PRELUDE

- ✓ Data warehouse is a repository of an organization's electronically stored data and are designed to facilitate reporting and analysis
- ✓ Also emphasizes on the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary.
- ✓ An expanded definition for data warehousing includes business intelligence tools, tools to extract, transform, and load data into the repository, and tools to manage and retrieve metadata (data about data).
- ✓ In contrast to data, warehouses are operational systems that perform day-to-day transaction processing.
- ✓ A data warehouse is a collection of computer-based information that is critical to successful execution of enterprise initiatives
- ✓ It provides a tool to satisfy the information needs of the employee's at all organizational levels-not just for complex data queries but as a general facility for getting quick, accurate and often insightful information.
- ✓ It is designed so that its users can recognize the information they want and access that information using simple tools.
- ✓ One of the principal reasons for developing a Data Warehouse is to integrate operational data from various sources into a single and consistent architecture that supports analysis and decision making with the enterprise.
- ✓ Some of the applications data warehousing can be used for are:
    - o Credit card churn analysis
    - o Insurance fraud analysis
    - o Call record analysis
    - o Logistics management (part of Supply Chain Management that plans, implements, and controls the efficient, effective, forward, and reverse flow and storage of goods, services, and related information between the point of origin and the point of consumption in order to meet customers' requirements)

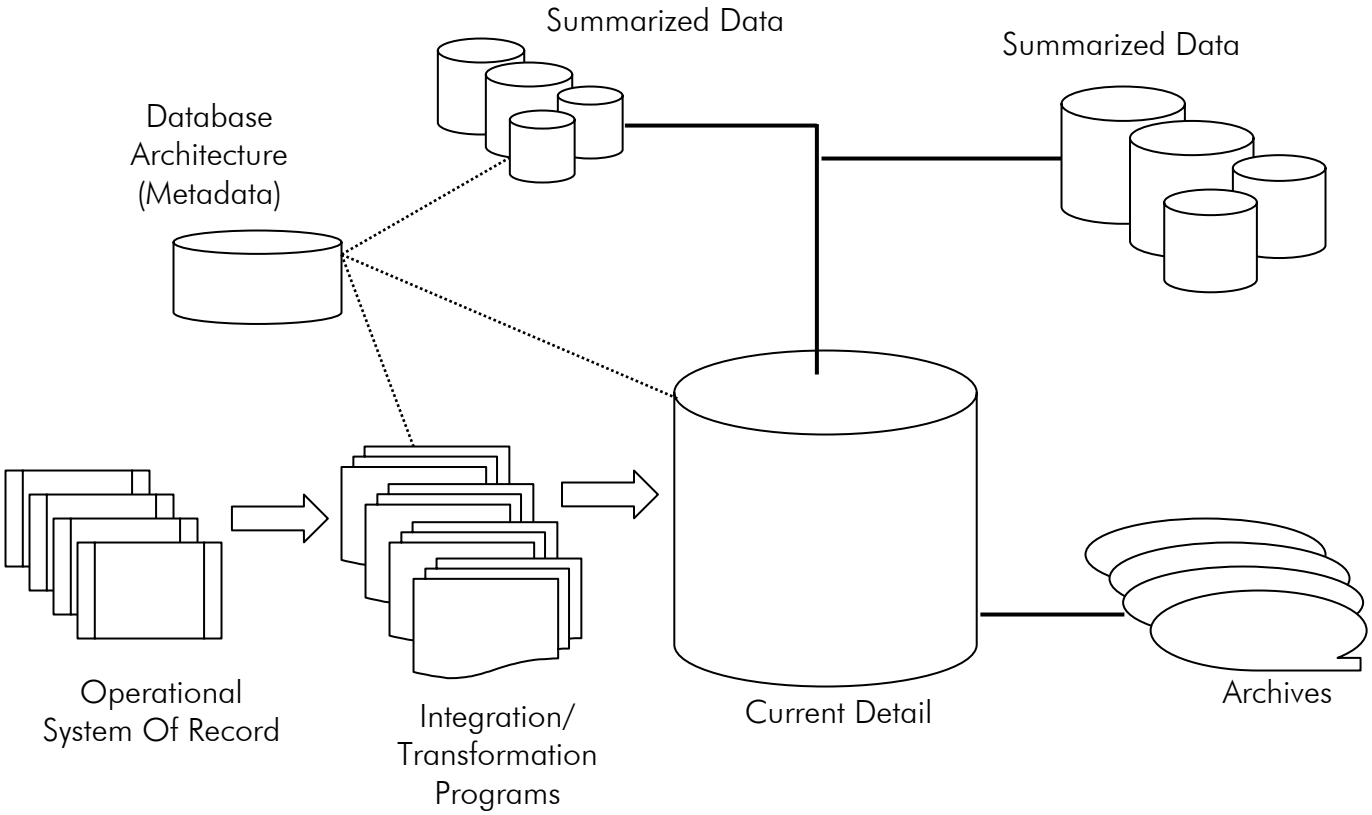
ADVANTAGES

- **More cost effective decision making:** A data warehouse allows reduction of staff and computer resources required to support queries and reports against operational and production database. This typically offers significant savings.
- **Better enterprise intelligence:** Increased quality and flexibility of enterprise analysis arises from the multi-level data structure which guarantees data accuracy and reliability ensuring that a Data Warehouse contains only "trusted" data.
- **Enhanced customer service:** An enterprise can maintain better customer relationships by correlating all customer data via a single Data Warehouse Architecture.
- **Business reengineering:** Allowing unlimited analysis of enterprise information often provides insights to enterprise processes that may yield breakthrough ideas for engineering those processes. Knowing what information is important to an enterprise will provide direction and priority for reengineering efforts.
- A data warehouse provides a common data model for all data of interest regardless of the data's source. This makes it easier to report and analyze information than it would be if

multiple data models were used to retrieve information such as sales invoices, order receipts, general ledger charges, etc.

- Prior to loading data into the data warehouse, inconsistencies are identified and resolved. This greatly simplifies reporting and analysis.
- Information in the data warehouse is under the control of data warehouse users so that, even if the source system data is purged (washed out) over time, the information in the warehouse can be stored safely for extended periods of time.
- Because they are separate from operational systems, data warehouses provide retrieval of data without slowing down operational systems.
- Data warehouses can work in conjunction with and, hence, enhance the value of operational business applications, notably customer relationship management (CRM) systems.
- Data warehouses facilitate decision support system applications such as trend reports (e.g., the items with the most sales in a particular area within the last two years), exception reports, and reports that show actual performance versus goals.

DATA WAREHOUSE COMPONENTS



Components Of A Data Warehouse

The major components of data warehouse are:

- Summarized Data
- Current Details
- Operational System of Record
- Integration/Transformation Programs
- Archives
- Data Warehouse Architecture or Metadata

Summarized Data:

- ✓ Classified into two categories: **lightly summarized** and **highly summarized**
- ✓ Lightly summarized data are the hallmark (trademark) of data warehouse as all enterprise elements do not have the same information requirement. They include less data than the total data stored in current detail.
- ✓ Highly summarized data are primarily for enterprise executives. They come from either the lightly summarized data used by enterprise elements or from current detail. Data volume at this level is much less than other levels.

Current Detail:

- ✓ It is the heart of the data warehouse where the whole bulk of data resides.
- ✓ Comes directly from operational systems of records and may be stored as raw data or as aggregations of raw data.
- ✓ It is the lowest level of data granularity which is typically two to five years old.

Operational System of Record:

- ✓ It is a source of the data that feeds the Data Warehouse.
- ✓ It is necessary for a Data Warehouse to be populated with the highest quality of data that is most timely, accurate, and complete and has the best structural conformance to the Data Warehouse.
- ✓ Often, these data are closest to the source of entry into the production environment. In other cases, a system of record may be containing already summarized data.

Integration/Transformation Programs

- ✓ As operational data items pass from their systems of record to a data warehouse, integration and transformation programs convert them from application-specific data into enterprise data. These integration and transformation programs functions such as:
  - o Reformatting, recalculating or modifying key structures
  - o Adding time elements
  - o Identifying default values
  - o Supplying logic to choose between multiple data sources
  - o Summarizing, tallying and merging data from multiple sources

Archives

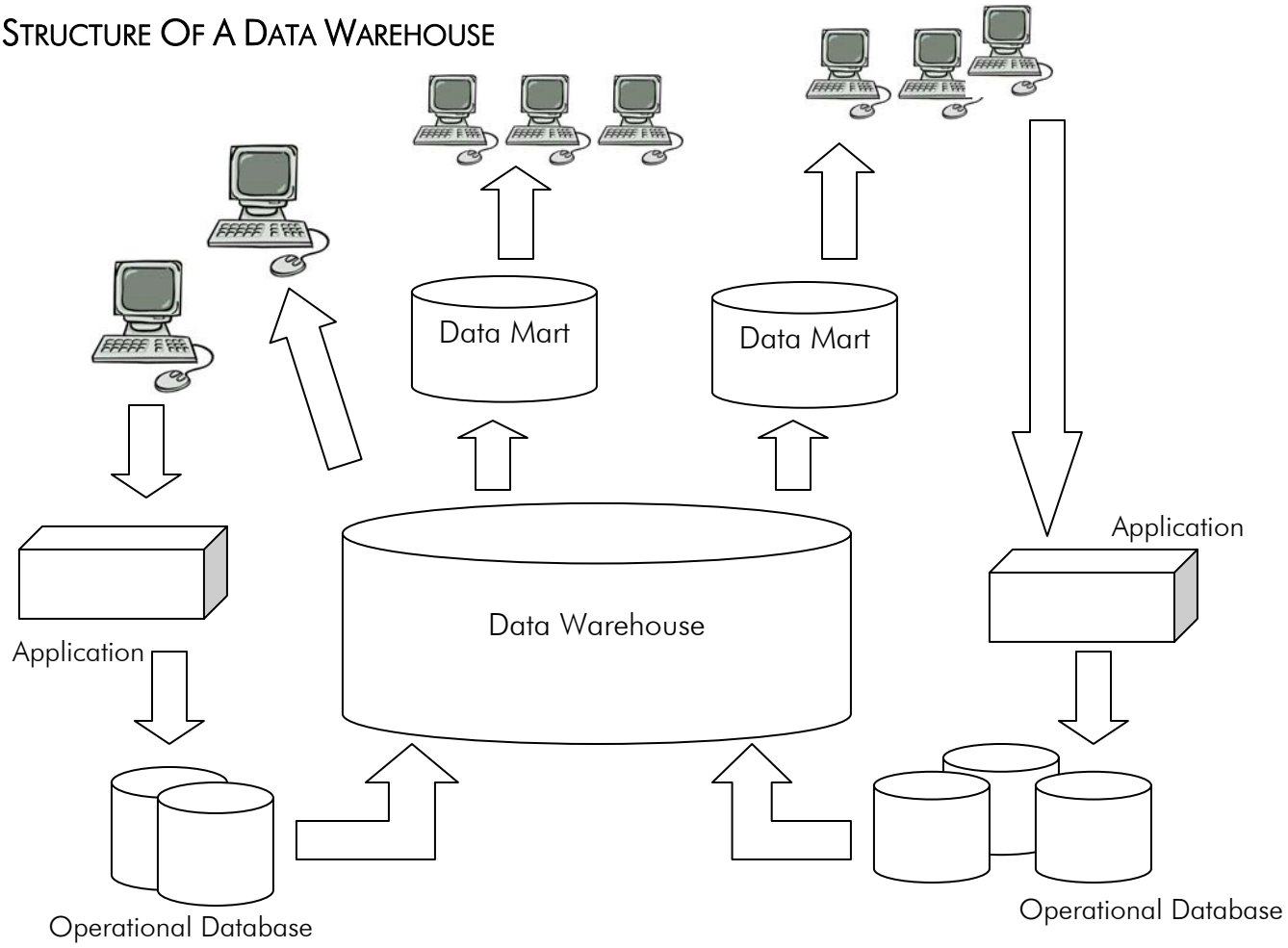
- ✓ Contains old data (normally over two years old) of significant, continuing interest and value to the enterprise.

- ✓ There is usually a massive amount of data stored in the Data Warehouse archives, with a low rate of access.
- ✓ Archives data are most often used for forecasting and trend analysis.
- ✓ Archives not only include old data, they also include the metadata that describes the old data's characteristic.

Metadata:

- ✓ Metadata is "data about other data", of any sort in any media.
- ✓ An item of metadata may describe an individual datum, or content item, or a collection of data including multiple content items and hierarchical levels.
- ✓ Metadata is definitional data that provides information about or documentation of other data managed within an application or environment.
- ✓ Metadata may include descriptive information about the context, quality and condition, or characteristics of the data. It may be recorded with high or low granularity.
- ✓ Data Warehouse developers use it to manage and control Data Warehouse creation and maintenance.
- ✓ For example, metadata would document data about data elements or attributes, (name, size, data type, etc) and data about records or data structures (length, fields, columns, etc) and data about data (where it is located, how it is associated, ownership, etc.).

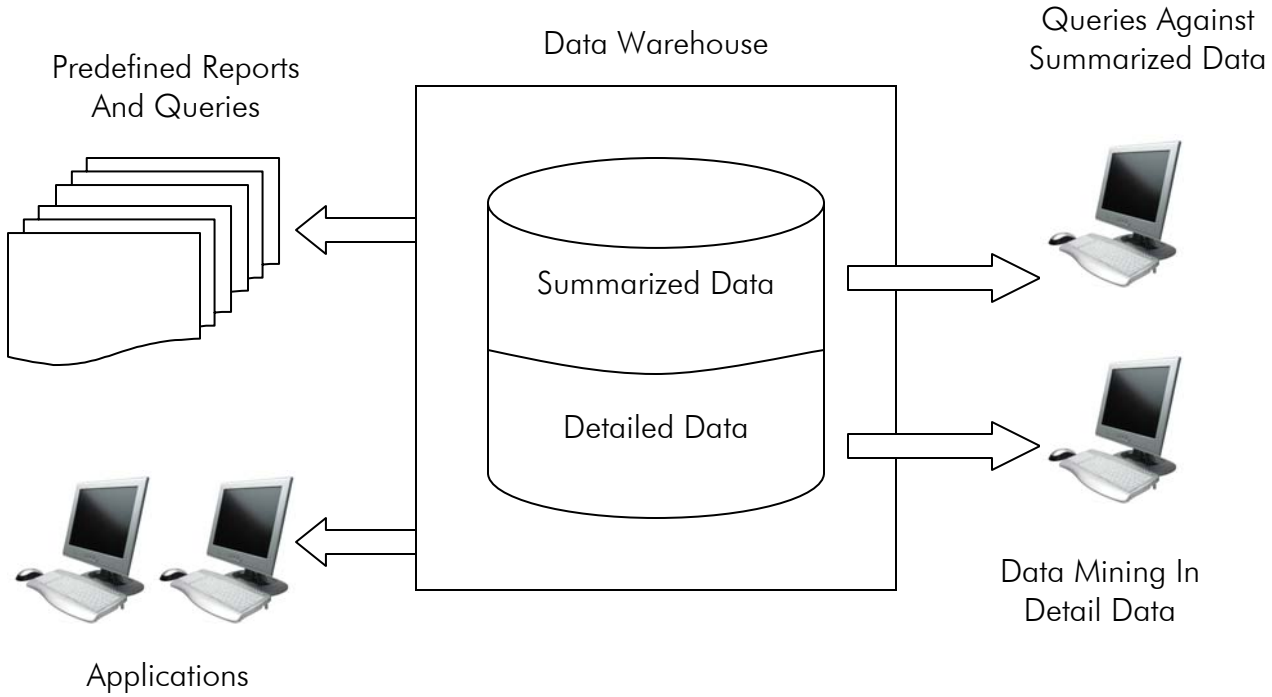
STRUCTURE OF A DATA WAREHOUSE



The structure of Data Warehouse consists of:

- Physical Data Warehouse:
  - ✓ Physical database in which all the data for the Data Warehouse is stored, along with metadata and processing logic for scrubbing, organizing, packaging and processing the detail data.
- Logical Data Warehouse:
  - ✓ It also contains metadata, including enterprise rules and processing logic for scrubbing, organizing, packaging and processing the data but does not contain actual data.
  - ✓ Instead it contains the information necessary to access the data wherever they reside.
  - ✓ This structure is effective only when there is a single source for the data and they are known to be accurate and timely.
- Data Marts:
  - ✓ Data Mart is a subset if an enterprise-wide Data Warehouse, which typically supports an enterprise element (department, region).
  - ✓ As part of an iterative Data Warehouse development process, an enterprise builds a series of physical (or logical) data marts over time and links them via an enterprise-wide logical data warehouse or feeds them from a single physical warehouse.
  - ✓ The data mart is directed at a partition of data (often called a subject area) that is created for the use of a dedicated group of users.
  - ✓ A data mart might, in fact, be a set of denormalized, summarized, or aggregated data.
  - ✓ In most instances, however, the data mart is a physically separate store of data and is resident on separate database server, often a local area network serving a dedicated user group.

USES OF A DATA WAREHOUSE



Standard Reports and Queries

Many users of the data warehouse need to access a set of standard reports and queries and hence it is desirable to periodically produce a set of standard reports that are required by many different users. When these users need a particular report, they can just view the report that has already run the data warehouse system rather than running it themselves. This facility can be particularly useful for reports that take a long time to run.

Queries Against Summarized Data

The summary views in the data warehouse can be object of a large majority of analysis in a data warehouse. These views contain predefined standard business analysis.

Data Mining

Data mining is the process of extracting hidden patterns from data. As more data are gathered, with the amount of data doubling every year, data mining is becoming an increasingly important tool to transform this data into information. It is commonly used in a wide range of profiling practices, such as marketing, fraud detection and scientific discovery. Data mining can be applied to data sets of any size.