TIE Model-paper-1 for BDA-18CS72

Disclaimer: The following question paper(s) are purely prepared/formed by the TIE review team under the strict guidance of certified teachers/mentors. The sole purpose of this is to give a thorough idea about the final assessment paper. We kindly request the students to use this for revision purposes only.

Module-1

1. a) How do you define data, web data, and Big data? How do you classify data as structured, semi-structured, multi-structured, or unstructured? What are the characteristics of big data?

   b) Define big data architecture. Describe the five architectural layers and their functions. What do you mean by "data management"?. Make a list of data management functions.

   c) Common data storage methods How can in-memory columnar format help OLAP? Give an example. What is an enterprise Server? What sets an enterprise server's data store from a web server's?

OR

2. a)Using a diagram, show how to export data from a data store to computers, web servers, web services, Amazon, Rackspace, and Hadoop cloud services.

b)How does a Big data stack help in analytics tasks? What are the reference models for traditional and big data analytics architectures?

c) What are the hazards of using Big data? So, how does Big data help manage credit risk? Describe how Big data analytics is used in healthcare and medicine.

Module-2

3. a) Describe Hadoop. Recognize Hadoop's two main parts. Explain their use. Draw a diagram of Hadoop's distributed storage, resource manager, processing framework, and API layers.

b) Describe how to acquire data streams using Apache Flume

OR

4. a) What is "MapReduce"? Big Data with MapReduce on Big Servers The app's resource needs. You can create new tasks and subtasks (threads). YARN resource control Keeping an eye on the Application Master? How does the YARN resource manager monitor the Application Master?

b) Explain in detail Apache pig?

Module-3

5. a). When should NoSQL replace SQL? Why use NoSQL for Big Data? Describe NoSQL data stores.

b)How are replication and Sharding used? Explain how does sharding help in minimizing downtime? Compare NoSQL data store with SQL databases.

c) Defining MongoDB and its features Explain MongoDB commands for querying the DBs? How can one achieve transaction and locking in MongoDB? (6 Marks)

OR

6. a) How can a NoSQL data storage add more data capacity? Analyze the differences between SQL databases in terms of database architecture, scalability, SQL joins, data size preferences, ACID features and support from prominent IT companies.

b) Identify the benefits and drawbacks of

1) the key-value data storage.

2) an object data store and

3) a graph database Describe the column family data storage. How do they work with OLAP? BigTable stores data in two ways. c) What are the

important design considerations when using a column family data store like Cassandra? List and explain usages of data types built into Cassandra.

Module-4

7. a) List and explain the features of the MapReduce programming model? How does the MapReduce program enable parallel processing? How does 'Group By' operate for creating Mapper output? What are the roles of partitioning and combining?

b) Describe the Hive architecture components. Why are HiveQL, SQL-like scripts used in place of RDBMS, such as MySQL for Big Data?

c) Describe the Features of PIG. What are the differences between the Pig programming model with MapReduce, relational database and Hive programming models?

OR

8. a) How does MapReduce implement the relational algebraic functions, union, projection, difference, intersection, natural join, grouping and aggregation? Explain each with an example.

b)Why should partitions be created in databases and tables in Hive data warehouse for very large datasets? What are aggregation commands provisioned in HiveQL? What are the partitioning commands?

c) Describe usages of Pig operations: parallel, split and defining a UDF Give an example of each.

Module-5

9. a) Describe the approaches used in linear, multivariate and multiple regression algorithms.

b)How the features are evaluated in the text documents? Explain five phases and steps in the phases during the text analytics.

OR

10. a) How are Jaccard distance, cosine distance and edit distance used for finding similar items? How are methods of collaborative filtering used?

b)What are the tasks in web data analytics? List and explain three phases of web usage mining?