# FACULTY OF EGINEERING

# DATA MINING & WAREHOUSEING
# LECTURE-23
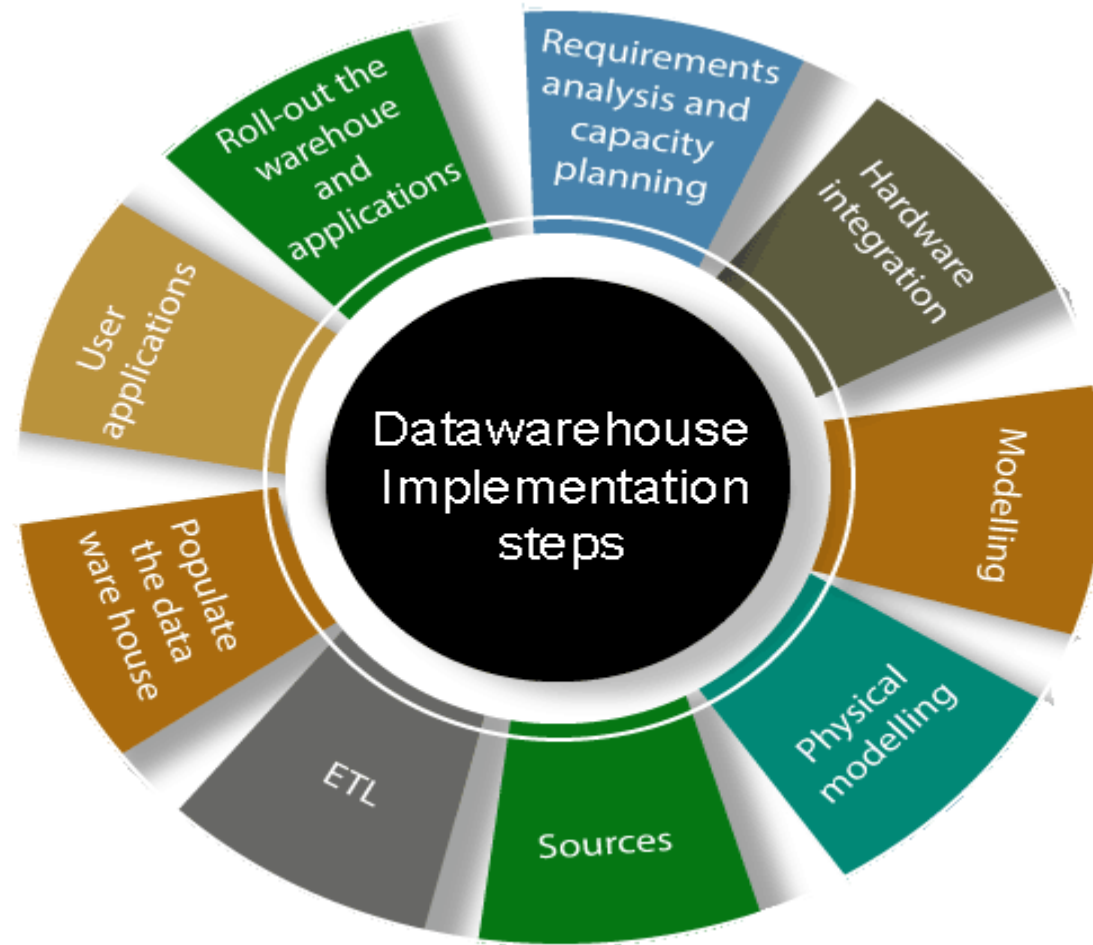
# MR. DHIRENDRA
## ASSISTANT PROFESSOR
## RAMA UNIVERSITY

# OUTLINE

- ❖ **DATA WAREHOUSE IMPLEMENTATION**

- ❖ **IMPLEMENTATION GUIDELINES**

- ❖ **MCQ**

- ❖ **REFERENCES**

# Data Warehouse Implementation

# Data Warehouse Implementation

**Requirements analysis and capacity planning:** The first process in data warehousing involves defining enterprise needs, defining architectures, carrying out capacity planning, and selecting the hardware and software tools. This step will contain be consulting senior management as well as the different stakeholder.

**Hardware integration:** Once the hardware and software has been selected, they require to be put by integrating the servers, the storage methods, and the user software tools.

**Modeling:** Modelling is a significant stage that involves designing the warehouse schema and views. This may contain using a modeling tool if the data warehouses are sophisticated.

**Physical modeling:** For the data warehouses to perform efficiently, physical modeling is needed. This contains designing the physical data warehouse organization, data placement, data partitioning, deciding on access techniques, and indexing.

# Data Warehouse Implementation

**Sources:** The information for the data warehouse is likely to come from several data sources. This step contains identifying and connecting the sources using the gateway, ODBC drives, or another wrapper.

**ETL:** The data from the source system will require to go through an ETL phase. The process of designing and implementing the ETL phase may contain defining a suitable ETL tool vendors and purchasing and implementing the tools. This may contains customize the tool to suit the need of the enterprises.

**Populate the data warehouses:** Once the ETL tools have been agreed upon, testing the tools will be needed, perhaps using a staging area. Once everything is working adequately, the ETL tools may be used in populating the warehouses given the schema and view definition.

**User applications:** For the data warehouses to be helpful, there must be end-user applications. This step contains designing and implementing applications required by the end-users.

**Roll-out the warehouses and applications:** Once the data warehouse has been populated and the end-client applications tested, the warehouse system and the operations may be rolled out for the user's community to use.
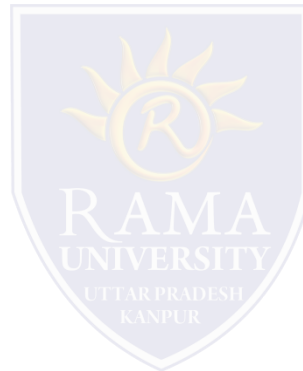
**Build incrementally:** Data warehouses must be built incrementally. Generally, it is recommended that a data marts may be created with one particular project in mind, and once it is implemented, several other sections of the enterprise may also want to implement similar systems. An enterprise data warehouses can then be implemented in an iterative manner allowing all data marts to extract information from the data warehouse.

**Need a champion:** A data warehouses project must have a champion who is active to carry out considerable researches into expected price and benefit of the project. Data warehousing projects requires inputs from many units in an enterprise and therefore needs to be driven by someone who is needed for interacting with people in the enterprises and can actively persuade colleagues.

**Senior management support:** A data warehouses project must be fully supported by senior management. Given the resource-intensive feature of such project and the time they can take to implement, a warehouse project signal for a sustained commitment from senior management.

# Implementation Guidelines

**Ensure quality:** The only record that has been cleaned and is of a quality that is implicit by the organizations should be loaded in the data warehouses.

**Corporate strategy:** A data warehouse project must be suitable for corporate strategies and business goals. The purpose of the project must be defined before the beginning of the projects.

**Business plan:** The financial costs (hardware, software, and peopleware), expected advantage, and a project plan for a data warehouses project must be clearly outlined and understood by all stakeholders. Without such understanding, rumors about expenditure and benefits can become the only sources of data, subversion the projects.

**Training:** Data warehouses projects must not overlook data warehouses training requirements. For a data warehouses project to be successful, the customers must be trained to use the warehouses and to understand its capabilities.

# Implementation Guidelines

**Adaptability:** The project should build in flexibility so that changes may be made to the data warehouses if and when required. Like any system, a data warehouse will require to change, as the needs of an enterprise change.

**Joint management:** The project must be handled by both IT and business professionals in the enterprise. To ensure that proper communication with the stakeholder and which the project is the target for assisting the enterprise's business, the business professional must be involved in the project along with technical professionals.

# Multiple Choice Question

1. _____ maps data into predefined groups.
a) Regression
b) Time series analysis
c) Prediction
d) Classification

2.. _____ is used to map a data item to a real valued prediction variable.
a) Regression
b) Time series analysis.
c) Prediction.
d) Classification

3. In _____, the value of an attribute is examined as it varies over time.
a) Regression
b) Time series analysis.
c) Sequence discovery.
d) Prediction

4. In _____ the groups are not predefined.
a) Association rules.
b) Summarization
c) Clustering
d) Prediction

5. Link Analysis is otherwise called as _____.
a) affinity analysis.
b) association rules.
c) both A & B.
d) Prediction.
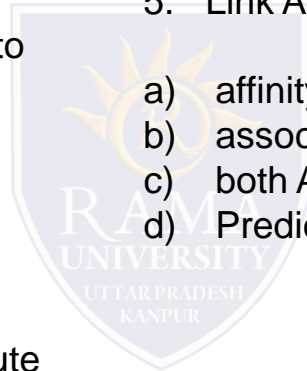
# REFERENCES

- https://www.tutorialspoint.com/dwh/dwh_overview.htm

- http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf   DATA MINING BOOK WRITTEN BY Micheline Kamber

- https://www.javatpoint.com/three-tier-data-warehouse-architecture

- M.H. Dunham, " Data Mining: Introductory & Advanced Topics" Pearson Education

- Jiawei Han, Micheline Kamber, " Data Mining Concepts & Techniques" Elsevier

- Sam Anahory, Denniss Murray," data warehousing in the Real World: A Practical Guide for Building Decision Support Systems, " Pearson Education

- Mallach," Data Warehousing System", TMH

- R. Agrawal, A. Gupta, and S. Sarawagi. Modeling multidimensional databases. ICDE'97 S. Chaudhuri and U. Dayal. An overview of data warehousing and OLAP technology. ACM SIGMOD Record, 26:65-74, 1997

- S. Agarwal, R. Agrawal, P. M. Deshpande, A. Gupta, J. F. Naughton, R. Ramakrishnan, and S. Sarawagi. On the computation of multidimensional aggregates. VLDB'96 D. Agrawal, A. E. Abbadi, A. Singh, and T. Yurek. Efficient view maintenance in data warehouses. SIGMOD'97

- E. F. Codd, S. B. Codd, and C. T. Salley. Beyond decision support. Computer World, 27, July 1993.

- J. Gray, et al. Data cube: A relational aggregation operator generalizing group-by, cross-tab and sub-totals. Data Mining and Knowledge Discovery, 1:29-54, 1997.