Objective Questions

1. Enlist various sources of data generation in today's era.
2. What is big data?
3. What is DFS? Enlist various services provided by DFS.
4. What is MapReduce?
5. _____ processing techniques processes data in chunks, whereas _____ processing techniques processes data in streams.
6. State true or false: "Clients should not know the number or locations of file servers and the storage devices refers to Structure transparency in distributed file system."
7. What is NoSQL?
8. Enlist various NoSQL databases available in market.
9. Define Collection in MongoDB.
10. State true or false: NoSQL database lacks ACID properties of transaction.
11. MongoDB is a _____ type of NoSQL database.
12. _____ method is used to display results in formatted way in MongoDB.

Introduction to Big data

Theory

1	What is big data? Explain characteristics of big data.	7
2	Write a short note on drivers (aspects that catalysed the growth of big data) of big data.	3
3	Explain four V's of big data.	3
4	Discuss and differentiate structured, unstructured and semi-structured data. Give proper examples.	3
5	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)	7
6	Discuss various types of analytics with proper example.	3
7	Explain working of following phases of Map Reduce with one common example. (i) Map Phase, (ii) Combiner Phase, (iii) Shuffle and Sort Phase, (iv) Reducer Phase	7
8.	Explain 4 'V's of big data with suitable example. Discuss how big data analytics can be useful in the development of smart transports.	7
9.		

Numerical/Derivable/Programs

1

Introduction to Hadoop and Hadoop Architecture

Theory

1	What are the advantages of Hadoop? Explain Hadoop architecture and its component with proper diagram. or Explain Hadoop architecture and its component with proper diagram.	7
2	Explain Job Scheduling in Map Reduce. How it is done in case of (i) The Fair Scheduler and (ii) The Capacity Scheduler	7
3	What is Hadoop Ecosystem? Discuss various components of Hadoop Ecosystem. or Write a short note on Hadoop Ecosystem.	7
4	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.	7
5	Explain Hadoop serialization with example.	3
6	Explain Hadoop data ingress and egress in brief. Enlist key design elements to be aware of when working with data ingress and egress.	3
7	Discuss role of JobTracker and TaskTracker in processing data with Hadoop.	3
8		

Numerical/Derivable/Programs

1	Write Map Reduce code for counting occurrences of specific words in the input text file(s). Also write the commands to compile and run the code.	7
2	What is MapReduce? Explain working of various phases of MapReduce with word count example.	4
3		

HDFS, Hive and HiveQL, HBase

Theory

1	Explain working of Hive with proper steps and diagram.	7
2	What is Zookeeper? List the benefits of it.	3
3	Differentiate: Apache pig Vs Map Reduce.	4
4	What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example.	7
5	Explain Metastore in Hive.	3
6	Explain Storage mechanism in HBase. Compare Raw oriented and Column Oriented database structures.	4
7	Draw HDFS Architecture. Explain any two commands of HDFS from following commands with syntax and at least one example of each. (i) copyFromLocal, (ii) setrep, (iii) checksum	7
8	With suitable block diagram explain architecture of HDFS. Discuss role of Data node and Name node in HDFS. Give commands with appropriate arguments to perform data transfer between local file system and HDFS.	7
9	Discuss role of Data node and Name node in HDFS.	3
	Explain detailed architecture of HDFS with block diagram. or With suitable block diagram explain architecture of HDFS.	7
	Give following commands for HDFS. a) Starting HDFS server b) Listing files in HDFS c) Inserting date into HDFS	3
	What is HiveQL? Explain various statements in HiveQL with example.	7
	What is HBase? Differentiate HBase and RDBMS.	4
	Write a short note on Apache Pig.	4

Numerical/Derivable/Programs

1	Give commands with appropriate arguments to perform data transfer between local file system and HDFS.	3

Apache Spark

Theory

1	Explain Spark components in detail. Also list the features of spark. or Write a short note on Spark stack. Give brief explanation of each component.	7
2	What are the problems related to Map Reduce data storage? How Apache Spark solves it using Resilient Distributed Dataset? Explain RDDs in detail. What is RDD? Explain role of RDD in Spark.	7 4
3	What is Apache Spark? What are the advantages of using Apache Spark over Hadoop? Explain in brief four major libraries of Apache Spark.	7
4	What is transformation and actions in Apache Spark? Discuss various commands available for this activities in Apache Spark?	7
5	Discuss how Spark is faster than Map Reduce.	3
6	What is transformation and actions in Spark? Explain with example. or What is transformation and actions in Spark? Explain with example.	4 3
7	Write a short note on Spark stack. Give brief explanation of each component.	7
8	Discuss various Machine learning algorithms available in Spark MLib.	4
9	Discuss limitations of Hadoop and how it is overcome in Apache Spark,	4
10		

Numerical/Derivable/Programs

1	Discuss Spark Streaming with suitable example such as analyzing tweets from Twitter.	7

NoSQL

Theory

1	Write a short note on NoSQL databases. List the differences between NoSQL and relational databases?	7
2	Differentiate between NoSQL and relational database. or Differentiate SQL and NoSQL databases. What are the applications of NoSQL database?	4
3	Describe applications of NoSQL databases in Industry. or Write a short note on use of NoSQL database in industry.	3
4	Write a short note on NoSQL databases. Explain various types in detail. or Discuss various types of NoSQL databases with example.	4
5	What is NoSQL database? Discuss key characteristics and advantages of NoSQL database.	4

Numerical/Derivable/Programs

1

Database for Modern Web

Theory

1	Explain Replication and scaling feature of MongoDB.	7
	What is MongoDB? Discuss important features of MongoDB.	3
2	Explain basic CRUD operations with example in MongoDB.	7
3	Explain following in brief with respect to MongoDB : • Collections and documents • Indexing and retrieval • Data aggregation	7
4	Discuss important features of MongoDB.	3
5	Define following terms in MongoDB. • Collection • Field • Document	3
6	Discuss model relationship between documents in schema design of MongoDB. Also give appropriate example.	7
	Explain database, collection, document and fields with respect to MongoDB. Also give its equivalent term in RDBMS.	3
	Explain use of aggregate function in MongoDB with suitable example.	4

Numerical/Derivable/Programs

1	Client need a database design for his blog with following specifications. • Every post has a unique title, description and url. • Every post can have one or more tags. • Every post has the name of its publisher and total number of likes. • Every post has comments given by users along with their name, message, data-time and likes. • On each post, there can be zero or more comments. For this set of requirements design a MongoDB schema. Justify your answer with proper detailing.	5
2	Requirement specification for a meeting dashboard application in an organization is as follows: • Any member in an organization can host a meeting and send an invitations to other members within an organization. • Invitees can accept or reject the meeting with proper reason. • Every meeting has the title, timestamp and place/location associated. • Every meeting has predefined agendas and documents associated. • Meeting discussion concludes with identifying tasks to accomplish. • Every task has title, priority, deadline and note associated with it. Task can be assigned to any attendee of meeting.	

For this set of requirements design a MongoDB schema.

3 A typical course feedback system functions as per following features: 7

- Course management.
- Subject management for course.
- Faculty subject engagement.
- Student registration for course.
- Student feedbacks for faculty for subject.

Design MongoDB schema for above application.

(Necessary assumptions could be made for detailed design.)