

Roll No:

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

The Assam Royal Global University, Guwahati
Royal School of Engineering Technology (RSET)
BTech CSE, 7th Semester
Special Supplementary Examination, August 2024
Course Title: Big Data Analytics
Course Code: CSE022D704

Time: 3 Hours

Maximum Marks: 70

Note: Attempt all questions as per instructions given.
The figures in the right-hand margin indicate marks.

Section – A

1. Attempt **all** questions. (Maximum word limit 50) **2 x 8**
- a. Describe any 2 major “V’s” in Big Data.
 - b. Distinguish between analytics and reporting.
 - c. Elaborate briefly about cluster analysis.
 - d. Write a note on Kernal methods and functions.
 - e. Write a note on sampling data in a stream.
 - f. In brief, describe trends and predictions that can be derived from big data.
 - g. Discuss and elaborate on NoSQL.
 - h. Write a note on Pig Latin and the Pig Data Model.

Section – B

2. Attempt **any two** of the following: **2 x 6**
- a. Define the characteristics and applications of Big Data.
 - b. Describe with a diagram the MapReduce function.
 - c. Discuss in details the Evolution of Analytic Scalability.
3. Attempt **any two** of the following: **2 x 7**
- a. Define statistical methods and regression modelling.
 - b. Describe in details SVM classification with suitable diagrams.
 - c. From the item list given below, using association rule mining, find the lift of the following: {Apple, Orange} => {Bread}

TID	Items
1	Apple, orange, milk, bread
2	Orange, tomato, bread
3	Apple, orange, milk, bread
4	Apple, orange, tomato, bread
5	Apple, orange, tomato, milk, bread
6	Orange, tomato, milk

4. Attempt **any two** of the following:

2 x 7

- a. In details, investigate about streams and stream data model architecture.
- b. Explain about mining data streams. Elaborate on the Real Time Analytics Platform and its applications.

(3+4)

- c. Explain about real time sentiment analysis and mining time series data.

5. Attempt **any two** of the following:

2 x 7

- a. Elaborate on Hive Data Definition and HiveQL Data Manipulation.
- b. Write in details about the Cassandra data model.
- c. In details, elaborate on HBase and its implementations.