

(4)

TCS-671

(b) Discuss the evolution of data storage from RDBMS through Data Warehouses to Data Lakes. (CO1, CO2)

(c) Distinguish between Structured and Unstructured Data. What are the characteristics of Bigdata? (CO1, CO2)

TCS-671

190

H

Roll No.

TCS-671

B. TECH. (CSE)

(SIXTH SEMESTER) END SEMESTER

EXAMINATION, 2022

BIGDATA STORAGE AND PROCESSING

Time :Three Hours

Maximum Marks : 100

Note : (i) All questions are compulsory.

(ii) Answer any *two* sub-questions among

(a), (b) and (c) in each main question.

(iii) Total marks in each main question are **twenty**.

(iv) Each question carries 10 marks.

P. T. O.

(2)

TCS-671

1. (a) What is the CAP theorem ? How are NoSQL databases classified based on the two CAP characteristics they support ?

(CO5)

- (b) Summarize the HBase Read and Write process. (CO5)

- (c) What are the components of HBase ? Explain the tasks of each component.

(CO5)

2. (a) Distinguish between Standalone Mode, Pseudo-distributed Mode and Fully Distributed Mode as related to Hadoop.

(CO5)

- (b) Enumerate the steps of Hadoop installation on any OS of your choice. (CO5)

- (c) Are hadoop clusters a good solution for all types of data analysis requirements ? Support your answer with examples.(CO5)

(3)

TCS-671

3. (a) Discuss all the phases of a Map-Reduce task. (CO4)

- (b) Write a MapReduce code for "Word Count". Also explain the code. (CO4)

- (c) What is Data Serialization ? Discuss Serialization aspects in Hadoop. (CO4)

4. (a) Discuss the architectural components of Hadoop. (CO3)

- (b) Summarize some of the most common file formats for the Hadoop ecosystem ? Why should you choose a good format ? (CO3)

- (c) What are the internal steps that happen during Hadoop read and write ? (CO3)

5. (a) Taking case studies discuss two different strategies of collecting data. (CO1, CO2)

P. T. O.