

USN

--	--	--	--	--	--	--	--	--	--

18AD644

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

Analysis on Big Data

Time: 3 hrs.

Max. Marks: 100

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

Module-1

- 1 a. Write notes on Analytics Scalability to Big data and Massive Parallel processing platforms. (10 Marks)
- b. What is grid computing? List and explain the features, draw backs of grid computing. (10 Marks)

OR

- 2 a. Illustrate the various phases involved in Big data Analytics with neat diagram. (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)

Module-2

- 3 a. Explain Hadoop Distributed file system. (10 Marks)
- b. Discuss the Apache Sqoop import and export methods with neat diagram. (10 Marks)

OR

- 4 a. How does the Hadoop MapReduce data flow work for a word count program? Give an example. (10 Marks)
- b. What is APACHE Flume? Describe the feature components and working of apache flume. (10 Marks)

Module-3

- 5 a. Discuss the usage of MangoDB, Cassandra, Couch DB, Oracle NoSQL and Riak. (10 Marks)
- b. Write down the steps to provide client to read and write values using Key-value store. What are the typical uses of key value stores? (10 Marks)

OR

- 6 a. Give the comparison between NoSQL and SQL/RDBMS. (05 Marks)
- b. List Pros and Cons of distribution using Sharding. (05 Marks)
- c. Discuss NoSQL data architecture pattern, with an example. (10 Marks)

Module-4

- 7 a. With a neat diagram, explain the process, in MapReduce when client submitting a Job. (10 Marks)
- b. Using Hive QL for the following :
 - i) Create a table with partition
 - ii) Add, rename and drop a partition to a table. (10 Marks)

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg. 42+8 = 50, will be treated as malpractice.

OR

- 8 a. How node failure can be handled in Hadoop? Discuss. (10 Marks)
b. Discuss the Pig Latin data types and examples. (10 Marks)

Module-5

- 9 a. Write a short note on text mining of webmining. (05 Marks)
b. List the limitations of link, Rank and web Graph Analysis. (05 Marks)
c. How are topic, identification, tracking and drift analysis done? (10 Marks)

OR

- 10 a. Write a note on :
i) Web content mining
ii) Web usage mining (10 Marks)
b. Define a Page Rank. How the cliques discover communities from social network analysis. (10 Marks)

* * * * *