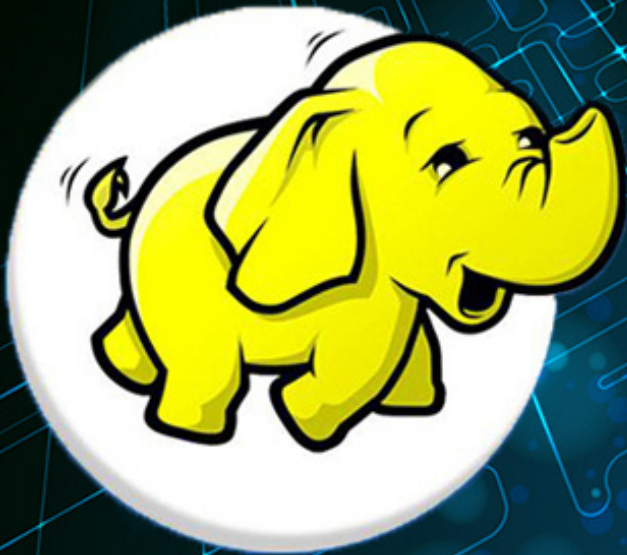




Fusion Software
Institute

WWW.FUSION-INSTITUTE.COM

BIG DATA HADOOP



Big Data
Hadoop

TABLE OF CONTENTS

1. BIGDATA AND HADOOP

- *Introduction to BigData and Hadoop*
- *Hadoop Storage:HDFS (Hadoop Distributed File System)*
- *Hadoop Processing: MapReduce Framework*

2. HADOOP ARCHITECTURE

- *Traditional processing Vs MapReduce Processing*
- *Hadoop 2.x clusterArchitecture[YARN]*
- *YARN in Details: - component, Architecture and Workflow*
- *Anatomy of MapReduce Program*
- *Federation and High Availability Architecture*
- *Hadoop clustermode*
- *Single Node and Multi-Node cluster setup*
- *HDFS commands*
- *Anatomy of MapReduce Program*
- *MapReduce Combiner and Partitioner*
- *Different file format in Hadoop*
- *Different compression format in Hadoop*
- *Basic Hadoop Administration*

TABLE OF CONTENTS

3.. APACHE HIVE

- *Introduction to Apache Hive*
- *Hive Architecture and Components*
- *Hive Limitation and Hive feature*
- *Hive Metastore*
- *Hive Data Types*
- *Hive partition and Bucketing*
- *Hive Tables (Managed and External Hive Table)*
- *Importing Data*
- *Querying Data and Managing Outputs*
- *Hive Scripts and Hive UDF*
- *Hive QL: Joining Tables, Dynamic Partitioning*
- *Hive Indexing and views*
- *Hive Query Optimization*
- *Hive Thrift server*
- *Hive UDF*
- *Hive SerDe*

4. APACHE HBASE

- *Apache HBase: Introduction to NoSQL Databases and HBase*
- *HBase v/s RDBMS*
- *HBase components*
- *HBase Architecture*
- *HBase Run Modes*
- *HBase Data Model*

TABLE OF CONTENTS

- *HBase shell*
- *HBase and schema Design*
- *HBase Client API*
- *Sqoop Introduction*

6. SQOOP

- *Sqoop Import*
- *Sqoop Import-all-Tables*
- *Sqoop Export*
- *Sqoop Job*
- *Sqoop Codegen*
- *Sqoop Eval*
- *Sqoop commands*

7. SPARK

- **Spark Basics** *What is Apache SparkSpark contextUsing Spark Shell*
- **Working with RDDs** *RDD Operations - Transformations and Action*
- **Spark on a Cluster** *OverviewA SparkStandalone ClusterThe Spark*
- *Writing Spark Application*
- *Caching and Persistence*
- *Improving Spark Performance*

TABLE OF CONTENTS

- *Spark API for different file formats and Compression Codecs*
- *Spark SQL*
- *Spark Streaming*

8. KAFKA

- *Overview*
- *Producer*
- *Consumer*
- *Broker*
- *Topics*
- *Partitions*
- *Kafka Integration with Spark Streaming*

9. OOZIE

- *Introduction to Oozie*
- *Oozie concepts*
- *Oozie workflow actions*
- *Workflow Applications*
- *Oozie coordinator*
- *Data triggercoordinator*
- *Oozie Bundles*

TABLE OF CONTENTS

10.SCALA

- *Introduction to Scala*
- *Scala Basics*
- *Data types and Variables*
- *Classes and Objects*
- *Access Modifiers*
- *Operators*
- *Conditional Statements*
- *Collections*
- *Pattern Matching*
- *Regular Expressions*
- *Exception Handling*
- *File IO*

11. CORE JAVA

12.CLOUD PLATFORM AWS

13.GIT HUB

14.JIRA

15.SVN

16.AGILE METHODOLOGY

17. TWO LIVE PROJECT

18.100% JOB PLACEMENT

TABLE OF CONTENTS

10. SCALA

- *Introduction to Scala*
- *Scala Basics*
- *Data types and Variables*
- *Classes and Objects*
- *Access Modifiers*
- *Operators*
- *Conditional Statements*
- *Collections*
- *Pattern Matching*
- *Regular Expressions*
- *Exception Handling*
- *File IO*

6. TWO LIVE PROJECT