HDP Overview: Apache Hadoop Essentials

Learn via: **Classroom / Virtual Classroom / Online**

Duration: **1 Day**

https://bilginc.com/en/training/hdp-overview-apache-hadoop-essentials-905-training/

Overview

This course provides a technical overview of Apache Hadoop. It includes high-level information about concepts, architecture, operation, and uses of the Hortonworks Data Platform (HDP) and the Hadoop ecosystem. The course provides an optional primer for those who plan to attend a hands-on, instructor-led course.

This course is offered in both Live Instructor-Led format, or get started now with our FREE self-paced Apache Hadoop Essentials course.

Prerequisites

No previous Hadoop or programming knowledge is required.Students will need browser access to the Internet.

Who Should Attend

Data architects, data integration architects, managers, C-level executives, decision makers, technical infrastructure team, and Hadoop administrators or developers who want to understand the fundamentals of Big Data and the Hadoop ecosystem.

What You Will Learn

- At the completion of the course students will be able to:
- Describe the case for Hadoop
- Identify the Hadoop Ecosystem architecture
- Data Management - HDFS, YARN
- Data Access - Pig, Hive, HBase, Storm, Solr, Spark
- Data Governance & Integration - Falcon, Flume, Sqoop, Kafka, Atlas
- Security - Kerberos, Falcon, Knox
- Operations - Ambari, Zookeeper, Oozie, Cloudbreak
- Observe popular data transformation and processing engines in action: Apache
- Hive , Apache Pig, Apache Spark
- Detail the architecture and features of YARN
- Describe backup and recovery options
- Describe how to secure Hadoop
- Explain the fundamentals of parallel processing
- Describe data ingestion options and frameworks for batch and real-time streaming
- Detail the HDFS architecture

Outline

- Operational overview with Ambari
- Loading data into HDFS
- Data manipulation with Hive
- Risk Analysis with Pig
- Risk Analysis with Spark and Zeppelin
- Securing Hive with Ranger