1904M114

Candidate's Seat No :_____

Integ. M.Sc DS Semester-8 Examination

CC-413

Big Data Analytics

April-2024

Time : 2-30 Hours] **[Max. Marks : 70**

Instructions: All questions are compulsory. Use of non-programmable scientific calculator is allowed.

Q.1 **(a)** Explain 5V's of Big Data with example.
(b) Explain Hadoop 1.x Architecture with neat sketch.

OR

(a) Explain Hadoop 2.x and 3.x Architecture with neat sketch and compare both.
(b) Give the name of types of data that we can store in NOSQL database and explain with Examples.

Q.2 **(a)** Explain Big Data Life cycle with diagram.
(b) Explain Spark Memory Management with neat sketch.

OR

(a) Explain SEMMA methodology in depth.
(b) Explain ETL Process vs ELT Process with diagram.

Q.3 **(a)** Explain HIVE Architecture with Diagram and example.
(b) Where can we store big data. Explain all with examples.

OR

(a) Provide the name of three file format that work in big data and why we use that?
(b) Explain Serialization and Deserialization with examples.

Q.4 **(a)** Explain Replication and Sharding process with diagram.
(b) State all operation that we can perform in NoSQL Database.

OR

(a) Explain Apache Spark Eco-System with Diagram.
(b) Explain ACID and CAP (BASE) theorem with CAP triangle.

Q.5 Attempt any **SEVEN** out of **TWELVE:**

(1) How a secondary name node differs from the name node in HDFS.
(2) How to create a table by using HIVEQL. (Write Syntax to create employee table)
(3) Specify the role of name node and data node in HDFS.
(4) What is Hadoop? Write the name of its components.
(5) Explain how map reduce jobs run on YARN.
(6) Difference between Hive and RDBMS.
(7) Explain Checkpointing in Hadoop.
(8) Give Difference between SQL and NOSQL database.
(9) Explain the MongoDB hierarchy or Data model of MongoDB.
(10) Explain the Feature of Spark.
(11) Provide five Big Data Application with example.
(12) Write the Difference between Data Science and Business Intelligence.

**** ****