



BIG Data:

- What is Big Data
- Necessity of Big Data and Hadoop in the industry
- Paradigm shift - why the industry is shifting to Big Data tools
- Different dimensions of Big Data
- Data explosion in the Big Data industry
- Various implementations of Big Data
- Different technologies to handle Big Data
- Traditional systems and associated problems
- Future of Big Data in the IT industry

BigData-Hadoop:

- Why Hadoop is at the heart of every Big Data solution
- Introduction to the Big Data Hadoop framework
- Hadoop architecture and design principles
- Ingredients of Hadoop
- Hadoop characteristics and data-flow
- Components of the Hadoop ecosystem
- Hadoop Flavors – Apache, Cludera, Hortonworks, and more

Hadoop Installation of Single Node Cluster:

- Hadoop environment setup and pre-requisites
- Hadoop Installation and configuration
- Working with Hadoop in pseudo-distributed mode
- Troubleshooting encountered problems

Hadoop Installation of Multi Node Cluster:

- Hadoop environment setup on the cloud (Amazon cloud)
- Installation of Hadoop pre-requisites on all nodes
- Configuration of masters and slaves on the cluster
- Playing with Hadoop in distributed mode

Hadoop HDFS – Storage layer

- What is HDFS (Hadoop Distributed File System)
- HDFS daemons and architecture
- HDFS data flow and storage mechanism
- Hadoop HDFS characteristics and design principles
- Responsibility of HDFS Master – NameNode
- Storage mechanism of Hadoop meta-data
- Work of HDFS Slaves – DataNodes
- Data Blocks and distributed storage
- Replication of blocks, reliability, and high availability
- Rack-awareness, scalability, and other features
- Different HDFS APIs and terminologies
- Commissioning of nodes and addition of more nodes
- Expanding clusters in real-time
- Hadoop HDFS Web UI and HDFS explorer
- HDFS best practices and hardware discussion

All About Map Reduce:

- What is MapReduce, the processing layer of Hadoop
- The need for a distributed processing framework
- Issues before MapReduce and its evolution
- List processing concepts
- Components of MapReduce – Mapper and Reducer
- MapReduce terminologies- keys, values, lists, and more
- Hadoop MapReduce execution flow
- Mapping and reducing data based on keys
- MapReduce word-count example to understand the flow
- Execution of Map and Reduce together
- Controlling the flow of mappers and reducers

All about Hadoop HIVE – Data Analysis Tool:

- The need for an adhoc SQL based solution – Apache Hive
- Introduction to and architecture of Hadoop Hive
- Playing with the Hive shell and running HQL queries
- Hive DDL and DML operations
- Hive execution flow
- Schema design and other Hive operations
- Schema-on-Read vs Schema-on-Write in Hive
- Meta-store management and the need for RDBMS
- Limitations of the default meta-store
- Using SerDe to handle different types of data
- Optimization of performance using partitioning
- Different Hive applications and use cases