

Hadoop Developer

1. Hadoop Ecosystem & Cluster

Available version Hadoop 1.x & 2

Available Distributions of Hadoop (Cloudera, Hortonworks) Hadoop Projects & Components

Architecture of Hadoop & Planning for cluster
The Hadoop Distributed File System (HDFS) Cluster Daemons & Its Functions

- Name Node
- Secondary Node
- Data Nodes
- Application Master and Task Tracker

YARN Responsibilities

Deployment of Hadoop Cluster

2. Cloudera Sandbox or Quick Start

Installation of cloudera quick start

Difference in sandbox and distributed environment
Overview of apache HUE

3. Map-Reduce, Map-Reduce Steaming (in Java)

All Map-Reduce API Concepts

Architecture of Map-Reduce

Writing Map-Reduce Drivers, Mappers, and Reducers in Java
Speeding Up Hadoop Development by Using Eclipse
Differences between the Old and New Map-Reduce APIs
Writing Mappers and Reducers with the Streaming API
Different question raised for Map-Reduce

4. Hadoop Shell and Commands

Hadoop Developer commands using shell
Map-Reduce job deployment
Oozie workflow design
Different Components Jobs design

5. HBase: The Hadoop Database

Problems with RDBMS
Introduction to HBase
Non-RDBMS, Not-Only SQL or No-SQL
Installation HBase & Deployment
Types CRUD & Batch Operations
Filters, Counters, Pool
Rest Interface & Web-UI

6. Hcatalog or MetaStore Tables

Introduction of apache Hcatalog
Creating tables using Hcatalog
Bulk uploads using MetaStore Tables
Play with semi-structured data
Integration of Hcatalog with Hive
Hive SQL query analysis

7 Hive

Problems with No-SQL Database
Introduction & Installation Hive
Schema and Data Storage Data Types &
Introduction to SQL Hive-SQL: DML &
DDL

Hive-SQL: Views & Indexes

Explain and use the various Hive file formats

Use Hive to run SQL-like queries to perform data analysis

Use Hive to join data sets using a variety of techniques, including Map-side joins and Sort-Merge-Bucket joins

Integration to HBase & Cassandra
Sentiment Analysis and N-Grams
Hive Thrift Service

8 Flume

Installation of Flume

Ingesting Data from External Sources with Flume

Configuration for flume

REST Interfaces

Best Practices for Importing Data

9 Sqoop

Installation of Sqoop

Ingesting Data from External (RDBMS) Sources with Sqoop
Ingesting Data from/to Relational Databases with Sqoop
Integration of Sqoop and Hbase

Integration of Sqoop and Hive

Best Practices for Importing Data

10 Conclusion & FAQs

Note:

- Every Topic has practical session
- Hadoop uses different components which discussed in required sessions

Hue

Cloudera Manager

Zookeeper
Oozie vetc

Prerequisites

This course is best suited to developers and engineers who have some or little bit programming experience. Knowledge of Java is not mandatory Any programming language can be used with Hadoop and is required to complete the hands-on exercises