CBCS SCHEME

USN | | | | | | | | | | |

18CS72

Seventh Semester B.E. Degree Examination, July/August 2022
Big Data Analytics

Time: 3 hrs. Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

Module-1

1 a. Define Data, Web data, Big data. Also explain structured, semistructured and unstructured data. **(10 Marks)**

b. List and explain the characteristics of big data. Illustrate by considering an example of E-commerce, how big data is used. **(10 Marks)**

OR

2 a. With a neat diagram, explain the function of each of the five layers in big data architecture design. **(12 Marks)**

b. How does Berkeley Data Analytics stack help in analytics tasks? **(08 Marks)**

Module-2

3 a. With a neat diagram, explain Hadoop main components and ecosystem components. **(08 Marks)**

b. Brief out the features of Hadoop HDFS? Also explain the functions of Name Node and Data Node. **(08 Marks)**

c. Explain any two HDFS commands with example. **(04 Marks)**

OR

4 a. Explain the following:
(i) HDFS block replication (ii) HDFS safe mode.
(iii) Rack awareness (iv) Name Node high availability. **(12 Marks)**

b. Discuss the Apache sqoop Import and Export methods with neat diagrams. **(08 Marks)**

Module-3

5 a. List and compare the features of Big Table, RC, ORC and Parquet data stores. **(10 Marks)**

b. With example explain key-value store. **(10 Marks)**

OR

6 a. Discuss the usage of MongoDB, Cassandra, CouchDB, Oracle NoSQL and Riak. **(10 Marks)**

b. List the Pros and Cons of distribution using sharding. **(05 Marks)**

c. Give the comparison between NoSQL and SQL/RDBMS. **(05 Marks)**

Module-4

7 a. Describe MapReduce Execution steps with a neat sketch. **(12 Marks)**

b. How node failure can be handled in Hadoop? Discuss. **(08 Marks)**

OR

8 a. With a neat diagram, describe Hive integration and work flow steps. **(10 Marks)**

b. Explain with Return type and Syntax the Hive built-in functions. **(10 Marks)**

Module-5

9 a. Discuss Regression Analysis using Linear and Non-linear regression models. **(10 Marks)**

b. Explain with an example Apriori algorithm to evaluate candidate key. **(10 Marks)**

OR

10 a. Write a note on:

(i) Web mining

(ii) Web content mining.

(iii) Web usage mining. **(12 Marks)**

b. How the Cliques discover communities from social network analysis? **(04 Marks)**

c. Define a Page Rank. **(04 Marks)**

* * * * *