b1Online trainings
Global Learning Portal

Hadoop Course Content:35-40hours

Course Outline

Introduction and Motivation of Hadoop

- What is Big Data
- Challenges in Big Data
- Challenges in Traditional Application
- New Requirements
- What is Hadoop
- Brief history of Hadoop
- Features of Hadoop
- Hadoop v/s RDBMS
- Hadoop Ecosystem's overview
- Overview of HDFS and MapReduce

Understanding Hadoop Distributed File System

- Understanding Configuration
- HDFS Concepts
- Blocks
- Replication
- Version File
- Safe mode
- Namespace IDs
- Reading and Writing in HDFS
- Understanding Name Node
- Understanding Data Node
- Understanding Secondary Name Node
- Understanding Job Tracker
- Understanding Task Tracker

HDFS Shell Commands

- Hands On Exercise

Accessing HDFS using API

- Understanding HDFS Java classes and methods
- Hands On Exercise

Map Reduce Programming

- Understanding block and input splits
- Common Input and Output Formats
- MapReduce Data types
- Understanding Writable and WritableComparable (Introduction)
- Data Flow in MapReduce Application
- Understanding WordCount problem
- Writing MapReduce Application
- Understanding Mapper function
- Understanding Reducer Function
- Understanding Driver
- Understanding Tool Runner
- Hands on Exercise

MapReduce Continued

- Using Combiner
- Using Distributed Cache
- Passing the parameters to mapper and reducer
- Hands On Exercise
- Writing Custom key values
- Writing Custom Partitioner
- Hands On Exercise

Introduction to PIG

- Terminology
- Understanding Pig Program, structure and Execution
- Pig Data types
- Loading and Dumping Data
- Filtering
- Group and Co-Group
- Joins
- Inner Join
- Left Outer Join
- Hands on
- Right Outer Join
- Full Outer Join

Introduction to Hive

- Motivation and Understanding Hive
- Using Hive Command line Interface
- Data types and File Formats
- Basic DDL operations
- Schema Design

Hadoop Ecosystem and Other Related Projects

- HBase
- Sqoop
- Flume
- Oozie