

Question Paper

Exam Date & Time: 20-Jun-2024 (02:30 PM - 05:30 PM)



MANIPAL ACADEMY OF HIGHER EDUCATION

SIXTH SEMESTER B.TECH MAKEUP EXAMINATIONS, JUNE 2024

BIG DATA ANALYTICS [CSE 3272]

Marks: 50

Duration: 180 mins.

Answer all the questions.

Instructions to Candidates: Answer ALL questions Missing data may be suitably assumed

- 1) Summarize different types of RDDs with code example. (5)
 - A)
 - B) Summarize properties of Hadoop. (3)
 - C) Draw a diagram depicting anatomy of HDFS read operation. (2)
- 2) List and explain pair RDD transformations. Write code example. (5)
 - A)
 - B) Outline advantages of matrix factorization. (3)
 - C) Write a program in PySpark to count the number of lines present in a file. (2)
- 3) How you choose best K in K-Means clustering. Write the code for the same. (5)
 - A)
 - B) What is one hot encoding? The weather can be sunny, rainy or windy represent it using one hot encoding. (3)
 - C) Compare count function with countByKey function. (2)
- 4) Outline the advantages of sliding window over state operations. Draw the diagram of a sliding window with window length of 2 and sliding interval of 2. Write the code for word count example with sliding window. (5)
 - A)
 - B) Summarize the components of GraphX. (3)
 - C) What is the role of inverse function in a sliding window? (2)
- 5) Outline different set transformations in PySpark. (5)
 - A)
 - B) Write the code for word count with Dstream transformation (3)
 - C) Write the code for decision tree. (2)

-----End-----