

Big Data Course Content

Overview: Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behavior analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set.

Day 1

Session 1: Introduction to Big Data

- Traditional Data Processing Technologies
- Apache Hadoop Architecture
 1. Hadoop Architecture
 2. Hadoop and RDBMS
 3. Hadoop Distribution

Session 2:

- HDFS Architecture
- Hadoop Ecosystem – MapReduce, Hadoop Streaming , Hive, Hbase
- Where Hadoop fits in the Enterprise
- Pig
 1. Introduction to Apache Pig
 2. Components of pig
 3. Map Reduce vs. Apache Pig
 4. Different Data Types in Pig
 5. Modes of Execution in Pig(Local Mode)
 6. Execution Mechanism
 7. Grunt Shell
 8. Pig Command
 9. Examples Of pig
 10. Word Count
 11. Batting Examples

Day 2

Session 1:

- Hadoop Setup and Installation
- HDFS Programming Basics
 1. Hadoop Streaming
 2. Performance Tuning
 3. Debugging Hadoop Programs
- MapReduce Architecture
 1. MapReduce Programming Basics
 2. MapReduce Programming Using Big Insights

Session 2:

- Accessing Hadoop Data Using Hive
- Hive Architecture.
- Downloading, Installing and Configuring Hive.
- Understand what Apache Hive is and Hive use cases.
- Make basic configuration changes in a Hive installation.
- Use DDL to create new Hive databases and tables.

Session 3:

- Query
- Competition
- Certificate Distribution