GUJARAT TECHNOLOGICAL UNIVERSITY

BE 7 SEMESTER – OLD PAPER – S22 TO S26 – QUESTION BANK

Subject Name & Code: Big Data Analytics - 3170722

Chapter 1: Introduction to Big Data

Repeated Questions:

1. Big Data and Its Sources:

- **Q.** What is Big data? Explain sources of big data. **(WINTER 2022, 3 Marks; SUMMER 2026, 3 Marks)**

2. V's of Big Data:

- **Q.** Explain 4 V's of Big data. / What is big data? Explain four V's of big data. **(SUMMER 2022, 4 Marks; WINTER 2023, 4 Marks; SUMMER 2024, 7 Marks; SUMMER 2026, 4 Marks)**
- **Q.** Explain 5V's of Big Data. / What is Big Data? Explain the five V's of Big Data. **(WINTER 2022, 3 Marks OR; SUMMER 2025, 3 Marks)**

3. Big Data Analytics: Applications, Advantages and Disadvantages:

- **Q.** What is Big data Analytics? List applications of it. / List various application of big data. **(SUMMER 2022, 3 Marks; SUMMER 2023, 3 Marks)**
- **Q.** Explain advantages and disadvantages of big data analytics. **(SUMMER 2023, 3 Marks; WINTER 2025, 4 Marks)**

4. Structured, Unstructured and Semi-structured Data:

- **Q.** Explain the difference between structured and unstructured data. / Structured, unstructured and semi-structured data differences. / Discuss structured, unstructured and semi-structured big data types. Give proper examples. / Discuss the key differences between structured, unstructured, and semi-structured data with examples. / How is Semi-structured data different from Unstructured and Structured Data? **(SUMMER 2023, 4 Marks; SUMMER 2023, 4 Marks; WINTER 2023, 7 Marks; SUMMER 2025, 4 Marks; WINTER 2025, 3 Marks)**

5. Traditional Data vs Big Data:

- **Q.** Give difference between traditional data and big data. / Describe Traditional vs. Big Data business approach. **(SUMMER 2024, 3 Marks; SUMMER 2026, 4 Marks)**

6. E-Commerce and Big Data:

- **Q.** Discuss about How E-Commerce is Using Big Data to Improve Business in detail. **(WINTER 2022, 7 Marks; SUMMER 2026, 7 Marks)**

Other Important Questions:

1. Definition, Characteristics and Types:

- **Q.** Define Big Data and also Explain Characteristic of Big Data. **(SUMMER 2023, 7 Marks)**
- **Q.** What is big data? Discuss it in terms of volume and velocity, veracity. **(WINTER 2025, 4 Marks)**
- **Q.** Explain types of Big data. **(WINTER 2022, 4 Marks)**

2. Analytics, Applications and Case Studies:

- **Q.** What is Big data Analytics? Explain any one case study of it. **(SUMMER 2024, 4 Marks)**
- **Q.** Discuss Big Data in Healthcare, Transportation & Medicine. **(SUMMER 2023, 7 Marks)**
- **Q.** Why is big data analytics important? **(WINTER 2022, 3 Marks)**
- **Q.** Explain advantages of Big Data Processing. **(WINTER 2022, 3 Marks OR)**

3. Challenges:

- **Q.** Explain different challenges of Big Data. **(WINTER 2022, 4 Marks)**
- **Q.** Explain various challenges of conventional system. **(WINTER 2023, 3 Marks)**

Chapter 2: Hadoop

Repeated Questions:

1. Hadoop and Its Core Components:

- **Q.** What is Hadoop? Briefly explain the core components of it. **(SUMMER 2022, 7 Marks; SUMMER 2024, 7 Marks)**

2. Hadoop Architecture:

- **Q.** Write a short note on Hadoop Architecture. **(WINTER 2022, 7 Marks; SUMMER 2026, 7 Marks)**
- **Q.** Explain Hadoop's architecture in detail with a diagram. How does the NameNode and DataNode work together in Hadoop? / Explain Hadoop architecture and its component with proper diagram. [Diagram required – refer to the original question paper] **(SUMMER 2025, 7 Marks; WINTER 2025, 7 Marks)**

3. HDFS:

- **Q.** Write a short note on HDFS (Hadoop Distributed File System). / Write a detailed note on HDFS and explain its significance in Big Data processing. **(WINTER 2022, 7 Marks; SUMMER 2025, 7 Marks; SUMMER 2026, 7 Marks)**
- **Q.** How HDFS is different from traditional NFS? **(SUMMER 2022, 4 Marks; SUMMER 2024, 4 Marks)**
- **Q.** Discuss role of Data node and Name node in HDFS. / Explain name node, data node and secondary node. **(SUMMER 2023, 4 Marks; WINTER 2023, 3 Marks; SUMMER 2024, 3 Marks; WINTER 2025, 4 Marks)**
- **Q.** Explain following commands of HDFS with syntax and at least one example of each. (i) copyFromLocal (ii) mv (iii) cat / Explain any three commands of HDFS with syntax along with example of each. **(WINTER 2023, 3 Marks; WINTER 2025, 3 Marks)**

4. MapReduce:

- **Q.** What is MapReduce? Explain working of various phases of MapReduce with appropriate example and diagram. / 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 / Map Reduce phases. [Diagram required – refer to the original question paper] **(SUMMER 2022, 7 Marks OR; WINTER 2023, 7 Marks OR; SUMMER 2024, 7 Marks OR; WINTER 2025, 7 Marks)**

5. Job Scheduling and YARN:

- **Q.** What do you mean by job scheduling in Hadoop? List different schedulers in Hadoop. **(SUMMER 2022, 3 Marks; SUMMER 2024, 3 Marks OR)**
- **Q.** Discuss Hadoop YARN in detail with failures in classic MapReduce. **(SUMMER 2023, 7 Marks OR; SUMMER 2026, 7 Marks OR)**

6. Hadoop Installation:

- **Q.** List various configuration files used in Hadoop Installation. **(SUMMER 2023, 3 Marks; SUMMER 2024, 3 Marks)**

Other Important Questions:

1. Hadoop Basics:

- **Q.** How is Big data and Hadoop related? **(SUMMER 2022, 3 Marks)**
- **Q.** Explain modules of Hadoop. **(WINTER 2022, 7 Marks)**
- **Q.** Write a short note on Hadoop ecosystem. **(WINTER 2023, 7 Marks)**
- **Q.** Define Hadoop. List and explain advantages of Hadoop. **(WINTER 2023, 4 Marks)**
- **Q.** Explain basic Components of Analyzing the Data with Hadoop. **(SUMMER 2026, 3 Marks)**

2. HDFS:

- **Q.** Draw and Explain HDFS architecture. How can you restart NameNode and all the daemons in Hadoop? [Diagram required – refer to the original question paper] **(SUMMER 2022, 7 Marks)**
- **Q.** How does HDFS ensure data Integrity in a Hadoop Cluster? **(WINTER 2025, 7 Marks OR)**
- **Q.** Discuss the concept of Data Locality in Hadoop. Why is it significant? **(SUMMER 2025, 4 Marks)**

3. MapReduce:

- **Q.** Explain usage of Map reduce. **(WINTER 2022, 3 Marks)**
- **Q.** Explain methods of Map reduce of Mapper class. **(WINTER 2022, 4 Marks)**
- **Q.** What is a MapReduce function? Explain the purpose of the Map and Reduce phases. **(SUMMER 2025, 3 Marks)**
- **Q.** Discuss how MapReduce handles failures during the execution of jobs. **(SUMMER 2025, 7 Marks OR)**

Chapter 3: NoSQL

Repeated Questions:

1. NoSQL Basics:

- **Q.** Explain NoSQL business drivers. **(WINTER 2023, 4 Marks; WINTER 2025, 3 Marks)**
- **Q.** Explain features and advantages of NoSQL. / What is NoSQL database? Discuss key characteristics and advantages of NoSQL database. **(WINTER 2022, 4 Marks; SUMMER 2024, 4 Marks; WINTER 2025, 7 Marks)**
- **Q.** Give difference between SQL and NoSQL. / Differences between NoSQL and relational databases. / Define NoSQL databases and explain how they differ from relational databases. / Differentiate SQL and NoSQL databases. What are the applications of NoSQL database? **(WINTER 2022, 3 Marks; SUMMER 2023, 4 Marks OR; WINTER 2023, 3 Marks; SUMMER 2024, 4 Marks OR; SUMMER 2025, 3 Marks; WINTER 2025, 7 Marks OR)**
- **Q.** Explain different types of NoSQL. / Elaborate Key-value stores, Graph stores, Column family stores & Document stores. **(SUMMER 2022, 7 Marks OR - includes types; WINTER 2022, 7 Marks; WINTER 2023, 7 Marks; SUMMER 2025, 7 Marks; SUMMER 2026, 4 Marks)**
- **Q.** Discuss four ways that NoSQL system handles big data problems. **(WINTER 2023, 7 Marks; WINTER 2025, 3 Marks OR; SUMMER 2026, 7 Marks OR)**
- **Q.** Give difference between master-slave model and peer-to-peer model. **(WINTER 2022, 3 Marks OR; SUMMER 2023, 3 Marks; SUMMER 2026, 3 Marks)**

2. MongoDB:

- **Q.** How MongoDB is better than SQL database? **(SUMMER 2022, 3 Marks OR; SUMMER 2026, 3 Marks OR)**
- **Q.** What are the features of MongoDB? **(SUMMER 2022, 4 Marks; SUMMER 2022, 3 Marks OR)**
- **Q.** Explain CRUD operations in MongoDB. **(SUMMER 2022, 7 Marks; SUMMER 2026, 7 Marks)**
- **Q.** Explain Replication and scaling feature of MongoDB. **(SUMMER 2022, 4 Marks OR; SUMMER 2026, 4 Marks OR)**
- **Q.** Write difference between MongoDB and Hadoop. **(SUMMER 2022, 4 Marks; SUMMER 2023, 4 Marks)**

Other Important Questions:

1. NoSQL Concepts:

- **Q.** Compare Row oriented and Column Oriented database structures. **(SUMMER 2023, 3 Marks OR)**
- **Q.** Define NoSQL and where is it used? (b) i) Document Oriented Database ii) Graph based Database. **(SUMMER 2023, 7 Marks)**
- **Q.** Explain NoSQL base property. **(WINTER 2023, 3 Marks OR)**
- **Q.** Explain the Document store model in NoSQL databases. **(SUMMER 2025, 4 Marks)**
- **Q.** Describe how NoSQL databases handle scalability and high availability. **(SUMMER 2025, 3 Marks OR)**
- **Q.** What is Sharding in NoSQL databases? How does it help in handling Big Data? **(SUMMER 2025, 4 Marks OR)**

Chapter 4: Mining Data Stream

Repeated Questions:

1. Stream Basics:

- **Q.** What is Stream Computing and Sampling Data in a Stream? **(SUMMER 2023, 4 Marks; WINTER 2025, 4 Marks OR; SUMMER 2026, 4 Marks)**
- **Q.** Discuss various benefits of stream processing. **(WINTER 2023, 3 Marks OR; WINTER 2025, 4 Marks)**
- **Q.** Write a short note on RTAP. **(SUMMER 2023, 4 Marks OR; WINTER 2025, 3 Marks OR)**

2. Sampling and Windows:

- **Q.** How samples are collected from data stream? **(WINTER 2023, 4 Marks OR; WINTER 2025, 4 Marks)**
- **Q.** Write a short note on Decaying Window. / Explain Counting Ones in a Window and Decaying Window of mining data stream in detail. **(SUMMER 2023, 3 Marks OR; SUMMER 2024, 7 Marks OR; WINTER 2025, 3 Marks; SUMMER 2026, 4 Marks OR)**

3. Stock Market Prediction:

- **Q.** What is Stock Market Prediction? / Write a short note: Stock Market Predictions. / Data analysis in stock market and weather forecasting predictions. **(WINTER 2022, 3 Marks; WINTER 2023, 4 Marks; SUMMER 2024, 7 Marks)**

Other Important Questions:

1. Stream Mining:

- **Q.** What is data stream mining? Explain stream data model and its architecture in detail. **(SUMMER 2024, 7 Marks)**
- **Q.** Explain filtering a stream in detail. **(WINTER 2022, 7 Marks OR)**
- **Q.** Give difference between DBMS and DSMS (Data Stream Management System). **(WINTER 2022, 4 Marks OR)**
- **Q.** Explain Bloom's Filter with the help of example. **(WINTER 2023, 7 Marks)**
- **Q.** Discuss Flajolet-Martin Algorithm to count Distinct elements in a stream with proper example. **(WINTER 2023, 7 Marks OR)**

2. Social Networks:

- **Q.** Define Social networks and Social Network Mining. **(WINTER 2022, 7 Marks OR)**

Chapter 5: Frameworks

Repeated Questions:

1. HBase:

- **Q.** Explain the architecture of HBase. / Explain briefly on HBase architecture with neat diagram. [Diagram required – refer to the original question paper] **(SUMMER 2022, 7 Marks; SUMMER 2024, 7 Marks OR; SUMMER 2026, 7 Marks)**
- **Q.** Explain the Data Model of HBase. **(SUMMER 2022, 4 Marks OR; SUMMER 2024, 3 Marks)**

2. Zookeeper:

- **Q.** What is Zookeeper? / What is Zookeeper? List the benefits of it. / What are the benefits of using a zookeeper? **(SUMMER 2022, 3 Marks OR; WINTER 2022, 7 Marks OR; SUMMER 2023, 7 Marks OR; WINTER 2025, 4 Marks OR; SUMMER 2026, 3 Marks OR)**

3. Pig:

- **Q.** What is Pig? **(SUMMER 2022, 3 Marks; SUMMER 2026, 3 Marks)**
- **Q.** Explain Pig data Model in detail and Discuss how it will help for effective data flow. **(SUMMER 2022, 7 Marks OR; SUMMER 2024, 7 Marks)**
- **Q.** Describe data processing operators in Pig. **(SUMMER 2023, 3 Marks OR; SUMMER 2026, 3 Marks OR)**

4. Hive:

- **Q.** Explain working of Hive with proper steps and diagram. [Diagram required – refer to the original question paper] **(SUMMER 2023, 7 Marks; WINTER 2023, 7 Marks; WINTER 2025, 7 Marks)**
- **Q.** Explain the HiveQL-Select-Order By with suitable example. **(SUMMER 2023, 7 Marks OR; SUMMER 2024, 3 Marks OR; SUMMER 2026, 7 Marks OR)**
- **Q.** Explain HIVE data types in brief. **(WINTER 2022, 4 Marks; SUMMER 2026, 4 Marks)**

5. Comparisons: Hive, HBase and Pig:

- **Q.** Differentiate between HIVE and HBASE. **(SUMMER 2022, 4 Marks OR; SUMMER 2026, 4 Marks OR)**
- **Q.** Differentiate between Pig vs Hive. **(WINTER 2023, 4 Marks; WINTER 2025, 4 Marks)**

Other Important Questions:

1. HBase:

- **Q.** What are WAL, MemStore, HFile and HLog in HBase? **(SUMMER 2022, 4 Marks)**
- **Q.** Explain the concept of regions in HBase and storing Big data with HBase. **(SUMMER 2022, 7 Marks)**
- **Q.** What is HBase and Hive? **(WINTER 2022, 4 Marks)**
- **Q.** Give list of HBase Commands. **(WINTER 2022, 4 Marks OR)**
- **Q.** How HBase uses Zookeeper to Build Applications? Explain in detail. **(SUMMER 2024, 4 Marks)**
- **Q.** Explain Storage mechanism in HBase. **(WINTER 2025, 3 Marks OR)**

2. Pig and Hive:

- **Q.** Differentiate: Apache Pig Vs Map Reduce. **(SUMMER 2023, 4 Marks OR)**
- **Q.** Define join and explain types of join. **(SUMMER 2023, 7 Marks OR)**
- **Q.** Compare Hive and Pig query language. **(SUMMER 2024, 4 Marks OR)**
- **Q.** Explain any three HiveQL DDL command with its syntax and example. **(WINTER 2023, 3 Marks OR)**

- **Q.** What is HiveQL? Explain various statements in HiveQL with example. **(WINTER 2025, 7 Marks OR)**

Chapter 6: Spark

Repeated Questions:

1. Spark Overview:

- **Q.** Write a short note on Spark. **(SUMMER 2022, 3 Marks; SUMMER 2026, 3 Marks)**
- **Q.** What is Apache Spark? What are the advantages of using Apache Spark over Hadoop? Explain in brief four major libraries of Apache Spark. **(SUMMER 2023, 7 Marks; WINTER 2025, 7 Marks OR)**
- **Q.** Describe important features of Apache Spark. / What are the components of spark? Also state the features of spark. / Explain Spark components in detail. Also list the features of Spark. **(WINTER 2023, 3 Marks; SUMMER 2024, 3 Marks OR; WINTER 2025, 7 Marks)**
- **Q.** Explain Spark Architecture. **(WINTER 2022, 7 Marks OR; SUMMER 2026, 7 Marks OR)**

2. RDD:

- **Q.** What is RDD? Explain role of RDD in Spark. **(WINTER 2023, 4 Marks OR; WINTER 2025, 3 Marks)**
- **Q.** What is RDD? State and Explain RDD operations. **(SUMMER 2022, 7 Marks OR; SUMMER 2024, 7 Marks OR)**

3. Transformations and Actions:

- **Q.** What is transformation and actions in Apache Spark? Discuss various commands available for this activity in Apache Spark? / Define various Actions in Spark. **(WINTER 2023, 7 Marks; SUMMER 2024, 4 Marks)**

Other Important Questions:

1. Spark Applications and Programs:

- **Q.** Write application of writing Spark. **(SUMMER 2023, 3 Marks)**
- **Q.** Write basic wordcount program using PySpark. **(WINTER 2023, 4 Marks)**
- **Q.** Justify spark is faster than MapReduce. **(SUMMER 2024, 3 Marks)**
- **Q.** What does a Spark Engine do? What is GraphX? **(SUMMER 2024, 4 Marks OR)**

****