

Reg. No. :

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

Question Paper Code: U8108

B.E./B.Tech. DEGREE EXAMINATION, APRIL/MAY 2025

Professional Elective

21ITV108- BIGDATA ANALYTICS

(Computer Science and Design)

(Regulations 2021)

Duration: Three hours

Maximum: 100 Marks

Answer ALL Questions

PART A - (10 x 2 = 20 Marks)

1. List out the main Characteristics of Bigdata? CO1- U
2. Write about Data Science. CO1- U
3. List the difference between NOSQL database and relational database. CO1-U
4. Explain aggregate data model with any two examples. CO1- U
5. Write a simple Map function in Python that counts the frequency of words in a text. CO2-App
6. What would be the output of the following MapReduce job for input hello world hello? CO2-App
7. What is the purpose of compression in Hadoop, and how does it benefit data processing? CO1- U
8. Name two key components of the Hadoop ecosystem and briefly describe their roles. CO1- U
9. What HiveQL command is used to add new records to an existing table, and how does it work? CO1- U
10. How does Pig simplify data processing in Hadoop? CO1- U

PART – B (5 x 16= 80 Marks)

11. (a) Explain storage considerations in bigdata. CO1-U (16)
Or
(b) (i) Illustrate the importance of big data. CO1-U (8)
(ii) List out the various challenges faced in big data in detail. CO1-U (8)

12. (a) Why NOSQL need bigdata solution,how to handle challenges of bigdata. CO1-U (16)
- Or
- (b) Describe the concept of aggregate and document data models. CO1-U (16)
13. (a) The classic Word Count program counts the number of occurrences of each word in a large text file. How would you modify this program to handle case sensitivity and ignore punctuation? Write the enhanced MapReduce code and explain your modifications. CO2-App (16)
- Or
- (b) Solve numerical problem of map reduce programming model with explanation of map reduce phases. CO2-App (16)
14. (a) What are the different scheduling policies available in YARN, and how do they impact the performance and fairness of application execution? CO1- U (16)
- Or
- (b) Describe the architecture of HDFS, focusing on the roles and responsibilities of the NameNode, DataNodes, and Secondary NameNode. How do these components interact to ensure data integrity and availability in a distributed environment? CO1- U (16)
15. (a) Design an HBase schema for a time-series database that stores sensor data from IoT devices. What factors would you consider in your design, and how would you optimize it for efficient data retrieval and storage? CO2-App (16)
- Or
- (b) You have been tasked with implementing a real-time recommendation engine using HBase to store user activity data. How would you design the data model to support fast reads and writes, and what strategies would you use to ensure data consistency and availability? CO2-App (16)