SRM VALLIAMMAI ENGINEERING COLLEGE

(An Autonomous Institution)

SRM Nagar, Kattankulathur – 603203

DEPARTMENT OF

COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK

II SEMESTER

M.E.-CSE

1912204-BIG DATA ANALYTICS

Regulation – 2019

Academic Year 2022-23 (Even)

Prepared by

Mr. T. Rajasekaran, Assistant Prof. / CSE

SRM VALLIAMMAI ENGNIEERING COLLEGE
SRM Nagar, Kattankulathur – 603203.

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

QUESTION BANK

SUBJECT: 1912204-BIG DATA ANALYTICS

SEM / YEAR: II/ I (M.E.-CSE)

UNIT I – INTRODUCTION TO BIG DATA & HADOOP
Types of Digital Data, Introduction to Big Data, Big Data Analytics, History of Hadoop, Apache Hadoop, Analysing Data with Unix tools, Analysing Data with Hadoop, Hadoop Streaming, Hadoop Echo System, IBM Big Data Strategy, Introduction to Infosphere BigInsights and Big Sheets.

PART A

S. No.	Question	BTL	Competence
1	What is Digital Data?	BTL -1	Remember
2	Define Big Data.	BTL -1	Remember
3	Evaluate the benefits of Big Data and Data Analytics.	BTL -5	Evaluate
4	List out various data formats?	BTL -1	Remember
5	State the characteristics of Big Data.	BTL -1	Remember
6	Analyze big data analytics.	BTL -4	Analyze
7	Summarize five ‘V’s of Big data.	BTL -2	Understand
8	Define Descriptive analytics.	BTL -1	Remember
9	Evaluate why is Big Data Important?	BTL -2	Understand
10	Pointout the difference between Predictive analytics and Prescriptive analytics.	BTL -4	Analyze
11	Illustrate Hadoop Ecosystem.	BTL -3	Apply
12	Compose notes on Apache Pig.	BTL -6	Create
13	Examine HBase.	BTL -3	Apply
14	Difference between Hive and RDBMS.	BTL -3	Apply
15	Generalize the applications of big data.	BTL -6	Create
16	Differentiate structure and unstructured data.	BTL -2	Understand
17	Differentiate HDFS and Hbase.	BTL -2	Evaluate
18	Evaluate Hadoop YARN.	BTL -5	Evaluate
19	Define Predictive analytics.	BTL -1	Remember
20	List the components of Hadoop.	BTL -2	Understand
21	Explain Characteristicis of Big Data.	BTL -2	Understand
22	What are the advantages of Hadoop?	BTL -3	Apply
23	Analyze the tools used in Hadoop.	BTL -4	Analyze
24	Evaluate Avro data serialization technique.	BTL -5	Evaluate

PART B

1	Describe the types of Digital data in detail. (13)	BTL -1	Remember
2	Why is Big Data Important? Explain in detail. (13)	BTL -2	Understand
3	i) Illustrate the big data analytics in detail .(7) ii) Examine the difference between structure and unstructured data. (6)	BTL -3	Apply
4	What is Big data? Discuss it in terms of five characteristics of Big Data. (13)	BTL -2	Understand
5	Explain core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail. (13)	BTL -1	Remember
6	Discuss the various formats of data and illustrate with real time examples? (13)	BTL -2	Understand
7	Illustrate about Apache Hadoop in detail. (13)	BTL-3	Apply
8	Analyze the following i) analyzing data with unix tools. (7) ii) analyzing data with Hadoop. (6)	BTL -4	Analyze
9	What is Big Data? Explain how big data processing differs from distributed processing. (13)	BTL -1	Remember
10	Develop how Hadoop streaming is suited with text processing explain. (13)	BTL -6	Create
11	Explain in detail about Hadoop streaming. (13)	BTL -5	Evaluate
12	Analyze IBM big data strategy. (13)	BTL -4	Analyze
13	Discuss the following in detail. i) Conventional challenges in big data. (7) ii) Nature of Data. (6)	BTL -1	Remember
14	What is Hadoop Ecosystem? Discuss various components of Hadoop Ecosystem. (13)	BTL -4	Analyze
15	Classify the Compare Row oriented and Column Oriented database structures. (13)	BTL -2	Understand
16	What are the advantages of Hadoop? Sketch and explain Hadoop Architecture and its Components with proper diagram. (13)	BTL -3	Apply
17	Criticize the various configuration files used in Hadoop Installation. What is use of mapredsite.xml? (13)	BTL -5	Evaluate

PART-C

1	Compose List various applications of big data. How it can be used to improve business for a superstore. (15)	BTL-6	Create
2	Summarize 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. (15)	BTL-5	Evaluate
3	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. (15)	BTL-6	Create
4	Appraise the IBM big data strategy in detail with an example for each. (15)	BTL-5	Evaluate
5	Discuss Big Data in Healthcare, Transportation & Medicine. (15)	BTL-5	Analyze

UNIT III - MAP REDUCE
Anatomy of a Map Reduce Job Run, Failures, Job Scheduling, Shuffle and Sort, Task Execution, Map Reduce Types and Formats, Map Reduce Features.

PART A

S. No.	Question	BTL	Competence
1	Define what is MapReduce?	BTL -1	Remember
2	List out the real time examples of MapReduce	BTL -1	Remember
3	Define what happens if the mapper output does not match the reducer input.	BTL -5	Evaluate
4	Explain the term YARN.	BTL -1	Remember
5	Discuss the term List Processing.	BTL -1	Remember
6	Generalise the term Record Reader/Writer?	BTL -4	Analyze
7	How to explore the Scale-out architecture?	BTL -4	Analyze
8	Define Partitioners.	BTL -1	Remember
9	Explain about Data Locality in MapReduce.	BTL -2	Understand
10	Combiner Phase" in MapReduce	BTL -4	Analyze
11	How to overcome the Faults and handling of Errors?	BTL -3	Apply
12	Design Map Phase.	BTL -6	Create
13	List the uses of MapReduce.	BTL -3	Apply
14	Illustrate the Capacity Scheduler.	BTL -3	Apply
15	Design Reducer Phase.	BTL -6	Create
16	Define Input Split.	BTL -2	Understand
17	Illustrate Combiner.	BTL -2	Understand
18	List out the trackers in Hadoop.	BTL -1	Remember
19	Define Container.	BTL -1	Remember
20	Describe short note on Input Format?	BTL -2	Understand
21	Summarize copyFromLocal.	BTL -2	Understand
22	Differentiate Container and Combiner.	BTL -3	Apply
23	Explain the Fair Scheduler.	BTL -4	Analyze
24	Summarize short on the FileInputFormat.	BTL -5	Evaluate

PART B

1	Describe the working principle of MapReduce? (13)	BTL -1	Remember
2	Estimate the entire process of data analysis conducted in the MapReduce programming model? (13)	BTL -4	Analyze
3	Define what conditions must be met to implement MapReduce application? (13)	BTL -3	Apply
4	Explain in detail the framework of MapReduce? (13)	BTL -2	Understand
5	Describe the uses of MapReduce? (13)	BTL -1	Remember
6	Describe "Map Phase" and "Combiner Phase" in MapReduce. (13)	BTL -2	Understand

7	Explain the description of MapReduce process for a specific case? (13)	BTL -5	Evaluate
8	Analyze in detail the concept of developing the Map Reduce Application. (13)	BTL -4	Analyze
9	What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram. (13)	BTL -1	Remember
10	Demonstrate the types of MapReduce applications? (13)	BTL -3	Apply
11	Explain Job Scheduling in Map Reduce. How it is done in case of (i) The Fair Scheduler (7) (ii) The Capacity Scheduler (6)	BTL -4	Analyze
12	Describe the following (i) Input Split. (7) (ii) FileInputFormat class. (6)	BTL -1	Remember
13	List and explain the main features of MapReduce? (13)	BTL -2	Understand
14	Compose the following commands with syntax and at least one example of each. (i) copyFromLocal (ii) showing the content of outputfile (iii) setrep (iv) Checksum (13)	BTL -6	Create
15	Discuss some techniques to optimize MapReduce jobs? (13)	BTL -2	Understand
16	Explain JobTracker and Task Tracker in Hadoop. (13)	BTL -3	Apply
17	Summarize note on Input Formats in MapReduce.	BTL -5	Analyze

PART-C

1	Infer the process of data analysis conducted in the MapReduce programming mode. (15)	BTL-5	Analyze
2	Can MapReduce be used to solve any kind of computational problems? if not, explain the cases where MapReduce is not applicable? (15)	BTL-5	Evaluate
3	Describe the working of the MapReduce algorithm? (15)	BTL-6	Create
4	Discuss the points you need to consider while designing a file system in MapReduce? (15)	BTL-5	Analyze

UNIT IV – HADOOP ECO SYSTEM
Pig : Introduction to PIG, Execution Modes of Pig, Comparison of Pig with Databases, Grunt, Pig Latin, User Defined Functions, Data Processing operators. Hive : Hive Shell, Hive Services, Hive Metastore, Comparison with Traditional Databases, HiveQL, Tables, Querying Data and User Defined Functions. Hbase : HBasics, Concepts, Clients, Example, Hbase Versus RDBMS. Big SQL : Introduction.

PART - A

S. No.	Question	BTL	Competence
1	Breifly explain the architecture of Pig.	BTL -1	Remember
2	List out the benefits of Pig.	BTL -1	Remember
3	Summarize what are the properties of Pig?	BTL -5	Understand
4	Explain about schema	BTL -1	Remember
5	Explain the Pig Latin application flow.	BTL -1	Remember
6	Discuss the modes of Pig scripts.	BTL -4	Analyze
7	Express DDL concepts in detail.	BTL -4	Understand
8	Define Hive.	BTL -1	Remember
9	List out the Data types in Hive.	BTL -2	Analyze
10	Classify Pig Latin commands in Pig.	BTL -4	Analyze
11	Interpret joins with examples.	BTL -3	Apply
12	Compose DISTINCT operator in Pig Latin.	BTL -6	Create
13	Demonstrate use of the FILTER operator in Pig Latin.	BTL -3	Apply
14	Illustrate Zookeeper?	BTL -3	Apply
15	Design HBase architecture.	BTL -6	Create
16	Differentiate: Hive and RDBMS.	BTL -2	Understand
17	Differentiate: Apache pig Vs Map Reduce.	BTL -2	Evaluate
18	Interpret what is SQL.	BTL -5	Evaluate
19	Define UDF.	BTL -1	Remember
20	In Hive, explain the term ‘aggregation’ and its uses?	BTL -2	Understand
21	Discuss the HiveQL.	BTL -2	Understand
22	Difference between HDFS and Hbase.	BTL -3	Apply
23	How will you query the data in HIVE?	BTL -4	Analyze
24	Compare Raw oriented and Column Oriented database structures.	BTL -5	Evaluate

PART - B

1	Discuss the two modes used for running the Pig scripts? (13)	BTL -1	Remember
2	i) What is Pig? Explain its installing process. (7)	BTL -5	Evaluate
	ii) Explain two execution types or modes in Pig. (6)		
3	Define what are the main reasons for developing Pig Latin? (13)	BTL -3	Apply

4	Analyze what do you understand by Pig Latin application flow? (13)	BTL -4	Analyze
5	Define the various Statements used in flow of data processing in Pig Latin? (13)	BTL -2	Understand
6	Discuss on Apache Pig and Enlist applications of Apache Pig. (13)	BTL -1	Remember
7	i) Discuss the use of the FOREACH and ASSERT operator in Pig Latin (7) ii) Discuss the use of the FILTER and DISTINCT operator in Pig Latin.(6)	BTL -2	Understand
8	Define what do you understand by introducing Hive? (13)	BTL -3	Apply
9	Analyze HiveQL-Select-Order By with suitable example. (13)	BTL -4	Analyze
10	Explain working of Hive with proper steps and diagram.	BTL -1	Remember
11	What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example (13)	BTL -6	Create
12	Discuss the various Hive services with examples? (13)	BTL -2	Understand
13	Analyze HBase architecture in neat diagram. (13)	BTL -4	Analyze
14	Explain the process of installing Hive & features of Hive. (13)	BTL -1	Remember
15	What is Zookeeper explain its features with applications. (13)	BTL -2	Understand
16	Illustrate the following (a) Metastore in Hive. (7) (b) Storage mechanism in HBase. (6)	BTL -3	Apply
17	Summarise joins. How many types of joins are there in Pig Latin with an examples? (13)	BTL -5	Analyze

PART - C

1	A start-up company want to use Hive for storing its data. List the collection types provided by Hive for this purpose? Write a shell command in Hive to list all the files in the current directory? (15)	BTL-5	Analyze
2	Illustrate the difference between the RDBMS versus Hadoop in detail?(15)	BTL-5	Evaluate
3	Design the Hive commands to create a table with four columns: First name, last name, age, and income? (15)	BTL-6	Create
4	Appraise HiveQL Data Definition Language. Explain any three HiveQL DDL command with its syntax and example. (15)	BTL-5	Evaluate