Business Intelligence (BI) and Data Warehousing (DW)

Introduction

Jackline Ssanyu

January 24, 2020

Brief History

- Data are crucial raw material in this information age, and **data storage and management** have become the focus of database design and implementation.
- Ultimately, the **reason** for collecting, storing, and managing data is **to generate information that becomes the basis for balanced decision making**.
- **Decision support systems** (DSSs) were originally developed to **facilitate the decision-making process**.
- However, **as the complexity and range of information requirements increased**, so did the **difficulty of extracting all the necessary information from** the data structures typically found in **operational databases**.
- Therefore, a **new data storage** facility, called **a data warehouse, was developed**.

Brief History.....

- The data warehouse **extracts or obtains** its **data from operational databases as well as from external sources**, providing a more comprehensive data pool.
- In parallel with data warehouses, **new ways to analyze and present decision support data** were developed.
  - **Online analytical processing** (OLAP) provides advanced data analysis and presentation tools (including multidimensional data analysis).
  - **Data mining** employs advanced statistical tools to analyze the wealth of data now available through data warehouses and other sources and to identify possible relationships and anomalies.

Business Drivers for DW and BI

There are many business drivers in play today that are motivating companies to establish data warehouses. Current, consistent and accurate business information, they believe is critical for strategic and tactical decision making. Some of the business drivers are:

- Single Version of the Truth
- Current and Accurate Information
- Rapidly Changing Information Needs
- Customer Service Excellence
- New Service Delivery Channels

Single Version of the Truth

- Fragmented, inconsistent and outdated data in multiple databases does not permit good strategic and tactical decision making.
- Companies require that business intelligence be consolidated and presented in a suitable format for decision making. Inconsistent information from disparate information systems is no longer acceptable.
- Data Warehouses help companies to achieve a single version of the truth by consolidating the most accurate and current data from the most reliable systems.

Current and Accurate Information

- In a highly competitive market place, businesses need to quickly identify problems and opportunities in order to respond to events expeditiously and appropriately.
- Up-to-date information on sales, profits, inventories and customers can help identify problems early and leverage opportunities that could otherwise be missed.
- Most application systems are too narrowly scoped and operate on cycles that don't support real-time or near real-time information access. A data warehouse, however, can be designed to deliver up-to-date accurate information to decision makers.

Rapidly Changing Information Needs

- It is very difficult for businesses to anticipate future information needs. Application systems often seem rigid and unable to adapt to evolving management information needs.
- Businesses need the flexibility to slice and dice data in many ways in order to identify and analyze changes in the market place or in the business itself.
- Data Warehouses are designed for online, analytical purposes and provide great flexibility.

Customer Service Excellence

- It is often said that 10% of a business's customers account for 90% of the business's profits. Identifying the good customers and providing them with excellent service helps retain good customers.
- Data Warehousing can help identify a company's best customers using a any number or criterion.

New Service Delivery Channels

- It is no longer sufficient to provide customers with just 9:00 AM to 5:00 PM in-store service. Customers want to do business 7 days a week, 24 hours per day using alternate service delivery channels such as via the Internet or telephone.
- By examining all customer transactions, regardless of the channel used, businesses can better understand their customers and serve them better.
- Data Warehousing is critical for profiling customers and their transactions, regardless of the channel used.

Technical Drivers for DW and BI

There are many technical drivers in play that are motivating companies to establish data warehouses for online queries and analytics. These are:

- Multiple Internal Databases
- Purchased Packages
- Increasing Complexity of Systems
- Application System Evolution
- Computer Networks and External Databases

Multiple Internal Databases

- Most medium and large businesses operate dozens, if not hundreds of un-integrated application systems. Individual departments in companies often focus on their own narrow system and information needs and don't see the corporate value of integrating data.
- When a lot of un-integrated data exist, data soon gets out of hand. Companies have a need for database that reflects a "single version of truth". Data Warehouses can help do that.

Purchased Packages

- "Out of the Box" purchased applications sometimes use underlying concepts and definitions that differ from those used by the business in existing custom built applications.
- For example, a "customer" in one system could encompass all current and past customers plus potential future customers. In another system, a customer might be defined more narrowly as someone who has purchased a product and service during the past 12 months.
- Such inconsistencies create problems from an analytical perspective. A count of customers done in the first database differ from a count done in the second.
- Companies have a need to align concepts and terminology. Data Warehouses help do this alignment.

Increasing Complexity of Systems

- The underlying data structures of application systems are often very complicated. To create what would intuitively might appear to be a simple query often requires complex programming logic that involves navigating multiple database tables and or applications systems.
- Writing reports or queries can consequently take time and money.
- Companies have a need for a reporting environment that allows reports and queries to be generated quickly, inexpensively and without expensive IT skills. Data Warehouses can simplify the reporting environment.

Application System Evolution

- Businesses are highly dynamic and applications systems are constantly needing to be enhanced to support new business requirements.
- When systems are changed, reports and queries that access any changed tables must also be updated. This maintenance work can be very costly.
- Businesses have a need to trim their application support costs. Data Warehouses can help shelter reports and queries from system changes that occur in "front end" operational systems.

Computer Networks and External Databases

- The rapid growth of computer networks has allowed companies to exchange data with their suppliers, consumers, government bodies and other groups.
- Businesses often have a need to integrate data from internal and external databases. Data Warehouse can be designed to to integrate corporate data with external data for reporting purposes.

Definition

- Business intelligence (BI) is a **comprehensive, cohesive (organized), and integrated set of tools** and **processes used to capture, collect, integrate, store, and analyze data**.
- **Purpose:** generate and present information used to support business decision making.
- **BI** allows a business to **transform data into information**, **information into knowledge**, and **knowledge into wisdom**.
- **Wisdom empowers users** to **make sound business decisions** based on the accumulated knowledge of the business as reflected on recorded facts (historic operational data).

Business Intelligence Framework

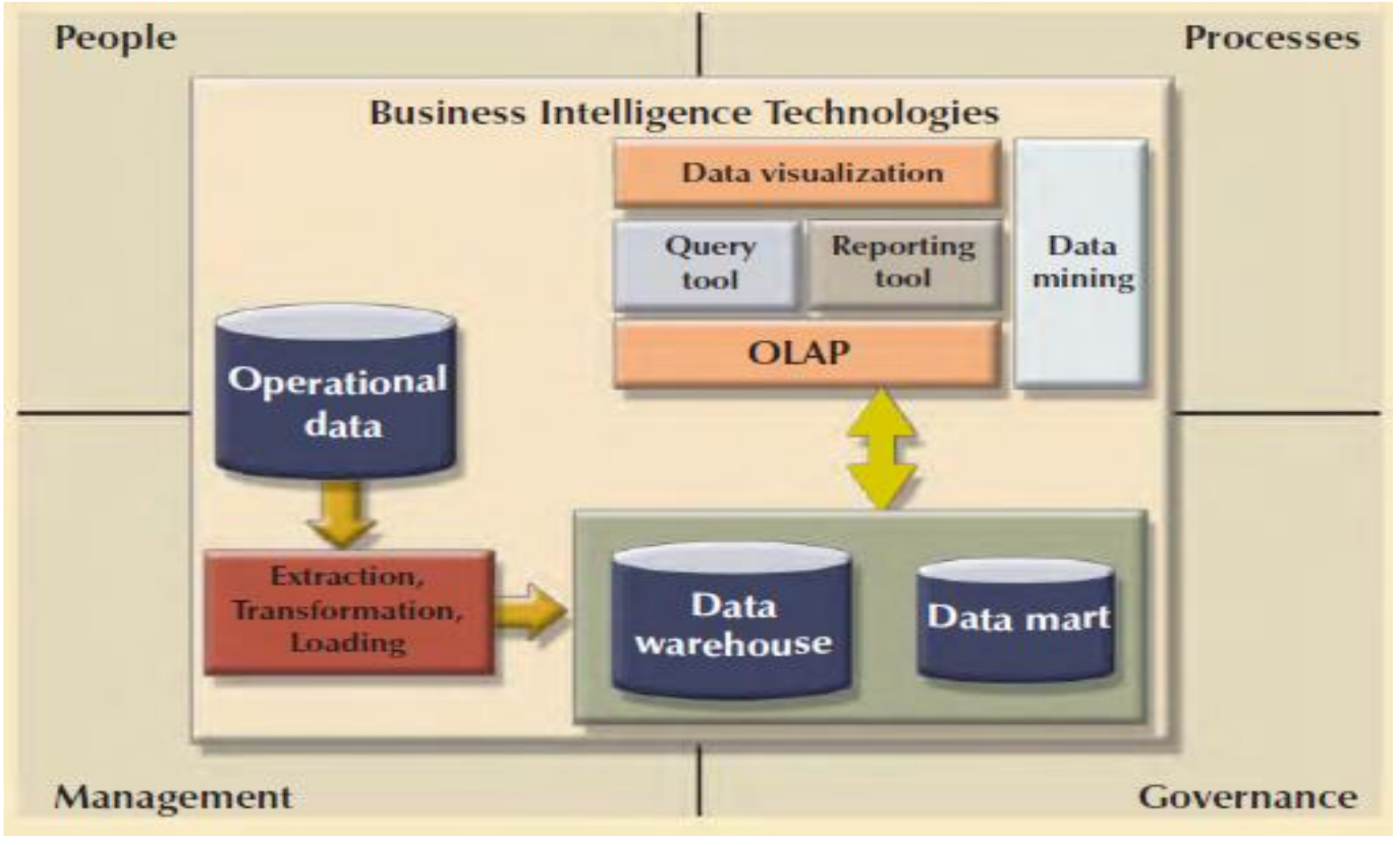


Figure: Business Intelligence Framework

Basic BI Architectural Components

- ETL tools -**Data extraction, transformation, and loading (ETL) tools** collect, filter, integrate, and aggregate operational data to be saved into a data store optimized for decision support.
  - **Example:** To determine the relative market share by selected product lines, you require data on competitors' products. Such data can be located in external databases provided by industry groups or by companies that market the data. As the name implies, this component (ETL) extracts the data, filters the extracted data to select the relevant records, and packages the data in the right format to be added to the data store component.

Basic BI Architectural Components.....

- Data Store - The data store is **optimized for decision support** and is generally represented by **a data warehouse or a data mart.** The data store contains two main types of data:
  - business data - are **extracted** from the **operational database** and from **external data sources** and is stored in structures that are **optimized** for **data analysis and query speed**. The **external data sources** provide data that cannot be found within the company but that are relevant to the business, such as stock prices, market indicators, marketing information (such as demographics), and competitors' data.
  - business model data - are **generated by special algorithms** that model the business to identify and enhance the understanding of business situations and problems.

Basic BI Architectural Components.....

- **Data query and analysis tools** - performs **data retrieval**, **data analysis**, and **data-mining tasks** using the data in the data store. This component is **used by the data analyst to create the queries that access the database**. Depending on the implementation, the query tool accesses either the operational database, or more commonly, the data store. This component is generally represented in the form of an OLAP tool.
- **Data presentation and visualization tools** - is **in charge of presenting the data to the end user in a variety of ways**. This component is used by the data analyst to organize and present the data. This tool helps the end user select the most appropriate presentation format, such as summary **report, map, pie or bar graph, or mixed graphs**. The query tool and the presentation tool are the front end to the BI environment.

Business Intelligence Framework.....

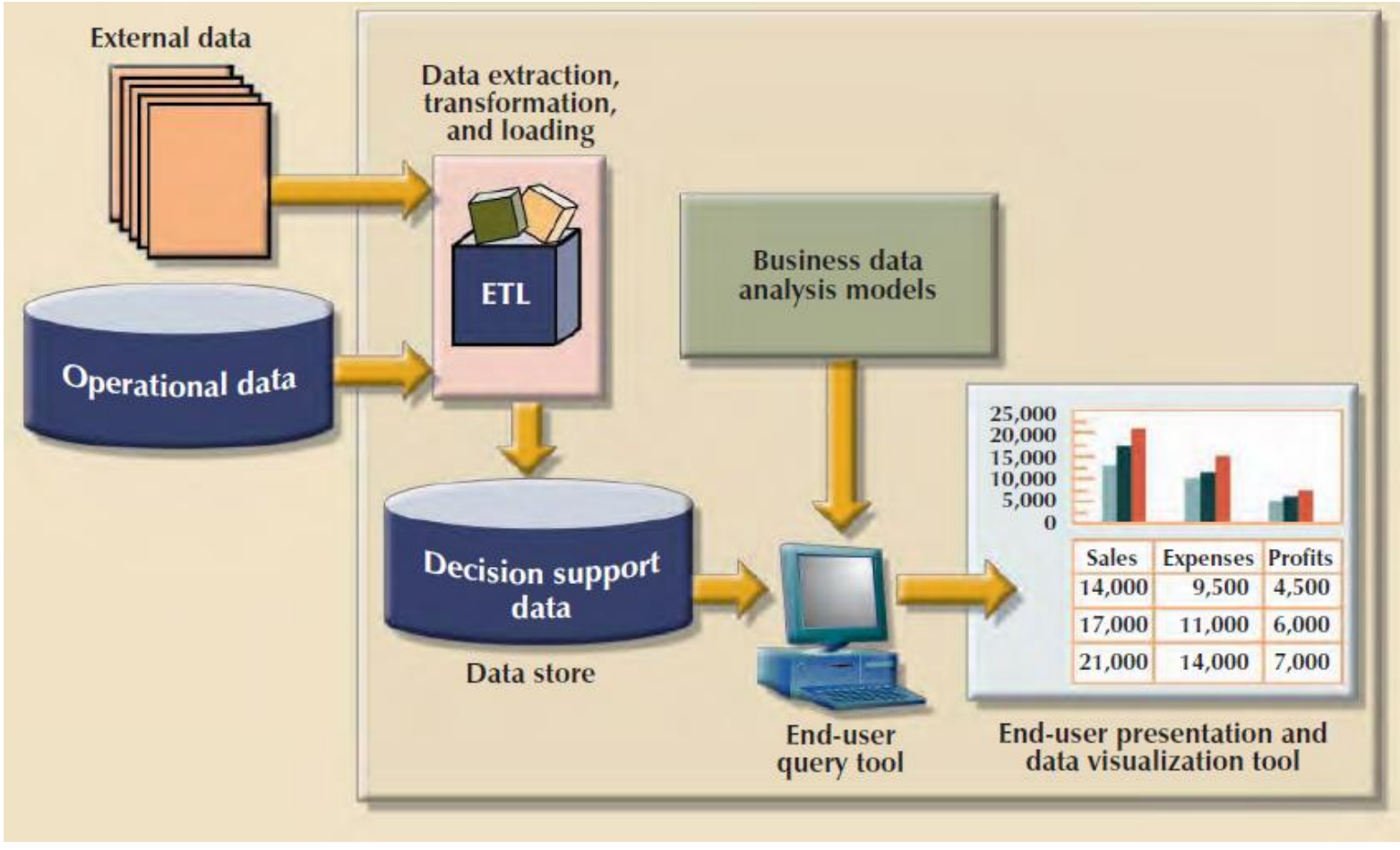


Figure: Business intelligence architectural component Illustration

Practices to Manage Data

BI uses an arrangement of the best management practices to manage data as a corporate asset.

- **Master data management (MDM)**: There must be an establishment of a collection of concepts, techniques, and processes for identification, definition, and management of data elements within an organization.
  - ensures that all company resources (people, procedures, and IT systems) that operate over data have uniform and consistent views of the company's data.
- **Governance**: There must be a method of government for monitoring and controlling business health and for consistent decision making

Practices to Manage Data....

- **Key performance indicators** (KPI): Numeric or scale-based measurements that assess company's effectiveness in reaching its goals. There are many different KPI used by different industries. Some examples of KPI are:
  - *General.* Year-to-year measurements of profit by line of business, same store sales, product turnovers, product recalls, sales by promotion, sales by employee, etc.
  - *Finance.* Earnings per share, profit margin, revenue per employee, percentage of sales to account receivables, assets to sales, etc.
  - *Human resources.* Applicants to job openings, employee turnover, employee longevity, etc.
  - *Education.* Graduation rates, number of incoming freshmen, student retention rates, etc.

Sample of Business Intelligence Tools

- **Spreadsheets** - These interactive computer applications manage information in a visual format, organized by rows and columns.
- **Decision Support Systems** - arrangement of computerized tools used to assist managerial decision making within a business.
- **Dashboards and business activity monitoring** - dashboards use Web-based technologies to present key business performance indicators or information in a single integrated view, generally using graphics in a clear, concise, and easy to understand manner.

Sample of Business Intelligence Tools...

- **Portals** - provide a unified, single point of entry for information distribution. Portals are a Web-based technology that uses a Web browser to integrate data from multiple sources into a single Web page. Many different types of BI functionality can be accessed through a portal.
- **Data analysis and reporting tools** - Advanced tools used to query multiple diverse data sources to create single integrated reports.
- **Data-mining tools** - Tools that provide advanced statistical analysis to uncover problems and opportunities hidden within business data.

Sample of Business Intelligence Tools...

- **Data warehouses** (DW) - the foundation on which a BI infrastructure is built. Data is captured from the OLTP system and placed in the DW on near-real-time basis. BI provides company-wide integration of data and the capability to respond to business issues in a timely manner.
- **OLAP tools** - Online analytical processing provides multidimensional data analysis.
- **Data visualization** - Tools that provide advanced visual analysis and techniques to enhance understanding of business data.

Operational Data Vs. Decision Support Data

- Operational data and decision support data serve different purposes.
  - **Operational data** storage is **optimized to support transactions that represent daily operations**.
  - **Decision Support Data** is optimized to **support decision making**.
- decision support data differ from operational data in three main areas: **time span, granularity, and dimensionality**.

Operational Data Vs. Decision Support Data....

- **Time span** - Operational data cover a short time frame. Decision support data tend to cover a longer time frame.
  - Example: Managers are seldom interested in a specific sales invoice to customer X; rather, they tend to focus on sales generated during the last month, the last year, or the last five years.

Operational Data Vs. Decision Support Data....

- **Granularity (level of aggregation)** - Decision support data must be presented at different levels of aggregation, from highly summarized to near-atomic.
  - Example, if managers must analyze sales by region, they must be able to access data showing the sales by region, by city within the region, by store within the city within the region, and so on.

Operational Data Vs. Decision Support Data....

- **Dimensionality.** - Operational data focus on representing individual transactions rather than on the effects of the transactions over time. In contrast, data analysts tend to include many data dimensions and are interested in how the data relate over those dimensions.
  - Example, an analyst might want to know how product X fared relative to product Z during the past six months by region, state, city, store, and customer. In that case, both place and time are part of the picture.

Operational Data Vs. Decision Support Data....

- The Figure in the next slide shows how decision support data can be examined from multiple dimensions (such as product, region, and year), using a variety of filters to produce each dimension.
- The ability to analyze, extract, and present information in meaningful ways is one of the differences between decision support data and transaction-at-a-time operational data.

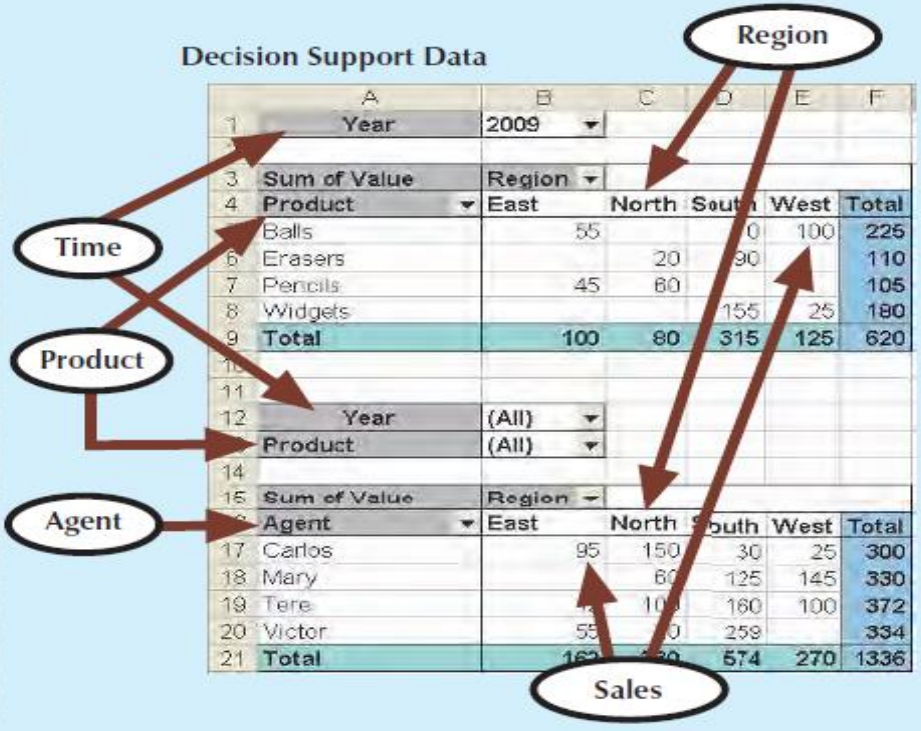
Operational Data Vs. Decision Support Data....

Operational Data

	A	B	C	D	E
3	Year	Region	Agent	Product	Value
4	2008	East	Carlos	Erasers	50
5	2008	East	Tere	Erasers	12
6	2008	North	Carlos	Widgets	120
7	2008	North	Tere	Widgets	100
8	2008	North	Carlos	Widgets	30
9	2008	South	Victor	Balls	145
10	2008	South	Victor	Balls	34
11	2008	South	Victor	Balls	80
12	2008	West	Mary	Pencils	89
13	2008	West	Mary	Pencils	56
14	2009	East	Carlos	Pencils	45
15	2009	East	Victor	Balls	55
16	2009	North	Mary	Pencils	60
17	2009	North	Victor	Erasers	20
18	2009	South	Carlos	Widgets	30
19	2009	South	Mary	Widgets	75
20	2009	South	Mary	Widgets	50
21	2009	South	Tere	Balls	70
22	2009	South	Tere	Erasers	90
23	2009	West	Carlos	Widgets	25
24	2009	West	Tere	Balls	100

Operational data have a narrow time span, low granularity, and single focus. Such data are usually presented in tabular format, in which each row represents a single transaction. This format often makes it difficult to derive useful information.

Decision Support Data



Decision support system (DSS) data focus on a broader timespan, tend to have high levels of granularity, and can be examined in multiple dimensions. For example, note these possible aggregations:

- Sales by product, region, agent, etc.
- Sales for all years or only a few selected years.
- Sales for all products or only a few selected products.

Figure: Transforming Operational Data into Decision Support Data

Operational Data Vs. Decision Support Data....

CHARACTERISTIC	OPERATIONAL DATA	DECISION SUPPORT DATA
Data currency	Current operations Real-time data	Historic data Snapshot of company data Time component (week/month/year)
Granularity	Atomic-detailed data	Summarized data
Summarization level	Low; some aggregate yields	High; many aggregation levels
Data model	Highly normalized Mostly relational DBMS	Non-normalized Complex structures Some relational, but mostly multidimensional DBMS
Transaction type	Mostly updates	Mostly query
Transaction volumes	High update volumes	Periodic loads and summary calculations
Transaction speed	Updates are critical	Retrievals are critical
Query activity	Low-to-medium	High
Query scope	Narrow range	Broad range
Query complexity	Simple-to-medium	Very complex
Data volumes	Hundreds of gigabytes	Terabytes to petabytes

Figure: Contrasting Operational and Decision Support Data Characteristics

Decision Support Database

- decision support database is a specialized DBMS tailored to provide fast answers to complex queries.
- There are four main requirements for a decision support database:
  - **database schema**, **data extraction and loading**, **end-user analytical interface**, and **database size**.

Decision Support Database Requirements

- **Database schema**
  - must support complex, non-normalized data representations
  - data must be aggregated and summarized
  - queries must be able to extract multidimensional time slices
- **Data extraction and loading**
  - database is created largely by extracting data from the operational database and by importing additional data from external sources.
  - data extraction capabilities should allow batch and scheduled data extraction.
  - data extraction capabilities should also support different data sources: flat files and hierarchical, network, and relational databases, as well as multiple vendors.
  - data-filtering capabilities must include the ability to check for inconsistent data or data validation rules and resolve them.
  - to filter and integrate the operational data into the decision support database, the DBMS must support advanced data integration, aggregation, and classification.

Decision Support Database Requirements.....

- **End-user analytical interface**
  - should permit the user to navigate through the data to simplify and accelerate the decision-making process.
  - the decision support DBMS must generate the necessary queries to retrieve the appropriate data from the decision support database.
  - because queries yield crucial information for decision makers, the queries must be optimized for speedy processing.
- **Database size should support**
  - very large databases (VLDBs)
  - advanced storage technologies
  - multiple-processor technologies

Business Intelligence Vs Artificial Intelligence

BI is different from Artificial Intelligence (AI)

- AI systems **make** decisions for the users.
- BI systems **help** the users make the right decisions, based on available data.

Business Intelligence and the Web

The Web makes BI even more useful

- Customers do not appear "physically" in a store; their behaviors cannot be observed by traditional methods
- A website log is used to capture the behavior of each customer, e.g., sequence of pages seen by a customer, the products viewed
- Idea: understand your customers using data and BI!
  - Utilize website logs, analyze customer behavior in more detail than before (e.g., what was not bought?)
  - Combine web data with traditional customer data

The Data Warehouse

- The **complex information requirements** and the **ever-growing demand for sophisticated data analysis** sparked the **creation of a new type of data repository**.
- This repository contains data in formats that facilitate data extraction, data analysis, and decision making.
- This data repository is known as a data warehouse and has become the foundation for a new generation of decision support systems.
- **Data warehouse** is an **integrated, subject-oriented, time-variant, nonvolatile** collection of data that provides **support for decision making**.

Definition of components

- **Integrated**
  - is a centralized, consolidated database that integrates data derived from the entire organization and from multiple sources with diverse formats.
  - data integration implies that all business entities, data elements, data characteristics, and business metrics are described in the same way throughout the enterprize.
- **Subject-oriented**
  - data warehouse data are organized and summarized by topic, such as sales, marketing, finance, distribution, and transportation.
  - for each topic, the data warehouse contains specific subjects of interest-products, customers, departments, regions, promotions, and so on.

Definition of components....

- **Time-variant**
  - warehouse data represent the flow of data through time.
  - the data warehouse contains a time ID that is used to generate summaries and aggregations by week, month, quarter, year, and so on.
- **Nonvolatile**
  - once data enter the data warehouse, they are never removed.
  - because data are never deleted and new data are continually added, the data warehouse is always growing.
  - that's why the DBMS must be able to support multigigabyte and even multiterabyte or greater databases, operating on multiprocessor hardware.

In Summary

- A data warehouse is a semantically consistent data store that serves as a physical implementation of a decision support data model and stores the information on which an enterprize needs to make strategic decisions.
- A data warehouse is also often viewed as an architecture, constructed by integrating data from multiple heterogeneous sources to support structured and/or ad hoc queries, analytical reporting, and decision making.

Data warehouses Vs Operational DBs

- The major task of online operational database systems is to perform online transaction and query processing. These systems are called **online transaction processing (OLTP) systems**. They cover most of the day-to-day operations of an organization, such as purchasing, inventory, manufacturing, banking, payroll, registration, and accounting.
- Data warehouse systems, on the other hand, serve users or knowledge workers in the role of data analysis and decision making. Such systems can organize and present data in various formats in order to accommodate the diverse needs of the different users. These systems are known as **online analytical processing (OLAP) systems**.

Data warehouses Vs Operational DBs...

The major distinguishing features between OLTP and OLAP are summarized as follows:

- **Users and system orientation:** An OLTP system is *customer-oriented* and is used for transaction and query processing by clerks, clients, and information technology professionals. An OLAP system is *market-oriented* and is used for data analysis by knowledge workers, including managers, executives, and analysts.
- **Data contents:** An OLTP system manages current data that, typically, are too detailed to be easily used for decision making. An OLAP system manages large amounts of historical data, provides facilities for summarization and aggregation, and stores and manages information at different levels of granularity. These features make the data easier to use in informed decision making.

Data warehouses Vs Operational DBs...

Distinguishing features between OLTP and OLAP (Cont.)

- **Database design:** An OLTP system usually adopts an **entity-relationship (ER) data model** and an application-oriented database design. An OLAP system typically adopts either a **star or snowflake model** and a subject-oriented database design.
- **View:** An OLTP system **focuses mainly on the current data within an enterprise or department**, without referring to historical data or data in different organizations. In contrast, an OLAP system often **spans multiple versions of a database schema**, due to the evolutionary process of an organization. OLAP systems also deal with information that originates from different organizations, integrating information from many data stores. Because of their huge volume, OLAP data are stored on multiple storage media.

Data warehouses Vs Operational DBs...

Distinguishing features between OLTP and OLAP (Cont.)

- **Access patterns:** The access patterns of an OLTP system consist mainly of short, atomic transactions. Such a system requires concurrency control and recovery mechanisms. However, accesses to OLAP systems are mostly read-only operations (because most data warehouses store historical rather than up-to-date information), although many could be complex queries.

Comparison between OLTP and OLAP systems

Table 4.1: Comparison between OLTP and OLAP systems.

Feature	*OLTP*	*OLAP*
Characteristic	operational processing	informational processing
Orientation	transaction	analysis
User	clerk, DBA, database professional	knowledge worker (e.g., manager, executive, analyst)
Function	day-to-day operations	long-term informational requirements, decision support
DB design	ER based, application-oriented	star/snowflake, subject-oriented
Data	current; guaranteed up-to-date	historical; accuracy maintained over time
Summarization	primitive, highly detailed	summarized, consolidated
View	detailed, flat relational	summarized, multidimensional
Unit of work	short, simple transaction	complex query
Access	read/write	mostly read
Focus	data in	information out
Operations	index/hash on primary key	lots of scans
Number of records accessed	tens	millions
Number of users	thousands	hundreds
DB size	GB to high-order GB	$\geq$ TB
Priority	high performance, high availability	high flexibility, end-user autonomy
Metric	transaction throughput	query throughput, response time

Figure: OLTP Vs OLAP

Why Have a Separate Data Warehouse?

Because operational databases store huge amounts of data, *"why not perform online analytical processing directly on such databases?"*. A major reason for such a separation is to help **promote the high performance of both systems.**

- An operational database is designed and tuned from known tasks and workloads, such as indexing and hashing using primary keys, searching for particular records, and optimizing queries.
- On the other hand, data warehouse queries are often complex. They involve the computation of large groups of data at summarized levels, and may require the use of special data organization, access, and implementation methods based on multidimensional views.

Processing OLAP queries in operational databases would substantially degrade the performance of operational tasks.

Why Have a Separate Data Warehouse?...

- In addition, the separation of operational databases from data warehouses is based on the different structures, contents, and uses of the data in these two systems. Decision support requires historical data, whereas operational databases do not typically maintain historical data.
- The data in operational databases, though abundant, is usually far from complete for decision making. Decision support requires consolidation (such as aggregation and summarization) of data from heterogeneous sources, resulting in high-quality, clean, and integrated data.
- In contrast, operational databases contain only detailed raw data, such as transactions, which need to be consolidated before analysis.

Data Warehousing: A Multi-Tiered Architecture

Data warehouses often adopt a three-tier architecture:

- The bottom tier is a **warehouse database server** that is almost always a relational database system. Back-end tools and utilities (commonly known as ETL) are used to feed data into the bottom tier from operational databases or other external sources. This tier also contains a metadata repository, which stores information about the data warehouse and its contents.
- The middle tier is an **OLAP server** that implements multidimensional data and operations.
- The top tier is a **front-end client layer**, which contains query and reporting tools, analysis tools, and/or data mining tools (e.g., trend analysis, prediction, and so on).

A three-tier data warehousing architecture.

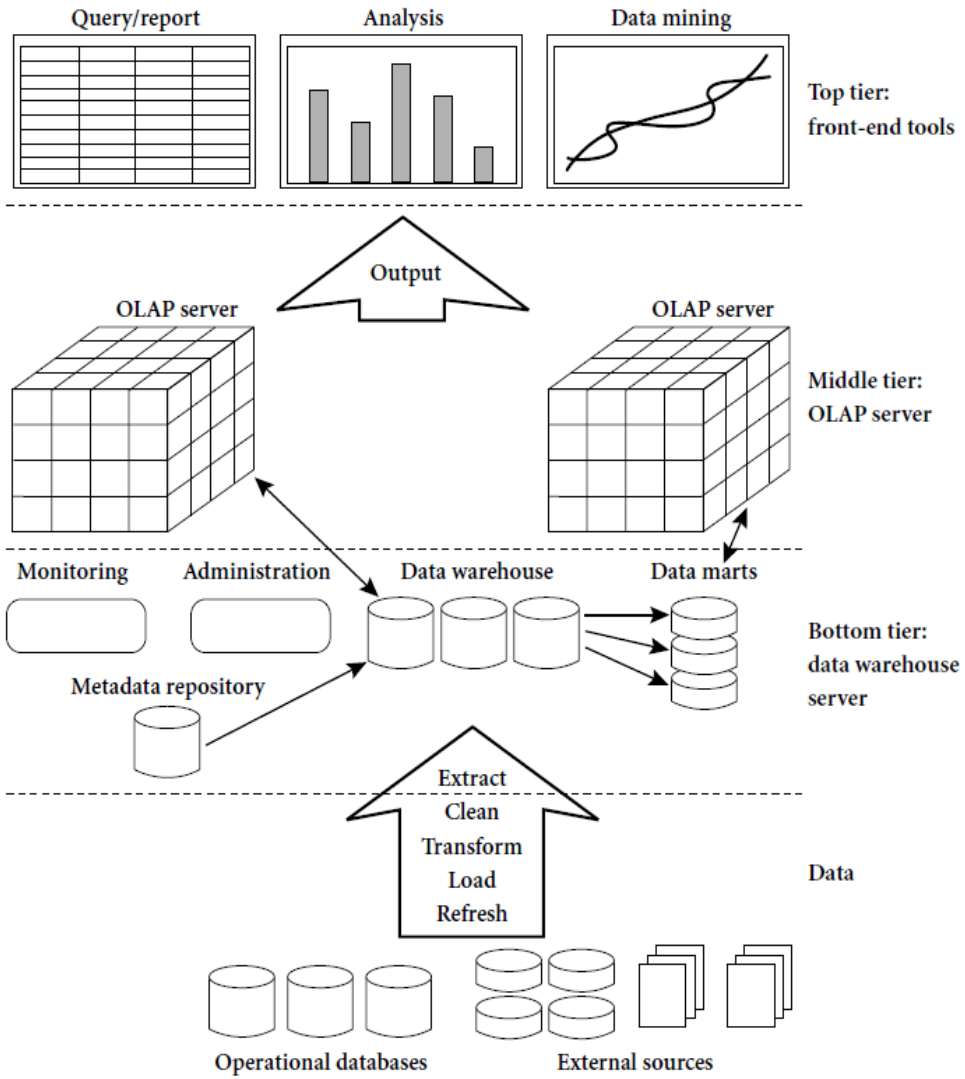


Figure: A three-tier data warehousing architecture.

Data Warehouse Models

From the architecture point of view, there are three data warehouse models:

- the enterprise warehouse
- the data mart
- the virtual warehouse

Enterprise warehouse

- An enterprise warehouse collects all the information and the subjects spanning an entire organization
- It provides us enterprise-wide data integration.
- The data is integrated from operational systems and external information providers.
- This information can vary from a few gigabytes to hundreds of gigabytes, terabytes or beyond.
- It requires extensive business modeling and may take years to design and build.

Data mart

- Data mart contains a subset of organization-wide data. This subset of data is valuable to specific groups of an organization.
- In other words, we can claim that data marts contain data specific to a particular group. For example, the marketing data mart may contain data related to items, customers, and sales. Data marts are confined to subjects.
- Data marts are usually implemented on low-cost departmental servers that are Unix/Linux- or Windows-based. The implementation cycle of a data mart is more likely to be measured in weeks rather than months or years. However, it may involve complex integration in the long run if its design and planning were not enterprise-wide.

Virtual warehouse

- A virtual data warehouse is a set of separate databases, which can be queried together, so a user can effectively access all the data as if it was stored in one data warehouse.
- A virtual warehouse is a set of views over operational databases.
- A virtual warehouse is easy to build but requires excess capacity on operational database servers.

Metadata Repository

- Metadata are data about data. When used in a data warehouse, metadata are the data that define warehouse objects.
- Metadata are created for the data names and definitions of the given warehouse. Additional metadata are created and captured for time stamping any extracted data, the source of the extracted data, and missing fields that have been added by data cleaning or integration processes.

Metadata Repository...

A metadata repository should contain the following:

- **Definition of data warehouse** - It includes the description of structure of data warehouse. The description is defined by schema, view, hierarchies, derived data definitions, and data mart locations and contents.
- **Business metadata** - It contains the data ownership information, business definition, and changing policies.
- **Operational Metadata** - It includes currency of data and data lineage. Currency of data means whether the data is active, archived, or purged. Lineage of data means the history of data migrated and transformation applied on it.

Metadata Repository...

- **Data for mapping from operational environment to data warehouse** - It includes the source databases and their contents, data extraction, data partition cleaning, transformation rules, data refresh and purging rules.
- **Algorithms for summarization** - It includes dimension algorithms, data on granularity, aggregation, summarizing, etc.

Role of Metadata

Metadata are used as a directory to help the decision support system analyst locate the contents of the data warehouse,:

- as a guide to the mapping of data when the data are transformed from the operational environment to the data warehouse environment
- as a guide to the algorithms used for summarization between the current detailed data and the lightly summarized data, and between the lightly summarized data and the highly summarized data.

Metadata should be stored and managed persistently (i.e., on disk).

The ETL process

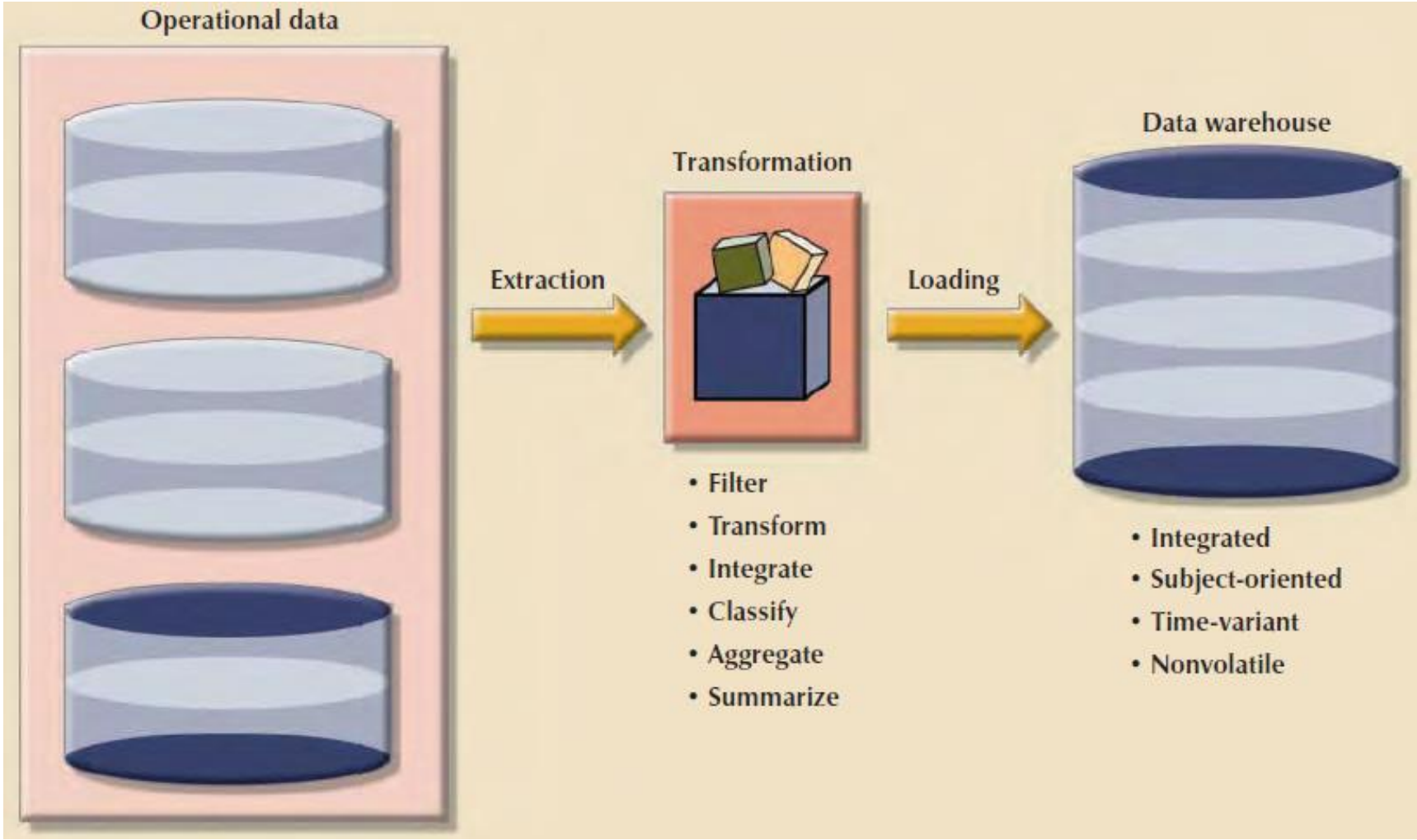


Figure: The ETL process

The ETL process...

- process of extracting, transforming, and loading the aggregated data into the data warehouse is known as ETL.
- ETL technology can migrate data from different types of data structures(e.g. databases and at files which may include the following:
  - Customer Relationship Management (CRM)
  - Enterprise Resource Planning (ERP)
  - Legacy Systems
  - E-Commerce
  - Supply Chain Management (SCM) and across different platforms (e.g. mainframe, server).
- the extracted data is converted from its previous form into a general form it needs to be in so that it can be loaded into the data warehouse.

Review Questions

- Explain web portals and dashboards and their role in Business Intelligence.
-