



# Hewlett Packard Enterprise

Course Datasheet

## **Big Data and Data Science (Analytics with R)**

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

Course length – 90 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 and also making you familiar with the field of analytics. 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 or in the analytics 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

#### Big Data – Programming and Development

##### 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

## Course Datasheet

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

## Course Datasheet

### Chapter 10 – Understanding Popular Big Data Platforms

- Cloudera, Hortonworks, Greenplum, Vertica

### Analytics with R

#### Chapter 1 – Introduction to Business Analytics

- Introduction to Business Analytics & its Features
- Types of Business Analytics
- Business Analytics Case Studies
- Business Decisions
- Business Intelligence
- Data Science and its importance

#### Chapter 2 - Introduction to R

- Introduction to R
- Understanding R
- Using R to illustrate the basic concepts
- Installing R and RStudio
- Integrated Development Environments (IDEs) for R
- Using R Console
- Scripting in R
- R Workplace and Packages
- Distributed R
  - Introduction
  - Installation
  - Programming Concepts

#### Chapter 3 - R Programming

- Introduction
- Operators in R (Arithmetic, Relational, Logical, Assignment)
- Basic and Advance Data Types
- Loops and Conditional Statement in R
- Commands to Run an R Script and a Batch Script
- Functions in R
- String Manipulation in R
- Dplyr Package – An Overview
- Installing Dplyr
- Functions of the Dplyr package

#### Chapter 4 - R Data Structure

- Types of Data Structures in R
- Vectors
- Scalars
- Matrices
- Arrays
- Data Frames
- Factors
- Lists
- Elements of the Different Data Structures in R

## Course Datasheet

- Acceptable Formats to Import and Export Data in R

### Chapter 5 - Data Visualization

- Graphics in R
- Types of Graphics
- Basic elements of graph
- Methods to Save Graphics as Files
- Procedure to Export Graphs in RStudio

### Chapter 6 - R Connection with Database

- Introduction to RDBMS
- Introduction to MySql
- R packages to connect to database
- Data analysis of data from database

### Chapter 7 - Debugging in R

- Introduction to Debugging
- Important Function to Debug

### Chapter 8 - Statistics in R

- Introduction to Statistics
- Types of Data
- Qualitative vs Quantitative Analysis
- Hypothesis Testing in R
- Need of Hypothesis Testing in Businesses
- Test of mean
- Test of variance
- Chi-square Test
- Non-parametric Test
- Linear Regression
- Basics of Classification
- Basics of Clustering