Hewlett Packard
Enterprise

Big Data – Programming and Development

HPE Certificate of Competency (CoC) Program - Associate Level
Exam Coverage

Modules

Module 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

Module 02 – Working with Hadoop and Its Components and Concepts

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

Module 03 –Scripting with Hive & HBase

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

Module 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

Module 05 – Programming using MapReduce for Big Data - 2

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

Module 06 – Distributed Resource synchronization using ZooKeeper

- ZooKeeper in detail

Course Datasheet

Module 07 – Data loading using Sqoop

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

Module 08 – Handling large log files using Flume

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

Module 09 – Handling workflows using Oozie

- Workflow scheduling using Oozie

Module 10 – Understanding Popular Big Data Platforms

- Cloudera, Hortonworks, Greenplum, Vertica