

# Cloudera Data Management



### **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 US: 1-888-789-1488 Intl: 1-650-362-0488 www.cloudera.com

#### Release Information

Version: Navigator 2.2.x Date: June 24, 2015

# **Table of Contents**

About Cloudera Data Management	4
Cloudera Navigator Auditing Component	5
Audit Log Properties	
Service Auditing Properties	
Auditing Impala Operations	
Audit Events and Audit Reports	
Viewing Audit Events	
Filtering Audit Events	
Creating Audit Reports	
Editing Audit Reports	
Downloading Audit Events	
Downloading HDFS Directory Access Permission Reports	
Downloading Tibro Directory Access Permission Reports	1 /
Cloudera Navigator Metadata Component	19
Cloudera Navigator Metadata Component	<b>19</b>
Cloudera Navigator Metadata Component  Metadata Search	1 <u>9</u>
Cloudera Navigator Metadata Component  Metadata  Metadata Search  Accessing Metadata	1 <u>9</u> 20
Cloudera Navigator Metadata Component  Metadata	19 20 22
Cloudera Navigator Metadata Component.  Metadata	19 20 24 22
Cloudera Navigator Metadata Component  Metadata	19 20 26 26
Cloudera Navigator Metadata Component.  Metadata	19 20 26 26
Cloudera Navigator Metadata Component  Metadata	1920243032
Cloudera Navigator Metadata Component.  Metadata	192020
Cloudera Navigator Metadata Component.  Metadata	
Cloudera Navigator Metadata Component.  Metadata.  Metadata Search.  Accessing Metadata.  Modifying Business Metadata.  Policies.  Policy Expressions.  Lineage Diagrams.  Displaying a Template Lineage Diagram.  Displaying an Instance Lineage Diagram.	
Cloudera Navigator Metadata Component.  Metadata	
Cloudera Navigator Metadata Component.  Metadata  Metadata Search.  Accessing Metadata.  Modifying Business Metadata.  Policies.  Policy Expressions.  Lineage Diagrams.  Displaying a Template Lineage Diagram.  Displaying an Instance Lineage Diagram.  Displaying the Template Lineage Diagram for an Instance Lineage Diagram.  Downloading a Lineage File.	

# About Cloudera Data Management

This guide describes how to perform data management using Cloudera Navigator. Data management activities include auditing access to data residing in HDFS and Hive metastores, reviewing and updating metadata, and discovering the lineage of data objects.

Important: This feature is available only with a Cloudera Enterprise license.

For other licenses, the following applies:

- Cloudera Express- The feature is not available.
- Cloudera Enterprise Data Hub Edition Trial The feature is available until you end the trial or the trial license expires.

To obtain a license for Cloudera Enterprise, fill in this form or call 866-843-7207. After you install a Cloudera Enterprise license, the feature will be available.

Cloudera Navigator is a fully integrated data management tool for the Hadoop platform. Data management capabilities are critical for enterprise customers that are in highly regulated industries and have stringent compliance requirements.

Cloudera Navigator provides two categories of functionality:

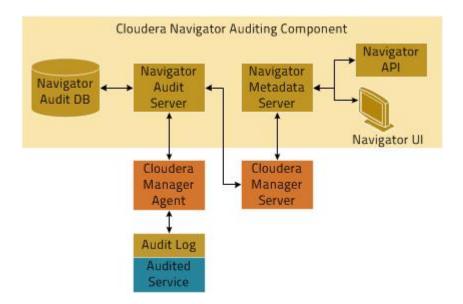
 Auditing data access and verifying access privileges - The goal of auditing is to capture a complete and immutable record of all activity within a system. While Hadoop has historically lacked centralized cross-component audit capabilities, products such as Cloudera Navigator add secured, real-time audit components to key data and access frameworks. Cloudera Navigator allows administrators to configure, collect, and view audit events, to understand who accessed what data and how. Cloudera Navigator also allows administrators to generate reports that list the HDFS access permissions granted to groups.

Cloudera Navigator tracks access permissions and actual accesses to all entities in HDFS, Hive, HBase, Impala, and Sentry to help answer questions such as - who has access to which entities, which entities were accessed by a user, when was an entity accessed and by whom, what entities were accessed using a service, which device was used to access, and so on. Cloudera Navigator auditing supports tracking access to:

- HDFS data accessed through HDFS, Hive, HBase, Cloudera Impala services
- HBase and Impala operations
- Hive metadata
- Sentry access
- Searching metadata and visualizing lineage Cloudera Navigator metadata management features allow DBAs, data modelers, business analysts, and data scientists to search for, amend the properties of, and tag data entities.

In addition, to satisfy risk and compliance audits and data retention policies, it supports the ability to answer questions such as: where did the data come from, where is it used, and what are the consequences of purging or modifying a set of data entities. Cloudera Navigator supports tracking the lineage of HDFS files and directories, Hive tables and columns, MapReduce and YARN jobs, Hive queries, Pig scripts, Sqoop jobs, and Oozie workflows.

The Cloudera Navigator auditing component provides data auditing and access features. The architecture of the Cloudera Navigator auditing component is illustrated below.



When the Cloudera Navigator auditing component is configured, plug-ins that enable collection of audit events are added to the HDFS, HBase, and Hive (that is, the HiveServer2 and Beeswax servers) services. The plug-ins write the audit events to an audit log on the local filesystem. Cloudera Impala and Sentry record audit events directly in an audit log file.

The Cloudera Manager Agent monitors the audit log files and sends these events to the Navigator Audit Server. The Cloudera Manager Agent retries any event that it fails to transmit. As there is no in-memory transient buffer involved, once the audit events are written to the audit log file, they are guaranteed to be delivered (as long as filesystem is available). The Cloudera Manager Agent keeps track of current audit event offset in the audit log that it has successfully transmitted, so on any crash/restart it picks up the audit event from the last successfully sent position and resumes. Audit logs are rotated and the Cloudera Manager Agent follows the rotation of the log. The Agent also takes care of purging old audit logs once they have been successfully transmitted to the Navigator Audit Server. If a plug-in fails to write audit event to audit log file, it can either drop the event or shut down the process in which they are running (depending on the configured queue policy).

The Navigator Audit DB stores audit events.

# **Audit Log Properties**

Describes auditing log properties and how to configure the log properties.

The following properties apply to the audit log file:

 Audit Log Directory - The directory in which audit event log files are written. By default, this property is not set if Cloudera Navigator is not installed.

- **Note:** If the value of this property is changed, and service is restarted, then the Cloudera Manager Agent will start monitoring the new log directory for audit events. In this case it is possible that not all events are published from the old audit log directory. To avoid loss of audit events, when this property is changed, perform the following steps:
  - **1.** Stop the service.
  - 2. Copy audit log files and (for Impala only) the impalad\_audit\_wal file from the old audit log directory to the new audit log directory. This need to be done on all the nodes where Impala daemons are running.
  - 3. Start the service.
- Maximum Audit Log File Size The maximum size of the audit event log file before a new file is created. The unit of the file size is service dependent:
  - HDFS, HBase, Hive MiB
  - Impala lines (queries)
- Number of Audit Logs to Retain Maximum number of rolled over audit logs to retain. The logs will not be deleted if they contain audit events that have not yet been propagated to Audit Server.

#### **Configuring Audit Logs**

- 1. Click a supported service.
- 2. Click the Configuration tab.
- 3. Configure the log properties in the following categories:
  - Impala Impala Daemon Default Group > Logs
  - HBase, HDFS, Hive, Sentry Service-Wide > Logs
- **4.** Edit the audit log properties.
- Click Save Changes.
- 6. Restart the service.

# **Service Auditing Properties**

Describes service auditing properties and how to configure the properties.

Each service (with exceptions noted) that supports auditing configuration has the following properties:

- Enable Collection Controls whether the Cloudera Manager Agent tracks a service's <u>audit log file</u>. A validation check is performed for all life cycle actions (stop/start/restart). If the Enable Collection flag is selected and the Audit Log Directory property is not set, the validator displays a message that says that the Audit Log Directory property must be set to enable auditing.
- Event Filter A set of rules that capture properties of auditable events and actions to be performed when an event matches those properties. This property is not supported for Sentry.
- Event Tracker A set of rules for tracking and coalescing events. This feature is used to define equivalency between different audit events. When events match, according to a set of configurable parameters, only one entry in the audit list is generated for all the matching events.
- Queue Policy The action to take when the audit event queue is full. The options are Drop or Shutdown. When a queue is full and the queue policy of the service is Shutdown, before shutting down the service, N audits will be discarded, where N is the size of the Cloudera Navigator Audit Server queue.

Note: If the queue policy is Shutdown, the Impala service is shut down only if Impala is unable to write to the audit log file. It is possible that an event may not appear in the audit event log due to an error in transfer to the Cloudera Manager Agent or database. In such cases Impala will not shut down and will keep writing to the log file. When the transfer problem is fixed the events will be transferred to the database.

The Event Filter and Event Tracker rules for filtering and coalescing events are expressed as JSON objects. For information on the structure of the objects, see the description on the configuration page within the Cloudera Manager Admin Console.

The default event filter discards events generated by the internal Cloudera and Hadoop users (cloudera-scm, hdfs, hbase, hive, mapred, solr, and dr. who) and that affect files in the /tmp directory.

#### **Configuring Service Auditing Properties**

# Required Role: Navigator Administrator Full Administrator

- 1. Click a service that supports auditing.
- 2. Click the Configuration tab.
- 3. Click the Cloudera Navigator category. The Service-Wide category displays.
- **4.** Edit the properties.
- 5. Click Save Changes.
- 6. Restart the service.

### Configuring Impala Daemon Logging

# Required Role: Configurator Cluster Administrator Full Administrator

To control whether the Impala daemon logs to the audit log:

- 1. Click the Impala service.
- 2. Click the Configuration tab.
- 3. Expand the Impala Daemon Default Group > Logs category.
- 4. Edit the Enable Impala Audit Event Generation checkbox setting.
- Click Save Changes.
- 6. Restart the service.

### Audit Logging to Syslog

# Required Role: Navigator Administrator Full Administrator

The Audit Server logs all audit records into a Log4j logger called auditstream. The log messages are logged at the TRACE level, with the attributes of the audit records. By default, the auditStream logger is inactive because the logger level is set to FATAL. It is also connected to a NullAppender, and does not forward to other appenders (additivity set to false).

To record the audit stream, configure the auditStream logger with the desired appender. For example, the standard SyslogAppender allows you to send the audit records to a remote syslog.

The Log4j SyslogAppender supports only UDP. An example syslog configuration would be:

```
$ModLoad imudp
$UDPServerRun 514
# Accept everything (even DEBUG messages) local2.* /my/audit/trail.log
```

It is also possible to attach other appenders to the auditStream to provide other integration behaviors.

You can audit events to syslog in two formats: JSON and RSA EnVision. To configure audit logging to syslog, do the following:

- 1. Do one of the following:
  - Select Clusters > Cloudera Management Service > Cloudera Management Service.
  - On the Status tab of the Home page, in Cloudera Management Service table, click the Cloudera Management Service link.
- 2. Click the Configuration tab.
- 3. Search for Navigator Audit Server Logging Advanced Configuration Snippet.
- **4.** Click the Value field and depending on the format type, enter:

```
log4j.logger.auditStream = TRACE,SYSLOG
log4j.appender.SYSLOG = org.apache.log4j.net.SyslogAppender
log4j.appender.SYSLOG.SyslogHost = hostname
log4j.appender.SYSLOG.Facility = Local2
log4j.appender.SYSLOG.FacilityPrinting = true
```

To configure the specific stream type, enter:

Format	Properties
JSON	log4j.additivity.auditStream = false
RSA EnVision	log4j.additivity.auditStreamEnVision = false

**5.** Click **Save Changes** to commit the changes.

#### Example Log Messages

Format	Log Message Example
JSON	Jul 23 11:05:15 hostname local2: {"type":"HDFS", "allowed":"true", "time":"1374602714758", "service":"HDFS-1", "user":"root", "ip":"10.20.93.93", "op":"mkdirs", "src":"/audit/root", "perms":"rwxr-xr-x"}
RSA EnVision	Cloudera Navigator 1 type="Hive",allowed="false",time="1382551146763", service="HIVE-1",user="systest",impersonator="",ip="/10.20.190.185",op="QUERY", opText="select count(*) from sample_07",db="default",table="sample_07",path="/user/hive/warehouse/sample_07",db]Type="TAHE"

If a particular field is not applicable for that audit event, it is omitted from the message.

# **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:

- On a system managed by Cloudera Manager, enable auditing by configuring Impala Daemon logging. On a system not managed by Cloudera Manager, include the option -audit\_event\_log\_dir=directory\_path in your impalad startup options. The log directory must be a local directory on the server, not an HDFS
- 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, configure Impala Daemon logging in Cloudera Manager, or include the option -max\_audit\_event\_log\_file\_size=number\_of\_queries in the impalad startup options for systems not managed by Cloudera Manager.
- Configure the Cloudera Navigator product to collect and consolidate the audit logs from all the nodes 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 Impala Daemon role 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 node serves as the coordinator for the query, that is, the node you connect to when you issue the guery. This might be the same node for all gueries, or different applications or users might connect to and issue queries through different nodes.

To avoid excessive I/O overhead on busy coordinator nodes, 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 the Impala Daemon role on the host where the auditing problem occurred.

# 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 nodes 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.

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
  - SOL 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. For CDH 5, see Audit Events and Audit Reports. For CDH 4, see the Cloudera Navigator documentation.

# **Audit Events and Audit Reports**

Required Role: Auditing Viewer Full Administrator

An *audit event* is an event that describes an action of accessing a service. An *audit report*, is a collection of audit events that satisfy a set of filters.

Audit events are recorded by the Cloudera Navigator Audit Server. Audit report metadata is recorded by the Cloudera Navigator Metadata Server.

The following properties can appear in an audit event entry:

- Timestamp Date and time the action was performed. The server stores the timestamp in the timezone of the server and the UI displays the timestamp converted to the local timezone.
- **Operation** The action performed.
  - HBase createTable, deleteTable, modifyTable, addColumn, modifyColumn, deleteColumn, enableTable, disableTable, move, assign, unassign, balance, balanceSwitch, shutdown, stopMaster, flush, split, compact, compactSelection, getClosestRowBefore, get, exists, put, delete, checkAndPut, checkAndDelete. incrementColumnValue, append, increment, scannerOpen, grant, revoke
  - HDFS setPermission, setOwner, open, concat, setTimes, createSymlink, setReplication, create, append, rename, delete, getfileinfo, mkdirs, listStatus, fsck
  - Hive EXPLAIN, LOAD, EXPORT, IMPORT, CREATEDATABASE, DROPDATABASE, SWITCHDATABASE, DROPTABLE, DESCTABLE, DESCFUNCTION, MSCK, ALTERTABLE ADDCOLS, ALTERTABLE REPLACECOLS, ALTERTABLE\_RENAMECOL, ALTERTABLE\_RENAMEPART, ALTERTABLE\_RENAME, ALTERTABLE DROPPARTS, ALTERTABLE ADDPARTS, ALTERTABLE TOUCH, ALTERTABLE ARCHIVE, ALTERTABLE UNARCHIVE, ALTERTABLE PROPERTIES, ALTERTABLE SERIALIZER, ALTERPARTITION\_SERIALIZER, ALTERTABLE\_SERDEPROPERTIES, ALTERPARTITION SERDEPROPERTIES. ALTERTABLE CLUSTER SORT, SHOWDATABASES, SHOWTABLES, SHOW TABLESTATUS, SHOW\_TBLPROPERTIES, SHOWFUNCTIONS, SHOWINDEXES, SHOWPARTITIONS, SHOWLOCKS, CREATEFUNCTION, DROPFUNCTION, CREATEVIEW, DROPVIEW, CREATEINDEX, DROPINDEX, ALTERINDEX\_REBUILD, ALTERVIEW\_PROPERTIES, LOCKTABLE, UNLOCKTABLE, ALTERTABLE PROTECTMODE, ALTERPARTITION PROTECTMODE, ALTERTABLE FILEFORMAT, ALTERPARTITION FILEFORMAT, ALTERTABLE LOCATION, ALTERPARTITION LOCATION, CREATETABLE, CREATETABLE AS SELECT, QUERY, ALTERINDEX PROPS, ALTERDATABASE, DESCDATABASE, ALTER\_TABLE\_MERGE, ALTER\_PARTITION\_MERGE, GRANT\_PRIVILEGE, REVOKE\_PRIVILEGE, SHOW\_GRANT, GRANT\_ROLE, REVOKE\_ROLE, SHOW\_ROLE\_GRANT, CREATEROLE, DROPROLE

- Impala Query, Insert, Update, Delete, GRANT PRIVILEGE, REVOKE PRIVILEGE, SHOW GRANT, GRANT\_ROLE, REVOKE\_ROLE, SHOW\_ROLE\_GRANT, CREATEROLE, DROPROLE
- Sentry GRANT PRIVILEGE, REVOKE PRIVILEGE, ADD ROLE TO GROUP, DELETE ROLE FROM GROUP, CREATE\_ROLE, DROP\_ROLE
- **Username** The name of the user that performed the action.
- Impersonator If the action was requested by another service, the name of the user that invoked the service action on behalf of the user.
  - When Sentry is not enabled, the Impersonator field always shows.
  - When Sentry is enabled, the Impersonator field shows for services other than Hive.
- IP Address The IP address of the host where the service action occurred.
- **Service Name** The name of the service that performed the service action.

# Viewing Audit Events

- 1. Start and log into the Navigator UI.
- 2. Click the Audits tab. The Audit Events report displays all audit events that occurred during the last hour.

# Filtering Audit Events

### Specifying a Time Range

- 1. Click the date-time range at the top right of the audits page.
- **2.** Do one of the following:
  - Click a Last n hours link.
  - Specify a custom range:
    - Click Custom range.
    - 2. In the Selected Range endpoints, click each endpoint and specify a date and time in the date control fields.
      - Date Click the down arrow to display a calendar and select a date, or click a subfield and click the spinner arrows or up and down arrow keys.
      - Time Click the hour, minute, and AM or PM fields and click the spinner arrows or up and down arrow keys to specify the value.
      - Move between fields using the right and left arrow keys.

#### 3. Click Apply.

#### Adding a Filter

- Click the Ticon that displays next to a property when you hover in one of the event entries. A filter containing the property, operator, and its value is added to the list of filters at the top and Cloudera Navigator redisplays all events that match the filter.
- Click the Filters link. The filters pane displaysand a filter control with property, operation, and value fields is added to the list of filters.
  - 1. Choose a property in the drop-down list. You can search by properties such as Username, Service Name, or Operation. The properties vary depending on the service or role.
  - 2. Choose an operator in the operator drop-down list.
  - 3. Type a property value in the value text field. To match a substring, use the like operator and specify % around the string. For example, to see all the audit events for files created in the folder /user/joe/out specify Source like %/user/joe/out%.

- 4. Click Apply. The property, operation, and value display above the list of audit events and the list of events displays all events that match the filter criteria.
- 5. Click Add New Filter to add more filters and repeat steps 1 through 4.

# Removing a Filter

- 1. Do one of the following:
  - Click the x next to the filter above the list of events.
  - Remove from the Filters pane:
    - 1. Click the **Filters** link. The filters pane displays.
    - **2.** Click the **=** at the right of the filter.
    - 3. Click Apply. The filter is removed from above the list of audit event and the list of events displays all events that match the filter criteria.

# **Creating Audit Reports**

- 1. Start and log into the Navigator UI.
- 2. Click the Audits tab. The Audit Events report displays all audit events that occurred during the last hour.
- 3. Do one of the following:
  - Save a filtered version of the Audit Events report:
    - 1. Optionally specify filters.
    - 2. Click Save As Report.
  - Create a new report:
    - 1. Click Create New Report.
- **4.** Enter a report name.
- 5. In the Default time range field, specify a relative time range. If you had specified a custom absolute time range before selecting **Save As Report**, the *custom absolute time range is discarded*.
- **6.** Optionally add <u>filters</u>.
- 7. Click Save.

## **Editing Audit Reports**

- 1. Start and log into the Navigator UI.
- 2. Click the Audits tab. The Audit Events report displays all audit events that occurred during the last hour.
- **3.** In the left pane, click a report name.
- 4. Click Edit Report.
- 5. In the Default time range field, specify a relative time range. If you had specified a custom absolute time range before selecting Save As Report, the custom absolute time range is discarded.
- **6.** Optionally add filters.
- 7. Click Save.

#### **Downloading Audit Events**

You can download audit events in the Audit UI or using the Audit API. An audit event contains the following fields: timestamp, service, username, ipAddress, command, resource, allowed, [operationText], serviceValues. The structure of the resource and serviceValues fields depends on the type of the service. Hive, Hue, Impala, and Sentry events have the operationText field, which contains the operation string.

#### Downloading Audit Events Using the Audit UI

- 1. Start and log into the Navigator UI.
- 2. Click the Audits tab. The Audit Events report displays all audit events that occurred during the last hour.
- 3. In the left pane, click a report name.
- Select Export > format, where format is CSV or JSON.

#### HDFS Audit Log Example

```
"items" : [ {
  "timestamp": "2014-10-10T16:39:25.656Z",
  "service": "HDFS-1",
  "username" : "admin",
  "ipAddress": "10.20.190.241",
  "command": "setPermission",
  "resource" : "/user/hive/warehouse/sample_09/000000_0",
  "allowed" : true,
  "serviceValues" :
    "dest" : null,
    "delegation_token_id" : null,
    "permissions" : "rwxrwxrwt",
    "src" : "/user/hive/warehouse/sample_09/000000_0"
}, {
  "timestamp" : "2014-10-10T16:39:25.632Z",
  "service": "HDFS-1",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "setPermission",
  "resource" : "/user/hive/warehouse/sample_09",
  "allowed" : true,
  "serviceValues" :
    "dest" : null,
    "delegation_token_id" : null,
    "permissions" : "rwxrwxrwt"
    "src": "/user/hive/warehouse/sample 09"
}, {
  "timestamp" : "2014-10-10T16:39:25.606Z",
  "service" : "HDFS-1",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "setOwner",
  "resource" : "/user/hive/warehouse/sample_09",
  "allowed" : false,
  "serviceValues" :
    "dest" : null,
    "delegation_token_id" : null,
    "permissions" : null,
    "src" : "/user/hive/warehouse/sample_09"
},'{
  "timestamp" : "2014-10-10T16:39:25.590Z",
 "service": "HDFS-1",
"username": "admin",
"ipAddress": "10.20.190.241",
  "command" : "delete"
  "resource" : "/user/hive/warehouse/sample_09",
  "allowed" : true,
  "serviceValues" :
    "dest" : null,
    "delegation_token_id" : null,
    "permissions" : null,
    "src" : "/user/hive/warehouse/sample_09"
}, '{
   "timestamp" : "2014-10-10T16:39:25.581Z",
  "service": "HDFS-1",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "getfileinfo",
```

```
"resource" : "/user/hive/warehouse",
"allowed" : true,
"serviceValues" :
  "dest" : null,
  "delegation_token_id" : null,
  "permissions" : null,
  "src" : "/user/hive/warehouse"
"timestamp" : "2014-10-10T16:39:25.575Z",
"service": "HDFS-1",
"username" : "admin",
"ipAddress" : "10.20.190.241",
"command" : "getfileinfo",
"resource" : "/user/hive/warehouse/sample_09",
"allowed" : true,
"serviceValues" :
  "dest" : null,
  "delegation_token_id" : null,
  "permissions" : null,
  "src" : "/user/hive/warehouse/sample_09"
```

In this example, the first event access was denied, and therefore the allowed field has the value false.

#### Hive Example - via downloaded JSON file

The following records list Hive operations to create and load a table:

```
[ {
  "timestamp" : "2014-10-10T16:39:26.184Z",
  "service": "HIVE-1",
  "username" : "admin",
  "ipAddress" : "10.20.190.241",
  "command" : "QUERY"
 "resource" : "default:sample_09",
  "operationText" : "INSERT OVERWRITE \n TABLE sample_09 \nSELECT \n
sample_07.code,sample_08.description \n FROM sample_07 \n JOIN sample_08 \n WHERE
sample_08.code = sample_07.code",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "TABLE",
    "database_name" : "default",
    "operation_text" : "INSERT OVERWRITE \n TABLE sample_09 \nSELECT \n
\verb|sample_07.code,sample_08.description \n FROM sample_07 \n JOIN sample_08 \n WHERE \\
sample_08.code = sample_07.code",
    "resource_path" : "/user/hive/warehouse/sample_09",
    "table_name" : "sample_09"
},'{
   "timestamp" : "2014-10-10T16:39:26.183Z",
 "service": "HIVE-1",
 "username" : "admin",
  "ipAddress" : "10.20.190.241",
  "command" : "QUERY"
  "resource" : "default:sample_07",
  "operationText" : "INSERT OVERWRITE \n TABLE sample_09 \nSELECT \n
sample_07.code,sample_08.description \n FROM sample_07 \n JOIN sample_08 \n WHERE
sample_08.code = sample_07.code",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "TABLE",
    "database_name" : "default",
    "operation_text" : "INSERT OVERWRITE \n TABLE sample_09 \nSELECT \n
sample_07.code,sample_08.description \n FROM sample_07 \n JOIN sample_08 \n WHERE
sample_08.code = sample_07.code",
    "resource_path" : "/user/hive/warehouse/sample_07",
    "table_name" : "sample_07"
}, {
```

```
"timestamp" : "2014-10-10T16:39:26.182Z",
  "service" : "HIVE-1",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "QUERY",
  "resource" : "default:sample_08",
  "operationText": "INSERT OVERWRITE \n TABLE sample_09 \nSELECT \n
sample_07.code,sample_08.description \n FROM sample_07 \n JOIN sample_08 \n WHERE
sample_08.code = sample_07.code",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "TABLE",
    "database_name" : "default".
    "operation_text" : "INSERT OVERWRITE \n TABLE sample_09 \nSELECT \n
sample_07.code,sample_08.description \n FROM sample_07 \n JOIN sample_08 \n WHERE
sample_08.code = sample_07.code",
    "resource_path" : "/user/hive/warehouse/sample_08",
"table_name" : "sample_08"
}, {
  "timestamp" : "2014-10-10T16:38:18.604Z",
  "service" : "HIVE-1",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command": "CREATETABLE",
  "resource" : "default:sample_09",
  "operationText": "CREATE TABLE sample_09 (code string, description string) ROW FORMAT
 DELIMITED FIELDS TERMINATED BY '\\t' STORED AS TextFile",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "TABLE",
    "database_name" : "default",
    "operation_text" : "CREATE TABLE sample_09 (code string, description string) ROW
FORMAT DELIMITED FIELDS TERMINATED BY '\\t' STORED AS TextFile",
    "resource_path" : "",
    "table_name" : "sample_09"
}, {
  "timestamp" : "2014-10-10T16:38:18.602Z",
  "service" : "HIVE-1",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "CREATETABLE",
  "resource" : "default:"
  "operationText" : "CREATE TABLE sample_09 (code string,description string) ROW FORMAT
 DELIMITED FIELDS TERMINATED BY '\\t' STORED AS TextFile",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "DATABASE",
    "database_name" : "default"
    "operation_text" : "CREATE TABLE sample_09 (code string, description string) ROW
FORMAT DELIMITED FIELDS TERMINATED BY '\\t' STORED AS TextFile",
    "resource_path" : "/user/hive/warehouse",
    "table_name" : ""
}, '{
   "timestamp" : "2014-10-10T16:37:06.836Z",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "LOAD",
  "resource" : ":",
  "operationText": "LOAD DATA INPATH\n '/user/admin/sample_08' OVERWRITE INTO
TABLE sample_08",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "DFS_DIR",
    "database_name" : ""
    "operation_text" : "LOAD DATA INPATH\n '/user/admin/sample_08' OVERWRITE INTO
 TABLE sample_08",
    "resource_path" : "/user/admin/sample_08",
"table_name" : ""
```

```
}, {
  "timestamp" : "2014-10-10T16:37:06.836Z",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "LOAD",
"resource" : "default:sample_08",
  "operationText": "LOAD DATA INPATH\n '/user/admin/sample_08' OVERWRITE INTO
TABLE sample_08",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "TABLE",
    "database_name" : "default"
   "operation_text" : "LOAD DATA INPATH\n
                                              '/user/admin/sample_08' OVERWRITE INTO
 TABLE sample_08",
    "resource_path" : "/user/hive/warehouse/sample_08",
    "table_name" : "sample_08"
  "timestamp" : "2014-10-10T16:37:05.752Z",
  "service": "HIVE-1"
  "username" : "admin",
  "ipAddress" : "10.20.190.241",
  "command" : "DESCTABLE",
  "resource" : "default:sample_08",
  "operationText" : "DESCRIBE EXTENDED sample_08",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "TABLE",
    "database_name" : "default",
    "operation_text" : "DESCRIBE EXTENDED sample_08"
    "resource_path" : "/user/hive/warehouse/sample_08",
    "table_name" : "sample_08"
}, '{
   "timestamp" : "2014-10-10T16:37:05.379Z",
  "service": "HIVE-1",
  "username" : "admin"
  "ipAddress" : "10.20.190.241",
  "command" : "LOAD",
"resource" : "default:sample_07",
  "operationText": "LOAD DATA INPATH\n '/user/admin/sample_07' OVERWRITE INTO
TABLE sample_07",
  "allowed" : true,
  "serviceValues" : {
    "object_type" : "TABLE",
    "database_name" : "default"
    "operation_text" : "LOAD DATA INPATH\n '/user/admin/sample_07' OVERWRITE INTO
 TABLE sample_07",
    "resource_path" : "/user/hive/warehouse/sample_07",
    "table_name" : "sample_07"
}, {
    "timestamp" : "2014-10-10T16:37:05.377Z",
  "service": "HIVE-1"
  "username" : "admin",
  "ipAddress" : "10.20.190.241",
  "command": "LOAD",
  "resource" : ":",
  "operationText" : "LOAD DATA INPATH\n '/user/admin/sample_07' OVERWRITE INTO
TABLE sample_07",
  "allowed" : true,
  "serviceValues" :
    "object_type" : "DFS_DIR",
    "database_name" : ""
    "operation_text" : "LOAD DATA INPATH\n
                                               '/user/admin/sample_07' OVERWRITE INTO
 TABLE sample_07"
    "resource_path" : "/user/admin/sample_07",
    "table_name" : ""
  "timestamp" : "2014-10-10T16:37:00.002Z",
 "service": "HIVE-1",
```

```
"username" : "admin",
"ipAddress" : "10.20.190.241",
"command" : "DESCTABLE",
"resource" : "default:sample_07",
"operationText" : "DESCRIBE EXTENDED sample_07",
"allowed" : true,
"serviceValues" :
  "object_type" : "TABLE",
  "database_name" : "default",
  "operation_text" : "DESCRIBE EXTENDED sample_07"
  "resource_path" : "/user/hive/warehouse/sample_07",
  "table_name" : "sample_07"
```

#### Downloading Audit Events Using the Audit API

You can filter and download audit events using the Cloudera Navigator API.

### Hive Example - via audit API

To download the audits events using the API, issue the request

http://host-1.ent.cloudera.com:7187/api/v3/audits?query=service==\*HIVE\*, which could return

```
"items" : [ {
  "timestamp": "2014-10-07T21:09:05.804Z",
  "service": "HIVE-1",
 "username" : "test",
 "impersonator" : ""
 "ipAddress": "20.10.191.128",
  "command" : "CREATEROLE",
 "resource" : ":",
  "operationText" : "CREATE ROLE bad_role",
  "allowed" : false,
  "serviceValues" : {
    "object_type" : "UNKNOWN",
"database_name" : "",
    "operation_text" : "CREATE ROLE bad_role",
    "resource_path" : "",
    "table_name" : ""
  "timestamp" : "2014-10-07T21:08:52.036Z",
 "service": "HIVE-1",
 "username" : "test",
 "ipAddress" : "20.10.191.128",
"command" : "DROPTABLE",
  "resource" : "default:ratings_sum",
  "operationText" : "DROP TABLE ratings_sum",
  "allowed" : true,
  "serviceValues" : {
    "object_type" : "TABLE",
    "database_name" : "default",
    "operation_text" : "DROP TABLE ratings_sum"
    "resource_path" : "/user/hive/warehouse/ratings_sum",
    "table_name" : "ratings_sum"
```

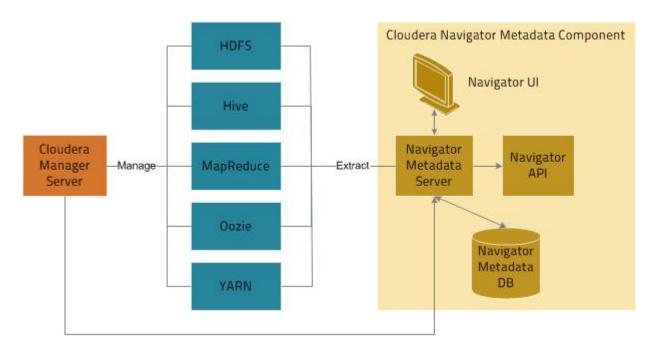
# Downloading HDFS Directory Access Permission Reports

Required Role: Cluster Administrator Full Administrator

For each HDFS service you can download a report that details the HDFS directories a group has permission to access.

- 1. In the Cloudera Manager Admin Console, click Clusters > ClusterName > General > Reports.
- **2.** In the Directory Access by Group row, click **CSV** or **XLS**. The Download User Access Report pop-up displays.
  - **a.** In the pop-up, type a group and directory.
  - **b.** Click **Download**. A report of the selected type will be generated containing the following information path, owner, permissions, and size – for each directory contained in the specified directory that the specified group has access to.

The Cloudera Navigator metadata component provides data discovery and data lineage management functions. The architecture of the Cloudera Navigator metadata component is illustrated below.



The Navigator Metadata Server performs the following functions:

- Obtains connection information about the services whose data it manages from the Cloudera Manager Server
- Extracts entity metadata from the services at periodic intervals
- Manages and applies metadata extraction policies
- Indexes and stores entity metadata
- Manages user authorization data
- Manages audit report metadata
- Implements the Navigator UI and API

The Navigator Metadata database stores entity metadata, policies, and user authorization and audit report metadata.

# Metadata

The Cloudera Navigator Metadata component manages metadata about the entities in a CDH cluster and relations between the entities.

The Navigator metadata schema defines the types of metadata that are available for each entity type it supports. The types of metadata defined by the Navigator Metadata component include: the name of an entity, the service that manages or uses the entity, type, path to the entity, date and time of creation, access, and modification, size, owner, purpose, and relations—parent-child, data flow, and instance of—between entities.

For example, the following shows the property sheet of a file entity:

#### sample\_07.csv

tags: source type: HDFS category: FILE path: /user/hdfs/sample\_07.csv owner: hdfs group: supergroup size: 44.98KiB last accessed: Oct 8 2013 1:33 PM last modified: Oct 8 2013 1:33 PM

There are two classes of metadata:

- technical metadata metadata defined when entities are extracted. You cannot modify technical metadata.
- business metadata metadata added to extracted entities. You can add and modify business metadata before or after entities are extracted.

#### Metadata Extraction

The Navigator Metadata Server extracts metadata for the following resource types from the listed servers:

- HDFS Extracts HDFS metadata at the next scheduled extraction run after an HDFS checkpoint. However, if you have high availability enabled, metadata is extracted as soon as it is written to the JournalNodes.
- Hive Extracts database and table metadata from the Hive Metastore Server.
- MapReduce Extracts job metadata from the JobTracker. The default setting in Cloudera Manager retains a maximum of five jobs, which means if you run more than five jobs between Navigator extractions, the Navigator Metadata Server would extract the five most recent jobs.
- Oozie Extracts Oozie workflows from the Oozie Server.
- **Pig** Extracts Pig script runs from the JobTracker or Job History Server.
- Sqoop 1 Extracts database and table metadata from the Hive Metastore Server.
- YARN Extracts job metadata from the Job History Server.

If an entity is created at time to in the system, that entity will be extracted and linked in Navigator after the extraction poll period (default 10 minutes) plus a service-specific interval as follows:

- HDFS: t0 + extraction poll period + HDFS checkpoint interval (default 1 hour)
- HDFS + HA: t0 + extraction poll period
- Hive: t0 + extraction poll period + Hive maximum wait time (default 60 minutes)

#### Metadata Indexing

After metadata is extracted it is indexed and made available for searching by an embedded Solr engine. The Solr schema indexes two types of metadata: entity properties and relationship between entities.

You can search entity metadata using the Navigator UI. Relationship metadata is implicitly visible in lineage diagrams and explicitly available in a lineage file.

#### Metadata Search

Search in the Navigator Metadata component is implemented by an embedded Solr engine that supports the syntax described in LuceneOParserPlugin.

# Search Syntax

You construct search strings by specifying the value of a default property, property name-value pairs, or user-defined name-value pairs using the syntax:

- Property name-value pairs propertyName: value, where
  - propertyName is one of the properties listed in Search Properties on page 21.
  - value is a single value or range of values specified as [value1 TO value2]. In a value, \* is a wildcard. In property name-value pairs you must escape special characters :, /, and \* with the backslash character \. For example, fileSystemPath:\/user\/admin.
- User-defined name-value pairs up\_propertyName:value.

To construct complex strings, join multiple property-value pairs using the or and and operators.

#### **Example Search Strings**

- Filesystem path /user/admin fileSystemPath:\/user\/admin
- Descriptions that start with the string "Banking" description: Banking\*
- Sources of type MapReduce or Hive sourceType:MAPREDUCE or sourceType:HIVE
- Directories owned by hdfs in the path /user/hdfs/input owner:HDFS and type:directory and fileSystemPath:\/user\/hdfs\/input
- Job started between 20:00 to 21:00 UTC started: [2013-10-21T20:00:00.000z TO 2013-10-21T21:00:00.000Z1
- User-defined key-value project-customer1 up\_project:customer1
  - Note: When viewing MapReduce jobs in the Cloudera Manager Activities page, the string that appear in a job's Name column equates to the original Name property. Therefore, to specify a MapReduce job's name in a search, use the following string: (resType:mapreduce) and (originalName: jobName), where jobName is the value in the job's Name column.

## Search Properties

A reference for the search schema properties.

#### **Default Properties**

The following properties can be searched by simply specifying a property value: type, fileSystemPath, inputs, jobId, mapper, mimeType, name, originalName, outputs, owner, principal, reducer, tags.

#### **Common Properties**

Name	Туре	Description
description	text	Description of the entity.
group	caseInsensitiveText	The group to which the owner of the entity belongs.
name	ngramedText	The overridden name of the entity. If the name has not been overridden, this value is empty. Names cannot contain spaces.
operationType	ngramedText	The type of an operation:
		■ Pig - SCRIPT
		■ Sqoop - Table Export, Query Import
originalName	ngramedText	The name of the entity when it was extracted.
originalDescription	text	The description of the entity when it was extracted.

Name	Туре	Description
owner	caselnsensitiveText	The owner of the entity.
principal	caseInsensitiveText	For entities with type OPERATION_EXECUTION, the initiator of the entity.
tags	ngramedText	A set of tags that describe the entity.
type	ngramedText	The type of the entity. The available types depend on the entity's source type:  HDFS - DIRECTORY, FILE  HIVE - DATABASE, TABLE, FIELD, OPERATION, OPERATION_EXECUTION, SUB_OPERATION, PARTITION, RESOURCE, UNKNOWN, VIEW  MAPREDUCE - OPERATION, OPERATION_EXECUTION  OOZIE - OPERATION, OPERATION_EXECUTION  PIG - OPERATION, OPERATION_EXECUTION  SQOOP - OPERATION, OPERATION_EXECUTION, SUB_OPERATION  YARN - OPERATION, OPERATION_EXECUTION
Query	J.	
queryText	string	The text of a Hive or Sqoop query.
Source	1	
clusterName	string	The name of the cluster in which the entity is stored.
sourceId	string	The ID of the source type.
sourceType	caseInsensitiveText	The source type of the entity: HDFS, HIVE, MAPREDUCE, OOZIE, PIG, SQOOP, YARN.
sourceUrl	string	The URL of the source type.
Timestamps		
The available timestamp fields vary by the source type:  HDFS - lastModified, lastAccessed HIVE - created, lastAccessed MAPREDUCE, PIG, SQOOP, and YARN - started, ended	date	<pre>Timestamps in the Solr Date Format. For example:     lastAccessed:[* TO NOW]     created:[1976-03-06T23:59:59.999Z TO *]     started:[1995-12-31T23:59:59.999Z TO 2007-03-06T00:00:00Z]     ended:[NOW-1YEAR/DAY TO NOW/DAY+1DAY]     created:[1976-03-06T23:59:59.999Z TO 1976-03-06T23:59:59.999Z+1YEAR]     lastAccessed:[1976-03-06T23:59:59.999Z/YEAR TO 1976-03-06T23:59:59.999Z]</pre>

# **HDFS Properties**

Name	Туре	Description
fileSystemPath	path	The path to the entity.

Name	Туре	Description
compressed	Boolean	Indicates whether the entity is compressed.
deleted	Boolean	Indicates whether the entity has been moved to the Trash folder.
deleteTime	date	The time the entity was moved to the Trash folder.
mimeType	ngramedText	The MIME type of the entity.
parentPath	string	The path to the parent entity of a child entity. For example: parent path:/default/sample_07 for the table sample_07 from the Hive database default.
permissions	string	The UNIX access permissions of the entity.
size	long	The exact size of the entity in bytes or a range of sizes. Range examples: size:[1000 TO *],size: [* TO 2000], and size:[* TO *] to find all fields with a size value.

# MAPREDUCE and YARN Properties

Name	Туре	Description
inputRecursive	Boolean	Indicates whether files are searched recursively under the input directories, or just files directly under the input directories are considered.
jobId	ngramedText	The ID of the job. For a job spawned by Oozie, the workflow ID.
mapper	string	The fully-qualified name of the mapper class.
outputKey	string	The fully-qualified name of the class of the output key.
outputValue	string	The fully-qualified name of the class of the output value.
reducer	string	The fully-qualified name of the reducer class.

# **OPERATION Properties**

Name	Туре	Description	
Operation			
inputFormat	string	The fully-qualified name of the class of the input format.	
outputFormat	string	The fully-qualified name of the class of the output format.	
Operation Exec	Operation Execution		
inputs	string	The name of the entity input to an operation execution. For entities of resource type MAPREDUCE, it is usually a directory. For entities of resource type Hive, it is usually a table.	
outputs	string	The name of the entity output from an operation execution. For entities of resource type MAPREDUCE, it is usually a directory. For entities of resource type Hive, it is usually a table.	

# **HIVE Properties**

Name	Туре	Description
Field		
dataType	ngramedText	The type of data stored in a field (column).

Name	Туре	Description
Table		
compressed	Boolean	Indicates whether a Hive table is compressed.
serDeLibName	string	The name of the library containing the SerDe class.
serDeName	string	The fully-qualified name of the SerDe class.
Partition		
partitionColNames	string	The table columns that define the partition.
partition(b)[Values	string	The table column values that define the partition.

## **Oozie Properties**

Name	Туре	Description
status	string	The status of the Oozie workflow: RUNNING, SUCCEEDED, or FAILED.

## **PIG Properties**

Name	Туре	Description
scriptId	string	The ID of the Pig script.

## **SQOOP Properties**

Name	Туре	Description
dburl	string	The URL of the database from or to which the data was imported or exported.
dbTable	string	The table from or to which the data was imported or exported.
dbUser	string	The database user.
dbWhere	string	A where clause that identifies which rows were imported.
doColumExpression	string	An expression that identifies which columns were imported.

# **Accessing Metadata**

Required Role: Lineage Viewer Policy Administrator | Metadata Administrator | Full Administrator

You can access metadata through the Navigator UI or through the Navigator API.

# Navigator Metadata UI

Required Role: Lineage Viewer Policy Administrator | Metadata Administrator Full Administrator

# Searching Metadata

You perform search in the Navigator UI by typing search strings or constructing search strings using UI controls.

- 1. Start and log into the Navigator UI.
- 2. Do one of the following:
  - Type a search string into the **Search** box that conforms to the <u>search syntax</u>. The Search Results page displays as soon as you start typing.

 Click the Query Builder link. The Query Builder landing page displays with the result of the wildcard search (\*). The Query Builder landing page displays Source Type and Type facets that match the search results with the number of results that match each value of those properties. You can filter the search results by clicking specific values for those properties or adding new properties.

The Full Query read-only box displays the search string constructed from the specified filters. Click Show **n** Results to display the Search Results page.

#### Search Results

The Search Results page has a Search box and two panes: the Query Builder pane and The Search Results pane.

1 to 25 of 83 results in pages listing 25 The Search Results pane displays the number of matching entries

entities per page. You can view the pages using the page control at the bottom of each page.

Each entry in the result list contains:

- Source type
- Entity name the name is a link to a page that displays the entity property editor and lineage diagram.
- Entity properties
- If Hue is running, a link at the far right labeled View in Hue that opens the Hue browser for the entity:
  - HDFS directories and files File Browser
  - Hive database and tables Metastore Manager
  - MapReduce, YARN, Pig Job Browser

#### For example:



#### Specifying Property Values in the Query Builder Pane

The Query Builder pane contains a Search box and a set of graphical controls that allow you to select property values to filter search results. You can filter using the Search box or the graphical controls.

In the **Search** box, type the values of default properties.

To filter on a property value for non-default properties, specify values as follows:

- Boolean Check the checkbox.
- Enumerated Start typing or click the field and then select from a drop-down list.
- Timestamps Specified in the format mm/dd/yyyy hh:mm [AM|PM] in a date control. The server stores the timestamp in UTC and the UI displays the timestamp converted to the local timezone. In the date control:
  - spinner arrows or up and down arrow keys.
  - Time Click the hour, minute, and AM or PM fields and click the spinner arrows 🗖 or up and down arrow keys to specify the value.
  - Move between fields using the right and left arrow keys.

To add a property, click **Add another filter...** and select a property name.

# Navigator API

The Navigator API allows you to search entity metadata using a REST API. For information about the API, see Cloudera Navigator API.

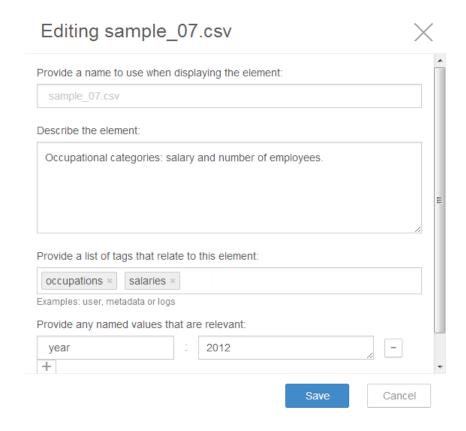
# Modifying Business Metadata

The Cloudera Navigator Metadata component allows you to add and modify the following business metadata associated with entities: display name, description, tags, and user-defined name-value pairs. You can modify business metadata using the Navigator Metadata UI, MapReduce service and job properties, Navigator metadata files, and the Navigator Metadata API.

Required Role: Policy Administrator Metadata Administrator **Full Administrator** 

#### Modifying Business Metadata Using the Navigator UI

- 1. Run a <u>search</u> in the Navigator UI.
- 2. Click an entity link returned in the search. The metadata pane displays on the left and the lineage page displays on the right.
- 3. In the top-right of the metadata pane, click . The Editing *entity* dialog drops down.
- 4. Edit any of the fields as instructed. Press Enter or Tab to create new tag entries. For example, a Description, the tags occupations and salaries, and property year with value 2012 have been added to the file sample\_07.csv:



- Note: You can specify special characters (for example, ".", "") in the name, but it will make searching for the entity more difficult as some characters collide with special characters in the search syntax.
- **5.** Click **Save**. The new metadata appears in the metadata pane:

# sample 07.csv Occupational categories: salary and number of employees. tags: occupations salaries source type: HDFS category: FILE path: /user/hdfs/sample 07.csv owner: hdfs group: supergroup size: 44.98KiB last accessed: Oct 8 2013 1:33 PM last modified: Oct 8 2013 1:33 PM year: 2012

#### Modifying MapReduce Business Metadata

You can set MapReduce job metadata statically for all jobs, or dynamically for a specific job or job instance.

To statically set metadata for all MapReduce jobs, do the following:

- 1. Do one of the following:
  - Select Clusters > Cloudera Management Service > Cloudera Management Service.
  - On the Status tab of the Home page, in Cloudera Management Service table, click the Cloudera Management Service link.
- 2. Click the Configuration tab.
- 3. Select Navigator Metadata Server Default Group > Advanced.
- 4. Click Navigator Metadata Server Advanced Configuration Snippet for cloudera-navigator.properties.
- **5.** Specify values for one or more of the following properties:
  - nav.user\_defined\_properties = comma-separated list of user-defined property names
  - nav.tags = comma-separated list of property names that serve as tags
- 6. Click Save Changes.
- 7. Click the Instances tab.
- 8. Check the checkbox next to the Navigator Metadata Server role.
- 9. Select Actions for Selected > Restart.

To modify parameters dynamically, specify one or more of the following properties in a job configuration:

- nav.job.user\_defined\_properties = comma-separated list of user-defined property names for a job (type:OPERATION)
- nav.job.tags = comma-separated list of property names that serve as tags for a job
- nav.jobexec.user\_defined\_properties = comma-separated list of user-defined property names for a job execution (type: OPERATION\_EXECUTION)

nav.jobexec.tags = comma-separated list of property names that serve as tags for a job execution

For example, to capture hive.mapred.partitioner in the MapReduce jobs that are launched through Hive, **Set** nav.user\_defined\_properties=hive.mapred.partitioner.

#### Modifying HDFS Business Metadata Using Metadata Files

You can add tags and properties to HDFS entities using metadata files. The reasons to use metadata files are to assign metadata to entities in bulk and to create metadata before the metadata is extracted. A metadata file is a JSON file with the following structure:

```
"name" : "aName",
"description" : "a description",
"properties" : {
   "prop1" : "value1", "prop2" : "value2"
"tags" : [ "tag1" ]
```

To add metadata files to files and directories, create a metadata file with the extension .navigator, naming the files as follows:

- File The path of the metadata file must be .filename.navigator. For example, to apply properties to the file /user/test/file1.txt, the metadata file path is /user/test/.file1.txt.navigator.
- **Directory** The path of the metadata file must be *dirpath* / .navigator. For example, to apply properties to the directory /user, the metadata path must be /user/.navigator.

The metadata file is applied to the entity metadata when the extractor runs.

#### Modifying HDFS and Hive Business Metadata Using the Navigator API

You can use the Cloudera Navigator API to modify the metadata of HDFS or Hive entities whether or not the entities have been extracted. If an entity has been extracted at the time the API is called, the metadata will be applied immediately. If the entity has not been extracted, you can preregister metadata which is then applied once the entity is extracted. Metadata is saved regardless of whether or not a matching entity is extracted, and Navigator does not perform any cleanup of unused metadata.

If you call the API before the entity is extracted, the metadata is stored with the entity's identity, source ID, metadata fields (name, description, tags, properties), and the fields relevant to the identifier. The rest of the entity fields (such as type) will not be present. To view all stored metadata, you can use the API to search for entities without an internal type:

```
curl http://hostname:port/api/v2/entities/?query=-internalType:* -u username:password
-X GET
```

The metadata provided via the API overwrites existing metadata. If, for example, you call the API with an empty name and description, empty array for tags, and empty dictionary for properties, the call removes this metadata. If you leave out the tags or properties fields, the existing values remain unchanged.

Modifying metadata using HDFS metadata files and the metadata API at the same time is not supported. You must use one or the other, because the two methods behave slightly differently. Metadata specified in files is merged with existing metadata whereas the API overwrites metadata. Also, the updates provided by metadata files wait in a queue before being merged, but API changes are committed immediately. This means there may be some inconsistency if a metadata file is being merged around the same time the API is in use.

You modify metadata using either the PUT or POST method. Use the PUT method if the entity has been extracted and the POST method to preregister metadata. The syntax of the methods are:

PUT

```
curl http://hostname:port/api/v2/entities/identity -u username:password -X PUT -H
"Content-Type: application/json" -d '{properties}'
```

#### where *properties* are:

- name: name metadata
- description: description metadata
- tags: tag metadata
- properties: property metadata

All existing naming rules apply, and if any value is invalid, the entire request will be denied.

POST

```
curl http://hostname:port/api/v2/entities/ -u username:password -X POST -H
"Content-Type: application/json" -d '{properties}'
```

#### where *properties* are:

 sourceId (required): The source ID must match an existing source ID. After the first extraction, you can retrieve the source IDs using the call

```
curl http://hostname:port/api/v2/entities/?query=type:SOURCE -u username:password
```

#### For example:

```
"identity" : "a09b0233cc58ff7d601eaa68673a20c6",
"originalName" : "HDFS-1",
"sourceId" : null,
"firstClassParentId" : null,
"parentPath" : null,
"extractorRunId" : null,
"name" : "HDFS-1",
"description" : null,
"tags" : null,
"properties" : null,
"clusterName" : "Cluster 1",
"sourceUrl" : "hdfs://hostname:8020",
"sourceType" : "HDFS",
"sourceExtractIteration": 4935,
"type" : "SOURCE",
"internalType" : "source"
```

If you have HDFS-1 and HDFS-2, you must specify the source that contains the entity you're expecting

- parentPath: The path of the parent entity, as defined below:
  - HDFS file or directory: fileSystemPath of the parent directory (do not provide this field if the entity being affected is the root directory). Example parentPath for /user/admin/input\_dir: /user/admin. If you apply metadata to a directory, the metadata does not propagate to any files and folders in that directory.
  - Hive database: If you are updating database metadata, you do not specify this field.
  - Hive table or view: Name of database containing this table or view. Example for table in default database: default.
  - Hive column: database name/ table/ view name. Example for column in sample 07 table: default/sample\_07.
- originalName (required): The name as defined by the source system.

- HDFS file or directory: name of file or directory (ROOT if the entity is the root directory). Example originalName for /user/admin/input\_dir: input\_dir.
- Hive database, table, view, or column: name of the database, table, view, or column.
  - Example for default database: default
  - Example for sample\_07 table: sample\_07
- name: name metadata
- description: description metadata
- tags: tag metadata
- properties: property metadata

All existing naming rules apply, and if any value is invalid, the entire request will be denied.

#### HDFS PUT Example for /user/admin/input dir Directory

```
http://Navigator_Metadata_Server_host:port/api/v4/entities/e461de8de38511a3ac6740dd7d51b8d0
 -u username:password -X PUT -H "Content-Type: application/json"
-d '{ "name": "my_name", "description": "My description",
"tags":["tag1","tag2"],"properties":{"property1":"value1","property2":"value2"}}'
```

#### HDFS POST Example for /user/admin/input\_dir Directory

```
curl http://Navigator_Metadata_Server_host:port/api/v4/entities/ -u username:password
-X POST -H "Content-Type: application/json"
-d '{"sourceId": "a09b0233cc58ff7d601eaa68673a20c6",
"parentPath": "/user/admin", "originalName": "input_dir", "name": "my_name", "description": "My
 description",
"tags":["tag1","tag2"],"properties":{"property1":"value1","property2":"value2"}}'
```

#### Hive POST Example for total\_emp Column

```
curl http://Navigator_Metadata_Server_host:port/api/v4/entities/ -u username:password
-X POST -H "Content-Type: application/json"
-d '{"sourceId":"4fbdadc6899638782fc8cb626176dc7b",
"parentPath": "default/sample_07", "originalName": "total_emp",
"name": "my_name", "description": "My description",
"tags":["tag1","tag2"],"properties":{"property1":"value1","property2":"value2"}}'
```

# **Policies**

A policy defines a set of actions performed when a class of entities is extracted. The following actions are supported:

- Adding business metadata such as tags and properties.
- Sending a message to a JMS message queue. The JSON format message contains the metadata of the entity to which the policy applies and the message text specified in the policy:

```
{"entity":entity_properties, "userMessage":"some message text"}
```

For each action, certain properties support specifying a value using a policy expression.

#### **Viewing Policies**

Required Role: Policy Viewer Policy Administrator **Full Administrator** 

- 1. Start and log into the Navigator UI.
- 2. Click the Policies tab.

**3.** In the left pane, click a policy.

#### **Creating Policies**

#### **Full Administrator Policy Administrator** Required Role:

- 1. Start and log into the Navigator UI.
- 2. Depending on the starting point, do one of the following:

Action	Procedure
Policies page	<ol> <li>Click the Policies tab.</li> <li>Click Create a New Policy.</li> </ol>
Search results page	1. In the Search results page, click <b>Create a policy</b> .

- **3.** Enter a name for the policy.
- 4. Check the Enable Policy checkbox.
- 5. Specify the search query that defines the class of entities to which the policy applies. If you arrive at the Policies page by clicking a search result, the query property is populated with the query that generated the result. To display a list of entities that satisfy a search query, click the Search Results link.
- **6.** Specify an optional description for the policy.
- 7. If policy expressions are enabled and you choose to use policy expressions in properties that support expressions, specify required imports in the Import Statements field. For example, if your policy expression uses policy enums, for example: entity.get(FSEntityProperties.ORIGINAL\_NAME, Object.class) your import would be: import com.cloudera.nav.hdfs.model.FSEntityProperties. If your expression uses another library, such as Date, add the required import for that library.
- 8. Choose the schedule for applying the policy: On Data Change, Immediate, Once, or Recurring.
- 9. Check the checkboxes next to the desired actions and follow the appropriate procedure:

Action	Procedure
Assign Metadata	1. Specify the <u>business metadata</u> . Optionally check the <b>Expression</b> checkbox and specify a policy expression for the indicated fields.
Send Notification to JMS	<ol> <li>If not already configured, configure a JMS server and queue.</li> <li>Specify the queue name and message. Optionally check the Expression checkbox and specify a policy expression for the message.</li> </ol>

#### 10. Click Save.

#### Cloning and Editing a Policy

#### Required Role: Policy Administrator **Full Administrator**

- 1. Start and log into the Navigator UI.
- 2. Click the Policies tab.
- 3. In the left pane, click a policy.
- 4. Click Clone Policy or Edit Policy.
- **5.** Edit the policy name, search query, or policy actions.
- 6. Click Save.

#### **Deleting Policies**

**Policy Administrator Full Administrator** Required Role:

- 1. Start and log into the Navigator UI.
- **2.** Click the **Policies** tab.
- **3.** In the left pane, click a policy.
- 4. Click Delete and click OK to confirm.

### Policy Expressions

Policy expressions allow certain policy properties to be specified programmatically using Java expressions instead of string literals.

**Note:** A policy expression must evaluate to a string.

Policy expressions are not enabled by default. To enable policy expressions, follow the procedure in Enabling and Disabling Policy Expressions.

The supported policy properties are entity name and description, key-value pairs, and JMS notification message.

#### Entity Properties in Policy Expressions

To include entity properties in property expressions, use the entity.get method, which takes a property and a return type:

```
entity.get(XXProperties.Property, return_type)
```

XXProperties. Property is the Java enumerated value representing an entity property, where

- XX is FSEntity, HiveColumn, HiveDatabase, HivePartition, HiveQueryExecution, HiveQueryPart, HiveQuery HiveTable, HiveView, JobExecution, Job, WorkflowInstance, Workflow, PigField, PigOperationExecution, PigOperation, PigRelation, SqoopExportSubOperation, SqoopImportSubOperation, SqoopOperationExecution, <u>SqoopQueryOperation</u>, <u>SqoopTableExportOperation</u>, <u>or SqoopTableImportOperation</u>.
- Property is one of the properties listed in Entity Property Enum Reference on page 33.

If you don't need to specify a return type, use Object.class as the return type. However, if you want to do type-specific operations with the result, set the return type to the type in the comment in the enum property reference. For example, in FSEntityProperties, the return type of the ORIGINAL\_NAME property is java.lang.String.If you use String.class as the return type, you can use the String method toLowerCase() to modify the returned value: entity.get(FSEntityProperties.ORIGINAL\_NAME, String.class).toLowerCase().

#### **Expression Examples**

Set a filesystem entity name to the original name concatenated with the entity type:

```
entity.get(FSEntityProperties.ORIGINAL NAME, Object.class) + " " +
entity.get(FSEntityProperties.TYPE, Object.class)
```

Import Statements:

```
import com.cloudera.nav.hdfs.model.FSEntityProperties;
```

Expression to add the CREATED date to the name:

```
entity.get(FSEntityProperties.ORIGINAL_NAME, Object.class) + " - "
+ new SimpleDateFormat("yyyy-MM-dd").format(entity.get(FSEntityProperties.CREATED,
Instant.class).toDate())
```

#### Import Statements:

```
import com.cloudera.nav.hdfs.model.FSEntityProperties; import
java.text.SimpleDateFormat; import org.joda.time.Instant;
```

#### **Entity Property Enum Reference**

The following reference lists the Java enumerated values for retrieving properties of each entity type.

```
com.cloudera.nav.hdfs.model.FSEntityProperties
public enum FSEntityProperties implements PropertyEnum {
   PERMISSIONS, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  SIZE, // Return type: java.lang.Long
OWNER, // Return type: java.lang.String
  LAST_MODIFIED, // Return type: org.joda.time.Instant
  SOURCE_TYPE, // Return type: java.lang.String
  DELETED, // Return type: java.lang.Boolean FILE_SYSTEM_PATH, // Return type: java.lang.String
  CREATED, // Return type: org.joda.time.Instant
  LAST_ACCESSED, // Return type: org.joda.time.Instant
  GROUP, // Return type: java.lang.String
  MIME_TYPE, // Return type: java.lang.String
  DELETE_TIME, // Return type: java.lang.Long
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HiveColumnProperties
public enum HiveColumnProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  DELETED, // Return type: java.lang.Boolean DATA_TYPE, // Return type: java.lang.String
  ORIGINAL_DESCRIPTION, // Return type: java.lang.String NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HiveDatabaseProperties
public enum HiveDatabaseProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String ORIGINAL_DESCRIPTION, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String DELETED, // Return type: java.lang.Boolean
  FILE_SYSTEM_PATH, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HivePartitionProperties
public enum HivePartitionProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  DELETED, // Return type: java.lang.Boolean
  FILE_SYSTEM_PATH, // Return type: java.lang.String
  CREATED, // Return type: org.joda.time.Instant
  LAST_ACCESSED, // Return type: org.joda.time.Instant COL_VALUES, // Return type: java.util.List
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
```

```
PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HiveQueryExecutionProperties
public enum HiveQueryExecutionProperties implements PropertyEnum {
  SOURCE_TYPE, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  ENDED, // Return type: org.joda.time.Instant
  INPUTS, // Return type: java.util.Collection
  OUTPUTS, // Return type: java.util.Collection
  STARTED, // Return type: org.joda.time.Instant
  PRINCIPAL, // Return type: java.lang.String
  WF_INST_ID, // Return type: java.lang.String NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HiveQueryPartProperties
public enum HiveQueryPartProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HiveQueryProperties
public enum HiveQueryProperties implements PropertyEnum {
  SOURCE_TYPE, // Return type: java.lang.String
  INPUTS, // Return type: java.util.Collection OUTPUTS, // Return type: java.util.Collection QUERY_TEXT, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  WF_IDS, // Return type: java.util.Collection
  NAME, // Return type: java.lang.String
ORIGINAL_NAME, // Return type: java.lang.String
USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HiveTableProperties
public enum HiveTableProperties implements PropertyEnum {
  OWNER, // Return type: java.lang.String
  INPUT_FORMAT, // Return type: java.lang.String
  OUTPUT_FORMAT, // Return type: java.lang.String
  DELETED, // Return type: java.lang.Boolean
  FILE_SYSTEM_PATH, // Return type: java.lang.String
  COMPRESSED, // Return type: java.lang.Boolean
  PARTITION_COL_NAMES, // Return type: java.util.List CLUSTERED_BY_COL_NAMES, // Return type: java.util.List
  SORT_BY_COL_NAMES, // Return type: java.util.List
  SER_DE_NAME, // Return type: java.lang.String
  SER_DE_LIB_NAME, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  CREATED, // Return type: org.joda.time.Instant
  LAST_ACCESSED, // Return type: org.joda.time.Instant
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
```

```
SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.hive.model.HiveViewProperties
public enum HiveViewProperties implements PropertyEnum {
  DELETED, // Return type: java.lang.Boolean
  QUERY_TEXT, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  CREATED, // Return type: org.joda.time.Instant
  LAST_ACCESSED, // Return type: org.joda.time.Instant
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.mapreduce.model.JobExecutionProperties
public enum JobExecutionProperties implements PropertyEnum {
  SOURCE_TYPE, // Return type: java.lang.String
  JOB_ID, // Return type: java.lang.String ENDED, // Return type: org.joda.time.Instant
  INPUT_RECURSIVE, // Return type: boolean
  TYPE, // Return type: java.lang.String
  INPUTS, // Return type: java.util.Collection OUTPUTS, // Return type: java.util.Collection
  STARTED, // Return type: org.joda.time.Instant
  PRINCIPAL, // Return type: java.lang.String WF_INST_ID, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.mapreduce.model.JobProperties
public enum JobProperties implements PropertyEnum {
  ORIGINAL_NAME, // Return type: java.lang.String
  INPUT_FORMAT, // Return type: java.lang.String
OUTPUT_FORMAT, // Return type: java.lang.String
  OUTPUT_KEY, // Return type: java.lang.String
  OUTPUT_VALUE, // Return type: java.lang.String
  MAPPER, // Return type: java.lang.String REDUCER, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  WF_IDS, // Return type: java.util.Collection
  NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.oozie.model.WorkflowInstanceProperties
public enum WorkflowInstanceProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  CREATED, // Return type: org.joda.time.Instant
  JOB_ID, // Return type: java.lang.String
  STATUS, // Return type: java.lang.String
  ENDED, // Return type: org.joda.time.Instant
  INPUTS, // Return type: java.util.Collection
```

```
OUTPUTS, // Return type: java.util.Collection
  STARTED, // Return type: org.joda.time.Instant
  PRINCIPAL, // Return type: java.lang.String WF_INST_ID, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.oozie.model.WorkflowProperties
public enum WorkflowProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  WF_IDS, // Return type: java.util.Collection
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.pig.model.PigFieldProperties
public enum PigFieldProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
  INDEX, // Return type: int
SOURCE_TYPE, // Return type: java.lang.String
  DATA_TYPE, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.pig.model.PigOperationExecutionProperties
public enum PigOperationExecutionProperties implements PropertyEnum {
  SOURCE_TYPE, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  ENDED, // Return type: org.joda.time.Instant
  INPUTS, // Return type: java.util.Collection OUTPUTS, // Return type: java.util.Collection
  STARTED, // Return type: org.joda.time.Instant
  PRINCIPAL, // Return type: java.lang.String WF_INST_ID, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.pig.model.PigOperationProperties
public enum PigOperationProperties implements PropertyEnum {
  SOURCE_TYPE, // Return type: java.lang.String
  OPERATION_TYPE, // Return type: java.lang.String
  SCRIPT_ID, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  WF_IDS, // Return type: java.util.Collection
  NAME, // Return type: java.lang.String
ORIGINAL_NAME, // Return type: java.lang.String
USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
```

```
PARENT_PATH; // Return type: java.lang.String
\verb|com.cloudera.nav.pig.model.PigRelationProperties|\\
public enum PigRelationProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
SOURCE_TYPE, // Return type: java.lang.String
  FILE_SYSTEM_PATH, // Return type: java.lang.String
  SCRIPT_ID, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.sqoop.model.SqoopExportSubOperationProperties
public enum SqoopExportSubOperationProperties implements PropertyEnum {
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  INPUTS, // Return type: java.util.Collection
  FIELD_INDEX, // Return type: int
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.sqoop.model.SqoopImportSubOperationProperties
public enum SqoopImportSubOperationProperties implements PropertyEnum {
  DB_COLUMN_EXPRESSION, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  INPUTS, // Return type: java.util.Collection
  FIELD_INDEX, // Return type: int
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String
  USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.sqoop.model.SqoopOperationExecutionProperties
public enum SqoopOperationExecutionProperties implements PropertyEnum {
  SOURCE_TYPE, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  ENDED, // Return type: org.joda.time.Instant
  INPUTS, // Return type: java.util.Collection
  OUTPUTS, // Return type: java.util.Collection
  STARTED, // Return type: org.joda.time.Instant
  PRINCIPAL, // Return type: java.lang.String WF_INST_ID, // Return type: java.lang.String
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.sqoop.model.SqoopQueryOperationProperties
public enum SqoopQueryOperationProperties implements PropertyEnum {
  SOURCE_TYPE, // Return type: java.lang.String
  INPUTS, // Return type: java.util.Collection
```

```
QUERY_TEXT, // Return type: java.lang.String
  DB_USER, // Return type: java.lang.String
  DB_URL, // Return type: java.lang.String
OPERATION_TYPE, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  WF_IDS, // Return type: java.util.Collection
  NAME, // Return type: java.lang.String
ORIGINAL_NAME, // Return type: java.lang.String
USER_ENTITY, // Return type: boolean
  SOURCE_ID, // Return type: java.lang.String EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.sqoop.model.SqoopTableExportOperationProperties
public enum SqoopTableExportOperationProperties implements PropertyEnum {
  DB_TABLE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  DB_USER, // Return type: java.lang.String
DB_URL, // Return type: java.lang.String
  OPERATION_TYPE, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  WF_IDS, // Return type: java.util.Collection
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
com.cloudera.nav.sqoop.model.SqoopTableImportOperationProperties
public enum SqoopTableImportOperationProperties implements PropertyEnum {
  DB_TABLE, // Return type: java.lang.String DB_WHERE, // Return type: java.lang.String
  SOURCE_TYPE, // Return type: java.lang.String
  DB_USER, // Return type: java.lang.String DB_URL, // Return type: java.lang.String
  OPERATION_TYPE, // Return type: java.lang.String
  TYPE, // Return type: java.lang.String
  WF_IDS, // Return type: java.util.Collection
  NAME, // Return type: java.lang.String
  ORIGINAL_NAME, // Return type: java.lang.String USER_ENTITY, // Return type: boolean SOURCE_ID, // Return type: java.lang.String
  EXTRACTOR_RUN_ID, // Return type: java.lang.String
  PARENT_PATH; // Return type: java.lang.String
```

# Lineage Diagrams

Required Role: Lineage Viewer Metadata Administrator Full Administrator

A *lineage diagram* is a directed graph that depicts an entity and its relations with other entities. A lineage diagram is limited to 1000 entities and 2000 relations.

There are two types of lineage diagrams:

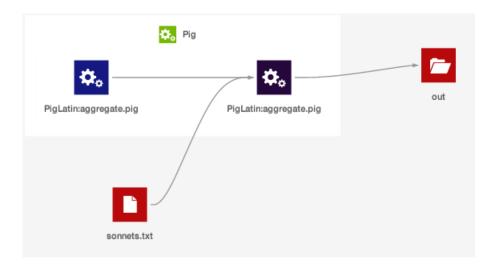
- **Template** represents an entity that is a model for other entities
- Instance represents an instance or execution of a template

#### Entities

In a lineage diagram, entity types are represented by icons:

HDFS		Oozie	Oozie	
<ul><li>File</li><li>Directory</li></ul>		<ul><li>Job template</li><li>Job execution</li></ul>	• 💠	
Hive		Pig	Pig	
<ul><li>Table</li><li>Query template</li><li>Query execution</li></ul>	· 📾	<ul><li>Table</li><li>Pig script</li><li>Pig script execution</li></ul>	· □ · . · . · . · . · . · . · . · . · .	
MapReduce		SQOOP	SQOOP	
<ul><li>Job template</li><li>Job execution</li></ul>	• 🐯	<ul><li>Job template</li><li>Job execution</li></ul>	<b>⇔</b> .	
		YARN	YARN	
		<ul><li>Job template</li><li>Job execution</li></ul>	• 0.	

Parent entities are represented by a white box enclosing other entities. The following lineage diagram illustrates the relations between the YARN job aggregate.pig and Pig script aggregate.pig invoked by the parent Pig Oozie workflow and its source file and destination folder:



# Cloudera Navigator Metadata Component

Note: In the following circumstances the entity type icon will appear as

will eventually be replaced with the correct entity Entities are not yet extracted. In this case icon after the entity is extracted and linked in Navigator. For information on how long it takes for newly created entities to be extracted, see Metadata Extraction on page 20.

Hive entities have been deleted from the system before they could be extracted by Navigator.

### Relations

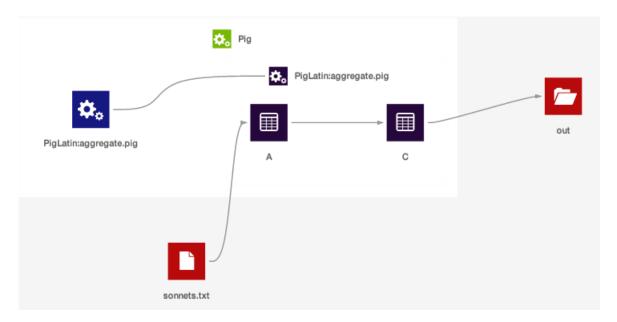
Relations between the entities are represented graphically by gray lines, with arrows indicating the direction of the data flow. There are the following types of relations:

Relation Type	Description	
Data flow	Describes a relation between data and a processing activity. For example, between a file and a MapReduce job or vice versa.	
Alias	Describes an alias relation. For example, from a table to a synonym.	
Parent-child	Describes a parent-child relation. For example, between a directory and a file.	
Logical-physical	Describes the relation between a logical entity and its physical entity. For example, between a Hive query and a MapReduce job.	
Conjoint	Describes a non-directional relation. For example, between an table and an index.	
Instance of	Describes the relation between a template and its instance. For example, an operation execution is an instance of operation.	
Control flow	Describes a relation where the source entity controls the data flow of the target entity. For example, between the columns used in an insert clause and the where clause of a Hive query.	

For lines connecting database columns, a dashed line indicates that the column is in the where clause; a solid line indicates that the column is in the select clause.

### Manipulating Lineage Diagrams

You can click a parent entity to display its child entities. For example, you can click the Pig script to display its child tables:



- To improve the layout of a lineage diagram you can drag and drop entities (in this case sonnets.txt and out) located outside a parent box.
- You can use the mouse scroll wheel to zoom the lineage diagram in and out.
- You can move the lineage diagram in the lineage pane by pressing the mouse button and dragging it.

## Displaying a Template Lineage Diagram

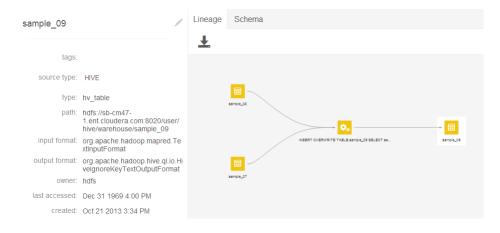
A template lineage diagram contains template entities, such as jobs and queries, that can be instantiated and the input and output entities to which they are related.

To display a template lineage diagram:

- 1. Perform a metadata search.
- 2. In the list of results, click a result entry. For example, when you click the sample\_09 result entry:

```
HIVE sample 09
                                                                                                               last accessed: Dec 31 19
type: hv_table
                 path: hdfs://sb-cm47-1.ent.cloudera.com:8020/user/hive/warehouse/sample_09
                                                                                                owner: hdfs
```

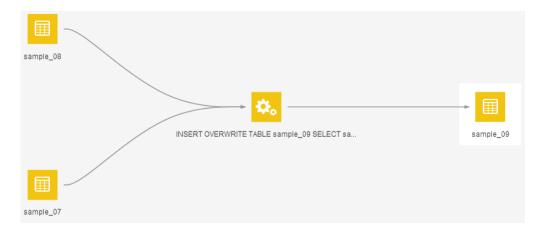
the Search screen is replaced with a page that displays the entity property sheet on the left and lineage diagram on the right:



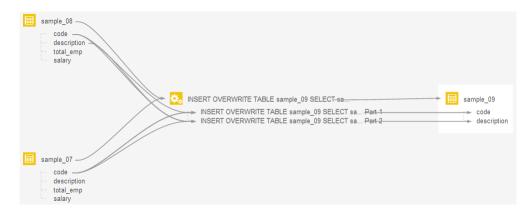
# Cloudera Navigator Metadata Component

The selected entity sample\_09 appears with a white box as a background.

This example lineage diagram illustrates the relations between a Hive query execution entity and its source and destination tables:



When you click each entity icon, columns and lines connecting the source and destination columns display:



If you hover over a part, the source and destination columns are highlighted:



# Displaying an Instance Lineage Diagram

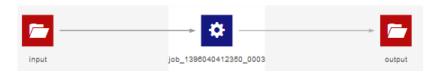
An instance lineage diagram displays instance entities, such as job and query executions, and the input and output entities to which they are related.

To display an instance lineage diagram:

1. Display a template lineage diagram. For example:



- 2. Click the Instances tab, which contains a list of links to instances of the template.
- 3. Click a link to display an instance lineage diagram. The job instance job\_1396040412350\_0003 replaces the wordcount job template.



# Displaying the Template Lineage Diagram for an Instance Lineage Diagram

You can navigate from an instance diagram to its template.

- 1. Display a instance lineage diagram.
- **2.** Click the value of the **template** property to navigate to the instance's template.

# Downloading a Lineage File

Lineage is externalized in a lineage file in JSON format.

- 1. Display a template or instance lineage diagram.
- Click the **icon** at the top left of the diagram.

A lineage file named lineage. json is downloaded. For example, the lineage file representing job\_1396040412350\_0003 from the preceding section is:

```
"entities":
 "d212538318276a6ad8abdd308a4487cc": {
   "level": 1,
   "physical": [],
   "logical": [],
   "aliasOf": [],
   "aliases": [],
   "instances": [],
"children": [],
    "workflows": [],
   "identity": "d212538318276a6ad8abdd308a4487cc",
    "originalName": "file0",
    "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
   "firstClassParentId": null,
    "parentPath": "/user/hdfs/input",
    "extractorRunId": null,
   "name": "file0",
    "description": null,
    "tags": null,
   "fileSystemPath": "/user/hdfs/input/file0",
   "type": "FILE",
    "size": 22,
    "created": "2014-07-08T22:03:32.126Z",
   "lastModified": "2014-07-08T22:03:32.126Z",
    "lastAccessed": "2014-07-08T22:03:31.612Z",
    "permissions": 420,
    "owner": "hdfs",
```

```
"group": "supergroup",
  "blockSize": null,
  "mimeType": "application/octet-stream",
  "deleted": false,
  "sourceType": "HDFS",
  "replication": null,
  "internalType": "fselement",
  "nameField": "originalName",
  "sourceName": "HDFS",
  "isScript": false,
  "hasDownstream": true,
 "parent": "e461de8de38511a3ac6740dd7d51b8d0",
  "activeChildren": []
"57c993fc305b553dae82210090d5da7a": {
  "level": 1,
  "physical":
 "logical": [],
 "aliasOf": [],
  "aliases": [],
  "instances": [
    "fc671272b2c7518ec738d6b17f08b7ed"
 "children": []
 "workflows": [],
  "identity": "57c993fc305b553dae82210090d5da7a",
  "originalName": "wordcount",
  "sourceId": "262c2128425eabeb29bb1c96ee0eb35f",
 "firstClassParentId": null,
  "parentPath": null,
  "extractorRunId": null,
  "name": "wordcount",
  "description": null,
  "tags": null,
  "wfIds": null,
  "inputFormat": "org.apache.hadoop.mapred.TextInputFormat"
  "outputFormat": "org.apache.hadoop.mapred.TextOutputFormat",
  "outputKey": "org.apache.hadoop.io.Text"
  "outputValue": "org.apache.hadoop.io.IntWritable",
  "mapper": "org.myorg.WordCount$Map"
  "reducer": "org.myorg.WordCount$Reduce",
  "sourceType": "YARN",
 "type": "OPERATION"
  "internalType": "mrjobspec",
 "nameField": "name",
"sourceName": "YARN (MR2 Included)",
  "isScript": false,
  "unorchestrated": true,
 "active": true,
  "column": 0,
 "renderOrdinal": 0,
 "activeChildren": [],
 "x": 0,
  "y": -52.3046875
},
"2b5a90798459c168d9fb5ecc2fc46aa5": {
  "level": 1,
  "physical": [],
 "logical": [],
  "aliasOf": [],
  "aliases": [],
  "instances": [],
 "children": [],
  "workflows": [],
  "identity": "2b5a90798459c168d9fb5ecc2fc46aa5",
  "originalName": "part-00001"
  "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
  "firstClassParentId": null,
  "parentPath": "/user/hdfs/output",
  "extractorRunId": null,
  "name": "part-00001",
  "description": null,
  "tags": null,
```

```
"fileSystemPath": "/user/hdfs/output/part-00001",
 "type": "FILE",
  "size": 6,
  "created": "2014-07-08T22:08:53.602Z"
 "lastModified": "2014-07-08T22:08:53.602Z",
  "lastAccessed": "2014-07-08T22:08:52.923Z",
  "permissions": 420,
 "owner": "hdfs",
 "group": "supergroup",
 "blockSize": null,
"mimeType": "application/octet-stream",
 "deleted": false,
 "sourceType": "HDFS" "replication": null,
 "internalType": "fselement",
 "nameField": "originalName",
  "sourceName": "HDFS",
  "isScript": false
 "parent": "e9b3d1f7aee29134338df3de4cae1100",
  "hasUpstream": true,
  "activeChildren": []
"fe5445fbd070d97c418d96200a218cae": {
 "level": 1,
  "physical": [],
 "logical": [],
 "aliasOf": [],
  "aliases": []
 "instances": [],
  "children": [],
  "workflows": [],
  "identity": "fe5445fbd070d97c418d96200a218cae",
  "originalName": "part-00002"
  "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs/output",
  "extractorRunId": null,
  "name": "part-00002",
 "description": null,
  "tags": null,
 "fileSystemPath": "/user/hdfs/output/part-00002",
 "type": "FILE",
 "size": 9,
  "created": "2014-07-08T22:08:53.619Z"
  "lastModified": "2014-07-08T22:08:53.619Z",
 "lastAccessed": "2014-07-08T22:08:52.464Z",
  "permissions": 420,
  "owner": "hdfs"
 "group": "supergroup",
 "blockSize": null,
"mimeType": "application/octet-stream",
  "deleted": false,
 "sourceType": "HDFS"
  "replication": null,
  "internalType": "fselement",
 "nameField": "originalName",
  "sourceName": "HDFS",
  "isScript": false
  "parent": "e9b3d1f7aee29134338df3de4cae1100",
  "hasUpstream": true,
  "activeChildren": []
"916b141fb9ce45094df8ef97ecdde41c": {
 "level": 1,
  "physical": [],
 "logical": [],
 "aliasOf": [],
"aliases": [],
  "instances": [],
 "children": [],
  "workflows": [],
  "identity": "916b141fb9ce45094df8ef97ecdde41c",
  "originalName": "file1._COPYING_",
```

```
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
  "firstClassParentId": null,
  "parentPath": "/user/hdfs/input",
  "extractorRunId": null,
 "name": "file1._COPYING_",
  "description": null,
 "tags": null,
 "fileSystemPath": "/user/hdfs/input/file1._COPYING_",
 "type": "FILE",
  "size": 28,
  "created": "2014-07-08T22:03:32.171Z"
 "lastModified": "2014-07-08T22:03:32.244Z",
  "lastAccessed": "2014-07-08T22:03:32.171Z",
  "permissions": 420,
 "owner": "hdfs",
 "group": "supergroup",
 "blockSize": null,
"mimeType": "application/octet-stream",
 "deleted": true,
 "sourceType": "HDFS"
"replication": null,
 "internalType": "fselement",
  "nameField": "originalName",
  "sourceName": "HDFS",
 "isScript": false,
 "parent": "e461de8de38511a3ac6740dd7d51b8d0",
  "activeChildren": []
"66bff6eeac3c17c5b7eb7c9035704eef": {
  "level": 1,
  "physical": [],
 "logical": [],
 "aliasOf": [],
  "aliases": [],
  "instances": [],
 "children": [],
  "workflows": [],
  "identity": "66bff6eeac3c17c5b7eb7c9035704eef",
 "originalName": "part-00005"
  "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs/output",
 "extractorRunId": null,
  "name": "part-00005",
  "description": null,
 "tags": null,
  "fileSystemPath": "/user/hdfs/output/part-00005",
  "type": "FILE",
 "size": 8,
  "created": "2014-07-08T22:08:53.668Z"
  "lastModified": "2014-07-08T22:08:53.668Z",
  "lastAccessed": "2014-07-08T22:08:52.419Z",
 "permissions": 420,
  "owner": "hdfs",
  "group": "supergroup",
 "blockSize": null,
  "mimeType": "application/octet-stream",
  "deleted": false,
  "sourceType": "HDFS",
  "replication": null,
  "internalType": "fselement",
  "nameField": "originalName",
 "sourceName": "HDFS",
  "isScript": false,
 "hasUpstream": true,
 "parent": "e9b3d1f7aee29134338df3de4cae1100",
  "activeChildren": []
"51db076ee11470b7f968bd5f33429e6b": {
 "level": 1,
  "physical": [],
  "logical": [],
 "aliasOf": [],
```

```
"aliases": [],
 "instances": [],
 "children": [],
 "workflows": []
 "identity": "51db076ee11470b7f968bd5f33429e6b",
 "originalName": "file0._COPYING_"
 "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs/input",
 "extractorRunId": null,
 "name": "file0._COPYING_"
 "description": null,
 "tags": null,
 "fileSystemPath": "/user/hdfs/input/file0._COPYING_",
 "type": "FILE",
 "size": 22,
"created": "2014-07-08T22:03:31.612Z"
 "lastModified": "2014-07-08T22:03:32.100Z",
 "lastAccessed": "2014-07-08T22:03:31.612Z",
 "permissions": 420,
 "owner": "hdfs",
 "group": "supergroup",
 "blockSize": null,
"mimeType": "application/octet-stream",
 "deleted": true,
 "sourceType": "HDFS",
 "replication": null,
 "internalType": "fselement",
 "nameField": "originalName",
 "sourceName": "HDFS",
 "isScript": false,
 "parent": "e461de8de38511a3ac6740dd7d51b8d0",
 "activeChildren": []
"c864e9c3bc3f5f7a99315c56830e811d": {
 "level": 1,
 "physical": [],
 "logical": [],
 "aliasOf": [],
 "aliases": [],
 "instances": [],
 "children": [],
 "workflows": [],
 "identity": "c864e9c3bc3f5f7a99315c56830e811d",
 "originalName": "_SUCCESS"
 "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs/output",
 "extractorRunId": null,
 "name": "_SUCCESS"
 "description": null,
 "tags": null,
 "fileSystemPath": "/user/hdfs/output/_SUCCESS",
 "type": "FILE",
 "size": 0,
 "created": "2014-07-08T22:08:53.689Z",
 "permissions": 420,
 "owner": "hdfs",
"group": "supergroup",
 "blockSize": null,
 "mimeType": "application/octet-stream",
 "deleted": false,
 "sourceType": "HDFS",
 "replication": null,
 "internalType": "fselement",
 "nameField": "originalName",
 "sourceName": "HDFS",
 "isScript": false,
 "parent": "e9b3d1f7aee29134338df3de4cae1100",
 "hasUpstream": true
 "activeChildren": []
```

```
"9d7ca9e46a9e4624df0d5d10949a5fc6": {
 "level": 1,
 "physical":
 "logical": [],
 "aliasOf": [],
"aliases": [],
 "instances": [],
 "children": [],
 "workflows": [],
 "identity": "9d7ca9e46a9e4624df0d5d10949a5fc6",
 "originalName": "part-00004"
 "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs/output",
 "extractorRunId": null,
 "name": "part-00004",
 "description": null,
 "tags": null,
 "fileSystemPath": "/user/hdfs/output/part-00004",
 "type": "FILE",
 "size": 10,
"created": "2014-07-08T22:08:53.651Z"
 "lastModified": "2014-07-08T22:08:53.651Z",
 "lastAccessed": "2014-07-08T22:08:52.856Z",
 "permissions": 420,
 "owner": "hdfs",
 "group": "supergroup",
 "blockSize": null,
 "mimeType": "application/octet-stream",
 "deleted": false,
 "sourceType": "HDFS",
 "replication": null,
 "internalType": "fselement",
 "nameField": "originalName",
 "sourceName": "HDFS",
 "isScript": false,
 "hasUpstream": true,
 "parent": "e9b3d1f7aee29134338df3de4cae1100",
 "activeChildren": []
"8c40fbf6b4584a0b213a257be2cae679": {
 "level": 1,
 "physical":
 "logical": [],
 "aliasOf": [],
 "aliases": [],
 "instances": [],
 "children": [],
 "workflows": [],
 "identity": "8c40fbf6b4584a0b213a257be2cae679",
 "originalName": "file1"
 "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs/input",
 "extractorRunId": null,
 "name": "file1"
 "description": null,
 "tags": null,
 "fileSystemPath": "/user/hdfs/input/file1",
 "type": "FILE",
 "size": 28,
"created": "2014-07-08T22:03:32.262Z",
 "lastModified": "2014-07-08T22:03:32.262Z", "lastAccessed": "2014-07-08T22:03:32.171Z",
 "permissions": 420,
 "owner": "hdfs",
"group": "supergroup",
 "blockSize": null,
 "mimeType": "application/octet-stream",
 "deleted": false,
  "sourceType": "HDFS",
 "replication": null,
```

```
"internalType": "fselement",
 "nameField": "originalName",
 "sourceName": "HDFS",
 "isScript": false,
 "hasDownstream": true,
 "parent": "e461de8de38511a3ac6740dd7d51b8d0",
 "activeChildren": []
"2690183198a2d6568188426ea9a82793": {
 "level": 1,
 "physical":
 "logical": [],
 "aliasOf": [],
"aliases": [],
 "instances": [],
 "children": [],
 "workflows": [],
 "identity": "2690183198a2d6568188426ea9a82793",
 "originalName": "part-00000"
 "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs/output",
 "extractorRunId": null,
 "name": "part-00000",
 "description": null,
 "tags": null,
 "fileSystemPath": "/user/hdfs/output/part-00000",
 "type": "FILE",
 "size": 0,
 "created": "2014-07-08T22:08:53.580Z"
 "lastModified": "2014-07-08T22:08:53.580Z",
 "lastAccessed": "2014-07-08T22:08:52.568Z",
 "permissions": 420,
 "owner": "hdfs",
 "group": "supergroup",
 "blockSize": null,
 "mimeType": "application/octet-stream",
 "deleted": false,
 "sourceType": "HDFS",
 "replication": null,
 "internalType": "fselement",
 "nameField": "originalName",
 "sourceName": "HDFS",
 "isScript": false,
 "hasUpstream": true,
 "parent": "e9b3d1f7aee29134338df3de4cae1100",
 "activeChildren": []
"e9b3d1f7aee29134338df3de4cae1100": {
 "level": 1,
 "physical":
 "logical": [],
 "aliasOf": [],
 "aliases": [],
 "instances": [],
 "children": [
    "fe5445fbd070d97c418d96200a218cae",
    "c864e9c3bc3f5f7a99315c56830e811d",
   "2b5a90798459c168d9fb5ecc2fc46aa5",
   "4374529a34441bc8db30ee001bbb3563",
   "2690183198a2d6568188426ea9a82793"
   "66bff6eeac3c17c5b7eb7c9035704eef",
   "9d7ca9e46a9e4624df0d5d10949a5fc6",
   "e5586cdbc89ecef7552e125acb91c2cf"
 "workflows": [],
 "identity": "e9b3d1f7aee29134338df3de4cae1100",
 "originalName": "output"
 "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
 "firstClassParentId": null,
 "parentPath": "/user/hdfs",
 "extractorRunId": null,
 "name": "output",
```

```
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/output",
"type": "DIRECTORY",
"size": null,
"created": "2014-07-08T22:08:31.569Z"
"lastModified": "2014-07-08T22:08:53.689Z",
"lastAccessed": null,
"permissions": 493,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": null,
"deleted": false,
"sourceType": "HDFS",
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"hasUpstream": true,
"column": 1,
"renderOrdinal": 2,
"activeChildren": [
    "level": 1,
    "physical": [],
    "logical": [],
    "aliasOf": [],
    "aliases": [],
    "instances": [],
    "children": [],
    "workflows": [],
"identity": "c864e9c3bc3f5f7a99315c56830e811d",
    "originalName": "_SUCCESS"
    "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
    "firstClassParentId": null,
"parentPath": "/user/hdfs/output",
    "extractorRunId": null,
    "name": "_SUCCESS"
    "description": null,
    "tags": null,
    "fileSystemPath": "/user/hdfs/output/_SUCCESS",
    "type": "FILE",
    "size": 0,
    "created": "2014-07-08T22:08:53.689Z",
    "lastModified": "2014-07-08T22:08:53.694Z",
    "lastAccessed": "2014-07-08T22:08:53.689Z",
    "permissions": 420,
    "owner": "hdfs",
"group": "supergroup",
    "blockSize": null,
    "mimeType": "application/octet-stream",
    "deleted": false,
    "sourceType": "HDFS"
    "replication": null,
    "internalType": "fselement",
"nameField": "originalName",
    "sourceName": "HDFS",
    "isScript": false,
    "parent": "e9b3d1f7aee29134338df3de4cae1100",
    "hasUpstream": true,
    "activeChildren": []
    "level": 1,
    "physical": [],
    "logical": [],
    "aliasOf": [],
    "aliases": [],
    "instances": [],
"children": [],
    "workflows": [],
```

```
"identity": "2690183198a2d6568188426ea9a82793",
"originalName": "part-00000",
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs/output",
"extractorRunId": null,
"name": "part-00000",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/output/part-00000",
"type": "FILE",
"size": 0,
"created": "2014-07-08T22:08:53.580Z",
"lastModified": "2014-07-08T22:08:53.580Z",
"lastAccessed": "2014-07-08T22:08:52.568Z",
"permissions": 420,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": "application/octet-stream",
"deleted": false,
"sourceType": "HDFS",
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"hasUpstream": true,
"parent": "e9b3d1f7aee29134338df3de4cae1100",
"activeChildren": []
"level": 1,
"physical": [],
"logical": [],
"aliasOf": [],
"aliases": [],
"instances": []
"children": [],
"workflows": [],
"identity": "2b5a90798459c168d9fb5ecc2fc46aa5",
"originalName": "part-00001",
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs/output",
"extractorRunId": null,
"name": "part-00001",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/output/part-00001",
"type": "FILE",
"size": 6,
"created": "2014-07-08T22:08:53.602Z",
"lastModified": "2014-07-08T22:08:53.602Z"
"lastAccessed": "2014-07-08T22:08:52.923Z",
"permissions": 420,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": "application/octet-stream",
"deleted": false,
"sourceType": "HDFS"
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"parent": "e9b3d1f7aee29134338df3de4cae1100",
"hasUpstream": true,
"activeChildren": []
"level": 1,
```

```
"physical": [],
"logical": [],
"aliasOf": [],
"aliases": [],
"instances": [],
"children": [],
"workflows": [],
"identity": "fe5445fbd070d97c418d96200a218cae",
"originalName": "part-00002"
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs/output",
"extractorRunId": null,
"name": "part-00002",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/output/part-00002",
"type": "FILE",
"size": 9,
"created": "2014-07-08T22:08:53.619Z",
"lastModified": "2014-07-08T22:08:53.619Z",
"lastAccessed": "2014-07-08T22:08:52.464Z",
"permissions": 420,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": "application/octet-stream",
"deleted": false,
"sourceType": "HDFS",
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"parent": "e9b3d1f7aee29134338df3de4cae1100",
"hasUpstream": true,
"activeChildren": []
"level": 1,
"physical": [],
"logical": [],
"aliasOf": [],
"aliases": [],
"instances": [],
"children": [],
"workflows": [],
"identity": "4374529a34441bc8db30ee001bbb3563",
"originalName": "part-00003",
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs/output",
"extractorRunId": null,
"name": "part-00003",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/output/part-00003",
"type": "FILE",
"size": 8,
"created": "2014-07-08T22:08:53.636Z",
"lastModified": "2014-07-08T22:08:53.636Z"
"lastAccessed": "2014-07-08T22:08:52.499Z",
"permissions": 420,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": "application/octet-stream",
"deleted": false,
"sourceType": "HDFS"
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
```

```
"isScript": false,
"parent": "e9b3d1f7aee29134338df3de4cae1100",
"hasUpstream": true,
"activeChildren": []
"level": 1,
"physical": [],
"logical": [],
"aliasOf": [],
"aliases": [],
"instances": [],
"children": [],
"workflows": []
"identity": "9d7ca9e46a9e4624df0d5d10949a5fc6",
"originalName": "part-00004"
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs/output",
"extractorRunId": null,
"name": "part-00004",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/output/part-00004",
"type": "FILE",
"size": 10,
"created": "2014-07-08T22:08:53.651Z",
"lastModified": "2014-07-08T22:08:53.651Z",
"lastAccessed": "2014-07-08T22:08:52.856Z",
"permissions": 420,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": "application/octet-stream",
"deleted": false,
"sourceType": "HDFS",
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"hasUpstream": true,
"parent": "e9b3d1f7aee29134338df3de4cae1100",
"activeChildren": []
"level": 1,
"physical": [],
"logical": [],
"aliasOf": [],
"aliases": [],
"instances": [],
"children": [],
"workflows": [],
"identity": "66bff6eeac3c17c5b7eb7c9035704eef",
"originalName": "part-00005",
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs/output",
"extractorRunId": null,
"name": "part-00005",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/output/part-00005",
"type": "FILE",
"size": 8,
"created": "2014-07-08T22:08:53.668Z",
"lastModified": "2014-07-08T22:08:53.668Z",
"lastAccessed": "2014-07-08T22:08:52.419Z",
"permissions": 420,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
```

```
"mimeType": "application/octet-stream",
      "deleted": false,
      "sourceType": "HDFS", "replication": null,
      "internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
      "isScript": false,
      "hasUpstream": true,
      "parent": "e9b3d1f7aee29134338df3de4cae1100",
      "activeChildren": []
      "level": 1,
      "physical": [],
      "logical": [],
      "aliasOf": [],
      "aliases": [],
      "instances": [],
      "children": [],
      "workflows": []
      "identity": "e5586cdbc89ecef7552e125acb91c2cf",
      "parent": "e9b3d1f7aee29134338df3de4cae1100",
      "activeChildren": []
  "x": 182,
  "y": -52.3046875
"4374529a34441bc8db30ee001bbb3563": {
  "level": 1,
  "physical": [],
  "logical": [],
  "aliasOf": [],
  "aliases": [],
  "instances": [],
  "children": [],
  "workflows": [],
"identity": "4374529a34441bc8db30ee001bbb3563",
  "originalName": "part-00003"
  "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
  "firstClassParentId": null,
  "parentPath": "/user/hdfs/output",
  "extractorRunId": null,
  "name": "part-00003",
  "description": null,
  "tags": null,
  "fileSystemPath": "/user/hdfs/output/part-00003",
  "type": "FILE",
  "size": 8,
"created": "2014-07-08T22:08:53.636Z"
  "lastModified": "2014-07-08T22:08:53.636Z",
  "lastAccessed": "2014-07-08T22:08:52.499Z",
  "permissions": 420,
  "owner": "hdfs",
  "group": "supergroup",
  "blockSize": null,
"mimeType": "application/octet-stream",
  "deleted": false,
  "sourceType": "HDFS"
  "replication": null,
  "internalType": "fselement",
  "nameField": "originalName",
  "sourceName": "HDFS",
  "isScript": false
  "parent": "e9b3d1f7aee29134338df3de4cae1100",
  "hasUpstream": true,
  "activeChildren": []
"e461de8de38511a3ac6740dd7d51b8d0": {
  "level": 1,
  "physical": [],
  "logical": [],
```

```
"aliasOf": [],
"aliases": [],
"instances": [],
"children": [
  "916b141fb9ce45094df8ef97ecdde41c",
  "8c40fbf6b4584a0b213a257be2cae679",
  "51db076ee11470b7f968bd5f33429e6b",
 "d212538318276a6ad8abdd308a4487cc"
"workflows": []
"identity": "e461de8de38511a3ac6740dd7d51b8d0",
"originalName": "input",
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs",
"extractorRunId": null,
"name": "input",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/input",
"type": "DIRECTORY",
"size": null,
"created": "2014-07-08T22:02:26.664Z"
"lastModified": "2014-07-08T22:03:32.262Z",
"lastAccessed": null,
"permissions": 493,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": null,
"deleted": false,
"sourceType": "HDFS",
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"hasDownstream": true,
"column": -1,
"renderOrdinal": 1
"activeChildren": [
    "level": 1,
    "physical":
    "logical": [],
    "aliasOf": [],
    "aliases": [],
    "instances": [],
    "children": [],
    "workflows": [],
"identity": "d212538318276a6ad8abdd308a4487cc",
    "originalName": "file0",
    "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
    "firstClassParentId": null,
    "parentPath": "/user/hdfs/input",
    "extractorRunId": null,
    "name": "file0",
    "description": null,
    "tags": null,
    "fileSystemPath": "/user/hdfs/input/file0",
    "type": "FILE",
    "size": 22,
"created": "2014-07-08T22:03:32.126Z",
    "lastModified": "2014-07-08T22:03:32.126Z",
    "lastAccessed": "2014-07-08T22:03:31.612Z",
    "permissions": 420,
    "owner": "hdfs",
"group": "supergroup",
    "blockSize": null,
    "mimeType": "application/octet-stream",
    "deleted": false,
    "sourceType": "HDFS",
    "replication": null,
```

```
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"hasDownstream": true,
"parent": "e461de8de38511a3ac6740dd7d51b8d0",
"activeChildren": []
"level": 1,
"physical": [],
"logical": [],
"aliasOf": [],
"aliases": [],
"instances": [],
"children": [],
"workflows": []
"identity": "51db076ee11470b7f968bd5f33429e6b",
"originalName": "file0._COPYING_"
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null
"parentPath": "/user/hdfs/input",
"extractorRunId": null,
"name": "file0._COPYING_",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/input/file0._COPYING_",
"type": "FILE",
"size": 22,
"created": "2014-07-08T22:03:31.612Z"
"lastModified": "2014-07-08T22:03:32.100Z",
"lastAccessed": "2014-07-08T22:03:31.612Z",
"permissions": 420,
"owner": "hdfs",
"group": "supergroup",
"blockSize": null,
"mimeType": "application/octet-stream",
"deleted": true,
"sourceType": "HDFS",
"replication": null,
"internalType": "fselement",
"nameField": "originalName",
"sourceName": "HDFS",
"isScript": false,
"parent": "e461de8de38511a3ac6740dd7d51b8d0",
"activeChildren": []
"level": 1,
"physical": [],
"logical": [],
"aliasOf": [],
"aliases": [],
"instances": [],
"children": [],
"workflows": [],
"identity": "8c40fbf6b4584a0b213a257be2cae679",
"originalName": "file1"
"sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
"firstClassParentId": null,
"parentPath": "/user/hdfs/input",
"extractorRunId": null,
"name": "file1",
"description": null,
"tags": null,
"fileSystemPath": "/user/hdfs/input/file1",
"type": "FILE",
"size": 28,
"created": "2014-07-08T22:03:32.262Z",
"lastModified": "2014-07-08T22:03:32.262Z",
"lastAccessed": "2014-07-08T22:03:32.171Z",
"permissions": 420,
"owner": "hdfs",
```

```
"group": "supergroup",
      "blockSize": null,
      "mimeType": "application/octet-stream",
"deleted": false,
      "sourceType": "HDFS",
      "replication": null,
      "internalType": "fselement",
      "nameField": "originalName",
      "sourceName": "HDFS",
      "isScript": false,
      "hasDownstream": true,
      "parent": "e461de8de38511a3ac6740dd7d51b8d0",
      "activeChildren": []
      "level": 1,
      "physical": [],
      "logical": [],
      "aliasOf": [],
      "aliases": [],
      "instances": [],
      "children": [],
      "workflows": [],
"identity": "916b141fb9ce45094df8ef97ecdde41c",
      "originalName": "file1._COPYING_"
      "sourceId": "012437f9eeb3c23dc69e679ac94a7fa2",
      "firstClassParentId": null
      "parentPath": "/user/hdfs/input",
      "extractorRunId": null,
      "name": "file1._COPYING_",
      "description": null,
      "tags": null,
      "fileSystemPath": "/user/hdfs/input/file1._COPYING_",
      "type": "FILE",
      "size": 28,
"created": "2014-07-08T22:03:32.171Z",
      "lastModified": "2014-07-08T22:03:32.244Z",
      "lastAccessed": "2014-07-08T22:03:32.171Z",
      "permissions": 420,
      "owner": "hdfs",
"group": "supergroup",
      "blockSize": null,
      "mimeType": "application/octet-stream",
      "deleted": true,
      "sourceType": "HDFS",
      "replication": null,
      "internalType": "fselement", "nameField": "originalName",
      "sourceName": "HDFS",
      "isScript": false,
"parent": "e461de8de38511a3ac6740dd7d51b8d0",
      "activeChildren": []
  ],
  "x": -182,
 "y": -52.3046875
"fc671272b2c7518ec738d6b17f08b7ed": {
 "level": 1,
  "physical": [],
 "logical": [],
  "aliasOf": [],
 "aliases": [],
  "instances": [],
  "children": [],
  "workflows": [],
  "identity": "fc671272b2c7518ec738d6b17f08b7ed",
  "originalName": "job_1396040412350_0003",
  "sourceId": "262c2128425eabeb29bb1c96ee0eb35f",
 "firstClassParentId": null,
  "parentPath": null,
  "extractorRunId": null,
  "name": "job_1396040412350_0003",
```

```
"description": null,
   "tags": null,
   "started": "2014-07-08T22:08:31.543Z",
   "ended": "2014-07-08T22:08:53.722Z",
   "principal": "hdfs",
   "inputs": [
     "hdfs://tcdn501-1.ent.cloudera.com:8020/user/hdfs/input"
   "outputs": [
     "hdfs://tcdn501-1.ent.cloudera.com:8020/user/hdfs/output"
   "wfInstId": null,
   "jobID": "job_1396040412350_0003",
   "sourceType": "YARN"
   "inputRecursive": false,
   "type": "OPERATION_EXECUTION",
   "internalType": "mrjobinstance",
   "nameField": "originalName",
   "sourceName": "YARN (MR2 Included)",
   "isScript": false,
   "hasUpstream": true
   "template": "57c993fc305b553dae82210090d5da7a",
   "hasDownstream": true
  e5586cdbc89ecef7552e125acb91c2cf": {
   "level": 1,
   "physical":
   "logical": [],
   "aliasOf": [],
   "aliases": [],
   "instances": [],
   "children": [],
   "workflows": [],
"identity": "e5586cdbc89ecef7552e125acb91c2cf",
   "parent": "e9b3d1f7aee29134338df3de4cae1100",
   "activeChildren": []
"relations": {
  "57bd1a6ddc6ff99ce878f9c024759bbf": {
   "identity": "57bd1a6ddc6ff99ce878f9c024759bbf",
   "type": "PARENT_CHILD",
   "propagatorId": null,
   "extractorRunId": "NOT_STORED",
   "parent":
      "entityId": "e9b3d1f7aee29134338df3de4cae1100"
   "entityIds": [
        "fe5445fbd070d97c418d96200a218cae"
   "unlinked": false,
   "propagatable": false
  "bc51b7d63696a428f0546e4d6088c936": {
   "identity": "bc51b7d63696a428f0546e4d6088c936",
   "type": "PARENT_CHILD",
   "propagatorId": null,
   "extractorRunId": "NOT_STORED",
   "parent":
      "entityId": "e461de8de38511a3ac6740dd7d51b8d0"
    "children":
     "entityIds": [
       "916b141fb9ce45094df8ef97ecdde41c"
   },
"unlinked": false,
"able": fa
   "propagatable": false
  d9568c3415368847afaecd4b88478e3f": {
    "identity": "d9568c3415368847afaecd4b88478e3f",
```

```
"type": "DATA_FLOW",
 "propagatorId": "0e1f4f973761cf05b3781934945b8098",
 "extractorRunId": "NOT STORED",
 "targets": {
   "entityIds": [
      "fc671272b2c7518ec738d6b17f08b7ed"
  "sources": {
   "entityIds": [
     "d212538318276a6ad8abdd308a4487cc",
     "8c40fbf6b4584a0b213a257be2cae679"
   ]
 },
"unlinked": false,
"able": fa
 "propagatable": false
f505a6ea5efe8d0d58f61e2e19e09644": {
 "identity": "f505a6ea5efe8d0d58f61e2e19e09644",
 "type": "PARENT_CHILD",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "parent":
   "entityId": "e9b3d1f7aee29134338df3de4cae1100"
  "children": {
   "entityIds": [
      "c864e9c3bc3f5f7a99315c56830e811d"
   ]
  "unlinked": false,
 "propagatable": false
"da1b64977bca48ed1d4e0f74bbcacea7": {
 "identity": "da1b64977bca48ed1d4e0f74bbcacea7",
 "type": "PARENT_CHILD",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
    "entityId": "e9b3d1f7aee29134338df3de4cae1100"
  "children": {
    "entityIds": [
     "2b5a90798459c168d9fb5ecc2fc46aa5"
   ]
  "unlinked": false,
 "propagatable": false
"74f685545f94873904c8a5ca5b7c9eaa":
 "identity": "74f685545f94873904c8a5ca5b7c9eaa",
 "type": "INSTANCE_OF",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "instances":
   "entityIds": [
      "fc671272b2c7518ec738d6b17f08b7ed"
   ]
  "template": {
   "entityId": "57c993fc305b553dae82210090d5da7a"
 "unlinked": false,
 "propagatable": false
"2e3c88897fcc1df2853fa4490c286207": {
 "identity": "2e3c88897fcc1df2853fa4490c286207",
 "type": "PARENT_CHILD",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "parent":
    "entityId": "e461de8de38511a3ac6740dd7d51b8d0"
```

```
"children": {
   "entityIds": [
      "8c40fbf6b4584a0b213a257be2cae679"
  "unlinked": false,
 "propagatable": false
"c35093d842e090c17b3374ef22fe5383": {
 "identity": "c35093d842e090c17b3374ef22fe5383",
 "type": "PARENT_CHILD"
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "parent":
    "entityId": "e9b3d1f7aee29134338df3de4cae1100"
  "children":
   "entityIds": [
     "4374529a34441bc8db30ee001bbb3563"
 "unlinked": false,
 "propagatable": false
"29caebf3d0a7a8d38d9a4a49bb2c016c": {
 "identity": "29caebf3d0a7a8d38d9a4a49bb2c016c",
 "type": "DATA_FLOW"
 "propagatorId": "cf0c7b96382b834021f51638b135da9e",
 "extractorRunId": "NOT_STORED",
 "targets": {
   "entityIds": [
     "2690183198a2d6568188426ea9a82793",
     "2b5a90798459c168d9fb5ecc2fc46aa5"
     "fe5445fbd070d97c418d96200a218cae"
     "4374529a34441bc8db30ee001bbb3563"
     "9d7ca9e46a9e4624df0d5d10949a5fc6",
     "66bff6eeac3c17c5b7eb7c9035704eef",
      "c864e9c3bc3f5f7a99315c56830e811d"
   ]
  "sources": {
   "entityIds": [
      "fc671272b2c7518ec738d6b17f08b7ed"
 "unlinked": false,
 "propagatable": false
"a2a35d94ba084495469290f8413e91ee": {
 "identity": "a2a35d94ba084495469290f8413e91ee",
 "type": "DATA_FLOW"
 "propagatorId": "cf0c7b96382b834021f51638b135da9e",
 "extractorRunId": "NOT_STORED",
 "targets": {
    "entityIds": [
     "fe5445fbd070d97c418d96200a218cae",
     "2b5a90798459c168d9fb5ecc2fc46aa5"
     "2690183198a2d6568188426ea9a82793"
     "4374529a34441bc8db30ee001bbb3563"
     "66bff6eeac3c17c5b7eb7c9035704eef",
      "9d7ca9e46a9e4624df0d5d10949a5fc6"
   ]
  "sources": {
   "entityIds": [
     "fc671272b2c7518ec738d6b17f08b7ed"
 },
"unlinked": false,
"able": fa
 "propagatable": false
,
"3a817d057f264d972c8c0e8dda4f9c3e": {
 "identity": "3a817d057f264d972c8c0e8dda4f9c3e",
```

```
"type": "PARENT_CHILD",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "parent":
   "entityId": "e9b3d1f7aee29134338df3de4cae1100"
 "children": {
   "entityIds": [
     "2690183198a2d6568188426ea9a82793"
   ]
 "unlinked": false,
 "propagatable": false
"e0dfacd87d398fdc8ffb91f30d675e43": {
 "identity": "e0dfacd87d398fdc8ffb91f30d675e43",
 "type": "PARENT_CHILD",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "parent":
    "entityId": "e9b3d1f7aee29134338df3de4cae1100"
  "children": {
   "entityIds": [
     "66bff6eeac3c17c5b7eb7c9035704eef"
 },
"unlinked": false,
"able": fa
 "propagatable": false
"cf0c7b96382b834021f51638b135da9e": {
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "targets": {
   "entityIds": [
     "e9b3d1f7aee29134338df3de4cae1100"
 "sources": {
   "entityIds": [
     "fc671272b2c7518ec738d6b17f08b7ed"
 "unlinked": false,
 "propagatable": false
"16384c6acdb0c1f4fd81624ec5e00de2": {
 "identity": "16384c6acdb0c1f4fd81624ec5e00de2",
 "type": "PARENT_CHILD",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "parent":
    "entityId": "e9b3d1f7aee29134338df3de4cae1100"
 "children": {
   "entityIds": [
     "9d7ca9e46a9e4624df0d5d10949a5fc6"
   ]
 "propagatable": false
"3745ae37cf1da17592754b9aed5d6f21": {
 "identity": "3745ae37cf1da17592754b9aed5d6f21",
 "type": "PARENT_CHILD",
 "propagatorId": null,
 "extractorRunId": "NOT_STORED",
 "parent":
    "entityId": "e461de8de38511a3ac6740dd7d51b8d0"
 "children": {
```

```
"entityIds": [
      "51db076ee11470b7f968bd5f33429e6b"
   1
  "unlinked": false,
  "propagatable": false
"0432c75f4c29379500dc51dc5887e196": {
  "identity": "0432c75f4c29379500dc5ldc5887e196",
  "type": "PARENT_CHILD",
  "propagatorId": null,
 "extractorRunId": "NOT_STORED",
  "parent":
    "entityId": "e461de8de38511a3ac6740dd7d51b8d0"
  "children": {
   "entityIds": [
      "d212538318276a6ad8abdd308a4487cc"
 "propagatable": false
"989443d7728154fb577d239ae0e0d677": {
  "identity": "989443d7728154fb577d239ae0e0d677",
  "type": "PARENT_CHILD",
  "propagatorId": null,
  "extractorRunId": "NOT_STORED",
 "parent":
    "entityId": "e9b3d1f7aee29134338df3de4cae1100"
  "children": {
    "entityIds": [
     "e5586cdbc89ecef7552e125acb91c2cf"
  "unlinked": false,
  "propagatable": false
3043bc0b3969889eaba53d74da1df08b": {
 "identity": "3043bc0b3969889eaba53d74da1df08b",
  "type": "DATA_FLOW"
  "propagatorId": "cf0c7b96382b834021f51638b135da9e",
  "extractorRunId": "NOT_STORED",
  "targets": {
    "entityIds": [
      "fe5445fbd070d97c418d96200a218cae",
      "2b5a90798459c168d9fb5ecc2fc46aa5",
     "2690183198a2d6568188426ea9a82793",
     "4374529a34441bc8db30ee001bbb3563",
      "66bff6eeac3c17c5b7eb7c9035704eef",
      "9d7ca9e46a9e4624df0d5d10949a5fc6"
   ]
 },
"sources": {
    '+vTds
   "entityIds": [
      "fc671272b2c7518ec738d6b17f08b7ed"
  "unlinked": false,
  "propagatable": false
"0e1f4f973761cf05b3781934945b8098": {
 "identity": "0elf4f973761cf05b3781934945b8098", "type": "DATA_FLOW",
 "propagatorId": null,
  "extractorRunId": "NOT_STORED",
  "targets": {
    "entityIds": [
      "fc671272b2c7518ec738d6b17f08b7ed"
  "sources": {
```

```
"entityIds": [
    "e461de8de38511a3ac6740dd7d51b8d0"
"unlinked": false,
"propagatable": false
```

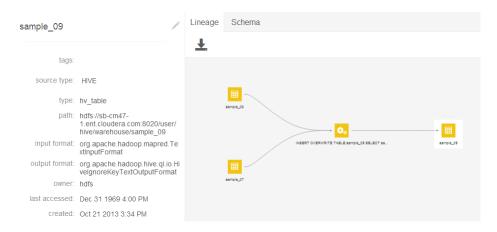
Tracing through the relations shows that job\_1396040412350\_0003, which has the identity fc671272b2c7518ec738d6b17f08b7ed, has the INSTANCE\_OF relation with wordcount and the DATA\_FLOW relation with /user/hdfs/input and /user/hdfs/output.

# **Tables**

## Displaying Table Schema

A table schema contains information about the names and types of the columns of a table.

- 1. Perform a metadata <u>search</u> for an entity of source type Hive.
- 2. In the list of results, click a result entry of type Table. The Search screen is replaced with a page that displays the entity property sheet on the left and lineage diagram on the right.



**3.** Click the **Schema** tab. The table schema displays.



# Displaying Pig Tables

A table contains information about the names and types of the columns of a Pig table.

- 1. Perform a metadata search for an entity of source type Pig.
  - In the list of results, click a result entry of type Table.
  - 1. In the list of results, click a result entry of type Operation\_Execution.

# Cloudera Navigator Metadata Component

2. Click the **Tables** tab. A list of links to tables involved in the operation displays.