

CBCS SCHEME



18CS72

Seventh Semester B.E. Degree Examination, June/July 2023

Big Data and 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 big data, and explain its characteristics. (06 Marks)
- b. List and explain different data sources. (04 Marks)
- c. Explain Big data designing architecture. (10 Marks)

OR

- 2 a. Explain the functions of each of the Big Query layers in big data architecture design using Big Query Cloud Service at Google Cloud Platform. (10 Marks)
- b. (i) Explain big data analytics applications. (05 Marks)
- (ii) Write a short note on data storage and analysis. (05 Marks)

Module-2

- 3 a. With a neat diagram, explain the components of HDFS (Hadoop Distribution File System). (10 Marks)
- b. How does the Hadoop MapReduce data flow work for a word count program? Give an example. (10 Marks)

OR

- 4 a. What is APACHE Flume? Describe the feature components and working of apache flume. (10 Marks)
- b. Explain the features and benefits of apache Hive in hadoop. (10 Marks)

Module-3

- 5 a. Discuss NOSQL data architecture pattern, with an example. (10 Marks)
- b. Explain four different ways for handling big data problems. (10 Marks)

OR

- 6 a. (i) Explain different components of Cassandra. (05 Marks)
- (ii) Explain different data types built into Cassandra. (05 Marks)
- b. (i) Describe different CQL commands and their functionalities. (05 Marks)
- (ii) Write a short note on NOSQL to Manage Big Data. (05 Marks)

Module-4

- 7 a. Describe the significance of apache pig in hadoop. (10 Marks)
- b. With a neat diagram, explain MapReduce Programming model. How does MapReduce enables query processing quickly in Big Data Problems? (10 Marks)

OR

- 8 a. Explain Hive architecture with a neat diagram. (10 Marks)
- b. (i) Differentiate between Pig and MapReduce. (05 Marks)
- (ii) Write a short note on Pig architecture design layers. (05 Marks)

Module-5

- 9 a. What are outliers? Describe the reasons for the presence of outliers in a relationship. (10 Marks)
- b. How can a university student's GPA be predicted from his/her – high school percentage (HSP) of marks? (Assume linear regression) Plot a graph for the same. (10 Marks)

OR

- 10 a. Discuss the different phases of text mining process. (10 Marks)
- b. Write a short note on text mining and web mining. (05 Marks)
- c. Discuss three phases for web usage mining. (05 Marks)

* * * * *