
BIG DATA ANALYTICS
Unit Wise Important Questions

UNIT - I

1. Identify the components of GFS Architecture and explain with an example.
2. Identify the building blocks of HDFS and explain with an example.
3. Choose the installation steps to install Hadoop in Simple Mode and explain with an example.
4. Build a distributed file system by installing Hadoop in pseudo Distributed Mode and explain with an example.
5. Build a distributed file system by installing Hadoop in Fully Distributed Mode and explain with an example.

UNIT – II

1. Examine the Classical mapreduce framework and explain it.
2. Examine the MapReduce-II framework and explain it.
3. Assume the weather data set and examine the mapreduce task on the dataset to find out the minimum temperature.
4. Assume the input data set and examine the mapreduce task on the dataset to perform wordcount.
5. Explain Mapper code, Reducer code, Driver code, combiner and partitioner for mapreduce paradigm.

UNIT – III

1. Define RDD. Explain the steps to create RDDs in Apache Spark.
2. Explain the working of SPARK with the help of its architecture.
3. List and explain RDD actions in spark with example.
4. List and explain RDD transformations in spark with example.
5. Explain in details about persistence in SPARK with an example

UNIT – IV

1. Analyze the architecture of a pig with a neat sketch.
2. List and categorize the operators supported by pig.
3. List the modes of running a pig script and explain with a neat sketch.
4. List and explain various data types supported by pig in its data model.
5. Analyze the syntax of a Pig script with suitable example.

UNIT – V

1. Explain about architecture of Hive.
2. Explain the process of creating and managing databases and tables in Hive.
3. Explain querying and analyzing data in Hive.
4. Explain in detail about the data types supported by HIVE.
5. Explain HIVEQL DDL and DML commands with their syntax and examples.