

MASTER OF COMPUTER APPLICATIONS (MCAOL)

1) ETL — Definition + Steps + Layered Implementation

Repeated 5 times

Appears as:

- What is ETL?
- Need of ETL
- Steps of ETL
- Layered ETL implementation
- ETL vs ELT

Years: 2023J, 2023D, 2024J, 2024D, 2025J

This is the MOST repeated topic in all 5 papers.

2) Apriori Algorithm / Frequent Pattern Mining

Repeated 5 times

- Write and explain Apriori algorithm
- Frequent pattern mining classifications
- Market basket analysis (connected)

Years: 2023J, 2023D, 2024D, 2024J, 2025J

3) OLAP — Cube Operations (Roll-up, Drill-down, Slice, Dice)

Repeated 4 times

Topics recurring:

- OLAP definition
- OLAP data cube operations
- Roll-up, Drill-down
- Slice, Dice
- Applications of OLAP

Years: 2023J, 2024J, 2024D, 2025J

MASTER OF COMPUTER APPLICATIONS (MCAOL)

4) Schema Modeling — Star, Snowflake, Fact Constellation

Repeated 4 times

Appears as:

- Draw star schema
- Draw snowflake schema
- Fact constellation
- Characteristics + advantages/limitations

Years: 2023D, 2024J, 2024D, 2025J

5) Clustering Algorithms (K-Means / K-NN / DBSCAN)

Repeated 5 times

What repeats:

- Define clustering
- K-means clustering
- K-NN classification
- DBSCAN clustering
- Advantages & disadvantages

Years: 2023J, 2023D, 2024J, 2024D, 2025J

6) Data Preprocessing Stages

Repeated 4 times

Includes:

- Data cleaning
- Data integration
- Data reduction
- Data transformation

MASTER OF COMPUTER APPLICATIONS (MCAOL)

- Noisy data + binning

Years: 2023J, 2023D, 2024D, 2025J (indirect)

7) Data Lake vs Data Warehouse + Data Lake Stages

Repeated 3 times

Appears as:

- Data lake vs Data warehouse
- Data lake stages (Puddle, Pond, Lake, Ocean)
- Data lake architecture

Years: 2023J, 2023D, 2025J

8) Data Mart — Structure / Difference from Warehouse

Repeated 3 times

Years: 2023J, 2024J, 2024D, 2025J

9) Classification Algorithms (Decision Tree, K-NN, Rule-based)

Repeated 4 times

Includes:

- Decision tree classifier
- K-NN
- Rule-based classification

Years: 2023D, 2024J, 2024D, 2025J (frequent pattern)

10) Text Mining / Vector Space Model / TF-IDF

Repeated 2–3 times

Years: 2024J, 2024D

MASTER OF COMPUTER APPLICATIONS (MCAOL)

11) Data Warehouse Architectures

Repeated 3 times

Includes:

- Real-time DW architecture
- Single-tier, three-tier architecture
- Hadoop DW conceptual architecture
- Top-down (Inmon) vs Bottom-up (Kimball)

Years: 2023D, 2024J, 2024D, 2025J (Inmon/Kimball)

12) Data Mining Basics & Techniques

Repeated 3 times

Topics:

- What is data mining
- ARM, Outlier detection, Regression
- Popular data mining tools

Years: 2023D, 2024D, 2025J

13) Data Integration Issues

Repeated 3 times

Years: 2024J, 2024D, 2025J (indirect under preprocessing)

14) K-Means / K-NN / K-Medoids

Repeated 3–4 times

Appears continuously 2023–2025.

You cannot skip clustering.

15) Metadata & Data Granularity (DW Concepts)

MASTER OF COMPUTER APPLICATIONS (MCAOL)

Repeated 2–3 times

Years: 2024J, 2024D, 2025J (metadata in DW lifecycle)

THE MOST IMPORTANT (REPEATED EVERY YEAR)

1. **ETL (steps, need, layered design, ETL vs ELT)**
2. **Apriori Algorithm + Frequent Pattern Mining**
3. **Star / Snowflake / Fact Constellation schemas**
4. **Clustering (K-means, K-NN, DBSCAN) + differences**
5. **OLAP operations (Roll-up, Drill-down, Slice, Dice)**
6. **Data Preprocessing Stages + Noisy Data + Binning**
7. **Data Lake vs Data Warehouse + lake stages**
8. **Data Mart + Architecture + Differences**