DATA WAREHOUSING
LECTURE 3

ENGR. MADEHA MUSHTAQ
DEPARTMENT OF COMPUTER SCIENCE
IQRA NATIONAL UNIVERSITY

MULTIPLE DATA TYPES OF DECISION SUPPORT SYSTEMS

- There are different types of data that need to be integrated in the data warehouse. It can be divided into two categories:
  - Structured Data
  - Unstructured Data

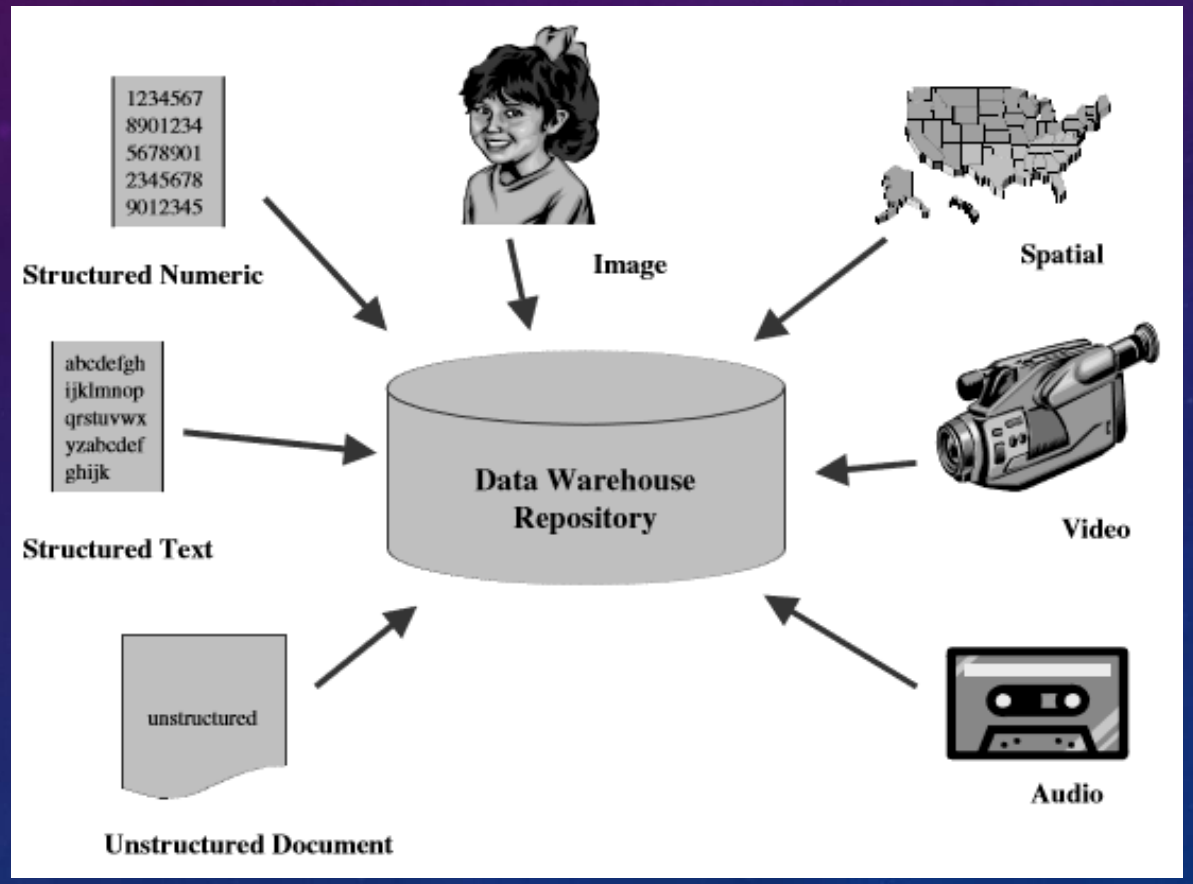
STRUCTURED VS UNSTRUCTURED DATA

- Structured data:
  - Structured data is comprised of clearly defined data types whose pattern makes them easily searchable.
  - Structured data usually resides in relational databases (RDBMS).
  - Examples of Structured Data are fields store length-delineated data phone numbers, social Security numbers, or ZIP codes.

STRUCTURED VS UNSTRUCTURED DATA

- Unstructured Data:
  - Unstructured data is essentially everything else.
  - Unstructured data is not easily searchable.
  - Unstructured data has internal structure but is not structured via pre-defined data models or schema. It may be textual or non-textual, and human- or machine-generated.
  - It may also be stored within a non-relational database like NoSQL.
  - Examples of Unstructured Data: Email messages, videos, photos, audio files.

STRUCTURED VS UNSTRUCTURED DATA



SEARCHING UNSTRUCTURED DATA

- After adding unstructured data, the next big challenge is the ability to search unstructured data.
- Vendors are now providing new search engines to find the information the user needs from unstructured data.
- Query by image content is an example of a search mechanism for images. The product allows you to pre-index images based on shapes, colors, and textures.

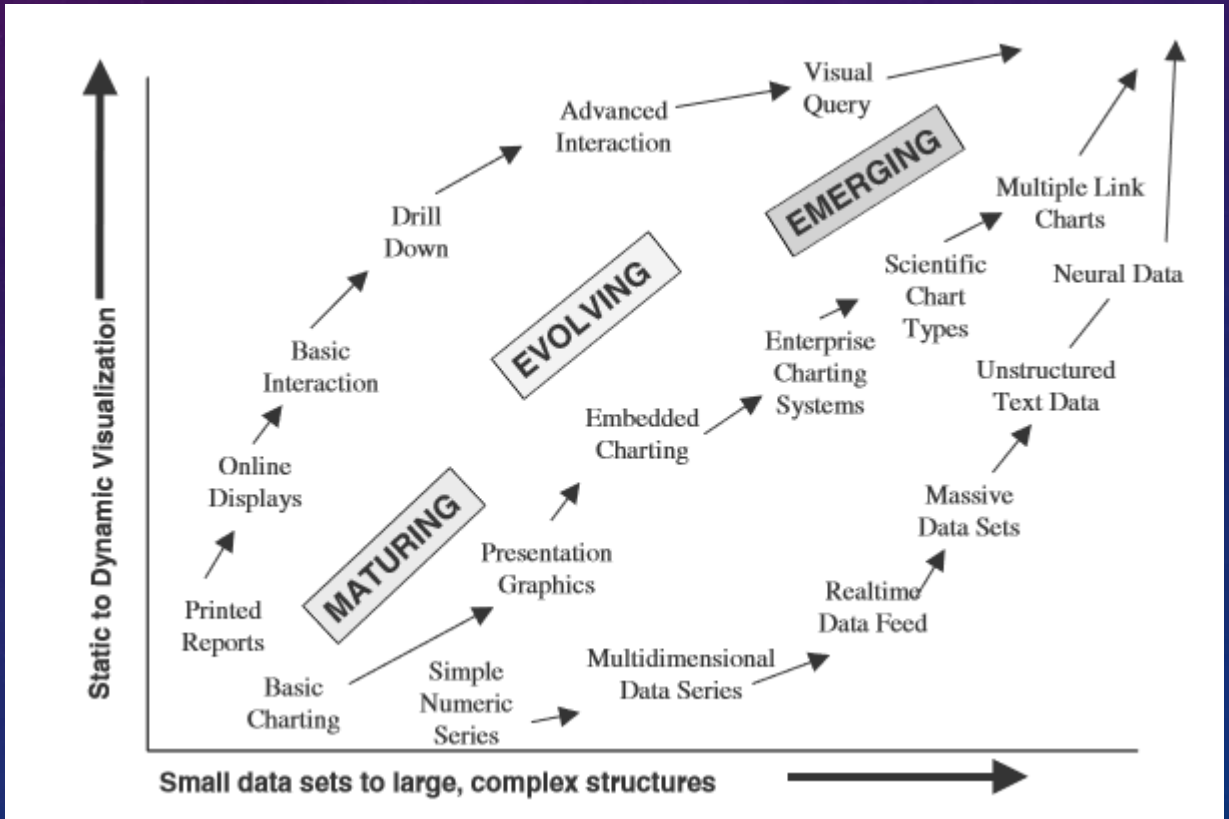
DATA VISUALIZATION

- Our DWH will be considered outdated, if it will display results only in the form of output lists or spreadsheets.
- We need to display results in the form of graphics and charts as well.
- Visualization of data in the result sets boosts the process of analysis for the user, especially when the user is looking for trends over time.
- Data visualization helps the user to interpret query results quickly and easily.

MAJOR VISUALIZATION TRENDS

- Most data visualizations are in the form of some standard chart type. The numerical results are converted into a pie chart, a scatter plot, or another chart type. Now the list of chart types supported by data visualization software has grown much longer.
- Visualizations are no longer static. Dynamic chart types are themselves user interfaces. Users can now review a result chart, manipulate it, and then see newer views online.
- Newer visualization software can visualize thousands of result points and complex data structures.

MAJOR VISUALIZATION TRENDS



ADVANCED VISUALIZATION TECHNIQUES

- The most remarkable advance in visualization techniques is the transition from static charts to dynamic interactive presentations.
  - Chart Manipulation.
  - Drill Down.
  - Advanced Interaction.

ADVANCED VISUALIZATION TECHNIQUES

- Chart Manipulation.
  - A user can rotate a chart or dynamically change the chart type to get a clearer view of the results.
  - With complex visualization types such as constellation and scatter plots, a user can select data points with a mouse and then move the points around to clarify the view.
- Drill Down.
  - The visualization first presents the results at the summary level.
  - The user can then drill down the visualization to display further visualizations at subsequent levels of detail.

ADVANCED VISUALIZATION TECHNIQUES

- Advanced Interaction.
  - The user simply double clicks a part of the visualization and then drags and drops representations of data entities.
  - Visual query is the most advanced of user interaction features.
  - For example, the user may see the outlying data points in a scatter plot, then select a few of them with the mouse and ask for a brand new visualization of just those selected points.

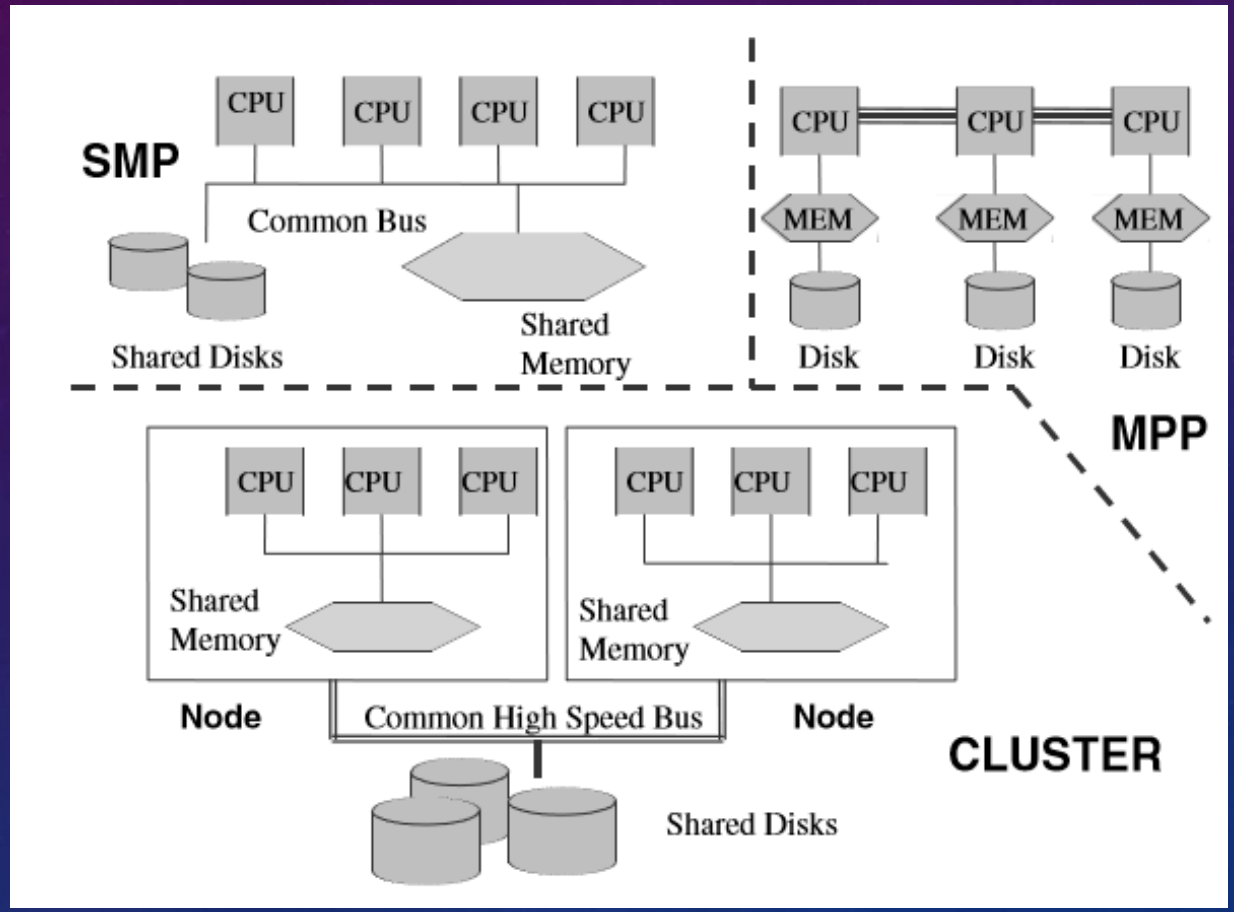
PARALLEL PROCESSING

- A task is divided into smaller units and these smaller units are executed concurrently.
- We need parallel processing to speed up query processing, data loading, and index creation.
- Both hardware configurations and software techniques go hand in hand to accomplish parallel processing.

PARALLEL PROCESSING HARDWARE OPTIONS

- In a parallel processing environment, we will find these characteristics: multiple CPUs, memory modules, one or more server nodes, and high-speed communication links between interconnected nodes.
- Essentially, we can choose from three architectural options.

PARALLEL PROCESSING HARDWARE OPTIONS



PARALLEL PROCESSING SOFTWARE IMPLEMENTATION

- Hardware alone would be worthless if the operating system and the database software cannot make use of the parallel features of the hardware.
- We will have to ensure that the software can allocate units of a larger task to the hardware components appropriately.

PARALLEL PROCESSING SOFTWARE IMPLEMENTATION

- Parallel processing software must be capable of performing the following steps:
  - Analyzing a large task to identify independent units that can be executed in parallel
  - Identifying which of the smaller units must be executed one after the other
  - Executing the independent units in parallel and the dependent units in the proper sequence
  - Collecting, collating, and consolidating the results returned by the smaller units.

DATA FUSION

- Data fusion is a technology dealing with the merging of data from disparate sources.
- The principles and techniques of data fusion technology have a direct application in data warehousing.
- In present-day warehouses, we tend to collect data in astronomical proportions.
- The more information stored, the more difficult it is to find the right information at the right time. Data fusion technology is expected to address this problem also.

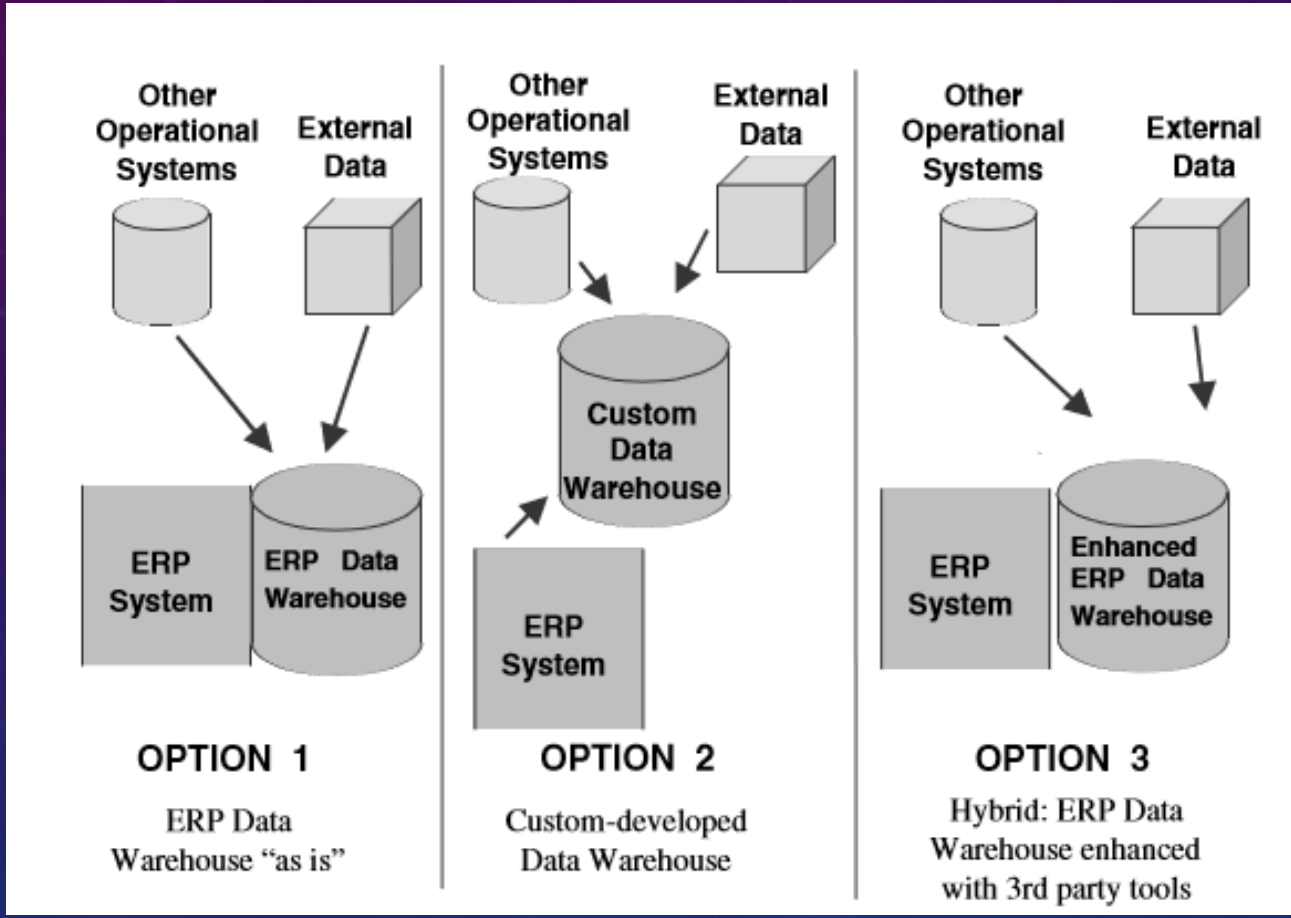
DATA WAREHOUSING AND ERP

- In the last few years, many businesses are adopting ERP (enterprise resource planning) application packages.
- A remarkable feature of an ERP package is that it supports practically every phase of the day-to-day business of an enterprise, from inventory control to customer billing, from human resources to production management.

DATA WAREHOUSING AND ERP

- However soon companies implementing ERP realized that the thousands of relational database tables, designed and normalized for running the business operations, were not at all suitable for providing strategic information.
- Moreover, ERP data repositories lacked data from external sources and from other operational systems in the company.
- If our company has ERP or is planning to get into ERP, we need to consider the integration of ERP with data warehousing.

DATA WAREHOUSING AND ERP



ERP and data warehouse integration options

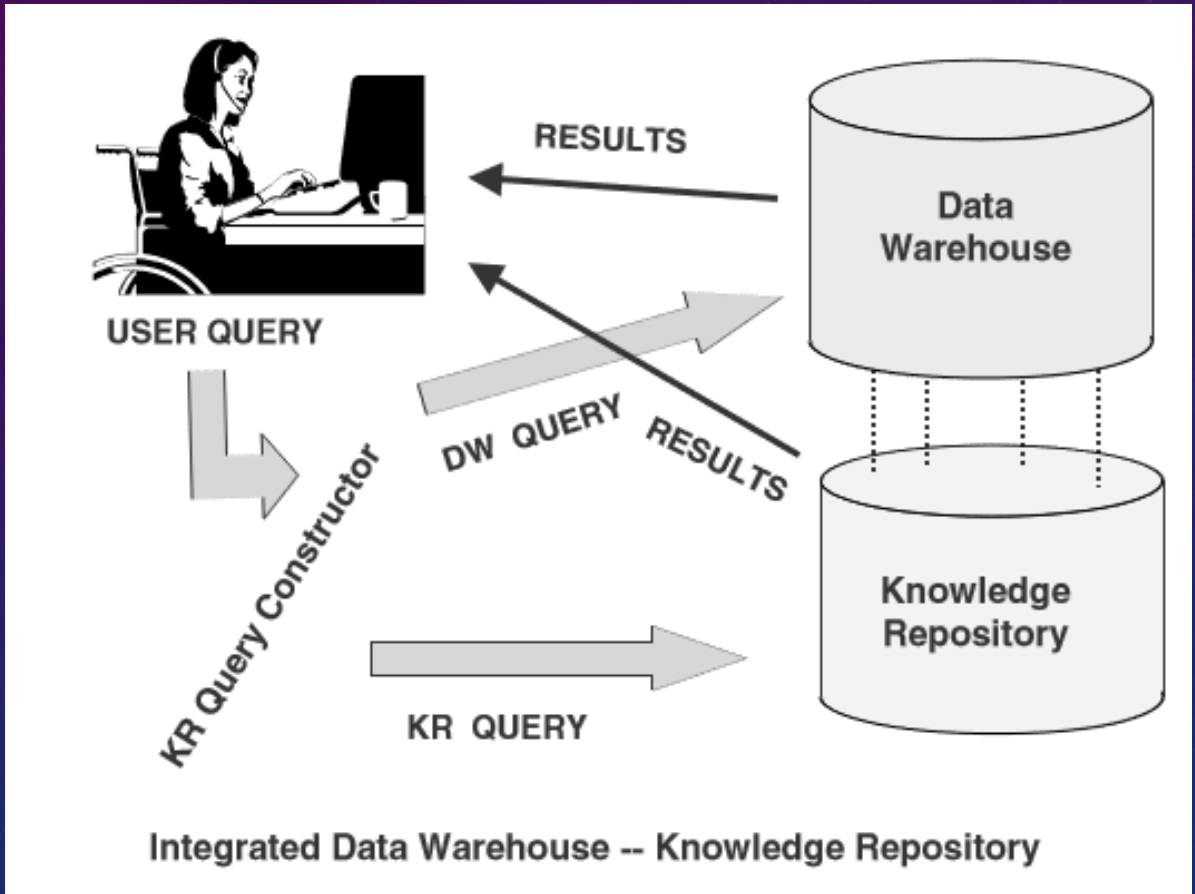
KNOWLEDGE MANAGEMENT

- Knowledge Management is a systematic process for capturing, integrating, organizing, and communicating knowledge accumulated by employees.
- It is a vehicle to share corporate knowledge so that the employees may be more effective and be productive in their work.
- Where does the knowledge exist in a corporation? Corporate procedures, documents, reports analyzing exception conditions, objects, math models, what-if cases, text streams, video clips—all of these and many more such instruments contain corporate knowledge.

DATA WAREHOUSING AND KM

- With technological advances in organizing, searching, and retrieval of unstructured data, more knowledge philosophy will enter into data warehousing.
- Figure shows how we can extend our data warehouse to include retrievals from the knowledge repository that is part of the knowledge management framework of our company.

DATA WAREHOUSING AND KM



Integration of KM and data warehouse

DATA WAREHOUSING AND CRM

- Companies are moving away from mass marketing to one-on-one marketing.
- Customer loyalty programs have become the norm.
- More and more companies are embracing customer relationship management (CRM) systems.
- When our company is gearing up to be more attuned to high levels of customer service, we will have to make our data warehouse CRM-ready, not an easy task by any means. In spite of the difficulties, the payoff from a CRM-ready data warehouse is substantial.

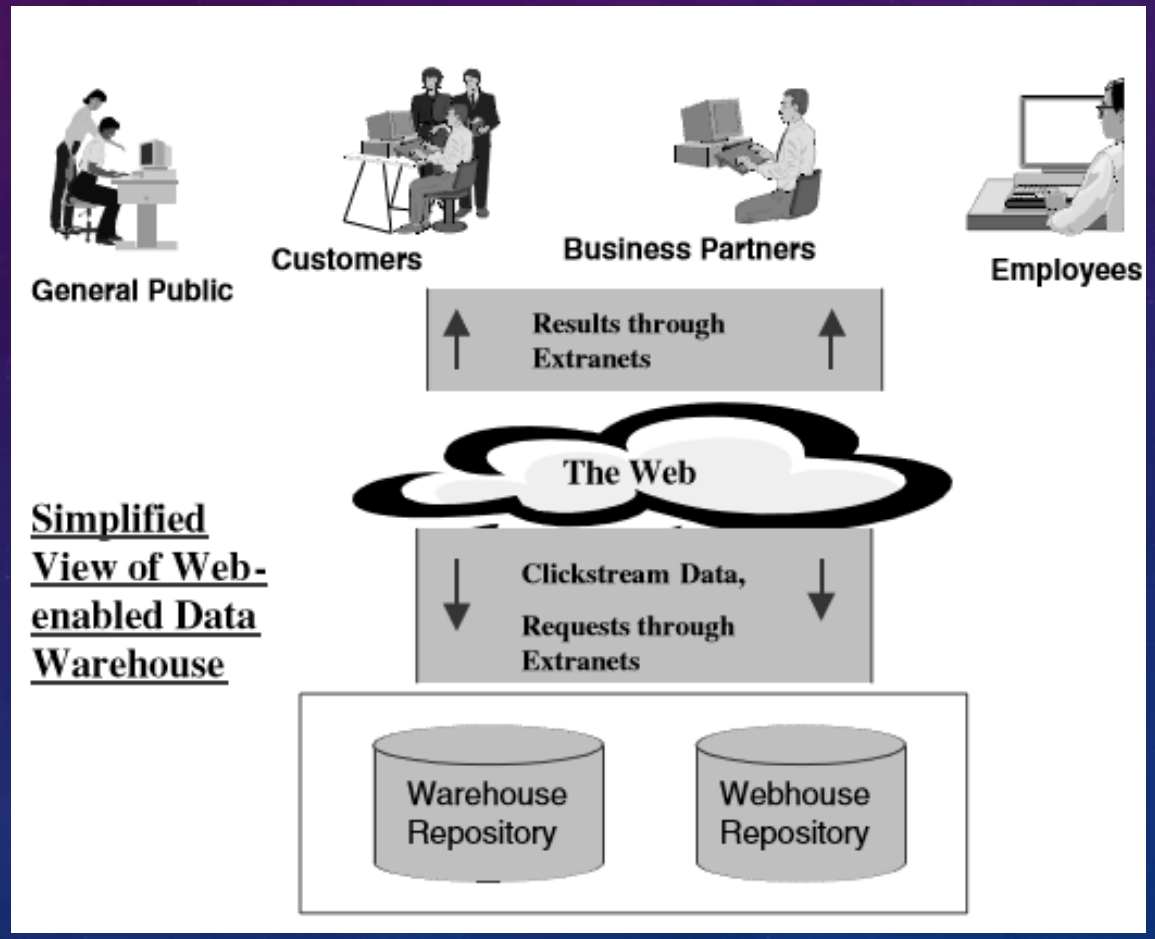
CRM-READY DATA WAREHOUSE

- Our data warehouse must hold details of every transaction at every touchpoint with each customer.
- This means every unit of every sale of every product to every customer must be gathered in the data warehouse repository.
- Making the data warehouse CRM-ready will increase the data volumes tremendously.
- For customer-related data, cleansing and transformation functions are more involved and complex.

WEB-ENABLED DATA WARE HOUSE

- Web-enabling the data warehouse means using the Web for information delivery and integrating the clickstream data from the corporate Web site for analysis.
- Notice the presence of the essential functional features of a traditional data warehouse.
- In addition to the data warehouse repository holding the usual types of information, the Web house repository contains clickstream data.

WEB-ENABLED DATA WARE HOUSE



END OF SLIDES