



BIGDATA-101

Classic Hadoop

Classic Hadoop

OVERVIEW

Skill Level : Beginner

Suitable for : Classic Hadoop was suitable for a specific set of people who were dealing with large-scale data processing and analytics challenges during the time when it was widely used.

Duration : 10 Days

Classic Hadoop is a pioneering open-source framework designed for distributed storage and processing of large datasets across clusters of commodity hardware. It introduced the MapReduce programming model, enabling parallel data processing tasks on the stored data.

At its core, Classic Hadoop consists of the Hadoop Distributed File System (HDFS), which stores data across a cluster, and the MapReduce processing framework for parallel computation. These components formed the foundation of the Hadoop ecosystem.

PREREQUISITES

- JAVA-101 – Java Fundamentals
- SQL-101 – SQL Fundamentals

LEARNING OUTCOMES

- Learning Hadoop exposes you to the MapReduce programming model, which involves breaking down complex tasks into smaller subtasks (map) and then aggregating the results (reduce). This approach enhances your understanding of parallel processing and fault tolerance.

- Hadoop's HDFS teaches you about distributed data storage, replication strategies, and data retrieval mechanisms. You'll understand how to manage data across a cluster for reliability and quick access.

COURSE OUTLINE

Day 1

- Introduction to Big Data & Hadoop
 - OLTP vs OLAP
 - Hadoop vs RDBMS
 - Data Sources
 - Data Lake
 - Data Warehouse
 - Data Marts
- Major Hadoop Vendors
 - Open Source
 - Cloudera Data Platform
 - Map/R
 - Amazon Web Services
 - Google Cloud Platform
 - Microsoft Azure
- Hadoop Architecture
 - HDFS
 - Namenode
 - Datanode
 - Blocks
 - Command Line
 - WebHDFS
 - Yarn
 - Resource Manager
 - Node Manager
 - Zookeeper
- Hue
- Working Group Formation

Day 2

- Algorithms – MapReduce
- Engine – Tez
- Sqoop
 - JDBC
- Hive
 - Commands
 - Data Types
 - DDL
 - DML
 - Metastore

- Partitions
- File Formats

Day 3

- Presto
- Flink

Day 4

- Oozie

Day 5

- Kafka

Day 6

- HBase
 - HMaster
 - Region Server
 - ZooKeeper
 - HDFS

Day 7

- HDFS
 - Maintenance
 - Rack Awareness
 - Tuning
 - Failover & Disaster Recovery
 - Rebalancing
- Security
 - Kerberos
 - Ranger
- YARN
 - Maintenance
 - Tuning
 - Failover & Disaster Recovery

Day 8 - 10

- Deployment on Amazon Web Services
- Sample Application
- Individual and Group Work
- Presentations and Final Exam



Engineering for the Real World

Enquiries



+63 2 5322 2307



training-sales@orangeandbronze.com