A   Reg. No. :

B.E. / B.Tech. DEGREE EXAMINATION, APRIL 2024

Professional Elective

Computer Science and Business Systems

21ADVG13 - DATA ANALYTICS

(Regulations 2021)

(Common to All Branches)

Duration: Three hours   Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 1 = 10 Marks)

1. Data Analysis is a process of   CO1 -U

   (a) inspecting data   (b) Cleaning data   (c) transforming data   (d) All the above

2. Which is not a type of NOSQL?   CO1 -U

   (a) HBase   (b) QBase   (c) MongoDB   (d) CouchDB

3. Which of the following constraints RDBS doesn't check before creating the tables?   CO1 -U

   (a) Not null   (b) Data Structure   (c) Primary keys   (d) Data integrity

4. Which of the following statements accurately describes the purpose of a column family in HBase?   CO1 -U

   (a) A column family represents a group of tables in HBase, allowing for logical organization of data.

   (b) A column family is a group of related columns stored together on disk for efficient retrieval

   (c) A column family defines the schema for a row in HBase, specifying the data types and constraints for each column.

   (d) A column family is a set of HBase instances deployed across multiple nodes for data redundancy and fault tolerance.

5. Hbase is what type of database?   CO1 -U

   (a) Schema – rigid   (b) Schema - flexi   (c) Schema - less   (d) Not a database

6. Which of the following command is used to show values to keys used in Pig?   CO1 -U

   (a) set   (b) display   (c) declare   (d) all of the above

7. Which of the following HQL statements is used to retrieve distinct values from a column in a Hive table?   CO1 -U

   (a) SELECT DISTINCT column_name FROM table_name;

   (b) SELECT UNIQUE column_name FROM table_name;

   (c) SELECT column_name UNIQUE FROM table_name;

   (d) SELECT column_name DISTINCT FROM table_name;

8. HBaseAdmin and _____ are the two important classes in this package that provide DDL functionalities.   CO1 -U

   (a) HTableDescriptor   (b) HDescriptor   (c) HTable   (d) HTabDescriptor

9. What is the default consistency level in Cassandra?   CO1 -U

   (a) QUORUM   (b) LOCAL_ONE   (c) ONE   (d) ALL

10. Which of the following constraints RDBS doesn't check before creating the tables?   CO1-U

   (a) Not null   (b) Data Structure   (c) Primary keys   (d) Data integrity

PART – B (10 x 3= 30 Marks)

11.	Define big data and its uses.	CO1-U
12.	What are the different Hadoop Configuration files?	CO1-U
13.	Differentiate Nosqlvs RDBMS	CO1-U
14.	What are the four V's of Big Data?	CO1 - U
15.	Anlayse the data storage component used by Hadoop with an example.	CO2 - Ana
16	How big data analysis helps businesses increase their revenue? Give example.	CO1 - U
17	Explain how indexing in HDFS is done?	CO1 - U
18	Does Pig differ from MapReduce? If yes, how?	CO2-Ana
19	What is Grunt?	CO1 - U
20	Compare and differentiate Pig and Hive.	CO1 - U

UCG13

21. (a) What is Bigdata? Describe the main features of a big data in detail   CO1- U   (20)

Or

     (b) (i) Generalize the list of tools related to Hadoop   CO1- U   (10)

        (ii) Prepare the list of works done by Hadoop working model   CO1- U   (10)

22 (a) Analyze HDFS and apply it in suitable scenario.   CO2- Ana   (20)

Or

     (b) Analyze Nature of Data and its applications   CO2- Ana   (20)

23. (a) You are working as a database administrator for a large e-commerce company that uses HBase to store product information. The company is experiencing a surge in traffic due to a major sale event, and the database is struggling to handle the increased load. As part of your role, you need to optimize the HBase deployment to ensure optimal performance during peak traffic periods   CO2- Ana   (20)

Or

     (b) You are a data engineer tasked with designing a data model for a social media platform that needs to handle a large volume of user-generated content. The platform requires efficient storage and retrieval of user posts, comments, likes, and other interactions. Outline your approach to designing a data model using Apache Cassandra for the social media platform   CO2- Ana   (20)

UCG13