**NAVRACHANA UNIVERSITY**
*a UGC recognized University*

| | |
|---|---|
| **School:** | School of Engineering and Technology |
| **Program/s:** | MTech – Computer Science and Engineering |
| **Year:** 2nd | **Semester:** 3rd |
| **Examination:** | End Semester Examination |
| **Examination Year:** | December - 2021 |

| | | | |
|---|---|---|---|
| **Course Code:** CS222 | **Course Name:** Data Analytics | | |
| **Date:** 03/12/2021 | | **Total Marks:** | 40 |
| **Time:** 08:30 am to 10:00 am | | **Total Pages:** | 3 |

---

**Instructions:**

→ Write each answer on a new page.
→ Use of a calculator is permitted/not permitted.
→ * COs=Course Outcome mapping. # BTL=Bloom's Taxonomy Level mapping

---

| Q. No. | Details | Marks | COs* | BTL# |
|---|---|---|---|---|
| Q.1 | **Choose the correct options for the following questions(All are compulsory)** | 10 | | |
| | 1. Data stored in DBMS tables is a/an<br>A. Structured Data<br>B. Unstructured Data<br>C. Semi-structured<br>D. Quasi-Structured | | CO2 | BT1, BT2 |
| | 2. Consider the following statements<br>**Statement-1**: HDFS performance gets degraded if file size is in KB's or MB's.<br>**Statement-2**: HDFS is not suitable if frequent edits are required than read operations.<br>A. Statement 1 and 2 are True.<br>B. Only statement 1 is True.<br>C. Only statement 2 is True.<br>D. Statement 1 and 2 are False. | | CO1 | BT2, BT3, BT4 |
| | 3. Scalability in Hadoop system refers to<br>A. In Hadoop a single machine hardware can be upgraded to have more primary and secondary memory.<br>B. In a Hadoop cluster, separate nodes can be added to upgrade its performance.<br>C. Extra set of software's can be added supporting Hadoop ecosystem.<br>D. Add nothing but increase only replication server for backup. | | CO3 | BT2, BT3, BT4 |

| | | | |
|---|---|---|---|
| 4. In HDFS, files metadata information is always stored in <br>    A. DataNode <br>    B. NameNode <br>    C. JobTracker <br>    D. Sharded Servers | | CO2 | BT1, BT2 |
| 5. _____ command is used to display contents in the files stored in root directory. <br>    A. hdfs dfs -ls /filename <br>    B. hdfs dfs -cat /filename <br>    C. hadoop hdfs -ls /filename <br>    D. hadoop -fs -display /filename | | CO3 | BT1, BT2 |
| 6. The purpose of Sqoop in Hadoop ecosystem is to <br>    A. Divide the task into subtask and sent it to DataNodes <br>    B. Divide the task into subtask and sent it to NameNodes <br>    C. Import/Export of Data between HDFS structured data and Sqoop. <br>    D. Sharding of Data between HDFS and Sqoop. | | CO4 | BT1, BT2 |
| 7. How many blocks would a file take whose size is 2098 MB. The default block size of HDFS is set to 512MB. <br>    A. 4 <br>    B. 4.5 <br>    C. 6 <br>    D. 5 | | CO4 | BT2 BT3, BT4, BT5, BT6 |
| 8. _____ Java library package is used to create an UDF in Pig platform. <br>    A. org.apache.pig.Evaluation <br>    B. org.apache.pig.EvalFunc <br>    C. org.apache.pig.UDF <br>    D. org.apache.pig.Functions | | CO3 | BT2 BT3, BT4, BT5, BT6 |
| 9. By default Pig always runs in a <br>    A. Local Mode <br>    B. Map-Reduce Mode <br>    C. Text Mode <br>    D. Binary Mode | | CO3 | BT1, BT2 |
| 10. HBase is a <br>    A. Key-Value, Vector Based Data <br>    B. Column Oriented, Sorted Map Data <br>    C. Graph Oriented, Tree Map Data <br>    D. Document Oriented, Sequential Data | | CO1 | BT1, BT2 |

| Q.2 | Answer the following question in brief (All are compulsory) | 5 | | |
|---|---|---|---|---|
| | 1. Which are the nested data-types in Pig ? | | CO3 | BT1, BT2 |
| | 2. What is the use of JDBC Server in Hive ? | | CO2 | BT2, BT3, BT4 |
| | 3. Which are the mechanism for organizing data in Hive ? | | CO1 | BT2, BT3, BT4 |
| | 4. Define Data-Locality. | | CO2 | BT2, BT3, BT4 |
| | 5. What kind of issues can be addressed by peer-to-peer replication? | | CO4 | BT2, BT3, BT4, BT5 |
| Q.3 | Answer the following questions in detail (Attempt any 5) | 25 | | |
| | 1. What is Big-Data ? Mention its categories with example. Also discuss its characteristics. | | CO1 | BT2, BT3, BT4 |
| | 2. Discuss HDFS and its Architecture. | | CO3 | BT1, BT2, BT3, BT4 |
| | 3. Discuss importance of NoSQL databases. Also discuss types of No-SQL databases. | | CO2 | BT2, BT3, BT4 |
| | 4. Write a note on Sharding and Master-Slave Architecture. | | CO2 | BT2, BT3, BT4 |
| | 5. Write characteristics of Pig in contrast with Map-Reduce. Explain execution cycle of pig script. | | CO4 | BT2, BT3, BT4 |
| | 6. Mention components of Hadoop ecosystem discussing each in brief. | | CO1 | BT1, BT2, BT3, BT4 |

***********End of Question Paper***********