



Vendor-Neutral Global IT Certifications

Star Big Data Programing



STRUCTURE DESIGN & BUILD ROBUST DATA TRACKNG MODELS



Big Data Programing

Exam Code: S07-116

LEARN



+

EXAM



=

CERTIFIED



 www.starcertification.org

 info@starcertification.org

Star Big Data Programming

To stay competitive a business needs to know as much as it can about people, the environment it's operating in, and who and where the competitors are. The amount of data companies collect keeps growing. There is an urgent need of a strategy to make sense of it all. Star Big Data Programming is a certification course that will help learners master the skills they need to establish a successful career as a data engineer. The program will help the learners master the skills on HDFS, MapReduce, HBase, Hive, Pig, Yarn, Oozie, Flume and Sqoop using real-time use cases from retail, social media, aviation, tourism, and finance industries. It equips the learners with in-depth knowledge of writing code using the MapReduce framework and managing large data sets with HBase.



Audience: Star Big Data Programming assumes that the learner has working knowledge of Java, or someone with good programming understanding, wanting to understand the complexities of Big Data development.

Course Objectives:

In this course, you will learn about:

- Big data and its business applications
- Apache Hadoop and its big data eco-system
- Deploying Hadoop in a clustered environment
- Interacting with No-SQL databases
- Managing key Hadoop components (HDFS, YARN and Hive)
- Spark - the next-generation computational framework
- Installing and working with Hadoop
- Hadoop related technologies – Avro, Flume, Sqoop, Pig, Oozie, etc
- Advanced topics like Hadoop security, Cloudera, IBM InfoSphere and more

Course Outcome :

After completing this course, you will be able to:

- Understand the finer nuances of the Big Data technology
- Deal with Big Data related tools, platforms, and their architecture to store, program, process, and manage the data
- Deploy Hadoop and its related technologies
- Use the Hadoop ecosystem to manage your data
- Deploy machine learning concepts with Mahout

Course Outline :

1. Introducing Data and Big Data
2. Identifying the Business Applications of Big Data
3. Big Data and Hadoop
4. HDFS - Storing Data in Hadoop
5. Introduction to MapReduce
6. YARN and MapReduce- Processing Data in Hadoop
7. Developing a First Application for MapReduce
8. Exploring the Working of a MapReduce Process
9. Avro
10. Parquet
11. Flume - Service for Streaming Event Data
12. Sqoop (MySQL to Hadoop)
13. Apache Pig
14. Hive – Data Warehouse
15. Oozie– Workflow Scheduler
16. Exploring Crunch- Joining and Data Integration
17. Exploring Spark and Scala
18. Exploring HBase- Big Data Store
19. Zookeeper - Coordination Service for Distributed Applications
20. Exploring Storm
21. Machine Learning with Mahout
22. Interacting with NoSQLDatabases
23. Hadoop and Security
24. Apache Drill and Google BigQuery
25. Exploring Cloudera
26. Exploring Hortonworks
27. HDInsight
28. IBMInfosphere
29. Hadoop and AWS
30. Appendix- Exploring Pivotal HD Case Studies

Labs

- Chapter 1. Setting up the required environment for Apache Hadoop installation
- Chapter 2. Installing the Single-Node Hadoop configuration on the system
- Chapter 3. Exploring the Web-Based User Interface of Hadoop Cluster
- Chapter 4. Implementing Map-Reduce Program for Word Count
- Chapter 5. Implementing Basic Pig Latin Script
- Chapter 6. Implementing Basic Hive Query Language Operations
- Chapter 7. Using Apache Flume to fetch open-source user tweets from Twitter

Exam Information:

Exam Code	: S07-116	Exam Pattern	: Multiple Choice
Exam Duration	: 2.5 Hrs	Exam Delivery	: AEPTC (ACADEMIC EDUCATION & PROFESSIONAL TESTING CENTER)
Passing Score	: 70%		

Course Duration : 72Hrs