BIG DATA HADOOP SYLLABUS

Master concepts of Hadoop 2.7 : Hbase, Zookeeper &Sqoop. Write complex Map Reduce programs & perform data analytics using Pig & Hive with real life projects to become a successful hadoop developer.

- ✓ 77% of companies consider Big Data a top priority Peer Research Survey
- ✓ Big Data needs 1.5 mn managers by 2018 McKinsey Global Report
- ✓ Avg salary of Big Data Hadoop Developers is $135k Indeed.com Salary Data

Key Features

- 150 hours of Instructor-Led Training program
- 24 hours of High Quality E-learning material
- 50 hours of industry projects
- Hands-on project execution with Cloud Lab
- Expert Assistant Premium Support
- Earn Hadoop 2.7 experience certificate

What are the course objectives?

After completing the Big Data and Hadoop course, you will:

Master Hadoop 2.7 framework's concepts, along with its deployment in a cluster environment

Learn to write complex Map Reduce programs

Perform Data Analytics using Pig and Hive Hadoop Components

Acquire an in-depth understanding of the Hadoop ecosystem including Flume, ApachieOozie workflow scheduler, and more

Master advanced Hadoop 2.7 concepts including Hbase, Zookeeper, and Sqoop

Get hands-on experience in setting up different Hadoop cluster configurations

Work on real-life, industry-based projects using Hadoop 2.7

Lesson 00 - Course Introduction

0.1 Course Introduction
0.2 Why Big Data
0.3 What is Big Data
0.4 What is Big Data (contd.)
0.5 Facts about Big Data
0.6 Evolution of Big Data
0.7 Case Study Netflix and the House of Cards
0.8 Market Trends
0.9 Course Objectives
0.10 Course Details
0.11 Project Submission and Certification
0.12 On Demand Support
0.13 Key Features
0.14 Conclusion

Lesson 01 - Introduction to Big Data and Hadoop

1.1 Introduction to Big Data and Hadoop

1.2 Objectives
1.3 Data Explosion
1.4 Types of Data
1.5 Need for Big Data
1.6 Big Data and Its Sources
1.7 Characteristics of Big Data
1.8 Characteristics of Big Data Technology
1.9 Knowledge Check
1.10 Leveraging Multiple Data Sources
1.11 Traditional IT Analytics Approach
1.12 Traditional IT Analytics Approach (contd.)
1.13 Big Data Technology Platform for Discovery and Exploration
1.14 Big Data Technology Platform for Discovery and Exploration (contd.)
1.15 Big Data Technology Capabilities
1.16 Big Data Use Cases
1.17 Handling Limitations of Big Data
1.18 Introduction to Hadoop
1.19 History and Milestones of Hadoop
1.20 Organizations Using Hadoop
1.21 VMware Player Introduction
1.22 VMware Player Hardware Requirements
1.23 Oracle Virtual Box to Open a VM
1.24 Installing VM using Oracle Virtual Box Demo
1.25 Opening a VM using Oracle Virtual Box Demo
1.26 Quiz
1.27 Summary
1.28 Conclusion

Lesson 02 - Hadoop Architecture

2.1 Hadoop Architecture
2.2 Objectives
2.3 Key Terms
2.4 Hadoop Cluster Using Commodity Hardware
2.5 Hadoop Configuration
2.6 Hadoop Core Services
2.7 Apache Hadoop Core Components
2.8 Why HDFS
2.9 What is HDFS
2.10 HDFS—Real-life Connect
2.11 Regular File System vs. HDFS
2.12 HDFS Characteristics
2.13 HDFS Key Features
2.14 HDFS Architecture
2.15 Name Node in HA mode
2.16 Name Node HA Architecture
2.17 HDFS Operation Principle

2.18 File System Namespace
2.19 Name Node Operation
2.20 Data Block Split
2.21 Benefits of Data Block Approach
2.22 HDFS Block Replication Architecture
2.23 Replication Method
2.24 Data Replication Topology
2.25 Data Replication Representation
2.26 HDFS Access
2.27 Business Scenario
2.28 Create a new Directory in HDFS Demo
2.29 Spot the Error
2.30 Quiz
2.31 Case Study
2.32 Case Study - Demo
2.33 Summary
2.34 Conclusion

Lesson 03 - Hadoop Deployment

3.1 Hadoop Deployment
3.2 Objectives
3.3 Ubuntu Server Introduction
3.4 Installation of Ubuntu Server 14.04
3.5 Business Scenario
3.6 Installing Ubuntu Server 14.04 Demo 01
3.7 Hadoop Installation Prerequisites
3.8 Hadoop Installation
3.9 Installing Hadoop 2.7 Demo 02
3.10 Hadoop Multi-Node Installation Prerequisites
3.11 Steps for Hadoop Multi-Node Installation
3.12 Single-Node Cluster vs. Multi-Node Cluster
3.13 Creating a Clone of Hadoop VM Demo
3.14 Performing Clustering of the Hadoop Environment Demo 04
3.15 Spot the Error
3.16 Quiz
3.17 Case Study
3.18 Case Study - Demo
3.19 Summary
3.20 Conclusion

Lesson 04 - Introduction to MapReduce

4.1 Introduction to YARN and MapReduce
4.2 Objectives
4.3 Why YARN
4.4 What is YARN
4.5 YARN Real Life Connect

4.6 YARN Infrastructure
4.7 YARN Infrastructure (contd.)
4.8 Resource Manager
4.9 Other Resource Manager Components
4.10 Resource Manager in HA Mode
4.11 Application Master
4.12 Node Manager
4.13 Container
4.14 Applications Running on YARN
4.15 Application Startup in YARN
4.16 Application Startup in YARN (contd.)
4.17 Role of AppMaster in Application Startup
4.18 Why MapReduce
4.19 What is MapReduce
4.20 MapReduce Real-life Connect
4.21 MapReduce Analogy
4.22 MapReduce Analogy (contd.)
4.23 MapReduce Example
4.24 Map Execution
4.25 Map Execution Distributed Two Node Environment 00:38
4.26 MapReduce Essentials
4.27 MapReduce Jobs
4.28 MapReduce and Associated Tasks
4.29 Hadoop Job Work Interaction
4.30 Characteristics of MapReduce
4.31 Real-time Uses of MapReduce
4.32 Prerequisites for Hadoop Installation in Ubuntu Desktop 14.04
4.33 Steps to Install Hadoop
4.34 Business Scenario
4.35 Set up Environment for MapReduce Development
4.36 Small Data and Big Data
4.37 Uploading Small Data and Big Data
4.38 Installing Ubuntu Desktop OS Demo 1
4.39 Build MapReduce Program
4.40 Build a MapReduce Program Demo 2
4.41 HadoopMapReduce Requirements
4.42 Steps of HadoopMapReduce
4.43 MapReduce Responsibilities
4.44 MapReduce Java Programming in Eclipse
4.45 Create a New Project 00:46
4.46 Checking Hadoop Environment for MapReduce
4.47 Build a Map Reduce Application using Eclipse and Run in Hadoop Cl Demo 3

4.48 Map Reduce v 2.7
4.49 Spot the Error
4.50 Quiz
4.51 Case Study
4.52 Case Study - Demo
4.53 Summary
4.54 Conclusion

Lesson 05 - Advanced HDFS and Map Reduce

5.1 Advanced HDFS and Map Reduce
5.2 Objectives
5.3 Advanced HDFS Introduction
5.4 HDFS Benchmarking
5.5 Setting Up HDFS Block Size
5.6 Decommissioning a DataNode
5.7 Business Scenario
5.8 HDFS Demo 01
5.9 Setting HDFS block size in Hadoop 2.7.1 Demo 02
5.10 Advanced MapReduce
5.11 Interfaces
5.12 Data Types in Hadoop
5.13 Data Types in Hadoop (contd.)
5.14 Input Formats in MapReduce
5.15 Output Formats in Map Reduce
5.16 Distributed Cache
5.17 Using Distributed Cache–Step 1
5.18 Using Distributed Cache–Step 2
5.19 Using Distributed Cache–Step 3
5.20 Joins in MapReduce
5.21 Reduce Side Join
5.22 Reduce Side Join (contd.)
5.23 Replicated Join
5.24 Replicated Join (contd.)
5.25 Composite Join
5.26 Composite Join (contd.)
5.27 Cartesian Product
5.28 Cartesian Product (contd.)
5.29 MapReduce program for Writable classes Demo 03
5.30 Spot the Error
5.31 Quiz
5.32 Case Study
5.33 Case Study - Demo
5.34 Summary

5.35 Conclusion

Lesson 06 - Pig

6.1 Pig
6.2 Objectives
6.3 Why Pig
6.4 What is Pig
6.5 Pig Real-life Connect
6.6 Components of Pig
6.7 How Pig Works
6.8 Data Model
6.9 Nested Data Model
6.10 Pig Execution Modes
6.11 Pig Interactive Modes
6.12 Salient Features
6.13 Pig vs. SQL
6.14 Pig vs. SQL Example
6.15 Additional Libraries for Pig
6.16 Installing Pig Engine
6.17 Steps to Installing Pig Engine
6.18 Business Scenario
6.19 Installing Pig in Ubuntu Server 14.04 LTS Demo 01
6.20 Steps to Run a Sample Program to Test Pig
6.21 Getting Datasets for Pig Development
6.22 Prerequisites to Set the Environment for Pig Latin
6.23 Loading and Storing Methods
6.24 Script Interpretation
6.25 Various Relations
6.26 Various Pig Command
6.27 Convert Unstructured Data into Equivalent Words Demo 02
6.28 Loading Files into Relations Demo 03
6.29 Finding the Number of Occurrences of a particular Word Demo 04
6.30 Performing Combining, Splitting, and Joining relations Demo 05
6.31 Performing Transforming and Shaping Relations Demo 06
6.32 Spot the Error
6.33 Quiz6.34 Case Study
6.35 Case Study - Demo
6.36 Summary
6.37 Conclusion

Lesson 07 - Hive

7.1 Hive
7.2 Objectives
7.3 Why Hive

7.4 What is Hive
7.5 Hive Characteristics
7.6 Hive Architecture and Components
7.7 Metastore
7.8 Driver
7.9 Hive Thrift Server
7.10 Client Components
7.11 Basics of Hive Query Language
7.12 Data Model—Tables
7.13 Data Model External Tables
7.14 Data Types in Hive
7.15 Data Model Partitions
7.16 Bucketing in Hive
7.17 Serialization and Deserialization
7.18 Hive File Formats
7.19 Hive Query Language
7.20 Running Hive
7.21 Programming in Hive
7.22 Hive Query Language Extensibility
7.23 User-Defined Function
7.24 Built-In Functions
7.25 Other Functions in Hive
7.26 MapReduce Scripts
7.27 UDF/ UDAF vs. MapReduce Scripts
7.28 New Features supported in Hive
7.29 Business Scenario
7.30 Installing Hive in Ubuntu Server 14.04 LTS Demo 01
7.31 Advanced Data Analytics Demo 02
7.32 Determining Word Count Demo 03
7.33 Partitioning with Hive Demo 04
7.34 Spot the Error
7.35 Quiz
7.36 Case Study
7.37 Case Study - Demo
7.38 Summary
7.39 Conclusion

Lesson 08 - HBase

8.1 Hbase
8.2 Objectives
8.3 Why HBase
8.4 What is HBase
8.5 HBase Real-life Connect

8.6 Characteristics of HBase
8.7 Companies Using HBase
8.8 HBase Architecture
8.9 HBase Components
8.10 Storage Model of HBase
8.11 Row Distribution of Data between RegionServers
8.12 Data Storage in HBase
8.13 Data Model
8.14 When to Use HBase
8.15 HBase vs. RDBMS
8.16 Installation of HBase
8.17 Configuration of HBase
8.18 Business Scenario
8.19 Installing and configuring HBase Demo 01
8.20 Connecting to HBase
8.21 HBase Shell Commands
8.22 Spot the Error
8.23 Quiz
8.24 Case Study
8.25 Case Study - Demo
8.26 Summary
8.27 Conclusion

Lesson 09 - Commercial Distribution of Hadoop

9.1 Commercial Distribution of Hadoop
9.2 Objectives
9.3 Cloudera Introduction
9.4 Cloudera CDH
9.5 Downloading the Cloudera VM
9.6 Starting the Cloudera VM
9.7 Logging into Hue
9.8 Cloudera Manager
9.9 Logging into Cloudera Manager
9.10 Business Scenario
9.11 Download,start and Work with Cloudera VM Demo 01
9.12 Eclipse with MapReduce in Cloudera'sQuickstart VM Demo 02
9.13 Hortonworks Data Platform
9.14 MapR Data Platform
9.15 Pivotal HD
9.16 Pivotal HD (contd.)
9.17 IBM InfoSphere Big Insights
9.18 IBM InfoSphere Big Insights (contd.)
9.19 Quiz

9.20 Summary
9.21 Conclusion
Lesson 10 - ZooKeeper, Sqoop, and Flume
10.1 ZooKeeper, Sqoop, and Flume
10.2 Objectives
10.3 Why ZooKeeper
10.4 What is ZooKeeper
10.5 Features of ZooKeeper
10.6 Challenges Faced in Distributed Applications
10.7 Coordination
10.8 Goals and Uses of ZooKeeper
10.9 ZooKeeper Entities
10.10 ZooKeeper Data Model
10.11 Znode
10.12 Client API Functions
10.13 Recipe 1 Cluster Management
10.14 Recipe 2 Leader Election
10.15 Recipe 3 Distributed Exclusive Lock
10.16 Business Scenario
10.17 View ZooKeeper Nodes Using C LI Demo 1
10.18 Why Sqoop
10.19 What is Sqoop
10.20 Sqoop Real-life Connect
10.21 Sqoop and Its Uses
10.22 Sqoop and Its Uses (contd.)
10.23 Benefits of Sqoop
10.24 Sqoop Processing
10.25 Sqoop Execution Process
10.26 Importing Data Using Sqoop
10.27 Sqoop Import Process
10.28 Sqoop Import Process (contd.)
10.29 Importing Data to Hive and HBase
10.30 Exporting Data from Hadoop Using Sqoop
10.31 Sqoop Connectors
10.32 Sample Sqoop Commands
10.33 Business Scenario
10.34 Install Sqoop Demo 2
10.35 Import Data on Sqoop Using MySQL Database Demo 3
10.36 Export Data Using Sqoop from Hadoop Demo 4
10.37 Why Flume
10.38 Apache Flume Introduction
10.39 Flume Model

10.40 Flume Goals
10.41 Scalability in Flume
10.42 Flume Sample Use Cases
10.43 Business Scenario
10.44 Configure and Run Flume Agents Demo 5
10.45 Spot the Error
10.46 Quiz
10.47 Case Study ZooKeeper
10.48 Case Study ZooKeeper Demo
10.49 Case Study Sqoop
10.50 Case Study Sqoop Demo
10.51 Case Study Flume
10.52 Case Study Flume Demo
10.53 Summary
10.54 Conclusion 00:07

Lesson 11 - Ecosystem and Its Components

11.1 Ecosystem and Its Components
11.2 Objectives
11.3 Apache Hadoop Ecosystem
11.4 File System Component
11.5 Data Store Components
11.6 Serialization Components
11.7 Job Execution Components
11.8 Work Management, Operations, and Development Components
11.9 Security Components
11.10 Data Transfer Components
11.11 Data Interactions Components
11.12 Data Interactions Components (contd.)
11.13 Analytics and Intelligence Components
11.14 Search Frameworks Components
11.15 Graph-Processing Framework Components
11.16 Apache Oozie
11.17 Apache Oozie Workflow
11.18 Apache Oozie Workflow (contd.)
11.19 Introduction to Mahout
11.20 Schedule workflow with Apache Oozie Demo 01
11.21 Introduction to Mahout (contd.)
11.22 Features of Mahout
11.23 Usage of Mahout
11.24 Usage of Mahout (contd.)
11.25 Apache Cassandra
11.26 Characteristics of Apache Cassandra

11.27 Apache Spark
11.28 Apache Spark Tools
11.29 Key Concepts of Apache Spark
11.30 Apache Spark Example
11.31 Building a program using Apache Spark Demo 02
11.32 Hadoop Integration
11.33 Spot the Error
11.34 Quiz
11.35 Case Study
11.36 Case Study - Demo
11.35 Summary
11.36 Conclusion
Lesson 12 - Hadoop Administration, Troubleshooting, and Security
12.1 Hadoop Administration, Troubleshooting, and Security
12.2 Objectives
12.3 Typical Hadoop Core Cluster
12.4 Load Balancer
12.5 Commands Used in Hadoop Programming
12.6 Different Configuration Files of Hadoop Cluster
12.7 Properties of hadoop-default.xml
12.8 Hadoop Cluster Critical Parameters
12.9 Hadoop DFS Operation Critical Parameters
12.10 Port Numbers for Individual Hadoop Services
12.11 Performance Monitoring
12.12 Performance Tuning
12.13 Parameters of Performance Tuning
12.14 Troubleshooting and Log Observation
12.15 Apache Ambari
12.16 Key Features of Apache Ambari
12.17 Business Scenario
12.18 Troubleshooting a Missing DataNode Issue Demo 01
12.19 Optimizing a Hadoop Cluster Demo 02
12.20 Hadoop Security Kerberos
12.21 Kerberos Authentication Mechanism
12.22 Kerberos Configuration Steps
12.23 Data Confidentiality
12.24 Spot the Error
12.25 Quiz
12.26 Case Study
12.27 Case Study - Demo
12.28 Summary
12.29 Thank you

12.30 Usage of Trademarks

Free CourseJava Essentials for Hadoop

Lesson 01 - Essentials of Java for Hadoop

1.1 Essentials of Java for Hadoop

1.2 Lesson Objectives

1.3 Java Definition

1.4 Java Virtual Machine (JVM)
1.5 Working of Java
1.6 Running a Basic Java Program
1.7 Running a Basic Java Program (contd.)
1.8 Running a Basic Java Program in NetBeans IDE
1.9 BASIC JAVA SYNTAX
1.10 Data Types in Java
1.11 Variables in Java
1.12 Naming Conventionsof Variables
1.13 Type Casting.
1.14 Operators
1.15 Mathematical Operators
1.16 Unary Operators.
1.17 Relational Operators
1.18 Logical or Conditional Operators
1.19 Bitwise Operators
1.20 Static Versus Non Static Variables
1.21 Static Versus Non Static Variables (contd.)
1.22 Statements and Blocks of Code
1.23 Flow Control
1.24 If Statement
1.25 Variants of if Statement
1.26 Nested If Statement
1.27 Switch Statement
1.28 Switch Statement (contd.)
1.29 Loop Statements
1.30 Loop Statements (contd.)
1.31 Break and Continue Statements
1.32 Basic Java Constructs
1.33 Arrays
1.34 Arrays (contd.)
1.35 JAVA CLASSES AND METHODS

1.36 Classes
1.37 Objects
1.38 Methods
1.39 Access Modifiers
1.40 Summary
1.41 Thank You
Lesson 02 - Java Constructors
2.1 Java Constructors
2.2 Objectives
2.3 Features of Java
2.4 Classes Objects and Constructors
2.5 Constructors
2.6 Constructor Overloading
2.7 Constructor Overloading (contd.)
2.8 PACKAGES
2.9 Definition of Packages
2.10 Advantages of Packages
2.11 Naming Conventions of Packages
2.12 INHERITANCE
2.13 Definition of Inheritance
2.14 Multilevel Inheritance
2.15 Hierarchical Inheritance
2.16 Method Overriding
2.17 Method Overriding(contd.)
2.18 Method Overriding(contd.)
2.19 ABSTRACT CLASSES
2.20 Definition of Abstract Classes
2.21 sage of Abstract Classes
2.22 INTERFACES
2.23 Features of Interfaces
2.24 Syntax for Creating Interfaces
2.25 Implementing an Interface
2.26 Implementing an Interface(contd.)
2.27 INPUT AND OUTPUT
2.28 Features of Input and Output
2.29 System.in.read() Method
2.30 Reading Input from the Console
2.31 Stream Objects
2.32 tringTokenizer Class
2.33 Scanner Class
2.34 Writing Output to the Console
2.35 Summary

2.36 Thank You
Lesson 03 - Essential Classes and Exceptions in Java
3.1 Essential Classes and Exceptions in Java
3.2 Objectives
3.3 The Enums in Java
3.4 Program Using Enum
3.5 ArrayList
3.6 ArrayList Constructors
3.7 Methods of ArrayList
3.8 ArrayList Insertion
3.9 ArrayList Insertion (contd.)
3.10 Iterator
3.11 terator (contd.)
3.12 ListIterator
3.13 ListIterator (contd.)
3.14 Displaying Items Using ListIterator
3.15 For-Each Loop
3.16 For-Each Loop (contd.)
3.17 Enumeration
3.18 Enumeration (contd.)
3.19 HASHMAPS
3.20 Features of Hashmaps
3.21 Hashmap Constructors
3.22 Hashmap Methods
3.23 Hashmap Insertion
3.24 HASHTABLE CLASS
3.25 Hashtable Class an Constructors
3.26 Hashtable Methods
3.27 Hashtable Methods
3.28 Hashtable Insertion and Display
3.29 Hashtable Insertion and Display (contd.)
3.30 EXCEPTIONS
3.31 Exception Handling
3.32 Exception Classes
3.33 User-Defined Exceptions
3.34 Types of Exceptions
3.35 Exception Handling Mechanisms
3.36 Try-Catch Block
3.37 Multiple Catch Blocks
3.38 Throw Statement
3.39 Throw Statement (contd.)
3.40 User-Defined Exceptions

3.41 Advantages of Using Exceptions
3.42 Error Handling and finally block
3.43 Summary

Exam & Certification

How can I become a Certified Big Data &Hadoop Developer?

To become a Certified Big Data & Hadoop Developer, you must fulfill both the following criteria:

Complete any one project out of the four projects given by SiSTech, within the maximum time allotted for the Big Data Hadoop developer course. When you have completed your project, you'll email it to the lead trainer who will evaluate it.

Passing the online examination with a minimum score of 80%. If you don't pass the exam the first time, you can re-attempt the exam one more time.

When you have completed the course, you will receive an experience certificate stating that you have 3 months experience in implementing Big Data and Hadoop Projects.

Note: You must fulfill both the criteria (completion of any one Project and passing the online exam with minimum score of 80%) to become a Certified Big Data & Hadoop Developer.