MCS–221

MASTER OF COMPUTER APPLICATIONS (MCAOL)

DATA WAVEHOUSING AND DATA MINING

Time: Three Hours

Maximum Marks: 100

Note: This question paper consists of three Sections.

Section–I (Short Answer) (5×4=20)

Note: Attempt any five questions from the following.

1. Describe the Top-Down Approach of building a Data Warehouse.

2. Discuss briefly the four essential characteristics of a Data Warehouse.

3. Explain briefly the cloud based Data Warehouse architecture.

4. What are Aggregate Tables ? Explain the need for building aggregate fact tables.

5. What is OLAP? List and briefly explain its characteristics.

6. What is Data Mining? Explain briefly the Data Mining Life cycle.

7. Explain the concept of Market Analysis.

Section–B (Medium Answer) (5×10=50)

Note: Attempt any five questions from the following.

8. What is Association Rule Mining (ARM)? With reference to ARM, explain the following concepts:

a) Frequent Itemset

b) Support count

c) Support

d) Confidence

9. What is Data Cleaning? Explain how the missing handled while data cleaning.

10. Explain the following attribute selection measures:

a) Entropy

b) Information Gain

c) Gain Ratio

d) Gini Index

11. Explain the Divisive Approach of Hierarchical method of clustering with the help of an example.

12. Explain the following four layers of data warehouse architecture:

a) Data source layer

b) Data staging layer

c) Data storage layer

d) Data presentation layer

13. Explain the star schema of Dimensional modelling. Give an example star-schema.

14. Explain the Anchor model and Data Vault models which are categorized as complex data modeling techniques.

Section–C (Long Answer) (2×15=30)

Note: Attempt any two questions from the following.

15. Explain the following OLAP architectures:

a) ROLAP

b) MOLAP

c) HOLAP

16. Explain the following text preprocessing techniques:

a) Segmentation

b) Tokenization

c) Normalization

17. Write and explain the k-means algorithm of clustering with the help of an example.