



Star Big Data Programming

Big Data Hadoop Training will make you an Expert in HDFS, Map Reduce, H base, Hive, Pig Yarn, Oazie, Flume and Sqoop using real time use cases on Retail, Social Media, Aviation, Tourism, Finance Domain. It equips you with in depth knowledge of writing codes using Map Reduce framework and managing Large Data Sets with H Base. The Topics covered in this course mainly includes-Hive, Pig and setup of Hadoop Cluster.

Audience: Star Certified – Big Data Developer assumes that the Learner is a Working Software Professional, or someone with Programming experience, wanting to understand the complexities of Big Data Development.

Course Objectives:

- About Hadoop and its eco-system.
- How to create program for Hadoop.
- How to create MapReduce Application.
- How to interact with related Technologies – Spark, Sqoop, Pig, Hive, Oozie etc.

- The flexibility and applications of Big Data to suit various Industries.
- About the complexities of Commercial Distribution Systems of Hadoop.
- About the various practical aspects of Installing and Working with Hadoop.

Course Outcome:

Understand the finer nuances of Big Data Technology and have answers to a lot of related questions. You will also be comfortable in dealing with Big Data related Tools, Platforms, and their Architecture to Store, Program, Process, and Manage the Data.

TOC Outline:

Part 1: Exploring Big Data and Hadoop

1. Introducing Data and Big Data
2. Identifying the Business Applications of Big Data
3. Big Data and Hadoop

Part 3: Programming for MapReduce

7. Developing a First Application for MapReduce
9. Exploring Additional Features of MapReduce

Part 4: Exploring Related Technologies

10. Exploring Avro – Data Serialization System
12. Exploring Flume – Service for streaming Event Data
14. Exploring Pig – Analysing Large Data
15. Exploring Hive – Data Warehouse
17. Exploring Crunch – Joining and Data Aggregation
19. Exploring HBase – Big Data Store
21. Exploring Storm – Real Time Computation
22. Machine Learning with Mahout

Part 5: Databases and Hadoop:

23. Interacting with No-SQL Databases

Part 6: Exploring Advanced Topics

24. Hadoop and Security
26. Exploring Cloudera
28. Exploring Pivotal HD
30. Exploring IBM InfoSphere

Part 7: Case Studies and Lab Exercises

Exam Information:

Exam Code	: 507-116	Exam Pattern	: Multiple Choice
Exam Duration	: 2.5 Hrs	Exam Delivery	: AEPTC (ACADEMIC EDUCATION & PROFESSIONAL TESTING CENTER)
Passing Score	: 75%		

Course Duration: 90 Hrs

Part 2: Discussing Hadoop Eco-system

4. Introduction to MapReduce
5. Exploring HDFS: Storing Data
6. Yarn and Map Reduce
8. Exploring the working of MapReduce Process
11. Exploring Parquet - Columnar Storage Format
13. Exploring Sqoop – Transferring Bulk Data (Hadoop → Relational Databases)
16. Exploring Oozie – Workflow Scheduler
18. Exploring Spark and Scala
20. Exploring ZooKeeper – Coordination Service for Distributed Applications
25. Exploring Apache Drill and Google BigQuery
27. Exploring Hortonworks
29. Exploring HDInsight
31. Exploring Hadoop and AWS



**BIG DATA
Programming**

