Hadoop Documentation

Level – 1 (Topics)

1. The Hadoop Distributed filesystem hadoop the definitive guide, O’ Reilly 41-73
   1. The design of HDFS
   2. HDFS concepts
   3. The command –line interface
   4. Hadoop filesystems
   5. The java interface
   6. Data flow
   7. Parallel copying with distcp
   8. Hadoop Archives
2. Hadoop I/O Hadoop the Definitive Guide, O’Reilly 75-103

2.1 Data Integrity

2.2 Compression

2.3 Serialization

1. Mapreduce Application Hadoop the Definitive Guide,O’Reilly 129-165

3.1 The Configuration API

3.2 Configuring the Development Environment

3.3 writing a unit Test

3.4 Running Locally on Test Data

3.5 Running on a Cluster

3.6 Tuning a job

3.7 MapReduce Workflows

1. Anatorny of a Mapreduce Hadoop the definitive Guide, O’ Reilly 167-186

4.1 Anatomy of a Mapreduce job run

4.2 Failures

4.3 job scheduling

4.4 Shuffle and sort

4.5 Task Execution

Hive L1

SLNO TOPIC REFERENCE PAGENO

1. Introduction to hive hadoop the Definitive Guide, O’Reilly 366-395
   1. The Hive shell
   2. An Example
   3. Running Hive
   4. Comparision with Traditional Databases
   5. Hiveql
   6. Tables

2) Installing & Executing Hive queries programming Hive, O’Reilly 21-38

2.2 What is inside Hive ?

2.3 Starting Hive

2.5 The Hive Command

2.6 The Command-Line interface

2.7 Executing Hive queries from Files

2.9 More on using the hive CU

2.10 Command History

2.11 Shell Execution

2.12 Hadoop dfs commands from inside Hive

2.13 Comments in hive scripts

2.14 Query Column Headers

3) Data Types and file Formats Programming Hiv, O’Reilly 41-48

3.1 Primitive Data Types

3.2 Collection Data types

3.3 Text file Encoding of data values

* 1. Schema on read

4)HiveQL: Data Definition programming Hive, O’Reilly 49-69

4.1 Databases in Hive

4.2 Alter Database

4.3 Creating Tables

4.4 partitioned, Managed Tables

4.5 Dropping Tables

4.6 Alter Table

5) HiveQL Data Manipulation Programming Hive, O’Reilly 71-76

5.1 Loading Data into Managed Tables

5.2 Inserting Data into Tables from Queries

5.3 Creating Tables and Loading them in one Query

5.4 Exporting Data

6) HiveQL Queries programming Hive, O’Reilly 79-112

6.1 SELECT\_FROM Clauses

6.2 WHERE Clauses

6.3 Group BY Clauses

6.4 HAVING Clauses

6.5 JOIN Statements

6.6 ORDER BY and SORT BY

6.7 DISTRIBUTE BY with SORT BY

6.8 CLUSTER BY

6.9 Casting

6.10 Queries that sample data

6.11 UNION ALL

7) HiveQL Views programming Hive, O’REILLY 113-115

7.1 Views to Reduce Query Complexity

7.2 Views that Restrict Data based on Conditions

7.3 Views and map type for Dynamic Tables

7.4 View Odds and Ends