

Lunar Technologies

Your inspiring Training partner

Hadoop Training

What is Hadoop?

Hadoop in a nutshell is a framework which provides a shared storage and analysis system. The storage is provided by HDFS and analysis by Mapreduce. There are other parts in Hadoop but these are the main ones which catch our eye. It allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage.

Why I need Apache Hadoop?

Hadoop brings the ability to cheaply process large amounts of data, regardless of its structure. By large, we mean from 10-100 gigabytes and above. Existing enterprise data warehouses and relational databases excel at processing structured data and can store massive amounts of data, though at a cost: This requirement for structure restricts the kinds of data that can be processed, and it imposes an inertia that makes data warehouses unsuited for agile exploration of massive heterogeneous data. The amount of effort required to warehouse data often means that valuable data sources in organizations are never mind. This is where Hadoop can make a big difference.

Who can learn Apache Hadoop?

Hadoop is a framework for Java Development. This framework helps you build the applications implementing the logic of map and reduce much easier. Anyone with prior understanding and acquaintance with the concepts of core Java is ready to start.

Job Prospects in Hadoop

New job opportunities are emerging for IT professionals in the field of "big data," the term used to describe how corporations gather vast amounts of real-time data about their customers and analyze that data to drive decision making and increase profitability.

Pre-requisites:

1. Coding experience using concepts of Core Java
2. Basic understanding of Linux Commands.

Lunar Technologies

Your inspiring Training partner

Training

35-40 hours of training

How Can We Help You

- Enhance your skill set if you are already working on a technology
- Help you as independent consultants
- Help your stay ahead and get prepared for what the future has to offer you.

Hadoop Syllabus

Hadoop Developer/Admin Training Course Content

Hadoop Architecture

- Introduction to Bigdata Hadoop.
- Parallel Computer vs. Distributed Computing
- Hadoop Daemons introduction: NameNode, DataNode, JobTracker, TaskTracker
- How to configure Hadoop on your system
- How to configure Hadoop cluster on multiple machines
- Exploring HDFS (Hadoop Distributed File System)
- Exploring the HDFS Apache Web UI
- NameNode architecture (EditLog, FsImage, location of replicas)
- Secondary NameNode architecture
- DataNode architecture

MapReduce Architecture

- Exploring JobTracker/TaskTracker
- How a client submits a Map-Reduce job
- Exploring Mapper/Reducer/Combiner
- Shuffle: Sort & Partition
- Input/output formats
- Job Scheduling (FIFO, Fair Scheduler, Capacity Scheduler)
- Exploring the Apache MapReduce Web UI

Lunar Technologies

Your inspiring Training partner

Hadoop Developer Tasks

- Writing a map-reduce programme
- Reading and writing data using Java
- Hadoop Eclipse integration
- Mapper in details
- Reducer in details
- Using Combiners
- Reducing Intermediate Data with Combiners
- Writing Partitioners for Better Load Balancing
- Sorting in HDFS
- Searching in HDFS
- Indexing in HDFS
- Hands-On Exercise

Hadoop Administrative Tasks

- Routine Administrative Procedures
- Understanding dfsadmin and mradmin
- Block Scanner, Balancer
- Health Check & Safe mode
- DataNode commissioning/decommissioning
- Monitoring and Debugging on a production cluster
- NameNode Back up and Recovery
- Upgrading Hadoop

HBase Architecture

- Introduction to Hbase
- HBase vs. RDBMS
- Exploring HBase Master and Region Servers
- Column Families and Regions
- Basic Hbase shell commands.
- Hbase table operations