

BMBM5023

**DIGITAL INTELLIGENCE AND
FUTURE DEVELOPMENT**

DATA WAREHOUSES



RE-CAP

What is Business Intelligence?

BI refers to a variety of software applications used to analyse an organisation's raw data.

What are the 4 major components of BI?

- ✦ The Data Warehouse
- ✦ Business Analytics
- ✦ Business Performance Management
- ✦ User Interface



SESSION OBJECTIVES

Understand the basic definitions and concepts of data warehouses

Learn different types of data warehousing architectures; their comparative advantages and disadvantages

Describe the processes used in developing and managing data warehouses

Explain data warehousing operations



SESSION OBJECTIVES

- Explain the role of data warehouses in decision support
- Explain data integration and the extraction, transformation, and load processes
- Describe real-time data warehousing
- Understand data warehouse administration and security issues



WHAT IS A DATA WAREHOUSE?

A large store of data accumulated from a wide range of sources within a company and used to guide management decisions.



DW COMPONENTS

Data Mart - A departmental small-scale "DW" that stores only limited/relevant data.

Two types of Data Marts:-

- 1. Dependant Data Mart** - A subset that is created directly from a data warehouse.
- 2. Independent Data Mart** - A small data warehouse designed for a strategic business unit or a department.



OTHER DW COMPONENTS

- ✦ **Operational Data Stores** - A type of database often used as an interim area for a data warehouse.
- ✦ **Oper Marts** - An operational data mart.
- ✦ **Enterprise Data Warehouse** - A data warehouse for the enterprise. EDW consolidates data from multiple sources in support of campus wide decision making and related information needs such as reporting, analysis, and planning.



OTHER DW COMPONENTS

Metadata – Data that describes other data.

Metadata summarises basic information about data, which can make finding and working with particular instances of data easier.



MAJOR COMPONENTS

Data sources.

- ❖ Data are sourced from operational systems and possibly from external data sources.

Data extraction and transformation.

- ❖ Data are extracted and properly transformed using custom-written or commercial software called ETL.

Data loading.

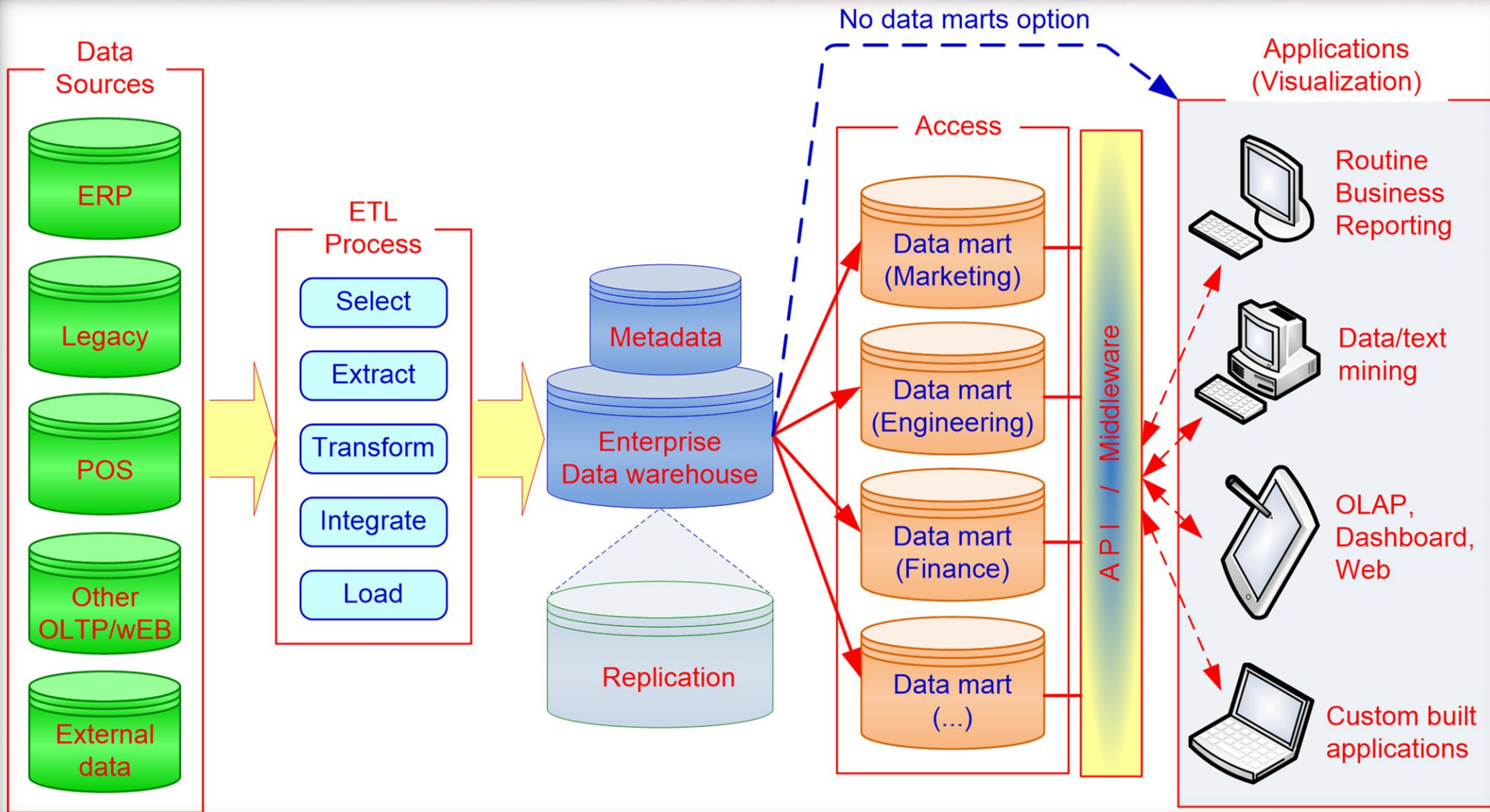
- ❖ Data are loaded into a staging area, where they are transformed and cleansed. The data are then ready to load into the data warehouse.



MAJOR COMPONENTS

- ✦ **Comprehensive database.** This is the EDW that supports decision analysis by providing relevant summarised and detailed information.
- ✦ **Metadata.** Metadata are maintained for access by IT personnel and users. Metadata include rules for organising data summaries that are easy to index and search.
- ✦ **Middleware tools.** Middleware tools enable access to the data warehouse from a variety of front-end applications.





DATA INTEGRATION AND THE EXTRACTION, TRANSFORMATION, AND LOAD (ETL) PROCESS

ETL = Extract Transform Load

Data integration

- Integration that comprises three major processes: data access, data federation, and change capture.

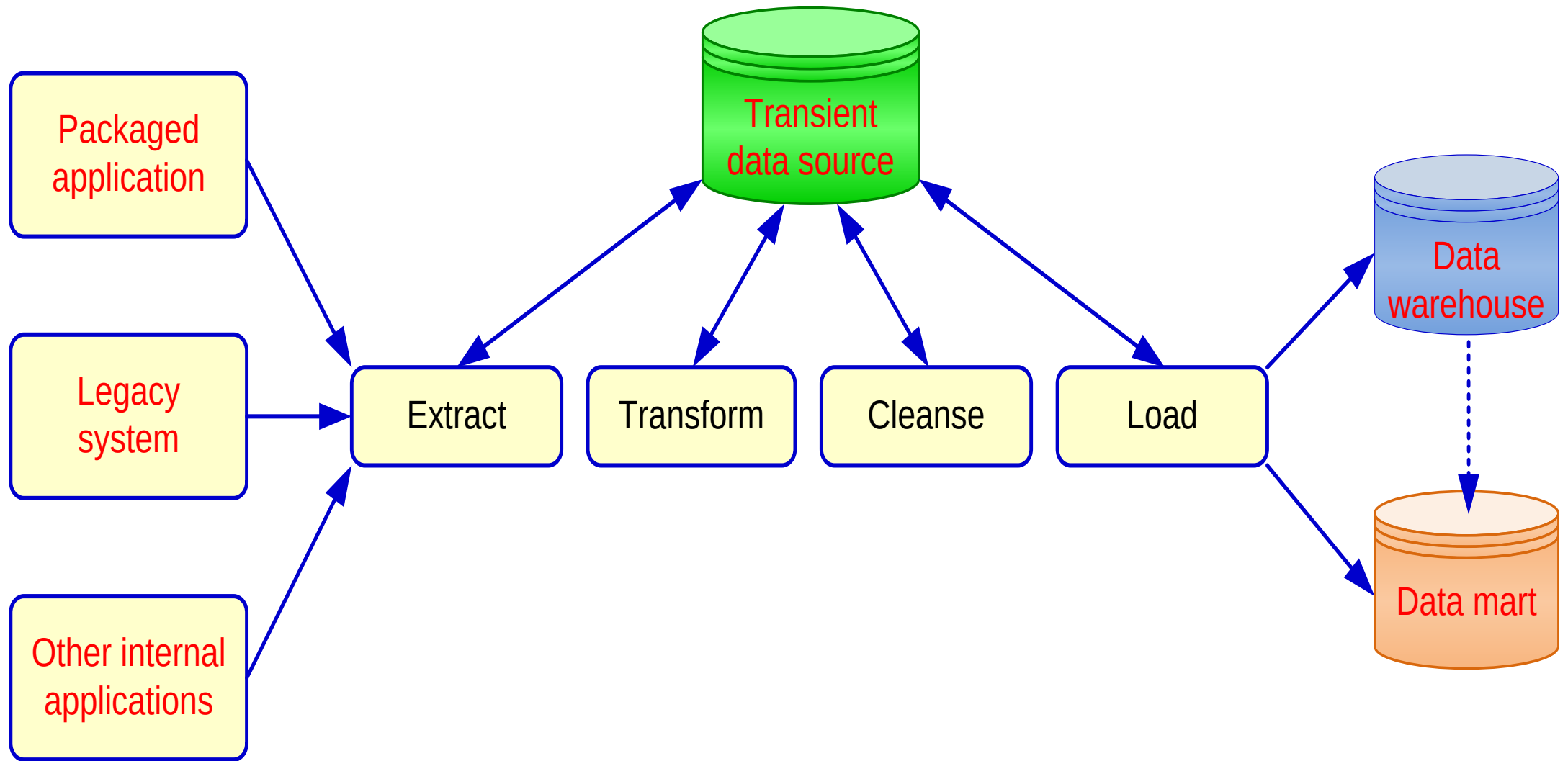
Enterprise application integration (EAI)

- A technology that provides a vehicle for pushing data from source systems into a data warehouse

Enterprise information integration (EII)

- An evolving tool space that promises real-time data integration from a variety of sources, such as relational or multidimensional databases, Web services, etc.





BENEFITS

- ❖ Requires minimal investment in infrastructure
- ❖ Frees up capacity on in-house systems
- ❖ Frees up cash flow
- ❖ Makes powerful solutions affordable
- ❖ Enables solutions that provide for growth
- ❖ Offers better quality equipment and software
- ❖ Provides faster connections



REPRESENTATION OF DATA IN DW

Dimensional Modeling

- ✦ A retrieval-based system that supports high-volume query access

Star schema

- ✦ The most commonly used and the simplest style of dimensional modeling
- ✦ Contain a **fact table** surrounded by and connected to several **dimension tables**

Snowflakes schema

- ✦ An extension of star schema where the diagram resembles a snowflake in shape



MULTIDIMENSIONALITY

The ability to organise, present, and analyse data by several dimensions, such as sales by region, by product, by salesperson, and by time (four dimensions)

Multidimensional presentation

- ▣ Dimensions: products, salespeople, market segments, business units, geographical locations, distribution channels, country, or industry
- ▣ Measures: money, sales volume, head count, inventory profit, actual versus forecast
- ▣ Time: daily, weekly, monthly, quarterly, or yearly



ANALYSIS OF DATA IN DW

OLTP vs. OLAP...

OLTP (Online Transaction Processing)

- ✦ Capturing and storing data
- ✦ The main focus is on efficiency of routine tasks

OLAP (Online Analytical Processing)

- ✦ Converting data into information for decision support
- ✦ Data cubes, drill-down / rollup, slice & dice, ...
- ✦ Requesting ad hoc reports
- ✦ Conducting statistical and other analyses
- ✦ Developing multimedia-based applications



OLAP vs. OLTP

TABLE 2.5 A Comparison Between OLTP and OLAP

Criteria	OLTP	OLAP
Purpose	To carry out day-to-day business functions	To support decision making and provide answers to business and management queries
Data source	Transaction database (a normalized data repository primarily focused on efficiency and consistency)	Data warehouse or data mart (a nonnormalized data repository primarily focused on accuracy and completeness)
Reporting	Routine, periodic, narrowly focused reports	Ad hoc, multidimensional, broadly focused reports and queries
Resource requirements	Ordinary relational databases	Multiprocessor, large-capacity, specialized databases
Execution speed	Fast (recording of business transactions and routine reports)	Slow (resource intensive, complex, large-scale queries)

OLAP OPERATIONS

Slice - a subset of a multidimensional array

Dice - a slice on more than two dimensions

Drill Down/Up - navigating among levels of data ranging from the most summarised (up) to the most detailed (down)

Roll Up - computing all of the data relationships for one or more dimensions

Pivot - used to change the dimensional orientation of a report or an ad hoc query-page display



DW IMPLEMENTATION ISSUES

- Identification of data sources and governance
- Data quality planning, data model design
- ETL tool selection
- Establishment of service-level agreements
- Data transport, data conversion
- Reconciliation process
- End-user support
- Political issues



THINGS TO AVOID

- ✦ Setting expectations that you cannot meet
- ✦ Engaging in politically naive behavior
- ✦ Loading the data warehouse with information just because it is available
- ✦ Believing that data warehousing database design is the same as transactional database design
- ✦ Choosing a data warehouse manager who is technology oriented rather than user oriented



FAILURE FACTORS IN DW PROJECTS

- ✦ Lack of executive sponsorship
- ✦ Unclear business objectives
- ✦ Cultural issues being ignored
- ✦ Change management
- ✦ Unrealistic expectations
- ✦ Inappropriate architecture
- ✦ Low data quality / missing information
- ✦ Loading data just because it is available



REAL-TIME / ACTIVE DW / BI

Enabling real-time data updates for real-time analysis and real-time decision making is growing rapidly.

- ✦ Push vs. Pull (of data)

Concerns about real-time BI

- ✦ Not all data should be updated continuously
- ✦ Mismatch of reports generated minutes apart
- ✦ May be cost prohibitive
- ✦ May also be infeasible



DW ADMINISTRATION AND SECURITY

Data warehouse administrator (DWA)

DWA should...

- ✦ have the knowledge of high-performance software, hardware and networking technologies
- ✦ possess solid business knowledge and insight
- ✦ be familiar with the decision-making processes so as to suitably design/maintain the DW structure
- ✦ possess excellent communications skills

Security and privacy is a pressing issue in DW

- ✦ Safeguarding the most valuable assets
- ✦ Government regulations (HIPAA, etc.)
- ✦ Must be explicitly planned and executed



THE FUTURE OF DW

Sourcing...

- ✦ Web, social media, and Big Data
- ✦ Open source software
- ✦ SaaS (software as a service)
- ✦ Cloud computing

Infrastructure...

- ✦ Columnar
- ✦ Real-time DW
- ✦ Data warehouse appliances
- ✦ Data management practices/technologies
- ✦ In-database & In-memory processing New DBMS
- ✦ Advanced analytics



ANY QUESTIONS ?

