simplilearn

Big Data and Hadoop Administrator Training – Course Agenda

Session I: Lesson 00—Course Overview

- About Simplilearn's Big Data and Hadoop Administrator course

Lesson 01—Introduction to Big Data and Hadoop

- Introduction to Big Data
- Introduction to Hadoop
- Why Hadoop
- Hadoop & Traditional RDBMS
- Components of Hadoop & Hadoop Architecture
- History and uses of Hadoop

Lesson 02—Planning Hadoop Cluster

- Overview of Hadoop Clusters
- Planning your Hadoop Cluster
- Overview of Hardware and other Network configurations
- Network Topology for Hadoop Clusters
- Overview of Cluster Management

Lesson 03—Hadoop Installation and Configuration

- Overview of various deployment types
- Installing and configuring Hadoop
- Configuring a single node Hadoop Cluster
- Configuring a multi node Hadoop Cluster
- Checking the correctness of Hadoop installation
- Demos:
  - Install Ubuntu Server 12.04
  - Hadoop 1.0 in Ubuntu Server 12.04
  - Create a Clone of Hadoop Virtual Machine
  - Perform Clustering of the Hadoop Environment
  - Install Hadoop 2.0 in Ubuntu Server 12.0

simplilearn

Lesson 04—Advanced Cluster Configuration Features

- Hadoop configuration overview and important configuration file
- Configuration parameters and values
- HDFS parameters MapReduce parameters
- Hadoop environment setup
- 'Include' and 'Exclude' configuration files
- Demo: Configuration Settings of Hadoop
- Lab Exercise

Lesson 05—Hadoop Distributed File System

- Introduction to HDFS
- Overview of HDFS Architecture
- Overview of HDFS Sorage mechanisms
- Overview of HDFS Rack
- Writing and reading files from HDFS
- Understanding the important commands of HDFS
- Introduction to Squoop
- Installing and configuring Sqoop
- Demos:
  - Install Sqoop
  - HDFS Demo
- Lab Exercise

Lesson 06—Overview of MapReduce and YARN

- Introduction to MapReduce
- MapReduce Architecture and working with MapReduce
- Development and Libraries of Map Reduce
- MapReduce components failures and recoveries
- Introduction to YARN
- YARN Architecture
- Installing and configuring YARN
- Working with YARN & YARN Web UI

simplilearn

Lesson 07—Important Hadoop Components

- Understanding Hive
- Installing and configuring Hive
- Understanding Pig
- Installing and configuring Pig
- Understanding Impala
- Installing and configuring Impala
- Demos:
  - Install Hive
  - Install Pig
- Lab Exercises

Lesson 08—Hadoop Administration and Maintenance

- Namenode/Datanode directory structures and files
- File system image and Edit log
- The Checkpoint Procedure
- Namenode failure and recovery procedure
- Safe Mode
- Metadata and Data backup
- Potential problems and solutions / what to look for
- Adding and removing nodes
- Lab Exercise

simplilearn

Lesson 09—Hadoop Ecosystem Components

- Eco system Component: Ganglia
  - Install and Configure Ganglia on a Cluster
  - Configure and Use Ganglia
  - Use Ganglia for Graphs
- Eco system Component: Nagios
  - Nagios Concepts
  - Install and Configure Nagios on Cluster
  - Use Nagios for Sample Alerts And Monitoring
- Eco system Component: Sqoop
  - Install and Configure Sqoop on Cluster
  - Import Data From Oracle/Mysql to Hive
- Overview of Other Eco system Components:
  - Oozie
  - Avro
  - Thrift
  - Rest
  - Mahout
  - Cassandra
  - YARN
  - MR2
- Hadoop Security
- Kerberos and Hadoop
- Why Hadoop Security is Important?
- Hadoop's Security System Concepts
- What Kerberos is and How it Works?
- Configuring Kerberos Security
- Securing a Hadoop Cluster with Kerberos
- Lab Exercise

Question papers

Feedback and Q & A session