REAL` and `DOUBLE` interchangeably, and behind the scenes Impala treats them always as `DOUBLE`.

```
[localhost:21000] > create table r1 (x real);
[localhost:21000] > describe r1;
+-----+-----+-----+
| name | type   | comment |
+-----+-----+-----+
| x    | double |          |
+-----+-----+-----+
[localhost:21000] > insert into r1 values (1.5), (cast (2.2 as double));
[localhost:21000] > select cast (1e6 as real);
+-----+
| cast(1000000.0 as double) |
+-----+
| 1000000                         |
+-----+
```

### SMALLINT Data Type

A 2-byte integer data type used in `CREATE TABLE` and `ALTER TABLE` statements.

#### Syntax:

In the column definition of a `CREATE TABLE` statement:

```
column_name SMALLINT
```

**Range:** -32768 .. 32767. There is no `UNSIGNED` subtype.

**Conversions:** Impala automatically converts to a larger integer type (`INT` or `BIGINT`) or a floating-point type (`FLOAT` or `DOUBLE`) automatically. Use `CAST()` to convert to `TINYINT`, `STRING`, or `TIMESTAMP`. Casting an integer or floating-point value `N` to `TIMESTAMP` produces a value that is `N` seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting `-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting `TIMESTAMP` represents a date and time in the local time zone.

#### Usage notes:

For a convenient and automated way to check the bounds of the `SMALLINT` type, call the functions `MIN_SMALLINT()` and `MAX_SMALLINT()`.

If an integer value is too large to be represented as a `SMALLINT`, use an `INT` instead.

**NULL considerations:** Casting any non-numeric value to this type produces a `NULL` value.

#### Examples:

```
CREATE TABLE t1 (x SMALLINT);
SELECT CAST(1000 AS SMALLINT);
```

#### Parquet considerations:

Physically, Parquet files represent `TINYINT` and `SMALLINT` values as 32-bit integers. Although Impala rejects attempts to insert out-of-range values into such columns, if you create a new table with the `CREATE TABLE ... LIKE PARQUET` syntax, any `TINYINT` or `SMALLINT` columns in the original table turn into `INT` columns in the new table.

**Partitioning:** Prefer to use this type for a partition key column. Impala can process the numeric type more efficiently than a `STRING` representation of the value.

**HBase considerations:** This data type is fully compatible with HBase tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Internal details:** Represented in memory as a 2-byte value.

**Added in:** Available in all versions of Impala.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

#### Related information:

[Numeric Literals](#) on page 139, [TINYINT Data Type](#) on page 135, [SMALLINT Data Type](#) on page 128, [INT Data Type](#) on page 126, [BIGINT Data Type](#) on page 112, [DECIMAL Data Type \(CDH 5.1 or higher only\)](#) on page 116, [Impala Mathematical Functions](#) on page 259

### STRING Data Type

A data type used in `CREATE TABLE` and `ALTER TABLE` statements.

#### Syntax:

In the column definition of a `CREATE TABLE` statement:

```
column_name STRING
```

**Length:** Maximum of 32,767 bytes. Do not use any length constraint when declaring `STRING` columns, as you might be familiar with from `VARCHAR`, `CHAR`, or similar column types from relational database systems. If you do need to manipulate string values with precise or maximum lengths, in Impala 2.0 and higher you can declare columns as `VARCHAR(max_length)` or `CHAR(length)`, but for best performance use `STRING` where practical.

**Character sets:** For full support in all Impala subsystems, restrict string values to the ASCII character set. UTF-8 character data can be stored in Impala and retrieved through queries, but UTF-8 strings containing non-ASCII characters are not guaranteed to work properly with string manipulation functions, comparison operators, or the `ORDER BY` clause. For any national language aspects such as collation order or interpreting extended ASCII variants such as ISO-8859-1 or ISO-8859-2 encodings, Impala does not include such metadata with the table definition. If you need to sort, manipulate, or display data depending on those national language characteristics of string data, use logic on the application side.

#### Conversions:

- Impala does not automatically convert `STRING` to any numeric type. Impala does automatically convert `STRING` to `TIMESTAMP` if the value matches one of the accepted `TIMESTAMP` formats; see [TIMESTAMP Data Type](#) on page 130 for details.
- You can use `CAST()` to convert `STRING` values to `TINYINT`, `SMALLINT`, `INT`, `BIGINT`, `FLOAT`, `DOUBLE`, or `TIMESTAMP`.
- You cannot directly cast a `STRING` value to `BOOLEAN`. You can use a `CASE` expression to evaluate string values such as `'T'`, `'true'`, and so on and return Boolean `true` and `false` values as appropriate.
- You can cast a `BOOLEAN` value to `STRING`, returning `'1'` for `true` values and `'0'` for `false` values.

#### Partitioning:

Although it might be convenient to use `STRING` columns for partition keys, even when those columns contain numbers, for performance and scalability it is much better to use numeric columns as partition keys whenever practical. Although the underlying HDFS directory name might be the same in either case, the in-memory storage for the partition key columns is more compact, and computations are faster, if partition key columns such as `YEAR`, `MONTH`, `DAY` and so on are declared as `INT`, `SMALLINT`, and so on.

**Zero-length strings:** For purposes of clauses such as `DISTINCT` and `GROUP BY`, Impala considers zero-length strings (" "), `NULL`, and space to all be different values.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Column statistics considerations:** Because the values of this type have variable size, none of the column statistics fields are filled in until you run the `COMPUTE STATS` statement.

#### Examples:

The following examples demonstrate double-quoted and single-quoted string literals, and required escaping for quotation marks within string literals:

```
SELECT 'I am a single-quoted string';
SELECT "I am a double-quoted string";
SELECT 'I\'m a single-quoted string with an apostrophe';
SELECT "I\'m a double-quoted string with an apostrophe";
SELECT 'I am a "short" single-quoted string containing quotes';
SELECT "I am a \"short\" double-quoted string containing quotes";
```

## Impala SQL Language Reference

The following examples demonstrate calls to string manipulation functions to concatenate strings, convert numbers to strings, or pull out substrings:

```
SELECT CONCAT("Once upon a time, there were ", CAST(3 AS STRING), ' little pigs.');
SELECT SUBSTR("hello world", 7, 5);
```

The following examples show how to perform operations on STRING columns within a table:

```
CREATE TABLE t1 (s1 STRING, s2 STRING);
INSERT INTO t1 VALUES ("hello", 'world'), (CAST(7 AS STRING), "wonders");
SELECT s1, s2, length(s1) FROM t1 WHERE s2 LIKE 'w%';
```

### Related information:

[String Literals](#) on page 140, [CHAR Data Type \(CDH 5.2 or higher only\)](#) on page 114, [VARCHAR Data Type \(CDH 5.2 or higher only\)](#) on page 136, [Impala String Functions](#) on page 279, [Impala Date and Time Functions](#) on page 268

## TIMESTAMP Data Type

A data type used in CREATE TABLE and ALTER TABLE statements, representing a point in time.

### Syntax:

In the column definition of a CREATE TABLE statement:

```
column_name TIMESTAMP
```

**Range:** Allowed date values range from 1400-01-01 to 9999-12-31; this range is different from the Hive TIMESTAMP type. Internally, the resolution of the time portion of a TIMESTAMP value is in nanoseconds.

### INTERVAL expressions:

You can perform date arithmetic by adding or subtracting a specified number of time units, using the INTERVAL keyword and the + and - operators or `date_add()` and `date_sub()` functions. You can specify units as YEAR[S], MONTH[S], WEEK[S], DAY[S], HOUR[S], MINUTE[S], SECOND[S], MILLISECOND[S], MICROSECOND[S], and NANOSECOND[S]. You can only specify one time unit in each interval expression, for example `INTERVAL 3 DAYS` or `INTERVAL 25 HOURS`, but you can produce any granularity by adding together successive INTERVAL values, such as `timestamp_value + INTERVAL 3 WEEKS - INTERVAL 1 DAY + INTERVAL 10 MICROSECONDS`.

For example:

```
select now() + interval 1 day;
select date_sub(now(), interval 5 minutes);
insert into auction_details
  select auction_id, auction_start_time, auction_start_time + interval 2 days + interval
    12 hours
  from new_auctions;
```

### Time zones:

By default, Impala does not store timestamps using the local timezone, to avoid undesired results from unexpected time zone issues. Timestamps are stored and interpreted relative to UTC, both when written to or read from data files, or when converted to or from Unix time values through functions such as `from_unixtime()` or `unix_timestamp()`. To convert such a TIMESTAMP value to one that represents the date and time in a specific time zone, convert the original value with the `from_utc_timestamp()` function.

Because Impala does not assume that TIMESTAMP values are in any particular time zone, you must be conscious of the time zone aspects of data that you query, insert, or convert.

For consistency with Unix system calls, the `TIMESTAMP` returned by the `now()` function represents the local time in the system time zone, rather than in UTC. To store values relative to the current time in a portable way, convert any `now()` return values using the `to_utc_timestamp()` function first. For example, the following example shows that the current time in California (where Cloudera HQ is located) is shortly after 2 PM. If that value was

written to a data file, and shipped off to a distant server to be analyzed alongside other data from far-flung locations, the dates and times would not match up precisely because of time zone differences. Therefore, the `to_utc_timestamp()` function converts it using a common reference point, the UTC time zone (descended from the old Greenwich Mean Time standard). The '`PDT`' argument indicates that the original value is from the Pacific time zone with Daylight Saving Time in effect. When servers in all geographic locations run the same transformation on any local date and time values (with the appropriate time zone argument), the stored data uses a consistent representation. Impala queries can use functions such as `EXTRACT()`, `MIN()`, `AVG()`, and so on to do time-series analysis on those timestamps.

```
[localhost:21000] > select now();
+-----+
| now() |
+-----+
| 2015-04-09 14:07:46.580465000 |
+-----+
[localhost:21000] > select to_utc_timestamp(now(), 'PDT');
+-----+
| to_utc_timestamp(now(), 'pdt') |
+-----+
| 2015-04-09 21:08:07.664547000 |
+-----+
```

The converse function, `from_utc_timestamp()`, lets you take stored `TIMESTAMP` data or calculated results and convert back to local date and time for processing on the application side. The following example shows how you might represent some future date (such as the ending date and time of an auction) in UTC, and then convert back to local time when convenient for reporting or other processing. The final query in the example tests whether this arbitrary UTC date and time has passed yet, by converting it back to the local time zone and comparing it against the current date and time.

```
[localhost:21000] > select to_utc_timestamp(now() + interval 2 weeks, 'PDT');
+-----+
| to_utc_timestamp(now() + interval 2 weeks, 'pdt') |
+-----+
| 2015-04-23 21:08:34.152923000 |
+-----+
[localhost:21000] > select from_utc_timestamp('2015-04-23 21:08:34.152923000', 'PDT');
+-----+
| from_utc_timestamp('2015-04-23 21:08:34.152923000', 'pdt') |
+-----+
| 2015-04-23 14:08:34.152923000 |
+-----+
[localhost:21000] > select from_utc_timestamp('2015-04-23 21:08:34.152923000', 'PDT')
< now();
+-----+
| from_utc_timestamp('2015-04-23 21:08:34.152923000', 'pdt') < now() |
+-----+
| false |
+-----+
```

If you have data files written by Hive, those `TIMESTAMP` values represent the local timezone of the host where the data was written, potentially leading to inconsistent results when processed by Impala. To avoid compatibility problems or having to code workarounds, you can specify one or both of these `impalad` startup flags:

`-use_local_tz_for_unix_timestamp_conversions=true`  
`-convert_legacy_hive_parquet_utc_timestamps=true`. Although  
`-convert_legacy_hive_parquet_utc_timestamps` is turned off by default, Cloudera recommends turning it on to avoid unexpected behavior when processing `TIMESTAMP` columns in Parquet files written by Hive.

The `-use_local_tz_for_unix_timestamp_conversions` setting affects conversions from `TIMESTAMP` to `BIGINT`, or from `BIGINT` to `TIMESTAMP`. By default, Impala treats all `TIMESTAMP` values as UTC, to simplify analysis of time-series data from different geographic regions. When you enable the `-use_local_tz_for_unix_timestamp_conversions` setting, these operations treat the input values as if they are in the local time zone of the host doing the processing. See [Impala Date and Time Functions](#) on page 268 for the list of functions affected by the `-use_local_tz_for_unix_timestamp_conversions` setting.

## Impala SQL Language Reference

The following sequence of examples shows how the interpretation of `TIMESTAMP` values in Parquet tables is affected by the setting of the `-convert_legacy_hive_parquet_utc_timestamps` setting.

Regardless of the `-convert_legacy_hive_parquet_utc_timestamps` setting, `TIMESTAMP` columns in text tables can be written and read interchangeably by Impala and Hive:

Impala DDL and queries for text table:

```
[localhost:21000] > create table t1 (x timestamp);
[localhost:21000] > insert into t1 values (now()), (now() + interval 1 day);
[localhost:21000] > select x from t1;
+-----+
| x |
+-----+
| 2015-04-07 15:43:02.892403000 |
| 2015-04-08 15:43:02.892403000 |
+-----+
[localhost:21000] > select to_utc_timestamp(x, 'PDT') from t1;
+-----+
| to_utc_timestamp(x, 'pdt') |
+-----+
| 2015-04-07 22:43:02.892403000 |
| 2015-04-08 22:43:02.892403000 |
+-----+
```

Hive query for text table:

```
hive> select * from t1;
OK
2015-04-07 15:43:02.892403
2015-04-08 15:43:02.892403
Time taken: 1.245 seconds, Fetched: 2 row(s)
```

When the table uses Parquet format, Impala expects any time zone adjustment to be applied prior to writing, while `TIMESTAMP` values written by Hive are adjusted to be in the UTC time zone. When Hive queries Parquet data files that it wrote, it adjusts the `TIMESTAMP` values back to the local time zone, while Impala does no conversion. Hive does no time zone conversion when it queries Impala-written Parquet files.

Impala DDL and queries for Parquet table:

```
[localhost:21000] > create table p1 stored as parquet as select x from t1;
+-----+
| summary |
+-----+
| Inserted 2 row(s) |
+-----+
[localhost:21000] > select x from p1;
+-----+
| x |
+-----+
| 2015-04-07 15:43:02.892403000 |
| 2015-04-08 15:43:02.892403000 |
+-----+
```

Hive DDL and queries for Parquet table:

```
hive> create table h1 (x timestamp) stored as parquet;
OK
hive> insert into h1 select * from p1;
...
OK
Time taken: 35.573 seconds
hive> select x from p1;
OK
2015-04-07 15:43:02.892403
2015-04-08 15:43:02.892403
Time taken: 0.324 seconds, Fetched: 2 row(s)
hive> select x from h1;
OK
2015-04-07 15:43:02.892403
```

```
2015-04-08 15:43:02.892403
Time taken: 0.197 seconds, Fetched: 2 row(s)
```

The discrepancy arises when Impala queries the Hive-created Parquet table. The underlying values in the `TIMESTAMP` column are different from the ones written by Impala, even though they were copied from one table to another by an `INSERT ... SELECT` statement in Hive. Hive did an implicit conversion from the local time zone to UTC as it wrote the values to Parquet.

```
Impala query for TIMESTAMP values from Impala-written and Hive-written data:

[localhost:21000] > select * from p1;
+-----+
| x |
+-----+
| 2015-04-07 15:43:02.892403000 |
| 2015-04-08 15:43:02.892403000 |
+-----+
Fetched 2 row(s) in 0.29s
[localhost:21000] > select * from h1;
+-----+
| x |
+-----+
| 2015-04-07 22:43:02.892403000 |
| 2015-04-08 22:43:02.892403000 |
+-----+
Fetched 2 row(s) in 0.41s

Underlying integer values for Impala-written and Hive-written data:

[localhost:21000] > select cast(x as bigint) from p1;
+-----+
| cast(x as bigint) |
+-----+
| 1428421382 |
| 1428507782 |
+-----+
Fetched 2 row(s) in 0.38s
[localhost:21000] > select cast(x as bigint) from h1;
+-----+
| cast(x as bigint) |
+-----+
| 1428446582 |
| 1428532982 |
+-----+
Fetched 2 row(s) in 0.20s
```

When the `-convert_legacy_hive_parquet_utc_timestamps` setting is enabled, Impala recognizes the Parquet data files written by Hive, and applies the same UTC-to-local-timezone conversion logic during the query as Hive uses, making the contents of the Impala-written `p1` table and the Hive-written `h1` table appear identical, whether represented as `TIMESTAMP` values or the underlying `BIGINT` integers:

```
[localhost:21000] > select x from p1;
+-----+
| x |
+-----+
| 2015-04-07 15:43:02.892403000 |
| 2015-04-08 15:43:02.892403000 |
+-----+
Fetched 2 row(s) in 0.37s
[localhost:21000] > select x from h1;
+-----+
| x |
+-----+
| 2015-04-07 15:43:02.892403000 |
| 2015-04-08 15:43:02.892403000 |
+-----+
Fetched 2 row(s) in 0.19s
[localhost:21000] > select cast(x as bigint) from p1;
+-----+
```

```
| cast(x as bigint) |
+-----+
| 1428446582
| 1428532982
+-----+
Fetched 2 row(s) in 0.29s
[localhost:21000] > select cast(x as bigint) from h1;
+-----+
| cast(x as bigint) |
+-----+
| 1428446582
| 1428532982
+-----+
Fetched 2 row(s) in 0.22s
```

### Conversions:

Impala automatically converts STRING literals of the correct format into TIMESTAMP values. Timestamp values are accepted in the format YYYY-MM-DD HH:MM:SS.aaaaaaaaaa, and can consist of just the date, or just the time, with or without the fractional second portion. For example, you can specify TIMESTAMP values such as '1966-07-30', '08:30:00', or '1985-09-25 17:45:30.005'. Casting an integer or floating-point value N to TIMESTAMP produces a value that is N seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting `-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting TIMESTAMP represents a date and time in the local time zone.

In Impala 1.3 and higher, the `FROM_UNIXTIME()` and `UNIX_TIMESTAMP()` functions allow a wider range of format strings, with more flexibility in element order, repetition of letter placeholders, and separator characters. See [Impala Date and Time Functions](#) on page 268 for details.

In Impala 2.2.0 and higher, built-in functions that accept or return integers representing TIMESTAMP values use the `BIGINT` type for parameters and return values, rather than `INT`. This change lets the date and time functions avoid an overflow error that would otherwise occur on January 19th, 2038 (known as the ["Year 2038 problem"](#) or ["Y2K38 problem"](#)). This change affects the `from_unixtime()` and `unix_timestamp()` functions. You might need to change application code that interacts with these functions, change the types of columns that store the return values, or add `CAST()` calls to SQL statements that call these functions.

### Partitioning:

Although you cannot use a TIMESTAMP column as a partition key, you can extract the individual years, months, days, hours, and so on and partition based on those columns. Because the partition key column values are represented in HDFS directory names, rather than as fields in the data files themselves, you can also keep the original TIMESTAMP values if desired, without duplicating data or wasting storage space. See [Partition Key Columns](#) on page 397 for more details on partitioning with date and time values.

```
[localhost:21000] > create table timeline (event string) partitioned by (happened timestamp);
ERROR: AnalysisException: Type 'TIMESTAMP' is not supported as partition-column type
in column: happened
```

### Examples:

```
select cast('1966-07-30' as timestamp);
select cast('1985-09-25 17:45:30.005' as timestamp);
select cast('08:30:00' as timestamp);
select hour('1970-01-01 15:30:00');           -- Succeeds, returns 15.
select hour('1970-01-01 15:30');            -- Returns NULL because seconds field
required.
select hour('1970-01-01 27:30:00');          -- Returns NULL because hour value out of
range.
select dayofweek('2004-06-13');             -- Returns 1, representing Sunday.
select dayname('2004-06-13');              -- Returns 'Sunday'.
select date_add('2004-06-13', 365);        -- Returns 2005-06-13 with zeros for hh:mm:ss
fields.
select day('2004-06-13');                  -- Returns 13.
```

```

select datediff('1989-12-31','1984-09-01'); -- How many days between these 2 dates?
select now();                                -- Returns current date and time in local
   timezone.

create table dates_and_times (t timestamp);
insert into dates_and_times values
  ('1966-07-30'), ('1985-09-25 17:45:30.005'), ('08:30:00'), (now());

```

**NULL considerations:** Casting any unrecognized STRING value to this type produces a NULL value.

**Partitioning:** Because this type potentially has so many distinct values, it is often not a sensible choice for a partition key column. For example, events 1 millisecond apart would be stored in different partitions. Consider using the TRUNC() function to condense the number of distinct values, and partition on a new column with the truncated values.

**HBase considerations:** This data type is fully compatible with HBase tables.

**Parquet considerations:** This type is fully compatible with Parquet tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Internal details:** Represented in memory as a 16-byte value.

**Added in:** Available in all versions of Impala.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the COMPUTE STATS statement.

#### Restrictions:

If you cast a STRING with an unrecognized format to a TIMESTAMP, the result is NULL rather than an error. Make sure to test your data pipeline to be sure any textual date and time values are in a format that Impala TIMESTAMP can recognize.

Currently, Avro tables cannot contain TIMESTAMP columns. If you need to store date and time values in Avro tables, as a workaround you can use a STRING representation of the values, convert the values to BIGINT with the UNIX\_TIMESTAMP() function, or create separate numeric columns for individual date and time fields using the EXTRACT() function.

#### Related information:

- [Timestamp Literals](#) on page 141.
- To convert to or from different date formats, or perform date arithmetic, use the date and time functions described in [Impala Date and Time Functions](#) on page 268. In particular, the from\_unixtime() function requires a case-sensitive format string such as "yyyy-MM-dd HH:mm:ss.SSSS", matching one of the allowed variations of a TIMESTAMP value (date plus time, only date, only time, optional fractional seconds).
- See [SQL Differences Between Impala and Hive](#) on page 323 for details about differences in TIMESTAMP handling between Impala and Hive.

## TINYINT Data Type

A 1-byte integer data type used in CREATE TABLE and ALTER TABLE statements.

#### Syntax:

In the column definition of a CREATE TABLE statement:

```

column_name TINYINT

```

**Range:** -128 .. 127. There is no UNSIGNED subtype.

**Conversions:** Impala automatically converts to a larger integer type (SMALLINT, INT, or BIGINT) or a floating-point type (FLOAT or DOUBLE) automatically. Use CAST() to convert to STRING or TIMESTAMP. Casting an integer or floating-point value N to TIMESTAMP produces a value that is N seconds past the start of the epoch date (January

1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting `-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting `TIMESTAMP` represents a date and time in the local time zone.

### Usage notes:

For a convenient and automated way to check the bounds of the `TINYINT` type, call the functions `MIN_TINYINT()` and `MAX_TINYINT()`.

If an integer value is too large to be represented as a `TINYINT`, use a `SMALLINT` instead.

**NULL considerations:** Casting any non-numeric value to this type produces a `NULL` value.

### Examples:

```
CREATE TABLE t1 (x TINYINT);
SELECT CAST(100 AS TINYINT);
```

### Parquet considerations:

Physically, Parquet files represent `TINYINT` and `SMALLINT` values as 32-bit integers. Although Impala rejects attempts to insert out-of-range values into such columns, if you create a new table with the `CREATE TABLE ... LIKE PARQUET` syntax, any `TINYINT` or `SMALLINT` columns in the original table turn into `INT` columns in the new table.

**HBase considerations:** This data type is fully compatible with HBase tables.

**Text table considerations:** Values of this type are potentially larger in text tables than in tables using Parquet or other binary formats.

**Internal details:** Represented in memory as a 1-byte value.

**Added in:** Available in all versions of Impala.

**Column statistics considerations:** Because this type has a fixed size, the maximum and average size fields are always filled in for column statistics, even before you run the `COMPUTE STATS` statement.

### Related information:

[Numeric Literals](#) on page 139, [TINYINT Data Type](#) on page 135, [SMALLINT Data Type](#) on page 128, [INT Data Type](#) on page 126, [BIGINT Data Type](#) on page 112, [DECIMAL Data Type \(CDH 5.1 or higher only\)](#) on page 116, [Impala Mathematical Functions](#) on page 259

## VARCHAR Data Type (CDH 5.2 or higher only)

A variable-length character type, truncated during processing if necessary to fit within the specified length.

### Syntax:

In the column definition of a `CREATE TABLE` statement:

```
column_name VARCHAR(max_length)
```

The maximum length you can specify is 65,535.

**Partitioning:** This type can be used for partition key columns. Because of the efficiency advantage of numeric values over character-based values, if the partition key is a string representation of a number, prefer to use an integer type with sufficient range (`INT`, `BIGINT`, and so on) where practical.

**HBase considerations:** This data type cannot be used with HBase tables.

### Parquet considerations:

- This type can be read from and written to Parquet files.
- There is no requirement for a particular level of Parquet.

- Parquet files generated by Impala and containing this type can be freely interchanged with other components such as Hive and MapReduce.
- Parquet data files can contain values that are longer than allowed by the `VARCHAR(n)` length limit. Impala ignores any extra trailing characters when it processes those values during a query.

#### Text table considerations:

Text data files can contain values that are longer than allowed by the `VARCHAR(n)` length limit. Any extra trailing characters are ignored when Impala processes those values during a query.

#### Schema evolution considerations:

You can use `ALTER TABLE ... CHANGE` to switch column data types to and from `VARCHAR`. You can convert from `STRING` to `VARCHAR(n)`, or from `VARCHAR(n)` to `STRING`, or from `CHAR(n)` to `VARCHAR(n)`, or from `VARCHAR(n)` to `CHAR(n)`. When switching back and forth between `VARCHAR` and `CHAR`, you can also change the length value. This schema evolution works the same for tables using any file format. If a table contains values longer than the maximum length defined for a `VARCHAR` column, Impala does not return an error. Any extra trailing characters are ignored when Impala processes those values during a query.

#### Compatibility:

This type is available using Impala 2.0 or higher under CDH 4, or with Impala on CDH 5.2 or higher. There are no compatibility issues with other components when exchanging data files or running Impala on CDH 4.

**Internal details:** Represented in memory as a byte array with the minimum size needed to represent each value.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

**Column statistics considerations:** Because the values of this type have variable size, none of the column statistics fields are filled in until you run the `COMPUTE STATS` statement.

#### Restrictions:

All data in `CHAR` and `VARCHAR` columns must be in a character encoding that is compatible with UTF-8. If you have binary data from another database system (that is, a `BLOB` type), use a `STRING` column to hold it.

#### Examples:

The following examples show how long and short `VARCHAR` values are treated. Values longer than the maximum specified length are truncated by `CAST()`, or when queried from existing data files. Values shorter than the maximum specified length are represented as the actual length of the value, with no extra padding as seen with `CHAR` values.

```
create table varchar_1 (s varchar(1));
create table varchar_4 (s varchar(4));
create table varchar_20 (s varchar(20));

insert into varchar_1 values (cast('a' as varchar(1))), (cast('b' as varchar(1))),
(cast('hello' as varchar(1))), (cast('world' as varchar(1)));
insert into varchar_4 values (cast('a' as varchar(4))), (cast('b' as varchar(4))),
(cast('hello' as varchar(4))), (cast('world' as varchar(4)));
insert into varchar_20 values (cast('a' as varchar(20))), (cast('b' as varchar(20))),
(cast('hello' as varchar(20))), (cast('world' as varchar(20)));

select * from varchar_1;
+---+
| s |
+---+
| a |
| b |
| h |
| w |
+---+
select * from varchar_4;
+-----+
| s   |
+-----+
| a   |
+-----+
```

```
+-----+  
| b   |  
| hell |  
| worl |  
+-----+  
[localhost:21000] > select * from varchar_20;  
+-----+  
| s   |  
+-----+  
| a   |  
| b   |  
| hello |  
| world |  
+-----+  
select concat('[' ,s ,']') as s from varchar_20;  
+-----+  
| s   |  
+-----+  
| [a] |  
| [b] |  
| [hello] |  
| [world] |  
+-----+
```

The following example shows how identical VARCHAR values compare as equal, even if the columns are defined with different maximum lengths. Both tables contain 'a' and 'b' values. The longer 'hello' and 'world' values from the VARCHAR\_20 table were truncated when inserted into the VARCHAR\_1 table.

```
+-----+  
| s   |  
+-----+  
| a   |  
| b   |  
+-----+
```

The following examples show how VARCHAR values are freely interchangeable with STRING values in contexts such as comparison operators and built-in functions:

```
+-----+  
| length |  
+-----+  
| 3      |  
+-----+  
select cast('xyz' as varchar(5)) > cast('abc' as varchar(10)) as greater;  
+-----+  
| greater |  
+-----+  
| true    |  
+-----+
```

**UDF considerations:** This type cannot be used for the argument or return type of a user-defined function (UDF) or user-defined aggregate function (UDA).

### Related information:

[STRING Data Type](#) on page 128, [CHAR Data Type \(CDH 5.2 or higher only\)](#) on page 114, [String Literals](#) on page 140, [Impala String Functions](#) on page 279

## Literals

Each of the Impala data types has corresponding notation for literal values of that type. You specify literal values in SQL statements, such as in the `SELECT` list or `WHERE` clause of a query, or as an argument to a function call. See [Data Types](#) on page 111 for a complete list of types, ranges, and conversion rules.

## Numeric Literals

To write literals for the integer types (`TINYINT`, `SMALLINT`, `INT`, and `BIGINT`), use a sequence of digits with optional leading zeros.

To write literals for the floating-point types (`DECIMAL`, `FLOAT`, and `DOUBLE`), use a sequence of digits with an optional decimal point (. character). To preserve accuracy during arithmetic expressions, Impala interprets floating-point literals as the `DECIMAL` type with the smallest appropriate precision and scale, until required by the context to convert the result to `FLOAT` or `DOUBLE`.

Integer values are promoted to floating-point when necessary, based on the context.

You can also use exponential notation by including an e character. For example, `1e6` is 1 times 10 to the power of 6 (1 million). A number in exponential notation is always interpreted as floating-point.

When Impala encounters a numeric literal, it considers the type to be the "smallest" that can accurately represent the value. The type is promoted to larger or more accurate types if necessary, based on subsequent parts of an expression.

For example, you can see by the types Impala defines for the following table columns how it interprets the corresponding numeric literals:

```
[localhost:21000] > create table ten as select 10 as x;
+-----+
| summary |
+-----+
| Inserted 1 row(s) |
+-----+
[localhost:21000] > desc ten;
+-----+-----+
| name | type   | comment |
+-----+-----+
| x   | tinyint |          |
+-----+-----+

[localhost:21000] > create table four_k as select 4096 as x;
+-----+
| summary |
+-----+
| Inserted 1 row(s) |
+-----+
[localhost:21000] > desc four_k;
+-----+-----+
| name | type   | comment |
+-----+-----+
| x   | smallint |          |
+-----+-----+

[localhost:21000] > create table one_point_five as select 1.5 as x;
+-----+
| summary |
+-----+
| Inserted 1 row(s) |
+-----+
[localhost:21000] > desc one_point_five;
+-----+-----+
| name | type   | comment |
+-----+-----+
| x   | decimal(2,1)|          |
+-----+-----+

[localhost:21000] > create table one_point_three_three_three as select 1.333 as x;
+-----+
| summary |
+-----+
| Inserted 1 row(s) |
+-----+
[localhost:21000] > desc one_point_three_three_three;
+-----+-----+
| name | type   | comment |
+-----+-----+
```

	x	decimal(4,3)	
+-----+			-----+

## String Literals

String literals are quoted using either single or double quotation marks. You can use either kind of quotes for string literals, even both kinds for different literals within the same statement.

Quoted literals are considered to be of type STRING. To use quoted literals in contexts requiring a CHAR or VARCHAR value, CAST( ) the literal to a CHAR or VARCHAR of the appropriate length.

### Escaping special characters:

To encode special characters within a string literal, precede them with the backslash (\) escape character:

- \t represents a tab.
- \n represents a newline or linefeed. This might cause extra line breaks in impala-shell output.
- \r represents a carriage return. This might cause unusual formatting (making it appear that some content is overwritten) in impala-shell output.
- \b represents a backspace. This might cause unusual formatting (making it appear that some content is overwritten) in impala-shell output.
- \0 represents an ASCII nul character (not the same as a SQL NULL). This might not be visible in impala-shell output.
- \z represents a DOS end-of-file character. This might not be visible in impala-shell output.
- \% and \\_ can be used to escape wildcard characters within the string passed to the LIKE operator.
- \ followed by 3 octal digits represents the ASCII code of a single character; for example, \101 is ASCII 65, the character A.
- Use two consecutive backslashes (\\) to prevent the backslash from being interpreted as an escape character.
- Use the backslash to escape single or double quotation mark characters within a string literal, if the literal is enclosed by the same type of quotation mark.
- If the character following the \ does not represent the start of a recognized escape sequence, the character is passed through unchanged.

### Quotes within quotes:

To include a single quotation character within a string value, enclose the literal with either single or double quotation marks, and optionally escape the single quote as a \' sequence. Earlier releases required escaping a single quote inside double quotes. Continue using escape sequences in this case if you also need to run your SQL code on older versions of Impala.

To include a double quotation character within a string value, enclose the literal with single quotation marks, no escaping is necessary in this case. Or, enclose the literal with double quotation marks and escape the double quote as a \" sequence.

```
[localhost:21000] > select "What\'s happening?" as single_within_double,
>          'I\'m not sure.' as single_within_single,
>          "Homer wrote \"The Iliad\"." as double_within_double,
>          'Homer also wrote "The Odyssey".' as double_within_single;
+-----+-----+-----+
| single_within_double | single_within_single | double_within_double | 
double_within_single   |
+-----+-----+-----+
| What's happening?    | I'm not sure.       | Homer wrote "The Iliad". | Homer also
wrote "The Odyssey". |
+-----+-----+-----+
```

### Field terminator character in CREATE TABLE:

- Note:** The CREATE TABLE clauses FIELDS TERMINATED BY, ESCAPED BY, and LINES TERMINATED BY have special rules for the string literal used for their argument, because they all require a single character. You can use a regular character surrounded by single or double quotation marks, an octal sequence such as '\054' (representing a comma), or an integer in the range '-127'..'128' (with quotation marks but no backslash), which is interpreted as a single-byte ASCII character. Negative values are subtracted from 256; for example, FIELDS TERMINATED BY '-2' sets the field delimiter to ASCII code 254, the "Icelandic Thorn" character used as a delimiter by some data formats.

### impala-shell considerations:

When dealing with output that includes non-ASCII or non-printable characters such as linefeeds and backspaces, use the `impala-shell` options to save to a file, turn off pretty printing, or both rather than relying on how the output appears visually. See [impala-shell Configuration Options](#) on page 332 for a list of `impala-shell` options.

## Boolean Literals

For BOOLEAN values, the literals are TRUE and FALSE, with no quotation marks and case-insensitive.

### Examples:

```
select true;
select * from t1 where assertion = false;
select case bool_col when true then 'yes' when false 'no' else 'null' end from t1;
```

## Timestamp Literals

Impala automatically converts STRING literals of the correct format into TIMESTAMP values. Timestamp values are accepted in the format YYYY-MM-DD HH:MM:SS.sssssssss, and can consist of just the date, or just the time, with or without the fractional second portion. For example, you can specify TIMESTAMP values such as '1966-07-30', '08:30:00', or '1985-09-25 17:45:30.005'. Casting an integer or floating-point value N to TIMESTAMP produces a value that is N seconds past the start of the epoch date (January 1, 1970). By default, the result value represents a date and time in the UTC time zone. If the setting `-use_local_tz_for_unix_timestamp_conversions=true` is in effect, the resulting TIMESTAMP represents a date and time in the local time zone.

You can also use INTERVAL expressions to add or subtract from timestamp literal values, such as '1966-07-30' + INTERVAL 5 YEARS + INTERVAL 3 DAYS. See [TIMESTAMP Data Type](#) on page 130 for details.

## NULL

The notion of NULL values is familiar from all kinds of database systems, but each SQL dialect can have its own behavior and restrictions on NULL values. For Big Data processing, the precise semantics of NULL values are significant: any misunderstanding could lead to inaccurate results or misformatted data, that could be time-consuming to correct for large data sets.

- NULL is a different value than an empty string. The empty string is represented by a string literal with nothing inside, "" or ''.
- In a delimited text file, the NULL value is represented by the special token \N.
- When Impala inserts data into a partitioned table, and the value of one of the partitioning columns is NULL or the empty string, the data is placed in a special partition that holds only these two kinds of values. When these values are returned in a query, the result is NULL whether the value was originally NULL or an empty string. This behavior is compatible with the way Hive treats NULL values in partitioned tables. Hive does not allow empty strings as partition keys, and it returns a string value such as \_\_HIVE\_DEFAULT\_PARTITION\_\_ instead of NULL when such values are returned from a query. For example:

```
create table t1 (i int) partitioned by (x int, y string);
-- Select an INT column from another table, with all rows going into a special
HDFS subdirectory
```

```
-- named __HIVE_DEFAULT_PARTITION__. Depending on whether one or both of the
partitioning keys
-- are null, this special directory name occurs at different levels of the physical
data directory
-- for the table.
insert into t1 partition(x=NULL, y=NULL) select c1 from some_other_table;
insert into t1 partition(x, y=NULL) select c1, c2 from some_other_table;
insert into t1 partition(x=NULL, y) select c1, c3 from some_other_table;
```

- There is no NOT NULL clause when defining a column to prevent NULL values in that column.
- There is no DEFAULT clause to specify a non-NULL default value.
- If an INSERT operation mentions some columns but not others, the unmentioned columns contain NULL for all inserted rows.
- In Impala 1.2.1 and higher, all NULL values come at the end of the result set for ORDER BY ... ASC queries, and at the beginning of the result set for ORDER BY ... DESC queries. In effect, NULL is considered greater than all other values for sorting purposes. The original Impala behavior always put NULL values at the end, even for ORDER BY ... DESC queries. The new behavior in Impala 1.2.1 makes Impala more compatible with other popular database systems. In Impala 1.2.1 and higher, you can override or specify the sorting behavior for NULL by adding the clause NULLS FIRST or NULLS LAST at the end of the ORDER BY clause.

▪ **Note:** Because the NULLS FIRST and NULLS LAST keywords are not currently available in Hive queries, any views you create using those keywords will not be available through Hive.

- In all other contexts besides sorting with ORDER BY, comparing a NULL to anything else returns NULL, making the comparison meaningless. For example, 10 > NULL produces NULL, 10 < NULL also produces NULL, 5 BETWEEN 1 AND NULL produces NULL, and so on.

Several built-in functions serve as shorthand for evaluating expressions and returning NULL, 0, or some other substitution value depending on the expression result: ifnull(), isnull(), nvl(), nullif(), nullifzero(), and zeroifnull(). See [Impala Conditional Functions](#) on page 275 for details.

## SQL Operators

SQL operators are a class of comparison functions that are widely used within the WHERE clauses of SELECT statements.

### Arithmetic Operators

The arithmetic operators use expressions with a left-hand argument, the operator, and then (in most cases) a right-hand argument.

#### Syntax:

```
left_hand_arg binary_operator right_hand_arg
unary_operator single_arg
```

- + and -: Can be used either as unary or binary operators.
  - With unary notation, such as +5, -2.5, or -col\_name, they multiply their single numeric argument by +1 or -1. Therefore, unary + returns its argument unchanged, while unary - flips the sign of its argument. Although you can double up these operators in expressions such as ++5 (always positive) or +-2 or +-2 (both always negative), you cannot double the unary minus operator because -- is interpreted as the start of a comment. (You can use a double unary minus operator if you separate the - characters, for example with a space or parentheses.)
  - With binary notation, such as 2+2, 5-2.5, or col1 + col2, they add or subtract respectively the right-hand argument to (or from) the left-hand argument. Both arguments must be of numeric types.

- \* and /: Multiplication and division respectively. Both arguments must be of numeric types.

When multiplying, the shorter argument is promoted if necessary (such as SMALLINT to INT or BIGINT, or FLOAT to DOUBLE), and then the result is promoted again to the next larger type. Thus, multiplying a TINYINT and an INT produces a BIGINT result. Multiplying a FLOAT and a FLOAT produces a DOUBLE result. Multiplying a FLOAT and a DOUBLE or a DOUBLE and a DOUBLE produces a DECIMAL(38,17), because DECIMAL values can represent much larger and more precise values than DOUBLE.

When dividing, Impala always treats the arguments and result as DOUBLE values to avoid losing precision. If you need to insert the results of a division operation into a FLOAT column, use the CAST() function to convert the result to the correct type.

- %: Modulo operator. Returns the remainder of the left-hand argument divided by the right-hand argument. Both arguments must be of one of the integer types.
- &, |, and ^: Bitwise operators that return the logical AND, logical OR, or logical XOR (exclusive OR) of their argument values. Both arguments must be of one of the integer types. If the arguments are of different type, the argument with the smaller type is implicitly extended to match the argument with the longer type.

You can chain a sequence of arithmetic expressions, optionally grouping them with parentheses.

The arithmetic operators generally do not have equivalent calling conventions using functional notation. For example, prior to Impala 2.2.0 / CDH 5.4.0, there is no MOD() function equivalent to the % modulo operator. Conversely, there are some arithmetic functions that do not have a corresponding operator. For example, for exponentiation you use the POW() function, but there is no \*\* exponentiation operator. See [Impala Mathematical Functions](#) on page 259 for the arithmetic functions you can use.

## BETWEEN Operator

In a WHERE clause, compares an expression to both a lower and upper bound. The comparison is successful if the expression is greater than or equal to the lower bound, and less than or equal to the upper bound. If the bound values are switched, so the lower bound is greater than the upper bound, does not match any values.

### Syntax:

```
expression BETWEEN lower_bound AND upper_bound
```

**Data types:** Typically used with numeric data types. Works with any data type, although not very practical for BOOLEAN values. (BETWEEN false AND true will match all BOOLEAN values.) Use CAST() if necessary to ensure the lower and upper bound values are compatible types. Call string or date/time functions if necessary to extract or transform the relevant portion to compare, especially if the value can be transformed into a number.

### Usage notes:

Be careful when using short string operands. A longer string that starts with the upper bound value will not be included, because it is considered greater than the upper bound. For example, BETWEEN 'A' and 'M' would not match the string value 'Midway'. Use functions such as upper(), lower(), substr(), trim(), and so on if necessary to ensure the comparison works as expected.

### Examples:

```
-- Retrieve data for January through June, inclusive.
select c1 from t1 where month between 1 and 6;

-- Retrieve data for names beginning with 'A' through 'M' inclusive.
-- Only test the first letter to ensure all the values starting with 'M' are matched.
-- Do a case-insensitive comparison to match names with various capitalization
conventions.
select last_name from customers where upper(substr(last_name,1,1)) between 'A' and 'M';

-- Retrieve data for only the first week of each month.
select count(distinct visitor_id) from web_traffic where dayofmonth(when_viewed)
between 1 and 7;
```

## Comparison Operators

Impala supports the familiar comparison operators for checking equality and sort order for the column data types:

### Syntax:

```
left_hand_expression comparison_operator right_hand_expression
```

- `=, !=, <>:` apply to all types.
- `<, <=, >, >=:` apply to all types; for BOOLEAN, TRUE is considered greater than FALSE.

### Alternatives:

The `IN` and `BETWEEN` operators provide shorthand notation for expressing combinations of equality, less than, and greater than comparisons with a single operator.

Because comparing any value to `NULL` produces `NULL` rather than `TRUE` or `FALSE`, use the `IS NULL` and `IS NOT NULL` operators to check if a value is `NULL` or not.

## EXISTS Operator

The `EXISTS` operator tests whether a subquery returns any results.

The converse, `NOT EXISTS`, helps to find all the values from one table that do not have any corresponding values in another table.

### Syntax:

```
EXISTS (subquery)  
NOT EXISTS (subquery)
```

### Usage notes:

The subquery can refer to a different table than the outer query block, or the same table. For example, you might use `EXISTS` or `NOT EXISTS` to check the existence of parent/child relationships between two columns of the same table.

You can also use operators and function calls within the subquery to test for other kinds of relationships other than strict equality. For example, you might use a call to `COUNT()` in the subquery to check whether the number of matching values is higher or lower than some limit. You might call a UDF in the subquery to check whether values in one table matches a hashed representation of those same values in a different table.

### NULL considerations:

If the subquery returns any value at all (even `NULL`), `EXISTS` returns `TRUE` and `NOT EXISTS` returns `false`.

The following example shows how even when the subquery returns only `NULL` values, `EXISTS` still returns `TRUE` and thus matches all the rows from the table in the outer query block.

```
[localhost:21000] > create table all_nulls (x int);  
[localhost:21000] > insert into all_nulls values (null), (null), (null);  
[localhost:21000] > select y from t2 where exists (select x from all_nulls);  
+---+  
| y |  
+---+  
| 2 |  
| 4 |  
| 6 |  
+---+
```

However, if the table in the subquery is empty and so the subquery returns an empty result set, `EXISTS` returns `FALSE`:

```
[localhost:21000] > create table empty (x int);
[localhost:21000] > select y from t2 where exists (select x from empty);
[localhost:21000] >
```

**Added in:** CDH 5.2.0 (Impala 2.0.0)

#### Restrictions:

Correlated subqueries used in `EXISTS` and `IN` operators cannot include a `LIMIT` clause.

The `NOT EXISTS` operator requires a correlated subquery.

#### Examples:

The following examples refer to these simple tables containing small sets of integers or strings:

```
[localhost:21000] > create table t1 (x int);
[localhost:21000] > insert into t1 values (1), (2), (3), (4), (5), (6);

[localhost:21000] > create table t2 (y int);
[localhost:21000] > insert into t2 values (2), (4), (6);

[localhost:21000] > create table t3 (z int);
[localhost:21000] > insert into t3 values (1), (3), (5);

[localhost:21000] > create table month_names (m string);
[localhost:21000] > insert into month_names values
    > ('January'), ('February'), ('March'),
    > ('April'), ('May'), ('June'), ('July'),
    > ('August'), ('September'), ('October'),
    > ('November'), ('December');
```

The following example shows a correlated subquery that finds all the values in one table that exist in another table. For each value `x` from `t1`, the query checks if the `y` column of `t2` contains an identical value, and the `EXISTS` operator returns `TRUE` or `FALSE` as appropriate in each case.

```
localhost:21000] > select x from t1 where exists (select y from t2 where t1.x = y);
+---+
| x |
+---+
| 2 |
| 4 |
| 6 |
+---+
```

An uncorrelated query is less interesting in this case. Because the subquery always returns `TRUE`, all rows from `t1` are returned. If the table contents were changed so that the subquery did not match any rows, none of the rows from `t1` would be returned.

```
localhost:21000] > select x from t1 where exists (select y from t2 where y > 5);
+---+
| x |
+---+
| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| 6 |
+---+
```

The following example shows how an uncorrelated subquery can test for the existence of some condition within a table. By using `LIMIT 1` or an aggregate function, the query returns a single result or no result based on

## Impala SQL Language Reference

whether the subquery matches any rows. Here, we know that t1 and t2 contain some even numbers, but t3 does not.

```
[localhost:21000] > select "contains an even number" from t1 where exists (select x
from t1 where x % 2 = 0) limit 1;
+-----+
| 'contains an even number' |
+-----+
| contains an even number |
+-----+
[localhost:21000] > select "contains an even number" as assertion from t1 where exists
(select x from t1 where x % 2 = 0) limit 1;
+-----+
| assertion |
+-----+
| contains an even number |
+-----+
[localhost:21000] > select "contains an even number" as assertion from t2 where exists
(select x from t2 where y % 2 = 0) limit 1;
ERROR: AnalysisException: couldn't resolve column reference: 'x'
[localhost:21000] > select "contains an even number" as assertion from t2 where exists
(select y from t2 where y % 2 = 0) limit 1;
+-----+
| assertion |
+-----+
| contains an even number |
+-----+
[localhost:21000] > select "contains an even number" as assertion from t3 where exists
(select z from t3 where z % 2 = 0) limit 1;
[localhost:21000] >
```

The following example finds numbers in one table that are 1 greater than numbers from another table. The EXISTS notation is simpler than an equivalent CROSS JOIN between the tables. (The example then also illustrates how the same test could be performed using an IN operator.)

```
[localhost:21000] > select x from t1 where exists (select y from t2 where x = y + 1);
+---+
| x |
+---+
| 3 |
| 5 |
+---+
[localhost:21000] > select x from t1 where x in (select y + 1 from t2);
+---+
| x |
+---+
| 3 |
| 5 |
+---+
```

The following example finds values from one table that do not exist in another table.

```
[localhost:21000] > select x from t1 where not exists (select y from t2 where x = y);
+---+
| x |
+---+
| 1 |
| 3 |
| 5 |
+---+
```

The following example uses the NOT EXISTS operator to find all the leaf nodes in tree-structured data. This simplified “tree of life” has multiple levels (class, order, family, and so on), with each item pointing upward through a PARENT pointer. The example runs an outer query and a subquery on the same table, returning only those items whose ID value is *not* referenced by the PARENT of any other item.

```
[localhost:21000] > create table tree (id int, parent int, name string);
[localhost:21000] > insert overwrite tree values
```

```

> (0, null, "animals"),
> (1, 0, "placentals"),
> (2, 0, "marsupials"),
> (3, 1, "bats"),
> (4, 1, "cats"),
> (5, 2, "kangaroos"),
> (6, 4, "lions"),
> (7, 4, "tigers"),
> (8, 5, "red kangaroo"),
> (9, 2, "wallabies");
[localhost:21000] > select name as "leaf node" from tree one
> where not exists (select parent from tree two where one.id =
two.parent);
+-----+
| leaf node |
+-----+
| bats      |
| lions     |
| tigers    |
| red kangaroo |
| wallabies |
+-----+

```

**Related information:**[Subqueries](#) on page 235

## IN Operator

The `IN` operator compares an argument value to a set of values, and returns `TRUE` if the argument matches any value in the set. The `NOT IN` operator reverses the comparison, and checks if the argument value is not part of a set of values.

**Syntax:**

```

expression IN (expression [, expression])
expression IN (subquery)

expression NOT IN (expression [, expression])
expression NOT IN (subquery)

```

The left-hand expression and the set of comparison values must be of compatible types.

The left-hand expression must consist only of a single value, not a tuple. Although the left-hand expression is typically a column name, it could also be some other value. For example, the `WHERE` clauses `WHERE id IN (5)` and `WHERE 5 IN (id)` produce the same results.

The set of values to check against can be specified as constants, function calls, column names, or other expressions in the query text. When the values are listed explicitly, the maximum number of expressions is 10,000.

In Impala 2.0 and higher, the set of values can also be generated by a subquery. `IN` can evaluate an unlimited number of results using a subquery.

**Usage notes:**

Any expression using the `IN` operator could be rewritten as a series of equality tests connected with `OR`, but the `IN` syntax is often clearer, more concise, and easier for Impala to optimize. For example, with partitioned tables, queries frequently use `IN` clauses to filter data by comparing the partition key columns to specific values.

**NULL considerations:**

If there really is a matching non-null value, `IN` returns `TRUE`:

```

[localhost:21000] > select 1 in (1,null,2,3);
+-----+
| 1 in (1, null, 2, 3) |
+-----+

```

```
| true
+-----+
[localhost:21000] > select 1 not in (1,null,2,3);
+-----+
| 1 not in (1, null, 2, 3) |
+-----+
| false
+-----+
```

If the searched value is not found in the comparison values, and the comparison values include `NULL`, the result is `NULL`:

```
[localhost:21000] > select 5 in (1,null,2,3);
+-----+
| 5 in (1, null, 2, 3) |
+-----+
| NULL
+-----+
[localhost:21000] > select 5 not in (1,null,2,3);
+-----+
| 5 not in (1, null, 2, 3) |
+-----+
| NULL
+-----+
[localhost:21000] > select 1 in (null);
+-----+
| 1 in (null) |
+-----+
| NULL
+-----+
[localhost:21000] > select 1 not in (null);
+-----+
| 1 not in (null) |
+-----+
| NULL
+-----+
```

If the left-hand argument is `NULL`, `IN` always returns `NULL`. This rule applies even if the comparison values include `NULL`.

```
[localhost:21000] > select null in (1,2,3);
+-----+
| null in (1, 2, 3) |
+-----+
| NULL
+-----+
[localhost:21000] > select null not in (1,2,3);
+-----+
| null not in (1, 2, 3) |
+-----+
| NULL
+-----+
[localhost:21000] > select null in (null);
+-----+
| null in (null) |
+-----+
| NULL
+-----+
[localhost:21000] > select null not in (null);
+-----+
| null not in (null) |
+-----+
| NULL
+-----+
```

**Added in:** Available in earlier Impala releases, but new capabilities were added in CDH 5.2.0 / Impala 2.0.0

### Restrictions:

Correlated subqueries used in `EXISTS` and `IN` operators cannot include a `LIMIT` clause.

**Examples:**

```
-- Using IN is concise and self-documenting.
SELECT * FROM t1 WHERE c1 IN (1,2,10);
-- Equivalent to series of = comparisons ORed together.
SELECT * FROM t1 WHERE c1 = 1 OR c1 = 2 OR c1 = 10;

SELECT c1 AS "starts with vowel" FROM t2 WHERE upper(substr(c1,1,1)) IN
('A','E','I','O','U');

SELECT COUNT(DISTINCT(visitor_id)) FROM web_traffic WHERE month IN
('January','June','July');
```

**Related information:**[Subqueries](#) on page 235

## IS NULL Operator

The `IS NULL` operator, and its converse the `IS NOT NULL` operator, test whether a specified value is `NULL`. Because using `NULL` with any of the other comparison operators such as `=` or `!=` also returns `NULL` rather than `TRUE` or `FALSE`, you use a special-purpose comparison operator to check for this special condition.

**Syntax:**

```
expression IS NULL
expression IS NOT NULL
```

**Usage notes:**

In many cases, `NULL` values indicate some incorrect or incomplete processing during data ingestion or conversion. You might check whether any values in a column are `NULL`, and if so take some followup action to fill them in.

With sparse data, often represented in “wide” tables, it is common for most values to be `NULL` with only an occasional non-`NULL` value. In those cases, you can use the `IS NOT NULL` operator to identify the rows containing any data at all for a particular column, regardless of the actual value.

With a well-designed database schema, effective use of `NULL` values and `IS NULL` and `IS NOT NULL` operators can save having to design custom logic around special values such as `0`, `-1`, `'N/A'`, empty string, and so on. `NULL` lets you distinguish between a value that is known to be `0`, `false`, or empty, and a truly unknown value.

**Examples:**

```
-- If this value is non-zero, something is wrong.
select count(*) from employees where employee_id is null;

-- With data from disparate sources, some fields might be blank.
-- Not necessarily an error condition.
select count(*) from census where household_income is null;

-- Sometimes we expect fields to be null, and followup action
-- is needed when they are not.
select count(*) from web_traffic where weird_http_code is not null;
```

## LIKE Operator

A comparison operator for `STRING` data, with basic wildcard capability using `_` to match a single character and `%` to match multiple characters. The argument expression must match the entire string value. Typically, it is more efficient to put any `%` wildcard match at the end of the string.

**Syntax:**

```
string_expression LIKE wildcard_expression
string_expression NOT LIKE wildcard_expression
```

## Examples:

```
select distinct c_last_name from customer where c_last_name like 'Mc%' or c_last_name like 'Mac%';
select count(c_last_name) from customer where c_last_name like 'M%';
select c_email_address from customer where c_email_address like '%.edu';

-- We can find 4-letter names beginning with 'M' by calling functions...
select distinct c_last_name from customer where length(c_last_name) = 4 and substr(c_last_name,1,1) = 'M';
-- ...or in a more readable way by matching M followed by exactly 3 characters.
select distinct c_last_name from customer where c_last_name like 'M___';
```

For a more general kind of search operator using regular expressions, see [REGEXP Operator](#) on page 152.

## Logical Operators

Logical operators return a `BOOLEAN` value, based on a binary or unary logical operation between arguments that are also Booleans. Typically, the argument expressions use [comparison operators](#).

### Syntax:

```
boolean_expression binary_logical_operator boolean_expression  
unary_logical_operator boolean_expression
```

The Impala logical operators are:

- `AND`: A binary operator that returns `true` if its left-hand and right-hand arguments both evaluate to `true`, `NULL` if either argument is `NULL`, and `false` otherwise.
- `OR`: A binary operator that returns `true` if either of its left-hand and right-hand arguments evaluate to `true`, `NULL` if one argument is `NULL` and the other is either `NULL` or `false`, and `false` otherwise.
- `NOT`: A unary operator that flips the state of a Boolean expression from `true` to `false`, or `false` to `true`. If the argument expression is `NULL`, the result remains `NULL`. (When `NOT` is used this way as a unary logical operator, it works differently than the `IS NOT NULL` comparison operator, which returns `true` when applied to a `NULL`.)

### Examples:

These examples demonstrate the `AND` operator:

```
[localhost:21000] > select true and true;
+-----+
| true and true |
+-----+
| true           |
+-----+
[localhost:21000] > select true and false;
+-----+
| true and false |
+-----+
| false          |
+-----+
[localhost:21000] > select false and false;
+-----+
| false and false |
+-----+
| false           |
+-----+
[localhost:21000] > select true and null;
+-----+
| true and null |
+-----+
| NULL           |
+-----+
[localhost:21000] > select (10 > 2) and (6 != 9);
+-----+
| (10 > 2) and (6 != 9) |
+-----+
```

```
+-----+
| true |
+-----+
```

These examples demonstrate the OR operator:

```
[localhost:21000] > select true or true;
+-----+
| true or true |
+-----+
| true |
+-----+
[localhost:21000] > select true or false;
+-----+
| true or false |
+-----+
| true |
+-----+
[localhost:21000] > select false or false;
+-----+
| false or false |
+-----+
| false |
+-----+
[localhost:21000] > select true or null;
+-----+
| true or null |
+-----+
| true |
+-----+
[localhost:21000] > select null or true;
+-----+
| null or true |
+-----+
| true |
+-----+
[localhost:21000] > select false or null;
+-----+
| false or null |
+-----+
| NULL |
+-----+
[localhost:21000] > select (1 = 1) or ('hello' = 'world');
+-----+
| (1 = 1) or ('hello' = 'world') |
+-----+
| true |
+-----+
[localhost:21000] > select (2 + 2 != 4) or (-1 > 0);
+-----+
| (2 + 2 != 4) or (-1 > 0) |
+-----+
| false |
+-----+
```

These examples demonstrate the NOT operator:

```
[localhost:21000] > select not true;
+-----+
| not true |
+-----+
| false |
+-----+
[localhost:21000] > select not false;
+-----+
| not false |
+-----+
| true |
+-----+
[localhost:21000] > select not null;
+-----+
```

```
| not null |
+-----+
| NULL   |
+-----+
[localhost:21000] > select not (1=1);
+-----+
| not (1 = 1) |
+-----+
| false        |
+-----+
```

## REGEXP Operator

Tests whether a value matches a regular expression. Uses the POSIX regular expression syntax where ^ and \$ match the beginning and end of the string, . represents any single character, \* represents a sequence of zero or more items, + represents a sequence of one or more items, ? produces a non-greedy match, and so on.

### Syntax:

```
string_expression REGEXP regular_expression
```

### Usage notes:

The regular expression must match the entire value, not just occur somewhere inside it. Use . \* at the beginning and/or the end if you only need to match characters anywhere in the middle. Thus, the ^ and \$ atoms are often redundant, although you might already have them in your expression strings that you reuse from elsewhere.

The RLIKE operator is a synonym for REGEXP.

The | symbol is the alternation operator, typically used within () to match different sequences. The () groups do not allow backreferences. To retrieve the part of a value matched within a () section, use the [regexp\\_extract\(\)](#) built-in function.

#### Note:

In Impala 1.3.1 and higher, the REGEXP and RLIKE operators now match a regular expression string that occurs anywhere inside the target string, the same as if the regular expression was enclosed on each side by . \*. See [REGEXP Operator](#) on page 152 for examples. Previously, these operators only succeeded when the regular expression matched the entire target string. This change improves compatibility with the regular expression support for popular database systems. There is no change to the behavior of the `regexp_extract()` and `regexp_replace()` built-in functions.

In Impala 2.0 and later, the Impala regular expression syntax conforms to the POSIX Extended Regular Expression syntax used by the Google RE2 library. For details, see [the RE2 documentation](#). It has most idioms familiar from regular expressions in Perl, Python, and so on, including . \*? for non-greedy matches.

In Impala 2.0 and later, a change in the underlying regular expression library could cause changes in the way regular expressions are interpreted by this function. Test any queries that use regular expressions and adjust the expression patterns if necessary. See [Incompatible Changes Introduced in Impala 2.0.0 / CDH 5.2.0](#) on page 489 for details.

### Examples:

The following examples demonstrate the identical syntax for the REGEXP and RLIKE operators.

```
-- Find all customers whose first name starts with 'J', followed by 0 or more of any character.
select c_first_name, c_last_name from customer where c_first_name regexp '^J.*';
select c_first_name, c_last_name from customer where c_first_name rlike '^J.*';

-- Find 'Macdonald', where the first 'a' is optional and the 'D' can be upper- or lowercase.
```

```
-- The ^...$ are required, to match the start and end of the value.
select c_first_name, c_last_name from customer where c_last_name regexp
'^Ma?c[Dd]onald$';
select c_first_name, c_last_name from customer where c_last_name rlike '^Ma?c[Dd]onald$';

-- Match multiple character sequences, either 'Mac' or 'Mc'.
select c_first_name, c_last_name from customer where c_last_name regexp
'^^(Mac|Mc)onald$';
select c_first_name, c_last_name from customer where c_last_name rlike
'^^(Mac|Mc)onald$';

-- Find names starting with 'S', then one or more vowels, then 'r', then any other
characters.
-- Matches 'Searcy', 'Sorenson', 'Sauer'.
select c_first_name, c_last_name from customer where c_last_name regexp '^S[aeiou]+r.*$';
select c_first_name, c_last_name from customer where c_last_name rlike '^S[aeiou]+r.*$';

-- Find names that end with 2 or more vowels: letters from the set a,e,i,o,u.
select c_first_name, c_last_name from customer where c_last_name regexp '.*[aeiou]{2,}$$';
select c_first_name, c_last_name from customer where c_last_name rlike '.*[aeiou]{2,}$$';

-- You can use letter ranges in the [] blocks, for example to find names starting with
A, B, or C.
select c_first_name, c_last_name from customer where c_last_name regexp '^[[A-C].*';
select c_first_name, c_last_name from customer where c_last_name rlike '^[[A-C].*';

-- If you are not sure about case, leading/trailing spaces, and so on, you can process
the
-- column using string functions first.
select c_first_name, c_last_name from customer where lower(trim(c_last_name)) regexp
'^de.*';
select c_first_name, c_last_name from customer where lower(trim(c_last_name)) rlike
'^de.*';
```

## RLIKE Operator

Synonym for the `REGEXP` operator. See [REGEXP Operator](#) on page 152 for details.

### Examples:

The following examples demonstrate the identical syntax for the `REGEXP` and `RLIKE` operators.

```
-- Find all customers whose first name starts with 'J', followed by 0 or more of any
character.
select c_first_name, c_last_name from customer where c_first_name regexp '^J.*';
select c_first_name, c_last_name from customer where c_first_name rlike '^J.*';

-- Find 'Macdonald', where the first 'a' is optional and the 'D' can be upper- or
lowercase.
-- The ^...$ are required, to match the start and end of the value.
select c_first_name, c_last_name from customer where c_last_name regexp
'^Ma?c[Dd]onald$';
select c_first_name, c_last_name from customer where c_last_name rlike '^Ma?c[Dd]onald$';

-- Match multiple character sequences, either 'Mac' or 'Mc'.
select c_first_name, c_last_name from customer where c_last_name regexp
'^^(Mac|Mc)onald$';
select c_first_name, c_last_name from customer where c_last_name rlike
'^^(Mac|Mc)onald$';

-- Find names starting with 'S', then one or more vowels, then 'r', then any other
characters.
-- Matches 'Searcy', 'Sorenson', 'Sauer'.
select c_first_name, c_last_name from customer where c_last_name regexp '^S[aeiou]+r.*$';
select c_first_name, c_last_name from customer where c_last_name rlike '^S[aeiou]+r.*$';

-- Find names that end with 2 or more vowels: letters from the set a,e,i,o,u.
select c_first_name, c_last_name from customer where c_last_name regexp '.*[aeiou]{2,}$$';
select c_first_name, c_last_name from customer where c_last_name rlike '.*[aeiou]{2,}$$';

-- You can use letter ranges in the [] blocks, for example to find names starting with
```

```
A, B, or C.  
select c_first_name, c_last_name from customer where c_last_name regexp '^[A-C].*';  
select c_first_name, c_last_name from customer where c_last_name rlike '^*[A-C].*';  
  
-- If you are not sure about case, leading/trailing spaces, and so on, you can process  
the  
-- column using string functions first.  
select c_first_name, c_last_name from customer where lower(trim(c_last_name)) regexp  
'^de.*';  
select c_first_name, c_last_name from customer where lower(trim(c_last_name)) rlike  
'^de.*';
```

## Schema Objects and Object Names

With Impala, you work schema objects that are familiar to database users: primarily databases, tables, views, and functions. The SQL syntax to work with these objects is explained in [SQL Statements](#) on page 162. This section explains the conceptual knowledge you need to work with these objects and the various ways to specify their names.

Within a table, partitions can also be considered a kind of object. Partitioning is an important subject for Impala, with its own documentation section covering use cases and performance considerations. See [Partitioning for Impala Tables](#) on page 394 for details.

Impala does not have a counterpart of the “tablespace” notion from some database systems. By default, all the data files for a database, table, or partition are located within nested folders within the HDFS file system. You can also specify a particular HDFS location for a given Impala table or partition. The raw data for these objects is represented as a collection of data files, providing the flexibility to load data by simply moving files into the expected HDFS location.

Information about the schema objects is held in the [metastore](#) database. This database is shared between Impala and Hive, allowing each to create, drop, and query each other's databases, tables, and so on. When Impala makes a change to schema objects through a `CREATE`, `ALTER`, `DROP`, `INSERT`, or `LOAD DATA` statement, it broadcasts those changes to all nodes in the cluster through the [catalog service](#). When you make such changes through Hive or directly through manipulating HDFS files, you use the [REFRESH](#) or [INVALIDATE METADATA](#) statements on the Impala side to recognize the newly loaded data, new tables, and so on.

## Aliases

When you write the names of tables, columns, or column expressions in a query, you can assign an alias at the same time. Then you can specify the alias rather than the original name when making other references to the table or column in the same statement. You typically specify aliases that are shorter, easier to remember, or both than the original names. The aliases are printed in the query header, making them useful for self-documenting output.

To set up an alias, add the `AS alias` clause immediately after any table, column, or expression name in the `SELECT` list or `FROM` list of a query. The `AS` keyword is optional; you can also specify the alias immediately after the original name.

To use an alias name that matches one of the Impala reserved keywords (listed in [Impala Reserved Words](#) on page 453), surround the identifier with either single or double quotation marks, or `` characters (backticks).

```
select c1 as name, c2 as address, c3 as phone from table_with_terse_columns;  
select sum(ss_xyz_dollars_net) as total_sales from table_with_cryptic_columns;  
select one.name, two.address, three.phone from  
  census one, building_directory two, phonebook three  
  where one.id = two.id and two.id = three.id;
```

Aliases follow the same rules as identifiers when it comes to case insensitivity. Aliases can be longer than identifiers (up to the maximum length of a Java string) and can include additional characters such as spaces and dashes when they are quoted using backtick characters.

**Alternatives:**

Another way to define different names for the same tables or columns is to create views. See [Views](#) on page 158 for details.

## Databases

In Impala, a database is a logical container for a group of tables. Each database defines a separate namespace. Within a database, you can refer to the tables inside it using their unqualified names. Different databases can contain tables with identical names.

Creating a database is a lightweight operation. There are no database-specific properties to configure. Therefore, there is no `ALTER DATABASE`

Typically, you create a separate database for each project or application, to avoid naming conflicts between tables and to make clear which tables are related to each other.

Each database is physically represented by a directory in HDFS.

There is a special database, named `default`, where you begin when you connect to Impala. Tables created in `default` are physically located one level higher in HDFS than all the user-created databases.

Impala includes another predefined database, `_impala_builtins`, that serves as the location for the [built-in functions](#). To see the built-in functions, use a statement like the following:

```
show functions in _impala_builtins;
show functions in _impala_builtins like '*substring*';
```

**Related statements:**

[CREATE DATABASE Statement](#) on page 175, [DROP DATABASE Statement](#) on page 192, [USE Statement](#) on page 256, [SHOW DATABASES](#) on page 248

## Functions

Functions let you apply arithmetic, string, or other computations and transformations to Impala data. You typically use them in `SELECT` lists and `WHERE` clauses to filter and format query results so that the result set is exactly what you want, with no further processing needed on the application side.

Scalar functions return a single result for each input row. See [Built-in Functions](#) on page 257.

Aggregate functions combine the results from multiple rows. See [Impala Aggregate Functions](#) on page 284.

User-defined functions let you code your own logic. They can be either scalar or aggregate functions. See [Impala User-Defined Functions \(UDFs\)](#) on page 308.

**Related statements:** [CREATE FUNCTION Statement](#) on page 177, [DROP FUNCTION Statement](#) on page 192

## Identifiers

Identifiers are the names of databases, tables, or columns that you specify in a SQL statement. The rules for identifiers govern what names you can give to things you create, the notation for referring to names containing unusual characters, and other aspects such as case sensitivity.

- The minimum length of an identifier is 1 character.
- The maximum length of an identifier is currently 128 characters, enforced by the metastore database.
- An identifier must start with an alphabetic character. The remainder can contain any combination of alphanumeric characters and underscores. Quoting the identifier with backticks has no effect on the allowed characters in the name.
- An identifier can contain only ASCII characters.

- To use an identifier name that matches one of the Impala reserved keywords (listed in [Impala Reserved Words](#) on page 453), surround the identifier with `` characters (backticks). Quote the reserved word even if it is part of a fully qualified name. The following example shows how a reserved word can be used as a column name if it is quoted with backticks in the CREATE TABLE statement, and how the column name must also be quoted with backticks in a query:

```
[localhost:21000] > create table reserved (`data` string);

[localhost:21000] > select data from reserved;
ERROR: AnalysisException: Syntax error in line 1:
  select data from reserved
    ^
Encountered: DATA
Expected: ALL, CASE, CAST, DISTINCT, EXISTS, FALSE, IF, INTERVAL, NOT, NULL,
STRAIGHT_JOIN, TRUE, IDENTIFIER
 CAUSED BY: Exception: Syntax error

[localhost:21000] > select reserved.data from reserved;
ERROR: AnalysisException: Syntax error in line 1:
  select reserved.data from reserved
    ^
Encountered: DATA
Expected: IDENTIFIER
 CAUSED BY: Exception: Syntax error

[localhost:21000] > select reserved.`data` from reserved;

[localhost:21000] >
```

**Important:** Because the list of reserved words grows over time as new SQL syntax is added, consider adopting coding conventions (especially for any automated scripts or in packaged applications) to always quote all identifiers with backticks. Quoting all identifiers protects your SQL from compatibility issues if new reserved words are added in later releases.

- Impala identifiers are always case-insensitive. That is, tables named `t1` and `T1` always refer to the same table, regardless of quote characters. Internally, Impala always folds all specified table and column names to lowercase. This is why the column headers in query output are always displayed in lowercase.

See [Aliases](#) on page 154 for how to define shorter or easier-to-remember aliases if the original names are long or cryptic identifiers. Aliases follow the same rules as identifiers when it comes to case insensitivity. Aliases can be longer than identifiers (up to the maximum length of a Java string) and can include additional characters such as spaces and dashes when they are quoted using backtick characters.

Another way to define different names for the same tables or columns is to create views. See [Views](#) on page 158 for details.

## Tables

Tables are the primary containers for data in Impala. They have the familiar row and column layout similar to other database systems, plus some features such as partitioning often associated with higher-end data warehouse systems.

Logically, each table has a structure based on the definition of its columns, partitions, and other properties.

Physically, each table that uses HDFS storage is associated with a directory in HDFS. The table data consists of all the data files underneath that directory:

- [Internal tables](#), managed by Impala, use directories inside the designated Impala work area.
- [External tables](#) use arbitrary HDFS directories, where the data files are typically shared between different Hadoop components.
- Large-scale data is usually handled by partitioned tables, where the data files are divided among different HDFS subdirectories.

Impala tables can also represent data that is stored in HBase, or in the Amazon S3 filesystem. See [Using Impala to Query HBase Tables](#) on page 426 and [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about those special kinds of tables.

Impala queries ignore files with extensions commonly used for temporary work files by Hadoop tools. Any files with extensions .tmp or .copying are not considered part of the Impala table. The suffix matching is case-insensitive, so for example Impala ignores both .copying and .COPYING suffixes.

**Related statements:** [CREATE TABLE Statement](#) on page 180, [DROP TABLE Statement](#) on page 197, [ALTER TABLE Statement](#) on page 164, [INSERT Statement](#) on page 202, [LOAD DATA Statement](#) on page 212, [SELECT Statement](#) on page 218

## Internal Tables

The default kind of table produced by the `CREATE TABLE` statement is known as an internal table. (Its counterpart is the external table, produced by the `CREATE EXTERNAL TABLE` syntax.)

- Impala creates a directory in HDFS to hold the data files.
- You can create data in internal tables by issuing `INSERT` or `LOAD DATA` statements.
- If you add or replace data using HDFS operations, issue the `REFRESH` command in `impala-shell` so that Impala recognizes the changes in data files, block locations, and so on.
- When you issue a `DROP TABLE` statement, Impala physically removes all the data files from the directory.
- To see whether a table is internal or external, and its associated HDFS location, issue the statement `DESCRIBE FORMATTED table_name`. The `Table Type` field displays `MANAGED_TABLE` for internal tables and `EXTERNAL_TABLE` for external tables. The `Location` field displays the path of the table directory as an HDFS URI.
- When you issue an `ALTER TABLE` statement to rename an internal table, all data files are moved into the new HDFS directory for the table. The files are moved even if they were formerly in a directory outside the Impala data directory, for example in an internal table with a `LOCATION` attribute pointing to an outside HDFS directory.

### Examples:

You can switch a table from internal to external, or from external to internal, by using the `ALTER TABLE` statement:

```
-- Switch a table from internal to external.
ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='TRUE');

-- Switch a table from external to internal.
ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='FALSE');
```

### Related information:

[External Tables](#) on page 157, [CREATE TABLE Statement](#) on page 180, [DROP TABLE Statement](#) on page 197, [ALTER TABLE Statement](#) on page 164, [DESCRIBE Statement](#) on page 188

## External Tables

The syntax `CREATE EXTERNAL TABLE` sets up an Impala table that points at existing data files, potentially in HDFS locations outside the normal Impala data directories.. This operation saves the expense of importing the data into a new table when you already have the data files in a known location in HDFS, in the desired file format.

- You can use Impala to query the data in this table.
- You can create data in external tables by issuing `INSERT` or `LOAD DATA` statements.
- If you add or replace data using HDFS operations, issue the `REFRESH` command in `impala-shell` so that Impala recognizes the changes in data files, block locations, and so on.

## Impala SQL Language Reference

- When you issue a `DROP TABLE` statement in Impala, that removes the connection that Impala has with the associated data files, but does not physically remove the underlying data. You can continue to use the data files with other Hadoop components and HDFS operations.
- To see whether a table is internal or external, and its associated HDFS location, issue the statement `DESCRIBE FORMATTED table_name`. The `Table Type` field displays `MANAGED_TABLE` for internal tables and `EXTERNAL_TABLE` for external tables. The `Location` field displays the path of the table directory as an HDFS URI.
- When you issue an `ALTER TABLE` statement to rename an external table, all data files are left in their original locations.
- You can point multiple external tables at the same HDFS directory by using the same `LOCATION` attribute for each one. The tables could have different column definitions, as long as the number and types of columns are compatible with the schema evolution considerations for the underlying file type. For example, for text data files, one table might define a certain column as a `STRING` while another defines the same column as a `BIGINT`.

### Examples:

You can switch a table from internal to external, or from external to internal, by using the `ALTER TABLE` statement:

```
-- Switch a table from internal to external.  
ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='TRUE') ;  
  
-- Switch a table from external to internal.  
ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='FALSE') ;
```

### Related information:

[Internal Tables](#) on page 157, [CREATE TABLE Statement](#) on page 180, [DROP TABLE Statement](#) on page 197, [ALTER TABLE Statement](#) on page 164, [DESCRIBE Statement](#) on page 188

## Views

Views are lightweight logical constructs that act as aliases for queries. You can specify a view name in a query (a `SELECT` statement or the `SELECT` portion of an `INSERT` statement) where you would usually specify a table name.

A view lets you:

- Set up fine-grained security where a user can query some columns from a table but not other columns. See [Controlling Access at the Column Level through Views](#) on page 97 for details.
- Issue complicated queries with compact and simple syntax:

```
-- Take a complicated reporting query, plug it into a CREATE VIEW statement...  
create view v1 as select c1, c2, avg(c3) from t1 group by c3 order by c1 desc limit  
10;  
-- ... and now you can produce the report with 1 line of code.  
select * from v1;
```

- Reduce maintenance, by avoiding the duplication of complicated queries across multiple applications in multiple languages:

```
create view v2 as select t1.c1, t1.c2, t2.c3 from t1 join t2 on (t1.id = t2.id);  
-- This simple query is safer to embed in reporting applications than the longer  
query above.  
-- The view definition can remain stable even if the structure of the underlying  
tables changes.  
select c1, c2, c3 from v2;
```

- Build a new, more refined query on top of the original query by adding new clauses, select-list expressions, function calls, and so on:

```
create view average_price_by_category as select category, avg(price) as avg_price
  from products group by category;
create view expensive_categories as select category, avg_price from
average_price_by_category order by avg_price desc limit 10000;
create view top_10_expensive_categories as select category, avg_price from
expensive_categories limit 10;
```

This technique lets you build up several more or less granular variations of the same query, and switch between them when appropriate.

- Set up aliases with intuitive names for tables, columns, result sets from joins, and so on:

```
-- The original tables might have cryptic names inherited from a legacy system.
create view action_items as select rrptsk as assignee, treq as due_date, dmisc as
notes from vxv_t1_br;
-- You can leave original names for compatibility, build new applications using
more intuitive ones.
select assignee, due_date, notes from action_items;
```

- Swap tables with others that use different file formats, partitioning schemes, and so on without any downtime for data copying or conversion:

```
create table slow (x int, s string) stored as textfile;
create view report as select s from slow where x between 20 and 30;
-- Query is kind of slow due to inefficient table definition, but it works.
select * from report;

create table fast (s string) partitioned by (x int) stored as parquet;
-- ...Copy data from SLOW to FAST. Queries against REPORT view continue to work...

-- After changing the view definition, queries will be faster due to partitioning,
-- binary format, and compression in the new table.
alter view report as select s from fast where x between 20 and 30;
select * from report;
```

- Avoid coding lengthy subqueries and repeating the same subquery text in many other queries.

The SQL statements that configure views are [CREATE VIEW Statement](#) on page 186, [ALTER VIEW Statement](#) on page 168, and [DROP VIEW Statement](#) on page 198. You can specify view names when querying data ([SELECT Statement](#) on page 218) and copying data from one table to another ([INSERT Statement](#) on page 202). The [WITH](#) clause creates an inline view, that only exists for the duration of a single query.

```
[localhost:21000] > create view trivial as select * from customer;
[localhost:21000] > create view some_columns as select c_first_name, c_last_name,
c_login from customer;
[localhost:21000] > select * from some_columns limit 5;
Query finished, fetching results ...
+-----+-----+
| c_first_name | c_last_name | c_login |
+-----+-----+
| Javier       | Lewis        |          |
| Amy          | Moses         |          |
| Latisha      | Hamilton     |          |
| Michael      | White         |          |
| Robert       | Moran         |          |
+-----+-----+
[localhost:21000] > create view ordered_results as select * from some_columns order by
c_last_name desc, c_first_name desc limit 1000;
[localhost:21000] > select * from ordered_results limit 5;
Query: select * from ordered_results limit 5
Query finished, fetching results ...
+-----+-----+
| c_first_name | c_last_name | c_login |
+-----+-----+
| Thomas       | Zuniga       |          |
| Sarah         | Zuniga       |          |
+-----+-----+
```

Norma	Zuniga		
Lloyd	Zuniga		
Lisa	Zuniga		

Returned 5 row(s) in 0.48s

The previous example uses descending order for ORDERED\_RESULTS because in the sample TPCD-H data, there are some rows with empty strings for both C\_FIRST\_NAME and C\_LAST\_NAME, making the lowest-ordered names unuseful in a sample query.

```
create view visitors_by_day as select day, count(distinct visitors) as howmany from
web_traffic group by day;
create view top_10_days as select day, howmany from visitors_by_day order by howmany
limit 10;
select * from top_10_days;
```

### Usage notes:

To see the definition of a view, issue a DESCRIBE FORMATTED statement, which shows the query from the original CREATE VIEW statement:

```
[localhost:21000] > create view v1 as select * from t1;
[localhost:21000] > describe formatted v1;
Query finished, fetching results ...
+-----+-----+-----+
| name | type | comment |
+-----+-----+-----+
# col_name | data_type | comment
x | int | None
y | int | None
s | string | None
# Detailed Table Information | NULL | NULL
Database: | views | NULL
Owner: | cloudera | NULL
CreateTime: | Mon Jul 08 15:56:27 EDT 2013 | NULL
LastAccessTime: | UNKNOWN | NULL
Protect Mode: | None | NULL
Retention: | 0 | NULL
Table Type: | VIRTUAL_VIEW | NULL
Table Parameters:
transient_lastDdlTime | 1373313387 | NULL
# Storage Information | NULL | NULL
SerDe Library: | null | NULL
InputFormat: | null | NULL
OutputFormat: | null | NULL
```

Compressed:	No	NULL
Num Buckets:	0	NULL
Bucket Columns:	[ ]	NULL
Sort Columns:	[ ]	NULL
	NULL	NULL
# View Information	NULL	NULL
View Original Text:	SELECT * FROM t1	NULL
View Expanded Text:	SELECT * FROM t1	NULL

Prior to Impala 1.4.0, it was not possible to use the `CREATE TABLE LIKE view_name` syntax. In Impala 1.4.0 and higher, you can create a table with the same column definitions as a view using the `CREATE TABLE LIKE` technique. Although `CREATE TABLE LIKE` normally inherits the file format of the original table, a view has no underlying file format, so `CREATE TABLE LIKE view_name` produces a text table by default. To specify a different file format, include a `STORED AS file_format` clause at the end of the `CREATE TABLE LIKE` statement.

#### Restrictions:

- You cannot insert into an Impala view. (In some database systems, this operation is allowed and inserts rows into the base table.) You can use a view name on the right-hand side of an `INSERT` statement, in the `SELECT` part.
- If a view applies to a partitioned table, any partition pruning is determined by the clauses in the original query. Impala does not prune additional columns if the query on the view includes extra `WHERE` clauses referencing the partition key columns.
- An `ORDER BY` clause without an additional `LIMIT` clause is ignored in any view definition. If you need to sort the entire result set from a view, use an `ORDER BY` clause in the `SELECT` statement that queries the view. You can still make a simple “top 10” report by combining the `ORDER BY` and `LIMIT` clauses in the same view definition:

```
[localhost:21000] > create table unsorted (x bigint);
[localhost:21000] > insert into unsorted values (1), (9), (3), (7), (5), (8), (4),
(6), (2);
[localhost:21000] > create view sorted_view as select x from unsorted order by x;
[localhost:21000] > select x from sorted_view; -- ORDER BY clause in view has no
effect.
+---+
| x |
+---+
| 1 |
| 9 |
| 3 |
| 7 |
| 5 |
| 8 |
| 4 |
| 6 |
| 2 |
+---+
[localhost:21000] > select x from sorted_view order by x; -- View query requires
ORDER BY at outermost level.
+---+
| x |
+---+
| 1 |
| 2 |
| 3 |
| 4 |
```

```
5  
6  
7  
8  
9  
+---+  
[localhost:21000] > create view top_3_view as select x from unsorted order by x  
limit 3;  
[localhost:21000] > select x from top_3_view; -- ORDER BY and LIMIT together in  
view definition are preserved.  
+---+  
| x |  
+---+  
| 1 |  
| 2 |  
| 3 |  
+---+
```

**Related statements:** [CREATE VIEW Statement](#) on page 186, [ALTER VIEW Statement](#) on page 168, [DROP VIEW Statement](#) on page 198

## SQL Statements

The Impala SQL dialect supports a range of standard elements, plus some extensions for Big Data use cases related to data loading and data warehousing.

▪ **Note:**

In the `impala-shell` interpreter, a semicolon at the end of each statement is required. Since the semicolon is not actually part of the SQL syntax, we do not include it in the syntax definition of each statement, but we do show it in examples intended to be run in `impala-shell`.

### DDL Statements

DDL refers to “Data Definition Language”, a subset of SQL statements that change the structure of the database schema in some way, typically by creating, deleting, or modifying schema objects such as databases, tables, and views. Most Impala DDL statements start with the keywords `CREATE`, `DROP`, or `ALTER`.

The Impala DDL statements are:

- [ALTER TABLE Statement](#) on page 164
- [ALTER VIEW Statement](#) on page 168
- [COMPUTE STATS Statement](#) on page 170
- [CREATE DATABASE Statement](#) on page 175
- [CREATE FUNCTION Statement](#) on page 177
- [CREATE ROLE Statement \(CDH 5.2 or higher only\)](#) on page 179
- [CREATE TABLE Statement](#) on page 180
- [CREATE VIEW Statement](#) on page 186
- [DROP DATABASE Statement](#) on page 192
- [DROP FUNCTION Statement](#) on page 192
- [DROP ROLE Statement \(CDH 5.2 or higher only\)](#) on page 193
- [DROP TABLE Statement](#) on page 197
- [DROP VIEW Statement](#) on page 198
- [GRANT Statement \(CDH 5.2 or higher only\)](#) on page 201
- [REVOKE Statement \(CDH 5.2 or higher only\)](#) on page 217

After Impala executes a DDL command, information about available tables, columns, views, partitions, and so on is automatically synchronized between all the Impala nodes in a cluster. (Prior to Impala 1.2, you had to issue a `REFRESH` or `INVALIDATE METADATA` statement manually on the other nodes to make them aware of the changes.)

If the timing of metadata updates is significant, for example if you use round-robin scheduling where each query could be issued through a different Impala node, you can enable the `SYNC_DDL` query option to make the DDL statement wait until all nodes have been notified about the metadata changes.

See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about how Impala DDL statements interact with tables and partitions stored in the Amazon S3 filesystem.

Although the `INSERT` statement is officially classified as a DML (data manipulation language) statement, it also involves metadata changes that must be broadcast to all Impala nodes, and so is also affected by the `SYNC_DDL` query option.

Because the `SYNC_DDL` query option makes each DDL operation take longer than normal, you might only enable it before the last DDL operation in a sequence. For example, if you are running a script that issues multiple of DDL operations to set up an entire new schema, add several new partitions, and so on, you might minimize the performance overhead by enabling the query option only before the last `CREATE`, `DROP`, `ALTER`, or `INSERT` statement. The script only finishes when all the relevant metadata changes are recognized by all the Impala nodes, so you could connect to any node and issue queries through it.

The classification of DDL, DML, and other statements is not necessarily the same between Impala and Hive. Impala organizes these statements in a way intended to be familiar to people familiar with relational databases or data warehouse products. Statements that modify the metastore database, such as `COMPUTE STATS`, are classified as DDL. Statements that only query the metastore database, such as `SHOW` or `DESCRIBE`, are put into a separate category of utility statements.

- **Note:** The query types shown in the Impala debug web user interface might not match exactly the categories listed here. For example, currently the `USE` statement is shown as DDL in the debug web UI. The query types shown in the debug web UI are subject to change, for improved consistency.

#### Related information:

The other major classifications of SQL statements are data manipulation language (see [DML Statements](#) on page 163) and queries (see [SELECT Statement](#) on page 218).

## DML Statements

DML refers to “Data Manipulation Language”, a subset of SQL statements that modify the data stored in tables. Because Impala focuses on query performance and leverages the append-only nature of HDFS storage, currently Impala only supports a small set of DML statements:

- [INSERT Statement](#) on page 202
- [LOAD DATA Statement](#) on page 212

`INSERT` in Impala is primarily optimized for inserting large volumes of data in a single statement, to make effective use of the multi-megabyte HDFS blocks. This is the way in Impala to create new data files. If you intend to insert one or a few rows at a time, such as using the `INSERT ... VALUES` syntax, that technique is much more efficient for Impala tables stored in HBase. See [Using Impala to Query HBase Tables](#) on page 426 for details.

`LOAD DATA` moves existing data files into the directory for an Impala table, making them immediately available for Impala queries. This is one way in Impala to work with data files produced by other Hadoop components. (`CREATE EXTERNAL TABLE` is the other alternative; with external tables, you can query existing data files, while the files remain in their original location.)

To simulate the effects of an `UPDATE` or `DELETE` statement in other database systems, typically you use `INSERT` or `CREATE TABLE AS SELECT` to copy data from one table to another, filtering out or changing the appropriate rows during the copy operation.

## Impala SQL Language Reference

Although Impala currently does not have an UPDATE statement, you can achieve a similar result by using Impala tables stored in HBase. When you insert a row into an HBase table, and the table already contains a row with the same value for the key column, the older row is hidden, effectively the same as a single-row UPDATE.

Currently, Impala cannot perform DML operations for tables or partitions stored in the Amazon S3 filesystem. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details.

### Related information:

The other major classifications of SQL statements are data definition language (see [DDL Statements](#) on page 162) and queries (see [SELECT Statement](#) on page 218).

## ALTER TABLE Statement

The ALTER TABLE statement changes the structure or properties of an existing table. In Impala, this is a logical operation that updates the table metadata in the metastore database that Impala shares with Hive; ALTER TABLE does not actually rewrite, move, and so on the actual data files. Thus, you might need to perform corresponding physical filesystem operations, such as moving data files to a different HDFS directory, rewriting the data files to include extra fields, or converting them to a different file format.

### Syntax:

```
ALTER TABLE [old_db_name.]old_table_name RENAME TO [new_db_name.]new_table_name  
  
ALTER TABLE name ADD COLUMNS (col_spec[, col_spec ...])  
ALTER TABLE name DROP [COLUMN] column_name  
ALTER TABLE name CHANGE column_name new_name new_type  
ALTER TABLE name REPLACE COLUMNS (col_spec[, col_spec ...])  
  
ALTER TABLE name { ADD | DROP } PARTITION (partition_spec)  
  
ALTER TABLE name [PARTITION (partition_spec)]  
  SET { FILEFORMAT file_format  
    | LOCATION 'hdfs_path_of_directory'  
    | TBLPROPERTIES (table_properties)  
    | SERDEPROPERTIES (serde_properties) }  
  
ALTER TABLE name [PARTITION (partition_spec)] SET { CACHED IN 'pool_name' [WITH  
REPLICATION = integer] | UNCACHED }  
  
new_name ::= [new_database.]new_table_name  
  
col_spec ::= col_name type_name  
  
partition_spec ::= partition_col=constant_value  
  
table_properties ::= 'name'='value'[, 'name'='value' ...]  
serde_properties ::= 'name'='value'[, 'name'='value' ...]  
file_format ::= { PARQUET | TEXTFILE | RCFILE | SEQUENCEFILE | AVRO }
```

### Statement type: DDL

### Usage notes:

Whenever you specify partitions in an ALTER TABLE statement, through the PARTITION (partition\_spec) clause, you must include all the partitioning columns in the specification.

Most of the ALTER TABLE operations work the same for internal tables (managed by Impala) as for external tables (with data files located in arbitrary locations). The exception is renaming a table; for an external table, the underlying data directory is not renamed or moved.

### Amazon S3 considerations:

You can specify an s3a:// prefix in the LOCATION attribute of a table or partition to make Impala query data from the Amazon S3 filesystem. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details.

## HDFS caching (CACHED IN clause):

If you specify the `CACHED IN` clause, any existing or future data files in the table directory or the partition subdirectories are designated to be loaded into memory with the HDFS caching mechanism. See [Using HDFS Caching with Impala \(CDH 5.1 or higher only\)](#) on page 372 for details about using the HDFS caching feature.

In Impala 2.2 / CDH 5.4 and higher, the optional `WITH REPLICATION` clause for `CREATE TABLE` and `ALTER TABLE` lets you specify a ***replication factor***, the number of hosts on which to cache the same data blocks. When Impala processes a cached data block, where the cache replication factor is greater than 1, Impala randomly selects a host that has a cached copy of that data block. This optimization avoids excessive CPU usage on a single host when the same cached data block is processed multiple times.

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

The following sections show examples of the use cases for various `ALTER TABLE` clauses.

## To rename a table (RENAME TO clause):

The `RENAME TO` clause lets you change the name of an existing table, and optionally which database it is located in.

For internal tables, this operation physically renames the directory within HDFS that contains the data files; the original directory name no longer exists. By qualifying the table names with database names, you can use this technique to move an internal table (and its associated data directory) from one database to another. For example:

```
create database d1;
create database d2;
create database d3;
use d1;
create table mobile (x int);
use d2;
-- Move table from another database to the current one.
alter table d1.mobile rename to mobile;
use d1;
-- Move table from one database to another.
alter table d2.mobile rename to d3.mobile;
```

For external tables,

## To change the physical location where Impala looks for data files associated with a table or partition:

```
ALTER TABLE table_name [PARTITION (partition_spec)] SET LOCATION
'hdbs_path_of_directory';
```

The path you specify is the full HDFS path where the data files reside, or will be created. Impala does not create any additional subdirectory named after the table. Impala does not move any data files to this new location or change any data files that might already exist in that directory.

To set the location for a single partition, include the `PARTITION` clause. Specify all the same partitioning columns for the table, with a constant value for each, to precisely identify the single partition affected by the statement:

```
create table p1 (s string) partitioned by (month int, day int);
-- Each ADD PARTITION clause creates a subdirectory in HDFS.
alter table p1 add partition (month=1, day=1);
alter table p1 add partition (month=1, day=2);
alter table p1 add partition (month=2, day=1);
alter table p1 add partition (month=2, day=2);
-- Redirect queries, INSERT, and LOAD DATA for one partition
-- to a specific different directory.
alter table p1 partition (month=1, day=1) set location
'/usr/external_data/new_years_day';
```

- **Note:** If you are creating a partition for the first time and specifying its location, for maximum efficiency, use a single `ALTER TABLE` statement including both the `ADD PARTITION` and `LOCATION` clauses, rather than separate statements with `ADD PARTITION` and `SET LOCATION` clauses.

### To change the key-value pairs of the `TBLPROPERTIES` and `SERDEPROPERTIES` fields:

```
ALTER TABLE table_name SET TBLPROPERTIES ('key1='value1', 'key2='value2'[, ...]);  
ALTER TABLE table_name SET SERDEPROPERTIES ('key1='value1', 'key2='value2'[, ...]);
```

The `TBLPROPERTIES` clause is primarily a way to associate arbitrary user-specified data items with a particular table.

The `SERDEPROPERTIES` clause sets up metadata defining how tables are read or written, needed in some cases by Hive but not used extensively by Impala. You would use this clause primarily to change the delimiter in an existing text table or partition, by setting the `'serialization.format'` and `'field.delim'` property values to the new delimiter character:

```
-- This table begins life as pipe-separated text format.  
create table change_to_csv (s1 string, s2 string) row format delimited fields terminated  
by '|';  
-- Then we change it to a CSV table.  
alter table change_to_csv set SERDEPROPERTIES ('serialization.format='',''  
'field.delim='','');  
insert overwrite change_to_csv values ('stop','go'), ('yes','no');  
!hdfs dfs -cat 'hdfs://hostname:8020/data_directory/dbname.db/change_to_csv/data_file';  
stop,go  
yes,no
```

Use the `DESCRIBE FORMATTED` statement to see the current values of these properties for an existing table. See [CREATE TABLE Statement](#) on page 180 for more details about these clauses. See [Setting Statistics Manually through ALTER TABLE](#) on page 368 for an example of using table properties to fine-tune the performance-related table statistics.

### To reorganize columns for a table:

```
ALTER TABLE table_name ADD COLUMNS (column_defs);  
ALTER TABLE table_name REPLACE COLUMNS (column_defs);  
ALTER TABLE table_name CHANGE column_name new_name new_type;  
ALTER TABLE table_name DROP column_name;
```

The `column_spec` is the same as in the `CREATE TABLE` statement: the column name, then its data type, then an optional comment. You can add multiple columns at a time. The parentheses are required whether you add a single column or multiple columns. When you replace columns, all the original column definitions are discarded. You might use this technique if you receive a new set of data files with different data types or columns in a different order. (The data files are retained, so if the new columns are incompatible with the old ones, use `INSERT OVERWRITE` or `LOAD DATA OVERWRITE` to replace all the data before issuing any further queries.)

You might use the `CHANGE` clause to rename a single column, or to treat an existing column as a different type than before, such as to switch between treating a column as `STRING` and `TIMESTAMP`, or between `INT` and `BIGINT`. You can only drop a single column at a time; to drop multiple columns, issue multiple `ALTER TABLE` statements, or define the new set of columns with a single `ALTER TABLE ... REPLACE COLUMNS` statement.

### To change the file format that Impala expects data to be in, for a table or partition:

Use an `ALTER TABLE ... SET FILEFORMAT` clause. You can include an optional `PARTITION (col1=val1, col2=val2, ...)` clause so that the file format is changed for a specific partition rather than the entire table.

Because this operation only changes the table metadata, you must do any conversion of existing data using regular Hadoop techniques outside of Impala. Any new data created by the Impala `INSERT` statement will be in the new format. You cannot specify the delimiter for Text files; the data files must be comma-delimited.

To set the file format for a single partition, include the `PARTITION` clause. Specify all the same partitioning columns for the table, with a constant value for each, to precisely identify the single partition affected by the statement:

```
create table p1 (s string) partitioned by (month int, day int);
-- Each ADD PARTITION clause creates a subdirectory in HDFS.
alter table p1 add partition (month=1, day=1);
alter table p1 add partition (month=1, day=2);
alter table p1 add partition (month=2, day=1);
alter table p1 add partition (month=2, day=2);
-- Queries and INSERT statements will read and write files
-- in this format for this specific partition.
alter table p1 partition (month=2, day=2) set fileformat parquet;
```

**To add or drop partitions for a table**, the table must already be partitioned (that is, created with a `PARTITIONED BY` clause). The partition is a physical directory in HDFS, with a name that encodes a particular column value (the **partition key**). The Impala `INSERT` statement already creates the partition if necessary, so the `ALTER TABLE ... ADD PARTITION` is primarily useful for importing data by moving or copying existing data files into the HDFS directory corresponding to a partition. (You can use the `LOAD DATA` statement to move files into the partition directory, or `ALTER TABLE ... PARTITION (...) SET LOCATION` to point a partition at a directory that already contains data files.)

The `DROP PARTITION` clause is used to remove the HDFS directory and associated data files for a particular set of partition key values; for example, if you always analyze the last 3 months worth of data, at the beginning of each month you might drop the oldest partition that is no longer needed. Removing partitions reduces the amount of metadata associated with the table and the complexity of calculating the optimal query plan, which can simplify and speed up queries on partitioned tables, particularly join queries. Here is an example showing the `ADD PARTITION` and `DROP PARTITION` clauses.

```
-- Create an empty table and define the partitioning scheme.
create table part_t (x int) partitioned by (month int);
-- Create an empty partition into which you could copy data files from some other
source.
alter table part_t add partition (month=1);
-- After changing the underlying data, issue a REFRESH statement to make the data
visible in Impala.
refresh part_t;
-- Later, do the same for the next month.
alter table part_t add partition (month=2);

-- Now you no longer need the older data.
alter table part_t drop partition (month=1);
-- If the table was partitioned by month and year, you would issue a statement like:
-- alter table part_t drop partition (year=2003,month=1);
-- which would require 12 ALTER TABLE statements to remove a year's worth of data.

-- If the data files for subsequent months were in a different file format,
-- you could set a different file format for the new partition as you create it.
alter table part_t add partition (month=3) set fileformat=parquet;
```

The value specified for a partition key can be an arbitrary constant expression, without any references to columns. For example:

```
alter table time_data add partition (month=concat('Decem','ber'));
alter table sales_data add partition (zipcode = cast(9021 * 10 as string));
```

- **Note:**

An alternative way to reorganize a table and its associated data files is to use `CREATE TABLE` to create a variation of the original table, then use `INSERT` to copy the transformed or reordered data to the new table. The advantage of `ALTER TABLE` is that it avoids making a duplicate copy of the data files, allowing you to reorganize huge volumes of data in a space-efficient way using familiar Hadoop techniques.

### To switch a table between internal and external:

You can switch a table from internal to external, or from external to internal, by using the `ALTER TABLE` statement:

```
-- Switch a table from internal to external.  
ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='TRUE') ;  
  
-- Switch a table from external to internal.  
ALTER TABLE table_name SET TBLPROPERTIES('EXTERNAL'='FALSE') ;
```

**Cancellation:** Cannot be cancelled.

### HDFS permissions:

Most `ALTER TABLE` clauses do not actually read or write any HDFS files, and so do not depend on specific HDFS permissions. For example, the `SET FILEFORMAT` clause does not actually check the file format existing data files or convert them to the new format, and the `SET LOCATION` clause does not require any special permissions on the new location. (Any permission-related failures would come later, when you actually query or insert into the table.)

In general, `ALTER TABLE` clauses that do touch HDFS files and directories require the same HDFS permissions as corresponding `CREATE`, `INSERT`, or `SELECT` statements. The permissions allow the user ID that the `impalad` daemon runs under, typically the `impala` user, to read or write files or directories, or (in the case of the execute bit) descend into a directory. The `RENAME TO` clause requires read, write, and execute permission in the source and destination database directories and in the table data directory, and read and write permission for the data files within the table. The `ADD PARTITION` and `DROP PARTITION` clauses require write and execute permissions for the associated partition directory.

### Related information:

[Tables](#) on page 156, [CREATE TABLE Statement](#) on page 180, [DROP TABLE Statement](#) on page 197, [Partitioning for Impala Tables](#) on page 394, [Internal Tables](#) on page 157, [External Tables](#) on page 157

## ALTER VIEW Statement

Changes the query associated with a view, or the associated database and/or name of the view.

Because a view is purely a logical construct (an alias for a query) with no physical data behind it, `ALTER VIEW` only involves changes to metadata in the metastore database, not any data files in HDFS.

### Syntax:

```
ALTER VIEW [database_name.]view_name AS select_statement  
ALTER VIEW [database_name.]view_name RENAME TO [database_name.]view_name
```

### Statement type: DDL

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

### Security considerations:

If these statements in your environment contain sensitive literal values such as credit card numbers or tax identifiers, Impala can redact this sensitive information when displaying the statements in log files and other administrative contexts. See [Sensitive Data Redaction](#) for details.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

**Examples:**

```
create table t1 (x int, y int, s string);
create table t2 like t1;
create view v1 as select * from t1;
alter view v1 as select * from t2;
alter view v1 as select x, upper(s) s from t2;
```

To see the definition of a view, issue a `DESCRIBE FORMATTED` statement, which shows the query from the original `CREATE VIEW` statement:

```
[localhost:21000] > create view v1 as select * from t1;
[localhost:21000] > describe formatted v1;
Query finished, fetching results ...
+-----+-----+-----+
| name | type | comment |
+-----+-----+-----+
# col_name | data_type | comment
| x | int | None |
| y | int | None |
| s | string | None |
|  | NULL | NULL |
# Detailed Table Information | NULL | NULL |
Database: | views | NULL |
Owner: | cloudera | NULL |
CreateTime: | Mon Jul 08 15:56:27 EDT 2013 | NULL |
LastAccessTime: | UNKNOWN | NULL |
Protect Mode: | None | NULL |
Retention: | 0 | NULL |
Table Type: | VIRTUAL_VIEW | NULL |
Table Parameters:
| transient_lastDdlTime | 1373313387 |
|  | NULL | NULL |
# Storage Information | NULL | NULL |
SerDe Library: | null | NULL |
InputFormat: | null | NULL |
OutputFormat: | null | NULL |
Compressed: | No | NULL |
Num Buckets: | 0 | NULL |
Bucket Columns: | [ ] | NULL |
Sort Columns: | [ ] | NULL |
|  | NULL | NULL |
# View Information | NULL | NULL |
```

View Original Text:	SELECT * FROM t1	NULL
View Expanded Text:	SELECT * FROM t1	NULL

### Related information:

[Views](#) on page 158, [CREATE VIEW Statement](#) on page 186, [DROP VIEW Statement](#) on page 198

## COMPUTE STATS Statement

Gathers information about volume and distribution of data in a table and all associated columns and partitions. The information is stored in the metastore database, and used by Impala to help optimize queries. For example, if Impala can determine that a table is large or small, or has many or few distinct values it can organize parallelize the work appropriately for a join query or insert operation. For details about the kinds of information gathered by this statement, see [How Impala Uses Statistics for Query Optimization](#) on page 363.

### Syntax:

```
COMPUTE STATS [db_name.]table_name  
COMPUTE INCREMENTAL STATS [db_name.]table_name [PARTITION (partition_spec)]  
partition_spec ::= partition_col=constant_value
```

The PARTITION clause is only allowed in combination with the INCREMENTAL clause. It is optional for COMPUTE INCREMENTAL STATS, and required for DROP INCREMENTAL STATS. Whenever you specify partitions through the PARTITION (partition\_spec) clause in a COMPUTE INCREMENTAL STATS or DROP INCREMENTAL STATS statement, you must include all the partitioning columns in the specification, and specify constant values for all the partition key columns.

### Usage notes:

Originally, Impala relied on users to run the Hive ANALYZE TABLE statement, but that method of gathering statistics proved unreliable and difficult to use. The Impala COMPUTE STATS statement is built from the ground up to improve the reliability and user-friendliness of this operation. COMPUTE STATS does not require any setup steps or special configuration. You only run a single Impala COMPUTE STATS statement to gather both table and column statistics, rather than separate Hive ANALYZE TABLE statements for each kind of statistics.

The COMPUTE INCREMENTAL STATS variation is a shortcut for partitioned tables that works on a subset of partitions rather than the entire table. The incremental nature makes it suitable for large tables with many partitions, where a full COMPUTE STATS operation takes too long to be practical each time a partition is added or dropped. See [Overview of Incremental Statistics](#) on page 364 for full usage details.

COMPUTE INCREMENTAL STATS only applies to partitioned tables. If you use the INCREMENTAL clause for an unpartitioned table, Impala automatically uses the original COMPUTE STATS statement. Such tables display false under the Incremental stats column of the SHOW TABLE STATS output.

- **Note:** Because many of the most performance-critical and resource-intensive operations rely on table and column statistics to construct accurate and efficient plans, `COMPUTE STATS` is an important step at the end of your ETL process. Run `COMPUTE STATS` on all tables as your first step during performance tuning for slow queries, or troubleshooting for out-of-memory conditions:
  - Accurate statistics help Impala construct an efficient query plan for join queries, improving performance and reducing memory usage.
  - Accurate statistics help Impala distribute the work effectively for insert operations into Parquet tables, improving performance and reducing memory usage.
  - Accurate statistics help Impala estimate the memory required for each query, which is important when you use resource management features, such as admission control and the YARN resource management framework. The statistics help Impala to achieve high concurrency, full utilization of available memory, and avoid contention with workloads from other Hadoop components.

### HBase considerations:

`COMPUTE STATS` works for HBase tables also. The statistics gathered for HBase tables are somewhat different than for HDFS-backed tables, but that metadata is still used for optimization when HBase tables are involved in join queries.

### Amazon S3 considerations:

`COMPUTE STATS` also works for tables where data resides in the Amazon Simple Storage Service (S3). See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details.

### Performance considerations:

The statistics collected by `COMPUTE STATS` are used to optimize join queries `INSERT` operations into Parquet tables, and other resource-intensive kinds of SQL statements. See [How Impala Uses Statistics for Query Optimization](#) on page 363 for details.

### Examples:

This example shows two tables, `T1` and `T2`, with a small number distinct values linked by a parent-child relationship between `T1.ID` and `T2.PARENT`. `T1` is tiny, while `T2` has approximately 100K rows. Initially, the statistics includes physical measurements such as the number of files, the total size, and size measurements for fixed-length columns such as with the `INT` type. Unknown values are represented by -1. After running `COMPUTE STATS` for each table, much more information is available through the `SHOW STATS` statements. If you were running a join query involving both of these tables, you would need statistics for both tables to get the most effective optimization for the query.

```
[localhost:21000] > show table stats t1;
Query: show table stats t1
+-----+-----+-----+
| #Rows | #Files | Size | Format |
+-----+-----+-----+
| -1   | 1     | 33B  | TEXT  |
+-----+-----+-----+
Returned 1 row(s) in 0.02s
[localhost:21000] > show table stats t2;
Query: show table stats t2
+-----+-----+-----+
| #Rows | #Files | Size      | Format |
+-----+-----+-----+
| -1   | 28    | 960.00KB | TEXT  |
+-----+-----+-----+
Returned 1 row(s) in 0.01s
[localhost:21000] > show column stats t1;
Query: show column stats t1
+-----+-----+-----+-----+-----+-----+
| Column | Type   | #Distinct Values | #Nulls | Max Size | Avg Size |
+-----+-----+-----+-----+-----+-----+
| id    | INT    | -1              | -1    | 4        | 4        |
| s     | STRING | -1              | -1    | -1       | -1       |
+-----+-----+-----+-----+-----+-----+
```

```

+-----+-----+-----+-----+-----+
Returned 2 row(s) in 1.71s
[localhost:21000] > show column stats t2;
Query: show column stats t2
+-----+-----+-----+-----+-----+
| Column | Type   | #Distinct Values | #Nulls | Max Size | Avg Size |
+-----+-----+-----+-----+-----+
| parent | INT    | -1              | -1     | 4        | 4        |
| s       | STRING  | -1              | -1     | -1      | -1      |
+-----+-----+-----+-----+-----+
Returned 2 row(s) in 0.01s
[localhost:21000] > compute stats t1;
Query: compute stats t1
+-----+
| summary
+-----+
| Updated 1 partition(s) and 2 column(s). |
+-----+
Returned 1 row(s) in 5.30s
[localhost:21000] > show table stats t1;
Query: show table stats t1
+-----+-----+-----+
| #Rows | #Files | Size   | Format |
+-----+-----+-----+
| 3     | 1      | 33B   | TEXT   |
+-----+-----+-----+
Returned 1 row(s) in 0.01s
[localhost:21000] > show column stats t1;
Query: show column stats t1
+-----+-----+-----+-----+-----+
| Column | Type   | #Distinct Values | #Nulls | Max Size | Avg Size |
+-----+-----+-----+-----+-----+
| id     | INT    | 3                | -1     | 4        | 4        |
| s       | STRING  | 3                | -1     | -1      | -1      |
+-----+-----+-----+-----+-----+
Returned 2 row(s) in 0.02s
[localhost:21000] > compute stats t2;
Query: compute stats t2
+-----+
| summary
+-----+
| Updated 1 partition(s) and 2 column(s). |
+-----+
Returned 1 row(s) in 5.70s
[localhost:21000] > show table stats t2;
Query: show table stats t2
+-----+-----+-----+
| #Rows | #Files | Size   | Format |
+-----+-----+-----+
| 98304 | 1      | 960.00KB | TEXT   |
+-----+-----+-----+
Returned 1 row(s) in 0.03s
[localhost:21000] > show column stats t2;
Query: show column stats t2
+-----+-----+-----+-----+-----+
| Column | Type   | #Distinct Values | #Nulls | Max Size | Avg Size |
+-----+-----+-----+-----+-----+
| parent | INT    | 3                | -1     | 4        | 4        |
| s       | STRING  | 6                | -1     | 14      | 9.3     |
+-----+-----+-----+-----+-----+
Returned 2 row(s) in 0.01s

```

The following example shows how to use the `INCREMENTAL` clause, available in Impala 2.1.0 and higher. The `COMPUTE INCREMENTAL STATS` syntax lets you collect statistics for newly added or changed partitions, without rescanning the entire table.

```

-- Initially the table has no incremental stats, as indicated
-- by -1 under #Rows and false under Incremental stats.
show table stats item_partitioned;
+-----+-----+-----+-----+-----+
| i_category | #Rows | #Files | Size   | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+

```

```

| Books      | -1   | 1    | 223.74KB | NOT CACHED | PARQUET | false
| Children   | -1   | 1    | 230.05KB | NOT CACHED | PARQUET | false
| Electronics | -1   | 1    | 232.67KB | NOT CACHED | PARQUET | false
| Home       | -1   | 1    | 232.56KB | NOT CACHED | PARQUET | false
| Jewelry    | -1   | 1    | 223.72KB | NOT CACHED | PARQUET | false
| Men        | -1   | 1    | 231.25KB | NOT CACHED | PARQUET | false
| Music      | -1   | 1    | 237.90KB | NOT CACHED | PARQUET | false
| Shoes      | -1   | 1    | 234.90KB | NOT CACHED | PARQUET | false
| Sports     | -1   | 1    | 227.97KB | NOT CACHED | PARQUET | false
| Women      | -1   | 1    | 226.27KB | NOT CACHED | PARQUET | false
| Total      | -1   | 10   | 2.25MB   | 0B          | PARQUET | false
+-----+-----+-----+-----+-----+-----+-----+
-- After the first COMPUTE INCREMENTAL STATS,
-- all partitions have stats.
compute incremental stats item_partitioned;
+-----+
| summary |
+-----+
| Updated 10 partition(s) and 21 column(s). |
+-----+
show table stats item_partitioned;
+-----+-----+-----+-----+-----+-----+-----+
| i_category | #Rows | #Files | Size      | Bytes Cached | Format | Incremental stats
+-----+-----+-----+-----+-----+-----+-----+
| Books      | 1733  | 1    | 223.74KB | NOT CACHED | PARQUET | true
| Children   | 1786  | 1    | 230.05KB | NOT CACHED | PARQUET | true
| Electronics | 1812  | 1    | 232.67KB | NOT CACHED | PARQUET | true
| Home       | 1807  | 1    | 232.56KB | NOT CACHED | PARQUET | true
| Jewelry    | 1740  | 1    | 223.72KB | NOT CACHED | PARQUET | true
| Men        | 1811  | 1    | 231.25KB | NOT CACHED | PARQUET | true
| Music      | 1860  | 1    | 237.90KB | NOT CACHED | PARQUET | true
| Shoes      | 1835  | 1    | 234.90KB | NOT CACHED | PARQUET | true
| Sports     | 1783  | 1    | 227.97KB | NOT CACHED | PARQUET | true
| Women      | 1790  | 1    | 226.27KB | NOT CACHED | PARQUET | true
| Total      | 17957 | 10   | 2.25MB   | 0B          | PARQUET | true
+-----+-----+-----+-----+-----+-----+-----+
-- Add a new partition...
alter table item_partitioned add partition (i_category='Camping');
-- Add or replace files in HDFS outside of Impala,
-- rendering the stats for a partition obsolete.
!import_data_into_sports_partition.sh
refresh item_partitioned;
drop incremental stats item_partitioned partition (i_category='Sports');
-- Now some partitions have incremental stats
-- and some don't.
show table stats item_partitioned;
+-----+-----+-----+-----+-----+-----+-----+
| i_category | #Rows | #Files | Size      | Bytes Cached | Format | Incremental stats
+-----+-----+-----+-----+-----+-----+-----+
| Books      | 1733  | 1    | 223.74KB | NOT CACHED | PARQUET | true
| Camping    | -1   | 1    | 408.02KB | NOT CACHED | PARQUET | false
| Children   | 1786  | 1    | 230.05KB | NOT CACHED | PARQUET | true
| Electronics | 1812  | 1    | 232.67KB | NOT CACHED | PARQUET | true
| Home       | 1807  | 1    | 232.56KB | NOT CACHED | PARQUET | true
| Jewelry    | 1740  | 1    | 223.72KB | NOT CACHED | PARQUET | true
| Men        | 1811  | 1    | 231.25KB | NOT CACHED | PARQUET | true
| Music      | 1860  | 1    | 237.90KB | NOT CACHED | PARQUET | true
| Shoes      | 1835  | 1    | 234.90KB | NOT CACHED | PARQUET | true
| Sports     | -1   | 1    | 227.97KB | NOT CACHED | PARQUET | false
| Women      | 1790  | 1    | 226.27KB | NOT CACHED | PARQUET | true
| Total      | 17957 | 11   | 2.65MB   | 0B          | PARQUET | true
+-----+-----+-----+-----+-----+-----+-----+
-- After another COMPUTE INCREMENTAL STATS,
-- all partitions have incremental stats, and only the 2
-- partitions without incremental stats were scanned.
compute incremental stats item_partitioned;
+-----+
| summary |
+-----+
| Updated 2 partition(s) and 21 column(s). |
+-----+

```

show table stats item_partitioned;							
i_category	#Rows	#Files	Size	Bytes Cached	Format	Incremental stats	
Books	1733	1	223.74KB	NOT CACHED	PARQUET	true	
Camping	5328	1	408.02KB	NOT CACHED	PARQUET	true	
Children	1786	1	230.05KB	NOT CACHED	PARQUET	true	
Electronics	1812	1	232.67KB	NOT CACHED	PARQUET	true	
Home	1807	1	232.56KB	NOT CACHED	PARQUET	true	
Jewelry	1740	1	223.72KB	NOT CACHED	PARQUET	true	
Men	1811	1	231.25KB	NOT CACHED	PARQUET	true	
Music	1860	1	237.90KB	NOT CACHED	PARQUET	true	
Shoes	1835	1	234.90KB	NOT CACHED	PARQUET	true	
Sports	1783	1	227.97KB	NOT CACHED	PARQUET	true	
Women	1790	1	226.27KB	NOT CACHED	PARQUET	true	
Total	17957	11	2.65MB	0B			

### File format considerations:

The `COMPUTE STATS` statement works with tables created with any of the file formats supported by Impala. See [How Impala Works with Hadoop File Formats](#) on page 399 for details about working with the different file formats. The following considerations apply to `COMPUTE STATS` depending on the file format of the table.

The `COMPUTE STATS` statement works with text tables with no restrictions. These tables can be created through either Impala or Hive.

The `COMPUTE STATS` statement works with Parquet tables. These tables can be created through either Impala or Hive.

- **Note:** Currently, a known issue ([IMPALA-488](#)) could cause excessive memory usage during a `COMPUTE STATS` operation on a Parquet table. As a workaround, issue the command `SET NUM_SCANNER_THREADS=2` in `impala-shell` before issuing the `COMPUTE STATS` statement. Then issue `UNSET NUM_SCANNER_THREADS` before continuing with queries.

The `COMPUTE STATS` statement works with Avro tables, as long as they are created with SQL-style column names and types rather than an Avro-style schema specification. These tables are currently always created through Hive rather than Impala.

The `COMPUTE STATS` statement works with RCFile tables with no restrictions. These tables can be created through either Impala or Hive.

The `COMPUTE STATS` statement works with SequenceFile tables with no restrictions. These tables can be created through either Impala or Hive.

The `COMPUTE STATS` statement works with partitioned tables, whether all the partitions use the same file format, or some partitions are defined through `ALTER TABLE` to use different file formats.

### Statement type: DDL

**Cancellation:** Certain multi-stage statements (`CREATE TABLE AS SELECT` and `COMPUTE STATS`) can be cancelled during some stages, when running `INSERT` or `SELECT` operations internally. To cancel this statement, use `Ctrl-C` from the `impala-shell` interpreter, the **Cancel** button from the **Watch** page in Hue, **Actions > Cancel** from the **Queries** list in Cloudera Manager, or **Cancel** from the list of in-flight queries (for a particular node) on the **Queries** tab in the Impala web UI (port 25000).

### Restrictions:

Currently, the `COMPUTE STATS` statement under CDH 4 does not store any statistics for `DECIMAL` columns. When Impala runs under CDH 5, which has better support for `DECIMAL` in the metastore database, `COMPUTE STATS` does collect statistics for `DECIMAL` columns and Impala uses the statistics to optimize query performance.

- **Note:** Prior to Impala 1.4.0, `COMPUTE STATS` counted the number of `NULL` values in each column and recorded that figure in the metastore database. Because Impala does not currently make use of the `NULL` count during query planning, Impala 1.4.0 and higher speeds up the `COMPUTE STATS` statement by skipping this `NULL` counting.

#### Internal details:

Behind the scenes, the `COMPUTE STATS` statement executes two statements: one to count the rows of each partition in the table (or the entire table if unpartitioned) through the `COUNT(*)` function, and another to count the approximate number of distinct values in each column through the `NDV()` function. You might see these queries in your monitoring and diagnostic displays. The same factors that affect the performance, scalability, and execution of other queries (such as parallel execution, memory usage, admission control, and timeouts) also apply to the queries run by the `COMPUTE STATS` statement.

#### HDFS permissions:

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read permission for all affected files in the source directory: all files in the case of an unpartitioned table or a partitioned table in the case of `COMPUTE STATS`; or all the files in partitions without incremental stats in the case of `COMPUTE INCREMENTAL STATS`. It must also have read and execute permissions for all relevant directories holding the data files. (Essentially, `COMPUTE STATS` requires the same permissions as the underlying `SELECT` queries it runs against the table.)

#### Related information:

[DROP STATS Statement](#) on page 193, [SHOW TABLE STATS Statement](#) on page 251, [SHOW COLUMN STATS Statement](#) on page 253, [How Impala Uses Statistics for Query Optimization](#) on page 363

## CREATE DATABASE Statement

Creates a new database.

In Impala, a database is both:

- A logical construct for grouping together related tables, views, and functions within their own namespace. You might use a separate database for each application, set of related tables, or round of experimentation.
- A physical construct represented by a directory tree in HDFS. Tables (internal tables), partitions, and data files are all located under this directory. You can back it up, measure space usage, or remove it (if it is empty) with a `DROP DATABASE` statement.

#### Syntax:

```
CREATE (DATABASE|SCHEMA) [IF NOT EXISTS] database_name[COMMENT 'database_comment']
[LOCATION hdfs_path];
```

**Statement type:** DDL

#### Usage notes:

A database is physically represented as a directory in HDFS, with a filename extension `.db`, under the main Impala data directory. If the associated HDFS directory does not exist, it is created for you. All databases and their associated directories are top-level objects, with no physical or logical nesting.

After creating a database, to make it the current database within an `impala-shell` session, use the `USE` statement. You can refer to tables in the current database without prepending any qualifier to their names.

When you first connect to Impala through `impala-shell`, the database you start in (before issuing any `CREATE DATABASE` or `USE` statements) is named `default`.

## Impala SQL Language Reference

Impala includes another predefined database, `_impala_builtins`, that serves as the location for the [built-in functions](#). To see the built-in functions, use a statement like the following:

```
show functions in _impala_builtins;
show functions in _impala_builtins like '*substring*';
```

After creating a database, your `impala-shell` session or another `impala-shell` connected to the same node can immediately access that database. To access the database through the Impala daemon on a different node, issue the `INVALIDATE METADATA` statement first while connected to that other node.

Setting the `LOCATION` attribute for a new database is a way to work with sets of files in an HDFS directory structure outside the default Impala data directory, as opposed to setting the `LOCATION` attribute for each individual table.

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

### Hive considerations:

When you create a database in Impala, the database can also be used by Hive.

### Examples:

```
create database first_db;
use first_db;
create table t1 (x int);

create database second_db;
use second_db;
-- Each database has its own namespace for tables.
-- You can reuse the same table names in each database.
create table t1 (s string);

create database temp;
-- You do not have to USE a database after creating it.
-- Just qualify the table name with the name of the database.
create table temp.t2 (x int, y int);
use database temp;
create table t3 (s string);
-- You cannot drop a database while it is selected by the USE statement.
drop database temp;
ERROR: AnalysisException: Cannot drop current default database: temp
-- The always-available database 'default' is a convenient one to USE.
use default;
-- Before dropping a database, first drop all the tables inside it.
drop database temp;
ERROR: ImpalaRuntimeException: Error making 'dropDatabase' RPC to Hive Metastore:
CAUSED BY: InvalidOperationException: Database temp is not empty
show tables in temp;
+-----+
| name |
+-----+
| t3   |
+-----+
drop table temp.t3;
drop database temp;
```

**Cancellation:** Cannot be cancelled.

### HDFS permissions:

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have write permission for the parent HDFS directory under which the database is located.

### Related information:

[Databases](#) on page 155, [DROP DATABASE Statement](#) on page 192, [USE Statement](#) on page 256, [SHOW DATABASES](#) on page 248, [Tables](#) on page 156

## CREATE FUNCTION Statement

Creates a user-defined function (UDF), which you can use to implement custom logic during `SELECT` or `INSERT` operations.

### Syntax:

The syntax is different depending on whether you create a scalar UDF, which is called once for each row and implemented by a single function, or a user-defined aggregate function (UDA), which is implemented by multiple functions that compute intermediate results across sets of rows.

To create a scalar UDF, issue a `CREATE FUNCTION` statement:

```
CREATE FUNCTION [IF NOT EXISTS] [db_name.]function_name([arg_type[, arg_type...]])
  RETURNS return_type
  LOCATION 'hdfs_path'
  SYMBOL='symbol_or_class'
```

To create a UDA, issue a `CREATE AGGREGATE FUNCTION` statement:

```
CREATE [AGGREGATE] FUNCTION [IF NOT EXISTS] [db_name.]function_name([arg_type[, arg_type...])
  RETURNS return_type
  LOCATION 'hdfs_path'
  [INIT_FN='function']
  UPDATE_FN='function'
  MERGE_FN='function'
  [PREPARE_FN='function']
  [CLOSEFN='function']
  [SERIALIZE_FN='function']
  [FINALIZE_FN='function']
```

**Statement type:** DDL

### Varargs notation:

If the underlying implementation of your function accepts a variable number of arguments:

- The variable arguments must go last in the argument list.
- The variable arguments must all be of the same type.
- You must include at least one instance of the variable arguments in every function call invoked from SQL.
- You designate the variable portion of the argument list in the `CREATE FUNCTION` statement by including ... immediately after the type name of the first variable argument. For example, to create a function that accepts an `INT` argument, followed by a `BOOLEAN`, followed by one or more `STRING` arguments, your `CREATE FUNCTION` statement would look like:

```
CREATE FUNCTION func_name (INT, BOOLEAN, STRING ...)
  RETURNS type LOCATION 'path' SYMBOL='entry_point';
```

See [Variable-Length Argument Lists](#) on page 314 for how to code the C++ or Java function to accept variable-length argument lists.

### Scalar and aggregate functions:

The simplest kind of user-defined function returns a single scalar value each time it is called, typically once for each row in the result set. This general kind of function is what is usually meant by UDF. User-defined aggregate functions (UDAs) are a specialized kind of UDF that produce a single value based on the contents of multiple rows. You usually use UDAs in combination with a `GROUP BY` clause to condense a large result set into a smaller one, or even a single row summarizing column values across an entire table.

You create UDAs by using the `CREATE AGGREGATE FUNCTION` syntax. The clauses `INIT_FN`, `UPDATE_FN`, `MERGE_FN`, `SERIALIZE_FN`, `FINALIZE_FN`, and `INTERMEDIATE` only apply when you create a UDA rather than a scalar UDF.

The \*\_FN clauses specify functions to call at different phases of function processing.

- **Initialize:** The function you specify with the `INIT_FN` clause does any initial setup, such as initializing member variables in internal data structures. This function is often a stub for simple UDAFs. You can omit this clause and a default (no-op) function will be used.
- **Update:** The function you specify with the `UPDATE_FN` clause is called once for each row in the original result set, that is, before any `GROUP BY` clause is applied. A separate instance of the function is called for each different value returned by the `GROUP BY` clause. The final argument passed to this function is a pointer, to which you write an updated value based on its original value and the value of the first argument.
- **Merge:** The function you specify with the `MERGE_FN` clause is called an arbitrary number of times, to combine intermediate values produced by different nodes or different threads as Impala reads and processes data files in parallel. The final argument passed to this function is a pointer, to which you write an updated value based on its original value and the value of the first argument.
- **Serialize:** The function you specify with the `SERIALIZE_FN` clause frees memory allocated to intermediate results. It is required if any memory was allocated by the `Allocate` function in the `Init`, `Update`, or `Merge` functions, or if the intermediate type contains any pointers. See [the UDA code samples](#) for details.
- **Finalize:** The function you specify with the `FINALIZE_FN` clause does any required teardown for resources acquired by your UDF, such as freeing memory, closing file handles if you explicitly opened any files, and so on. This function is often a stub for simple UDAFs. You can omit this clause and a default (no-op) function will be used. It is required in UDAFs where the final return type is different than the intermediate type, or if any memory was allocated by the `Allocate` function in the `Init`, `Update`, or `Merge` functions. See [the UDA code samples](#) for details.

If you use a consistent naming convention for each of the underlying functions, Impala can automatically determine the names based on the first such clause, so the others are optional.

For end-to-end examples of UDAFs, see [Impala User-Defined Functions \(UDFs\)](#) on page 308.

### Usage notes:

- You can write Impala UDFs in either C++ or Java. C++ UDFs are new to Impala, and are the recommended format for high performance utilizing native code. Java-based UDFs are compatible between Impala and Hive, and are most suited to reusing existing Hive UDFs. (Impala can run Java-based Hive UDFs but not Hive UDAFs.)
- The body of the UDF is represented by a `.so` or `.jar` file, which you store in HDFS and the `CREATE FUNCTION` statement distributes to each Impala node.
- Impala calls the underlying code during SQL statement evaluation, as many times as needed to process all the rows from the result set. All UDFs are assumed to be deterministic, that is, to always return the same result when passed the same argument values. Impala might or might not skip some invocations of a UDF if the result value is already known from a previous call. Therefore, do not rely on the UDF being called a specific number of times, and do not return different result values based on some external factor such as the current time, a random number function, or an external data source that could be updated while an Impala query is in progress.
- The names of the function arguments in the UDF are not significant, only their number, positions, and data types.
- You can overload the same function name by creating multiple versions of the function, each with a different argument signature. For security reasons, you cannot make a UDF with the same name as any built-in function.
- In the UDF code, you represent the function return result as a `struct`. This `struct` contains 2 fields. The first field is a `boolean` representing whether the value is `NULL` or not. (When this field is `true`, the return value is interpreted as `NULL`.) The second field is the same type as the specified function return type, and holds the return value when the function returns something other than `NULL`.
- In the UDF code, you represent the function arguments as an initial pointer to a UDF context structure, followed by references to zero or more `structs`, corresponding to each of the arguments. Each `struct` has the same 2 fields as with the return value, a `boolean` field representing whether the argument is `NULL`, and a field of the appropriate type holding any non-`NULL` argument value.
- For sample code and build instructions for UDFs, see [the sample UDFs in the Impala github repo](#).
- Because the file representing the body of the UDF is stored in HDFS, it is automatically available to all the Impala nodes. You do not need to manually copy any UDF-related files between servers.

- Because Impala currently does not have any `ALTER FUNCTION` statement, if you need to rename a function, move it to a different database, or change its signature or other properties, issue a `DROP FUNCTION` statement for the original function followed by a `CREATE FUNCTION` with the desired properties.
- Because each UDF is associated with a particular database, either issue a `USE` statement before doing any `CREATE FUNCTION` statements, or specify the name of the function as `db_name.function_name`.

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

#### **Compatibility:**

Impala can run UDFs that were created through Hive, as long as they refer to Impala-compatible data types (not composite or nested column types). Hive can run Java-based UDFs that were created through Impala, but not Impala UDFs written in C++.

#### **Restrictions:**

Currently, Impala UDFs and UDAFs are not persisted in the metastore database. Information about these functions is held in the memory of the `catalogd` daemon. You must reload them by running the `CREATE FUNCTION` statements again each time you restart the `catalogd` daemon.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch and HDFS files or directories, therefore no HDFS permissions are required.

#### **Related information:**

[Impala User-Defined Functions \(UDFs\)](#) on page 308 for more background information, usage instructions, and examples for Impala UDFs; [DROP FUNCTION Statement](#) on page 192

## [CREATE ROLE Statement \(CDH 5.2 or higher only\)](#)

The `CREATE ROLE` statement creates a role to which privileges can be granted. Privileges can be granted to roles, which can then be assigned to users. A user that has been assigned a role will only be able to exercise the privileges of that role. Only users that have administrative privileges can create/drop roles. By default, the `hive`, `impala` and `hue` users have administrative privileges in Sentry.

#### **Syntax:**

```
CREATE ROLE role_name
```

#### **Required privileges:**

Only administrative users (those with `ALL` privileges on the server, defined in the Sentry policy file) can use this statement.

#### **Compatibility:**

Impala makes use of any roles and privileges specified by the `GRANT` and `REVOKE` statements in Hive, and Hive makes use of any roles and privileges specified by the `GRANT` and `REVOKE` statements in Impala. The Impala `GRANT` and `REVOKE` statements for privileges do not require the `ROLE` keyword to be repeated before each role name, unlike the equivalent Hive statements.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch and HDFS files or directories, therefore no HDFS permissions are required.

#### **Related information:**

[Enabling Sentry Authorization for Impala](#) on page 91, [GRANT Statement \(CDH 5.2 or higher only\)](#) on page 201, [REVOKE Statement \(CDH 5.2 or higher only\)](#) on page 217, [DROP ROLE Statement \(CDH 5.2 or higher only\)](#) on page 193, [SHOW Statement](#) on page 242

## CREATE TABLE Statement

Creates a new table and specifies its characteristics. While creating a table, you optionally specify aspects such as:

- Whether the table is internal or external.
- The columns and associated data types.
- The columns used for physically partitioning the data.
- The file format for data files.
- The HDFS directory where the data files are located.

### Syntax:

The general syntax for creating a table and specifying its columns is as follows:

#### Explicit column definitions:

```
CREATE [EXTERNAL] TABLE [IF NOT EXISTS] [db_name.]table_name  
[ (col_name data_type [COMMENT 'col_comment'], ...) ]  
[ COMMENT 'table_comment' ]  
[PARTITIONED BY (col_name data_type [COMMENT 'col_comment'], ...) ]  
[WITH SERDEPROPERTIES ('key1'='value1', 'key2'='value2', ...) ]  
[  
    [ROW FORMAT row_format] [STORED AS file_format]  
]  
[LOCATION 'hdfs_path']  
[TBLPROPERTIES ('key1'='value1', 'key2'='value2', ...) ]  
[CACHED IN 'pool_name' [WITH REPLICATION = integer] | UNCACHED]
```

#### Column definitions inferred from data file:

```
CREATE [EXTERNAL] TABLE [IF NOT EXISTS] [db_name.]table_name  
LIKE PARQUET 'hdfs_path_of_parquet_file'  
[ COMMENT 'table_comment' ]  
[PARTITIONED BY (col_name data_type [COMMENT 'col_comment'], ...) ]  
[WITH SERDEPROPERTIES ('key1'='value1', 'key2'='value2', ...) ]  
[  
    [ROW FORMAT row_format] [STORED AS file_format]  
]  
[LOCATION 'hdfs_path']  
[TBLPROPERTIES ('key1'='value1', 'key2'='value2', ...) ]  
[CACHED IN 'pool_name' [WITH REPLICATION = integer] | UNCACHED]  
data_type  
: primitive_type
```

#### CREATE TABLE AS SELECT:

```
CREATE [EXTERNAL] TABLE [IF NOT EXISTS] db_name.table_name  
[ COMMENT 'table_comment' ]  
[WITH SERDEPROPERTIES ('key1'='value1', 'key2'='value2', ...) ]  
[  
    [ROW FORMAT row_format] [STORED AS file_format]  
]  
[LOCATION 'hdfs_path']  
[TBLPROPERTIES ('key1'='value1', 'key2'='value2', ...) ]  
[CACHED IN 'pool_name' [WITH REPLICATION = integer] | UNCACHED]  
AS  
select_statement
```

```
primitive_type  
: TINYINT  
| SMALLINT  
| INT  
| BIGINT  
| BOOLEAN  
| FLOAT  
| DOUBLE  
| DECIMAL
```

```

    STRING
    CHAR
    VARCHAR
    TIMESTAMP

row_format
  : DELIMITED [FIELDS TERMINATED BY 'char' [ESCAPED BY 'char']]]
  [LINES TERMINATED BY 'char']

file_format:
  PARQUET
  TEXTFILE
  AVRO
  SEQUENCEFILE
  RCFILE

```

**Statement type:** DDL**Internal and external tables (EXTERNAL and LOCATION clauses):**

By default, Impala creates an “internal” table, where Impala manages the underlying data files for the table, and physically deletes the data files when you drop the table. If you specify the `EXTERNAL` clause, Impala treats the table as an “external” table, where the data files are typically produced outside Impala and queried from their original locations in HDFS, and Impala leaves the data files in place when you drop the table. For details about internal and external tables, see [Tables](#) on page 156.

Typically, for an external table you include a `LOCATION` clause to specify the path to the HDFS directory where Impala reads and writes files for the table. For example, if your data pipeline produces Parquet files in the HDFS directory `/user/etl/destination`, you might create an external table as follows:

```
CREATE EXTERNAL TABLE external_parquet (c1 INT, c2 STRING, c3 TIMESTAMP)
  STORED AS PARQUET LOCATION '/user/etl/destination';
```

Although the `EXTERNAL` and `LOCATION` clauses are often specified together, `LOCATION` is optional for external tables, and you can also specify `LOCATION` for internal tables. The difference is all about whether Impala “takes control” of the underlying data files and moves them when you rename the table, or deletes them when you drop the table. For more about internal and external tables and how they interact with the `LOCATION` attribute, see [Internal Tables](#) on page 157.

**Partitioned tables (PARTITIONED BY clause):**

The `PARTITIONED BY` clause divides the data files based on the values from one or more specified columns. Impala queries can use the partition metadata to minimize the amount of data that is read from disk or transmitted across the network, particularly during join queries. For details about partitioning, see [Partitioning for Impala Tables](#) on page 394.

**Specifying file format (STORED AS and ROW FORMAT clauses):**

The `STORED AS` clause identifies the format of the underlying data files. Currently, Impala can query more types of file formats than it can create or insert into. Use Hive to perform any create or data load operations that are not currently available in Impala. For example, Impala can create a SequenceFile table but cannot insert data into it. There are also Impala-specific procedures for using compression with each kind of file format. For details about working with data files of various formats, see [How Impala Works with Hadoop File Formats](#) on page 399.

- **Note:** In Impala 1.4.0 and higher, Impala can create Avro tables, which formerly required doing the `CREATE TABLE` statement in Hive. See [Using the Avro File Format with Impala Tables](#) on page 417 for details and examples.

By default (when no `STORED AS` clause is specified), data files in Impala tables are created as text files with Ctrl-A (hex 01) characters as the delimiter. Specify the `ROW FORMAT DELIMITED` clause to produce or ingest data files that use a different delimiter character such as tab or |, or a different line end character such as carriage return or newline. When specifying delimiter and line end characters with the `FIELDS TERMINATED BY` and `LINES TERMINATED BY` clauses, use '\t' for tab, '\n' for newline or linefeed, '\r' for carriage return, and '\0

for ASCII nul (hex 00). For more examples of text tables, see [Using Text Data Files with Impala Tables](#) on page 400.

The `ESCAPED BY` clause applies both to text files that you create through an `INSERT` statement to an Impala `TEXTFILE` table, and to existing data files that you put into an Impala table directory. (You can ingest existing data files either by creating the table with `CREATE EXTERNAL TABLE ... LOCATION`, the `LOAD DATA` statement, or through an HDFS operation such as `hdfs dfs -put file hdfs_path`.) Choose an escape character that is not used anywhere else in the file, and put it in front of each instance of the delimiter character that occurs within a field value. Surrounding field values with quotation marks does not help Impala to parse fields with embedded delimiter characters; the quotation marks are considered to be part of the column value. If you want to use \ as the escape character, specify the clause in `impala-shell` as `ESCAPED BY '\\\\'`.

- **Note:** The `CREATE TABLE` clauses `FIELDS TERMINATED BY`, `ESCAPED BY`, and `LINES TERMINATED BY` have special rules for the string literal used for their argument, because they all require a single character. You can use a regular character surrounded by single or double quotation marks, an octal sequence such as '`\054`' (representing a comma), or an integer in the range '-127'..128' (with quotation marks but no backslash), which is interpreted as a single-byte ASCII character. Negative values are subtracted from 256; for example, `FIELDS TERMINATED BY '-2'` sets the field delimiter to ASCII code 254, the "Icelandic Thorn" character used as a delimiter by some data formats.

### Cloning tables (LIKE clause):

To create an empty table with the same columns, comments, and other attributes as another table, use the following variation. The `CREATE TABLE ... LIKE` form allows a restricted set of clauses, currently only the `LOCATION`, `COMMENT`, and `STORED AS` clauses.

```
CREATE [EXTERNAL] TABLE [IF NOT EXISTS] [db_name.]table_name  
  LIKE { [db_name.]table_name | PARQUET 'hdfs_path_of_parquet_file' }  
  [COMMENT 'table_comment']  
  [STORED AS file_format]  
  [LOCATION 'hdfs_path']
```

- **Note:** To clone the structure of a table and transfer data into it in a single operation, use the `CREATE TABLE AS SELECT` syntax described in the next subsection.

When you clone the structure of an existing table using the `CREATE TABLE ... LIKE` syntax, the new table keeps the same file format as the original one, so you only need to specify the `STORED AS` clause if you want to use a different file format, or when specifying a view as the original table. (Creating a table "like" a view produces a text table by default.)

Although normally Impala cannot create an HBase table directly, Impala can clone the structure of an existing HBase table with the `CREATE TABLE ... LIKE` syntax, preserving the file format and metadata from the original table.

There are some exceptions to the ability to use `CREATE TABLE ... LIKE` with an Avro table. For example, you cannot use this technique for an Avro table that is specified with an Avro schema but no columns. When in doubt, check if a `CREATE TABLE ... LIKE` operation works in Hive; if not, it typically will not work in Impala either.

If the original table is partitioned, the new table inherits the same partition key columns. Because the new table is initially empty, it does not inherit the actual partitions that exist in the original one. To create partitions in the new table, insert data or issue `ALTER TABLE ... ADD PARTITION` statements.

Prior to Impala 1.4.0, it was not possible to use the `CREATE TABLE LIKE view_name` syntax. In Impala 1.4.0 and higher, you can create a table with the same column definitions as a view using the `CREATE TABLE LIKE` technique. Although `CREATE TABLE LIKE` normally inherits the file format of the original table, a view has no underlying file format, so `CREATE TABLE LIKE view_name` produces a text table by default. To specify a different file format, include a `STORED AS file_format` clause at the end of the `CREATE TABLE LIKE` statement.

Because `CREATE TABLE ... LIKE` only manipulates table metadata, not the physical data of the table, issue `INSERT INTO TABLE` statements afterward to copy any data from the original table into the new one, optionally converting the data to a new file format. (For some file formats, Impala can do a `CREATE TABLE ... LIKE` to create the table, but Impala cannot insert data in that file format; in these cases, you must load the data in Hive. See [How Impala Works with Hadoop File Formats](#) on page 399 for details.)

### `CREATE TABLE AS SELECT`:

The `CREATE TABLE AS SELECT` syntax is a shorthand notation to create a table based on column definitions from another table, and copy data from the source table to the destination table without issuing any separate `INSERT` statement. This idiom is so popular that it has its own acronym, "CTAS".

See [SELECT Statement](#) on page 218 for details about query syntax for the `SELECT` portion of a `CREATE TABLE AS SELECT` statement.

The newly created table inherits the column names that you select from the original table, which you can override by specifying column aliases in the query. Any column or table comments from the original table are not carried over to the new table.

**Sorting considerations:** Although you can specify an `ORDER BY` clause in an `INSERT ... SELECT` statement, any `ORDER BY` clause is ignored and the results are not necessarily sorted. An `INSERT ... SELECT` operation potentially creates many different data files, prepared on different data nodes, and therefore the notion of the data being stored in sorted order is impractical.

For example, the following statements show how you can clone all the data in a table, or a subset of the columns and/or rows, or reorder columns, rename them, or construct them out of expressions:

```
-- Create new table and copy all data.
CREATE TABLE clone_of_t1 AS SELECT * FROM t1;
-- Same idea as CREATE TABLE LIKE, don't copy any data.
CREATE TABLE empty_clone_of_t1 AS SELECT * FROM t1 WHERE 1=0;
-- Copy some data.
CREATE TABLE subset_of_t1 AS SELECT * FROM t1 WHERE x > 100 AND y LIKE 'A%';
CREATE TABLE summary_of_t1 AS SELECT c1, sum(c2) AS total, avg(c2) AS average FROM t1
  GROUP BY c2;
-- Switch file format.
CREATE TABLE parquet_version_of_t1 STORED AS PARQUET AS SELECT * FROM t1;
-- Create tables with different column order, names, or types than the original.
CREATE TABLE some_columns_from_t1 AS SELECT c1, c3, c5 FROM t1;
CREATE TABLE reordered_columns_from_t1 AS SELECT c4, c3, c1, c2 FROM t1;
CREATE TABLE synthesized_columns AS SELECT upper(c1) AS all_caps, c2+c3 AS total,
  "California" AS state FROM t1;
```

As part of a CTAS operation, you can convert the data to any file format that Impala can write (currently, `TEXTFILE` and `PARQUET`). You cannot specify the lower-level properties of a text table, such as the delimiter. Although you can use a partitioned table as the source and copy data from it, you cannot specify any partitioning clauses for the new table.

### `CREATE TABLE LIKE PARQUET`:

The variation `CREATE TABLE ... LIKE PARQUET 'hdfs_path_of_parquet_file'` lets you skip the column definitions of the `CREATE TABLE` statement. The column names and data types are automatically configured based on the organization of the specified Parquet data file, which must already reside in HDFS. You can use a data file located outside the Impala database directories, or a file from an existing Impala Parquet table; either way, Impala only uses the column definitions from the file and does not use the HDFS location for the `LOCATION` attribute of the new table. (Although you can also specify the enclosing directory with the `LOCATION` attribute, to both use the same schema as the data file and point the Impala table at the associated directory for querying.)

The following considerations apply when you use the `CREATE TABLE LIKE PARQUET` technique:

- Any column comments from the original table are not preserved in the new table. Each column in the new table has a comment stating the low-level Parquet field type used to deduce the appropriate SQL column type.
- If you use a data file from a partitioned Impala table, any partition key columns from the original table are left out of the new table, because they are represented in HDFS directory names rather than stored in the

data file. To preserve the partition information, repeat the same PARTITION clause as in the original CREATE TABLE statement.

- The file format of the new table defaults to text, as with other kinds of CREATE TABLE statements. To make the new table also use Parquet format, include the clause STORED AS PARQUET in the CREATE TABLE LIKE PARQUET statement.
- If the Parquet data file comes from an existing Impala table, currently, any TINYINT or SMALLINT columns are turned into INT columns in the new table. Internally, Parquet stores such values as 32-bit integers.
- When the destination table uses the Parquet file format, the CREATE TABLE AS SELECT and INSERT ... SELECT statements always create at least one data file, even if the SELECT part of the statement does not match any rows. You can use such an empty Parquet data file as a template for subsequent CREATE TABLE LIKE PARQUET statements.

For more details about creating Parquet tables, and examples of the CREATE TABLE LIKE PARQUET syntax, see [Using the Parquet File Format with Impala Tables](#) on page 407.

### Visibility and Metadata (TBLPROPERTIES and WITH SERDEPROPERTIES clauses):

You can associate arbitrary items of metadata with a table by specifying the TBLPROPERTIES clause. This clause takes a comma-separated list of key-value pairs and stores those items in the metastore database. You can also change the table properties later with an ALTER TABLE statement. You can observe the table properties for different delimiter and escape characters using the DESCRIBE FORMATTED command, and change those settings for an existing table with ALTER TABLE ... SET TBLPROPERTIES.

You can also associate SerDes properties with the table by specifying key-value pairs through the WITH SERDEPROPERTIES clause. This metadata is not used by Impala, which has its own built-in serializer and deserializer for the file formats it supports. Particular property values might be needed for Hive compatibility with certain variations of file formats, particularly Avro.

Some DDL operations that interact with other Hadoop components require specifying particular values in the SERDEPROPERTIES or TBLPROPERTIES fields, such as creating an Avro table or an HBase table. (You typically create HBase tables in Hive, because they require additional clauses not currently available in Impala.)

To see the column definitions and column comments for an existing table, for example before issuing a CREATE TABLE ... LIKE or a CREATE TABLE ... AS SELECT statement, issue the statement DESCRIBE *table\_name*. To see even more detail, such as the location of data files and the values for clauses such as ROW FORMAT and STORED AS, issue the statement DESCRIBE FORMATTED *table\_name*. DESCRIBE FORMATTED is also needed to see any overall table comment (as opposed to individual column comments).

After creating a table, your impala-shell session or another impala-shell connected to the same node can immediately query that table. There might be a brief interval (one statestore heartbeat) before the table can be queried through a different Impala node. To make the CREATE TABLE statement return only when the table is recognized by all Impala nodes in the cluster, enable the SYNC\_DDL query option.

### HDFS caching (CACHED IN clause):

If you specify the CACHED IN clause, any existing or future data files in the table directory or the partition subdirectories are designated to be loaded into memory with the HDFS caching mechanism. See [Using HDFS Caching with Impala \(CDH 5.1 or higher only\)](#) on page 372 for details about using the HDFS caching feature.

In Impala 2.2 / CDH 5.4 and higher, the optional WITH REPLICATION clause for CREATE TABLE and ALTER TABLE lets you specify a *replication factor*, the number of hosts on which to cache the same data blocks. When Impala processes a cached data block, where the cache replication factor is greater than 1, Impala randomly selects a host that has a cached copy of that data block. This optimization avoids excessive CPU usage on a single host when the same cached data block is processed multiple times.

### Hive considerations:

Impala queries can make use of metadata about the table and columns, such as the number of rows in a table or the number of different values in a column. Prior to Impala 1.2.2, to create this metadata, you issued the ANALYZE TABLE statement in Hive to gather this information, after creating the table and loading representative data into it. In Impala 1.2.2 and higher, the COMPUTE STATS statement produces these statistics within Impala, without needing to use Hive at all.

### Column order:

If you intend to use the table to hold data files produced by some external source, specify the columns in the same order as they appear in the data files.

If you intend to insert or copy data into the table through Impala, or if you have control over the way externally produced data files are arranged, use your judgment to specify columns in the most convenient order:

- If certain columns are often `NULL`, specify those columns last. You might produce data files that omit these trailing columns entirely. Impala automatically fills in the `NULL` values if so.
- If an unpartitioned table will be used as the source for an `INSERT ... SELECT` operation into a partitioned table, specify `last` in the unpartitioned table any columns that correspond to partition key columns in the partitioned table, and in the same order as the partition key columns are declared in the partitioned table. This technique lets you use `INSERT ... SELECT *` when copying data to the partitioned table, rather than specifying each column name individually.
- If you specify columns in an order that you later discover is suboptimal, you can sometimes work around the problem without recreating the table. You can create a view that selects columns from the original table in a permuted order, then do a `SELECT *` from the view. When inserting data into a table, you can specify a permuted order for the inserted columns to match the order in the destination table.

### HBase considerations:

- **Note:**

The Impala `CREATE TABLE` statement cannot create an HBase table, because it currently does not support the `STORED BY` clause needed for HBase tables. Create such tables in Hive, then query them through Impala. For information on using Impala with HBase tables, see [Using Impala to Query HBase Tables](#) on page 426.

### Amazon S3 considerations:

To create a table where the data resides in the Amazon Simple Storage Service (S3), specify a `s3a://` prefix `LOCATION` attribute pointing to the data files in S3. You can use this special `LOCATION` syntax when creating an empty table, but not as part of a `CREATE TABLE AS SELECT` statement. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details.

**Sorting considerations:** Although you can specify an `ORDER BY` clause in an `INSERT ... SELECT` statement, any `ORDER BY` clause is ignored and the results are not necessarily sorted. An `INSERT ... SELECT` operation potentially creates many different data files, prepared on different data nodes, and therefore the notion of the data being stored in sorted order is impractical.

### HDFS considerations:

The `CREATE TABLE` statement for an internal table creates a directory in HDFS. The `CREATE EXTERNAL TABLE` statement associates the table with an existing HDFS directory, and does not create any new directory in HDFS. To locate the HDFS data directory for a table, issue a `DESCRIBE FORMATTED` *table* statement. To examine the contents of that HDFS directory, use an OS command such as `hdfs dfs -ls hdfs://path`, either from the OS command line or through the `shell` or `! commands` in `impala-shell`.

The `CREATE TABLE AS SELECT` syntax creates data files under the table data directory to hold any data copied by the `INSERT` portion of the statement. (Even if no data is copied, Impala might create one or more empty data files.)

### HDFS permissions:

#### Security considerations:

If these statements in your environment contain sensitive literal values such as credit card numbers or tax identifiers, Impala can redact this sensitive information when displaying the statements in log files and other administrative contexts. See [Sensitive Data Redaction](#) for details.

**Cancellation:** Certain multi-stage statements (CREATE TABLE AS SELECT and COMPUTE STATS) can be cancelled during some stages, when running INSERT or SELECT operations internally. To cancel this statement, use Ctrl-C from the impala-shell interpreter, the Cancel button from the Watch page in Hue, Actions > Cancel from the Queries list in Cloudera Manager, or Cancel from the list of in-flight queries (for a particular node) on the Queries tab in the Impala web UI (port 25000).

### Related information:

[Tables](#) on page 156, [ALTER TABLE Statement](#) on page 164, [DROP TABLE Statement](#) on page 197, [Partitioning for Impala Tables](#) on page 394, [Internal Tables](#) on page 157, [External Tables](#) on page 157, [COMPUTE STATS Statement](#) on page 170, [SYNC\\_DDL Query Option](#) on page 352, [SHOW TABLES Statement](#) on page 249, [SHOW CREATE TABLE Statement](#) on page 250, [DESCRIBE Statement](#) on page 188

## CREATE VIEW Statement

The CREATE VIEW statement lets you create a shorthand abbreviation for a more complicated query. The base query can involve joins, expressions, reordered columns, column aliases, and other SQL features that can make a query hard to understand or maintain.

Because a view is purely a logical construct (an alias for a query) with no physical data behind it, ALTER VIEW only involves changes to metadata in the metastore database, not any data files in HDFS.

### Syntax:

```
CREATE VIEW [IF NOT EXISTS] view_name [(column_list)]
    AS select_statement
```

**Statement type:** DDL

### Usage notes:

The CREATE VIEW statement can be useful in scenarios such as the following:

- To turn even the most lengthy and complicated SQL query into a one-liner. You can issue simple queries against the view from applications, scripts, or interactive queries in impala-shell. For example:

```
select * from view_name;
select * from view_name order by c1 desc limit 10;
```

The more complicated and hard-to-read the original query, the more benefit there is to simplifying the query using a view.

- To hide the underlying table and column names, to minimize maintenance problems if those names change. In that case, you re-create the view using the new names, and all queries that use the view rather than the underlying tables keep running with no changes.
- To experiment with optimization techniques and make the optimized queries available to all applications. For example, if you find a combination of WHERE conditions, join order, join hints, and so on that works the best for a class of queries, you can establish a view that incorporates the best-performing techniques. Applications can then make relatively simple queries against the view, without repeating the complicated and optimized logic over and over. If you later find a better way to optimize the original query, when you re-create the view, all the applications immediately take advantage of the optimized base query.
- To simplify a whole class of related queries, especially complicated queries involving joins between multiple tables, complicated expressions in the column list, and other SQL syntax that makes the query difficult to understand and debug. For example, you might create a view that joins several tables, filters using several WHERE conditions, and selects several columns from the result set. Applications might issue queries against this view that only vary in their LIMIT, ORDER BY, and similar simple clauses.

For queries that require repeating complicated clauses over and over again, for example in the select list, ORDER BY, and GROUP BY clauses, you can use the WITH clause as an alternative to creating a view.

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

#### **Security considerations:**

If these statements in your environment contain sensitive literal values such as credit card numbers or tax identifiers, Impala can redact this sensitive information when displaying the statements in log files and other administrative contexts. See [Sensitive Data Redaction](#) for details.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

#### **Examples:**

```
-- Create a view that is exactly the same as the underlying table.
create view v1 as select * from t1;

-- Create a view that includes only certain columns from the underlying table.
create view v2 as select c1, c3, c7 from t1;

-- Create a view that filters the values from the underlying table.
create view v3 as select distinct c1, c3, c7 from t1 where c1 is not null and c5 > 0;

-- Create a view that reorders and renames columns from the underlying table.
create view v4 as select c4 as last_name, c6 as address, c2 as birth_date from t1;

-- Create a view that runs functions to convert or transform certain columns.
create view v5 as select c1, cast(c3 as string) c3, concat(c4,c5) c5, trim(c6) c6,
"Constant" c8 from t1;

-- Create a view that hides the complexity of a view query.
create view v6 as select t1.c1, t2.c2 from t1 join t2 on t1.id = t2.id;
```

The following example creates a series of views and then drops them. These examples illustrate how views are associated with a particular database, and both the view definitions and the view names for `CREATE VIEW` and `DROP VIEW` can refer to a view in the current database or a fully qualified view name.

```
-- Create and drop a view in the current database.
CREATE VIEW few_rows_from_t1 AS SELECT * FROM t1 LIMIT 10;
DROP VIEW few_rows_from_t1;

-- Create and drop a view referencing a table in a different database.
CREATE VIEW table_from_other_db AS SELECT x FROM db1.foo WHERE x IS NOT NULL;
DROP VIEW table_from_other_db;

USE db1;
-- Create a view in a different database.
CREATE VIEW db2.v1 AS SELECT * FROM db2.foo;
-- Switch into the other database and drop the view.
USE db2;
DROP VIEW v1;

USE db1;
-- Create a view in a different database.
CREATE VIEW db2.v1 AS SELECT * FROM db2.foo;
-- Drop a view in the other database.
DROP VIEW db2.v1;
```

#### **Related information:**

[Views](#) on page 158, [ALTER VIEW Statement](#) on page 168, [DROP VIEW Statement](#) on page 198

## DESCRIBE Statement

The `DESCRIBE` statement displays metadata about a table, such as the column names and their data types. Its syntax is:

```
DESCRIBE [FORMATTED] [db_name.]table_name
```

You can use the abbreviation `DESC` for the `DESCRIBE` statement.

The `DESCRIBE FORMATTED` variation displays additional information, in a format familiar to users of Apache Hive. The extra information includes low-level details such as whether the table is internal or external, when it was created, the file format, the location of the data in HDFS, whether the object is a table or a view, and (for views) the text of the query from the view definition.

- **Note:** The `Compressed` field is not a reliable indicator of whether the table contains compressed data. It typically always shows `No`, because the compression settings only apply during the session that loads data and are not stored persistently with the table metadata.

### Usage notes:

After the `impalad` daemons are restarted, the first query against a table can take longer than subsequent queries, because the metadata for the table is loaded before the query is processed. This one-time delay for each table can cause misleading results in benchmark tests or cause unnecessary concern. To "warm up" the Impala metadata cache, you can issue a `DESCRIBE` statement in advance for each table you intend to access later.

When you are dealing with data files stored in HDFS, sometimes it is important to know details such as the path of the data files for an Impala table, and the host name for the namenode. You can get this information from the `DESCRIBE FORMATTED` output. You specify HDFS URLs or path specifications with statements such as `LOAD DATA` and the `LOCATION` clause of `CREATE TABLE` or `ALTER TABLE`. You might also use HDFS URLs or paths with Linux commands such as `hadoop` and `hdfs` to copy, rename, and so on, data files in HDFS.

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

Each table can also have associated table statistics and column statistics. To see these categories of information, use the `SHOW TABLE STATS table_name` and `SHOW COLUMN STATS table_name` statements. See [SHOW Statement](#) on page 242 for details.

- **Important:** After adding or replacing data in a table used in performance-critical queries, issue a `COMPUTE STATS` statement to make sure all statistics are up-to-date. Consider updating statistics for a table after any `INSERT`, `LOAD DATA`, or `CREATE TABLE AS SELECT` statement in Impala, or after loading data through Hive and doing a `REFRESH table_name` in Impala. This technique is especially important for tables that are very large, used in join queries, or both.

### Examples:

The following example shows the results of both a standard `DESCRIBE` and `DESCRIBE FORMATTED` for different kinds of schema objects:

- `DESCRIBE` for a table or a view returns the name, type, and comment for each of the columns. For a view, if the column value is computed by an expression, the column name is automatically generated as `_c0`, `_c1`, and so on depending on the ordinal number of the column.
- A table created with no special format or storage clauses is designated as a `MANAGED_TABLE` (an "internal table" in Impala terminology). Its data files are stored in an HDFS directory under the default Hive data directory. By default, it uses Text data format.
- A view is designated as `VIRTUAL_VIEW` in `DESCRIBE FORMATTED` output. Some of its properties are `NULL` or blank because they are inherited from the base table. The text of the query that defines the view is part of the `DESCRIBE FORMATTED` output.

- A table with additional clauses in the CREATE TABLE statement has differences in DESCRIBE FORMATTED output. The output for t2 includes the EXTERNAL\_TABLE keyword because of the CREATE EXTERNAL TABLE syntax, and different InputFormat and OutputFormat fields to reflect the Parquet file format.

```
[localhost:21000] > create table t1 (x int, y int, s string);
Query: create table t1 (x int, y int, s string)
[localhost:21000] > describe t1;
Query: describe t1
Query finished, fetching results ...
+-----+-----+
| name | type   | comment |
+-----+-----+
| x    | int    |
| y    | int    |
| s    | string |
+-----+-----+
Returned 3 row(s) in 0.13s
[localhost:21000] > describe formatted t1;
Query: describe formatted t1
Query finished, fetching results ...
+-----+-----+
| name           | comment          | type      |
+-----+-----+
| # col_name     | comment          | data_type |
|                 | NULL             | NULL      |
| x              | NULL             | int       |
| y              | None             | int       |
| s              | None             | string    |
|                 | NULL             | NULL      |
| # Detailed Table Information | NULL            |
| Database:      | NULL             | describe_formatted |
| Owner:         | NULL             | cloudera  |
| CreateTime:    | NULL             | Mon Jul 22 17:03:16 EDT 2013 |
| LastAccessTime: | NULL             | UNKNOWN    |
| Protect Mode:  | NULL             | None      |
| Retention:     | NULL             | 0         |
| Location:      | NULL             |
hdfs://127.0.0.1:8020/user/hive/warehouse/describe_formatted.db/t1 | NULL
| Table Type:    | MANAGED_TABLE   |
| Table Parameters: | NULL            |
|                 | transient_lastDdlTime |
|                 | 1374526996     |
|                 | NULL             |
| # Storage Information | NULL            |
| SerDe Library: | org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe |
| InputFormat:   | org.apache.hadoop.mapred.TextInputFormat |
| OutputFormat: | org.apache.hadoop.hive.io.HiveIgnoreKeyTextOutputFormat | NULL
| Compressed:   | No               |
| Num Buckets: | 0                |
```

```

| NULL           | []
| Bucket Columns: | []
| NULL           | []
| Sort Columns:  | []
+-----+
Returned 26 row(s) in 0.03s
[localhost:21000] > create view v1 as select x, upper(s) from t1;
Query: create view v1 as select x, upper(s) from t1
[localhost:21000] > describe v1;
Query: describe v1
Query finished, fetching results ...
+-----+-----+
| name | type   | comment |
+-----+-----+
| x    | int    |          |
| _cl  | string |          |
+-----+-----+
Returned 2 row(s) in 0.10s
[localhost:21000] > describe formatted v1;
Query: describe formatted v1
Query finished, fetching results ...
+-----+-----+-----+
| name           | type      | comment |
+-----+-----+-----+
| # col_name     | data_type | comment |
|                 | NULL       | NULL     |
| x              | int        | None     |
| _cl            | string     | None     |
|                 | NULL       | NULL     |
| # Detailed Table Information | NULL      | NULL     |
| Database:      | describe_formatted | NULL     |
| Owner:         | cloudera  | NULL     |
| CreateTime:    | Mon Jul 22 16:56:38 EDT 2013 | NULL     |
| LastAccessTime: | UNKNOWN    | NULL     |
| Protect Mode:  | None       | NULL     |
| Retention:     | 0          | NULL     |
| Table Type:    | VIRTUAL_VIEW | NULL     |
| Table Parameters: | NULL      | NULL     |
|                   | transient_lastDdlTime | 1374526598 |
|                   | NULL       | NULL     |
| # Storage Information | NULL      | NULL     |
| SerDe Library:  | null      | NULL     |
| InputFormat:    | null      | NULL     |
| OutputFormat:   | null      | NULL     |
| Compressed:     | No        | NULL     |
| Num Buckets:   | 0          | NULL     |
| Bucket Columns: | []         | NULL     |
| Sort Columns:   | []         | NULL     |

```

```

| NULL           | NULL           |
| # View Information | NULL           |
| View Original Text: | SELECT x, upper(s) FROM t1 | NULL
| View Expanded Text: | SELECT x, upper(s) FROM t1 | NULL
+-----+
Returned 28 row(s) in 0.03s
[localhost:21000] > create external table t2 (x int, y int, s string) stored as parquet
location '/user/cloudera/sample_data';
[localhost:21000] > describe formatted t2;
Query: describe formatted t2
Query finished, fetching results ...
+-----+-----+-----+
| name          | type           |
| comment       |                |
+-----+-----+-----+
| # col_name    | data_type      |
| comment       |                |
| NULL          | NULL           |
| x             | int            |
| None          |                |
| y             | int            |
| None          |                |
| s             | string          |
| None          |                |
| NULL          | NULL           |
| # Detailed Table Information | NULL           |
| NULL          |                |
| Database:     | describe_formatted |
| NULL          |                |
| Owner:        | cloudera      |
| NULL          |                |
| CreateTime:   | Mon Jul 22 17:01:47 EDT 2013 |
| NULL          |                |
| LastAccessTime: | UNKNOWN        |
| NULL          |                |
| Protect Mode: | None           |
| NULL          |                |
| Retention:    | 0              |
| NULL          |                |
| Location:     | hdfs://127.0.0.1:8020/user/cloudera/sample_data |
| NULL          |                |
| Table Type:   | EXTERNAL_TABLE |
| NULL          |                |
| Table Parameters: | NULL           |
| NULL          |                |
| TRUE          | EXTERNAL       |
| 1374526907   | transient_lastDdlTime |
| NULL          | NULL           |
| # Storage Information | NULL           |
| NULL          |                |
| SerDe Library: | org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe |
| NULL          |                |
| InputFormat:  | com.cloudera.impala.hive.serde.ParquetInputFormat |
| NULL          |                |
| OutputFormat: | com.cloudera.impala.hive.serde.ParquetOutputFormat |
| NULL          |                |
| Compressed:   | No             |
| NULL          |                |
| Num Buckets: | 0              |
| NULL          |                |
| Bucket Columns: | [ ]            |
| NULL          |                |
| Sort Columns: | [ ]            |
+-----+

```

```
|  
+-----+  
| NULL |  
+-----+  
Returned 27 row(s) in 0.17s
```

**Cancellation:** Cannot be cancelled.

**HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read and execute permissions for all directories that are part of the table. (A table could span multiple different HDFS directories if it is partitioned. The directories could be widely scattered because a partition can reside in an arbitrary HDFS directory based on its `LOCATION` attribute.)

**Related information:**

[Tables](#) on page 156, [CREATE TABLE Statement](#) on page 180, [SHOW TABLES Statement](#) on page 249, [SHOW CREATE TABLE Statement](#) on page 250

## DROP DATABASE Statement

Removes a database from the system, and deletes the corresponding `*.db` directory from HDFS. The database must be empty before it can be dropped, to avoid losing any data.

**Syntax:**

```
DROP (DATABASE|SCHEMA) [IF EXISTS] database_name;
```

**Statement type:** DDL

**Usage notes:**

Before dropping a database, use a combination of `DROP TABLE`, `DROP VIEW`, `ALTER TABLE`, and `ALTER VIEW` statements, to drop all the tables and views in the database or move them to other databases.

**Hive considerations:**

When you drop a database in Impala, the database can no longer be used by Hive.

**Examples:**

See [CREATE DATABASE Statement](#) on page 175 for examples covering `CREATE DATABASE`, `USE`, and `DROP DATABASE`.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have write permission for the directory associated with the database.

**Related information:**

[Databases](#) on page 155, [CREATE DATABASE Statement](#) on page 175, [USE Statement](#) on page 256, [SHOW DATABASES](#) on page 248

## DROP FUNCTION Statement

Removes a user-defined function (UDF), so that it is not available for execution during Impala `SELECT` or `INSERT` operations.

**Syntax:**

```
DROP [AGGREGATE] FUNCTION [IF EXISTS] [db_name.]function_name(type[, type...])
```

**Statement type:** DDL

**Usage notes:**

Because the same function name could be overloaded with different argument signatures, you specify the argument types to identify the exact function to drop.

#### **Restrictions:**

Currently, Impala UDFs and UDAFs are not persisted in the metastore database. Information about these functions is held in the memory of the `catalogd` daemon. You must reload them by running the `CREATE FUNCTION` statements again each time you restart the `catalogd` daemon.

**Cancellation:** Cannot be cancelled.

#### **HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, does not need any particular HDFS permissions to perform this statement. All read and write operations are on the metastore database, not HDFS files and directories.

#### **Related information:**

[Impala User-Defined Functions \(UDFs\)](#) on page 308, [CREATE FUNCTION Statement](#) on page 177

## DROP ROLE Statement (CDH 5.2 or higher only)

The `DROP ROLE` statement removes a role from the metastore database. Once dropped, the role is revoked for all users to whom it was previously assigned, and all privileges granted to that role are revoked. Queries that are already executing are not affected. Impala verifies the role information approximately every 60 seconds, so the effects of `DROP ROLE` might not take effect for new Impala queries for a brief period.

#### **Syntax:**

```
DROP ROLE role_name
```

#### **Required privileges:**

Only administrative users (initially, a predefined set of users specified in the Sentry service configuration file) can use this statement.

#### **Compatibility:**

Impala makes use of any roles and privileges specified by the `GRANT` and `REVOKE` statements in Hive, and Hive makes use of any roles and privileges specified by the `GRANT` and `REVOKE` statements in Impala. The Impala `GRANT` and `REVOKE` statements for privileges do not require the `ROLE` keyword to be repeated before each role name, unlike the equivalent Hive statements.

#### **Related information:**

[Enabling Sentry Authorization for Impala](#) on page 91, [GRANT Statement \(CDH 5.2 or higher only\)](#) on page 201, [REVOKE Statement \(CDH 5.2 or higher only\)](#) on page 217, [CREATE ROLE Statement \(CDH 5.2 or higher only\)](#) on page 179, [SHOW Statement](#) on page 242

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

## DROP STATS Statement

Removes the specified statistics from a table or partition. The statistics were originally created by the `COMPUTE STATS` or `COMPUTE INCREMENTAL STATS` statement.

#### **Syntax:**

```
DROP STATS [database_name.]table_name
DROP INCREMENTAL STATS [database_name.]table_name PARTITION (partition_spec)
```

```
partition_spec ::= partition_col=constant_value
```

The PARTITION clause is only allowed in combination with the INCREMENTAL clause. It is optional for COMPUTE INCREMENTAL STATS, and required for DROP INCREMENTAL STATS. Whenever you specify partitions through the PARTITION (*partition\_spec*) clause in a COMPUTE INCREMENTAL STATS or DROP INCREMENTAL STATS statement, you must include all the partitioning columns in the specification, and specify constant values for all the partition key columns.

DROP STATS removes all statistics from the table, whether created by COMPUTE STATS or COMPUTE INCREMENTAL STATS.

DROP INCREMENTAL STATS only affects incremental statistics for a single partition, specified through the PARTITION clause. The incremental stats are marked as outdated, so that they are recomputed by the next COMPUTE INCREMENTAL STATS statement.

#### Usage notes:

You typically use this statement when the statistics for a table or a partition have become stale due to data files being added to or removed from the associated HDFS data directories, whether by manual HDFS operations or INSERT, INSERT OVERWRITE, or LOAD DATA statements, or adding or dropping partitions.

When a table or partition has no associated statistics, Impala treats it as essentially zero-sized when constructing the execution plan for a query. In particular, the statistics influence the order in which tables are joined in a join query. To ensure proper query planning and good query performance and scalability, make sure to run COMPUTE STATS or COMPUTE INCREMENTAL STATS on the table or partition after removing any stale statistics.

Dropping the statistics is not required for an unpartitioned table or a partitioned table covered by the original type of statistics. A subsequent COMPUTE STATS statement replaces any existing statistics with new ones, for all partitions, regardless of whether the old ones were outdated. Therefore, this statement was rarely used before the introduction of incremental statistics.

Dropping the statistics is required for a partitioned table containing incremental statistics, to make a subsequent COMPUTE INCREMENTAL STATS statement rescan an existing partition. See [How Impala Uses Statistics for Query Optimization](#) on page 363 for information about incremental statistics, a new feature available in Impala 2.1.0 and higher.

#### Statement type: DDL

**Cancellation:** Cannot be cancelled.

#### HDFS permissions:

The user ID that the impalad daemon runs under, typically the impala user, does not need any particular HDFS permissions to perform this statement. All read and write operations are on the metastore database, not HDFS files and directories.

#### Examples:

The following example shows a partitioned table that has associated statistics produced by the COMPUTE INCREMENTAL STATS statement, and how the situation evolves as statistics are dropped from specific partitions, then the entire table.

Initially, all table and column statistics are filled in.

show table stats item_partitioned;						
i_category	#Rows	#Files	Size	Bytes Cached	Format	Incremental stats
Books	1733	1	223.74KB	NOT CACHED	PARQUET	true
Children	1786	1	230.05KB	NOT CACHED	PARQUET	true
Electronics	1812	1	232.67KB	NOT CACHED	PARQUET	true
Home	1807	1	232.56KB	NOT CACHED	PARQUET	true
Jewelry	1740	1	223.72KB	NOT CACHED	PARQUET	true
Men	1811	1	231.25KB	NOT CACHED	PARQUET	true

Music	1860	1	237.90KB	NOT CACHED	PARQUET	true
Shoes	1835	1	234.90KB	NOT CACHED	PARQUET	true
Sports	1783	1	227.97KB	NOT CACHED	PARQUET	true
Women	1790	1	226.27KB	NOT CACHED	PARQUET	true
Total	17957	10	2.25MB	0B		

show column stats item_partitioned;						
Column	Type	#Distinct Values	#Nulls	Max Size	Avg Size	
i_item_sk	INT	19443	-1	4	4	
i_item_id	STRING	9025	-1	16	16	
i_rec_start_date	TIMESTAMP	4	-1	16	16	
i_rec_end_date	TIMESTAMP	3	-1	16	16	
i_item_desc	STRING	13330	-1	200	100.3028030395	
i_current_price	FLOAT	2807	-1	4	4	
i_wholesale_cost	FLOAT	2105	-1	4	4	
i_brand_id	INT	965	-1	4	4	
i_brand	STRING	725	-1	22	16.17760086059	
i_class_id	INT	16	-1	4	4	
i_class	STRING	101	-1	15	7.767499923706	
i_category_id	INT	10	-1	4	4	
i_manufact_id	INT	1857	-1	4	4	
i_manufact	STRING	1028	-1	15	11.32950019836	
i_size	STRING	8	-1	11	4.334599971771	
i_formulation	STRING	12884	-1	20	19.97999954223	
i_color	STRING	92	-1	10	5.380899906158	
i_units	STRING	22	-1	7	4.186900138854	
i_container	STRING	2	-1	7	6.992599964141	
i_manager_id	INT	105	-1	4	4	
i_product_name	STRING	19094	-1	25	18.02330017089	
i_category	STRING	10	0	-1	-1	

To remove statistics for particular partitions, use the `DROP INCREMENTAL STATS` statement. After removing statistics for two partitions, the table-level statistics reflect that change in the `#Rows` and `Incremental stats` fields. The counts, maximums, and averages of the column-level statistics are unaffected.

- Note:** (It is possible that the row count might be preserved in future after a `DROP INCREMENTAL STATS` statement. Check the resolution of the issue [IMPALA-1615](#).)

drop incremental stats item_partitioned partition (i_category='Sports');						
drop incremental stats item_partitioned partition (i_category='Electronics');						
show table stats item_partitioned						
+-----+-----+-----+-----+-----+-----+-----+						
i_category	#Rows	#Files	Size	Bytes Cached	Format	Incremental stats
Books	1733	1	223.74KB	NOT CACHED	PARQUET	true
Children	1786	1	230.05KB	NOT CACHED	PARQUET	true
Electronics	-1	1	232.67KB	NOT CACHED	PARQUET	false
Home	1807	1	232.56KB	NOT CACHED	PARQUET	true
Jewelry	1740	1	223.72KB	NOT CACHED	PARQUET	true
Men	1811	1	231.25KB	NOT CACHED	PARQUET	true
Music	1860	1	237.90KB	NOT CACHED	PARQUET	true
Shoes	1835	1	234.90KB	NOT CACHED	PARQUET	true
Sports	-1	1	227.97KB	NOT CACHED	PARQUET	false
Women	1790	1	226.27KB	NOT CACHED	PARQUET	true
Total	17957	10	2.25MB	0B		
+-----+-----+-----+-----+-----+-----+-----+						
show column stats item_partitioned						
+-----+-----+-----+-----+-----+-----+-----+						
Column	Type	#Distinct Values	#Nulls	Max Size	Avg Size	
i_item_sk	INT	19443	-1	4	4	
i_item_id	STRING	9025	-1	16	16	
i_rec_start_date	TIMESTAMP	4	-1	16	16	
i_rec_end_date	TIMESTAMP	3	-1	16	16	
i_item_desc	STRING	13330	-1	200	100.3028030395	

i_current_price	FLOAT	2807	-1	4	4
i_wholesale_cost	FLOAT	2105	-1	4	4
i_brand_id	INT	965	-1	4	4
i_brand	STRING	725	-1	22	16.17760086059
i_class_id	INT	16	-1	4	4
i_class	STRING	101	-1	15	7.767499923706
i_category_id	INT	10	-1	4	4
i_manufact_id	INT	1857	-1	4	4
i_manufact	STRING	1028	-1	15	11.32950019836
i_size	STRING	8	-1	11	4.334599971771
i_formulation	STRING	12884	-1	20	19.97999954223
i_color	STRING	92	-1	10	5.380899906158
i_units	STRING	22	-1	7	4.186900138854
i_container	STRING	2	-1	7	6.992599964141
i_manager_id	INT	105	-1	4	4
i_product_name	STRING	19094	-1	25	18.02330017089
i_category	STRING	10	0	-1	-1

To remove all statistics from the table, whether produced by COMPUTE STATS or COMPUTE INCREMENTAL STATS, use the DROP STATS statement without the INCREMENTAL clause). Now, both table-level and column-level statistics are reset.

drop stats item_partitioned;
show table stats item_partitioned
+-----+-----+-----+-----+-----+-----+-----+-----+
i_category   #Rows   #Files   Size   Bytes Cached   Format   Incremental stats
+-----+-----+-----+-----+-----+-----+-----+-----+
Books -1 1 223.74KB NOT CACHED PARQUET false
Children -1 1 230.05KB NOT CACHED PARQUET false
Electronics -1 1 232.67KB NOT CACHED PARQUET false
Home -1 1 232.56KB NOT CACHED PARQUET false
Jewelry -1 1 223.72KB NOT CACHED PARQUET false
Men -1 1 231.25KB NOT CACHED PARQUET false
Music -1 1 237.90KB NOT CACHED PARQUET false
Shoes -1 1 234.90KB NOT CACHED PARQUET false
Sports -1 1 227.97KB NOT CACHED PARQUET false
Women -1 1 226.27KB NOT CACHED PARQUET false
Total -1 10 2.25MB 0B
+-----+-----+-----+-----+-----+-----+-----+-----+
show column stats item_partitioned
+-----+-----+-----+-----+-----+-----+-----+
Column   Type   #Distinct Values   #Nulls   Max Size   Avg Size
+-----+-----+-----+-----+-----+-----+-----+
i_item_sk INT -1 -1 4 4
i_item_id STRING -1 -1 -1 -1
i_rec_start_date TIMESTAMP -1 -1 16 16
i_rec_end_date TIMESTAMP -1 -1 16 16
i_item_desc STRING -1 -1 -1 -1
i_current_price FLOAT -1 -1 4 4
i_wholesale_cost FLOAT -1 -1 4 4
i_brand_id INT -1 -1 4 4
i_brand STRING -1 -1 -1 -1
i_class_id INT -1 -1 4 4
i_class STRING -1 -1 -1 -1
i_category_id INT -1 -1 4 4
i_manufact_id INT -1 -1 4 4
i_manufact STRING -1 -1 -1 -1
i_size STRING -1 -1 -1 -1
i_formulation STRING -1 -1 -1 -1
i_color STRING -1 -1 -1 -1
i_units STRING -1 -1 -1 -1
i_container STRING -1 -1 -1 -1
i_manager_id INT -1 -1 4 4
i_product_name STRING -1 -1 -1 -1
i_category STRING 10 0 -1 -1
+-----+-----+-----+-----+-----+-----+-----+-----+

#### Related information:

[COMPUTE STATS Statement](#) on page 170, [SHOW TABLE STATS Statement](#) on page 251, [SHOW COLUMN STATS Statement](#) on page 253, [How Impala Uses Statistics for Query Optimization](#) on page 363

## DROP TABLE Statement

Removes an Impala table. Also removes the underlying HDFS data files for internal tables, although not for external tables.

### Syntax:

```
DROP TABLE [IF EXISTS] [db_name.]table_name
```

**Statement type:** DDL

### Usage notes:

By default, Impala removes the associated HDFS directory and data files for the table. If you issue a `DROP TABLE` and the data files are not deleted, it might be for the following reasons:

- If the table was created with the [EXTERNAL](#) clause, Impala leaves all files and directories untouched. Use external tables when the data is under the control of other Hadoop components, and Impala is only used to query the data files from their original locations.
- Impala might leave the data files behind unintentionally, if there is no HDFS location available to hold the HDFS trashcan for the `impala` user. See [User Account Requirements](#) on page 24 for the procedure to set up the required HDFS home directory.

Make sure that you are in the correct database before dropping a table, either by issuing a `USE` statement first or by using a fully qualified name `db_name.table_name`.

The optional `IF EXISTS` clause makes the statement succeed whether or not the table exists. If the table does exist, it is dropped; if it does not exist, the statement has no effect. This capability is useful in standardized setup scripts that remove existing schema objects and create new ones. By using some combination of `IF EXISTS` for the `DROP` statements and `IF NOT EXISTS` clauses for the `CREATE` statements, the script can run successfully the first time you run it (when the objects do not exist yet) and subsequent times (when some or all of the objects do already exist).

If you intend to issue a `DROP DATABASE` statement, first issue `DROP TABLE` statements to remove all the tables in that database.

### Examples:

```
create database temporary;
use temporary;
create table unimportant (x int);
create table trivial (s string);
-- Drop a table in the current database.
drop table unimportant;
-- Switch to a different database.
use default;
-- To drop a table in a different database...
drop table trivial;
ERROR: AnalysisException: Table does not exist: default.trivial
-- ...use a fully qualified name.
drop table temporary.trivial;
```

For other tips about managing and reclaiming Impala disk space, see [Managing Disk Space for Impala Data](#) on page 85.

### Amazon S3 considerations:

Although Impala cannot write new data to a table stored in the Amazon S3 filesystem, the `DROP TABLE` statement can remove data files from S3 if the associated S3 table is an internal table. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about working with S3 tables.

**Cancellation:** Cannot be cancelled.

### HDFS permissions:

For an internal table, the user ID that the `impalad` daemon runs under, typically the `impala` user, must have write permission for all the files and directories that make up the table.

For an external table, dropping the table only involves changes to metadata in the metastore database. Because Impala does not remove any HDFS files or directories when external tables are dropped, no particular permissions are needed for the associated HDFS files or directories.

### Related information:

[Tables](#) on page 156, [ALTER TABLE Statement](#) on page 164, [CREATE TABLE Statement](#) on page 180, [Partitioning for Impala Tables](#) on page 394, [Internal Tables](#) on page 157, [External Tables](#) on page 157

## DROP VIEW Statement

Removes the specified view, which was originally created by the `CREATE VIEW` statement. Because a view is purely a logical construct (an alias for a query) with no physical data behind it, `DROP VIEW` only involves changes to metadata in the metastore database, not any data files in HDFS.

### Syntax:

```
DROP VIEW [ IF EXISTS ] [db_name.]view_name
```

### Statement type:

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

### Examples:

The following example creates a series of views and then drops them. These examples illustrate how views are associated with a particular database, and both the view definitions and the view names for `CREATE VIEW` and `DROP VIEW` can refer to a view in the current database or a fully qualified view name.

```
-- Create and drop a view in the current database.  
CREATE VIEW few_rows_from_t1 AS SELECT * FROM t1 LIMIT 10;  
DROP VIEW few_rows_from_t1;  
  
-- Create and drop a view referencing a table in a different database.  
CREATE VIEW table_from_other_db AS SELECT x FROM db1.foo WHERE x IS NOT NULL;  
DROP VIEW table_from_other_db;  
  
USE db1;  
-- Create a view in a different database.  
CREATE VIEW db2.v1 AS SELECT * FROM db2.foo;  
-- Switch into the other database and drop the view.  
USE db2;  
DROP VIEW v1;  
  
USE db1;  
-- Create a view in a different database.  
CREATE VIEW db2.v1 AS SELECT * FROM db2.foo;  
-- Drop a view in the other database.  
DROP VIEW db2.v1;
```

### Related information:

[Views](#) on page 158, [CREATE VIEW Statement](#) on page 186, [ALTER VIEW Statement](#) on page 168

## EXPLAIN Statement

Returns the execution plan for a statement, showing the low-level mechanisms that Impala will use to read the data, divide the work among nodes in the cluster, and transmit intermediate and final results across the network. Use `explain` followed by a complete `SELECT` query. For example:

### Syntax:

```
EXPLAIN { select_query | ctas_stmt | insert_stmt }
```

The `select_query` is a `SELECT` statement, optionally prefixed by a `WITH` clause. See [SELECT Statement](#) on page 218 for details.

The `insert_stmt` is an `INSERT` statement that inserts into or overwrites an existing table. It can use either the `INSERT ... SELECT` or `INSERT ... VALUES` syntax. See [INSERT Statement](#) on page 202 for details.

The `ctas_stmt` is a `CREATE TABLE` statement using the `AS SELECT` clause, typically abbreviated as a "CTAS" operation. See [CREATE TABLE Statement](#) on page 180 for details.

### Usage notes:

You can interpret the output to judge whether the query is performing efficiently, and adjust the query and/or the schema if not. For example, you might change the tests in the `WHERE` clause, add hints to make join operations more efficient, introduce subqueries, change the order of tables in a join, add or change partitioning for a table, collect column statistics and/or table statistics in Hive, or any other performance tuning steps.

The `EXPLAIN` output reminds you if table or column statistics are missing from any table involved in the query. These statistics are important for optimizing queries involving large tables or multi-table joins. See [COMPUTE STATS Statement](#) on page 170 for how to gather statistics, and [How Impala Uses Statistics for Query Optimization](#) on page 363 for how to use this information for query tuning.

Read the `EXPLAIN` plan from bottom to top:

- The last part of the plan shows the low-level details such as the expected amount of data that will be read, where you can judge the effectiveness of your partitioning strategy and estimate how long it will take to scan a table based on total data size and the size of the cluster.
- As you work your way up, next you see the operations that will be parallelized and performed on each Impala node.
- At the higher levels, you see how data flows when intermediate result sets are combined and transmitted from one node to another.
- See [EXPLAIN\\_LEVEL Query Option](#) on page 343 for details about the `EXPLAIN_LEVEL` query option, which lets you customize how much detail to show in the `EXPLAIN` plan depending on whether you are doing high-level or low-level tuning, dealing with logical or physical aspects of the query.

If you come from a traditional database background and are not familiar with data warehousing, keep in mind that Impala is optimized for full table scans across very large tables. The structure and distribution of this data is typically not suitable for the kind of indexing and single-row lookups that are common in OLTP environments. Seeing a query scan entirely through a large table is common, not necessarily an indication of an inefficient query. Of course, if you can reduce the volume of scanned data by orders of magnitude, for example by using a query that affects only certain partitions within a partitioned table, then you might be able to optimize a query so that it executes in seconds rather than minutes.

For more information and examples to help you interpret `EXPLAIN` output, see [Using the EXPLAIN Plan for Performance Tuning](#) on page 379.

### Extended EXPLAIN output:

For performance tuning of complex queries, and capacity planning (such as using the admission control and resource management features), you can enable more detailed and informative output for the `EXPLAIN` statement. In the `impala-shell` interpreter, issue the command `SET EXPLAIN_LEVEL=level`, where `level` is an integer from 0 to 3 or corresponding mnemonic values `minimal`, `standard`, `extended`, or `verbose`.

## Impala SQL Language Reference

When extended EXPLAIN output is enabled, EXPLAIN statements print information about estimated memory requirements, minimum number of virtual cores, and so on that you can use to fine-tune the resource management options explained in [impalad Startup Options for Resource Management](#) on page 80. (The estimated memory requirements are intentionally on the high side, to allow a margin for error, to avoid cancelling a query unnecessarily if you set the MEM\_LIMIT option to the estimated memory figure.)

See [EXPLAIN\\_LEVEL Query Option](#) on page 343 for details and examples.

### Examples:

This example shows how the standard EXPLAIN output moves from the lowest (physical) level to the higher (logical) levels. The query begins by scanning a certain amount of data; each node performs an aggregation operation (evaluating COUNT(\*)) on some subset of data that is local to that node; the intermediate results are transmitted back to the coordinator node (labelled here as the EXCHANGE node); lastly, the intermediate results are summed to display the final result.

```
[impalad-host:21000] > explain select count(*) from customer_address;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=42.00MB Vcores=1
      |
      03:AGGREGATE [MERGE FINALIZE]
      |   output: sum(count(*))
      |
      02:EXCHANGE [PARTITION=UNPARTITIONED]
      |
      01:AGGREGATE
      |   output: count(*)
      |
      00:SCAN HDFS [default.customer_address]
      |   partitions=1/1 size=5.25MB
+-----+
```

These examples show how the extended EXPLAIN output becomes more accurate and informative as statistics are gathered by the COMPUTE STATS statement. Initially, much of the information about data size and distribution is marked “unavailable”. Impala can determine the raw data size, but not the number of rows or number of distinct values for each column without additional analysis. The COMPUTE STATS statement performs this analysis, so a subsequent EXPLAIN statement has additional information to use in deciding how to optimize the distributed query.

```
[localhost:21000] > set explain_level=extended;
EXPLAIN_LEVEL set to extended
[localhost:21000] > explain select x from t1;
[localhost:21000] > explain select x from t1;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=32.00MB Vcores=1
      |
      01:EXCHANGE [PARTITION=UNPARTITIONED]
      |   hosts=1 per-host-mem=unavailable
      |   tuple-ids=0 row-size=4B cardinality=unavailable
      |
      00:SCAN HDFS [default.t2, PARTITION=RANDOM]
      |   partitions=1/1 size=36B
      |   table stats: unavailable
      |   column stats: unavailable
      |   hosts=1 per-host-mem=32.00MB
      |   tuple-ids=0 row-size=4B cardinality=unavailable
+-----+
```

```
[localhost:21000] > compute stats t1;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 1 column(s). |
+-----+
```

```
+-----+
[localhost:21000] > explain select x from t1;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=64.00MB VCores=1
01:EXCHANGE [PARTITION=UNPARTITIONED]
  hosts=1 per-host-mem=unavailable
    tuple-ids=0 row-size=4B cardinality=0
00:SCAN HDFS [default.t1, PARTITION=RANDOM]
  partitions=1/1 size=36B
    table stats: 0 rows total
    column stats: all
    hosts=1 per-host-mem=64.00MB
      tuple-ids=0 row-size=4B cardinality=0
+-----+
```

**Security considerations:**

If these statements in your environment contain sensitive literal values such as credit card numbers or tax identifiers, Impala can redact this sensitive information when displaying the statements in log files and other administrative contexts. See [Sensitive Data Redaction](#) for details.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read and execute permissions for all applicable directories in all source tables for the query that is being explained. (A `SELECT` operation could read files from multiple different HDFS directories if the source table is partitioned.)

**Related information:**

[SELECT Statement](#) on page 218, [INSERT Statement](#) on page 202, [CREATE TABLE Statement](#) on page 180, [Understanding Impala Query Performance - EXPLAIN Plans and Query Profiles](#) on page 378

## GRANT Statement (CDH 5.2 or higher only)

The `GRANT` statement grants roles or privileges on specified objects to groups. Only Sentry administrative users can grant roles to a group.

**Syntax:**

```
GRANT ROLE role_name TO GROUP group_name

GRANT privilege ON object_type object_name
  TO [ROLE] roleName
  [WITH GRANT OPTION]

privilege ::= SELECT | INSERT | ALL
object_type ::= TABLE | DATABASE | SERVER | URI
```

Typically, the object name is an identifier. For URIs, it is a string literal.

**Required privileges:**

Only administrative users (initially, a predefined set of users specified in the Sentry service configuration file) can use this statement.

The `WITH GRANT OPTION` clause allows members of the specified role to issue `GRANT` and `REVOKE` statements for those same privileges. Hence, if a role has the `ALL` privilege on a database and the `WITH GRANT OPTION` set, users granted that role can execute `GRANT/REVOKE` statements only for that database or child tables of the database. This means a user could revoke the privileges of the user that provided them the `GRANT OPTION`.

## Impala SQL Language Reference

Impala does not currently support revoking only the WITH GRANT OPTION from a privilege previously granted to a role. To remove the WITH GRANT OPTION, revoke the privilege and grant it again without the WITH GRANT OPTION flag.

### Compatibility:

- The Impala GRANT and REVOKE statements are available in CDH 5.2 and later.
- In CDH 5.1 and later, Impala can make use of any roles and privileges specified by the GRANT and REVOKE statements in Hive, when your system is configured to use the Sentry service instead of the file-based policy mechanism.
- The Impala GRANT and REVOKE statements for privileges do not require the ROLE keyword to be repeated before each role name, unlike the equivalent Hive statements.
- Currently, each Impala GRANT or REVOKE statement can only grant or revoke a single privilege to or from a single role.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

### Related information:

[Enabling Sentry Authorization for Impala](#) on page 91, [REVOKE Statement \(CDH 5.2 or higher only\)](#) on page 217, [CREATE ROLE Statement \(CDH 5.2 or higher only\)](#) on page 179, [DROP ROLE Statement \(CDH 5.2 or higher only\)](#) on page 193, [SHOW Statement](#) on page 242

## INSERT Statement

Impala supports inserting into tables and partitions that you create with the Impala CREATE TABLE statement, or pre-defined tables and partitions created through Hive.

### Syntax:

```
[with_clause]
INSERT { INTO | OVERWRITE } [TABLE] table_name
  [(column_list)]
  [ PARTITION (partition_clause) ]
{
  [hint_clause] select_statement
  | VALUES (value [, value ...]) [, (value [, value ...]) ...]
}
partition_clause ::= col_name [= constant] [, col_name [= constant] ...]
hint_clause ::= [SHUFFLE] | [NOSHUFFLE]      (Note: the square brackets are part of the
   syntax.)
```

### Appending or replacing (INTO and OVERWRITE clauses):

The INSERT INTO syntax appends data to a table. The existing data files are left as-is, and the inserted data is put into one or more new data files.

The INSERT OVERWRITE syntax replaces the data in a table. Currently, the overwritten data files are deleted immediately; they do not go through the HDFS trash mechanism.

### Usage notes:

Impala currently supports:

- Copy data from another table using SELECT query. In Impala 1.2.1 and higher, you can combine CREATE TABLE and INSERT operations into a single step with the CREATE TABLE AS SELECT syntax, which bypasses the actual INSERT keyword.
- An optional [WITH clause](#) before the INSERT keyword, to define a subquery referenced in the SELECT portion.
- Create one or more new rows using constant expressions through VALUES clause. (The VALUES clause was added in Impala 1.0.1.)

- By default, the first column of each newly inserted row goes into the first column of the table, the second column into the second column, and so on.

You can also specify the columns to be inserted, an arbitrarily ordered subset of the columns in the destination table, by specifying a column list immediately after the name of the destination table. This feature lets you adjust the inserted columns to match the layout of a `SELECT` statement, rather than the other way around. (This feature was added in Impala 1.1.)

The number of columns mentioned in the column list (known as the "column permutation") must match the number of columns in the `SELECT` list or the `VALUES` tuples. The order of columns in the column permutation can be different than in the underlying table, and the columns of each input row are reordered to match. If the number of columns in the column permutation is less than in the destination table, all unmentioned columns are set to `NULL`.

- For a partitioned table, the optional `PARTITION` clause identifies which partition or partitions the new values go into. If a partition key column is given a constant value such as `PARTITION (year=2012)` or `PARTITION (year=2012, month=2)`, all the inserted rows use those same values for those partition key columns and you omit any corresponding columns in the source table from the `SELECT` list. This form is known as "static partitioning".

If a partition key column is mentioned but not assigned a value, such as in `PARTITION (year, region)` (both columns unassigned) or `PARTITION(year, region='CA')` (`year` column unassigned), the unassigned columns are filled in with the final columns of the `SELECT` list. In this case, the number of columns in the `SELECT` list must equal the number of columns in the column permutation plus the number of partition key columns not assigned a constant value. This form is known as "dynamic partitioning".

See [Static and Dynamic Partitioning Clauses](#) on page 395 for examples and performance characteristics of static and dynamic partitioned inserts.

- An optional hint clause immediately before the `SELECT` keyword, to fine-tune the behavior when doing an `INSERT ... SELECT` operation into partitioned Parquet tables. The hint keywords are `[SHUFFLE]` and `[NOSHUFFLE]`, including the square brackets. Inserting into partitioned Parquet tables can be a resource-intensive operation because it potentially involves many files being written to HDFS simultaneously, and separate large memory buffers being allocated to buffer the data for each partition. For usage details, see [Loading Data into Parquet Tables](#) on page 409.

▪ **Note:**

- Insert commands that partition or add files result in changes to Hive metadata. Because Impala uses Hive metadata, such changes may necessitate a metadata refresh. For more information, see the [REFRESH](#) function.
- Currently, Impala can only insert data into tables that use the text and Parquet formats. For other file formats, insert the data using Hive and use Impala to query it.
- As an alternative to the `INSERT` statement, if you have existing data files elsewhere in HDFS, the `LOAD DATA` statement can move those files into a table. This statement works with tables of any file format.

**Statement type:** DML (but still affected by [SYNC\\_DDL](#) query option)

**Usage notes:**

When you insert the results of an expression, particularly of a built-in function call, into a small numeric column such as `INT`, `SMALLINT`, `TINYINT`, or `FLOAT`, you might need to use a `CAST()` expression to coerce values into the appropriate type. Impala does not automatically convert from a larger type to a smaller one. For example, to insert cosine values into a `FLOAT` column, write `CAST(COS(angle) AS FLOAT)` in the `INSERT` statement to make the conversion explicit.

Any `INSERT` statement for a Parquet table requires enough free space in the HDFS filesystem to write one block. Because Parquet data files use a block size of 1 GB by default, an `INSERT` might fail (even for a very small amount of data) if your HDFS is running low on space.

## Impala SQL Language Reference

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

- **Important:** After adding or replacing data in a table used in performance-critical queries, issue a `COMPUTE STATS` statement to make sure all statistics are up-to-date. Consider updating statistics for a table after any `INSERT`, `LOAD DATA`, or `CREATE TABLE AS SELECT` statement in Impala, or after loading data through Hive and doing a `REFRESH table_name` in Impala. This technique is especially important for tables that are very large, used in join queries, or both.

### Examples:

The following example sets up new tables with the same definition as the `TAB1` table from the [Tutorial](#) section, using different file formats, and demonstrates inserting data into the tables created with the `STORED AS TEXTFILE` and `STORED AS PARQUET` clauses:

```
CREATE DATABASE IF NOT EXISTS file_formats;
USE file_formats;

DROP TABLE IF EXISTS text_table;
CREATE TABLE text_table
( id INT, col_1 BOOLEAN, col_2 DOUBLE, col_3 TIMESTAMP )
STORED AS TEXTFILE;

DROP TABLE IF EXISTS parquet_table;
CREATE TABLE parquet_table
( id INT, col_1 BOOLEAN, col_2 DOUBLE, col_3 TIMESTAMP )
STORED AS PARQUET;
```

With the `INSERT INTO TABLE` syntax, each new set of inserted rows is appended to any existing data in the table. This is how you would record small amounts of data that arrive continuously, or ingest new batches of data alongside the existing data. For example, after running 2 `INSERT INTO TABLE` statements with 5 rows each, the table contains 10 rows total:

```
[localhost:21000] > insert into table text_table select * from default.tab1;
Inserted 5 rows in 0.41s

[localhost:21000] > insert into table text_table select * from default.tab1;
Inserted 5 rows in 0.46s

[localhost:21000] > select count(*) from text_table;
+-----+
| count(*) |
+-----+
| 10       |
+-----+
Returned 1 row(s) in 0.26s
```

With the `INSERT OVERWRITE TABLE` syntax, each new set of inserted rows replaces any existing data in the table. This is how you load data to query in a data warehousing scenario where you analyze just the data for a particular day, quarter, and so on, discarding the previous data each time. You might keep the entire set of data in one raw table, and transfer and transform certain rows into a more compact and efficient form to perform intensive analysis on that subset.

For example, here we insert 5 rows into a table using the `INSERT INTO` clause, then replace the data by inserting 3 rows with the `INSERT OVERWRITE` clause. Afterward, the table only contains the 3 rows from the final `INSERT` statement.

```
[localhost:21000] > insert into table parquet_table select * from default.tab1;
Inserted 5 rows in 0.35s

[localhost:21000] > insert overwrite table parquet_table select * from default.tab1
limit 3;
Inserted 3 rows in 0.43s
```

```
[localhost:21000] > select count(*) from parquet_table;
+-----+
| count(*) |
+-----+
| 3         |
+-----+
Returned 1 row(s) in 0.43s
```

The [VALUES](#) clause lets you insert one or more rows by specifying constant values for all the columns. The number, types, and order of the expressions must match the table definition.

- **Note:** The `INSERT ... VALUES` technique is not suitable for loading large quantities of data into HDFS-based tables, because the insert operations cannot be parallelized, and each one produces a separate data file. Use it for setting up small dimension tables or tiny amounts of data for experimenting with SQL syntax, or with HBase tables. Do not use it for large ETL jobs or benchmark tests for load operations. Do not run scripts with thousands of `INSERT ... VALUES` statements that insert a single row each time. If you do run `INSERT ... VALUES` operations to load data into a staging table as one stage in an ETL pipeline, include multiple row values if possible within each `VALUES` clause, and use a separate database to make cleanup easier if the operation does produce many tiny files.

The following example shows how to insert one row or multiple rows, with expressions of different types, using literal values, expressions, and function return values:

```
create table val_test_1 (c1 int, c2 float, c3 string, c4 boolean, c5 timestamp);
insert into val_test_1 values (100, 99.9/10, 'abc', true, now());
create table val_test_2 (id int, token string);
insert overwrite val_test_2 values (1, 'a'), (2, 'b'), (-1,'xyzzy');
```

These examples show the type of “not implemented” error that you see when attempting to insert data into a table with a file format that Impala currently does not write to:

```
DROP TABLE IF EXISTS sequence_table;
CREATE TABLE sequence_table
( id INT, col_1 BOOLEAN, col_2 DOUBLE, col_3 TIMESTAMP )
STORED AS SEQUENCEFILE;

DROP TABLE IF EXISTS rc_table;
CREATE TABLE rc_table
( id INT, col_1 BOOLEAN, col_2 DOUBLE, col_3 TIMESTAMP )
STORED AS RCFILE;

[localhost:21000] > insert into table rc_table select * from default.tab1;
Remote error
Backend 0:RC_FILE not implemented.

[localhost:21000] > insert into table sequence_table select * from default.tab1;
Remote error
Backend 0:SEQUENCE_FILE not implemented.
```

Inserting data into partitioned tables requires slightly different syntax that divides the partitioning columns from the others:

```
create table t1 (i int) partitioned by (x int, y string);
-- Select an INT column from another table.
-- All inserted rows will have the same x and y values, as specified in the INSERT statement.
-- This technique of specifying all the partition key values is known as static partitioning.
insert into t1 partition(x=10, y='a') select c1 from some_other_table;
-- Select two INT columns from another table.
-- All inserted rows will have the same y value, as specified in the INSERT statement.
-- Values from c2 go into t1.x.
-- Any partitioning columns whose value is not specified are filled in
-- from the columns specified last in the SELECT list.
```

```
-- This technique of omitting some partition key values is known as dynamic partitioning.  
insert into t1 partition(x, y='b') select c1, c2 from some_other_table;  
-- Select an INT and a STRING column from another table.  
-- All inserted rows will have the same x value, as specified in the INSERT statement.  
-- Values from c3 go into t1.y.  
insert into t1 partition(x=20, y) select c1, c3 from some_other_table;
```

The following examples show how you can copy the data in all the columns from one table to another, copy the data from only some columns, or specify the columns in the select list in a different order than they actually appear in the table:

```
-- Start with 2 identical tables.  
create table t1 (c1 int, c2 int);  
create table t2 like t1;  
  
-- If there is no () part after the destination table name,  
-- all columns must be specified, either as * or by name.  
insert into t2 select * from t1;  
insert into t2 select c1, c2 from t1;  
  
-- With the () notation following the destination table name,  
-- you can omit columns (all values for that column are NULL  
-- in the destination table), and/or reorder the values  
-- selected from the source table. This is the "column permutation" feature.  
insert into t2 (c1) select c1 from t1;  
insert into t2 (c2, c1) select c1, c2 from t1;  
  
-- The column names can be entirely different in the source and destination tables.  
-- You can copy any columns, not just the corresponding ones, from the source table.  
-- But the number and type of selected columns must match the columns mentioned in the  
() part.  
alter table t2 replace columns (x int, y int);  
insert into t2 (y) select c1 from t1;  
  
-- For partitioned tables, all the partitioning columns must be mentioned in the ()  
column list  
-- or a PARTITION clause; these columns cannot be defaulted to NULL.  
create table pt1 (x int, y int) partitioned by (z int);  
-- The values from c1 are copied into the column x in the new table,  
-- all in the same partition based on a constant value for z.  
-- The values of y in the new table are all NULL.  
insert into pt1 (x) partition (z=5) select c1 from t1;  
-- Again we omit the values for column y so they are all NULL.  
-- The inserted x values can go into different partitions, based on  
-- the different values inserted into the partitioning column z.  
insert into pt1 (x,z) select x, z from t2;
```

SELECT \* for a partitioned table requires that all partition key columns in the source table be declared as the last columns in the CREATE TABLE statement. You still include a PARTITION BY clause listing all the partition key columns. These partition columns are automatically mapped to the last columns from the SELECT \* list.

```
create table source (x int, y int, year int, month int, day int);  
create table destination (x int, y int) partitioned by (year int, month int, day int);  
...load some data into the unpartitioned source table...  
-- Insert a single partition of data.  
-- The SELECT * means you cannot specify partition (year=2014, month, day).  
insert overwrite destination partition (year, month, day) select * from source where  
year=2014;  
-- Insert the data for all year/month/day combinations.  
insert overwrite destination partition (year, month, day) select * from source;  
  
-- If one of the partition columns is omitted from the source table,  
-- then you can specify a specific value for that column in the PARTITION clause.  
-- Here the source table holds only data from 2014, and so does not include a year  
column.  
create table source_2014 (x int, y int, month, day);  
...load some data into the unpartitioned source_2014 table...  
insert overwrite destination partition (year=2014, month, day) select * from source_2014;
```

**Sorting considerations:** Although you can specify an `ORDER BY` clause in an `INSERT ... SELECT` statement, any `ORDER BY` clause is ignored and the results are not necessarily sorted. An `INSERT ... SELECT` operation potentially creates many different data files, prepared on different data nodes, and therefore the notion of the data being stored in sorted order is impractical.

**Concurrency considerations:** Each `INSERT` operation creates new data files with unique names, so you can run multiple `INSERT INTO` statements simultaneously without filename conflicts. While data is being inserted into an Impala table, the data is staged temporarily in a subdirectory inside the data directory; during this period, you cannot issue queries against that table in Hive. If an `INSERT` operation fails, the temporary data file and the subdirectory could be left behind in the data directory. If so, remove the relevant subdirectory and any data files it contains manually, by issuing an `hdfs dfs -rm -r` command, specifying the full path of the work subdirectory, whose name ends in `_dir`.

## VALUES Clause

The `VALUES` clause is a general-purpose way to specify the columns of one or more rows, typically within an `INSERT` statement.

- **Note:** The `INSERT ... VALUES` technique is not suitable for loading large quantities of data into HDFS-based tables, because the insert operations cannot be parallelized, and each one produces a separate data file. Use it for setting up small dimension tables or tiny amounts of data for experimenting with SQL syntax, or with HBase tables. Do not use it for large ETL jobs or benchmark tests for load operations. Do not run scripts with thousands of `INSERT ... VALUES` statements that insert a single row each time. If you do run `INSERT ... VALUES` operations to load data into a staging table as one stage in an ETL pipeline, include multiple row values if possible within each `VALUES` clause, and use a separate database to make cleanup easier if the operation does produce many tiny files.

The following examples illustrate:

- How to insert a single row using a `VALUES` clause.
- How to insert multiple rows using a `VALUES` clause.
- How the row or rows from a `VALUES` clause can be appended to a table through `INSERT INTO`, or replace the contents of the table through `INSERT OVERWRITE`.
- How the entries in a `VALUES` clause can be literals, function results, or any other kind of expression. See [Literals](#) on page 138 for the notation to use for literal values, especially [String Literals](#) on page 140 for quoting and escaping conventions for strings. See [SQL Operators](#) on page 142 and [Built-in Functions](#) on page 257 for other things you can include in expressions with the `VALUES` clause.

```
[localhost:21000] > describe val_example;
Query: describe val_example
Query finished, fetching results ...
+-----+-----+-----+
| name | type   | comment |
+-----+-----+-----+
| id   | int    |          |
| col_1| boolean|          |
| col_2| double |          |
+-----+-----+-----+
[localhost:21000] > insert into val_example values (1,true,100.0);
Inserted 1 rows in 0.30s
[localhost:21000] > select * from val_example;
+-----+-----+
| id | col_1 | col_2 |
+-----+-----+
| 1  | true  | 100   |
+-----+-----+
[localhost:21000] > insert overwrite val_example values (10,false,pow(2,5)),
(50,true,10/3);
Inserted 2 rows in 0.16s
[localhost:21000] > select * from val_example;
```

id	col_1	col_2
10	false	32
50	true	3.333333333333333

When used in an `INSERT` statement, the Impala `VALUES` clause can specify some or all of the columns in the destination table, and the columns can be specified in a different order than they actually appear in the table. To specify a different set or order of columns than in the table, use the syntax:

```
INSERT INTO destination
  (col_x, col_y, col_z)
VALUES
  (val_x, val_y, val_z);
```

Any columns in the table that are not listed in the `INSERT` statement are set to `NULL`.

To use a `VALUES` clause like a table in other statements, wrap it in parentheses and use `AS` clauses to specify aliases for the entire object and any columns you need to refer to:

```
[localhost:21000] > select * from (values(4,5,6),(7,8,9)) as t;
+---+---+---+
| 4 | 5 | 6 |
+---+---+---+
| 4 | 5 | 6 |
| 7 | 8 | 9 |
+---+---+
[localhost:21000] > select * from (values(1 as c1, true as c2, 'abc' as
c3),(100,false,'xyz')) as t;
+---+---+---+
| c1 | c2 | c3 |
+---+---+---+
| 1  | true | abc |
| 100 | false | xyz |
+---+---+
```

For example, you might use a tiny table constructed like this from constant literals or function return values as part of a longer statement involving joins or `UNION ALL`.

### HDFS considerations:

Impala physically writes all inserted files under the ownership of its default user, typically `impala`. Therefore, this user must have HDFS write permission in the corresponding table directory.

The permission requirement is independent of the authorization performed by the Sentry framework. (If the connected user is not authorized to insert into a table, Sentry blocks that operation immediately, regardless of the privileges available to the `impala` user.) Files created by Impala are not owned by and do not inherit permissions from the connected user.

The number of data files produced by an `INSERT` statement depends on the size of the cluster, the number of data blocks that are processed, the partition key columns in a partitioned table, and the mechanism Impala uses for dividing the work in parallel. Do not assume that an `INSERT` statement will produce some particular number of output files. In case of performance issues with data written by Impala, check that the output files do not suffer from issues such as many tiny files or many tiny partitions. (In the Hadoop context, even files or partitions of a few tens of megabytes are considered "tiny".)

The `INSERT` statement has always left behind a hidden work directory inside the data directory of the table. Formerly, this hidden work directory was named `.impala_insert_staging`. In Impala 2.0.1 and later, this directory name is changed to `_impala_insert_staging`. (While HDFS tools are expected to treat names beginning either with underscore and dot as hidden, in practice names beginning with an underscore are more widely supported.) If you have any scripts, cleanup jobs, and so on that rely on the name of this work directory, adjust them to use the new name.

### HBase considerations:

You can use the `INSERT` statement with HBase tables as follows:

- You can insert a single row or a small set of rows into an HBase table with the `INSERT ... VALUES` syntax. This is a good use case for HBase tables with Impala, because HBase tables are not subject to the same kind of fragmentation from many small insert operations as HDFS tables are.
- You can insert any number of rows at once into an HBase table using the `INSERT ... SELECT` syntax.
- If more than one inserted row has the same value for the HBase key column, only the last inserted row with that value is visible to Impala queries. You can take advantage of this fact with `INSERT ... VALUES` statements to effectively update rows one at a time, by inserting new rows with the same key values as existing rows. Be aware that after an `INSERT ... SELECT` operation copying from an HDFS table, the HBase table might contain fewer rows than were inserted, if the key column in the source table contained duplicate values.
- You cannot `INSERT OVERWRITE` into an HBase table. New rows are always appended.
- When you create an Impala or Hive table that maps to an HBase table, the column order you specify with the `INSERT` statement might be different than the order you declare with the `CREATE TABLE` statement. Behind the scenes, HBase arranges the columns based on how they are divided into column families. This might cause a mismatch during insert operations, especially if you use the syntax `INSERT INTO hbase_table SELECT * FROM hdfs_table`. Before inserting data, verify the column order by issuing a `DESCRIBE` statement for the table, and adjust the order of the select list in the `INSERT` statement.

See [Using Impala to Query HBase Tables](#) on page 426 for more details about using Impala with HBase.

#### **Amazon S3 considerations:**

Currently, Impala cannot insert or load data into a table or partition that resides in the Amazon Simple Storage Service (S3). Bring data into S3 using the normal S3 transfer mechanisms, then use Impala to query the S3 data. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about using Impala with S3.

#### **Security considerations:**

If these statements in your environment contain sensitive literal values such as credit card numbers or tax identifiers, Impala can redact this sensitive information when displaying the statements in log files and other administrative contexts. See [Sensitive Data Redaction](#) for details.

**Cancellation:** Can be cancelled. To cancel this statement, use Ctrl-C from the `impala-shell` interpreter, the **Cancel** button from the **Watch** page in Hue, **Actions > Cancel** from the **Queries** list in Cloudera Manager, or **Cancel** from the list of in-flight queries (for a particular node) on the **Queries** tab in the Impala web UI (port 25000).

#### **HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read permission for the files in the source directory of an `INSERT ... SELECT` operation, and write permission for all affected directories in the destination table. (An `INSERT` operation could write files to multiple different HDFS directories if the destination table is partitioned.) This user must also have write permission to create a temporary work directory in the top-level HDFS directory of the destination table. An `INSERT OVERWRITE` operation does not require write permission on the original data files in the table, only on the table directories themselves.

#### **Restrictions:**

For `INSERT` operations into `CHAR` or `VARCHAR` columns, you must cast all `STRING` literals or expressions returning `STRING` to a `CHAR` or `VARCHAR` type with the appropriate length.

#### **Related startup options:**

By default, if an `INSERT` statement creates any new subdirectories underneath a partitioned table, those subdirectories are assigned default HDFS permissions for the `impala` user. To make each subdirectory have the same permissions as its parent directory in HDFS, specify the `--insert_inherit_permissions` startup option for the `impalad` daemon.

### INVALIDATE METADATA Statement

Marks the metadata for one or all tables as stale. Required after a table is created through the Hive shell, before the table is available for Impala queries. The next time the current Impala node performs a query against a table whose metadata is invalidated, Impala reloads the associated metadata before the query proceeds. This is a relatively expensive operation compared to the incremental metadata update done by the REFRESH statement, so in the common scenario of adding new data files to an existing table, prefer REFRESH rather than INVALIDATE METADATA. If you are not familiar with the way Impala uses metadata and how it shares the same metastore database as Hive, see [Overview of Impala Metadata and the Metastore](#) on page 21 for background information.

#### Syntax:

```
INVALIDATE METADATA [ [db_name.]table_name ]
```

By default, the cached metadata for all tables is flushed. If you specify a table name, only the metadata for that one table is flushed. Even for a single table, INVALIDATE METADATA is more expensive than REFRESH, so prefer REFRESH in the common case where you add new data files for an existing table.

#### Internal details:

To accurately respond to queries, Impala must have current metadata about those databases and tables that clients query directly. Therefore, if some other entity modifies information used by Impala in the metastore that Impala and Hive share, the information cached by Impala must be updated. However, this does not mean that all metadata updates require an Impala update.

##### Note:

In Impala 1.2.4 and higher, you can specify a table name with INVALIDATE METADATA after the table is created in Hive, allowing you to make individual tables visible to Impala without doing a full reload of the catalog metadata. Impala 1.2.4 also includes other changes to make the metadata broadcast mechanism faster and more responsive, especially during Impala startup. See [New Features in Impala Version 1.2.4](#) on page 479 for details.

In Impala 1.2 and higher, a dedicated daemon (`catalogd`) broadcasts DDL changes made through Impala to all Impala nodes. Formerly, after you created a database or table while connected to one Impala node, you needed to issue an INVALIDATE METADATA statement on another Impala node before accessing the new database or table from the other node. Now, newly created or altered objects are picked up automatically by all Impala nodes. You must still use the INVALIDATE METADATA technique after creating or altering objects through Hive. See [The Impala Catalog Service](#) on page 19 for more information on the catalog service.

The INVALIDATE METADATA statement is new in Impala 1.1 and higher, and takes over some of the use cases of the Impala 1.0 REFRESH statement. Because REFRESH now requires a table name parameter, to flush the metadata for all tables at once, use the INVALIDATE METADATA statement.

Because REFRESH `table_name` only works for tables that the current Impala node is already aware of, when you create a new table in the Hive shell, you must enter INVALIDATE METADATA with no table parameter before you can see the new table in `impala-shell`. Once the table is known by the Impala node, you can issue REFRESH `table_name` after you add data files for that table.

INVALIDATE METADATA and REFRESH are counterparts: INVALIDATE METADATA waits to reload the metadata when needed for a subsequent query, but reloads all the metadata for the table, which can be an expensive operation, especially for large tables with many partitions. REFRESH reloads the metadata immediately, but only loads the block location data for newly added data files, making it a less expensive operation overall. If data was altered in some more extensive way, such as being reorganized by the HDFS balancer, use INVALIDATE METADATA to avoid a performance penalty from reduced local reads. If you used Impala version 1.0, the INVALIDATE METADATA statement works just like the Impala 1.0 REFRESH statement did, while the Impala 1.1 REFRESH is optimized for the common use case of adding new data files to an existing table, thus the table name argument is now required.

**Usage notes:**

A metadata update for an `impalad` instance **is** required if:

- A metadata change occurs.
- **and** the change is made from another `impalad` instance in your cluster, or through Hive.
- **and** the change is made to a database to which clients such as the Impala shell or ODBC directly connect.

A metadata update for an Impala node **is not** required when you issue queries from the same Impala node where you ran `ALTER TABLE`, `INSERT`, or other table-modifying statement.

Database and table metadata is typically modified by:

- Hive - via `ALTER`, `CREATE`, `DROP` or `INSERT` operations.
- Impalad - via `CREATE TABLE`, `ALTER TABLE`, and `INSERT` operations.

`INVALIDATE METADATA` causes the metadata for that table to be marked as stale, and reloaded the next time the table is referenced. For a huge table, that process could take a noticeable amount of time; thus you might prefer to use `REFRESH` where practical, to avoid an unpredictable delay later, for example if the next reference to the table is during a benchmark test.

**Examples:**

The following example shows how you might use the `INVALIDATE METADATA` statement after creating new tables (such as SequenceFile or HBase tables) through the Hive shell. Before the `INVALIDATE METADATA` statement was issued, Impala would give a “table not found” error if you tried to refer to those table names. The `DESCRIBE` statements cause the latest metadata to be immediately loaded for the tables, avoiding a delay the next time those tables are queried.

```
[impalad-host:21000] > invalidate metadata;
[impalad-host:21000] > describe t1;
...
[impalad-host:21000] > describe t2;
...
```

For more examples of using `REFRESH` and `INVALIDATE METADATA` with a combination of Impala and Hive operations, see [Switching Back and Forth Between Impala and Hive](#) on page 65.

If you need to ensure that the metadata is up-to-date when you start an `impala-shell` session, run `impala-shell` with the `-r` or `--refresh_after_connect` command-line option. Because this operation adds a delay to the next query against each table, potentially expensive for large tables with many partitions, try to avoid using this option for day-to-day operations in a production environment.

**HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have execute permissions for all the relevant directories holding table data. (A table could have data spread across multiple directories, or in unexpected paths, if it uses partitioning or specifies a `LOCATION` attribute for individual partitions or the entire table.) Issues with permissions might not cause an immediate error for this statement, but subsequent statements such as `SELECT` or `SHOW TABLE STATS` could fail.

**HDFS considerations:**

By default, the `INVALIDATE METADATA` command checks HDFS permissions of the underlying data files and directories, caching this information so that a statement can be cancelled immediately if for example the `impala` user does not have permission to write to the data directory for the table. (This checking does not apply if you have set the catalogd configuration option `--load_catalog_in_background=false`.) Impala reports any lack of write permissions as an `INFO` message in the log file, in case that represents an oversight. If you change HDFS permissions to make data readable or writeable by the Impala user, issue another `INVALIDATE METADATA` to make Impala aware of the change.

**Usage notes:**

## Impala SQL Language Reference

This example illustrates creating a new database and new table in Hive, then doing an `INVALIDATE METADATA` statement in Impala using the fully qualified table name, after which both the new table and the new database are visible to Impala. The ability to specify `INVALIDATE METADATA table_name` for a table created in Hive is a new capability in Impala 1.2.4. In earlier releases, that statement would have returned an error indicating an unknown table, requiring you to do `INVALIDATE METADATA` with no table name, a more expensive operation that reloaded metadata for all tables and databases.

```
$ hive
hive> create database new_db_from_hive;
OK
Time taken: 4.118 seconds
hive> create table new_db_from_hive.new_table_from_hive (x int);
OK
Time taken: 0.618 seconds
hive> quit;
$ impala-shell
[localhost:21000] > show databases like 'new*';
[localhost:21000] > refresh new_db_from_hive.new_table_from_hive;
ERROR: AnalysisException: Database does not exist: new_db_from_hive
[localhost:21000] > invalidate metadata new_db_from_hive.new_table_from_hive;
[localhost:21000] > show databases like 'new*';
+-----+
| name      |
+-----+
| new_db_from_hive |
+-----+
[localhost:21000] > show tables in new_db_from_hive;
+-----+
| name      |
+-----+
| new_table_from_hive |
+-----+
```

### Amazon S3 considerations:

The `REFRESH` and `INVALIDATE METADATA` statements also cache metadata for tables where the data resides in the Amazon Simple Storage Service (S3). In particular, issue a `REFRESH` for a table after adding or removing files in the associated S3 data directory. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about working with S3 tables.

**Cancellation:** Cannot be cancelled.

#### Related information:

[Overview of Impala Metadata and the Metastore](#) on page 21, [REFRESH Statement](#) on page 215

## LOAD DATA Statement

The `LOAD DATA` statement streamlines the ETL process for an internal Impala table by moving a data file or all the data files in a directory from an HDFS location into the Impala data directory for that table.

#### Syntax:

```
LOAD DATA INPATH 'hdfs_file_or_directory_path' [OVERWRITE] INTO TABLE tablename
  [PARTITION (partcol1=val1, partcol2=val2 ...)]
```

When the `LOAD DATA` statement operates on a partitioned table, it always operates on one partition at a time. Specify the `PARTITION` clauses and list all the partition key columns, with a constant value specified for each.

**Statement type:** DML (but still affected by `SYNC_DDL` query option)

#### Usage notes:

- The loaded data files are moved, not copied, into the Impala data directory.
- You can specify the HDFS path of a single file to be moved, or the HDFS path of a directory to move all the files inside that directory. You cannot specify any sort of wildcard to take only some of the files from a

directory. When loading a directory full of data files, keep all the data files at the top level, with no nested directories underneath.

- Currently, the Impala `LOAD DATA` statement only imports files from HDFS, not from the local filesystem. It does not support the `LOCAL` keyword of the Hive `LOAD DATA` statement. You must specify a path, not an `hdfs://` URI.
- In the interest of speed, only limited error checking is done. If the loaded files have the wrong file format, different columns than the destination table, or other kind of mismatch, Impala does not raise any error for the `LOAD DATA` statement. Querying the table afterward could produce a runtime error or unexpected results. Currently, the only checking the `LOAD DATA` statement does is to avoid mixing together uncompressed and LZO-compressed text files in the same table.
- When you specify an HDFS directory name as the `LOAD DATA` argument, any hidden files in that directory (files whose names start with a `.`) are not moved to the Impala data directory.
- The loaded data files retain their original names in the new location, unless a name conflicts with an existing data file, in which case the name of the new file is modified slightly to be unique. (The name-mangling is a slight difference from the Hive `LOAD DATA` statement, which replaces identically named files.)
- By providing an easy way to transport files from known locations in HDFS into the Impala data directory structure, the `LOAD DATA` statement lets you avoid memorizing the locations and layout of HDFS directory tree containing the Impala databases and tables. (For a quick way to check the location of the data files for an Impala table, issue the statement `DESCRIBE FORMATTED table_name`.)
- The `PARTITION` clause is especially convenient for ingesting new data for a partitioned table. As you receive new data for a time period, geographic region, or other division that corresponds to one or more partitioning columns, you can load that data straight into the appropriate Impala data directory, which might be nested several levels down if the table is partitioned by multiple columns. When the table is partitioned, you must specify constant values for all the partitioning columns.

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [SYNC\\_DDL Query Option](#) on page 352 for details.

- **Important:** After adding or replacing data in a table used in performance-critical queries, issue a `COMPUTE STATS` statement to make sure all statistics are up-to-date. Consider updating statistics for a table after any `INSERT`, `LOAD DATA`, or `CREATE TABLE AS SELECT` statement in Impala, or after loading data through Hive and doing a `REFRESH table_name` in Impala. This technique is especially important for tables that are very large, used in join queries, or both.

## Examples:

First, we use a trivial Python script to write different numbers of strings (one per line) into files stored in the cloudera HDFS user account. (Substitute the path for your own HDFS user account when doing `hdfs dfs` operations like these.)

```
$ random_strings.py 1000 | hdfs dfs -put - /user/cloudera/thousand_strings.txt
$ random_strings.py 100 | hdfs dfs -put - /user/cloudera/hundred_strings.txt
$ random_strings.py 10 | hdfs dfs -put - /user/cloudera/ten_strings.txt
```

Next, we create a table and load an initial set of data into it. Remember, unless you specify a `STORED AS` clause, Impala tables default to `TEXTFILE` format with Ctrl-A (hex 01) as the field delimiter. This example uses a single-column table, so the delimiter is not significant. For large-scale ETL jobs, you would typically use binary format data files such as Parquet or Avro, and load them into Impala tables that use the corresponding file format.

```
[localhost:21000] > create table t1 (s string);
[localhost:21000] > load data inpath '/user/cloudera/thousand_strings.txt' into table
t1;
Query finished, fetching results ...
+-----+
| summary |
+-----+
```

## Impala SQL Language Reference

```
| Loaded 1 file(s). Total files in destination location: 1 |
+-----+
Returned 1 row(s) in 0.61s
[kilo2-202-961.cs1cloud.internal:21000] > select count(*) from t1;
Query finished, fetching results ...
+-----+
| _c0 |
+-----+
| 1000 |
+-----+
Returned 1 row(s) in 0.67s
[localhost:21000] > load data inpath '/user/cloudera/thousand_strings.txt' into table
t1;
ERROR: AnalysisException: INPATH location '/user/cloudera/thousand_strings.txt' does
not exist.
```

As indicated by the message at the end of the previous example, the data file was moved from its original location. The following example illustrates how the data file was moved into the Impala data directory for the destination table, keeping its original filename:

```
$ hdfs dfs -ls /user/hive/warehouse/load_data_testing.db/t1
Found 1 items
-rw-r--r-- 1 cloudera cloudera 13926 2013-06-26 15:40
/usr/hive/warehouse/load_data_testing.db/t1/thousand_strings.txt
```

The following example demonstrates the difference between the `INTO TABLE` and `OVERWRITE TABLE` clauses. The table already contains 1000 rows. After issuing the `LOAD DATA` statement with the `INTO TABLE` clause, the table contains 100 more rows, for a total of 1100. After issuing the `LOAD DATA` statement with the `OVERWRITE INTO TABLE` clause, the former contents are gone, and now the table only contains the 10 rows from the just-loaded data file.

```
[localhost:21000] > load data inpath '/user/cloudera/hundred_strings.txt' into table
t1;
Query finished, fetching results ...
+-----+
| summary |
+-----+
| Loaded 1 file(s). Total files in destination location: 2 |
+-----+
Returned 1 row(s) in 0.24s
[localhost:21000] > select count(*) from t1;
Query finished, fetching results ...
+-----+
| _c0 |
+-----+
| 1100 |
+-----+
Returned 1 row(s) in 0.55s
[localhost:21000] > load data inpath '/user/cloudera/ten_strings.txt' overwrite into
table t1;
Query finished, fetching results ...
+-----+
| summary |
+-----+
| Loaded 1 file(s). Total files in destination location: 1 |
+-----+
Returned 1 row(s) in 0.26s
[localhost:21000] > select count(*) from t1;
Query finished, fetching results ...
+-----+
| _c0 |
+-----+
| 10 |
+-----+
Returned 1 row(s) in 0.62s
```

### Amazon S3 considerations:

Currently, Impala cannot insert or load data into a table or partition that resides in the Amazon Simple Storage Service (S3). Bring data into S3 using the normal S3 transfer mechanisms, then use Impala to query the S3 data. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about using Impala with S3.

**Cancellation:** Cannot be cancelled.

#### HDFS permissions:

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read and write permissions for the files in the source directory, and write permission for the destination directory.

#### Related information:

The `LOAD DATA` statement is an alternative to the `INSERT` statement. Use `LOAD DATA` when you have the data files in HDFS but outside of any Impala table.

The `LOAD DATA` statement is also an alternative to the `CREATE EXTERNAL TABLE` statement. Use `LOAD DATA` when it is appropriate to move the data files under Impala control rather than querying them from their original location.

## REFRESH Statement

To accurately respond to queries, the Impala node that acts as the coordinator (the node to which you are connected through `impala-shell`, JDBC, or ODBC) must have current metadata about those databases and tables that are referenced in Impala queries. If you are not familiar with the way Impala uses metadata and how it shares the same metastore database as Hive, see [Overview of Impala Metadata and the Metastore](#) on page 21 for background information.

#### Syntax:

```
REFRESH [db_name.]table_name
```

#### Usage notes:

Use the `REFRESH` statement to load the latest metastore metadata and block location data for a particular table in these scenarios:

- After loading new data files into the HDFS data directory for the table. (Once you have set up an ETL pipeline to bring data into Impala on a regular basis, this is typically the most frequent reason why metadata needs to be refreshed.)
- After issuing `ALTER TABLE`, `INSERT`, `LOAD DATA`, or other table-modifying SQL statement in Hive.

You only need to issue the `REFRESH` statement on the node to which you connect to issue queries. The coordinator node divides the work among all the Impala nodes in a cluster, and sends read requests for the correct HDFS blocks without relying on the metadata on the other nodes.

`REFRESH` reloads the metadata for the table from the metastore database, and does an incremental reload of the low-level block location data to account for any new data files added to the HDFS data directory for the table. It is a low-overhead, single-table operation, specifically tuned for the common scenario where new data files are added to HDFS.

Only the metadata for the specified table is flushed. The table must already exist and be known to Impala, either because the `CREATE TABLE` statement was run in Impala rather than Hive, or because a previous `INVALIDATE METADATA` statement caused Impala to reload its entire metadata catalog.

- **Note:**

In Impala 1.2 and higher, the catalog service broadcasts any changed metadata as a result of Impala `ALTER TABLE`, `INSERT` and `LOAD DATA` statements to all Impala nodes. Thus, the `REFRESH` statement is only required if you load data through Hive or by manipulating data files in HDFS directly. See [The Impala Catalog Service](#) on page 19 for more information on the catalog service.

In Impala 1.2.1 and higher, another way to avoid inconsistency across nodes is to enable the `SYNC_DDL` query option before performing a DDL statement or an `INSERT` or `LOAD DATA`.

The functionality of the `REFRESH` statement has changed in Impala 1.1 and higher. Now the table name is a required parameter. To flush the metadata for all tables, use the [`INVALIDATE\_METADATA`](#) command.

Because `REFRESH table_name` only works for tables that Impala is already aware of, when you create a new table in the Hive shell, you must enter `INVALIDATE_METADATA` with no table parameter before you can see the new table in `impala-shell`. Once the table is known to Impala, you can issue `REFRESH table_name` as needed after you add more data files for that table.

`INVALIDATE_METADATA` and `REFRESH` are counterparts: `INVALIDATE_METADATA` waits to reload the metadata when needed for a subsequent query, but reloads all the metadata for the table, which can be an expensive operation, especially for large tables with many partitions. `REFRESH` reloads the metadata immediately, but only loads the block location data for newly added data files, making it a less expensive operation overall. If data was altered in some more extensive way, such as being reorganized by the HDFS balancer, use `INVALIDATE_METADATA` to avoid a performance penalty from reduced local reads. If you used Impala version 1.0, the `INVALIDATE_METADATA` statement works just like the Impala 1.0 `REFRESH` statement did, while the Impala 1.1 `REFRESH` is optimized for the common use case of adding new data files to an existing table, thus the table name argument is now required.

A metadata update for an `impalad` instance **is** required if:

- A metadata change occurs.
- **and** the change is made through Hive.
- **and** the change is made to a database to which clients such as the Impala shell or ODBC directly connect.

A metadata update for an Impala node is **not** required after you run `ALTER TABLE`, `INSERT`, or other table-modifying statement in Impala rather than Hive. Impala handles the metadata synchronization automatically through the catalog service.

Database and table metadata is typically modified by:

- Hive - through `ALTER`, `CREATE`, `DROP` or `INSERT` operations.
- Impalad - through `CREATE TABLE`, `ALTER TABLE`, and `INSERT` operations. In Impala 1.2 and higher, such changes are propagated to all Impala nodes by the Impala catalog service.

`REFRESH` causes the metadata for that table to be immediately reloaded. For a huge table, that process could take a noticeable amount of time; but doing the refresh up front avoids an unpredictable delay later, for example if the next reference to the table is during a benchmark test.

If you connect to different Impala nodes within an `impala-shell` session for load-balancing purposes, you can enable the `SYNC_DDL` query option to make each DDL statement wait before returning, until the new or changed metadata has been received by all the Impala nodes. See [`SYNC\_DDL` Query Option](#) on page 352 for details.

### Examples:

### Examples:

The following example shows how you might use the `REFRESH` statement after manually adding new HDFS data files to the Impala data directory for a table:

```
[impalad-host:21000] > refresh t1;
[impalad-host:21000] > refresh t2;
```

```
[impalad-host:21000] > select * from t1;
...
[impalad-host:21000] > select * from t2;
...
```

For more examples of using `REFRESH` and `INVALIDATE METADATA` with a combination of Impala and Hive operations, see [Switching Back and Forth Between Impala and Hive](#) on page 65.

#### Related impalad options:

In Impala 1.0, the `-r` option of `impala-shell` issued `REFRESH` to reload metadata for all tables.

In Impala 1.1 and higher, this option issues `INVALIDATE METADATA` because `REFRESH` now requires a table name parameter. Due to the expense of reloading the metadata for all tables, the `impala-shell -r` option is not recommended for day-to-day use in a production environment.

In Impala 1.2 and higher, the `-r` option is needed even less frequently, because metadata changes caused by SQL statements in Impala are automatically broadcast to all Impala nodes.

#### HDFS permissions:

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have execute permissions for all the relevant directories holding table data. (A table could have data spread across multiple directories, or in unexpected paths, if it uses partitioning or specifies a `LOCATION` attribute for individual partitions or the entire table.) Issues with permissions might not cause an immediate error for this statement, but subsequent statements such as `SELECT` or `SHOW TABLE STATS` could fail.

#### HDFS considerations:

The `REFRESH` command checks HDFS permissions of the underlying data files and directories, caching this information so that a statement can be cancelled immediately if for example the `impala` user does not have permission to write to the data directory for the table. Impala reports any lack of write permissions as an `INFO` message in the log file, in case that represents an oversight. If you change HDFS permissions to make data readable or writeable by the Impala user, issue another `REFRESH` to make Impala aware of the change.

- **Important:** After adding or replacing data in a table used in performance-critical queries, issue a `COMPUTE STATS` statement to make sure all statistics are up-to-date. Consider updating statistics for a table after any `INSERT`, `LOAD DATA`, or `CREATE TABLE AS SELECT` statement in Impala, or after loading data through Hive and doing a `REFRESH table_name` in Impala. This technique is especially important for tables that are very large, used in join queries, or both.

#### Amazon S3 considerations:

The `REFRESH` and `INVALIDATE METADATA` statements also cache metadata for tables where the data resides in the Amazon Simple Storage Service (S3). In particular, issue a `REFRESH` for a table after adding or removing files in the associated S3 data directory. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about working with S3 tables.

**Cancellation:** Cannot be cancelled.

#### Related information:

[Overview of Impala Metadata and the Metastore](#) on page 21, [INVALIDATE METADATA Statement](#) on page 210

## REVOKE Statement (CDH 5.2 or higher only)

The `REVOKE` statement revokes roles or privileges on a specified object from groups. Only Sentry administrative users can revoke the role from a group. The revocation has a cascading effect. For example, revoking the `ALL` privilege on a database also revokes the same privilege for all the tables in that database.

## Syntax:

```
REVOKE ROLE role_name FROM GROUP group_name  
REVOKE privilege ON object_type object_name  
    FROM [ROLE] role_name  
  
privilege ::= SELECT | INSERT | ALL  
object_type ::= TABLE | DATABASE | SERVER | URI
```

Typically, the object name is an identifier. For URIs, it is a string literal.

## Required privileges:

Only administrative users (those with `ALL` privileges on the server, defined in the Sentry policy file) can use this statement.

## Compatibility:

- The Impala `GRANT` and `REVOKE` statements are available in CDH 5.2 and higher.
- In CDH 5.1 and higher, Impala makes use of any roles and privileges specified by the `GRANT` and `REVOKE` statements in Hive, when your system is configured to use the Sentry service instead of the file-based policy mechanism.
- The Impala `GRANT` and `REVOKE` statements do not require the `ROLE` keyword to be repeated before each role name, unlike the equivalent Hive statements.
- Currently, each Impala `GRANT` or `REVOKE` statement can only grant or revoke a single privilege to or from a single role.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

## Related information:

[Enabling Sentry Authorization for Impala](#) on page 91, [GRANT Statement \(CDH 5.2 or higher only\)](#) on page 201, [CREATE ROLE Statement \(CDH 5.2 or higher only\)](#) on page 179, [DROP ROLE Statement \(CDH 5.2 or higher only\)](#) on page 193, [SHOW Statement](#) on page 242

## SELECT Statement

The `SELECT` statement performs queries, retrieving data from one or more tables and producing result sets consisting of rows and columns.

The Impala [INSERT](#) statement also typically ends with a `SELECT` statement, to define data to copy from one table to another.

## Syntax:

```
[WITH name AS (select_expression) [, ...] ]  
SELECT  
  [ALL | DISTINCT]  
  [STRAIGHT_JOIN]  
  expression [, expression ...]  
FROM table_reference [, table_reference ...]  
[[[FULL | [LEFT | RIGHT] INNER | [LEFT | RIGHT] OUTER | [LEFT | RIGHT] SEMI | [LEFT |  
RIGHT] ANTI | CROSS]  
  JOIN table_reference  
  [ON join_equality_clauses | USING (col1[, col2 ...])] ...  
WHERE conditions  
GROUP BY { column | expression [ASC | DESC] [NULLS FIRST | NULLS LAST] [, ...] }  
HAVING conditions  
GROUP BY { column | expression [ASC | DESC] [, ...] }  
LIMIT expression [OFFSET expression]  
[UNION [ALL] select_statement] ...]
```

Impala SELECT queries support:

- SQL data types: [BOOLEAN](#), [TINYINT](#), [SMALLINT](#), [INT](#), [BIGINT](#), [DECIMAL](#), [FLOAT](#), [DOUBLE](#), [TIMESTAMP](#), [STRING](#), [VARCHAR](#), [CHAR](#).
- An optional [WITH clause](#) before the `SELECT` keyword, to define a subquery whose name or column names can be referenced from later in the main query. This clause lets you abstract repeated clauses, such as aggregation functions, that are referenced multiple times in the same query.
- By default, one `DISTINCT` clause per query. See [DISTINCT Operator](#) on page 238 for details. See [APPX\\_COUNT\\_DISTINCT Query Option](#) on page 340 for a query option to allow multiple `COUNT(DISTINCT)` impressions in the same query.
- Subqueries in a `FROM` clause.
- `WHERE`, `GROUP BY`, `HAVING` clauses.
- [ORDER BY](#). Prior to Impala 1.4.0, Impala required that queries using an `ORDER BY` clause also include a [LIMIT](#) clause. In Impala 1.4.0 and higher, this restriction is lifted; sort operations that would exceed the Impala memory limit automatically use a temporary disk work area to perform the sort.
- Impala supports a wide variety of `JOIN` clauses. Left, right, semi, full, and outer joins are supported in all Impala versions. The `CROSS JOIN` operator is available in Impala 1.2.2 and higher. During performance tuning, you can override the reordering of join clauses that Impala does internally by including the keyword `STRAIGHT_JOIN` immediately after the `SELECT` keyword

See [Joins](#) on page 220 for details and examples of join queries.

- `UNION ALL`.
- `LIMIT`.
- External tables.
- Relational operators such as greater than, less than, or equal to.
- Arithmetic operators such as addition or subtraction.
- Logical/Boolean operators `AND`, `OR`, and `NOT`. Impala does not support the corresponding symbols `&&`, `||`, and `!`.
- Common SQL built-in functions such as `COUNT`, `SUM`, `CAST`, `LIKE`, `IN`, `BETWEEN`, and `COALESCE`. Impala specifically supports built-ins described in [Built-in Functions](#) on page 257.

Impala queries ignore files with extensions commonly used for temporary work files by Hadoop tools. Any files with extensions `.tmp` or `.copying` are not considered part of the Impala table. The suffix matching is case-insensitive, so for example Impala ignores both `.copying` and `.COPYING` suffixes.

#### **Security considerations:**

If these statements in your environment contain sensitive literal values such as credit card numbers or tax identifiers, Impala can redact this sensitive information when displaying the statements in log files and other administrative contexts. See [Sensitive Data Redaction](#) for details.

**Cancellation:** Can be cancelled. To cancel this statement, use Ctrl-C from the `impala-shell` interpreter, the **Cancel** button from the **Watch** page in Hue, **Actions > Cancel** from the **Queries** list in Cloudera Manager, or **Cancel** from the list of in-flight queries (for a particular node) on the **Queries** tab in the Impala web UI (port 25000).

#### **HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read permissions for the files in all applicable directories in all source tables, and read and execute permissions for the relevant data directories. (A `SELECT` operation could read files from multiple different HDFS directories if the source table is partitioned.) If a query attempts to read a data file and is unable to because of an HDFS permission error, the query halts and does not return any further results.

#### **Related information:**

The `SELECT` syntax is so extensive that it forms its own category of statements: queries. The other major classifications of SQL statements are data definition language (see [DDL Statements](#) on page 162) and data manipulation language (see [DML Statements](#) on page 163).

## Impala SQL Language Reference

Because the focus of Impala is on fast queries with interactive response times over huge data sets, query performance and scalability are important considerations. See [Tuning Impala for Performance](#) on page 354 and [Scalability Considerations for Impala](#) on page 388 for details.

### Joins

A join query is a `SELECT` statement that combines data from two or more tables, and returns a result set containing items from some or all of those tables. It is a way to cross-reference and correlate related data that is organized into multiple tables, typically using identifiers that are repeated in each of the joined tables.

#### Syntax:

Impala supports a wide variety of `JOIN` clauses. Left, right, semi, full, and outer joins are supported in all Impala versions. The `CROSS JOIN` operator is available in Impala 1.2.2 and higher. During performance tuning, you can override the reordering of join clauses that Impala does internally by including the keyword `STRAIGHT_JOIN` immediately after the `SELECT` keyword

```
SELECT select_list FROM
  table_or_subquery1 [INNER] JOIN table_or_subquery2 |
  table_or_subquery1 {LEFT [OUTER] | RIGHT [OUTER] | FULL [OUTER]} JOIN
  table_or_subquery2 |
  table_or_subquery1 {LEFT | RIGHT} SEMI JOIN table_or_subquery2 |
  table_or_subquery1 {LEFT | RIGHT} ANTI JOIN table_or_subquery2 |
  [ ON col1 = col2 [AND col3 = col4 ...] |
    USING (col1 [, col2 ...]) ]
  [other_join_clause ...]
[ WHERE where_clauses ]
```

```
SELECT select_list FROM
  table_or_subquery1, table_or_subquery2 [, table_or_subquery3 ...]
  [other_join_clause ...]
WHERE
  col1 = col2 [AND col3 = col4 ...]
```

```
SELECT select_list FROM
  table_or_subquery1 CROSS JOIN table_or_subquery2
  [other_join_clause ...]
[ WHERE where_clauses ]
```

#### SQL-92 and SQL-89 Joins:

Queries with the explicit `JOIN` keywords are known as SQL-92 style joins, referring to the level of the SQL standard where they were introduced. The corresponding `ON` or `USING` clauses clearly show which columns are used as the join keys in each case:

```
SELECT t1.c1, t2.c2 FROM t1 JOIN t2
  ON t1.id = t2.id and t1.type_flag = t2.type_flag
  WHERE t1.c1 > 100;
```

```
SELECT t1.c1, t2.c2 FROM t1 JOIN t2
  USING (id, type_flag)
  WHERE t1.c1 > 100;
```

The `ON` clause is a general way to compare columns across the two tables, even if the column names are different. The `USING` clause is a shorthand notation for specifying the join columns, when the column names are the same in both tables. You can code equivalent `WHERE` clauses that compare the columns, instead of `ON` or `USING` clauses, but that practice is not recommended because mixing the join comparisons with other filtering clauses is typically less readable and harder to maintain.

Queries with a comma-separated list of tables and subqueries are known as SQL-89 style joins. In these queries, the equality comparisons between columns of the joined tables go in the `WHERE` clause alongside other kinds

of comparisons. This syntax is easy to learn, but it is also easy to accidentally remove a WHERE clause needed for the join to work correctly.

```
SELECT t1.c1, t2.c2 FROM t1, t2
  WHERE
    t1.id = t2.id AND t1.type_flag = t2.type_flag
    AND t1.c1 > 100;
```

### Self-joins:

Impala can do self-joins, for example to join on two different columns in the same table to represent parent-child relationships or other tree-structured data. There is no explicit syntax for this; just use the same table name for both the left-hand and right-hand table, and assign different table aliases to use when referring to the fully qualified column names:

```
-- Combine fields from both parent and child rows.
SELECT lhs.id, rhs.parent, lhs.c1, rhs.c2 FROM tree_data lhs, tree_data rhs WHERE lhs.id
= rhs.parent;
```

### Cartesian joins:

To avoid producing huge result sets by mistake, Impala does not allow Cartesian joins of the form:

```
SELECT ... FROM t1 JOIN t2;
SELECT ... FROM t1, t2;
```

If you intend to join the tables based on common values, add ON or WHERE clauses to compare columns across the tables. If you truly intend to do a Cartesian join, use the CROSS JOIN keyword as the join operator. The CROSS JOIN form does not use any ON clause, because it produces a result set with all combinations of rows from the left-hand and right-hand tables. The result set can still be filtered by subsequent WHERE clauses. For example:

```
SELECT ... FROM t1 CROSS JOIN t2;
SELECT ... FROM t1 CROSS JOIN t2 WHERE tests_on_non_join_columns;
```

### Inner and outer joins:

An inner join is the most common and familiar type: rows in the result set contain the requested columns from the appropriate tables, for all combinations of rows where the join columns of the tables have identical values. If a column with the same name occurs in both tables, use a fully qualified name or a column alias to refer to the column in the select list or other clauses. Impala performs inner joins by default for both SQL-89 and SQL-92 join syntax:

```
-- The following 3 forms are all equivalent.
SELECT t1.id, c1, c2 FROM t1, t2 WHERE t1.id = t2.id;
SELECT t1.id, c1, c2 FROM t1 JOIN t2 ON t1.id = t2.id;
SELECT t1.id, c1, c2 FROM t1 INNER JOIN t2 ON t1.id = t2.id;
```

An outer join retrieves all rows from the left-hand table, or the right-hand table, or both; wherever there is no matching data in the table on the other side of the join, the corresponding columns in the result set are set to NULL. To perform an outer join, include the OUTER keyword in the join operator, along with either LEFT, RIGHT, or FULL:

```
SELECT * FROM t1 LEFT OUTER JOIN t2 ON t1.id = t2.id;
SELECT * FROM t1 RIGHT OUTER JOIN t2 ON t1.id = t2.id;
SELECT * FROM t1 FULL OUTER JOIN t2 ON t1.id = t2.id;
```

For outer joins, Impala requires SQL-92 syntax; that is, the JOIN keyword instead of comma-separated table names. Impala does not support vendor extensions such as (+) or \*= notation for doing outer joins with SQL-89 query syntax.

### Equijoins and Non-Equijoins:

## Impala SQL Language Reference

By default, Impala requires an equality comparison between the left-hand and right-hand tables, either through `ON`, `USING`, or `WHERE` clauses. These types of queries are classified broadly as equijoins. Inner, outer, full, and semi joins can all be equijoins based on the presence of equality tests between columns in the left-hand and right-hand tables.

In Impala 1.2.2 and higher, non-equijoin queries are also possible, with comparisons such as `!=` or `<` between the join columns. These kinds of queries require care to avoid producing huge result sets that could exceed resource limits. Once you have planned a non-equijoin query that produces a result set of acceptable size, you can code the query using the `CROSS JOIN` operator, and add the extra comparisons in the `WHERE` clause:

```
SELECT ... FROM t1 CROSS JOIN t2 WHERE t1.total > t2.maximum_price;
```

### Semi-joins:

Semi-joins are a relatively rarely used variation. With the left semi-join, only data from the left-hand table is returned, for rows where there is matching data in the right-hand table, based on comparisons between join columns in `ON` or `WHERE` clauses. Only one instance of each row from the left-hand table is returned, regardless of how many matching rows exist in the right-hand table. A right semi-join (available in Impala 2.0 and higher) reverses the comparison and returns data from the right-hand table.

```
SELECT t1.c1, t1.c2, t1.c3 FROM t1 LEFT SEMI JOIN t2 ON t1.id = t2.id;
```

### Natural joins (not supported):

Impala does not support the `NATURAL JOIN` operator, again to avoid inconsistent or huge result sets. Natural joins do away with the `ON` and `USING` clauses, and instead automatically join on all columns with the same names in the left-hand and right-hand tables. This kind of query is not recommended for rapidly evolving data structures such as are typically used in Hadoop. Thus, Impala does not support the `NATURAL JOIN` syntax, which can produce different query results as columns are added to or removed from tables.

If you do have any queries that use `NATURAL JOIN`, make sure to rewrite them with explicit `USING` clauses, because Impala could interpret the `NATURAL` keyword as a table alias:

```
-- 'NATURAL' is interpreted as an alias for 't1' and Impala attempts an inner join,
-- resulting in an error because inner joins require explicit comparisons between
-- columns.
SELECT t1.c1, t2.c2 FROM t1 NATURAL JOIN t2;
ERROR: NotImplementedException: Join with 't2' requires at least one conjunctive equality
predicate.
To perform a Cartesian product between two tables, use a CROSS JOIN.

-- If you expect the tables to have identically named columns with matching values,
-- list the corresponding column names in a USING clause.
SELECT t1.c1, t2.c2 FROM t1 JOIN t2 USING (id, type_flag, name, address);
```

### Anti-joins (Impala 2.0 / CDH 5.2 and higher only):

Impala supports the `LEFT ANTI JOIN` and `RIGHT ANTI JOIN` clauses in Impala 2.0 and higher on CDH 4, or CDH 5.2 and higher on CDH 5. The `LEFT` or `RIGHT` keyword is required for this kind of join. For `LEFT ANTI JOIN`, this clause returns those values from the left-hand table that have no matching value in the right-hand table. `RIGHT ANTI JOIN` reverses the comparison and returns values from the right-hand table. You can express this negative relationship either through the `ANTI JOIN` clause or through a `NOT EXISTS` operator with a subquery.

### Usage notes:

You typically use join queries in situations like these:

- When related data arrives from different sources, with each data set physically residing in a separate table. For example, you might have address data from business records that you cross-check against phone listings or census data.

- **Note:** Impala can join tables of different file formats, including Impala-managed tables and HBase tables. For example, you might keep small dimension tables in HBase, for convenience of single-row lookups and updates, and for the larger fact tables use Parquet or other binary file format optimized for scan operations. Then, you can issue a join query to cross-reference the fact tables with the dimension tables.
- When data is normalized, a technique for reducing data duplication by dividing it across multiple tables. This kind of organization is often found in data that comes from traditional relational database systems. For example, instead of repeating some long string such as a customer name in multiple tables, each table might contain a numeric customer ID. Queries that need to display the customer name could "join" the table that specifies which customer ID corresponds to which name.
- When certain columns are rarely needed for queries, so they are moved into separate tables to reduce overhead for common queries. For example, a biography field might be rarely needed in queries on employee data. Putting that field in a separate table reduces the amount of I/O for common queries on employee addresses or phone numbers. Queries that do need the biography column can retrieve it by performing a join with that separate table.

When comparing columns with the same names in ON or WHERE clauses, use the fully qualified names such as `db_name.table_name`, or assign table aliases, column aliases, or both to make the code more compact and understandable:

```
select t1.c1 as first_id, t2.c2 as second_id from
  t1 join t2 on first_id = second_id;

select fact.custno, dimension.custno from
  customer_data as fact join customer_address as dimension
  using (custno)
```

▪ **Note:**

Performance for join queries is a crucial aspect for Impala, because complex join queries are resource-intensive operations. An efficient join query produces much less network traffic and CPU overhead than an inefficient one. For best results:

- Make sure that both [table and column statistics](#) are available for all the tables involved in a join query, and especially for the columns referenced in any join conditions. Impala uses the statistics to automatically deduce an efficient join order. Use `SHOW TABLE STATS table_name` and `SHOW COLUMN STATS table_name` to check if statistics are already present. Issue the `COMPUTE STATS table_name` for a nonpartitioned table, or (in Impala 2.1.0 and higher) `COMPUTE INCREMENTAL STATS table_name` for a partitioned table, to collect the initial statistics at both the table and column levels, and to keep the statistics up to date after any substantial `INSERT` or `LOAD DATA` operations.
- If table or column statistics are not available, join the largest table first. You can check the existence of statistics with the `SHOW TABLE STATS table_name` and `SHOW COLUMN STATS table_name` statements.
- If table or column statistics are not available, join subsequent tables according to which table has the most selective filter, based on overall size and WHERE clauses. Joining the table with the most selective filter results in the fewest number of rows being returned.

For more information and examples of performance for join queries, see [Performance Considerations for Join Queries](#) on page 357.

To control the result set from a join query, include the names of corresponding column names in both tables in an ON or USING clause, or by coding equality comparisons for those columns in the WHERE clause.

```
[localhost:21000] > select c_last_name, ca_city from customer join customer_address
  where c_customer_sk = ca_address_sk;
+-----+-----+

```

```

+-----+-----+
| c_last_name | ca_city      |
+-----+-----+
| Lewis       | Fairfield    |
| Moses       | Fairview     |
| Hamilton    | Pleasant Valley|
| White       | Oak Ridge    |
| Moran       | Glendale    |
+-----+-----+
| Richards    | Lakewood     |
| Day          | Lebanon      |
| Painter      | Oak Hill     |
| Bentley      | Greenfield   |
| Jones        | Stringtown   |
+-----+-----+
Returned 50000 row(s) in 9.82s

```

One potential downside of joins is the possibility of excess resource usage in poorly constructed queries. Impala imposes restrictions on join queries to guard against such issues. To minimize the chance of runaway queries on large data sets, Impala requires every join query to contain at least one equality predicate between the columns of the various tables. For example, if t1 contains 1000 rows and t2 contains 1,000,000 rows, a query `SELECT columns FROM t1 JOIN t2` could return up to 1 billion rows ( $1000 * 1,000,000$ ); Impala requires that the query include a clause such as `ON t1.c1 = t2.c2` or `WHERE t1.c1 = t2.c2`.

Because even with equality clauses, the result set can still be large, as we saw in the previous example, you might use a `LIMIT` clause to return a subset of the results:

```

[localhost:21000] > select c_last_name, ca_city from customer, customer_address where
  c_customer_sk = ca_address_sk limit 10;
+-----+-----+
| c_last_name | ca_city      |
+-----+-----+
| Lewis       | Fairfield    |
| Moses       | Fairview     |
| Hamilton    | Pleasant Valley|
| White       | Oak Ridge    |
| Moran       | Glendale    |
| Sharp        | Lakeview     |
| Wiles        | Farmington   |
| Shipman     | Union        |
| Gilbert      | New Hope    |
| Brunson     | Martinsville |
+-----+-----+
Returned 10 row(s) in 0.63s

```

Or you might use additional comparison operators or aggregation functions to condense a large result set into a smaller set of values:

```

[localhost:21000] > -- Find the names of customers who live in one particular town.
[localhost:21000] > select distinct c_last_name from customer, customer_address where
  c_customer_sk = ca_address_sk
  and ca_city = "Green Acres";
+-----+
| c_last_name |
+-----+
| Hensley    |
| Pearson    |
| Mayer      |
| Montgomery |
| Ricks      |
...
| Barrett    |
| Price      |
| Hill       |
| Hansen    |
| Meeks      |
+-----+
Returned 332 row(s) in 0.97s

[localhost:21000] > -- See how many different customers in this town have names starting

```

```

with "A".
[localhost:21000] > select count(distinct c_last_name) from customer, customer_address
  where
    c_customer_sk = ca_address_sk
    and ca_city = "Green Acres"
    and substr(c_last_name,1,1) = "A";
+-----+
| count(distinct c_last_name) |
+-----+
| 12                         |
+-----+
Returned 1 row(s) in 1.00s

```

Because a join query can involve reading large amounts of data from disk, sending large amounts of data across the network, and loading large amounts of data into memory to do the comparisons and filtering, you might do benchmarking, performance analysis, and query tuning to find the most efficient join queries for your data set, hardware capacity, network configuration, and cluster workload.

The two categories of joins in Impala are known as **partitioned joins** and **broadcast joins**. If inaccurate table or column statistics, or some quirk of the data distribution, causes Impala to choose the wrong mechanism for a particular join, consider using query hints as a temporary workaround. For details, see [Hints](#) on page 239.

### Examples:

The following examples refer to these simple tables containing small sets of integers:

```

[localhost:21000] > create table t1 (x int);
[localhost:21000] > insert into t1 values (1), (2), (3), (4), (5), (6);

[localhost:21000] > create table t2 (y int);
[localhost:21000] > insert into t2 values (2), (4), (6);

[localhost:21000] > create table t3 (z int);
[localhost:21000] > insert into t3 values (1), (3), (5);

```

The following example demonstrates an anti-join, returning the values from `t1` that do not exist in `t2` (in this case, the odd numbers 1, 3, and 5):

```

[localhost:21000] > select x from t1 left anti join t2 on (t1.x = t2.y);
+---+
| x |
+---+
| 1 |
| 3 |
| 5 |
+---+

```

### Related information:

See these tutorials for examples of different kinds of joins:

- [Cross Joins and Cartesian Products with the CROSS JOIN Operator](#) on page 66

## ORDER BY Clause

The familiar `ORDER BY` clause of a `SELECT` statement sorts the result set based on the values from one or more columns.

For distributed queries, this is a relatively expensive operation, because the entire result set must be produced and transferred to one node before the sorting can happen. This can require more memory capacity than a query without `ORDER BY`. Even if the query takes approximately the same time to finish with or without the `ORDER BY` clause, subjectively it can appear slower because no results are available until all processing is finished, rather than results coming back gradually as rows matching the `WHERE` clause are found. Therefore, if you only need the first N results from the sorted result set, also include the `LIMIT` clause, which reduces network overhead and the memory requirement on the coordinator node.

- **Note:**

In Impala 1.4.0 and higher, the `LIMIT` clause is now optional (rather than required) for queries that use the `ORDER BY` clause. Impala automatically uses a temporary disk work area to perform the sort if the sort operation would otherwise exceed the Impala memory limit for a particular data node.

## Syntax:

The full syntax for the `ORDER BY` clause is:

```
ORDER BY col_ref [, col_ref ...] [ASC | DESC] [NULLS FIRST | NULLS LAST]  
col_ref ::= column_name | integer_literal
```

Although the most common usage is `ORDER BY column_name`, you can also specify `ORDER BY 1` to sort by the first column of the result set, `ORDER BY 2` to sort by the second column, and so on. The number must be a numeric literal, not some other kind of constant expression. (If the argument is some other expression, even a `STRING` value, the query succeeds but the order of results is undefined.)

`ORDER BY column_number` can only be used when the query explicitly lists the columns in the `SELECT` list, not with `SELECT *` queries.

### Ascending and descending sorts:

The default sort order (the same as using the `ASC` keyword) puts the smallest values at the start of the result set, and the largest values at the end. Specifying the `DESC` keyword reverses that order.

### Sort order for NULL values:

See [NULL](#) on page 141 for details about how `NULL` values are positioned in the sorted result set, and how to use the `NULLS FIRST` and `NULLS LAST` clauses. (The sort position for `NULL` values in `ORDER BY ... DESC` queries is changed in Impala 1.2.1 and higher to be more standards-compliant, and the `NULLS FIRST` and `NULLS LAST` keywords are new in Impala 1.2.1.)

Prior to Impala 1.4.0, Impala required any query including an `ORDER BY` clause to also use a `LIMIT` clause. In Impala 1.4.0 and higher, the `LIMIT` clause is optional for `ORDER BY` queries. In cases where sorting a huge result set requires enough memory to exceed the Impala memory limit for a particular node, Impala automatically uses a temporary disk work area to perform the sort operation.

### Usage notes:

Although the `LIMIT` clause is now optional on `ORDER BY` queries, if your query only needs some number of rows that you can predict in advance, use the `LIMIT` clause to reduce unnecessary processing. For example, if the query has a clause `LIMIT 10`, each data node sorts its portion of the relevant result set and only returns 10 rows to the coordinator node. The coordinator node picks the 10 highest or lowest row values out of this small intermediate result set.

If an `ORDER BY` clause is applied to an early phase of query processing, such as a subquery or a view definition, Impala ignores the `ORDER BY` clause. To get ordered results from a subquery or view, apply an `ORDER BY` clause to the outermost or final `SELECT` level.

`ORDER BY` is often used in combination with `LIMIT` to perform "top-N" queries:

```
SELECT user_id as "Top 10 Visitors", SUM(page_views) FROM web_stats  
GROUP BY page_views, user_id  
ORDER BY SUM(page_views) DESC LIMIT 10;
```

`ORDER BY` is sometimes used in combination with `OFFSET` and `LIMIT` to paginate query results, although it is relatively inefficient to issue multiple queries like this against the large tables typically used with Impala:

```
SELECT page_title as "Page 1 of search results", page_url FROM search_content  
WHERE LOWER(page_title) LIKE '%game%'  
ORDER BY page_title LIMIT 10 OFFSET 0;
```

```

SELECT page_title as "Page 2 of search results", page_url FROM search_content
  WHERE LOWER(page_title) LIKE '%game%')
    ORDER BY page_title LIMIT 10 OFFSET 10;
SELECT page_title as "Page 3 of search results", page_url FROM search_content
  WHERE LOWER(page_title) LIKE '%game%')
    ORDER BY page_title LIMIT 10 OFFSET 20;

```

**Internal details:**

Impala sorts the intermediate results of an `ORDER BY` clause in memory whenever practical. In a cluster of N data nodes, each node sorts roughly 1/Nth of the result set, the exact proportion varying depending on how the data matching the query is distributed in HDFS.

If the size of the sorted intermediate result set on any data node would cause the query to exceed the Impala memory limit, Impala sorts as much as practical in memory, then writes partially sorted data to disk. (This technique is known in industry terminology as “external sorting” and “spilling to disk.”) As each 8 MB batch of data is written to disk, Impala frees the corresponding memory to sort a new 8 MB batch of data. When all the data has been processed, a final merge sort operation is performed to correctly order the in-memory and on-disk results as the result set is transmitted back to the coordinator node. When external sorting becomes necessary, Impala requires approximately 60 MB of RAM at a minimum for the buffers needed to read, write, and sort the intermediate results. If more RAM is available on the data node, Impala will use the additional RAM to minimize the amount of disk I/O for sorting.

This external sort technique is used as appropriate on each data node (possibly including the coordinator node) to sort the portion of the result set that is processed on that node. When the sorted intermediate results are sent back to the coordinator node to produce the final result set, the coordinator node uses a merge sort technique to produce a final sorted result set without using any extra resources on the coordinator node.

**Configuration for disk usage:**

By default, intermediate files used during large sort, join, aggregation, or analytic function operations are stored in the directory `/tmp/impala-scratch`. These files are removed when the operation finishes. (Multiple concurrent queries can perform operations that use the “spill to disk” technique, without any name conflicts for these temporary files.) You can specify a different location by starting the `impalad` daemon with the `--scratch_dirs="path_to_directory"` configuration option or the equivalent configuration option in the Cloudera Manager user interface. You can specify a single directory, or a comma-separated list of directories. The scratch directories must be on the local filesystem, not in HDFS. You might specify different directory paths for different hosts, depending on the capacity and speed of the available storage devices. Impala will not start if it cannot create or read and write files in the “scratch” directory. If there is less than 1 GB free on the filesystem where that directory resides, Impala still runs, but writes a warning message to its log.

**Sorting considerations:** Although you can specify an `ORDER BY` clause in an `INSERT ... SELECT` statement, any `ORDER BY` clause is ignored and the results are not necessarily sorted. An `INSERT ... SELECT` operation potentially creates many different data files, prepared on different data nodes, and therefore the notion of the data being stored in sorted order is impractical.

An `ORDER BY` clause without an additional `LIMIT` clause is ignored in any view definition. If you need to sort the entire result set from a view, use an `ORDER BY` clause in the `SELECT` statement that queries the view. You can still make a simple “top 10” report by combining the `ORDER BY` and `LIMIT` clauses in the same view definition:

```

[localhost:21000] > create table unsorted (x bigint);
[localhost:21000] > insert into unsorted values (1), (9), (3), (7), (5), (8), (4), (6),
(2);
[localhost:21000] > create view sorted_view as select x from unsorted order by x;
[localhost:21000] > select x from sorted_view; -- ORDER BY clause in view has no effect.
+---+
| x |
+---+
| 1 |
| 9 |
| 3 |
| 7 |
| 5 |

```

```

+---+
| 8 |
| 4 |
| 6 |
| 2 |
+---+
[localhost:21000] > select x from sorted_view order by x; -- View query requires ORDER
    BY at outermost level.
+---+
| x |
+---+
| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
| 6 |
| 7 |
| 8 |
| 9 |
+---+
[localhost:21000] > create view top_3_view as select x from unsorted order by x limit
    3;
[localhost:21000] > select x from top_3_view; -- ORDER BY and LIMIT together in view
    definition are preserved.
+---+
| x |
+---+
| 1 |
| 2 |
| 3 |
+---+

```

With the lifting of the requirement to include a `LIMIT` clause in every `ORDER BY` query (in Impala 1.4 and higher):

- Now the use of scratch disk space raises the possibility of an “out of disk space” error on a particular data node, as opposed to the previous possibility of an “out of memory” error. Make sure to keep at least 1 GB free on the filesystem used for temporary sorting work.
- The query options `DEFAULT_ORDER_BY_LIMIT` and `ABORT_ON_DEFAULT_LIMIT_EXCEEDED`, which formerly controlled the behavior of `ORDER BY` queries with no limit specified, are now ignored.

In Impala 1.2.1 and higher, all `NULL` values come at the end of the result set for `ORDER BY ... ASC` queries, and at the beginning of the result set for `ORDER BY ... DESC` queries. In effect, `NULL` is considered greater than all other values for sorting purposes. The original Impala behavior always put `NULL` values at the end, even for `ORDER BY ... DESC` queries. The new behavior in Impala 1.2.1 makes Impala more compatible with other popular database systems. In Impala 1.2.1 and higher, you can override or specify the sorting behavior for `NULL` by adding the clause `NULLS FIRST` or `NULLS LAST` at the end of the `ORDER BY` clause.

```

[localhost:21000] > create table numbers (x int);
[localhost:21000] > insert into numbers values (1), (null), (2), (null), (3);
[localhost:21000] > select x from numbers order by x nulls first;
+---+
| x |
+---+
| NULL |
| NULL |
| 1 |
| 2 |
| 3 |
+---+
[localhost:21000] > select x from numbers order by x desc nulls first;
+---+
| x |
+---+
| NULL |
| NULL |
| 3 |
| 2 |
| 1 |
+---+

```

```
+----+
[localhost:21000] > select x from numbers order by x nulls last;
+----+
| x |
+----+
| 1 |
| 2 |
| 3 |
| NULL |
| NULL |
+----+
[localhost:21000] > select x from numbers order by x desc nulls last;
+----+
| x |
+----+
| 3 |
| 2 |
| 1 |
| NULL |
| NULL |
+----+
```

**Related information:**

See [SELECT Statement](#) on page 218 for further examples of queries with the `ORDER BY` clause.

Analytic functions use the `ORDER BY` clause in a different context to define the sequence in which rows are analyzed. See [Impala Analytic Functions](#) on page 296 for details.

**GROUP BY Clause**

Specify the `GROUP BY` clause in queries that use aggregation functions, such as `COUNT()`, `SUM()`, `AVG()`, `MIN()`, and `MAX()`. Specify in the `GROUP BY` clause the names of all the columns that do not participate in the aggregation operation.

For example, the following query finds the 5 items that sold the highest total quantity (using the `SUM()` function, and also counts the number of sales transactions for those items (using the `COUNT()` function). Because the column representing the item IDs is not used in any aggregation functions, we specify that column in the `GROUP BY` clause.

```
select
  ss_item_sk as Item,
  count(ss_item_sk) as Times_Purchased,
  sum(ss_quantity) as Total_Quantity_Purchased
from store_sales
group by ss_item_sk
order by sum(ss_quantity) desc
limit 5;
```

item	times_purchased	total_quantity_purchased
9325	372	19072
4279	357	18501
7507	371	18475
5953	369	18451
16753	375	18446

The `HAVING` clause lets you filter the results of aggregate functions, because you cannot refer to those expressions in the `WHERE` clause. For example, to find the 5 lowest-selling items that were included in at least 100 sales transactions, we could use this query:

```
select
  ss_item_sk as Item,
  count(ss_item_sk) as Times_Purchased,
  sum(ss_quantity) as Total_Quantity_Purchased
from store_sales
group by ss_item_sk
```

```
having times_purchased >= 100
order by sum(ss_quantity)
limit 5;
+-----+-----+-----+
| item | times_purchased | total_quantity_purchased |
+-----+-----+-----+
| 13943 | 105 | 4087 |
| 2992 | 101 | 4176 |
| 4773 | 107 | 4204 |
| 14350 | 103 | 4260 |
| 11956 | 102 | 4275 |
+-----+-----+-----+
```

When performing calculations involving scientific or financial data, remember that columns with type `FLOAT` or `DOUBLE` are stored as true floating-point numbers, which cannot precisely represent every possible fractional value. Thus, if you include a `FLOAT` or `DOUBLE` column in a `GROUP BY` clause, the results might not precisely match literal values in your query or from an original Text data file. Use rounding operations, the `BETWEEN` operator, or another arithmetic technique to match floating-point values that are “near” literal values you expect. For example, this query on the `ss_wholesale_cost` column returns cost values that are close but not identical to the original figures that were entered as decimal fractions.

```
select ss_wholesale_cost, avg(ss_quantity * ss_sales_price) as avg_revenue_per_sale
from sales
group by ss_wholesale_cost
order by avg_revenue_per_sale desc
limit 5;
+-----+-----+
| ss_wholesale_cost | avg_revenue_per_sale |
+-----+-----+
| 96.94000244140625 | 4454.351539300434 |
| 95.93000030517578 | 4423.119941283189 |
| 98.37999725341797 | 4332.516490316291 |
| 97.97000122070312 | 4330.480601655014 |
| 98.52999877929688 | 4291.316953108634 |
+-----+-----+
```

Notice how wholesale cost values originally entered as decimal fractions such as 96.94 and 98.38 are slightly larger or smaller in the result set, due to precision limitations in the hardware floating-point types. The imprecise representation of `FLOAT` and `DOUBLE` values is why financial data processing systems often store currency using data types that are less space-efficient but avoid these types of rounding errors.

**Zero-length strings:** For purposes of clauses such as `DISTINCT` and `GROUP BY`, Impala considers zero-length strings (" "), `NULL`, and space to all be different values.

### Related information:

[SELECT Statement](#) on page 218, [Impala Aggregate Functions](#) on page 284

## HAVING Clause

Performs a filter operation on a `SELECT` query, by examining the results of aggregation functions rather than testing each individual table row. Therefore, it is always used in conjunction with a function such as [`COUNT\(\)`](#), [`SUM\(\)`](#), [`AVG\(\)`](#), [`MIN\(\)`](#), or [`MAX\(\)`](#), and typically with the `GROUP BY` clause also.

### Restrictions:

The filter expression in the `HAVING` clause cannot include a scalar subquery.

### Related information:

[SELECT Statement](#) on page 218, [GROUP BY Clause](#) on page 229, [Impala Aggregate Functions](#) on page 284

## LIMIT Clause

The `LIMIT` clause in a `SELECT` query sets a maximum number of rows for the result set. Pre-selecting the maximum size of the result set helps Impala to optimize memory usage while processing a distributed query.

**Syntax:**

```
LIMIT constant_integer_expression
```

The argument to the `LIMIT` clause must evaluate to a constant value. It can be a numeric literal, or another kind of numeric expression involving operators, casts, and function return values. You cannot refer to a column or use a subquery.

**Usage notes:**

This clause is useful in contexts such as:

- To return exactly N items from a top-N query, such as the 10 highest-rated items in a shopping category or the 50 hostnames that refer the most traffic to a web site.
- To demonstrate some sample values from a table or a particular query. (To display some arbitrary items, use a query with no `ORDER BY` clause. An `ORDER BY` clause causes additional memory and/or disk usage during the query.)
- To keep queries from returning huge result sets by accident if a table is larger than expected, or a `WHERE` clause matches more rows than expected.

Originally, the value for the `LIMIT` clause had to be a numeric literal. In Impala 1.2.1 and higher, it can be a numeric expression.

Prior to Impala 1.4.0, Impala required any query including an `ORDER BY` clause to also use a `LIMIT` clause. In Impala 1.4.0 and higher, the `LIMIT` clause is optional for `ORDER BY` queries. In cases where sorting a huge result set requires enough memory to exceed the Impala memory limit for a particular node, Impala automatically uses a temporary disk work area to perform the sort operation.

See [ORDER BY Clause](#) on page 225 for details.

In Impala 1.2.1 and higher, you can combine a `LIMIT` clause with an `OFFSET` clause to produce a small result set that is different from a top-N query, for example, to return items 11 through 20. This technique can be used to simulate “paged” results. Because Impala queries typically involve substantial amounts of I/O, use this technique only for compatibility in cases where you cannot rewrite the application logic. For best performance and scalability, wherever practical, query as many items as you expect to need, cache them on the application side, and display small groups of results to users using application logic.

**Restrictions:**

Correlated subqueries used in `EXISTS` and `IN` operators cannot include a `LIMIT` clause.

**Examples:**

The following example shows how the `LIMIT` clause caps the size of the result set, with the limit being applied after any other clauses such as `WHERE`.

```
[localhost:21000] > create database limits;
[localhost:21000] > use limits;
[localhost:21000] > create table numbers (x int);
[localhost:21000] > insert into numbers values (1), (3), (4), (5), (2);
Inserted 5 rows in 1.34s
[localhost:21000] > select x from numbers limit 100;
+---+
| x |
+---+
| 1 |
| 3 |
| 4 |
| 5 |
| 2 |
+---+
Returned 5 row(s) in 0.26s
[localhost:21000] > select x from numbers limit 3;
+---+
| x |
+---+
```

```
| 1 |
| 3 |
| 4 |
+---+
Returned 3 row(s) in 0.27s
[localhost:21000] > select x from numbers where x > 2 limit 2;
+---+
| x |
+---+
| 3 |
| 4 |
+---+
Returned 2 row(s) in 0.27s
```

For top-N and bottom-N queries, you use the `ORDER BY` and `LIMIT` clauses together:

```
[localhost:21000] > select x as "Top 3" from numbers order by x desc limit 3;
+-----+
| top 3 |
+-----+
| 5 |
| 4 |
| 3 |
+-----+
[localhost:21000] > select x as "Bottom 3" from numbers order by x limit 3;
+-----+
| bottom 3 |
+-----+
| 1 |
| 2 |
| 3 |
+-----+
```

You can use constant values besides integer literals as the `LIMIT` argument:

```
-- Other expressions that yield constant integer values work too.
SELECT x FROM t1 LIMIT 1e6;           -- Limit is one million.
SELECT x FROM t1 LIMIT length('hello world');    -- Limit is 11.
SELECT x FROM t1 LIMIT 2+2;           -- Limit is 4.
SELECT x FROM t1 LIMIT cast(truncate(9.9) AS INT); -- Limit is 9.
```

### OFFSET Clause

The `OFFSET` clause in a `SELECT` query causes the result set to start some number of rows after the logical first item. The result set is numbered starting from zero, so `OFFSET 0` produces the same result as leaving out the `OFFSET` clause. Always use this clause in combination with `ORDER BY` (so that it is clear which item should be first, second, and so on) and `LIMIT` (so that the result set covers a bounded range, such as items 0-9, 100-199, and so on).

In Impala 1.2.1 and higher, you can combine a `LIMIT` clause with an `OFFSET` clause to produce a small result set that is different from a top-N query, for example, to return items 11 through 20. This technique can be used to simulate “paged” results. Because Impala queries typically involve substantial amounts of I/O, use this technique only for compatibility in cases where you cannot rewrite the application logic. For best performance and scalability, wherever practical, query as many items as you expect to need, cache them on the application side, and display small groups of results to users using application logic.

#### Examples:

The following example shows how you could run a “paging” query originally written for a traditional database application. Because typical Impala queries process megabytes or gigabytes of data and read large data files from disk each time, it is inefficient to run a separate query to retrieve each small group of items. Use this

technique only for compatibility while porting older applications, then rewrite the application code to use a single query with a large result set, and display pages of results from the cached result set.

```
[localhost:21000] > create table numbers (x int);
[localhost:21000] > insert into numbers select x from very_long_sequence;
Inserted 1000000 rows in 1.34s
[localhost:21000] > select x from numbers order by x limit 5 offset 0;
+---+
| x |
+---+
| 1 |
| 2 |
| 3 |
| 4 |
| 5 |
+---+
[localhost:21000] > select x from numbers order by x limit 5 offset 5;
+---+
| x |
+---+
| 6 |
| 7 |
| 8 |
| 9 |
| 10 |
+---+
```

## UNION Clause

The `UNION` clause lets you combine the result sets of multiple queries. By default, the result sets are combined as if the `DISTINCT` operator was applied.

### Syntax:

```
query_1 UNION [DISTINCT | ALL] query_2
```

### Usage notes:

The `UNION` keyword by itself is the same as `UNION DISTINCT`. Because eliminating duplicates can be a memory-intensive process for a large result set, prefer `UNION ALL` where practical. (That is, when you know the different queries in the union will not produce any duplicates, or where the duplicate values are acceptable.)

When an `ORDER BY` clause applies to a `UNION ALL` or `UNION` query, in Impala 1.4 and higher, the `LIMIT` clause is no longer required. To make the `ORDER BY` and `LIMIT` clauses apply to the entire result set, turn the `UNION` query into a subquery, `SELECT` from the subquery, and put the `ORDER BY` clause at the end, outside the subquery.

### Examples:

First, we set up some sample data, including duplicate 1 values.

```
[localhost:21000] > create table few_ints (x int);
[localhost:21000] > insert into few_ints values (1), (1), (2), (3);
[localhost:21000] > set default_order_by_limit=1000;
```

This example shows how `UNION ALL` returns all rows from both queries, without any additional filtering to eliminate duplicates. For the large result sets common with Impala queries, this is the most memory-efficient technique.

```
[localhost:21000] > select x from few_ints order by x;
+---+
| x |
+---+
| 1 |
| 1 |
| 2 |
| 3 |
+---+
```

## Impala SQL Language Reference

```
Returned 4 row(s) in 0.41s
[localhost:21000] > select x from few_ints union all select x from few_ints;
+---+
| x |
+---+
| 1 |
| 1 |
| 2 |
| 3 |
| 1 |
| 1 |
| 2 |
| 3 |
+---+
Returned 8 row(s) in 0.42s
[localhost:21000] > select * from (select x from few_ints union all select x from
few_ints) as t1 order by x;
+---+
| x |
+---+
| 1 |
| 1 |
| 1 |
| 1 |
| 2 |
| 2 |
| 3 |
| 3 |
+---+
Returned 8 row(s) in 0.53s
[localhost:21000] > select x from few_ints union all select 10;
+---+
| x |
+---+
| 10 |
| 1 |
| 1 |
| 2 |
| 3 |
+---+
Returned 5 row(s) in 0.38s
```

This example shows how the `UNION` clause without the `ALL` keyword condenses the result set to eliminate all duplicate values, making the query take more time and potentially more memory. The extra processing typically makes this technique not recommended for queries that return result sets with millions or billions of values.

```
[localhost:21000] > select x from few_ints union select x+1 from few_ints;
+---+
| x |
+---+
| 3 |
| 4 |
| 1 |
| 2 |
+---+
Returned 4 row(s) in 0.51s
[localhost:21000] > select x from few_ints union select 10;
+---+
| x |
+---+
| 2 |
| 10 |
| 1 |
| 3 |
+---+
Returned 4 row(s) in 0.49s
[localhost:21000] > select * from (select x from few_ints union select x from few_ints)
as t1 order by x;
+---+
| x |
+---+
| 1 |
```

```

| 2 |
| 3 |
+---+
Returned 3 row(s) in 0.53s

```

## Subqueries

A **subquery** is a query that is nested within another query. Subqueries let queries on one table dynamically adapt based on the contents of another table. This technique provides great flexibility and expressive power for SQL queries.

A subquery can return a result set for use in the `FROM` or `WITH` clauses, or with operators such as `IN` or `EXISTS`.

A **scalar subquery** produces a result set with a single row containing a single column, typically produced by an aggregation function such as `MAX()` or `SUM()`. This single result value can be substituted in scalar contexts such as arguments to comparison operators. If the result set is empty, the value of the scalar subquery is `NULL`. For example, the following query finds the maximum value of `t2.y` and then substitutes that value into the `WHERE` clause of the outer block that queries `t1`:

```
SELECT x FROM t1 WHERE x > (SELECT MAX(y) FROM t2);
```

**Uncorrelated subqueries** do not refer to any tables from the outer block of the query. The same value or set of values produced by the subquery is used when evaluating each row from the outer query block. In this example, the subquery returns an arbitrary number of values from `t2.y`, and each value of `t1.x` is tested for membership in that same set of values:

```
SELECT x FROM t1 WHERE x IN (SELECT y FROM t2);
```

**Correlated subqueries** compare one or more values from the outer query block to values referenced in the `WHERE` clause of the subquery. Each row evaluated by the outer `WHERE` clause can be evaluated using a different set of values. These kinds of subqueries are restricted in the kinds of comparisons they can do between columns of the inner and outer tables. (See the following **Restrictions** item.)

For example, the following query finds all the employees with salaries that are higher than average for their department. The subquery potentially computes a different `AVG()` value for each employee.

```
SELECT employee_name, employee_id FROM employees one WHERE
    salary > (SELECT avg(salary) FROM employees two WHERE one.dept_id = two.dept_id);
```

### Syntax:

#### Subquery in the `FROM` clause:

```

SELECT select_list FROM table_ref [, table_ref ...]
table_ref ::= table_name | (select_statement)

```

#### Subqueries in `WHERE` clause:

```

WHERE value comparison_operator (scalar_select_statement)
WHERE value [NOT] IN (select_statement)
WHERE [NOT] EXISTS (correlated_select_statement)
WHERE NOT EXISTS (correlated_select_statement)

```

`comparison_operator` is a numeric comparison such as `=`, `<=`, `!=`, and so on, or a string comparison operator such as `LIKE` or `REGEXP`.

Although you can use non-equality comparison operators such as `<` or `>=`, the subquery must include at least one equality comparison between the columns of the inner and outer query blocks.

All syntax is available for both correlated and uncorrelated queries, except that the `NOT EXISTS` clause cannot be used with an uncorrelated subquery.

## Impala SQL Language Reference

Impala subqueries can be nested arbitrarily deep.

**Standards compliance:** Introduced in [SQL:1999](#).

### Examples:

This example illustrates how subqueries can be used in the `FROM` clause to organize the table names, column names, and column values by producing intermediate result sets, especially for join queries.

```
SELECT avg(t1.x), max(t2.y) FROM
  (SELECT id, cast(a AS DECIMAL(10,5)) AS x FROM raw_data WHERE a BETWEEN 0 AND 100)
AS t1
JOIN
  (SELECT id, length(s) AS y FROM raw_data WHERE s LIKE 'A%') AS t2;
USING (id);
```

These examples show how a query can test for the existence of values in a separate table using the `EXISTS()` operator with a subquery.

The following examples show how a value can be compared against a set of values returned by a subquery.

```
SELECT count(x) FROM t1 WHERE EXISTS(SELECT 1 FROM t2 WHERE t1.x = t2.y * 10);
SELECT x FROM t1 WHERE x IN (SELECT y FROM t2 WHERE state = 'CA');
```

The following examples demonstrate scalar subqueries. When a subquery is known to return a single value, you can substitute it where you would normally put a constant value.

```
SELECT x FROM t1 WHERE y = (SELECT max(z) FROM t2);
SELECT x FROM t1 WHERE y > (SELECT count(z) FROM t2);
```

### Usage notes:

If the same table is referenced in both the outer and inner query blocks, construct a table alias in the outer query block and use a fully qualified name to distinguish the inner and outer table references:

```
SELECT * FROM t1 one WHERE id IN (SELECT parent FROM t1 two WHERE t1.parent = t2.id);
```

### Internal details:

Internally, subqueries involving `IN`, `NOT IN`, `EXISTS`, or `NOT EXISTS` clauses are rewritten into join queries. Depending on the syntax, the subquery might be rewritten to an outer join, semi join, cross join, or anti join.

A query is processed differently depending on whether the subquery calls any aggregation functions. There are correlated and uncorrelated forms, with and without calls to aggregation functions. Each of these four categories is rewritten differently.

### Column statistics considerations:

Because queries that include correlated and uncorrelated subqueries in the `WHERE` clause are written into join queries, to achieve best performance, follow the same guidelines for running the `COMPUTE STATS` statement as you do for tables involved in regular join queries. Run the `COMPUTE STATS` statement for each associated tables after loading or substantially changing the data in that table. See [How Impala Uses Statistics for Query Optimization](#) on page 363 for details.

**Added in:** Subqueries are substantially enhanced starting in Impala 2.0 for CDH 4, and CDH 5.2.0. Now, they can be used in the `WHERE` clause, in combination with clauses such as `EXISTS` and `IN`, rather than just in the `FROM` clause.

### Restrictions:

The initial Impala support for nested subqueries addresses the most common use cases. Some restrictions remain:

- Although you can use subqueries in a query involving `UNION` or `UNION ALL` in Impala 2.1.0 and higher, currently you cannot construct a union of two subqueries (for example, in the argument of an `IN` or `EXISTS` operator).
- Subqueries returning scalar values cannot be used with the operators `ANY` or `ALL`. (Impala does not currently have a `SOME` operator, but if it did, the same restriction would apply.)
- For the `EXISTS` and `NOT EXISTS` clauses, any subquery comparing values from the outer query block to another table must use at least one equality comparison, not exclusively other kinds of comparisons such as less than, greater than, `BETWEEN`, or `!=`.
- Currently, a scalar subquery cannot be used as the first or second argument to the `BETWEEN` operator.
- A subquery cannot be used inside an `OR` conjunction. Expressions inside a subquery, for example in the `WHERE` clause, can use `OR` conjunctions; the restriction only applies to parts of the query "above" the subquery.
- Scalar subqueries are only supported in numeric contexts. You cannot use a scalar subquery as an argument to the `LIKE`, `REGEXP`, or `RLIKE` operators, or compare it to a value of a non-numeric type such as `TIMESTAMP` or `BOOLEAN`.
- You cannot use subqueries with the `CASE` function to generate the comparison value, the values to be compared against, or the return value.
- A subquery is not allowed in the filter condition for the `HAVING` clause. (Strictly speaking, a subquery cannot appear anywhere outside the `WITH`, `FROM`, and `WHERE` clauses.)
- You must use a fully qualified name (`table_name.column_name` or `database_name.table_name.column_name`) when referring to any column from the outer query block within a subquery.

#### Related information:

[EXISTS Operator](#) on page 144, [IN Operator](#) on page 147

## WITH Clause

A clause that can be added before a `SELECT` statement, to define aliases for complicated expressions that are referenced multiple times within the body of the `SELECT`. Similar to `CREATE VIEW`, except that the table and column names defined in the `WITH` clause do not persist after the query finishes, and do not conflict with names used in actual tables or views. Also known as "subquery factoring".

You can rewrite a query using subqueries to work the same as with the `WITH` clause. The purposes of the `WITH` clause are:

- Convenience and ease of maintenance from less repetition with the body of the query. Typically used with queries involving `UNION`, joins, or aggregation functions where the similar complicated expressions are referenced multiple times.
- SQL code that is easier to read and understand by abstracting the most complex part of the query into a separate block.
- Improved compatibility with SQL from other database systems that support the same clause (primarily Oracle Database).

▪ **Note:**

The Impala `WITH` clause does not support recursive queries in the `WITH`, which is supported in some other database systems.

**Standards compliance:** Introduced in [SQL:1999](#).

## Examples:

```
-- Define 2 subqueries that can be referenced from the body of a longer query.  
with t1 as (select 1), t2 as (select 2) insert into tab select * from t1 union all  
select * from t2;  
  
-- Define one subquery at the outer level, and another at the inner level as part of  
the  
-- initial stage of the UNION ALL query.  
with t1 as (select 1) (with t2 as (select 2) select * from t2) union all select * from  
t1;
```

## DISTINCT Operator

The DISTINCT operator in a SELECT statement filters the result set to remove duplicates:

```
-- Returns the unique values from one column.  
-- NULL is included in the set of values if any rows have a NULL in this column.  
select distinct c_birth_country from customer;  
-- Returns the unique combinations of values from multiple columns.  
select distinct c_salutation, c_last_name from customer;
```

You can use DISTINCT in combination with an aggregation function, typically COUNT(), to find how many different values a column contains:

```
-- Counts the unique values from one column.  
-- NULL is not included as a distinct value in the count.  
select count(distinct c_birth_country) from customer;  
-- Counts the unique combinations of values from multiple columns.  
select count(distinct c_salutation, c_last_name) from customer;
```

One construct that Impala SQL does *not* support is using DISTINCT in more than one aggregation function in the same query. For example, you could not have a single query with both COUNT(DISTINCT c\_first\_name) and COUNT(DISTINCT c\_last\_name) in the SELECT list.

**Zero-length strings:** For purposes of clauses such as DISTINCT and GROUP BY, Impala considers zero-length strings (" "), NULL, and space to all be different values.

▪ **Note:**

By default, Impala only allows a single COUNT(DISTINCT *columns*) expression in each query.

If you do not need precise accuracy, you can produce an estimate of the distinct values for a column by specifying NDV(*column*); a query can contain multiple instances of NDV(*column*). To make Impala automatically rewrite COUNT(DISTINCT) expressions to NDV(), enable the APPX\_COUNT\_DISTINCT query option.

To produce the same result as multiple COUNT(DISTINCT) expressions, you can use the following technique for queries involving a single table:

```
select v1.c1 result1, v2.c1 result2 from  
(select count(distinct col1) as c1 from t1) v1  
cross join  
(select count(distinct col2) as c1 from t1) v2;
```

Because CROSS JOIN is an expensive operation, prefer to use the NDV() technique wherever practical.

▪ **Note:**

In contrast with some database systems that always return DISTINCT values in sorted order, Impala does not do any ordering of DISTINCT values. Always include an ORDER BY clause if you need the values in alphabetical or numeric sorted order.

## Hints

The Impala SQL dialect supports query hints, for fine-tuning the inner workings of queries. Specify hints as a temporary workaround for expensive queries, where missing statistics or other factors cause inefficient performance.

Hints are most often used for the most resource-intensive kinds of Impala queries:

- Join queries involving large tables, where intermediate result sets are transmitted across the network to evaluate the join conditions.
- Inserting into partitioned Parquet tables, where many memory buffers could be allocated on each host to hold intermediate results for each partition.

### Syntax:

You can represent the hints as keywords surrounded by [ ] square brackets; include the brackets in the text of the SQL statement.

```
SELECT STRAIGHT_JOIN select_list FROM
join_left_hand_table
  JOIN [ {BROADCAST|SHUFFLE} ]
join_right_hand_table
remainder_of_query;

INSERT insert_clauses
[ {SHUFFLE|NOSHUFFLE} ]
  SELECT remainder_of_query;
```

In Impala 2.0 and higher, or CDH 5.2 and higher, you can also specify the hints inside comments that use either the /\* \*/ or -- notation. Specify a + symbol immediately before the hint name.

```
SELECT STRAIGHT_JOIN select_list FROM
join_left_hand_table
  JOIN /* +BROADCAST|SHUFFLE */
join_right_hand_table
remainder_of_query;

SELECT select_list FROM
join_left_hand_table
  JOIN -- +BROADCAST|SHUFFLE
join_right_hand_table
remainder_of_query;

INSERT insert_clauses
/* +SHUFFLE|NOSHUFFLE */
  SELECT remainder_of_query;

INSERT insert_clauses
-- +SHUFFLE|NOSHUFFLE
  SELECT remainder_of_query;
```

### Usage notes:

With both forms of hint syntax, include the STRAIGHT\_JOIN keyword immediately after the SELECT keyword to prevent Impala from reordering the tables in a way that makes the hint ineffective.

To reduce the need to use hints, run the COMPUTE STATS statement against all tables involved in joins, or used as the source tables for INSERT ... SELECT operations where the destination is a partitioned Parquet table. Do this operation after loading data or making substantial changes to the data within each table. Having up-to-date statistics helps Impala choose more efficient query plans without the need for hinting. See [How Impala Uses Statistics for Query Optimization](#) on page 363 for details and examples.

To see which join strategy is used for a particular query, examine the EXPLAIN output for that query. See [Using the EXPLAIN Plan for Performance Tuning](#) on page 379 for details and examples.

### Hints for join queries:

## Impala SQL Language Reference

The `[BROADCAST]` and `[SHUFFLE]` hints control the execution strategy for join queries. Specify one of the following constructs immediately after the `JOIN` keyword in a query:

- `[SHUFFLE]` - Makes that join operation use the “partitioned” technique, which divides up corresponding rows from both tables using a hashing algorithm, sending subsets of the rows to other nodes for processing. (The keyword `SHUFFLE` is used to indicate a “partitioned join”, because that type of join is not related to “partitioned tables”.) Since the alternative “broadcast” join mechanism is the default when table and index statistics are unavailable, you might use this hint for queries where broadcast joins are unsuitable; typically, partitioned joins are more efficient for joins between large tables of similar size.
- `[BROADCAST]` - Makes that join operation use the “broadcast” technique that sends the entire contents of the right-hand table to all nodes involved in processing the join. This is the default mode of operation when table and index statistics are unavailable, so you would typically only need it if stale metadata caused Impala to mistakenly choose a partitioned join operation. Typically, broadcast joins are more efficient in cases where one table is much smaller than the other. (Put the smaller table on the right side of the `JOIN` operator.)

### Hints for `INSERT ... SELECT` queries:

When inserting into partitioned tables, especially using the Parquet file format, you can include a hint in the `INSERT` statement to fine-tune the overall performance of the operation and its resource usage:

- These hints are available in Impala 1.2.2 and higher.
- You would only use these hints if an `INSERT` into a partitioned Parquet table was failing due to capacity limits, or if such an `INSERT` was succeeding but with less-than-optimal performance.
- To use these hints, put the hint keyword `[SHUFFLE]` or `[NOSHUFFLE]` (including the square brackets) after the `PARTITION` clause, immediately before the `SELECT` keyword.
- `[SHUFFLE]` selects an execution plan that minimizes the number of files being written simultaneously to HDFS, and the number of memory buffers holding data for individual partitions. Thus it reduces overall resource usage for the `INSERT` operation, allowing some `INSERT` operations to succeed that otherwise would fail. It does involve some data transfer between the nodes so that the data files for a particular partition are all constructed on the same node.
- `[NOSHUFFLE]` selects an execution plan that might be faster overall, but might also produce a larger number of small data files or exceed capacity limits, causing the `INSERT` operation to fail. Use `[SHUFFLE]` in cases where an `INSERT` statement fails or runs inefficiently due to all nodes attempting to construct data for all partitions.
- Impala automatically uses the `[SHUFFLE]` method if any partition key column in the source table, mentioned in the `INSERT ... SELECT` query, does not have column statistics. In this case, only the `[NOSHUFFLE]` hint would have any effect.
- If column statistics are available for all partition key columns in the source table mentioned in the `INSERT ... SELECT` query, Impala chooses whether to use the `[SHUFFLE]` or `[NOSHUFFLE]` technique based on the estimated number of distinct values in those columns and the number of nodes involved in the `INSERT` operation. In this case, you might need the `[SHUFFLE]` or the `[NOSHUFFLE]` hint to override the execution plan selected by Impala.

### Suggestions versus directives:

In early Impala releases, hints were always obeyed and so acted more like directives. Once Impala gained join order optimizations, sometimes join queries were automatically reordered in a way that made a hint irrelevant. Therefore, the hints act more like suggestions in Impala 1.2.2 and higher.

To force Impala to follow the hinted execution mechanism for a join query, include the `STRAIGHT_JOIN` keyword in the `SELECT` statement. See [Overriding Join Reordering with `STRAIGHT\_JOIN`](#) on page 358 for details. When you use this technique, Impala does not reorder the joined tables at all, so you must be careful to arrange the join order to put the largest table (or subquery result set) first, then the smallest, second smallest, third smallest, and so on. This ordering lets Impala do the most I/O-intensive parts of the query using local reads on the data nodes, and then reduce the size of the intermediate result set as much as possible as each subsequent table or subquery result set is joined.

### Restrictions:

Queries that include subqueries in the WHERE clause can be rewritten internally as join queries. Currently, you cannot apply hints to the joins produced by these types of queries.

Because hints can prevent queries from taking advantage of new metadata or improvements in query planning, use them only when required to work around performance issues, and be prepared to remove them when they are no longer required, such as after a new Impala release or bug fix.

In particular, the [BROADCAST] and [SHUFFLE] hints are expected to be needed much less frequently in Impala 1.2.2 and higher, because the join order optimization feature in combination with the COMPUTE STATS statement now automatically choose join order and join mechanism without the need to rewrite the query and add hints. See [Performance Considerations for Join Queries](#) on page 357 for details.

#### Compatibility:

The hints embedded within -- comments are compatible with Hive queries. The hints embedded within /\* \*/ comments or [ ] square brackets are not recognized by or not compatible with Hive. For example, Hive raises an error for Impala hints within /\* \*/ comments because it does not recognize the Impala hint names.

#### Considerations for views:

If you use a hint in the query that defines a view, the hint is preserved when you query the view. Impala internally rewrites all hints in views to use the -- comment notation, so that Hive can query such views without errors due to unrecognized hint names.

#### Examples:

For example, this query joins a large customer table with a small lookup table of less than 100 rows. The right-hand table can be broadcast efficiently to all nodes involved in the join. Thus, you would use the [broadcast] hint to force a broadcast join strategy:

```
select straight_join customer.address, state_lookup.state_name
  from customer join [broadcast] state_lookup
    on customer.state_id = state_lookup.state_id;
```

This query joins two large tables of unpredictable size. You might benchmark the query with both kinds of hints and find that it is more efficient to transmit portions of each table to other nodes for processing. Thus, you would use the [shuffle] hint to force a partitioned join strategy:

```
select straight_join weather.wind_velocity, geospatial.altitude
  from weather join [shuffle] geospatial
    on weather.lat = geospatial.lat and weather.long = geospatial.long;
```

For joins involving three or more tables, the hint applies to the tables on either side of that specific JOIN keyword. The STRAIGHT\_JOIN keyword ensures that joins are processed in a predictable order from left to right. For example, this query joins t1 and t2 using a partitioned join, then joins that result set to t3 using a broadcast join:

```
select straight_join t1.name, t2.id, t3.price
  from t1 join [shuffle] t2 join [broadcast] t3
    on t1.id = t2.id and t2.id = t3.id;
```

#### Related information:

For more background information about join queries, see [Joins](#) on page 220. For performance considerations, see [Performance Considerations for Join Queries](#) on page 357.

## SET Statement

Specifies values for query options that control the runtime behavior of other statements within the same session.

#### Syntax:

```
SET [query_option(option_value)]
```

## Impala SQL Language Reference

SET with no arguments returns a result set consisting of all available query options and their current values.

The query option name and any string argument values are case-insensitive.

Each query option has a specific allowed notation for its arguments. Boolean options can be enabled and disabled by assigning values of either `true` and `false`, or `1` and `0`. Some numeric options accept a final character signifying the unit, such as `2g` for 2 gigabytes or `100m` for 100 megabytes. See [Query Options for the SET Statement](#) on page 339 for the details of each query option.

### Usage notes:

`MEM_LIMIT` is probably the most commonly used query option. You can specify a high value to allow a resource-intensive query to complete. For testing how queries would work on memory-constrained systems, you might specify an artificially low value.

### Examples:

The following example sets some numeric and some Boolean query options to control usage of memory, disk space, and timeout periods, then runs a query whose success could depend on the options in effect:

```
set mem_limit=64g;
set DISABLE_UNSAFE_SPILLS=true;
set parquet_file_size=400m;
set RESERVATION_REQUEST_TIMEOUT=900000;
insert overwrite parquet_table select c1, c2, count(c3) from text_table group by c1,
c2, c3;
```

**Added in:** CDH 5.2.0 (Impala 2.0.0)

SET has always been available as an `impala-shell` command. Promoting it to a SQL statement lets you use this feature in client applications through the JDBC and ODBC APIs.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch and HDFS files or directories, therefore no HDFS permissions are required.

### Related information:

See [Query Options for the SET Statement](#) on page 339 for the query options you can adjust using this statement.

## SHOW Statement

The SHOW statement is a flexible way to get information about different types of Impala objects.

### Syntax:

```
SHOW DATABASES [[LIKE] 'pattern']
SHOW SCHEMAS [[LIKE] 'pattern'] - an alias for SHOW DATABASES
SHOW TABLES [IN database_name] [[LIKE] 'pattern']
SHOW [AGGREGATE | ANALYTIC] FUNCTIONS [IN database_name] [[LIKE] 'pattern']
SHOW CREATE TABLE [database_name].table_name
SHOW TABLE STATS [database_name.]table_name
SHOW COLUMN STATS [database_name.]table_name
SHOW PARTITIONS [database_name.]table_name
SHOW FILES IN [database_name.]table_name [PARTITION (key_col=value [, key_col=value])]
SHOW ROLES
SHOW CURRENT ROLES
SHOW ROLE GRANT GROUP group_name
SHOW GRANT ROLE role_name
```

Issue a `SHOW object_type` statement to see the appropriate objects in the current database, or `SHOW object_type IN database_name` to see objects in a specific database.

The optional *pattern* argument is a quoted string literal, using Unix-style \* wildcards and allowing | for alternation. The preceding `LIKE` keyword is also optional. All object names are stored in lowercase, so use all lowercase letters in the pattern string. For example:

```
show databases 'a*';
show databases like 'a*';
show tables in some_db like '*fact*';
use some_db;
show tables '*dim*|*fact*';
```

**Cancellation:** Cannot be cancelled.

## SHOW FILES Statement

The `SHOW FILES` statement displays the files that constitute a specified table, or a partition within a partitioned table. This syntax is available in CDH 5.4 and higher only. The output includes the names of the files, the size of each file, and the applicable partition for a partitioned table. The size includes a suffix of `B` for bytes, `MB` for megabytes, and `GB` for gigabytes.

- **Note:** This statement applies to tables and partitions stored on HDFS, or in the Amazon Simple Storage System (S3). It does not apply to views. It does not apply to tables mapped onto HBase, because HBase does not use the same file-based storage layout.

### Usage notes:

You can use this statement to verify the results of your ETL process: that is, that the expected files are present, with the expected sizes. You can examine the file information to detect conditions such as empty files, missing files, or inefficient layouts due to a large number of small files. When you use `INSERT` statements to copy from one table to another, you can see how the file layout changes due to file format conversions, compaction of small input files into large data blocks, and multiple output files from parallel queries and partitioned inserts.

The output from this statement does not include files that Impala considers to be hidden or invisible, such as those whose names start with a dot or an underscore, or that end with the suffixes `.copying` or `.tmp`.

The information for partitioned tables complements the output of the `SHOW PARTITIONS` statement, which summarizes information about each partition. `SHOW PARTITIONS` produces some output for each partition, while `SHOW FILES` does not produce any output for empty partitions because they do not include any data files.

### HDFS permissions:

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read permission for all the table files, read and execute permission for all the directories that make up the table, and execute permission for the database directory and all its parent directories.

### Examples:

The following example constructs `SHOW FILES` statements for an unpartitioned tables using text format:

```
[localhost:21000] > create table unpartitioned_text (x bigint, s string);
[localhost:21000] > insert into unpartitioned_text (x, s) select id, name from
oreilly.sample_data limit 20e6;
[localhost:21000] > show files in unpartitioned_text;
+-----+
| path
| size      | partition |
+-----+
| hdfs://impala_data_dir/show_files.db/unpartitioned_text/35665776ef85cfaf_1012432410_data.0.
| 448.31MB   |           |
+-----+
[localhost:21000] > insert into unpartitioned_text (x, s) select id, name from
oreilly.sample_data limit 100e6;
[localhost:21000] > show files in unpartitioned_text;
+-----+
| path
```

	size	partition
hdfs://impala_data_dir/show_files.db/unpartitioned_text/35665776ef85cfaf_1012432410_data.0.	448.31MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_text/ac3dba252a8952b8_1663177415_data.0.	2.19GB	

This example illustrates how, after issuing some `INSERT ... VALUES` statements, the table now contains some tiny files of just a few bytes. Such small files could cause inefficient processing of parallel queries that are expecting multi-megabyte input files. The example shows how you might compact the small files by doing an `INSERT ... SELECT` into a different table, possibly converting the data to Parquet in the process:

[localhost:21000] > insert into unpartitioned_text values (10,'hello'), (20, 'world');		
[localhost:21000] > insert into unpartitioned_text values (-1,'foo'), (-1000, 'bar');		
[localhost:21000] > show files in unpartitioned_text;		
path	size	partition
hdfs://impala_data_dir/show_files.db/unpartitioned_text/4f11b8bdf8b6aa92_238145083_data.0.	18B	
hdfs://impala_data_dir/show_files.db/unpartitioned_text/35665776ef85cfaf_1012432410_data.0.	448.31MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_text/ac3dba252a8952b8_1663177415_data.0.	2.19GB	
hdfs://impala_data_dir/show_files.db/unpartitioned_text/cfb8252452445682_1868457216_data.0.	17B	
[localhost:21000] > create table unpartitioned_parquet stored as parquet as select * from unpartitioned_text;		
summary		
Inserted 120000002 row(s)		
[localhost:21000] > show files in unpartitioned_parquet;		
path	size	partition
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630184_549959007_data.0.parq	255.36MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630184_549959007_data.1.parq	178.52MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630185_549959007_data.0.parq	255.37MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630185_549959007_data.1.parq	57.71MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630186_2141167244_data.0.parq	255.40MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630186_2141167244_data.1.parq	175.52MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630187_1006832086_data.0.parq	255.40MB	
hdfs://impala_data_dir/show_files.db/unpartitioned_parquet/60798d96ba630187_1006832086_data.1.parq		

214.61MB		
----------	--	--

The following example shows a `SHOW FILES` statement for a partitioned text table with data in two different partitions, and two empty partitions. The partitions with no data are not represented in the `SHOW FILES` output.

```
[localhost:21000] > create table partitioned_text (x bigint, y int, s string) partitioned
  by (year bigint, month bigint, day bigint);
[localhost:21000] > insert overwrite partitioned_text (x, y, s) partition
  (year=2014,month=1,day=1) select id, val, name from oreilly.normalized_parquet
  where id between 1 and 1000000;
[localhost:21000] > insert overwrite partitioned_text (x, y, s) partition
  (year=2014,month=1,day=2) select id, val, name from oreilly.normalized_parquet
  where id between 1000001 and 2000000;
[localhost:21000] > alter table partitioned_text add partition (year=2014,month=1,day=3);
[localhost:21000] > alter table partitioned_text add partition (year=2014,month=1,day=4);
[localhost:21000] > show partitions partitioned_text;
```

Format	year	month	day	#Rows	#Files	Size	Bytes Cached	Cache Replication
	Incremental							
TEXT	2014	1	1	-1	4	25.16MB	NOT CACHED	NOT CACHED
TEXT	2014	1	2	-1	4	26.22MB	NOT CACHED	NOT CACHED
TEXT	2014	1	3	-1	0	0B	NOT CACHED	NOT CACHED
TEXT	2014	1	4	-1	0	0B	NOT CACHED	NOT CACHED
Total				-1	8	51.38MB	0B	

```
[localhost:21000] > show files in partitioned_text;
```

path	size	partition
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=1/80732d9dc80689f_1418645991_data.0	5.77MB	year=2014/month=1/day=1
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=1/80732d9dc8068a0_1418645991_data.0	6.25MB	year=2014/month=1/day=1
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=1/80732d9dc8068a1_147082319_data.0	7.16MB	year=2014/month=1/day=1
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=1/80732d9dc8068a2_2111411753_data.0	5.98MB	year=2014/month=1/day=1
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=2/21a828cf494b5bbb_501271652_data.0	6.42MB	year=2014/month=1/day=2
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=2/21a828cf494b5bbc_501271652_data.0	6.62MB	year=2014/month=1/day=2
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=2/21a828cf494b5bbd_1393490200_data.0	6.98MB	year=2014/month=1/day=2
hdfs://impala_data_dir/show_files.db/partitioned_text/year=2014/month=1/day=2/21a828cf494b5bbe_1393490200_data.0	6.20MB	year=2014/month=1/day=2

The following example shows a `SHOW FILES` statement for a partitioned Parquet table. The number and sizes of files are different from the equivalent partitioned text table used in the previous example, because `INSERT`

## Impala SQL Language Reference

operations for Parquet tables are parallelized differently than for text tables. (Also, the amount of data is so small that it can be written to Parquet without involving all the hosts in this 4-node cluster.)

```
[localhost:21000] > create table partitioned_parquet (x bigint, y int, s string)
partitioned by (year bigint, month bigint, day bigint) stored as parquet;
[localhost:21000] > insert into partitioned_parquet partition (year,month,day) select
  x, y, s, year, month, day from partitioned_text;
[localhost:21000] > show partitions partitioned_parquet;
+-----+-----+-----+-----+-----+-----+-----+-----+
| year | month | day | #Rows | #Files | Size      | Bytes Cached | Cache Replication |
| Format| Incremental stats |                         |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 2014  | 1     | 1   | -1    | 3      | 17.89MB | NOT CACHED | NOT CACHED      |
| PARQUET| false          |                         |
| 2014  | 1     | 2   | -1    | 3      | 17.89MB | NOT CACHED | NOT CACHED      |
| PARQUET| false          |                         |
| Total  |        |     | -1    | 6      | 35.79MB | 0B           |                  |
+-----+-----+-----+-----+-----+-----+-----+-----+
[localhost:21000] > show files in partitioned_parquet;
+-----+-----+-----+
| path                         | size   | partition |
+-----+-----+-----+
| hdfs://impala_data_dir/show_files.db/partitioned_parquet/year=2014/month=1/day=1/1134113650_data.0.parq
| 4.49MB | year=2014/month=1/day=1 |
|
| hdfs://impala_data_dir/show_files.db/partitioned_parquet/year=2014/month=1/day=1/617567880_data.0.parq
| 5.14MB | year=2014/month=1/day=1 |
|
| hdfs://impala_data_dir/show_files.db/partitioned_parquet/year=2014/month=1/day=1/2099499416_data.0.parq
| 8.27MB | year=2014/month=1/day=1 |
|
| hdfs://impala_data_dir/show_files.db/partitioned_parquet/year=2014/month=1/day=2/945567189_data.0.parq
| 8.80MB | year=2014/month=1/day=2 |
|
| hdfs://impala_data_dir/show_files.db/partitioned_parquet/year=2014/month=1/day=2/2145850112_data.0.parq
| 4.80MB | year=2014/month=1/day=2 |
|
| hdfs://impala_data_dir/show_files.db/partitioned_parquet/year=2014/month=1/day=2/665613448_data.0.parq
| 4.29MB | year=2014/month=1/day=2 |
+-----+-----+-----+
```

The following example shows output from the `SHOW FILES` statement for a table where the data files are stored in Amazon S3:

```
[localhost:21000] > show files in s3_testing.sample_data_s3;
+-----+-----+-----+
| path                         | size   |
| partition |
+-----+-----+-----+
| s3a://impala-demo/sample_data/e065453cba1988a6_1733868553_data.0.parq | 24.84MB |
+-----+-----+
```

## SHOW ROLES Statement

The `SHOW ROLES` statement displays roles. This syntax is available in CDH 5.2 and later only, when you are using the Sentry authorization framework along with the Sentry service, as described in [Using Impala with the Sentry Service \(CDH 5.1 or higher only\)](#) on page 93. It does not apply when you use the Sentry framework with privileges defined in a policy file.

### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe

an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

#### Examples:

Depending on the roles set up within your organization by the `CREATE ROLE` statement, the output might look something like this:

```
show roles;
+-----+
| role_name |
+-----+
| analyst    |
| role1      |
| sales       |
| superuser   |
| test_role   |
+-----+
```

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

#### Related information:

[Enabling Sentry Authorization for Impala](#) on page 91

## SHOW CURRENT ROLE

The `SHOW CURRENT ROLE` statement displays roles assigned to the current user. This syntax is available in CDH 5.2 and later only, when you are using the Sentry authorization framework along with the Sentry service, as described in [Using Impala with the Sentry Service \(CDH 5.1 or higher only\)](#) on page 93. It does not apply when you use the Sentry framework with privileges defined in a policy file.

#### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

#### Examples:

Depending on the roles set up within your organization by the `CREATE ROLE` statement, the output might look something like this:

```
show current roles;
+-----+
| role_name |
+-----+
| role1    |
| superuser |
+-----+
```

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

#### Related information:

[Enabling Sentry Authorization for Impala](#) on page 91

## SHOW ROLE GRANT Statement

The `SHOW ROLE GRANT` statement lists all the roles assigned to the specified group. This statement is only allowed for Sentry administrative users and others users that are part of the specified group. This syntax is

## Impala SQL Language Reference

available in CDH 5.2 and later only, when you are using the Sentry authorization framework along with the Sentry service, as described in [Using Impala with the Sentry Service \(CDH 5.1 or higher only\)](#) on page 93. It does not apply when you use the Sentry framework with privileges defined in a policy file.

### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

**HDFS permissions:** This statement does not touch and HDFS files or directories, therefore no HDFS permissions are required.

### Related information:

[Enabling Sentry Authorization for Impala](#) on page 91

## SHOW GRANT ROLE Statement

The `SHOW GRANT ROLE` statement lists all the grants for the given role name. This statement is only allowed for Sentry administrative users and other users that have been granted the specified role. This syntax is available in CDH 5.2 and later only, when you are using the Sentry authorization framework along with the Sentry service, as described in [Using Impala with the Sentry Service \(CDH 5.1 or higher only\)](#) on page 93. It does not apply when you use the Sentry framework with privileges defined in a policy file.

### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

**HDFS permissions:** This statement does not touch and HDFS files or directories, therefore no HDFS permissions are required.

### Related information:

[Enabling Sentry Authorization for Impala](#) on page 91

## SHOW DATABASES

The `SHOW DATABASES` statement is often the first one you issue when connecting to an instance for the first time. You typically issue `SHOW DATABASES` to see the names you can specify in a `USE db_name` statement, then after switching to a database you issue `SHOW TABLES` to see the names you can specify in `SELECT` and `INSERT` statements.

The output of `SHOW DATABASES` includes the special `_impala_builtin` database, which lets you view definitions of built-in functions, as described under `SHOW FUNCTIONS`.

### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

### Examples:

This example shows how you might locate a particular table on an unfamiliar system. The `DEFAULT` database is the one you initially connect to; a database with that name is present on every system. You can issue `SHOW TABLES IN db_name` without going into a database, or `SHOW TABLES` once you are inside a particular database.

```
[localhost:21000] > show databases;
+-----+
| name |
+-----+
| _impala_builtins |
| analyze_testing |
| avro |
| ctas |
| d1 |
| d2 |
| d3 |
| default |
| file_formats |
| hbase |
| load_data |
| partitioning |
| regexp_testing |
| reports |
| temporary |
+-----+
Returned 14 row(s) in 0.02s
[localhost:21000] > show tables in file_formats;
+-----+
| name |
+-----+
| parquet_table |
| rcfile_table |
| sequencefile_table |
| textfile_table |
+-----+
Returned 4 row(s) in 0.01s
[localhost:21000] > use file_formats;
[localhost:21000] > show tables like '*parq*';
+-----+
| name |
+-----+
| parquet_table |
+-----+
Returned 1 row(s) in 0.01s
```

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

#### Related information:

[Databases](#) on page 155, [CREATE DATABASE Statement](#) on page 175, [DROP DATABASE Statement](#) on page 192, [USE Statement](#) on page 256 [SHOW TABLES Statement](#) on page 249, [SHOW FUNCTIONS Statement](#) on page 256

## SHOW TABLES Statement

Displays the names of tables. By default, lists tables in the current database, or with the `IN` clause, in a specified database. By default, lists all tables, or with the `LIKE` clause, only those whose name match a pattern with \* wildcards.

#### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other databases, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read and execute permissions for all directories that are part of the table. (A table could span multiple different HDFS directories)

if it is partitioned. The directories could be widely scattered because a partition can reside in an arbitrary HDFS directory based on its `LOCATION` attribute.)

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

### Related information:

[Tables](#) on page 156, [CREATE TABLE Statement](#) on page 180, [ALTER TABLE Statement](#) on page 164, [DROP TABLE Statement](#) on page 197, [DESCRIBE Statement](#) on page 188, [SHOW CREATE TABLE Statement](#) on page 250, [SHOW TABLE STATS Statement](#) on page 251, [SHOW DATABASES](#) on page 248, [SHOW FUNCTIONS Statement](#) on page 256

## SHOW CREATE TABLE Statement

As a schema changes over time, you might run a `CREATE TABLE` statement followed by several `ALTER TABLE` statements. To capture the cumulative effect of all those statements, `SHOW CREATE TABLE` displays a `CREATE TABLE` statement that would reproduce the current structure of a table. You can use this output in scripts that set up or clone a group of tables, rather than trying to reproduce the original sequence of `CREATE TABLE` and `ALTER TABLE` statements. When creating variations on the original table, or cloning the original table on a different system, you might need to edit the `SHOW CREATE TABLE` output to change things such as the database name, `LOCATION` field, and so on that might be different on the destination system.

### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other databases, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

### Examples:

The following example shows how various clauses from the `CREATE TABLE` statement are represented in the output of `SHOW CREATE TABLE`.

```
create table show_create_table_demo (id int comment "Unique ID", y double, s string)
  partitioned by (year smallint)
  stored as parquet;

show create table show_create_table_demo;
+-----+
| result
|   |
| +-----+
| | CREATE TABLE scratch.show_create_table_demo (
| |   id INT COMMENT 'Unique ID',
| |   y DOUBLE,
| |   s STRING
| | )
| | PARTITIONED BY (
| |   year SMALLINT
| | )
| | STORED AS PARQUET
| | LOCATION 'hdfs://127.0.0.1:8020/user/hive/warehouse/scratch.db/show_create_table_demo'
| +-----+
```

```
+-----+
| TBLPROPERTIES ('transient_lastDdlTime='1418152582')
| |
+-----+
```

The following example shows how, after a sequence of `ALTER TABLE` statements, the output from `SHOW CREATE TABLE` represents the current state of the table. This output could be used to create a matching table rather than executing the original `CREATE TABLE` and sequence of `ALTER TABLE` statements.

```
+-----+
alter table show_create_table_demo drop column s;
alter table show_create_table_demo set fileformat textfile;

show create table show_create_table_demo;
+-----+
| result
| |
+-----+
| CREATE TABLE scratch.show_create_table_demo (
|   id INT COMMENT 'Unique ID',
|   y DOUBLE
| )
| PARTITIONED BY (
|   year SMALLINT
| )
| STORED AS TEXTFILE
| LOCATION 'hdfs://127.0.0.1:8020/user/hive/warehouse/demo.db/show_create_table_demo'
| TBLPROPERTIES ('transient_lastDdlTime='1418152638')
+-----+
```

#### Related information:

[CREATE TABLE Statement](#) on page 180, [DESCRIBE Statement](#) on page 188, [SHOW TABLES Statement](#) on page 249

## SHOW TABLE STATS Statement

The `SHOW TABLE STATS` and `SHOW COLUMN STATS` variants are important for tuning performance and diagnosing performance issues, especially with the largest tables and the most complex join queries.

Any values that are not available (because the `COMPUTE STATS` statement has not been run yet) are displayed as -1.

`SHOW TABLE STATS` provides some general information about the table, such as the number of files, overall size of the data, whether some or all of the data is in the HDFS cache, and the file format, that is useful whether or not you have run the `COMPUTE STATS` statement. A -1 in the `#Rows` output column indicates that the `COMPUTE STATS` statement has never been run for this table. If the table is partitioned, `SHOW TABLE STATS` provides this information for each partition. (It produces the same output as the `SHOW PARTITIONS` statement in this case.)

The output of `SHOW COLUMN STATS` is primarily only useful after the `COMPUTE STATS` statement has been run on the table. A -1 in the `#Distinct Values` output column indicates that the `COMPUTE STATS` statement has never been run for this table. Currently, Impala always leaves the `#Nulls` column as -1, even after `COMPUTE STATS` has been run.

These `SHOW` statements work on actual tables only, not on views.

#### Security considerations:

## Impala SQL Language Reference

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

### Examples:

The following examples show how the `SHOW TABLE STATS` statement displays physical information about a table and the associated data files:

```
show table stats store_sales;
+-----+-----+-----+-----+-----+-----+
| #Rows | #Files | Size      | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| -1    | 1      | 370.45MB | NOT CACHED   | TEXT   | false            |
+-----+-----+-----+-----+-----+-----+

show table stats customer;
+-----+-----+-----+-----+-----+-----+
| #Rows | #Files | Size      | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| -1    | 1      | 12.60MB  | NOT CACHED   | TEXT   | false            |
+-----+-----+-----+-----+-----+-----+
```

The following example shows how, after a `COMPUTE STATS` or `COMPUTE INCREMENTAL STATS` statement, the `#Rows` field is now filled in. Because the `STORE_SALES` table in this example is not partitioned, the `COMPUTE INCREMENTAL STATS` statement produces regular stats rather than incremental stats, therefore the `Incremental stats` field remains false.

```
compute stats customer;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 18 column(s). |
+-----+-----+-----+-----+-----+-----+
| #Rows | #Files | Size      | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| 100000 | 1      | 12.60MB  | NOT CACHED   | TEXT   | false            |
+-----+-----+-----+-----+-----+-----+

show table stats customer;
+-----+-----+-----+-----+-----+-----+
| #Rows | #Files | Size      | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| 100000 | 1      | 12.60MB  | NOT CACHED   | TEXT   | false            |
+-----+-----+-----+-----+-----+-----+

compute incremental stats store_sales;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 23 column(s). |
+-----+-----+-----+-----+-----+-----+
| #Rows | #Files | Size      | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| 2880404 | 1      | 370.45MB | NOT CACHED   | TEXT   | false            |
+-----+-----+-----+-----+-----+-----+
```

### HDFS permissions:

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read and execute permissions for all directories that are part of the table. (A table could span multiple different HDFS directories if it is partitioned. The directories could be widely scattered because a partition can reside in an arbitrary HDFS directory based on its `LOCATION` attribute.) The `Impala` user must also have execute permission for the database directory, and any parent directories of the database directory in HDFS.

### Related information:

[COMPUTE STATS Statement](#) on page 170, [SHOW COLUMN STATS Statement](#) on page 253

See [How Impala Uses Statistics for Query Optimization](#) on page 363 for usage information and examples.

## SHOW COLUMN STATS Statement

The SHOW TABLE STATS and SHOW COLUMN STATS variants are important for tuning performance and diagnosing performance issues, especially with the largest tables and the most complex join queries.

### Security considerations:

When authorization is enabled, the output of the SHOW statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the SHOW output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

### Examples:

The following examples show the output of the SHOW COLUMN STATS statement for some tables, before the COMPUTE STATS statement is run. Impala deduces some information, such as maximum and average size for fixed-length columns, and leaves and unknown values as -1.

show column stats customer;						
Column	Type	#Distinct Values	#Nulls	Max Size	Avg Size	
c_customer_sk	INT	-1	-1	4	4	
c_customer_id	STRING	-1	-1	-1	-1	
c_current_cdemo_sk	INT	-1	-1	4	4	
c_current_hdemo_sk	INT	-1	-1	4	4	
c_current_addr_sk	INT	-1	-1	4	4	
c_first_shipto_date_sk	INT	-1	-1	4	4	
c_first_sales_date_sk	INT	-1	-1	4	4	
c_salutation	STRING	-1	-1	-1	-1	
c_first_name	STRING	-1	-1	-1	-1	
c_last_name	STRING	-1	-1	-1	-1	
c_preferred_cust_flag	STRING	-1	-1	-1	-1	
c_birth_day	INT	-1	-1	4	4	
c_birth_month	INT	-1	-1	4	4	
c_birth_year	INT	-1	-1	4	4	
c_birth_country	STRING	-1	-1	-1	-1	
c_login	STRING	-1	-1	-1	-1	
c_email_address	STRING	-1	-1	-1	-1	
c_last_review_date	STRING	-1	-1	-1	-1	

show column stats store_sales;						
Column	Type	#Distinct Values	#Nulls	Max Size	Avg Size	
ss_sold_date_sk	INT	-1	-1	4	4	
ss_sold_time_sk	INT	-1	-1	4	4	
ss_item_sk	INT	-1	-1	4	4	
ss_customer_sk	INT	-1	-1	4	4	
ss_cdemo_sk	INT	-1	-1	4	4	
ss_hdemo_sk	INT	-1	-1	4	4	
ss_addr_sk	INT	-1	-1	4	4	
ss_store_sk	INT	-1	-1	4	4	
ss_promo_sk	INT	-1	-1	4	4	
ss_ticket_number	INT	-1	-1	4	4	
ss_quantity	INT	-1	-1	4	4	
ss_wholesale_cost	FLOAT	-1	-1	4	4	
ss_list_price	FLOAT	-1	-1	4	4	
ss_sales_price	FLOAT	-1	-1	4	4	
ss_ext_discount_amt	FLOAT	-1	-1	4	4	
ss_ext_sales_price	FLOAT	-1	-1	4	4	
ss_ext_wholesale_cost	FLOAT	-1	-1	4	4	
ss_ext_list_price	FLOAT	-1	-1	4	4	
ss_ext_tax	FLOAT	-1	-1	4	4	

ss_coupon_amt	FLOAT	-1	-1	4	4
ss_net_paid	FLOAT	-1	-1	4	4
ss_net_paid_inc_tax	FLOAT	-1	-1	4	4
ss_net_profit	FLOAT	-1	-1	4	4

The following examples show the output of the `SHOW COLUMN STATS` statement for some tables, after the `COMPUTE STATS` statement is run. Now most of the -1 values are changed to reflect the actual table data. The `#Nulls` column remains -1 because Impala does not use the number of NULL values to influence query planning.

compute stats customer;						
summary						
Updated 1 partition(s) and 18 column(s).						
compute stats store_sales;						
summary						
Updated 1 partition(s) and 23 column(s).						
show column stats customer;						
Column	Type	#Distinct Values	#Nulls	Max Size	Avg Size	
c_customer_sk	INT	139017	-1	4	4	
c_customer_id	STRING	111904	-1	16	16	
c_current_cdemo_sk	INT	95837	-1	4	4	
c_current_hdemo_sk	INT	8097	-1	4	4	
c_current_addr_sk	INT	57334	-1	4	4	
c_first_shipto_date_sk	INT	4374	-1	4	4	
c_first_sales_date_sk	INT	4409	-1	4	4	
c_salutation	STRING	7	-1	4	3.1308	
c_first_name	STRING	3887	-1	11	5.6356	
c_last_name	STRING	4739	-1	13	5.9106	
c_preferred_cust_flag	STRING	3	-1	1	0.9656	
c_birth_day	INT	31	-1	4	4	
c_birth_month	INT	12	-1	4	4	
c_birth_year	INT	71	-1	4	4	
c_birth_country	STRING	205	-1	20	8.4001	
c_login	STRING	1	-1	0	0	
c_email_address	STRING	94492	-1	46	26.485	
c_last_review_date	STRING	349	-1	7	6.7561	
show column stats store_sales;						
Column	Type	#Distinct Values	#Nulls	Max Size	Avg Size	
ss_sold_date_sk	INT	4395	-1	4	4	
ss_sold_time_sk	INT	63617	-1	4	4	
ss_item_sk	INT	19463	-1	4	4	
ss_customer_sk	INT	122720	-1	4	4	
ss_cdemo_sk	INT	242982	-1	4	4	
ss_hdemo_sk	INT	8097	-1	4	4	
ss_addr_sk	INT	70770	-1	4	4	
ss_store_sk	INT	6	-1	4	4	
ss_promo_sk	INT	355	-1	4	4	
ss_ticket_number	INT	304098	-1	4	4	
ss_quantity	INT	105	-1	4	4	
ss_wholesale_cost	FLOAT	9600	-1	4	4	
ss_list_price	FLOAT	22191	-1	4	4	
ss_sales_price	FLOAT	20693	-1	4	4	
ss_ext_discount_amt	FLOAT	228141	-1	4	4	
ss_ext_sales_price	FLOAT	433550	-1	4	4	
ss_ext_wholesale_cost	FLOAT	406291	-1	4	4	
ss_ext_list_price	FLOAT	574871	-1	4	4	
ss_ext_tax	FLOAT	91806	-1	4	4	
ss_coupon_amt	FLOAT	228141	-1	4	4	

ss_net_paid	FLOAT	493107	-1	4	4	
ss_net_paid_inc_tax	FLOAT	653523	-1	4	4	
ss_net_profit	FLOAT	611934	-1	4	4	

**HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read and execute permissions for all directories that are part of the table. (A table could span multiple different HDFS directories if it is partitioned. The directories could be widely scattered because a partition can reside in an arbitrary HDFS directory based on its `LOCATION` attribute.) The Impala user must also have execute permission for the database directory, and any parent directories of the database directory in HDFS.

**Related information:**

[COMPUTE STATS Statement](#) on page 170, [SHOW TABLE STATS Statement](#) on page 251

See [How Impala Uses Statistics for Query Optimization](#) on page 363 for usage information and examples.

## SHOW PARTITIONS Statement

`SHOW PARTITIONS` displays information about each partition for a partitioned table. (The output is the same as the `SHOW TABLE STATS` statement, but `SHOW PARTITIONS` only works on a partitioned table.) Because it displays table statistics for all partitions, the output is more informative if you have run the `COMPUTE STATS` statement after creating all the partitions. See [COMPUTE STATS Statement](#) on page 170 for details. For example, on a `CENSUS` table partitioned on the `YEAR` column:

**Security considerations:**

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

**Examples:**

[localhost:21000] > show partitions census;				
year	#Rows	#Files	Size	Format
2000	-1	0	0B	TEXT
2004	-1	0	0B	TEXT
2008	-1	0	0B	TEXT
2010	-1	0	0B	TEXT
2011	4	1	22B	TEXT
2012	4	1	22B	TEXT
2013	1	1	231B	PARQUET
Total	9	3	275B	

**HDFS permissions:**

The user ID that the `impalad` daemon runs under, typically the `impala` user, must have read and execute permissions for all directories that are part of the table. (A table could span multiple different HDFS directories if it is partitioned. The directories could be widely scattered because a partition can reside in an arbitrary HDFS directory based on its `LOCATION` attribute.) The Impala user must also have execute permission for the database directory, and any parent directories of the database directory in HDFS.

**Related information:**

See [How Impala Uses Statistics for Query Optimization](#) on page 363 for usage information and examples.

[SHOW TABLE STATS Statement](#) on page 251, [Partitioning for Impala Tables](#) on page 394

### SHOW FUNCTIONS Statement

By default, `SHOW FUNCTIONS` displays user-defined functions (UDFs) and `SHOW AGGREGATE FUNCTIONS` displays user-defined aggregate functions (UDAFs) associated with a particular database. The output from `SHOW FUNCTIONS` includes the argument signature of each function. You specify this argument signature as part of the `DROP FUNCTION` statement. You might have several UDFs with the same name, each accepting different argument data types.

#### Security considerations:

When authorization is enabled, the output of the `SHOW` statement is limited to those objects for which you have some privilege. There might be other database, tables, and so on, but their names are concealed. If you believe an object exists but you cannot see it in the `SHOW` output, check with the system administrator if you need to be granted a new privilege for that object. See [Enabling Sentry Authorization for Impala](#) on page 91 for how to set up authorization and add privileges for specific kinds of objects.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

#### Examples:

To display Impala built-in functions, specify the special database name `_impala_builtins`:

```
show functions in _impala_builtin;
+-----+-----+
| return type | signature
+-----+-----+
| BOOLEAN      | ifnull(BOOLEAN, BOOLEAN)
| TINYINT       | ifnull(TINYINT, TINYINT)
| SMALLINT      | ifnull(SMALLINT, SMALLINT)
| INT           | ifnull(INT, INT)
...
show functions in _impala_builtin like '*week*';
+-----+-----+
| return type | signature
+-----+-----+
| INT          | weekofyear(TIMESTAMP)
| TIMESTAMP    | weeks_add(TIMESTAMP, INT)
| TIMESTAMP    | weeks_add(TIMESTAMP, BIGINT)
| TIMESTAMP    | weeks_sub(TIMESTAMP, INT)
| TIMESTAMP    | weeks_sub(TIMESTAMP, BIGINT)
| INT          | dayofweek(TIMESTAMP)
```

To search for functions that use a particular data type, specify a case-sensitive data type name in all capitals:

```
show functions in _impala_builtin like '*BIGINT*';
+-----+
| name
+-----+
| adddate(TIMESTAMP, BIGINT)
| bin(BIGINT)
| coalesce(BIGINT...)
...
```

#### Related information:

[Functions](#) on page 155, [Built-in Functions](#) on page 257, [Impala User-Defined Functions \(UDFs\)](#) on page 308, [SHOW DATABASES](#) on page 248, [SHOW TABLES Statement](#) on page 249

### USE Statement

Switches the current session to a specified database. The **current database** is where any `CREATE TABLE`, `INSERT`, `SELECT`, or other statements act when you specify a table or other object name, without prefixing it with a database name. The new current database applies for the duration of the session or until another `USE` statement is executed.

**Syntax:**

```
USE db_name
```

By default, when you connect to an Impala instance, you begin in a database named `default`.

**Usage notes:**

Switching the default database is convenient in the following situations:

- To avoid qualifying each reference to a table with the database name. For example, `SELECT * FROM t1 JOIN t2` rather than `SELECT * FROM db.t1 JOIN db.t2`.
- To do a sequence of operations all within the same database, such as creating a table, inserting data, and querying the table.

To start the `impala-shell` interpreter and automatically issue a `USE` statement for a particular database, specify the option `-d db_name` for the `impala-shell` command. The `-d` option is useful to run SQL scripts, such as setup or test scripts, against multiple databases without hardcoding a `USE` statement into the SQL source.

**Examples:**

See [CREATE DATABASE Statement](#) on page 175 for examples covering `CREATE DATABASE`, `USE`, and `DROP DATABASE`.

**Cancellation:** Cannot be cancelled.

**HDFS permissions:** This statement does not touch any HDFS files or directories, therefore no HDFS permissions are required.

**Related information:**

[CREATE DATABASE Statement](#) on page 175, [DROP DATABASE Statement](#) on page 192, [SHOW DATABASES](#) on page 248

## Built-in Functions

Impala supports several categories of built-in functions. These functions let you perform mathematical calculations, string manipulation, date calculations, and other kinds of data transformations directly in `SELECT` statements. The built-in functions let a SQL query return results with all formatting, calculating, and type conversions applied, rather than performing time-consuming postprocessing in another application. By applying function calls where practical, you can make a SQL query that is as convenient as an expression in a procedural programming language or a formula in a spreadsheet.

The categories of functions supported by Impala are:

- [Impala Mathematical Functions](#) on page 259
- [Impala Type Conversion Functions](#) on page 267
- [Impala Date and Time Functions](#) on page 268
- [Impala Conditional Functions](#) on page 275
- [Impala String Functions](#) on page 279
- Aggregation functions, explained in [Impala Aggregate Functions](#) on page 284.

You call any of these functions through the `SELECT` statement. For most functions, you can omit the `FROM` clause and supply literal values for any required arguments:

```
select abs(-1);
+-----+
| abs(-1) |
+-----+
| 1        |
+-----+

select concat('The rain ', 'in Spain');
+-----+
```

```
| concat('the rain ', 'in spain') |
+-----+
| The rain in Spain |
+-----+

select power(2,5);
+-----+
| power(2, 5) |
+-----+
| 32           |
+-----+
```

When you use a `FROM` clause and specify a column name as a function argument, the function is applied for each item in the result set:

```
select concat('Country = ',country_code) from all_countries where population > 100000000;
select round(price) as dollar_value from product_catalog where price between 0.0 and
100.0;
```

Typically, if any argument to a built-in function is `NULL`, the result value is also `NULL`:

```
select cos(null);
+-----+
| cos(null) |
+-----+
| NULL       |
+-----+

select power(2,null);
+-----+
| power(2, null) |
+-----+
| NULL           |
+-----+

select concat('a',null,'b');
+-----+
| concat('a', null, 'b') |
+-----+
| NULL                 |
+-----+
```

Aggregate functions are a special category with different rules. These functions calculate a return value across all the items in a result set, so they require a `FROM` clause in the query:

```
select count(product_id) from product_catalog;
select max(height), avg(height) from census_data where age > 20;
```

Aggregate functions also ignore `NULL` values rather than returning a `NULL` result. For example, if some rows have `NULL` for a particular column, those rows are ignored when computing the `AVG()` for that column. Likewise, specifying `COUNT(col_name)` in a query counts only those rows where `col_name` contains a non-`NULL` value.

Aggregate functions are a special category with different rules. These functions calculate a return value across all the items in a result set, so they do require a `FROM` clause in the query:

```
select count(product_id) from product_catalog;
select max(height), avg(height) from census_data where age > 20;
```

Aggregate functions also ignore `NULL` values rather than returning a `NULL` result. For example, if some rows have `NULL` for a particular column, those rows are ignored when computing the `AVG()` for that column. Likewise, specifying `COUNT(col_name)` in a query counts only those rows where `col_name` contains a non-`NULL` value.

Analytic functions are a variation on aggregate functions. Instead of returning a single value, or an identical value for each group of rows, they can compute values that vary based on a “window” consisting of other rows around them in the result set.

## Impala Mathematical Functions

Mathematical functions, or arithmetic functions, perform numeric calculations that are typically more complex than basic addition, subtraction, multiplication, and division. For example, these functions include trigonometric, logarithmic, and base conversion operations.

- **Note:** In Impala, exponentiation uses the `pow( )` function rather than an exponentiation operator such as `**`.

### Related information:

The mathematical functions operate mainly on these data types: [INT Data Type](#) on page 126, [BIGINT Data Type](#) on page 112, [SMALLINT Data Type](#) on page 128, [TINYINT Data Type](#) on page 135, [DOUBLE Data Type](#) on page 125, [FLOAT Data Type](#) on page 125, and [DECIMAL Data Type \(CDH 5.1 or higher only\)](#) on page 116. For the operators that perform the standard operations such as addition, subtraction, multiplication, and division, see [Arithmetic Operators](#) on page 142.

### Function reference:

Impala supports the following mathematical functions:

`abs(numeric_type a)`

**Purpose:** Returns the absolute value of the argument.

**Return type:** Same as the input value

**Usage notes:** Use this function to ensure all return values are positive. This is different than the `positive()` function, which returns its argument unchanged (even if the argument was negative).

`acos(double a)`

**Purpose:** Returns the arccosine of the argument.

**Return type:** `double`

`asin(double a)`

**Purpose:** Returns the arcsine of the argument.

**Return type:** `double`

`atan(double a)`

**Purpose:** Returns the arctangent of the argument.

**Return type:** `double`

`bin(bigint a)`

**Purpose:** Returns the binary representation of an integer value, that is, a string of 0 and 1 digits.

**Return type:** `string`

`ceil(double a), ceiling(double a), ceil(decimal(p,s) a), ceiling(decimal(p,s) a)`

**Purpose:** Returns the smallest integer that is greater than or equal to the argument.

**Return type:** `int` or `decimal(p,s)` depending on the type of the input argument

`conv(bigint num, int from_base, int to_base), conv(string num, int from_base, int to_base)`

**Purpose:** Returns a string representation of an integer value in a particular base. The input value can be a string, for example to convert a hexadecimal number such as `fce2` to decimal. To use the return value as a number (for example, when converting to base 10), use `CAST( )` to convert to the appropriate type.

**Return type:** `string`

`cos(double a)`

**Purpose:** Returns the cosine of the argument.

**Return type:** double

`degrees(double a)`

**Purpose:** Converts argument value from radians to degrees.

**Return type:** double

`e()`

**Purpose:** Returns the [mathematical constant e](#).

**Return type:** double

`exp(double a)`

**Purpose:** Returns the [mathematical constant e](#) raised to the power of the argument.

**Return type:** double

`floor(double a)`

**Purpose:** Returns the largest integer that is less than or equal to the argument.

**Return type:** bigint or decimal(p,s) depending on the type of the input argument

`fmod(double a, double b), fmod(float a, float b)`

**Purpose:** Returns the modulus of a floating-point number. Equivalent to the % arithmetic operator.

**Return type:** float or double, depending on type of arguments

**Added in:** Impala 1.1.1

### Usage notes:

Because this function operates on DOUBLE or FLOAT values, it is subject to potential rounding errors for values that cannot be represented precisely. Prefer to use whole numbers, or values that you know can be represented precisely by the DOUBLE or FLOAT types.

### Examples:

The following examples show equivalent operations with the `fmod()` function and the % arithmetic operator, for values not subject to any rounding error.

```
select fmod(10,3);
+-----+
| fmod(10, 3) |
+-----+
| 1           |
+-----+

select fmod(5.5,2);
+-----+
| fmod(5.5, 2) |
+-----+
| 1.5          |
+-----+

select 10 % 3;
+-----+
| 10 % 3    |
+-----+
| 1           |
+-----+

select 5.5 % 2;
+-----+
| 5.5 % 2   |
+-----+
| 1.5         |
+-----+
```

The following examples show operations with the `fmod()` function for values that cannot be represented precisely by the `DOUBLE` or `FLOAT` types, and thus are subject to rounding error. `fmod(9.9, 3.0)` returns a value slightly different than the expected 0.9 because of rounding. `fmod(9.9, 3.3)` returns a value quite different from the expected value of 0 because of rounding error during intermediate calculations.

```
select fmod(9.9,3.0);
+-----+
| fmod(9.9, 3.0) |
+-----+
| 0.8999996185302734 |
+-----+

select fmod(9.9,3.3);
+-----+
| fmod(9.9, 3.3) |
+-----+
| 3.299999713897705 |
+-----+
```

#### `fnv_hash(type v)`

**Purpose:** Returns a consistent 64-bit value derived from the input argument, for convenience of implementing hashing logic in an application.

**Return type:** `BIGINT`

#### Usage notes:

You might use the return value in an application where you perform load balancing, bucketing, or some other technique to divide processing or storage.

Because the result can be any 64-bit value, to restrict the value to a particular range, you can use an expression that includes the `ABS()` function and the `%` (modulo) operator. For example, to produce a hash value in the range 0-9, you could use the expression `ABS(FNV_HASH(x)) % 10`.

This function implements the same algorithm that Impala uses internally for hashing, on systems where the CRC32 instructions are not available.

This function implements the [Fowler–Noll–Vo hash function](#), in particular the FNV-1a variation. This is not a perfect hash function: some combinations of values could produce the same result value. It is not suitable for cryptographic use.

Similar input values of different types could produce different hash values, for example the same numeric value represented as `SMALLINT` or `BIGINT`, `FLOAT` or `DOUBLE`, or `DECIMAL(5,2)` or `DECIMAL(20,5)`.

#### Examples:

```
[localhost:21000] > create table h (x int, s string);
[localhost:21000] > insert into h values (0, 'hello'), (1,'world'),
(1234567890,'antidisestablishmentarianism');
[localhost:21000] > select x, fnv_hash(x) from h;
+-----+-----+
| x      | fnv_hash(x)   |
+-----+-----+
| 0      | -2611523532599129963 |
| 1      | 4307505193096137732 |
| 1234567890 | 3614724209955230832 |
+-----+-----+
[localhost:21000] > select s, fnv_hash(s) from h;
+-----+-----+
| s          | fnv_hash(s)   |
+-----+-----+
| hello      | 6414202926103426347 |
| world      | 6535280128821139475 |
| antidisestablishmentarianism | -209330013948433970 |
+-----+-----+
[localhost:21000] > select s, abs(fnv_hash(s)) % 10 from h;
+-----+-----+
| s          | abs(fnv_hash(s)) % 10.0 |
+-----+-----+
```

hello	8
world	6
antidisestablishmentarianism	4

For short argument values, the high-order bits of the result have relatively low entropy:

[localhost:21000] > create table b (x boolean);
[localhost:21000] > insert into b values (true), (true), (false), (false);
[localhost:21000] > select x, fnv_hash(x) from b;
+-----+-----+
x   fnv_hash(x)
+-----+-----+
true   2062020650953872396
true   2062020650953872396
false   2062021750465500607
false   2062021750465500607
+-----+-----+

**Added in:** Impala 1.2.2

`greatest(bigint a[, bigint b ...]), greatest(double a[, double b ...]), greatest(decimal(p,s) a[, decimal(p,s) b ...]), greatest(string a[, string b ...]), greatest(timestamp a[, timestamp b ...])`

**Purpose:** Returns the largest value from a list of expressions.

**Return type:** same as the initial argument value, except that integer values are promoted to BIGINT and floating-point values are promoted to DOUBLE; use CAST( ) when inserting into a smaller numeric column

`hex(bigint a), hex(string a)`

**Purpose:** Returns the hexadecimal representation of an integer value, or of the characters in a string.

**Return type:** string

`is_inf(double a),`

**Purpose:** Tests whether a value is equal to the special value "inf", signifying infinity.

**Return type:** boolean

**Usage notes:**

Infinity and NaN can be specified in text data files as `inf` and `nan` respectively, and Impala interprets them as these special values. They can also be produced by certain arithmetic expressions; for example, `pow(-1, 0.5)` returns infinity and `1/0` returns NaN. Or you can cast the literal values, such as `CAST('nan' AS DOUBLE)` or `CAST('inf' AS DOUBLE)`.

`is_nan(double a),`

**Purpose:** Tests whether a value is equal to the special value "NaN", signifying "not a number".

**Return type:** boolean

**Usage notes:**

Infinity and NaN can be specified in text data files as `inf` and `nan` respectively, and Impala interprets them as these special values. They can also be produced by certain arithmetic expressions; for example, `pow(-1, 0.5)` returns infinity and `1/0` returns NaN. Or you can cast the literal values, such as `CAST('nan' AS DOUBLE)` or `CAST('inf' AS DOUBLE)`.

`least(bigint a[, bigint b ...]), least(double a[, double b ...]), least(decimal(p,s) a[, decimal(p,s) b ...]), least(string a[, string b ...]), least(timestamp a[, timestamp b ...])`

**Purpose:** Returns the smallest value from a list of expressions.

**Return type:** same as the initial argument value, except that integer values are promoted to BIGINT and floating-point values are promoted to DOUBLE; use CAST( ) when inserting into a smaller numeric column

`ln(double a)`

**Purpose:** Returns the [natural logarithm](#) of the argument.

**Return type:** double

`log(double base, double a)`

**Purpose:** Returns the logarithm of the second argument to the specified base.

**Return type:** double

`log10(double a)`

**Purpose:** Returns the logarithm of the argument to the base 10.

**Return type:** double

`log2(double a)`

**Purpose:** Returns the logarithm of the argument to the base 2.

**Return type:** double

`max_int(), max_tinyint(), max_smallint(), max_bigint()`

**Purpose:** Returns the largest value of the associated integral type.

**Return type:** The same as the integral type being checked.

**Usage notes:** Use the corresponding `min_` and `max_` functions to check if all values in a column are within the allowed range, before copying data or altering column definitions. If not, switch to the next higher integral type or to a `DECIMAL` with sufficient precision.

`min_int(), min_tinyint(), min_smallint(), min_bigint()`

**Purpose:** Returns the smallest value of the associated integral type (a negative number).

**Return type:** The same as the integral type being checked.

**Usage notes:** Use the corresponding `min_` and `max_` functions to check if all values in a column are within the allowed range, before copying data or altering column definitions. If not, switch to the next higher integral type or to a `DECIMAL` with sufficient precision.

`mod(numeric_type a, same_type b)`

**Purpose:** Returns the modulus of a number. Equivalent to the `%` arithmetic operator. Works with any size integer type, any size floating-point type, and `DECIMAL` with any precision and scale.

**Return type:** Same as the input value

**Added in:** CDH 5.4.0 (Impala 2.2.0)

**Usage notes:**

Because this function works with `DECIMAL` values, prefer it over `fmod()` when working with fractional values. It is not subject to the rounding errors that make `fmod()` problematic with floating-point numbers. The `%` arithmetic operator now uses the `mod()` function in cases where its arguments can be interpreted as `DECIMAL` values, increasing the accuracy of that operator.

**Examples:**

The following examples show how the `mod()` function works for whole numbers and fractional values, and how the `%` operator works the same way. In the case of `mod(9.9, 3)`, the type conversion for the second argument results in the first argument being interpreted as `DOUBLE`, so to produce an accurate `DECIMAL` result requires casting the second argument or writing it as a `DECIMAL` literal, `3.0`.

```
select mod(10,3);
+-----+
| fmod(10, 3) |
+-----+
| 1           |
```

```
+-----+
select mod(5.5,2);
+-----+
| fmod(5.5, 2) |
+-----+
| 1.5           |
+-----+

select 10 % 3;
+-----+
| 10 % 3 |
+-----+
| 1          |
+-----+

select 5.5 % 2;
+-----+
| 5.5 % 2   |
+-----+
| 1.5        |
+-----+

select mod(9.9,3.3);
+-----+
| mod(9.9, 3.3) |
+-----+
| 0.0          |
+-----+

select mod(9.9,3);
+-----+
| mod(9.9, 3) |
+-----+
| 0.8999996185302734 |
+-----+

select mod(9.9, cast(3 as decimal(2,1)));
+-----+
| mod(9.9, cast(3 as decimal(2,1))) |
+-----+
| 0.9          |
+-----+

select mod(9.9,3.0);
+-----+
| mod(9.9, 3.0) |
+-----+
| 0.9          |
+-----+
```

### `negative(numeric_type a)`

**Purpose:** Returns the argument with the sign reversed; returns a positive value if the argument was already negative.

**Return type:** Same as the input value

**Usage notes:** Use `-abs(a)` instead if you need to ensure all return values are negative.

### `pi()`

**Purpose:** Returns the constant pi.

**Return type:** `double`

### `pmod(bigint a, bigint b), pmod(double a, double b)`

**Purpose:** Returns the positive modulus of a number. Primarily for [HiveQL compatibility](#).

**Return type:** `int` or `double`, depending on type of arguments

**Examples:**

The following examples show how the `fmod( )` function sometimes returns a negative value depending on the sign of its arguments, and the `pmod( )` function returns the same value as `fmod( )`, but sometimes with the sign flipped.

```
select fmod(-5,2);
+-----+
| fmod(-5, 2) |
+-----+
| -1          |
+-----+

select pmod(-5,2);
+-----+
| pmod(-5, 2) |
+-----+
| 1           |
+-----+

select fmod(-5,-2);
+-----+
| fmod(-5, -2) |
+-----+
| -1          |
+-----+

select pmod(-5,-2);
+-----+
| pmod(-5, -2) |
+-----+
| -1          |
+-----+

select fmod(5,-2);
+-----+
| fmod(5, -2) |
+-----+
| 1           |
+-----+

select pmod(5,-2);
+-----+
| pmod(5, -2) |
+-----+
| -1          |
+-----+
```

`positive(numeric_type a)`

**Purpose:** Returns the original argument unchanged (even if the argument is negative).

**Return type:** Same as the input value

**Usage notes:** Use `abs( )` instead if you need to ensure all return values are positive.

`pow(double a, double p), power(double a, double p)`

**Purpose:** Returns the first argument raised to the power of the second argument.

**Return type:** `double`

`precision(numeric_expression)`

**Purpose:** Computes the precision (number of decimal digits) needed to represent the type of the argument expression as a `DECIMAL` value.

**Usage notes:**

Typically used in combination with the `scale( )` function, to determine the appropriate `DECIMAL(precision,scale)` type to declare in a `CREATE TABLE` statement or `CAST( )` function.

**Return type:** `int`

**Examples:**

The following examples demonstrate how to check the precision and scale of numeric literals or other numeric expressions. Impala represents numeric literals in the smallest appropriate type. 5 is a `TINYINT` value, which ranges from -128 to 127, therefore 3 decimal digits are needed to represent the entire range, and because it is an integer value there are no fractional digits. 1.333 is interpreted as a `DECIMAL` value, with 4 digits total and 3 digits after the decimal point.

```
[localhost:21000] > select precision(5), scale(5);
+-----+-----+
| precision(5) | scale(5) |
+-----+-----+
| 3           | 0       |
+-----+-----+
[localhost:21000] > select precision(1.333), scale(1.333);
+-----+-----+
| precision(1.333) | scale(1.333) |
+-----+-----+
| 4             | 3       |
+-----+-----+
[localhost:21000] > with t1 as
  ( select cast(12.34 as decimal(20,2)) x union select cast(1 as decimal(8,6)) x
  )
  select precision(x), scale(x) from t1 limit 1;
+-----+-----+
| precision(x) | scale(x) |
+-----+-----+
| 24           | 6       |
+-----+-----+
```

### `quotient(int numerator, int denominator)`

**Purpose:** Returns the first argument divided by the second argument, discarding any fractional part. Avoids promoting arguments to `DOUBLE` as happens with the `/` SQL operator.

**Return type:** `int`

### `radians(double a)`

**Purpose:** Converts argument value from degrees to radians.

**Return type:** `double`

### `rand(), rand(int seed)`

**Purpose:** Returns a random value between 0 and 1. After `rand()` is called with a seed argument, it produces a consistent random sequence based on the seed value.

**Return type:** `double`

**Usage notes:** Currently, the random sequence is reset after each query, and multiple calls to `rand()` within the same query return the same value each time. For different number sequences that are different for each query, pass a unique seed value to each call to `rand()`. For example, `select rand(unix_timestamp()) from ...`

### `round(double a), round(double a, int d), round(decimal a, int_type d)`

**Purpose:** Rounds a floating-point value. By default (with a single argument), rounds to the nearest integer. Values ending in .5 are rounded up for positive numbers, down for negative numbers (that is, away from zero). The optional second argument specifies how many digits to leave after the decimal point; values greater than zero produce a floating-point return value rounded to the requested number of digits to the right of the decimal point.

**Return type:** `bigint` for single `float` argument. `double` for `double` argument when second argument greater than zero. For `DECIMAL` values, the smallest `DECIMAL(p,s)` type with appropriate precision and scale.

### `scale(numeric_expression)`

**Purpose:** Computes the scale (number of decimal digits to the right of the decimal point) needed to represent the type of the argument expression as a `DECIMAL` value.

**Usage notes:**

Typically used in combination with the `precision()` function, to determine the appropriate `DECIMAL(precision,scale)` type to declare in a `CREATE TABLE` statement or `CAST()` function.

**Return type:** int

**Examples:**

The following examples demonstrate how to check the precision and scale of numeric literals or other numeric expressions. Impala represents numeric literals in the smallest appropriate type. 5 is a `TINYINT` value, which ranges from -128 to 127, therefore 3 decimal digits are needed to represent the entire range, and because it is an integer value there are no fractional digits. 1.333 is interpreted as a `DECIMAL` value, with 4 digits total and 3 digits after the decimal point.

```
[localhost:21000] > select precision(5), scale(5);
+-----+-----+
| precision(5) | scale(5) |
+-----+-----+
| 3           | 0       |
+-----+-----+
[localhost:21000] > select precision(1.333), scale(1.333);
+-----+-----+
| precision(1.333) | scale(1.333) |
+-----+-----+
| 4           | 3       |
+-----+-----+
[localhost:21000] > with t1 as
  ( select cast(12.34 as decimal(20,2)) x union select cast(1 as decimal(8,6)) x
  )
  select precision(x), scale(x) from t1 limit 1;
+-----+-----+
| precision(x) | scale(x) |
+-----+-----+
| 24          | 6       |
+-----+-----+
```

**sign(double a)**

**Purpose:** Returns -1, 0, or 1 to indicate the signedness of the argument value.

**Return type:** int

**sin(double a)**

**Purpose:** Returns the sine of the argument.

**Return type:** double

**sqrt(double a)**

**Purpose:** Returns the square root of the argument.

**Return type:** double

**tan(double a)**

**Purpose:** Returns the tangent of the argument.

**Return type:** double

**unhex(string a)**

**Purpose:** Returns a string of characters with ASCII values corresponding to pairs of hexadecimal digits in the argument.

**Return type:** string

## Impala Type Conversion Functions

Impala supports the following type conversion functions:

- `cast(expr as type)`

Conversion functions are usually used in combination with other functions, to explicitly pass the expected data types. Impala has strict rules regarding data types for function parameters. For example, Impala does not automatically convert a `DOUBLE` value to `FLOAT`, a `BIGINT` value to `INT`, or other conversion where precision could be lost or overflow could occur. Use `CAST` when passing a column value or literal to a function that expects a parameter with a different type. For example:

```
select concat('Here are the first ',10,' results.');// Fails  
select concat('Here are the first ',cast(10 as string),' results.');// Succeeds
```

The way this function deals with time zones when converting to or from `TIMESTAMP` values is affected by the `-use_local_tz_for_unix_timestamp_conversions` startup flag for the `impalad` daemon. See [TIMESTAMP Data Type](#) on page 130 for details about how Impala handles time zone considerations for the `TIMESTAMP` data type.

### Related information:

For details of casts from each kind of data type, see the description of the appropriate type: [TINYINT Data Type](#) on page 135, [SMALLINT Data Type](#) on page 128, [INT Data Type](#) on page 126, [BIGINT Data Type](#) on page 112, [FLOAT Data Type](#) on page 125, [DOUBLE Data Type](#) on page 125, [DECIMAL Data Type \(CDH 5.1 or higher only\)](#) on page 116, [STRING Data Type](#) on page 128, [CHAR Data Type \(CDH 5.2 or higher only\)](#) on page 114, [VARCHAR Data Type \(CDH 5.2 or higher only\)](#) on page 136, [TIMESTAMP Data Type](#) on page 130, [BOOLEAN Data Type](#) on page 113

## Impala Date and Time Functions

The underlying Impala data type for date and time data is `TIMESTAMP`, which has both a date and a time portion. Functions that extract a single field, such as `hour()` or `minute()`, typically return an integer value. Functions that format the date portion, such as `date_add()` or `to_date()`, typically return a string value.

You can also adjust a `TIMESTAMP` value by adding or subtracting an `INTERVAL` expression. See [TIMESTAMP Data Type](#) on page 130 for details. `INTERVAL` expressions are also allowed as the second argument for the `date_add()` and `date_sub()` functions, rather than integers.

Some of these functions are affected by the setting of the `-use_local_tz_for_unix_timestamp_conversions` startup flag for the `impalad` daemon. This setting is off by default, meaning that functions such as `from_unixtime()` and `unix_timestamp()` consider the input values to always represent the UTC time zone. This setting also applies when you `CAST()` a `BIGINT` value to `TIMESTAMP`, or a `TIMESTAMP` value to `BIGINT`. When this setting is enabled, these functions and operations convert to and from values representing the local time zone. See [TIMESTAMP Data Type](#) on page 130 for details about how Impala handles time zone considerations for the `TIMESTAMP` data type.

Impala supports the following data and time functions:

`add_months(timestamp date, int months), add_months(timestamp date, bigint months)`

**Purpose:** Returns the specified date and time plus some number of months.

**Return type:** `timestamp`

**Usage notes:**

Same as `months_add()`. Available in Impala 1.4 and higher. For compatibility when porting code with vendor extensions.

`adddate(timestamp startdate, int days), adddate(timestamp startdate, bigint days),`

**Purpose:** Adds a specified number of days to a `TIMESTAMP` value. Similar to `date_add()`, but starts with an actual `TIMESTAMP` value instead of a string that is converted to a `TIMESTAMP`.

**Return type:** `timestamp`

`current_timestamp()`

**Purpose:** Alias for the `now()` function.

**Return type:** timestamp

`date_add(timestamp startdate, int days), date_add(timestamp startdate, interval_expression)`

**Purpose:** Adds a specified number of days to a TIMESTAMP value. The first argument can be a string, which is automatically cast to TIMESTAMP if it uses the recognized format, as described in [TIMESTAMP Data Type](#) on page 130. With an INTERVAL expression as the second argument, you can calculate a delta value using other units such as weeks, years, hours, seconds, and so on; see [TIMESTAMP Data Type](#) on page 130 for details.

**Return type:** timestamp

`date_part(string, timestamp)`

**Purpose:** Similar to EXTRACT(), with the argument order reversed. Supports the same date and time units as EXTRACT(). For compatibility with SQL code containing vendor extensions.

**Return type:** int

`date_sub(timestamp startdate, int days), date_sub(timestamp startdate, interval_expression)`

**Purpose:** Subtracts a specified number of days from a TIMESTAMP value. The first argument can be a string, which is automatically cast to TIMESTAMP if it uses the recognized format, as described in [TIMESTAMP Data Type](#) on page 130. With an INTERVAL expression as the second argument, you can calculate a delta value using other units such as weeks, years, hours, seconds, and so on; see [TIMESTAMP Data Type](#) on page 130 for details.

**Return type:** timestamp

`datediff(string enddate, string startdate)`

**Purpose:** Returns the number of days between two dates represented as strings.

**Return type:** int

`day(string date), dayofmonth(string date)`

**Purpose:** Returns the day field from a date represented as a string.

**Return type:** int

`dayname(string date)`

**Purpose:** Returns the day field from a date represented as a string, converted to the string corresponding to that day name. The range of return values is 'Sunday' to 'Saturday'. Used in report-generating queries, as an alternative to calling dayofweek() and turning that numeric return value into a string using a CASE expression.

**Return type:** string

`dayofweek(string date)`

**Purpose:** Returns the day field from a date represented as a string, corresponding to the day of the week. The range of return values is 1 (Sunday) to 7 (Saturday).

**Return type:** int

`dayofyear(timestamp date)`

**Purpose:** Returns the day field from a TIMESTAMP value, corresponding to the day of the year. The range of return values is 1 (January 1) to 366 (December 31 of a leap year).

**Return type:** int

`days_add(timestamp startdate, int days), days_add(timestamp startdate, bigint days)`

**Purpose:** Adds a specified number of days to a TIMESTAMP value. Similar to date\_add(), but starts with an actual TIMESTAMP value instead of a string that is converted to a TIMESTAMP.

**Return type:** timestamp

```
days_sub(timestamp startdate, int days), days_sub(timestamp startdate, bigint days)
```

**Purpose:** Subtracts a specified number of days from a `TIMESTAMP` value. Similar to `date_sub()`, but starts with an actual `TIMESTAMP` value instead of a string that is converted to a `TIMESTAMP`.

**Return type:** `timestamp`

```
extract(timestamp, string unit) extract(unit FROM timestamp)
```

**Purpose:** Returns one of the numeric date or time fields from a `TIMESTAMP` value.

**Unit argument:** The `unit` string can be one of `year`, `month`, `day`, `hour`, `minute`, `second`, or `millisecond`. This argument value is case-insensitive.

In Impala 2.0 and higher, you can use special syntax rather than a regular function call, for compatibility with code that uses the SQL-99 format with the `FROM` keyword. With this style, the unit names are identifiers rather than STRING literals. For example, the following calls are both equivalent:

```
extract(year from now());
extract(now(), "year");
```

### Usage notes:

Typically used in `GROUP BY` queries to arrange results by hour, day, month, and so on. You can also use this function in an `INSERT ... SELECT` into a partitioned table to split up `TIMESTAMP` values into individual parts, if the partitioned table has separate partition key columns representing year, month, day, and so on. If you need to divide by more complex units of time, such as by week or by quarter, use the `TRUNC()` function instead.

**Return type:** `int`

```
from_unixtime(bigint unixtime[, string format])
```

**Purpose:** Converts the number of seconds from the Unix epoch to the specified time into a string in the local time zone.

**Return type:** `string`

In Impala 2.2.0 and higher, built-in functions that accept or return integers representing `TIMESTAMP` values use the `BIGINT` type for parameters and return values, rather than `INT`. This change lets the date and time functions avoid an overflow error that would otherwise occur on January 19th, 2038 (known as the ["Year 2038 problem"](#) or ["Y2K38 problem"](#)). This change affects the `from_unixtime()` and `unix_timestamp()` functions. You might need to change application code that interacts with these functions, change the types of columns that store the return values, or add `CAST()` calls to SQL statements that call these functions.

### Usage notes:

The format string accepts the variations allowed for the `TIMESTAMP` data type: date plus time, date by itself, time by itself, and optional fractional seconds for the time. See [TIMESTAMP Data Type](#) on page 130 for details.

Currently, the format string is case-sensitive, especially to distinguish `m` for minutes and `M` for months. In Impala 1.3 and later, you can switch the order of elements, use alternative separator characters, and use a different number of placeholders for each unit. Adding more instances of `y`, `d`, `H`, and so on produces output strings zero-padded to the requested number of characters. The exception is `M` for months, where `M` produces a non-padded value such as `3`, `MM` produces a zero-padded value such as `03`, `MMM` produces an abbreviated month name such as `Mar`, and sequences of 4 or more `M` are not allowed. A date string including all fields could be `"yyyy-MM-dd HH:mm:ss.SSSSSS"`, `"dd/MM/yyyy HH:mm:ss.SSSSSS"`, `"MMM dd, yyyy HH.mm.ss (SSSSSS)"` or other combinations of placeholders and separator characters.

The way this function deals with time zones when converting to or from `TIMESTAMP` values is affected by the `-use_local_tz_for_unix_timestamp_conversions` startup flag for the `impalad` daemon. See [TIMESTAMP Data Type](#) on page 130 for details about how Impala handles time zone considerations for the `TIMESTAMP` data type.

- Note:** The more flexible format strings allowed with the built-in functions do not change the rules about using `CAST()` to convert from a string to a `TIMESTAMP` value. Strings being converted through `CAST()` must still have the elements in the specified order and use the specified delimiter characters, as described in [TIMESTAMP Data Type](#) on page 130.

**Examples:**

```
[localhost:21000] > select from_unixtime(1392394861, "yyyy-MM-dd HH:mm:ss.SSSS");
+-----+
| from_unixtime(1392394861, 'yyyy-mm-dd hh:mm:ss.sssss') |
+-----+
| 2014-02-14 16:21:01.0000 |
+-----+
[localhost:21000] > select from_unixtime(1392394861, "yyyy-MM-dd");
+-----+
| from_unixtime(1392394861, 'yyyy-mm-dd') |
+-----+
| 2014-02-14 |
+-----+
[localhost:21000] > select from_unixtime(1392394861, "HH:mm:ss.SSSS");
+-----+
| from_unixtime(1392394861, 'hh:mm:ss.sssss') |
+-----+
| 16:21:01.0000 |
+-----+
[localhost:21000] > select from_unixtime(1392394861, "HH:mm:ss");
+-----+
| from_unixtime(1392394861, 'hh:mm:ss') |
+-----+
| 16:21:01 |
+-----+
```

`unix_timestamp()` and `from_unixtime()` are often used in combination to convert a `TIMESTAMP` value into a particular string format. For example:

```
select from_unixtime(unix_timestamp(now() + interval 3 days), 'yyyy/MM/dd HH:mm');
```

**from\_utc\_timestamp(timestamp, string timezone)**

**Purpose:** Converts a specified UTC timestamp value into the appropriate value for a specified time zone.

**Return type:** `timestamp`

**Usage notes:** Often used to translate UTC time zone data stored in a table back to the local date and time for reporting. The opposite of the `to_utc_timestamp()` function.

**Examples:** See discussion of time zones in [TIMESTAMP Data Type](#) on page 130 for information about using this function for conversions between the local time zone and UTC.

**hour(string date)**

**Purpose:** Returns the hour field from a date represented as a string.

**Return type:** `int`

**hours\_add(timestamp date, int hours), hours\_add(timestamp date, bigint hours)**

**Purpose:** Returns the specified date and time plus some number of hours.

**Return type:** `timestamp`

**hours\_sub(timestamp date, int hours), hours\_sub(timestamp date, bigint hours)**

**Purpose:** Returns the specified date and time minus some number of hours.

**Return type:** `timestamp`

## Impala SQL Language Reference

```
microseconds_add(timestamp date, int microseconds),microseconds_add(timestamp date, bigint
microseconds)

Purpose: Returns the specified date and time plus some number of microseconds.

Return type: timestamp

microseconds_sub(timestamp date, int microseconds),microseconds_sub(timestamp date, bigint
microseconds)

Purpose: Returns the specified date and time minus some number of microseconds.

Return type: timestamp

milliseconds_add(timestamp date, int milliseconds),milliseconds_add(timestamp date, bigint
milliseconds)

Purpose: Returns the specified date and time plus some number of milliseconds.

Return type: timestamp

milliseconds_sub(timestamp date, int milliseconds),milliseconds_sub(timestamp date, bigint
milliseconds)

Purpose: Returns the specified date and time minus some number of milliseconds.

Return type: timestamp

minute(string date)

Purpose: Returns the minute field from a date represented as a string.

Return type: int

minutes_add(timestamp date, int minutes),minutes_add(timestamp date, bigint minutes)

Purpose: Returns the specified date and time plus some number of minutes.

Return type: timestamp

minutes_sub(timestamp date, int minutes),minutes_sub(timestamp date, bigint minutes)

Purpose: Returns the specified date and time minus some number of minutes.

Return type: timestamp

month(string date)

Purpose: Returns the month field from a date represented as a string.

Return type: int

months_add(timestamp date, int months),months_add(timestamp date, bigint months)

Purpose: Returns the specified date and time plus some number of months.

Return type: timestamp

months_sub(timestamp date, int months),months_sub(timestamp date, bigint months)

Purpose: Returns the specified date and time minus some number of months.

Return type: timestamp

nanoseconds_add(timestamp date, int nanoseconds),nanoseconds_add(timestamp date, bigint
nanoseconds)

Purpose: Returns the specified date and time plus some number of nanoseconds.

Return type: timestamp

nanoseconds_sub(timestamp date, int nanoseconds),nanoseconds_sub(timestamp date, bigint
nanoseconds)

Purpose: Returns the specified date and time minus some number of nanoseconds.
```

**Return type:** timestamp

`now()`

**Purpose:** Returns the current date and time (in the local time zone) as a timestamp value.

**Return type:** timestamp

**Usage notes:**

To find a date/time value in the future or the past relative to the current date and time, add or subtract an INTERVAL expression to the return value of `now()`. See [TIMESTAMP Data Type](#) on page 130 for examples.

`second(string date)`

**Purpose:** Returns the second field from a date represented as a string.

**Return type:** int

`seconds_add(timestamp date, int seconds), seconds_add(timestamp date, bigint seconds)`

**Purpose:** Returns the specified date and time plus some number of seconds.

**Return type:** timestamp

`seconds_sub(timestamp date, int seconds), seconds_sub(timestamp date, bigint seconds)`

**Purpose:** Returns the specified date and time minus some number of seconds.

**Return type:** timestamp

`subdate(timestamp startdate, int days), subdate(timestamp startdate, bigint days),`

**Purpose:** Subtracts a specified number of days from a TIMESTAMP value. Similar to `date_sub()`, but starts with an actual TIMESTAMP value instead of a string that is converted to a TIMESTAMP.

**Return type:** timestamp

`to_date(timestamp)`

**Purpose:** Returns a string representation of the date field from a timestamp value.

**Return type:** string

`to_utc_timestamp(timestamp, string timezone)`

**Purpose:** Converts a specified timestamp value in a specified time zone into the corresponding value for the UTC time zone.

**Return type:** timestamp

**Usage notes:** Often used in combination with the `now()` function, to translate local date and time values to the UTC time zone for consistent representation on disk. The opposite of the `from_utc_timestamp()` function.

**Examples:** See discussion of time zones in [TIMESTAMP Data Type](#) on page 130 for information about using this function for conversions between the local time zone and UTC.

`trunc(timestamp, string unit)`

**Purpose:** Strips off fields from a TIMESTAMP value.

**Unit argument:** The unit argument value is case-sensitive. This argument string can be one of:

- SYYYY, YYYY, YEAR, SYEAR, YYY, YY, y: Year.
- Q: Quarter.
- MONTH, MON, MM, RM: Month.
- ww, w: Same day of the week as the first day of the month.
- DDD, DD, J: Day.
- DAY, DY, D: Starting day of the week. (Not necessarily the current day.)
- HH, HH12, HH24: Hour. A TIMESTAMP value truncated to the hour is always represented in 24-hour notation, even for the HH12 argument string.

- MI: Minute.

### Usage notes:

Typically used in `GROUP BY` queries to aggregate results from the same hour, day, week, month, quarter, and so on. You can also use this function in an `INSERT ... SELECT` into a partitioned table to divide `TIMESTAMP` values into the correct partition.

Because the return value is a `TIMESTAMP`, if you cast the result of `TRUNC()` to `STRING`, you will often see zeroed-out portions such as `00:00:00` in the time field. If you only need the individual units such as hour, day, month, or year, use the `EXTRACT()` function instead. If you need the individual units from a truncated `TIMESTAMP` value, run the `TRUNCATE()` function on the original value, then run `EXTRACT()` on the result.

**Return type:** `timestamp`

```
unix_timestamp(), unix_timestamp(string datetime), unix_timestamp(string datetime, string format), unix_timestamp(timestamp datetime)
```

**Purpose:** Returns an integer value representing the current date and time as a delta from the Unix epoch, or converts from a specified date and time value represented as a `TIMESTAMP` or `STRING`.

**Return type:** `bigint`

### Usage notes:

See `from_unixtime()` for details about the patterns you can use in the `format` string to represent the position of year, month, day, and so on in the `date` string. In Impala 1.3 and higher, you have more flexibility to switch the positions of elements and use different separator characters.

In Impala 2.2.0 and higher, built-in functions that accept or return integers representing `TIMESTAMP` values use the `BIGINT` type for parameters and return values, rather than `INT`. This change lets the date and time functions avoid an overflow error that would otherwise occur on January 19th, 2038 (known as the ["Year 2038 problem"](#) or ["Y2K38 problem"](#)). This change affects the `from_unixtime()` and `unix_timestamp()` functions. You might need to change application code that interacts with these functions, change the types of columns that store the return values, or add `CAST()` calls to SQL statements that call these functions.

`unix_timestamp()` and `from_unixtime()` are often used in combination to convert a `TIMESTAMP` value into a particular string format. For example:

```
select from_unixtime(unix_timestamp(now() + interval 3 days), 'yyyy/MM/dd HH:mm');
```

The way this function deals with time zones when converting to or from `TIMESTAMP` values is affected by the `-use_local_tz_for_unix_timestamp_conversions` startup flag for the `impalad` daemon. See [TIMESTAMP Data Type](#) on page 130 for details about how Impala handles time zone considerations for the `TIMESTAMP` data type.

```
weekofyear(string date)
```

**Purpose:** Returns the corresponding week (1-53) from a date represented as a string.

**Return type:** `int`

```
weeks_add(timestamp date, int weeks), weeks_add(timestamp date, bigint weeks)
```

**Purpose:** Returns the specified date and time plus some number of weeks.

**Return type:** `timestamp`

```
weeks_sub(timestamp date, int weeks), weeks_sub(timestamp date, bigint weeks)
```

**Purpose:** Returns the specified date and time minus some number of weeks.

**Return type:** `timestamp`

```
year(string date)
  Purpose: Returns the year field from a date represented as a string.
  Return type: int

years_add(timestamp date, int years), years_add(timestamp date, bigint years)
  Purpose: Returns the specified date and time plus some number of years.
  Return type: timestamp

years_sub(timestamp date, int years), years_sub(timestamp date, bigint years)
  Purpose: Returns the specified date and time minus some number of years.
  Return type: timestamp
```

## Impala Conditional Functions

Impala supports the following conditional functions for testing equality, comparison operators, and nullity:

`CASE a WHEN b THEN c [WHEN d THEN e]... [ELSE f] END`

**Purpose:** Compares an expression to one or more possible values, and returns a corresponding result when a match is found.

**Return type:** same as the initial argument value, except that integer values are promoted to `BIGINT` and floating-point values are promoted to `DOUBLE`; use `CAST( )` when inserting into a smaller numeric column

**Usage notes:**

In this form of the `CASE` expression, the initial value `A` being evaluated for each row it typically a column reference, or an expression involving a column. This form can only compare against a set of specified values, not ranges, multi-value comparisons such as `BETWEEN` or `IN`, regular expressions, or `NULL`.

**Examples:**

Although this example is split across multiple lines, you can put any or all parts of a `CASE` expression on a single line, with no punctuation or other separators between the `WHEN`, `ELSE`, and `END` clauses.

```
select case x
  when 1 then 'one'
  when 2 then 'two'
  when 0 then 'zero'
  else 'out of range'
end
from t1;
```

`CASE WHEN a THEN b [WHEN c THEN d]... [ELSE e] END`

**Purpose:** Tests whether any of a sequence of expressions is true, and returns a corresponding result for the first true expression.

**Return type:** same as the initial argument value, except that integer values are promoted to `BIGINT` and floating-point values are promoted to `DOUBLE`; use `CAST( )` when inserting into a smaller numeric column

**Usage notes:**

`CASE` expressions without an initial test value have more flexibility. For example, they can test different columns in different `WHEN` clauses, or use comparison operators such as `BETWEEN`, `IN` and `IS NULL` rather than comparing against discrete values.

`CASE` expressions are often the foundation of long queries that summarize and format results for easy-to-read reports. For example, you might use a `CASE` function call to turn values from a numeric column into category strings corresponding to integer values, or labels such as "Small", "Medium" and "Large" based on ranges. Then subsequent parts of the query might aggregate based on the transformed

## Impala SQL Language Reference

values, such as how many values are classified as small, medium, or large. You can also use CASE to signal problems with out-of-bounds values, NULL values, and so on.

By using operators such as OR, IN, REGEXP, and so on in CASE expressions, you can build extensive tests and transformations into a single query. Therefore, applications that construct SQL statements often rely heavily on CASE calls in the generated SQL code.

Because this flexible form of the CASE expressions allows you to perform many comparisons and call multiple functions when evaluating each row, be careful applying elaborate CASE expressions to queries that process large amounts of data. For example, when practical, evaluate and transform values through CASE after applying operations such as aggregations that reduce the size of the result set; transform numbers to strings after performing joins with the original numeric values.

### Examples:

Although this example is split across multiple lines, you can put any or all parts of a CASE expression on a single line, with no punctuation or other separators between the WHEN, ELSE, and END clauses.

```
select case
    when dayname(now()) in ('Saturday', 'Sunday') then 'result undefined on
weekends'
    when x > y then 'x greater than y'
    when x = y then 'x and y are equal'
    when x is null or y is null then 'one of the columns is null'
    else null
end
from t1;

coalesce(type v1, type v2, ...)
```

**Purpose:** Returns the first specified argument that is not NULL, or NULL if all arguments are NULL.

**Return type:** same as the initial argument value, except that integer values are promoted to BIGINT and floating-point values are promoted to DOUBLE; use CAST( ) when inserting into a smaller numeric column

```
decode(type expression, type search1, type result1 [, type search2, type result2 ...] [,,
type default] )
```

**Purpose:** Compares an expression to one or more possible values, and returns a corresponding result when a match is found.

**Return type:** same as the initial argument value, except that integer values are promoted to BIGINT and floating-point values are promoted to DOUBLE; use CAST( ) when inserting into a smaller numeric column

### Usage notes:

Can be used as shorthand for a CASE expression.

The original expression and the search expressions must of the same type or convertible types. The result expression can be a different type, but all result expressions must be of the same type.

Returns a successful match if the original expression is NULL and a search expression is also NULL. The

Returns NULL if the final default value is omitted and none of the search expressions match the original expression.

### Examples:

The following example translates numeric day values into descriptive names:

```
SELECT event, decode(day_of_week, 1, "Monday", 2, "Tuesday", 3, "Wednesday",
4, "Thursday", 5, "Friday", 6, "Saturday", 7, "Sunday", "Unknown day")
FROM calendar;
```

```
if(boolean condition, type ifTrue, type ifFalseOrNull)
```

**Purpose:** Tests an expression and returns a corresponding result depending on whether the result is true, false, or NULL.

**Return type:** Same as the ifTrue argument value

```
ifnull(type a, type ifNotNull)
```

**Purpose:** Alias for the isnull() function, with the same behavior. To simplify porting SQL with vendor extensions to Impala.

**Added in:** Impala 1.3.0

```
isfalse(boolean)
```

**Purpose:** Tests if a Boolean expression is false or not. Returns true if so. If the argument is NULL, returns false. Identical to isnottrue(), except it returns the opposite value for a NULL argument.

**Return type:** BOOLEAN

**Added in:** CDH 5.4.0 (Impala 2.2.0)

```
isnotfalse(boolean)
```

**Purpose:** Tests if a Boolean expression is not false (that is, either true or NULL). Returns true if so. If the argument is NULL, returns true. Identical to istrue(), except it returns the opposite value for a NULL argument.

**Return type:** BOOLEAN

**Usage notes:** Primarily for compatibility with code containing industry extensions to SQL.

**Added in:** CDH 5.4.0 (Impala 2.2.0)

```
isnottrue(boolean)
```

**Purpose:** Tests if a Boolean expression is not true (that is, either false or NULL). Returns true if so. If the argument is NULL, returns true. Identical to isfalse(), except it returns the opposite value for a NULL argument.

**Return type:** BOOLEAN

**Added in:** CDH 5.4.0 (Impala 2.2.0)

```
isnull(type a, type ifNotNull)
```

**Purpose:** Tests if an expression is NULL, and returns the expression result value if not. If the first argument is NULL, returns the second argument.

**Compatibility notes:** Equivalent to the nvl() function from Oracle Database or ifnull() from MySQL. The nvl() and ifnull() functions are also available in Impala.

**Return type:** Same as the first argument value

```
istrue(boolean)
```

**Purpose:** Tests if a Boolean expression is true or not. Returns true if so. If the argument is NULL, returns false. Identical to isnotfalse(), except it returns the opposite value for a NULL argument.

**Return type:** BOOLEAN

**Usage notes:** Primarily for compatibility with code containing industry extensions to SQL.

**Added in:** CDH 5.4.0 (Impala 2.2.0)

```
notnullvalue(expression)
```

**Purpose:** Tests if an expression (of any type) is NULL or not. Returns false if so. The converse of nullvalue().

**Return type:** BOOLEAN

**Usage notes:** Primarily for compatibility with code containing industry extensions to SQL.

**Added in:** CDH 5.4.0 (Impala 2.2.0)

`nullif(expr1,expr2)`

**Purpose:** Returns `NULL` if the two specified arguments are equal. If the specified arguments are not equal, returns the value of `expr1`. The data types of the expressions must be compatible, according to the conversion rules from [Data Types](#) on page 111. You cannot use an expression that evaluates to `NULL` for `expr1`; that way, you can distinguish a return value of `NULL` from an argument value of `NULL`, which would never match `expr2`.

**Usage notes:** This function is effectively shorthand for a `CASE` expression of the form:

```
CASE
  WHEN expr1 = expr2 THEN NULL
  ELSE expr1
END
```

It is commonly used in division expressions, to produce a `NULL` result instead of a divide-by-zero error when the divisor is equal to zero:

```
select 1.0 / nullif(c1,0) as reciprocal from t1;
```

You might also use it for compatibility with other database systems that support the same `NULIF()` function.

**Return type:** same as the initial argument value, except that integer values are promoted to `BIGINT` and floating-point values are promoted to `DOUBLE`; use `CAST()` when inserting into a smaller numeric column

**Added in:** Impala 1.3.0

`nullifzero(numeric_expr)`

**Purpose:** Returns `NULL` if the numeric expression evaluates to 0, otherwise returns the result of the expression.

**Usage notes:** Used to avoid error conditions such as divide-by-zero in numeric calculations. Serves as shorthand for a more elaborate `CASE` expression, to simplify porting SQL with vendor extensions to Impala.

**Return type:** same as the initial argument value, except that integer values are promoted to `BIGINT` and floating-point values are promoted to `DOUBLE`; use `CAST()` when inserting into a smaller numeric column

**Added in:** Impala 1.3.0

`nullvalue(expression)`

**Purpose:** Tests if an expression (of any type) is `NULL` or not. Returns `true` if so. The converse of `notnullvalue()`.

**Return type:** `BOOLEAN`

**Usage notes:** Primarily for compatibility with code containing industry extensions to SQL.

**Added in:** CDH 5.4.0 (Impala 2.2.0)

`nvl(type a, type ifNotNull)`

**Purpose:** Alias for the `isnull()` function. Tests if an expression is `NULL`, and returns the expression result value if not. If the first argument is `NULL`, returns the second argument. Equivalent to the `nvl()` function from Oracle Database or `ifnull()` from MySQL.

**Return type:** Same as the first argument value

**Added in:** Impala 1.1

**`zeroifnull(numeric_expr)`**

**Purpose:** Returns 0 if the numeric expression evaluates to `NULL`, otherwise returns the result of the expression.

**Usage notes:** Used to avoid unexpected results due to unexpected propagation of `NULL` values in numeric calculations. Serves as shorthand for a more elaborate `CASE` expression, to simplify porting SQL with vendor extensions to Impala.

**Return type:** same as the initial argument value, except that integer values are promoted to `BIGINT` and floating-point values are promoted to `DOUBLE`; use `CAST( )` when inserting into a smaller numeric column

**Added in:** Impala 1.3.0

## Impala String Functions

String functions are classified as those primarily accepting or returning `STRING`, `VARCHAR`, or `CHAR` data types, for example to measure the length of a string or concatenate two strings together.

- All the functions that accept `STRING` arguments also accept the `VARCHAR` and `CHAR` types introduced in Impala 2.0.
- Whenever `VARCHAR` or `CHAR` values are passed to a function that returns a string value, the return type is normalized to `STRING`. For example, a call to `concat( )` with a mix of `STRING`, `VARCHAR`, and `CHAR` arguments produces a `STRING` result.

### Related information:

The string functions operate mainly on these data types: [STRING Data Type](#) on page 128, [VARCHAR Data Type \(CDH 5.2 or higher only\)](#) on page 136, and [CHAR Data Type \(CDH 5.2 or higher only\)](#) on page 114.

### Function reference:

Impala supports the following string functions:

**`ascii(string str)`**

**Purpose:** Returns the numeric ASCII code of the first character of the argument.

**Return type:** `int`

**`char_length(string a), character_length(string a)`**

**Purpose:** Returns the length in characters of the argument string. Aliases for the `length()` function.

**Return type:** `int`

**`concat(string a, string b...)`**

**Purpose:** Returns a single string representing all the argument values joined together.

**Return type:** `string`

**Usage notes:** `concat( )` and `concat_ws( )` are appropriate for concatenating the values of multiple columns within the same row, while `group_concat( )` joins together values from different rows.

**`concat_ws(string sep, string a, string b...)`**

**Purpose:** Returns a single string representing the second and following argument values joined together, delimited by a specified separator.

**Return type:** `string`

**Usage notes:** `concat( )` and `concat_ws( )` are appropriate for concatenating the values of multiple columns within the same row, while `group_concat( )` joins together values from different rows.

**`find_in_set(string str, string strList)`**

**Purpose:** Returns the position (starting from 1) of the first occurrence of a specified string within a comma-separated string. Returns `NULL` if either argument is `NULL`, 0 if the search string is not found, or 0 if the search string contains a comma.

**Return type:** int

```
group_concat(string s [, string sep])
```

**Purpose:** Returns a single string representing the argument value concatenated together for each row of the result set. If the optional separator string is specified, the separator is added between each pair of concatenated values.

**Return type:** string

**Usage notes:** concat() and concat\_ws() are appropriate for concatenating the values of multiple columns within the same row, while group\_concat() joins together values from different rows.

By default, returns a single string covering the whole result set. To include other columns or values in the result set, or to produce multiple concatenated strings for subsets of rows, include a GROUP BY clause in the query.

```
initcap(string str)
```

**Purpose:** Returns the input string with the first letter capitalized.

**Return type:** string

```
instr(string str, string substr)
```

**Purpose:** Returns the position (starting from 1) of the first occurrence of a substring within a longer string.

**Return type:** int

```
length(string a)
```

**Purpose:** Returns the length in characters of the argument string.

**Return type:** int

```
locate(string substr, string str[, int pos])
```

**Purpose:** Returns the position (starting from 1) of the first occurrence of a substring within a longer string, optionally after a particular position.

**Return type:** int

```
lower(string a), lcase(string a)
```

**Purpose:** Returns the argument string converted to all-lowercase.

**Return type:** string

```
lpad(string str, int len, string pad)
```

**Purpose:** Returns a string of a specified length, based on the first argument string. If the specified string is too short, it is padded on the left with a repeating sequence of the characters from the pad string. If the specified string is too long, it is truncated on the right.

**Return type:** string

```
ltrim(string a)
```

**Purpose:** Returns the argument string with any leading spaces removed from the left side.

**Return type:** string

```
parse_url(string urlString, string partToExtract [, string keyToExtract])
```

**Purpose:** Returns the portion of a URL corresponding to a specified part. The part argument can be 'PROTOCOL', 'HOST', 'PATH', 'REF', 'AUTHORITY', 'FILE', 'USERINFO', or 'QUERY'. Uppercase is required for these literal values. When requesting the QUERY portion of the URL, you can optionally specify a key to retrieve just the associated value from the key-value pairs in the query string.

**Return type:** string

**Usage notes:** This function is important for the traditional Hadoop use case of interpreting web logs. For example, if the web traffic data features raw URLs not divided into separate table columns, you can count visitors to a particular page by extracting the 'PATH' or 'FILE' field, or analyze search terms by extracting the corresponding key from the 'QUERY' field.

```
regexp_extract(string subject, string pattern, int index)
```

**Purpose:** Returns the specified () group from a string based on a regular expression pattern. Group 0 refers to the entire extracted string, while group 1, 2, and so on refers to the first, second, and so on ( . . . ) portion.

**Return type:** string

In Impala 2.0 and later, the Impala regular expression syntax conforms to the POSIX Extended Regular Expression syntax used by the Google RE2 library. For details, see [the RE2 documentation](#). It has most idioms familiar from regular expressions in Perl, Python, and so on, including `.*?` for non-greedy matches.

In Impala 2.0 and later, a change in the underlying regular expression library could cause changes in the way regular expressions are interpreted by this function. Test any queries that use regular expressions and adjust the expression patterns if necessary. See [Incompatible Changes Introduced in Impala 2.0.0 / CDH 5.2.0](#) on page 489 for details.

Because the `impala-shell` interpreter uses the \ character for escaping, use `\\"` to represent the regular expression escape character in any regular expressions that you submit through `impala-shell`. You might prefer to use the equivalent character class names, such as `[[:digit:]]` instead of `\d` which you would have to escape as `\\\d`.

### Examples:

This example shows how group 0 matches the full pattern string, including the portion outside any () group:

```
[localhost:21000] > select regexp_extract('abcdef123ghi456jkl','.*?(\\\\d+)',0);
+-----+
| regexp_extract('abcdef123ghi456jkl', '.*?(\\\\d+)', 0) |
+-----+
| abcdef123ghi456
+-----+
Returned 1 row(s) in 0.11s
```

This example shows how group 1 matches just the contents inside the first () group in the pattern string:

```
[localhost:21000] > select regexp_extract('abcdef123ghi456jkl','.*?(\\\\d+)',1);
+-----+
| regexp_extract('abcdef123ghi456jkl', '.*?(\\\\d+)', 1) |
+-----+
| 456
+-----+
Returned 1 row(s) in 0.11s
```

Unlike in earlier Impala releases, the regular expression library used in Impala 2.0 and later supports the `.*?` idiom for non-greedy matches. This example shows how a pattern string starting with `.*?` matches the shortest possible portion of the source string, returning the rightmost set of lowercase letters. A pattern string both starting and ending with `.*?` finds two potential matches of equal length, and returns the first one found (the leftmost set of lowercase letters).

```
[localhost:21000] > select regexp_extract('AbcdBCdefGHI','.*?([[:lower:]]+)',1);
+-----+
| regexp_extract('abcdBCdefghi', '.*?([[:lower:]]+)', 1) |
+-----+
| def
+-----+
[localhost:21000] > select regexp_extract('AbcdBCdefGHI','.*?([[:lower:]]+).*?',1);
+-----+
| regexp_extract('abcdBCdefghi', '.*?([[:lower:]]+).*?', 1) |
+-----+
```

bcd	
+-----+	+-----+

```
regexp_replace(string initial, string pattern, string replacement)
```

**Purpose:** Returns the initial argument with the regular expression pattern replaced by the final argument string.

**Return type:** string

In Impala 2.0 and later, the Impala regular expression syntax conforms to the POSIX Extended Regular Expression syntax used by the Google RE2 library. For details, see [the RE2 documentation](#). It has most idioms familiar from regular expressions in Perl, Python, and so on, including `. *?` for non-greedy matches.

In Impala 2.0 and later, a change in the underlying regular expression library could cause changes in the way regular expressions are interpreted by this function. Test any queries that use regular expressions and adjust the expression patterns if necessary. See [Incompatible Changes Introduced in Impala 2.0.0 / CDH 5.2.0](#) on page 489 for details.

Because the `impala-shell` interpreter uses the `\` character for escaping, use `\\\` to represent the regular expression escape character in any regular expressions that you submit through `impala-shell`. You might prefer to use the equivalent character class names, such as `[:digit:]` instead of `\d` which you would have to escape as `\\\d`.

### Examples:

These examples show how you can replace parts of a string matching a pattern with replacement text, which can include backreferences to any `( )` groups in the pattern string. The backreference numbers start at 1, and any `\` characters must be escaped as `\\\`.

Replace a character pattern with new text:

```
[localhost:21000] > select regexp_replace('aaabbbaaa', 'b+', 'xyz');
+-----+
| regexp_replace('aaabbbaaa', 'b+', 'xyz') |
+-----+
| aaaxyzaaa
+-----+
Returned 1 row(s) in 0.11s
```

Replace a character pattern with substitution text that includes the original matching text:

```
[localhost:21000] > select regexp_replace('aaabbbaaa', '(b+)', '<\\1>');
+-----+
| regexp_replace('aaabbbaaa', '(b+)', '<\\1>') |
+-----+
| aaa<bbb>aaa
+-----+
Returned 1 row(s) in 0.11s
```

Remove all characters that are not digits:

```
[localhost:21000] > select regexp_replace('123-456-789', '[^[:digit:]]', '');
+-----+
| regexp_replace('123-456-789', '[^[:digit:]]', '') |
+-----+
| 123456789
+-----+
Returned 1 row(s) in 0.12s
```

```
repeat(string str, int n)
```

**Purpose:** Returns the argument string repeated a specified number of times.

**Return type:** string

```
reverse(string a)
Purpose: Returns the argument string with characters in reversed order.

Return type: string

rpad(string str, int len, string pad)
Purpose: Returns a string of a specified length, based on the first argument string. If the specified string is too short, it is padded on the right with a repeating sequence of the characters from the pad string. If the specified string is too long, it is truncated on the right.

Return type: string

rtrim(string a)
Purpose: Returns the argument string with any trailing spaces removed from the right side.

Return type: string

space(int n)
Purpose: Returns a concatenated string of the specified number of spaces. Shorthand for repeat(' ', n).

Return type: string

strleft(string a, int num_chars)
Purpose: Returns the leftmost characters of the string. Shorthand for a call to substr() with 2 arguments.

Return type: string

strright(string a, int num_chars)
Purpose: Returns the rightmost characters of the string. Shorthand for a call to substr() with 2 arguments.

Return type: string

substr(string a, int start [, int len]), substring(string a, int start [, int len])
Purpose: Returns the portion of the string starting at a specified point, optionally with a specified maximum length. The characters in the string are indexed starting at 1.

Return type: string

translate(string input, string from, string to)
Purpose: Returns the input string with a set of characters replaced by another set of characters.

Return type: string

trim(string a)
Purpose: Returns the input string with both leading and trailing spaces removed. The same as passing the string through both ltrim() and rtrim().

Return type: string

upper(string a), ucase(string a)
Purpose: Returns the argument string converted to all-uppercase.

Return type: string
```

### Impala Miscellaneous Functions

Impala supports the following utility functions that do not operate on a particular column or data type:

`current_database()`

**Purpose:** Returns the database that the session is currently using, either default if no database has been selected, or whatever database the session switched to through a `USE` statement or the `impalad-d` option.

**Return type:** `string`

`pid()`

**Purpose:** Returns the process ID of the `impalad` daemon that the session is connected to. You can use it during low-level debugging, to issue Linux commands that trace, show the arguments, and so on on the `impalad` process.

**Return type:** `int`

`user()`

**Purpose:** Returns the username of the Linux user who is connected to the `impalad` daemon. Typically called a single time, in a query without any `FROM` clause, to understand how authorization settings apply in a security context; once you know the logged-in user name, you can check which groups that user belongs to, and from the list of groups you can check which roles are available to those groups through the authorization policy file.

In Impala 2.0 and later, `user()` returns the full Kerberos principal string, such as `user@example.com`, in a Kerberized environment.

**Return type:** `string`

`version()`

**Purpose:** Returns information such as the precise version number and build date for the `impalad` daemon that you are currently connected to. Typically used to confirm that you are connected to the expected level of Impala to use a particular feature, or to connect to several nodes and confirm they are all running the same level of `impalad`.

**Return type:** `string` (with one or more embedded newlines)

### Impala Aggregate Functions

Aggregate functions are a special category with different rules. These functions calculate a return value across all the items in a result set, so they require a `FROM` clause in the query:

```
select count(product_id) from product_catalog;
select max(height), avg(height) from census_data where age > 20;
```

Aggregate functions also ignore `NULL` values rather than returning a `NULL` result. For example, if some rows have `NULL` for a particular column, those rows are ignored when computing the `AVG()` for that column. Likewise, specifying `COUNT(col_name)` in a query counts only those rows where *col\_name* contains a non-`NULL` value.

### APPX\_MEDIAN Function

An aggregate function that returns a value that is approximately the median (midpoint) of values in the set of input values.

**Syntax:**

```
APPX_MEDIAN([DISTINCT | ALL] expression)
```

This function works with any input type, because the only requirement is that the type supports less-than and greater-than comparison operators.

**Usage notes:**

Because the return value represents the estimated midpoint, it might not reflect the precise midpoint value, especially if the cardinality of the input values is very high. If the cardinality is low (up to approximately 20,000), the result is more accurate because the sampling considers all or almost all of the different values.

**Return type:** Same as the input value, except for CHAR and VARCHAR arguments which produce a STRING result  
The return value is always the same as one of the input values, not an "in-between" value produced by averaging.

**Restrictions:**

The OVER( ) clause is not allowed at all with this function.

**Examples:**

The following example uses a table of a million random floating-point numbers ranging up to approximately 50,000. The average is approximately 25,000. Because of the random distribution, we would expect the median to be close to this same number. Computing the precise median is a more intensive operation than computing the average, because it requires keeping track of every distinct value and how many times each occurs. The APPX\_MEDIAN( ) function uses a sampling algorithm to return an approximate result, which in this case is close to the expected value. To make sure that the value is not substantially out of range due to a skewed distribution, subsequent queries confirm that there are approximately 500,000 values higher than the APPX\_MEDIAN( ) value, and approximately 500,000 values lower than the APPX\_MEDIAN( ) value.

```
[localhost:21000] > select min(x), max(x), avg(x) from million_numbers;
+-----+-----+-----+
| min(x) | max(x) | avg(x) |
+-----+-----+-----+
| 4.725693727250069 | 49994.56852674231 | 24945.38563793553 |
+-----+-----+-----+
[localhost:21000] > select appx_median(x) from million_numbers;
+-----+
| appx_median(x) |
+-----+
| 24721.6 |
+-----+
[localhost:21000] > select count(x) as higher from million_numbers where x > (select
appx_median(x) from million_numbers);
+-----+
| higher |
+-----+
| 502013 |
+-----+
[localhost:21000] > select count(x) as lower from million_numbers where x < (select
appx_median(x) from million_numbers);
+-----+
| lower |
+-----+
| 497987 |
+-----+
```

The following example computes the approximate median using a subset of the values from the table, and then confirms that the result is a reasonable estimate for the midpoint.

```
[localhost:21000] > select appx_median(x) from million_numbers where x between 1000
and 5000;
+-----+
| appx_median(x) |
+-----+
| 3013.107787358159 |
+-----+
[localhost:21000] > select count(x) as higher from million_numbers where x between 1000
and 5000 and x > 3013.107787358159;
+-----+
| higher |
+-----+
| 37692 |
+-----+
```

```
[localhost:21000] > select count(x) as lower from million_numbers where x between 1000  
and 5000 and x < 3013.107787358159;  
+-----+  
| lower |  
+-----+  
| 37089 |  
+-----+
```

### AVG Function

An aggregate function that returns the average value from a set of numbers or `TIMESTAMP` values. Its single argument can be numeric column, or the numeric result of a function or expression applied to the column value. Rows with a `NULL` value for the specified column are ignored. If the table is empty, or all the values supplied to `AVG` are `NULL`, `AVG` returns `NULL`.

#### Syntax:

```
AVG([DISTINCT | ALL] expression) [OVER (analytic_clause)]
```

When the query contains a `GROUP BY` clause, returns one value for each combination of grouping values.

**Return type:** `DOUBLE` for numeric values; `TIMESTAMP` for `TIMESTAMP` values

#### Examples:

```
-- Average all the non-NULL values in a column.  
insert overwrite avg_t values (2),(4),(6),(null),(null);  
-- The average of the above values is 4: (2+4+6) / 3. The 2 NULL values are ignored.  
select avg(x) from avg_t;  
-- Average only certain values from the column.  
select avg(x) from t1 where month = 'January' and year = '2013';  
-- Apply a calculation to the value of the column before averaging.  
select avg(x/3) from t1;  
-- Apply a function to the value of the column before averaging.  
-- Here we are substituting a value of 0 for all NULLs in the column,  
-- so that those rows do factor into the return value.  
select avg(isnull(x,0)) from t1;  
-- Apply some number-returning function to a string column and average the results.  
-- If column s contains any NULLs, length(s) also returns NULL and those rows are  
ignored.  
select avg(length(s)) from t1;  
-- Can also be used in combination with DISTINCT and/or GROUP BY.  
-- Return more than one result.  
select month, year, avg(page_visits) from web_stats group by month, year;  
-- Filter the input to eliminate duplicates before performing the calculation.  
select avg(distinct x) from t1;  
-- Filter the output after performing the calculation.  
select avg(x) from t1 group by y having avg(x) between 1 and 20;
```

The following example shows how to use `AVG()` in an analytic context. It uses a table containing integers from 1 to 10. Notice how the `AVG()` is reported for each input value, as opposed to the `GROUP BY` clause which condenses the result set.

```
select x, property, avg(x) over (partition by property) as avg from int_t where property  
in ('odd','even');  
+-----+  
| x   | property | avg  |  
+-----+  
| 2   | even    | 6    |  
| 4   | even    | 6    |  
| 6   | even    | 6    |  
| 8   | even    | 6    |  
| 10  | even    | 6    |  
| 1   | odd     | 5    |  
| 3   | odd     | 5    |  
| 5   | odd     | 5    |  
| 7   | odd     | 5    |
```

9	odd	5
---	-----	---

**Restrictions:**

Due to the way arithmetic on `FLOAT` and `DOUBLE` columns uses high-performance hardware instructions, and distributed queries can perform these operations in different order for each query, results can vary slightly for aggregate function calls such as `SUM()` and `AVG()` for `FLOAT` and `DOUBLE` columns, particularly on large data sets where millions or billions of values are summed or averaged. For perfect consistency and repeatability, use the `DECIMAL` data type for such operations instead of `FLOAT` or `DOUBLE`.

**Related information:**

[Impala Analytic Functions](#) on page 296, [MAX Function](#) on page 289, [MIN Function](#) on page 290

## COUNT Function

An aggregate function that returns the number of rows, or the number of non-NULL rows.

**Syntax:**

```
COUNT([DISTINCT | ALL] expression) [OVER (analytic_clause)]
```

Depending on the argument, `COUNT()` considers rows that meet certain conditions:

- The notation `COUNT(*)` includes `NULL` values in the total.
- The notation `COUNT(column_name)` only considers rows where the column contains a non-`NULL` value.
- You can also combine `COUNT` with the `DISTINCT` operator to eliminate duplicates before counting, and to count the combinations of values across multiple columns.

When the query contains a `GROUP BY` clause, returns one value for each combination of grouping values.

**Return type:** `BIGINT`

**Examples:**

```
-- How many rows total are in the table, regardless of NULL values?
select count(*) from t1;
-- How many rows are in the table with non-NULL values for a column?
select count(c1) from t1;
-- Count the rows that meet certain conditions.
-- Again, * includes NULLs, so COUNT(*) might be greater than COUNT(col).
select count(*) from t1 where x > 10;
select count(c1) from t1 where x > 10;
-- Can also be used in combination with DISTINCT and/or GROUP BY.
-- Combine COUNT and DISTINCT to find the number of unique values.
-- Must use column names rather than * with COUNT(DISTINCT ...) syntax.
-- Rows with NULL values are not counted.
select count(distinct c1) from t1;
-- Rows with a NULL value in _either_ column are not counted.
select count(distinct c1, c2) from t1;
-- Return more than one result.
select month, year, count(distinct visitor_id) from web_stats group by month, year;
```

The following example shows how to use `COUNT()` in an analytic context. It uses a table containing integers from 1 to 10. Notice how the `COUNT()` is reported for each input value, as opposed to the `GROUP BY` clause which condenses the result set.

```
select x, property, count(x) over (partition by property) as count from int_t where
property in ('odd','even');
+-----+-----+
| x   | property | count |
+-----+-----+
| 2   | even     | 5      |
| 4   | even     | 5      |
| 6   | even     | 5      |
```

8	even	5
10	even	5
1	odd	5
3	odd	5
5	odd	5
7	odd	5
9	odd	5

### Related information:

[Impala Analytic Functions](#) on page 296

#### Note:

By default, Impala only allows a single COUNT(DISTINCT *columns*) expression in each query.

If you do not need precise accuracy, you can produce an estimate of the distinct values for a column by specifying NDV(*column*); a query can contain multiple instances of NDV(*column*). To make Impala automatically rewrite COUNT(DISTINCT) expressions to NDV(), enable the APPX\_COUNT\_DISTINCT query option.

To produce the same result as multiple COUNT(DISTINCT) expressions, you can use the following technique for queries involving a single table:

```
select v1.c1 result1, v2.c1 result2 from
  (select count(distinct col1) as c1 from t1) v1
  cross join
  (select count(distinct col2) as c1 from t1) v2;
```

Because CROSS JOIN is an expensive operation, prefer to use the NDV() technique wherever practical.

## GROUP\_CONCAT Function

An aggregate function that returns a single string representing the argument value concatenated together for each row of the result set. If the optional separator string is specified, the separator is added between each pair of concatenated values. The default separator is a comma followed by a space.

#### Syntax:

```
GROUP_CONCAT([ALL] expression [, separator])
```

**Usage notes:** concat() and concat\_ws() are appropriate for concatenating the values of multiple columns within the same row, while group\_concat() joins together values from different rows.

By default, returns a single string covering the whole result set. To include other columns or values in the result set, or to produce multiple concatenated strings for subsets of rows, include a GROUP BY clause in the query.

**Return type:** STRING

#### Restrictions:

You cannot apply the DISTINCT operator to the argument of this function.

This function does not support the OVER clause, and so cannot be used as an analytic function.

Currently, Impala returns an error if the result value grows larger than 1 GiB.

#### Examples:

The following examples illustrate various aspects of the GROUP\_CONCAT() function.

You can call the function directly on a `STRING` column. To use it with a numeric column, cast the value to `STRING`.

```
[localhost:21000] > create table t1 (x int, s string);
[localhost:21000] > insert into t1 values (1, "one"), (3, "three"), (2, "two"), (1, "one");
[localhost:21000] > select group_concat(s) from t1;
+-----+
| group_concat(s) |
+-----+
| one, three, two, one |
+-----+
[localhost:21000] > select group_concat(cast(x as string)) from t1;
+-----+
| group_concat(cast(x as string)) |
+-----+
| 1, 3, 2, 1 |
+-----+
```

The optional separator lets you format the result in flexible ways. The separator can be an arbitrary string expression, not just a single character.

```
[localhost:21000] > select group_concat(s, '|') from t1;
+-----+
| group_concat(s, '|') |
+-----+
| one|three|two|one |
+-----+
[localhost:21000] > select group_concat(s,'---') from t1;
+-----+
| group_concat(s, '---') |
+-----+
| one---three---two---one |
+-----+
```

The default separator is a comma followed by a space. To get a comma-delimited result without extra spaces, specify a delimiter character that is only a comma.

```
[localhost:21000] > select group_concat(s,',') from t1;
+-----+
| group_concat(s, ',') |
+-----+
| one,three,two,one |
+-----+
```

Including a `GROUP BY` clause lets you produce a different concatenated result for each group in the result set. In this example, the only `x` value that occurs more than once is `1`, so that is the only row in the result set where `GROUP_CONCAT()` returns a delimited value. For groups containing a single value, `GROUP_CONCAT()` returns the original value of its `STRING` argument.

```
[localhost:21000] > select x, group_concat(s) from t1 group by x;
+-----+
| x | group_concat(s) |
+-----+
| 2 | two |
| 3 | three |
| 1 | one, one |
+-----+
```

## MAX Function

An aggregate function that returns the maximum value from a set of numbers. Opposite of the `MIN` function. Its single argument can be numeric column, or the numeric result of a function or expression applied to the column value. Rows with a `NULL` value for the specified column are ignored. If the table is empty, or all the values supplied to `MAX` are `NULL`, `MAX` returns `NULL`.

### Syntax:

```
MAX([DISTINCT | ALL] expression) [OVER (analytic_clause)]
```

When the query contains a GROUP BY clause, returns one value for each combination of grouping values.

**Restrictions:** In Impala 2.0 and higher, this function can be used as an analytic function, but with restrictions on any window clause. For MAX() and MIN(), the window clause is only allowed if the start bound is UNBOUNDED PRECEDING

**Return type:** Same as the input value, except for CHAR and VARCHAR arguments which produce a STRING result

### Examples:

```
-- Find the largest value for this column in the table.  
select max(c1) from t1;  
-- Find the largest value for this column from a subset of the table.  
select max(c1) from t1 where month = 'January' and year = '2013';  
-- Find the largest value from a set of numeric function results.  
select max(length(s)) from t1;  
-- Can also be used in combination with DISTINCT and/or GROUP BY.  
-- Return more than one result.  
select month, year, max(purchase_price) from store_stats group by month, year;  
-- Filter the input to eliminate duplicates before performing the calculation.  
select max(distinct x) from t1;
```

The following example shows how to use MAX() in an analytic context. It uses a table containing integers from 1 to 10. Notice how the MAX() is reported for each input value, as opposed to the GROUP BY clause which condenses the result set.

```
select x, property, max(x) over (partition by property) as max from int_t where property  
in ('odd','even');  
+-----+-----+  
| x | property | max |  
+-----+-----+  
| 2 | even     | 10 |  
| 4 | even     | 10 |  
| 6 | even     | 10 |  
| 8 | even     | 10 |  
| 10| even    | 10 |  
| 1 | odd      | 9 |  
| 3 | odd      | 9 |  
| 5 | odd      | 9 |  
| 7 | odd      | 9 |  
| 9 | odd      | 9 |  
+-----+-----+
```

### Related information:

[Impala Analytic Functions](#) on page 296, [MIN Function](#) on page 290, [AVG Function](#) on page 286

## MIN Function

An aggregate function that returns the minimum value from a set of numbers. Opposite of the MAX function. Its single argument can be numeric column, or the numeric result of a function or expression applied to the column value. Rows with a NULL value for the specified column are ignored. If the table is empty, or all the values supplied to MIN are NULL, MIN returns NULL.

### Syntax:

```
MIN([DISTINCT | ALL] expression) [OVER (analytic_clause)]
```

When the query contains a GROUP BY clause, returns one value for each combination of grouping values.

**Restrictions:** In Impala 2.0 and higher, this function can be used as an analytic function, but with restrictions on any window clause. For `MAX()` and `MIN()`, the window clause is only allowed if the start bound is `UNBOUNDED PRECEDING`.

**Return type:** Same as the input value, except for `CHAR` and `VARCHAR` arguments which produce a `STRING` result.

#### Examples:

```
-- Find the smallest value for this column in the table.
select min(c1) from t1;
-- Find the smallest value for this column from a subset of the table.
select min(c1) from t1 where month = 'January' and year = '2013';
-- Find the smallest value from a set of numeric function results.
select min(length(s)) from t1;
-- Can also be used in combination with DISTINCT and/or GROUP BY.
-- Return more than one result.
select month, year, min(purchase_price) from store_stats group by month, year;
-- Filter the input to eliminate duplicates before performing the calculation.
select min(distinct x) from t1;
```

The following example shows how to use `MIN()` in an analytic context. It uses a table containing integers from 1 to 10. Notice how the `MIN()` is reported for each input value, as opposed to the `GROUP BY` clause which condenses the result set.

```
select x, property, min(x) over (partition by property) as min from int_t where property
in ('odd','even');
+-----+
| x | property | min |
+-----+
| 2 | even     | 2 |
| 4 | even     | 2 |
| 6 | even     | 2 |
| 8 | even     | 2 |
| 10| even    | 2 |
| 1 | odd      | 1 |
| 3 | odd      | 1 |
| 5 | odd      | 1 |
| 7 | odd      | 1 |
| 9 | odd      | 1 |
+-----+
```

#### Related information:

[Impala Analytic Functions](#) on page 296, [MAX Function](#) on page 289, [AVG Function](#) on page 286

## NDV Function

An aggregate function that returns an approximate value similar to the result of `COUNT(DISTINCT col)`, the "number of distinct values". It is much faster than the combination of `COUNT` and `DISTINCT`, and uses a constant amount of memory and thus is less memory-intensive for columns with high cardinality.

#### Syntax:

```
NDV([DISTINCT | ALL] expression)
```

This is the mechanism used internally by the `COMPUTE STATS` statement for computing the number of distinct values in a column.

#### Usage notes:

Because this number is an estimate, it might not reflect the precise number of different values in the column, especially if the cardinality is very low or very high. If the estimated number is higher than the number of rows in the table, Impala adjusts the value internally during query planning.

**Return type:** `DOUBLE` in Impala 2.0 and higher; `STRING` in earlier releases

#### Restrictions:

## Impala SQL Language Reference

The OVER( ) clause is not allowed at all with this function.

### Examples:

The following example queries a billion-row table to illustrate the relative performance of COUNT(DISTINCT) and NDV( ). It shows how COUNT(DISTINCT) gives a precise answer, but is inefficient for large-scale data where an approximate result is sufficient. The NDV( ) function gives an approximate result but is much faster.

```
select count(distinct col1) from sample_data;
+-----+
| count(distinct col1)|
+-----+
| 100000
+-----+
Fetched 1 row(s) in 20.13s

select cast(ndv(col1) as bigint) as col1 from sample_data;
+-----+
| col1 |
+-----+
| 139017 |
+-----+
Fetched 1 row(s) in 8.91s
```

The following example shows how you can code multiple NDV( ) calls in a single query, to easily learn which columns have substantially more or fewer distinct values. This technique is faster than running a sequence of queries with COUNT(DISTINCT) calls.

```
select cast(ndv(col1) as bigint) as col1, cast(ndv(col2) as bigint) as col2,
       cast(ndv(col3) as bigint) as col3, cast(ndv(col4) as bigint) as col4
  from sample_data;
+-----+-----+-----+-----+
| col1 | col2 | col3 | col4 |
+-----+-----+-----+-----+
| 139017 | 282 | 46 | 145636240 |
+-----+-----+-----+-----+
Fetched 1 row(s) in 34.97s

select count(distinct col1) from sample_data;
+-----+
| count(distinct col1)|
+-----+
| 100000
+-----+
Fetched 1 row(s) in 20.13s

select count(distinct col2) from sample_data;
+-----+
| count(distinct col2) |
+-----+
| 278
+-----+
Fetched 1 row(s) in 20.09s

select count(distinct col3) from sample_data;
+-----+
| count(distinct col3) |
+-----+
| 46
+-----+
Fetched 1 row(s) in 19.12s

select count(distinct col4) from sample_data;
+-----+
| count(distinct col4) |
+-----+
| 147135880
+-----+
Fetched 1 row(s) in 266.95s
```

## STDDEV, STDDEV\_SAMP, STDDEV\_POP Functions

An aggregate function that [standard deviation](#) of a set of numbers.

### Syntax:

```
{ STDDEV | STDDEV_SAMP | STDDEV_POP } ([DISTINCT | ALL] expression)
```

This function works with any numeric data type.

**Return type:** DOUBLE in Impala 2.0 and higher; STRING in earlier releases

This function is typically used in mathematical formulas related to probability distributions.

The `STDDEV_POP()` and `STDDEV_SAMP()` functions compute the population standard deviation and sample standard deviation, respectively, of the input values. (`STDDEV()` is an alias for `STDDEV_SAMP()`.) Both functions evaluate all input rows matched by the query. The difference is that `STDDEV_SAMP()` is scaled by  $1/(N-1)$  while `STDDEV_POP()` is scaled by  $1/N$ .

If no input rows match the query, the result of any of these functions is `NULL`. If a single input row matches the query, the result of any of these functions is "0.0".

### Examples:

This example demonstrates how `STDDEV()` and `STDDEV_SAMP()` return the same result, while `STDDEV_POP()` uses a slightly different calculation to reflect that the input data is considered part of a larger "population".

```
[localhost:21000] > select stddev(score) from test_scores;
+-----+
| stddev(score) |
+-----+
| 28.5          |
+-----+
[localhost:21000] > select stddev_samp(score) from test_scores;
+-----+
| stddev_samp(score) |
+-----+
| 28.5            |
+-----+
[localhost:21000] > select stddev_pop(score) from test_scores;
+-----+
| stddev_pop(score) |
+-----+
| 28.4858         |
+-----+
```

This example demonstrates that, because the return value of these aggregate functions is a STRING, you must currently convert the result with `CAST`.

```
[localhost:21000] > create table score_stats as select cast(stddev(score) as decimal(7,4)) `standard_deviation`, cast(variance(score) as decimal(7,4)) `variance` from test_scores;
+-----+
| summary |
+-----+
| Inserted 1 row(s) |
+-----+
[localhost:21000] > desc score_stats;
+-----+-----+-----+
| name      | type       | comment |
+-----+-----+-----+
| standard_deviation | decimal(7,4) |           |
| variance     | decimal(7,4) |           |
+-----+-----+-----+
```

**Restrictions:** This function cannot be used as an analytic function; it does not currently support the `OVER()` clause.

### Related information:

The `STDDEV()`, `STDDEV_POP()`, and `STDDEV_SAMP()` functions compute the standard deviation (square root of the variance) based on the results of `VARIANCE()`, `VARIANCE_POP()`, and `VARIANCE_SAMP()` respectively. See [VARIANCE, VARIANCE\\_SAMP, VARIANCE\\_POP, VAR\\_SAMP, VAR\\_POP Functions](#) on page 295 for details about the variance property.

### SUM Function

An aggregate function that returns the sum of a set of numbers. Its single argument can be numeric column, or the numeric result of a function or expression applied to the column value. Rows with a `NULL` value for the specified column are ignored. If the table is empty, or all the values supplied to `MIN` are `NULL`, `SUM` returns `NULL`.

#### Syntax:

```
SUM([DISTINCT | ALL] expression) [OVER (analytic_clause)]
```

When the query contains a `GROUP BY` clause, returns one value for each combination of grouping values.

**Return type:** `BIGINT` for integer arguments, `DOUBLE` for floating-point arguments

#### Examples:

The following example shows how to use `SUM()` to compute the total for all the values in the table, a subset of values, or the sum for each combination of values in the `GROUP BY` clause:

```
-- Total all the values for this column in the table.  
select sum(c1) from t1;  
-- Find the total for this column from a subset of the table.  
select sum(c1) from t1 where month = 'January' and year = '2013';  
-- Find the total from a set of numeric function results.  
select sum(length(s)) from t1;  
-- Often used with functions that return predefined values to compute a score.  
select sum(case when grade = 'A' then 1.0 when grade = 'B' then 0.75 else 0) as  
class_honors from test_scores;  
-- Can also be used in combination with DISTINCT and/or GROUP BY.  
-- Return more than one result.  
select month, year, sum(purchase_price) from store_stats group by month, year;  
-- Filter the input to eliminate duplicates before performing the calculation.  
select sum(distinct x) from t1;
```

The following example shows how to use `SUM()` in an analytic context. It uses a table containing integers from 1 to 10. Notice how the `SUM()` is reported for each input value, as opposed to the `GROUP BY` clause which condenses the result set.

```
select x, property, sum(x) over (partition by property) as sum from int_t where property  
in ('odd','even');  
+-----+-----+  
| x   | property | sum  |  
+-----+-----+  
| 2   | even    | 30   |  
| 4   | even    | 30   |  
| 6   | even    | 30   |  
| 8   | even    | 30   |  
| 10  | even    | 30   |  
| 1   | odd     | 25   |  
| 3   | odd     | 25   |  
| 5   | odd     | 25   |  
| 7   | odd     | 25   |  
| 9   | odd     | 25   |  
+-----+-----+
```

#### Related information:

[Impala Analytic Functions](#) on page 296

#### Restrictions:

Due to the way arithmetic on `FLOAT` and `DOUBLE` columns uses high-performance hardware instructions, and distributed queries can perform these operations in different order for each query, results can vary slightly for

aggregate function calls such as `SUM()` and `AVG()` for `FLOAT` and `DOUBLE` columns, particularly on large data sets where millions or billions of values are summed or averaged. For perfect consistency and repeatability, use the `DECIMAL` data type for such operations instead of `FLOAT` or `DOUBLE`.

## VARIANCE, VARIANCE\_SAMP, VARIANCE\_POP, VAR\_SAMP, VAR\_POP Functions

An aggregate function that returns the [variance](#) of a set of numbers. This is a mathematical property that signifies how far the values spread apart from the mean. The return value can be zero (if the input is a single value, or a set of identical values), or a positive number otherwise.

### Syntax:

```
{ VARIANCE | VAR[IANCE]_SAMP | VAR[IANCE]_POP } ([DISTINCT | ALL] expression)
```

This function works with any numeric data type.

**Return type:** `DOUBLE` in Impala 2.0 and higher; `STRING` in earlier releases

This function is typically used in mathematical formulas related to probability distributions.

The `VARIANCE_SAMP()` and `VARIANCE_POP()` functions compute the sample variance and population variance, respectively, of the input values. (`VARIANCE()` is an alias for `VARIANCE_SAMP()`.) Both functions evaluate all input rows matched by the query. The difference is that `STDDEV_SAMP()` is scaled by  $1/(N-1)$  while `STDDEV_POP()` is scaled by  $1/N$ .

The functions `VAR_SAMP()` and `VAR_POP()` are the same as `VARIANCE_SAMP()` and `VARIANCE_POP()`, respectively. These aliases are available in Impala 2.0 and later.

If no input rows match the query, the result of any of these functions is `NULL`. If a single input row matches the query, the result of any of these functions is "0.0".

### Examples:

This example demonstrates how `VARIANCE()` and `VARIANCE_SAMP()` return the same result, while `VARIANCE_POP()` uses a slightly different calculation to reflect that the input data is considered part of a larger "population".

```
[localhost:21000] > select variance(score) from test_scores;
+-----+
| variance(score) |
+-----+
| 812.25 |
+-----+
[localhost:21000] > select variance_samp(score) from test_scores;
+-----+
| variance_samp(score) |
+-----+
| 812.25 |
+-----+
[localhost:21000] > select variance_pop(score) from test_scores;
+-----+
| variance_pop(score) |
+-----+
| 811.438 |
+-----+
```

This example demonstrates that, because the return value of these aggregate functions is a `STRING`, you convert the result with `CAST` if you need to do further calculations as a numeric value.

```
[localhost:21000] > create table score_stats as select cast(stddev(score) as decimal(7,4)) `standard_deviation`, cast(variance(score) as decimal(7,4)) `variance` from test_scores;
+-----+
| summary |
+-----+
| Inserted 1 row(s) |
+-----+
```

```
[localhost:21000] > desc score_stats;
+-----+-----+-----+
| name | type | comment |
+-----+-----+-----+
| standard_deviation | decimal(7,4) |
| variance | decimal(7,4) |
+-----+-----+
```

**Restrictions:** This function cannot be used as an analytic function; it does not currently support the `OVER()` clause.

### Related information:

The `STDDEV()`, `STDDEV_POP()`, and `STDDEV_SAMP()` functions compute the standard deviation (square root of the variance) based on the results of `VARIANCE()`, `VARIANCE_POP()`, and `VARIANCE_SAMP()` respectively. See [STDDEV, STDDEV\\_SAMP, STDDEV\\_POP Functions](#) on page 293 for details about the standard deviation property.

## Impala Analytic Functions

Analytic functions (also known as window functions) are a special category of built-in functions. Like aggregate functions, they examine the contents of multiple input rows to compute each output value. However, rather than being limited to one result value per `GROUP BY` group, they operate on **windows** where the input rows are ordered and grouped using flexible conditions expressed through an `OVER()` clause.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

Some functions, such as `LAG()` and `RANK()`, can only be used in this analytic context. Some aggregate functions do double duty: when you call the aggregation functions such as `MAX()`, `SUM()`, `AVG()`, and so on with an `OVER()` clause, they produce an output value for each row, based on computations across other rows in the window.

Although analytic functions often compute the same value you would see from an aggregate function in a `GROUP BY` query, the analytic functions produce a value for each row in the result set rather than a single value for each group. This flexibility lets you include additional columns in the `SELECT` list, offering more opportunities for organizing and filtering the result set.

Analytic function calls are only allowed in the `SELECT` list and in the outermost `ORDER BY` clause of the query. During query processing, analytic functions are evaluated after other query stages such as joins, `WHERE`, and `GROUP BY`,

The rows that are part of each partition are analyzed by computations across an ordered or unordered set of rows. For example, `COUNT()` and `SUM()` might be applied to all the rows in the partition, in which case the order of analysis does not matter. The `ORDER BY` clause might be used inside the `OVER()` clause to defines the ordering that applies to functions such as `LAG()` and `FIRST_VALUE()`.

Analytic functions are frequently used in fields such as finance and science to provide trend, outlier, and bucketed analysis for large data sets. You might also see the term "window functions" in database literature, referring to the sequence of rows (the "window") that the function call applies to, particularly when the `OVER` clause includes a `ROWS` or `RANGE` keyword.

The following sections describe the analytic query clauses and the pure analytic functions provided by Impala. For usage information about aggregate functions in an analytic context, see [Impala Aggregate Functions](#) on page 284.

### OVER Clause

The `OVER` clause is required for calls to pure analytic functions such as `LEAD()`, `RANK()`, and `FIRST_VALUE()`. When you include an `OVER` clause with calls to aggregate functions such as `MAX()`, `COUNT()`, or `SUM()`, they operate as analytic functions.

#### Syntax:

```
function(args) OVER([partition_by_clause] [order_by_clause [window_clause]])
```

```

partition_by_clause ::= PARTITION BY expr [, expr ...]
order_by_clause ::= ORDER BY expr [ASC | DESC] [NULLS FIRST | NULLS LAST] [, expr [ASC
| DESC] [NULLS FIRST | NULLS LAST] ...]
window_clause: See Window Clause

```

**PARTITION BY clause:**

The `PARTITION BY` clause acts much like the `GROUP BY` clause in the outermost block of a query. It divides the rows into groups containing identical values in one or more columns. These logical groups are known as ***partitions***. Throughout the discussion of analytic functions, “partitions” refers to the groups produced by the `PARTITION BY` clause, not to partitioned tables.

Analytic operations such as `MAX()` or `ROW_NUMBER()` then apply to each partition independently. Omit the `PARTITION BY` clause to apply the analytic operation to all the rows in the table.

**ORDER BY clause:**

The `ORDER BY` clause works much like the `ORDER BY` clause in the outermost block of a query. It defines the order in which rows are evaluated for the entire input set, or for each group produced by a `PARTITION BY` clause. You can order by one or multiple expressions, and for each expression optionally choose ascending or descending order and whether nulls come first or last in the sort order. Because this `ORDER BY` clause only defines the order in which rows are evaluated, if you want the results to be output in a specific order, also include an `ORDER BY` clause in the outer block of the query.

When the `ORDER BY` clause is omitted, the analytic function applies to all items in the group produced by the `PARTITION BY` clause. When the `ORDER BY` clause is included, the analysis can apply to all or a subset of the items in the group, depending on the optional window clause.

The order in which the rows are analyzed is only defined for those columns specified in `ORDER BY` clauses.

One difference between the analytic and outer uses of the `ORDER BY` clause: inside the `OVER` clause, `ORDER BY 1` or other integer value is interpreted as a constant sort value (effectively a no-op) rather than referring to column 1.

**Window clause:**

The window clause is only allowed in combination with an `ORDER BY` clause. If the `ORDER BY` clause is specified but the window clause is not, the default window is `RANGE BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW`. See [Window Clause](#) on page 298 for full details.

**HBase considerations:**

Because HBase tables are optimized for single-row lookups rather than full scans, analytic functions using the `OVER()` clause are not recommended for HBase tables. Although such queries work, their performance is lower than on comparable tables using HDFS data files.

**Parquet considerations:**

Analytic functions are very efficient for Parquet tables. The data that is examined during evaluation of the `OVER()` clause comes from a specified set of columns, and the values for each column are arranged sequentially within each data file.

**Text table considerations:**

Analytic functions are convenient to use with text tables for exploratory business intelligence. When the volume of data is substantial, prefer to use Parquet tables for performance-critical analytic queries.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

**Examples:**

The following example shows how to synthesize a numeric sequence corresponding to all the rows in a table. The new table has the same columns as the old one, plus an additional column `ID` containing the integers 1, 2, 3, and so on, corresponding to the order of a `TIMESTAMP` column in the original table.

```
CREATE TABLE events_with_id AS
  SELECT
    row_number() OVER (ORDER BY date_and_time) AS id,
    c1, c2, c3, c4
  FROM events;
```

The following example shows how to determine the number of rows containing each value for a column. Unlike a corresponding `GROUP BY` query, this one can analyze a single column and still return all values (not just the distinct ones) from the other columns.

```
SELECT x, y, z,
       count() OVER (PARTITION BY x) AS how_many_x
  FROM t1;
```

### Restrictions:

You cannot directly combine the `DISTINCT` operator with analytic function calls. You can put the analytic function call in a `WITH` clause or an inline view, and apply the `DISTINCT` operator to its result set.

```
WITH t1 AS (SELECT x, sum(x) OVER (PARTITION BY x) AS total FROM t1)
  SELECT DISTINCT x, total FROM t1;
```

## Window Clause

Certain analytic functions accept an optional **window clause**, which makes the function analyze only certain rows "around" the current row rather than all rows in the partition. For example, you can get a moving average by specifying some number of preceding and following rows, or a running count or running total by specifying all rows up to the current position. This clause can result in different analytic results for rows within the same partition.

The window clause is supported with the `AVG()`, `COUNT()`, `FIRST_VALUE()`, `LAST_VALUE()`, and `SUM()` functions. For `MAX()` and `MIN()`, the window clause only allowed if the start bound is `UNBOUNDED PRECEDING`

### Syntax:

```
ROWS BETWEEN [ { m | UNBOUNDED } PRECEDING | CURRENT ROW] [ AND [CURRENT ROW | {
  UNBOUNDED | n } FOLLOWING] ]
RANGE BETWEEN [ {m | UNBOUNDED } PRECEDING | CURRENT ROW] [ AND [CURRENT ROW | {
  UNBOUNDED | n } FOLLOWING] ]
```

`ROWS BETWEEN` defines the size of the window in terms of the indexes of the rows in the result set. The size of the window is predictable based on the clauses the position within the result set.

`RANGE BETWEEN` does not currently support numeric arguments to define a variable-size sliding window.

Currently, Impala supports only some combinations of arguments to the `RANGE` clause:

- `RANGE BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW` (the default when `ORDER BY` is specified and the window clause is omitted)
- `RANGE BETWEEN CURRENT ROW AND UNBOUNDED FOLLOWING`
- `RANGE BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING`

When `RANGE` is used, `CURRENT ROW` includes not just the current row but all rows that are tied with the current row based on the `ORDER BY` expressions.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

### Examples:

The following examples show financial data for a fictional stock symbol JDR. The closing price moves up and down each day.

```
create table stock_ticker (stock_symbol string, closing_price decimal(8,2), closing_date
    timestamp);
...load some data...
select * from stock_ticker order by stock_symbol, closing_date
+-----+-----+-----+
| stock_symbol | closing_price | closing_date |
+-----+-----+-----+
| JDR          | 12.86       | 2014-10-02 00:00:00 |
| JDR          | 12.89       | 2014-10-03 00:00:00 |
| JDR          | 12.94       | 2014-10-04 00:00:00 |
| JDR          | 12.55       | 2014-10-05 00:00:00 |
| JDR          | 14.03       | 2014-10-06 00:00:00 |
| JDR          | 14.75       | 2014-10-07 00:00:00 |
| JDR          | 13.98       | 2014-10-08 00:00:00 |
+-----+-----+-----+
```

The queries use analytic functions with window clauses to compute moving averages of the closing price. For example, `ROWS BETWEEN 1 PRECEDING AND 1 FOLLOWING` produces an average of the value from a 3-day span, producing a different value for each row. The first row, which has no preceding row, only gets averaged with the row following it. If the table contained more than one stock symbol, the `PARTITION BY` clause would limit the window for the moving average to only consider the prices for a single stock.

```
select stock_symbol, closing_date, closing_price,
    avg(closing_price) over (partition by stock_symbol order by closing_date
        rows between 1 preceding and 1 following) as moving_average
from stock_ticker;
+-----+-----+-----+-----+
| stock_symbol | closing_date      | closing_price | moving_average |
+-----+-----+-----+-----+
| JDR          | 2014-10-02 00:00:00 | 12.86         | 12.87          |
| JDR          | 2014-10-03 00:00:00 | 12.89         | 12.89          |
| JDR          | 2014-10-04 00:00:00 | 12.94         | 12.79          |
| JDR          | 2014-10-05 00:00:00 | 12.55         | 13.17          |
| JDR          | 2014-10-06 00:00:00 | 14.03         | 13.77          |
| JDR          | 2014-10-07 00:00:00 | 14.75         | 14.25          |
| JDR          | 2014-10-08 00:00:00 | 13.98         | 14.36          |
+-----+-----+-----+-----+
```

The clause `ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW` produces a cumulative moving average, from the earliest data up to the value for each day.

```
select stock_symbol, closing_date, closing_price,
    avg(closing_price) over (partition by stock_symbol order by closing_date
        rows between unbounded preceding and current row) as moving_average
from stock_ticker;
+-----+-----+-----+-----+
| stock_symbol | closing_date      | closing_price | moving_average |
+-----+-----+-----+-----+
| JDR          | 2014-10-02 00:00:00 | 12.86         | 12.86          |
| JDR          | 2014-10-03 00:00:00 | 12.89         | 12.87          |
| JDR          | 2014-10-04 00:00:00 | 12.94         | 12.89          |
| JDR          | 2014-10-05 00:00:00 | 12.55         | 12.81          |
| JDR          | 2014-10-06 00:00:00 | 14.03         | 13.05          |
| JDR          | 2014-10-07 00:00:00 | 14.75         | 13.33          |
| JDR          | 2014-10-08 00:00:00 | 13.98         | 13.42          |
+-----+-----+-----+-----+
```

## DENSE\_RANK() Function

Returns an ascending sequence of integers, starting with 1. The output sequence produces duplicate integers for duplicate values of the `ORDER BY` expressions. After generating duplicate output values for the “tied” input values, the function continues the sequence with the next higher integer. Therefore, the sequence contains duplicates but no gaps when the input contains duplicates. Starts the sequence over for each group produced by the `PARTITIONED BY` clause.

### Syntax:

```
DENSE_RANK() OVER([partition_by_clause] order_by_clause)
```

The PARTITION BY clause is optional. The ORDER BY clause is required. The window clause is not allowed.

### Usage notes:

Often used for top-N and bottom-N queries. For example, it could produce a “top 10” report including all the items with the 10 highest values, even if several items tied for 1st place.

Similar to ROW\_NUMBER and RANK. These functions differ in how they treat duplicate combinations of values.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

### Examples:

The following example demonstrates how the DENSE\_RANK() function identifies where each value “places” in the result set, producing the same result for duplicate values, but with a strict sequence from 1 to the number of groups. For example, when results are ordered by the x column, both 1 values are tied for first; both 2 values are tied for second; and so on.

```
select x, dense_rank() over(order by x) as rank, property from int_t;
+-----+-----+
| x | rank | property |
+-----+-----+
| 1 | 1 | square |
| 1 | 1 | odd |
| 2 | 2 | even |
| 2 | 2 | prime |
| 3 | 3 | prime |
| 3 | 3 | odd |
| 4 | 4 | even |
| 4 | 4 | square |
| 5 | 5 | odd |
| 5 | 5 | prime |
| 6 | 6 | even |
| 6 | 6 | perfect |
| 7 | 7 | lucky |
| 7 | 7 | lucky |
| 7 | 7 | lucky |
| 7 | 7 | odd |
| 7 | 7 | prime |
| 8 | 8 | even |
| 9 | 9 | square |
| 9 | 9 | odd |
| 10 | 10 | round |
| 10 | 10 | even |
+-----+-----+
```

The following examples show how the DENSE\_RANK() function is affected by the PARTITION property within the ORDER BY clause.

Partitioning by the PROPERTY column groups all the even, odd, and so on values together, and DENSE\_RANK() returns the place of each value within the group, producing several ascending sequences.

```
select x, dense_rank() over(partition by property order by x) as rank, property from int_t;
+-----+-----+
| x | rank | property |
+-----+-----+
| 2 | 1 | even |
| 4 | 2 | even |
| 6 | 3 | even |
| 8 | 4 | even |
| 10 | 5 | even |
| 7 | 1 | lucky |
| 7 | 1 | lucky |
| 7 | 1 | lucky |
+-----+-----+
```

1	1	odd
3	2	odd
5	3	odd
7	4	odd
9	5	odd
6	1	perfect
2	1	prime
3	2	prime
5	3	prime
7	4	prime
10	1	round
1	1	square
4	2	square
9	3	square

Partitioning by the x column groups all the duplicate numbers together and returns the place each each value within the group; because each value occurs only 1 or 2 times, DENSE\_RANK() designates each x value as either first or second within its group.

x	rank	property
1	1	odd
1	2	square
2	1	even
2	2	prime
3	1	odd
3	2	prime
4	1	even
4	2	square
5	1	odd
5	2	prime
6	1	even
6	2	perfect
7	1	lucky
7	1	lucky
7	1	lucky
7	2	odd
7	3	prime
8	1	even
9	1	odd
9	2	square
10	1	even
10	2	round

The following example shows how DENSE\_RANK() produces a continuous sequence while still allowing for ties. In this case, Croesus and Midas both have the second largest fortune, while Crassus has the third largest. (In [RANK\(\) Function](#) on page 305, you see a similar query with the RANK() function that shows that while Crassus has the third largest fortune, he is the fourth richest person.)

placement	name	net_worth
1	Solomon	2000000000.00
2	Croesus	1000000000.00
2	Midas	1000000000.00
3	Crassus	500000000.00
4	Scrooge	80000000.00

#### Related information:

[RANK\(\) Function](#) on page 305, [ROW\\_NUMBER\(\) Function](#) on page 307

### FIRST\_VALUE() Function

Returns the expression value from the first row in the window. The return value is NULL if the input expression is NULL.

#### Syntax:

```
FIRST_VALUE(expr) OVER([partition_by_clause] order_by_clause [window_clause])
```

The PARTITION BY clause is optional. The ORDER BY clause is required. The window clause is optional.

#### Usage notes:

If any duplicate values occur in the tuples evaluated by the ORDER BY clause, the result of this function is not deterministic. Consider adding additional ORDER BY columns to ensure consistent ordering.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

#### Examples:

The following example shows a table with a wide variety of country-appropriate greetings. For consistency, we want to standardize on a single greeting for each country. The FIRST\_VALUE() function helps to produce a mail merge report where every person from the same country is addressed with the same greeting.

```
select name, country, greeting from mail_merge
+-----+-----+-----+
| name | country | greeting |
+-----+-----+-----+
| Pete | USA    | Hello   |
| John | USA    | Hi      |
| Boris| Germany | Guten tag |
| Michael| Germany | Guten morgen |
| Bjorn | Sweden  | Hej     |
| Mats  | Sweden  | Tja    |
+-----+-----+-----+
select country, name,
       first_value(greeting)
      over (partition by country order by name, greeting) as greeting
     from mail_merge;
+-----+-----+-----+
| country | name   | greeting |
+-----+-----+-----+
| Germany | Boris  | Guten tag |
| Germany | Michael| Guten tag |
| Sweden  | Bjorn  | Hej      |
| Sweden  | Mats   | Hej      |
| USA     | John   | Hi       |
| USA     | Pete   | Hi       |
+-----+-----+-----+
```

Changing the order in which the names are evaluated changes which greeting is applied to each group.

```
select country, name,
       first_value(greeting)
      over (partition by country order by name desc, greeting) as greeting
     from mail_merge;
+-----+-----+-----+
| country | name   | greeting |
+-----+-----+-----+
| Germany | Michael| Guten morgen |
| Germany | Boris  | Guten morgen |
| Sweden  | Mats   | Tja      |
| Sweden  | Bjorn  | Tja      |
| USA     | Pete   | Hello    |
| USA     | John   | Hello    |
+-----+-----+-----+
```

#### Related information:

[LAST\\_VALUE\(\) Function](#) on page 304

## LAG() Function

This function returns the value of an expression using column values from a preceding row. You specify an integer offset, which designates a row position some number of rows previous to the current row. Any column references in the expression argument refer to column values from that prior row. Typically, the table contains a time sequence or numeric sequence column that clearly distinguishes the ordering of the rows.

### Syntax:

```
LAG (expr [, offset] [, default])
    OVER ([partition_by_clause] order_by_clause)
```

The ORDER BY clause is required. The PARTITION BY clause is optional. The window clause is not allowed.

### Usage notes:

Sometimes used an an alternative to doing a self-join.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

### Examples:

The following example uses the same stock data created in [Window Clause](#) on page 298. For each day, the query prints the closing price alongside the previous day's closing price. The first row for each stock symbol has no previous row, so that LAG() value is NULL.

```
select stock_symbol, closing_date, closing_price,
       lag(closing_price,1) over (partition by stock_symbol order by closing_date) as
       "yesterday closing"
  from stock_ticker
 order by closing_date;
```

stock_symbol	closing_date	closing_price	yesterday closing
JDR	2014-09-13 00:00:00	12.86	NULL
JDR	2014-09-14 00:00:00	12.89	12.86
JDR	2014-09-15 00:00:00	12.94	12.89
JDR	2014-09-16 00:00:00	12.55	12.94
JDR	2014-09-17 00:00:00	14.03	12.55
JDR	2014-09-18 00:00:00	14.75	14.03
JDR	2014-09-19 00:00:00	13.98	14.75

The following example does an arithmetic operation between the current row and a value from the previous row, to produce a delta value for each day. This example also demonstrates how ORDER BY works independently in the different parts of the query. The ORDER BY closing\_date in the OVER clause makes the query analyze the rows in chronological order. Then the outer query block uses ORDER BY closing\_date DESC to present the results with the most recent date first.

```
select stock_symbol, closing_date, closing_price,
       cast(
         closing_price - lag(closing_price,1) over
           (partition by stock_symbol order by closing_date)
         as decimal(8,2)
       )
     as "change from yesterday"
  from stock_ticker
 order by closing_date desc;
```

stock_symbol	closing_date	closing_price	change from yesterday
JDR	2014-09-19 00:00:00	13.98	-0.76
JDR	2014-09-18 00:00:00	14.75	0.72
JDR	2014-09-17 00:00:00	14.03	1.47
JDR	2014-09-16 00:00:00	12.55	-0.38
JDR	2014-09-15 00:00:00	12.94	0.04

JDR	2014-09-14 00:00:00	12.89	0.03
JDR	2014-09-13 00:00:00	12.86	NULL

### Related information:

This function is the converse of [LEAD\(\) Function](#) on page 304.

## LAST\_VALUE() Function

Returns the expression value from the last row in the window. This same value is repeated for all result rows for the group. The return value is `NULL` if the input expression is `NULL`.

### Syntax:

```
LAST_VALUE(expr) OVER([partition_by_clause] order_by_clause [window_clause])
```

The `PARTITION BY` clause is optional. The `ORDER BY` clause is required. The window clause is optional.

### Usage notes:

If any duplicate values occur in the tuples evaluated by the `ORDER BY` clause, the result of this function is not deterministic. Consider adding additional `ORDER BY` columns to ensure consistent ordering.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

### Examples:

The following example uses the same `MAIL_MERGE` table as in the example for [FIRST\\_VALUE\(\) Function](#) on page 302. Because the default window when `ORDER BY` is used is `BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW`, the query requires the `UNBOUNDED FOLLOWING` to look ahead to subsequent rows and find the last value for each country.

```
select country, name,
       last_value(greeting) over (
         partition by country order by name, greeting
         rows between unbounded preceding and unbounded following
       ) as greeting
  from mail_merge
```

country	name	greeting
Germany	Boris	Guten morgen
Germany	Michael	Guten morgen
Sweden	Bjorn	Tja
Sweden	Mats	Tja
USA	John	Hello
USA	Pete	Hello

### Related information:

[FIRST\\_VALUE\(\) Function](#) on page 302

## LEAD() Function

This function returns the value of an expression using column values from a following row. You specify an integer offset, which designates a row position some number of rows after to the current row. Any column references in the expression argument refer to column values from that later row. Typically, the table contains a time sequence or numeric sequence column that clearly distinguishes the ordering of the rows.

### Syntax:

```
LEAD (expr [, offset] [, default])
      OVER ([partition_by_clause] order_by_clause)
```

The `ORDER BY` clause is required. The `PARTITION BY` clause is optional. The window clause is not allowed.

#### Usage notes:

Sometimes used as an alternative to doing a self-join.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

#### Examples:

The following example uses the same stock data created in [Window Clause](#) on page 298. The query analyzes the closing price for a stock symbol, and for each day evaluates if the closing price for the following day is higher or lower.

```
select stock_symbol, closing_date, closing_price,
  case
    (lead(closing_price,1)
      over (partition by stock_symbol order by closing_date)
      - closing_price) > 0
      when true then "higher"
      when false then "flat or lower"
    end as "trending"
  from stock_ticker
  order by closing_date;
```

stock_symbol	closing_date	closing_price	trending
JDR	2014-09-13 00:00:00	12.86	higher
JDR	2014-09-14 00:00:00	12.89	higher
JDR	2014-09-15 00:00:00	12.94	flat or lower
JDR	2014-09-16 00:00:00	12.55	higher
JDR	2014-09-17 00:00:00	14.03	higher
JDR	2014-09-18 00:00:00	14.75	flat or lower
JDR	2014-09-19 00:00:00	13.98	NULL

#### Related information:

This function is the converse of [LAG\(\) Function](#) on page 303.

## RANK() Function

Returns an ascending sequence of integers, starting with 1. The output sequence produces duplicate integers for duplicate values of the `ORDER BY` expressions. After generating duplicate output values for the “tied” input values, the function increments the sequence by the number of tied values. Therefore, the sequence contains both duplicates and gaps when the input contains duplicates. Starts the sequence over for each group produced by the `PARTITIONED BY` clause.

#### Syntax:

```
RANK() OVER([partition_by_clause] order_by_clause)
```

The `PARTITION BY` clause is optional. The `ORDER BY` clause is required. The window clause is not allowed.

#### Usage notes:

Often used for top-N and bottom-N queries. For example, it could produce a “top 10” report including several items that were tied for 10th place.

Similar to `ROW_NUMBER` and `DENSE_RANK`. These functions differ in how they treat duplicate combinations of values.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

#### Examples:

The following example demonstrates how the `RANK()` function identifies where each value “places” in the result set, producing the same result for duplicate values, and skipping values in the sequence to account for the

## Impala SQL Language Reference

number of duplicates. For example, when results are ordered by the `x` column, both 1 values are tied for first; both 2 values are tied for third; and so on.

x	rank()	property
1	1	square
1	1	odd
2	3	even
2	3	prime
3	5	prime
3	5	odd
4	7	even
4	7	square
5	9	odd
5	9	prime
6	11	even
6	11	perfect
7	13	lucky
7	13	lucky
7	13	lucky
7	13	odd
7	13	prime
8	18	even
9	19	square
9	19	odd
10	21	round
10	21	even

The following examples show how the `RANK()` function is affected by the `PARTITION` property within the `ORDER BY` clause.

Partitioning by the `PROPERTY` column groups all the even, odd, and so on values together, and `RANK()` returns the place of each value within the group, producing several ascending sequences.

x	rank()	property
2	1	even
4	2	even
6	3	even
8	4	even
10	5	even
7	1	lucky
7	1	lucky
7	1	lucky
1	1	odd
3	2	odd
5	3	odd
7	4	odd
9	5	odd
6	1	perfect
2	1	prime
3	2	prime
5	3	prime
7	4	prime
10	1	round
1	1	square
4	2	square
9	3	square

Partitioning by the `x` column groups all the duplicate numbers together and returns the place each value within the group; because each value occurs only 1 or 2 times, `RANK()` designates each `x` value as either first or second within its group.

<code>x</code>	<code>rank</code>	<code>property</code>
1	1	odd
1	2	square
2	1	even
2	2	prime
3	1	odd
3	2	prime
4	1	even
4	2	square
5	1	odd
5	2	prime
6	1	even
6	2	perfect
7	1	lucky
7	1	lucky
7	1	lucky
7	4	odd
7	5	prime
8	1	even
9	1	odd
9	2	square
10	1	even
10	2	round

The following example shows how a magazine might prepare a list of history's wealthiest people. Croesus and Midas are tied for second, then Crassus is fourth.

<code>rank</code>	<code>name</code>	<code>net_worth</code>
1	Solomon	2000000000.00
2	Croesus	1000000000.00
2	Midas	1000000000.00
4	Crassus	500000000.00
5	Scrooge	80000000.00

#### Related information:

[DENSE\\_RANK\(\) Function](#) on page 299, [ROW\\_NUMBER\(\) Function](#) on page 307

### ROW\_NUMBER() Function

Returns an ascending sequence of integers, starting with 1. Starts the sequence over for each group produced by the `PARTITIONED BY` clause. The output sequence includes different values for duplicate input values. Therefore, the sequence never contains any duplicates or gaps, regardless of duplicate input values.

#### Syntax:

```
ROW_NUMBER() OVER([partition_by_clause] order_by_clause)
```

The `ORDER BY` clause is required. The `PARTITION BY` clause is optional. The window clause is not allowed.

#### Usage notes:

Often used for top-N and bottom-N queries where the input values are known to be unique, or precisely N rows are needed regardless of duplicate values.

## Impala SQL Language Reference

Because its result value is different for each row in the result set (when used without a `PARTITION BY` clause), `ROW_NUMBER()` can be used to synthesize unique numeric ID values, for example for result sets involving unique values or tuples.

Similar to `RANK` and `DENSE_RANK`. These functions differ in how they treat duplicate combinations of values.

**Added in:** CDH 5.2.0 (Impala 2.0.0)

### Examples:

The following example demonstrates how `ROW_NUMBER()` produces a continuous numeric sequence, even though some values of `x` are repeated.

select x, row_number() over(order by x, property) as row_number, property from int_t;		
x	row_number	property
1	1	odd
1	2	square
2	3	even
2	4	prime
3	5	odd
3	6	prime
4	7	even
4	8	square
5	9	odd
5	10	prime
6	11	even
6	12	perfect
7	13	lucky
7	14	lucky
7	15	lucky
7	16	odd
7	17	prime
8	18	even
9	19	odd
9	20	square
10	21	even
10	22	round

The following example shows how a financial institution might assign customer IDs to some of history's wealthiest figures. Although two of the people have identical net worth figures, unique IDs are required for this purpose. `ROW_NUMBER()` produces a sequence of five different values for the five input rows.

select row_number() over (order by net_worth desc) as account_id, name, net_worth from wealth order by account_id, name;		
account_id	name	net_worth
1	Solomon	2000000000.00
2	Croesus	1000000000.00
3	Midas	1000000000.00
4	Crassus	500000000.00
5	Scrooge	80000000.00

### Related information:

[RANK\(\) Function](#) on page 305, [DENSE\\_RANK\(\) Function](#) on page 299

## Impala User-Defined Functions (UDFs)

User-defined functions (frequently abbreviated as UDFs) let you code your own application logic for processing column values during an Impala query. For example, a UDF could perform calculations using an external math library, combine several column values into one, do geospatial calculations, or other kinds of tests and transformations that are outside the scope of the built-in SQL operators and functions.

You can use UDFs to simplify query logic when producing reports, or to transform data in flexible ways when copying from one table to another with the `INSERT ... SELECT` syntax.

You might be familiar with this feature from other database products, under names such as stored functions or stored routines.

Impala support for UDFs is available in Impala 1.2 and higher:

- In Impala 1.1, using UDFs in a query required using the Hive shell. (Because Impala and Hive share the same metastore database, you could switch to Hive to run just those queries requiring UDFs, then switch back to Impala.)
- Starting in Impala 1.2, Impala can run both high-performance native code UDFs written in C++, and Java-based Hive UDFs that you might already have written.
- Impala can run scalar UDFs that return a single value for each row of the result set, and user-defined aggregate functions (UDAFs) that return a value based on a set of rows. Currently, Impala does not support user-defined table functions (UDTFs) or window functions.

## UDF Concepts

Depending on your use case, you might write all-new functions, reuse Java UDFs that you have already written for Hive, or port Hive Java UDF code to higher-performance native Impala UDFs in C++. You can code either scalar functions for producing results one row at a time, or more complex aggregate functions for doing analysis across. The following sections discuss these different aspects of working with UDFs.

### UDFs and UDAFs

Depending on your use case, the user-defined functions (UDFs) you write might accept or produce different numbers of input and output values:

- The most general kind of user-defined function (the one typically referred to by the abbreviation UDF) takes a single input value and produces a single output value. When used in a query, it is called once for each row in the result set. For example:

```
select customer_name, is_frequent_customer(customer_id) from customers;
select obfuscate(sensitive_column) from sensitive_data;
```

- A user-defined aggregate function (UDAF) accepts a group of values and returns a single value. You use UDAFs to summarize and condense sets of rows, in the same style as the built-in `COUNT`, `MAX()`, `SUM()`, and `AVG()` functions. When called in a query that uses the `GROUP BY` clause, the function is called once for each combination of `GROUP BY` values. For example:

```
-- Evaluates multiple rows but returns a single value.
select closest_restaurant(latitude, longitude) from places;

-- Evaluates batches of rows and returns a separate value for each batch.
select most_profitable_location(store_id, sales, expenses, tax_rate, depreciation)
from franchise_data group by year;
```

- Currently, Impala does not support other categories of user-defined functions, such as user-defined table functions (UDTFs) or window functions.

### Native Impala UDFs

Impala supports UDFs written in C++, in addition to supporting existing Hive UDFs written in Java. Cloudera recommends using C++ UDFs because the compiled native code can yield higher performance, with UDF execution time often 10x faster for a C++ UDF than the equivalent Java UDF.

### Using Hive UDFs with Impala

Impala can run Java-based user-defined functions (UDFs), originally written for Hive, with no changes, subject to the following conditions:

- The parameters and return value must all use data types supported by Impala. For example, nested or composite types are not supported.

## Impala SQL Language Reference

- Currently, Hive UDFs that accept or return the `TIMESTAMP` type are not supported.
- The return type must be a "Writable" type such as `Text` or `IntWritable`, rather than a Java primitive type such as `String` or `int`. Otherwise, the UDF will return `NULL`.
- Hive UDAFs and UDTFs are not supported.
- Typically, a Java UDF will execute several times slower in Impala than the equivalent native UDF written in C++.

To take full advantage of the Impala architecture and performance features, you can also write Impala-specific UDFs in C++.

For background about Java-based Hive UDFs, see the [Hive documentation for UDFs](#). For examples or tutorials for writing such UDFs, search the web for related blog posts.

The ideal way to understand how to reuse Java-based UDFs (originally written for Hive) with Impala is to take some of the Hive built-in functions (implemented as Java UDFs) and take the applicable JAR files through the UDF deployment process for Impala, creating new UDFs with different names:

1. Take a copy of the Hive JAR file containing the Hive built-in functions. For example, the path might be like `/usr/lib/hive/lib/hive-exec-0.10.0-cdh4.2.0.jar`, with different version numbers corresponding to your specific level of CDH.
2. Use `jar tf jar_file` to see a list of the classes inside the JAR. You will see names like `org.apache.hadoop.hive ql.udf.UDFLower.class` and `org.apache.hadoop.hive ql.udf.UDFOPNegative.class`. Make a note of the names of the functions you want to experiment with. When you specify the entry points for the Impala `CREATE FUNCTION` statement, change the slash characters to dots and strip off the `.class` suffix, for example `org.apache.hadoop.hive.ql.udf.UDFLower` and `org.apache.hadoop.hive.ql.udf.UDFOPNegative`.
3. Copy that file to an HDFS location that Impala can read. (In the examples here, we renamed the file to `hive-builtins.jar` in HDFS for simplicity.)
4. For each Java-based UDF that you want to call through Impala, issue a `CREATE FUNCTION` statement, with a `LOCATION` clause containing the full HDFS path of the JAR file, and a `SYMBOL` clause with the fully qualified name of the class, using dots as separators and without the `.class` extension. Remember that user-defined functions are associated with a particular database, so issue a `USE` statement for the appropriate database first, or specify the SQL function name as `db_name.function_name`. Use completely new names for the SQL functions, because Impala UDFs cannot have the same name as Impala built-in functions.
5. Call the function from your queries, passing arguments of the correct type to match the function signature. These arguments could be references to columns, arithmetic or other kinds of expressions, the results of `CAST` functions to ensure correct data types, and so on.

### Java UDF Example: Reusing `lower()` Function

For example, the following `impala-shell` session creates an Impala UDF `my_lower()` that reuses the Java code for the Hive `lower()`: built-in function. We cannot call it `lower()` because Impala does not allow UDFs to have the same name as built-in functions. From SQL, we call the function in a basic way (in a query with no `WHERE` clause), directly on a column, and on the results of a string expression:

```
[localhost:21000] > create database udfs;
[localhost:21000] > use udfs;
localhost:21000 > create function lower(string) returns string location
'/user/hive/udfs/hive.jar' symbol='org.apache.hadoop.hive.ql.udf.UDFLower';
ERROR: AnalysisException: Function cannot have the same name as a builtin: lower
[localhost:21000] > create function my_lower(string) returns string location
'/user/hive/udfs/hive.jar' symbol='org.apache.hadoop.hive.ql.udf.UDFLower';
[localhost:21000] > select my_lower('Some String NOT ALREADY LOWERCASE');
+-----+
| udfs.my_lower('some string not already lowercase') |
+-----+
| some string not already lowercase                      |
+-----+
Returned 1 row(s) in 0.11s
[localhost:21000] > create table t2 (s string);
[localhost:21000] > insert into t2 values ('lower'),('UPPER'),('Init cap'),('CamelCase');
Inserted 4 rows in 2.28s
```

```
[localhost:21000] > select * from t2;
+-----+
| s |
+-----+
| lower |
| UPPER |
| Init cap |
| CamelCase |
+-----+
Returned 4 row(s) in 0.47s
[localhost:21000] > select my_lower(s) from t2;
+-----+
| udfs.my_lower(s) |
+-----+
| lower |
| upper |
| init cap |
| camelcase |
+-----+
Returned 4 row(s) in 0.54s
[localhost:21000] > select my_lower(concat('ABC ',s,' XYZ')) from t2;
+-----+
| udfs.my_lower(concat('abc ', s, ' xyz')) |
+-----+
| abc lower xyz |
| abc upper xyz |
| abc init cap xyz |
| abc camelcase xyz |
+-----+
Returned 4 row(s) in 0.22s
```

### Java UDF Example: Reusing negative() Function

Here is an example that reuses the Hive Java code for the `negative()` built-in function. This example demonstrates how the data types of the arguments must match precisely with the function signature. At first, we create an Impala SQL function that can only accept an integer argument. Impala cannot find a matching function when the query passes a floating-point argument, although we can call the integer version of the function by casting the argument. Then we overload the same function name to also accept a floating-point argument.

```
[localhost:21000] > create table t (x int);
[localhost:21000] > insert into t values (1), (2), (4), (100);
Inserted 4 rows in 1.43s
[localhost:21000] > create function my_neg(bigint) returns bigint location
'./user/hive/udfs/hive.jar' symbol='org.apache.hadoop.hive.ql.udf.UDFOPNegative';
[localhost:21000] > select my_neg(4);
+-----+
| udfs.my_neg(4) |
+-----+
| -4 |
+-----+
[localhost:21000] > select my_neg(x) from t;
+-----+
| udfs.my_neg(x) |
+-----+
| -2 |
| -4 |
| -100 |
+-----+
Returned 3 row(s) in 0.60s
[localhost:21000] > select my_neg(4.0);
ERROR: AnalysisException: No matching function with signature: udfs.my_neg(FLOAT).
[localhost:21000] > select my_neg(cast(4.0 as int));
+-----+
| udfs.my_neg(cast(4.0 as int)) |
+-----+
| -4 |
+-----+
Returned 1 row(s) in 0.11s
[localhost:21000] > create function my_neg(double) returns double location
'./user/hive/udfs/hive.jar' symbol='org.apache.hadoop.hive.ql.udf.UDFOPNegative';
```

```
[localhost:21000] > select my_neg(4.0);
+-----+
| udfs.my_neg(4.0) |
+-----+
| -4
+-----+
Returned 1 row(s) in 0.11s
```

You can find the sample files mentioned here in [the Impala github repo](#).

### Runtime Environment for UDFs

By default, Impala copies UDFs into `/tmp`, and you can configure this location through the `--local_library_dir` startup flag for the `impalad` daemon.

### Installing the UDF Development Package

To develop UDFs for Impala, download and install the `impala-udf-devel` package containing header files, sample source, and build configuration files.

1. Start at <http://archive.cloudera.com/cdh5/> for the CDH 5 package, or <http://archive.cloudera.com/impala/> for the CDH 4 package.
2. Locate the appropriate `.repo` or `list` file for your operating system version, such as [the .repo file for CDH 4 on RHEL 6](#).
3. Use the familiar `yum`, `zypper`, or `apt-get` commands depending on your operating system, with `impala-udf-devel` for the package name.

- **Note:** The UDF development code does not rely on Impala being installed on the same machine. You can write and compile UDFs on a minimal development system, then deploy them on a different one for use with Impala. If you develop UDFs on a server managed by Cloudera Manager through the parcel mechanism, you still install the UDF development kit through the package mechanism; this small standalone package does not interfere with the parcels containing the main Impala code.

When you are ready to start writing your own UDFs, download the sample code and build scripts from [the Cloudera sample UDF github](#). Then see [Writing User-Defined Functions \(UDFs\)](#) on page 312 for how to code UDFs, and [Examples of Creating and Using UDFs](#) on page 317 for how to build and run UDFs.

### Writing User-Defined Functions (UDFs)

Before starting UDF development, make sure to install the development package and download the UDF code samples, as described in [Installing the UDF Development Package](#) on page 312.

When writing UDFs:

- Keep in mind the data type differences as you transfer values from the high-level SQL to your lower-level UDF code. For example, in the UDF code you might be much more aware of how many bytes different kinds of integers require.
- Use best practices for function-oriented programming: choose arguments carefully, avoid side effects, make each function do a single thing, and so on.

### Getting Started with UDF Coding

To understand the layout and member variables and functions of the predefined UDF data types, examine the header file `/usr/include/impala_udf/udf.h`:

```
// This is the only Impala header required to develop UDFs and UDA. This header
// contains the types that need to be used and the FunctionContext object. The context
// object serves as the interface object between the UDF/UDA and the impala process.
```

For the basic declarations needed to write a scalar UDF, see the header file `udf-sample.h` within the sample build environment, which defines a simple function named `AddUdf()`:

```
#ifndef IMPALA_UDF_SAMPLE_UDF_H
#define IMPALA_UDF_SAMPLE_UDF_H

#include <impala_udf/udf.h>

using namespace impala_udf;

IntVal AddUdf(FunctionContext* context, const IntVal& arg1, const IntVal& arg2);

#endif
```

For sample C++ code for a simple function named `AddUdf()`, see the source file `udf-sample.cc` within the sample build environment:

```
#include "udf-sample.h"

// In this sample we are declaring a UDF that adds two ints and returns an int.
IntVal AddUdf(FunctionContext* context, const IntVal& arg1, const IntVal& arg2) {
    if (arg1.is_null || arg2.is_null) return IntVal::null();
    return IntVal(arg1.val + arg2.val);
}

// Multiple UDFs can be defined in the same file
```

## Data Types for Function Arguments and Return Values

Each value that a user-defined function can accept as an argument or return as a result value must map to a SQL data type that you could specify for a table column.

Each data type has a corresponding structure defined in the C++ and Java header files, with two member fields and some predefined comparison operators and constructors:

- `is_null` indicates whether the value is `NONE` or not. `val` holds the actual argument or return value when it is non-`NONE`.
- Each struct also defines a `null()` member function that constructs an instance of the struct with the `is_null` flag set.
- The built-in SQL comparison operators and clauses such as `<`, `>=`, `BETWEEN`, and `ORDER BY` all work automatically based on the SQL return type of each UDF. For example, Impala knows how to evaluate `BETWEEN 1 AND udf_returning_int(col1)` or `ORDER BY udf_returning_string(col2)` without you declaring any comparison operators within the UDF itself.

For convenience within your UDF code, each struct defines `==` and `!=` operators for comparisons with other structs of the same type. These are for typical C++ comparisons within your own code, not necessarily reproducing SQL semantics. For example, if the `is_null` flag is set in both structs, they compare as equal. That behavior of `null` comparisons is different from SQL (where `NULL == NULL` is `NULL` rather than `true`), but more in line with typical C++ behavior.

- Each kind of struct has one or more constructors that define a filled-in instance of the struct, optionally with default values.
- Each kind of struct has a `null()` member function that returns an instance of the struct with the `is_null` flag set.
- Because Impala currently does not support composite or nested types, Impala cannot process UDFs that accept such types as arguments or return them as result values. This limitation applies both to Impala UDFs written in C++ and Java-based Hive UDFs.
- You can overload functions by creating multiple functions with the same SQL name but different argument types. For overloaded functions, you must use different C++ or Java entry point names in the underlying functions.

## Impala SQL Language Reference

The data types defined on the C++ side (in `/usr/include/impala_udf/udf.h`) are:

- `IntVal` represents an `INT` column.
- `BigIntVal` represents a `BIGINT` column. Even if you do not need the full range of a `BIGINT` value, it can be useful to code your function arguments as `BigIntVal` to make it convenient to call the function with different kinds of integer columns and expressions as arguments. Impala automatically casts smaller integer types to larger ones when appropriate, but does not implicitly cast large integer types to smaller ones.
- `Smallintval` represents a `SIMALLINT` column.
- `Tinyintval` represents a `TINYINT` column.
- `StringVal` represents a `STRING` column. It has a `len` field representing the length of the string, and a `ptr` field pointing to the string data. It has constructors that create a new `StringVal` struct based on a null-terminated C-style string, or a pointer plus a length; these new structs still refer to the original string data rather than allocating a new buffer for the data. It also has a constructor that takes a pointer to a `FunctionContext` struct and a length, that does allocate space for a new copy of the string data, for use in UDFs that return string values.
- `BooleanVal` represents a `BOOLEAN` column.
- `FloatVal` represents a `FLOAT` column.
- `DoubleVal` represents a `DOUBLE` column.
- `TimestampVal` represents a `TIMESTAMP` column. It has a `date` field, a 32-bit integer representing the Gregorian date, that is, the days past the epoch date. It also has a `time_of_day` field, a 64-bit integer representing the current time of day in nanoseconds.

### Variable-Length Argument Lists

UDFs typically take a fixed number of arguments, with each one named explicitly in the signature of your C++ function. Your function can also accept additional optional arguments, all of the same type. For example, you can concatenate two strings, three strings, four strings, and so on. Or you can compare two numbers, three numbers, four numbers, and so on.

To accept a variable-length argument list, code the signature of your function like this:

```
StringVal Concat(FunctionContext* context, const StringVal& separator,
    int num_var_args, const StringVal* args);
```

In the `CREATE FUNCTION` statement, after the type of the first optional argument, include `...` to indicate it could be followed by more arguments of the same type. For example, the following function accepts a `STRING` argument, followed by one or more additional `STRING` arguments:

```
[localhost:21000] > create function my_concat(string, string ...) returns string location
  '/user/test_user/udfs/sample.so' symbol='Concat';
```

The call from the SQL query must pass at least one argument to the variable-length portion of the argument list.

When Impala calls the function, it fills in the initial set of required arguments, then passes the number of extra arguments and a pointer to the first of those optional arguments.

### Handling NULL Values

For correctness, performance, and reliability, it is important for each UDF to handle all situations where any `NULL` values are passed to your function. For example, when passed a `NULL`, UDFs typically also return `NULL`. In an aggregate function, which could be passed a combination of real and `NULL` values, you might make the final value into a `NULL` (as in `CONCAT()`), ignore the `NULL` value (as in `AVG()`), or treat it the same as a numeric zero or empty string.

Each parameter type, such as `IntVal` or `StringVal`, has an `is_null` Boolean member. Test this flag immediately for each argument to your function, and if it is set, do not refer to the `val` field of the argument structure. The `val` field is undefined when the argument is `NULL`, so your function could go into an infinite loop or produce incorrect results if you skip the special handling for `NULL`.

If your function returns `NULL` when passed a `NULL` value, or in other cases such as when a search string is not found, you can construct a null instance of the return type by using its `null()` member function.

### Memory Allocation for UDFs

By default, memory allocated within a UDF is deallocated when the function exits, which could be before the query is finished. The input arguments remain allocated for the lifetime of the function, so you can refer to them in the expressions for your return values. If you use temporary variables to construct all-new string values, use the `StringVal()` constructor that takes an initial `FunctionContext*` argument followed by a length, and copy the data into the newly allocated memory buffer.

### Thread-Safe Work Area for UDFs

One way to improve performance of UDFs is to specify the optional `PREPARE_FN` and `CLOSE_FN` clauses on the `CREATE FUNCTION` statement. The “prepare” function sets up a thread-safe data structure in memory that you can use as a work area. The “close” function deallocates that memory. Each subsequent call to the UDF within the same thread can access that same memory area. There might be several such memory areas allocated on the same host, as UDFs are parallelized using multiple threads.

Within this work area, you can set up predefined lookup tables, or record the results of complex operations on data types such as `STRING` or `TIMESTAMP`. Saving the results of previous computations rather than repeating the computation each time is an optimization known as <http://en.wikipedia.org/wiki/Memoization>. For example, if your UDF performs a regular expression match or date manipulation on a column that repeats the same value over and over, you could store the last-computed value or a hash table of already-computed values, and do a fast lookup to find the result for subsequent iterations of the UDF.

Each such function must have the signature:

```
void function_name(impala_udf::FunctionContext*,
    impala_udf::FunctionContext::FunctionScope)
```

Currently, only `THREAD_SCOPE` is implemented, not `FRAGMENT_SCOPE`. See `udf.h` for details about the scope values.

### Error Handling for UDFs

To handle errors in UDFs, you call functions that are members of the initial `FunctionContext*` argument passed to your function.

A UDF can record one or more warnings, for conditions that indicate minor, recoverable problems that do not cause the query to stop. The signature for this function is:

```
bool AddWarning(const char* warning_msg);
```

For a serious problem that requires cancelling the query, a UDF can set an error flag that prevents the query from returning any results. The signature for this function is:

```
void SetError(const char* error_msg);
```

### Writing User-Defined Aggregate Functions (UDAFs)

User-defined aggregate functions (UDAFs or UDAs) are a powerful and flexible category of user-defined functions. If a query processes N rows, calling a UDAF during the query condenses the result set, anywhere from a single value (such as with the `SUM` or `MAX` functions), or some number less than or equal to N (as in queries using the `GROUP BY` or `HAVING` clause).

## The Underlying Functions for a UDA

A UDAF must maintain a state value across subsequent calls, so that it can accumulate a result across a set of calls, rather than derive it purely from one set of arguments. For that reason, a UDAF is represented by multiple underlying functions:

- An initialization function that sets any counters to zero, creates empty buffers, and does any other one-time setup for a query.
- An update function that processes the arguments for each row in the query result set and accumulates an intermediate result for each node. For example, this function might increment a counter, append to a string buffer, or set flags.
- A merge function that combines the intermediate results from two different nodes.
- A serialize function that flattens any intermediate values containing pointers, and frees any memory allocated during the init, update, and merge phases.
- A finalize function that either passes through the combined result unchanged, or does one final transformation.

In the SQL syntax, you create a UDAF by using the statement `CREATE AGGREGATE FUNCTION`. You specify the entry points of the underlying C++ functions using the clauses `INIT_FN`, `UPDATE_FN`, `MERGE_FN`, `SERIALIZE_FN`, and `FINALIZE_FN`.

For convenience, you can use a naming convention for the underlying functions and Impala automatically recognizes those entry points. Specify the `UPDATE_FN` clause, using an entry point name containing the string `update` or `Update`. When you omit the other `_FN` clauses from the SQL statement, Impala looks for entry points with names formed by substituting the `update` or `Update` portion of the specified name.

`uda-sample.h`:

See this file online at: <https://github.com/cloudera/impala-udf-samples/blob/master/uda-sample.cc>

`uda-sample.cc`:

See this file online at: <https://github.com/cloudera/impala-udf-samples/blob/master/uda-sample.h>

## Building and Deploying UDFs

This section explains the steps to compile Impala UDFs from C++ source code, and deploy the resulting libraries for use in Impala queries.

Impala ships with a sample build environment for UDFs, that you can study, experiment with, and adapt for your own use. This sample build environment starts with the `cmake` configuration command, which reads the file `CMakeLists.txt` and generates a `Makefile` customized for your particular directory paths. Then the `make` command runs the actual build steps based on the rules in the `Makefile`.

Impala loads the shared library from an HDFS location. After building a shared library containing one or more UDFs, use `hdfs dfs` or `hadoop fs` commands to copy the binary file to an HDFS location readable by Impala.

The final step in deployment is to issue a `CREATE FUNCTION` statement in the `impala-shell` interpreter to make Impala aware of the new function. See [CREATE FUNCTION Statement](#) on page 177 for syntax details. Because each function is associated with a particular database, always issue a `USE` statement to the appropriate database before creating a function, or specify a fully qualified name, that is, `CREATE FUNCTION db_name.function_name`.

As you update the UDF code and redeploy updated versions of a shared library, use `DROP FUNCTION` and `CREATE FUNCTION` to let Impala pick up the latest version of the code.

### Note:

Currently, Impala UDFs and UDAFs are not persisted in the metastore database. Information about these functions is held in the memory of the `catalogd` daemon. You must reload them by running the `CREATE FUNCTION` statements again each time you restart the `catalogd` daemon.

Prerequisites for the build environment are:

```
# Use the appropriate package installation command for your Linux distribution.
sudo yum install gcc-c++ cmake boost-devel
sudo yum install impala-udf-devel
```

Then, unpack the sample code in `udf_samples.tar.gz` and use that as a template to set up your build environment.

To build the original samples:

```
# Process CMakeLists.txt and set up appropriate Makefiles.
cmake .
# Generate shared libraries from UDF and UDAF sample code,
# udf_samples/libudfsample.so and udf_samples/libudasample.so
make
```

The sample code to examine, experiment with, and adapt is in these files:

- `udf-sample.h`: Header file that declares the signature for a scalar UDF (`AddUDF`).
- `udf-sample.cc`: Sample source for a simple UDF that adds two integers. Because Impala can reference multiple function entry points from the same shared library, you could add other UDF functions in this file and add their signatures to the corresponding header file.
- `udf-sample-test.cc`: Basic unit tests for the sample UDF.
- `uda-sample.h`: Header file that declares the signature for sample aggregate functions. The SQL functions will be called `COUNT`, `AVG`, and `STRINGCONCAT`. Because aggregate functions require more elaborate coding to handle the processing for multiple phases, there are several underlying C++ functions such as `CountInit`, `AvgUpdate`, and `StringConcatFinalize`.
- `uda-sample.cc`: Sample source for simple UDAFs that demonstrate how to manage the state transitions as the underlying functions are called during the different phases of query processing.
  - The UDAF that imitates the `COUNT` function keeps track of a single incrementing number; the merge functions combine the intermediate count values from each Impala node, and the combined number is returned verbatim by the finalize function.
  - The UDAF that imitates the `AVG` function keeps track of two numbers, a count of rows processed and the sum of values for a column. These numbers are updated and merged as with `COUNT`, then the finalize function divides them to produce and return the final average value.
  - The UDAF that concatenates string values into a comma-separated list demonstrates how to manage storage for a string that increases in length as the function is called for multiple rows.
- `uda-sample-test.cc`: basic unit tests for the sample UDAFs.

## Performance Considerations for UDFs

Because a UDF typically processes each row of a table, potentially being called billions of times, the performance of each UDF is a critical factor in the speed of the overall ETL or ELT pipeline. Tiny optimizations you can make within the function body can pay off in a big way when the function is called over and over when processing a huge result set.

## Examples of Creating and Using UDFs

This section demonstrates how to create and use all kinds of user-defined functions (UDFs).

For downloadable examples that you can experiment with, adapt, and use as templates for your own functions, see [the Cloudera sample UDF github](#). You must have already installed the appropriate header files, as explained in [Installing the UDF Development Package](#) on page 312.

### Sample C++ UDFs: HasVowels, CountVowels, StripVowels

This example shows 3 separate UDFs that operate on strings and return different data types. In the C++ code, the functions are `HasVowels()` (checks if a string contains any vowels), `CountVowels()` (returns the number of vowels in a string), and `StripVowels()` (returns a new string with vowels removed).

First, we add the signatures for these functions to `udf-sample.h` in the demo build environment:

```
BooleanVal HasVowels(FunctionContext* context, const StringVal& input);
IntVal CountVowels(FunctionContext* context, const StringVal& arg1);
StringVal StripVowels(FunctionContext* context, const StringVal& arg1);
```

Then, we add the bodies of these functions to `udf-sample.cc`:

```
BooleanVal HasVowels(FunctionContext* context, const StringVal& input)
{
    if (input.is_null) return BooleanVal::null();

    int index;
    uint8_t *ptr;

    for (ptr = input.ptr, index = 0; index <= input.len; index++, ptr++)
    {
        uint8_t c = tolower(*ptr);
        if (c == 'a' || c == 'e' || c == 'i' || c == 'o' || c == 'u')
        {
            return BooleanVal(true);
        }
    }
    return BooleanVal(false);
}

IntVal CountVowels(FunctionContext* context, const StringVal& arg1)
{
    if (arg1.is_null) return IntVal::null();

    int count;
    int index;
    uint8_t *ptr;

    for (ptr = arg1.ptr, count = 0, index = 0; index <= arg1.len; index++, ptr++)
    {
        uint8_t c = tolower(*ptr);
        if (c == 'a' || c == 'e' || c == 'i' || c == 'o' || c == 'u')
        {
            count++;
        }
    }
    return IntVal(count);
}

StringVal StripVowels(FunctionContext* context, const StringVal& arg1)
{
    if (arg1.is_null) return StringVal::null();

    int index;
    std::string original((const char *)arg1.ptr, arg1.len);
    std::string shorter("");

    for (index = 0; index < original.length(); index++)
    {
        uint8_t c = original[index];
        uint8_t l = tolower(c);

        if (l == 'a' || l == 'e' || l == 'i' || l == 'o' || l == 'u')
        {
            ;
        }
        else
        {
            shorter.append(1, (char)c);
        }
    }
}
```

```

        }
        // The modified string is stored in 'shorter', which is destroyed when this function
        // ends. We need to make a string val
        // and copy the contents.
        StringVal result(context, shorter.size()); // Only the version of the ctor that
        // takes a context object allocates new memory
        memcpy(result.ptr, shorter.c_str(), shorter.size());
        return result;
    }
}

```

We build a shared library, libudfsample.so, and put the library file into HDFS where Impala can read it:

```

$ make
[  0%] Generating udf_samples/uda-sample.ll
[ 16%] Built target uda-sample-ir
[ 33%] Built target udasample
[ 50%] Built target uda-sample-test
[ 50%] Generating udf_samples/udf-sample.ll
[ 66%] Built target udf-sample-ir
Scanning dependencies of target udfsamples
[ 83%] Building CXX object CMakeFiles/udfsamples.dir/udf-sample.o
Linking CXX shared library udf_samples/libudfsample.so
[ 83%] Built target udfsamples
Linking CXX executable udf_samples/udf-sample-test
[100%] Built target udf-sample-test
$ hdfs dfs -put ./udf_samples/libudfsample.so /user/hive/udfs/libudfsample.so

```

Finally, we go into the impala-shell interpreter where we set up some sample data, issue CREATE FUNCTION statements to set up the SQL function names, and call the functions in some queries:

```

[localhost:21000] > create database udf_testing;
[localhost:21000] > use udf_testing;

[localhost:21000] > create function has_vowels (string) returns boolean location
'/user/hive/udfs/libudfsample.so' symbol='HasVowels';
[localhost:21000] > select has_vowels('abc');
+-----+
| udfs.has_vowels('abc') |
+-----+
| true |
+-----+
Returned 1 row(s) in 0.13s
[localhost:21000] > select has_vowels('zxcvbnm');
+-----+
| udfs.has_vowels('zxcvbnm') |
+-----+
| false |
+-----+
Returned 1 row(s) in 0.12s
[localhost:21000] > select has_vowels(null);
+-----+
| udfs.has_vowels(null) |
+-----+
| NULL |
+-----+
Returned 1 row(s) in 0.11s
[localhost:21000] > select s, has_vowels(s) from t2;
+-----+-----+
| s      | udfs.has_vowels(s) |
+-----+-----+
| lower   | true   |
| UPPER   | true   |
| Init cap | true   |
| CamelCase | true   |
+-----+-----+
Returned 4 row(s) in 0.24s

[localhost:21000] > create function count_vowels (string) returns int location
'/user/hive/udfs/libudfsample.so' symbol='CountVowels';
[localhost:21000] > select count_vowels('cat in the hat');
+-----+

```

```
| udfs.count_vowels('cat in the hat') |
+-----+
| 4
+-----+
Returned 1 row(s) in 0.12s
[localhost:21000] > select s, count_vowels(s) from t2;
+-----+
| s      | udfs.count_vowels(s) |
+-----+
| lower  | 2
| UPPER  | 2
| Init cap | 3
| CamelCase | 4
+-----+
Returned 4 row(s) in 0.23s
[localhost:21000] > select count_vowels(null);
+-----+
| udfs.count_vowels(null) |
+-----+
| NULL
+-----+
Returned 1 row(s) in 0.12s

[localhost:21000] > create function strip_vowels (string) returns string location
'/user/hive/udfs/libudfsample.so' symbol='StripVowels';
[localhost:21000] > select strip_vowels('abcdefg');
+-----+
| udfs.strip_vowels('abcdefg') |
+-----+
| bcdfg
+-----+
Returned 1 row(s) in 0.11s
[localhost:21000] > select strip_vowels('ABCDEFG');
+-----+
| udfs.strip_vowels('abcdefg') |
+-----+
| BCFG
+-----+
Returned 1 row(s) in 0.12s
[localhost:21000] > select strip_vowels(null);
+-----+
| udfs.strip_vowels(null) |
+-----+
| NULL
+-----+
Returned 1 row(s) in 0.16s
[localhost:21000] > select s, strip_vowels(s) from t2;
+-----+
| s      | udfs.strip_vowels(s) |
+-----+
| lower  | lwr
| UPPER  | PPR
| Init cap | nt cp
| CamelCase | CmlCs
+-----+
Returned 4 row(s) in 0.24s
```

### Sample C++ UDA: SumOfSquares

This example demonstrates a user-defined aggregate function (UDA) that produces the sum of the squares of its input values.

The coding for a UDA is a little more involved than a scalar UDF, because the processing is split into several phases, each implemented by a different function. Each phase is relatively straightforward: the “update” and “merge” phases, where most of the work is done, read an input value and combine it with some accumulated intermediate value.

As in our sample UDF from the previous example, we add function signatures to a header file (in this case, `uda-sample.h`). Because this is a math-oriented UDA, we make two versions of each function, one accepting an integer value and the other accepting a floating-point value.

```
void SumOfSquaresInit(FunctionContext* context, BigIntVal* val);
void SumOfSquaresInit(FunctionContext* context, DoubleVal* val);

void SumOfSquaresUpdate(FunctionContext* context, const BigIntVal& input, BigIntVal*
val);
void SumOfSquaresUpdate(FunctionContext* context, const DoubleVal& input, DoubleVal*
val);

void SumOfSquaresMerge(FunctionContext* context, const BigIntVal& src, BigIntVal* dst);
void SumOfSquaresMerge(FunctionContext* context, const DoubleVal& src, DoubleVal* dst);

BigIntVal SumOfSquaresFinalize(FunctionContext* context, const BigIntVal& val);
DoubleVal SumOfSquaresFinalize(FunctionContext* context, const DoubleVal& val);
```

We add the function bodies to a C++ source file (in this case, `uda-sample.cc`):

```
void SumOfSquaresInit(FunctionContext* context, BigIntVal* val) {
    val->is_null = false;
    val->val = 0;
}
void SumOfSquaresInit(FunctionContext* context, DoubleVal* val) {
    val->is_null = false;
    val->val = 0.0;
}

void SumOfSquaresUpdate(FunctionContext* context, const BigIntVal& input, BigIntVal*
val) {
    if (input.is_null) return;
    val->val += input.val * input.val;
}
void SumOfSquaresUpdate(FunctionContext* context, const DoubleVal& input, DoubleVal*
val) {
    if (input.is_null) return;
    val->val += input.val * input.val;
}

void SumOfSquaresMerge(FunctionContext* context, const BigIntVal& src, BigIntVal* dst)
{
    dst->val += src.val;
}
void SumOfSquaresMerge(FunctionContext* context, const DoubleVal& src, DoubleVal* dst)
{
    dst->val += src.val;
}

BigIntVal SumOfSquaresFinalize(FunctionContext* context, const BigIntVal& val) {
    return val;
}
DoubleVal SumOfSquaresFinalize(FunctionContext* context, const DoubleVal& val) {
    return val;
}
```

As with the sample UDF, we build a shared library and put it into HDFS:

```
$ make
[  0%] Generating udf_samples/uda-sample.ll
[ 16%] Built target uda-sample-ir
Scanning dependencies of target udasample
[ 33%] Building CXX object CMakeFiles/udasample.dir/uda-sample.o
Linking CXX shared library udf_samples/libudasample.so
[ 33%] Built target udasample
Scanning dependencies of target uda-sample-test
[ 50%] Building CXX object CMakeFiles/uda-sample-test.dir/uda-sample-test.o
Linking CXX executable udf_samples/uda-sample-test
[ 50%] Built target uda-sample-test
[ 50%] Generating udf_samples/udf-sample.ll
```

```
[ 66%] Built target udf-sample-ir
[ 83%] Built target udfs
[100%] Built target udf-sample-test
$ hdfs dfs -put ./udf_samples/libudasample.so /user/hive/udfs/libudasample.so
```

To create the SQL function, we issue a `CREATE AGGREGATE FUNCTION` statement and specify the underlying C++ function names for the different phases:

```
[localhost:21000] > use udf_testing;
[localhost:21000] > create table sos (x bigint, y double);
[localhost:21000] > insert into sos values (1, 1.1), (2, 2.2), (3, 3.3), (4, 4.4);
Inserted 4 rows in 1.10s

[localhost:21000] > create aggregate function sum_of_squares(bigint) returns bigint
> location '/user/hive/udfs/libudasample.so'
> init_fn='SumOfSquaresInit'
> update_fn='SumOfSquaresUpdate'
> merge_fn='SumOfSquaresMerge'
> finalize_fn='SumOfSquaresFinalize';

[localhost:21000] > -- Compute the same value using literals or the UDA;
[localhost:21000] > select 1*1 + 2*2 + 3*3 + 4*4;
+-----+
| 1 * 1 + 2 * 2 + 3 * 3 + 4 * 4 |
+-----+
| 30
+-----+
Returned 1 row(s) in 0.12s
[localhost:21000] > select sum_of_squares(x) from sos;
+-----+
| udfs.sum_of_squares(x) |
+-----+
| 30
+-----+
Returned 1 row(s) in 0.35s
```

Until we create the overloaded version of the UDA, it can only handle a single data type. To allow it to handle `DOUBLE` as well as `BIGINT`, we issue another `CREATE AGGREGATE FUNCTION` statement:

```
[localhost:21000] > select sum_of_squares(y) from sos;
ERROR: AnalysisException: No matching function with signature:
udfs.sum_of_squares(DOUBLE).

[localhost:21000] > create aggregate function sum_of_squares(double) returns double
> location '/user/hive/udfs/libudasample.so'
> init_fn='SumOfSquaresInit'
> update_fn='SumOfSquaresUpdate'
> merge_fn='SumOfSquaresMerge'
> finalize_fn='SumOfSquaresFinalize';

[localhost:21000] > -- Compute the same value using literals or the UDA;
[localhost:21000] > select 1.1*1.1 + 2.2*2.2 + 3.3*3.3 + 4.4*4.4;
+-----+
| 1.1 * 1.1 + 2.2 * 2.2 + 3.3 * 3.3 + 4.4 * 4.4 |
+-----+
| 36.3
+-----+
Returned 1 row(s) in 0.12s
[localhost:21000] > select sum_of_squares(y) from sos;
+-----+
| udfs.sum_of_squares(y) |
+-----+
| 36.3
+-----+
Returned 1 row(s) in 0.35s
```

Typically, you use a UDA in queries with `GROUP BY` clauses, to produce a result set with a separate aggregate value for each combination of values from the `GROUP BY` clause. Let's change our sample table to use 0 to

indicate rows containing even values, and 1 to flag rows containing odd values. Then the GROUP BY query can return two values, the sum of the squares for the even values, and the sum of the squares for the odd values:

```
[localhost:21000] > insert overwrite sos values (1, 1), (2, 0), (3, 1), (4, 0);
Inserted 4 rows in 1.24s

[localhost:21000] > -- Compute 1 squared + 3 squared, and 2 squared + 4 squared;
[localhost:21000] > select y, sum_of_squares(x) from sos group by y;
+-----+
| y | udfs.sum_of_squares(x) |
+-----+
| 1 | 10
| 0 | 20
+-----+
Returned 2 row(s) in 0.43s
```

## Security Considerations for User-Defined Functions

When the Impala authorization feature is enabled:

- To call a UDF in a query, you must have the required read privilege for any databases and tables used in the query.
- Because incorrectly coded UDFs could cause performance or capacity problems, for example by going into infinite loops or allocating excessive amounts of memory, only an administrative user can create UDFs. That is, to execute the CREATE FUNCTION statement requires the ALL privilege on the server.

See [Enabling Sentry Authorization for Impala](#) on page 91 for details about authorization in Impala.

## Limitations and Restrictions for Impala UDFs

The following limitations and restrictions apply to Impala UDFs in the current release:

- Impala does not support Hive UDFs that accept or return composite or nested types, or other types not available in Impala tables.
- All Impala UDFs must be deterministic, that is, produce the same output each time when passed the same argument values. For example, an Impala UDF must not call functions such as `rand()` to produce different values for each invocation. It must not retrieve data from external sources, such as from disk or over the network.
- An Impala UDF must not spawn other threads or processes.
- When the `catalogd` process is restarted, all UDFs become undefined and must be reloaded.
- Impala currently does not support user-defined table functions (UDTFs).
- The `CHAR` and `VARCHAR` types cannot be used as input arguments or return values for UDFs.

## SQL Differences Between Impala and Hive

Impala's SQL syntax follows the SQL-92 standard, and includes many industry extensions in areas such as built-in functions. See [Porting SQL from Other Database Systems to Impala](#) on page 326 for a general discussion of adapting SQL code from a variety of database systems to Impala.

Because Impala and Hive share the same metastore database and their tables are often used interchangeably, the following section covers differences between Impala and Hive in detail.

## HiveQL Features not Available in Impala

The current release of Impala does not support the following SQL features that you might be familiar with from HiveQL:

- Non-scalar data types such as maps, arrays, structs.
- Extensibility mechanisms such as `TRANSFORM`, custom file formats, or custom SerDes.
- XML and JSON functions.

## Impala SQL Language Reference

- Certain aggregate functions from HiveQL: `covar_pop`, `covar_samp`, `corr`, `percentile`, `percentile_approx`, `histogram_numeric`, `collect_set`; Impala supports the set of aggregate functions listed in [Impala Aggregate Functions](#) on page 284 and analytic functions listed in [Impala Analytic Functions](#) on page 296.
- Sampling.
- Lateral views.
- Multiple `DISTINCT` clauses per query, although Impala includes some workarounds for this limitation.

▪ **Note:**

By default, Impala only allows a single `COUNT(DISTINCT columns)` expression in each query.

If you do not need precise accuracy, you can produce an estimate of the distinct values for a column by specifying `NDV(column)`; a query can contain multiple instances of `NDV(column)`. To make Impala automatically rewrite `COUNT(DISTINCT)` expressions to `NDV()`, enable the `APPX_COUNT_DISTINCT` query option.

To produce the same result as multiple `COUNT(DISTINCT)` expressions, you can use the following technique for queries involving a single table:

```
select v1.c1 result1, v2.c1 result2 from
  (select count(distinct col1) as c1 from t1) v1
  cross join
  (select count(distinct col2) as c1 from t1) v2;
```

Because `CROSS JOIN` is an expensive operation, prefer to use the `NDV()` technique wherever practical.

User-defined functions (UDFs) are supported starting in Impala 1.2. See [Impala User-Defined Functions \(UDFs\)](#) on page 308 for full details on Impala UDFs.

- Impala supports high-performance UDFs written in C++, as well as reusing some Java-based Hive UDFs.
- Impala supports scalar UDFs and user-defined aggregate functions (UDAFs). Impala does not currently support user-defined table generating functions (UDTFs).
- Only Impala-supported column types are supported in Java-based UDFs.

Impala does not currently support these HiveQL statements:

- `ANALYZE TABLE` (the Impala equivalent is `COMPUTE STATS`)
- `DESCRIBE COLUMN`
- `DESCRIBE DATABASE`
- `EXPORT TABLE`
- `IMPORT TABLE`
- `SHOW TABLE EXTENDED`
- `SHOW INDEXES`
- `SHOW COLUMNS`

## Semantic Differences Between Impala and HiveQL Features

This section covers instances where Impala and Hive have similar functionality, sometimes including the same syntax, but there are differences in the runtime semantics of those features.

### Security:

Impala utilizes the [Apache Sentry \(incubating\)](#) authorization framework, which provides fine-grained role-based access control to protect data against unauthorized access or tampering.

The Hive component included in CDH 5.1 and higher now includes Sentry-enabled `GRANT`, `REVOKE`, and `CREATE/DROP ROLE` statements. Earlier Hive releases had a privilege system with `GRANT` and `REVOKE` statements.

that were primarily intended to prevent accidental deletion of data, rather than a security mechanism to protect against malicious users.

Impala can make use of privileges set up through Hive GRANT and REVOKE statements. Impala has its own GRANT and REVOKE statements in Impala 2.0 and higher. See [Enabling Sentry Authorization for Impala](#) on page 91 for the details of authorization in Impala, including how to switch from the original policy file-based privilege model to the Sentry service using privileges stored in the metastore database.

#### **SQL statements and clauses:**

The semantics of Impala SQL statements varies from HiveQL in some cases where they use similar SQL statement and clause names:

- Impala uses different syntax and names for query hints, [SHUFFLE] and [NOSHUFFLE] rather than MapJoin or StreamJoin. See [Joins](#) on page 220 for the Impala details.
- Impala does not expose MapReduce specific features of SORT BY, DISTRIBUTE BY, or CLUSTER BY.
- Impala does not require queries to include a FROM clause.

#### **Data types:**

- Impala supports a limited set of implicit casts. This can help avoid undesired results from unexpected casting behavior.
  - Impala does not implicitly cast between string and numeric or Boolean types. Always use CAST() for these conversions.
  - Impala does perform implicit casts among the numeric types, when going from a smaller or less precise type to a larger or more precise one. For example, Impala will implicitly convert a SMALLINT to a BIGINT or FLOAT, but to convert from DOUBLE to FLOAT or INT to TINYINT requires a call to CAST() in the query.
  - Impala does perform implicit casts from string to timestamp. Impala has a restricted set of literal formats for the TIMESTAMP data type and the from\_unixtime() format string; see [TIMESTAMP Data Type](#) on page 130 for details.

See [Data Types](#) on page 111 for full details on implicit and explicit casting for all types, and [Impala Type Conversion Functions](#) on page 267 for details about the CAST() function.

- Impala does not store or interpret timestamps using the local timezone, to avoid undesired results from unexpected time zone issues. Timestamps are stored and interpreted relative to UTC. This difference can produce different results for some calls to similarly named date/time functions between Impala and Hive. See [Impala Date and Time Functions](#) on page 268 for details about the Impala functions. See [TIMESTAMP Data Type](#) on page 130 for a discussion of how Impala handles time zones, and configuration options you can use to make Impala match the Hive behavior more closely when dealing with Parquet-encoded TIMESTAMP data or when converting between the local time zone and UTC.
- The Impala TIMESTAMP type can represent dates ranging from 1400-01-01 to 9999-12-31. This is different from the Hive date range, which is 0000-01-01 to 9999-12-31.
- Impala does not return column overflows as NULL, so that customers can distinguish between NULL data and overflow conditions similar to how they do so with traditional database systems. Impala returns the largest or smallest value in the range for the type. For example, valid values for a tinyint range from -128 to 127. In Impala, a tinyint with a value of -200 returns -128 rather than NULL. A tinyint with a value of 200 returns 127.

#### **Miscellaneous features:**

- Impala does not provide virtual columns.
- Impala does not expose locking.
- Impala does not expose some configuration properties.

## Porting SQL from Other Database Systems to Impala

Although Impala uses standard SQL for queries, you might need to modify SQL source when bringing applications to Impala, due to variations in data types, built-in functions, vendor language extensions, and Hadoop-specific syntax. Even when SQL is working correctly, you might make further minor modifications for best performance.

### Porting DDL and DML Statements

When adapting SQL code from a traditional database system to Impala, expect to find a number of differences in the DDL statements that you use to set up the schema. Clauses related to physical layout of files, tablespaces, and indexes have no equivalent in Impala. You might restructure your schema considerably to account for the Impala partitioning scheme and Hadoop file formats.

Expect SQL queries to have a much higher degree of compatibility. With modest rewriting to address vendor extensions and features not yet supported in Impala, you might be able to run identical or almost-identical query text on both systems.

Therefore, consider separating out the DDL into a separate Impala-specific setup script. Focus your reuse and ongoing tuning efforts on the code for SQL queries.

### Porting Data Types from Other Database Systems

- Change any VARCHAR, VARCHAR2, and CHAR columns to STRING. Remove any length constraints from the column declarations; for example, change VARCHAR(32) or CHAR(1) to STRING. Impala is very flexible about the length of string values; it does not impose any length constraints or do any special processing (such as blank-padding) for STRING columns. (In Impala 2.0 and higher, there are data types VARCHAR and CHAR, with length constraints for both types and blank-padding for CHAR. However, for performance reasons, it is still preferable to use STRING columns where practical.)
- For national language character types such as NCHAR, NVARCHAR, or NCLOB, be aware that while Impala can store and query UTF-8 character data, currently some string manipulation operations only work correctly with ASCII data. See [STRING Data Type](#) on page 128 for details.
- Change any DATE, DATETIME, or TIME columns to TIMESTAMP. Remove any precision constraints. Remove any timezone clauses, and make sure your application logic or ETL process accounts for the fact that Impala expects all TIMESTAMP values to be in [Coordinated Universal Time \(UTC\)](#). See [TIMESTAMP Data Type](#) on page 130 for information about the TIMESTAMP data type, and [Impala Date and Time Functions](#) on page 268 for conversion functions for different date and time formats.

You might also need to adapt date- and time-related literal values and format strings to use the supported Impala date and time formats. If you have date and time literals with different separators or different numbers of YY, MM, and so on placeholders than Impala expects, consider using calls to `regexp_replace()` to transform those values to the Impala-compatible format. See [TIMESTAMP Data Type](#) on page 130 for information about the allowed formats for date and time literals, and [Impala String Functions](#) on page 279 for string conversion functions such as `regexp_replace()`.

Instead of SYSDATE, call the function NOW().

Instead of adding or subtracting directly from a date value to produce a value *N* days in the past or future, use an INTERVAL expression, for example `NOW() + INTERVAL 30 DAYS`.

- Although Impala supports INTERVAL expressions for datetime arithmetic, as shown in [TIMESTAMP Data Type](#) on page 130, INTERVAL is not available as a column data type in Impala. For any INTERVAL values stored in tables, convert them to numeric values that you can add or subtract using the functions in [Impala Date and Time Functions](#) on page 268. For example, if you had a table DEADLINES with an INT column TIME\_PERIOD, you could construct dates *N* days in the future like so:

```
SELECT NOW() + INTERVAL time_period DAYS from deadlines;
```

- For `YEAR` columns, change to the smallest Impala integer type that has sufficient range. See [Data Types](#) on page 111 for details about ranges, casting, and so on for the various numeric data types.
- Change any `DECIMAL` and `NUMBER` types. If fixed-point precision is not required, you can use `FLOAT` or `DOUBLE` on the Impala side depending on the range of values. For applications that require precise decimal values, such as financial data, you might need to make more extensive changes to table structure and application logic, such as using separate integer columns for dollars and cents, or encoding numbers as string values and writing UDFs to manipulate them. See [Data Types](#) on page 111 for details about ranges, casting, and so on for the various numeric data types.
- `FLOAT`, `DOUBLE`, and `REAL` types are supported in Impala. Remove any precision and scale specifications. (In Impala, `REAL` is just an alias for `DOUBLE`; columns declared as `REAL` are turned into `DOUBLE` behind the scenes.) See [Data Types](#) on page 111 for details about ranges, casting, and so on for the various numeric data types.
- Most integer types from other systems have equivalents in Impala, perhaps under different names such as `BIGINT` instead of `INT8`. For any that are unavailable, for example `MEDIUMINT`, switch to the smallest Impala integer type that has sufficient range. Remove any precision specifications. See [Data Types](#) on page 111 for details about ranges, casting, and so on for the various numeric data types.
- Remove any `UNSIGNED` constraints. All Impala numeric types are signed. See [Data Types](#) on page 111 for details about ranges, casting, and so on for the various numeric data types.
- For any types holding bitwise values, use an integer type with enough range to hold all the relevant bits within a positive integer. See [Data Types](#) on page 111 for details about ranges, casting, and so on for the various numeric data types.

For example, `TINYINT` has a maximum positive value of 127, not 256, so to manipulate 8-bit bitfields as positive numbers switch to the next largest type `SMALLINT`.

```
[localhost:21000] > select cast(127*2 as tinyint);
+-----+
| cast(127 * 2 as tinyint) |
+-----+
| -2
+-----+
[localhost:21000] > select cast(128 as tinyint);
+-----+
| cast(128 as tinyint) |
+-----+
| -128
+-----+
[localhost:21000] > select cast(127*2 as smallint);
+-----+
| cast(127 * 2 as smallint) |
+-----+
| 254
+-----+
```

Impala does not support notation such as `b'0101'` for bit literals.

- For `BLOB` values, use `STRING` to represent `CLOB` or `TEXT` types (character based large objects) up to 32 KB in size. Binary large objects such as `BLOB`, `RAW BINARY`, and `VARBINARY` do not currently have an equivalent in Impala.
- For Boolean-like types such as `BOOL`, use the Impala `BOOLEAN` type.
- Because Impala currently does not support composite or nested types, any spatial data types in other database systems do not have direct equivalents in Impala. You could represent spatial values in string format and write UDFs to process them. See [Impala User-Defined Functions \(UDFs\)](#) on page 308 for details. Where practical, separate spatial types into separate tables so that Impala can still work with the non-spatial data.
- Take out any `DEFAULT` clauses. Impala can use data files produced from many different sources, such as Pig, Hive, or MapReduce jobs. The fast import mechanisms of `LOAD DATA` and external tables mean that Impala

## Impala SQL Language Reference

is flexible about the format of data files, and Impala does not necessarily validate or cleanse data before querying it. When copying data through Impala `INSERT` statements, you can use conditional functions such as `CASE` or `NVL` to substitute some other value for `NULL` fields; see [Impala Conditional Functions](#) on page 275 for details.

- Take out any constraints from your `CREATE TABLE` and `ALTER TABLE` statements, for example `PRIMARY KEY`, `FOREIGN KEY`, `UNIQUE`, `NOT NULL`, `UNSIGNED`, or `CHECK` constraints. Impala can use data files produced from many different sources, such as Pig, Hive, or MapReduce jobs. Therefore, Impala expects initial data validation to happen earlier during the ETL or ELT cycle. After data is loaded into Impala tables, you can perform queries to test for `NULL` values. When copying data through Impala `INSERT` statements, you can use conditional functions such as `CASE` or `NVL` to substitute some other value for `NULL` fields; see [Impala Conditional Functions](#) on page 275 for details.

Do as much verification as practical before loading data into Impala. After data is loaded into Impala, you can do further verification using SQL queries to check if values have expected ranges, if values are `NULL` or not, and so on. If there is a problem with the data, you will need to re-run earlier stages of the ETL process, or do an `INSERT ... SELECT` statement in Impala to copy the faulty data to a new table and transform or filter out the bad values.

- Take out any `CREATE INDEX`, `DROP INDEX`, and `ALTER INDEX` statements, and equivalent `ALTER TABLE` statements. Remove any `INDEX`, `KEY`, or `PRIMARY KEY` clauses from `CREATE TABLE` and `ALTER TABLE` statements. Impala is optimized for bulk read operations for data warehouse-style queries, and therefore does not support indexes for its tables.
- Calls to built-in functions with out-of-range or otherwise incorrect arguments, return `NULL` in Impala as opposed to raising exceptions. (This rule applies even when the `ABORT_ON_ERROR=true` query option is in effect.) Run small-scale queries using representative data to doublecheck that calls to built-in functions are returning expected values rather than `NULL`. For example, unsupported `CAST` operations do not raise an error in Impala:

```
select cast('foo' as int);
+-----+
| cast('foo' as int) |
+-----+
| NULL             |
+-----+
```

- For any other type not supported in Impala, you could represent their values in string format and write UDFs to process them. See [Impala User-Defined Functions \(UDFs\)](#) on page 308 for details.
- To detect the presence of unsupported or unconvertable data types in data files, do initial testing with the `ABORT_ON_ERROR=true` query option in effect. This option causes queries to fail immediately if they encounter disallowed type conversions. See [ABORT\\_ON\\_ERROR Query Option](#) on page 339 for details. For example:

```
set abort_on_error=true;
select count(*) from (select * from t1);
-- The above query will fail if the data files for T1 contain any
-- values that can't be converted to the expected Impala data types.
-- For example, if T1.C1 is defined as INT but the column contains
-- floating-point values like 1.1, the query will return an error.
```

## SQL Statements to Remove or Adapt

Some SQL statements or clauses that you might be familiar with are not currently supported in Impala:

- Impala has no `DELETE` statement. Impala is intended for data warehouse-style operations where you do bulk moves and transforms of large quantities of data. Instead of using `DELETE`, use `INSERT OVERWRITE` to entirely replace the contents of a table or partition, or use `INSERT ... SELECT` to copy a subset of data (everything but the rows you intended to delete) from one table to another. See [DML Statements](#) on page 163 for an overview of Impala DML statements.

- Impala has no `UPDATE` statement. Impala is intended for data warehouse-style operations where you do bulk moves and transforms of large quantities of data. Instead of using `UPDATE`, do all necessary transformations early in the ETL process, such as in the job that generates the original data, or when copying from one table to another to convert to a particular file format or partitioning scheme. See [DML Statements](#) on page 163 for an overview of Impala DML statements.
- Impala has no transactional statements, such as `COMMIT` or `ROLLBACK`. Impala effectively works like the `AUTOCOMMIT` mode in some database systems, where changes take effect as soon as they are made.
- If your database, table, column, or other names conflict with Impala reserved words, use different names or quote the names with backticks. See [Impala Reserved Words](#) on page 453 for the current list of Impala reserved words.

Conversely, if you use a keyword that Impala does not recognize, it might be interpreted as a table or column alias. For example, in `SELECT * FROM t1 NATURAL JOIN t2`, Impala does not recognize the `NATURAL` keyword and interprets it as an alias for the table `t1`. If you experience any unexpected behavior with queries, check the list of reserved words to make sure all keywords in join and `WHERE` clauses are recognized.

- Impala supports subqueries only in the `FROM` clause of a query, not within the `WHERE` clauses. Therefore, you cannot use clauses such as `WHERE column IN (subquery)`. Also, Impala does not allow `EXISTS` or `NOT EXISTS` clauses (although `EXISTS` is a reserved keyword).
- Impala supports `UNION` and `UNION ALL` set operators, but not `INTERSECT`. Prefer `UNION ALL` over `UNION` when you know the data sets are disjoint or duplicate values are not a problem; `UNION ALL` is more efficient because it avoids materializing and sorting the entire result set to eliminate duplicate values.
- Within queries, Impala requires query aliases for any subqueries:

```
-- Without the alias 'contents_of_t1' at the end, query gives syntax error.
select count(*) from (select * from t1) contents_of_t1;
```

- When an alias is declared for an expression in a query, that alias cannot be referenced again within the same query block:

```
-- Can't reference AVERAGE twice in the SELECT list where it's defined.
select avg(x) as average, average+1 from t1 group by x;
ERROR: AnalysisException: couldn't resolve column reference: 'average'

-- Although it can be referenced again later in the same query.
select avg(x) as average from t1 group by x having average > 3;
```

For Impala, either repeat the expression again, or abstract the expression into a `WITH` clause, creating named columns that can be referenced multiple times anywhere in the base query:

```
-- The following 2 query forms are equivalent.
select avg(x) as average, avg(x)+1 from t1 group by x;
with avg_t as (select avg(x) average from t1 group by x) select average, average+1
from avg_t;
```

- Impala does not support certain rarely used join types that are less appropriate for high-volume tables used for data warehousing. In some cases, Impala supports join types but requires explicit syntax to ensure you do not do inefficient joins of huge tables by accident. For example, Impala does not support natural joins or anti-joins, and requires the `CROSS JOIN` operator for Cartesian products. See [Joins](#) on page 220 for details on the syntax for Impala join clauses.
- Impala has a limited choice of partitioning types. Partitions are defined based on each distinct combination of values for one or more partition key columns. Impala does not redistribute or check data to create evenly distributed partitions; you must choose partition key columns based on your knowledge of the data volume and distribution. Adapt any tables that use range, list, hash, or key partitioning to use the Impala partition syntax for `CREATE TABLE` and `ALTER TABLE` statements. Impala partitioning is similar to range partitioning where every range has exactly one value, or key partitioning where the hash function produces a separate

bucket for every combination of key values. See [Partitioning for Impala Tables](#) on page 394 for usage details, and [CREATE TABLE Statement](#) on page 180 and [ALTER TABLE Statement](#) on page 164 for syntax.

- **Note:** Because the number of separate partitions is potentially higher than in other database systems, keep a close eye on the number of partitions and the volume of data in each one; scale back the number of partition key columns if you end up with too many partitions with a small volume of data in each one. Remember, to distribute work for a query across a cluster, you need at least one HDFS block per node. HDFS blocks are typically multiple megabytes, especially for Parquet files. Therefore, if each partition holds only a few megabytes of data, you are unlikely to see much parallelism in the query because such a small amount of data is typically processed by a single node.
- For "top-N" queries, Impala uses the `LIMIT` clause rather than comparing against a pseudocolumn named `ROWNUM` or `ROW_NUM`. See [LIMIT Clause](#) on page 230 for details.

## SQL Constructs to Doublecheck

Some SQL constructs that are supported have behavior or defaults more oriented towards convenience than optimal performance. Also, sometimes machine-generated SQL, perhaps issued through JDBC or ODBC applications, might have inefficiencies or exceed internal Impala limits. As you port SQL code, be alert and change these things where appropriate:

- A `CREATE TABLE` statement with no `STORED AS` clause creates data files in plain text format, which is convenient for data interchange but not a good choice for high-volume data with high-performance queries. See [How Impala Works with Hadoop File Formats](#) on page 399 for why and how to use specific file formats for compact data and high-performance queries. Especially see [Using the Parquet File Format with Impala Tables](#) on page 407, for details about the file format most heavily optimized for large-scale data warehouse queries.
- A `CREATE TABLE` statement with no `PARTITIONED BY` clause stores all the data files in the same physical location, which can lead to scalability problems when the data volume becomes large.

On the other hand, adapting tables that were already partitioned in a different database system could produce an Impala table with a high number of partitions and not enough data in each one, leading to underutilization of Impala's parallel query features.

See [Partitioning for Impala Tables](#) on page 394 for details about setting up partitioning and tuning the performance of queries on partitioned tables.

- The `INSERT ... VALUES` syntax is suitable for setting up toy tables with a few rows for functional testing, but because each such statement creates a separate tiny file in HDFS, it is not a scalable technique for loading megabytes or gigabytes (let alone petabytes) of data. Consider revising your data load process to produce raw data files outside of Impala, then setting up Impala external tables or using the `LOAD DATA` statement to use those data files instantly in Impala tables, with no conversion or indexing stage. See [External Tables](#) on page 157 and [LOAD DATA Statement](#) on page 212 for details about the Impala techniques for working with data files produced outside of Impala; see [Data Loading and Querying Examples](#) on page 61 for examples of ETL workflow for Impala.
- If your ETL process is not optimized for Hadoop, you might end up with highly fragmented small data files, or a single giant data file that cannot take advantage of distributed parallel queries or partitioning. In this case, use an `INSERT ... SELECT` statement to copy the data into a new table and reorganize into a more efficient layout in the same operation. See [INSERT Statement](#) on page 202 for details about the `INSERT` statement.

You can do `INSERT ... SELECT` into a table with a more efficient file format (see [How Impala Works with Hadoop File Formats](#) on page 399) or from an unpartitioned table into a partitioned one (see [Partitioning for Impala Tables](#) on page 394).

- The number of expressions allowed in an Impala query might be smaller than for some other database systems, causing failures for very complicated queries (typically produced by automated SQL generators). Where practical, keep the number of expressions in the WHERE clauses to approximately 2000 or fewer. As a workaround, set the query option `DISABLE_CODEGEN=true` if queries fail for this reason. See [DISABLE\\_CODEGEN Query Option](#) on page 342 for details.
- If practical, rewrite UNION queries to use the `UNION ALL` operator instead. Prefer `UNION ALL` over `UNION` when you know the data sets are disjoint or duplicate values are not a problem; `UNION ALL` is more efficient because it avoids materializing and sorting the entire result set to eliminate duplicate values.

## Next Porting Steps after Verifying Syntax and Semantics

Throughout this section, some of the decisions you make during the porting process also have a substantial impact on performance. After your SQL code is ported and working correctly, doublecheck the performance-related aspects of your schema design, physical layout, and queries to make sure that the ported application is taking full advantage of Impala's parallelism, performance-related SQL features, and integration with Hadoop components.

- Have you run the `COMPUTE STATS` statement on each table involved in join queries? Have you also run `COMPUTE STATS` for each table used as the source table in an `INSERT ... SELECT` or `CREATE TABLE AS SELECT` statement?
- Are you using the most efficient file format for your data volumes, table structure, and query characteristics?
- Are you using partitioning effectively? That is, have you partitioned on columns that are often used for filtering in WHERE clauses? Have you partitioned at the right granularity so that there is enough data in each partition to parallelize the work for each query?
- Does your ETL process produce a relatively small number of multi-megabyte data files (good) rather than a huge number of small files (bad)?

See [Tuning Impala for Performance](#) on page 354 for details about the whole performance tuning process.

# Using the Impala Shell (impala-shell Command)

You can use the Impala shell tool (`impala-shell`) to set up databases and tables, insert data, and issue queries. For ad hoc queries and exploration, you can submit SQL statements in an interactive session. To automate your work, you can specify command-line options to process a single statement or a script file. The `impala-shell` interpreter accepts all the same SQL statements listed in [SQL Statements](#) on page 162, plus some shell-only commands that you can use for tuning performance and diagnosing problems.

The `impala-shell` command fits into the familiar Unix toolchain:

- The `-q` option lets you issue a single query from the command line, without starting the interactive interpreter. You could use this option to run `impala-shell` from inside a shell script or with the command invocation syntax from a Python, Perl, or other kind of script.
- The `-o` option lets you save query output to a file.
- The `-B` option turns off pretty-printing, so that you can produce comma-separated, tab-separated, or other delimited text files as output. (Use the `--output_delimiter` option to choose the delimiter character; the default is the tab character.)
- In non-interactive mode, query output is printed to `stdout` or to the file specified by the `-o` option, while incidental output is printed to `stderr`, so that you can process just the query output as part of a Unix pipeline.
- In interactive mode, `impala-shell` uses the `readline` facility to recall and edit previous commands.

For information on installing the Impala shell, see [Installing Impala](#) on page 30. In Cloudera Manager 4.1 and higher, Cloudera Manager installs `impala-shell` automatically. You might install `impala-shell` manually on other systems not managed by Cloudera Manager, so that you can issue queries from client systems that are not also running the Impala daemon or other Apache Hadoop components.

For information about establishing a connection to a DataNode running the `impalad` daemon through the `impala-shell` command, see [Connecting to impalad through impala-shell](#) on page 335.

For a list of the `impala-shell` command-line options, see [impala-shell Configuration Options](#) on page 332. For reference information about the `impala-shell` interactive commands, see [impala-shell Command Reference](#) on page 337.

## impala-shell Configuration Options

You can specify the following options when starting the `impala-shell` command to change how shell commands are executed. The table shows the format to use when specifying each option on the command line, or through the `$HOME/.impalarc` configuration file.

▪ **Note:**

These options are different than the configuration options for the `impalad` daemon itself. For the `impalad` options, see [Modifying Impala Startup Options](#) on page 48.

## Summary of impala-shell Configuration Options

The following table shows the names and allowed arguments for the `impala-shell` configuration options. You can specify options on the command line, or in a configuration file as described in [impala-shell Configuration File](#) on page 334.

Command-Line Option	Configuration File Setting	Explanation
<code>-B</code> or <code>--delimited</code>	<code>write_delimited=true</code>	Causes all query results to be printed in plain format as a delimited text file. Useful for producing data files to be used with other Hadoop components. Also useful

Command-Line Option	Configuration File Setting	Explanation
		for avoiding the performance overhead of pretty-printing all output, especially when running benchmark tests using queries returning large result sets. Specify the delimiter character with the <code>--output_delimiter</code> option. Store all query results in a file rather than printing to the screen with the <code>-B</code> option. Added in Impala 1.0.1.
<code>--print_header</code>	<code>print_header=true</code>	
<code>-o filename</code> or <code>--output_file filename</code>	<code>output_file=filename</code>	Stores all query results in the specified file. Typically used to store the results of a single query issued from the command line with the <code>-q</code> option. Also works for interactive sessions; you see the messages such as number of rows fetched, but not the actual result set. To suppress these incidental messages when combining the <code>-q</code> and <code>-o</code> options, redirect <code>stderr</code> to <code>/dev/null</code> . Added in Impala 1.0.1.
<code>--output_delimiter=character</code>	<code>output_delimiter=character</code>	Specifies the character to use as a delimiter between fields when query results are printed in plain format by the <code>-B</code> option. Defaults to tab (' <code>\t</code> '). If an output value contains the delimiter character, that field is quoted and/or escaped. Added in Impala 1.0.1.
<code>-p</code> or <code>--show_profiles</code>	<code>show_profiles=true</code>	Displays the query execution plan (same output as the <code>EXPLAIN</code> statement) and a more detailed low-level breakdown of execution steps, for every query executed by the shell.
<code>-h</code> or <code>--help</code>	N/A	Displays help information.
<code>-i hostname</code> or <code>--impalad=hostname[:portnum]</code>	<code>impalad=hostname[:portnum]</code>	Connects to the <code>impalad</code> daemon on the specified host. The default port of 21000 is assumed unless you provide another value. You can connect to any host in your cluster that is running <code>impalad</code> . If you connect to an instance of <code>impalad</code> that was started with an alternate port specified by the <code>--fe_port</code> flag, provide that alternative port.
<code>-q query</code> or <code>--query=query</code>	<code>query=query</code>	Passes a query or other <code>impala-shell</code> command from the command line. The <code>impala-shell</code> interpreter immediately exits after processing the statement. It is limited to a single statement, which could be a <code>SELECT</code> , <code>CREATE TABLE</code> , <code>SHOW TABLES</code> , or any other statement recognized in <code>impala-shell</code> . Because you cannot pass a <code>USE</code> statement and another query, fully qualify the names for any tables outside the default database. (Or use the <code>-f</code> option to pass a file with a <code>USE</code> statement followed by other queries.)
<code>-f query_file</code> or <code>--query_file=query_file</code>	<code>query_file=path_to_query_file</code>	Passes a SQL query from a file. Multiple statements must be semicolon ( <code>;</code> ) delimited.
<code>-k</code> or <code>--kerberos</code>	<code>use_kerberos=true</code>	Kerberos authentication is used when the shell connects to <code>impalad</code> . If Kerberos is not enabled on the

## Using the Impala Shell (impala-shell Command)

Command-Line Option	Configuration File Setting	Explanation
		instance of <code>impalad</code> to which you are connecting, errors are displayed.
<code>-s kerberos_service_name</code> or <code>--kerberos_service_name=name</code>	<code>kerberos_service_name=name</code>	Instructs <code>impala-shell</code> to authenticate to a particular <code>impalad</code> service principal. If a <code>kerberos_service_name</code> is not specified, <code>impala</code> is used by default. If this option is used in conjunction with a connection in which Kerberos is not supported, errors are returned.
<code>-V</code> or <code>--verbose</code>	<code>verbose=true</code>	Enables verbose output.
<code>--quiet</code>	<code>verbose=false</code>	Disables verbose output.
<code>-v</code> or <code>--version</code>	<code>version=true</code>	Displays version information.
<code>-c</code>	<code>ignore_query_failure=true</code>	Continues on query failure.
<code>-r</code> or <code>--refresh_after_connect</code>	<code>refresh_after_connect=true</code>	Updates Impala metadata upon connection. Same as running the <code>INVALIDATE_METADATA</code> statement after connecting. (This option was originally named when the <code>REFRESH</code> statement did the extensive metadata updates now performed by <code>INVALIDATE_METADATA</code> .)
<code>-d default_db</code> or <code>--database=default_db</code>	<code>default_db=default_db</code>	Specifies the database to be used on startup. Same as running the <code>USE</code> statement after connecting. If not specified, a database named <code>DEFAULT</code> is used.
<code>-ssl</code>	<code>ssl=true</code>	Enables SSL for <code>impala-shell</code> .
<code>--ca_cert=path_to_certificate</code>	<code>ca_cert=path_to_certificate</code>	The local pathname pointing to the third-party CA certificate, or to a copy of the server certificate for self-signed server certificates. If <code>--ca_cert</code> is not set, <code>impala-shell</code> enables SSL, but does not validate the server certificate. This is useful for connecting to a known-good Impala that is only running over SSL, when a copy of the certificate is not available (such as when debugging customer installations).
<code>-l</code>	<code>use_ldap=true</code>	Enables LDAP authentication.
<code>-u</code>	<code>user=user_name</code>	Supplies the user name, when LDAP authentication is enabled by the <code>-l</code> option. (Specify the short user name, not the full LDAP distinguished name.) The shell then prompts interactively for the password.
<code>--config_file=path_to_config_file</code>	N/A	Specifies the path of the file containing <code>impala-shell</code> configuration settings. The default is <code>\$HOME/.impalarc</code> . This setting can only be specified on the command line.

### impala-shell Configuration File

You can define a set of default options for your `impala-shell` environment, stored in the file `$HOME/.impalarc`. This file consists of key-value pairs, one option per line. Everything after a `#` character on a line is treated as a comment and ignored.

The configuration file must contain a header label [`impala`], followed by the options specific to `impala-shell`. (This standard convention for configuration files lets you use a single file to hold configuration options for multiple applications.)

To specify a different filename or path for the configuration file, specify the argument `--config_file=path_to_config_file` on the `impala-shell` command line.

The names of the options in the configuration file are similar (although not necessarily identical) to the long-form command-line arguments to the `impala-shell` command. For the names to use, see [Summary of impala-shell Configuration Options](#) on page 332.

Any options you specify on the `impala-shell` command line override any corresponding options within the configuration file.

The following example shows a configuration file that you might use during benchmarking tests. It sets verbose mode, so that the output from each SQL query is followed by timing information. `impala-shell` starts inside the database containing the tables with the benchmark data, avoiding the need to issue a `USE` statement or use fully qualified table names.

In this example, the query output is formatted as delimited text rather than enclosed in ASCII art boxes, and is stored in a file rather than printed to the screen. Those options are appropriate for benchmark situations, so that the overhead of `impala-shell` formatting and printing the result set does not factor into the timing measurements. It also enables the `show_profiles` option. That option prints detailed performance information after each query, which might be valuable in understanding the performance of benchmark queries.

```
[impala]
verbose=true
default_db=tpc_benchmarking
write_delimited=true
output_delimiter=,
output_file=/home/tester1/benchmark_results.csv
show_profiles=true
```

The following example shows a configuration file that connects to a specific remote Impala node, runs a single query within a particular database, then exits. You would typically use this kind of single-purpose configuration setting with the `impala-shell` command-line option `--config_file=path_to_config_file`, to easily select between many predefined queries that could be run against different databases, hosts, or even different clusters. To run a sequence of statements instead of a single query, specify the configuration option `query_file=path_to_query_file` instead.

```
[impala]
impalad=impala-test-node1.example.com
default_db=site_stats
# Issue a predefined query and immediately exit.
query=select count(*) from web_traffic where event_date = trunc(now(), 'dd')
```

## Connecting to impalad through impala-shell

Within an `impala-shell` session, you can only issue queries while connected to an instance of the `impalad` daemon. You can specify the connection information through command-line options when you run the `impala-shell` command, or during an `impala-shell` session by issuing a `CONNECT` command. You can connect to any DataNode where an instance of `impalad` is running, and that node coordinates the execution of all queries sent to it.

For simplicity, you might always connect to the same node, perhaps running `impala-shell` on the same node as `impalad` and specifying the host name as `localhost`. Routing all SQL statements to the same node can help to avoid issuing frequent `REFRESH` statements, as is necessary when table data or metadata is updated through a different node.

For load balancing or general flexibility, you might connect to an arbitrary node for each `impala-shell` session. In this case, depending on whether table data or metadata might have been updated through another node, you

## Using the Impala Shell (impala-shell Command)

might issue a `REFRESH` statement to bring the metadata for all tables up to date on this node (for a long-lived session that will query many tables) or issue specific `REFRESH table_name` statements just for the tables you intend to query.

To connect the Impala shell to any DataNode with an `impalad` daemon:

1. Start the Impala shell with no connection:

```
$ impala-shell
```

You should see a prompt like the following:

```
Welcome to the Impala shell. Press TAB twice to see a list of available commands.  
Copyright (c) year Cloudera, Inc. All rights reserved.  
(Shell  
    build version: Impala Shell v2.2.x (hash) built on  
    date)  
[Not connected] >
```

2. Use the `connect` command to connect to an Impala instance. Enter a command of the form:

```
[Not connected] > connect impalad-host  
[impalad-host:21000] >
```

- **Note:** Replace `impalad-host` with the host name you have configured for any DataNode running Impala in your environment. The changed prompt indicates a successful connection.

## Running Commands in impala-shell

For information on available commands, see [impala-shell Command Reference](#) on page 337. You can see the full set of available commands by pressing TAB twice, for example:

```
[impalad-host:21000] >  
connect  describe  explain  help      history  insert   quit      refresh  select  
        set      shell    show     use       version  
[impalad-host:21000] >
```

- **Note:** Commands must be terminated by a semi-colon. A command can span multiple lines.

For example:

```
[localhost:21000] > select *  
      > from t1  
      > limit 5;  
+-----+-----+  
| s1   | s2   |  
+-----+-----+  
| hello | world |  
| goodbye | cleveland |  
+-----+-----+
```

A comment is considered part of the statement it precedes, so when you enter a `--` or `/* */` comment, you get a continuation prompt until you finish entering a statement ending with a semicolon:

```
[localhost:21000] > -- This is a test comment  
      > show tables like 't*';  
+-----+  
| name |
```

t1
t2
tab1
tab2
tab3
text_t

Use the up-arrow and down-arrow keys to cycle through and edit previous commands. `impala-shell` uses the readline library and so supports a standard set of keyboard shortcuts for editing and cursor movement, such as `Ctrl-A` for beginning of line and `Ctrl-E` for end of line.

## impala-shell Command Reference

Use the following commands within `impala-shell` to pass requests to the `impalad` daemon that the shell is connected to. You can enter a command interactively at the prompt, or pass it as the argument to the `-q` option of `impala-shell`. Most of these commands are passed to the Impala daemon as SQL statements; refer to the corresponding [SQL language reference sections](#) for full syntax details.

Command	Explanation
alter	Changes the underlying structure or settings of an Impala table, or a table shared between Impala and Hive. See <a href="#">ALTER TABLE Statement</a> on page 164 and <a href="#">ALTER VIEW Statement</a> on page 168 for details.
compute stats	Gathers important performance-related information for a table, used by Impala to optimize queries. See <a href="#">COMPUTE STATS Statement</a> on page 170 for details.
connect	Connects to the specified instance of <code>impalad</code> . The default port of 21000 is assumed unless you provide another value. You can connect to any host in your cluster that is running <code>impalad</code> . If you connect to an instance of <code>impalad</code> that was started with an alternate port specified by the <code>--fe_port</code> flag, you must provide that alternate port. See <a href="#">Connecting to impalad through impala-shell</a> on page 335 for examples.  The <code>SET</code> statement has no effect until the <code>impala-shell</code> interpreter is connected to an Impala server. Once you are connected, any query options you set remain in effect as you issue a subsequent <code>CONNECT</code> command to connect to a different Impala host.
describe	Shows the columns, column data types, and any column comments for a specified table. <code>DESCRIBE FORMATTED</code> shows additional information such as the HDFS data directory, partitions, and internal properties for the table. See <a href="#">DESCRIBE Statement</a> on page 188 for details about the basic <code>DESCRIBE</code> output and the <code>DESCRIBE FORMATTED</code> variant. You can use <code>DESC</code> as shorthand for the <code>DESCRIBE</code> command.
drop	Removes a schema object, and in some cases its associated data files. See <a href="#">DROP TABLE Statement</a> on page 197, <a href="#">DROP VIEW Statement</a> on page 198, <a href="#">DROP DATABASE Statement</a> on page 192, and <a href="#">DROP FUNCTION Statement</a> on page 192 for details.
explain	Provides the execution plan for a query. <code>EXPLAIN</code> represents a query as a series of steps. For example, these steps might be map/reduce stages, metastore operations, or file system operations such as move or rename. See <a href="#">EXPLAIN Statement</a> on page 199 and <a href="#">Using the EXPLAIN Plan for Performance Tuning</a> on page 379 for details.
help	Help provides a list of all available commands and options.
history	Maintains an enumerated cross-session command history. This history is stored in the <code>~/.impalahistory</code> file.

## Using the Impala Shell (impala-shell Command)

Command	Explanation
insert	Writes the results of a query to a specified table. This either overwrites table data or appends data to the existing table content. See <a href="#">INSERT Statement</a> on page 202 for details.
invalidate metadata	Updates impalad metadata for table existence and structure. Use this command after creating, dropping, or altering databases, tables, or partitions in Hive. See <a href="#">INVALIDATE METADATA Statement</a> on page 210 for details.
profile	Displays low-level information about the most recent query. Used for performance diagnosis and tuning. The report starts with the same information as produced by the EXPLAIN statement and the SUMMARY command. See <a href="#">Using the Query Profile for Performance Tuning</a> on page 380 for details.
quit	Exits the shell. Remember to include the final semicolon so that the shell recognizes the end of the command.
refresh	Refreshes impalad metadata for the locations of HDFS blocks corresponding to Impala data files. Use this command after loading new data files into an Impala table through Hive or through HDFS commands. See <a href="#">REFRESH Statement</a> on page 215 for details.
select	Specifies the data set on which to complete some action. All information returned from select can be sent to some output such as the console or a file or can be used to complete some other element of query. See <a href="#">SELECT Statement</a> on page 218 for details.
set	Manages query options for an impala-shell session. The available options are the ones listed in <a href="#">Query Options for the SET Statement</a> on page 339. These options are used for query tuning and troubleshooting. Issue SET with no arguments to see the current query options, either based on the impalad defaults, as specified by you at impalad startup, or based on earlier SET statements in the same session. To modify option values, issue commands with the syntax set option=value. To restore an option to its default, use the unset command. Some options take Boolean values of true and false. Others take numeric arguments, or quoted string values.  The SET statement has no effect until the impala-shell interpreter is connected to an Impala server. Once you are connected, any query options you set remain in effect as you issue a subsequent CONNECT command to connect to a different Impala host.  In Impala 2.0 and later, SET is available as a SQL statement for any kind of application, not only through impala-shell. See <a href="#">SET Statement</a> on page 241 for details.
shell	Executes the specified command in the operating system shell without exiting impala-shell. You can use the ! character as shorthand for the shell command. <div style="border: 1px solid #ccc; padding: 10px; margin-top: 10px;"><p>▪ <b>Note:</b> Quote any instances of the -- or /* tokens to avoid them being interpreted as the start of a comment. To embed comments within source or ! commands, use the shell comment character # before the comment portion of the line.</p></div>
show	Displays metastore data for schema objects created and accessed through Impala, Hive, or both. show can be used to gather information about objects such as databases, tables, and functions. See <a href="#">SHOW Statement</a> on page 242 for details.
summary	Summarizes the work performed in various stages of a query. It provides a higher-level view of the information displayed by the EXPLAIN command. Added in Impala 1.4.0. See <a href="#">Using the SUMMARY Report for Performance Tuning</a> on page 379 for details.

Command	Explanation
unset	Removes any user-specified value for a query option and returns the option to its default value. See <a href="#">Query Options for the SET Statement</a> on page 339 for the available query options.
use	Indicates the database against which to execute subsequent commands. Lets you avoid using fully qualified names when referring to tables in databases other than default. See <a href="#">USE Statement</a> on page 256 for details. Not effective with the -q option, because that option only allows a single statement in the argument.
version	Returns Impala version information.

## Query Options for the SET Statement

You can specify the following options using the `SET` statement, and those settings affect all queries issued from that session.

Some query options are useful in day-to-day operations for improving usability, performance, or flexibility.

Other query options control special-purpose aspects of Impala operation and are intended primarily for advanced debugging or troubleshooting.

Options with Boolean parameters can be set to 1 or `true` to enable, or 0 or `false` to turn off.

- **Note:** In Impala 2.0 and later, you can set query options directly through the JDBC and ODBC interfaces by using the `SET` statement. Formerly, `SET` was only available as a command within the `impala-shell` interpreter.

### Related information:

[SET Statement](#) on page 241

## ABORT\_ON\_DEFAULT\_LIMIT\_EXCEEDED Query Option

Now that the `ORDER BY` clause no longer requires an accompanying `LIMIT` clause in Impala 1.4.0 and higher, this query option is deprecated and has no effect.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of `SET` statement)

## ABORT\_ON\_ERROR Query Option

When this option is enabled, Impala cancels a query immediately when any of the nodes encounters an error, rather than continuing and possibly returning incomplete results. This option is disabled by default, to help gather maximum diagnostic information when an error occurs, for example, whether the same problem occurred on all nodes or only a single node. Currently, the errors that Impala can skip over involve data corruption, such as a column that contains a string value when expected to contain an integer value.

To control how much logging Impala does for non-fatal errors when `ABORT_ON_ERROR` is turned off, use the `MAX_ERRORS` option.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of `SET` statement)

### Related information:

[MAX\\_ERRORS Query Option](#) on page 349, [Using Impala Logging](#) on page 443

## Using the Impala Shell (impala-shell Command)

### ALLOW\_UNSUPPORTED\_FORMATS Query Option

An obsolete query option from early work on support for file formats. Do not use. Might be removed in the future.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of `SET` statement)

### APPX\_COUNT\_DISTINCT Query Option

Allows multiple `COUNT(DISTINCT)` operations within a single query, by internally rewriting each `COUNT(DISTINCT)` to use the `NDV()` function. The resulting count is approximate rather than precise.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of `SET` statement)

#### Examples:

The following examples show how the `APPX_COUNT_DISTINCT` lets you work around the restriction where a query can only evaluate `COUNT(DISTINCT col_name)` for a single column. By default, you can count the distinct values of one column or another, but not both in a single query:

```
[localhost:21000] > select count(distinct x) from int_t;
+-----+
| count(distinct x) |
+-----+
| 10               |
+-----+
[localhost:21000] > select count(distinct property) from int_t;
+-----+
| count(distinct property) |
+-----+
| 7                 |
+-----+
[localhost:21000] > select count(distinct x), count(distinct property) from int_t;
ERROR: AnalysisException: all DISTINCT aggregate functions need to have the same set
of parameters
as count(DISTINCT x); deviating function: count(DISTINCT property)
```

When you enable the `APPX_COUNT_DISTINCT` query option, now the query with multiple `COUNT(DISTINCT)` works. The reason this behavior requires a query option is that each `COUNT(DISTINCT)` is rewritten internally to use the `NDV()` function instead, which provides an approximate result rather than a precise count.

```
[localhost:21000] > set APPX_COUNT_DISTINCT=true;
[localhost:21000] > select count(distinct x), count(distinct property) from int_t;
+-----+-----+
| count(distinct x) | count(distinct property) |
+-----+-----+
| 10            | 7              |
+-----+-----+
```

#### Related information:

[COUNT Function](#) on page 287, [DISTINCT Operator](#) on page 238, [NDV Function](#) on page 291

### BATCH\_SIZE Query Option

Number of rows evaluated at a time by SQL operators. Unspecified or a size of 0 uses a predefined default size. Primarily for Cloudera testing.

**Type:** numeric

**Default:** 0 (meaning 1024)

## COMPRESSION\_CODEC Query Option

When Impala writes Parquet data files using the `INSERT` statement, the underlying compression is controlled by the `COMPRESSION_CODEC` query option.

- **Note:** Prior to Impala 2.0, this option was named `PARQUET_COMPRESSION_CODEC`. In Impala 2.0 and later, the `PARQUET_COMPRESSION_CODEC` name is not recognized. Use the more general name `COMPRESSION_CODEC` for new code.

### Syntax:

```
SET COMPRESSION_CODEC=codec_name;
```

The allowed values for this query option are `SNAPPY` (the default), `GZIP`, and `NONE`.

- **Note:** A Parquet file created with `COMPRESSION_CODEC=NONE` is still typically smaller than the original data, due to encoding schemes such as run-length encoding and dictionary encoding that are applied separately from compression.

The option value is not case-sensitive.

If the option is set to an unrecognized value, all kinds of queries will fail due to the invalid option setting, not just queries involving Parquet tables. (The value `BZIP2` is also recognized, but is not compatible with Parquet tables.)

**Type:** STRING

**Default:** SNAPPY

### Examples:

```
set compression_codec=gzip;
insert into parquet_table_highly_compressed select * from t1;

set compression_codec=snappy;
insert into parquet_table_compression_plus_fast_queries select * from t1;

set compression_codec=none;
insert into parquet_table_no_compression select * from t1;

set compression_codec=foo;
select * from t1 limit 5;
ERROR: Invalid compression codec: foo
```

### Related information:

For information about how compressing Parquet data files affects query performance, see [Snappy and GZip Compression for Parquet Data Files](#) on page 411.

## DEBUG\_ACTION Query Option

Introduces artificial problem conditions within queries. For internal Cloudera debugging and troubleshooting.

**Type:** STRING

**Default:** empty string

## DEFAULT\_ORDER\_BY\_LIMIT Query Option

Now that the `ORDER BY` clause no longer requires an accompanying `LIMIT` clause in Impala 1.4.0 and higher, this query option is deprecated and has no effect.

Prior to Impala 1.4.0, Impala queries that use the `ORDER_BY` clause must also include a `LIMIT` clause, to avoid accidentally producing huge result sets that must be sorted. Sorting a huge result set is a memory-intensive

## Using the Impala Shell (impala-shell Command)

operation. In Impala 1.4.0 and higher, Impala uses a temporary disk work area to perform the sort if that operation would otherwise exceed the Impala memory limit on a particular host.

**Type:** numeric

**Default:** -1 (no default limit)

### DISABLE\_CODEGEN Query Option

This is a debug option, intended for diagnosing and working around issues that cause crashes. If a query fails with an “illegal instruction” or other hardware-specific message, try setting `DISABLE_CODEGEN=true` and running the query again. If the query succeeds only when the `DISABLE_CODEGEN` option is turned on, submit the problem to Cloudera support and include that detail in the problem report. Do not otherwise run with this setting turned on, because it results in lower overall performance.

Because the code generation phase adds a small amount of overhead for each query, you might turn on the `DISABLE_CODEGEN` option to achieve maximum throughput when running many short-lived queries against small tables.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of `SET` statement)

### DISABLE\_UNSAFE\_SPILLS Query Option

Enable this option if you prefer to have queries fail when they exceed the Impala memory limit, rather than write temporary data to disk.

Queries that “spill” to disk typically complete successfully, when in earlier Impala releases they would have failed. However, queries with exorbitant memory requirements due to missing statistics or inefficient join clauses could become so slow as a result that you would rather have them cancelled automatically and reduce the memory usage through standard Impala tuning techniques.

This option prevents only “unsafe” spill operations, meaning that one or more tables are missing statistics or the query does not include a hint to set the most efficient mechanism for a join or `INSERT ... SELECT` into a partitioned table. These are the tables most likely to result in suboptimal execution plans that could cause unnecessary spilling. Therefore, leaving this option enabled is a good way to find tables on which to run the `COMPUTE STATS` statement.

See [SQL Operations that Spill to Disk](#) on page 389 for information about the “spill to disk” feature for queries processing large result sets with joins, `ORDER BY`, `GROUP BY`, `DISTINCT`, aggregation functions, or analytic functions.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of `SET` statement)

### EXEC\_SINGLE\_NODE\_ROWS\_THRESHOLD Query Option

This setting controls the cutoff point (in terms of number of rows scanned) below which Impala treats a query as a “small” query, turning off optimizations such as parallel execution and native code generation. The overhead for these optimizations is applicable for queries involving substantial amounts of data, but it makes sense to skip them for queries involving tiny amounts of data. Reducing the overhead for small queries allows Impala to complete them more quickly, keeping YARN resources, admission control slots, and so on available for data-intensive queries.

**Syntax:**

```
SET EXEC_SINGLE_NODE_ROWS_THRESHOLD=number_of_rows
```

**Type:** numeric

**Default:** 100

**Internal details:**

This setting applies to query fragments where the amount of data to scan can be accurately determined, either through table and column statistics, or by the presence of a `LIMIT` clause. If Impala cannot accurately estimate the size of the input data, this setting does not apply.

For a query that is determined to be “small”, all work is performed on the coordinator node. This might result in some I/O being performed by remote reads. The savings from not distributing the query work and not generating native code are expected to outweigh any overhead from the remote reads.

**Examples:**

A common use case is to query just a few rows from a table to inspect typical data values. In this example, Impala does not parallelize the query or perform native code generation because the result set is guaranteed to be smaller than the threshold value from this query option:

```
SET EXEC_SINGLE_NODE_ROWS_THRESHOLD=500;
SELECT * FROM enormous_table LIMIT 300;
```

## EXPLAIN\_LEVEL Query Option

Controls the amount of detail provided in the output of the `EXPLAIN` statement. The basic output can help you identify high-level performance issues such as scanning a higher volume of data or more partitions than you expect. The higher levels of detail show how intermediate results flow between nodes and how different SQL operations such as `ORDER BY`, `GROUP BY`, joins, and `WHERE` clauses are implemented within a distributed query.

**Type:** STRING or INT

**Default:** 1

**Arguments:**

The allowed range of numeric values for this option is 0 to 3:

- 0 or `MINIMAL`: A barebones list, one line per operation. Primarily useful for checking the join order in very long queries where the regular `EXPLAIN` output is too long to read easily.
- 1 or `STANDARD`: The default level of detail, showing the logical way that work is split up for the distributed query.
- 2 or `EXTENDED`: Includes additional detail about how the query planner uses statistics in its decision-making process, to understand how a query could be tuned by gathering statistics, using query hints, adding or removing predicates, and so on.
- 3 or `VERBOSE`: The maximum level of detail, showing how work is split up within each node into “query fragments” that are connected in a pipeline. This extra detail is primarily useful for low-level performance testing and tuning within Impala itself, rather than for rewriting the SQL code at the user level.

**Note:** Prior to Impala 1.3, the allowed argument range for `EXPLAIN_LEVEL` was 0 to 1: level 0 had the mnemonic `NORMAL`, and level 1 was `VERBOSE`. In Impala 1.3 and higher, `NORMAL` is not a valid mnemonic value, and `VERBOSE` still applies to the highest level of detail but now corresponds to level 3. You might need to adjust the values if you have any older `impala-shell` script files that set the `EXPLAIN_LEVEL` query option.

Changing the value of this option controls the amount of detail in the output of the `EXPLAIN` statement. The extended information from level 2 or 3 is especially useful during performance tuning, when you need to confirm whether the work for the query is distributed the way you expect, particularly for the most resource-intensive operations such as join queries against large tables, queries against tables with large numbers of partitions, and insert operations for Parquet tables. The extended information also helps to check estimated resource usage when you use the admission control or resource management features explained in [Integrated Resource Management with YARN](#) on page 78. See [EXPLAIN Statement](#) on page 199 for the syntax of the `EXPLAIN` statement, and [Using the EXPLAIN Plan for Performance Tuning](#) on page 379 for details about how to use the extended information.

## Using the Impala Shell (impala-shell Command)

### Usage notes:

As always, read the EXPLAIN output from bottom to top. The lowest lines represent the initial work of the query (scanning data files), the lines in the middle represent calculations done on each node and how intermediate results are transmitted from one node to another, and the topmost lines represent the final results being sent back to the coordinator node.

The numbers in the left column are generated internally during the initial planning phase and do not represent the actual order of operations, so it is not significant if they appear out of order in the EXPLAIN output.

At all EXPLAIN levels, the plan contains a warning if any tables in the query are missing statistics. Use the COMPUTE STATS statement to gather statistics for each table and suppress this warning. See [How Impala Uses Statistics for Query Optimization](#) on page 363 for details about how the statistics help query performance.

The PROFILE command in impala-shell always starts with an explain plan showing full detail, the same as with EXPLAIN\_LEVEL=3. After the explain plan comes the executive summary, the same output as produced by the SUMMARY command in impala-shell.

### Examples:

These examples use a trivial, empty table to illustrate how the essential aspects of query planning are shown in EXPLAIN output:

```
[localhost:21000] > create table t1 (x int, s string);
[localhost:21000] > set explain_level=1;
[localhost:21000] > explain select count(*) from t1;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=10.00MB Vcores=1
WARNING: The following tables are missing relevant table and/or column statistics.
explain_plan.t1

03:AGGREGATE [MERGE FINALIZE]
|   output: sum(count(*))
|
02:EXCHANGE [PARTITION=UNPARTITIONED]
|
01:AGGREGATE
|   output: count(*)
|
00:SCAN HDFS [explain_plan.t1]
    partitions=1/1 size=0B
+-----+
[localhost:21000] > explain select * from t1;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=-9223372036854775808B Vcores=0
WARNING: The following tables are missing relevant table and/or column statistics.
explain_plan.t1
```

```

01:EXCHANGE [PARTITION=UNPARTITIONED]
|
00:SCAN HDFS [explain_plan.t1]
    partitions=1/1 size=0B
+-----+
[localhost:21000] > set explain_level=2;
[localhost:21000] > explain select * from t1;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=-9223372036854775808B Vcores=0
WARNING: The following tables are missing relevant table and/or column statistics.
explain_plan.t1

01:EXCHANGE [PARTITION=UNPARTITIONED]
|   hosts=0 per-host-mem=unavailable
|   tuple-ids=0 row-size=19B cardinality=unavailable
|
00:SCAN HDFS [explain_plan.t1, PARTITION=RANDOM]
    partitions=1/1 size=0B
    table stats: unavailable
    column stats: unavailable
    hosts=0 per-host-mem=0B
    tuple-ids=0 row-size=19B cardinality=unavailable
+-----+
[localhost:21000] > set explain_level=3;
[localhost:21000] > explain select * from t1;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=-9223372036854775808B Vcores=0
WARNING: The following tables are missing relevant table and/or column statistics.
explain_plan.t1

F01:PLAN FRAGMENT [PARTITION=UNPARTITIONED]
01:EXCHANGE [PARTITION=UNPARTITIONED]
    hosts=0 per-host-mem=unavailable
    tuple-ids=0 row-size=19B cardinality=unavailable

F00:PLAN FRAGMENT [PARTITION=RANDOM]
    DATASTREAM SINK [FRAGMENT=F01, EXCHANGE=01, PARTITION=UNPARTITIONED]

```

## Using the Impala Shell (impala-shell Command)

```
00:SCAN HDFS [explain_plan.t1, PARTITION=RANDOM]
  partitions=1/1 size=0B
  table stats: unavailable
  column stats: unavailable
  hosts=0 per-host-mem=0B
  tuple-ids=0 row-size=19B cardinality=unavailable
```

As the warning message demonstrates, most of the information needed for Impala to do efficient query planning, and for you to understand the performance characteristics of the query, requires running the `COMPUTE STATS` statement for the table:

```
[localhost:21000] > compute stats t1;
+-----+
| summary |
+-----+
| Updated 1 partition(s) and 2 column(s). |
+-----+
[localhost:21000] > explain select * from t1;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=-9223372036854775808B VCores=0

F01:PLAN FRAGMENT [PARTITION=UNPARTITIONED]
  01:EXCHANGE [PARTITION=UNPARTITIONED]
    hosts=0 per-host-mem=unavailable
    tuple-ids=0 row-size=20B cardinality=0

F00:PLAN FRAGMENT [PARTITION=RANDOM]
  DATASTREAM SINK [FRAGMENT=F01, EXCHANGE=01, PARTITION=UNPARTITIONED]
  00:SCAN HDFS [explain_plan.t1, PARTITION=RANDOM]
    partitions=1/1 size=0B
    table stats: 0 rows total
    column stats: all
    hosts=0 per-host-mem=0B
    tuple-ids=0 row-size=20B cardinality=0
```

Joins and other complicated, multi-part queries are the ones where you most commonly need to examine the `EXPLAIN` output and customize the amount of detail in the output. This example shows the default `EXPLAIN` output for a three-way join query, then the equivalent output with a `[SHUFFLE]` hint to change the join mechanism between the first two tables from a broadcast join to a shuffle join.

```
[localhost:21000] > set explain_level=1;
[localhost:21000] > explain select one.*, two.*, three.* from t1 one, t1 two, t1 three
  where one.x = two.x and two.x = three.x;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=4.00GB VCores=3

07:EXCHANGE [PARTITION=UNPARTITIONED]
|
04:HASH JOIN [INNER JOIN, BROADCAST]
|   hash predicates: two.x = three.x
```

```
|  
|  
|--06:EXCHANGE [BROADCAST]  
|  
|  
| 02:SCAN HDFS [explain_plan.t1 three]  
|    partitions=1/1 size=0B  
|  
|  
03:HASH JOIN [INNER JOIN, BROADCAST]  
|  hash predicates: one.x = two.x  
|  
|  
|--05:EXCHANGE [BROADCAST]  
|  
|  
|  01:SCAN HDFS [explain_plan.t1 two]  
|    partitions=1/1 size=0B  
|  
|  
00:SCAN HDFS [explain_plan.t1 one]  
|  partitions=1/1 size=0B  
+-----+[localhost:21000] > explain select one.*, two.*, three.* from t1 one join [shuffle] t1  
two join t1 three where one.x = two.x and two.x = three.x;  
+-----+  
| Explain String  
+-----+  
Estimated Per-Host Requirements: Memory=4.00GB Vcores=3  
  
08:EXCHANGE [PARTITION=UNPARTITIONED]  
|  
|  
04:HASH JOIN [INNER JOIN, BROADCAST]  
|  hash predicates: two.x = three.x  
|  
|  
|--07:EXCHANGE [BROADCAST]  
|  
|  
|  02:SCAN HDFS [explain_plan.t1 three]  
|    partitions=1/1 size=0B  
|  
|  
03:HASH JOIN [INNER JOIN, PARTITIONED]  
|  hash predicates: one.x = two.x  
|  
|  
|--06:EXCHANGE [PARTITION=HASH(two.x)]  
|  
|
```

## Using the Impala Shell (impala-shell Command)

```
|   01:SCAN HDFS [explain_plan.t1 two]
|     partitions=1/1 size=0B
|
| 05:EXCHANGE [PARTITION=HASH(one.x)]
|
| 00:SCAN HDFS [explain_plan.t1 one]
|   partitions=1/1 size=0B
```

For a join involving many different tables, the default EXPLAIN output might stretch over several pages, and the only details you care about might be the join order and the mechanism (broadcast or shuffle) for joining each pair of tables. In that case, you might set EXPLAIN\_LEVEL to its lowest value of 0, to focus on just the join order and join mechanism for each stage. The following example shows how the rows from the first and second joined tables are hashed and divided among the nodes of the cluster for further filtering; then the entire contents of the third table are broadcast to all nodes for the final stage of join processing.

```
[localhost:21000] > set explain_level=0;
[localhost:21000] > explain select one.*, two.*, three.* from t1 one join [shuffle] t1
  two join t1 three where one.x = two.x and two.x = three.x;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=4.00GB Vcores=3
+-----+
08:EXCHANGE [PARTITION=UNPARTITIONED]
04:HASH JOIN [INNER JOIN, BROADCAST]
|--07:EXCHANGE [BROADCAST]
|  02:SCAN HDFS [explain_plan.t1 three]
03:HASH JOIN [INNER JOIN, PARTITIONED]
|--06:EXCHANGE [PARTITION=HASH(two.x)]
|  01:SCAN HDFS [explain_plan.t1 two]
05:EXCHANGE [PARTITION=HASH(one.x)]
00:SCAN HDFS [explain_plan.t1 one]
```

## HBASE\_CACHE\_BLOCKS Query Option

Setting this option is equivalent to calling the `setCacheBlocks` method of the class [org.apache.hadoop.hbase.client.Scan](#), in an HBase Java application. Helps to control the memory pressure on the HBase region server, in conjunction with the HBASE\_CACHING query option.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of SET statement)

**Related information:**

[Using Impala to Query HBase Tables](#) on page 426, [HBASE\\_CACHING Query Option](#) on page 348

## HBASE\_CACHING Query Option

Setting this option is equivalent to calling the `setCaching` method of the class [org.apache.hadoop.hbase.client.Scan](#), in an HBase Java application. Helps to control the memory pressure on the HBase region server, in conjunction with the HBASE\_CACHE\_BLOCKS query option.

**Type:** BOOLEAN

**Default:** 0

**Related information:**

[Using Impala to Query HBase Tables](#) on page 426, [HBASE\\_CACHE\\_BLOCKS Query Option](#) on page 348

## MAX\_ERRORS Query Option

Maximum number of non-fatal errors for any particular query that are recorded in the Impala log file. For example, if a billion-row table had a non-fatal data error in every row, you could diagnose the problem without all billion errors being logged. Unspecified or 0 indicates the built-in default value of 1000.

This option only controls how many errors are reported. To specify whether Impala continues or halts when it encounters such errors, use the `ABORT_ON_ERROR` option.

**Type:** numeric

**Default:** 0 (meaning 1000 errors)

**Related information:**

[ABORT\\_ON\\_ERROR Query Option](#) on page 339, [Using Impala Logging](#) on page 443

## MAX\_IO\_BUFFERS Query Option

Deprecated query option. Currently has no effect.

**Type:** numeric

**Default:** 0

## MAX\_SCAN\_RANGE\_LENGTH Query Option

Maximum length of the scan range. Interacts with the number of HDFS blocks in the table to determine how many CPU cores across the cluster are involved with the processing for a query. (Each core processes one scan range.)

Lowering the value can sometimes increase parallelism if you have unused CPU capacity, but a too-small value can limit query performance because each scan range involves extra overhead.

Only applicable to HDFS tables. Has no effect on Parquet tables. Unspecified or 0 indicates backend default, which is the same as the HDFS block size for each table.

Although the scan range can be arbitrarily long, Impala internally uses an 8 MB read buffer so that it can query tables with huge block sizes without allocating equivalent blocks of memory.

**Type:** numeric

**Default:** 0

## MEM\_LIMIT Query Option

When resource management is not enabled, defines the maximum amount of memory a query can allocate on each node. If query processing exceeds the specified memory limit on any node, Impala cancels the query automatically. Memory limits are checked periodically during query processing, so the actual memory in use might briefly exceed the limit without the query being cancelled.

When resource management is enabled in CDH 5, the mechanism for this option changes. If set, it overrides the automatic memory estimate from Impala. Impala requests this amount of memory from YARN on each node, and the query does not proceed until that much memory is available. The actual memory used by the query could be lower, since some queries use much less memory than others. With resource management, the `MEM_LIMIT` setting acts both as a hard limit on the amount of memory a query can use on any node (enforced by YARN) and a guarantee that that much memory will be available on each node while the query is being executed. When resource management is enabled but no `MEM_LIMIT` setting is specified, Impala estimates the amount of memory needed on each node for each query, requests that much memory from YARN before starting the query, and then internally sets the `MEM_LIMIT` on each node to the requested amount of memory during

## Using the Impala Shell (impala-shell Command)

the query. Thus, if the query takes more memory than was originally estimated, Impala detects that the `MEM_LIMIT` is exceeded and cancels the query itself.

**Type:** numeric

**Default:** 0

### NUM\_NODES Query Option

Limit the number of nodes that process a query, typically during debugging.

**Type:** numeric

**Allowed values:** Only accepts the values 0 (meaning all nodes) or 1 (meaning all work is done on the coordinator node).

**Default:** 0

**Usage notes:**

If you are diagnosing a problem that you suspect is due to a timing issue due to distributed query processing, you can set `NUM_NODES=1` to verify if the problem still occurs when all the work is done on a single node.

You might set the `NUM_NODES` option to 1 briefly, during `INSERT OR CREATE TABLE AS SELECT` statements. Normally, those statements produce one or more data files per data node. If the write operation involves small amounts of data, a Parquet table, and/or a partitioned table, the default behavior could produce many small files when intuitively you might expect only a single output file. `SET NUM_NODES=1` turns off the "distributed" aspect of the write operation, making it more likely to produce only one or a few data files.

### NUM\_SCANNER\_THREADS Query Option

Maximum number of scanner threads (on each node) used for each query. By default, Impala uses as many cores as are available (one thread per core). You might lower this value if queries are using excessive resources on a busy cluster. Impala imposes a maximum value automatically, so a high value has no practical effect.

**Type:** numeric

**Default:** 0

- **Note:** Currently, a known issue ([IMPALA-488](#)) could cause excessive memory usage during a `COMPUTE STATS` operation on a Parquet table. As a workaround, issue the command `SET NUM_SCANNER_THREADS=2` in `impala-shell` before issuing the `COMPUTE STATS` statement. Then issue `UNSET NUM_SCANNER_THREADS` before continuing with queries.

### PARQUET\_COMPRESSION\_CODEC Query Option

Deprecated. Use `COMPRESSION_CODEC` in Impala 2.0 and later. See [COMPRESSION\\_CODEC Query Option](#) on page 341 for details.

### PARQUET\_FILE\_SIZE Query Option

Specifies the maximum size of each Parquet data file produced by Impala `INSERT` statements.

**Syntax:**

Specify the size in bytes, or with a trailing `m` or `g` character to indicate megabytes or gigabytes. For example:

```
-- 128 megabytes.  
set PARQUET_FILE_SIZE=134217728  
INSERT OVERWRITE parquet_table SELECT * FROM text_table;  
  
-- 512 megabytes.  
set PARQUET_FILE_SIZE=512m;
```

```
INSERT OVERWRITE parquet_table SELECT * FROM text_table;
-- 1 gigabyte.
set PARQUET_FILE_SIZE=1g;
INSERT OVERWRITE parquet_table SELECT * FROM text_table;
```

**Usage notes:**

With tables that are small or finely partitioned, the default Parquet block size (formerly 1 GB, now 256 MB in Impala 2.0 and later) could be much larger than needed for each data file. For `INSERT` operations into such tables, you can increase parallelism by specifying a smaller `PARQUET_FILE_SIZE` value, resulting in more HDFS blocks that can be processed by different nodes.

**Type:** numeric, with optional unit specifier

- **Important:**

Currently, the maximum value for this setting is 1 gigabyte (`1g`). Setting a value higher than 1 gigabyte could result in errors during an `INSERT` operation.

**Default:** 0 (produces files with a target size of 256 MB; files might be larger for very wide tables)

**Isilon considerations:**

Because the EMC Isilon storage devices use a global value for the block size rather than a configurable value for each file, the `PARQUET_FILE_SIZE` query option has no effect when Impala inserts data into a table or partition residing on Isilon storage. Use the `isi` command to set the default block size globally on the Isilon device. For example, to set the Isilon default block size to 256 MB, the recommended size for Parquet data files for Impala, issue the following command:

```
isi hdfs settings modify --default-block-size=256MB
```

**Related information:**

For information about the Parquet file format, and how the number and size of data files affects query performance, see [Using the Parquet File Format with Impala Tables](#) on page 407.

## QUERY\_TIMEOUT\_S Query Option

Sets the idle query timeout value for the session, in seconds. Queries that sit idle for longer than the timeout value are automatically cancelled. If the system administrator specified the `--idle_query_timeout` startup option, `QUERY_TIMEOUT_S` must be smaller than or equal to the `--idle_query_timeout` value.

- **Note:**

The timeout clock for queries and sessions only starts ticking when the query or session is idle. For queries, this means the query has results ready but is waiting for a client to fetch the data. A query can run for an arbitrary time without triggering a timeout, because the query is computing results rather than sitting idle waiting for the results to be fetched. The timeout period is intended to prevent unclosed queries from consuming resources and taking up slots in the admission count of running queries, potentially preventing other queries from starting.

For sessions, this means that no query has been submitted for some period of time.

**Syntax:**

```
SET QUERY_TIMEOUT_S=seconds;
```

**Type:** numeric

## Using the Impala Shell (impala-shell Command)

**Default:** 0 (no timeout if --idle\_query\_timeout not in effect; otherwise, use --idle\_query\_timeout value)

**Related information:**

[Setting Timeout Periods for Daemons, Queries, and Sessions](#) on page 81

### REQUEST\_POOL Query Option

The pool or queue name that queries should be submitted to. Only applies when you enable the Impala admission control feature (CDH 4 or CDH 5; see [Admission Control and Query Queuing](#) on page 69), or the YARN resource management feature (CDH 5 only; see [Integrated Resource Management with YARN](#) on page 78). Specifies the name of the pool used by requests from Impala to the resource manager.

Formerly known as `YARN_POOL` during the CDH 5 beta period. Renamed to reflect that it can be used both with YARN and with the lightweight admission control feature introduced in Impala 1.3.

**Type:** STRING

**Default:** empty (use the user-to-pool mapping defined by an `impalad` startup option in the Impala configuration file)

### RESERVATION\_REQUEST\_TIMEOUT Query Option (CDH 5 only)

Maximum number of milliseconds Impala will wait for a reservation to be completely granted or denied. Used in conjunction with the Impala resource management feature in Impala 1.2 and higher with CDH 5.

**Type:** numeric

**Default:** 300000 (5 minutes)

### SUPPORT\_START\_OVER Query Option

Leave this setting at its default value. It is a read-only setting, tested by some client applications such as Hue.

If you accidentally change it through `impala-shell`, subsequent queries encounter errors until you undo the change by issuing `UNSET support_start_over`.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false`

### SYNC\_DDL Query Option

When enabled, causes any DDL operation such as `CREATE TABLE` or `ALTER TABLE` to return only when the changes have been propagated to all other Impala nodes in the cluster by the Impala catalog service. That way, if you issue a subsequent `CONNECT` statement in `impala-shell` to connect to a different node in the cluster, you can be sure that other node will already recognize any added or changed tables. (The catalog service automatically broadcasts the DDL changes to all nodes automatically, but without this option there could be a period of inconsistency if you quickly switched to another node, such as by issuing a subsequent query through a load-balancing proxy.)

Although `INSERT` is classified as a DML statement, when the `SYNC_DDL` option is enabled, `INSERT` statements also delay their completion until all the underlying data and metadata changes are propagated to all Impala nodes. Internally, Impala inserts have similarities with DDL statements in traditional database systems, because they create metadata needed to track HDFS block locations for new files and they potentially add new partitions to partitioned tables.

- **Note:** Because this option can introduce a delay after each write operation, if you are running a sequence of `CREATE DATABASE`, `CREATE TABLE`, `ALTER TABLE`, `INSERT`, and similar statements within a setup script, to minimize the overall delay you can enable the `SYNC_DDL` query option only near the end, before the final DDL statement.

**Type:** Boolean; recognized values are 1 and 0, or `true` and `false`; any other value interpreted as `false`

**Default:** `false` (shown as 0 in output of `SET` statement)

**Related information:**

[DDL Statements](#) on page 162

### V\_CPU\_CORES Query Option (CDH 5 only)

The number of per-host virtual CPU cores to request from YARN. If set, the query option overrides the automatic estimate from Impala. Used in conjunction with the Impala resource management feature in Impala 1.2 and higher and CDH 5.

**Type:** numeric

**Default:** 0 (use automatic estimates)

# Tuning Impala for Performance

The following sections explain the factors affecting the performance of Impala features, and procedures for tuning, monitoring, and benchmarking Impala queries and other SQL operations.

This section also describes techniques for maximizing Impala scalability. Scalability is tied to performance: it means that performance remains high as the system workload increases. For example, reducing the disk I/O performed by a query can speed up an individual query, and at the same time improve scalability by making it practical to run more queries simultaneously. Sometimes, an optimization technique improves scalability more than performance. For example, reducing memory usage for a query might not change the query performance much, but might improve scalability by allowing more Impala queries or other kinds of jobs to run at the same time without running out of memory.

■ **Note:**

Before starting any performance tuning or benchmarking, make sure your system is configured with all the recommended minimum hardware requirements from [Hardware Requirements](#) on page 24 and software settings from [Post-Installation Configuration for Impala](#) on page 34.

- [Partitioning for Impala Tables](#) on page 394. This technique physically divides the data based on the different values in frequently queried columns, allowing queries to skip reading a large percentage of the data in a table.
- [Performance Considerations for Join Queries](#) on page 357. Joins are the main class of queries that you can tune at the SQL level, as opposed to changing physical factors such as the file format or the hardware configuration. The related topics [Overview of Column Statistics](#) on page 364 and [Overview of Table Statistics](#) on page 363 are also important primarily for join performance.
- [Overview of Table Statistics](#) on page 363 and [Overview of Column Statistics](#) on page 364. Gathering table and column statistics, using the `COMPUTE STATS` statement, helps Impala automatically optimize the performance for join queries, without requiring changes to SQL query statements. (This process is greatly simplified in Impala 1.2.2 and higher, because the `COMPUTE STATS` statement gathers both kinds of statistics in one operation, and does not require any setup and configuration as was previously necessary for the `ANALYZE TABLE` statement in Hive.)
- [Testing Impala Performance](#) on page 377. Do some post-setup testing to ensure Impala is using optimal settings for performance, before conducting any benchmark tests.
- [Benchmarking Impala Queries](#) on page 371. The configuration and sample data that you use for initial experiments with Impala is often not appropriate for doing performance tests.
- [Controlling Impala Resource Usage](#) on page 372. The more memory Impala can utilize, the better query performance you can expect. In a cluster running other kinds of workloads as well, you must make tradeoffs to make sure all Hadoop components have enough memory to perform well, so you might cap the memory that Impala can use.
- [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436. Queries against data stored in the Amazon Simple Storage Service (S3) have different performance characteristics than when the data is stored in HDFS.

A good source of tips related to scalability and performance tuning is the [Impala Cookbook](#) presentation. These slides are updated periodically as new features come out and new benchmarks are performed.

## Impala Performance Guidelines and Best Practices

Here are performance guidelines and best practices that you can use during planning, experimentation, and performance tuning for an Impala-enabled CDH cluster. All of this information is also available in more detail elsewhere in the Impala documentation; it is gathered together here to serve as a cookbook and emphasize which performance techniques typically provide the highest return on investment.

### Choose the appropriate file format for the data.

Typically, for large volumes of data (multiple gigabytes per table or partition), the Parquet file format performs best because of its combination of columnar storage layout, large I/O request size, and compression and encoding. See [How Impala Works with Hadoop File Formats](#) on page 399 for comparisons of all file formats supported by Impala, and [Using the Parquet File Format with Impala Tables](#) on page 407 for details about the Parquet file format.

- **Note:** For smaller volumes of data, a few gigabytes or less for each table or partition, you might not see significant performance differences between file formats. At small data volumes, reduced I/O from an efficient compressed file format can be counterbalanced by reduced opportunity for parallel execution. When planning for a production deployment or conducting benchmarks, always use realistic data volumes to get a true picture of performance and scalability.

### Avoid data ingestion processes that produce many small files.

When producing data files outside of Impala, prefer either text format or Avro, where you can build up the files row by row. Once the data is in Impala, you can convert it to the more efficient Parquet format and split into multiple data files using a single `INSERT ... SELECT` statement. Or, if you have the infrastructure to produce multi-megabyte Parquet files as part of your data preparation process, do that and skip the conversion step inside Impala.

Always use `INSERT ... SELECT` to copy significant volumes of data from table to table within Impala. Avoid `INSERT ... VALUES` for any substantial volume of data or performance-critical tables, because each such statement produces a separate tiny data file. See [INSERT Statement](#) on page 202 for examples of the `INSERT ... SELECT` syntax.

For example, if you have thousands of partitions in a Parquet table, each with less than 256 MB of data, consider partitioning in a less granular way, such as by year / month rather than year / month / day. If an inefficient data ingestion process produces thousands of data files in the same table or partition, consider compacting the data by performing an `INSERT ... SELECT` to copy all the data to a different table; the data will be reorganized into a smaller number of larger files by this process.

### Choose partitioning granularity based on actual data volume.

Partitioning is a technique that physically divides the data based on values of one or more columns, such as by year, month, day, region, city, section of a web site, and so on. When you issue queries that request a specific value or range of values for the partition key columns, Impala can avoid reading the irrelevant data, potentially yielding a huge savings in disk I/O.

When deciding which column(s) to use for partitioning, choose the right level of granularity. For example, should you partition by year, month, and day, or only by year and month? Choose a partitioning strategy that puts at least 256 MB of data in each partition, to take advantage of HDFS bulk I/O and Impala distributed queries.

Over-partitioning can also cause query planning to take longer than necessary, as Impala prunes the unnecessary partitions. Ideally, keep the number of partitions in the table under 30 thousand.

When preparing data files to go in a partition directory, create several large files rather than many small ones. If you receive data in the form of many small files and have no control over the input format, consider using the `INSERT ... SELECT` syntax to copy data from one table or partition to another, which compacts the files into a relatively small number (based on the number of nodes in the cluster).

If you need to reduce the overall number of partitions and increase the amount of data in each partition, first look for partition key columns that are rarely referenced or are referenced in non-critical queries (not subject to an SLA). For example, your web site log data might be partitioned by year, month, day, and hour, but if most queries roll up the results by day, perhaps you only need to partition by year, month, and day.

If you need to reduce the granularity even more, consider creating "buckets", computed values corresponding to different sets of partition key values. For example, you can use the `TRUNC( )` function with a `TIMESTAMP`

## Tuning Impala for Performance

column to group date and time values based on intervals such as week or quarter. See [Impala Date and Time Functions](#) on page 268 for details.

See [Partitioning for Impala Tables](#) on page 394 for full details and performance considerations for partitioning.

### Use smallest appropriate integer types for partition key columns.

Although it is tempting to use strings for partition key columns, since those values are turned into HDFS directory names anyway, you can minimize memory usage by using numeric values for common partition key fields such as YEAR, MONTH, and DAY. Use the smallest integer type that holds the appropriate range of values, typically TINYINT for MONTH and DAY, and SMALLINT for YEAR. Use the EXTRACT( ) function to pull out individual date and time fields from a TIMESTAMP value, and CAST( ) the return value to the appropriate integer type.

### Choose an appropriate Parquet block size.

By default, the Impala `INSERT ... SELECT` statement creates Parquet files with a 256 MB block size. (This default was changed in Impala 2.0. Formerly, the limit was 1 GB, but Impala made conservative estimates about compression, resulting in files that were smaller than 1 GB.)

Each file is a single block, allowing the whole file to be processed by a single node. As you copy Parquet files into HDFS or between HDFS filesystems, use `hdfs dfs -pb` to preserve the original block size.

If there is only one or a few data files in your Parquet table, or in a partition that is the only one accessed by a query, then you might experience a slowdown for a different reason: not enough data to take advantage of Impala's parallel distributed queries. Each data file is processed by a single core on one of the data nodes. In a 100-node cluster of 16-core machines, you could potentially process thousands of data files simultaneously. You want to find a sweet spot between "many tiny files" and "single giant file" that balances bulk I/O and parallel processing. You can set the `PARQUET_FILE_SIZE` query option before doing an `INSERT ... SELECT` statement to reduce the size of each generated Parquet file. (Specify the file size as an absolute number of bytes, or in Impala 2.0 and later, in units ending with `m` for megabytes or `g` for gigabytes.) Run benchmarks with different file sizes to find the right balance point for your particular data volume.

### Gather statistics for all tables used in performance-critical or high-volume join queries.

Gather the statistics with the `COMPUTE STATS` statement. See [Performance Considerations for Join Queries](#) on page 357 for details.

### Minimize the overhead of transmitting results back to the client.

Use techniques such as:

- Aggregation. If you need to know how many rows match a condition, the total values of matching values from some column, the lowest or highest matching value, and so on, call aggregate functions such as `COUNT()`, `SUM()`, and `MAX()` in the query rather than sending the result set to an application and doing those computations there. Remember that the size of an unaggregated result set could be huge, requiring substantial time to transmit across the network.
- Filtering. Use all applicable tests in the `WHERE` clause of a query to eliminate rows that are not relevant, rather than producing a big result set and filtering it using application logic.
- `LIMIT` clause. If you only need to see a few sample values from a result set, or the top or bottom values from a query using `ORDER BY`, include the `LIMIT` clause to reduce the size of the result set rather than asking for the full result set and then throwing most of the rows away.
- Avoid overhead from pretty-printing the result set and displaying it on the screen. When you retrieve the results through `impala-shell`, use `impala-shell` options such as `-B` and `--output_delimiter` to produce results without special formatting, and redirect output to a file rather than printing to the screen. Consider using `INSERT ... SELECT` to write the results directly to new files in HDFS. See [impala-shell Configuration Options](#) on page 332 for details about the `impala-shell` command-line options.

### Verify that your queries are planned in an efficient logical manner.

Examine the `EXPLAIN` plan for a query before actually running it. See [EXPLAIN Statement](#) on page 199 and [Using the EXPLAIN Plan for Performance Tuning](#) on page 379 for details.

### Verify performance characteristics of queries.

Verify that the low-level aspects of I/O, memory usage, network bandwidth, CPU utilization, and so on are within expected ranges by examining the query profile for a query after running it. See [Using the Query Profile for Performance Tuning](#) on page 380 for details.

## Performance Considerations for Join Queries

Queries involving join operations often require more tuning than queries that refer to only one table. The maximum size of the result set from a join query is the product of the number of rows in all the joined tables. When joining several tables with millions or billions of rows, any missed opportunity to filter the result set, or other inefficiency in the query, could lead to an operation that does not finish in a practical time and has to be cancelled.

The simplest technique for tuning an Impala join query is to collect statistics on each table involved in the join using the `COMPUTE STATS` statement, and then let Impala automatically optimize the query based on the size of each table, number of distinct values of each column, and so on. The `COMPUTE STATS` statement and the join optimization are new features introduced in Impala 1.2.2. For accurate statistics about each table, issue the `COMPUTE STATS` statement after loading the data into that table, and again if the amount of data changes substantially due to an `INSERT`, `LOAD DATA`, adding a partition, and so on.

If statistics are not available for all the tables in the join query, or if Impala chooses a join order that is not the most efficient, you can override the automatic join order optimization by specifying the `STRAIGHT_JOIN` keyword immediately after the `SELECT` keyword. In this case, Impala uses the order the tables appear in the query to guide how the joins are processed.

When you use the `STRAIGHT_JOIN` technique, you must order the tables in the join query manually instead of relying on the Impala optimizer. The optimizer uses sophisticated techniques to estimate the size of the result set at each stage of the join. For manual ordering, use this heuristic approach to start with, and then experiment to fine-tune the order:

- Specify the largest table first. This table is read from disk by each Impala node and so its size is not significant in terms of memory usage during the query.
- Next, specify the smallest table. The contents of the second, third, and so on tables are all transmitted across the network. You want to minimize the size of the result set from each subsequent stage of the join query. The most likely approach involves joining a small table first, so that the result set remains small even as subsequent larger tables are processed.
- Join the next smallest table, then the next smallest, and so on.
- For example, if you had tables `BIG`, `MEDIUM`, `SMALL`, and `TINY`, the logical join order to try would be `BIG, TINY, SMALL, MEDIUM`.

The terms “largest” and “smallest” refers to the size of the intermediate result set based on the number of rows and columns from each table that are part of the result set. For example, if you join one table `sales` with another table `customers`, a query might find results from 100 different customers who made a total of 5000 purchases. In that case, you would specify `SELECT ... FROM sales JOIN customers ...`, putting `customers` on the right side because it is smaller in the context of this query.

The Impala query planner chooses between different techniques for performing join queries, depending on the absolute and relative sizes of the tables. **Broadcast joins** are the default, where the right-hand table is considered to be smaller than the left-hand table, and its contents are sent to all the other nodes involved in the query. The alternative technique is known as a **partitioned join** (not related to a partitioned table), which is more suitable for large tables of roughly equal size. With this technique, portions of each table are sent to appropriate other

## Tuning Impala for Performance

nodes where those subsets of rows can be processed in parallel. The choice of broadcast or partitioned join also depends on statistics being available for all tables in the join, gathered by the `COMPUTE STATS` statement.

To see which join strategy is used for a particular query, issue an `EXPLAIN` statement for the query. If you find that a query uses a broadcast join when you know through benchmarking that a partitioned join would be more efficient, or vice versa, add a hint to the query to specify the precise join mechanism to use. See [Hints](#) on page 239 for details.

### How Joins Are Processed when Statistics Are Unavailable

If table or column statistics are not available for some tables in a join, Impala still reorders the tables using the information that is available. Tables with statistics are placed on the left side of the join order, in descending order of cost based on overall size and cardinality. Tables without statistics are treated as zero-size, that is, they are always placed on the right side of the join order.

### Overriding Join Reordering with `STRAIGHT_JOIN`

If an Impala join query is inefficient because of outdated statistics or unexpected data distribution, you can keep Impala from reordering the joined tables by using the `STRAIGHT_JOIN` keyword immediately after the `SELECT` keyword. The `STRAIGHT_JOIN` keyword turns off the reordering of join clauses that Impala does internally, and produces a plan that relies on the join clauses being ordered optimally in the query text. In this case, rewrite the query so that the largest table is on the left, followed by the next largest, and so on until the smallest table is on the right.

In this example, the subselect from the `BIG` table produces a very small result set, but the table might still be treated as if it were the biggest and placed first in the join order. Using `STRAIGHT_JOIN` for the last join clause prevents the final table from being reordered, keeping it as the rightmost table in the join order.

```
select straight_join x from medium join small join (select * from big where c1 < 10)
as big
  where medium.id = small.id and small.id = big.id;
```

### Examples of Join Order Optimization

Here are examples showing joins between tables with 1 billion, 200 million, and 1 million rows. (In this case, the tables are unpartitioned and using Parquet format.) The smaller tables contain subsets of data from the largest one, for convenience of joining on the unique `ID` column. The smallest table only contains a subset of columns from the others.

```
[localhost:21000] > create table big stored as parquet as select * from raw_data;
+-----+
| summary |
+-----+
| Inserted 10000000000 row(s) |
+-----+
Returned 1 row(s) in 671.56s
[localhost:21000] > desc big;
+-----+-----+-----+
| name      | type      | comment |
+-----+-----+-----+
| id        | int       |
| val       | int       |
| zfill     | string    |
| name      | string    |
| assertion | boolean   |
+-----+-----+-----+
Returned 5 row(s) in 0.01s
[localhost:21000] > create table medium stored as parquet as select * from big limit
200 * floor(1e6);
+-----+
| summary |
+-----+
| Inserted 200000000 row(s) |
```

```
+-----+
Returned 1 row(s) in 138.31s
[localhost:21000] > create table small stored as parquet as select id,val,name from
big where assertion = true limit 1 * floor(1e6);
+-----+
| summary |
+-----+
| Inserted 1000000 row(s) |
+-----+
Returned 1 row(s) in 6.32s
```

For any kind of performance experimentation, use the EXPLAIN statement to see how any expensive query will be performed without actually running it, and enable verbose EXPLAIN plans containing more performance-oriented detail: The most interesting plan lines are highlighted in bold, showing that without statistics for the joined tables, Impala cannot make a good estimate of the number of rows involved at each stage of processing, and is likely to stick with the BROADCAST join mechanism that sends a complete copy of one of the tables to each node.

```
[localhost:21000] > set explain_level=verbose;
EXPLAIN_LEVEL set to verbose
[localhost:21000] > explain select count(*) from big join medium where big.id =
medium.id;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=2.10GB VCores=2

PLAN FRAGMENT 0
PARTITION: UNPARTITIONED

6:AGGREGATE (merge finalize)
  output: SUM(COUNT(*))
  cardinality: 1
  per-host memory: unavailable
  tuple ids: 2

5:EXCHANGE
  cardinality: 1
  per-host memory: unavailable
  tuple ids: 2

PLAN FRAGMENT 1
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 5
UNPARTITIONED

3:AGGREGATE
  output: COUNT(*)
  cardinality: 1
  per-host memory: 10.00MB
  tuple ids: 2

2:HASH JOIN
  join op: INNER JOIN (BROADCAST)
  hash predicates:
    big.id = medium.id
  cardinality: unavailable
  per-host memory: 2.00GB
  tuple ids: 0 1

----4:EXCHANGE
  cardinality: unavailable
  per-host memory: 0B
  tuple ids: 1

0:SCAN HDFS
  table=join_order.big #partitions=1/1 size=23.12GB
  table stats: unavailable
  column stats: unavailable
```

## Tuning Impala for Performance

```
cardinality: unavailable
per-host memory: 88.00MB
tuple ids: 0

PLAN FRAGMENT 2
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 4
UNPARTITIONED

1:SCAN HDFS
  table=join_order.medium #partitions=1/1 size=4.62GB
  table stats: unavailable
  column stats: unavailable
  cardinality: unavailable
  per-host memory: 88.00MB
  tuple ids: 1
+
+-----+
Returned 64 row(s) in 0.04s
```

Gathering statistics for all the tables is straightforward, one COMPUTE STATS statement per table:

```
[localhost:21000] > compute stats small;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 3 column(s). |
+-----+
Returned 1 row(s) in 4.26s
[localhost:21000] > compute stats medium;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 5 column(s). |
+-----+
Returned 1 row(s) in 42.11s
[localhost:21000] > compute stats big;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 5 column(s). |
+-----+
Returned 1 row(s) in 165.44s
```

With statistics in place, Impala can choose a more effective join order rather than following the left-to-right sequence of tables in the query, and can choose BROADCAST OR PARTITIONED join strategies based on the overall sizes and number of rows in the table:

```
[localhost:21000] > explain select count(*) from medium join big where big.id =
medium.id;
Query: explain select count(*) from medium join big where big.id = medium.id
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=937.23MB VCores=2

PLAN FRAGMENT 0
PARTITION: UNPARTITIONED

6:AGGREGATE (merge finalize)
  output: SUM(COUNT(*))
  cardinality: 1
  per-host memory: unavailable
  tuple ids: 2

5:EXCHANGE
  cardinality: 1
  per-host memory: unavailable
  tuple ids: 2
```

```

PLAN FRAGMENT 1
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 5
UNPARTITIONED

3:AGGREGATE
  output: COUNT(*)
  cardinality: 1
  per-host memory: 10.00MB
  tuple ids: 2

2:HASH JOIN
  join op: INNER JOIN (BROADCAST)
  hash predicates:
    big.id = medium.id
  cardinality: 1443004441
  per-host memory: 839.23MB
  tuple ids: 1 0

----4:EXCHANGE
  cardinality: 200000000
  per-host memory: 0B
  tuple ids: 0

1:SCAN HDFS
  table=join_order.big #partitions=1/1 size=23.12GB
  table stats: 1000000000 rows total
  column stats: all
  cardinality: 1000000000
  per-host memory: 88.00MB
  tuple ids: 1

PLAN FRAGMENT 2
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 4
UNPARTITIONED

0:SCAN HDFS
  table=join_order.medium #partitions=1/1 size=4.62GB
  table stats: 200000000 rows total
  column stats: all
  cardinality: 200000000
  per-host memory: 88.00MB
  tuple ids: 0
+-----+
Returned 64 row(s) in 0.04s

[localhost:21000] > explain select count(*) from small join big where big.id = small.id;
Query: explain select count(*) from small join big where big.id = small.id
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=101.15MB VCores=2

PLAN FRAGMENT 0
PARTITION: UNPARTITIONED

6:AGGREGATE (merge finalize)
  output: SUM(COUNT(*))
  cardinality: 1
  per-host memory: unavailable
  tuple ids: 2

5:EXCHANGE
  cardinality: 1
  per-host memory: unavailable
  tuple ids: 2

PLAN FRAGMENT 1

```

## Tuning Impala for Performance

```
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 5
UNPARTITIONED

3:AGGREGATE
    output: COUNT(*)
    cardinality: 1
    per-host memory: 10.00MB
    tuple ids: 2

2:HASH JOIN
    join op: INNER JOIN (BROADCAST)
    hash predicates:
        big.id = small.id
    cardinality: 1000000000
    per-host memory: 3.15MB
    tuple ids: 1 0

----4:EXCHANGE
    cardinality: 1000000
    per-host memory: 0B
    tuple ids: 0

1:SCAN HDFS
    table=join_order.big #partitions=1/1 size=23.12GB
    table stats: 1000000000 rows total
    column stats: all
    cardinality: 1000000000
    per-host memory: 88.00MB
    tuple ids: 1

PLAN FRAGMENT 2
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 4
UNPARTITIONED

0:SCAN HDFS
    table=join_order.small #partitions=1/1 size=17.93MB
    table stats: 1000000 rows total
    column stats: all
    cardinality: 1000000
    per-host memory: 32.00MB
    tuple ids: 0

+-----+
| Returned 64 row(s) in 0.03s
```

When queries like these are actually run, the execution times are relatively consistent regardless of the table order in the query text. Here are examples using both the unique ID column and the VAL column containing duplicate values:

```
[localhost:21000] > select count(*) from big join small on (big.id = small.id);
Query: select count(*) from big join small on (big.id = small.id)
+-----+
| count(*) |
+-----+
| 1000000 |
+-----+
Returned 1 row(s) in 21.68s
[localhost:21000] > select count(*) from small join big on (big.id = small.id);
Query: select count(*) from small join big on (big.id = small.id)
+-----+
| count(*) |
+-----+
| 1000000 |
+-----+
Returned 1 row(s) in 20.45s

[localhost:21000] > select count(*) from big join small on (big.val = small.val);
```

```
+-----+
| count(*) |
+-----+
| 2000948962 |
+-----+
Returned 1 row(s) in 108.85s
[localhost:21000] > select count(*) from small join big on (big.val = small.val);
+-----+
| count(*) |
+-----+
| 2000948962 |
+-----+
Returned 1 row(s) in 100.76s
```

- **Note:** When examining the performance of join queries and the effectiveness of the join order optimization, make sure the query involves enough data and cluster resources to see a difference depending on the query plan. For example, a single data file of just a few megabytes will reside in a single HDFS block and be processed on a single node. Likewise, if you use a single-node or two-node cluster, there might not be much difference in efficiency for the broadcast or partitioned join strategies.

## How Impala Uses Statistics for Query Optimization

Impala can do better optimization for complex or multi-table queries when it has access to statistics about the volume of data and how the values are distributed. Impala uses this information to help parallelize and distribute the work for a query. The following sections describe the categories of statistics Impala can work with, and how to produce them and keep them up to date.

Originally, Impala relied on the Hive mechanism for collecting statistics, through the Hive `ANALYZE TABLE` statement which initiates a MapReduce job. For better user-friendliness and reliability, Impala implements its own `COMPUTE STATS` statement in Impala 1.2.2 and higher, along with the `DROP STATS`, `SHOW TABLE STATS`, and `SHOW COLUMN STATS` statements.

### Overview of Table Statistics

The Impala query planner can make use of statistics about entire tables and partitions when that metadata is available in the metastore database. This metadata is used on its own for certain optimizations, and used in combination with column statistics for other optimizations.

To gather table statistics after loading data into a table or partition, use one of the following techniques:

- Issue the statement `COMPUTE STATS` in Impala. This statement, available in Impala 1.2.2 and higher, is the preferred method because:
  - It gathers table statistics and statistics for all partitions and columns in a single operation.
  - It does not rely on any special Hive settings, metastore configuration, or separate database to hold the statistics.
  - If you need to adjust statistics incrementally for an existing table, such as after adding a partition or inserting new data, you can use an `ALTER TABLE` statement such as:

```
alter table analysis_data set tblproperties('numRows'='new_value',
'STATS_GENERATED_VIA_STATS_TASK' = 'true');
```

to update that one value numeric property rather than re-processing the whole table. (The requirement to include the `STATS_GENERATED_VIA_STATS_TASK` property is relatively new, as a result of the issue [HIVE-8648](#) for the Hive metastore.)

- Load the data through the `INSERT OVERWRITE` statement in Hive, while the Hive setting `hive.stats.autogather` is enabled.

## Tuning Impala for Performance

- Issue an `ANALYZE TABLE` statement in Hive, for the entire table or a specific partition.

```
ANALYZE TABLE tablename [PARTITION(partcol1[=val1], partcol2[=val2], ...)] COMPUTE
STATISTICS [NOSCAN];
```

For example, to gather statistics for a non-partitioned table:

```
ANALYZE TABLE customer COMPUTE STATISTICS;
```

To gather statistics for a `store` table partitioned by state and city, and both of its partitions:

```
ANALYZE TABLE store PARTITION(s_state, s_county) COMPUTE STATISTICS;
```

To gather statistics for the `store` table and only the partitions for California:

```
ANALYZE TABLE store PARTITION(s_state='CA', s_county) COMPUTE STATISTICS;
```

To check that table statistics are available for a table, and see the details of those statistics, use the statement `SHOW TABLE STATS table_name`. See [SHOW Statement](#) on page 242 for details.

If you use the Hive-based methods of gathering statistics, see [the Hive wiki](#) for information about the required configuration on the Hive side. Cloudera recommends using the Impala `COMPUTE STATS` statement to avoid potential configuration and scalability issues with the statistics-gathering process.

## Overview of Column Statistics

The Impala query planner can make use of statistics about individual columns when that metadata is available in the metastore database. This technique is most valuable for columns compared across tables in [join queries](#), to help estimate how many rows the query will retrieve from each table. These statistics are also important for correlated subqueries using the `EXISTS()` or `IN()` operators, which are processed internally the same way as join queries.

■ **Note:**

For column statistics to be effective in Impala, you also need to have table statistics for the applicable tables, as described in [Overview of Table Statistics](#) on page 363. When you use the Impala `COMPUTE STATS` statement, both table and column statistics are automatically gathered at the same time, for all columns in the table.

Currently, the `COMPUTE STATS` statement under CDH 4 does not store any statistics for `DECIMAL` columns. When Impala runs under CDH 5, which has better support for `DECIMAL` in the metastore database, `COMPUTE STATS` does collect statistics for `DECIMAL` columns and Impala uses the statistics to optimize query performance.

■ **Note:** Prior to Impala 1.4.0, `COMPUTE STATS` counted the number of `NULL` values in each column and recorded that figure in the metastore database. Because Impala does not currently make use of the `NULL` count during query planning, Impala 1.4.0 and higher speeds up the `COMPUTE STATS` statement by skipping this `NULL` counting.

To check whether column statistics are available for a particular set of columns, use the `SHOW COLUMN STATS table_name` statement, or check the extended `EXPLAIN` output for a query against that table that refers to those columns. See [SHOW Statement](#) on page 242 and [EXPLAIN Statement](#) on page 199 for details.

## Overview of Incremental Statistics

In Impala 2.1.0 and higher, you can use the syntax `COMPUTE INCREMENTAL STATS` and `DROP INCREMENTAL STATS`. The `INCREMENTAL` clauses work with incremental statistics, a specialized feature for partitioned tables that are large or frequently updated with new partitions.

When you compute incremental statistics for a partitioned table, by default Impala only processes those partitions that do not yet have incremental statistics. By processing only newly added partitions, you can keep statistics up to date for large partitioned tables, without incurring the overhead of reprocessing the entire table each time.

You can also compute or drop statistics for a single partition by including a `PARTITION` clause in the `COMPUTE INCREMENTAL STATS` or `DROP INCREMENTAL STATS` statement.

The metadata for incremental statistics is handled differently from the original style of statistics:

- If you have an existing partitioned table for which you have already computed statistics, issuing `COMPUTE INCREMENTAL STATS` without a partition clause causes Impala to rescan the entire table. Once the incremental statistics are computed, any future `COMPUTE INCREMENTAL STATS` statements only scan any new partitions and any partitions where you performed `DROP INCREMENTAL STATS`.
- The `SHOW TABLE STATS` and `SHOW PARTITIONS` statements now include an additional column showing whether incremental statistics are available for each column. A partition could already be covered by the original type of statistics based on a prior `COMPUTE STATS` statement, as indicated by a value other than `-1` under the `#Rows` column. Impala query planning uses either kind of statistics when available.
- `COMPUTE INCREMENTAL STATS` takes more time than `COMPUTE STATS` for the same volume of data. Therefore it is most suitable for tables with large data volume where new partitions are added frequently, making it impractical to run a full `COMPUTE STATS` operation for each new partition. For unpartitioned tables, or partitioned tables that are loaded once and not updated with new partitions, use the original `COMPUTE STATS` syntax.
- `COMPUTE INCREMENTAL STATS` uses some memory in the `catalogd` process, proportional to the number of partitions and number of columns in the applicable table. The memory overhead is approximately 400 bytes for each column in each partition. This memory is reserved in the `catalogd` daemon, the `statestored` daemon, and in each instance of the `impalad` daemon.
- In cases where new files are added to an existing partition, issue a `REFRESH` statement for the table, followed by a `DROP INCREMENTAL STATS` and `COMPUTE INCREMENTAL STATS` sequence for the changed partition.
- The `DROP INCREMENTAL STATS` statement operates only on a single partition at a time. To remove statistics (whether incremental or not) from all partitions of a table, issue a `DROP STATS` statement with no `INCREMENTAL` or `PARTITION` clauses.

The following considerations apply to incremental statistics when the structure of an existing table is changed (known as *schema evolution*):

- If you use an `ALTER TABLE` statement to drop a column, the existing statistics remain valid and `COMPUTE INCREMENTAL STATS` does not rescan any partitions.
- If you use an `ALTER TABLE` statement to add a column, Impala rescans all partitions and fills in the appropriate column-level values the next time you run `COMPUTE INCREMENTAL STATS`.
- If you use an `ALTER TABLE` statement to change the data type of a column, Impala rescans all partitions and fills in the appropriate column-level values the next time you run `COMPUTE INCREMENTAL STATS`.
- If you use an `ALTER TABLE` statement to change the file format of a table, the existing statistics remain valid and a subsequent `COMPUTE INCREMENTAL STATS` does not rescan any partitions.

See [COMPUTE STATS Statement](#) on page 170 and [DROP STATS Statement](#) on page 193 for syntax details.

## Detecting Missing Statistics

You can check whether a specific table has statistics using the `SHOW TABLE STATS` statement (for any table) or the `SHOW PARTITIONS` statement (for a partitioned table). Both statements display the same information. If a table or a partition does not have any statistics, the `#Rows` field contains `-1`. Once you compute statistics for the table or partition, the `#Rows` field changes to an accurate value.

## Tuning Impala for Performance

The following example shows a table that initially does not have any statistics. The `SHOW TABLE STATS` statement displays different values for `#Rows` before and after the `COMPUTE STATS` operation.

```
[localhost:21000] > create table no_stats (x int);
[localhost:21000] > show table stats no_stats;
+-----+-----+-----+-----+-----+-----+
| #Rows | #Files | Size | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| -1   | 0     | 0B   | NOT CACHED    | TEXT    | false      |
+-----+-----+-----+-----+-----+-----+
[localhost:21000] > compute stats no_stats;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 1 column(s).
+-----+
[localhost:21000] > show table stats no_stats;
+-----+-----+-----+-----+-----+-----+
| #Rows | #Files | Size | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| 0    | 0     | 0B   | NOT CACHED    | TEXT    | false      |
+-----+-----+-----+-----+-----+-----+
```

The following example shows a similar progression with a partitioned table. Initially, `#Rows` is `-1`. After a `COMPUTE STATS` operation, `#Rows` changes to an accurate value. Any newly added partition starts with no statistics, meaning that you must collect statistics after adding a new partition.

```
[localhost:21000] > create table no_stats_partitioned (x int) partitioned by (year
smallint);
[localhost:21000] > show table stats no_stats_partitioned;
+-----+-----+-----+-----+-----+-----+
| year | #Rows | #Files | Size | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| Total | -1   | 0     | 0B   | 0B          |          |           |
+-----+-----+-----+-----+-----+-----+
[localhost:21000] > show partitions no_stats_partitioned;
+-----+-----+-----+-----+-----+-----+
| year | #Rows | #Files | Size | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| Total | -1   | 0     | 0B   | 0B          |          |           |
+-----+-----+-----+-----+-----+-----+
[localhost:21000] > alter table no_stats_partitioned add partition (year=2013);
[localhost:21000] > compute stats no_stats_partitioned;
+-----+
| summary
+-----+
| Updated 1 partition(s) and 1 column(s).
+-----+
[localhost:21000] > alter table no_stats_partitioned add partition (year=2014);
[localhost:21000] > show partitions no_stats_partitioned;
+-----+-----+-----+-----+-----+-----+
| year | #Rows | #Files | Size | Bytes Cached | Format | Incremental stats |
+-----+-----+-----+-----+-----+-----+
| 2013 | 0     | 0     | 0B   | NOT CACHED  | TEXT   | false      |
| 2014 | -1   | 0     | 0B   | NOT CACHED  | TEXT   | false      |
| Total| 0     | 0     | 0B   | 0B          |          |           |
+-----+-----+-----+-----+-----+-----+
```

- **Note:** Because the default `COMPUTE STATS` statement creates and updates statistics for all partitions in a table, if you expect to frequently add new partitions, use the `COMPUTE INCREMENTAL STATS` syntax instead, which lets you compute stats for a single specified partition, or only for those partitions that do not already have incremental stats.

If checking each individual table is impractical, due to a large number of tables or views that hide the underlying base tables, you can also check for missing statistics for a particular query. Use the `EXPLAIN` statement to preview query efficiency before actually running the query. Use the query profile output available through the `PROFILE` command in `impala-shell` or the web UI to verify query execution and timing after running the query.

Both the EXPLAIN plan and the PROFILE output display a warning if any tables or partitions involved in the query do not have statistics.

```
[localhost:21000] > create table no_stats (x int);
[localhost:21000] > explain select count(*) from no_stats;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=10.00MB VCores=1
WARNING: The following tables are missing relevant table and/or column statistics.
incremental_stats.no_stats

03:AGGREGATE [FINALIZE]
|   output: count:merge(*)
|
02:EXCHANGE [UNPARTITIONED]
|
01:AGGREGATE
|   output: count(*)
|
00:SCAN HDFS [incremental_stats.no_stats]
    partitions=1/1 files=0 size=0B
+-----+
```

Because Impala uses the ***partition pruning*** technique when possible to only evaluate certain partitions, if you have a partitioned table with statistics for some partitions and not others, whether or not the EXPLAIN statement shows the warning depends on the actual partitions used by the query. For example, you might see warnings or not for different queries against the same table:

```
-- No warning because all the partitions for the year 2012 have stats.
EXPLAIN SELECT ... FROM t1 WHERE year = 2012;

-- Missing stats warning because one or more partitions in this range
-- do not have stats.
EXPLAIN SELECT ... FROM t1 WHERE year BETWEEN 2006 AND 2009;
```

To confirm if any partitions at all in the table are missing statistics, you might explain a query that scans the entire table, such as `SELECT COUNT(*) FROM table_name`.

## Keeping Statistics Up to Date

When the contents of a table or partition change significantly, recompute the stats for the relevant table or partition. The degree of change that qualifies as "significant" varies, depending on the absolute and relative sizes of the tables. Typically, if you add more than 30% more data to a table, it is worthwhile to recompute stats, because the differences in number of rows and number of distinct values might cause Impala to choose a different join order when that table is used in join queries. This guideline is most important for the largest tables. For example, adding 30% new data to a table containing 1 TB has a greater effect on join order than adding 30% to a table containing only a few megabytes, and the larger table has a greater effect on query performance if Impala chooses a suboptimal join order as a result of outdated statistics.

## Tuning Impala for Performance

If you reload a complete new set of data for a table, but the number of rows and number of distinct values for each column is relatively unchanged from before, you do not need to recompute stats for the table.

If the statistics for a table are out of date, and the table's large size makes it impractical to recompute new stats immediately, you can use the `DROP STATS` statement to remove the obsolete statistics, making it easier to identify tables that need a new `COMPUTE STATS` operation. [Hints](#) on page 239

For a large partitioned table, consider using the incremental stats feature available in Impala 2.1.0 and higher, as explained in [Overview of Incremental Statistics](#) on page 364. If you add a new partition to a table, it is worthwhile to recompute incremental stats, because the operation only scans the data for that one new partition.

### Setting Statistics Manually through ALTER TABLE

The most crucial piece of data in all the statistics is the number of rows in the table (for an unpartitioned table) or for each partition (for a partitioned table). The `COMPUTE STATS` statement always gathers statistics about all columns, as well as overall table statistics. If it is not practical to do an entire `COMPUTE STATS` operation after adding a partition or inserting data, or if you can see that Impala would produce a more efficient plan if the number of rows was different, you can manually set the number of rows through an `ALTER TABLE` statement:

```
create table analysis_data stored as parquet as select * from raw_data;
Inserted 1000000000 rows in 181.98s
compute stats analysis_data;
insert into analysis_data select * from smaller_table_we_forgot_before;
Inserted 1000000 rows in 15.32s
-- Now there are 1001000000 rows. We can update this single data point in the stats.
alter table analysis_data set tblproperties(' numRows'='1001000000');
```

For a partitioned table, update both the per-partition number of rows and the number of rows for the whole table:

```
-- If the table originally contained 1000000 rows, and we add another partition,
-- change the numRows property for the partition and the overall table.
alter table partitioned_data partition(year=2009, month=4) set tblproperties
  ('numRows'='30000');
alter table partitioned_data set tblproperties ('numRows'='1030000');
```

In practice, the `COMPUTE STATS` statement should be fast enough that this technique is not needed. It is most useful as a workaround for in case of performance issues where you might adjust the `numRows` value higher or lower to produce the ideal join order.

### Examples of Using Table and Column Statistics with Impala

The following examples walk through a sequence of `SHOW TABLE STATS`, `SHOW COLUMN STATS`, `ALTER TABLE`, and `SELECT` and `INSERT` statements to illustrate various aspects of how Impala uses statistics to help optimize queries.

This example shows table and column statistics for the `STORE` column used in the [TPC-DS benchmarks for decision support](#) systems. It is a tiny table holding data for 12 stores. Initially, before any statistics are gathered by a `COMPUTE STATS` statement, most of the numeric fields show placeholder values of -1, indicating that the figures are unknown. The figures that are filled in are values that are easily countable or deducible at the physical level, such as the number of files, total data size of the files, and the maximum and average sizes for data types that have a constant size such as `INT`, `FLOAT`, and `TIMESTAMP`.

```
[localhost:21000] > show table stats store;
+-----+-----+-----+
| #Rows | #Files | Size   | Format |
+-----+-----+-----+
| -1   | 1     | 3.08KB | TEXT   |
+-----+-----+-----+
Returned 1 row(s) in 0.03s
[localhost:21000] > show column stats store;
+-----+-----+-----+-----+-----+
| Column          | Type      | #Distinct Values | #Nulls | Max Size | Avg Size |
+-----+-----+-----+-----+-----+
```

s_store_sk	INT	-1	-1	4	4	
s_store_id	STRING	-1	-1	-1	-1	
s_rec_start_date	TIMESTAMP	-1	-1	16	16	
s_rec_end_date	TIMESTAMP	-1	-1	16	16	
s_closed_date_sk	INT	-1	-1	4	4	
s_store_name	STRING	-1	-1	-1	-1	
s_number_employees	INT	-1	-1	4	4	
s_floor_space	INT	-1	-1	4	4	
s_hours	STRING	-1	-1	-1	-1	
s_manager	STRING	-1	-1	-1	-1	
s_market_id	INT	-1	-1	4	4	
s_geography_class	STRING	-1	-1	-1	-1	
s_market_desc	STRING	-1	-1	-1	-1	
s_market_manager	STRING	-1	-1	-1	-1	
s_division_id	INT	-1	-1	4	4	
s_division_name	STRING	-1	-1	-1	-1	
s_company_id	INT	-1	-1	4	4	
s_company_name	STRING	-1	-1	-1	-1	
s_street_number	STRING	-1	-1	-1	-1	
s_street_name	STRING	-1	-1	-1	-1	
s_street_type	STRING	-1	-1	-1	-1	
s_suite_number	STRING	-1	-1	-1	-1	
s_city	STRING	-1	-1	-1	-1	
s_county	STRING	-1	-1	-1	-1	
s_state	STRING	-1	-1	-1	-1	
s_zip	STRING	-1	-1	-1	-1	
s_country	STRING	-1	-1	-1	-1	
s_gmt_offset	FLOAT	-1	-1	4	4	
s_tax_percentage	FLOAT	-1	-1	4	4	

Returned 29 row(s) in 0.04s

With the Hive ANALYZE TABLE statement for column statistics, you had to specify each column for which to gather statistics. The Impala COMPUTE STATS statement automatically gathers statistics for all columns, because it reads through the entire table relatively quickly and can efficiently compute the values for all the columns. This example shows how after running the COMPUTE STATS statement, statistics are filled in for both the table and all its columns:

[localhost:21000] > compute stats store;
+-----+   summary   +-----+
Updated 1 partition(s) and 29 column(s).
+-----+
Returned 1 row(s) in 1.88s
[localhost:21000] > show table stats store;
+-----+-----+-----+   #Rows   #Files   Size   Format   +-----+-----+-----+   12   1   3.08KB   TEXT   +-----+-----+-----+
Returned 1 row(s) in 0.02s
[localhost:21000] > show column stats store;
+-----+-----+-----+-----+-----+-----+   Column   Type   #Distinct Values   #Nulls   Max Size   Avg Size   +-----+-----+-----+-----+-----+-----+   s_store_sk   INT   12   -1   4   4     s_store_id   STRING   6   -1   16   16     s_rec_start_date   TIMESTAMP   4   -1   16   16     s_rec_end_date   TIMESTAMP   3   -1   16   16     s_closed_date_sk   INT   3   -1   4   4     s_store_name   STRING   8   -1   5   4.25     s_number_employees   INT   9   -1   4   4

## Tuning Impala for Performance

	s_floor_space	INT	10	-1	4	4	
	s_hours	STRING	2	-1	8		
7.083300113677979	s_manager	STRING	7	-1	15	12	
	s_market_id	INT	7	-1	4	4	
	s_geography_class	STRING	1	-1	7	7	
	s_market_desc	STRING	10	-1	94	55.5	
	s_market_manager	STRING	7	-1	16	14	
	s_division_id	INT	1	-1	4	4	
	s_division_name	STRING	1	-1	7	7	
	s_company_id	INT	1	-1	4	4	
	s_company_name	STRING	1	-1	7	7	
	s_street_number	STRING	9	-1	3		
2.833300113677979	s_street_name	STRING	12	-1	11		
6.583300113677979	s_street_type	STRING	8	-1	9		
4.833300113677979	s_suite_number	STRING	11	-1	9	8.25	
	s_city	STRING	2	-1	8	6.5	
	s_county	STRING	1	-1	17	17	
	s_state	STRING	1	-1	2	2	
	s_zip	STRING	2	-1	5	5	
	s_country	STRING	1	-1	13	13	
	s_gmt_offset	FLOAT	1	-1	4	4	
	s_tax_percentage	FLOAT	5	-1	4	4	
+-----+-----+-----+-----+-----+-----+-----+							
Returned 29 row(s) in 0.04s							

The following example shows how statistics are represented for a partitioned table. In this case, we have set up a table to hold the world's most trivial census data, a single `STRING` field, partitioned by a `YEAR` column. The table statistics include a separate entry for each partition, plus final totals for the numeric fields. The column statistics include some easily deducible facts for the partitioning column, such as the number of distinct values (the number of partition subdirectories).

localhost:21000] > describe census;				
name	type	comment		
name	string			
year	smallint			
+-----+-----+-----+-----+-----+				
Returned 2 row(s) in 0.02s				
[localhost:21000] > show table stats census;				
year	#Rows	#Files	Size	Format
2000	-1	0	0B	TEXT
2004	-1	0	0B	TEXT
2008	-1	0	0B	TEXT
2010	-1	0	0B	TEXT
2011	0	1	22B	TEXT

```

+-----+
| 2012 | -1   | 1     | 22B  | TEXT
| 2013 | -1   | 1     | 231B | PARQUET
| Total| 0    | 3     | 275B |
+-----+
Returned 8 row(s) in 0.02s
[localhost:21000] > show column stats census;
+-----+
| Column | Type      | #Distinct Values | #Nulls | Max Size | Avg Size |
+-----+
| name   | STRING    | -1              | -1    | -1       | -1
| year   | SMALLINT  | 7               | -1    | 2        | 2
+-----+
Returned 2 row(s) in 0.02s

```

The following example shows how the statistics are filled in by a `COMPUTE STATS` statement in Impala.

```

[localhost:21000] > compute stats census;
+-----+
| summary
+-----+
| Updated 3 partition(s) and 1 column(s).
+-----+
Returned 1 row(s) in 2.16s
[localhost:21000] > show table stats census;
+-----+
| year | #Rows | #Files | Size | Format |
+-----+
| 2000 | -1   | 0     | 0B   | TEXT
| 2004 | -1   | 0     | 0B   | TEXT
| 2008 | -1   | 0     | 0B   | TEXT
| 2010 | -1   | 0     | 0B   | TEXT
| 2011 | 4    | 1     | 22B  | TEXT
| 2012 | 4    | 1     | 22B  | TEXT
| 2013 | 1    | 1     | 231B | PARQUET
| Total| 9    | 3     | 275B |
+-----+
Returned 8 row(s) in 0.02s
[localhost:21000] > show column stats census;
+-----+
| Column | Type      | #Distinct Values | #Nulls | Max Size | Avg Size |
+-----+
| name   | STRING    | 4               | -1    | 5        | 4.5
| year   | SMALLINT  | 7               | -1    | 2        | 2
+-----+
Returned 2 row(s) in 0.02s

```

For examples showing how some queries work differently when statistics are available, see [Examples of Join Order Optimization](#) on page 358. You can see how Impala executes a query differently in each case by observing the `EXPLAIN` output before and after collecting statistics. Measure the before and after query times, and examine the throughput numbers in before and after `SUMMARY` or `PROFILE` output, to verify how much the improved plan speeds up performance.

## Benchmarking Impala Queries

Because Impala, like other Hadoop components, is designed to handle large data volumes in a distributed environment, conduct any performance tests using realistic data and cluster configurations. Use a multi-node cluster rather than a single node; run queries against tables containing terabytes of data rather than tens of gigabytes. The parallel processing techniques used by Impala are most appropriate for workloads that are beyond the capacity of a single server.

When you run queries returning large numbers of rows, the CPU time to pretty-print the output can be substantial, giving an inaccurate measurement of the actual query time. Consider using the `-B` option on the `impala-shell` command to turn off the pretty-printing, and optionally the `-o` option to store query results in a file rather than printing to the screen. See [impala-shell Configuration Options](#) on page 332 for details.

# Controlling Impala Resource Usage

Sometimes, balancing raw query performance against scalability requires limiting the amount of resources, such as memory or CPU, used by a single query or group of queries. Impala can use several mechanisms that help to smooth out the load during heavy concurrent usage, resulting in faster overall query times and sharing of resources across Impala queries, MapReduce jobs, and other kinds of workloads across a CDH cluster:

- The Impala admission control feature uses a fast, distributed mechanism to hold back queries that exceed limits on the number of concurrent queries or the amount of memory used. The queries are queued, and executed as other queries finish and resources become available. You can control the concurrency limits, and specify different limits for different groups of users to divide cluster resources according to the priorities of different classes of users. This feature is new in Impala 1.3, and works with both CDH 4 and CDH 5. See [Admission Control and Query Queuing](#) on page 69 for details.
- You can restrict the amount of memory Impala reserves during query execution by specifying the `-mem_limit` option for the `impalad` daemon. See [Modifying Impala Startup Options](#) on page 48 for details. This limit applies only to the memory that is directly consumed by queries; Impala reserves additional memory at startup, for example to hold cached metadata.
- For production deployment, Cloudera recommends that you implement resource isolation using mechanisms such as cgroups, which you can configure using Cloudera Manager. For details, see the [Static Resource Pools](#) in the Cloudera Manager documentation.
- When you use Impala in combination with CDH 5, you can use the YARN resource management framework in combination with the Llama service, as explained in [Integrated Resource Management with YARN](#) on page 78.

## Using HDFS Caching with Impala (CDH 5.1 or higher only)

HDFS caching provides performance and scalability benefits in production environments where Impala queries and other Hadoop jobs operate on quantities of data much larger than the physical RAM on the data nodes, making it impractical to rely on the Linux OS cache, which only keeps the most recently used data in memory. Data read from the HDFS cache avoids the overhead of checksumming and memory-to-memory copying involved when using data from the Linux OS cache.

For background information about how to set up and manage HDFS caching for a CDH cluster, see [the CDH documentation](#).

### Overview of HDFS Caching for Impala

On CDH 5.1 and higher, Impala can use the HDFS caching feature to make more effective use of RAM, so that repeated queries can take advantage of data “pinned” in memory regardless of how much data is processed overall. The HDFS caching feature lets you designate a subset of frequently accessed data to be pinned permanently in memory, remaining in the cache across multiple queries and never being evicted. This technique is suitable for tables or partitions that are frequently accessed and are small enough to fit entirely within the HDFS memory cache. For example, you might designate several dimension tables to be pinned in the cache, to speed up many different join queries that reference them. Or in a partitioned table, you might pin a partition holding data from the most recent time period because that data will be queried intensively; then when the next set of data arrives, you could unpin the previous partition and pin the partition holding the new data.

### Setting Up HDFS Caching for Impala

To use HDFS caching with Impala, first set up that feature for your CDH cluster:

- Decide how much memory to devote to the HDFS cache on each host. Remember that the total memory available for cached data is the sum of the cache sizes on all the hosts. By default, any data block is only cached on one host, although you can cache a block across multiple hosts by increasing the replication factor.

- Issue `hdfs cacheadmin` commands to set up one or more cache pools, owned by the same user as the `impalad` daemon (typically `impala`). For example:

```
hdfs cacheadmin -addPool four_gig_pool -owner impala -limit 4000000000
```

For details about the `hdfs cacheadmin` command, see [the CDH documentation](#).

Once HDFS caching is enabled and one or more pools are available, see [Enabling HDFS Caching for Impala Tables and Partitions](#) on page 373 for how to choose which Impala data to load into the HDFS cache. On the Impala side, you specify the cache pool name defined by the `hdfs cacheadmin` command in the Impala DDL statements that enable HDFS caching for a table or partition, such as `CREATE TABLE ... CACHED IN pool` or `ALTER TABLE ... SET CACHED IN pool`.

## Enabling HDFS Caching for Impala Tables and Partitions

Begin by choosing which tables or partitions to cache. For example, these might be lookup tables that are accessed by many different join queries, or partitions corresponding to the most recent time period that are analyzed by different reports or ad hoc queries.

In your SQL statements, you specify logical divisions such as tables and partitions to be cached. Impala translates these requests into HDFS-level directives that apply to particular directories and files. For example, given a partitioned table `CENSUS` with a partition key column `YEAR`, you could choose to cache all or part of the data as follows:

In Impala 2.2 / CDH 5.4 and higher, the optional `WITH REPLICATION` clause for `CREATE TABLE` and `ALTER TABLE` lets you specify a ***replication factor***, the number of hosts on which to cache the same data blocks. When Impala processes a cached data block, where the cache replication factor is greater than 1, Impala randomly selects a host that has a cached copy of that data block. This optimization avoids excessive CPU usage on a single host when the same cached data block is processed multiple times.

```
-- Cache the entire table (all partitions).
alter table census set cached in 'pool_name';

-- Remove the entire table from the cache.
alter table census set uncached;

-- Cache a portion of the table (a single partition).
-- If the table is partitioned by multiple columns (such as year, month, day),
-- the ALTER TABLE command must specify values for all those columns.
alter table census partition (year=1960) set cached in 'pool_name';

-- Cache the data from one partition on up to 4 hosts, to minimize CPU load on any
single host
-- when the same data block is processed multiple times.
alter table census partition (year=1970) set cached in 'pool_name' with replication =
4;

-- At each stage, check the volume of cached data.
-- For large tables or partitions, the background loading might take some time,
-- so you might have to wait and reissue the statement until all the data
-- has finished being loaded into the cache.
show table stats census;
+-----+-----+-----+-----+-----+
| year | #Rows | #Files | Size | Bytes Cached | Format |
+-----+-----+-----+-----+-----+
| 1900 | -1 | 1 | 11B | NOT CACHED | TEXT |
| 1940 | -1 | 1 | 11B | NOT CACHED | TEXT |
| 1960 | -1 | 1 | 11B | 11B | TEXT |
| 1970 | -1 | 1 | 11B | NOT CACHED | TEXT |
| Total | -1 | 4 | 44B | 11B | 
```

### CREATE TABLE considerations:

The HDFS caching feature affects the Impala `CREATE TABLE` statement as follows:

## Tuning Impala for Performance

- You can put a `CACHED IN 'pool_name'` clause and optionally a `WITH REPLICATION = number_of_hosts` clause at the end of a `CREATE TABLE` statement to automatically cache the entire contents of the table, including any partitions added later. The `pool_name` is a pool that you previously set up with the `hdfs cacheadmin` command.
- Once a table is designated for HDFS caching through the `CREATE TABLE` statement, if new partitions are added later through `ALTER TABLE ... ADD PARTITION` statements, the data in those new partitions is automatically cached in the same pool.
- If you want to perform repetitive queries on a subset of data from a large table, and it is not practical to designate the entire table or specific partitions for HDFS caching, you can create a new cached table with just a subset of the data by using `CREATE TABLE ... CACHED IN 'pool_name' AS SELECT ... WHERE ...`. When you are finished with generating reports from this subset of data, drop the table and both the data files and the data cached in RAM are automatically deleted.

See [CREATE TABLE Statement](#) on page 180 for the full syntax.

### Other memory considerations:

Certain DDL operations, such as `ALTER TABLE ... SET LOCATION`, are blocked while the underlying HDFS directories contain cached files. You must uncache the files first, before changing the location, dropping the table, and so on.

When data is requested to be pinned in memory, that process happens in the background without blocking access to the data while the caching is in progress. Loading the data from disk could take some time. Impala reads each HDFS data block from memory if it has been pinned already, or from disk if it has not been pinned yet. When files are added to a table or partition whose contents are cached, Impala automatically detects those changes and performs a `REFRESH` automatically once the relevant data is cached.

The amount of data that you can pin on each node through the HDFS caching mechanism is subject to a quota that is enforced by the underlying HDFS service. Before requesting to pin an Impala table or partition in memory, check that its size does not exceed this quota.

- **Note:** Because the HDFS cache consists of combined memory from all the data nodes in the cluster, cached tables or partitions can be bigger than the amount of HDFS cache memory on any single host.

## Loading and Removing Data with HDFS Caching Enabled

When HDFS caching is enabled, extra processing happens in the background when you add or remove data through statements such as `INSERT` and `DROP TABLE`.

### Inserting or loading data:

- When Impala performs an `INSERT` or `LOAD DATA` statement for a table or partition that is cached, the new data files are automatically cached and Impala recognizes that fact automatically.
- If you perform an `INSERT` or `LOAD DATA` through Hive, as always, Impala only recognizes the new data files after a `REFRESH table_name` statement in Impala.
- If the cache pool is entirely full, or becomes full before all the requested data can be cached, the Impala DDL statement returns an error. This is to avoid situations where only some of the requested data could be cached.
- When HDFS caching is enabled for a table or partition, new data files are cached automatically when they are added to the appropriate directory in HDFS, without the need for a `REFRESH` statement in Impala. Impala automatically performs a `REFRESH` once the new data is loaded into the HDFS cache.

### Dropping tables, partitions, or cache pools:

The HDFS caching feature interacts with the Impala `DROP TABLE` and `ALTER TABLE ... DROP PARTITION` statements as follows:

- When you issue a `DROP TABLE` for a table that is entirely cached, or has some partitions cached, the `DROP TABLE` succeeds and all the cache directives Impala submitted for that table are removed from the HDFS cache system.
- The same applies to `ALTER TABLE ... DROP PARTITION`. The operation succeeds and any cache directives are removed.
- As always, the underlying data files are removed if the dropped table is an internal table, or the dropped partition is in its default location underneath an internal table. The data files are left alone if the dropped table is an external table, or if the dropped partition is in a non-default location.
- If you designated the data files as cached through the `hdfs cacheadmin` command, and the data files are left behind as described in the previous item, the data files remain cached. Impala only removes the cache directives submitted by Impala through the `CREATE TABLE` or `ALTER TABLE` statements. It is OK to have multiple redundant cache directives pertaining to the same files; the directives all have unique IDs and owners so that the system can tell them apart.
- If you drop an HDFS cache pool through the `hdfs cacheadmin` command, all the Impala data files are preserved, just no longer cached. After a subsequent `REFRESH`, `SHOW TABLE STATS` reports 0 bytes cached for each associated Impala table or partition.

#### Relocating a table or partition:

The HDFS caching feature interacts with the Impala `ALTER TABLE ... SET LOCATION` statement as follows:

- If you have designated a table or partition as cached through the `CREATE TABLE` or `ALTER TABLE` statements, subsequent attempts to relocate the table or partition through an `ALTER TABLE ... SET LOCATION` statement will fail. You must issue an `ALTER TABLE ... SET UNCACHED` statement for the table or partition first. Otherwise, Impala would lose track of some cached data files and have no way to uncache them later.

## Administration for HDFS Caching with Impala

Here are the guidelines and steps to check or change the status of HDFS caching for Impala data:

#### `hdfs cacheadmin` command:

- If you drop a cache pool with the `hdfs cacheadmin` command, Impala queries against the associated data files will still work, by falling back to reading the files from disk. After performing a `REFRESH` on the table, Impala reports the number of bytes cached as 0 for all associated tables and partitions.
- You might use `hdfs cacheadmin` to get a list of existing cache pools, or detailed information about the pools, as follows:

```
hdfs cacheadmin -listDirectives          # Basic info
Found 122 entries
  ID POOL      REPL EXPIRY PATH
  123 testPool    1 never  /user/hive/warehouse/tpcds.store_sales
  124 testPool    1 never  /user/hive/warehouse/tpcds.store_sales/ss_date=1998-01-15
  125 testPool    1 never  /user/hive/warehouse/tpcds.store_sales/ss_date=1998-02-01
  ...
hdfs cacheadmin -listDirectives -stats  # More details
Found 122 entries
  ID POOL      REPL EXPIRY PATH
  BYTES_NEEDED  BYTES_CACHED FILES_NEEDED FILES_CACHED
  123 testPool    1 never  /user/hive/warehouse/tpcds.store_sales
                  0         0         0         0
  124 testPool    1 never  /user/hive/warehouse/tpcds.store_sales/ss_date=1998-01-15      143169
  143169          1           1
  125 testPool    1 never  /user/hive/warehouse/tpcds.store_sales/ss_date=1998-02-01      112447
  112447          1           1
  ...

```

#### Impala SHOW statement:

## Tuning Impala for Performance

- For each table or partition, the `SHOW TABLE STATS` or `SHOW PARTITIONS` statement displays the number of bytes currently cached by the HDFS caching feature. If there are no cache directives in place for that table or partition, the result set displays `NOT CACHED`. A value of 0, or a smaller number than the overall size of the table or partition, indicates that the cache request has been submitted but the data has not been entirely loaded into memory yet. See [SHOW Statement](#) on page 242 for details.

### Cloudera Manager:

- You can enable or disable HDFS caching through Cloudera Manager, using the configuration setting **Maximum Memory Used for Caching** for the HDFS service. This control sets the HDFS configuration parameter `dfs_datanode_max_locked_memory`, which specifies the upper limit of HDFS cache size on each node.
- All the other manipulation of the HDFS caching settings, such as what files are cached, is done through the command line, either Impala DDL statements or the Linux `hdfs cacheadmin` command.

### Impala memory limits:

The Impala HDFS caching feature interacts with the Impala memory limits as follows:

- The maximum size of each HDFS cache pool is specified externally to Impala, through the `hdfs cacheadmin` command.
- All the memory used for HDFS caching is separate from the `impalad` daemon address space and does not count towards the limits of the `--mem_limit` startup option, `MEM_LIMIT` query option, or further limits imposed through YARN resource management or the Linux `cgroups` mechanism.
- Because accessing HDFS cached data avoids a memory-to-memory copy operation, queries involving cached data require less memory on the Impala side than the equivalent queries on uncached data. In addition to any performance benefits in a single-user environment, the reduced memory helps to improve scalability under high-concurrency workloads.

## Performance Considerations for HDFS Caching with Impala

In Impala 1.4.0 and higher, Impala supports efficient reads from data that is pinned in memory through HDFS caching. Impala takes advantage of the HDFS API and reads the data from memory rather than from disk whether the data files are pinned using Impala DDL statements, or using the command-line mechanism where you specify HDFS paths.

When you examine the output of the `impala-shell SUMMARY` command, or look in the metrics report for the `impalad` daemon, you see how many bytes are read from the HDFS cache. For example, this excerpt from a query profile illustrates that all the data read during a particular phase of the query came from the HDFS cache, because the `BytesRead` and `BytesReadDataNodeCache` values are identical.

```
HDFS_SCAN_NODE (id=0):(Total: 11s114ms, non-child: 11s114ms, % non-child: 100.00%)
  - AverageHdfsReadThreadConcurrency: 0.00
  - AverageScannerThreadConcurrency: 32.75
  - BytesRead: 10.47 GB (11240756479)
  - BytesReadDataNodeCache: 10.47 GB (11240756479)
  - BytesReadLocal: 10.47 GB (11240756479)
  - BytesReadShortCircuit: 10.47 GB (11240756479)
  - DecompressionTime: 27s572ms
```

For queries involving smaller amounts of data, or in single-user workloads, you might not notice a significant difference in query response time with or without HDFS caching. Even with HDFS caching turned off, the data for the query might still be in the Linux OS buffer cache. The benefits become clearer as data volume increases, and especially as the system processes more concurrent queries. HDFS caching improves the scalability of the overall system. That is, it prevents query performance from declining when the workload outstrips the capacity of the Linux OS cache.

### SELECT considerations:

The Impala HDFS caching feature interacts with the `SELECT` statement and query performance as follows:

- Impala automatically reads from memory any data that has been designated as cached and actually loaded into the HDFS cache. (It could take some time after the initial request to fully populate the cache for a table

with large size or many partitions.) The speedup comes from two aspects: reading from RAM instead of disk, and accessing the data straight from the cache area instead of copying from one RAM area to another. This second aspect yields further performance improvement over the standard OS caching mechanism, which still results in memory-to-memory copying of cached data.

- For small amounts of data, the query speedup might not be noticeable in terms of wall clock time. The performance might be roughly the same with HDFS caching turned on or off, due to recently used data being held in the Linux OS cache. The difference is more pronounced with:

- Data volumes (for all queries running concurrently) that exceed the size of the Linux OS cache.
- A busy cluster running many concurrent queries, where the reduction in memory-to-memory copying and overall memory usage during queries results in greater scalability and throughput.
- Thus, to really exercise and benchmark this feature in a development environment, you might need to simulate realistic workloads and concurrent queries that match your production environment.
- One way to simulate a heavy workload on a lightly loaded system is to flush the OS buffer cache (on each data node) between iterations of queries against the same tables or partitions:

```
$ sync
$ echo 1 > /proc/sys/vm/drop_caches
```

- Impala queries take advantage of HDFS cached data regardless of whether the cache directive was issued by Impala or externally through the `hdfs cacheadmin` command, for example for an external table where the cached data files might be accessed by several different Hadoop components.
- If your query returns a large result set, the time reported for the query could be dominated by the time needed to print the results on the screen. To measure the time for the underlying query processing, query the `COUNT()` of the big result set, which does all the same processing but only prints a single line to the screen.

## Testing Impala Performance

Test to ensure that Impala is configured for optimal performance. If you have installed Impala without Cloudera Manager, complete the processes described in this topic to help ensure a proper configuration. Even if you installed Impala with Cloudera Manager, which automatically applies appropriate configurations, these procedures can be used to verify that Impala is set up correctly.

### Checking Impala Configuration Values

You can inspect Impala configuration values by connecting to your Impala server using a browser.

#### To check Impala configuration values:

1. Use a browser to connect to one of the hosts running `impalad` in your environment. Connect using an address of the form `http://hostname:port/varz`.

**Note:** In the preceding example, replace `hostname` and `port` with the name and port of your Impala server. The default port is 25000.

2. Review the configured values.

For example, to check that your system is configured to use block locality tracking information, you would check that the value for `dfs.datanode.hdfs-blocks-metadata.enabled` is `true`.

#### To check data locality:

1. Execute a query on a dataset that is available across multiple nodes. For example, for a table named `MyTable` that has a reasonable chance of being spread across multiple DataNodes:

```
[impalad-host:21000] > SELECT COUNT (*) FROM MyTable
```

## Tuning Impala for Performance

2. After the query completes, review the contents of the Impala logs. You should find a recent message similar to the following:

```
Total remote scan volume = 0
```

The presence of remote scans may indicate `impalad` is not running on the correct nodes. This can be because some DataNodes do not have `impalad` running or it can be because the `impalad` instance that is starting the query is unable to contact one or more of the `impalad` instances.

### To understand the causes of this issue:

1. Connect to the debugging web server. By default, this server runs on port 25000. This page lists all `impalad` instances running in your cluster. If there are fewer instances than you expect, this often indicates some DataNodes are not running `impalad`. Ensure `impalad` is started on all DataNodes.
2. If you are using multi-homed hosts, ensure that the Impala daemon's hostname resolves to the interface on which `impalad` is running. The hostname Impala is using is displayed when `impalad` starts. If you need to explicitly set the hostname, use the `--hostname` flag.
3. Check that `statestored` is running as expected. Review the contents of the state store log to ensure all instances of `impalad` are listed as having connected to the state store.

### Reviewing Impala Logs

You can review the contents of the Impala logs for signs that short-circuit reads or block location tracking are not functioning. Before checking logs, execute a simple query against a small HDFS dataset. Completing a query task generates log messages using current settings. Information on starting Impala and executing queries can be found in [Starting Impala](#) on page 47 and [Using the Impala Shell \(impala-shell Command\)](#) on page 332. Information on logging can be found in [Using Impala Logging](#) on page 443. Log messages and their interpretations are as follows:

Log Message	Interpretation
Unknown disk id. This will negatively affect performance. Check your hdfs settings to enable block location metadata	Tracking block locality is not enabled.
Unable to load native-hadoop library for your platform... using builtin-java classes where applicable	Native checksumming is not enabled.

## Understanding Impala Query Performance - EXPLAIN Plans and Query Profiles

To understand the high-level performance considerations for Impala queries, read the output of the `EXPLAIN` statement for the query. You can get the `EXPLAIN` plan without actually running the query itself.

For an overview of the physical performance characteristics for a query, issue the `SUMMARY` statement in `impala-shell` immediately after executing a query. This condensed information shows which phases of execution took the most time, and how the estimates for memory usage and number of rows at each phase compare to the actual values.

To understand the detailed performance characteristics for a query, issue the `PROFILE` statement in `impala-shell` immediately after executing a query. This low-level information includes physical details about memory, CPU, I/O, and network usage, and thus is only available after the query is actually run.

Also, see [Performance Considerations for the Impala-HBase Integration](#) on page 427 and [Understanding and Tuning Impala Query Performance for S3 Data](#) on page 440 for examples of interpreting `EXPLAIN` plans for queries against HBase tables and data stored in the Amazon Simple Storage System (S3).

## Using the EXPLAIN Plan for Performance Tuning

The [EXPLAIN](#) statement gives you an outline of the logical steps that a query will perform, such as how the work will be distributed among the nodes and how intermediate results will be combined to produce the final result set. You can see these details before actually running the query. You can use this information to check that the query will not operate in some very unexpected or inefficient way.

```
[impalad-host:21000] > explain select count(*) from customer_address;
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=42.00MB Vcores=1
03:AGGREGATE [MERGE FINALIZE]
|   output: sum(count(*))
02:EXCHANGE [PARTITION=UNPARTITIONED]
01:AGGREGATE
|   output: count(*)
00:SCAN HDFS [default.customer_address]
|   partitions=1/1 size=5.25MB
+-----+
```

Read the EXPLAIN plan from bottom to top:

- The last part of the plan shows the low-level details such as the expected amount of data that will be read, where you can judge the effectiveness of your partitioning strategy and estimate how long it will take to scan a table based on total data size and the size of the cluster.
- As you work your way up, next you see the operations that will be parallelized and performed on each Impala node.
- At the higher levels, you see how data flows when intermediate result sets are combined and transmitted from one node to another.
- See [EXPLAIN\\_LEVEL Query Option](#) on page 343 for details about the EXPLAIN\_LEVEL query option, which lets you customize how much detail to show in the EXPLAIN plan depending on whether you are doing high-level or low-level tuning, dealing with logical or physical aspects of the query.

The EXPLAIN plan is also printed at the beginning of the query profile report described in [Using the Query Profile for Performance Tuning](#) on page 380, for convenience in examining both the logical and physical aspects of the query side-by-side.

The amount of detail displayed in the EXPLAIN output is controlled by the [EXPLAIN\\_LEVEL](#) query option. You typically increase this setting from `normal` to `verbose` (or from 0 to 1) when doublechecking the presence of table and column statistics during performance tuning, or when estimating query resource usage in conjunction with the resource management features in CDH 5.

## Using the SUMMARY Report for Performance Tuning

The [SUMMARY](#) command within the `impala-shell` interpreter gives you an easy-to-digest overview of the timings for the different phases of execution for a query. Like the EXPLAIN plan, it is easy to see potential performance bottlenecks. Like the PROFILE output, it is available after the query is run and so displays actual timing numbers.

The SUMMARY report is also printed at the beginning of the query profile report described in [Using the Query Profile for Performance Tuning](#) on page 380, for convenience in examining high-level and low-level aspects of the query side-by-side.

For example, here is a query involving an aggregate function, on a single-node VM. The different stages of the query and their timings are shown (rolled up for all nodes), along with estimated and actual values used in planning the query. In this case, the `AVG()` function is computed for a subset of data on each node (stage 01) and then the aggregated results from all nodes are combined at the end (stage 03). You can see which stages took the most time, and whether any estimates were substantially different than the actual data distribution.

## Tuning Impala for Performance

(When examining the time values, be sure to consider the suffixes such as `us` for microseconds and `ms` for milliseconds, rather than just looking for the largest numbers.)

```
[localhost:21000] > select avg(ss_sales_price) from store_sales where ss_coupon_amt = 0;
+-----+
| avg(ss_sales_price) |
+-----+
| 37.80770926328327 |
+-----+
[localhost:21000] > summary;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Operator | #Hosts | Avg Time | Max Time | #Rows | Est. #Rows | Peak Mem | Est. |
| Peak Mem | Detail | | | | | | |
+-----+-----+-----+-----+-----+-----+-----+-----+
| 03:AGGREGATE | 1 | 1.03ms | 1.03ms | 1 | 1 | 48.00 KB | -1 B |
| | MERGE FINALIZE | | | | | | |
| 02:EXCHANGE | 1 | 0ns | 0ns | 1 | 1 | 0 B | -1 B |
| | UNPARTITIONED | | | | | | |
| 01:AGGREGATE | 1 | 30.79ms | 30.79ms | 1 | 1 | 80.00 KB | 10.00 |
| MB | | | | | | |
| 00:SCAN HDFS | 1 | 5.45s | 5.45s | 2.21M | -1 | 64.05 MB | 432.00 |
| MB | tpc.store_sales | | | | | | |
+-----+-----+-----+-----+-----+-----+-----+-----+
```

Notice how the longest initial phase of the query is measured in seconds (s), while later phases working on smaller intermediate results are measured in milliseconds (ms) or even nanoseconds (ns).

Here is an example from a more complicated query, as it would appear in the `PROFILE` output:

Operator Peak Mem Detail	#Hosts	Avg Time	Max Time	#Rows	Est.	#Rows	Peak Mem	Est.
09:MERGING-EXCHANGE -1.00 B UNPARTITIONED	1	79.738us	79.738us	5		5	0	
05:TOP-N 120.00 B	3	84.693us	88.810us	5		5	12.00 KB	
04:AGGREGATE 10.00 MB MERGE FINALIZE	3	5.263ms	6.432ms	5		5	44.00 KB	
08:AGGREGATE 15.11 MB MERGE	3	16.659ms	27.444ms	52.52K	600.12K	3.20 MB		
07:EXCHANGE 0 HASH(o_orderpriority)	3	2.644ms	5.1ms	52.52K	600.12K	0		
03:AGGREGATE 15.11 MB	3	342.913ms	966.291ms	52.52K	600.12K	10.80 MB		
02:HASH JOIN 941.01 KB INNER JOIN, BROADCAST	3	2s165ms	2s171ms	144.87K	600.12K	13.63 MB		
--06:EXCHANGE 0 BROADCAST	3	8.296ms	8.692ms	57.22K	15.00K	0		
01:SCAN HDFS 176.00 MB tpch.orders o	2	1s412ms	1s978ms	57.22K	15.00K	24.21 MB		
00:SCAN HDFS 264.00 MB tpch.lineitem l	3	8s032ms	8s558ms	3.79M	600.12K	32.29 MB		

## Using the Query Profile for Performance Tuning

The `PROFILE` statement, available in the `impala-shell` interpreter, produces a detailed low-level report showing how the most recent query was executed. Unlike the `EXPLAIN` plan described in [Using the EXPLAIN Plan for Performance Tuning](#) on page 379, this information is only available after the query has finished. It shows physical details such as the number of bytes read, maximum memory usage, and so on for each node. You can use this information to determine if the query is I/O-bound or CPU-bound, whether some network condition is imposing a bottleneck, whether a slowdown is affecting some nodes but not others, and to check that recommended configuration settings such as short-circuit local reads are in effect.

The `EXPLAIN` plan is also printed at the beginning of the query profile report, for convenience in examining both the logical and physical aspects of the query side-by-side. The `EXPLAIN_LEVEL` query option also controls the verbosity of the `EXPLAIN` output printed by the `PROFILE` command.

Here is an example of a query profile, from a relatively straightforward query on a single-node pseudo-distributed cluster to keep the output relatively brief.

```
[localhost:21000] > profile;
Query Runtime Profile:
Query (id=6540a03d4bee0691:4963d6269b210ebd):
  Summary:
    Session ID: ea4a197f1c7bf858:c74e66f72e3a33ba
    Session Type: BEESWAX
    Start Time: 2013-12-02 17:10:30.263067000
    End Time: 2013-12-02 17:10:50.932044000
    Query Type: QUERY
    Query State: FINISHED
    Query Status: OK
    Impala Version: impalad version 1.2.1 RELEASE (build
edb5af1bcad63d410bc5d47cc203df3a880e9324)
    User: cloudera
    Network Address: 127.0.0.1:49161
    Default Db: stats_testing
    Sql Statement: select t1.s, t2.s from t1 join t2 on (t1.id = t2.parent)
    Plan:
-----
Estimated Per-Host Requirements: Memory=2.09GB Vcores=2

PLAN FRAGMENT 0
PARTITION: UNPARTITIONED

4:EXCHANGE
  cardinality: unavailable
  per-host memory: unavailable
  tuple ids: 0 1

PLAN FRAGMENT 1
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 4
UNPARTITIONED

2:HASH JOIN
  join op: INNER JOIN (BROADCAST)
  hash predicates:
    t1.id = t2.parent
  cardinality: unavailable
  per-host memory: 2.00GB
  tuple ids: 0 1

----3:EXCHANGE
  cardinality: unavailable
  per-host memory: 0B
  tuple ids: 1

0:SCAN HDFS
  table=stats_testing.t1 #partitions=1/1 size=33B
  table stats: unavailable
  column stats: unavailable
  cardinality: unavailable
  per-host memory: 32.00MB
  tuple ids: 0

PLAN FRAGMENT 2
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 3
UNPARTITIONED

1:SCAN HDFS
  table=stats_testing.t2 #partitions=1/1 size=960.00KB
  table stats: unavailable
  column stats: unavailable
  cardinality: unavailable
  per-host memory: 96.00MB
```

## Tuning Impala for Performance

```
tuple ids: 1
-----
Query Timeline: 20s670ms
  - Start execution: 2.559ms (2.559ms)
  - Planning finished: 23.587ms (21.27ms)
  - Rows available: 666.199ms (642.612ms)
  - First row fetched: 668.919ms (2.719ms)
  - Unregister query: 20s668ms (20s000ms)
ImpalaServer:
  - ClientFetchWaitTimer: 19s637ms
  - RowMaterializationTimer: 167.121ms
Execution Profile 6540a03d4bee0691:4963d6269b210ebd:(Active: 837.815ms, % non-child: 0.00%)
  Per Node Peak Memory Usage: impala-1.example.com:22000(7.42 MB)
    - FinalizationTimer: 0ns
  Coordinator Fragment:(Active: 195.198ms, % non-child: 0.00%)
    MemoryUsage(500.0ms): 16.00 KB, 7.42 MB, 7.33 MB, 7.10 MB, 6.94 MB, 6.71 MB, 6.56 MB, 6.40 MB, 6.17 MB, 6.02 MB, 5.79 MB, 5.63 MB, 5.48 MB, 5.25 MB, 5.09 MB, 4.86 MB, 4.71 MB, 4.47 MB, 4.32 MB, 4.09 MB, 3.93 MB, 3.78 MB, 3.55 MB, 3.39 MB, 3.16 MB, 3.01 MB, 2.78 MB, 2.62 MB, 2.39 MB, 2.24 MB, 2.08 MB, 1.85 MB, 1.70 MB, 1.54 MB, 1.31 MB, 1.16 MB, 948.00 KB, 790.00 KB, 553.00 KB, 395.00 KB, 237.00 KB
    ThreadUsage(500.0ms): 1
      - AverageThreadTokens: 1.00
      - PeakMemoryUsage: 7.42 MB
      - PrepareTime: 36.144us
      - RowsProduced: 98.30K (98304)
      - TotalCpuTime: 20s449ms
      - TotalNetworkWaitTime: 191.630ms
      - TotalStorageWaitTime: 0ns
    CodeGen:(Active: 150.679ms, % non-child: 77.19%)
      - CodegenTime: 0ns
      - CompileTime: 139.503ms
      - LoadTime: 10.7ms
      - ModuleFileSize: 95.27 KB
    EXCHANGE_NODE (id=4):(Active: 194.858ms, % non-child: 99.83%)
      - BytesReceived: 2.33 MB
      - ConvertRowBatchTime: 2.732ms
      - DataArrivalWaitTime: 191.118ms
      - DeserializeRowBatchTimer: 14.943ms
      - FirstBatchArrivalWaitTime: 191.117ms
      - PeakMemoryUsage: 7.41 MB
      - RowsReturned: 98.30K (98304)
      - RowsReturnedRate: 504.49 K/sec
      - SendersBlockedTimer: 0ns
      - SendersBlockedTotalTimer(*): 0ns
    Averaged Fragment 1:(Active: 442.360ms, % non-child: 0.00%)
      split sizes: min: 33.00 B, max: 33.00 B, avg: 33.00 B, stddev: 0.00
      completion times: min:443.720ms max:443.720ms mean: 443.720ms stddev:0ns
      execution rates: min:74.00 B/sec max:74.00 B/sec mean:74.00 B/sec stddev:0.00
    /sec
      num instances: 1
        - AverageThreadTokens: 1.00
        - PeakMemoryUsage: 6.06 MB
        - PrepareTime: 7.291ms
        - RowsProduced: 98.30K (98304)
        - TotalCpuTime: 784.259ms
        - TotalNetworkWaitTime: 388.818ms
        - TotalStorageWaitTime: 3.934ms
      CodeGen:(Active: 312.862ms, % non-child: 70.73%)
        - CodegenTime: 2.669ms
        - CompileTime: 302.467ms
        - LoadTime: 9.231ms
        - ModuleFileSize: 95.27 KB
      DataStreamSender (dst_id=4):(Active: 80.63ms, % non-child: 18.10%)
        - BytesSent: 2.33 MB
        - NetworkThroughput(*): 35.89 MB/sec
        - OverallThroughput: 29.06 MB/sec
        - PeakMemoryUsage: 5.33 KB
        - SerializeBatchTime: 26.487ms
        - ThriftTransmitTime(*): 64.814ms
        - UncompressedRowBatchSize: 6.66 MB
      HASH_JOIN_NODE (id=2):(Active: 362.25ms, % non-child: 3.92%)
        - BuildBuckets: 1.02K (1024)
```

```

- BuildRows: 98.30K (98304)
- BuildTime: 12.622ms
- LoadFactor: 0.00
- PeakMemoryUsage: 6.02 MB
- ProbeRows: 3
- ProbeTime: 3.579ms
- RowsReturned: 98.30K (98304)
- RowsReturnedRate: 271.54 K/sec
EXCHANGE_NODE (id=3):(Active: 344.680ms, % non-child: 77.92%)
- BytesReceived: 1.15 MB
- ConvertRowBatchTime: 2.792ms
- DataArrivalWaitTime: 339.936ms
- DeserializeRowBatchTimer: 9.910ms
- FirstBatchArrivalWaitTime: 199.474ms
- PeakMemoryUsage: 156.00 KB
- RowsReturned: 98.30K (98304)
- RowsReturnedRate: 285.20 K/sec
- SendersBlockedTimer: 0ns
- SendersBlockedTotalTimer(*): 0ns
HDFS_SCAN_NODE (id=0):(Active: 13.616us, % non-child: 0.00%)
- AverageHdfsReadThreadConcurrency: 0.00
- AverageScannerThreadConcurrency: 0.00
- BytesRead: 33.00 B
- BytesReadLocal: 33.00 B
- BytesReadShortCircuit: 33.00 B
- NumDisksAccessed: 1
- NumScannerThreadsStarted: 1
- PeakMemoryUsage: 46.00 KB
- PerReadThreadRawHdfsThroughput: 287.52 KB/sec
- RowsRead: 3
- RowsReturned: 3
- RowsReturnedRate: 220.33 K/sec
- ScanRangesComplete: 1
- ScannerThreadsInvoluntaryContextSwitches: 26
- ScannerThreadsTotalWallClockTime: 55.199ms
- DelimiterParseTime: 2.463us
- MaterializeTupleTime(*): 1.226us
- ScannerThreadsSysTime: 0ns
- ScannerThreadsUserTime: 42.993ms
- ScannerThreadsVoluntaryContextSwitches: 1
- TotalRawHdfsReadTime(*): 112.86us
- TotalReadThroughput: 0.00 /sec
Averaged Fragment 2:(Active: 190.120ms, % non-child: 0.00%)
split sizes: min: 960.00 KB, max: 960.00 KB, avg: 960.00 KB, stddev: 0.00
completion times: min:191.736ms max:191.736ms mean: 191.736ms stddev:0ns
execution rates: min:4.89 MB/sec max:4.89 MB/sec mean:4.89 MB/sec stddev:0.00
/sec
num instances: 1
- AverageThreadTokens: 0.00
- PeakMemoryUsage: 906.33 KB
- PrepareTime: 3.67ms
- RowsProduced: 98.30K (98304)
- TotalCpuTime: 403.351ms
- TotalNetworkWaitTime: 34.999ms
- TotalStorageWaitTime: 108.675ms
CodeGen:(Active: 162.57ms, % non-child: 85.24%)
- CodegenTime: 3.133ms
- CompileTime: 148.316ms
- LoadTime: 12.317ms
- ModuleFileSize: 95.27 KB
DataStreamSender (dst_id=3):(Active: 70.620ms, % non-child: 37.14%)
- BytesSent: 1.15 MB
- NetworkThroughput(*): 23.30 MB/sec
- OverallThroughput: 16.23 MB/sec
- PeakMemoryUsage: 5.33 KB
- SerializeBatchTime: 22.69ms
- ThriftTransmitTime(*): 49.178ms
- UncompressedRowBatchSize: 3.28 MB
HDFS_SCAN_NODE (id=1):(Active: 118.839ms, % non-child: 62.51%)
- AverageHdfsReadThreadConcurrency: 0.00
- AverageScannerThreadConcurrency: 0.00
- BytesRead: 960.00 KB
- BytesReadLocal: 960.00 KB

```

## Tuning Impala for Performance

```
- BytesReadShortCircuit: 960.00 KB
- NumDisksAccessed: 1
- NumScannerThreadsStarted: 1
- PeakMemoryUsage: 869.00 KB
- PerReadThreadRawHdfsThroughput: 130.21 MB/sec
- RowsRead: 98.30K (98304)
- RowsReturned: 98.30K (98304)
- RowsReturnedRate: 827.20 K/sec
- ScanRangesComplete: 15
- ScannerThreadsInvoluntaryContextSwitches: 34
- ScannerThreadsTotalWallClockTime: 189.774ms
  - DelimiterParseTime: 15.703ms
  - MaterializeTupleTime(*): 3.419ms
  - ScannerThreadsSysTime: 1.999ms
  - ScannerThreadsUserTime: 44.993ms
- ScannerThreadsVoluntaryContextSwitches: 118
  - TotalRawHdfsReadTime(*): 7.199ms
  - TotalReadThroughput: 0.00 /sec
Fragment 1:
  Instance 6540a03d4bee0691:4963d6269b210ebf
  (host=impala-1.example.com:22000):(Active: 442.360ms, % non-child: 0.00%)
    Hdfs split stats (<volume id>:<# splits>/<split lengths>): 0:1/33.00 B
    MemoryUsage(500.0ms): 69.33 KB
    ThreadUsage(500.0ms): 1
      - AverageThreadTokens: 1.00
      - PeakMemoryUsage: 6.06 MB
      - PrepareTime: 7.291ms
      - RowsProduced: 98.30K (98304)
      - TotalCpuTime: 784.259ms
      - TotalNetworkWaitTime: 388.818ms
      - TotalStorageWaitTime: 3.934ms
    CodeGen:(Active: 312.862ms, % non-child: 70.73%)
      - CodegenTime: 2.669ms
      - CompileTime: 302.467ms
      - LoadTime: 9.231ms
      - ModuleFileSize: 95.27 KB
    DataStreamSender (dst_id=4):(Active: 80.63ms, % non-child: 18.10%)
      - BytesSent: 2.33 MB
      - NetworkThroughput(*): 35.89 MB/sec
      - OverallThroughput: 29.06 MB/sec
      - PeakMemoryUsage: 5.33 KB
      - SerializeBatchTime: 26.487ms
      - ThriftTransmitTime(*): 64.814ms
      - UncompressedRowBatchSize: 6.66 MB
  HASH_JOIN_NODE (id=2):(Active: 362.25ms, % non-child: 3.92%)
    ExecOption: Build Side Codegen Enabled, Probe Side Codegen Enabled, Hash
Table Built Asynchronously
  - BuildBuckets: 1.02K (1024)
  - BuildRows: 98.30K (98304)
  - BuildTime: 12.622ms
  - LoadFactor: 0.00
  - PeakMemoryUsage: 6.02 MB
  - ProbeRows: 3
  - ProbeTime: 3.579ms
  - RowsReturned: 98.30K (98304)
  - RowsReturnedRate: 271.54 K/sec
EXCHANGE_NODE (id=3):(Active: 344.680ms, % non-child: 77.92%)
  - BytesReceived: 1.15 MB
  - ConvertRowBatchTime: 2.792ms
  - DataArrivalWaitTime: 339.936ms
  - DeserializeRowBatchTimer: 9.910ms
  - FirstBatchArrivalWaitTime: 199.474ms
  - PeakMemoryUsage: 156.00 KB
  - RowsReturned: 98.30K (98304)
  - RowsReturnedRate: 285.20 K/sec
  - SendersBlockedTimer: 0ns
  - SendersBlockedTotalTimer(*): 0ns
HDFS_SCAN_NODE (id=0):(Active: 13.616us, % non-child: 0.00%)
  Hdfs split stats (<volume id>:<# splits>/<split lengths>): 0:1/33.00 B
  Hdfs Read Thread Concurrency Bucket: 0:0% 1:0%
  File Formats: TEXT/NONE:1
  ExecOption: Codegen enabled: 1 out of 1
    - AverageHdfsReadThreadConcurrency: 0.00
```

```

- AverageScannerThreadConcurrency: 0.00
- BytesRead: 33.00 B
- BytesReadLocal: 33.00 B
- BytesReadShortCircuit: 33.00 B
- NumDisksAccessed: 1
- NumScannerThreadsStarted: 1
- PeakMemoryUsage: 46.00 KB
- PerReadThreadRawHdfsThroughput: 287.52 KB/sec
- RowsRead: 3
- RowsReturned: 3
- RowsReturnedRate: 220.33 K/sec
- ScanRangesComplete: 1
- ScannerThreadsInvoluntaryContextSwitches: 26
- ScannerThreadsTotalWallClockTime: 55.199ms
  - DelimiterParseTime: 2.463us
  - MaterializeTupleTime(*): 1.226us
  - ScannerThreadsSysTime: 0ns
  - ScannerThreadsUserTime: 42.993ms
- ScannerThreadsVoluntaryContextSwitches: 1
- TotalRawHdfsReadTime(*): 112.86us
- TotalReadThroughput: 0.00 /sec
Fragment 2:
  Instance 6540a03d4bee0691:4963d6269b210ec0
  (host=impala-1.example.com:22000):(Active: 190.120ms, % non-child: 0.00%)
    Hdfs split stats (<volume id>:<# splits>/<split lengths>): 0:15/960.00 KB
      - AverageThreadTokens: 0.00
      - PeakMemoryUsage: 906.33 KB
      - PrepareTime: 3.67ms
      - RowsProduced: 98.30K (98304)
      - TotalCpuTime: 403.351ms
      - TotalNetworkWaitTime: 34.999ms
      - TotalStorageWaitTime: 108.675ms
  CodeGen:(Active: 162.57ms, % non-child: 85.24%)
    - CodegenTime: 3.133ms
    - CompileTime: 148.316ms
    - LoadTime: 12.317ms
    - ModuleFileSize: 95.27 KB
  DataStreamSender (dst_id=3):(Active: 70.620ms, % non-child: 37.14%)
    - BytesSent: 1.15 MB
    - NetworkThroughput(*): 23.30 MB/sec
    - OverallThroughput: 16.23 MB/sec
    - PeakMemoryUsage: 5.33 KB
    - SerializeBatchTime: 22.69ms
    - ThriftTransmitTime(*): 49.178ms
    - UncompressedRowBatchSize: 3.28 MB
  HDFS_SCAN_NODE (id=1):(Active: 118.839ms, % non-child: 62.51%)
    Hdfs split stats (<volume id>:<# splits>/<split lengths>): 0:15/960.00 KB
    Hdfs Read Thread Concurrency Bucket: 0:0% 1:0%
    File Formats: TEXT/NONE:15
    ExecOption: Codegen enabled: 15 out of 15
      - AverageHdfsReadThreadConcurrency: 0.00
      - AverageScannerThreadConcurrency: 0.00
      - BytesRead: 960.00 KB
      - BytesReadLocal: 960.00 KB
      - BytesReadShortCircuit: 960.00 KB
      - NumDisksAccessed: 1
      - NumScannerThreadsStarted: 1
      - PeakMemoryUsage: 869.00 KB
      - PerReadThreadRawHdfsThroughput: 130.21 MB/sec
      - RowsRead: 98.30K (98304)
      - RowsReturned: 98.30K (98304)
      - RowsReturnedRate: 827.20 K/sec
      - ScanRangesComplete: 15
      - ScannerThreadsInvoluntaryContextSwitches: 34
      - ScannerThreadsTotalWallClockTime: 189.774ms
        - DelimiterParseTime: 15.703ms
        - MaterializeTupleTime(*): 3.419ms
        - ScannerThreadsSysTime: 1.999ms
        - ScannerThreadsUserTime: 44.993ms
      - ScannerThreadsVoluntaryContextSwitches: 118
      - TotalRawHdfsReadTime(*): 7.199ms
      - TotalReadThroughput: 0.00 /sec

```

### Detecting and Correcting HDFS Block Skew Conditions

For best performance of Impala parallel queries, the work is divided equally across hosts in the cluster, and all hosts take approximately equal time to finish their work. If one host takes substantially longer than others, the extra time needed for the slow host can become the dominant factor in query performance. Therefore, one of the first steps in performance tuning for Impala is to detect and correct such conditions.

The main cause of uneven performance that you can correct within Impala is **skew** in the number of HDFS data blocks processed by each host, where some hosts process substantially more data blocks than others. This condition can occur because of uneven distribution of the data values themselves, for example causing certain data files or partitions to be large while others are very small. (Although it is possible to have unevenly distributed data without any problems with the distribution of HDFS blocks.) Block skew could also be due to the underlying block allocation policies within HDFS, the replication factor of the data files, and the way that Impala chooses the host to process each data block.

The most convenient way to detect block skew, or slow-host issues in general, is to examine the "executive summary" information from the query profile after running a query:

- In `impala-shell`, issue the `SUMMARY` command immediately after the query is complete, to see just the summary information. If you detect issues involving skew, you might switch to issuing the `PROFILE` command, which displays the summary information followed by a detailed performance analysis.
- In the Cloudera Manager interface or the Impala debug web UI, click on the **Profile** link associated with the query after it is complete. The executive summary information is displayed early in the profile output.

For each phase of the query, you see an **Avg Time** and a **Max Time** value, along with **#Hosts** indicating how many hosts are involved in that query phase. For all the phases with **#Hosts** greater than one, look for cases where the maximum time is substantially greater than the average time. Focus on the phases that took the longest, for example, those taking multiple seconds rather than milliseconds or microseconds.

If you detect that some hosts take longer than others, first rule out non-Impala causes. One reason that some hosts could be slower than others is if those hosts have less capacity than the others, or if they are substantially busier due to unevenly distributed non-Impala workloads:

- For clusters running Impala, keep the relative capacities of all hosts roughly equal. Any cost savings from including some underpowered hosts in the cluster will likely be outweighed by poor or uneven performance, and the time spent diagnosing performance issues.
- If non-Impala workloads cause slowdowns on some hosts but not others, use the appropriate load-balancing techniques for the non-Impala components to smooth out the load across the cluster.

If the hosts on your cluster are evenly powered and evenly loaded, examine the detailed profile output to determine which host is taking longer than others for the query phase in question. Examine how many bytes are processed during that phase on that host, how much memory is used, and how many bytes are transmitted across the network.

The most common symptom is a higher number of bytes read on one host than others, due to one host being requested to process a higher number of HDFS data blocks. This condition is more likely to occur when the number of blocks accessed by the query is relatively small. For example, if you have a 10-node cluster and the query processes 10 HDFS blocks, each node might not process exactly one block. If one node sits idle while another node processes two blocks, the query could take twice as long as if the data was perfectly distributed.

Possible solutions in this case include:

- If the query is artificially small, perhaps for benchmarking purposes, scale it up to process a larger data set. For example, if some nodes read 10 HDFS data blocks while others read 11, the overall effect of the uneven distribution is much lower than when some nodes did twice as much work as others. As a guideline, aim for a "sweet spot" where each node reads 2 GB or more from HDFS per query. Queries that process lower volumes than that could experience inconsistent performance that smooths out as queries become more data-intensive.

- If the query processes only a few large blocks, so that many nodes sit idle and cannot help to parallelize the query, consider reducing the overall block size. For example, you might adjust the `PARQUET_FILE_SIZE` query option before copying or converting data into a Parquet table. Or you might adjust the granularity of data files produced earlier in the ETL pipeline by non-Impala components. In Impala 2.0 and later, the default Parquet block size is 256 MB, reduced from 1 GB, to improve parallelism for common cluster sizes and data volumes.
- Reduce the amount of compression applied to the data. For text data files, the highest degree of compression (gzip) produces unsplittable files that are more difficult for Impala to process in parallel, and require extra memory during processing to hold the compressed and uncompressed data simultaneously. For binary formats such as Parquet and Avro, compression can result in fewer data blocks overall, but remember that when queries process relatively few blocks, there is less opportunity for parallel execution and many nodes in the cluster might sit idle. Note that when Impala writes Parquet data with the query option `COMPRESSION_CODEC=NONE` enabled, the data is still typically compact due to the encoding schemes used by Parquet, independent of the final compression step.

# Scalability Considerations for Impala

This section explains how the size of your cluster and the volume of data influences SQL performance and schema design for Impala tables. Typically, adding more cluster capacity reduces problems due to memory limits or disk throughput. On the other hand, larger clusters are more likely to have other kinds of scalability issues, such as a single slow node that causes performance problems for queries.

A good source of tips related to scalability and performance tuning is the [Impala Cookbook](#) presentation. These slides are updated periodically as new features come out and new benchmarks are performed.

## Scalability Considerations for the Impala Statestore

Before CDH 5.3, the statestore sent only one kind of message to its subscribers. This message contained all updates for any topics that a subscriber had subscribed to. It also served to let subscribers know that the statestore had not failed, and conversely the statestore used the success of sending a heartbeat to a subscriber to decide whether or not the subscriber had failed.

Combining topic updates and failure detection in a single message led to bottlenecks in clusters with large numbers of tables, partitions, and HDFS data blocks. When the statestore was overloaded with metadata updates to transmit, heartbeat messages were sent less frequently, sometimes causing subscribers to time out their connection with the statestore. Increasing the subscriber timeout and decreasing the frequency of statestore heartbeats worked around the problem, but reduced responsiveness when the statestore failed or restarted.

As of CDH 5.3, the statestore now sends topic updates and heartbeats in separate messages. This allows the statestore to send and receive a steady stream of lightweight heartbeats, and removes the requirement to send topic updates according to a fixed schedule, reducing statestore network overhead.

The statestore now has the following relevant configuration flags for the `statestored` daemon:

**-statestore\_num\_update\_threads**

The number of threads inside the statestore dedicated to sending topic updates. You should not typically need to change this value.

**Default:** 10

**-statestore\_update\_frequency\_ms**

The frequency, in milliseconds, with which the statestore tries to send topic updates to each subscriber. This is a best-effort value; if the statestore is unable to meet this frequency, it sends topic updates as fast as it can. You should not typically need to change this value.

**Default:** 2000

**-statestore\_num\_heartbeat\_threads**

The number of threads inside the statestore dedicated to sending heartbeats. You should not typically need to change this value.

**Default:** 10

**-statestore\_heartbeat\_frequency\_ms**

The frequency, in milliseconds, with which the statestore tries to send heartbeats to each subscriber. This value should be good for large catalogs and clusters up to approximately 150 nodes. Beyond that, you might need to increase this value to make the interval longer between heartbeat messages.

**Default:** 1000 (one heartbeat message every second)

As of CDH 5.3, not all of these flags are present in the Cloudera Manager user interface. Some must be set using the **Advanced Configuration Snippet** fields for the statestore component.

If it takes a very long time for a cluster to start up, and `impala-shell` consistently displays `This Impala daemon is not ready to accept user requests, the statestore might be taking too long to send the entire catalog topic to the cluster.` In this case, consider adding `--load_catalog_in_background=false` to your catalog service configuration. This setting stops the statestore from loading the entire catalog into memory at cluster startup. Instead, metadata for each table is loaded when the table is accessed for the first time.

## SQL Operations that Spill to Disk

Certain memory-intensive operations write temporary data to disk (known as *spilling* to disk) when Impala is close to exceeding its memory limit on a particular host.

The result is a query that completes successfully, rather than failing with an out-of-memory error. The tradeoff is decreased performance due to the extra disk I/O to write the temporary data and read it back in. The slowdown could be potentially be significant. Thus, while this feature improves reliability, you should optimize your queries, system parameters, and hardware configuration to make this spilling a rare occurrence.

### What kinds of queries might spill to disk:

For example, when a query uses a `GROUP BY` clause for columns with millions or billions of distinct values, Impala keeps a similar number of temporary results in memory, to accumulate the aggregate results for each value in the group. When large tables are joined together, Impala keeps the values of the join columns from one table in memory, to compare them to incoming values from the other table. When a large result set is sorted by the `ORDER BY` clause, each node sorts its portion of the result set in memory. The `DISTINCT` and `UNION` operators also build in-memory data structures to represent all values found so far, to eliminate duplicates as the query progresses.

### Memory usage for SQL operators:

The infrastructure of the spilling feature affects the way the affected SQL operators, such as `GROUP BY`, `DISTINCT`, and joins, use memory. On each host that participates in the query, each such operator in a query accumulates memory while building the data structure to process the aggregation or join operation. The amount of memory used depends on the portion of the data being handled by that host, and thus might be different from one host to another. When the amount of memory being used for the operator on a particular host reaches a threshold amount, Impala reserves an additional memory buffer to use as a work area in case that operator causes the query to exceed the memory limit for that host. After allocating the memory buffer, the memory used by that operator remains essentially stable or grows only slowly, until the point where the memory limit is reached and the query begins writing temporary data to disk.

Prior to Impala 2.2 (CDH 5.4), the extra memory buffer for an operator that might spill to disk is allocated when the data structure used by the applicable SQL operator reaches 16 MB in size, and the memory buffer itself is 512 MB. In Impala 2.2 and higher, these values are halved: the threshold value is 8 MB and the memory buffer is 256 MB. A query that uses multiple such operators might allocate multiple such memory buffers, as the size of the data structure for each operator crosses the threshold on a particular host.

Therefore, a query that processes a relatively small amount of data on each host would likely never reach the threshold for any operator, and would never allocate any extra memory buffers. A query that did process millions of groups, distinct values, join keys, and so on might cross the threshold, causing its memory requirement to rise suddenly and then flatten out. The larger the cluster, less data is processed on any particular host, thus reducing the chance of requiring the extra memory allocation.

**Added in:** This feature was added to the `ORDER BY` clause in Impala 1.4 for CDH 4, and in CDH 5.1. This feature was extended to cover join queries, aggregation functions, and analytic functions in Impala 2.0 for CDH 4, and in CDH 5.2. The size of the memory work area required by each operator that spills was reduced from 512 megabytes to 256 megabytes in Impala 2.2 (CDH 5.4).

### Avoiding queries that spill to disk:

Because the extra I/O can impose significant performance overhead on these types of queries, try to avoid this situation by using the following steps:

## Scalability Considerations for Impala

1. Detect how often queries spill to disk, and how much temporary data is written. Refer to the following sources:
  - The output of the `PROFILE` command in the `impala-shell` interpreter. This data shows the memory usage for each host and in total across the cluster. The `BlockMgr.BytesWritten` counter reports how much data was written to disk during the query.
  - The **Impala Queries** dialog in Cloudera Manager. You can see the peak memory usage for a query, combined across all nodes in the cluster.
  - The **Queries** tab in the Impala debug web user interface. Select the query to examine and click the corresponding **Profile** link. This data breaks down the memory usage for a single host within the cluster, the host whose web interface you are connected to.
2. Use one or more techniques to reduce the possibility of the queries spilling to disk:
  - Increase the Impala memory limit if practical, for example, if you can increase the available memory by more than the amount of temporary data written to disk on a particular node. Remember that in Impala 2.0 and later, you can issue `SET MEM_LIMIT` as a SQL statement, which lets you fine-tune the memory usage for queries from JDBC and ODBC applications.
  - Increase the number of nodes in the cluster, to increase the aggregate memory available to Impala and reduce the amount of memory required on each node.
  - Increase the overall memory capacity of each data node at the hardware level.
  - On a cluster with resources shared between Impala and other Hadoop components, use resource management features to allocate more memory for Impala. See [Integrated Resource Management with YARN](#) on page 78 for details.
  - If the memory pressure is due to running many concurrent queries rather than a few memory-intensive ones, consider using the Impala admission control feature to lower the limit on the number of concurrent queries. By spacing out the most resource-intensive queries, you can avoid spikes in memory usage and improve overall response times. See [Admission Control and Query Queuing](#) on page 69 for details.
  - Tune the queries with the highest memory requirements, using one or more of the following techniques:
    - Run the `COMPUTE STATS` statement for all tables involved in large-scale joins and aggregation queries.
    - Minimize your use of `STRING` columns in join columns. Prefer numeric values instead.
    - Examine the `EXPLAIN` plan to understand the execution strategy being used for the most resource-intensive queries. See [Using the EXPLAIN Plan for Performance Tuning](#) on page 379 for details.
    - If Impala still chooses a suboptimal execution strategy even with statistics available, or if it is impractical to keep the statistics up to date for huge or rapidly changing tables, add hints to the most resource-intensive queries to select the right execution strategy. See [Hints](#) on page 239 for details.
  - If your queries experience substantial performance overhead due to spilling, enable the `DISABLE_UNSAFE_SPILLS` query option. This option prevents queries whose memory usage is likely to be exorbitant from spilling to disk. See [DISABLE\\_UNSAFE\\_SPILLS Query Option](#) on page 342 for details. As you tune problematic queries using the preceding steps, fewer and fewer will be cancelled by this option setting.

### Testing performance implications of spilling to disk:

To artificially provoke spilling, to test this feature and understand the performance implications, use a test environment with a memory limit of at least 2 GB. Issue the `SET` command with no arguments to check the current setting for the `MEM_LIMIT` query option. Set the query option `DISABLE_UNSAFE_SPILLS=true`. This option limits the spill-to-disk feature to prevent runaway disk usage from queries that are known in advance to be suboptimal. Within `impala-shell`, run a query that you expect to be memory-intensive, based on the criteria explained earlier. A self-join of a large table is a good candidate:

```
select count(*) from big_table a join big_table b using (column_with_many_values);
```

Issue the `PROFILE` command to get a detailed breakdown of the memory usage on each node during the query. The crucial part of the profile output concerning memory is the `BlockMgr` portion. For example, this profile shows that the query did not quite exceed the memory limit.

```
BlockMgr:
  - BlockWritesIssued: 1
  - BlockWritesOutstanding: 0
  - BlocksCreated: 24
  - BlocksRecycled: 1
  - BufferedPins: 0
  - MaxBlockSize: 8.00 MB (8388608)
  - MemoryLimit: 200.00 MB (209715200)
  - PeakMemoryUsage: 192.22 MB (201555968)
  - TotalBufferWaitTime: 0ns
  - TotalEncryptionTime: 0ns
  - TotalIntegrityCheckTime: 0ns
  - TotalReadBlockTime: 0ns
```

In this case, because the memory limit was already below any recommended value, I increased the volume of data for the query rather than reducing the memory limit any further.

Set the `MEM_LIMIT` query option to a value that is smaller than the peak memory usage reported in the profile output. Do not specify a memory limit lower than about 300 MB, because with such a low limit, queries could fail to start for other reasons. Now try the memory-intensive query again.

Check if the query fails with a message like the following:

```
WARNINGS: Spilling has been disabled for plans that do not have stats and are not hinted
to prevent potentially bad plans from using too many cluster resources. Compute stats
on
these tables, hint the plan or disable this behavior via query options to enable
spilling.
```

If so, the query could have consumed substantial temporary disk space, slowing down so much that it would not complete in any reasonable time. Rather than rely on the spill-to-disk feature in this case, issue the `COMPUTE STATS` statement for the table or tables in your sample query. Then run the query again, check the peak memory usage again in the `PROFILE` output, and adjust the memory limit again if necessary to be lower than the peak memory usage.

At this point, you have a query that is memory-intensive, but Impala can optimize it efficiently so that the memory usage is not exorbitant. You have set an artificial constraint through the `MEM_LIMIT` option so that the query would normally fail with an out-of-memory error. But the automatic spill-to-disk feature means that the query should actually succeed, at the expense of some extra disk I/O to read and write temporary work data.

Try the query again, and confirm that it succeeds. Examine the `PROFILE` output again. This time, look for lines of this form:

```
- SpilledPartitions: N
```

If you see any such lines with *N* greater than 0, that indicates the query would have failed in Impala releases prior to 2.0, but now it succeeded because of the spill-to-disk feature. Examine the total time taken by the `AGGREGATION_NODE` or other query fragments containing non-zero `SpilledPartitions` values. Compare the times to similar fragments that did not spill, for example in the `PROFILE` output when the same query is run with a higher memory limit. This gives you an idea of the performance penalty of the spill operation for a particular query with a particular memory limit. If you make the memory limit just a little lower than the peak memory usage, the query only needs to write a small amount of temporary data to disk. The lower you set the memory limit, the more temporary data is written and the slower the query becomes.

Now repeat this procedure for actual queries used in your environment. Use the `DISABLE_UNSAFE_SPILLS` setting to identify cases where queries used more memory than necessary due to lack of statistics on the relevant tables and columns, and issue `COMPUTE STATS` where necessary.

#### When to use `DISABLE_UNSAFE_SPILLS`:

## Scalability Considerations for Impala

You might wonder, why not leave `DISABLE_UNSAFE_SPILLS` turned on all the time. Whether and how frequently to use this option depends on your system environment and workload.

`DISABLE_UNSAFE_SPILLS` is suitable for an environment with ad hoc queries whose performance characteristics and memory usage are not known in advance. It prevents “worst-case scenario” queries that use large amounts of memory unnecessarily. Thus, you might turn this option on within a session while developing new SQL code, even though it is turned off for existing applications.

Organizations where table and column statistics are generally up-to-date might leave this option turned on all the time, again to avoid worst-case scenarios for untested queries or if a problem in the ETL pipeline results in a table with no statistics. Turning on `DISABLE_UNSAFE_SPILLS` lets you “fail fast” in this case and immediately gather statistics or tune the problematic queries.

Some organizations might leave this option turned off. For example, you might have tables large enough that the `COMPUTE STATS` takes substantial time to run, making it impractical to re-run after loading new data. If you have examined the `EXPLAIN` plans of your queries and know that they are operating efficiently, you might leave `DISABLE_UNSAFE_SPILLS` turned off. In that case, you know that any queries that spill will not go overboard with their memory consumption.

### Turning off the spill-to-disk feature:

You might turn off the spill-to-disk feature if you are in an environment with constraints on disk space, or if you prefer for queries that exceed the memory capacity in your cluster to “fail fast” so that you can tune and retry them.

To turn off this feature, set the following configuration options for each `impalad` daemon, either through the `impalad` advanced configuration snippet in Cloudera Manager, or during `impalad` startup on each data node on systems not managed by Cloudera Manager:

```
-enable_partitioned_aggregation=false  
-enable_partitioned_hash_join=false
```

## Limits on Query Size and Complexity

There are hardcoded limits on the maximum size and complexity of queries. Currently, the maximum number of expressions in a query is 2000. You might exceed the limits with large or deeply nested queries produced by business intelligence tools or other query generators.

If you have the ability to customize such queries or the query generation logic that produces them, replace sequences of repetitive expressions with single operators such as `IN` or `BETWEEN` that can represent multiple values or ranges. For example, instead of a large number of `OR` clauses:

```
WHERE val = 1 OR val = 2 OR val = 6 OR val = 100 ...
```

use a single `IN` clause:

```
WHERE val IN (1,2,6,100,...)
```

## Scalability Considerations for Impala I/O

Impala parallelizes its I/O operations aggressively, therefore the more disks you can attach to each host, the better. Impala retrieves data from disk so quickly using bulk read operations on large blocks, that most queries are CPU-bound rather than I/O-bound.

Because the kind of sequential scanning typically done by Impala queries does not benefit much from the random-access capabilities of SSDs, spinning disks typically provide the most cost-effective kind of storage for Impala data, with little or no performance penalty as compared to SSDs.

Resource management features such as YARN, Llama, and admission control typically constrain the amount of memory, CPU, or overall number of queries in a high-concurrency environment. Currently, there is no throttling mechanism for Impala I/O.

## Scalability Considerations for Table Layout

Due to the overhead of retrieving and updating table metadata in the metastore database, try to limit the number of columns in a table to a maximum of approximately 2000. Although Impala can handle wider tables than this, the metastore overhead can become significant, leading to query performance that is slower than expected based on the actual data volume.

To minimize overhead related to the metastore database and Impala query planning, try to limit the number of partitions for any partitioned table to a few tens of thousands.

# Partitioning for Impala Tables

By default, all the data files for a table are located in a single directory. Partitioning is a technique for physically dividing the data during loading, based on values from one or more columns, to speed up queries that test those columns. For example, with a `school_records` table partitioned on a `year` column, there is a separate data directory for each different year value, and all the data for that year is stored in a data file in that directory. A query that includes a `WHERE` condition such as `YEAR=1966`, `YEAR IN (1989,1999)`, or `YEAR BETWEEN 1984 AND 1989` can examine only the data files from the appropriate directory or directories, greatly reducing the amount of data to read and test.

See [Attaching an External Partitioned Table to an HDFS Directory Structure](#) on page 63 for an example that illustrates the syntax for creating partitioned tables, the underlying directory structure in HDFS, and how to attach a partitioned Impala external table to data files stored elsewhere in HDFS.

Parquet is a popular format for partitioned Impala tables because it is well suited to handle huge data volumes. See [Query Performance for Impala Parquet Tables](#) on page 410 for performance considerations for partitioned Parquet tables.

See [NULL](#) on page 141 for details about how `NULL` values are represented in partitioned tables.

See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details about setting up tables where some or all partitions reside on the Amazon Simple Storage Service (S3).

## When to Use Partitioned Tables

Partitioning is typically appropriate for:

- Tables that are very large, where reading the entire data set takes an impractical amount of time.
- Tables that are always or almost always queried with conditions on the partitioning columns. In our example of a table partitioned by year, `SELECT COUNT(*) FROM school_records WHERE year = 1985` is efficient, only examining a small fraction of the data; but `SELECT COUNT(*) FROM school_records` has to process a separate data file for each year, resulting in more overall work than in an unpartitioned table. You would probably not partition this way if you frequently queried the table based on last name, student ID, and so on without testing the year.
- Columns that have reasonable cardinality (number of different values). If a column only has a small number of values, for example `Male` or `Female`, you do not gain much efficiency by eliminating only about 50% of the data to read for each query. If a column has only a few rows matching each value, the number of directories to process can become a limiting factor, and the data file in each directory could be too small to take advantage of the Hadoop mechanism for transmitting data in multi-megabyte blocks. For example, you might partition census data by year, store sales data by year and month, and web traffic data by year, month, and day. (Some users with high volumes of incoming data might even partition down to the individual hour and minute.)
- Data that already passes through an extract, transform, and load (ETL) pipeline. The values of the partitioning columns are stripped from the original data files and represented by directory names, so loading data into a partitioned table involves some sort of transformation or preprocessing.

## SQL Statements for Partitioned Tables

In terms of Impala SQL syntax, partitioning affects these statements:

- [`CREATE TABLE`](#): you specify a `PARTITIONED BY` clause when creating the table to identify names and data types of the partitioning columns. These columns are not included in the main list of columns for the table.
- [`ALTER TABLE`](#): you can add or drop partitions, to work with different portions of a huge data set. You can designate the HDFS directory that holds the data files for a specific partition. With data partitioned by date values, you might “age out” data that is no longer relevant.

- **Note:** If you are creating a partition for the first time and specifying its location, for maximum efficiency, use a single `ALTER TABLE` statement including both the `ADD PARTITION` and `LOCATION` clauses, rather than separate statements with `ADD PARTITION` and `SET LOCATION` clauses.

- [INSERT](#): When you insert data into a partitioned table, you identify the partitioning columns. One or more values from each inserted row are not stored in data files, but instead determine the directory where that row value is stored. You can also specify which partition to load a set of data into, with `INSERT OVERWRITE` statements; you can replace the contents of a specific partition but you cannot append data to a specific partition.

By default, if an `INSERT` statement creates any new subdirectories underneath a partitioned table, those subdirectories are assigned default HDFS permissions for the `impala` user. To make each subdirectory have the same permissions as its parent directory in HDFS, specify the `--insert_inherit_permissions` startup option for the `impalad` daemon.

- Although the syntax of the [SELECT](#) statement is the same whether or not the table is partitioned, the way queries interact with partitioned tables can have a dramatic impact on performance and scalability. The mechanism that lets queries skip certain partitions during a query is known as partition pruning; see [Partition Pruning for Queries](#) on page 396 for details.
- In Impala 1.4 and later, there is a `SHOW PARTITIONS` statement that displays information about each partition in a table. See [SHOW Statement](#) on page 242 for details.

## Static and Dynamic Partitioning Clauses

Specifying all the partition columns in a SQL statement is called “static partitioning”, because the statement affects a single predictable partition. For example, you use static partitioning with an `ALTER TABLE` statement that affects only one partition, or with an `INSERT` statement that inserts all values into the same partition:

```
insert into t1 partition(x=10, y='a') select c1 from some_other_table;
```

When you specify some partition key columns in an `INSERT` statement, but leave out the values, Impala determines which partition to insert. This technique is called “dynamic partitioning”:

```
insert into t1 partition(x, y='b') select c1, c2 from some_other_table;
-- Create new partition if necessary based on variable year, month, and day; insert a
single value.
insert into weather partition (year, month, day) select 'cloudy',2014,4,21;
-- Create new partition if necessary for specified year and month but variable day;
insert a single value.
insert into weather partition (year=2014, month=04, day) select 'sunny',22;
```

The more key columns you specify in the `PARTITION` clause, the fewer columns you need in the `SELECT` list. The trailing columns in the `SELECT` list are substituted in order for the partition key columns with no specified value.

## Permissions for Partition Subdirectories

By default, if an `INSERT` statement creates any new subdirectories underneath a partitioned table, those subdirectories are assigned default HDFS permissions for the `impala` user. To make each subdirectory have the same permissions as its parent directory in HDFS, specify the `--insert_inherit_permissions` startup option for the `impalad` daemon.

### Partition Pruning for Queries

Partition pruning refers to the mechanism where a query can skip reading the data files corresponding to one or more partitions. If you can arrange for queries to prune large numbers of unnecessary partitions from the query execution plan, the queries use fewer resources and are thus proportionally faster and more scalable.

For example, if a table is partitioned by columns `YEAR`, `MONTH`, and `DAY`, then `WHERE` clauses such as `WHERE year = 2013`, `WHERE year < 2010`, or `WHERE year BETWEEN 1995 AND 1998` allow Impala to skip the data files in all partitions outside the specified range. Likewise, `WHERE year = 2013 AND month BETWEEN 1 AND 3` could prune even more partitions, reading the data files for only a portion of one year.

#### Checking if Partition Pruning Happens for a Query

To check the effectiveness of partition pruning for a query, check the `EXPLAIN` output for the query before running it. For example, this example shows a table with 3 partitions, where the query only reads 1 of them. The notation `#partitions=1/3` in the `EXPLAIN` plan confirms that Impala can do the appropriate partition pruning.

```
[localhost:21000] > insert into census partition (year=2010) values ('Smith'),('Jones');
[localhost:21000] > insert into census partition (year=2011) values
('Smith'),('Jones'),('Doe');
[localhost:21000] > insert into census partition (year=2012) values ('Smith'),('Doe');
[localhost:21000] > select name from census where year=2010;
+-----+
| name   |
+-----+
| Smith  |
| Jones  |
+-----+
[localhost:21000] > explain select name from census where year=2010;
+-----+
| Explain String
+-----+
PLAN FRAGMENT 0
PARTITION: UNPARTITIONED

1:EXCHANGE

PLAN FRAGMENT 1
PARTITION: RANDOM

STREAM DATA SINK
EXCHANGE ID: 1
UNPARTITIONED

0:SCAN HDFS
  table=predicate_propagation.census #partitions=1/3 size=12B
+
```

For a report of the volume of data that was actually read and processed at each stage of the query, check the output of the `SUMMARY` command immediately after running the query. For a more detailed analysis, look at the output of the `PROFILE` command; it includes this same summary report near the start of the profile output.

#### What SQL Constructs Work with Partition Pruning

Impala can even do partition pruning in cases where the partition key column is not directly compared to a constant, by applying the transitive property to other parts of the `WHERE` clause. This technique is known as predicate propagation, and is available in Impala 1.2.2 and later. In this example, the `census` table includes another column indicating when the data was collected, which happens in 10-year intervals. Even though the query does not compare the partition key column (`YEAR`) to a constant value, Impala can deduce that only the partition `YEAR=2010` is required, and again only reads 1 out of 3 partitions.

```
[localhost:21000] > drop table census;
[localhost:21000] > create table census (name string, census_year int) partitioned by
```

```

(year int);
[localhost:21000] > insert into census partition (year=2010) values
('Smith',2010),('Jones',2010);
[localhost:21000] > insert into census partition (year=2011) values
('Smith',2020),('Jones',2020),('Doe',2020);
[localhost:21000] > insert into census partition (year=2012) values
('Smith',2020),('Doe',2020);
[localhost:21000] > select name from census where year = census_year and
census_year=2010;
+-----+
| name |
+-----+
| Smith |
| Jones |
+-----+
[localhost:21000] > explain select name from census where year = census_year and
census_year=2010;
+-----+
| Explain String
+-----+
| PLAN FRAGMENT 0
  PARTITION: UNPARTITIONED
    1:EXCHANGE
| PLAN FRAGMENT 1
  PARTITION: RANDOM
    STREAM DATA SINK
      EXCHANGE ID: 1
      UNPARTITIONED
    0:SCAN HDFS
      table=predicate_propagation.census #partitions=1/3 size=22B
      predicates: census_year = 2010, year = census_year
+-----+

```

If a view applies to a partitioned table, any partition pruning is determined by the clauses in the original query. Impala does not prune additional columns if the query on the view includes extra WHERE clauses referencing the partition key columns.

## Partition Key Columns

The columns you choose as the partition keys should be ones that are frequently used to filter query results in important, large-scale queries. Popular examples are some combination of year, month, and day when the data has associated time values, and geographic region when the data is associated with some place.

- For time-based data, split out the separate parts into their own columns, because Impala cannot partition based on a `TIMESTAMP` column.
- The data type of the partition columns does not have a significant effect on the storage required, because the values from those columns are not stored in the data files, rather they are represented as strings inside HDFS directory names.
- Remember that when Impala queries data stored in HDFS, it is most efficient to use multi-megabyte files to take advantage of the HDFS block size. For Parquet tables, the block size (and ideal size of the data files) is 256 MB in Impala 2.0 and later. Therefore, avoid specifying too many partition key columns, which could result in individual partitions containing only small amounts of data. For example, if you receive 1 GB of data per day, you might partition by year, month, and day; while if you receive 5 GB of data per minute, you might partition by year, month, day, hour, and minute. If you have data with a geographic component, you might partition based on postal code if you have many megabytes of data for each postal code, but if not, you might partition by some larger region such as city, state, or country. state

### Setting Different File Formats for Partitions

Partitioned tables have the flexibility to use different file formats for different partitions. (For background information about the different file formats Impala supports, see [How Impala Works with Hadoop File Formats](#) on page 399.) For example, if you originally received data in text format, then received new data in RCFile format, and eventually began receiving data in Parquet format, all that data could reside in the same table for queries. You just need to ensure that the table is structured so that the data files that use different file formats reside in separate partitions.

For example, here is how you might switch from text to Parquet data as you receive data for different years:

```
[localhost:21000] > create table census (name string) partitioned by (year smallint);
[localhost:21000] > alter table census add partition (year=2012); -- Text format;

[localhost:21000] > alter table census add partition (year=2013); -- Text format switches
    to Parquet before data loaded;
[localhost:21000] > alter table census partition (year=2013) set fileformat parquet;

[localhost:21000] > insert into census partition (year=2012) values
    ('Smith'),('Jones'),('Lee'),('Singh');
[localhost:21000] > insert into census partition (year=2013) values
    ('Flores'),('Bogomolov'),('Cooper'),('Appiah');
```

At this point, the HDFS directory for `year=2012` contains a text-format data file, while the HDFS directory for `year=2013` contains a Parquet data file. As always, when loading non-trivial data, you would use `INSERT ... SELECT` or `LOAD DATA` to import data in large batches, rather than `INSERT ... VALUES` which produces small files that are inefficient for real-world queries.

For other file types that Impala cannot create natively, you can switch into Hive and issue the `ALTER TABLE ... SET FILEFORMAT` statements and `INSERT` or `LOAD DATA` statements there. After switching back to Impala, issue a `REFRESH table_name` statement so that Impala recognizes any partitions or new data added through Hive.

### Managing Partitions

You can add, drop, set the expected file format, or set the HDFS location of the data files for individual partitions within an Impala table. See [ALTER TABLE Statement](#) on page 164 for syntax details, and [Setting Different File Formats for Partitions](#) on page 398 for tips on managing tables containing partitions with different file formats.

- **Note:** If you are creating a partition for the first time and specifying its location, for maximum efficiency, use a single `ALTER TABLE` statement including both the `ADD PARTITION` and `LOCATION` clauses, rather than separate statements with `ADD PARTITION` and `SET LOCATION` clauses.

What happens to the data files when a partition is dropped depends on whether the partitioned table is designated as internal or external. For an internal (managed) table, the data files are deleted. For example, if data in the partitioned table is a copy of raw data files stored elsewhere, you might save disk space by dropping older partitions that are no longer required for reporting, knowing that the original data is still available if needed later. For an external table, the data files are left alone. For example, dropping a partition without deleting the associated files lets Impala consider a smaller set of partitions, improving query efficiency and reducing overhead for DDL operations on the table; if the data is needed again later, you can add the partition again. See [Tables](#) on page 156 for details and examples.

## How Impala Works with Hadoop File Formats

Impala supports several familiar file formats used in Apache Hadoop. Impala can load and query data files produced by other Hadoop components such as Pig or MapReduce, and data files produced by Impala can be used by other components also. The following sections discuss the procedures, limitations, and performance considerations for using each file format with Impala.

The file format used for an Impala table has significant performance consequences. Some file formats include compression support that affects the size of data on the disk and, consequently, the amount of I/O and CPU resources required to deserialize data. The amounts of I/O and CPU resources required can be a limiting factor in query performance since querying often begins with moving and decompressing data. To reduce the potential impact of this part of the process, data is often compressed. By compressing data, a smaller total number of bytes are transferred from disk to memory. This reduces the amount of time taken to transfer the data, but a tradeoff occurs when the CPU decompresses the content.

Impala can query files encoded with most of the popular file formats and compression codecs used in Hadoop. Impala can create and insert data into tables that use some file formats but not others; for file formats that Impala cannot write to, create the table in Hive, issue the `INVALIDATE METADATA table_name` statement in `impala-shell`, and query the table through Impala. File formats can be structured, in which case they may include metadata and built-in compression. Supported formats include:

**Table 4: File Format Support in Impala**

File Type	Format	Compression Codecs	Impala Can CREATE?	Impala Can INSERT?
<a href="#">Parquet</a>	Structured	Snappy, gzip; currently Snappy by default	Yes.	Yes: <code>CREATE TABLE</code> , <code>INSERT</code> , <code>LOAD DATA</code> , and query.
<a href="#">Text</a>	Unstructured	LZO, gzip, bzip2, Snappy	Yes. For <code>CREATE TABLE</code> with no <code>STORED AS</code> clause, the default file format is uncompressed text, with values separated by ASCII 0x01 characters (typically represented as Ctrl-A).	Yes: <code>CREATE TABLE</code> , <code>INSERT</code> , <code>LOAD DATA</code> , and query. If LZO compression is used, you must create the table and load data in Hive. If other kinds of compression are used, you must load data through <code>LOAD DATA</code> , Hive, or manually in HDFS.
<a href="#">Avro</a>	Structured	Snappy, gzip, deflate, bzip2	Yes, in Impala 1.4.0 and higher. Before that, create the table using Hive.	No. Import data by using <code>LOAD DATA</code> on data files already in the right format, or use <code>INSERT</code> in Hive followed by <code>REFRESH table_name</code> in Impala.
<a href="#">RCFile</a>	Structured	Snappy, gzip, deflate, bzip2	Yes.	No. Import data by using <code>LOAD DATA</code> on data files already in the right format, or use <code>INSERT</code> in Hive followed by <code>REFRESH table_name</code> in Impala.
<a href="#">SequenceFile</a>	Structured	Snappy, gzip, deflate, bzip2	Yes.	No. Import data by using <code>LOAD DATA</code> on data files already in the right format, or use <code>INSERT</code> in Hive followed by <code>REFRESH table_name</code> in Impala.

Impala supports the following compression codecs:

## How Impala Works with Hadoop File Formats

- Snappy. Recommended for its effective balance between compression ratio and decompression speed. Snappy compression is very fast, but gzip provides greater space savings. Supported for text files in Impala 2.0 and higher.
- Gzip. Recommended when achieving the highest level of compression (and therefore greatest disk-space savings) is desired. Supported for text files in Impala 2.0 and higher.
- Deflate. Not supported for text files.
- Bzip2. Supported for text files in Impala 2.0 and higher.
- LZO, for text files only. Impala can query LZO-compressed Text tables, but currently cannot create them or insert data into them; perform these operations in Hive.

## Choosing the File Format for a Table

Different file formats and compression codecs work better for different data sets. While Impala typically provides performance gains regardless of file format, choosing the proper format for your data can yield further performance improvements. Use the following considerations to decide which combination of file format and compression to use for a particular table:

- If you are working with existing files that are already in a supported file format, use the same format for the Impala table where practical. If the original format does not yield acceptable query performance or resource usage, consider creating a new Impala table with different file format or compression characteristics, and doing a one-time conversion by copying the data to the new table using the `INSERT` statement. Depending on the file format, you might run the `INSERT` statement in `impala-shell` or in Hive.
- Text files are convenient to produce through many different tools, and are human-readable for ease of verification and debugging. Those characteristics are why text is the default format for an Impala `CREATE TABLE` statement. When performance and resource usage are the primary considerations, use one of the other file formats and consider using compression. A typical workflow might involve bringing data into an Impala table by copying CSV or TSV files into the appropriate data directory, and then using the `INSERT ... SELECT` syntax to copy the data into a table using a different, more compact file format.
- If your architecture involves storing data to be queried in memory, do not compress the data. There is no I/O savings since the data does not need to be moved from disk, but there is a CPU cost to decompress the data.

## Using Text Data Files with Impala Tables

Cloudera Impala supports using text files as the storage format for input and output. Text files are a convenient format to use for interchange with other applications or scripts that produce or read delimited text files, such as CSV or TSV with commas or tabs for delimiters.

Text files are also very flexible in their column definitions. For example, a text file could have more fields than the Impala table, and those extra fields are ignored during queries; or it could have fewer fields than the Impala table, and those missing fields are treated as `NULL` values in queries. You could have fields that were treated as numbers or timestamps in a table, then use `ALTER TABLE ... REPLACE COLUMNS` to switch them to strings, or the reverse.

**Table 5: Text Format Support in Impala**

File Type	Format	Compression Codecs	Impala Can CREATE?	Impala Can INSERT?
Text	Unstructured	LZO, gzip, bzip2, Snappy	Yes. For <code>CREATE TABLE</code> with no <code>STORED AS</code> clause, the default file format is uncompressed text, with values separated by ASCII <code>0x01</code> characters (typically represented as Ctrl-A).	Yes: <code>CREATE TABLE</code> , <code>INSERT</code> , <code>LOAD DATA</code> , and <code>query</code> . If LZO compression is used, you must create the table and load data in Hive. If other kinds of compression are used, you must

File Type	Format	Compression Codecs	Impala Can CREATE?	Impala Can INSERT?
				load data through <code>LOAD DATA</code> , Hive, or manually in HDFS.

## Query Performance for Impala Text Tables

Data stored in text format is relatively bulky, and not as efficient to query as binary formats such as Parquet. You typically use text tables with Impala if that is the format you receive the data and you do not have control over that process, or if you are a relatively new Hadoop user and not familiar with techniques to generate files in other formats. (Because the default format for `CREATE TABLE` is text, you might create your first Impala tables as text without giving performance much thought.) Either way, look for opportunities to use more efficient file formats for the tables used in your most performance-critical queries.

For frequently queried data, you might load the original text data files into one Impala table, then use an `INSERT` statement to transfer the data to another table that uses the Parquet file format; the data is converted automatically as it is stored in the destination table.

For more compact data, consider using LZO compression for the text files. LZO is the only compression codec that Impala supports for text data, because the “splittable” nature of LZO data files lets different nodes work on different parts of the same file in parallel. See [Using LZO-Compressed Text Files](#) on page 404 for details.

In Impala 2.0 and later, you can also use text data compressed in the gzip, bzip2, or Snappy formats. Because these compressed formats are not “splittable” in the way that LZO is, there is less opportunity for Impala to parallelize queries on them. Therefore, use these types of compressed data only for convenience if that is the format in which you receive the data. Prefer to use LZO compression for text data if you have the choice, or convert the data to Parquet using an `INSERT ... SELECT` statement to copy the original data into a Parquet table.

- **Note:**

Impala supports bzip files created by the `bzip2` command, but not bzip files with multiple streams created by the `pbzip2` command. Impala decodes only the data from the first part of such files, leading to incomplete results.

The maximum size that Impala can accommodate for an individual bzip file is 1 GB (after uncompression).

## Creating Text Tables

### To create a table using text data files:

If the exact format of the text data files (such as the delimiter character) is not significant, use the `CREATE TABLE` statement with no extra clauses at the end to create a text-format table. For example:

```
create table my_table(id int, s string, n int, t timestamp, b boolean);
```

The data files created by any `INSERT` statements will use the Ctrl-A character (hex 01) as a separator between each column value.

A common use case is to import existing text files into an Impala table. The syntax is more verbose; the significant part is the `FIELDS TERMINATED BY` clause, which must be preceded by the `ROW FORMAT DELIMITED` clause. The statement can end with a `STORED AS TEXTFILE` clause, but that clause is optional because text format tables are the default. For example:

```
create table csv(id int, s string, n int, t timestamp, b boolean)
  row format delimited
    fields terminated by ',';

create table tsv(id int, s string, n int, t timestamp, b boolean)
```

## How Impala Works with Hadoop File Formats

```
row format delimited  
fields terminated by '\t';  
  
create table pipe_separated(id int, s string, n int, t timestamp, b boolean)  
row format delimited  
fields terminated by '|'  
stored as textfile;
```

You can create tables with specific separator characters to import text files in familiar formats such as CSV, TSV, or pipe-separated. You can also use these tables to produce output data files, by copying data into them through the `INSERT ... SELECT` syntax and then extracting the data files from the Impala data directory.

In Impala 1.3.1 and higher, you can specify a delimiter character '`\0`' to use the ASCII 0 (`nul`) character for text tables:

```
create table nul_separated(id int, s string, n int, t timestamp, b boolean)  
row format delimited  
fields terminated by '\0'  
stored as textfile;
```

■ **Note:**

Do not surround string values with quotation marks in text data files that you construct. If you need to include the separator character inside a field value, for example to put a string value with a comma inside a CSV-format data file, specify an escape character on the `CREATE TABLE` statement with the `ESCAPED BY` clause, and insert that character immediately before any separator characters that need escaping.

Issue a `DESCRIBE FORMATTED table_name` statement to see the details of how each table is represented internally in Impala.

## Data Files for Text Tables

When Impala queries a table with data in text format, it consults all the data files in the data directory for that table, with some exceptions:

- Impala ignores any hidden files, that is, files whose names start with a dot or an underscore.
- Impala ignores files with extensions commonly used for temporary work files by Hadoop tools. Any files with extensions `.tmp` or `.copying` are not considered part of the Impala table. The suffix matching is case-insensitive, so for example Impala ignores both `.copying` and `.COPYING` suffixes.
- Impala uses suffixes to recognize when text data files are compressed text. For Impala to recognize the compressed text files, they must have the appropriate file extension corresponding to the compression codec, either `.gz`, `.bz2`, or `.snappy`. The extensions can be in uppercase or lowercase.
- Otherwise, the file names are not significant. When you put files into an HDFS directory through ETL jobs, or point Impala to an existing HDFS directory with the `CREATE EXTERNAL TABLE` statement, or move data files under external control with the `LOAD DATA` statement, Impala preserves the original filenames.

Filenames for data produced through Impala `INSERT` statements are given unique names to avoid filename conflicts.

An `INSERT ... SELECT` statement produces one data file from each node that processes the `SELECT` part of the statement. An `INSERT ... VALUES` statement produces a separate data file for each statement; because Impala is more efficient querying a small number of huge files than a large number of tiny files, the `INSERT ... VALUES` syntax is not recommended for loading a substantial volume of data. If you find yourself with a table that is inefficient due to too many small data files, reorganize the data into a few large files by doing `INSERT ... SELECT` to transfer the data to a new table.

### Special values within text data files:

- Impala recognizes the literal strings `inf` for infinity and `nan` for "Not a Number", for `FLOAT` and `DOUBLE` columns.
- Impala recognizes the literal string `\N` to represent `NULL`. When using Sqoop, specify the options `--null-non-string` and `--null-string` to ensure all `NULL` values are represented correctly in the Sqoop output files. By default, Sqoop writes `NULL` values using the string `null`, which causes a conversion error when such rows are evaluated by Impala. (A workaround for existing tables and data files is to change the table properties through `ALTER TABLE name SET TBLPROPERTIES("serialization.null.format"="null")`.)

## Loading Data into Impala Text Tables

To load an existing text file into an Impala text table, use the `LOAD DATA` statement and specify the path of the file in HDFS. That file is moved into the appropriate Impala data directory.

To load multiple existing text files into an Impala text table, use the `LOAD DATA` statement and specify the HDFS path of the directory containing the files. All non-hidden files are moved into the appropriate Impala data directory.

To convert data to text from any other file format supported by Impala, use a SQL statement such as:

```
-- Text table with default delimiter, the hex 01 character.
CREATE TABLE text_table AS SELECT * FROM other_file_format_table;

-- Text table with user-specified delimiter. Currently, you cannot specify
-- the delimiter as part of CREATE TABLE LIKE or CREATE TABLE AS SELECT.
-- But you can change an existing text table to have a different delimiter.
CREATE TABLE csv LIKE other_file_format_table;
ALTER TABLE csv SET SERDEPROPERTIES ('serialization.format'=',', 'field.delim'=',');
INSERT INTO csv SELECT * FROM other_file_format_table;
```

This can be a useful technique to see how Impala represents special values within a text-format data file. Use the `DESCRIBE FORMATTED` statement to see the HDFS directory where the data files are stored, then use Linux commands such as `hdfs dfs -ls hdfs_directory` and `hdfs dfs -cat hdfs_file` to display the contents of an Impala-created text file.

To create a few rows in a text table for test purposes, you can use the `INSERT ... VALUES` syntax:

```
INSERT INTO text_table VALUES ('string_literal',100,hex('hello world'));
```

- **Note:** Because Impala and the HDFS infrastructure are optimized for multi-megabyte files, avoid the `INSERT ... VALUES` notation when you are inserting many rows. Each `INSERT ... VALUES` statement produces a new tiny file, leading to fragmentation and reduced performance. When creating any substantial volume of new data, use one of the bulk loading techniques such as `LOAD DATA` or `INSERT ... SELECT`. Or, [use an HBase table](#) for single-row `INSERT` operations, because HBase tables are not subject to the same fragmentation issues as tables stored on HDFS.

When you create a text file for use with an Impala text table, specify `\N` to represent a `NULL` value. For the differences between `NULL` and empty strings, see [NULL](#) on page 141.

If a text file has fewer fields than the columns in the corresponding Impala table, all the corresponding columns are set to `NULL` when the data in that file is read by an Impala query.

If a text file has more fields than the columns in the corresponding Impala table, the extra fields are ignored when the data in that file is read by an Impala query.

You can also use manual HDFS operations such as `hdfs dfs -put` or `hdfs dfs -cp` to put data files in the data directory for an Impala table. When you copy or move new data files into the HDFS directory for the Impala table, issue a `REFRESH table_name` statement in `impala-shell` before issuing the next query against that table, to make Impala recognize the newly added files.

## How Impala Works with Hadoop File Formats

### Using LZO-Compressed Text Files

Cloudera Impala supports using text data files that employ LZO compression. Cloudera recommends compressing text data files when practical. Impala queries are usually I/O-bound; reducing the amount of data read from disk typically speeds up a query, despite the extra CPU work to uncompress the data in memory.

Impala can work with LZO-compressed text files are preferable to files compressed by other codecs, because LZO-compressed files are “splittable”, meaning that different portions of a file can be uncompressed and processed independently by different nodes.

Impala does not currently support writing LZO-compressed text files.

Because Impala can query LZO-compressed files but currently cannot write them, you use Hive to do the initial `CREATE TABLE` and load the data, then switch back to Impala to run queries. For instructions on setting up LZO compression for Hive `CREATE TABLE` and `INSERT` statements, see [the LZO page on the Hive wiki](#). Once you have created an LZO text table, you can also manually add LZO-compressed text files to it, produced by the `lzoop` command or similar method.

#### Preparing to Use LZO-Compressed Text Files

Before using LZO-compressed tables in Impala, do the following one-time setup for each machine in the cluster. Install the necessary packages using either the Cloudera public repository, a private repository you establish, or by using packages. You must do these steps manually, whether or not the cluster is managed by the Cloudera Manager product.

##### 1. Prepare your systems to work with LZO using Cloudera repositories:

###### On systems managed by Cloudera Manager, using parcels:

See the setup instructions for the LZO parcel in the Cloudera Manager documentation for [Cloudera Manager 5](#).

###### On systems managed by Cloudera Manager, using packages, or not managed by Cloudera Manager:

Download and install the appropriate file to each machine on which you intend to use LZO with Impala. These files all come from the Cloudera [GPL extras](#) download site. Install the:

- [Red Hat 5 repo file](#) to `/etc/yum.repos.d/`.
- [Red Hat 6 repo file](#) to `/etc/yum.repos.d/`.
- [SUSE repo file](#) to `/etc/zypp/repos.d/`.
- [Ubuntu 10.04 list file](#) to `/etc/apt/sources.list.d/`.
- [Ubuntu 12.04 list file](#) to `/etc/apt/sources.list.d/`.
- [Debian list file](#) to `/etc/apt/sources.list.d/`.

##### 2. Configure Impala to use LZO:

Use **one** of the following sets of commands to refresh your package management system's repository information, install the base LZO support for Hadoop, and install the LZO support for Impala.

- **Note:** The name of the Hadoop LZO package changes between CDH 4 and CDH 5. In CDH 4, the package name is `hadoop-lzo-cdh4`. In CDH 5, the package name is `hadoop-lzo`. Use the appropriate package name depending on the level of CDH in your cluster.

###### For RHEL/CentOS systems:

```
$ sudo yum update
$ sudo yum install hadoop-lzo-cdh4 # For clusters running CDH 4.
$ sudo yum install hadoop-lzo      # For clusters running CDH 5 or higher.
$ sudo yum install impala-lzo
```

**For SUSE systems:**

```
$ sudo apt-get update
$ sudo zypper install hadoop-lzo-cdh4 # For clusters running CDH 4.
$ sudo zypper install hadoop-lzo      # For clusters running CDH 5 or higher.
$ sudo zypper install impala-lzo
```

**For Debian/Ubuntu systems:**

```
$ sudo zypper update
$ sudo apt-get install hadoop-lzo-cdh4 # For clusters running CDH 4.
$ sudo apt-get install hadoop-lzo      # For clusters running CDH 5 or higher.
$ sudo apt-get install impala-lzo
```

**Note:**

The level of the `impala-lzo-cdh4` package is closely tied to the version of Impala you use. Any time you upgrade Impala, re-do the installation command for `impala-lzo` on each applicable machine to make sure you have the appropriate version of that package.

3. For `core-site.xml` on the client **and** server (that is, in the configuration directories for both Impala and Hadoop), append `com.hadoop.compression.lzo.LzopCodec` to the comma-separated list of codecs. For example:

```
<property>
  <name>io.compression.codecs</name>
  <value>org.apache.hadoop.io.compress.DefaultCodec,org.apache.hadoop.io.compress.GzipCodec,
        org.apache.hadoop.io.compress.BZip2Codec,org.apache.hadoop.io.compress.DeflateCodec,
        org.apache.hadoop.io.compress.SnappyCodec,com.hadoop.compression.lzo.LzopCodec</value>
</property>
```

**Note:**

If this is the first time you have edited the Hadoop `core-site.xml` file, note that the `/etc/hadoop/conf` directory is typically a symbolic link, so the canonical `core-site.xml` might reside in a different directory:

```
$ ls -l /etc/hadoop
total 8
lrwxrwxrwx. 1 root root 29 Feb 26 2013 conf ->
/etc/alternatives/hadoop-conf
lrwxrwxrwx. 1 root root 10 Feb 26 2013 conf.dist -> conf.empty
drwxr-xr-x. 2 root root 4096 Feb 26 2013 conf.empty
drwxr-xr-x. 2 root root 4096 Oct 28 15:46 conf.pseudo
```

If the `io.compression.codecs` property is missing from `core-site.xml`, only add `com.hadoop.compression.lzo.LzopCodec` to the new property value, not all the names from the preceding example.

4. Restart the MapReduce and Impala services.

## How Impala Works with Hadoop File Formats

### Creating LZO Compressed Text Tables

A table containing LZO-compressed text files must be created in Hive with the following storage clause:

```
STORED AS  
  INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  
  OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat'
```

Also, certain Hive settings need to be in effect. For example:

```
hive> SET mapreduce.output.fileoutputformat.compress=true;  
hive> SET hive.exec.compress.output=true;  
hive> SET  
mapreduce.output.fileoutputformat.compress.codec=com.hadoop.compression.lzo.LzopCodec;  
hive> CREATE TABLE lzo_t (s string) STORED AS  
  > INPUTFORMAT 'com.hadoop.mapred.DeprecatedLzoTextInputFormat'  
  > OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.HiveIgnoreKeyTextOutputFormat';  
hive> INSERT INTO TABLE lzo_t SELECT col1, col2 FROM uncompressed_text_table;
```

Once you have created LZO-compressed text tables, you can convert data stored in other tables (regardless of file format) by using the `INSERT ... SELECT` statement in Hive.

Files in an LZO-compressed table must use the `.lzo` extension. Examine the files in the HDFS data directory after doing the `INSERT` in Hive, to make sure the files have the right extension. If the required settings are not in place, you end up with regular uncompressed files, and Impala cannot access the table because it finds data files with the wrong (uncompressed) format.

After loading data into an LZO-compressed text table, index the files so that they can be split. You index the files by running a Java class, `com.hadoop.compression.lzo.DistributedLzoIndexer`, through the Linux command line. This Java class is included in the `hadoop-lzo` package.

Run the indexer using a command like the following:

```
$ hadoop jar /usr/lib/hadoop/lib/hadoop-lzo-cdh4-0.4.15-gplextras.jar  
com.hadoop.compression.lzo.DistributedLzoIndexer /hdfs_location_of_table/
```

- **Note:** If the path of the JAR file in the preceding example is not recognized, do a `find` command to locate `hadoop-lzo-*-gplextras.jar` and use that path.

Indexed files have the same name as the file they index, with the `.index` extension. If the data files are not indexed, Impala queries still work, but the queries read the data from remote DataNodes, which is very inefficient.

Once the LZO-compressed tables are created, and data is loaded and indexed, you can query them through Impala. As always, the first time you start `impala-shell` after creating a table in Hive, issue an `INVALIDATE METADATA` statement so that Impala recognizes the new table. (In Impala 1.2 and higher, you only have to run `INVALIDATE METADATA` on one node, rather than on all the Impala nodes.)

### Using gzip, bzip2, or Snappy-Compressed Text Files

In Impala 2.0 and later, Impala supports using text data files that employ gzip, bzip2, or Snappy compression. These compression types are primarily for convenience within an existing ETL pipeline rather than maximum performance. Although it requires less I/O to read compressed text than the equivalent uncompressed text, files compressed by these codecs are not “splittable” and therefore cannot take full advantage of the Impala parallel query capability.

As each bzip2- or Snappy-compressed text file is processed, the node doing the work reads the entire file into memory and then decompresses it. Therefore, the node must have enough memory to hold both the compressed and uncompressed data from the text file. The memory required to hold the uncompressed data is difficult to estimate in advance, potentially causing problems on systems with low memory limits or with resource management enabled. In Impala 2.1 and higher, this memory overhead is reduced for gzip-compressed text files. The gzipped data is decompressed as it is read, rather than all at once.

To create a table to hold gzip, bzip2, or Snappy-compressed text, create a text table with no special compression options. Specify the delimiter and escape character if required, using the `ROW FORMAT` clause.

Because Impala can query compressed text files but currently cannot write them, produce the compressed text files outside Impala and use the `LOAD DATA` statement, manual HDFS commands to move them to the appropriate Impala data directory. (Or, you can use `CREATE EXTERNAL TABLE` and point the `LOCATION` attribute at a directory containing existing compressed text files.)

For Impala to recognize the compressed text files, they must have the appropriate file extension corresponding to the compression codec, either `.gz`, `.bz2`, or `.snappy`. The extensions can be in uppercase or lowercase.

The following example shows how you can create a regular text table, put different kinds of compressed and uncompressed files into it, and Impala automatically recognizes and decompresses each one based on their file extensions:

```
create table csv_compressed (a string, b string, c string)
  row format delimited fields terminated by ",";

insert into csv_compressed values
  ('one - uncompressed', 'two - uncompressed', 'three - uncompressed'),
  ('abc - uncompressed', 'xyz - uncompressed', '123 - uncompressed');
...make equivalent .gz, .bz2, and .snappy files and load them into same table
directory...

select * from csv_compressed;
+-----+-----+-----+
| a    | b    | c    |
+-----+-----+-----+
| one - snappy | two - snappy | three - snappy |
| one - uncompressed | two - uncompressed | three - uncompressed |
| abc - uncompressed | xyz - uncompressed | 123 - uncompressed |
| one - bz2 | two - bz2 | three - bz2 |
| abc - bz2 | xyz - bz2 | 123 - bz2 |
| one - gzip | two - gzip | three - gzip |
| abc - gzip | xyz - gzip | 123 - gzip |
+-----+-----+-----+

$ hdfs dfs -ls
'hdfs://127.0.0.1:8020/user/hive/warehouse/file_formats.db/csv_compressed/';
...truncated for readability...
75
hdfs://127.0.0.1:8020/user/hive/warehouse/file_formats.db/csv_compressed/csv_compressed.snappy
79
hdfs://127.0.0.1:8020/user/hive/warehouse/file_formats.db/csv_compressed/csv_compressed_bz2.csv.bz2
80
hdfs://127.0.0.1:8020/user/hive/warehouse/file_formats.db/csv_compressed/csv_compressed_gzip.csv.gz
116
hdfs://127.0.0.1:8020/user/hive/warehouse/file_formats.db/csv_compressed/dd414df64d67d49b_data.0.
```

## Using the Parquet File Format with Impala Tables

Impala helps you to create, manage, and query Parquet tables. Parquet is a column-oriented binary file format intended to be highly efficient for the types of large-scale queries that Impala is best at. Parquet is especially good for queries scanning particular columns within a table, for example to query "wide" tables with many columns, or to perform aggregation operations such as `SUM()` and `AVG()` that need to process most or all of the values from a column. Each data file contains the values for a set of rows (the "row group"). Within a data file, the values from each column are organized so that they are all adjacent, enabling good compression for the values from that column. Queries against a Parquet table can retrieve and analyze these values from any column quickly and with minimal I/O.

## How Impala Works with Hadoop File Formats

**Table 6: Parquet Format Support in Impala**

File Type	Format	Compression Codecs	Impala Can CREATE?	Impala Can INSERT?
Parquet	Structured	Snappy, gzip; currently Snappy by default	Yes.	Yes: CREATE TABLE, INSERT, LOAD DATA, and query.

### Creating Parquet Tables in Impala

To create a table named PARQUET\_TABLE that uses the Parquet format, you would use a command like the following, substituting your own table name, column names, and data types:

```
[impala-host:21000] > create table parquet_table_name (x INT, y STRING) STORED AS PARQUET;
```

Or, to clone the column names and data types of an existing table:

```
[impala-host:21000] > create table parquet_table_name LIKE other_table_name STORED AS PARQUET;
```

In Impala 1.4.0 and higher, you can derive column definitions from a raw Parquet data file, even without an existing Impala table. For example, you can create an external table pointing to an HDFS directory, and base the column definitions on one of the files in that directory:

```
CREATE EXTERNAL TABLE ingest_existing_files LIKE PARQUET  
'/user/etl/destination/datafile1.dat'  
STORED AS PARQUET  
LOCATION '/user/etl/destination';
```

Or, you can refer to an existing data file and create a new empty table with suitable column definitions. Then you can use `INSERT` to create new data files or `LOAD DATA` to transfer existing data files into the new table.

```
CREATE TABLE columns_from_data_file LIKE PARQUET '/user/etl/destination/datafile1.dat'  
STORED AS PARQUET;
```

The default properties of the newly created table are the same as for any other `CREATE TABLE` statement. For example, the default file format is text; if you want the new table to use the Parquet file format, include the `STORED AS PARQUET` file also.

In this example, the new table is partitioned by year, month, and day. These partition key columns are not part of the data file, so you specify them in the `CREATE TABLE` statement:

```
CREATE TABLE columns_from_data_file LIKE PARQUET '/user/etl/destination/datafile1.dat'  
PARTITION (year INT, month TINYINT, day TINYINT)  
STORED AS PARQUET;
```

See [CREATE TABLE Statement](#) on page 180 for more details about the `CREATE TABLE LIKE PARQUET` syntax.

Once you have created a table, to insert data into that table, use a command similar to the following, again with your own table names:

```
[impala-host:21000] > insert overwrite table parquet_table_name select * from other_table_name;
```

If the Parquet table has a different number of columns or different column names than the other table, specify the names of columns from the other table rather than `*` in the `SELECT` statement.

## Loading Data into Parquet Tables

Choose from the following techniques for loading data into Parquet tables, depending on whether the original data is already in an Impala table, or exists as raw data files outside Impala.

If you already have data in an Impala or Hive table, perhaps in a different file format or partitioning scheme, you can transfer the data to a Parquet table using the Impala `INSERT . . . SELECT` syntax. You can convert, filter, repartition, and do other things to the data as part of this same `INSERT` statement. See [Snappy and GZip Compression for Parquet Data Files](#) on page 411 for some examples showing how to insert data into Parquet tables.

When inserting into partitioned tables, especially using the Parquet file format, you can include a hint in the `INSERT` statement to fine-tune the overall performance of the operation and its resource usage:

- These hints are available in Impala 1.2.2 and higher.
- You would only use these hints if an `INSERT` into a partitioned Parquet table was failing due to capacity limits, or if such an `INSERT` was succeeding but with less-than-optimal performance.
- To use these hints, put the hint keyword `[SHUFFLE]` or `[NOSHUFFLE]` (including the square brackets) after the `PARTITION` clause, immediately before the `SELECT` keyword.
- `[SHUFFLE]` selects an execution plan that minimizes the number of files being written simultaneously to HDFS, and the number of memory buffers holding data for individual partitions. Thus it reduces overall resource usage for the `INSERT` operation, allowing some `INSERT` operations to succeed that otherwise would fail. It does involve some data transfer between the nodes so that the data files for a particular partition are all constructed on the same node.
- `[NOSHUFFLE]` selects an execution plan that might be faster overall, but might also produce a larger number of small data files or exceed capacity limits, causing the `INSERT` operation to fail. Use `[SHUFFLE]` in cases where an `INSERT` statement fails or runs inefficiently due to all nodes attempting to construct data for all partitions.
- Impala automatically uses the `[SHUFFLE]` method if any partition key column in the source table, mentioned in the `INSERT . . . SELECT` query, does not have column statistics. In this case, only the `[NOSHUFFLE]` hint would have any effect.
- If column statistics are available for all partition key columns in the source table mentioned in the `INSERT . . . SELECT` query, Impala chooses whether to use the `[SHUFFLE]` or `[NOSHUFFLE]` technique based on the estimated number of distinct values in those columns and the number of nodes involved in the `INSERT` operation. In this case, you might need the `[SHUFFLE]` or the `[NOSHUFFLE]` hint to override the execution plan selected by Impala.

Any `INSERT` statement for a Parquet table requires enough free space in the HDFS filesystem to write one block. Because Parquet data files use a block size of 1 GB by default, an `INSERT` might fail (even for a very small amount of data) if your HDFS is running low on space.

Avoid the `INSERT . . . VALUES` syntax for Parquet tables, because `INSERT . . . VALUES` produces a separate tiny data file for each `INSERT . . . VALUES` statement, and the strength of Parquet is in its handling of data (compressing, parallelizing, and so on) in large chunks.

If you have one or more Parquet data files produced outside of Impala, you can quickly make the data queryable through Impala by one of the following methods:

- The `LOAD DATA` statement moves a single data file or a directory full of data files into the data directory for an Impala table. It does no validation or conversion of the data. The original data files must be somewhere in HDFS, not the local filesystem.
- The `CREATE TABLE` statement with the `LOCATION` clause creates a table where the data continues to reside outside the Impala data directory. The original data files must be somewhere in HDFS, not the local filesystem. For extra safety, if the data is intended to be long-lived and reused by other applications, you can use the `CREATE EXTERNAL TABLE` syntax so that the data files are not deleted by an Impala `DROP TABLE` statement.
- If the Parquet table already exists, you can copy Parquet data files directly into it, then use the `REFRESH` statement to make Impala recognize the newly added data. Remember to preserve the block size of the Parquet data files by using the `hadoop distcp -pb` command rather than a `-put` or `-cp` operation on the Parquet files. See [Example of Copying Parquet Data Files](#) on page 413 for an example of this kind of operation.

## How Impala Works with Hadoop File Formats

- **Note:**

Currently, Impala always decodes the column data in Parquet files based on the ordinal position of the columns, not by looking up the position of each column based on its name. Parquet files produced outside of Impala must write column data in the same order as the columns are declared in the Impala table. Any optional columns that are omitted from the data files must be the rightmost columns in the Impala table definition.

If you created compressed Parquet files through some tool other than Impala, make sure that any compression codecs are supported in Parquet by Impala. For example, Impala does not currently support LZO compression in Parquet files. Also doublecheck that you used any recommended compatibility settings in the other tool, such as `spark.sql.parquet.binaryAsString` when writing Parquet files through Spark.

If the data exists outside Impala and is in some other format, combine both of the preceding techniques. First, use a `LOAD DATA OR CREATE EXTERNAL TABLE ... LOCATION` statement to bring the data into an Impala table that uses the appropriate file format. Then, use an `INSERT ... SELECT` statement to copy the data to the Parquet table, converting to Parquet format as part of the process.

Loading data into Parquet tables is a memory-intensive operation, because the incoming data is buffered until it reaches one data block in size, then that chunk of data is organized and compressed in memory before being written out. The memory consumption can be larger when inserting data into partitioned Parquet tables, because a separate data file is written for each combination of partition key column values, potentially requiring several large chunks to be manipulated in memory at once.

When inserting into a partitioned Parquet table, Impala redistributes the data among the nodes to reduce memory consumption. You might still need to temporarily increase the memory dedicated to Impala during the insert operation, or break up the load operation into several `INSERT` statements, or both.

- **Note:** All the preceding techniques assume that the data you are loading matches the structure of the destination table, including column order, column names, and partition layout. To transform or reorganize the data, start by loading the data into a Parquet table that matches the underlying structure of the data, then use one of the table-copying techniques such as `CREATE TABLE AS SELECT` or `INSERT ... SELECT` to reorder or rename columns, divide the data among multiple partitions, and so on. For example to take a single comprehensive Parquet data file and load it into a partitioned table, you would use an `INSERT ... SELECT` statement with dynamic partitioning to let Impala create separate data files with the appropriate partition values; for an example, see [INSERT Statement](#) on page 202.

## Query Performance for Impala Parquet Tables

Query performance for Parquet tables depends on the number of columns needed to process the `SELECT` list and `WHERE` clauses of the query, the way data is divided into large data files with block size equal to file size, the reduction in I/O by reading the data for each column in compressed format, which data files can be skipped (for partitioned tables), and the CPU overhead of decompressing the data for each column.

For example, the following is an efficient query for a Parquet table:

```
select avg(income) from census_data where state = 'CA';
```

The query processes only 2 columns out of a large number of total columns. If the table is partitioned by the `STATE` column, it is even more efficient because the query only has to read and decode 1 column from each data file, and it can read only the data files in the partition directory for the state '`CA`', skipping the data files for all the other states, which will be physically located in other directories.

The following is a relatively inefficient query for a Parquet table:

```
select * from census_data;
```

Impala would have to read the entire contents of each large data file, and decompress the contents of each column for each row group, negating the I/O optimizations of the column-oriented format. This query might still be faster for a Parquet table than a table with some other file format, but it does not take advantage of the unique strengths of Parquet data files.

Impala can optimize queries on Parquet tables, especially join queries, better when statistics are available for all the tables. Issue the `COMPUTE STATS` statement for each table after substantial amounts of data are loaded into or appended to it. See [COMPUTE STATS Statement](#) on page 170 for details.

- **Note:** Currently, a known issue ([IMPALA-488](#)) could cause excessive memory usage during a `COMPUTE STATS` operation on a Parquet table. As a workaround, issue the command `SET NUM_SCANNER_THREADS=2` in `impala-shell` before issuing the `COMPUTE STATS` statement. Then issue `UNSET NUM_SCANNER_THREADS` before continuing with queries.

## Partitioning for Parquet Tables

As explained in [Partitioning for Impala Tables](#) on page 394, partitioning is an important performance technique for Impala generally. This section explains some of the performance considerations for partitioned Parquet tables.

The Parquet file format is ideal for tables containing many columns, where most queries only refer to a small subset of the columns. As explained in [How Parquet Data Files Are Organized](#) on page 414, the physical layout of Parquet data files lets Impala read only a small fraction of the data for many queries. The performance benefits of this approach are amplified when you use Parquet tables in combination with partitioning. Impala can skip the data files for certain partitions entirely, based on the comparisons in the `WHERE` clause that refer to the partition key columns. For example, queries on partitioned tables often analyze data for time intervals based on columns such as `YEAR`, `MONTH`, and/or `DAY`, or for geographic regions. Remember that Parquet data files use a large block size, so when deciding how finely to partition the data, try to find a granularity where each partition contains 256 MB or more of data, rather than creating a large number of smaller files split among many partitions.

Inserting into a partitioned Parquet table can be a resource-intensive operation, because each Impala node could potentially be writing a separate data file to HDFS for each combination of different values for the partition key columns. The large number of simultaneous open files could exceed the HDFS “transceivers” limit. To avoid exceeding this limit, consider the following techniques:

- Load different subsets of data using separate `INSERT` statements with specific values for the `PARTITION` clause, such as `PARTITION (year=2010)`.
- Increase the “transceivers” value for HDFS, sometimes spelled “xcievers” (sic). The property value in the `hdfs-site.xml` configuration file is `dfs.datanode.max.transfer.threads`. For example, if you were loading 12 years of data partitioned by year, month, and day, even a value of 4096 might not be high enough. This [blog post](#) explores the considerations for setting this value higher or lower, using HBase examples for illustration.
- Use the `COMPUTE STATS` statement to collect [column statistics](#) on the source table from which data is being copied, so that the Impala query can estimate the number of different values in the partition key columns and distribute the work accordingly.

- **Note:** Currently, a known issue ([IMPALA-488](#)) could cause excessive memory usage during a `COMPUTE STATS` operation on a Parquet table. As a workaround, issue the command `SET NUM_SCANNER_THREADS=2` in `impala-shell` before issuing the `COMPUTE STATS` statement. Then issue `UNSET NUM_SCANNER_THREADS` before continuing with queries.

## Snappy and GZip Compression for Parquet Data Files

When Impala writes Parquet data files using the `INSERT` statement, the underlying compression is controlled by the `COMPRESSION_CODEC` query option. (Prior to Impala 2.0, the query option name was `PARQUET_COMPRESSION_CODEC`.) The allowed values for this query option are `snappy` (the default), `gzip`, and

## How Impala Works with Hadoop File Formats

none. The option value is not case-sensitive. If the option is set to an unrecognized value, all kinds of queries will fail due to the invalid option setting, not just queries involving Parquet tables.

### Example of Parquet Table with Snappy Compression

By default, the underlying data files for a Parquet table are compressed with Snappy. The combination of fast compression and decompression makes it a good choice for many data sets. To ensure Snappy compression is used, for example after experimenting with other compression codecs, set the COMPRESSION\_CODEC query option to snappy before inserting the data:

```
[localhost:21000] > create database parquet_compression;
[localhost:21000] > use parquet_compression;
[localhost:21000] > create table parquet_snappy like raw_text_data;
[localhost:21000] > set COMPRESSION_CODEC=snappy;
[localhost:21000] > insert into parquet_snappy select * from raw_text_data;
Inserted 1000000000 rows in 181.98s
```

### Example of Parquet Table with GZip Compression

If you need more intensive compression (at the expense of more CPU cycles for uncompressing during queries), set the COMPRESSION\_CODEC query option to gzip before inserting the data:

```
[localhost:21000] > create table parquet_gzip like raw_text_data;
[localhost:21000] > set COMPRESSION_CODEC=gzip;
[localhost:21000] > insert into parquet_gzip select * from raw_text_data;
Inserted 1000000000 rows in 1418.24s
```

### Example of Uncompressed Parquet Table

If your data compresses very poorly, or you want to avoid the CPU overhead of compression and decompression entirely, set the COMPRESSION\_CODEC query option to none before inserting the data:

```
[localhost:21000] > create table parquet_none like raw_text_data;
[localhost:21000] > set COMPRESSION_CODEC=none;
[localhost:21000] > insert into parquet_none select * from raw_text_data;
Inserted 1000000000 rows in 146.90s
```

### Examples of Sizes and Speeds for Compressed Parquet Tables

Here are some examples showing differences in data sizes and query speeds for 1 billion rows of synthetic data, compressed with each kind of codec. As always, run similar tests with realistic data sets of your own. The actual compression ratios, and relative insert and query speeds, will vary depending on the characteristics of the actual data.

In this case, switching from Snappy to GZip compression shrinks the data by an additional 40% or so, while switching from Snappy compression to no compression expands the data also by about 40%:

```
$ hdfs dfs -du -h /user/hive/warehouse/parquet_compression.db
23.1 G  /user/hive/warehouse/parquet_compression.db/parquet_snappy
13.5 G  /user/hive/warehouse/parquet_compression.db/parquet_gzip
32.8 G  /user/hive/warehouse/parquet_compression.db/parquet_none
```

Because Parquet data files are typically large, each directory will have a different number of data files and the row groups will be arranged differently.

At the same time, the less aggressive the compression, the faster the data can be decompressed. In this case using a table with a billion rows, a query that evaluates all the values for a particular column runs faster with no compression than with Snappy compression, and faster with Snappy compression than with Gzip compression.

Query performance depends on several other factors, so as always, run your own benchmarks with your own data to determine the ideal tradeoff between data size, CPU efficiency, and speed of insert and query operations.

```
[localhost:21000] > desc parquet_snappy;
Query finished, fetching results ...
+-----+-----+-----+
| name | type | comment |
+-----+-----+-----+
| id | int |
| val | int |
| zfill | string |
| name | string |
| assertion | boolean |
+-----+-----+-----+
Returned 5 row(s) in 0.14s
[localhost:21000] > select avg(val) from parquet_snappy;
Query finished, fetching results ...
+-----+
| _c0 |
+-----+
| 250000.93577915 |
+-----+
Returned 1 row(s) in 4.29s
[localhost:21000] > select avg(val) from parquet_gzip;
Query finished, fetching results ...
+-----+
| _c0 |
+-----+
| 250000.93577915 |
+-----+
Returned 1 row(s) in 6.97s
[localhost:21000] > select avg(val) from parquet_none;
Query finished, fetching results ...
+-----+
| _c0 |
+-----+
| 250000.93577915 |
+-----+
Returned 1 row(s) in 3.67s
```

## Example of Copying Parquet Data Files

Here is a final example, to illustrate how the data files using the various compression codecs are all compatible with each other for read operations. The metadata about the compression format is written into each data file, and can be decoded during queries regardless of the `COMPRESSION_CODEC` setting in effect at the time. In this example, we copy data files from the `PARQUET_SNAPPY`, `PARQUET_GZIP`, and `PARQUET_NONE` tables used in the previous examples, each containing 1 billion rows, all to the data directory of a new table `PARQUET_EVERYTHING`. A couple of sample queries demonstrate that the new table now contains 3 billion rows featuring a variety of compression codecs for the data files.

First, we create the table in Impala so that there is a destination directory in HDFS to put the data files:

```
[localhost:21000] > create table parquet_everything like parquet_snappy;
Query: create table parquet_everything like parquet_snappy
```

Then in the shell, we copy the relevant data files into the data directory for this new table. Rather than using `hdfs dfs -cp` as with typical files, we use `hadoop distcp -pb` to ensure that the special block size of the Parquet data files is preserved.

```
$ hadoop distcp -pb /user/hive/warehouse/parquet_compression.db/parquet_snappy \
  /user/hive/warehouse/parquet_compression.db/parquet_everything
...MapReduce output...
$ hadoop distcp -pb /user/hive/warehouse/parquet_compression.db/parquet_gzip \
  /user/hive/warehouse/parquet_compression.db/parquet_everything
...MapReduce output...
$ hadoop distcp -pb /user/hive/warehouse/parquet_compression.db/parquet_none \
  /user/hive/warehouse/parquet_compression.db/parquet_everything
...MapReduce output...
```

## How Impala Works with Hadoop File Formats

Back in the `impala-shell` interpreter, we use the `REFRESH` statement to alert the Impala server to the new data files for this table, then we can run queries demonstrating that the data files represent 3 billion rows, and the values for one of the numeric columns match what was in the original smaller tables:

```
[localhost:21000] > refresh parquet_everything;
Query finished, fetching results ...

Returned 0 row(s) in 0.32s
[localhost:21000] > select count(*) from parquet_everything;
Query finished, fetching results ...
+-----+
| _c0   |
+-----+
| 3000000000 |
+-----+
Returned 1 row(s) in 8.18s
[localhost:21000] > select avg(val) from parquet_everything;
Query finished, fetching results ...
+-----+
| _c0   |
+-----+
| 250000.93577915 |
+-----+
Returned 1 row(s) in 13.35s
```

## Exchanging Parquet Data Files with Other Hadoop Components

Starting in CDH 4.5, you can read and write Parquet data files from Hive, Pig, and MapReduce. See the [CDH 4 Installation Guide](#) for details.

Previously, it was not possible to create Parquet data through Impala and reuse that table within Hive. Now that Parquet support is available for Hive in CDH 4.5, reusing existing Impala Parquet data files in Hive requires updating the table metadata. Use the following command if you are already running Impala 1.1.1 or higher:

```
ALTER TABLE table_name SET FILEFORMAT PARQUET;
```

If you are running a level of Impala that is older than 1.1.1, do the metadata update through Hive:

```
ALTER TABLE table_name SET SERDE 'parquet.hive.serde.ParquetHiveSerDe';
ALTER TABLE table_name SET FILEFORMAT
  INPUTFORMAT "parquet.hive.DeprecatedParquetInputFormat"
  OUTPUTFORMAT "parquet.hive.DeprecatedParquetOutputFormat";
```

Impala 1.1.1 and higher can reuse Parquet data files created by Hive, without any action required.

Impala supports the scalar data types that you can encode in a Parquet data file, but not composite or nested types such as maps or arrays. In Impala 2.2.0 / CDH 5.4.0 and higher, Impala can query Parquet data files that include composite or nested types, as long as the query only refers to columns with scalar types.

If you copy Parquet data files between nodes, or even between different directories on the same node, make sure to preserve the block size by using the command `hadoop distcp -pb`. To verify that the block size was preserved, issue the command `hdfs fsck -blocks HDFS_path_of_impala_table_dir` and check that the average block size is at or near 256 MB (or whatever other size is defined by the `PARQUET_FILE_SIZE` query option).. (The `hadoop distcp` operation typically leaves some directories behind, with names matching `_distcp_logs_*`, that you can delete from the destination directory afterward.) Issue the command `hadoop distcp` for details about `distcp` command syntax.

## How Parquet Data Files Are Organized

Although Parquet is a column-oriented file format, do not expect to find one data file for each column. Parquet keeps all the data for a row within the same data file, to ensure that the columns for a row are always available on the same node for processing. What Parquet does is to set a large HDFS block size and a matching maximum data file size, to ensure that I/O and network transfer requests apply to large batches of data.

Within that data file, the data for a set of rows is rearranged so that all the values from the first column are organized in one contiguous block, then all the values from the second column, and so on. Putting the values from the same column next to each other lets Impala use effective compression techniques on the values in that column.

- **Note:**

Impala `INSERT` statements write Parquet data files using an HDFS block size that matches the data file size, to ensure that each data file is represented by a single HDFS block, and the entire file can be processed on a single node without requiring any remote reads.

If you create Parquet data files outside of Impala, such as through a MapReduce or Pig job, ensure that the HDFS block size is greater than or equal to the file size, so that the “one file per block” relationship is maintained. Set the `dfs.block.size` or the `dfs.blocksize` property large enough that each file fits within a single HDFS block, even if that size is larger than the normal HDFS block size.

If the block size is reset to a lower value during a file copy, you will see lower performance for queries involving those files, and the `PROFILE` statement will reveal that some I/O is being done suboptimally, through remote reads. See [Example of Copying Parquet Data Files](#) on page 413 for an example showing how to preserve the block size when copying Parquet data files.

When Impala retrieves or tests the data for a particular column, it opens all the data files, but only reads the portion of each file containing the values for that column. The column values are stored consecutively, minimizing the I/O required to process the values within a single column. If other columns are named in the `SELECT` list or `WHERE` clauses, the data for all columns in the same row is available within that same data file.

If an `INSERT` statement brings in less than one Parquet block’s worth of data, the resulting data file is smaller than ideal. Thus, if you do split up an ETL job to use multiple `INSERT` statements, try to keep the volume of data for each `INSERT` statement to approximately 256 MB, or a multiple of 256 MB.

## RLE and Dictionary Encoding for Parquet Data Files

Parquet uses some automatic compression techniques, such as run-length encoding (RLE) and dictionary encoding, based on analysis of the actual data values. Once the data values are encoded in a compact form, the encoded data can optionally be further compressed using a compression algorithm. Parquet data files created by Impala can use Snappy, GZip, or no compression; the Parquet spec also allows LZO compression, but currently Impala does not support LZO-compressed Parquet files.

RLE and dictionary encoding are compression techniques that Impala applies automatically to groups of Parquet data values, in addition to any Snappy or GZip compression applied to the entire data files. These automatic optimizations can save you time and planning that are normally needed for a traditional data warehouse. For example, dictionary encoding reduces the need to create numeric IDs as abbreviations for longer string values.

Run-length encoding condenses sequences of repeated data values. For example, if many consecutive rows all contain the same value for a country code, those repeating values can be represented by the value followed by a count of how many times it appears consecutively.

Dictionary encoding takes the different values present in a column, and represents each one in compact 2-byte form rather than the original value, which could be several bytes. (Additional compression is applied to the compacted values, for extra space savings.) This type of encoding applies when the number of different values for a column is less than  $2^{**}16$  (16,384). It does not apply to columns of data type `BOOLEAN`, which are already very short. `TIMESTAMP` columns sometimes have a unique value for each row, in which case they can quickly exceed the  $2^{**}16$  limit on distinct values. The  $2^{**}16$  limit on different values within a column is reset for each data file, so if several different data files each contained 10,000 different city names, the city name column in each data file could still be condensed using dictionary encoding.

## How Impala Works with Hadoop File Formats

### Compacting Data Files for Parquet Tables

If you reuse existing table structures or ETL processes for Parquet tables, you might encounter a “many small files” situation, which is suboptimal for query efficiency. For example, statements like these might produce inefficiently organized data files:

```
-- In an N-node cluster, each node produces a data file
-- for the INSERT operation. If you have less than
-- N GB of data to copy, some files are likely to be
-- much smaller than the default Parquet block size.
insert into parquet_table select * from text_table;

-- Even if this operation involves an overall large amount of data,
-- when split up by year/month/day, each partition might only
-- receive a small amount of data. Then the data files for
-- the partition might be divided between the N nodes in the cluster.
-- A multi-gigabyte copy operation might produce files of only
-- a few MB each.
insert into partitioned_parquet_table partition (year, month, day)
  select year, month, day, url, referer, user_agent, http_code, response_time
  from web_stats;
```

Here are techniques to help you produce large data files in Parquet `INSERT` operations, and to compact existing too-small data files:

- When inserting into a partitioned Parquet table, use statically partitioned `INSERT` statements where the partition key values are specified as constant values. Ideally, use a separate `INSERT` statement for each partition.
- You might set the `NUM_NODES` option to 1 briefly, during `INSERT` or `CREATE TABLE AS SELECT` statements. Normally, those statements produce one or more data files per data node. If the write operation involves small amounts of data, a Parquet table, and/or a partitioned table, the default behavior could produce many small files when intuitively you might expect only a single output file. `SET NUM_NODES=1` turns off the “distributed” aspect of the write operation, making it more likely to produce only one or a few data files.
- Be prepared to reduce the number of partition key columns from what you are used to with traditional analytic database systems.
- Do not expect Impala-written Parquet files to fill up the entire Parquet block size. Impala estimates on the conservative side when figuring out how much data to write to each Parquet file. Typically, the of uncompressed data in memory is substantially reduced on disk by the compression and encoding techniques in the Parquet file format. The final data file size varies depending on the compressibility of the data. Therefore, it is not an indication of a problem if 256 MB of text data is turned into 2 Parquet data files, each less than 256 MB.
- If you accidentally end up with a table with many small data files, consider using one or more of the preceding techniques and copying all the data into a new Parquet table, either through `CREATE TABLE AS SELECT` or `INSERT ... SELECT` statements.

To avoid rewriting queries to change table names, you can adopt a convention of always running important queries against a view. Changing the view definition immediately switches any subsequent queries to use the new underlying tables:

```
create view production_table as select * from table_with_many_small_files;
-- CTAS or INSERT...SELECT all the data into a more efficient layout...
alter view production_table as select * from table_with_few_big_files;
select * from production_table where c1 = 100 and c2 < 50 and ...;
```

### Schema Evolution for Parquet Tables

Schema evolution refers to using the statement `ALTER TABLE ... REPLACE COLUMNS` to change the names, data type, or number of columns in a table. You can perform schema evolution for Parquet tables as follows:

- The Impala `ALTER TABLE` statement never changes any data files in the tables. From the Impala side, schema evolution involves interpreting the same data files in terms of a new table definition. Some types of schema changes make sense and are represented correctly. Other types of changes cannot be represented in a sensible way, and produce special result values or conversion errors during queries.
- The `INSERT` statement always creates data using the latest table definition. You might end up with data files with different numbers of columns or internal data representations if you do a sequence of `INSERT` and `ALTER TABLE ... REPLACE COLUMNS` statements.
- If you use `ALTER TABLE ... REPLACE COLUMNS` to define additional columns at the end, when the original data files are used in a query, these final columns are considered to be all `NULL` values.
- If you use `ALTER TABLE ... REPLACE COLUMNS` to define fewer columns than before, when the original data files are used in a query, the unused columns still present in the data file are ignored.
- Parquet represents the `TINYINT`, `SMALLINT`, and `INT` types the same internally, all stored in 32-bit integers.
  - That means it is easy to promote a `TINYINT` column to `SMALLINT` or `INT`, or a `SMALLINT` column to `INT`. The numbers are represented exactly the same in the data file, and the columns being promoted would not contain any out-of-range values.
  - If you change any of these column types to a smaller type, any values that are out-of-range for the new type are returned incorrectly, typically as negative numbers.
  - You cannot change a `TINYINT`, `SMALLINT`, or `INT` column to `BIGINT`, or the other way around. Although the `ALTER TABLE` succeeds, any attempt to query those columns results in conversion errors.
  - Any other type conversion for columns produces a conversion error during queries. For example, `INT` to `STRING`, `FLOAT` to `DOUBLE`, `TIMESTAMP` to `STRING`, `DECIMAL(9, 0)` to `DECIMAL(5, 2)`, and so on.

## Data Type Considerations for Parquet Tables

The Parquet format defines a set of data types whose names differ from the names of the corresponding Impala data types. If you are preparing Parquet files using other Hadoop components such as Pig or MapReduce, you might need to work with the type names defined by Parquet. The following figure lists the Parquet-defined types and the equivalent types in Impala.

### Primitive types:

BINARY -> STRING
BOOLEAN -> BOOLEAN
DOUBLE -> DOUBLE
FLOAT -> FLOAT
INT32 -> INT
INT64 -> BIGINT
INT96 -> TIMESTAMP

### Logical types:

BINARY + OriginalType UTF8 -> STRING
BINARY + OriginalType DECIMAL -> DECIMAL

## Using the Avro File Format with Impala Tables

Cloudera Impala supports using tables whose data files use the Avro file format. Impala can query Avro tables, and in Impala 1.4.0 and higher can create them, but currently cannot insert data into them. For insert operations, use Hive, then switch back to Impala to run queries.

## How Impala Works with Hadoop File Formats

**Table 7: Avro Format Support in Impala**

File Type	Format	Compression Codecs	Impala Can CREATE?	Impala Can INSERT?
Avro	Structured	Snappy, gzip, deflate, bzip2	Yes, in Impala 1.4.0 and higher. Before that, create the table using Hive.	No. Import data by using <code>LOAD DATA</code> on data files already in the right format, or use <code>INSERT</code> in Hive followed by <code>REFRESH table_name</code> in Impala.

### Creating Avro Tables

To create a new table using the Avro file format, issue the `CREATE TABLE` statement through Impala with the `STORED AS AVRO` clause, or through Hive. If you create the table through Impala, you must include column definitions that match the fields specified in the Avro schema. With Hive, you can omit the columns and just specify the Avro schema.

▪ **Note:**

Currently, Avro tables cannot contain `TIMESTAMP` columns. If you need to store date and time values in Avro tables, as a workaround you can use a `STRING` representation of the values, convert the values to `BIGINT` with the `UNIX_TIMESTAMP()` function, or create separate numeric columns for individual date and time fields using the `EXTRACT()` function.

The following examples demonstrate creating an Avro table in Impala, using either an inline column specification or one taken from a JSON file stored in HDFS:

```
[localhost:21000] > CREATE TABLE impala_avro_table
    > (bool_col BOOLEAN, int_col INT, long_col BIGINT, float_col FLOAT,
double_col DOUBLE, string_col STRING, nullable_int INT)
    > STORED AS AVRO
    > TBLPROPERTIES ('avro.schema.literal'='{
    >     "name": "my_record",
    >     "type": "record",
    >     "fields": [
    >         {"name": "bool_col", "type": "boolean"},
    >         {"name": "int_col", "type": "int"},
    >         {"name": "long_col", "type": "long"},
    >         {"name": "float_col", "type": "float"},
    >         {"name": "double_col", "type": "double"},
    >         {"name": "string_col", "type": "string"},
    >         {"name": "nullable_int", "type": ["null", "int"]}]}');
[localhost:21000] > CREATE TABLE avro_examples_of_all_types (
    >     id INT,
    >     bool_col BOOLEAN,
    >     tinyint_col TINYINT,
    >     smallint_col SMALLINT,
    >     int_col INT,
    >     bigint_col BIGINT,
    >     float_col FLOAT,
    >     double_col DOUBLE,
    >     date_string_col STRING,
    >     string_col STRING
    > )
    > STORED AS AVRO
    > TBLPROPERTIES
('avro.schema.url'='hdfs://localhost:8020/avro_schemas/alltypes.json');
```

The following example demonstrates creating an Avro table in Hive:

```
hive> CREATE TABLE hive_avro_table
    > ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
    > STORED AS INPUTFORMAT 'org.apache.hadoop.hive.ql.io.avro.AvroContainerInputFormat'
```

```
> OUTPUTFORMAT 'org.apache.hadoop.hive.io.avro.AvroContainerOutputFormat'
> TBLPROPERTIES ('avro.schema.literal'='{
>   "name": "my_record",
>   "type": "record",
>   "fields": [
>     {"name": "bool_col", "type": "boolean"},
>     {"name": "int_col", "type": "int"},
>     {"name": "long_col", "type": "long"},
>     {"name": "float_col", "type": "float"},
>     {"name": "double_col", "type": "double"},
>     {"name": "string_col", "type": "string"},
>     {"name": "nullable_int", "type": ["null", "int"]}]})');
```

Each field of the record becomes a column of the table. Note that any other information, such as the record name, is ignored.

- **Note:** For nullable Avro columns, make sure to put the "null" entry before the actual type name. In Impala, all columns are nullable; Impala currently does not have a NOT NULL clause. Any non-nullable property is only enforced on the Avro side.

Most column types map directly from Avro to Impala under the same names. These are the exceptions and special cases to consider:

- The DECIMAL type is defined in Avro as a BYTE type with the logicalType property set to "decimal" and a specified precision and scale. Use DECIMAL in Avro tables only under CDH 5. The infrastructure and components under CDH 4 do not have reliable DECIMAL support.
- The Avro long type maps to BIGINT in Impala.

If you create the table through Hive, switch back to `impala-shell` and issue an `INVALIDATE METADATA table_name` statement. Then you can run queries for that table through `impala-shell`.

## Using a Hive-Created Avro Table in Impala

If you have an Avro table created through Hive, you can use it in Impala as long as it contains only Impala-compatible data types. It cannot contain:

- Complex types: array, map, record, struct, union other than `[supported_type, null]` or `[null, supported_type]`
- The Avro-specific types enum, bytes, and fixed
- Any scalar type other than those listed in [Data Types](#) on page 111

Because Impala and Hive share the same metastore database, Impala can directly access the table definitions and data for tables that were created in Hive.

If you create an Avro table in Hive, issue an `INVALIDATE METADATA` the next time you connect to Impala through `impala-shell`. This is a one-time operation to make Impala aware of the new table. You can issue the statement while connected to any Impala node, and the catalog service broadcasts the change to all other Impala nodes.

If you load new data into an Avro table through Hive, either through a Hive `LOAD DATA` or `INSERT` statement, or by manually copying or moving files into the data directory for the table, issue a `REFRESH table_name` statement the next time you connect to Impala through `impala-shell`. You can issue the statement while connected to any Impala node, and the catalog service broadcasts the change to all other Impala nodes. If you issue the `LOAD DATA` statement through Impala, you do not need a `REFRESH` afterward.

Impala only supports fields of type boolean, int, long, float, double, and string, or unions of these types with null; for example, `["string", "null"]`. Unions with null essentially create a nullable type.

## Specifying the Avro Schema through JSON

While you can embed a schema directly in your `CREATE TABLE` statement, as shown above, column width restrictions in the Hive metastore limit the length of schema you can specify. If you encounter problems with

## How Impala Works with Hadoop File Formats

long schema literals, try storing your schema as a JSON file in HDFS instead. Specify your schema in HDFS using table properties similar to the following:

```
tblproperties ('avro.schema.url'='hdfs://your-name-node:port/path/to/schema.json');
```

### Loading Data into an Avro Table

If you already have data files in Avro format, you can also issue `LOAD DATA` in either Impala or Hive.

To copy data from another table, issue any `INSERT` statements through Hive. For information about loading data into Avro tables through Hive, see [Avro page on the Hive wiki](#).

### Enabling Compression for Avro Tables

To enable compression for Avro tables, specify settings in the Hive shell to enable compression and to specify a codec, then issue a `CREATE TABLE` statement as in the preceding examples. Impala supports the `snappy` and `deflate` codecs for Avro tables.

For example:

```
hive> set hive.exec.compress.output=true;
hive> set avro.output.codec=snappy;
```

### How Impala Handles Avro Schema Evolution

Starting in Impala 1.1, Impala can deal with Avro data files that employ ***schema evolution***, where different data files within the same table use slightly different type definitions. (You would perform the schema evolution operation by issuing an `ALTER TABLE` statement in the Hive shell.) The old and new types for any changed columns must be compatible, for example a column might start as an `int` and later change to a `bigint` or `float`.

As with any other tables where the definitions are changed or data is added outside of the current `impalad` node, ensure that Impala loads the latest metadata for the table if the Avro schema is modified through Hive. Issue a `REFRESH table_name` or `INVALIDATE METADATA table_name` statement. `REFRESH` reloads the metadata immediately, `INVALIDATE METADATA` reloads the metadata the next time the table is accessed.

When Avro data files or columns are not consulted during a query, Impala does not check for consistency. Thus, if you issue `SELECT c1, c2 FROM t1`, Impala does not return any error if the column `c3` changed in an incompatible way. If a query retrieves data from some partitions but not others, Impala does not check the data files for the unused partitions.

In the Hive DDL statements, you can specify an `avro.schema.literal` table property (if the schema definition is short) or an `avro.schema.url` property (if the schema definition is long, or to allow convenient editing for the definition).

For example, running the following SQL code in the Hive shell creates a table using the Avro file format and puts some sample data into it:

```
CREATE TABLE avro_table (a string, b string)
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.avro.AvroSerDe'
STORED AS INPUTFORMAT 'org.apache.hadoop.hive.ql.io.avro.AvroContainerInputFormat'
OUTPUTFORMAT 'org.apache.hadoop.hive.ql.io.avro.AvroContainerOutputFormat'
TBLPROPERTIES (
  'avro.schema.literal'='{
    "type": "record",
    "name": "my_record",
    "fields": [
      {"name": "a", "type": "int"},
      {"name": "b", "type": "string"}
    ]
  }';
  INSERT OVERWRITE TABLE avro_table SELECT 1, "avro" FROM functional.alltypes LIMIT 1;
```

Once the Avro table is created and contains data, you can query it through the `impala-shell` command:

```
[localhost:21000] > select * from avro_table;
+-----+
| a | b |
+-----+
| 1 | avro |
+-----+
```

Now in the Hive shell, you change the type of a column and add a new column with a default value:

```
-- Promote column "a" from INT to FLOAT (no need to update Avro schema)
ALTER TABLE avro_table CHANGE A A FLOAT;

-- Add column "c" with default
ALTER TABLE avro_table ADD COLUMNS (c int);
ALTER TABLE avro_table SET TBLPROPERTIES (
  'avro.schema.literal'='{
    "type": "record",
    "name": "my_record",
    "fields": [
      {"name": "a", "type": "int"},
      {"name": "b", "type": "string"},
      {"name": "c", "type": "int", "default": 10}
    ]
}');
```

Once again in `impala-shell`, you can query the Avro table based on its latest schema definition. Because the table metadata was changed outside of Impala, you issue a `REFRESH` statement first so that Impala has up-to-date metadata for the table.

```
[localhost:21000] > refresh avro_table;
[localhost:21000] > select * from avro_table;
+-----+
| a | b | c |
+-----+
| 1 | avro | 10 |
+-----+
```

## Data Type Considerations for Avro Tables

The Avro format defines a set of data types whose names differ from the names of the corresponding Impala data types. If you are preparing Avro files using other Hadoop components such as Pig or MapReduce, you might need to work with the type names defined by Avro. The following figure lists the Avro-defined types and the equivalent types in Impala.

<b>Primitive Types</b>
-----
STRING -> STRING
INT -> INT
BOOLEAN -> BOOLEAN
LONG -> BIGINT
FLOAT -> FLOAT
DOUBLE -> DOUBLE
<b>Logical Types</b>
-----
BYTES + logicalType = "decimal" -> DECIMAL
<b>Avro Types with No Impala Equivalent</b>
-----
RECORD, MAP, ARRAY, UNION, ENUM, FIXED, NULL
<b>Impala Types with No Avro Equivalent</b>
-----
TIMESTAMP

### Using the RCFile File Format with Impala Tables

Cloudera Impala supports using RCFile data files.

**Table 8: RCFile Format Support in Impala**

File Type	Format	Compression Codecs	Impala Can CREATE?	Impala Can INSERT?
RCFile	Structured	Snappy, gzip, deflate, bzip2	Yes.	No. Import data by using <code>LOAD DATA</code> on data files already in the right format, or use <code>INSERT</code> in Hive followed by <code>REFRESH table_name</code> in Impala.

### Creating RCFile Tables and Loading Data

If you do not have an existing data file to use, begin by creating one in the appropriate format.

#### To create an RCFile table:

In the `impala-shell` interpreter, issue a command similar to:

```
create table rcfie_table (column_specs) stored as rcfie;
```

Because Impala can query some kinds of tables that it cannot currently write to, after creating tables of certain file formats, you might use the Hive shell to load the data. See [How Impala Works with Hadoop File Formats](#) on page 399 for details. After loading data into a table through Hive or other mechanism outside of Impala, issue a `REFRESH table_name` statement the next time you connect to the Impala node, before querying the table, to make Impala recognize the new data.

- Important: See [Known Issues in the Current Production Release \(Impala 2.2.x / CDH 5.4.x\)](#) on page 496 for potential compatibility issues with RCFile tables created in Hive 0.12, due to a change in the default RCFile SerDe for Hive.

For example, here is how you might create some RCFile tables in Impala (by specifying the columns explicitly, or cloning the structure of another table), load data through Hive, and query them through Impala:

```
$ impala-shell -i localhost
[localhost:21000] > create table rcfie_table (x int) stored as rcfie;
[localhost:21000] > create table rcfie_clone like some_other_table stored as rcfie;
[localhost:21000] > quit;

$ hive
hive> insert into table rcfie_table select x from some_other_table;
3 Rows loaded to rcfie_table
Time taken: 19.015 seconds
hive> quit;

$ impala-shell -i localhost
[localhost:21000] > select * from rcfie_table;
Returned 0 row(s) in 0.23s
[localhost:21000] > -- Make Impala recognize the data loaded through Hive;
[localhost:21000] > refresh rcfie_table;
[localhost:21000] > select * from rcfie_table;
+---+
| x |
+---+
| 1 |
| 2 |
| 3 |
+---+
Returned 3 row(s) in 0.23s
```

## Enabling Compression for RCFile Tables

You may want to enable compression on existing tables. Enabling compression provides performance gains in most cases and is supported for RCFile tables. For example, to enable Snappy compression, you would specify the following additional settings when loading data through the Hive shell:

```
hive> SET hive.exec.compress.output=true;
hive> SET mapred.max.split.size=256000000;
hive> SET mapred.output.compression.type=BLOCK;
hive> SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
hive> INSERT OVERWRITE TABLE new_table SELECT * FROM old_table;
```

If you are converting partitioned tables, you must complete additional steps. In such a case, specify additional settings similar to the following:

```
hive> CREATE TABLE new_table (your_cols) PARTITIONED BY (partition_cols) STORED AS
new_format;
hive> SET hive.exec.dynamic.partition.mode=nonstrict;
hive> SET hive.exec.dynamic.partition=true;
hive> INSERT OVERWRITE TABLE new_table PARTITION(comma_separated_partition_cols) SELECT
* FROM old_table;
```

Remember that Hive does not require that you specify a source format for it. Consider the case of converting a table with two partition columns called year and month to a Snappy compressed RCFile. Combining the components outlined previously to complete this table conversion, you would specify settings similar to the following:

```
hive> CREATE TABLE tbl_rc (int_col INT, string_col STRING) STORED AS RCFILE;
hive> SET hive.exec.compress.output=true;
hive> SET mapred.max.split.size=256000000;
hive> SET mapred.output.compression.type=BLOCK;
hive> SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
hive> SET hive.exec.dynamic.partition.mode=nonstrict;
hive> SET hive.exec.dynamic.partition=true;
hive> INSERT OVERWRITE TABLE tbl_rc SELECT * FROM tbl;
```

To complete a similar process for a table that includes partitions, you would specify settings similar to the following:

```
hive> CREATE TABLE tbl_rc (int_col INT, string_col STRING) PARTITIONED BY (year INT)
STORED AS RCFILE;
hive> SET hive.exec.compress.output=true;
hive> SET mapred.max.split.size=256000000;
hive> SET mapred.output.compression.type=BLOCK;
hive> SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
hive> SET hive.exec.dynamic.partition.mode=nonstrict;
hive> SET hive.exec.dynamic.partition=true;
hive> INSERT OVERWRITE TABLE tbl_rc PARTITION(year) SELECT * FROM tbl;
```

- **Note:**

The compression type is specified in the following command:

```
SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
```

You could elect to specify alternative codecs such as GzipCodec here.

## Using the SequenceFile File Format with Impala Tables

Cloudera Impala supports using SequenceFile data files.

## How Impala Works with Hadoop File Formats

**Table 9: SequenceFile Format Support in Impala**

File Type	Format	Compression Codecs	Impala Can CREATE?	Impala Can INSERT?
<a href="#">SequenceFile</a>	Structured	Snappy, gzip, deflate, bzip2	Yes.	No. Import data by using <code>LOAD DATA</code> on data files already in the right format, or use <code>INSERT</code> in Hive followed by <code>REFRESH table_name</code> in Impala.

## Creating SequenceFile Tables and Loading Data

If you do not have an existing data file to use, begin by creating one in the appropriate format.

### To create a SequenceFile table:

In the `impala-shell` interpreter, issue a command similar to:

```
create table sequencefile_table (column_specs) stored as sequencefile;
```

Because Impala can query some kinds of tables that it cannot currently write to, after creating tables of certain file formats, you might use the Hive shell to load the data. See [How Impala Works with Hadoop File Formats](#) on page 399 for details. After loading data into a table through Hive or other mechanism outside of Impala, issue a `REFRESH table_name` statement the next time you connect to the Impala node, before querying the table, to make Impala recognize the new data.

For example, here is how you might create some SequenceFile tables in Impala (by specifying the columns explicitly, or cloning the structure of another table), load data through Hive, and query them through Impala:

```
$ impala-shell -i localhost
[localhost:21000] > create table seqfile_table (x int) stored as sequencefile;
[localhost:21000] > create table seqfile_clone like some_other_table stored as
sequencefile;
[localhost:21000] > quit;

$ hive
hive> insert into table seqfile_table select x from some_other_table;
3 Rows loaded to seqfile_table
Time taken: 19.047 seconds
hive> quit;

$ impala-shell -i localhost
[localhost:21000] > select * from seqfile_table;
Returned 0 row(s) in 0.23s
[localhost:21000] > -- Make Impala recognize the data loaded through Hive;
[localhost:21000] > refresh seqfile_table;
[localhost:21000] > select * from seqfile_table;
+---+
| x |
+---+
| 1 |
| 2 |
| 3 |
+---+
Returned 3 row(s) in 0.23s
```

## Enabling Compression for SequenceFile Tables

You may want to enable compression on existing tables. Enabling compression provides performance gains in most cases and is supported for SequenceFile tables. For example, to enable Snappy compression, you would specify the following additional settings when loading data through the Hive shell:

```
hive> SET hive.exec.compress.output=true;
hive> SET mapred.max.split.size=256000000;
```

```
hive> SET mapred.output.compression.type=BLOCK;
hive> SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
hive> insert overwrite table new_table select * from old_table;
```

If you are converting partitioned tables, you must complete additional steps. In such a case, specify additional settings similar to the following:

```
hive> create table new_table (your_cols) partitioned by (partition_cols) stored as
new_format;
hive> SET hive.exec.dynamic.partition.mode=nonstrict;
hive> SET hive.exec.dynamic.partition=true;
hive> insert overwrite table new_table partition(comma_separated_partition_cols) select
* from old_table;
```

Remember that Hive does not require that you specify a source format for it. Consider the case of converting a table with two partition columns called `year` and `month` to a Snappy compressed SequenceFile. Combining the components outlined previously to complete this table conversion, you would specify settings similar to the following:

```
hive> create table TBL_SEQ (int_col int, string_col string) STORED AS SEQUENCEFILE;
hive> SET hive.exec.compress.output=true;
hive> SET mapred.max.split.size=256000000;
hive> SET mapred.output.compression.type=BLOCK;
hive> SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
hive> SET hive.exec.dynamic.partition.mode=nonstrict;
hive> SET hive.exec.dynamic.partition=true;
hive> INSERT OVERWRITE TABLE tbl_seq SELECT * FROM tbl;
```

To complete a similar process for a table that includes partitions, you would specify settings similar to the following:

```
hive> CREATE TABLE tbl_seq (int_col INT, string_col STRING) PARTITIONED BY (year INT)
STORED AS SEQUENCEFILE;
hive> SET hive.exec.compress.output=true;
hive> SET mapred.max.split.size=256000000;
hive> SET mapred.output.compression.type=BLOCK;
hive> SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
hive> SET hive.exec.dynamic.partition.mode=nonstrict;
hive> SET hive.exec.dynamic.partition=true;
hive> INSERT OVERWRITE TABLE tbl_seq PARTITION(year) SELECT * FROM tbl;
```

■ **Note:**

The compression type is specified in the following command:

```
SET mapred.output.compression.codec=org.apache.hadoop.io.compress.SnappyCodec;
```

You could elect to specify alternative codecs such as GzipCodec here.

# Using Impala to Query HBase Tables

You can use Impala to query HBase tables. This capability allows convenient access to a storage system that is tuned for different kinds of workloads than the default with Impala. The default Impala tables use data files stored on HDFS, which are ideal for bulk loads and queries using full-table scans. In contrast, HBase can do efficient queries for data organized for OLTP-style workloads, with lookups of individual rows or ranges of values.

From the perspective of an Impala user, coming from an RDBMS background, HBase is a kind of key-value store where the value consists of multiple fields. The key is mapped to one column in the Impala table, and the various fields of the value are mapped to the other columns in the Impala table.

For background information on HBase, see the snapshot of the Apache HBase site (including documentation) for the level of HBase that comes with [CDH 4](#) or [CDH 5](#). To install HBase on a CDH cluster, see the installation instructions for [CDH 4](#) or [CDH 5](#).

## Overview of Using HBase with Impala

When you use Impala with HBase:

- You create the tables on the Impala side using the Hive shell, because the Impala `CREATE TABLE` statement currently does not support custom SerDes and some other syntax needed for these tables:
  - You designate it as an HBase table using the `STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'` clause on the Hive `CREATE TABLE` statement.
  - You map these specially created tables to corresponding tables that exist in HBase, with the clause `TBLPROPERTIES( "hbase.table.name" = "table_name_in_hbase" )` on the Hive `CREATE TABLE` statement.
  - See [Examples of Querying HBase Tables from Impala](#) on page 433 for a full example.
- You define the column corresponding to the HBase row key as a string with the `#string` keyword, or map it to a `STRING` column.
- Because Impala and Hive share the same metastore database, once you create the table in Hive, you can query or insert into it through Impala. (After creating a new table through Hive, issue the `INVALIDATE METADATA` statement in `impala-shell` to make Impala aware of the new table.)
- You issue queries against the Impala tables. For efficient queries, use `WHERE` clauses to find a single key value or a range of key values wherever practical, by testing the Impala column corresponding to the HBase row key. Avoid queries that do full-table scans, which are efficient for regular Impala tables but inefficient in HBase.

To work with an HBase table from Impala, ensure that the `impala` user has read/write privileges for the HBase table, using the `GRANT` command in the HBase shell. For details about HBase security, see the [Security chapter in the HBase Reference Guide](#).

## Configuring HBase for Use with Impala

HBase works out of the box with Impala. There is no mandatory configuration needed to use these two components together.

To avoid delays if HBase is unavailable during Impala startup or after an `INVALIDATE METADATA` statement, Cloudera recommends setting timeout values as follows in `/etc/impala/conf/hbase-site.xml` (for environments not managed by Cloudera Manager):

```
<property>
  <name>hbase.client.retries.number</name>
  <value>3</value>
</property>
<property>
  <name>hbase.rpc.timeout</name>
  <value>3000</value>
</property>
```

Currently, Cloudera Manager does not have an Impala-only override for HBase settings, so any HBase configuration change you make through Cloudera Manager would affect for all HBase applications. Therefore, this change is not recommended on systems managed by Cloudera Manager.

## Supported Data Types for HBase Columns

To understand how Impala column data types are mapped to fields in HBase, you should have some background knowledge about HBase first. You set up the mapping by running the `CREATE TABLE` statement in the Hive shell. See [the Hive wiki](#) for a starting point, and [Examples of Querying HBase Tables from Impala](#) on page 433 for examples.

HBase works as a kind of "bit bucket", in the sense that HBase does not enforce any typing for the key or value fields. All the type enforcement is done on the Impala side.

For best performance of Impala queries against HBase tables, most queries will perform comparisons in the `WHERE` against the column that corresponds to the HBase row key. When creating the table through the Hive shell, use the `STRING` data type for the column that corresponds to the HBase row key. Impala can translate conditional tests (through operators such as `=`, `<`, `BETWEEN`, and `IN`) against this column into fast lookups in HBase, but this optimization ("predicate pushdown") only works when that column is defined as `STRING`.

Starting in Impala 1.1, Impala also supports reading and writing to columns that are defined in the Hive `CREATE TABLE` statement using binary data types, represented in the Hive table definition using the `#binary` keyword, often abbreviated as `#b`. Defining numeric columns as binary can reduce the overall data volume in the HBase tables. You should still define the column that corresponds to the HBase row key as a `STRING`, to allow fast lookups using those columns.

## Performance Considerations for the Impala-HBase Integration

To understand the performance characteristics of SQL queries against data stored in HBase, you should have some background knowledge about how HBase interacts with SQL-oriented systems first. See [the Hive wiki](#) for a starting point; because Impala shares the same metastore database as Hive, the information about mapping columns from Hive tables to HBase tables is generally applicable to Impala too.

Impala uses the HBase client API via Java Native Interface (JNI) to query data stored in HBase. This querying does not read HFiles directly. The extra communication overhead makes it important to choose what data to store in HBase or in HDFS, and construct efficient queries that can retrieve the HBase data efficiently:

- Use HBase table for queries that return a single row or a range of rows, not queries that scan the entire table. (If a query has no `WHERE` clause, that is a strong indicator that it is an inefficient query for an HBase table.)
- If you have join queries that do aggregation operations on large fact tables and join the results against small dimension tables, consider using Impala for the fact tables and HBase for the dimension tables. (Because Impala does a full scan on the HBase table in this case, rather than doing single-row HBase lookups based on the join column, only use this technique where the HBase table is small enough that doing a full table scan does not cause a performance bottleneck for the query.)

## Using Impala to Query HBase Tables

Query predicates are applied to row keys as start and stop keys, thereby limiting the scope of a particular lookup. If row keys are not mapped to string columns, then ordering is typically incorrect and comparison operations do not work. For example, if row keys are not mapped to string columns, evaluating for greater than (>) or less than (<) cannot be completed.

Predicates on non-key columns can be sent to HBase to scan as `SingleColumnValueFilters`, providing some performance gains. In such a case, HBase returns fewer rows than if those same predicates were applied using Impala. While there is some improvement, it is not as great when start and stop rows are used. This is because the number of rows that HBase must examine is not limited as it is when start and stop rows are used. As long as the row key predicate only applies to a single row, HBase will locate and return that row. Conversely, if a non-key predicate is used, even if it only applies to a single row, HBase must still scan the entire table to find the correct result.

### Interpreting EXPLAIN Output for HBase Queries

For example, here are some queries against the following Impala table, which is mapped to an HBase table. The examples show excerpts from the output of the `EXPLAIN` statement, demonstrating what things to look for to indicate an efficient or inefficient query against an HBase table.

The first column (`cust_id`) was specified as the key column in the `CREATE EXTERNAL TABLE` statement; for performance, it is important to declare this column as `STRING`. Other columns, such as `BIRTH_YEAR` and `NEVER_LOGGED_ON`, are also declared as `STRING`, rather than their “natural” types of `INT` or `BOOLEAN`, because Impala can optimize those types more effectively in HBase tables. For comparison, we leave one column, `YEAR_REGISTERED`, as `INT` to show that filtering on this column is inefficient.

```
describe hbase_table;
Query: describe hbase_table
+-----+-----+-----+
| name | type | comment |
+-----+-----+-----+
| cust_id | string |          |
| birth_year | string |          |
| never_logged_on | string |          |
| private_email_address | string |          |
| year_registered | int |          |
+-----+-----+-----+
```

The best case for performance involves a single row lookup using an equality comparison on the column defined as the row key:

```
explain select count(*) from hbase_table where cust_id = 'some_user@example.com';
+-----+
| Explain String
+-----+
Estimated Per-Host Requirements: Memory=1.01GB Vcores=1
WARNING: The following tables are missing relevant table and/or column statistics.
hbase.hbase_table

03:AGGREGATE [MERGE FINALIZE]
|   output: sum(count(*))
|
02:EXCHANGE [PARTITION=UNPARTITIONED]
|
01:AGGREGATE
|   output: count(*)
```

```

00:SCAN HBASE [hbase.hbase_table]
  start key: some_user@example.com
  stop key: some_user@example.com\0

```

Another type of efficient query involves a range lookup on the row key column, using SQL operators such as greater than (or equal), less than (or equal), or BETWEEN. This example also includes an equality test on a non-key column; because that column is a STRING, Impala can let HBase perform that test, indicated by the hbase filters: line in the EXPLAIN output. Doing the filtering within HBase is more efficient than transmitting all the data to Impala and doing the filtering on the Impala side.

```

explain select count(*) from hbase_table where cust_id between 'a' and 'b'
  and never_logged_on = 'true';
+-----+
| Explain String
+-----+
...
01:AGGREGATE
|   output: count(*)
|
00:SCAN HBASE [hbase.hbase_table]
  start key: a
  stop key: b\0
  hbase filters: cols:never_logged_on EQUAL 'true'

```

The query is less efficient if Impala has to evaluate any of the predicates, because Impala must scan the entire HBase table. Impala can only push down predicates to HBase for columns declared as STRING. This example tests a column declared as INT, and the predicates: line in the EXPLAIN output indicates that the test is performed after the data is transmitted to Impala.

```

explain select count(*) from hbase_table where year_registered = 2010;
+-----+
| Explain String
+-----+
...
01:AGGREGATE
|   output: count(*)
|
00:SCAN HBASE [hbase.hbase_table]
  predicates: year_registered = 2010

```

## Using Impala to Query HBase Tables

The same inefficiency applies if the key column is compared to any non-constant value. Here, even though the key column is a `STRING`, and is tested using an equality operator, Impala must scan the entire HBase table because the key column is compared to another column value rather than a constant.

```
explain select count(*) from hbase_table where cust_id = private_email_address;
+-----+
| Explain String
+-----+
...
01:AGGREGATE
|   output: count(*)
|
00:SCAN HBASE [hbase.hbase_table]
  predicates: cust_id = private_email_address
+-----+
```

Currently, tests on the row key using `OR` or `IN` clauses are not optimized into direct lookups either. Such limitations might be lifted in the future, so always check the `EXPLAIN` output to be sure whether a particular SQL construct results in an efficient query or not for HBase tables.

```
explain select count(*) from hbase_table where
  cust_id = 'some_user@example.com' OR cust_id = 'other_user@example.com';
+-----+
| Explain String
+-----+
...
01:AGGREGATE
|   output: count(*)
|
00:SCAN HBASE [hbase.hbase_table]
  predicates: cust_id = 'some_user@example.com' OR cust_id = 'other_user@example.com'
+-----+
```

```
explain select count(*) from hbase_table where
  cust_id in ('some_user@example.com', 'other_user@example.com');
+-----+
| Explain String
+-----+
...
01:AGGREGATE
|   output: count(*)
|
00:SCAN HBASE [hbase.hbase_table]
  predicates: cust_id IN ('some_user@example.com', 'other_user@example.com')
+-----+
```

Either rewrite into separate queries for each value and combine the results in the application, or combine the single-row queries using UNION ALL:

```

select count(*) from hbase_table where cust_id = 'some_user@example.com';
select count(*) from hbase_table where cust_id = 'other_user@example.com';

explain
  select count(*) from hbase_table where cust_id = 'some_user@example.com'
  union all
  select count(*) from hbase_table where cust_id = 'other_user@example.com';
+-----+
| Explain String
+-----+
...
|   04:AGGREGATE
|   |   output: count(*)
|   |
|   03:SCAN HBASE [hbase.hbase_table]
|       start key: other_user@example.com
|       stop key: other_user@example.com\0
|
|   10:MERGE
|
...
|   02:AGGREGATE
|   |   output: count(*)
|
|   01:SCAN HBASE [hbase.hbase_table]
|       start key: some_user@example.com
|       stop key: some_user@example.com\0
+-----+

```

### Configuration Options for Java HBase Applications

If you have an HBase Java application that calls the `setCacheBlocks` or `setCaching` methods of the class [org.apache.hadoop.hbase.client.Scan](#), you can set these same caching behaviors through Impala query options, to control the memory pressure on the HBase region server. For example, when doing queries in HBase that result in full-table scans (which by default are inefficient for HBase), you can reduce memory usage and speed up the queries by turning off the `HBASE_CACHE_BLOCKS` setting and specifying a large number for the `HBASE_CACHING` setting.

To set these options, issue commands like the following in `impala-shell`:

```

-- Same as calling setCacheBlocks(true) or setCacheBlocks(false).
set hbase_cache_blocks=true;
set hbase_cache_blocks=false;

-- Same as calling setCaching(rows).
set hbase_caching=1000;

```

## Using Impala to Query HBase Tables

Or update the `impalad` defaults file `/etc/default/impala` and include settings for `HBASE_CACHE_BLOCKS` and/or `HBASE_CACHING` in the `-default_query_options` setting for `IMPALA_SERVER_ARGS`. See [Modifying Impala Startup Options](#) on page 48 for details.

- **Note:** In Impala 2.0 and later, these options are settable through the JDBC or ODBC interfaces using the `SET` statement.

## Use Cases for Querying HBase through Impala

The following are popular use cases for using Impala to query HBase tables:

- Keeping large fact tables in Impala, and smaller dimension tables in HBase. The fact tables use Parquet or other binary file format optimized for scan operations. Join queries scan through the large Impala fact tables, and cross-reference the dimension tables using efficient single-row lookups in HBase.
- Using HBase to store rapidly incrementing counters, such as how many times a web page has been viewed, or on a social network, how many connections a user has or how many votes a post received. HBase is efficient for capturing such changeable data: the append-only storage mechanism is efficient for writing each change to disk, and a query always returns the latest value. An application could query specific totals like these from HBase, and combine the results with a broader set of data queried from Impala.
- Storing very wide tables in HBase. Wide tables have many columns, possibly thousands, typically recording many attributes for an important subject such as a user of an online service. These tables are also often sparse, that is, most of the columns values are `NULL`, `0`, `false`, empty string, or other blank or placeholder value. (For example, any particular web site user might have never used some site feature, filled in a certain field in their profile, visited a particular part of the site, and so on.) A typical query against this kind of table is to look up a single row to retrieve all the information about a specific subject, rather than summing, averaging, or filtering millions of rows as in typical Impala-managed tables.

Or the HBase table could be joined with a larger Impala-managed table. For example, analyze the large Impala table representing web traffic for a site and pick out 50 users who view the most pages. Join that result with the wide user table in HBase to look up attributes of those users. The HBase side of the join would result in 50 efficient single-row lookups in HBase, rather than scanning the entire user table.

## Loading Data into an HBase Table

The Impala `INSERT` statement works for HBase tables. The `INSERT ... VALUES` syntax is ideally suited to HBase tables, because inserting a single row is an efficient operation for an HBase table. (For regular Impala tables, with data files in HDFS, the tiny data files produced by `INSERT ... VALUES` are extremely inefficient, so you would not use that technique with tables containing any significant data volume.)

When you use the `INSERT ... SELECT` syntax, the result in the HBase table could be fewer rows than you expect. HBase only stores the most recent version of each unique row key, so if an `INSERT ... SELECT` statement copies over multiple rows containing the same value for the key column, subsequent queries will only return one row with each key column value:

Although Impala does not have an `UPDATE` statement, you can achieve the same effect by doing successive `INSERT` statements using the same value for the key column each time:

## Limitations and Restrictions of the Impala and HBase Integration

The Impala integration with HBase has the following limitations and restrictions, some inherited from the integration between HBase and Hive, and some unique to Impala:

- If you issue a `DROP TABLE` for an internal (Impala-managed) table that is mapped to an HBase table, the underlying table is not removed in HBase. The Hive `DROP TABLE` statement also removes the HBase table in this case.
- The `INSERT OVERWRITE` statement is not available for HBase tables. You can insert new data, or modify an existing row by inserting a new row with the same key value, but not replace the entire contents of the table. You can do an `INSERT OVERWRITE` in Hive if you need this capability.
- If you issue a `CREATE TABLE LIKE` statement for a table mapped to an HBase table, the new table is also an HBase table, but inherits the same underlying HBase table name as the original. The new table is effectively an alias for the old one, not a new table with identical column structure. Avoid using `CREATE TABLE LIKE` for HBase tables, to avoid any confusion.
- Copying data into an HBase table using the Impala `INSERT ... SELECT` syntax might produce fewer new rows than are in the query result set. If the result set contains multiple rows with the same value for the key column, each row supercedes any previous rows with the same key value. Because the order of the inserted rows is unpredictable, you cannot rely on this technique to preserve the "latest" version of a particular key value.

## Examples of Querying HBase Tables from Impala

The following examples create an HBase table with four column families, create a corresponding table through Hive, then insert and query the table through Impala.

Note that in HBase shell, the table name is quoted in `CREATE` and `DROP` statements. Tables created in HBase begin in "enabled" state; before dropping them through the HBase shell, you must issue a `disable 'table_name'` statement.

```
$ hbase shell
15/02/10 16:07:45
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 0.94.2-cdh4.2.0, rUnknown, Fri Feb 15 11:51:18 PST 2013

hbase(main):001:0> create 'hbasealltypessmall', 'boolsCF', 'intsCF', 'floatsCF',
'stringsCF'
0 row(s) in 4.6520 seconds

=> Hbase:::Table - hbasealltypessmall
hbase(main):006:0> quit
```

Issue the following `CREATE TABLE` statement in the Hive shell. (The Impala `CREATE TABLE` statement currently does not support the `STORED BY` clause, so you switch into Hive to create the table, then back to Impala and the `impala-shell` interpreter to issue the queries.)

This example creates an external table mapped to the HBase table, usable by both Impala and Hive. It is defined as an external table so that when dropped by Impala or Hive, the original HBase table is not touched at all.

The `WITH SERDEPROPERTIES` clause specifies that the first column (`ID`) represents the row key, and maps the remaining columns of the SQL table to HBase column families. The mapping relies on the ordinal order of the columns in the table, not the column names in the `CREATE TABLE` statement. The first column is defined to be the lookup key; the `STRING` data type produces the fastest key-based lookups for HBase tables.

- Note:** For Impala with HBase tables, the most important aspect to ensure good performance is to use a `STRING` column as the row key, as shown in this example.

```
$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
Hive history file=/tmp/cloudera/hive_job_log_cloudera_201502101610_1980712808.txt
hive> use hbase;
```

## Using Impala to Query HBase Tables

```
OK
Time taken: 4.095 seconds
hive> CREATE EXTERNAL TABLE hbasestringids (
    >     id string,
    >     bool_col boolean,
    >     tinyint_col tinyint,
    >     smallint_col smallint,
    >     int_col int,
    >     bigint_col bigint,
    >     float_col float,
    >     double_col double,
    >     date_string_col string,
    >     string_col string,
    >     timestamp_col timestamp)
    > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
    > WITH SERDEPROPERTIES (
    >     "hbase.columns.mapping" =
    >
":key,boolsCF:bool_col,intsCF:tinyint_col,intsCF:smallint_col,intsCF:int_col,intsCF:\\"+
    >     bigint_col,floatsCF:float_col,floatsCF:double_col,stringsCF:date_string_col,\\"+
    >     stringsCF:string_col,stringsCF:timestamp_col"
    > )
    > TBLPROPERTIES("hbase.table.name" = "hbasealltypessmall");
OK
Time taken: 2.879 seconds
hive> quit;
```

Once you have established the mapping to an HBase table, you can issue DML statements and queries from Impala. The following example shows a series of `INSERT` statements followed by a query. The ideal kind of query from a performance standpoint retrieves a row from the table based on a row key mapped to a string column. An initial `INVALIDATE METADATA table_name` statement makes the table created through Hive visible to Impala.

```
$ impala-shell -i localhost -d hbase
Starting Impala Shell without Kerberos authentication
Connected to localhost:21000
Server version: impalad version 2.1.0-cdh4 RELEASE (build
d520a9cdea2fc97e8d5da9fbb0244e60ee416bfa)
Welcome to the Impala shell. Press TAB twice to see a list of available commands.

Copyright (c) 2012 Cloudera, Inc. All rights reserved.

(Shell build version: Impala Shell v2.1.0-cdh4 (d520a9c) built on Mon Dec  8 21:41:17
PST 2014)
Query: use `hbase`
[localhost:21000] > invalidate metadata hbasestringids;
Fetched 0 row(s) in 0.09s
[localhost:21000] > desc hbasestringids;
+-----+-----+-----+
| name      | type       | comment |
+-----+-----+-----+
| id        | string     |
| bool_col  | boolean    |
| double_col| double     |
| float_col | float      |
| bigint_col| bigint    |
| int_col   | int        |
| smallint_col| smallint |
| tinyint_col| tinyint   |
| date_string_col| string  |
| string_col| string    |
| timestamp_col| timestamp|
+-----+-----+-----+
Fetched 11 row(s) in 0.02s
[localhost:21000] > insert into hbasestringids values
('0001',true,3.141,9.94,1234567,32768,4000,76,'2014-12-31','Hello world',now());
Inserted 1 row(s) in 0.26s
[localhost:21000] > insert into hbasestringids values
('0002',false,2.004,6.196,1500,8000,129,127,'2014-01-01','Foo bar',now());
Inserted 1 row(s) in 0.12s
[localhost:21000] > select * from hbasestringids where id = '0001';
```

id	bool_col	double_col	float_col	bigint_col	int_col	smallint_col
tinyint_col	date_string_col	string_col	timestamp_col			
0001	true	3.141	9.939999580383301	1234567	32768	4000
	76	2014-12-31	Hello world	2015-02-10 16:36:59.764838000		

Fetched 1 row(s) in 0.54s

- **Note:** After you create a table in Hive, such as the HBase mapping table in this example, issue an `INVALIDATE METADATA table_name` statement the next time you connect to Impala, make Impala aware of the new table. (Prior to Impala 1.2.4, you could not specify the table name if Impala was not aware of the table yet; in Impala 1.2.4 and higher, specifying the table name avoids reloading the metadata for other tables that are not changed.)

# Using Impala to Query the Amazon S3 Filesystem (Unsupported Preview)

- **Important:**

Impala query support for Amazon S3 is included in CDH 5.4.0, but is not currently supported or recommended for production use. If you're interested in this feature, try it out in a test environment until we address the issues and limitations needed for production-readiness.

You can use Impala to query data residing on the Amazon S3 filesystem. This capability allows convenient access to a storage system that is remotely managed, accessible from anywhere, and integrated with various cloud-based services. Impala can query files in any supported file format from S3. The S3 storage location can be for an entire table or individual partitions in a partitioned table.

The default Impala tables use data files stored on HDFS, which are ideal for bulk loads and queries using full-table scans. In contrast, queries against S3 data are less performant, making S3 suitable for holding "cold" data that is only queried occasionally, while more frequently accessed "hot" data resides in HDFS. In a partitioned table, you can set the `LOCATION` attribute for individual partitions to put some partitions on HDFS and others on S3, typically depending on the age of the data.

## Specifying Impala Credentials to Access Data in S3

To allow Impala to access data in S3, specify values for the following configuration settings in your `hdfs-site.xml` file:

```
<property>
<name>fs.s3a.access.key</name>
<value>your_access_key</value>
</property>
<property>
<name>fs.s3a.secret.key</name>
<value>your_secret_key</value>
</property>
```

As of CDH 5.4.0, these settings do not have corresponding controls in the Cloudera Manager user interface. Specify them in the **HDFS Client Advanced Configuration Snippet (Safety Valve)** for `hdfs-site.xml` field. After specifying the credentials, restart both the Impala and Hive services. (Restarting Hive is required because Impala queries, `CREATE TABLE` statements, and so on go through the Hive metastore.)

- **Important:** Although you can specify the access key ID and secret key as part of the `s3a://` URL in the `LOCATION` attribute, doing so makes this sensitive information visible in many places, such as `DESCRIBE FORMATTED` output and Impala log files. Therefore, specify this information centrally in the `hdfs-site.xml` file, and restrict read access to that file to only trusted users.

## Loading Data into S3 for Impala Queries

Use the Amazon-provided methods to bring data files into S3 for querying through Impala. See [the Amazon S3 web site](#) for details. For example, you might use one or more of the following techniques:

- The [Amazon AWS / S3 web interface](#) to upload from a web browser.
- The [Amazon AWS CLI](#) to manipulate files from the command line.
- Other S3-enabled software, such as [the S3Tools client software](#).

After you upload data files to a location already mapped to an Impala table or partition, or if you delete files in S3 from such a location, issue the `REFRESH table_name` statement to make Impala aware of the new set of data files.

## Creating Impala Databases, Tables, and Partitions for Data Stored on S3

Impala reads data for a table or partition from S3 based on the `LOCATION` attribute for the table or partition. Specify the S3 details in the `LOCATION` clause of a `CREATE TABLE` or `ALTER TABLE` statement. The notation for the `LOCATION` clause is `s3a://bucket_name/path/to/file`. The filesystem prefix is always `s3a://`; Impala does not support the `s3://` or `s3n://` prefixes.

For a partitioned table, either specify a separate `LOCATION` clause for each new partition, or specify a base `LOCATION` for the table and set up a directory structure in S3 to mirror the way Impala partitioned tables are structured in HDFS. Although, strictly speaking, S3 filenames do not have directory paths, Impala treats S3 filenames with / characters the same as HDFS pathnames that include directories.

To point a nonpartitioned table or an individual partition at S3 involves specifying a single directory path in S3, which could be any arbitrary directory. To replicate the structure of an entire Impala partitioned table or database in S3 requires more care, with directories and subdirectories nested and named to match the equivalent directory tree in HDFS. Consider setting up an empty staging area if necessary in HDFS, and recording the complete directory structure so that you can replicate it in S3.

For convenience when working with multiple tables with data files stored in S3, you can create a database with a `LOCATION` attribute pointing to an S3 path. Specify a URL of the form `s3a://bucket/root/path/for/database` for the `LOCATION` attribute of the database. Any tables created inside that database automatically create directories underneath the one specified by the database `LOCATION` attribute.

For example, the following session creates a partitioned table where only a single partition resides on S3. The partitions for years 2013 and 2014 are located on HDFS. The partition for year 2015 includes a `LOCATION` attribute with an `s3a://` URL, and so refers to data residing on S3, under a specific path underneath the bucket `impala-demo`.

```
[localhost:21000] > create database db_on_hdfs;
[localhost:21000] > use db_on_hdfs;
[localhost:21000] > create table mostly_on_hdfs (x int) partitioned by (year int);
[localhost:21000] > alter table mostly_on_hdfs add partition (year=2013);
[localhost:21000] > alter table mostly_on_hdfs add partition (year=2014);
[localhost:21000] > alter table mostly_on_hdfs add partition (year=2015) location
's3a://impala-demo/dir1/dir2/dir3/t1';
```

The following session creates a database and two partitioned tables residing entirely on S3, one partitioned by a single column and the other partitioned by multiple columns. Because a `LOCATION` attribute with an `s3a://` URL is specified for the database, the tables inside that database are automatically created on S3 underneath the database directory. To see the names of the associated subdirectories, including the partition key values, we use an S3 client tool to examine how the directory structure is organized on S3. For example, Impala partition directories such as `month=1` do not include leading zeroes such sometimes appear in partitioned tables created through Hive.

```
[localhost:21000] > create database db_on_s3 location 's3a://impala-demo/dir1/dir2/dir3';
[localhost:21000] > use db_on_s3;

[localhost:21000] > create table partitioned_on_s3 (x int) partitioned by (year int);
[localhost:21000] > alter table partitioned_on_s3 add partition (year=2013);
[localhost:21000] > alter table partitioned_on_s3 add partition (year=2014);
[localhost:21000] > alter table partitioned_on_s3 add partition (year=2015);

[localhost:21000] > !aws s3 ls s3://impala-demo/dir1/dir2/dir3 --recursive;
2015-03-17 13:56:34          0 dir1/dir2/dir3/
2015-03-17 16:43:28          0 dir1/dir2/dir3/partitioned_on_s3/
2015-03-17 16:43:49          0 dir1/dir2/dir3/partitioned_on_s3/year=2013/
2015-03-17 16:43:53          0 dir1/dir2/dir3/partitioned_on_s3/year=2014/
2015-03-17 16:43:58          0 dir1/dir2/dir3/partitioned_on_s3/year=2015/
```

## Using Impala to Query the Amazon S3 Filesystem (Unsupported Preview)

```
[localhost:21000] > create table partitioned_multiple_keys (x int) partitioned by (year
  smallint, month tinyint, day tinyint);
[localhost:21000] > alter table partitioned_multiple_keys add partition
  (year=2015,month=1,day=1);
[localhost:21000] > alter table partitioned_multiple_keys add partition
  (year=2015,month=1,day=31);
[localhost:21000] > alter table partitioned_multiple_keys add partition
  (year=2015,month=2,day=28);

[localhost:21000] > !aws s3 ls s3://impala-demo/dir1/dir2/dir3 --recursive;
2015-03-17 13:56:34      0 dir1/dir2/dir3/
2015-03-17 16:47:13      0 dir1/dir2/dir3/partitioned_multiple_keys/
2015-03-17 16:47:44      0
dir1/dir2/dir3/partitioned_multiple_keys/year=2015/month=1/day=1/
2015-03-17 16:47:50      0
dir1/dir2/dir3/partitioned_multiple_keys/year=2015/month=1/day=31/
2015-03-17 16:47:57      0
dir1/dir2/dir3/partitioned_multiple_keys/year=2015/month=2/day=28/
2015-03-17 16:43:28      0 dir1/dir2/dir3/partitioned_on_s3/
2015-03-17 16:43:49      0 dir1/dir2/dir3/partitioned_on_s3/year=2013/
2015-03-17 16:43:53      0 dir1/dir2/dir3/partitioned_on_s3/year=2014/
2015-03-17 16:43:58      0 dir1/dir2/dir3/partitioned_on_s3/year=2015/
```

The `CREATE DATABASE` and `CREATE TABLE` statements create the associated directory paths if they do not already exist. You can specify multiple levels of directories, and the `CREATE` statement creates all appropriate levels, similar to using `mkdir -p`.

Use the standard S3 file upload methods to actually put the data files into the right locations. You can also put the directory paths and data files in place before creating the associated Impala databases or tables, and Impala automatically uses the data from the appropriate location after the associated databases and tables are created.

You can switch whether an existing table or partition points to data in HDFS or S3. For example, if you have an Impala table or partition pointing to data files in HDFS or S3, and you later transfer those data files to the other filesystem, use an `ALTER TABLE` statement to adjust the `LOCATION` attribute of the corresponding table or partition to reflect that change. Because Impala does not have an `ALTER DATABASE` statement, this location-switching technique is not practical for entire databases that have a custom `LOCATION` attribute.

## Internal and External Tables Located on S3

Just as with tables located on HDFS storage, you can designate S3-based tables as either internal (managed by Impala) or external, by using the syntax `CREATE TABLE` or `CREATE EXTERNAL TABLE` respectively. When you drop an internal table, the files associated with the table are removed, even if they are on S3 storage. When you drop an external table, the files associated with the table are left alone, and are still available for access by other tools or components. See [Tables](#) on page 156 for details.

If the data on S3 is intended to be long-lived and accessed by other tools in addition to Impala, create any associated S3 tables with the `CREATE EXTERNAL TABLE` syntax, so that the files are not deleted from S3 when the table is dropped.

If the data on S3 is only needed for querying by Impala and can be safely discarded once the Impala workflow is complete, create the associated S3 tables using the `CREATE TABLE` syntax, so that dropping the table also deletes the corresponding data files on S3.

For example, this session creates a table in S3 with the same column layout as a table in HDFS, then examines the S3 table and queries some data from it. The table in S3 works the same as a table in HDFS as far as the expected file format of the data, table and column statistics, and other table properties. The only indication that it is not an HDFS table is the `s3a://` URL in the `LOCATION` property. Many data files can reside in the S3 directory, and their combined contents form the table data. Because the data in this example is uploaded after the table is created, a `REFRESH` statement prompts Impala to update its cached information about the data files.

```
[localhost:21000] > create table usa_cities_s3 like usa_cities location
  's3a://impala-demo/usa_cities';
```

## Using Impala to Query the Amazon S3 Filesystem (Unsupported Preview)

```
[localhost:21000] > desc usa_cities_s3;
+-----+-----+-----+
| name | type   | comment |
+-----+-----+-----+
| id   | smallint |
| city | string   |
| state| string   |
+-----+-----+-----+
-- Now from a web browser, upload the same data file(s) to S3 as in the HDFS table,
under the relevant bucket and path.
-- If you already have the data in S3, you would point the table LOCATION at an existing
path.

[localhost:21000] > refresh usa_cities_s3;
[localhost:21000] > select count(*) from usa_cities_s3;
+-----+
| count(*) |
+-----+
| 289      |
+-----+
[localhost:21000] > select distinct state from sample_data_s3 limit 5;
+-----+
| state   |
+-----+
| Louisiana|
| Minnesota|
| Georgia   |
| Alaska    |
| Ohio     |
+-----+
[localhost:21000] > desc formatted usa_cities_s3;
+-----+-----+-----+
| name           | type        | comment |
+-----+-----+-----+
| # col_name     | data_type   |          |
| id             | NULL        |          |
| city           | smallint    |          |
| state          | string      |          |
|                 | string      |          |
|                 | NULL        |          |
|                 | s3_testing  |          |
|                 | jrussell    |          |
|                 | Mon Mar 16 11:36:25 PDT 2015 |          |
|                 | UNKNOWN     |          |
|                 | None        |          |
|                 | 0           |          |
|                 | s3a://impala-demo/usa_cities |          |
|                 | MANAGED_TABLE|          |
| ...            |           |          |
+-----+-----+-----+
```

In this case, we have already uploaded a Parquet file with a million rows of data to the `sample_data` directory underneath the `impala-demo` bucket on S3. This session creates a table with matching column settings pointing to the corresponding location in S3, then queries the table. Because the data is already in place on S3 when the table is created, no `REFRESH` statement is required.

```
[localhost:21000] > create table sample_data_s3 (id int, id bigint, val int, zerofill
  string,
          > name string, assertion boolean, city string, state string)
          > stored as parquet location 's3a://impala-demo/sample_data';
[localhost:21000] > select count(*) from sample_data_s3;
+-----+
| count(*) |
+-----+
| 1000000  |
+-----+
[localhost:21000] > select count(*) howmany, assertion from sample_data_s3 group by
  assertion;
+-----+
```

## Using Impala to Query the Amazon S3 Filesystem (Unsupported Preview)

howmany	assertion
667149	true
332851	false

## Running and Tuning Impala Queries for Data Stored on S3

Once the appropriate `LOCATION` attributes are set up at the table or partition level, you query data stored in S3 exactly the same as data stored on HDFS or in HBase:

- Queries against S3 data support all the same file formats as for HDFS data.
- Tables can be unpartitioned or partitioned. For partitioned tables, either manually construct paths in S3 corresponding to the HDFS directories representing partition key values, or use `ALTER TABLE ... ADD PARTITION` to set up the appropriate paths in S3.
- HDFS and HBase tables can be joined to S3 tables, or S3 tables can be joined with each other.
- Authorization using the Sentry framework to control access to databases, tables, or columns works the same whether the data is in HDFS or in S3.
- The `catalogd` daemon caches metadata for both HDFS and S3 tables. Use `REFRESH` and `INVALIDATE METADATA` for S3 tables in the same situations where you would issue those statements for HDFS tables.
- Queries against S3 tables are subject to the same kinds of admission control and resource management as HDFS tables.
- Metadata about S3 tables is stored in the same metastore database as for HDFS tables.
- You can set up views referring to S3 tables, the same as for HDFS tables.
- The `COMPUTE STATS`, `SHOW TABLE STATS`, and `SHOW COLUMN STATS` statements work for S3 tables also.

## Understanding and Tuning Impala Query Performance for S3 Data

Although Impala queries for data stored in S3 might be less performant than queries against the equivalent data stored in HDFS, you can still do some tuning. Here are techniques you can use to interpret explain plans and profiles for queries against S3 data, and tips to achieve the best performance possible for such queries.

All else being equal, performance is expected to be lower for queries running against data on S3 rather than HDFS. The actual mechanics of the `SELECT` statement are somewhat different when the data is in S3. Although the work is still distributed across the datanodes of the cluster, Impala might parallelize the work for a distributed query differently for data on HDFS and S3. S3 does not have the same block notion as HDFS, so Impala uses heuristics to determine how to split up large S3 files for processing in parallel. Because all hosts can access any S3 data file with equal efficiency, the distribution of work might be different than for HDFS data, where the data blocks are physically read using short-circuit local reads by hosts that contain the appropriate block replicas. Although the I/O to read the S3 data might be spread evenly across the hosts of the cluster, the fact that all data is initially retrieved across the network means that the overall query performance is likely to be lower for S3 data than for HDFS data.

When optimizing aspects of complex queries such as the join order, Impala treats tables on HDFS and S3 the same way. Therefore, follow all the same tuning recommendations for S3 tables as for HDFS ones, such as using the `COMPUTE STATS` statement to help Impala construct accurate estimates of row counts and cardinality. See [Tuning Impala for Performance](#) on page 354 for details.

In query profile reports, the numbers for `BytesReadLocal`, `BytesReadShortCircuit`, `BytesReadDataNodeCached`, and `BytesReadRemoteUnexpected` are blank because those metrics come from HDFS.

## Restrictions on Impala Support for S3

Impala requires that the default filesystem for the cluster be HDFS. You cannot use S3 as the only filesystem in the cluster.

Currently, Impala cannot perform DML operations (`INSERT`, `LOAD DATA`, or `CREATE TABLE AS SELECT`) where the destination is a table or partition located on an S3 filesystem. This restriction is because S3 does not support file renaming, which Impala uses during DML operations.

Impala does not support the old `s3://` block-based and `s3n://` filesystem schemes, only `s3a://`.

Although S3 is often used to store JSON-formatted data, the current Impala support for S3 does not include directly querying JSON data. For Impala queries, use data files in one of the file formats listed in [How Impala Works with Hadoop File Formats](#) on page 399. If you have data in JSON format, you can prepare a flattened version of that data for querying by Impala as part of your ETL cycle.

You cannot use the `ALTER TABLE ... SET CACHED` statement for tables or partitions that are located in S3.

## Using Impala with Isilon Storage

You can use Impala to query data files that reside on EMC Isilon storage devices, rather than in HDFS. This capability allows convenient query access to a storage system where you might already be managing large volumes of data. The combination of the Impala query engine and Isilon storage is certified on CDH 5.4.3.

Because the EMC Isilon storage devices use a global value for the block size rather than a configurable value for each file, the `PARQUET_FILE_SIZE` query option has no effect when Impala inserts data into a table or partition residing on Isilon storage. Use the `isi` command to set the default block size globally on the Isilon device. For example, to set the Isilon default block size to 256 MB, the recommended size for Parquet data files for Impala, issue the following command:

```
isi hdfs settings modify --default-block-size=256MB
```

The typical use case for Impala and Isilon together is to use Isilon for the default filesystem, replacing HDFS entirely. In this configuration, when you create a database, table, or partition, the data always resides on Isilon storage and you do not need to specify any special `LOCATION` attribute. If you do specify a `LOCATION` attribute, its value refers to a path within the Isilon filesystem. For example:

```
-- If the default filesystem is Isilon, all Impala data resides there  
-- and all Impala databases and tables are located there.  
CREATE TABLE t1 (x INT, s STRING);  
  
-- You can specify LOCATION for database, table, or partition,  
-- using values from the Isilon filesystem.  
CREATE DATABASE d1 LOCATION '/some/path/on/isilon/server/d1.db';  
CREATE TABLE d1.t2 (a TINYINT, b BOOLEAN);
```

Impala can write to, delete, and rename data files and database, table, and partition directories on Isilon storage. Therefore, Impala statements such as `CREATE TABLE`, `DROP TABLE`, `CREATE DATABASE`, `DROP DATABASE`, `ALTER TABLE`, and `INSERT` work the same with Isilon storage as with HDFS.

When the Impala spill-to-disk feature is activated by a query that approaches the memory limit, Impala writes all the temporary data to a local (not Isilon) storage device. Because the I/O bandwidth for the temporary data depends on the number of local disks, and clusters using Isilon storage might not have as many local disks attached, pay special attention on Isilon-enabled clusters to any queries that use the spill-to-disk feature. Where practical, tune the queries or allocate extra memory for Impala to avoid spilling. Although you can specify an Isilon storage device as the destination for the temporary data for the spill-to-disk feature, that configuration is not recommended due to the need to transfer the data both ways using remote I/O.

When tuning Impala queries on HDFS, you typically try to avoid any remote reads. When the data resides on Isilon storage, all the I/O consists of remote reads. Do not be alarmed when you see non-zero numbers for remote read measurements in query profile output. The benefit of the Impala and Isilon integration is primarily convenience of not having to move or copy large volumes of data to HDFS, rather than raw query performance. You can increase the performance of Impala I/O for Isilon systems by increasing the value for the `num_remote_hdfs_io_threads` configuration parameter, in the Cloudera Manager user interface for clusters using Cloudera Manager, or through the `--num_remote_hdfs_io_threads` startup option for the `impalad` daemon on clusters not using Cloudera Manager.

For information about managing Isilon storage devices through Cloudera Manager, see [Managing Isilon](#).

# Using Impala Logging

The Impala logs record information about:

- Any errors Impala encountered. If Impala experienced a serious error during startup, you must diagnose and troubleshoot that problem before you can do anything further with Impala.
- How Impala is configured.
- Jobs Impala has completed.

■ **Note:**

Formerly, the logs contained the query profile for each query, showing low-level details of how the work is distributed among nodes and how intermediate and final results are transmitted across the network. To save space, those query profiles are now stored in zlib-compressed files in `/var/log/impala/profiles`. You can access them through the Impala web user interface. For example, at `http://impalad-node-hostname:25000/queries`, each query is followed by a `Profile` link leading to a page showing extensive analytical data for the query execution.

The auditing feature introduced in Impala 1.1.1 produces a separate set of audit log files when enabled. See [Auditing Impala Operations](#) on page 108 for details.

The lineage feature introduced in Impala 2.2.0 produces a separate lineage log file when enabled. See [Viewing Lineage Information for Impala Data](#) on page 110 for details.

## Locations and Names of Impala Log Files

- By default, the log files are under the directory `/var/log/impala`. To change log file locations, modify the defaults file described in [Starting Impala](#) on page 47.
- The significant files for the `impalad` process are `impalad.INFO`, `impalad.WARNING`, and `impalad.ERROR`. You might also see a file `impalad.FATAL`, although this is only present in rare conditions.
- The significant files for the `statestored` process are `statestored.INFO`, `statestored.WARNING`, and `statestored.ERROR`. You might also see a file `statestored.FATAL`, although this is only present in rare conditions.
- The significant files for the `catalogd` process are `catalogd.INFO`, `catalogd.WARNING`, and `catalogd.ERROR`. You might also see a file `catalogd.FATAL`, although this is only present in rare conditions.
- Examine the `.INFO` files to see configuration settings for the processes.
- Examine the `.WARNING` files to see all kinds of problem information, including such things as suboptimal settings and also serious runtime errors.
- Examine the `.ERROR` and/or `.FATAL` files to see only the most serious errors, if the processes crash, or queries fail to complete. These messages are also in the `.WARNING` file.
- A new set of log files is produced each time the associated daemon is restarted. These log files have long names including a timestamp. The `.INFO`, `.WARNING`, and `.ERROR` files are physically represented as symbolic links to the latest applicable log files.
- The init script for the `impala-server` service also produces a consolidated log file `/var/logs/impalad/impala-server.log`, with all the same information as the corresponding `.INFO`, `.WARNING`, and `.ERROR` files.
- The init script for the `impala-state-store` service also produces a consolidated log file `/var/logs/impalad/impala-state-store.log`, with all the same information as the corresponding `.INFO`, `.WARNING`, and `.ERROR` files.

Impala stores information using the `glog_v` logging system. You will see some messages referring to C++ file names. Logging is affected by:

## Using Impala Logging

- The `GLOG_v` environment variable specifies which types of messages are logged. See [Setting Logging Levels](#) on page 445 for details.
- The `-logbuflevel` startup flag for the `impalad` daemon specifies how often the log information is written to disk. The default is 0, meaning that the log is immediately flushed to disk when Impala outputs an important messages such as a warning or an error, but less important messages such as informational ones are buffered in memory rather than being flushed to disk immediately.
- Cloudera Manager has an Impala configuration setting that sets the `-logbuflevel` startup option.

## Managing Impala Logs through Cloudera Manager or Manually

Cloudera recommends installing Impala through the Cloudera Manager administration interface. To assist with troubleshooting, Cloudera Manager collects front-end and back-end logs together into a single view, and let you do a search across log data for all the managed nodes rather than examining the logs on each node separately. If you installed Impala using Cloudera Manager, refer to the topics on Monitoring Services ([CDH 5](#), [CDH 4](#)) or Logs ([CDH 5](#), [CDH 4](#)).

If you are using Impala in an environment not managed by Cloudera Manager, review Impala log files on each host, when you have traced an issue back to a specific system.

## Rotating Impala Logs

Impala periodically switches the physical files representing the current log files, after which it is safe to remove the old files if they are no longer needed.

Impala can automatically remove older unneeded log files, a feature known as *log rotation*.

In Impala 2.2 and higher, the `-max_log_files` configuration option specifies how many log files to keep at each severity level. You can specify an appropriate setting for each Impala-related daemon (`impalad`, `statestored`, and `catalogd`). The default value is 10, meaning that Impala preserves the latest 10 log files for each severity level (`INFO`, `WARNING`, `ERROR`, and `FATAL`). Impala checks to see if any old logs need to be removed based on the interval specified in the `logbufsecs` setting, every 5 seconds by default.

A value of 0 preserves all log files, in which case you would set up manual log rotation using your Linux tool or technique of choice. A value of 1 preserves only the very latest log file.

To set up log rotation on a system managed by Cloudera Manager 5.4.0 and higher, search for the `max_log_files` option name and set the appropriate value for the `Maximum Log Files` field for each Impala configuration category (Impala, Catalog Server, and StateStore). Then restart the Impala service. In earlier Cloudera Manager releases, specify the `-max_log_files=maximum` option in the **Command Line Argument Advanced Configuration Snippet (Safety Valve)** field for each Impala configuration category.

## Reviewing Impala Logs

By default, the Impala log is stored at `/var/logs/impalad/`. The most comprehensive log, showing informational, warning, and error messages, is in the file name `impalad.INFO`. View log file contents by using the web interface or by examining the contents of the log file. (When you examine the logs through the file system, you can troubleshoot problems by reading the `impalad.WARNING` and/or `impalad.ERROR` files, which contain the subsets of messages indicating potential problems.)

On a machine named `impala.example.com` with default settings, you could view the Impala logs on that machine by using a browser to access `http://impala.example.com:25000/logs`.

- **Note:**

The web interface limits the amount of logging information displayed. To view every log entry, access the log files directly through the file system.

You can view the contents of the `impalad.INFO` log file in the file system. With the default configuration settings, the start of the log file appears as follows:

```
[user@example impalad]$ pwd
/var/log/impalad
[user@example impalad]$ more impalad.INFO
Log file created at: 2013/01/07 08:42:12
Running on machine: impala.example.com
Log line format: [IWEF]mmdd hh:mm:ss.uuuuuu threadid file:line] msg
I0107 08:42:12.292155 14876 daemon.cc:34] impalad version 0.4 RELEASE (build
9d7fadca0461ab40b9e9df8cda47107ec6b27cff)
Built on Fri, 21 Dec 2012 12:55:19 PST
I0107 08:42:12.292484 14876 daemon.cc:35] Using hostname: impala.example.com
I0107 08:42:12.292706 14876 logging.cc:76] Flags (see also /varz are on debug webserver):
--dump_ir=false
--module_output=
--be_port=22000
--classpath=
--hostname=impala.example.com
```

- **Note:** The preceding example shows only a small part of the log file. Impala log files are often several megabytes in size.

## Understanding Impala Log Contents

The logs store information about Impala startup options. This information appears once for each time Impala is started and may include:

- Machine name.
- Impala version number.
- Flags used to start Impala.
- CPU information.
- The number of available disks.

There is information about each job Impala has run. Because each Impala job creates an additional set of data about queries, the amount of job specific data may be very large. Logs may contain detailed information on jobs. These detailed log entries may include:

- The composition of the query.
- The degree of data locality.
- Statistics on data throughput and response times.

## Setting Logging Levels

Impala uses the GLOG system, which supports three logging levels. You can adjust the logging levels using the Cloudera Manager Admin Console. You can adjust logging levels without going through the Cloudera Manager Admin Console by exporting variable settings. To change logging settings manually, use a command similar to the following on each node before starting `impalad`:

```
export GLOG_v=1
```

## Using Impala Logging

- **Note:** For performance reasons, Cloudera highly recommends not enabling the most verbose logging level of 3.

For more information on how to configure GLOG, including how to set variable logging levels for different system components, see [How To Use Google Logging Library \(glog\)](#).

### Understanding What is Logged at Different Logging Levels

As logging levels increase, the categories of information logged are cumulative. For example, GLOG\_v=2 records everything GLOG\_v=1 records, as well as additional information.

Increasing logging levels imposes performance overhead and increases log size. Cloudera recommends using GLOG\_v=1 for most cases: this level has minimal performance impact but still captures useful troubleshooting information.

Additional information logged at each level is as follows:

- GLOG\_v=1 - The default level. Logs information about each connection and query that is initiated to an impalad instance, including runtime profiles.
- GLOG\_v=2 - Everything from the previous level plus information for each RPC initiated. This level also records query execution progress information, including details on each file that is read.
- GLOG\_v=3 - Everything from the previous level plus logging of every row that is read. This level is only applicable for the most serious troubleshooting and tuning scenarios, because it can produce exceptionally large and detailed log files, potentially leading to its own set of performance and capacity problems.

## Redacting Sensitive Information from Impala Log Files

**Log redaction** is a security feature that prevents sensitive information from being displayed in locations used by administrators for monitoring and troubleshooting, such as log files, the Cloudera Manager user interface, and the Impala debug web user interface. You configure regular expressions that match sensitive types of information processed by your system, such as credit card numbers or tax IDs, and literals matching these patterns are obfuscated wherever they would normally be recorded in log files or displayed in administration or debugging user interfaces.

In a security context, the log redaction feature is complementary to the Sentry authorization framework. Sentry prevents unauthorized users from being able to directly access table data. Redaction prevents administrators or support personnel from seeing the smaller amounts of sensitive or personally identifying information (PII) that might appear in queries issued by those authorized users.

See [Sensitive Data Redaction](#) for details about how to enable this feature and set up the regular expressions to detect and redact sensitive information within SQL statement text.

# Troubleshooting Impala

Use the following steps to diagnose and debug problems with any aspect of Impala.

## Troubleshooting Impala SQL Syntax Issues

In general, if queries issued against Impala fail, you can try running these same queries against Hive.

- If a query fails against both Impala and Hive, it is likely that there is a problem with your query or other elements of your CDH environment:
  - Review the [Language Reference](#) to ensure your query is valid.
  - Check [Impala Reserved Words](#) on page 453 to see if any database, table, column, or other object names in your query conflict with Impala reserved words. Quote those names with backticks (`) if so.
  - Check [Built-in Functions](#) on page 257 to confirm whether Impala supports all the built-in functions being used by your query, and whether argument and return types are the same as you expect.
  - Review the [contents of the Impala logs](#) for any information that may be useful in identifying the source of the problem.
- If a query fails against Impala but not Hive, it is likely that there is a problem with your Impala installation.

## Impala Troubleshooting Quick Reference

The following table lists common problems and potential solutions.

Symptom	Explanation	Recommendation
Impala takes a long time to start.	Impala instances with large numbers of tables, partitions, or data files take longer to start because the metadata for these objects is broadcast to all <code>impalad</code> nodes and cached.	Adjust timeout and synchronicity settings.
Joins fail to complete.	There may be insufficient memory. During a join, data from the second, third, and so on sets to be joined is loaded into memory. If Impala chooses an inefficient join order or join mechanism, the query could exceed the total memory available.	Start by gathering statistics with the <code>COMPUTE STATS</code> statement for each table involved in the join. Consider specifying the <code>[SHUFFLE]</code> hint so that data from the joined tables is split up between nodes rather than broadcast to each node. If tuning at the SQL level is not sufficient, add more memory to your system or join smaller data sets.
Queries return incorrect results.	Impala metadata may be outdated after changes are performed in Hive.	Where possible, use the appropriate Impala statement ( <code>INSERT</code> , <code>LOAD DATA</code> , <code>CREATE TABLE</code> , <code>ALTER TABLE</code> , <code>COMPUTE STATS</code> , and so on) rather than switching back and forth between Impala and Hive. Impala automatically broadcasts the results of DDL and DML operations to all Impala nodes in the cluster, but does not automatically recognize when such changes are made through Hive. After inserting data, adding a partition, or other operation in Hive, refresh the metadata for the table as described in <a href="#">REFRESH Statement</a> on page 215.

## Troubleshooting Impala

Symptom	Explanation	Recommendation
Queries are slow to return results.	<p>Some <code>impalad</code> instances may not have started. Using a browser, connect to the host running the Impala state store. Connect using an address of the form <code>http://hostname:port/metrics</code>.</p> <p><b>Note:</b> Replace <code>hostname</code> and <code>port</code> with the hostname and port of your Impala state store host machine and web server port. The default port is 25010.</p> <p>The number of <code>impalad</code> instances listed should match the expected number of <code>impalad</code> instances installed in the cluster. There should also be one <code>impalad</code> instance installed on each DataNode</p>	Ensure Impala is installed on all DataNodes. Start any <code>impalad</code> instances that are not running.
Queries are slow to return results.	Impala may not be configured to use native checksumming. Native checksumming uses machine-specific instructions to compute checksums over HDFS data very quickly. Review Impala logs. If you find instances of "INFO util.NativeCodeLoader: Loaded the native-hadoop" messages, native checksumming is not enabled.	Ensure Impala is configured to use native checksumming as described in <a href="#">Post-Installation Configuration for Impala</a> on page 34.
Queries are slow to return results.	Impala may not be configured to use data locality tracking.	Test Impala for data locality tracking and make configuration changes as necessary. Information on this process can be found in <a href="#">Post-Installation Configuration for Impala</a> on page 34.
Attempts to complete Impala tasks such as executing INSERT-SELECT actions fail. The Impala logs include notes that files could not be opened due to permission denied.	This can be the result of permissions issues. For example, you could use the Hive shell as the hive user to create a table. After creating this table, you could attempt to complete some action, such as an INSERT-SELECT on the table. Because the table was created using one user and the INSERT-SELECT is attempted by another, this action may fail due to permissions issues.	In general, ensure the Impala user has sufficient permissions. In the preceding example, ensure the Impala user has sufficient permissions to the table that the Hive user created.
Impala fails to start up, with the <code>impalad</code> logs referring to errors connecting to the statestore	A large number of databases, tables, partitions, and so on can require metadata synchronization, particularly on startup, that takes longer than the default timeout for the statestore service.	Configure the statestore timeout value and possibly other settings related to the frequency of statestore updates and metadata loading. See <a href="#">Increasing the Statestore Timeout</a> on page 81 and <a href="#">Scalability Considerations for the Impala Statestore</a> on page 388.

Symptom	Explanation	Recommendation
service and attempts to re-register.		

## Impala Web User Interface for Debugging

The web user interface is primarily for problem diagnosis and troubleshooting. The items listed and their format is subject to change. To monitor Impala health, particularly across the entire cluster at once, use the Cloudera Manager interface.

### Debug Web UI for impalad

To debug and troubleshoot the `impalad` daemon using a web-based interface, open the URL `http://impala-server-hostname:25000/` in a browser. (For secure clusters, use the prefix `https://` instead of `http://`.) Because each Impala node produces its own set of debug information, choose a specific node that you are curious about or suspect is having problems.

- **Note:** To get a convenient picture of the health of all Impala nodes in a cluster, use the Cloudera Manager interface, which collects the low-level operational information from all Impala nodes, and presents a unified view of the entire cluster.

#### Main Page

By default, the main page of the debug web UI is at `http://impala-server-hostname:25000/` (non-secure cluster) or `https://impala-server-hostname:25000/` (secure cluster).

This page lists the version of the `impalad` daemon, plus basic hardware and software information about the corresponding host, such as information about the CPU, memory, disks, and operating system version.

#### Backends Page

By default, the **backends** page of the debug web UI is at `http://impala-server-hostname:25000/backends` (non-secure cluster) or `https://impala-server-hostname:25000/backends` (secure cluster).

This page lists the host and port info for each of the `impalad` nodes in the cluster. Because each `impalad` daemon knows about every other `impalad` daemon through the statestore, this information should be the same regardless of which node you select. Links take you to the corresponding debug web pages for any of the other nodes in the cluster.

#### Catalog Page

By default, the **catalog** page of the debug web UI is at `http://impala-server-hostname:25000/catalog` (non-secure cluster) or `https://impala-server-hostname:25000/catalog` (secure cluster).

This page displays a list of databases and associated tables recognized by this instance of `impalad`. You can use this page to locate which database a table is in, check the exact spelling of a database or table name, look for identical table names in multiple databases, and so on.

#### Logs Page

By default, the **logs** page of the debug web UI is at `http://impala-server-hostname:25000/logs` (non-secure cluster) or `https://impala-server-hostname:25000/logs` (secure cluster).

This page shows the last portion of the `impalad.INFO` log file, the most detailed of the info, warning, and error logs for the `impalad` daemon. You can refer here to see the details of the most recent operations, whether the

## Troubleshooting Impala

operations succeeded or encountered errors. This central page can be more convenient than looking around the filesystem for the log files, which could be in different locations on clusters that use Cloudera Manager or not.

### Memz Page

By default, the **memz** page of the debug web UI is at `http://impala-server-hostname:25000/memz` (non-secure cluster) or `https://impala-server-hostname:25000/memz` (secure cluster).

This page displays summary and detailed information about memory usage by the `impalad` daemon. You can see the memory limit in effect for the node, and how much of that memory Impala is currently using.

### Metrics Page

By default, the **metrics** page of the debug web UI is at `http://impala-server-hostname:25000/metrics` (non-secure cluster) or `https://impala-server-hostname:25000/metrics` (secure cluster).

This page displays the current set of metrics: counters and flags representing various aspects of `impalad` internal operation. For the meanings of these metrics, see [Impala Metrics](#) in the Cloudera Manager documentation.

### Queries Page

By default, the **queries** page of the debug web UI is at `http://impala-server-hostname:25000/queries` (non-secure cluster) or `https://impala-server-hostname:25000/queries` (secure cluster).

This page lists all currently running queries, plus any completed queries whose details still reside in memory. The queries are listed in reverse chronological order, with the most recent at the top. (You can control the amount of memory devoted to completed queries by specifying the `--query_log_size` startup option for `impalad`.)

On this page, you can see at a glance how many SQL statements are failing (state value of `EXCEPTION`), how large the result sets are (# rows fetched), and how long each statement took (Start Time and End Time).

Each query has an associated link that displays the detailed query profile, which you can examine to understand the performance characteristics of that query. See [Using the Query Profile for Performance Tuning](#) on page 380 for details.

### Sessions Page

By default, the **sessions** page of the debug web UI is at `http://impala-server-hostname:25000/sessions` (non-secure cluster) or `https://impala-server-hostname:25000/sessions` (secure cluster).

This page displays information about the sessions currently connected to this `impalad` instance. For example, sessions could include connections from the `impala-shell` command, JDBC or ODBC applications, or the Impala Query UI in the Hue web interface.

### Threadz Page

By default, the **threadz** page of the debug web UI is at `http://impala-server-hostname:25000/threadz` (non-secure cluster) or `https://impala-server-hostname:25000/threadz` (secure cluster).

This page displays information about the threads used by this instance of `impalad`, and shows which categories they are grouped into. Making use of this information requires substantial knowledge about Impala internals.

### Varz Page

By default, the **varz** page of the debug web UI is at `http://impala-server-hostname:25000/varz` (non-secure cluster) or `https://impala-server-hostname:25000/varz` (secure cluster).

This page shows the configuration settings in effect when this instance of `impalad` communicates with other Hadoop components such as HDFS and YARN. These settings are collected from a set of configuration files; Impala might not actually make use of all settings.

The bottom of this page also lists all the command-line settings in effect for this instance of `impalad`. See [Modifying Impala Startup Options](#) on page 48 for information about modifying these values.

## Ports Used by Impala

Impala uses the TCP ports listed in the following table. Before deploying Impala, ensure these ports are open on each system.

Component	Service	Port	Access Requirement	Comment
Impala Daemon	Impala Daemon Frontend Port	21000	External	Used to transmit commands and receive results by <code>impala-shell</code> , Beeswax, and version 1.2 of the Cloudera ODBC driver.
Impala Daemon	Impala Daemon Frontend Port	21050	External	Used to transmit commands and receive results by applications, such as Business Intelligence tools, using JDBC and version 2.0 or higher of the Cloudera ODBC driver.
Impala Daemon	Impala Daemon Backend Port	22000	Internal	Internal use only. Impala daemons use this port to communicate with each other.
Impala Daemon	StateStoreSubscriber Service Port	23000	Internal	Internal use only. Impala daemons listen on this port for updates from the statestore daemon.
Catalog Daemon	StateStoreSubscriber Service Port	23020	Internal	Internal use only. The catalog daemon listens on this port for updates from the statestore daemon.
Impala Daemon	Impala Daemon HTTP Server Port	25000	External	Impala web interface for administrators to monitor and troubleshoot.
Impala StateStore Daemon	StateStore HTTP Server Port	25010	External	StateStore web interface for administrators to monitor and troubleshoot.
Impala Catalog Daemon	Catalog HTTP Server Port	25020	External	Catalog service web interface for administrators to monitor and troubleshoot. New in Impala 1.2 and higher.
Impala StateStore Daemon	StateStore Service Port	24000	Internal	Internal use only. The statestore daemon listens on this port for registration/unregistration requests.

## Ports Used by Impala

Component	Service	Port	Access Requirement	Comment
Impala Catalog Daemon	StateStore Service Port	26000	Internal	Internal use only. The catalog service uses this port to communicate with the Impala daemons. New in Impala 1.2 and higher.
Impala Daemon	Llama Callback Port	28000	Internal	Internal use only. Impala daemons use to communicate with Llama. New in CDH 5.0.0 and higher.
Impala Llama ApplicationMaster	Llama Thrift Admin Port	15002	Internal	Internal use only. New in CDH 5.0.0 and higher.
Impala Llama ApplicationMaster	Llama Thrift Port	15000	Internal	Internal use only. New in CDH 5.0.0 and higher.
Impala Llama ApplicationMaster	Llama HTTP Port	15001	External	Llama service web interface for administrators to monitor and troubleshoot. New in CDH 5.0.0 and higher.

# Impala Reserved Words

The following are the reserved words for the current release of Cloudera Impala. A reserved word is one that cannot be used directly as an identifier; you must quote it with backticks. For example, a statement `CREATE TABLE select (x INT)` fails, while `CREATE TABLE `select` (x INT)` succeeds. Impala does not reserve the names of aggregate or scalar built-in functions. (Formerly, Impala did reserve the names of some aggregate functions.)

Because different database systems have different sets of reserved words, and the reserved words change from release to release, carefully consider database, table, and column names to ensure maximum compatibility between products and versions.

```
add
aggregate
all
alter
analytic
and
anti
api_version
as
asc
avro
between
bigint
binary
boolean
by
cached
case
cast
change
char
class
close_fn
column
columns
comment
compute
create
cross
current
data
database
databases
date
datetime
decimal
delimited
desc
describe
distinct
div
double
drop
else
end
escaped
exists
explain
external
false
fields
fileformat
finalize_fn
first
float
following
```

## Impala Reserved Words

```
for  
format  
formatted  
from  
full  
function  
functions  
grant  
group  
having  
if  
in  
incremental  
init_fn  
inner  
inpath  
insert  
int  
integer  
intermediate  
interval  
into  
invalidate  
is  
join  
last  
left  
like  
limit  
lines  
load  
location  
merge_fn  
metadata  
not  
null  
nulls  
offset  
on  
or  
order  
outer  
over  
overwrite  
parquet  
parquetfile  
partition  
partitioned  
partitions  
preceding  
prepare_fn  
produced  
range  
rcfile  
real  
refresh  
regexp  
rename  
replace  
returns  
revoke  
right  
rlike  
role  
roles  
row  
rows  
schema  
schemas  
select  
semi  
sequencefile  
serdeproperties
```

```
serialize_fn  
set  
show  
smallint  
stats  
stored  
straight_join  
string  
symbol  
table  
tables  
tblproperties  
terminated  
textfile  
then  
timestamp  
tinyint  
to  
true  
unbounded  
uncached  
union  
update_fn  
use  
using  
values  
varchar  
view  
when  
where  
with
```

# Cloudera Impala Frequently Asked Questions

Here are the categories of frequently asked questions for the Cloudera Impala product.

## Trying Impala

### How do I try Cloudera Impala out?

To look at the core features and functionality on Impala, the easiest way to try out Impala is to download the Cloudera QuickStart VM and start the Impala service through Cloudera Manager, then use `impala-shell` in a terminal window or the Impala Query UI in the Hue web interface.

To do performance testing and try out the management features for Impala on a cluster, you need to move beyond the QuickStart VM with its virtualized single-node environment. Ideally, download the Cloudera Manager software to set up the cluster, then install the Impala software through Cloudera Manager.

### Does Cloudera offer a VM for demonstrating Impala?

Cloudera offers a demonstration VM called the QuickStart VM, available in VMWare, VirtualBox, and KVM formats. For more information, see [the Cloudera QuickStart VM](#). After booting the QuickStart VM, many services are turned off by default; in the Cloudera Manager UI that appears automatically, turn on Impala and any other components that you want to try out.

### Where can I find Impala documentation?

Starting with Impala 1.3.0, Impala documentation is integrated with the CDH 5 documentation, in addition to the standalone Impala documentation for use with CDH 4. For CDH 5, the core Impala developer and administrator information remains in the associated [Impala documentation](#) portion. Information about Impala release notes, installation, configuration, startup, and security is embedded in the corresponding CDH 5 guides.

- [New features](#)
- [Known and fixed issues](#)
- [Incompatible changes](#)
- [Installing Impala](#)
- [Upgrading Impala](#)
- [Configuring Impala](#)
- [Starting Impala](#)
- [Security for Impala](#)
- [CDH Version and Packaging Information](#)

Information about the latest CDH 4-compatible Impala release remains at the [Impala for CDH 4 Documentation](#) page.

### Where can I get more information about Impala?

More product information is available here:

- O'Reilly introductory e-book: [Cloudera Impala: Bringing the SQL and Hadoop Worlds Together](#)
- O'Reilly getting started guide for developers: [Getting Started with Impala: Interactive SQL for Apache Hadoop](#)
- Blog: [Cloudera Impala: Real-Time Queries in Apache Hadoop, For Real](#)
- Webinar: [Introduction to Impala](#)
- Product website page: [Cloudera Enterprise RTQ](#)

To see the latest release announcements for Impala, see the [Cloudera Announcements](#) forum.

### How can I ask questions and provide feedback about Impala?

- Join the [Impala discussion forum](#) and the [Impala mailing list](#) to ask questions and provide feedback.
- Use the [Impala Jira project](#) to log bug reports and requests for features.

### Where can I get sample data to try?

You can get scripts that produce data files and set up an environment for TPC-DS style benchmark tests from [this Github repository](#). In addition to being useful for experimenting with performance, the tables are suited to experimenting with many aspects of SQL on Impala: they contain a good mixture of data types, data distributions, partitioning, and relational data suitable for join queries.

## Impala System Requirements

### What are the software and hardware requirements for running Impala?

For information on Impala requirements, see [Cloudera Impala Requirements](#) on page 23. Note that there is often a minimum required level of Cloudera Manager for any given Impala version.

### How much memory is required?

Although Impala is not an in-memory database, when dealing with large tables and large result sets, you should expect to dedicate a substantial portion of physical memory for the `impalad` daemon. Recommended physical memory for an Impala node is 128 GB or higher. If practical, devote approximately 80% of physical memory to Impala.

The amount of memory required for an Impala operation depends on several factors:

- The file format of the table. Different file formats represent the same data in more or fewer data files. The compression and encoding for each file format might require a different amount of temporary memory to decompress the data for analysis.
- Whether the operation is a `SELECT` or an `INSERT`. For example, Parquet tables require relatively little memory to query, because Impala reads and decompresses data in 8 MB chunks. Inserting into a Parquet table is a more memory-intensive operation because the data for each data file (potentially hundreds of megabytes, depending on the value of the `PARQUET_FILE_SIZE` query option) is stored in memory until encoded, compressed, and written to disk.
- Whether the table is partitioned or not, and whether a query against a partitioned table can take advantage of partition pruning.
- Whether the final result set is sorted by the `ORDER BY` clause. Each Impala node scans and filters a portion of the total data, and applies the `LIMIT` to its own portion of the result set. In Impala 1.4.0 and higher, if the sort operation requires more memory than is available on any particular host, Impala uses a temporary disk work area to perform the sort. The intermediate result sets are all sent back to the coordinator node, which does the final sorting and then applies the `LIMIT` clause to the final result set.

For example, if you execute the query:

```
select * from giant_table order by some_column limit 1000;
```

and your cluster has 50 nodes, then each of those 50 nodes will transmit a maximum of 1000 rows back to the coordinator node. The coordinator node needs enough memory to sort (`LIMIT * cluster_size`) rows, although in the end the final result set is at most `LIMIT` rows, 1000 in this case.

Likewise, if you execute the query:

```
select * from giant_table where test_val > 100 order by some_column;
```

## Cloudera Impala Frequently Asked Questions

then each node filters out a set of rows matching the `WHERE` conditions, sorts the results (with no size limit), and sends the sorted intermediate rows back to the coordinator node. The coordinator node might need substantial memory to sort the final result set, and so might use a temporary disk work area for that final phase of the query.

- Whether the query contains any join clauses, `GROUP BY` clauses, analytic functions, or `DISTINCT` operators. These operations all require some in-memory work areas that vary depending on the volume and distribution of data. In Impala 2.0 and later, these kinds of operations utilize temporary disk work areas if memory usage grows too large to handle. See [SQL Operations that Spill to Disk](#) on page 389 for details.
- The size of the result set. When intermediate results are being passed around between nodes, the amount of data depends on the number of columns returned by the query. For example, it is more memory-efficient to query only the columns that are actually needed in the result set rather than always issuing `SELECT *`.
- The mechanism by which work is divided for a join query. You use the `COMPUTE STATS` statement, and query hints in the most difficult cases, to help Impala pick the most efficient execution plan. See [Performance Considerations for Join Queries](#) on page 357 for details.

See [Hardware Requirements](#) on page 24 for more details and recommendations about Impala hardware prerequisites.

### What processor type and speed does Cloudera recommend?

Impala makes use of SSE 4.1 instructions.

### What EC2 instances are recommended for Impala?

For large storage capacity and large I/O bandwidth, consider the `hs1.8xlarge` and `cc2.8xlarge` instance types. Impala I/O patterns typically do not benefit enough from SSD storage to make up for the lower overall size. For performance and security considerations for deploying CDH and its components on AWS, see [Cloudera Enterprise Reference Architecture for AWS Deployments](#).

## Supported and Unsupported Functionality In Impala

### What are the main features of Impala?

- A large set of SQL statements, including `SELECT` and `INSERT`, with [joins](#), [Subqueries](#) on page 235, and [Impala Analytic Functions](#) on page 296. Highly compatible with HiveQL, and also including some vendor extensions. For more information, see [Impala SQL Language Reference](#) on page 111.
- Distributed, high-performance queries. See [Tuning Impala for Performance](#) on page 354 for information about Impala performance optimizations and tuning techniques for queries.
- Using Cloudera Manager, you can deploy and manage your Impala services. Cloudera Manager is the best way to get started with Impala on your cluster.
- Using Hue for queries.
- Appending and inserting data into tables through the `INSERT` statement. See [How Impala Works with Hadoop File Formats](#) on page 399 for the details about which operations are supported for which file formats.
- ODBC: Impala is certified to run against MicroStrategy and Tableau, with restrictions. For more information, see [Configuring Impala to Work with ODBC](#) on page 37.
- Querying data stored in HDFS and HBase in a single query. See [Using Impala to Query HBase Tables](#) on page 426 for details.
- In Impala 2.2.0 and higher, querying data stored in the Amazon Simple Storage Service (S3). See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details.
- Concurrent client requests. Each Impala daemon can handle multiple concurrent client requests. The effects on performance depend on your particular hardware and workload.
- Kerberos authentication. For more information, see [Impala Security](#) on page 87.

- Partitions. With Impala SQL, you can create partitioned tables with the `CREATE TABLE` statement, and add and drop partitions with the `ALTER TABLE` statement. Impala also takes advantage of the partitioning present in Hive tables. See [Partitioning for Impala Tables](#) on page 394 for details.

#### What features from relational databases or Hive are not available in Impala?

- Querying streaming data.
- Deleting individual rows. You delete data in bulk by overwriting an entire table or partition, or by dropping a table.
- Indexing (not currently). LZO-compressed text files can be indexed outside of Impala, as described in [Using LZO-Compressed Text Files](#) on page 404.
- Full text search on text fields. The Cloudera Search product is appropriate for this use case.
- Custom Hive Serializer/Deserializer classes (SerDes). Impala supports a set of common native file formats that have built-in SerDes in CDH. See [How Impala Works with Hadoop File Formats](#) on page 399 for details.
- Checkpointing within a query. That is, Impala does not save intermediate results to disk during long-running queries. Currently, Impala cancels a running query if any host on which that query is executing fails. When one or more hosts are down, Impala reroutes future queries to only use the available hosts, and Impala detects when the hosts come back up and begins using them again. Because a query can be submitted through any Impala node, there is no single point of failure. In the future, we will consider adding additional work allocation features to Impala, so that a running query would complete even in the presence of host failures.
- Encryption of data transmitted between Impala daemons.
- Hive indexes.
- Non-Hadoop data stores, such as relational databases.

For the detailed list of features that are different between Impala and HiveQL, see [SQL Differences Between Impala and Hive](#) on page 323.

#### Does Impala support generic JDBC?

Impala supports the HiveServer2 JDBC driver.

#### Is Avro supported?

Yes, Avro is supported. Impala has always been able to query Avro tables. You can use the Impala `LOAD DATA` statement to load existing Avro data files into a table. Starting with Impala 1.4, you can create Avro tables with Impala. Currently, you still use the `INSERT` statement in Hive to copy data from another table into an Avro table. See [Using the Avro File Format with Impala Tables](#) on page 417 for details.

## How do I?

#### How do I prevent users from seeing the text of SQL queries?

For instructions on making the Impala log files unreadable by unprivileged users, see [Securing Impala Data and Log Files](#) on page 88.

For instructions on password-protecting the web interface to the Impala log files and other internal server information, see [Securing the Impala Web User Interface](#) on page 89.

In Impala 2.2 / CDH 5.4 and higher, you can use the log redaction feature to obfuscate sensitive information in Impala log files. See [Sensitive Data Redaction](#) for details.

## Cloudera Impala Frequently Asked Questions

### How do I know how many Impala nodes are in my cluster?

The Impala statestore keeps track of how many `impalad` nodes are currently available. You can see this information through the statestore web interface. For example, at the URL

`http://statestore_host:25010/metrics` you might see lines like the following:

```
statestore.live-backends:3
statestore.live-backends.list:[host1:22000, host1:26000, host2:22000]
```

The number of `impalad` nodes is the number of list items referring to port 22000, in this case two. (Typically, this number is one less than the number reported by the `statestore.live-backends` line.) If an `impalad` node became unavailable or came back after an outage, the information reported on this page would change appropriately.

## Impala Performance

### Are results returned as they become available, or all at once when a query completes?

Impala streams results whenever they are available, when possible. Certain SQL operations (aggregation or `ORDER BY`) require all of the input to be ready before Impala can return results.

### Why does my query run slowly?

There are many possible reasons why a given query could be slow. Use the following checklist to diagnose performance issues with existing queries, and to avoid such issues when writing new queries, setting up new nodes, creating new tables, or loading data.

- Immediately after the query finishes, issue a `SUMMARY` command in `impala-shell`. You can check which phases of execution took the longest, and compare estimated values for memory usage and number of rows with the actual values.
- Immediately after the query finishes, issue a `PROFILE` command in `impala-shell`. The numbers in the `BytesRead`, `BytesReadLocal`, and `BytesReadShortCircuit` should be identical for a specific node. For example:

```
- BytesRead: 180.33 MB
- BytesReadLocal: 180.33 MB
- BytesReadShortCircuit: 180.33 MB
```

If `BytesReadLocal` is lower than `BytesRead`, something in your cluster is misconfigured, such as the `impalad` daemon not running on all the data nodes. If `BytesReadShortCircuit` is lower than `BytesRead`, short-circuit reads are not enabled properly on that node; see [Post-Installation Configuration for Impala](#) on page 34 for instructions.

- If the table was just created, or this is the first query that accessed the table after an `INVALIDATE METADATA` statement or after the `impalad` daemon was restarted, there might be a one-time delay while the metadata for the table is loaded and cached. Check whether the slowdown disappears when the query is run again. When doing performance comparisons, consider issuing a `DESCRIBE table_name` statement for each table first, to make sure any timings only measure the actual query time and not the one-time wait to load the table metadata.
- Is the table data in uncompressed text format? Check by issuing a `DESCRIBE FORMATTED table_name` statement. A text table is indicated by the line:

```
InputFormat: org.apache.hadoop.mapred.TextInputFormat
```

Although uncompressed text is the default format for a `CREATE TABLE` statement with no `STORED AS` clauses, it is also the bulkiest format for disk storage and consequently usually the slowest format for queries. For data where query performance is crucial, particularly for tables that are frequently queried, consider starting

with or converting to a compact binary file format such as Parquet, Avro, RCFile, or SequenceFile. For details, see [How Impala Works with Hadoop File Formats](#) on page 399.

- If your table has many columns, but the query refers to only a few columns, consider using the Parquet file format. Its data files are organized with a column-oriented layout that lets queries minimize the amount of I/O needed to retrieve, filter, and aggregate the values for specific columns. See [Using the Parquet File Format with Impala Tables](#) on page 407 for details.
- If your query involves any joins, are the tables in the query ordered so that the tables or subqueries are ordered with the one returning the largest number of rows on the left, followed by the smallest (most selective), the second smallest, and so on? That ordering allows Impala to optimize the way work is distributed among the nodes and how intermediate results are routed from one node to another. For example, all other things being equal, the following join order results in an efficient query:

```
select some_col from
  huge_table join big_table join small_table join medium_table
  where
    huge_table.id = big_table.id
    and big_table.id = medium_table.id
    and medium_table.id = small_table.id;
```

See [Performance Considerations for Join Queries](#) on page 357 for performance tips for join queries.

- Also for join queries, do you have table statistics for the table, and column statistics for the columns used in the join clauses? Column statistics let Impala better choose how to distribute the work for the various pieces of a join query. See [How Impala Uses Statistics for Query Optimization](#) on page 363 for details about gathering statistics.
- Does your table consist of many small data files? Impala works most efficiently with data files in the multi-megabyte range; Parquet, a format optimized for data warehouse-style queries, uses large files (originally 1 GB, now 256 MB in Impala 2.0 and higher) with a block size matching the file size. Use the DESCRIBE FORMATTED *table\_name* statement in `impala-shell` to see where the data for a table is located, and use the `hadoop fs -ls` or `hdfs dfs -ls` Unix commands to see the files and their sizes. If you have thousands of small data files, that is a signal that you should consolidate into a smaller number of large files. Use an `INSERT ... SELECT` statement to copy the data to a new table, reorganizing into new data files as part of the process. Prefer to construct large data files and import them in bulk through the `LOAD DATA` or `CREATE EXTERNAL TABLE` statements, rather than issuing many `INSERT ... VALUES` statements; each `INSERT ... VALUES` statement creates a separate tiny data file. If you have thousands of files all in the same directory, but each one is megabytes in size, consider using a partitioned table so that each partition contains a smaller number of files. See the following point for more on partitioning.
- If your data is easy to group according to time or geographic region, have you partitioned your table based on the corresponding columns such as `YEAR`, `MONTH`, and/or `DAY`? Partitioning a table based on certain columns allows queries that filter based on those same columns to avoid reading the data files for irrelevant years, postal codes, and so on. (Do not partition down to too fine a level; try to structure the partitions so that there is still sufficient data in each one to take advantage of the multi-megabyte HDFS block size.) See [Partitioning for Impala Tables](#) on page 394 for details.

## Why does my SELECT statement fail?

When a `SELECT` statement fails, the cause usually falls into one of the following categories:

- A timeout because of a performance, capacity, or network issue affecting one particular node.
- Excessive memory use for a join query, resulting in automatic cancellation of the query.
- A low-level issue affecting how native code is generated on each node to handle particular `WHERE` clauses in the query. For example, a machine instruction could be generated that is not supported by the processor of a certain node. If the error message in the log suggests the cause was an illegal instruction, consider turning off native code generation temporarily, and trying the query again.
- Malformed input data, such as a text data file with an enormously long line, or with a delimiter that does not match the character specified in the `FIELDS TERMINATED BY` clause of the `CREATE TABLE` statement.

## Cloudera Impala Frequently Asked Questions

### Why does my INSERT statement fail?

When an `INSERT` statement fails, it is usually the result of exceeding some limit within a Hadoop component, typically HDFS.

- An `INSERT` into a partitioned table can be a strenuous operation due to the possibility of opening many files and associated threads simultaneously in HDFS. Impala 1.1.1 includes some improvements to distribute the work more efficiently, so that the values for each partition are written by a single node, rather than as a separate data file from each node.
- Certain expressions in the `SELECT` part of the `INSERT` statement can complicate the execution planning and result in an inefficient `INSERT` operation. Try to make the column data types of the source and destination tables match up, for example by doing `ALTER TABLE ... REPLACE COLUMNS` on the source table if necessary. Try to avoid `CASE` expressions in the `SELECT` portion, because they make the result values harder to predict than transferring a column unchanged or passing the column through a built-in function.
- Be prepared to raise some limits in the HDFS configuration settings, either temporarily during the `INSERT` or permanently if you frequently run such `INSERT` statements as part of your ETL pipeline.
- The resource usage of an `INSERT` statement can vary depending on the file format of the destination table. Inserting into a Parquet table is memory-intensive, because the data for each partition is buffered in memory until it reaches 1 gigabyte, at which point the data file is written to disk. Impala can distribute the work for an `INSERT` more efficiently when statistics are available for the source table that is queried during the `INSERT` statement. See [How Impala Uses Statistics for Query Optimization](#) on page 363 for details about gathering statistics.

### Does Impala performance improve as it is deployed to more hosts in a cluster in much the same way that Hadoop performance does?

Yes. Impala scales with the number of hosts. It is important to install Impala on all the data nodes in the cluster, because otherwise some of the nodes must do remote reads to retrieve data not available for local reads. Data locality is an important architectural aspect for Impala performance. See [this Impala performance blog post](#) for background. Note that this blog post refers to benchmarks with Impala 1.1.1; Impala has added even more performance features in the 1.2.x series.

### Is the HDFS block size reduced to achieve faster query results?

No. Impala does not make any changes to the HDFS or HBase data sets.

The default Parquet block size is relatively large (256 MB in Impala 2.0 and later; 1 GB in earlier releases). You can control the block size when creating Parquet files using the `PARQUET_FILE_SIZE` query option.

### Does Impala use caching?

Impala does not cache table data. It does cache some table and file metadata. Although queries might run faster on subsequent iterations because the data set was cached in the OS buffer cache, Impala does not explicitly control this.

Impala takes advantage of the HDFS caching feature in CDH 5. You can designate which tables or partitions are cached through the `CACHED` and `UNCACHED` clauses of the `CREATE TABLE` and `ALTER TABLE` statements. Impala can also take advantage of data that is pinned in the HDFS cache through the `hdfscacheadmin` command. See [Using HDFS Caching with Impala \(CDH 5.1 or higher only\)](#) on page 372 for details.

## Impala Use Cases

### What are good use cases for Impala as opposed to Hive or MapReduce?

Impala is well-suited to executing SQL queries for interactive exploratory analytics on large data sets. Hive and MapReduce are appropriate for very long running, batch-oriented tasks such as ETL.

### [Is MapReduce required for Impala? Will Impala continue to work as expected if MapReduce is stopped?](#)

Impala does not use MapReduce at all.

### [Can Impala be used for complex event processing?](#)

For example, in an industrial environment, many agents may generate large amounts of data. Can Impala be used to analyze this data, checking for notable changes in the environment?

Complex Event Processing (CEP) is usually performed by dedicated stream-processing systems. Impala is not a stream-processing system, as it most closely resembles a relational database.

### [Is Impala intended to handle real time queries in low-latency applications or is it for ad hoc queries for the purpose of data exploration?](#)

Ad-hoc queries are the primary use case for Impala. We anticipate it being used in many other situations where low-latency is required. Whether Impala is appropriate for any particular use-case depends on the workload, data size and query volume. See [Impala Benefits](#) on page 16 for the primary benefits you can expect when using Impala.

## Questions about Impala And Hive

### [How does Impala compare to Hive and Pig?](#)

Impala is different from Hive and Pig because it uses its own daemons that are spread across the cluster for queries. Because Impala does not rely on MapReduce, it avoids the startup overhead of MapReduce jobs, allowing Impala to return results in real time.

### [Can I do transforms or add new functionality?](#)

Impala adds support for UDFs in Impala 1.2. You can write your own functions in C++, or reuse existing Java-based Hive UDFs. The UDF support includes scalar functions and user-defined aggregate functions (UDAs). User-defined table functions (UDTFs) are not currently supported.

Impala does not currently support an extensible serialization-deserialization framework (SerDes), and so adding extra functionality to Impala is not as straightforward as for Hive or Pig.

### [Can any Impala query also be executed in Hive?](#)

Yes. There are some minor differences in how some queries are handled, but Impala queries can also be completed in Hive. Impala SQL is a subset of HiveQL, with some functional limitations such as transforms. For details of the Impala SQL dialect, see [SQL Statements](#) on page 162. For the Impala built-in functions, see [Built-in Functions](#) on page 257. For the detailed list of differences between Impala and HiveQL, see [SQL Differences Between Impala and Hive](#) on page 323.

### [Can I use Impala to query data already loaded into Hive and HBase?](#)

There are no additional steps to allow Impala to query tables managed by Hive, whether they are stored in HDFS or HBase. Make sure that Impala is configured to access the Hive metastore correctly and you should be ready to go. Keep in mind that `impalad`, by default, runs as the `impala` user, so you might need to adjust some file permissions depending on how strict your permissions are currently.

See [Using Impala to Query HBase Tables](#) on page 426 for details about querying data in HBase.

### [Is Hive an Impala requirement?](#)

The Hive metastore service is a requirement. Impala shares the same metastore database as Hive, allowing Impala and Hive to access the same tables transparently.

## Cloudera Impala Frequently Asked Questions

Hive itself is optional, and does not need to be installed on the same nodes as Impala. Currently, Impala supports a wider variety of read (query) operations than write (insert) operations; you use Hive to insert data into tables that use certain file formats. See [How Impala Works with Hadoop File Formats](#) on page 399 for details.

## Impala Availability

### Is Impala production ready?

Impala has finished its beta release cycle, and the 1.0, 1.1, and 1.2 GA releases are production ready. The 1.1.x series includes additional security features for authorization, an important requirement for production use in many organizations. The 1.2.x series includes important performance features, particularly for large join queries. Some Cloudera customers are already using Impala for large workloads.

The Impala 1.3.0 and higher releases are bundled with corresponding levels of CDH 5. The number of new features grows with each release. See [New Features in Impala](#) on page 470 for a full list.

### How do I configure Hadoop high availability (HA) for Impala?

You can set up a proxy server to relay requests back and forth to the Impala servers, for load balancing and high availability. See [Using Impala through a Proxy for High Availability](#) on page 82 for details.

You can enable HDFS HA for the Hive metastore. See the [CDH5 High Availability Guide](#) or the [CDH4 High Availability Guide](#) for details.

### What happens if there is an error in Impala?

There is not a single point of failure in Impala. All Impala daemons are fully able to handle incoming queries. If a machine fails however, all queries with fragments running on that machine will fail. Because queries are expected to return quickly, you can just rerun the query if there is a failure. See [Impala Concepts and Architecture](#) on page 18 for details about the Impala architecture.

The longer answer: Impala must be able to connect to the Hive metastore. Impala aggressively caches metadata so the metastore host should have minimal load. Impala relies on the HDFS NameNode, and, in CDH4, you can configure HA for HDFS. Impala also has centralized services, known as the [statestore](#) and [catalog](#) services, that run on one host only. Impala continues to execute queries if the statestore host is down, but it will not get state updates. For example, if a host is added to the cluster while the statestore host is down, the existing instances of `impalad` running on the other hosts will not find out about this new host. Once the statestore process is restarted, all the information it serves is automatically reconstructed from all running Impala daemons.

### What is the maximum number of rows in a table?

There is no defined maximum. Some customers have used Impala to query a table with over a trillion rows.

### Can Impala and MapReduce jobs run on the same cluster without resource contention?

Yes. See [Controlling Impala Resource Usage](#) on page 372 for how to control Impala resource usage using the Linux cgroup mechanism, and [Integrated Resource Management with YARN](#) on page 78 for how to use Impala with the YARN resource management framework. Impala is designed to run on the DataNode hosts. Any contention depends mostly on the cluster setup and workload.

For a detailed example of configuring a cluster to share resources between Impala queries and MapReduce jobs, see [Setting up a Multi-tenant Cluster for Impala and MapReduce](#)

## Impala Internals

### [On which hosts does Impala run?](#)

Cloudera strongly recommends running the `impalad` daemon on each DataNode for good performance. Although this topology is not a hard requirement, if there are data blocks with no Impala daemons running on any of the hosts containing replicas of those blocks, queries involving that data could be very inefficient. In that case, the data must be transmitted from one host to another for processing by “remote reads”, a condition Impala normally tries to avoid. See [Impala Concepts and Architecture](#) on page 18 for details about the Impala architecture. Impala schedules query fragments on all hosts holding data relevant to the query, if possible.

### [How are joins performed in Impala?](#)

By default, Impala automatically determines the most efficient order in which to join tables using a cost-based method, based on their overall size and number of rows. (This is a new feature in Impala 1.2.2 and higher.) The `COMPUTE STATS` statement gathers information about each table that is crucial for efficient join performance. Impala chooses between two techniques for join queries, known as “broadcast joins” and “partitioned joins”. See [Joins](#) on page 220 for syntax details and [Performance Considerations for Join Queries](#) on page 357 for performance considerations.

### [How does Impala process join queries for large tables?](#)

Impala utilizes multiple strategies to allow joins between tables and result sets of various sizes. When joining a large table with a small one, the data from the small table is transmitted to each node for intermediate processing. When joining two large tables, the data from one of the tables is divided into pieces, and each node processes only selected pieces. See [Joins](#) on page 220 for details about join processing, [Performance Considerations for Join Queries](#) on page 357 for performance considerations, and [Hints](#) on page 239 for how to fine-tune the join strategy.

### [What is Impala's aggregation strategy?](#)

Impala currently only supports in-memory hash aggregation. In Impala 2.0 and higher, if the memory requirements for a join or aggregation operation exceed the memory limit for a particular host, Impala uses a temporary work area on disk to help the query complete successfully.

### [How is Impala metadata managed?](#)

Impala uses two pieces of metadata: the catalog information from the Hive metastore and the file metadata from the NameNode. Currently, this metadata is lazily populated and cached when an `impalad` needs it to plan a query.

The `REFRESH` statement updates the metadata for a particular table after loading new data through Hive. The [INVALIDATE METADATA Statement](#) on page 210 statement refreshes all metadata, so that Impala recognizes new tables or other DDL and DML changes performed through Hive.

In Impala 1.2 and higher, a dedicated `catalogd` daemon broadcasts metadata changes due to Impala DDL or DML statements to all nodes, reducing or eliminating the need to use the `REFRESH` and `INVALIDATE_METADATA` statements.

### [What load do concurrent queries produce on the NameNode?](#)

The load Impala generates is very similar to MapReduce. Impala contacts the NameNode during the planning phase to get the file metadata (this is only run on the host the query was sent to). Every `impalad` will read files as part of normal processing of the query.

## Cloudera Impala Frequently Asked Questions

### How does Impala achieve its performance improvements?

These are the main factors in the performance of Impala versus that of other Hadoop components and related technologies.

Impala avoids MapReduce. While MapReduce is a great general parallel processing model with many benefits, it is not designed to execute SQL. Impala avoids the inefficiencies of MapReduce in these ways:

- Impala does not materialize intermediate results to disk. SQL queries often map to multiple MapReduce jobs with all intermediate data sets written to disk.
- Impala avoids MapReduce start-up time. For interactive queries, the MapReduce start-up time becomes very noticeable. Impala runs as a service and essentially has no start-up time.
- Impala can more naturally disperse query plans instead of having to fit them into a pipeline of map and reduce jobs. This enables Impala to parallelize multiple stages of a query and avoid overheads such as sort and shuffle when unnecessary.

Impala uses a more efficient execution engine by taking advantage of modern hardware and technologies:

- Impala generates runtime code. Impala uses LLVM to generate assembly code for the query that is being run. Individual queries do not have to pay the overhead of running on a system that needs to be able to execute arbitrary queries.
- Impala uses available hardware instructions when possible. Impala uses the supplemental SSE3 (SSSE3) instructions which can offer tremendous speedups in some cases. (Impala 2.0 and 2.1 required the SSE4.1 instruction set; Impala 2.2 and higher relax the restriction again so only SSSE3 is required.)
- Impala uses better I/O scheduling. Impala is aware of the disk location of blocks and is able to schedule the order to process blocks to keep all disks busy.
- Impala is designed for performance. A lot of time has been spent in designing Impala with sound performance-oriented fundamentals, such as tight inner loops, inlined function calls, minimal branching, better use of cache, and minimal memory usage.

### What happens when the data set exceeds available memory?

Currently, if the memory required to process intermediate results on a node exceeds the amount available to Impala on that node, the query is cancelled. You can adjust the memory available to Impala on each node, and you can fine-tune the join strategy to reduce the memory required for the biggest queries. We do plan on supporting external joins and sorting in the future.

Keep in mind though that the memory usage is not directly based on the input data set size. For aggregations, the memory usage is the number of rows *after* grouping. For joins, the memory usage is the combined size of the tables *excluding* the biggest table, and Impala can use join strategies that divide up large joined tables among the various nodes rather than transmitting the entire table to each node.

### What are the most memory-intensive operations?

If a query fails with an error indicating “memory limit exceeded”, you might suspect a memory leak. The problem could actually be a query that is structured in a way that causes Impala to allocate more memory than you expect, exceeded the memory allocated for Impala on a particular node. Some examples of query or table structures that are especially memory-intensive are:

- `INSERT` statements using dynamic partitioning, into a table with many different partitions. (Particularly for tables using Parquet format, where the data for each partition is held in memory until it reaches the full block size in size before it is written to disk.) Consider breaking up such operations into several different `INSERT` statements, for example to load data one year at a time rather than for all years at once.
- `GROUP BY` on a unique or high-cardinality column. Impala allocates some handler structures for each different value in a `GROUP BY` query. Having millions of different `GROUP BY` values could exceed the memory limit.
- Queries involving very wide tables, with thousands of columns, particularly with many `STRING` columns. Because Impala allows a `STRING` value to be up to 32 KB, the intermediate results during such queries could require substantial memory allocation.

## When does Impala hold on to or return memory?

Impala allocates memory using `tcmalloc`, a memory allocator that is optimized for high concurrency. Once Impala allocates memory, it keeps that memory reserved to use for future queries. Thus, it is normal for Impala to show high memory usage when idle. If Impala detects that it is about to exceed its memory limit (defined by the `-mem_limit` startup option or the `MEM_LIMIT` query option), it deallocates memory not needed by the current queries.

When issuing queries through the JDBC or ODBC interfaces, make sure to call the appropriate close method afterwards. Otherwise, some memory associated with the query is not freed.

## SQL

### Is there an UPDATE statement?

Impala does not currently have an `UPDATE` statement, which would typically be used to change a single row, a small group of rows, or a specific column. The HDFS-based files used by typical Impala queries are optimized for bulk operations across many megabytes of data at a time, making traditional `UPDATE` operations inefficient or impractical.

You can use the following techniques to achieve the same goals as the familiar `UPDATE` statement, in a way that preserves efficient file layouts for subsequent queries:

- Replace the entire contents of a table or partition with updated data that you have already staged in a different location, either using `INSERT OVERWRITE`, `LOAD DATA`, or manual HDFS file operations followed by a `REFRESH` statement for the table. Optionally, you can use built-in functions and expressions in the `INSERT` statement to transform the copied data in the same way you would normally do in an `UPDATE` statement, for example to turn a mixed-case string into all uppercase or all lowercase.
- To update a single row, use an HBase table, and issue an `INSERT ... VALUES` statement using the same key as the original row. Because HBase handles duplicate keys by only returning the latest row with a particular key value, the newly inserted row effectively hides the previous one.

### Can Impala do user-defined functions (UDFs)?

Impala 1.2 and higher does support UDFs and UDAFs. You can either write native Impala UDFs and UDAFs in C++, or reuse UDFs (but not UDAFs) originally written in Java for use with Hive. See [Impala User-Defined Functions \(UDFs\)](#) on page 308 for details.

### Why do I have to use REFRESH and INVALIDATE METADATA, what do they do?

In Impala 1.2 and higher, there is much less need to use the `REFRESH` and `INVALIDATE METADATA` statements:

- The new `impala-catalog` service, represented by the `catalogd` daemon, broadcasts the results of Impala DDL statements to all Impala nodes. Thus, if you do a `CREATE TABLE` statement in Impala while connected to one node, you do not need to do `INVALIDATE METADATA` before issuing queries through a different node.
- The catalog service only recognizes changes made through Impala, so you must still issue a `REFRESH` statement if you load data through Hive or by manipulating files in HDFS, and you must issue an `INVALIDATE METADATA` statement if you create a table, alter a table, add or drop partitions, or do other DDL statements in Hive.
- Because the catalog service broadcasts the results of `REFRESH` and `INVALIDATE METADATA` statements to all nodes, in the cases where you do still need to issue those statements, you can do that on a single node rather than on every node, and the changes will be automatically recognized across the cluster, making it more convenient to load balance by issuing queries through arbitrary Impala nodes rather than always using the same coordinator node.

## Cloudera Impala Frequently Asked Questions

### Why is space not freed up when I issue DROP TABLE?

Impala deletes data files when you issue a `DROP TABLE` on an internal table, but not an external one. By default, the `CREATE TABLE` statement creates internal tables, where the files are managed by Impala. An external table is created with a `CREATE EXTERNAL TABLE` statement, where the files reside in a location outside the control of Impala. Issue a `DESCRIBE FORMATTED` statement to check whether a table is internal or external. The keyword `MANAGED_TABLE` indicates an internal table, from which Impala can delete the data files. The keyword `EXTERNAL_TABLE` indicates an external table, where Impala will leave the data files untouched when you drop the table.

Even when you drop an internal table and the files are removed from their original location, you might not get the hard drive space back immediately. By default, files that are deleted in HDFS go into a special trashcan directory, from which they are purged after a period of time (by default, 6 hours). For background information on the trashcan mechanism, see

<http://archive.cloudera.com/cdh4/cdh/4/hadoop/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html>. For

information on purging files from the trashcan, see

<http://archive.cloudera.com/cdh4/cdh/4/hadoop/hadoop-project-dist/hadoop-common/FileSystemShell.html>.

When Impala deletes files and they are moved to the HDFS trashcan, they go into an HDFS directory owned by the `impala` user. If the `impala` user does not have an HDFS home directory where a trashcan can be created, the files are not deleted or moved, as a safety measure. If you issue a `DROP TABLE` statement and find that the table data files are left in their original location, create an HDFS directory `/user/impala`, owned and writeable by the `impala` user. For example, you might find that `/user/impala` is owned by the `hdfs` user, in which case you would switch to the `hdfs` user and issue a command such as:

```
hdfs dfs -chown -R impala /user/impala
```

### Is there a DUAL table?

You might be used to running queries against a single-row table named `DUAL` to try out expressions, built-in functions, and UDFs. Impala does not have a `DUAL` table. To achieve the same result, you can issue a `SELECT` statement without any table name:

```
select 2+2;
select substr('hello',2,1);
select pow(10,6);
```

## Partitioned Tables

### How do I load a big CSV file into a partitioned table?

To load a data file into a partitioned table, when the data file includes fields like year, month, and so on that correspond to the partition key columns, use a two-stage process. First, use the `LOAD DATA` or `CREATE EXTERNAL TABLE` statement to bring the data into an unpartitioned text table. Then use an `INSERT ... SELECT` statement to copy the data from the unpartitioned table to a partitioned one. Include a `PARTITION` clause in the `INSERT` statement to specify the partition key columns. The `INSERT` operation splits up the data into separate data files for each partition. For examples, see [Partitioning for Impala Tables](#) on page 394. For details about loading data into partitioned Parquet tables, a popular choice for high-volume data, see [Loading Data into Parquet Tables](#) on page 409.

### Can I do `INSERT ... SELECT *` into a partitioned table?

When you use the `INSERT ... SELECT *` syntax to copy data into a partitioned table, the columns corresponding to the partition key columns must appear last in the columns returned by the `SELECT *`. You can create the table with the partition key columns defined last. Or, you can use the `CREATE VIEW` statement to create a view

that reorders the columns: put the partition key columns last, then do the `INSERT ... SELECT * from the view.`

## HBase

### What kinds of Impala queries or data are best suited for HBase?

HBase tables are ideal for queries where normally you would use a key-value store. That is, where you retrieve a single row or a few rows, by testing a special unique key column using the = or IN operators.

HBase tables are not suitable for queries that produce large result sets with thousands of rows. HBase tables are also not suitable for queries that perform full table scans because the WHERE clause does not request specific values from the unique key column.

Use HBase tables for data that is inserted one row or a few rows at a time, such as by the `INSERT ... VALUES` syntax. Loading data piecemeal like this into an HDFS-backed table produces many tiny files, which is a very inefficient layout for HDFS data files.

If the lack of an UPDATE statement in Impala is a problem for you, you can simulate single-row updates by doing an `INSERT ... VALUES` statement using an existing value for the key column. The old row value is hidden; only the new row value is seen by queries.

HBase tables are often wide (containing many columns) and sparse (with most column values NULL). For example, you might record hundreds of different data points for each user of an online service, such as whether the user had registered for an online game or enabled particular account features. With Impala and HBase, you could look up all the information for a specific customer efficiently in a single query. For any given customer, most of these columns might be NULL, because a typical customer might not make use of most features of an online service.

## Cloudera Impala Release Notes

These release notes provide information on the [new features](#) and [known issues and limitations](#) for Impala versions up to Impala 2.2.x / CDH 5.4.x. For users upgrading from earlier Impala releases, or using Impala in combination with specific versions of other Cloudera software, [Incompatible Changes in Impala](#) on page 487 lists any changes to file formats, SQL syntax, or software dependencies to take into account.

Once you are finished reviewing these release notes, for more information about using Impala, see [Cloudera Impala Guide](#).

## Cloudera Impala Release Notes

These release notes provide information on the [new features](#) and [known issues and limitations](#) for Impala versions up to Impala 2.2.x / CDH 5.4.x. For users upgrading from earlier Impala releases, or using Impala in combination with specific versions of other Cloudera software, [Incompatible Changes in Impala](#) on page 487 lists any changes to file formats, SQL syntax, or software dependencies to take into account.

Once you are finished reviewing these release notes, for more information about using Impala, see [Cloudera Impala Guide](#).

## New Features in Impala

This release of Impala contains the following changes and enhancements from previous releases.

### New Features in Impala Version 2.2.3 / CDH 5.4.3

No new features added to the Impala code. The certification of Impala with EMC Isilon under CDH 5.4.3 means that now you can query data stored on Isilon storage devices through Impala. See [Managing Isilon](#) for details.

- **Note:** The Impala 2.2.x maintenance releases now use the CDH 5.4.x numbering system rather than increasing the Impala version numbers. Impala 2.2 and higher are not available under CDH 4.

### New Features in Impala for CDH 5.4.x

No new features. CDH maintenance releases such as 5.4.1, 5.4.2, and so on are exclusively bug fix releases. See [New Features in Impala Version 2.2.0 / CDH 5.4.0](#) on page 470 for the most recent set of new Impala features.

- **Note:** The Impala 2.2.x maintenance releases now use the CDH 5.4.x numbering system rather than increasing the Impala version numbers. Impala 2.2 and higher are not available under CDH 4.

### New Features in Impala Version 2.2.0 / CDH 5.4.0

- **Note:** Impala 2.2.0 is available as part of CDH 5.4.0 and is not available for CDH 4. Cloudera does not intend to release future versions of Impala for CDH 4 outside patch and maintenance releases if required. Given the upcoming end-of-maintenance for CDH 4, Cloudera recommends all customers to migrate to a recent CDH 5 release.

The following are the major new features in Impala 2.2.0. This major release, available as part of CDH 5.4.0, contains improvements to performance, manageability, security, and SQL syntax.

- Several improvements to date and time features enable higher interoperability with Hive and other database systems, provide more flexibility for handling time zones, and future-proof the handling of `TIMESTAMP` values:

- Startup flags for the `impalad` daemon enable a higher level of compatibility with `TIMESTAMP` values written by Hive, and more flexibility for working with date and time data using the local time zone instead of UTC. To enable these features, set the `impalad` startup flags  
`-use_local_tz_for_unix_timestamp_conversions=true` and  
`-convert_legacy_hive_parquet_utc_timestamps=true`.

The `-use_local_tz_for_unix_timestamp_conversions` setting controls how the `unix_timestamp()`, `from_unixtime()`, and `now()` functions handle time zones. By default (when this setting is turned off), Impala considers all `TIMESTAMP` values to be in the UTC time zone when converting to or from Unix time values. When this setting is enabled, Impala treats `TIMESTAMP` values passed to or returned from these functions to be in the local time zone. When this setting is enabled, take particular care that all hosts in the cluster have the same timezone settings, to avoid inconsistent results depending on which host reads or writes `TIMESTAMP` data.

The `-convert_legacy_hive_parquet_utc_timestamps` setting causes Impala to convert `TIMESTAMP` values to the local time zone when it reads them from Parquet files written by Hive. This setting only applies to data using the Parquet file format, where Impala can use metadata in the files to reliably determine that the files were written by Hive. If in the future Hive changes the way it writes `TIMESTAMP` data in Parquet, Impala will automatically handle that new `TIMESTAMP` encoding.

See [TIMESTAMP Data Type](#) on page 130 for details about time zone handling and the configuration options for Impala / Hive compatibility with Parquet format.

- In Impala 2.2.0 and higher, built-in functions that accept or return integers representing `TIMESTAMP` values use the `BIGINT` type for parameters and return values, rather than `INT`. This change lets the date and time functions avoid an overflow error that would otherwise occur on January 19th, 2038 (known as the ["Year 2038 problem" or "Y2K38 problem"](#)). This change affects the `from_unixtime()` and `unix_timestamp()` functions. You might need to change application code that interacts with these functions, change the types of columns that store the return values, or add `CAST()` calls to SQL statements that call these functions.

See [Impala Date and Time Functions](#) on page 268 for the current function signatures.

- The `SHOW FILES` statement lets you view the names and sizes of the files that make up an entire table or a specific partition. See [SHOW FILES Statement](#) on page 243 for details.
- Impala can now run queries against Parquet data containing columns with composite or nested types, as long as the query only refers to columns with scalar types.
- Performance improvements for queries that include `IN()` operators and involve partitioned tables.
- The new `-max_log_files` configuration option specifies how many log files to keep at each severity level. The default value is 10, meaning that Impala preserves the latest 10 log files for each severity level (`INFO`, `WARNING`, and `ERROR`) for each Impala-related daemon (`impalad`, `statestored`, and `catalogd`). Impala checks to see if any old logs need to be removed based on the interval specified in the `logbufsecs` setting, every 5 seconds by default. See [Rotating Impala Logs](#) on page 444 for details.
- Redaction of sensitive data from Impala log files. This feature protects details such as credit card numbers or tax IDs from administrators who see the text of SQL statements in the course of monitoring and troubleshooting a Hadoop cluster. See [Redacting Sensitive Information from Impala Log Files](#) on page 446 for background information for Impala users, and [Sensitive Data Redaction](#) for usage details.
- Lineage information is available for data created or queried by Impala. This feature lets you track who has accessed data through Impala SQL statements, down to the level of specific columns, and how data has been propagated between tables. See [Viewing Lineage Information for Impala Data](#) on page 110 for background information for Impala users, [Impala Lineage Properties](#) for usage details, and [Lineage Diagrams](#) for how to interpret the lineage information.
- Impala tables and partitions can now be located on the Amazon Simple Storage Service (S3) filesystem, for convenience in cases where data is already located in S3 and you prefer to query it in-place. Queries might have lower performance than when the data files reside on HDFS, because Impala uses some HDFS-specific

optimizations. Impala can query data in S3, but cannot write to S3. Therefore, statements such as `INSERT` and `LOAD DATA` are not available when the destination table or partition is in S3. See [Using Impala to Query the Amazon S3 Filesystem \(Unsupported Preview\)](#) on page 436 for details.

▪ **Important:**

Impala query support for Amazon S3 is included in CDH 5.4.0, but is not currently supported or recommended for production use. If you're interested in this feature, try it out in a test environment until we address the issues and limitations needed for production-readiness.

- Improved support for HDFS encryption. The `LOAD DATA` statement now works when the source directory and destination table are in different encryption zones.
- Additional arithmetic function `mod()`. See [Impala Mathematical Functions](#) on page 259 for details.
- Flexibility to interpret `TIMESTAMP` values using the UTC time zone (the traditional Impala behavior) or using the local time zone (for compatibility with `TIMESTAMP` values produced by Hive).
- Enhanced support for ETL using tools such as Flume. Impala ignores temporary files typically produced by these tools (filenames with suffixes `.copying` and `.tmp`).
- The CPU requirement for Impala, which had become more restrictive in Impala 2.0.x and 2.1.x, has now been relaxed.

The prerequisite for CPU architecture has been relaxed in Impala 2.2.0 and higher. From this release onward, Impala works on CPUs that have the SSSE3 instruction set. The SSE4 instruction set is no longer required. This relaxed requirement simplifies the upgrade planning from Impala 1.x releases, which also worked on SSSE3-enabled processors.

- Enhanced support for `CHAR` and `VARCHAR` types in the `COMPUTE STATS` statement.
- The amount of memory required during setup for “spill to disk” operations is greatly reduced. This enhancement reduces the chance of a memory-intensive join or aggregation query failing with an out-of-memory error.
- Several new conditional functions provide enhanced compatibility when porting code that uses industry extensions. The new functions are: `isfalse()`, `isnotfalse()`, `isnottrue()`, `istrue()`, `notnullvalue()`, and `nullvalue()`. See [Impala Conditional Functions](#) on page 275 for details.
- The Impala debug web UI now can display a visual representation of the query plan. On the `/queries` tab, select **Details** for a particular query. The **Details** page includes a **Plan** tab with a plan diagram that you can zoom in or out (using scroll gestures through mouse wheel or trackpad).

### New Features in Impala Version 2.1.4 / CDH 5.3.4

No new features. This point release is exclusively a bug fix release. Because CDH 5.3.5 does not include any code changes for Impala, Impala 2.1.4 is included with both CDH 5.3.4 and 5.3.5.

- **Note:** This Impala maintenance release is only available as part of CDH 5, not under CDH 4.

### New Features in Impala Version 2.1.3 / CDH 5.3.3

No new features. This point release is exclusively a bug fix release.

- **Note:** Impala 2.1.3 is available as part of CDH 5.3.3, not under CDH 4.

### New Features in Impala Version 2.1.2 / CDH 5.3.2

No new features. This point release is exclusively a bug fix release.

- **Note:** Impala 2.1.2 is available as part of CDH 5.3.2, not under CDH 4.

## New Features in Impala Version 2.1.1 / CDH 5.3.1

No new features. This point release is exclusively a bug fix release.

## New Features in Impala Version 2.1.0 / CDH 5.3.0

This release contains the following enhancements to query performance and system scalability:

- Impala can now collect statistics for individual partitions in a partitioned table, rather than processing the entire table for each `COMPUTE STATS` statement. This feature is known as incremental statistics, and is controlled by the `COMPUTE INCREMENTAL STATS` syntax. (You can still use the original `COMPUTE STATS` statement for nonpartitioned tables or partitioned tables that are unchanging or whose contents are entirely replaced all at once.) See [COMPUTE STATS Statement](#) on page 170 and [How Impala Uses Statistics for Query Optimization](#) on page 363 for details.
- Optimization for small queries lets Impala process queries that process very few rows without the unnecessary overhead of parallelizing and generating native code. Reducing this overhead lets Impala clear small queries quickly, keeping YARN resources and admission control slots available for data-intensive queries. The number of rows considered to be a “small” query is controlled by the `EXEC_SINGLE_NODE_ROWS_THRESHOLD` query option. See [EXEC\\_SINGLE\\_NODE\\_ROWS\\_THRESHOLD Query Option](#) on page 342 for details.
- An enhancement to the statestore component lets it transmit heartbeat information independently of broadcasting metadata updates. This optimization improves reliability of health checking on large clusters with many tables and partitions.
- The memory requirement for querying gzip-compressed text is reduced. Now Impala decompresses the data as it is read, rather than reading the entire gzipped file and decompressing it in memory.

## New Features in Impala Version 2.0.5 / CDH 5.2.6

No new features. This point release is exclusively a bug fix release.

- **Note:** Impala 2.0.5 is available as part of CDH 5.2.6, not under CDH 4.

## New Features in Impala Version 2.0.4 / CDH 5.2.5

No new features. This point release is exclusively a bug fix release.

- **Note:** Impala 2.0.4 is available as part of CDH 5.2.5, not under CDH 4.

## New Features in Impala Version 2.0.3 / CDH 5.2.4

No new features. This point release is exclusively a bug fix release.

- **Note:** Impala 2.0.3 is available as part of CDH 5.2.4, not under CDH 4.

## New Features in Impala Version 2.0.2 / CDH 5.2.3

No new features. This point release is exclusively a bug fix release.

- **Note:** Impala 2.0.2 is available as part of CDH 5.2.3, not under CDH 4.

## New Features in Impala Version 2.0.1 / CDH 5.2.1

No new features. This point release is exclusively a bug fix release.

## New Features in Impala Version 2.0.0 / CDH 5.2.0

The following are the major new features in Impala 2.0. This major release, available both with CDH 5.2 and for CDH 4, contains improvements to performance, scalability, security, and SQL syntax.

- Queries with joins or aggregation functions involving high volumes of data can now use temporary work areas on disk, reducing the chance of failure due to out-of-memory errors. When the required memory for the intermediate result set exceeds the amount available on a particular node, the query automatically uses a temporary work area on disk. This “spill to disk” mechanism is similar to the `ORDER BY` improvement from Impala 1.4. For details, see [SQL Operations that Spill to Disk](#) on page 389.
- Subquery enhancements:
  - Subqueries are now allowed in the `WHERE` clause, for example with the `IN` operator.
  - The `EXISTS` and `NOT EXISTS` operators are available. They are always used in conjunction with subqueries.
  - The `IN` and `NOT IN` queries can now operate on the result set from a subquery, not just a hardcoded list of values.
  - Uncorrelated subqueries let you compare against one or more values for equality, `IN`, and `EXISTS` comparisons. For example, you might use `WHERE` clauses such as `WHERE column = (SELECT MAX(some_other_column FROM table)` or `WHERE column IN (SELECT some_other_column FROM table WHERE conditions)`.
  - Correlated subqueries let you cross-reference values from the outer query block and the subquery.
  - Scalar subqueries let you substitute the result of single-value aggregate functions such as `MAX()`, `MIN()`, `COUNT()`, or `AVG()`, where you would normally use a numeric value in a `WHERE` clause.

For details about subqueries, see [Subqueries](#) on page 235 For information about new and improved operators, see [EXISTS Operator](#) on page 144 and [IN Operator](#) on page 147.

- Analytic functions such as `RANK()`, `LAG()`, `LEAD()`, and `FIRST_VALUE()` let you analyze sequences of rows with flexible ordering and grouping. Existing aggregate functions such as `MAX()`, `SUM()`, and `COUNT()` can also be used in an analytic context. See [Impala Analytic Functions](#) on page 296 for details. See [Impala Aggregate Functions](#) on page 284 for enhancements to existing aggregate functions.
- New data types provide greater compatibility with source code from traditional database systems:
  - `VARCHAR` is like the `STRING` data type, but with a maximum length. See [VARCHAR Data Type \(CDH 5.2 or higher only\)](#) on page 136 for details.
  - `CHAR` is like the `STRING` data type, but with a precise length. Short values are padded with spaces on the right. See [CHAR Data Type \(CDH 5.2 or higher only\)](#) on page 114 for details.
- Security enhancements:
  - Formerly, Impala was restricted to using either Kerberos or LDAP / Active Directory authentication within a cluster. Now, Impala can freely accept either kind of authentication request, allowing you to set up some hosts with Kerberos authentication and others with LDAP or Active Directory. See [Using Multiple Authentication Methods with Impala](#) on page 108 for details.
  - `GRANT` statement. See [GRANT Statement \(CDH 5.2 or higher only\)](#) on page 201 for details.
  - `REVOKE` statement. See [REVOKE Statement \(CDH 5.2 or higher only\)](#) on page 217 for details.
  - `CREATE ROLE` statement. See [CREATE ROLE Statement \(CDH 5.2 or higher only\)](#) on page 179 for details.
  - `DROP ROLE` statement. See [DROP ROLE Statement \(CDH 5.2 or higher only\)](#) on page 193 for details.
  - `SHOW ROLES` and `SHOW ROLE GRANT` statements. See [SHOW Statement](#) on page 242 for details.
  - To complement the HDFS encryption feature, a new Impala configuration option, `--disk_spill_encryption` secures sensitive data from being observed or tampered with when temporarily stored on disk.

The new security-related SQL statements work along with the Sentry authorization framework. See [Enabling Sentry Authorization for Impala](#) on page 91 for details.

- Impala can now read compressed text files compressed by gzip, bzip, or Snappy. These files do not require any special table settings to work in an Impala text table. Impala recognizes the compression type automatically based on file extensions of .gz, .bz2, and .snappy respectively. These types of compressed text files are intended for convenience with existing ETL pipelines. Their non-splitable nature means they are not optimal for high-performance parallel queries. See [Using gzip, bzip2, or Snappy-Compressed Text Files](#) on page 406 for details.
- Query hints can now use comment notation, /\* +hint\_name \*/ or -- +hint\_name, at the same places in the query where the hints enclosed by [ ] are recognized. This enhancement makes it easier to reuse Impala queries on other database systems. See [Hints](#) on page 239 for details.
- A new query option, QUERY\_TIMEOUT\_S, lets you specify a timeout period in seconds for individual queries.

The working of the --idle\_query\_timeout configuration option is extended. If no QUERY\_OPTION\_S query option is in effect, --idle\_query\_timeout works the same as before, setting the timeout interval. When the QUERY\_OPTION\_S query option is specified, its maximum value is capped by the value of the --idle\_query\_timeout option.

That is, the system administrator sets the default and maximum timeout through the --idle\_query\_timeout startup option, and then individual users or applications can set a lower timeout value if desired through the QUERY\_TIMEOUT\_S query option. See [Setting Timeout Periods for Daemons, Queries, and Sessions](#) on page 81 and [QUERY\\_TIMEOUT\\_S Query Option](#) on page 351 for details.

- New functions VAR\_SAMP() and VAR\_POP() are aliases for the existing VARIANCE\_SAMP() and VARIANCE\_POP() functions.
- A new date and time function, DATE\_PART(), provides similar functionality to EXTRACT(). You can also call the EXTRACT() function using the SQL-99 syntax, EXTRACT(unit FROM timestamp). These enhancements simplify the porting process for date-related code from other systems. See [Impala Date and Time Functions](#) on page 268 for details.
- New approximation features provide a fast way to get results when absolute precision is not required:
  - The APPX\_COUNT\_DISTINCT query option lets Impala rewrite COUNT(DISTINCT) calls to use NDV() instead, which speeds up the operation and allows multiple COUNT(DISTINCT) operations in a single query. See [APPX\\_COUNT\\_DISTINCT Query Option](#) on page 340 for details.

The APPX\_MEDIAN() aggregate function produces an estimate for the median value of a column by using sampling. See [APPX\\_MEDIAN Function](#) on page 284 for details.

- Impala now supports a DECODE() function. This function works as a shorthand for a CASE() expression, and improves compatibility with SQL code containing vendor extensions. See [Impala Conditional Functions](#) on page 275 for details.
- The STDDEV(), STDDEV\_POP(), STDDEV\_SAMP(), VARIANCE(), VARIANCE\_POP(), VARIANCE\_SAMP(), and NDV() aggregate functions now all return DOUBLE results rather than STRING. Formerly, you were required to CAST() the result to a numeric type before using it in arithmetic operations.
- The default settings for Parquet block size, and the associated PARQUET\_FILE\_SIZE query option, are changed. Now, Impala writes Parquet files with a size of 256 MB and an HDFS block size of 256 MB. Previously, Impala attempted to write Parquet files with a size of 1 GB and an HDFS block size of 1 GB. In practice, Impala used a conservative estimate of the disk space needed for each Parquet block, leading to files that were typically 512 MB anyway. Thus, this change will make the file size more accurate if you specify a value for the PARQUET\_FILE\_SIZE query option. It also reduces the amount of memory reserved during INSERT into Parquet tables, potentially avoiding out-of-memory errors and improving scalability when inserting data into Parquet tables.

## Cloudera Impala Release Notes

- Anti-joins are now supported, expressed using the `LEFT ANTI JOIN` and `RIGHT ANTI JOIN` clauses. These clauses returns results from one table that have no match in the other table. You might use this type of join in the same sorts of use cases as the `NOT EXISTS` and `NOT IN` operators. See [Joins](#) on page 220 for details.
- The `SET` command in `impala-shell` has been promoted to a real SQL statement. You can now set query options such as `PARQUET_FILE_SIZE`, `MEM_LIMIT`, and `SYNC_DDL` within JDBC, ODBC, or any other kind of application that submits SQL without going through the `impala-shell` interpreter. See [SET Statement](#) on page 241 for details.
- The `impala-shell` interpreter now reads settings from an optional configuration file, named `$HOME/.impalarc` by default. See [impala-shell Configuration Options](#) on page 332 for details.
- The library used for regular expression parsing has changed from Boost to Google RE2. This implementation change adds support for non-greedy matches using the `.*?` notation. This and other changes in the way regular expressions are interpreted means you might need to re-test queries that use functions such as `regexp_extract()` or `regexp_replace()`, or operators such as `REGEXP` or `RLIKE`. See [Incompatible Changes in Impala](#) on page 487 for those details.

### New Features in Impala Version 1.4.4 / CDH 5.1.5

No new features. This point release is exclusively a bug fix release.

- **Note:** Impala 1.4.4 is available as part of CDH 5.1.5, not under CDH 4.

### New Features in Impala Version 1.4.3 / CDH 5.1.4

No new features. This point release is exclusively a bug fix release for an SSL security issue.

- **Note:** Impala 1.4.3 is available as part of CDH 5.1.4, and under CDH 4.

### New Features in Impala Version 1.4.2 / CDH 5.1.3

Impala 1.4.2 is purely a bug-fix release. It does not include any new features.

- **Note:** Impala 1.4.2 is only available as part of CDH 5.1.3, not under CDH 4.

### New Features in Impala Version 1.4.1 / CDH 5.1.2

Impala 1.4.1 is purely a bug-fix release. It does not include any new features.

### New Features in Impala Version 1.4.0 / CDH 5.1.0

- The `DECIMAL` data type lets you store fixed-precision values, for working with currency or other fractional values where it is important to represent values exactly and avoid rounding errors. This feature includes enhancements to built-in functions, numeric literals, and arithmetic expressions.
- On CDH 5, Impala can take advantage of the HDFS caching feature to “pin” entire tables or individual partitions in memory, to speed up queries on frequently accessed data and reduce the CPU overhead of memory-to-memory copying. When HDFS files are cached in memory, Impala can read the cached data without any disk reads, and without making an additional copy of the data in memory. Other Hadoop components that read the same data files also experience a performance benefit.
- Impala can now use Sentry-based authorization based either on the original policy file, or on rules defined by `GRANT` and `REVOKE` statements issued through Hive.

- For interoperability with Parquet files created through other Hadoop components, such as Pig or MapReduce jobs, you can create an Impala table that automatically sets up the column definitions based on the layout of an existing Parquet data file.
- ORDER BY queries no longer require a LIMIT clause. If the size of the result set to be sorted exceeds the memory available to Impala, Impala uses a temporary work space on disk to perform the sort operation.
- LDAP connections can be secured through either SSL or TLS.
- The following new built-in scalar and aggregate functions are available:
  - A new built-in function, EXTRACT( ), returns one date or time field from a TIMESTAMP value.
  - A new built-in function, TRUNC( ), truncates date/time values to a particular granularity, such as year, month, day, hour, and so on.
  - ADD\_MONTHS( ) built-in function, an alias for the existing MONTHS\_ADD( ) function.
  - A new built-in function, ROUND( ), rounds DECIMAL values to a specified number of fractional digits.
  - Several built-in aggregate functions for computing properties for statistical distributions: STDDEV( ), STDDEV\_SAMP( ), STDDEV\_POP( ), VARIANCE( ), VARIANCE\_SAMP( ), and VARIANCE\_POP( ).
  - Several new built-in functions, such as MAX\_INT( ), MIN\_SMALLINT( ), and so on, let you conveniently check whether data values are in an expected range. You might be able to switch a column to a smaller type, saving memory during processing.
  - New built-in functions, IS\_INF( ) and IS\_NAN( ), check for the special values infinity and "not a number". These values could be specified as inf or nan in text data files, or be produced by certain arithmetic expressions.
- The SHOW PARTITIONS statement displays information about the structure of a partitioned table.
- New configuration options for the impalad daemon let you specify initial memory usage for all queries. The initial resource requests handled by Llama and YARN can be expanded later if needed, avoiding unnecessary over-allocation and reducing the chance of out-of-memory conditions.
- Impala can take advantage of the Llama high availability feature in CDH 5.1, for improved reliability of resource management through YARN.
- The Impala CREATE TABLE statement now has a STORED AS AVRO clause, allowing you to create Avro tables through Impala.
- New impalad configuration options let you fine-tune the calculations Impala makes to estimate resource requirements for each query. These options can help avoid problems due to overconsumption due to too-low estimates, or underutilization due to too-high estimates.
- A new SUMMARY command in the impala-shell interpreter provides a high-level summary of the work performed at each stage of the explain plan. The summary is also included in output from the PROFILE command.
- Performance improvements for the COMPUTE STATS statement:
  - The NDV function is speeded up through native code generation.
  - Because the NULL count is not currently used by the Impala query planner, in Impala 1.4.0 and higher, COMPUTE STATS does not count the NULL values for each column. (The #Nulls field of the stats table is left as -1, signifying that the value is unknown.)
- Performance improvements for partition pruning. This feature reduces the time spent in query planning, for partitioned tables with thousands of partitions. Previously, Impala typically queried tables with up to approximately 3000 partitions. With the performance improvement in partition pruning, now Impala can comfortably handle tables with tens of thousands of partitions.
- The documentation provides additional guidance for planning tasks.

## Cloudera Impala Release Notes

- The `impala-shell` interpreter now supports UTF-8 characters for input and output. You can control whether `impala-shell` ignores invalid Unicode code points through the `--strict_unicode` option. (Although this option is removed in Impala 2.0.)

### New Features in Impala Version 1.3.3 / CDH 5.0.5

No new features. This point release is exclusively a bug fix release for an SSL security issue.

- Note:** Impala 1.3.3 is only available as part of CDH 5.0.5, not under CDH 4.

### New Features in Impala Version 1.3.2 / CDH 5.0.4

No new features. This point release is exclusively a bug fix release for the IMPALA-1019 issue related to HDFS caching.

- Note:** Impala 1.3.2 is only available as part of CDH 5.0.4, not under CDH 4.

### New Features in Impala Version 1.3.1 / CDH 5.0.3

This point release is primarily a vehicle to deliver bug fixes. Any new features are minor changes resulting from fixes for performance, reliability, or usability issues.

Because 1.3.1 is the first 1.3.x release for CDH 4, if you are on CDH 4, also consult [New Features in Impala Version 1.3.0 / CDH 5.0.0](#) on page 478 for more features that are new to you.

- Note:**

- The Impala 1.3.1 release is available for both CDH 4 and CDH 5. This is the first release in the 1.3.x series for CDH 4.

- A new `impalad` startup option, `--insert_inherit_permissions`, causes Impala `INSERT` statements to create each new partition with the same HDFS permissions as its parent directory. By default, `INSERT` statements create directories for new partitions using default HDFS permissions. See [INSERT Statement](#) on page 202 for examples of `INSERT` statements for partitioned tables.
- The `SHOW FUNCTIONS` statement now displays the return type of each function, in addition to the types of its arguments. See [SHOW Statement](#) on page 242 for examples.
- You can now specify the clause `FIELDS TERMINATED BY '\0'` with a `CREATE TABLE` statement to use text data files that use ASCII 0 (`\nul`) characters as a delimiter. See [Using Text Data Files with Impala Tables](#) on page 400 for details.
- In Impala 1.3.1 and higher, the `REGEXP` and `RLIKE` operators now match a regular expression string that occurs anywhere inside the target string, the same as if the regular expression was enclosed on each side by `.*`. See [REGEXP Operator](#) on page 152 for examples. Previously, these operators only succeeded when the regular expression matched the entire target string. This change improves compatibility with the regular expression support for popular database systems. There is no change to the behavior of the `regexp_extract()` and `regexp_replace()` built-in functions.

### New Features in Impala Version 1.3.0 / CDH 5.0.0

- Note:**

- The Impala 1.3.1 release is available for both CDH 4 and CDH 5. This is the first release in the 1.3.x series for CDH 4.

- The admission control feature lets you control and prioritize the volume and resource consumption of concurrent queries. This mechanism reduces spikes in resource usage, helping Impala to run alongside other kinds of workloads on a busy cluster. It also provides more user-friendly conflict resolution when multiple memory-intensive queries are submitted concurrently, avoiding resource contention that formerly resulted in out-of-memory errors. See [Admission Control and Query Queuing](#) on page 69 for details.
- Enhanced EXPLAIN plans provide more detail in an easier-to-read format. Now there are four levels of verbosity: the EXPLAIN\_LEVEL option can be set from 0 (most concise) to 3 (most verbose). See [EXPLAIN Statement](#) on page 199 for syntax and [Understanding Impala Query Performance - EXPLAIN Plans and Query Profiles](#) on page 378 for usage information.
- The TIMESTAMP data type accepts more kinds of input string formats through the UNIX\_TIMESTAMP function, and produces more varieties of string formats through the FROM\_UNIXTIME function. The documentation now also lists more functions for date arithmetic, used for adding and subtracting INTERVAL expressions from TIMESTAMP values. See [Impala Date and Time Functions](#) on page 268 for details.
- New conditional functions, NULLIF( ), NULLIFZERO( ), and ZEROIFNULL( ), simplify porting SQL containing vendor extensions to Impala. See [Impala Conditional Functions](#) on page 275 for details.
- New utility function, CURRENT\_DATABASE( ). See [Impala Miscellaneous Functions](#) on page 284 for details.
- Integration with the YARN resource management framework. Only available in combination with CDH 5. This feature makes use of the underlying YARN service, plus an additional service (Llama) that coordinates requests to YARN for Impala resources, so that the Impala query only proceeds when all requested resources are available. See [Integrated Resource Management with YARN](#) on page 78 for full details.

On the Impala side, this feature involves some new startup options for the impalad daemon:

- –enable\_rm
- –llama\_host
- –llama\_port
- –llama\_callback\_port
- –cgroup\_hierarchy\_path

For details of these startup options, see [Modifying Impala Startup Options](#) on page 48.

This feature also involves several new or changed query options that you can set through the impala-shell interpreter and apply within a specific session:

- MEM\_LIMIT: the function of this existing option changes when Impala resource management is enabled.
- REQUEST\_POOL: a new option. (Renamed to RESOURCE\_POOL in Impala 1.3.0.)
- V\_CPU\_CORES: a new option.
- RESERVATION\_REQUEST\_TIMEOUT: a new option.

For details of these query options, see [impala-shell Query Options for Resource Management](#) on page 80.

## New Features in Impala Version 1.2.4

**Note:** Impala 1.2.4 works with CDH 4. It is primarily a bug fix release for Impala 1.2.3, plus some performance enhancements for the catalog server to minimize startup and DDL wait times for Impala deployments with large numbers of databases, tables, and partitions.

- On Impala startup, the metadata loading and synchronization mechanism has been improved and optimized, to give more responsiveness when starting Impala on a system with a large number of databases, tables, or partitions. The initial metadata loading happens in the background, allowing queries to be run before the entire process is finished. When a query refers to a table whose metadata is not yet loaded, the query waits until the metadata for that table is loaded, and the load operation for that table is prioritized to happen first.
- Formerly, if you created a new table in Hive, you had to issue the INVALIDATE\_METADATA statement (with no table name) which was an expensive operation that reloaded metadata for all tables. Impala did not

recognize the name of the Hive-created table, so you could not do `INVALIDATE METADATA new_table` to get the metadata for just that one table. Now, when you issue `INVALIDATE METADATA table_name`, Impala checks to see if that name represents a table created in Hive, and if so recognizes the new table and loads the metadata for it. Additionally, if the new table is in a database that was newly created in Hive, Impala also recognizes the new database.

- If you issue `INVALIDATE METADATA table_name` and the table has been dropped through Hive, Impala will recognize that the table no longer exists.
- New startup options let you control the parallelism of the metadata loading during startup for the catalogd daemon:
  - `--load_catalog_in_background` makes Impala load and cache metadata using background threads after startup. It is true by default. Previously, a system with a large number of databases, tables, or partitions could be unresponsive or even time out during startup.
  - `--num_metadata_loading_threads` determines how much parallelism Impala devotes to loading metadata in the background. The default is 16. You might increase this value for systems with huge numbers of databases, tables, or partitions. You might lower this value for busy systems that are CPU-constrained due to jobs from components other than Impala.

### New Features in Impala Version 1.2.3

- **Note:** Impala 1.2.3 works with CDH 4 and with CDH 5 beta 2. The resource management feature requires CDH 5 beta.

Impala 1.2.3 contains exactly the same feature set as Impala 1.2.2. Its only difference is one additional fix for compatibility with Parquet files generated outside of Impala by components such as Hive, Pig, or MapReduce. If you are upgrading from Impala 1.2.1 or earlier, see [New Features in Impala Version 1.2.2](#) on page 480 for the latest added features.

### New Features in Impala Version 1.2.2

- **Note:** Impala 1.2.2 works with CDH 4. Its feature set is a superset of features in the Impala 1.2.0 beta, with the exception of resource management, which relies on CDH 5.

Impala 1.2.2 includes new features for performance, security, and flexibility. The major enhancements over 1.2.1 are performance related, primarily for join queries.

New user-visible features include:

- Join order optimizations. This highly valuable feature automatically distributes and parallelizes the work for a join query to minimize disk I/O and network traffic. The automatic optimization reduces the need to use query hints or to rewrite join queries with the tables in a specific order based on size or cardinality. The new `COMPUTE STATS` statement gathers statistical information about each table that is crucial for enabling the join optimizations. See [Performance Considerations for Join Queries](#) on page 357 for details.
- `COMPUTE STATS` statement to collect both table statistics and column statistics with a single statement. Intended to be more comprehensive, efficient, and reliable than the corresponding Hive `ANALYZE TABLE` statement, which collects statistics in multiple phases through MapReduce jobs. These statistics are important for query planning for join queries, queries on partitioned tables, and other types of data-intensive operations. For optimal planning of join queries, you need to collect statistics for each table involved in the join. See [COMPUTE STATS Statement](#) on page 170 for details.
- Reordering of tables in a join query can be overridden by the `STRAIGHT_JOIN` operator, allowing you to fine-tune the planning of the join query if necessary, by using the original technique of ordering the joined tables in descending order of size. See [Overriding Join Reordering with STRAIGHT\\_JOIN](#) on page 358 for details.

- The `CROSS JOIN` clause in the `SELECT` statement to allow Cartesian products in queries, that is, joins without an equality comparison between columns in both tables. Because such queries must be carefully checked to avoid accidental overconsumption of memory, you must use the `CROSS JOIN` operator to explicitly select this kind of join. See [Cross Joins and Cartesian Products with the CROSS JOIN Operator](#) on page 66 for examples.
- The `ALTER TABLE` statement has new clauses that let you fine-tune table statistics. You can use this technique as a less-expensive way to update specific statistics, in case the statistics become stale, or to experiment with the effects of different data distributions on query planning.
- LDAP username/password authentication in JDBC/ODBC. See [Enabling LDAP Authentication for Impala](#) on page 106 for details.
- `GROUP_CONCAT()` aggregate function to concatenate column values across all rows of a result set.
- The `INSERT` statement now accepts hints, `[SHUFFLE]` and `[NOSHUFFLE]`, to influence the way work is redistributed during `INSERT...SELECT` operations. The hints are primarily useful for inserting into partitioned Parquet tables, where using the `[SHUFFLE]` hint can avoid problems due to memory consumption and simultaneous open files in HDFS, by collecting all the new data for each partition on a specific node.
- Several built-in functions and operators are now overloaded for more numeric data types, to reduce the requirement to use `CAST()` for type coercion in `INSERT` statements. For example, the expression `2+2` in an `INSERT` statement formerly produced a `BIGINT` result, requiring a `CAST()` to be stored in an `INT` variable. Now, addition, subtraction, and multiplication only produce a result that is one step “bigger” than their arguments, and numeric and conditional functions can return `SMALLINT`, `FLOAT`, and other smaller types rather than always `BIGINT` or `DOUBLE`.
- New `fnv_hash()` built-in function for constructing hashed values. See [Impala Mathematical Functions](#) on page 259 for details.
- The clause `STORED AS PARQUET` is accepted as an equivalent for `STORED AS PARQUETFILE`. This more concise form is recommended for new code.

Because Impala 1.2.2 builds on a number of features introduced in 1.2.1, if you are upgrading from an older 1.1.x release straight to 1.2.2, also review [New Features in Impala Version 1.2.1](#) on page 481 to see features such as the `SHOW TABLE STATS` and `SHOW COLUMN STATS` statements, and user-defined functions (UDFs).

## New Features in Impala Version 1.2.1

- **Note:** Impala 1.2.1 works with CDH 4. Its feature set is a superset of features in the Impala 1.2.0 beta, with the exception of resource management, which relies on CDH 5.

Impala 1.2.1 includes new features for security, performance, and flexibility.

New user-visible features include:

- `SHOW TABLE STATS` *table\_name* and `SHOW COLUMN STATS` *table\_name* statements, to verify that statistics are available and to see the values used during query planning.
- `CREATE TABLE AS SELECT` syntax, to create a new table and transfer data into it in a single operation.
- `OFFSET` clause, for use with the `ORDER BY` and `LIMIT` clauses to produce “paged” result sets such as items 1-10, then 11-20, and so on.
- `NULLS FIRST` and `NULLS LAST` clauses to ensure consistent placement of `NULL` values in `ORDER BY` queries.
- New [built-in functions](#): `least()`, `greatest()`, `initcap()`.
- New aggregate function: `ndv()`, a fast alternative to `COUNT(DISTINCT col)` returning an approximate result.
- The `LIMIT` clause can now accept a numeric expression as an argument, rather than only a literal constant.

- The `SHOW CREATE TABLE` statement displays the end result of all the `CREATE TABLE` and `ALTER TABLE` statements for a particular table. You can use the output to produce a simplified setup script for a schema.
- The `--idle_query_timeout` and `--idle_session_timeout` options for `impalad` control the time intervals after which idle queries are cancelled, and idle sessions expire. See [Setting Timeout Periods for Daemons, Queries, and Sessions](#) on page 81 for details.
- User-defined functions (UDFs). This feature lets you transform data in very flexible ways, which is important when using Impala as part of an ETL or ELT pipeline. Prior to Impala 1.2, using UDFs required switching into Hive. Impala 1.2 can run scalar UDFs and user-defined aggregate functions (UDAs). Impala can run high-performance functions written in C++, or you can reuse existing Hive functions written in Java.

You create UDFs through the `CREATE FUNCTION` statement and drop them through the `DROP FUNCTION` statement. See [Impala User-Defined Functions \(UDFs\)](#) on page 308 for instructions about coding, building, and deploying UDFs, and [CREATE FUNCTION Statement](#) on page 177 and [DROP FUNCTION Statement](#) on page 192 for related SQL syntax.

- A new service automatically propagates changes to table data and metadata made by one Impala node, sending the new or updated metadata to all the other Impala nodes. The automatic synchronization mechanism eliminates the need to use the `INVALIDATE METADATA` and `REFRESH` statements after issuing Impala statements such as `CREATE TABLE`, `ALTER TABLE`, `DROP TABLE`, `INSERT`, and `LOAD DATA`.

For even more precise synchronization, you can enable the `SYNC_DDL` query option before issuing a DDL, `INSERT`, or `LOAD DATA` statement. This option causes the statement to wait, returning only after the catalog service has broadcast the applicable changes to all Impala nodes in the cluster.

▪ **Note:**

Because the catalog service only monitors operations performed through Impala, `INVALIDATE METADATA` and `REFRESH` are still needed on the Impala side after creating new tables or loading data through the Hive shell or by manipulating data files directly in HDFS. Because the catalog service broadcasts the result of the `REFRESH` and `INVALIDATE METADATA` statements to all Impala nodes, when you do need to use those statements, you can do so a single time rather than on every Impala node.

This service is implemented by the `catalogd` daemon. See [The Impala Catalog Service](#) on page 19 for details.

- `CREATE TABLE ... AS SELECT` syntax, to create a table and copy data into it in a single operation. See [CREATE TABLE Statement](#) on page 180 for details.
- The `CREATE TABLE` and `ALTER TABLE` statements have new clauses `TBLPROPERTIES` and `WITH SERDEPROPERTIES`. The `TBLPROPERTIES` clause lets you associate arbitrary items of metadata with a particular table as key-value pairs. The `WITH SERDEPROPERTIES` clause lets you specify the serializer/deserializer (SerDes) classes that read and write data for a table; although Impala does not make use of these properties, sometimes particular values are needed for Hive compatibility. See [CREATE TABLE Statement](#) on page 180 and [ALTER TABLE Statement](#) on page 164 for details.
- Impersonation support lets you authorize certain OS users associated with applications (for example, `hue`), to submit requests using the credentials of other users. Only available in combination with CDH 5. See [Configuring Per-User Access for Hue](#) on page 102 for details.
- Enhancements to `EXPLAIN` output. In particular, when you enable the new `EXPLAIN_LEVEL` query option, the `EXPLAIN` and `PROFILE` statements produce more verbose output showing estimated resource requirements and whether table and column statistics are available for the applicable tables and columns. See [EXPLAIN Statement](#) on page 199 for details.
- `SHOW CREATE TABLE` summarizes the effects of the original `CREATE TABLE` statement and any subsequent `ALTER TABLE` statements, giving you a `CREATE TABLE` statement that will re-create the current structure and layout for a table.

- The `LIMIT` clause for queries now accepts an arithmetic expression, in addition to numeric literals.

## New Features in Impala Version 1.2.0 (Beta)

- Note:** The Impala 1.2.0 beta release only works in combination with the beta version of CDH 5. The Impala 1.2.0 software is bundled together with the CDH 5 beta 1 download.

The Impala 1.2.0 beta includes new features for security, performance, and flexibility.

New user-visible features include:

- User-defined functions (UDFs). This feature lets you transform data in very flexible ways, which is important when using Impala as part of an ETL or ELT pipeline. Prior to Impala 1.2, using UDFs required switching into Hive. Impala 1.2 can run scalar UDFs and user-defined aggregate functions (UDAs). Impala can run high-performance functions written in C++, or you can reuse existing Hive functions written in Java.

You create UDFs through the `CREATE FUNCTION` statement and drop them through the `DROP FUNCTION` statement. See [Impala User-Defined Functions \(UDFs\)](#) on page 308 for instructions about coding, building, and deploying UDFs, and [CREATE FUNCTION Statement](#) on page 177 and [DROP FUNCTION Statement](#) on page 192 for related SQL syntax.

- A new service automatically propagates changes to table data and metadata made by one Impala node, sending the new or updated metadata to all the other Impala nodes. The automatic synchronization mechanism eliminates the need to use the `INVALIDATE METADATA` and `REFRESH` statements after issuing Impala statements such as `CREATE TABLE`, `ALTER TABLE`, `DROP TABLE`, `INSERT`, and `LOAD DATA`.

- Note:**

Because this service only monitors operations performed through Impala, `INVALIDATE METADATA` and `REFRESH` are still needed on the Impala side after creating new tables or loading data through the Hive shell or by manipulating data files directly in HDFS. Because the catalog service broadcasts the result of the `REFRESH` and `INVALIDATE METADATA` statements to all Impala nodes, when you do need to use those statements, you can do so a single time rather than on every Impala node.

This service is implemented by the `catalogd` daemon. See [The Impala Catalog Service](#) on page 19 for details.

- Integration with the YARN resource management framework. Only available in combination with CDH 5. This feature makes use of the underlying YARN service, plus an additional service (Llama) that coordinates requests to YARN for Impala resources, so that the Impala query only proceeds when all requested resources are available. See [Integrated Resource Management with YARN](#) on page 78 for full details.

On the Impala side, this feature involves some new startup options for the `impalad` daemon:

- `-enable_rm`
- `-llama_host`
- `-llama_port`
- `-llama_callback_port`
- `-cgroup_hierarchy_path`

For details of these startup options, see [Modifying Impala Startup Options](#) on page 48.

This feature also involves several new or changed query options that you can set through the `impala-shell` interpreter and apply within a specific session:

- `MEM_LIMIT`: the function of this existing option changes when Impala resource management is enabled.
- `YARN_POOL`: a new option. (Renamed to `RESOURCE_POOL` in Impala 1.3.0.)
- `V_CPU_CORES`: a new option.
- `RESERVATION_REQUEST_TIMEOUT`: a new option.

For details of these query options, see [impala-shell Query Options for Resource Management](#) on page 80.

- `CREATE TABLE ... AS SELECT` syntax, to create a table and copy data into it in a single operation. See [CREATE TABLE Statement](#) on page 180 for details.
- The `CREATE TABLE` and `ALTER TABLE` statements have a new `TBLPROPERTIES` clause that lets you associate arbitrary items of metadata with a particular table as key-value pairs. See [CREATE TABLE Statement](#) on page 180 and [ALTER TABLE Statement](#) on page 164 for details.
- Impersonation support lets you authorize certain OS users associated with applications (for example, `hue`), to submit requests using the credentials of other users. Only available in combination with CDH 5. See [Configuring Per-User Access for Hue](#) on page 102 for details.
- Enhancements to `EXPLAIN` output. In particular, when you enable the new `EXPLAIN_LEVEL` query option, the `EXPLAIN` and `PROFILE` statements produce more verbose output showing estimated resource requirements and whether table and column statistics are available for the applicable tables and columns. See [EXPLAIN Statement](#) on page 199 for details.

### New Features in Impala Version 1.1.1

Impala 1.1.1 includes new features for security and stability.

New user-visible features include:

- Additional security feature: auditing. New startup options for `impalad` let you capture information about Impala queries that succeed or are blocked due to insufficient privileges. To take full advantage of this feature with Cloudera Manager, upgrade to Cloudera Manager 4.7 or higher. For details, see [Impala Security](#) on page 87.
- Parquet data files generated by Impala 1.1.1 are now compatible with the Parquet support in Hive. See [Incompatible Changes in Impala](#) on page 487 for the procedure to update older Impala-created Parquet files to be compatible with the Hive Parquet support.
- Additional improvements to stability and resource utilization for Impala queries.
- Additional enhancements for compatibility with existing file formats.

### New Features in Impala Version 1.1

Impala 1.1 includes new features for security, performance, and usability.

New user-visible features include:

- Extensive new security features, built on top of the Sentry open source project. Impala now supports fine-grained authorization based on roles. A policy file determines which privileges on which schema objects (servers, databases, tables, and HDFS paths) are available to users based on their membership in groups. By assigning privileges for views, you can control access to table data at the column level. For details, see [Impala Security](#) on page 87.
- Impala 1.1 works with Cloudera Manager 4.6 or higher. To use Cloudera Manager to manage authorization for the Impala web UI (the web pages served from port 25000 by default), use Cloudera Manager 4.6.2 or higher.
- Impala can now create, alter, drop, and query views. Views provide a flexible way to set up simple aliases for complex queries; hide query details from applications and users; and simplify maintenance as you rename or reorganize databases, tables, and columns. See the overview section [Views](#) on page 158 and the statements [CREATE VIEW Statement](#) on page 186, [ALTER VIEW Statement](#) on page 168, and [DROP VIEW Statement](#) on page 198.
- Performance is improved through a number of automatic optimizations. Resource consumption is also reduced for Impala queries. These improvements apply broadly across all kinds of workloads and file formats. The major areas of performance enhancement include:
  - Improved disk and thread scheduling, which applies to all queries.

- Improved hash join and aggregation performance, which applies to queries with large build tables or a large number of groups.
- Dictionary encoding with Parquet, which applies to Parquet tables with short string columns.
- Improved performance on systems with SSDs, which applies to all queries and file formats.
- Some new built-in functions are implemented: [translate\(\)](#) to substitute characters within strings, [user\(\)](#) to check the login ID of the connected user.
- The new `WITH` clause for `SELECT` statements lets you simplify complicated queries in a way similar to creating a view. The effects of the `WITH` clause only last for the duration of one query, unlike views, which are persistent schema objects that can be used by multiple sessions or applications. See [WITH Clause](#) on page 237.
- An enhancement to `DESCRIBE` statement, `DESCRIBE FORMATTED table_name`, displays more detailed information about the table. This information includes the file format, location, delimiter, ownership, external or internal, creation and access times, and partitions. The information is returned as a result set that can be interpreted and used by a management or monitoring application. See [DESCRIBE Statement](#) on page 188.
- You can now insert a subset of columns for a table, with other columns being left as all `NULL` values. Or you can specify the columns in any order in the destination table, rather than having to match the order of the corresponding columns in the source. `VALUES` clause. This feature is known as "column permutation". See [INSERT Statement](#) on page 202.
- The new `LOAD DATA` statement lets you load data into a table directly from an HDFS data file. This technique lets you minimize the number of steps in your ETL process, and provides more flexibility. For example, you can bring data into an Impala table in one step. Formerly, you might have created an external table where the data files are not entirely under your control, or copied the data files to Impala data directories manually, or loaded the original data into one table and then used the `INSERT` statement to copy it to a new table with a different file format, partitioning scheme, and so on. See [LOAD DATA Statement](#) on page 212.
- Improvements to Impala-HBase integration:
  - New query options for HBase performance: [HBASE\\_CACHE\\_BLOCKS](#) and [HBASE\\_CACHING](#).
  - Support for binary data types in HBase tables. See [Supported Data Types for HBase Columns](#) on page 427 for details.
- You can issue `REFRESH` as a SQL statement through any of the programming interfaces that Impala supports. `REFRESH` formerly had to be issued as a command through the `impala-shell` interpreter, and was not available through a JDBC or ODBC API call. As part of this change, the functionality of the `REFRESH` statement is divided between two statements. In Impala 1.1, `REFRESH` requires a table name argument and immediately reloads the metadata; the new `INVALIDATE METADATA` statement works the same as the Impala 1.0 `REFRESH` did: the table name argument is optional, and the metadata for one or all tables is marked as stale, but not actually reloaded until the table is queried. When you create a new table in the Hive shell or through a different Impala node, you must enter `INVALIDATE METADATA` with no table parameter before you can see the new table in `impala-shell`. See [REFRESH Statement](#) on page 215 and [INVALIDATE METADATA Statement](#) on page 210.

## New Features in Impala Version 1.0.1

The primary enhancements in Impala 1.0.1 are internal, for compatibility with the new Cloudera Manager 4.6 release. Try out the new **Impala Query Monitoring** feature in Cloudera Manager 4.6, which requires Impala 1.0.1.

New user-visible features include:

- The `VALUES` clause lets you `INSERT` one or more rows using literals, function return values, or other expressions. For performance and scalability, you should still use `INSERT ... SELECT` for bringing large quantities of data into an Impala table. The `VALUES` clause is a convenient way to set up small tables, particularly for initial testing of SQL features that do not require large amounts of data. See [VALUES Clause](#) on page 207 for details.
- The `-B` and `-o` options of the `impala-shell` command can turn query results into delimited text files and store them in an output file. The plain text results are useful for using with other Hadoop components or Unix tools. In benchmark tests, it is also faster to produce plain rather than pretty-printed results, and write to a file rather than to the screen, giving a more accurate picture of the actual query time.
- Several bug fixes. See [Issues Fixed in the 1.0.1 Release](#) on page 525 for details.

## New Features in Impala Version 1.0

This version has multiple performance improvements and adds the following functionality:

- Several bug fixes. See [Issues Fixed in the 1.0 GA Release](#) on page 526.
- [ALTER\\_TABLE](#) statement.
- [Hints](#) to allow specifying a particular join strategy.
- [REFRESH](#) for a single table.
- Dynamic resource management, allowing high concurrency for Impala queries.

## New Features in Version 0.7 of the Cloudera Impala Beta Release

This version has multiple performance improvements and adds the following functionality:

- Several bug fixes. See [Issues Fixed in Version 0.7 of the Beta Release](#) on page 529.
- Support for the Parquet file format. For more information on file formats, see [How Impala Works with Hadoop File Formats](#) on page 399.
- Added support for Avro.
- Support for the memory limits. For more information, see the example on modifying memory limits in [Modifying Impala Startup Options](#) on page 48.
- Bigger and faster joins through the addition of partitioned joins to the already supported broadcast joins.
- Fully distributed aggregations.
- Fully distributed top-n computation.
- Support for creating and altering tables.
- Support for GROUP BY with floats and doubles.

In this version, both CDH 4.1 and 4.2 are supported, but due to performance improvements added, we highly recommend you use CDH 4.2 or higher to see the full benefit. If you are using Cloudera Manager, version 4.5 is required.

## New Features in Version 0.6 of the Cloudera Impala Beta Release

- Several bug fixes. See [Issues Fixed in Version 0.6 of the Beta Release](#) on page 530.
- Added support for Impala on SUSE and Debian/Ubuntu. Impala is now supported on:
  - RHEL5.7/6.2 and Centos5.7/6.2
  - SUSE 11 with Service Pack 1 or higher
  - Ubuntu 10.04/12.04 and Debian 6.03
- Cloudera Manager 4.5 and CDH 4.2 support Impala 0.6.
- Support for the RCFile file format. For more information on file formats, see [Understanding File Formats](#).

## New Features in Version 0.5 of the Cloudera Impala Beta Release

- Several bug fixes. See [Issues Fixed in Version 0.5 of the Beta Release](#) on page 531.
- Added support for a JDBC driver that allows you to access Impala from a Java client. To use this feature, follow the instructions in [Configuring Impala to Work with JDBC](#) on page 39 to install the JDBC driver JARs on the client machine and modify the `CLASSPATH` on the client to include the JARs.

## New Features in Version 0.4 of the Cloudera Impala Beta Release

- Several bug fixes. See [Issues Fixed in Version 0.4 of the Beta Release](#) on page 532.
- Added support for Impala on RHEL5.7/Centos5.7. Impala is now supported on RHEL5.7/6.2 and Centos5.7/6.2.
- Cloudera Manager 4.1.3 supports Impala 0.4.
- The Impala debug webserver now has the ability to serve static files from  `${IMPALA_HOME} /www`. This can be disabled by setting `--enable_webserver_doc_root=false` on the command line. As a result, Impala

now uses the Twitter Bootstrap library to style its debug webpages, and the /queries page now tracks the last 25 queries run by each Impala daemon.

- Additional metrics available on the Impala Debug Webpage.

## New Features in Version 0.3 of the Cloudera Impala Beta Release

- Several bug fixes. See [Issues Fixed in Version 0.3 of the Beta Release](#) on page 533.
- The state-store-service binary has been renamed statestored.
- The location of the Impala configuration files has changed from the /usr/lib/impala/conf directory to the /etc/impala/conf directory.

## New Features in Version 0.2 of the Cloudera Impala Beta Release

- Several bug fixes. See [Issues Fixed in Version 0.2 of the Beta Release](#) on page 534.
- **Added Default Query Options** Default query options override all default QueryOption values when starting impalad. The format is:

```
-default_query_options='key=value;key=value'
```

## Incompatible Changes in Impala

The Impala version covered by this documentation library contains the following incompatible changes. These are things such as file format changes, removed features, or changes to implementation, default configuration, dependencies, or prerequisites that could cause issues during or after an Impala upgrade.

Even added SQL statements or clauses can produce incompatibilities, if you have databases, tables, or columns whose names conflict with the new keywords.

### Incompatible Changes Introduced in Impala for CDH 5.4.x

No incompatible changes. CDH maintenance releases such as 5.4.1, 5.4.2, and so on are exclusively bug fix releases. See [Incompatible Changes Introduced in Impala 2.2.0 / CDH 5.4.0](#) on page 487 for the most recent set of Impala incompatible changes.

- **Note:** The Impala 2.2.x maintenance releases now use the CDH 5.4.x numbering system rather than increasing the Impala version numbers. Impala 2.2 and higher are not available under CDH 4.

### Incompatible Changes Introduced in Impala 2.2.0 / CDH 5.4.0

- **Note:** Impala 2.2.0 is available as part of CDH 5.4.0 and is not available for CDH 4. Cloudera does not intend to release future versions of Impala for CDH 4 outside patch and maintenance releases if required. Given the upcoming end-of-maintenance for CDH 4, Cloudera recommends all customers to migrate to a recent CDH 5 release.

#### Changes to File Handling

Impala queries ignore files with extensions commonly used for temporary work files by Hadoop tools. Any files with extensions .tmp or .copying are not considered part of the Impala table. The suffix matching is case-insensitive, so for example Impala ignores both .copying and .COPYING suffixes.

The log rotation feature in Impala 2.2.0 and higher means that older log files are now removed by default. The default is to preserve the latest 10 log files for each severity level, for each Impala-related daemon. If you have set up your own log rotation processes that expect older files to be present, either adjust your procedures or change the Impala -max\_log\_files setting. See [Rotating Impala Logs](#) on page 444 for details.

## Cloudera Impala Release Notes

### Changes to Prerequisites

The prerequisite for CPU architecture has been relaxed in Impala 2.2.0 and higher. From this release onward, Impala works on CPUs that have the SSSE3 instruction set. The SSE4 instruction set is no longer required. This relaxed requirement simplifies the upgrade planning from Impala 1.x releases, which also worked on SSSE3-enabled processors.

### Incompatible Changes Introduced in Cloudera Impala 2.1.4 / CDH 5.3.4

No incompatible changes. Because CDH 5.3.5 does not include any code changes for Impala, Impala 2.1.4 is included with both CDH 5.3.4 and 5.3.5.

- **Note:** This Impala maintenance release is only available as part of CDH 5, not under CDH 4.

### Incompatible Changes Introduced in Impala 2.1.3 / CDH 5.3.3

No incompatible changes.

- **Note:** Impala 2.1.3 is available as part of CDH 5.3.3, not under CDH 4.

### Incompatible Changes Introduced in Impala 2.1.2 / CDH 5.3.2

No incompatible changes.

- **Note:** Impala 2.1.2 is available as part of CDH 5.3.2, not under CDH 4.

### Incompatible Changes Introduced in Impala 2.1.1 / CDH 5.3.1

No incompatible changes.

### Incompatible Changes Introduced in Impala 2.1.0 / CDH 5.3.0

#### Changes to Prerequisites

Currently, Impala 2.1.x does not function on CPUs without the SSE4.1 instruction set. This minimum CPU requirement is higher than in previous versions, which relied on the older SSSE3 instruction set. Check the CPU level of the hosts in your cluster before upgrading to Impala 2.1.x or CDH 5.3.x.

#### Changes to Output Format

The “small query” optimization feature introduces some new information in the EXPLAIN plan, which you might need to account for if you parse the text of the plan output.

#### New Reserved Words

New SQL syntax introduces additional reserved words: FOR, GRANT, REVOKE, ROLE, ROLES, INCREMENTAL.

### Incompatible Changes Introduced in Cloudera Impala 2.0.5 / CDH 5.2.6

No incompatible changes.

- **Note:** Impala 2.0.5 is available as part of CDH 5.2.6, not under CDH 4.

### Incompatible Changes Introduced in Impala 2.0.4 / CDH 5.2.5

No incompatible changes.

- **Note:** Impala 2.0.4 is available as part of CDH 5.2.5, not under CDH 4.

## Incompatible Changes Introduced in Impala 2.0.3 / CDH 5.2.4

- **Note:** Impala 2.0.3 is available as part of CDH 5.2.4, not under CDH 4.

## Incompatible Changes Introduced in Impala 2.0.2 / CDH 5.2.3

No incompatible changes.

- **Note:** Impala 2.0.2 is available as part of CDH 5.2.3, not under CDH 4.

## Incompatible Changes Introduced in Impala 2.0.1 / CDH 5.2.1

- The `INSERT` statement has always left behind a hidden work directory inside the data directory of the table. Formerly, this hidden work directory was named `.impala_insert_staging`. In Impala 2.0.1 and later, this directory name is changed to `_impala_insert_staging`. (While HDFS tools are expected to treat names beginning either with underscore and dot as hidden, in practice names beginning with an underscore are more widely supported.) If you have any scripts, cleanup jobs, and so on that rely on the name of this work directory, adjust them to use the new name.
- The `abs()` function now takes a broader range of numeric types as arguments, and the return type is the same as the argument type.
- Shorthand notation for character classes in regular expressions, such as `\d` for digit, are now available again in regular expression operators and functions such as `regexp_extract()` and `regexp_replace()`. Some other differences in regular expression behavior remain between Impala 1.x and Impala 2.x releases. See [Incompatible Changes Introduced in Impala 2.0.0 / CDH 5.2.0](#) on page 489 for details.

## Incompatible Changes Introduced in Impala 2.0.0 / CDH 5.2.0

### Changes to Prerequisites

Currently, Impala 2.0.x does not function on CPUs without the SSE4.1 instruction set. This minimum CPU requirement is higher than in previous versions, which relied on the older SSSE3 instruction set. Check the CPU level of the hosts in your cluster before upgrading to Impala 2.0.x or CDH 5.2.x.

### Changes to Query Syntax

The new syntax where query hints are allowed in comments causes some changes in the way comments are parsed in the `impala-shell` interpreter. Previously, you could end a -- comment line with a semicolon and `impala-shell` would treat that as a no-op statement. Now, a comment line ending with a semicolon is passed as an empty statement to the Impala daemon, where it is flagged as an error.

Impala 2.0 and later uses a different support library for regular expression parsing than in earlier Impala versions. Now, Impala uses the [Google RE2 library](#) rather than Boost for evaluating regular expressions. This implementation change causes some differences in the allowed regular expression syntax, and in the way certain regex operators are interpreted. The following are some of the major differences (not necessarily a complete list):

- `. *?` notation for non-greedy matches is now supported, where it was not in earlier Impala releases.
- By default, `^` and `$` now match only begin/end of buffer, not begin/end of each line. This behavior can be overridden in the regex itself using the `m` flag.
- By default, `.` does not match newline. This behavior can be overridden in the regex itself using the `s` flag.

## Cloudera Impala Release Notes

- \z is not supported.
- < and > for start of word and end of word are not supported.
- Lookahead and lookbehind are not supported.
- Shorthand notation for character classes, such as \d for digit, is not recognized. (This restriction is lifted in Impala 2.0.1, which restores the shorthand notation.)

### Changes to Output Format

In Impala 2.0 and later, `user()` returns the full Kerberos principal string, such as `user@example.com`, in a Kerberized environment.

The changed format for the user name in secure environments is also reflected where the user name is displayed in the output of the `PROFILE` command.

In the output from `SHOW FUNCTIONS`, `SHOW AGGREGATE FUNCTIONS`, and `SHOW ANALYTIC FUNCTIONS`, arguments and return types of arbitrary `DECIMAL` scale and precision are represented as `DECIMAL(*,*)`. Formerly, these items were displayed as `DECIMAL(-1,-1)`.

### Changes to Query Options

The `PARQUET_COMPRESSION_CODEC` query option has been replaced by the `COMPRESSION_CODEC` query option. See [COMPRESSION\\_CODEC Query Option](#) on page 341 for details.

### Changes to Configuration Options

The meaning of the `--idle_query_timeout` configuration option is changed, to accommodate the new `QUERY_TIMEOUT_S` query option. Rather than setting an absolute timeout period that applies to all queries, it now sets a maximum timeout period, which can be adjusted downward for individual queries by specifying a value for the `QUERY_TIMEOUT_S` query option. In sessions where no `QUERY_TIMEOUT_S` query option is specified, the `--idle_query_timeout` period applies the same as in earlier versions.

The `--strict_unicode` option of `impala-shell` was removed. To avoid problems with Unicode values in `impala-shell`, define the following locale setting before running `impala-shell`:

```
-----  
| export LC_CTYPE=en_US.UTF-8  
-----
```

### New Reserved Words

Some new SQL syntax requires the addition of new reserved words: `ANTI`, `ANALYTIC`, `OVER`, `PRECEDING`, `UNBOUNDED`, `FOLLOWING`, `CURRENT`, `ROWS`, `RANGE`, `CHAR`, `VARCHAR`.

### Changes to Data Files

The default Parquet block size for Impala is changed from 1 GB to 256 MB. This change could have implications for the sizes of Parquet files produced by `INSERT` and `CREATE TABLE AS SELECT` statements.

Although older Impala releases typically produced files that were smaller than the old default size of 1 GB, now the file size matches more closely whatever value is specified for the `PARQUET_FILE_SIZE` query option. Thus, if you use a non-default value for this setting, the output files could be larger than before. They still might be somewhat smaller than the specified value, because Impala makes conservative estimates about the space needed to represent each column as it encodes the data.

When you do not specify an explicit value for the `PARQUET_FILE_SIZE` query option, Impala tries to keep the file size within the 256 MB default size, but Impala might adjust the file size to be somewhat larger if needed to accommodate the layout for `wide` tables, that is, tables with hundreds or thousands of columns.

This change is unlikely to affect memory usage while writing Parquet files, because Impala does not pre-allocate the memory needed to hold the entire Parquet block.

## Incompatible Changes Introduced in Impala 1.4.4 / CDH 5.1.5

No incompatible changes.

- **Note:** Impala 1.4.4 is available as part of CDH 5.1.5, not under CDH 4.

## Incompatible Changes Introduced in Impala 1.4.3 / CDH 5.1.4

No incompatible changes. The SSL security fix does not require any change in the way you interact with Impala.

- **Note:** Impala 1.4.3 is available as part of CDH 5.1.4, and under CDH 4.

## Incompatible Changes Introduced in Impala 1.4.2 / CDH 5.1.3

None. Impala 1.4.2 is purely a bug-fix release. It does not include any incompatible changes.

- **Note:** Impala 1.4.2 is only available as part of CDH 5.1.3, not under CDH 4.

## Incompatible Changes Introduced in Impala 1.4.1 / CDH 5.1.2

None. Impala 1.4.1 is purely a bug-fix release. It does not include any incompatible changes.

## Incompatible Changes Introduced in Impala 1.4.0 / CDH 5.1.0

- There is a slight change to required security privileges in the Sentry framework. To create a new object, now you need the `ALL` privilege on the parent object. For example, to create a new table, view, or function requires having the `ALL` privilege on the database containing the new object.
- With the ability of `ORDER BY` queries to process unlimited amounts of data with no `LIMIT` clause, the query options `DEFAULT_ORDER_BY_LIMIT` and `ABORT_ON_DEFAULT_LIMIT_EXCEEDED` are now deprecated and have no effect.
- There are some changes to the list of reserved words. The following keywords are new:

- `API_VERSION`
- `BINARY`
- `CACHED`
- `CLASS`
- `PARTITIONS`
- `PRODUCED`
- `UNCACHED`

The following were formerly reserved keywords, but are no longer reserved:

- `COUNT`
- `GROUP_CONCAT`
- `NDV`
- `SUM`

- The fix for issue [IMPALA-973](#) changes the behavior of the `INVALIDATE_METADATA` statement regarding nonexistent tables. In Impala 1.4.0 and higher, the statement returns an error if the specified table is not in the metastore database at all. It completes successfully if the specified table is in the metastore database but not yet recognized by Impala, for example if the table was created through Hive. Formerly, you could issue this statement for a completely nonexistent table, with no error.

### Incompatible Changes Introduced in Impala 1.3.3 / CDH 5.0.5

No incompatible changes. The SSL security fix does not require any change in the way you interact with Impala.

- **Note:** Impala 1.3.3 is only available as part of CDH 5.0.5, not under CDH 4.

### Incompatible Changes Introduced in Impala 1.3.2 / CDH 5.0.4

With the fix for IMPALA-1019, you can use HDFS caching for files that are accessed by Impala.

- **Note:** Impala 1.3.2 is only available as part of CDH 5.0.4, not under CDH 4.

### Incompatible Changes Introduced in Impala 1.3.1 / CDH 5.0.3

- In Impala 1.3.1 and higher, the `REGEXP` and `RLIKE` operators now match a regular expression string that occurs anywhere inside the target string, the same as if the regular expression was enclosed on each side by `.*`. See [REGEXP Operator](#) on page 152 for examples. Previously, these operators only succeeded when the regular expression matched the entire target string. This change improves compatibility with the regular expression support for popular database systems. There is no change to the behavior of the `regexp_extract()` and `regexp_replace()` built-in functions.
- The result set for the `SHOW FUNCTIONS` statement includes a new first column, with the data type of the return value.

### Incompatible Changes Introduced in Impala 1.3.0 / CDH 5.0.0

- The `EXPLAIN_LEVEL` query option now accepts numeric options from 0 (most concise) to 3 (most verbose), rather than only 0 or 1. If you formerly used `SET EXPLAIN_LEVEL=1` to get detailed explain plans, switch to `SET EXPLAIN_LEVEL=3`. If you used the mnemonic keyword (`SET EXPLAIN_LEVEL=verbose`), you do not need to change your code because now level 3 corresponds to `verbose`.
- The keyword `DECIMAL` is now a reserved word. If you have any databases, tables, columns, or other objects already named `DECIMAL`, quote any references to them using backticks (`\``) to avoid name conflicts with the keyword.

- **Note:** Although the `DECIMAL` keyword is a reserved word, currently Impala does not support `DECIMAL` as a data type for columns.

- The query option named `YARN_POOL` during the CDH 5 beta period is now named `REQUEST_POOL` to reflect its broader use with the Impala admission control feature.
- There are some changes to the list of reserved words.
  - The names of aggregate functions are no longer reserved words, so you can have databases, tables, columns, or other objects named `AVG`, `MIN`, and so on without any name conflicts.
  - The internal function names `DISTINCTPC` and `DISTINCTPCSA` are no longer reserved words, although `DISTINCT` is still a reserved word.
  - The keywords `CLOSE_FN` and `PREPARE_FN` are now reserved words.
- The HDFS property `dfs.client.file-block-storage-locations.timeout` was renamed to `dfs.client.file-block-storage-locations.timeout.millis`, to emphasize that the unit of measure is milliseconds, not seconds. Impala requires a timeout of at least 10 seconds, making the minimum value for this setting 10000. On systems not managed by Cloudera Manager, you might need to edit the `hdfs-site.xml` file in the Impala configuration directory for the new name and minimum value.

## Incompatible Changes Introduced in Impala 1.2.4

There are no incompatible changes introduced in Impala 1.2.4.

Previously, after creating a table in Hive, you had to issue the `INVALIDATE METADATA` statement with no table name, a potentially expensive operation on clusters with many databases, tables, and partitions. Starting in Impala 1.2.4, you can issue the statement `INVALIDATE METADATA table_name` for a table newly created through Hive. Loading the metadata for only this one table is faster and involves less network overhead. Therefore, you might revisit your setup DDL scripts to add the table name to `INVALIDATE METADATA` statements, in cases where you create and populate the tables through Hive before querying them through Impala.

## Incompatible Changes Introduced in Impala 1.2.3

Because the feature set of Impala 1.2.3 is identical to Impala 1.2.2, there are no new incompatible changes. See [Incompatible Changes Introduced in Impala 1.2.2](#) on page 493 if you are upgrading from Impala 1.2.1 or 1.1.x.

## Incompatible Changes Introduced in Impala 1.2.2

The following changes to SQL syntax and semantics in Impala 1.2.2 could require updates to your SQL code, or schema objects such as tables or views:

- With the addition of the `CROSS JOIN` keyword, you might need to rewrite any queries that refer to a table named `CROSS` or use the name `CROSS` as a table alias:
 

```
-- Formerly, 'cross' in this query was an alias for t1
-- and it was a normal join query.
-- In 1.2.2 and higher, CROSS JOIN is a keyword, so 'cross'
-- is not interpreted as a table alias, and the query
-- uses the special CROSS JOIN processing rather than a
-- regular join.
select * from t1 cross join t2...

-- Now if CROSS is used in other context such as a table or column name,
-- use backticks to escape it.
create table `cross` (x int);
select * from `cross`;
```
- Formerly, a `DROP DATABASE` statement in Impala would not remove the top-level HDFS directory for that database. The `DROP DATABASE` has been enhanced to remove that directory. (You still need to drop all the tables inside the database first; this change only applies to the top-level directory for the entire database.)
- The keyword `PARQUET` is introduced as a synonym for `PARQUETFILE` in the `CREATE TABLE` and `ALTER TABLE` statements, because that is the common name for the file format. (As opposed to `SequenceFile` and `RCFile` where the "File" suffix is part of the name.) Documentation examples have been changed to prefer the new shorter keyword. The `PARQUETFILE` keyword is still available for backward compatibility with older Impala versions.
- New overloads are available for several operators and built-in functions, allowing you to insert their result values into smaller numeric columns such as `INT`, `SMALLINT`, `TINYINT`, and `FLOAT` without using a `CAST()` call. If you remove the `CAST()` calls from `INSERT` statements, those statements might not work with earlier versions of Impala.

Because many users are likely to upgrade straight from Impala 1.x to Impala 1.2.2, also read [Incompatible Changes Introduced in Impala 1.2.1](#) on page 494 for things to note about upgrading to Impala 1.2.x in general.

In a Cloudera Manager environment, the catalog service is not recognized or managed by Cloudera Manager versions prior to 4.8. Cloudera Manager 4.8 and higher require the catalog service to be present for Impala. Therefore, if you upgrade to Cloudera Manager 4.8 or higher, you must also upgrade Impala to 1.2.1 or higher. Likewise, if you upgrade Impala to 1.2.1 or higher, you must also upgrade Cloudera Manager to 4.8 or higher.

### Incompatible Changes Introduced in Impala 1.2.1

The following changes to SQL syntax and semantics in Impala 1.2.1 could require updates to your SQL code, or schema objects such as tables or views:

- In Impala 1.2.1 and higher, all `NULL` values come at the end of the result set for `ORDER BY ... ASC` queries, and at the beginning of the result set for `ORDER BY ... DESC` queries. In effect, `NULL` is considered greater than all other values for sorting purposes. The original Impala behavior always put `NULL` values at the end, even for `ORDER BY ... DESC` queries. The new behavior in Impala 1.2.1 makes Impala more compatible with other popular database systems. In Impala 1.2.1 and higher, you can override or specify the sorting behavior for `NULL` by adding the clause `NULLS FIRST` or `NULLS LAST` at the end of the `ORDER BY` clause.

Impala 1.2.1 goes along with CDH 4.5 and Cloudera Manager 4.8. If you used the beta version Impala 1.2.0 that came with the beta of CDH 5, Impala 1.2.1 includes all the features of Impala 1.2.0 except for resource management, which relies on the YARN framework from CDH 5.

The new `catalogd` service might require changes to any user-written scripts that stop, start, or restart Impala services, install or upgrade Impala packages, or issue `REFRESH` or `INVALIDATE_METADATA` statements:

- See [Installing Impala](#) on page 30, [Upgrading Impala](#) on page 43 and [Starting Impala](#) on page 47, for usage information for the `catalogd` daemon.
- The `REFRESH` and `INVALIDATE_METADATA` statements are no longer needed when the `CREATE TABLE`, `INSERT`, or other table-changing or data-changing operation is performed through Impala. These statements are still needed if such operations are done through Hive or by manipulating data files directly in HDFS, but in those cases the statements only need to be issued on one Impala node rather than on all nodes. See [REFRESH Statement](#) on page 215 and [INVALIDATE\\_METADATA Statement](#) on page 210 for the latest usage information for those statements.
- See [The Impala Catalog Service](#) on page 19 for background information on the `catalogd` service.

In a Cloudera Manager environment, the catalog service is not recognized or managed by Cloudera Manager versions prior to 4.8. Cloudera Manager 4.8 and higher require the catalog service to be present for Impala. Therefore, if you upgrade to Cloudera Manager 4.8 or higher, you must also upgrade Impala to 1.2.1 or higher. Likewise, if you upgrade Impala to 1.2.1 or higher, you must also upgrade Cloudera Manager to 4.8 or higher.

### Incompatible Changes Introduced in Impala 1.2.0 (Beta)

There are no incompatible changes to SQL syntax in Impala 1.2.0 (beta).

Because Impala 1.2.0 is bundled with the CDH 5 beta download and depends on specific levels of Apache Hadoop components supplied with CDH 5, you can only install it in combination with the CDH 5 beta.

The new `catalogd` service might require changes to any user-written scripts that stop, start, or restart Impala services, install or upgrade Impala packages, or issue `REFRESH` or `INVALIDATE_METADATA` statements:

- See [Installing Impala](#) on page 30, [Upgrading Impala](#) on page 43 and [Starting Impala](#) on page 47, for usage information for the `catalogd` daemon.
- The `REFRESH` and `INVALIDATE_METADATA` statements are no longer needed when the `CREATE TABLE`, `INSERT`, or other table-changing or data-changing operation is performed through Impala. These statements are still needed if such operations are done through Hive or by manipulating data files directly in HDFS, but in those cases the statements only need to be issued on one Impala node rather than on all nodes. See [REFRESH Statement](#) on page 215 and [INVALIDATE\\_METADATA Statement](#) on page 210 for the latest usage information for those statements.
- See [The Impala Catalog Service](#) on page 19 for background information on the `catalogd` service.

The new resource management feature interacts with both YARN and Llama services, which are available in CDH 5. These services are set up for you automatically in a Cloudera Manager (CM) environment. For information

about setting up the YARN and Llama services, see the instructions for [YARN](#) and [Llama](#) in the *CDH 5 Documentation*.

## Incompatible Changes Introduced in Impala 1.1.1

There are no incompatible changes in Impala 1.1.1.

Previously, it was not possible to create Parquet data through Impala and reuse that table within Hive. Now that Parquet support is available for Hive 10, reusing existing Impala Parquet data files in Hive requires updating the table metadata. Use the following command if you are already running Impala 1.1.1:

```
ALTER TABLE table_name SET FILEFORMAT PARQUETFILE;
```

If you are running a level of Impala that is older than 1.1.1, do the metadata update through Hive:

```
ALTER TABLE table_name SET SERDE 'parquet.hive.serde.ParquetHiveSerDe';
ALTER TABLE table_name SET FILEFORMAT
  INPUTFORMAT "parquet.hive.DeprecatedParquetInputFormat"
  OUTPUTFORMAT "parquet.hive.DeprecatedParquetOutputFormat";
```

Impala 1.1.1 and higher can reuse Parquet data files created by Hive, without any action required.

As usual, make sure to upgrade the `impala-lzo-cdh4` package to the latest level at the same time as you upgrade the Impala server.

## Incompatible Change Introduced in Cloudera Impala 1.1

- The `REFRESH` statement now requires a table name; in Impala 1.0, the table name was optional. This syntax change is part of the internal rework to make `REFRESH` a true Impala SQL statement so that it can be called through the JDBC and ODBC APIs. `REFRESH` now reloads the metadata immediately, rather than marking it for update the next time any affected table is accessed. The previous behavior, where omitting the table name caused a refresh of the entire Impala metadata catalog, is available through the new `INVALIDATE METADATA` statement. `INVALIDATE METADATA` can be specified with a table name to affect a single table, or without a table name to affect the entire metadata catalog; the relevant metadata is reloaded the next time it is requested during the processing for a SQL statement. See [REFRESH Statement](#) on page 215 and [INVALIDATE METADATA Statement](#) on page 210 for the latest details about these statements.

## Incompatible Changes Introduced in Impala 1.0

- If you use LZO-compressed text files, when you upgrade Impala to version 1.0, also update the `impala-lzo-cdh4` to the latest level. See [Using LZO-Compressed Text Files](#) on page 404 for details.
- Cloudera Manager 4.5.2 and higher only supports Impala 1.0 and higher, and vice versa. If you upgrade to Impala 1.0 or higher managed by Cloudera Manager, you must also upgrade Cloudera Manager to version 4.5.2 or higher. If you upgrade from an earlier version of Cloudera Manager, and were using Impala, you must also upgrade Impala to version 1.0 or higher. The beta versions of Impala are no longer supported as of the release of Impala 1.0.

## Incompatible Change Introduced in Version 0.7 of the Cloudera Impala Beta Release

- The defaults for the `-nn` and `-nn_port` flags have changed and are now read from `core-site.xml`. Impala prints the values of `-nn` and `-nn_port` to the log when it starts. The ability to set `-nn` and `-nn_port` on the command line is deprecated in 0.7 and may be removed in Impala 0.8.

## Incompatible Change Introduced in Version 0.6 of the Cloudera Impala Beta Release

- Cloudera Manager 4.5 supports only version 0.6 of the Cloudera Impala Beta Release. It does not support the earlier beta versions. If you upgrade your Cloudera Manager installation, you must also upgrade Impala to beta version 0.6. If you upgrade Impala to beta version 0.6, you must upgrade Cloudera Manager to 4.5.

### Incompatible Change Introduced in Version 0.4 of the Cloudera Impala Beta Release

- Cloudera Manager 4.1.3 supports only version 0.4 of the Cloudera Impala Beta Release. It does not support the earlier beta versions. If you upgrade your Cloudera Manager installation, you must also upgrade Impala to beta version 0.4. If you upgrade Impala to beta version 0.4, you must upgrade Cloudera Manager to 4.1.3.

### Incompatible Change Introduced in Version 0.3 of the Cloudera Impala Beta Release

- Cloudera Manager 4.1.2 supports only version 0.3 of the Cloudera Impala Beta Release. It does not support the earlier beta versions. If you upgrade your Cloudera Manager installation, you must also upgrade Impala to beta version 0.3. If you upgrade Impala to beta version 0.3, you must upgrade Cloudera Manager to 4.1.2.

## Known Issues and Workarounds in Impala

The following sections describe known issues and workarounds in Impala.

For issues fixed in various Impala releases, see [Fixed Issues in Impala](#) on page 500.

### Known Issues in the Current Production Release (Impala 2.2.x / CDH 5.4.x)

These known issues affect the current release. Any workarounds are listed here. The bug links take you to the Impala issues site, where you can see the diagnosis and whether a fix is in the pipeline.

#### invalid tuple\_idx when combining INSERT INTO with analytic subquery

An `INSERT ... SELECT` statement could encounter an error if the `SELECT` portion included an analytic function call.

**Bug:** [IMPALA-1737](#)

**Severity:** Major

**Workaround:** Rewrite the statement as a `CREATE TABLE AS SELECT` statement.

#### Support individual memory allocations larger than 1 GB

The largest single block of memory that Impala can allocate during a query is 1 GiB. Therefore, a query could fail or Impala could crash if a compressed text file resulted in more than 1 GiB of data in uncompressed form, or if a string function such as `group_concat()` returned a value greater than 1 GiB.

**Bug:** [IMPALA-1619](#)

**Severity:** Major

#### Can't update stats manually via alter table after upgrading to CDH 5.2

**Bug:** [IMPALA-1420](#)

**Severity:** High

**Workaround:** On CDH 5.2, when adjusting table statistics manually by setting the `numRows`, you must also enable the Boolean property `STATS_GENERATED_VIA_STATS_TASK`. For example, use a statement like the following to set both properties with a single `ALTER TABLE` statement:

```
-----  
| ALTER TABLE table_name SET TBLPROPERTIES('numRows'='new_value',  
|   'STATS_GENERATED_VIA_STATS_TASK' = 'true');  
-----
```

**Resolution:** The underlying cause is the issue [HIVE-8648](#) that affects the metastore in Hive 0.13. The workaround is only needed until the fix for this issue is incorporated into a CDH release.

## ORDER BY rand() does not work.

Because the value for `rand()` is computed early in a query, using an `ORDER BY` expression involving a call to `rand()` does not actually randomize the results.

**Bug:** [IMPALA-397](#)

**Severity:** High

## Impala BE cannot parse Avro schema that contains a trailing semi-colon

If an Avro table has a schema definition with a trailing semicolon, Impala encounters an error when the table is queried.

**Bug:** [IMPALA-1024](#)

**Severity:** High

## Process mem limit does not account for the JVM's memory usage

Some memory allocated by the JVM used internally by Impala is not counted against the memory limit for the `impalad` daemon.

**Bug:** [IMPALA-691](#)

**Severity:** High

**Workaround:** To monitor overall memory usage, use the `top` command, or add the memory figures in the Impala web UI `/memz` tab to JVM memory usage shown on the `/metrics` tab.

## Impala Parser issue when using fully qualified table names that start with a number.

A fully qualified table name starting with a number could cause a parsing error. In a name such as `db. 571_market`, the decimal point followed by digits is interpreted as a floating-point number.

**Bug:** [IMPALA-941](#)

**Severity:** High

**Workaround:** Surround each part of the fully qualified name with backticks (`\``).

## CatalogServer should not require HBase to be up to reload its metadata

If HBase is unavailable during Impala startup or after an `INVALIDATE METADATA` statement, the `catalogd` daemon could go into an error loop, making Impala unresponsive.

**Bug:** [IMPALA-788](#)

**Severity:** High

**Workaround:** For systems not managed by Cloudera Manager, add the following settings to `/etc/impala/conf/hbase-site.xml`:

```
<property>
  <name>hbase.client.retries.number</name>
  <value>3</value>
</property>
<property>
  <name>hbase.rpc.timeout</name>
  <value>3000</value>
</property>
```

Currently, Cloudera Manager does not have an Impala-only override for HBase settings, so any HBase configuration change you make through Cloudera Manager would affect for all HBase applications. Therefore, this change is not recommended on systems managed by Cloudera Manager.

### Kerberos tickets must be renewable

In a Kerberos environment, the `impalad` daemon might not start if Kerberos tickets are not renewable.

**Workaround:** Configure your KDC to allow tickets to be renewed, and configure `krb5.conf` to request renewable tickets.

### Avro Scanner fails to parse some schemas

Querying certain Avro tables could cause a crash or return no rows, even though Impala could `DESCRIBE` the table.

**Bug:** [IMPALA-635](#)

**Severity:** High

**Workaround:** Swap the order of the fields in the schema specification. For example, [ "null", "string" ] instead of [ "string", "null" ].

**Resolution:** Not allowing this syntax agrees with the Avro specification, so it may still cause an error even when the crashing issue is resolved.

### Configuration needed for Flume to be compatible with Impala

For compatibility with Impala, the value for the Flume HDFS Sink `hdfs.writeFormat` must be set to `Text`, rather than its default value of `Writable`. The `hdfs.writeFormat` setting must be changed to `Text` before creating data files with Flume; otherwise, those files cannot be read by either Impala or Hive.

**Severity:** High

**Resolution:** This information has been requested to be added to the upstream Flume documentation.

### Impala does not support running on clusters with federated namespaces

Impala does not support running on clusters with federated namespaces. The `impalad` process will not start on a node running such a filesystem based on the `org.apache.hadoop.fs.viewfs.ViewFs` class.

**Bug:** [IMPALA-77](#)

**Severity:** Undetermined

**Anticipated Resolution:** Limitation

**Workaround:** Use standard HDFS on all Impala nodes.

### Deviation from Hive behavior: Out of range values float/double values are returned as maximum allowed value of type (Hive returns NULL)

Impala behavior differs from Hive with respect to out of range float/double values. Out of range values are returned as maximum allowed value of type (Hive returns NULL).

**Severity:** Low

**Workaround:** None

### Deviation from Hive behavior: Impala does not do implicit casts between string and numeric and boolean types.

**Severity:** Low

**Anticipated Resolution:** None

**Workaround:** Use explicit casts.

## If Hue and Impala are installed on the same host, and if you configure Hue Beeswax in CDH 4.1 to execute Impala queries, Beeswax cannot list Hive tables and shows an error on Beeswax startup.

Hue requires Beeswaxd to be running in order to list the Hive tables. Because of a port conflict bug in Hue in CDH4.1, when Hue and Impala are installed on the same host, an error page is displayed when you start the Beeswax application, and when you open the **Tables** page in Beeswax.

**Severity:** High

**Anticipated Resolution:** Fixed in an upcoming CDH4 release

**Workarounds:** Choose one of the following workarounds (but only one):

- Install Hue and Impala on different hosts. *OR*
- Upgrade to CDH4.1.2 and add the following property in the `beeswax` section of the `/etc/hue/hue.ini` configuration file:

```
beeswax_meta_server_only=9004
```

*OR*

- If you are using CDH4.1.1 and you want to install Hue and Impala on the same host, change the code in this file:

```
/usr/share/hue/apps/beeswax/src/beeswax/management/commands/beeswax_server.py
```

Replace line 66:

```
str(beeswax.conf.BEESWAX_SERVER_PORT.get()),
```

With this line:

```
'8004',
```

Beeswaxd will then use port 8004.

■ **Note:**

If you used Cloudera Manager to install Impala, refer to the Cloudera Manager release notes for information about using an equivalent workaround by specifying the `beeswax_meta_server_only=9004` configuration value in the options field for Hue. In Cloudera Manager 4, these fields are labelled **Safety Valve**; in Cloudera Manager 5, they are called **Advanced Configuration Snippet**.

## Impala should tolerate bad locale settings

If the `LC_*` environment variables specify an unsupported locale, Impala does not start.

**Bug:** [IMPALA-532](#)

**Severity:** Low

**Workaround:** Add `LC_ALL="C"` to the environment settings for both the Impala daemon and the Statestore daemon. See [Modifying Impala Startup Options](#) on page 48 for details about modifying these environment settings.

**Resolution:** Fixing this issue would require an upgrade to Boost 1.47 in the Impala distribution.

### Log Level 3 Not Recommended for Impala

The extensive logging produced by log level 3 can cause serious performance overhead and capacity issues.

**Severity:** Low

**Workaround:** Reduce the log level to its default value of 1, that is, `GLOG_v=1`. See [Setting Logging Levels](#) on page 445 for details about the effects of setting different logging levels.

## Fixed Issues in Impala

The following sections describe the major issues fixed in each Impala release.

For known issues that are currently unresolved, see [Known Issues and Workarounds in Impala](#) on page 496.

### Issues Fixed in Impala for CDH 5.4.3

This section lists the most frequently encountered customer issues fixed in Impala for CDH 5.4.3.

- **Note:** The Impala 2.2.x maintenance releases now use the CDH 5.4.x numbering system rather than increasing the Impala version numbers. Impala 2.2 and higher are not available under CDH 4.

For the full list of fixed issues, see [the CDH 5.4.1 release notes](#).

#### Enable using Isilon as the underlying filesystem.

The certification of CDH and Impala with the Isilon filesystem involves a number of fixes to performance and flexibility for dealing with I/O using remote reads. See [Using Impala with Isilon Storage](#) on page 442 for details on using Impala and Isilon together.

**Bug:** [IMPALA-1968](#), [IMPALA-1730](#)

**Severity:** High

#### Expand set of supported timezones.

The set of timezones recognized by Impala was expanded. You can always find the latest list of supported timezones in the Impala source code, in the file `timezone_db.cc`.

**Bug:** [IMPALA-1381](#)

**Severity:** High

#### Impala Timestamp ISO-8601 Support.

Impala can now process `TIMESTAMP` literals including a trailing `z`, signifying "Zulu" time, a synonym for UTC.

**Bug:** [IMPALA-1963](#)

**Severity:** High

#### Fix wrong warning when insert overwrite to empty table

An `INSERT OVERWRITE` operation would encounter an error if the `SELECT` portion of the statement returned zero rows, such as with a `LIMIT 0` clause.

**Bug:** [IMPALA-2008](#)

**Severity:** High

## Expand parsing of decimals to include scientific notation

`DECIMAL` literals can now include e scientific notation. For example, now `CAST(1e3 AS DECIMAL(5,3))` is a valid expression. Formerly it returned `NULL`. Some scientific expressions might have worked before in `DECIMAL` context, but only when the scale was 0.

**Bug:** <https://issues.cloudera.org/browse/>

**Severity:** High

## Issues Fixed in Impala for CDH 5.4.1

This section lists the most frequently encountered customer issues fixed in Impala for CDH 5.4.1.

- **Note:** The Impala 2.2.x maintenance releases now use the CDH 5.4.x numbering system rather than increasing the Impala version numbers. Impala 2.2 and higher are not available under CDH 4.

For the full list of fixed issues, see [the CDH 5.4.1 release notes](#).

## Issues Fixed in the 2.2.0 Release / CDH 5.4.0

This section lists the most frequently encountered customer issues fixed in Impala 2.2.0.

For the full list of fixed issues in Impala 2.2.0, including over 40 critical issues, see [this report in the JIRA system](#).

- **Note:** Impala 2.2.0 is available as part of CDH 5.4.0 and is not available for CDH 4. Cloudera does not intend to release future versions of Impala for CDH 4 outside patch and maintenance releases if required. Given the upcoming end-of-maintenance for CDH 4, Cloudera recommends all customers to migrate to a recent CDH 5 release.

## Altering a column's type causes column stats to stop sticking for that column

When the type of a column was changed in either Hive or Impala through `ALTER TABLE CHANGE COLUMN`, the metastore database did not correctly propagate that change to the table that contains the column statistics. The statistics (particularly the `NDV`) for that column were permanently reset and could not be changed by Impala's `COMPUTE STATS` command. The underlying cause is a Hive bug (HIVE-9866).

**Bug:** [IMPALA-1607](#)

**Severity:** Major

**Resolution:** Resolved by incorporating the fix for [HIVE-9866](#).

**Workaround:** On systems without the corresponding Hive fix, change the column back to its original type. The stats reappear and you can recompute or drop them.

## Impala may leak or use too many file descriptors

If a file was truncated in HDFS without a corresponding `REFRESH` in Impala, Impala could allocate memory for file descriptors and not free that memory.

**Bug:** [IMPALA-1854](#)

**Severity:** High

## Spurious stale block locality messages

Impala could issue messages stating the block locality metadata was stale, when the metadata was actually fine. The internal "remote bytes read" counter was not being reset properly. This issue did not cause an actual slowdown in query execution, but the spurious error could result in unnecessary debugging work and unnecessary use of the `INVALIDATE METADATA` statement.

## Cloudera Impala Release Notes

**Bug:** [IMPALA-1712](#)

**Severity:** High

### DROP TABLE fails after COMPUTE STATS and ALTER TABLE RENAME to a different database.

When a table was moved from one database to another, the column statistics were not pointed to the new database. This could result in lower performance for queries due to unavailable statistics, and also an inability to drop the table.

**Bug:** [IMPALA-1711](#)

**Severity:** High

### IMPALA-1556 causes memory leak with secure connections

impalad daemons could experience a memory leak on clusters using Kerberos authentication, with memory usage growing as more data is transferred across the secure channel, either to the client program or between Impala nodes. The same issue affected LDAP-secured clusters to a lesser degree, because the LDAP security only covers data transferred back to client programs.

**Bug:** [IMPALA-1674](#)

**Severity:** High

### unix\_timestamp() does not return correct time

The unix\_timestamp( ) function could return an incorrect value (a constant value of 1).

**Bug:** [IMPALA-1623](#)

**Severity:** High

### Impala incorrectly handles text data missing a newline on the last line

Some queries did not recognize the final line of a text data file if the line did not end with a newline character. This could lead to inconsistent results, such as a different number of rows for SELECT COUNT(\*) as opposed to SELECT \*.

**Bug:** [IMPALA-1476](#)

**Severity:** High

### Impala's ACLs check do not consider all group ACLs, only checked first one.

If the HDFS user ID associated with the impalad process had read or write access in HDFS based on group membership, Impala statements could still fail with HDFS permission errors if that group was not the first listed group for that user ID.

**Bug:** [IMPALA-1805](#)

**Severity:** High

### Fix infinite loop opening or closing file with invalid metadata

Truncating a file in HDFS, after Impala had cached the file metadata, could produce a hang when Impala queried a table containing that file.

**Bug:** [IMPALA-1794](#)

**Severity:** High

### Cannot write Parquet files when values are larger than 64KB

Impala could sometimes fail to INSERT into a Parquet table if a column value such as a STRING was larger than 64 KB.

**Bug:** [IMPALA-1705](#)

**Severity:** High

### [Impala Will Not Run on Certain Intel CPUs](#)

This fix relaxes the CPU requirement for Impala. Now only the SSSE3 instruction set is required. Formerly, SSE4.1 instructions were generated, making Impala refuse to start on some older CPUs.

**Bug:** [IMPALA-1646](#)

**Severity:** High

### [Issues Fixed in the 2.1.4 Release / CDH 5.3.4](#)

This section lists the most significant Impala issues fixed in Impala 2.1.4 for CDH 5.3.4. Because CDH 5.3.5 does not include any code changes for Impala, Impala 2.1.4 is included with both CDH 5.3.4 and 5.3.5.

For the full list of Impala fixed issues in Impala 2.1.4 for CDH 5.3.4, see [Issues Fixed in CDH 5.3.4](#).

- **Note:** This Impala maintenance release is only available as part of CDH 5, not under CDH 4.

### [Crash: impala::TupleIsNotNullPredicate::Prepare](#)

When expressions that tested for `NULL` were used in combination with analytic functions, an error could occur. (The original crash issue was fixed by an earlier patch.)

**Bug:** [IMPALA-1519](#)

**Severity:** High

### [Expand parsing of decimals to include scientific notation](#)

`DECIMAL` literals could include e scientific notation. For example, now `CAST(1e3 AS DECIMAL(5,3))` is a valid expression. Formerly it returned `NULL`. Some scientific expressions might have worked before in `DECIMAL` context, but only when the scale was 0.

**Bug:** [IMPALA-1952](#)

**Severity:** High

### [INSERT/CTAS evaluates and applies constant predicates.](#)

An `INSERT OVERWRITE` statement would write new data, even if a constant clause such as `WHERE 1 = 0` should have prevented it from writing any rows.

**Bug:** [IMPALA-1860](#)

**Severity:** High

### [Assign predicates below analytic functions with a compatible partition by clause](#)

If the `PARTITION BY` clause in an analytic function refers to partition key columns in a partitioned table, now Impala can perform partition pruning during the analytic query.

**Bug:** [IMPALA-1900](#)

**Severity:** High

### [FIRST\\_VALUE may produce incorrect results with preceding windows](#)

A query using the `FIRST_VALUE` analytic function and a window defined with the `PRECEDING` keyword could produce wrong results.

**Bug:** [IMPALA-1888](#)

**Severity:** High

## [FIRST\\_VALUE rewrite fn type might not match slot type](#)

A query referencing a DECIMAL column with the FIRST\_VALUE analytic function could encounter an error.

**Bug:** [IMPALA-1559](#)

**Severity:** High

## [AnalyticEvalNode cannot handle partition/order by exprs with NaN](#)

A query using an analytic function could encounter an error if the evaluation of an analytic ORDER BY or PARTITION expression resulted in a NaN value, for example if the ORDER BY or PARTITION contained a division operation where both operands were zero.

**Bug:** [IMPALA-1808](#)

**Severity:** High

## [AnalyticEvalNode not properly handling nullable tuples](#)

An analytic function containing only an OVER clause could encounter an error if another part of the query (specifically an outer join) produced all-NULL tuples.

**Bug:** [IMPALA-1562](#)

**Severity:** High

## [Issues Fixed in the 2.1.3 Release / CDH 5.3.3](#)

This section lists the most significant issues fixed in Impala 2.1.3.

For the full list of fixed issues in Impala 2.1.3, see the [CDH 5.3.3 release notes](#).

- **Note:** Impala 2.1.3 is available as part of CDH 5.3.3, not under CDH 4.

## [Add compatibility flag for Hive-Parquet-Timestamps](#)

When Hive writes TIMESTAMP values, it represents them in the local time zone of the server. Impala expects TIMESTAMP values to always be in the UTC time zone, possibly leading to inconsistent results depending on which component created the data files. This patch introduces a new startup flag,

`-convert_legacy_hive_parquet_utc_timestamps` for the `impalad` daemon. Specify `-convert_legacy_hive_parquet_utc_timestamps=true` to make Impala recognize Parquet data files written by Hive and automatically adjust TIMESTAMP values read from those files into the UTC time zone for compatibility with other Impala TIMESTAMP processing. Although this setting is currently turned off by default, consider enabling it if practical in your environment, for maximum interoperability with Hive-created Parquet files.

**Bug:** [IMPALA-1658](#)

**Severity:** High

## [Use sprintf\(\) instead of lexical\\_cast\(\) in float-to-string casts](#)

Converting a floating-point value to a STRING could be slower than necessary.

**Bug:** [IMPALA-1738](#)

**Severity:** High

## Fix partition spilling cleanup when new stream OOMs

Certain calls to aggregate functions with STRING arguments could encounter a serious error when the system ran low on memory and attempted to activate the spill-to-disk mechanism. The error message referenced the function `impala::AggregateFunctions::StringValGetValue`.

**Bug:** [IMPALA-1865](#)

**Severity:** High

## Impala's ACLs check do not consider all group ACLs, only checked first one.

If the HDFS user ID associated with the `impalad` process had read or write access in HDFS based on group membership, Impala statements could still fail with HDFS permission errors if that group was not the first listed group for that user ID.

**Bug:** [IMPALA-1805](#)

**Severity:** High

## Fix infinite loop opening or closing file with invalid metadata

Truncating a file in HDFS, after Impala had cached the file metadata, could produce a hang when Impala queried a table containing that file.

**Bug:** [IMPALA-1794](#)

**Severity:** High

## external-data-source-executor leaking global jni refs

Successive calls to the data source API could result in excessive memory consumption, with memory allocated but never freed.

**Bug:** [IMPALA-1801](#)

**Severity:** High

## Spurious stale block locality messages

Impala could issue messages stating the block locality metadata was stale, when the metadata was actually fine. The internal “remote bytes read” counter was not being reset properly. This issue did not cause an actual slowdown in query execution, but the spurious error could result in unnecessary debugging work and unnecessary use of the `INVALIDATE_METADATA` statement.

**Bug:** [IMPALA-1712](#)

**Severity:** High

## Issues Fixed in the 2.1.2 Release / CDH 5.3.2

This section lists the most significant issues fixed in Impala 2.1.2.

For the full list of fixed issues in Impala 2.1.2, see [this report in the JIRA system](#).

- **Note:** Impala 2.1.2 is available as part of CDH 5.3.2, not under CDH 4.

## Impala incorrectly handles double numbers with more than 19 significant decimal digits

When a floating-point value was read from a text file and interpreted as a `FLOAT` or `DOUBLE` value, it could be incorrectly interpreted if it included more than 19 significant digits.

**Bug:** [IMPALA-1622](#)

**Severity:** High

## Cloudera Impala Release Notes

### [unix\\_timestamp\(\) does not return correct time](#)

The `unix_timestamp()` function could return an incorrect value (a constant value of 1).

**Bug:** [IMPALA-1623](#)

**Severity:** High

### [Row Count Mismatch: Partition pruning with NULL](#)

A query against a partitioned table could return incorrect results if the `WHERE` clause compared the partition key to `NULL` using operators such as `=` or `!=`.

**Bug:** [IMPALA-1535](#)

**Severity:** High

### [Fetch column stats in bulk using new \(Hive .13\) HMS APIs](#)

The performance of the `COMPUTE STATS` statement and queries was improved, particularly for wide tables.

**Bug:** [IMPALA-1120](#)

**Severity:** High

## [Issues Fixed in the 2.1.1 Release / CDH 5.3.1](#)

This section lists the most significant issues fixed in Impala 2.1.1.

For the full list of fixed issues in Impala 2.1.1, see [this report in the JIRA system](#).

### [IMPALA-1556 causes memory leak with secure connections](#)

`impalad` daemons could experience a memory leak on clusters using Kerberos authentication, with memory usage growing as more data is transferred across the secure channel, either to the client program or between Impala nodes. The same issue affected LDAP-secured clusters to a lesser degree, because the LDAP security only covers data transferred back to client programs.

**Bug:** <https://issues.cloudera.org/browse/IMPALA-1674> IMPALA-1674

**Severity:** High

### [TSaslServerTransport::Factory::getTransport\(\) leaks transport map entries](#)

`impalad` daemons in clusters secured by Kerberos or LDAP could experience a slight memory leak on each connection. The accumulation of unreleased memory could cause problems on long-running clusters.

**Bug:** [IMPALA-1668](#)

**Severity:** High

## [Issues Fixed in the 2.1.0 Release / CDH 5.3.0](#)

This section lists the most significant issues fixed in Impala 2.1.0.

For the full list of fixed issues in Impala 2.1.0, see [this report in the JIRA system](#).

### [Kerberos fetches 3x slower](#)

Transferring large result sets back to the client application on Kerberos

**Bug:** [IMPALA-1455](#)

**Severity:** High

## Compressed file needs to be held entirely in Memory

Queries on gzipped text files required holding the entire data file and its uncompressed representation in memory at the same time. `SELECT` and `COMPUTE STATS` statements could fail or perform inefficiently as a result. The fix enables streaming reads for gzipped text, so that the data is uncompressed as it is read.

**Bug:** [IMPALA-1556](#)

**Severity:** High

## Cannot read hbase metadata with NullPointerException: null

Impala might not be able to access HBase tables, depending on the associated levels of Impala and HBase on the system.

**Bug:** [IMPALA-1611](#)

**Severity:** High

## Serious errors / crashes

Improved code coverage in Impala testing uncovered a number of potentially serious errors that could occur with specific query syntax. These errors are resolved in Impala 2.1.

**Bug:** [IMPALA-1553](#), [IMPALA-1528](#), [IMPALA-1526](#), [IMPALA-1524](#), [IMPALA-1508](#), [IMPALA-1493](#), [IMPALA-1501](#), [IMPALA-1483](#)

**Severity:** High

## Issues Fixed in the 2.0.5 Release / CDH 5.2.6

For the full list of fixed issues in Impala 2.0.5, see [this report in the JIRA system](#).

- **Note:** Impala 2.0.5 is available as part of CDH 5.2.6, not under CDH 4.

## Issues Fixed in the 2.0.4 Release / CDH 5.2.5

This section lists the most significant issues fixed in Impala 2.0.4.

For the full list of fixed issues in Impala 2.0.4, see [this report in the JIRA system](#).

- **Note:** Impala 2.0.4 is available as part of CDH 5.2.5, not under CDH 4.

## Add compatibility flag for Hive-Parquet-Timestamps

When Hive writes `TIMESTAMP` values, it represents them in the local time zone of the server. Impala expects `TIMESTAMP` values to always be in the UTC time zone, possibly leading to inconsistent results depending on which component created the data files. This patch introduces a new startup flag, `-convert_legacy_hive_parquet_utc_timestamps` for the `impalad` daemon. Specify `-convert_legacy_hive_parquet_utc_timestamps=true` to make Impala recognize Parquet data files written by Hive and automatically adjust `TIMESTAMP` values read from those files into the UTC time zone for compatibility with other Impala `TIMESTAMP` processing. Although this setting is currently turned off by default, consider enabling it if practical in your environment, for maximum interoperability with Hive-created Parquet files.

**Bug:** [IMPALA-1658](#)

**Severity:** High

## IoMgr infinite loop opening/closing file when shorter than cached metadata size

If a table data file was replaced by a shorter file outside of Impala, such as with `INSERT OVERWRITE` in Hive producing an empty output file, subsequent Impala queries could hang.

## Cloudera Impala Release Notes

**Bug:** [IMPALA-1794](#)

**Severity:** High

### Issues Fixed in the 2.0.3 Release / CDH 5.2.4

This section lists the most significant issues fixed in Impala 2.0.3.

For the full list of fixed issues in Impala 2.0.3, see [this report in the JIRA system](#).

- **Note:** Impala 2.0.3 is available as part of CDH 5.2.4, not under CDH 4.

#### Anti join could produce incorrect results when spilling

An anti-join query (or a NOT EXISTS operation that was rewritten internally into an anti-join) could produce incorrect results if Impala reached its memory limit, causing the query to write temporary results to disk.

**Bug:** [IMPALA-1471](#)

**Severity:** High

#### Row Count Mismatch: Partition pruning with NULL

A query against a partitioned table could return incorrect results if the WHERE clause compared the partition key to NULL using operators such as = or !=.

**Bug:** [IMPALA-1535](#)

**Severity:** High

#### Fetch column stats in bulk using new (Hive .13) HMS APIs

The performance of the COMPUTE STATS statement and queries was improved, particularly for wide tables.

**Bug:** [IMPALA-1120](#)

**Severity:** High

### Issues Fixed in the 2.0.2 Release / CDH 5.2.3

This section lists the most significant issues fixed in Impala 2.0.2.

For the full list of fixed issues in Impala 2.0.2, see [this report in the JIRA system](#).

- **Note:** Impala 2.0.2 is available as part of CDH 5.2.3, not under CDH 4.

#### GROUP BY on STRING column produces inconsistent results

Some operations in queries submitted through Hue or other HiveServer2 clients could produce inconsistent results.

**Bug:** [IMPALA-1453](#)

**Severity:** High

#### Fix leaked file descriptor and excessive file descriptor use

Impala could encounter an error from running out of file descriptors. The fix reduces the amount of time file descriptors are kept open, and avoids leaking file descriptors when read operations encounter errors.

**Severity:** High

## [unix\\_timestamp\(\) does not return correct time](#)

The `unix_timestamp()` function could return a constant value 1 instead of a representation of the time.

**Bug:** [IMPALA-1623](#)

**Severity:** High

## [Impala should randomly select cached replica](#)

To avoid putting too heavy a load on any one node, Impala now randomizes which scan node processes each HDFS data block rather than choosing the first cached block replica.

**Bug:** [IMPALA-1586](#)

**Severity:** High

## [Impala does not always give short name to Llama.](#)

In clusters secured by Kerberos or LDAP, a discrepancy in internal transmission of user names could cause a communication error with Llama.

**Bug:** [IMPALA-1606](#)

**Severity:** High

## [accept unmangled native UDF symbols](#)

The `CREATE FUNCTION` statement could report that it could not find a function entry point within the `.so` file for a UDF written in C++, even if the corresponding function was present.

**Bug:** [IMPALA-1475](#)

**Severity:** High

## [Issues Fixed in the 2.0.1 Release / CDH 5.2.1](#)

This section lists the most significant issues fixed in Impala 2.0.1.

For the full list of fixed issues in Impala 2.0.1, see [this report in the JIRA system](#).

## [Queries fail with metastore exception after upgrade and compute stats](#)

After running the `COMPUTE STATS` statement on an Impala table, subsequent queries on that table could fail with the exception message Failed to load metadata for table: `default.stats_test`.

**Bug:** <https://issues.cloudera.org/browse/IMPALA-1416> IMPALA-1416

**Severity:** High

**Workaround:** Upgrading to CDH 5.2.1, or another level of CDH that includes the fix for HIVE-8627, prevents the problem from affecting future `COMPUTE STATS` statements. On affected levels of CDH, or for Impala tables that have become inaccessible, the workaround is to disable the `hive.metastore.try.direct.sql` setting in the Hive metastore `hive-site.xml` file and issue the `INVALIDATE METADATA` statement for the affected table. You do not need to rerun the `COMPUTE STATS` statement for the table.

## [Issues Fixed in the 2.0.0 Release / CDH 5.2.0](#)

This section lists the most significant issues fixed in Impala 2.0.0.

For the full list of fixed issues in Impala 2.0.0, see [this report in the JIRA system](#).

### Join Hint is dropped when used inside a view

Hints specified within a view query did not take effect when the view was queried, leading to slow performance. As part of this fix, Impala now supports hints embedded within comments.

**Bug:** [IMPALA-995](#)

**Severity:** High

### WHERE condition ignored in simple query with RIGHT JOIN

Potential wrong results for some types of queries.

**Bug:** [IMPALA-1101](#)

**Severity:** High

### Query with self joined table may produce incorrect results

Potential wrong results for some types of queries.

**Bug:** [IMPALA-1102](#)

**Severity:** High

### Incorrect plan after reordering predicates (inner join following outer join)

Potential wrong results for some types of queries.

**Bug:** [IMPALA-1118](#)

**Severity:** High

### Combining fragments with compatible data partitions can lead to incorrect results due to type incompatibilities (missing casts).

Potential wrong results for some types of queries.

**Bug:** [IMPALA-1123](#)

**Severity:** High

### Predicate dropped: Inline view + DISTINCT aggregate in outer query

Potential wrong results for some types of queries.

**Bug:** [IMPALA-1165](#)

**Severity:** High

### Reuse of a column in JOIN predicate may lead to incorrect results

Potential wrong results for some types of queries.

**Bug:** [IMPALA-1353](#)

**Severity:** High

### Usage of TRUNC with string timestamp reliably crashes node

Serious error for certain combinations of function calls and data types.

**Bug:** [IMPALA-1105](#)

**Severity:** High

## Timestamp Cast Returns invalid TIMESTAMP

Serious error for certain combinations of function calls and data types.

**Bug:** [IMPALA-1109](#)

**Severity:** High

## IllegalStateException upon JOIN of DECIMAL columns with different precision

DECIMAL columns with different precision could not be compared in join predicates.

**Bug:** [IMPALA-1121](#)

**Severity:** High

## Allow creating Avro tables without column definitions. Allow COMPUTE STATS to always work on Impala-created Avro tables.

Hive-created Avro tables with columns specified by a JSON file or literal could produce errors when queried in Impala, and could not be used with the COMPUTE STATS statement. Now you can create such tables in Impala to avoid such errors.

**Bug:** [IMPALA-1104](#)

**Severity:** High

## Ensure all webserver output is escaped

The Impala debug web UI did not properly encode all output.

**Bug:** [IMPALA-1133](#)

**Severity:** High

## Queries with union in inline view have empty resource requests

Certain queries could run without obeying the limits imposed by resource management.

**Bug:** [IMPALA-1236](#)

**Severity:** High

## Impala does not employ ACLs when checking path permissions for LOAD and INSERT

Certain INSERT and LOAD DATA statements could fail unnecessarily, if the target directories in HDFS had restrictive HDFS permissions, but those permissions were overridden by HDFS extended ACLs.

**Bug:** [IMPALA-1279](#)

**Severity:** High

## Impala does not map principals to lowercase, affecting Sentry authorisation

In a Kerberos environment, the principal name was not mapped to lowercase, causing issues when a user logged in with an uppercase principal name and Sentry authorization was enabled.

**Bug:** [IMPALA-1334](#)

**Severity:** High

## Issues Fixed in the 1.4.4 Release / CDH 5.1.5

For the list of fixed issues, see [Issues Fixed in CDH 5.1.5](#) in the *CDH 5 Release Notes*.

- **Note:** Impala 1.4.4 is available as part of CDH 5.1.5, not under CDH 4.

## Issues Fixed in the 1.4.3 Release / CDH 5.1.4

Impala 1.4.3 includes fixes to address what is known as the POODLE vulnerability in SSLv3. SSLv3 access is disabled in the Impala debug web UI.

- **Note:** Impala 1.4.3 is available as part of CDH 5.1.4, and under CDH 4.

## Issues Fixed in the 1.4.2 Release / CDH 5.1.3

This section lists the most significant issues fixed in Impala 1.4.2.

For the full list of fixed issues in Impala 1.4.2, see [this report in the JIRA system](#).

- **Note:** Impala 1.4.3 is available as part of CDH 5.1.4, and under CDH 4.

## Issues Fixed in the 1.4.1 Release / CDH 5.1.2

This section lists the most significant issues fixed in Impala 1.4.1.

For the full list of fixed issues in Impala 1.4.1, see [this report in the JIRA system](#).

- **Note:** Impala 1.4.1 is only available as part of CDH 5.1.2, not under CDH 4.

### impalad terminating with Boost exception

Occasionally, a non-trivial query run through Llama could encounter a serious error. The detailed error in the log was:

```
boost::exception_detail::clone_impl
<boost::exception_detail::error_info_injector<boost::lock_error> >
```

**Severity:** High

### Impalad uses wrong string format when writing logs

Impala log files could contain internal error messages due to a problem formatting certain strings. The messages consisted of a Java call stack starting with:

```
[jni-util.cc:177] java.util.MissingFormatArgumentException: Format specifier 's'
```

**Severity:** High

### Update HS2 client API.

A downlevel version of the HiveServer2 API could cause difficulty retrieving the precision and scale of a DECIMAL value.

**Bug:** [IMPALA-1107](#)

**Severity:** High

### Impalad catalog updates can fail with error: "IllegalArgumentException: fromKey out of range" at com.cloudera.impala.catalog.CatalogDeltaLog

The error in the title could occur following a DDL statement. This issue was discovered during internal testing and has not been reported in customer environments.

**Bug:** [IMPALA-1093](#)

**Severity:** High

## "Total" time counter does not capture all the network transmit time

The time for some network operations was not counted in the report of total time for a query, making it difficult to diagnose network-related performance issues.

**Bug:** [IMPALA-1131](#)

**Severity:** High

## Impala will crash when reading certain Avro files containing bytes data

Certain Avro fields for byte data could cause Impala to be unable to read an Avro data file, even if the field was not part of the Impala table definition. With this fix, Impala can now read these Avro data files, although Impala queries cannot refer to the "bytes" fields.

**Bug:** [IMPALA-1149](#)

**Severity:** High

## Support specifying a custom AuthorizationProvider in Impala

The `--authorization_policy_provider_class` option for `impalad` was added back. This option specifies a custom `AuthorizationProvider` class rather than the default `HadoopGroupAuthorizationProvider`. It had been used for internal testing, then removed in Impala 1.4.0, but it was considered useful by some customers.

**Bug:** [IMPALA-1142](#)

**Severity:** High

## Issues Fixed in the 1.4.0 Release / CDH 5.1.0

This section lists the most significant issues fixed in Impala 1.4.0.

For the full list of fixed issues in Impala 1.4.0, see [this report in the JIRA system](#).

### Failed DCHECK in disk-io-mgr-reader-context.cc:174

The serious error in the title could occur, with the supplemental message:

```
-----[ num_used_buffers_ < 0: #used=-1 during cancellation HDFS cached data ]-----
```

The issue was due to the use of HDFS caching with data files accessed by Impala. Support for HDFS caching in Impala was introduced in Impala 1.4.0 for CDH 5.1.0. The fix for this issue was backported to Impala 1.3.x, and is the only change in Impala 1.3.2 for CDH 5.0.4.

**Bug:** [IMPALA-1019](#)

**Severity:** High

**Workaround:** On CDH 5.0.x, upgrade to CDH 5.0.4 with Impala 1.3.2, where this issue is fixed. In Impala 1.3.0 or 1.3.1 on CDH 5.0.x, do not use HDFS caching for Impala data files in Impala internal or external tables. If some of these data files are cached (for example because they are used by other components that take advantage of HDFS caching), set the query option `DISABLE_CACHED_READS=true`. To set that option for all Impala queries across all sessions, start `impalad` with the `-default_query_options` option and include this setting in the option argument, or on a cluster managed by Cloudera Manager, fill in this option setting on the **Impala Daemon** options page.

**Resolution:** This issue is fixed in Impala 1.3.2 for CDH 5.0.4. The addition of HDFS caching support in Impala 1.4 means that this issue does not apply to any new level of Impala on CDH 5.

### impala-shell only works with ASCII characters

The `impala-shell` interpreter could encounter errors processing SQL statements containing non-ASCII characters.

**Bug:** [IMPALA-489](#)

**Severity:** High

### The extended view definition SQL text in Views created by Impala should always have fully-qualified table names

When a view was accessed while inside a different database, references to tables were not resolved unless the names were fully qualified when the view was created.

**Bug:** [IMPALA-962](#)

**Severity:** High

### Impala forgets about partitions with non-existant locations

If an `ALTER TABLE` specified a non-existent HDFS location for a partition, afterwards Impala would not be able to access the partition at all.

**Bug:** [IMPALA-741](#)

**Severity:** High

### `CREATE TABLE LIKE` fails if source is a view

The `CREATE TABLE LIKE` clause was enhanced to be able to create a table with the same column definitions as a view. The resulting table is a text table unless the `STORED AS` clause is specified, because a view does not have an associated file format to inherit.

**Bug:** [IMPALA-834](#)

**Severity:** High

### Improve partition pruning time

Operations on tables with many partitions could be slow due to the time to evaluate which partitions were affected. The partition pruning code was speeded up substantially.

**Bug:** [IMPALA-887](#)

**Severity:** High

### Improve compute stats performance

The performance of the `COMPUTE STATS` statement was improved substantially. The efficiency of its internal operations was improved, and some statistics are no longer gathered because they are not currently used for planning Impala queries.

**Bug:** [IMPALA-1003](#)

**Severity:** High

### When I run `CREATE TABLE new_table LIKE avro_table`, the schema does not get mapped properly from an avro schema to a hive schema

After a `CREATE TABLE LIKE` statement using an Avro table as the source, the new table could have incorrect metadata and be inaccessible, depending on how the original Avro table was created.

**Bug:** [IMPALA-185](#)

**Severity:** High

### Race condition in IoMgr. Blocked ranges enqueued after cancel.

Impala could encounter a serious error after a query was cancelled.

**Bug:** [IMPALA-1046](#)

**Severity:** High

### Deadlock in scan node

A deadlock condition could make all `impalad` daemons hang, making the cluster unresponsive for Impala queries.

**Bug:** [IMPALA-1083](#)

**Severity:** High

## Issues Fixed in the 1.3.3 Release / CDH 5.0.5

Impala 1.3.3 includes fixes to address what is known as the POODLE vulnerability in SSLv3. SSLv3 access is disabled in the Impala debug web UI.

- **Note:** Impala 1.3.3 is only available as part of CDH 5.0.5, not under CDH 4.

## Issues Fixed in the 1.3.2 Release / CDH 5.0.4

This backported bug fix is the only change between Impala 1.3.1 and Impala 1.3.2.

- **Note:** Impala 1.3.3 is only available as part of CDH 5.0.5, not under CDH 4.

### Failed DCHECK in disk-io-mgr-reader-context.cc:174

The serious error in the title could occur, with the supplemental message:

```
num_used_buffers_ < 0: #used=-1 during cancellation HDFS cached data
```

The issue was due to the use of HDFS caching with data files accessed by Impala. Support for HDFS caching in Impala was introduced in Impala 1.4.0 for CDH 5.1.0. The fix for this issue was backported to Impala 1.3.x, and is the only change in Impala 1.3.2 for CDH 5.0.4.

**Bug:** [IMPALA-1019](#)

**Severity:** High

**Workaround:** On CDH 5.0.x, upgrade to CDH 5.0.4 with Impala 1.3.2, where this issue is fixed. In Impala 1.3.0 or 1.3.1 on CDH 5.0.x, do not use HDFS caching for Impala data files in Impala internal or external tables. If some of these data files are cached (for example because they are used by other components that take advantage of HDFS caching), set the query option `DISABLE_CACHED_READS=true`. To set that option for all Impala queries across all sessions, start `impalad` with the `-default_query_options` option and include this setting in the option argument, or on a cluster managed by Cloudera Manager, fill in this option setting on the **Impala Daemon** options page.

**Resolution:** This issue is fixed in Impala 1.3.2 for CDH 5.0.4. The addition of HDFS caching support in Impala 1.4 means that this issue does not apply to any new level of Impala on CDH 5.

## Issues Fixed in the 1.3.1 Release / CDH 5.0.3

This section lists the most significant issues fixed in Impala 1.3.1.

For the full list of fixed issues in Impala 1.3.1, see [this report in the JIRA system](#). Because 1.3.1 is the first 1.3.x release for CDH 4, if you are on CDH 4, also consult [Issues Fixed in the 1.3.0 Release / CDH 5.0.0](#) on page 517.

### Impalad crashes when left joining inline view that has aggregate using distinct

Impala could encounter a severe error in a query combining a left outer join with an inline view containing a `COUNT(DISTINCT)` operation.

**Bug:** [IMPALA-904](#)

**Severity:** High

### Incorrect result with group by query with null value in group by data

If the result of a `GROUP BY` operation is `NULL`, the resulting row might be omitted from the result set. This issue depends on the data values and data types in the table.

**Bug:** [IMPALA-901](#)

**Severity:** High

### Drop Function does not clear local library cache

When a UDF is dropped through the `DROP FUNCTION` statement, and then the UDF is re-created with a new `.so` library or JAR file, the original version of the UDF is still used when the UDF is called from queries.

**Bug:** [IMPALA-786](#)

**Severity:** High

**Workaround:** Restart the `impalad` daemon on all nodes.

### Compute stats doesn't propagate underlying error correctly

If a `COMPUTE STATS` statement encountered an error, the error message is "Query aborted" with no further detail. Common reasons why a `COMPUTE STATS` statement might fail include network errors causing the coordinator node to lose contact with other `impalad` instances, and column names that match Impala [reserved words](#). (Currently, if a column name is an Impala reserved word, `COMPUTE STATS` always returns an error.)

**Bug:** [IMPALA-762](#)

**Severity:** High

### Inserts should respect changes in partition location

After an `ALTER TABLE` statement that changes the `LOCATION` property of a partition, a subsequent `INSERT` statement would always use a path derived from the base data directory for the table.

**Bug:** [IMPALA-624](#)

**Severity:** High

### Text data with carriage returns generates wrong results for count(\*)

A `COUNT(*)` operation could return the wrong result for text tables using nul characters (ASCII value 0) as delimiters.

**Bug:** [IMPALA-13](#)

**Severity:** High

**Workaround:** Impala adds support for ASCII 0 characters as delimiters through the clause `FIELDS TERMINATED BY '\0'`.

### IO Mgr should take instance memory limit into account when creating io buffers

Impala could allocate more memory than necessary during certain operations.

**Bug:** [IMPALA-488](#)

**Severity:** High

**Workaround:** Before issuing a `COMPUTE STATS` statement for a Parquet table, reduce the number of threads used in that operation by issuing `SET NUM_SCANNER_THREADS=2` in `impala-shell`. Then issue `UNSET NUM_SCANNER_THREADS` before continuing with queries.

## [Impala should provide an option for new sub directories to automatically inherit the permissions of the parent directory](#)

When new subdirectories are created underneath a partitioned table by an `INSERT` statement, previously the new subdirectories always used the default HDFS permissions for the `impala` user, which might not be suitable for directories intended to be read and written by other components also.

**Bug:** [IMPALA-827](#)

**Severity:** High

**Resolution:** In Impala 1.3.1 and higher, you can specify the `--insert_inherit_permissions` configuration when starting the `impalad` daemon.

## [Illegal state exception \(or crash\) in query with UNION in inline view](#)

Impala could encounter a severe error in a query where the `FROM` list contains an inline view that includes a `UNION`. The exact type of the error varies.

**Bug:** [IMPALA-888](#)

**Severity:** High

## [INSERT column reordering doesn't work with SELECT clause](#)

The ability to specify a subset of columns in an `INSERT` statement, with order different than in the target table, was not working as intended.

**Bug:** [IMPALA-945](#)

**Severity:** High

## [Issues Fixed in the 1.3.0 Release / CDH 5.0.0](#)

This section lists the most significant issues fixed in Impala 1.3.0, primarily issues that could cause wrong results, or cause problems running the `COMPUTE STATS` statement, which is very important for performance and scalability.

For the full list of fixed issues, see [this report in the JIRA system](#).

## [Inner join after right join may produce wrong results](#)

The automatic join reordering optimization could incorrectly reorder queries with an outer join or semi join followed by an inner join, producing incorrect results.

**Bug:** [IMPALA-860](#)

**Severity:** High

**Workaround:** Including the `STRAIGHT_JOIN` keyword in the query prevented the issue from occurring.

## [Incorrect results with codegen on multi-column group by with NULLs.](#)

A query with a `GROUP BY` clause referencing multiple columns could introduce incorrect `NULL` values in some columns of the result set. The incorrect `NULL` values could appear in rows where a different `GROUP BY` column actually did return `NULL`.

**Bug:** [IMPALA-850](#)

**Severity:** High

## [Using distinct inside aggregate function may cause incorrect result when using having clause](#)

A query could return incorrect results if it combined an aggregate function call, a `DISTINCT` operator, and a `HAVING` clause, without a `GROUP BY` clause.

## Cloudera Impala Release Notes

**Bug:** [IMPALA-845](#)

**Severity:** High

### Aggregation on union inside (inline) view not distributed properly.

An aggregation query or a query with `ORDER BY` and `LIMIT` could be executed on a single node in some cases, rather than distributed across the cluster. This issue affected queries whose `FROM` clause referenced an inline view containing a `UNION`.

**Bug:** [IMPALA-831](#)

**Severity:** High

### Wrong expression may be used in aggregate query if there are multiple similar expressions

If a `GROUP BY` query referenced the same columns multiple times using different operators, result rows could contain multiple copies of the same expression.

**Bug:** [IMPALA-817](#)

**Severity:** High

### Incorrect results when changing the order of aggregates in the select list with codegen enabled

Referencing the same columns in both a `COUNT()` and a `SUM()` call in the same query, or some other combinations of aggregate function calls, could incorrectly return a result of 0 from one of the aggregate functions. This issue affected references to `TINYINT` and `SMALLINT` columns, but not `INT` or `BIGINT` columns.

**Bug:** [IMPALA-765](#)

**Severity:** High

**Workaround:** Setting the query option `DISABLE_CODEGEN=TRUE` prevented the incorrect results. Switching the order of the function calls could also prevent the issue from occurring.

### Union queries give Wrong result in a UNION followed by SIGSEGV in another union

A `UNION` query could produce a wrong result, followed by a serious error for a subsequent `UNION` query.

**Bug:** [IMPALA-723](#)

**Severity:** High

### String data in MR-produced parquet files may be read incorrectly

Impala could return incorrect string results when reading uncompressed Parquet data files containing multiple row groups. This issue only affected Parquet data files produced by MapReduce jobs.

**Bug:** [IMPALA-729](#)

**Severity:** High

### Compute stats need to use quotes with identifiers that are Impala keywords

Using a column or table name that conflicted with Impala keywords could prevent running the `COMPUTE STATS` statement for the table.

**Bug:** [IMPALA-777](#)

**Severity:** High

### COMPUTE STATS child queries do not inherit parent query options.

The `COMPUTE STATS` statement did not use the setting of the `MEM_LIMIT` query option in `impala-shell`, potentially causing problems gathering statistics for wide Parquet tables.

**Bug:** [IMPALA-903](#)

**Severity:** High

### COMPUTE STATS should update partitions in batches

The COMPUTE STATS statement could be slow or encounter a timeout while analyzing a table with many partitions.

**Bug:** [IMPALA-880](#)

**Severity:** High

### Fail early (in analysis) when COMPUTE STATS is run against Avro table with no columns

If the columns for an Avro table were all defined in the TBLPROPERTIES or SERDEPROPERTIES clauses, the COMPUTE STATS statement would fail after completely analyzing the table, potentially causing a long delay. Although the COMPUTE STATS statement still does not work for such tables, now the problem is detected and reported immediately.

**Bug:** [IMPALA-867](#)

**Severity:** High

**Workaround:** Re-create the Avro table with columns defined in SQL style, using the output of SHOW CREATE TABLE. (See the JIRA page for detailed steps.)

## Issues Fixed in the 1.2.4 Release

This section lists the most significant issues fixed in Impala 1.2.4. For the full list of fixed issues, see [this report in the JIRA system](#).

### The Catalog Server exits with an OOM error after a certain number of CREATE statements

A large number of concurrent CREATE TABLE statements can cause the catalogd process to consume excessive memory, and potentially be killed due to an out-of-memory condition.

**Bug:** [IMPALA-818](#)

**Severity:** High

**Workaround:** Restart the catalogd service and re-try the DDL operations that failed.

### Catalog Server consumes excessive cpu cycle

A large number of tables and partitions could result in unnecessary CPU overhead during Impala idle time and background operations.

**Bug:** [IMPALA-821](#)

**Severity:** High

**Resolution:** Catalog server processing was optimized in several ways.

### Query against Avro table crashes Impala with codegen enabled

A query against a TIMESTAMP column in an Avro table could encounter a serious issue.

**Bug:** [IMPALA-828](#)

**Severity:** High

**Workaround:** Set the query option DISABLE\_CODEGEN=TRUE

**Statestore seems to send concurrent heartbeats to the same subscriber leading to repeated "Subscriber '*hostname*' is registering with statestore, ignoring update" messages**

Impala nodes could produce repeated error messages after recovering from a communication error with the statestore service.

**Bug:** [IMPALA-809](#)

**Severity:** High

### Join predicate incorrectly ignored

A join query could produce wrong results if multiple equality comparisons between the same tables referred to the same column.

**Bug:** [IMPALA-805](#)

**Severity:** High

### Query result differing between Impala and Hive

Certain outer join queries could return wrong results. If one of the tables involved in the join was an inline view, some tests from the WHERE clauses could be applied to the wrong phase of the query.

**Severity:** High

### ArrayIndexOutOfBoundsException / Invalid query handle when reading large HBase cell

An HBase cell could contain a value larger than 32 KB, leading to a serious error when Impala queries that table. The error could occur even if the applicable row is not part of the result set.

**Bug:** [IMPALA-715](#)

**Severity:** High

**Workaround:** Use smaller values in the HBase table, or exclude the column containing the large value from the result set.

### select with distinct and full outer join, impalad coredump

A query involving a DISTINCT operator combined with a FULL OUTER JOIN could encounter a serious error.

**Bug:** [IMPALA-735](#)

**Severity:** High

**Workaround:** Set the query option DISABLE\_CODEGEN=TRUE

### Impala cannot load tables with more than Short.MAX\_VALUE number of partitions

If a table had more than 32,767 partitions, Impala would not recognize the partitions above the 32K limit and query results could be incomplete.

**Bug:** [IMPALA-749](#)

**Severity:** High

### Various issues with HBase row key specification

Queries against HBase tables could fail with an error if the row key was compared to a function return value rather than a string constant. Also, queries against HBase tables could fail if the WHERE clause contained combinations of comparisons that could not possibly match any row key.

**Severity:** High

**Resolution:** Queries now return appropriate results when function calls are used in the row key comparison. For queries involving non-existent row keys, such as `WHERE row_key IS NULL` or where the lower bound is greater than the upper bound, the query succeeds and returns an empty result set.

## Issues Fixed in the 1.2.3 Release

This release is a fix release that supercedes Impala 1.2.2, with the same features and fixes as 1.2.2 plus one additional fix for compatibility with Parquet files generated outside of Impala by components such as Hive, Pig, or MapReduce.

### Impala cannot read Parquet files with multiple row groups

The `parquet-mr` library included with CDH4.5 writes files that are not readable by Impala, due to the presence of multiple row groups. Queries involving these data files might result in a crash or a failure with an error such as "Column chunk should not contain two dictionary pages".

This issue does not occur for Parquet files produced by Impala `INSERT` statements, because Impala only produces files with a single row group.

**Bug:** [IMPALA-720](#)

**Severity:** High

## Issues Fixed in the 1.2.2 Release

This section lists the most significant issues fixed in Impala 1.2.2. For the full list of fixed issues, see [this report in the JIRA system](#).

### Order of table references in FROM clause is critical for optimal performance

Impala does not currently optimize the join order of queries; instead, it joins tables in the order in which they are listed in the `FROM` clause. Queries that contain one or more large tables on the right hand side of joins (either an explicit join expressed as a `JOIN` statement or a join implicit in the list of table references in the `FROM` clause) may run slowly or crash Impala due to out-of-memory errors. For example:

```
SELECT ... FROM small_table JOIN large_table
```

**Severity:** Medium

**Anticipated Resolution:** Fixed in Impala 1.2.2.

**Workaround:** In Impala 1.2.2 and higher, use the `COMPUTE STATS` statement to gather statistics for each table involved in the join query, after data is loaded. Prior to Impala 1.2.2, modify the query, if possible, to join the largest table first. For example:

```
SELECT ... FROM small_table JOIN large_table
```

should be modified to:

```
SELECT ... FROM large_table JOIN small_table
```

### Parquet in CDH4.5 writes data files that are sometimes unreadable by Impala

Some Parquet files could be generated by other components that Impala could not read.

**Bug:** [IMPALA-694](#)

**Severity:** High

**Resolution:** The underlying issue is being addressed by a fix in the CDH Parquet libraries. Impala 1.2.2 works around the problem and reads the existing data files.

### Deadlock in statestore when unregistering a subscriber and building a topic update

The statestore service could experience an internal error leading to a hang.

**Bug:** [IMPALA-699](#)

**Severity:** High

### IllegalStateException when doing a union involving a group by

A UNION query where one side involved a GROUP BY operation could cause a serious error.

**Bug:** [IMPALA-687](#)

**Severity:** High

### Impala Parquet Writer hit DCHECK in RleEncoder

A serious error could occur when doing an INSERT into a Parquet table.

**Bug:** [IMPALA-689](#)

**Severity:** High

### Hive UDF jars cannot be loaded by the FE

If the JAR file for a Java-based Hive UDF was not in the CLASSPATH, the UDF could not be called during a query.

**Bug:** [IMPALA-695](#)

**Severity:** High

## Issues Fixed in the 1.2.1 Release

This section lists the most significant issues fixed in Impala 1.2.1. For the full list of fixed issues, see [this report in the JIRA system](#).

### Scanners use too much memory when reading past scan range

While querying a table with long column values, Impala could over-allocate memory leading to an out-of-memory error. This problem was observed most frequently with tables using uncompressed RCFile or text data files.

**Bug:** [IMPALA-525](#)

**Severity:** High

**Resolution:** Fixed in 1.2.1

### Join node consumes memory way beyond mem-limit

A join query could allocate a temporary work area that was larger than needed, leading to an out-of-memory error. The fix makes Impala return unused memory to the system when the memory limit is reached, avoiding unnecessary memory errors.

**Bug:** [IMPALA-657](#)

**Severity:** High

**Resolution:** Fixed in 1.2.1

### Excessive memory consumption when query tables with 1k columns (Parquet file)

Impala could encounter an out-of-memory condition setting up work areas for Parquet tables with many columns. The fix reduces the size of the allocated memory when not actually needed to hold table data.

**Bug:** [IMPALA-652](#)

**Severity:** High

**Resolution:** Fixed in 1.2.1

## Issues Fixed in the 1.2.0 Beta Release

This section lists the most significant issues fixed in Impala 1.2 (beta). For the full list of fixed issues, see [this report in the JIRA system](#).

## Issues Fixed in the 1.1.1 Release

This section lists the most significant issues fixed in Impala 1.1.1. For the full list of fixed issues, see [this report in the JIRA system](#).

### Unexpected LLVM Crash When Querying Doubles on CentOS 5.x

Certain queries involving `DOUBLE` columns could fail with a serious error. The fix improves the generation of native machine instructions for certain chipsets.

**Bug:** [IMPALA-477](#)

**Severity:** High

### "block size is too big" error with Snappy-compressed RCFile containing null

Queries could fail with a "block size is too big" error, due to `NULL` values in RCFile tables using Snappy compression.

**Bug:** [IMPALA-482](#)

**Severity:** High

### Cannot query RC file for table that has more columns than the data file

Queries could fail if an Impala RCFile table was defined with more columns than in the corresponding RCFile data files.

**Bug:** [IMPALA-510](#)

**Severity:** High

### Views Sometimes Not Utilizing Partition Pruning

Certain combinations of clauses in a view definition for a partitioned table could result in inefficient performance and incorrect results.

**Bug:** [IMPALA-495](#)

**Severity:** High

### Update the serde name we write into the metastore for Parquet tables

The SerDes class string written into Parquet data files created by Impala was updated for compatibility with Parquet support in Hive. See [Incompatible Changes Introduced in Impala 1.1.1](#) on page 495 for the steps to update older Parquet data files for Hive compatibility.

**Bug:** [IMPALA-485](#)

**Severity:** High

### Selective queries over large tables produce unnecessary memory consumption

A query returning a small result sets from a large table could tie up memory unnecessarily for the duration of the query.

**Bug:** [IMPALA-534](#)

**Severity:** High

### [Impala stopped to query AVRO tables](#)

Queries against Avro tables could fail depending on whether the Avro schema URL was specified in the TBLPROPERTIES or SERDEPROPERTIES field. The fix causes Impala to check both fields for the schema URL.

**Bug:** [IMPALA-538](#)

**Severity:** High

### [Impala continues to allocate more memory even though it has exceed its mem-limit](#)

Queries could allocate substantially more memory than specified in the `impalad -mem_limit` startup option. The fix causes more frequent checking of the limit during query execution.

**Bug:** [IMPALA-520](#)

**Severity:** High

## [Issues Fixed in the 1.1.0 Release](#)

This section lists the most significant issues fixed in Impala 1.1. For the full list of fixed issues, see [this report in the JIRA system](#).

### [10-20% perf regression for most queries across all table formats](#)

This issue is due to a performance tradeoff between systems running many queries concurrently, and systems running a single query. Systems running only a single query could experience lower performance than in early beta releases. Systems running many queries simultaneously should experience higher performance than in the beta releases.

**Severity:** High

### [planner fails with "Join requires at least one equality predicate between the two tables" when "from" table order does not match "where" join order](#)

A query could fail if it involved 3 or more tables and the last join table was specified as a subquery.

**Bug:** [IMPALA-85](#)

**Severity:** High

### [Parquet writer uses excessive memory with partitions](#)

`INSERT` statements against partitioned tables using the Parquet format could use excessive amounts of memory as the number of partitions grew large.

**Bug:** [IMPALA-257](#)

**Severity:** High

### [Comments in impala-shell in interactive mode are not handled properly causing syntax errors or wrong results](#)

The `impala-shell` interpreter did not accept comment entered at the command line, making it problematic to copy and paste from scripts or other code examples.

**Bug:** [IMPALA-192](#)

**Severity:** Low

### [Cancelled queries sometimes aren't removed from the inflight query list](#)

The Impala web UI would sometimes display a query as if it were still running, after the query was cancelled.

**Bug:** [IMPALA-364](#)

**Severity:** High

### Impala's 1.0.1 Shell Broke Python 2.4 Compatibility (AttributeError: 'module' object has no attribute 'field\_size\_limit')

The `impala-shell` command in Impala 1.0.1 does not work with Python 2.4, which is the default on Red Hat 5.

For the `impala-shell` command in Impala 1.0, the `-o` option (pipe output to a file) does not work with Python 2.4.

**Bug:** [IMPALA-396](#)

**Severity:** High

## Issues Fixed in the 1.0.1 Release

This section lists the most significant issues fixed in Impala 1.0.1. For the full list of fixed issues, see [this report in the JIRA system](#).

### Impala parquet scanner can not read all data files generated by other frameworks

Impala might issue an erroneous error message when processing a Parquet data file produced by a non-Impala Hadoop component.

**Bug:** [IMPALA-333](#)

**Severity:** High

**Resolution:** Fixed

### Impala is unable to query RCFile tables which describe fewer columns than the file's header.

If an RCFile table definition had fewer columns than the fields actually in the data files, queries would fail.

**Bug:** [IMPALA-293](#)

**Severity:** High

**Resolution:** Fixed

### Impala does not correctly substitute \_HOST with hostname in --principal

The `_HOST` placeholder in the `--principal` startup option was not substituted with the correct hostname, potentially leading to a startup error in setups using Kerberos authentication.

**Bug:** [IMPALA-351](#)

**Severity:** High

**Resolution:** Fixed

### HBase query missed the last region

A query for an HBase table could omit data from the last region.

**Bug:** [IMPALA-356](#)

**Severity:** High

**Resolution:** Fixed

### Hbase region changes are not handled correctly

After a region in an HBase table was split or moved, an Impala query might return incomplete or out-of-date results.

## Cloudera Impala Release Notes

**Bug:** [IMPALA-300](#)

**Severity:** High

**Resolution:** Fixed

### Query state for successful create table is EXCEPTION

After a successful CREATE TABLE statement, the corresponding query state would be incorrectly reported as EXCEPTION.

**Bug:** [IMPALA-349](#)

**Severity:** High

**Resolution:** Fixed

### Double check release of JNI-allocated byte-strings

Operations involving calls to the Java JNI subsystem (for example, queries on HBase tables) could allocate memory but not release it.

**Bug:** [IMPALA-358](#)

**Severity:** High

**Resolution:** Fixed

### Impala returns 0 for bad time values in UNIX\_TIMESTAMP, Hive returns NULL

Impala returns 0 for bad time values in UNIX\_TIMESTAMP, Hive returns NULL.

**Impala:**

```
impala> select UNIX_TIMESTAMP('10:02:01') ;
impala> 0
```

**Hive:**

```
hive> select UNIX_TIMESTAMP('10:02:01') FROM tmp;
hive> NULL
```

**Bug:** [IMPALA-16](#)

**Severity:** Low

**Anticipated Resolution:** Fixed

### INSERT INTO TABLE SELECT <constant> does not work.

Insert INTO TABLE SELECT <constant> will not insert any data and may return an error.

**Severity:** Low

**Anticipated Resolution:** Fixed

## Issues Fixed in the 1.0 GA Release

Here are the major user-visible issues fixed in Impala 1.0. For a full list of fixed issues, see [this report in the public issue tracker](#).

### Undeterministically receive "ERROR: unknown row batch destination..." and "ERROR: Invalid query handle" from impala shell when running union query

A query containing both UNION and LIMIT clauses could intermittently cause the impalad process to halt with a segmentation fault.

**Bug:** [IMPALA-183](#)

**Severity:** High

**Resolution:** Fixed

### Insert with NULL partition keys results in SIGSEGV.

An `INSERT` statement specifying a `NULL` value for one of the partitioning columns could cause the `impalad` process to halt with a segmentation fault.

**Bug:** [IMPALA-190](#)

**Severity:** High

**Resolution:** Fixed

### INSERT queries don't show completed profiles on the debug webpage

In the Impala web user interface, the profile page for an `INSERT` statement showed obsolete information for the statement once it was complete.

**Bug:** [IMPALA-217](#)

**Severity:** High

**Resolution:** Fixed

### Impala HBase scan is very slow

Queries involving an HBase table could be slower than expected, due to excessive memory usage on the Impala nodes.

**Bug:** [IMPALA-231](#)

**Severity:** High

**Resolution:** Fixed

### Add some library version validation logic to impalad when loading impala-lzo shared library

No validation was done to check that the `impala-lzo` shared library was compatible with the version of Impala, possibly leading to a crash when using LZO-compressed text files.

**Bug:** [IMPALA-234](#)

**Severity:** High

**Resolution:** Fixed

**Workaround:** Always upgrade the `impala-lzo` library at the same time as you upgrade Impala itself.

### Problems inserting into tables with TIMESTAMP partition columns leading table metadata loading failures and failed dchecks

`INSERT` statements for tables partitioned on columns involving datetime types could appear to succeed, but cause errors for subsequent queries on those tables. The problem was especially serious if an improperly formatted timestamp value was specified for the partition key.

**Bug:** [IMPALA-238](#)

**Severity:** Critical

**Resolution:** Fixed

### Ctrl-C sometimes interrupts shell in system call, rather than cancelling query

Pressing Ctrl-C in the `impala-shell` interpreter could sometimes display an error and return control to the shell, making it impossible to cancel the query.

**Bug:** [IMPALA-243](#)

**Severity:** Critical

**Resolution:** Fixed

### Empty string partition value causes metastore update failure

Specifying an empty string or `NULL` for a partition key in an `INSERT` statement would fail.

**Bug:** [IMPALA-252](#)

**Severity:** High

**Resolution:** Fixed. The behavior for empty partition keys was made more compatible with the corresponding Hive behavior.

### Round() does not output the right precision

The `round()` function did not always return the correct number of significant digits.

**Bug:** [IMPALA-266](#)

**Severity:** High

**Resolution:** Fixed

### Cannot cast string literal to string

Casting from a string literal back to the same type would cause an "invalid type cast" error rather than leaving the original value unchanged.

**Bug:** [IMPALA-267](#)

**Severity:** High

**Resolution:** Fixed

### Excessive mem usage for certain queries which are very selective

Some queries that returned very few rows experienced unnecessary memory usage.

**Bug:** [IMPALA-288](#)

**Severity:** High

**Resolution:** Fixed

### HdfsScanNode crashes in UpdateCounters

A serious error could occur for relatively small and inexpensive queries.

**Bug:** [IMPALA-289](#)

**Severity:** High

**Resolution:** Fixed

### Parquet performance issues on large dataset

Certain aggregation queries against Parquet tables were inefficient due to lower than required thread utilization.

**Bug:** [IMPALA-292](#)

**Severity:** High

**Resolution:** Fixed

### impala not populating hive metadata correctly for create table

The Impala `CREATE TABLE` command did not fill in the `owner` and `tbl_type` columns in the Hive metastore database.

**Bug:** [IMPALA-295](#)

**Severity:** High

**Resolution:** Fixed. The metadata was made more Hive-compatible.

### impala daemons die if statestore goes down

The `impalad` instances in a cluster could halt when the `statestored` process became unavailable.

**Bug:** [IMPALA-312](#)

**Severity:** High

**Resolution:** Fixed

### Constant SELECT clauses do not work in subqueries

A subquery would fail if the `SELECT` statement inside it returned a constant value rather than querying a table.

**Bug:** [IMPALA-67](#)

**Severity:** High

**Resolution:** Fixed

### Right outer Join includes NULLs as well and hence wrong result count

The result set from a right outer join query could include erroneous rows containing `NULL` values.

**Bug:** [IMPALA-90](#)

**Severity:** High

**Resolution:** Fixed

### Parquet scanner hangs for some queries

The Parquet scanner non-deterministically hangs when executing some queries.

**Bug:** [IMPALA-204](#)

**Severity:** Medium

**Resolution:** Fixed

## Issues Fixed in Version 0.7 of the Beta Release

### Impala does not gracefully handle unsupported Hive table types (INDEX and VIEW tables)

When attempting to load metadata from an unsupported Hive table type (INDEX and VIEW tables), Impala fails with an unclear error message.

**Bug:** [IMPALA-167](#)

**Severity:** Low

**Resolution:** Fixed in 0.7

### DDL statements (CREATE/ALTER/DROP TABLE) are not supported in the Impala Beta Release

**Severity:** Medium

**Resolution:** Fixed in 0.7

### Avro is not supported in the Impala Beta Release

**Severity:** Medium

**Resolution:** Fixed in 0.7

**Workaround:** None

### Impala does not currently allow limiting the memory consumption of a single query

It is currently not possible to limit the memory consumption of a single query. All tables on the right hand side of JOIN statements need to be able to fit in memory. If they do not, Impala may crash due to out of memory errors.

**Severity:** High

**Resolution:** Fixed in 0.7

### Aggregate of a subquery result set returns wrong results if the subquery contains a 'limit' and data is distributed across multiple nodes

Aggregate of a subquery result set returns wrong results if the subquery contains a 'limit' clause and data is distributed across multiple nodes. From the query plan, it looks like we are just summing the results from each slave.

**Bug:** [IMPALA-20](#)

**Severity:** Low

**Resolution:** Fixed in 0.7

### Partition pruning for arbitrary predicates that are fully bound by a particular partition column

We currently can't utilize a predicate like "country\_code in ('DE', 'FR', 'US')" to do partitioning pruning, because that requires an equality predicate or a binary comparison.

We should create a superclass of planner.ValueRange, ValueSet, that can be constructed with an arbitrary predicate, and whose isInRange(analyzer, valueExpr) constructs a literal predicate by substitution of the valueExpr into the predicate.

**Bug:** [IMPALA-144](#)

**Severity:** Medium

**Resolution:** Fixed in 0.7

## Issues Fixed in Version 0.6 of the Beta Release

### Impala reads the NameNode address and port as command line parameters

Impala reads the NameNode address and port as command line parameters rather than reading them from core-site.xml. Updating the NameNode address in the core-site.xml file does not propagate to Impala.

**Severity:** Low

**Resolution:** Fixed in 0.6 - Impala reads the namenode location and port from the Hadoop configuration files, though setting -nn and -nn\_port overrides this. Users are advised not to set -nn or -nn\_port.

## Queries may fail on secure environment due to impalad Kerberos ticket expiration

Queries may fail on secure environment due to `impalad` Kerberos tickets expiring. This can happen if the `Impala-kerberos_reinit_interval` flag is set to a value ten minutes or less. This may lead to an `impalad` requesting a ticket with a lifetime that is less than the time to the next ticket renewal.

**Bug:** [IMPALA-64](#)

**Severity:** Medium

**Resolution:** Fixed in 0.6

## Concurrent queries may fail when Impala uses Thrift to communicate with the Hive Metastore

Concurrent queries may fail when Impala is using Thrift to communicate with part of the Hive Metastore such as the Hive Metastore Service. In such a case, the error `get_fields failed: out of sequence response`" may occur because Impala shared a single Hive Metastore Client connection across threads. With Impala 0.6, a separate connection is used for each metadata request.

**Bug:** [IMPALA-48](#)

**Severity:** Low

**Resolution:** Fixed in 0.6

## impalad fails to start if unable to connect to the Hive Metastore

Impala fails to start if it is unable to establish a connection with the Hive Metastore. This behavior was fixed, allowing Impala to start, even when no Metastore is available.

**Bug:** [IMPALA-58](#)

**Severity:** Low

**Resolution:** Fixed in 0.6

## Impala treats database names as case-sensitive in some contexts

In some queries (including "USE database" statements), database names are treated as case-sensitive. This may lead queries to fail with an `IllegalStateException`.

**Bug:** [IMPALA-44](#)

**Severity:** Medium

**Resolution:** Fixed in 0.6

## Impala does not ignore hidden HDFS files

Impala does not ignore hidden HDFS files, meaning those files prefixed with a period '.' or underscore '\_'. This diverges from Hive/MapReduce, which skips these files.

**Bug:** [IMPALA-18](#)

**Severity:** Low

**Resolution:** Fixed in 0.6

## Issues Fixed in Version 0.5 of the Beta Release

### Impala may have reduced performance on tables that contain a large number of partitions

Impala may have reduced performance on tables that contain a large number of partitions. This is due to extra overhead reading/parsing the partition metadata.

**Severity:** High

## Cloudera Impala Release Notes

**Resolution:** Fixed in 0.5

### Backend client connections not getting cached causes an observable latency in secure clusters

Backend impalads do not cache connections to the coordinator. On a secure cluster, this introduces a latency proportional to the number of backend clients involved in query execution, as the cost of establishing a secure connection is much higher than in the non-secure case.

**Bug:** [IMPALA-38](#)

**Severity:** Medium

**Resolution:** Fixed in 0.5

### Concurrent queries may fail with error: "Table object has not been initialised : `PARTITIONS`"

Concurrent queries may fail with error: "Table object has not been initialised : `PARTITIONS`". This was due to a lack of locking in the Impala table/database metadata cache.

**Bug:** [IMPALA-30](#)

**Severity:** Medium

**Resolution:** Fixed in 0.5

### UNIX\_TIMESTAMP format behaviour deviates from Hive when format matches a prefix of the time value

The Impala UNIX\_TIMESTAMP(val, format) operation compares the length of format and val and returns NULL if they do not match. Hive instead effectively truncates val to the length of the format parameter.

**Bug:** [IMPALA-15](#)

**Severity:** Medium

**Resolution:** Fixed in 0.5

## Issues Fixed in Version 0.4 of the Beta Release

### Impala fails to refresh the Hive metastore if a Hive temporary configuration file is removed

Impala is impacted by Hive bug [HIVE-3596](#) which may cause metastore refreshes to fail if a Hive temporary configuration file is deleted (normally located at `/tmp/hive-<user>-<tmp_number>.xml`). Additionally, the `impala-shell` will incorrectly report that the failed metadata refresh completed successfully.

**Severity:** Medium

**Anticipated Resolution:** To be fixed in a future release

**Workaround:** Restart the `impalad` service. Use the `impalad` log to check for metadata refresh errors.

### Ipad/rpad builtin functions is not correct.

The Ipad/rpad builtin functions generate the wrong results.

**Severity:** Mild

**Resolution:** Fixed in 0.4

### Files with .gz extension reported as 'not supported'

Compressed files with extensions incorrectly generate an exception.

**Bug:** [IMPALA-14](#)

**Severity:** High

**Resolution:** Fixed in 0.4

### [Queries with large limits would hang.](#)

Some queries with large limits were hanging.

**Severity:** High

**Resolution:** Fixed in 0.4

### [Order by on a string column produces incorrect results if there are empty strings](#)

**Severity:** Low

**Resolution:** Fixed in 0.4

## [Issues Fixed in Version 0.3 of the Beta Release](#)

### [All table loading errors show as unknown table](#)

If Impala is unable to load the metadata for a table for any reason, a subsequent query referring to that table will return an `unknown table` error message, even if the table is known.

**Severity:** Mild

**Resolution:** Fixed in 0.3

### [A table that cannot be loaded will disappear from SHOW TABLES](#)

After failing to load metadata for a table, Impala removes that table from the list of known tables returned in `SHOW TABLES`. Subsequent attempts to query the table returns 'unknown table', even if the metadata for that table is fixed.

**Severity:** Mild

**Resolution:** Fixed in 0.3

### [Impala cannot read from HBase tables that are not created as external tables in the hive metastore.](#)

Attempting to select from these tables fails.

**Severity:** Medium

**Resolution:** Fixed in 0.3

### [Certain queries that contain OUTER JOINs may return incorrect results](#)

Queries that contain OUTER JOINs may not return the correct results if there are predicates referencing any of the joined tables in the WHERE clause.

**Severity:** Medium

**Resolution:** Fixed in 0.3.

### Issues Fixed in Version 0.2 of the Beta Release

#### [Subqueries which contain aggregates cannot be joined with other tables or Impala may crash](#)

Subqueries that contain an aggregate cannot be joined with another table or Impala may crash. For example:

```
SELECT * FROM (SELECT sum(coll) FROM some_table GROUP BY coll) t1 JOIN other_table ON  
(...);
```

**Severity:** Medium

**Resolution:** Fixed in 0.2

#### [An insert with a limit that runs as more than one query fragment inserts more rows than the limit.](#)

For example:

```
INSERT OVERWRITE TABLE test SELECT * FROM test2 LIMIT 1;
```

**Severity:** Medium

**Resolution:** Fixed in 0.2

#### [Query with limit clause might fail.](#)

For example:

```
SELECT * FROM test2 LIMIT 1;
```

**Severity:** Medium

**Resolution:** Fixed in 0.2

#### [Files in unsupported compression formats are read as plain text.](#)

Attempting to read such files does not generate a diagnostic.

**Severity:** Medium

**Resolution:** Fixed in 0.2

#### [Impala server raises a null pointer exception when running an HBase query.](#)

When querying an HBase table whose row-key is string type, the Impala server may raise a null pointer exception.

**Severity:** Medium

**Resolution:** Fixed in 0.2