Hadoop Developer Course Contents

Lesson 1: Introduction to Big Data and Hadoop

1. Introduction to Big Data and Hadoop
2. Traditional IT Analytics Approach
3. Data explosion and the need for Big Data
4. Concept of Big Data
5. Basics of Hadoop
6. History and milestones of Hadoop
7. Market Trends
8. Course Objectives
9. How to use Oracle Virtual Box to open a VM
10. Quiz

Lesson 2: Hadoop Architecture

1. Use of Hadoop in commodity hardware
2. Various configurations and services of Hadoop
3. Apache Hadoop Core Components
4. Difference between a regular and a Hadoop Distributed File System
5. Why HDFS & What is HDFS?
6. HDFS architecture
7. Name Node in HA mode
8. Name Node HA Architecture
9. HDFS Operation Principle
10. Data Block Split and it Benefits
11. Data Replication Topology and Representation
12. HDFS Access and Demo on How to Use
13. Quiz
14. Case Study

Lesson 3: Hadoop Deployment

1. Steps to install Ubuntu Server 14.04 for Hadoop
2. Steps involved in single and multi-node Hadoop installation on Ubuntu server
3. Steps to perform clustering of the Hadoop environment
4. Hadoop Multi-Node Installation Prerequisites
5. Single-Node Cluster vs. Multi-Node Cluster
6. Performing Clustering of the Hadoop Environment Demo
7. Quiz
8. Case Study

- by Chitra Sundararajan

Hadoop Developer Course Contents

Lesson 4: Introduction to YARN and MapReduce

1. Why & What is YARN?
2. YARN architecture
3. Different components of YARN
4. ResourceManager
5. ResourceManager in HA Mode
6. ApplicationMaster
7. NodeManager
8. Container
9. Why & what is MapReduce?
10. Concepts of MapReduce
11. Set up Environment for MapReduce Development
12. Steps to install Hadoop in Ubuntu machine
13. Build MapReduce Program
14. Build a MapReduce Program Demo
15. Steps of Hadoop MapReduce
16. MapReduce Responsibilities
17. MapReduce Java Programming in Eclipse
18. Roles of user and system
19. Quiz
20. Case Study

Lesson 5: Advanced HDFS and MapReduce

1. Advanced HDFS and related concepts
2. HDFS Benchmarking
3. Decommissioning a DataNode
4. HDFS Demo
5. Steps to decommission a DataNode
6. Advanced MapReduce concepts
7. InputFormats in MapReduce
8. OutputFormats in MapReduce
9. Distributed Cache
10. Various joins in MapReduce
    o Reduce Side Join
    o Replicated Join
    o Composite Join
    o Cartesian Product
11. Quiz
12. Case Study

Lesson 6: Pig

1. Concepts of Pig
2. Installation of a Pig engine
3. Prerequisites for the preparation of the environment for Pig Latin
4. How Pig Works

- by Chitra Sundararajan

Hadoop Developer Course Contents

5. Nested Data Model
6. Pig Execution Modes
7. Pig Interactive Modes
8. Pig vs. SQL
9. Additional Libraries for Pig
10. Installing Pig in Ubuntu Server 14.04 LTS Demo
11. Quiz
12. Case Study

-   by Chitra Sundararajan

Hadoop Developer Course Contents

Lesson 7: Hive
1. HiveArchitecture and Components
2. Hive and its importance
3. Hive architecture and its components
   - Metastore
   - Driver
   - Hive Thrift Server
   - Client Components
4. Steps to install and configure Hive
5. Basics of Hive programming
   - Data Model Tables
   - Data Model External Tables
   - Data ModelPartitions
   - Bucketing in Hive
   - Serialization and Deserialization
   - Hive File Formats
6. Programming in Hive
   - User-Defined Function
   - Built-In Functions
   - Other Functions in Hive
   - MapReduce Scripts
7. Quiz
8. Case Study

Lesson 8: HBase
1. HBase architecture
2. HBase data model
3. Steps to install HBase
4. How to insert data and query data from HBase
5. Quiz
6. Case Study

Lesson 9: Commercial Distribution of Hadoop
1. Major commercial distributions of Hadoop
2. Cloudera Quickstart Virtual Machine or VM
3. Hue interface
4. Cloudera Manager interface

Lesson 10: ZooKeeper, Sqoop, and Flume
1. ZooKeeper and its role
2. Challenges faced in distributed processing
3. Install and configure ZooKeeper
4. Concept of Sqoop
5. Configure Sqoop

- by Chitra Sundararajan

Hadoop Developer Course Contents

6. Sample Sqoop Commands
7. Importing Exporting Data from Hadoop Using Sqoop
8. Concept of Flume
9. Configure and run Flume
10. Flume Sample Use Cases
11. Quiz
12. Case Studies for each

Lesson 11: Ecosystem and its Components
1. Hadoop ecosystem structure
2. Different components and their roles in the ecosystem
3. Apache Oozie Workflow
4. Schedule workflow with Apache Oozie Demo
5. Introduction to Mahout
6. Apache Spark Tools
7. Building a program using Apache Spark Demo
8. Quiz
9. Case Study

Lesson 12: Hadoop Administration, Troubleshooting, and Security
1. Command used in Hadoop programming
2. Different configurations of Hadoop cluster
3. Different parameters for performance monitoring and tuning
4. Configuration of security parameters in Hadoop
5. Quiz
6. Case Study

- by Chitra Sundararajan