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

18CS72

Seventh Semester B.E. Degree Examination, June/July 2025
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 discuss structured, semi-structured and unstructured data with examples. **(05 Marks)**

b. How do you classify Big Data? **(05 Marks)**

c. Describe the process of designing a data architecture using five layers model. **(10 Marks)**

OR

2 a. Explain data preprocessing and also discuss the factors which affects data quality with suitable examples. **(10 Marks)**

b. Analyze different phases in data analytics. Discuss with a neat diagram the traditional and big data analytics architecture reference model with suitable example. **(10 Marks)**

Module-2

3 a. Why are the Hadoop system and ecosystem components shown in 4 layers? Justify with an example. **(10 Marks)**

b. Explain Hadoop physical architecture with a neat diagram and various system roles in an HDFS deployment. **(10 Marks)**

OR

4 a. Discuss any five essential Hadoop tools in brief with relevant diagrams. **(10 Marks)**

b. i) Explain the working of mapReduce framework. **(04 Marks)**

ii) Discuss the details of Hadoop2 execution model. **(06 Marks)**

Module-3

5 a. Demonstrate the concept of increasing flexibility in NOSQL DBS with database for the students in various courses as an example. **(05 Marks)**

b. How do ACID and BASE properties differ? Explain each with suitable example. **(07 Marks)**

c. Explain when will you use the following:
MongoDB, Cassandra, CouchDB, OracleNOSQL and Riak with example. **(08 Marks)**

OR

6 a. Briefly describe NOSQL data architecture with relevant example. **(12 Marks)**

b. How will you consider MongoDB as a complete query language? Justify with an example. **(08 Marks)**

Module-4

7 a. Show mapReduce process diagrammatically to depict a client submitting a job, the workflow of Jobtracker and Tasktracker, and Tasktrackers creating the outputs. **(10 Marks)**

b. How does the data convert to (key, value) pairs before passing to the mapper? How do the input split and record reader function? **(10 Marks)**

OR

8 a. Describe the mapReduce processing steps to illustrate how the ACPAMS receive alerts/messages. **(08 Marks)**

b. Explain in detail Hive integration and workflow steps. **(06 Marks)**

c. Describe the pig architecture for scripts dataflow and processing also list the steps for installing pig. **(06 Marks)**

Module-5

9 a. When will you perform t-test and F-test? **(05 Marks)**

b. Write applications of Euclidean, Jaccard, Cosine, Edit and Hamming distance measures. **(08 Marks)**

c. Why does Apriori principle that 'if an item set is frequent, these all of its subsets must also be frequent' hold true? Show the apiori algorithm process. **(07 Marks)**

OR

10 a. Explain with a neat diagram the five phases of processing text. **(12 Marks)**

b. Explain in details the algorithms used to compute page rank. **(08 Marks)**

* * * * *