Chapter 1

1. Define Big Data. List and discuss the 5V's of Big Data. (5 Marks)
2. Define Big Data. List and discuss types of Big Data. (5 Marks)
3. Compare Traditional Data vs Big Data. (5 Marks)
4. Define Big Data Analytics. List and explain different types of Big Data Analytics. (5 Marks)
5. Explain the desired properties of a Big Data System. (5 Marks)
6. List and explain the ethical and privacy issues in Big Data. (5 Marks)
7. List and explain various applications of Big Data. (5 Marks)

Chapter 2

1. List and explain the goals of Hadoop. (5 Marks)
2. Explain the core components of Hadoop. (5 Marks)
3. Explain HDFS architecture in detail. (5 Marks)
4. Explain MapReduce architecture in brief. (5 Marks)
5. Apply MapReduce on the following document to count the frequency of words. Show all the phases properly. (5 Marks)

Welcome to Hadoop class Hadoop is good Hadoop is bad

Dog Cat Mouse Dog Dog Cat Dog Cat Duck

Bus Car Train Train Plane Car Bus Bus Plane

Deer Bear River Car Car River Deer Car Bear

The quick brown fox The fox ate the mouse Now how brown cow

6. Write the algorithm for Matrix Multiplication using MapReduce. Perform matrix multiplication using the MapReduce algorithm. (10 Marks)

$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ B = $\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 4 \end{bmatrix}$ $B = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}$

$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ $B = \begin{bmatrix} 6 & 3 \\ 5 & 2 \\ 4 & 1 \end{bmatrix}$

$$A = \begin{bmatrix} 10 & 0 & 20 \\ 0 & 30 & 40 \\ 50 & 60 & 70 \end{bmatrix} \quad B = \begin{bmatrix} -1 & 0 \\ -2 & -3 \\ 0 & -4 \end{bmatrix}$$

7. Explain Hadoop Ecosystem in detail with neat diagram. (10 Marks)
8. Explain the Physical Architecture of Hadoop and also explain its working. (10 Marks)
9. State the advantages and limitations of Hadoop. (5 Marks)
10. Explain MapReduce Execution Pipeline in detail with neat diagram. (10 Marks)
11. Explain different Input and Output Formats in MapReduce. (10 Marks)
12. Write the HDFS commands for the following with suitable example: (5 Marks)
    i. To display recursively the contents in the directory.
    ii. To copy files/folders from local file system to HDFS store.
    iii. To copy files/folders from HDFS store to local file system.
    iv. To give the size of each file in directory.
    v. To change the replication factor of a file/directory in HDFS.

Chapter 3

1. State the features of Hive. Also, explain the architecture of Hive. (10 Marks)
2. Compare Hive Vs Traditional Database. (5 Marks)
3. Consider the following database table: (10 Marks)

Id	Name	Branch	Mobile
101	Darsheel	Computer	9876502314
102	Ayush	Mechanical	9932416578
103	Yash	Computer	9810324576

    i. Create the following table in Hive with transactional property = true and partitions.
    ii. Insert the following values in the table.
    iii. Display the count of students with respect to branch.
    iv. Update the name to “Yashvi” where ID =103.
    v. Alter the table to add new column “CGPA” with float datatype.
4. Explain different Joins in Hive with suitable example. (5 Marks)
5. Explain the concept of Views in Hive with suitable example. (5 Marks)
6. Compare and contrast HBase with RDBMS. (5 Marks)
7. Explain HBase Architecture with neat diagram. (10 Marks)
8. Explain the concepts in HBase Schema with suitable example. (5 Marks)
9. Write the HBase Commands for the following: (5 Marks)
    a. Create a table with name Product.
    b. Add column family ‘Shoe’ and ‘Tshirt’.
    c. In column family Shoe, add the columns as below:
    Brand = Nike, Price = $50, Size = 9, Description = For Men
    d. Display the contents of table ‘Product’
    e. Check whether the table ‘Product’ is enabled or disabled.
10. State the features and applications of Pig. (5 Marks)
11. Explain the execution modes of Pig. (5 Marks)
12. Compare Apache Pig vs MapReduce. (5 Marks)
13. Explain the significance of Pig Grunt and Pig Latin. (5 Marks)
14. Explain Pig Architecture. Also explain different Pig Data Models. (10 Marks)
15. State the features of Zookeeper. Also explain the architecture of Zookeeper. (10 Marks)

16. Explain the working of Zookeeper. Also state the benefits of Zookeeper. (10 Marks)

Chapter 4

1. State the features of Spark. Also explain the architecture of Spark. (10 Marks)
2. Explain in detail the components of Spark. (10 Marks)
3. Compare Apache Hadoop Vs Apache Spark. (5 Marks)
4. Explain RDD with its features, advantages and imitations. (5 Marks)
5. Explain Spark SQL with its features, advantages and disadvantages. (5 Marks)
6. Explain different types of schedulers supported by Apache Spark. (5 Marks)
7. Write a short note on: Shared Variables in Apache Spark. (5 Marks)

Chapter 5

1. What is NoSQL? Explain the features of NoSQL. (5 Marks)
2. Compare SQL Vs NoSQL. (5 Marks)
3. Explain the business drivers of NoSQL. (5 Marks)
4. Explain in detail different architectural patterns of NoSQL. (10 Marks)
5. What is MongoDB? State the features of MongoDB. (10 Marks)
6. Compare MongoDB Vs RDBMS. (5 Marks)
7. Explain MongoDB architecture with neat diagram. (5 Marks)
8. Consider the following collection "Movie": (10 Marks)

ID	Title	Director	Year	Budget
1	12 Angry Men	Sidney Lumet	1957	350000
2	Shutter Island	Martin Scorsese	2010	80000000
3	Interstellar	Christopher Nolan	2014	1650000000

Write MongoDB queries for the following:

(a) Create a collection named "Movie".
(b) Insert all the documents as shown in the collection in a single pass.
(c) Insert new document with the following values {4, "Arrival", "Denis Villeneuve", 2014}
(d) Update the Movie "Arrival" and set Budget = 47000000.
(e) Retrieve all the movies released after year 2000.
(f) Find the movies which were released before 2015 and have a *budget* greater than 50 million.
(g) Find the sum of the *budgets* of the movies released after 2000.
(h) Delete the document having the movie name as "Interstellar".
(i) Remove all documents in the collection.

9. Consider the following collection "Employee": (10 Marks)

ID	Name	Role	Year	Salary
101	Sahil	Programmer	2001	250000
102	Vansh	Developer	2005	180000
103	Sayali	Designer	2008	200000

Write MongoDB queries for the following:

(a) Create a collection named "Employee".
(b) Insert all the documents as shown in the collection in a single pass.
(c) Insert new document with the following values {104, "Shubh", "Engineer", 2014}

(d) Update the Employee “Shubh” and set Salary = 150000.
(e) Retrieve all the employees joined after year 2003.
(f) Find the employees which have joined before 2015 and have a salary greater than 190000.
(g) Find the sum of the salaries of the employees joined after 2002.
(h) Delete the document having the role as “Developer”.
(i) Remove all documents in the collection.

Chapter 6

1. Define Data Streaming. Describe the characteristics of streaming data. Also give examples of data stream. (5 Marks)
2. What is Apache Kafka? Explain the components of Apache Kafka with a neat diagram. (5 Marks)
3. Explain the Cluster Architecture of Apache Kafka. Also explain the workflow of Kafka. (10 Marks)
4. Compare Hadoop Vs Kafka. (5 Marks)
5. Compare Spark Vs Kafka. (5 Marks)
6. What is Apache Storm? Explain the components of Apache Storm with a neat diagram. (5 Marks)
7. Explain the Cluster Architecture of Apache Storm. Also explain the workflow of Storm. (10 Marks)
8. Compare Hadoop Vs Storm. (5 Marks)
9. Explain Spark Streaming with its architecture. (5 Marks)
10. Give the application use cases of Apache Kafka and Apache Storm. (5 Marks)