



Hewlett Packard Enterprise

Course Datasheet

Big Data – Programming and Development

Education Services course product number – HPE-BD-v1.0

Course length – 60 Hrs.

Delivery mode – Instructor Led Training (ILT)

Virtual Instructor Led Training (vILT)

Course Objective

This course is an ideal package for individuals who want to understand the basic concepts of Big Data and Hadoop. Completing this course will make learner able to construe what goes behind the processing of huge volumes of data and preparing an individual for a job in Big Data Programming and Development space.

Prerequisite

- Basics of programming language
 - Concepts of OOP
 - Basics of scripting language
- Basics of Linux/Unix operating systems
- Good understanding of Java programming language
 - Core Java
- Understanding of basic SQL statements

Course Modules

Chapter 01 – Introduction to Big Data

- Introduction to Big Data
- Applicability of Big Data
- Introduction to Big Data technologies
- Introduction to Hadoop
- Distributed Computing Basics
- Evolution of Distributed Systems

Chapter 02 – Working with Hadoop and Its Components and Concepts

- Analysis of Hadoop
- HDFS and Hadoop Commands
- Introduction to MapReduce
- How MapReduce Works

Course Datasheet

- Pig
- Hive

Chapter 03 –Scripting with Hive & HBase

- Hive Data Types and File Formats
- Hive Query Language
- HBase Architecture Details
- Working with HBase

Chapter 04 – Programming using MapReduce for Big Data - 1

- Programming Concepts in Mapreduce
- HDFS programming in Java
- MapReduce programming in Java
- Executing a MapReduce program
- Debugging & Diagnosing Mapreduce program

Chapter 05 – Programming using MapReduce for Big Data - 2

- Job Chaining & Merging
- Input & Output patterns
- NextGen MapReduce using YARN & REST

Chapter 06 – Distributed Resource synchronization using ZooKeeper

- ZooKeeper in detail

Chapter 07 – Data loading using Sqoop

- Sqoop in detail
- Introduction to ETL and CDC
- TelenD
 - Introduction
 - Components
 - ETL Perspective
 - Installation
 - Basic Operations

Chapter 08 – Handling large log files using Flume

- Flume in detail
- Kafka
 - Introduction
 - Architecture and workflow
 - Installation
 - Basic operations

Chapter 09 – Handling workflows using Oozie

- Workflow scheduling using Oozie

Chapter 10 – Understanding Popular Big Data Platforms

- Cloudera, Hortonworks, Greenplum, Vertica