What is Big Data? Discuss Four V's of Big Data. Explain characteristics of Big Data. Explain how big data processing differs from distributed processing.

List various application of big data. How it can be used to improve business for a superstore.

Explain "Map Phase" and "Combiner Phase" in Map-Reduce. Explain working of reduce phase of Map-Reduce with an example. Explain "Shuffle & Sort" phase and "Reducer Phase" in Map-Reduce.

List various configuration files used in Hadoop Installation. What is use of mapred-site.xml?

How to create collection in MongoDB? Explain with its syntax. Explain MongoDB sharding process. Which terms are used for table, row, column and table-join in MongoDB?

Explain structured, semi structured and unstructured data in terms of big data analytics.

Differentiate between: SQL and NoSQL

Write down the differences between Apache Pig and Map-Reduce.

What is Apache Pig and why do we need it?

Explain the components of SPARK. Explain about the major libraries that constitute the Spark Ecosystem. How can you minimize data transfers when working with Spark?

Write down the goals of HDFS.

Explain Avro data serialization technique in Map-Reduce.

What is Zookeeper? What are the benefits of Zookeeper?

How Big Data Analytics can be useful in the development of smart cities

Compare Raw oriented and Column Oriented database structures.

Discuss how will Pig data model help in effective data flow

What is Compute and Storage nodes in Hadoop?

Explain the core methods of the Reducer?

What are primary and secondary replica sets in MongoDB?

What is the role of a profiler in MongoDB? Where does the writes all the data?

What are the tenV's of big data?

How are big data and Hadoop related to each other?

How is big data analysis helpful in increasing business revenue?

What are the applications of Big data?

Explain the steps to be followed to deploy a Big Data solution.

Define Big Data And Explain The Five Vs of Big Data.

Define HDFS and YARN, and talk about their respective components.

Explain the term ‘Commodity Hardware.’

Define and describe the term FSCK.

What is the purpose of the JPS command in Hadoop?

Name the different commands for starting up and shutting down Hadoop Daemons.

Explain structured, semi structured and unstructured data in terms of big data analytics.

Explain working of reduce phase of MapReduce with an example.

Define HDFS. Describe namenode, datanode and block. Explain HDFS operations in detail.

What is Big Data? Explain how big data processing differs from distributed processing.

Explain characteristics of Big Data.

What is Big data? Discuss it in terms of volume and velocity.

Why do we need Hadoop for Big Data Analytics?

Talk about the different tombstone markers used for deletion purposes in HBase.

Differentiate SQL and NoSQL.

What is HBase? Explain storage mechanism of HBase with an example.

How to create collection in MongoDB? Explain with its syntax.

Explain CRUD operations of MongoDB with an example..Explain MongoDB sharding process.

What is NoSQL? List out the features of NoSQL. Explain types of NoSQL databases in brief

Define Term Frequency and Inverse Document Frequency

"Explain following commands with syntax and at least one example of

each. (1) copyFromLocal (2) showing the content of outputfile."

"Write Map Reduce steps for counting occurrences of specific numbers in the input text file(s). Also write the commands to compile and run the code."

Explain “Shuffle & Sort” phase and “Reducer Phase” in MapReduce.

Explain following in brief with respect to Mongo DB : 1) Collections and documents 2) Indexing and retrieval

Write difference between MangoDB and Hadoop.

What is NoSQL database? List the differences between NoSQL and relational databases. Differentiate between SQL and NoSQL Explain in brief various types of NoSQL databases in practice.

Explain scaling in MangoDB

Explain CRUD operations in MongoDB.

What is Resilient Distributed Dataset in Apache Spark? Explain in detail. Make a note on why RDD is better than Map Reduce data storage?

Which terms are used for table, row, column and table-join in MongoDB?

What are the common input formats in Hadoop?

Explain the different modes in which Hadoop run.

Explain the core components of Hadoop. What are the different configuration files in Hadoop?

How can you achieve security in Hadoop?

What are the differences between Hadoop 2 and Hadoop 3?

How is NFS different from HDFS?

Explain the process that overwrites the replication factors in HDFS.

What will happen with a NameNode that doesn't have any data?

Explain NameNode recovery process.

Define respective components of HDFS and YARN

What is fsck?

What are the main differences between NAS (Network-attached storage) and HDFS?

What is the Command to format the NameNode?

How to recover a NameNode when it is down?

What do you understand by Rack Awareness in Hadoop?

What is the difference between "HDFS Block" and "Input Split"?

Explain the difference between Hadoop and RDBMS.

What are the configuration parameters in a "MapReduce" program?

What is Distributed Cache in a MapReduce Framework?

What are the three running modes of Hadoop?

How can you achieve security in Hadoop?

How is NFS different from HDFS?

What are the Port Numbers for NameNode, Task Tracker, and Job Tracker?

What are the different file permissions in HDFS for files or directory levels?

What are the basic parameters of a Mapper?

Explain the process that overwrites the replication factors in HDFS.

How Is Hadoop CLASSPATH essential to start or stop Hadoop daemons?

Why is HDFS only suitable for large data sets and not the correct tool to use for many small files?

Why do we need Data Locality in Hadoop? Explain.

DFS can handle a large volume of data then why do we need Hadoop framework?

How does HDFS Index Data blocks? Explain.

What are Edge Nodes in Hadoop?

Explain the core methods of a Reducer.

What are the advantages of Hadoop? Draw Hadoop ecosystem and explain its components.

"Write the use and syntax of following HDFS commands:

i. put ii. expunge iii. chmod iv. get"

Write down the goals of HDFS.

Explain core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail.

What is Hadoop Ecosystem? Discuss various components of Hadoop Ecosystem.

"What is data serialization? With proper examples discuss and differentiate structured, unstructured and semi-structured data. Make a note on how

type of data affects data serialization."

List various configuration files used in Hadoop Installation. What is use of mapred-site.xml?

Define HDFS. Discuss the HDFS Architecture and HDFS Commands in brief.

What is RDD? Explain transformations and actions in RDD. Explain RDD operations in brief.

Write down the differences between Apache Pig and MapReduce.

Explain Five V's of Big Data in brief

Justify: "SPARK is faster than MapReduce".

What is Apache Pig and why do we need it?

Explain the components of SPARK.

Explain the architecture and features of HIVE. Explain Metastore in Hive.

Explain Spark components in detail. Also list the features of spark

Write Short note on Hadoop Ecosystem also explain various elements of hadoop.

Write a brief short note on: Spark Unified Stack

Explain Avro data serialization technique in MapReduce.

What is Zookeeper? What are the benefits of Zookeeper?

Draw architecture of APACHE PIG and explain in short.

Draw and explain Architecture of APACHE HIVE. Explain various data insertion techniques in HIVE with example.\Explain benefits of ZooKeeper

Discuss Machine Learning with MLlib in SPARK

Discuss the applications of big data analytics in weather forecasting. List various application of big data. How it can be used to improve business for a superstore. How can Big Data Analytics be useful in the development of smart cities?

What are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities.(Discuss one application)

What is Big Data? What do you mean by Big Data Analytics? What is importance of Business Analytics and Predictive Analytics?

What is Big Data? Explain how big data processing differs from distributed processing.

What is/are difference(s) between structured data, unstructured data and semi-structured data? Give the example of each and explain in brief.

List various application of big data. How it can be used to improve business for a superstore.

Compare Distributed File System (DFS), Google File System (GFS) with Hadoop Distributed File System (HDFS)? Give various criteria for comparison also.

Explain in brief the framework and architecture of Hadoop.

Elaborate the working of Map-Reduce Algorithm. What do you mean by heartbeat and replica in Hadoop?

Explain characteristics of Big Data. Discuss Ten V's of Big Data.

Differentiate between: SQL and NoSQL

Explain working of reduce phase of MapReduce with an example.

Explain working of map phase of MapReduce with an example. Explain “Map Phase” and “Combiner Phase” in MapReduce.

Explain “Shuffle & Sort” phase and “Reducer Phase” in MapReduce.

How to create collection in MongoDB? Explain with its syntax.

What is Apache Pig and why do we need it? Write down the differences between Apache Pig and MapReduce.

Write down the goals of HDFS.

Explain MongoDB sharding process

Which terms are used for table, row, column and table-join in MongoDB?

Explain Avro data serialization technique in MapReduce.

What is Zookeeper? What are the benefits of Zookeeper?

Explain Collections and documents, Indexing and retrieval wrt to MongoDB.

Compare Raw oriented and Column Oriented database structures.

Compare Cassandra with HBase and MongoDB.

Compare HDFS and HBase.

Compare RDBMS with Cassandra

Compare RDBMS with Neo4j

What is Graph Database? Give the applications of it.

Explain the features of HDFS

Explain the commands for HDFS

Which are the advantages of Hadoop? Explain Hadoop architecture and its components.

What is Map-Reduce? Which are its advantages?

What is NoSQL Databases? Which are the major categories of each? Explain in brief with examples.

Explain the working HBase with proper steps and diagram.

Explain the working Cassandra with proper steps and diagram.

Explain the working Neo4j with proper steps and diagram.

Explain the working MongoDB with proper steps and diagram.

Which are the features of BigTable and Titan InitGraph have been combined in Cassandra?

What is Zookeper? List the benefits of it?

Write a note on Apache pig and enlist the applications of Apache pig.

Differentiate between Apache Pig and Map-Reduce.

Differentiate between Apache Pig and Hive.

Which are the problems related to Map-Reduce data storage?

Explain NewSQL. Explain the characteristics of NewSQL.

Show basic CRUD operations in MongoDB with proper examples.

Explain principles of schema design in MongoDB.

Explain the advantages of MongoDB over RDBMS.

Explain Job Scheduling in Map-Reduce.

What is Map-Reduce? Explain working of various phases of M-R with appropriate example & diagram.