



**RAMA**  
**UNIVERSITY**

[www.ramauniversity.ac.in](http://www.ramauniversity.ac.in)

**FACULTY OF ENGINEERING**

**DATA MINING & WAREHOUSEING**  
**LECTURE-09**

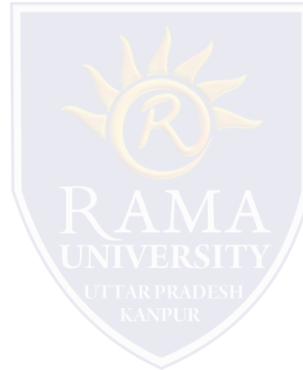
**MR. DHIRENDRA**

ASSISTANT PROFESSOR

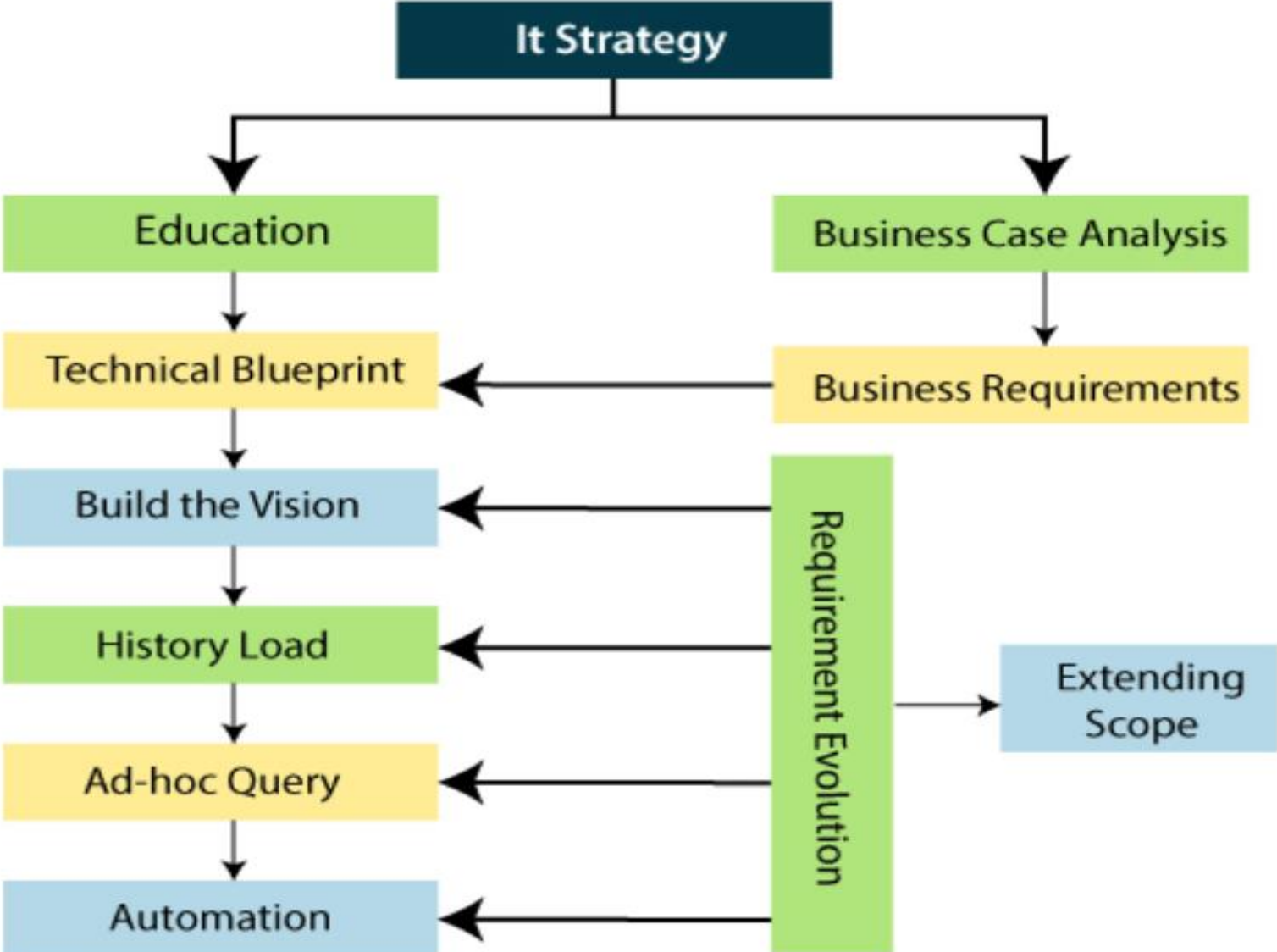
RAMA UNIVERSITY

# OUTLINE

- ❖ THE DATA WAREHOUSE DELIVERY PROCESS
- ❖ THREE DATA WAREHOUSE MODELS
- ❖ DATA WAREHOUSE BACK-END TOOLS AND UTILITIES
- ❖ MCQ
- ❖ REFERENCES



# The Data Warehouse Delivery process



# The Data Warehouse Delivery process

- **IT Strategy** : DWH project must include IT strategy for procuring and retaining funding.
- **Business Case Analysis** : It is necessary to understand the level of investment that can be justified and to identify the projected business benefits that should be derived from using DWH.
- **Education & Prototyping** : Organizations will experiment with the concept of data analysis and educate themselves on the value of DWH. This is valuable and should be considered if this is the organization's first exposure to the benefits of DS information. The prototyping activity can progress the growth of education. It is better than working models. Prototyping require business requirement, technical blueprint ,architecture.
- **Business Requirement** : It includes such as
  - The logical model for information within the DWH.
  - The source system that provide this data( mapping rules)
  - The business rules to be applied to data.
  - The query profiles for the immediate requirement

# The Data Warehouse Delivery process

- **Technical Blueprint** : It must identify :

  - The overall system architecture

  - The server & data mart architecture for both data & applications.

  - The essential components of the data base design.

  - The data retention strategy

  - The backup & recovery strategy

  - The capacity plan for H/w and infrastructure



- **Building the vision** : It is the stage where the first production deliverable is produced. This stage will probably build the major infrastructure components for extracting and loading data, but limit them to the extraction & load of data sources.
- **History Load** : In most cases, the next phase is one where the remainder of the required history is loaded into the DWH. This means that new entities would not be added to the DWH, but additional physical tables would probably be created to store the increased data volumes.
- **AD-Hoc Query** : In this process we configure an ad-hoc query tool to operate against the DWH. These end-user access tools are capable of automatically generating the database query that answer any question posed by the user.

# The Data Warehouse Delivery process

- **Automation** : The automation phase is where many of the operational management processes are fully automated

within the DWH. These would include:

- Extracting & loading the data from a variety of sources systems
- Transforming the data into a form suitable for analysis
- Backing up, restoring & archiving data
- Generating aggregations from predefined definitions within DWH
- Monitoring query profiles & determining the appropriate aggregations to
- maintain system performance.



- **Extending Scope** : In this phase, the scope of DWH is extended to address a new set of business requirements.

This involves the loading of additional data sources into the DWH i.e. introduction of new data marts .

- **Requirements evolution** : The most important aspect of the delivery process is that the requirements are never static. Business requirements will constantly change during the life of the DWH, so it is imperative that the process support this, and allows these changes to be reflected within the system.

# Three Data Warehouse Models

- **Enterprise warehouse**

- collects all of the information about subjects spanning the entire organization

- **Data Mart**

- a subset of corporate-wide data that is of value to a specific groups of users. Its scope is confined to specific, selected groups, such as marketing data mart Independent vs. dependent (directly from warehouse) data mart

- **Virtual warehouse**

- A set of views over operational databases Only some of the possible summary views may be materialized



# Data Warehouse Back-End Tools and Utilities

- **Data extraction**

get data from multiple, heterogeneous, and external sources

- **Data cleaning**

detect errors in the data and rectify them when possible

- **Data transformation**

convert data from legacy or host format to warehouse format

- **Load**

sort, summarize, consolidate, compute views, check integrity, and build indices and partitions

- **Refresh**

propagate the updates from the data sources to the warehouse





# Multiple Choice Question

1. The extract process is \_\_\_\_\_.
  - a) capturing all of the data contained in various operational systems.
  - b) capturing a subset of the data contained in various operational systems.
  - c) capturing all of the data contained in various decision support systems.
  - d) capturing a subset of the data contained in various decision support systems.
- 2.. Data scrubbing is \_\_\_\_\_.
  - a) a process to reject data from the data warehouse and to create the necessary indexes.
  - b) a process to load the data in the data warehouse and to create the necessary indexes.
  - c) a process to upgrade the quality of data after it is moved into a data warehouse.
  - d) a process to upgrade the quality of data before it is moved into a data warehouse
3. The load and index is \_\_\_\_\_.
  - a) a process to reject data from the data warehouse and to create the necessary indexes.
  - b) a process to load the data in the data warehouse and to create the necessary indexes.
  - c) a process to upgrade the quality of data after it is moved into a data warehouse.
  - d) a process to upgrade the quality of data before it is moved into a data warehouse.
4. Data transformation includes \_\_\_\_\_.
  - a) a process to change data from a detailed level to a summary level.
  - b) a process to change data from a summary level to a detailed level.
  - c) joining data from one source into various sources of data.
  - d) separating data from one source into various sources of data.
5. \_\_\_\_\_ is called a multifield transformation.
  - a) Converting data from one field into multiple fields.
  - b) Converting data from fields into field.
  - c) Converting data from double fields into multiple fields.
  - d) Converting data from one field to one field.

# REFERENCES

- [https://www.tutorialspoint.com/dwh/dwh\\_overview.htm](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

