SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar, Kattankulathur – 603 203

DEPARTMENT OF COMPUTER APPLICATIONS

QUESTION BANK

II SEMESTER

PMC304- BIG DATA ANALYTICS

Regulation – 2024

Academic Year 2024 – 25 (Even)

Prepared by

Mr. M. Nagarajan, Assistant Professor (O.G)/MCA

SRM VALLIAMMAI ENGINEERING COLLEGE

SRM Nagar,Kattankulathur- 603 203

DEPARTMENT OF COMPUTER APPLICATIONS

Academic Year 2025-2026

QUESTION BANK- ODD SEMESTER

SUBJECT : PMC304 - BIG DATA ANALYTICS
YEAR / SEM: I / II SEM MCA

UNIT-1 UNDERSTANDING BIG DATA

Introduction to big data – convergence of key trends – unstructured data – industry Examples of big data – web analytics – big data applications – big data technologies – introduction to Hadoop – open source technologies – cloud and big data – mobile business intelligence – Crowd sourcing analytics inter and trans firewall analytics.

PART – A

Q. No	Question	Level	Competence
1	**Define** Big Data?	BTL 1	Remember
2	**Compare** Big Data and Conventional Data.	BTL 2	Understand
3	**List** the various dimensions of growth of Big Data.	BTL 1	Remember
4	**What** is unstructured data?	BTL 1	Remember
5	**Define** web analytics.	BTL 1	Remember
6	**List** any two key trends that led to the emergence of Big Data	BTL 1	Remember
7	**Identify** two examples of unstructured data.	BTL 1	Remember
8	**Describe** the term "convergence of key trends" in Big Data.	BTL 2	Understand
9	**Name** any two Big Data technologies.	BTL 1	Remember
10	**Describe** the role of web analytics in Big Data applications.	BTL 2	Understand
11	**State** the purpose of Hadoop.	BTL 1	Remember
12	**Compare** between structured and unstructured data.	BTL 2	Understand
13	**State** how cloud computing is used to manage Big Data.	BTL 1	Remember
14	**Outline** the benefits of using open-source technologies in Big Data.	BTL 2	Understand
15	**Describe** the usage of mobile business.	BTL 2	Understand
16	**Describe** the different applications of Big Data	BTL 2	Understand
17	**Describe** the concept of crowdsourcing analytics.	BTL 2	Understand
18	**Classify** the role of cloud computing Big Data application	BTL 2	Understand
19	**Identify** scenarios where inter-firewall analytics is essential.	BTL 3	Apply
20	**What** is the use of mobile business intelligence in decision-making.	BTL 1	Analyze

21	**Define** Cloud and Big Data Analytics.		BTL 1	Remember
22	**Describe** about HDFS		BTL 2	Understand
23	**State** the function of MapReduce		BTL 1	Remember
24	**Lis**t the sources of Big Data.		BTL 1	Remember
	PART-B			
Q.No.	**Question**	**Marks**	**Level**	**Competence**
1	**Develop** the evolution of Big Data and discuss the convergence of key trends that have made it indispensable in today's industries.	16	BTL 3	Apply
2	**Examine** the characteristics of Big Data and the key trends that have contributed to its evolution	16	BTL 4	Analyze
3(a)	**a. Analyze** the differences between structured and unstructured data with examples.	8	BTL 4	Analyze
3(b)	**Examine** a solution to manage unstructured data in a social media platform.	8	BTL 4	Analyze
4	**Compare** traditional database systems and Big Data technologies in terms of scalability, cost, and efficiency	16	BTL 4	Analyze
5(a)	**Identify** the role of web analytics	8	BTL 3	Apply
5(b)	**Apply** the effectiveness of web analytics tools	8	BTL 3	Apply
6	**Develop** the Industry examples of Big Data	16	BTL 3	Apply
7	**Identify** how Big Data is used in the healthcare industry.	16	BTL 3	Apply
8	**Analyse** the role of unstructured data in Big Data processing and discuss its impact on decision-making in organisations.	16	BTL 4	Analyze
9(a)	**Construct** about Big Data	8	BTL 3	Apply
9(b)	**Compare** the features of Cloud and Big Data	8	BTL 4	Analyze
10(a)	**Analyze** Applications of Big Data in Social and Affiliate Marketing	8	BTL 4	Analyze
10(b)	**Identify patterns** of Fraud Detection with Near Real-Time Event Processing Framework Diagram	8	BTL 4	Analyze
11	**Identify** the implication of crowdsourcing analytics with real-world examples	16	BTL 3	Apply
12(a)	**Analyze** the key components of Hadoop with diagrammatic representation of the Hadoop Distributed File System (HDFS)	8	BTL 4	Analyze
12(b)	**Construct** about MapReduce	8	BTL 3	Apply
13	**Build** open-source technologies in Big Data and describe their functionalities.	16	BTL 3	Apply
14	**Analyze** how cloud platforms enable scalability and flexibility in Big Data solutions	16	BTL 4	Analyze
15	**Analyse** the benefits and limitations of crowdsourcing in data collection and analysis.	16)	BTL 4	Analyze

16(a)	**Analyze** mobile business intelligence	8	BTL 4	Analyze
16(b)	**Discover** key features of mobile business intelligence	8	BTL 4	Analyze
17(a)	**Identify** the significance of inter-firewall analytics	8	BTL 3	Apply
17(b)	**Choose** the Advantages of inter-firewall and trans-firewall	8	BTL 3	Apply

UNIT-2 NOSQL DATA MANAGEMENT

Introduction to NoSQL – Aggregate data models – Key-value and document data models – Relationships graph databases – Schemeless databases – Materialized views – Distribution models – Master-Slave Replication – Consistency – Cassandra – Cassandra data model – Cassandra examples – Cassandra clients.

PART – A

Q. No	Question	Level	Competence
1	**List** the primary characteristics of NoSQL databases	BTL 1	Remember
2	**Define** NoSQL and its significance in modern databases	BTL 1	Remember
3	**Identify** two advantages of NoSQL over relational databases	BTL 2	Understand
4	**Compare** SQL and NoSQL databases.	BTL 2	Understand
5	**Describe** aggregate data models in NoSQL	BTL 2	Understand
6	**Give** examples of two aggregate data models.	BTL 1	Remember
7	**Identify** the advantages of using aggregate data models.	BTL 2	Understand
8	**Define** the key-value data model.	BTL 1	Remember
9	**Describe** the document data model in NoSQL.	BTL 2	Understand
10	**Compare** between key-value and document data models.	BTL 2	Understand
11	**Describe** graph databases.	BTL 2	Understand
12	**Define** a scenario where graph databases are ideal for use.	BTL 1	Remember
13	**Define** a schemeless database**.**	BTL 1	Remember
14	**State** two benefits of schemeless databases.	BTL 1	Remember
15	**Define** materialized views.	BTL 1	Remember
16	**Describe** two advantages of using materialized views.	BTL 2	Understand
17	**Compare** materialized views with regular database views.	BTL 2	Understand
18	**Define** the term aggregate data models in NoSQL databases.	BTL 1	Remember
19	**Classify** the three types of aggregate data models used in NoSQL	BTL 2	Remember
20	**Describe** master-slave replication.	BTL 2	Understand
21	**What** are the two advantages of master-slave replication	BTL 1	Remember
22	**Describe** the Cassandra data model.	BTL 2	Understand
23	**What is** the significance of a Cassandra data model	BTL 1	Remember

24	**Show** the use of Cassandra for large-scale applications.		BTL 1	Remember
PART-B				
Q.No.	**Question**	**Marks**	**Level**	**Competence**
1	**Examine** the characteristics and significance of NoSQL databases	16	BTL 4	Analyze
2(a)	**Illustrate** the differences between relational databases and NoSQL databases with examples.	8	BTL 3	Apply
2(b)	**Analyze** the role of NoSQL in handling unstructured and semi-structured data.	8	BTL 4	Analyze
3	**Organize** the NoSQL databases and explain their core features**.**	16	BTL 3	Apply
4	**Analyze** the aggregate data models and explain their relevance in NoSQL systems	16	BTL 4	Analyze
5	**Illustrate** the implementation of aggregate data models with real-world examples.	16	BTL 3	Apply
6(a)	**Identify** the key-value data model.	8	BTL 3	Apply
6(b)	**Organize** the document data model.	8	BTL 3	Apply
7(a)	**Compare** the key-value and document data models	8	BTL 4	Analyze
7(b)	**Examine** the graph data model.	8	BTL 4	Analyze
8	**Examine** how graph databases manage complex relationships with examples.	16	BTL 4	Analyze
9(a)	**Analyse** the use of graph databases in social media applications	10	BTL 3	Apply
9(b)	**Evaluate** the advantages and limitations of graph databases	6	BTL 4	Analyze
10	**Classify** the concept of master-slave replication in NoSQL databases**.**	16	BTL 4	Analyze
11	**Investigate** the schemeless databases and state their key benefits.	16	BTL 4	Analyze
12(a)	**Examine** how schemeless databases adapt to rapidly changing data structures	8	BTL 4	Analyze
12(b)	**Compare** master-slave replication with peer-to-peer replication models.	8	BTL 4	Analyze
13(a)	**Compare** about Consistency and Read Consistency	8	BTL 4	Analyze
13(b)	**Examine** Replication consistency with an Example Diagram	8	BTL 4	Analyze
14	**Classify** materialized views and analyze their role in query performance optimization.	16	BTL 4	Analyze

15	Illustrate the Cassandra architecture and list the features of Cassandra.	16	BTL 3	Apply
16(a)	**Analyse** the advantages of Cassandra.	6	BTL 3	Apply
16(b)	**Investigate** the impact of the CAP theorem in NoSQL databases.	10	BTL 4	Analyze
17	**Examine** the scalability and fault tolerance of Cassandra in large-scale applications.	16	BTL 4	Analyze

UNIT-3 MAP REDUCE APPLICATIONS

MapReduce workflows – unit tests with MRUnit – test data and local tests – anatomy of Map Reduce job run – classic Map-reduce – YARN – failures in classic Map-reduce and YARN – job scheduling – shuffle and sort – task execution – MapReduce types – input formats – output formats.

PART – A

Q. No	Question	Level	Competence
1	**Define** MapReduce.	BTL 1	Remember
2	**Infer** the primary phases of a MapReduce workflow.	BTL 2	Understand
3	**Explain** the role of the Mapper function in a MapReduce.	BTL 2	Understand
4	**Identify** the role of the Reducer function in MapReduce.	BTL 1	Remember
5	**Define** MRUnit and its purpose.	BTL 1	Remember
6	**List** the benefits of using MRUnit for unit testing MapReduce jobs	BTL 1	Remember
7	**Explain** the importance of unit testing in MapReduce workflows.	BTL 2	Understand
8	**Compare** the advantages of MRUnit over manual testing	BTL 2	Understand
9	**Define** the anatomy of a MapReduce job run	BTL 1	Remember
10	**List** the key components involved in a MapReduce job run.	BTL 1	Remember
11	**Explain** the lifecycle of a MapReduce job run.	BTL 2	Understand
12	**Define** YARN and its purpose**.**	BTL 1	Remember
13	**Illustrate** the architecture of YARN with a diagram.	BTL 2	Understand
14	**Compare** Classic MapReduce and YARN.	BTL 2	Understand
15	**Explain** how YARN handles task failures.	BTL 2	Understand
16	**Define** job scheduling in MapReduce.	BTL 1	Remember

17	**List** different job scheduling algorithms used in MapReduce.	BTL 1	Remember
18	**What** is the purpose of the shuffle and sort phase?	BTL 1	Remember
19	**Explain** how sorting is performed in MapReduce.	BTL 2	Understand
20	**Define** task execution in MapReduce	BTL 1	Remember
21	**Illustrate** the role of task trackers in Classic MapReduce	BTL 2	Understand
22	**List** the Phases of Data Flow in HADOOP.	BTL 2	Understand
23	**Define** output formats in MapReduce and their significance	BTL 1	Remember
24	**Illustrate** how to implement a custom output format.	BTL 2	Understand

PART-B

Q.No.	Question	Marks	Level	Competence
1	**Construct** the architecture of the MapReduce framework in Hadoop	16	BTL 3	Apply
2	**Illustrate** the phases of a MapReduce job, focusing on the roles of mappers and reducers	16	BTL 3	Apply
3(a)	**Choose** the significance of the shuffle and sort phases in the MapReduce workflow.	8	BTL 3	Apply
3(b)	**Compare** MRUnit testing with traditional testing methods for MapReduce.	8	BTL 3	Apply
4	**Examine** the importance of unit testing in MapReduce workflows with MRUnit.	16	BTL 4	Analyze
5	**Evaluate** the process of writing MRUnit test cases for a word count MapReduce program.	16	BTL 4	Analyze
6	**Differentiate** the benefits and limitations of using MRUnit for validating MapReduce jobs.	16	BTL 4	Analyze
7	**Examine** a test plan for a MapReduce application using MRUnit.	16	BTL 3	Apply
8	**Develop** the anatomy of a MapReduce job run with detailed steps and a diagram	16	BTL 3	Apply
9	**Differentiate** the job execution flow in classic MapReduce and YARN.	16	BTL 3	Apply
10(a)	**Classify** the key features of the classic MapReduce architecture.	8	BTL 4	Analyze
10(b)	**Examine** the advantages of YARN over classic MapReduce in handling failures	8	BTL 4	Analyze
11(a)	**Investigate** the fault-tolerance mechanisms implemented in YARN	8	BTL 4	Analyze
11(b)	**Differentiate** the handling of node failures in classic MapReduce and YARN.	8	BTL 4	Analyze
12	**Identify** the importance of job scheduling in distributed computing systems.	16	BTL 3	Apply

13(a)	**Analyze** the impact of poor scheduling on the performance of MapReduce jobs.	16	BTL 4	Analyze
13(b)	**Summarize** the enhancements to the job scheduling mechanism in YARN to improve efficiency.	8	BTL 3	Apply
14	**Solve** the shuffle and sort phases in MapReduce with an example.	8	BTL 3	Apply
15	**Examine** he challenges in implementing the shuffle and sort operations efficiently.	16	BTL 4	Analyze
16	**Analyze** the challenges in monitoring and debugging a MapReduce job run**.**	16	BTL 4	Analyze
17	**Examine** the role of the Fair Scheduler in YARN with practical examples.	16	BTL 4	Analyze

UNIT-4 BASICS OF HADOOP

Data format – analyzing data with Hadoop – scaling out – Hadoop streaming – Haddop pipes design of Haddop distributed file system (HDFS) – HDFS concepts – Java interface – data flow – Hadoop I/O – data integrity – compression – serialization – Avro – file- based data structures – Cassandra – Hadoop integration.

PART – A

Q. No	Question	Level	Competence
1	**Define** Hadoop and its primary use.	BTL 1	Remember
2	**Describe** the purpose of Hadoop Pipes?	BTL 2	Understand
3	**Define** serialization in Hadoop.	BTL 1	Remember
4	**State** the function of Avro in Hadoop?	BTL 2	Understand
5	**Show** the Cassandra in the context of big data?	BTL 2	Understand
6	**Define** data compression in Hadoop I/O.	BTL 1	Remember
7	**Define** the concept of scaling out in Hadoop.	BTL 1	Remember
8	**Summarize** the design principles of HDFS.	BTL 2	Understand
9	**State** the data integrity important in Hadoop I/O?	BTL 1	Remember
10	**Explain** how Hadoop integrates with Cassandra**.**	BTL 2	Understand
11	**Explain** the concept of data flow in Hadoop systems.	BTL 2	Understand
12	**Infer** the purpose of file-based data structures in Hadoop?	BTL 2	Understand
13	**Outline** Avro serialization for a given data set.	BTL 2	Understand
14	**Define** the data integrity in HDFS?	BTL 1	Remember
15	**Interpret** how Cassandra managing a large-scale distributed database.	BTL 2	Understand
16	**Classify** Hadoop I/O operations for handling unstructured data.	BTL 2	Understand
17	**How** Hadoop Streaming with Hadoop Pipes.	BTL 1	Remember
18	**Compare** between HDFS and traditional file systems.	BTL 2	Understand

19	**Break down** the data flow process in HDFS.	BTL 1	Remember
20	**List** the challenges in integrating Hadoop with Cassandra?	BTL 1	Remember
21	**Define** Avro with other file-based data structures in Hadoop.	BTL 1	Remember
22	**How** the performance of Hadoop I/O when handling large datasets.	BTL 1	Remember
23	**Tell** the pros and cons of using the Java interface for HDFS operations**.**	BTL 1	Remember
24	**Describe** the significance of the Java interface in HDFS.	BTL 2	Understand

PART-B

Q.No.	Question	Marks	Level	Competence
1(a)	**Applty** the concept of data flow in Hadoop.	8	BTL 3	Apply
1(b)	**Identify** the features of the Hadoop Distributed File System (HDFS).	8	BTL 3	Apply
2	**Examine** the primary components of Hadoop architecture.	16	BTL 4	Analyze
3	**Investigate** the Hadoop pipes, and how are they used?	16	BTL 4	Analyze
4	**Break Down** the complete details of Avro, and why is it used in Hadoop?	16	BTL 4	Analyze
5(a)	**Compare** Hadoop Streaming and Hadoop Pipes in terms of design and functionality.	8	BTL 4	Analyze
5(b)	Illustrate the Java interface of HDFS with examples.	8	BTL 3	Apply
6	**Develop** how Cassandra integrates with Hadoop.	16	BTL 3	Apply
7	**Analyze** the process of Compression, Serialization and its importance in Hadoop.	16	BTL 4	Analyze
8	**Classify** Avro for data serialization in a Hadoop-based system	16	BTL 4	Analyze
9(a)	**Contrast** HDFS with other distributed file systems.	8	BTL 4	Analyze
9(b)	**Identify** the advantages and disadvantages of using Avro in Hadoop.	8	BTL 3	Apply
10	**Analyze** the significance of scaling out in Hadoop.	16	BTL 4	Analyze
11	**Break Down** the design of Hadoop's Distributed File System (HDFS).	16	BTL 4	Analyze
12	**Examine** the process of serialization and its importance in Hadoop.	16	BTL 4	Analyze
13(a)	**Differentiate** between serialization and deserialization in Hadoop I/O.	8	BTL 4	Analyze
13(b)	**Break down** the integration process of Cassandra with Hadoop.	8	BTL 4	Analyze
14	**Examine** the architecture and workflow of Hadoop in managing big data.	16	BTL 4	Analyze

15	**Examine** the MapReduce program to analyze a dataset using Hadoop.	16	BTL 4	Analyze
16(a)	**Build** the implications of using Avro over other serialization frameworks.	8	BTL 3	Apply
16(b)	**Analyze** the role of file-based data structures in Hadoop.	8	BTL 4	Analyze
17	Apply HDFS concepts to design a storage system for large-scale data.	16	BTL 3	Apply

UNIT-5 HADOOP RELATED TOOLS

Hbase – data model and implementations – Hbase clients – Hbase examples – praxis. Pig – Grunt – pig data model – Pig Latin – developing and testing Pig Latin scripts. Hive – data types and file formats – HiveQL data definition – HiveQL data manipulation– HiveQL queries.

PART – A

Q. No	Question	Level	Competence
1	**Define** what HBase is and explain how it differs from relational databases.	BTL 1	Remember
2	**Show** HBase differs from traditional relational databases.	BTL 2	Understand
3	**List** the key features of the HBase data model.	BTL 1	Remember
4	**Illustrate** the key features of the HBase data model with examples.	BTL 2	Understand
5	**Recall** the role of HBase clients in data access.	BTL 1	Remember
6	**Summarize** the function of HBase clients in facilitating database access.	BTL 2	Understand
7	**Define** Praxis in the context of HBase.	BTL 1	Remember
8	**Infer** the concept of Praxis in HBase and its significance.	BTL 2	Understand
9	**What** is Pig in the Hadoop ecosystem.	BTL 1	Remember
10	**Outline** the role of Pig in the Hadoop ecosystem and its applications.	BTL 2	Understand
11	**State** the purpose of Grunt in Pig.	BTL 1	Remember
12	**Demonstrate** how Grunt is used in Pig scripting.	BTL 2	Understand
13	**Define** Pig Latin and mention its primary use.	BTL 1	Remember
14	**Clarify** the purpose and primary use of Pig Latin in data processing.	BTL 2	Understand
15	**List** the common data types available in Hive.	BTL 1	Remember
16	**Compare** the common data types in Hive with examples of their use.	BTL 2	Understand
17	**Recall** the default file format used in Hive.	BTL 1	Remember
18	**Classify** why the default file format in Hive is significant for data storage.	BTL 2	Understand
19	**Define** how HBase clients interact with the database.	BTL 1	Remember
20	**Infer** the interaction process of HBase clients with the database.	BTL 2	Understand

21	**List** the components of the Pig data model used in data processing.	BTL 1	Remember
22	**Classify** how the Pig data model contributes to effective data processing.	BTL 2	Understand
23	**State** the importance of Grunt in Pig scripting.	BTL 1	Remember
24	**Summarize** the importance of Grunt as a tool in scripting with Pig.	BTL 2	Understand

PART – B

Q. No	Question	Marks	Level	Competence
1	**Construct** how to set up an HBase database with a basic schema and perform data insertion using HBase clients.	16	BTL 3	Apply
2	**Develop** a Pig Latin script to analyze and filter large datasets based on a specific condition.	16	BTL 3	Apply
3	**Analyze** how to configure and use Grunt to debug and test Pig Latin scripts effectively.	16	BTL 4	Analyze
4(a)	**Examine** a Hive table with appropriate data types and file formats for storing IoT sensor data.	8	BTL 4	Analyze
4(b)	**Analyze** a HiveQL query to filter and extract data from a partitioned table.	8	BTL 4	Analyze
5	**Develop** a workflow using HiveQL to create, insert, and query a Hive table for e-commerce sales data	16	BTL 3	Apply
6(a)	**Analyze** how Pig Latin's data model accommodates both structured and unstructured data in workflows.	8	BTL 4	Analyze
6(b)	**Compare** the performance of Grunt and HiveCLI for managing big data workflows.	8	BTL 4	Analyze
7	**Construct** a HiveQL query to transform unstructured log data into a structured format and analyze error trends.	16	BTL 3	Apply
8(a)	**Identify** how the choice of data types in Hive affects storage and computation efficiency.	8	BTL 3	Apply
8(b)	**Analyze** the differences in data manipulation techniques between HiveQL and Pig Latin.	8	BTL 4	Analyze
9	**Apply** the concept of Praxis in HBase to optimize database read and write operations for a banking application.	16	BTL 3	Apply
10(a)	**Analyze** the effectiveness of Praxis in real-world HBase applications, particularly in high-volume data processing.	8	BTL 4	Analyze
10(b)	**Examine** the performance implications of using nested data structures in the Pig data model.	8	BTL 4	Analyze
11	**Apply** how to integrate Hive with HBase to process real-time streaming data.	16	BTL 3	Apply
12	**Develop** a Pig Latin script to group and sort data by a specified attribute.	16	BTL 3	Apply
13	**Develop** and execute a Pig Latin script for performing a word count operation on large datasets.	16	BTL 3	Apply
14(a)	**Analyze** the differences between the HBase data model and traditional relational database models in practical scenarios.	8	BTL 4	Analyze
14(b)	**Examine** the scalability of HBase for storing and retrieving time-series data in a distributed environment.	8	BTL 4	Analyze

15	**Compare** and contrast the implementation of joins in Pig Latin and HiveQL with examples.	16	BTL 4	Analyze
16(a)	**Identify** the role of Grunt in enhancing the development and debugging of Pig Latin scripts.	8	BTL 3	Apply
16(b)	**Analyze** how Pig Latin scripts handle large-scale data processing compared to HiveQL queries.	8	BTL 4	Analyze
17	**Classify** the advantages and limitations of using HiveQL for handling structured and unstructured data	16	BTL 4	Analyze