

# Question Paper

Exam Date & Time: 11-May-2023 (09:30 AM - 12:30 PM)



## MANIPAL ACADEMY OF HIGHER EDUCATION

### INTERNATIONAL CENTRE FOR APPLIED SCIENCES END SEMESTER THEORY EXAMINATION-MAY 2023 IV SEMESTER B.Sc.(APPLIED SCIENCES) IN ENGG.

#### BIG DATA ANALYTICS [ICS 245]

**Marks: 50**

**Duration: 180 mins.**

**Answer all the questions.**

**Missing data if any, may be suitably assumed.**

- 1) With a neat diagram, explain the anatomy of File write in HDFS. (5)
  - A)
  - B) With a suitable diagram, explain the anatomy of YARN Application run. (5)
- 2) With a neat diagram, list the steps of Spark application running in YARN client mode. (5)
  - A)
  - B) Given an input file containing temperature recorded in a city, Write a pyspark program to find the maximum and minimum temperature. (5)
- 3) With a brief introduction to matrix factorization model, discuss the Alternating Least Squares recommender algorithm. (5)
  - A)
  - B) What is a feature vector? Illustrate with an example. (5)
- 4) What is anomaly detection? Explain. Mention three real life applications of anomaly detection. (5)
  - A)
  - B) What are the design goals of Spark Streaming? Explain. Also mention the advantage of Spark Streaming. (5)
- 5) Compare and contrast Map Reduce with traditional RDBMS. (5)
  - A)
  - B) Write short notes on following: (5)
    - i) Approaches for calculating the financial statistic, VaR
    - ii) Degree Distribution

-----End-----