

IT 775

Database Technology

Data Stores

Data Warehouses

INTRODUCTION

A typical organization maintains and utilizes a number of operational data sources.

The operational data sources include the databases and other data repositories which are used to support the organization's day-to-day operations

A data warehouse is created within an organization as a separate data store whose primary purpose is data analysis.

INTRODUCTION

Two main reasons for the creation of a data warehouse as a separate analytical database

The performance of operational day-to-day tasks involving data use can be severely diminished if such tasks have to compete for computing resources with analytical queries

It is often impossible to structure a database which can be used in an efficient manner for both operational and analytical purposes

INTRODUCTION

Operational information (transactional information) - the information collected and used in support of day to day operational needs in businesses and other organizations

Analytical information - the information collected and used in support of analytical tasks

Analytical information is based on operational (transactional) information

OPERATIONAL VS. ANALYTICAL INFORMATION

Operational Data

Analytical Data

Data Makeup Differences

Typical Time-Horizon: Days/Months
Detailed
Current

Typical Time-Horizon: Years
Summarized (and/or Detailed)
Values over time (Snapsho

Technical Differences

Small Amounts used in a Process
High frequency of Access
Can be Updated
Non-Redundant

Large Amounts used in a Process
Low/Modest frequency of Access
Read (and Append) Only
Redundancy not an Issue

Functional Differences

Used by all types of employees
for tactical purposes
Application Oriented

Used by a narrower set of
users for decision making
Subject Oriented

Application Oriented vs. Subject Oriented– Example

An **application-oriented** database serving the Vitality Health Club Visits and Payments Application

HEALTH CLUB MEMBER

<u>MemberID</u>	MemberName	MemberGender	MLevelID	DateMembershipPaid
111	Joe	M	A	1/1/2013
222	Sue	F	B	1/1/2013
333	Pam	F	A	1/2/2013
...

MEMBERSHIP LEVEL

<u>MLevelID</u>	MLevelType	MLevelFee	MLevelDescription
A	Gold	\$100	Includes the Pool Usage
B	Basic	\$50	No Pool Usage

DAILY VISIT FROM NONMEMBERS

<u>DVisitTID</u>	DVisitLevelID	DVisitDate	DVisitorGender
11xx22	YP	1/1/2013	M
11xx23	NP	1/2/2013	M
11xx24	YP	1/2/2013	F
...

VISIT LEVEL

<u>DVisitLevelID</u>	DVisitLevelFee	DVisitLevelType
YP	\$15	With Pool Usage
NP	\$10	Without Pool Usage

Application Oriented vs. Subject Oriented– Example

A **subject-oriented** database for the analysis of the subject *revenue* in the Vitality Health Club

REVENUE					
<u>RevenueRecordID</u>	Date	GeneratedBy	ClientGender	Pool Use Included in Purchase	Amount
1000	1/1/2013	Member	M	Yes	\$100
1001	1/1/2013	Member	F	No	\$50
1002	1/1/2013	Nonmember	M	Yes	\$15
1003	1/2/2013	Member	F	Yes	\$100
1004	1/2/2013	Nonmember	M	No	\$10
1005	1/2/2013	Nonmember	F	Yes	\$15
...			

THE DATA WAREHOUSE DEFINITION

The data warehouse is a **structured repository** of **integrated, subject-oriented, enterprise-wide, historical, and time-variant** data. The purpose of the data warehouse is the **retrieval of analytical information**. A data warehouse can store **detailed and/or summarized data**.

THE DATA WAREHOUSE DEFINITION

Structured repository

The data warehouse is a *database* containing analytically useful information

Any database is a *structured repository* with its structure represented in its metadata

THE DATA WAREHOUSE DEFINITION

Integrated

The data warehouse integrates the analytically useful data from the various operational databases (and possibly other sources)

Integration refers to this process of bringing the data from multiple data sources into a singular data warehouse.

THE DATA WAREHOUSE DEFINITION

Subject-oriented

The term *subject-oriented* refers to the fundamental difference in the purpose of an operational database system and a data warehouse.

- An operational database system is developed in order to support a specific business operation
- A data warehouse is developed to analyze specific business subject areas

THE DATA WAREHOUSE DEFINITION

Enterprise-wide

The term *enterprise-wide* refers to the fact that the data warehouse provides an organization-wide view of the analytically useful information it contains

THE DATA WAREHOUSE DEFINITION

Historical

The term *historical* refers to the larger time horizon in the data warehouse than in the operational databases

THE DATA WAREHOUSE DEFINITION

Time variant

The term *time variant* refers to the fact that a data warehouse contains slices or snapshots of data from different periods of time across its time horizon

- With the data slices, the user can create reports for various periods of time within the time horizon

THE DATA WAREHOUSE DEFINITION

Retrieval of analytical information

A data warehouse is developed for the *retrieval of analytical information*, and it is not meant for direct data entry by the users.

- The only functionality available to the users of the data warehouse is retrieval
- The data in the data warehouse is not subject to changes.
- The data in the data warehouse is referred to as *non-volatile, static, or read-only*

THE DATA WAREHOUSE DEFINITION

Detailed and/or summarized data

A data warehouse, depending on its purpose, may include the *detailed data* or *summary data* or *both*

A data warehouse that contains the data at the finest level of detail is the most powerful

DATA WAREHOUSE COMPONENTS

Data warehouse components

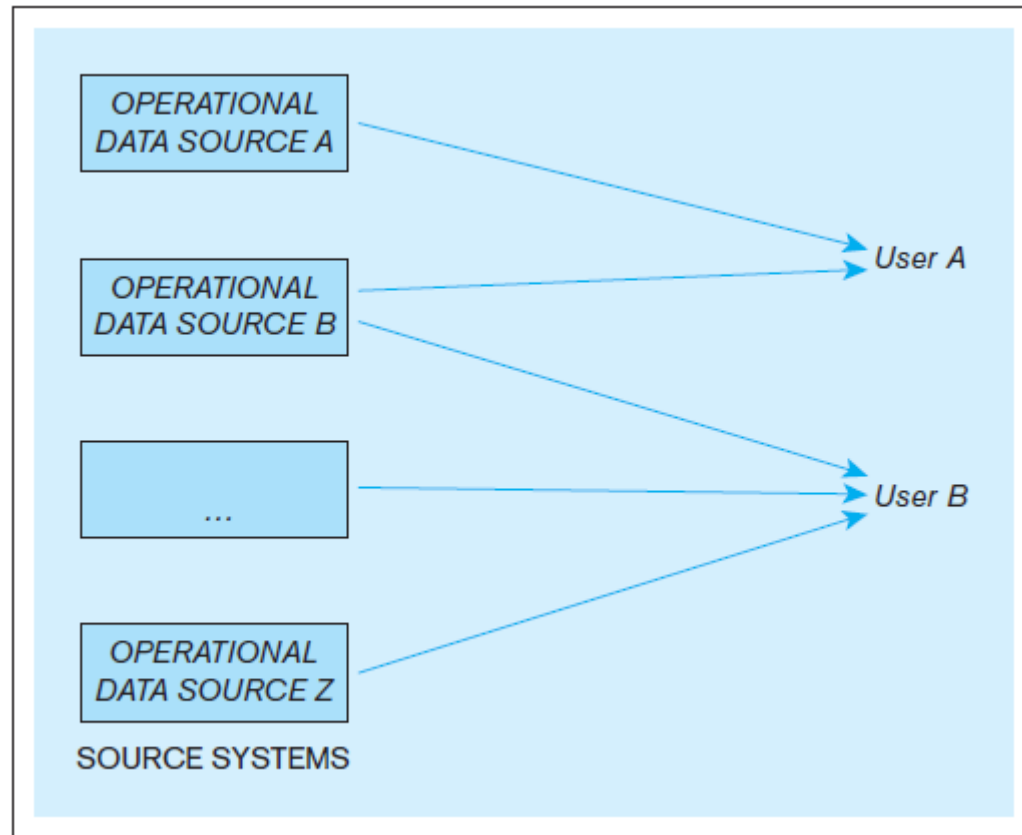
Source systems

Extraction-transformation-load (ETL) infrastructure

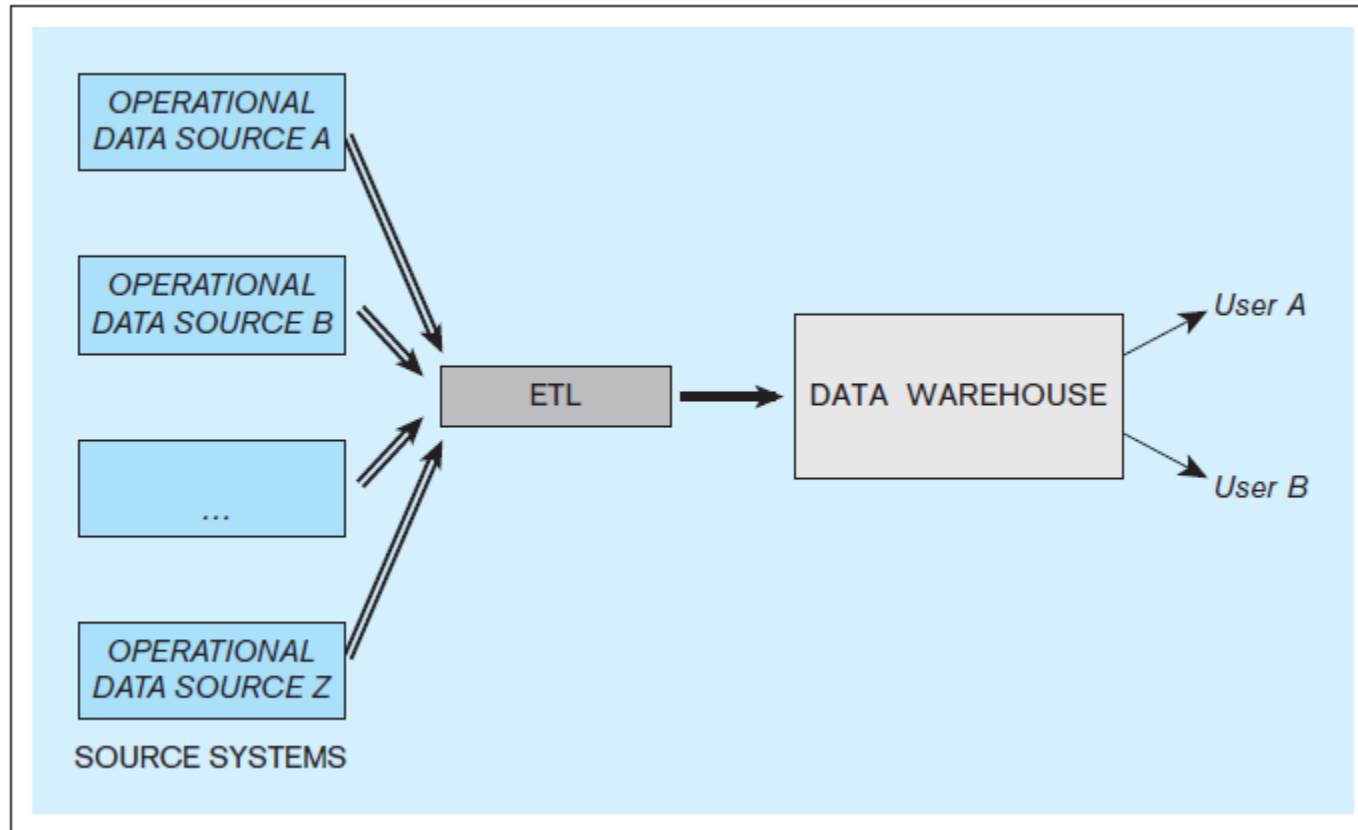
Data warehouse

Front-end applications

Example - The use of operational data sources for operational purposes in an organization



Example - The core components of a data warehousing system



DATA WAREHOUSE COMPONENTS

Source systems

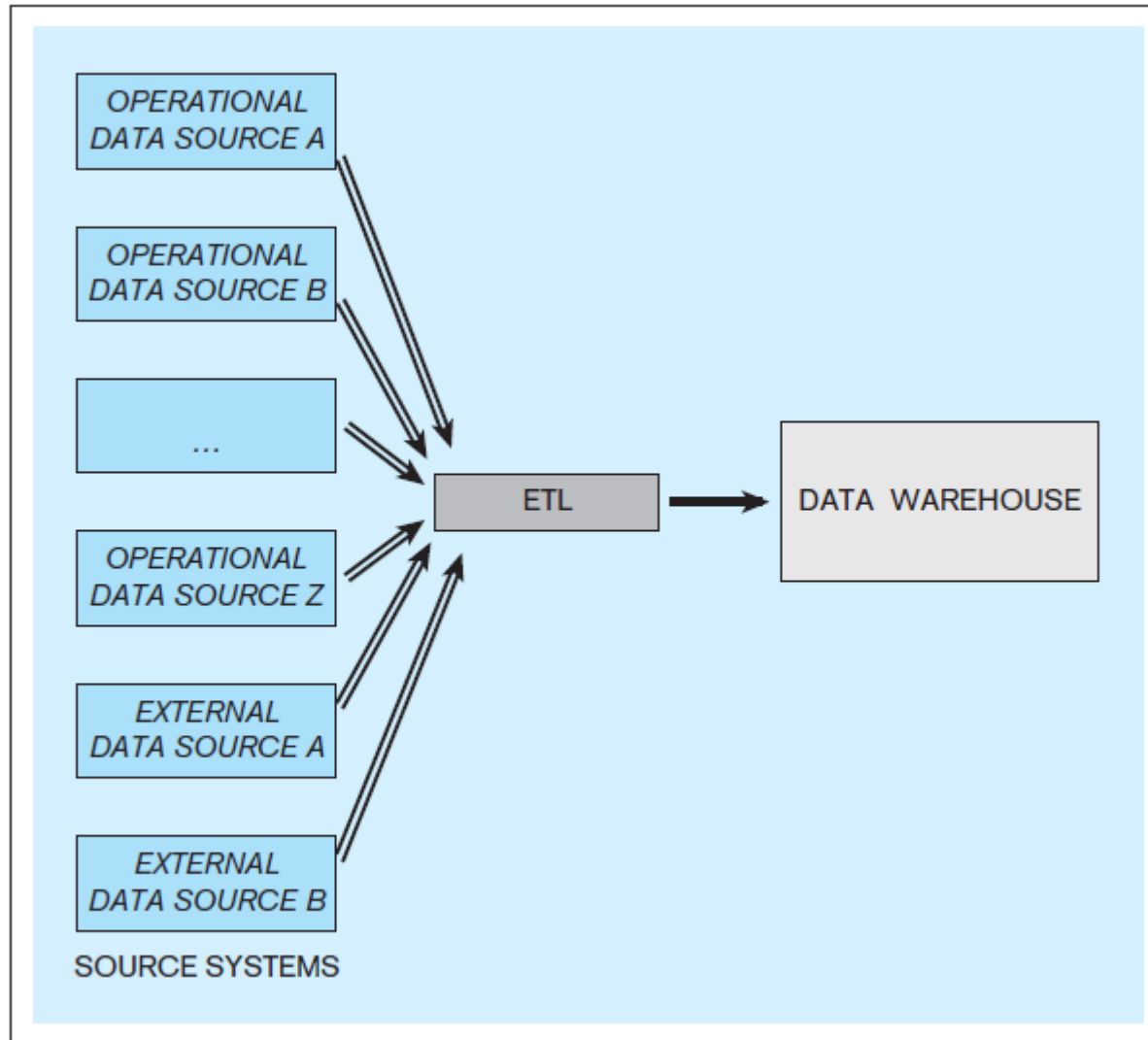
In the context of data warehousing, *source systems* are operational databases and other operational data repositories (in other words, any sets of data used for operational purposes) that provide analytically useful information for the data warehouse's subjects of analysis

Every operational data store that is used as a source system for the data warehouse has two purposes:

- The original operational purpose
- As a source system for the data warehouse

Source systems can include *external data sources*

Example - A data warehouse with internal and external source systems



DATA WAREHOUSE COMPONENTS

Data warehouse

The *data warehouse* is sometimes referred to as the *target system*, to indicate the fact that it is a destination for the data from the source systems

A typical data warehouse periodically retrieves selected analytically useful data from the operational data sources

DATA WAREHOUSE COMPONENTS

ETL infrastructure

The infrastructure that facilitates the retrieval of data from operational databases into the data warehouses

ETL includes the following tasks:

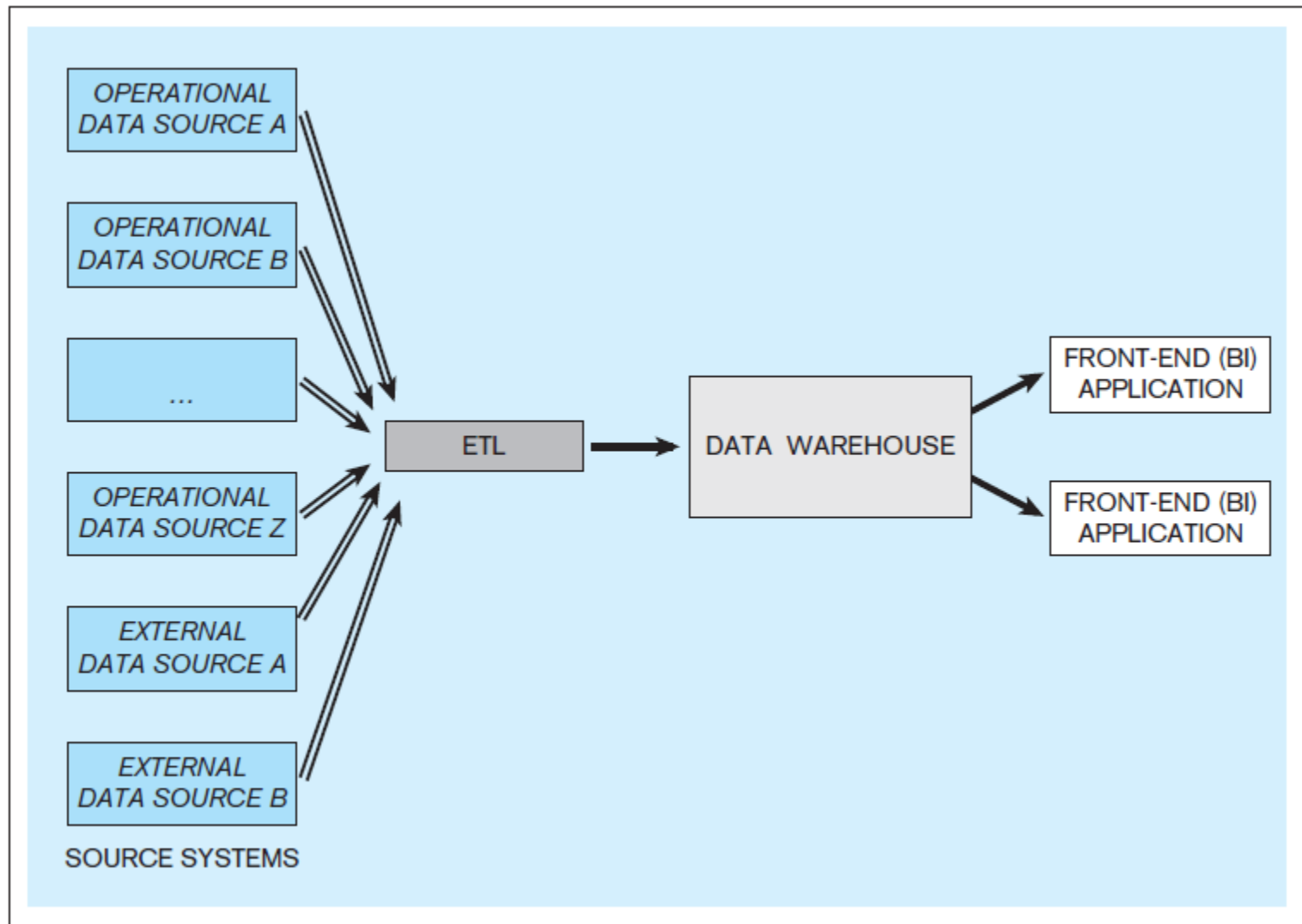
- *Extracting* analytically useful data from the operational data sources
- *Transforming* such data so that it conforms to the structure of the subject-oriented target data warehouse model (while ensuring the quality of the transformed data)
- *Loading* the transformed and quality assured data into the target data warehouse

DATA WAREHOUSE COMPONENTS

Data warehouse front-end (BI) applications

Used to provide access to the data warehouse for users who are engaging in indirect use

Example - A data warehouse with front-end applications



DATA MARTS

Data mart

A data store based on the same principles as a data warehouse, but with a more limited scope

	DATA WAREHOUSE	DATA MART
Subjects	<i>Multiple</i>	<i>Single</i>
Data Sources	<i>Many</i>	<i>Fewer</i>
Typical Size	<i>Very big (routinely terabytes of data and larger)</i>	<i>Not as big</i>
Implementation Time	<i>Relatively long (months, years)</i>	<i>Not as long</i>
Focus	<i>Organization-wide</i>	<i>Often narrower than organization-wide</i>

DATA MARTS

Independent data mart

Stand-alone data mart, created in the same fashion as the data warehouse

Independent data mart has its own source systems and ETL infrastructure

DATA MARTS

Dependent data mart

Does not have its own source systems

The data comes from the data warehouse

STEPS IN THE DEVELOPMENT OF DATA WAREHOUSES

