QP CODE: 2242202946

2242202946

Reg No : .....

Name : .....

M.C.A DEGREE EXAMINATION, NOVEMBER 2022

Second Semester

MASTER OF COMPUTER APPLICATION

CORE - MCACT204 - DATA SCIENCE & BIG DATA ANALYSIS

2020 Admission Onwards

3B3DACB6

Time: 3 Hours Maximum: 75 Marks

Part A

*Answer any **ten** questions*

*Each question carries **3** marks*

1. What motivated Data Mining? why is it important?
2. Illustrate the technologies used in datamining
3. Basic Methods of Attribute subset selection
4. What are strong association rules?
5. Briefly explain closed frequent itemset mining.
6. Discuss partitioning method of clustering.
7. What is the purpose of Web Analytics?
8. Explain the different applications of Big Data?
9. Compare distributed and parallel system?
10. Define Name node and Data Node ?
11. What is the relationship between big data and a data warehouse?
12. Write any 3 built-in functions and 3 aggregate functions of Hive

(10×3=30 marks)

Part B

*Answer **all** questions*

*Each question carries **9** marks*

13. a) Explain the Architecture of a typical data mining system with diagram

OR

b) What is Data Reduction ? Briefly explain different methods used in data Reduction

14. a) Illustrate the Apriori Principle with an example

OR

b) Discuss the requirements of cluster analysis in Data mining

15. a) Explain Web Analytics and Credit risk management

OR

b) Explain Big Data in healthcare?

16. a) Explain:- i) Structuring Big data ii) Elements of Big data iii) Big Data Analytics

OR

b) what is HDFS? Diagrammatically explain the architecture of HDFS along with Heartbeat Mechanism.

17. a) What is YARN? Explain the working of YARN.

OR

b) Write notes on a) Any 9 built-in functions of Hive b) Any 9 aggregate functions of Hive

(5×9=45 marks)