



ADMISSION NUMBER

## School of Business

Bachelor of Business Administration  
Semester End Examination - Aug 2024

Duration : 180 Minutes  
Max Marks : 100

### Sem VI - D1UF602T - Big Data Analytics

General Instructions

Answer to the specific question asked

Draw neat, labelled diagrams wherever necessary

Approved data hand books are allowed subject to verification by the Invigilator

- |     |   |        |
|-----|---|--------|
| 1)  | Identify the main configuration parameters are specified in MapReduce.  | K3(6)  |
| 2)  | A tech startup is considering migrating its traditional relational database to a more scalable and flexible NoSQL solution. What factors would you consider when evaluating different database options, and how would you recommend implementing the transition to ensure minimal disruption to the company's operations?   | K3(9)  |
| 3)  | Which one is the master node in HDFS? Can it be commodity?  | K4(4)  |
| 4)  | Analyze the different types of ingestions and recommend the best one for the real-time data analysis process.   | K4(8)  |
| 5)  | consider your company needs to analyze large volumes of streaming data from IoT devices in real-time to detect anomalies. Explain how Spark can be used to efficiently process this data and provide insights with minimal latency.   | K4(8)  |
| 6)  | Compile the role of RecordReader, Combiner, and Partitioner play in a MapReduce operation.  | K5(10) |
| 7)  | Determine the security framework for protecting data stored in Hadoop and HBase from unauthorized access.   | K5(10) |
| 8)  | Client need a database design for his blog with following specifications.<br>1 Every post has a unique title, description and url.<br>2 Every post can have one or more tags.<br>3 Every post has the name of its publisher and total number of likes.<br>4 Every post has comments given by users along with their name, message, data-time and likes.<br>5 On each post, there can be zero or more comments.<br>For this set of requirements design a schema. Justify your answer with proper detailing | K5(15) |
| 9)  | If you have an input file of 350 MB, how many input splits would HDFS create and what would be the size of each input split?  | K6(12) |
| 10) | Case Study: Real-Time Analytics & Recommendation Systems  | K6(18) |

An online streaming service wants to enhance its user experience by providing real-time recommendations to its subscribers. They are exploring the implementation of real-time analytics and recommendation systems.

### Real-Time Analytics & Recommendation Systems:

The streaming service adopts a collaborative filtering approach combined with real-time analytics to generate personalized recommendations for users based on their viewing history and preferences. They use technologies like Apache Flink and Apache Mahout for real-time data processing and recommendation generation.

#### Question:

- a) How does collaborative filtering work in generating personalized recommendations for users?(6Marks)
- b) Explain the role of real-time analytics in improving the effectiveness of recommendation systems for the streaming service.(6Marks)
- c) What are the ethical considerations the streaming service should take into account when implementing real-time recommendation systems?(6Marks)