



Innovations in Business Solutions

COURSE OUTLINE – Big-Data and Hadoop Developer Certification Training

Course Duration- 40 Hrs

Main Topics:

1. Course Introduction

- ❖ Introduction Preview
- ❖ Objectives Preview
- ❖ Overview
- ❖ Professional Values

2. Lesson 1

- ❖ Introduction to Big Data
- ❖ Introduction to Hadoop
- ❖ Why Hadoop
- ❖ Difference B/w Hadoop and traditional RDBMS
- ❖ Components of Hadoop and its Architecture
- ❖ Evolution of Hadoop

3. Lesson 2- Hadoop Cluster planning

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

4. Lesson 3 - Installation and configuration

- ❖ Installing and configuring Hadoop
- ❖ Configuring a single node Hadoop Cluster
- ❖ Configuring a multi node Hadoop Cluster
- ❖ Checking the correctness of Hadoop installation
- ❖ Demo and Exercise

5. Lesson 4 - Advance configuration of cluster features

- ❖ Hadoop configuration overview and important configuration file
- ❖ Configuration parameters and values

A new beginning in your Career!



Innovations in Business Solutions

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

6. Lesson 5-Hadoop Distributed File System

- ❖ Introduction to HDFS
- ❖ Overview of HDFS Architecture
- ❖ Overview of HDFS Storage mechanisms
- ❖ Overview of HDFS Rack
- ❖ Writing and reading files from HDFS
- ❖ Understanding the important commands of HDFS
- ❖ Introduction to Sqoop
- ❖ Installing and configuring Sqoop
- ❖ Lab Exercise

7. Lesson 6 - 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
- ❖ Exercises

8. Lesson 7 - 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



Innovations in Business Solutions

9. Lesson 8 - Maintenance and Administration

- ❖ 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

10. Lesson 9 -Ecosystem Components

- ❖ Ecosystem Component: Ganglia
 - Install and Configure Ganglia on a Cluster
 - Configure and Use Ganglia
 - Use Ganglia for Graphs
- ❖ Ecosystem Component: Nagios
 - Nagios Concepts
 - Install and Configure Nagios on Cluster
 - Use Nagios for Sample Alerts and Monitoring
- ❖ Ecosystem Component: Sqoop
 - Install and Configure Sqoop on Cluster
 - Import Data from Oracle/MySQL to Hive
- ❖ Overview of Other Ecosystem 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