BIG DATA INTRODUCTION	1. 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. What are the applications of Big data? 2. Explain types of Big Data 3. Explain structured, semi structured and unstructured data in terms of big data analytics. 4. How Big Data Analytics can be useful in the development of • smart cities • business revenue • business for a superstore • weather forecasting 5. List various application of big data. 6. How are big data and Hadoop related to each other? 7. Explain the steps to be followed to deploy a Big Data solution.
DIFFERENCES	1. Write difference between MongoDB and Hadoop. 2. Write down the differences between Apache Pig and Map-Reduce. 3. Differentiate between: SQL and NoSQL 4. What are the differences between Hadoop 2 and Hadoop 3? 5. How is NFS different from HDFS? 6. What are the main differences between NAS (Network-attached storage) and HDFS? 7. What is the difference between "HDFS Block" and "Input Split"? 8. Explain the difference between Hadoop and RDBMS. 9. What is/are difference(s) between structured data, unstructured data and semi-structured data? Give the example of each and explain in brief. 10. Compare Distributed File System (DFS), Google File System (GFS) with Hadoop Distributed File System (HDFS)? Give various criteria for comparison also.

	11. Compare Cassandra with HBase and MongoDB. 12. Compare HDFS and HBase. 13. Compare RDBMS with Cassandra 14. Compare RDBMS with Neo4j 15. Differentiate between Apache Pig and Hive.
HADOOP	1. Explain core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail. 2. List various configuration files used in Hadoop Installation. What is use of mapred-site.xml? 3. What is Compute and Storage nodes in Hadoop? 4. Explain the term 'Commodity Hardware.' 5. What are the advantages of Hadoop? Draw Hadoop ecosystem and explain its components. 6. What is the purpose of the JPS command in Hadoop? 7. Name the different commands for starting up and shutting down Hadoop Daemons. 8. How Is Hadoop CLASSPATH essential to start or stop Hadoop daemons? 9. What are the common input formats in Hadoop? 10. Explain the different modes in which Hadoop run. 11. Explain the core components of Hadoop. What are the different configuration files in Hadoop? 12. How can you achieve security in Hadoop? 13. What do you understand by Rack Awareness in Hadoop? 14. What are the Port Numbers for NameNode, Task Tracker, and Job Tracker? 15. Why do we need Data Locality in Hadoop? Explain.

	16. DFS can handle a large volume of data then why do we need Hadoop framework? 17. What are Edge Nodes in Hadoop?
HDFS	1. Write down the goals of HDFS and explain hdfs architecture. 2. Define HDFS and YARN, and talk about their respective components. 3. Define HDFS. Describe namenode, datanode and block. Explain HDFS operations in detail. 4. Explain the process that overwrites the replication factors in HDFS. 5. What will happen with a NameNode that doesn't have any data? 6. Explain NameNode recovery process. 7. What is the Command to format the NameNode? 8. How to recover a NameNode when it is down? 9. What are the different file permissions in HDFS for files or directory levels? 10. Why is HDFS only suitable for large data sets and not the correct tool to use for many small files? 11. How does HDFS Index Data blocks? Explain. 12. Write the use and syntax of following HDFS commands: • put • expunge • chmod • get
MAP REDUCE	1. 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. 2. Explain Avro data serialization technique in Map-Reduce. 3. Explain the core methods of the Reducer?

	4. 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 5. What are the configuration parameters in a “MapReduce” program? 6. What is Distributed Cache in a MapReduce Framework? 7. What are the basic parameters of a Mapper? 8. What do you mean by heartbeat and replica in Hadoop? 9. Which are the problems related to Map-Reduce data storage? 10. Explain Job Scheduling in Map-Reduce.
APACHE PIG	1. What is Apache Pig and why do we need it? 2. Discuss how will Pig data model help in effective data flow 3. Draw architecture of APACHE PIG and explain in short.
DATABASES	1. Explain the working MongoDB with proper steps and diagram. 2. 3. How to create collection in MongoDB? Explain with its syntax. Explain CRUD operations of MongoDB with an example. Explain MongoDB sharding process. Which terms are used for table, row, column and table-join in MongoDB? 4. Compare Raw oriented and Column Oriented database structures. 5. What are primary and secondary replica sets in MongoDB? 6. What is the role of a profiler in MongoDB? Where does the writes all the data? 7. How to create collection in MongoDB? Explain with its syntax. 8. What is NoSQL? List out the features of NoSQL. Explain types of NoSQL databases in brief

	9. Explain following in brief with respect to Mongo DB : 1) Collections and documents 2) Indexing and retrieval 10. Explain scaling in MongoDB 11. Which terms are used for table, row, column and table-join in MongoDB?
SPARK	1. 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? 2. What is Resilient Distributed Dataset in Apache Spark? Explain in detail. Make a note on why RDD is better than Map Reduce data storage? 3. Justify: "SPARK is faster than MapReduce". 4. Write a brief short note on: Spark Unified Stack 5. Discuss Machine Learning with MLlib in SPARK
HBASE	1. Talk about the different tombstone markers used for deletion purposes in HBase. 2. What is HBase? Explain storage mechanism of HBase with an example. 3. Explain the working HBase with proper steps and diagram.
HIVE	1. Explain the architecture and features of HIVE. Explain Metastore in Hive. Explain various data insertion techniques in HIVE with example.
CASSANDRA	1. Explain the working Cassandra with proper steps and diagram. 2. Which are the features of BigTable and Titan InitGraph have been combined in Cassandra?
NEO4J	1. Explain the working Neo4j with proper steps and diagram.

ZOOKEEPER	1. What is Zookeper? List the benefits of it?
NEWSQL	1. Explain NewSQL. Explain the characteristics of NewSQL.