Data and Knowledge Management

Pertemuan 3

Universitas Bunda Mulia

Kompetensi Khusus

- Mahasiswa dapat menjelaskan konsep penggunaan basisdata serta manajemen pengetahuan dalam perusahaan (C2)

Materi :

1. Managing Data
2. The Database Approach
3. Big Data
4. Data Warehouse and Data Marts
5. Knowledge Management

1. Managing Data

1.1 Managing Data

- Data is a collection of facts, such as numbers, words, measurements, observations or even just descriptions of things.

1.1 Managing Data (Lanj)

- All IT applications require data. These data should be of high quality, meaning that they should be accurate, complete, timely, consistent, accessible, relevant, and concise. Unfortunately, the process of acquiring, keeping, and managing data is becoming increasingly difficult.

1.2 The Difficulties of Managing Data

- The amount of data increases exponentially with time.
- Data are also scattered throughout organizations
- Data collected by many individuals using various methods and devices.
- Data frequently stored in numerous servers and location and in different computing systems, database, format and human and computer languages.
- Data are generated from multiple sources (internal sources, personal sources & external sources.

1.3 Data Governance

- To address the numerous problems associated with data, organizations are turning to data governance.
- What is data governance?

1.3 Data Governance (Lanj)

- Data governance is an approach to managing information across an entire organization.
- Involves a formal set of business processes and policies that are designed to ensure that data are handled in a certain, well-defined fashion.
- The objective is to make information available, transparent, and useful for the people who are authorized to access it, from the moment it enters an organization until it is outdated and deleted.

1.4 Implementing Data Governance

- One strategy, for implementing data governance is master data management.
- **Master Data Management** is a process that spans all organizational business processes and application.
- **Master Data Management** provides companies with the ability store maintain exchange and synchronize a consistent, accurate and timely" single version of the truth" for the company's **master data.**

1.5 Master Data

- Master data are a set of core data, such as customer, product, employee, vendor, geographic location, that span the enterprise information system
- Its important to distinguish between master data and transaction data.
- Transaction data which are generated and captured by operational systems, describe the business's activities or transaction. In contrast, master data are applied to multiple transactions and are used to categorize, aggregate and evaluate the transaction data

2. The Database Approach

2.1 Data File

- A data file is a collection of logically related records. In a file management environment each application has a specific data file related to it.
- This file contains all of the data records the application requires.

2.1 Data File Example

- Imagine that most of your information is store in your university's central database, but a club to which you belong maintains its own files, the athletics department has separate files for student athletes, and your instructors maintain grade data on their personal computer.
- Its easy for your name to be misspelled in one of databases or files but not in others. Similarly if you move, then your address might be updated correctly in one database or file but in others.

2.3 Database System

- Database system minimize the following problems:

1. Data redundancy : the same data are stored in multiple location
2. Data isolation : Applications cannot access data associated with other applications
3. Data inconsistency : Various copies of data do not agree.

2.4 Database System

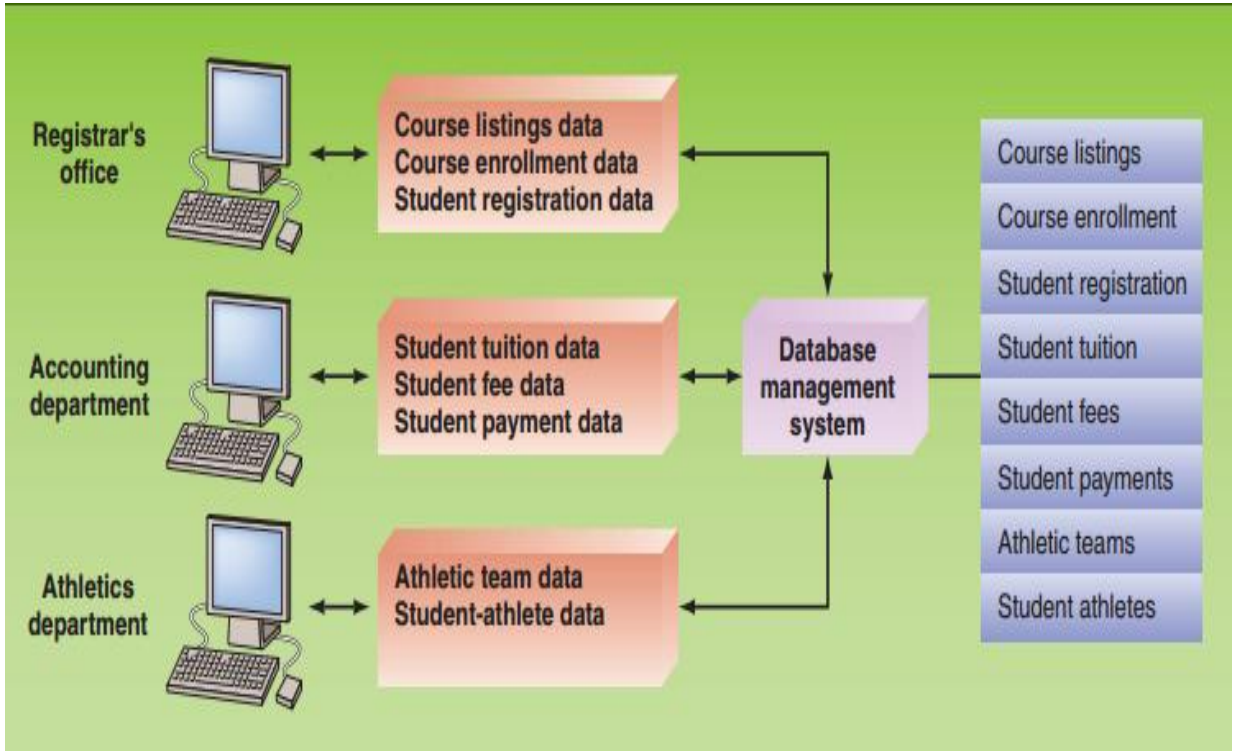
- In additional, database system maximize the following

1. Data security : Because data “put in one place” in database, there is risk of losing a lot of data at one time. Therefore, database must have extremely high security measures in place to minimize mistakes and deter attacks.
2. Data integrity : data meet certain constraints: There are no alphabetic characters in a Social Security number field.

2.4 Database System (Lanj)

3. Data independence : applications and data are independent of one another: that is applications and data are not linked to each other, so all applications are able to access the same data.

2.5. Database Management System

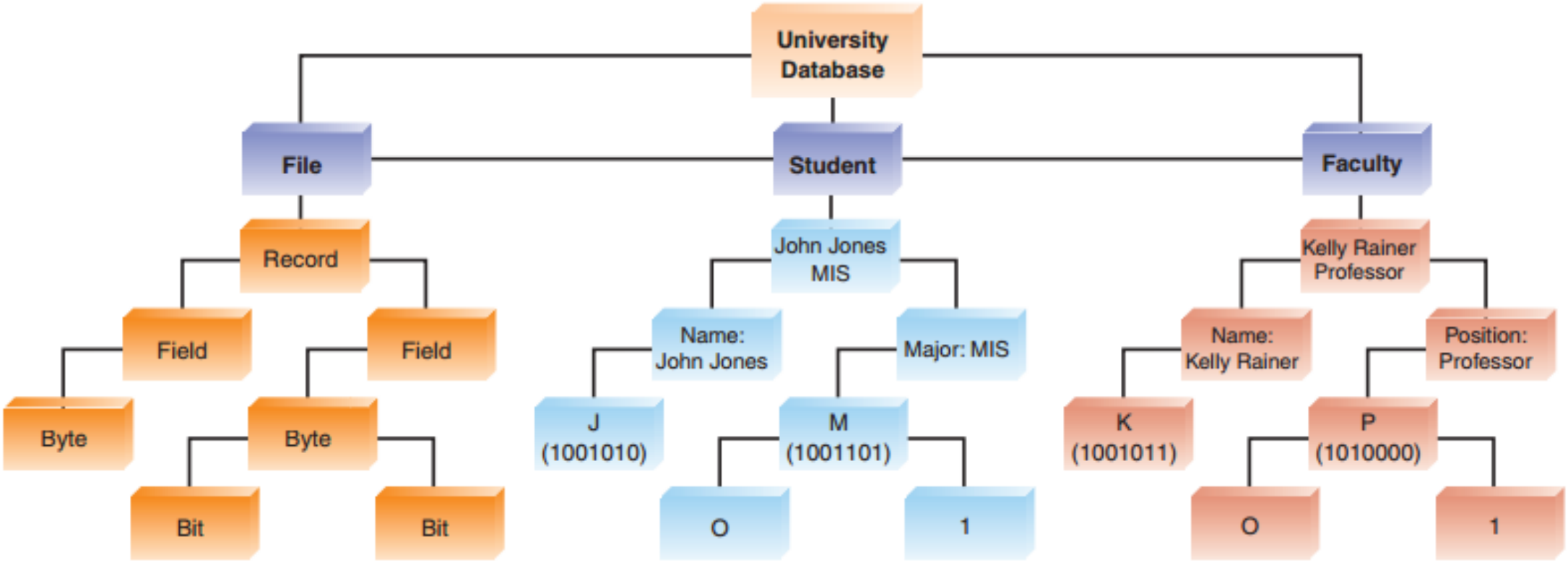


- Figure 1. **Illustrates a university database.**
- Note that university application from the registrar's office, the accounting department and the athletics department access data through the database management system.

2.6 The Data Hierarchy

- Data are organized in a hierarchy that begins with bits and proceeds all the way to databases.
- For examples a student name in a university computer files would appear in the “name” field, and her or his social security number would appear in the social security number field.
- Field can also contain data other than text and number, they can contain an image, or any other type of multimedia. Examples are a motor vehicle departments licensing database that contains a driver’s photograph and a field that contains a voice sample to authorize access to a secure facility

2.7 Hierarchy Of Data for a Computer Based File



2.8 Database Management System

- A database management system (DBMS) is a set of program that provide users with tools to create and manage a database.
- Managing a database refers to the processes of adding, deleting, accessing, modifying, and analyzing data stored in a database.
- DBMS also provide the mechanisms for maintaining the integrity of stored data, managing security and user access and recovering information if the system fails.
- Database and DBMS are essential to all areas of business, they must be carefully managed.

3. Big Data

3.1 Introduction

- Organization and individuals must process a vast amount of data that continues to rapidly increase.
- According to IDC the world generates exabytes of data each year (one trillion terabytes)
- Furthermore, the amount of data produced worldwide in increasing by 50% each year.

3.2 Big Data

- Big data is a collection of data so large and complex that it is difficult to manage using traditional database management systems.
- Essentially big data is about predictions. Predictions do not come from “teaching” computer to “think” like human. Instead prediction come from applying mathematics to huge quantities of data to infer probabilities.

3.3 Characteristics of Big Data

- Big data has three district characteristics:

1. Volume
2. Velocity
3. Variety

- These characteristics distinguish Big Data from traditional data.

3.3.1 Volume

- We have noted the huge volume of big data. Consider machine-generated data which are generated in much larger quantities than nontraditional data. For examples sensor in a single jet engine can generate 10 terabytes of data in 30 minutes.

3.3.2 Velocity

- The rate at which data flow into an organization is rapidly increasing. Velocity is critical because it increases the speed of the feedback loop between a company, its customers, its suppliers, and its business partner.
- Examples: the internet and mobile technology enable online retailers to compile histories not only on final sales but also in their customers every click and interaction.

3.3.3 Variety

- Traditional data formats tend to be structured and relativity well described and they change slowly.
- Traditional data include financial market data, point-of-sale transactions, and much more.
- Big data format change rapidly. They include satellite imagery, broadcast audio streams, digital music files, web page content, scans of government documents and comments posted on social networks.

3.4 Issues with Big Data

- Despite its extreme value, big data does have issues, in this section we take a look at data integrity, data quality, and the nuances of analysis that are worth noting

1. Big data can come from untrusted sources
2. Big data is dirty
3. Big data changes especially in data streams.

4. Data Warehouses and Data Marts

4.1 What is Data Warehouse?

- Defined in many different ways, but not rigorously.
  - A decision support database that is maintained **separately** from the organization's operational database
  - Support information processing by providing a solid platform of consolidated, historical data for analysis.
- "A data warehouse is a **subject-oriented**, **integrated**, **time-variant**, and **nonvolatile** collection of data in support of management's decision-making process." W. H. Inmon
- Data warehousing:
  - The process of constructing and using data warehouses

4.1 What is Data Warehouse? (Cont..)

- A *data warehouse* is a pool of data organized in a *format* that enables users to ***interpret*** data and convert it into useful ***information*** to gain ***knowledge*** from this interpretation.
- It is a single place that contains ***complete*** and ***consistent*** data from ***multiple*** sources.
- Data warehousing is the act of a business person extracting ***business value*** from the data stored in the data warehouse.

4.2 Data Warehouse—Subject-Oriented

- Organized around major subjects, such as customer, product, sales.
- Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing.
- Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.

4.2 Data Warehouse Subject-Oriented

- **DW organized around major *subjects***
  - Insurance company: customer, premium, claim
- **Conventional database organized around *applications***
  - Insurance company: auto, health, life

4.3 Data Warehouse—Integrated

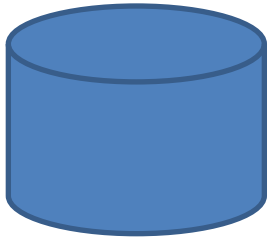
- **Constructed by integrating multiple, heterogeneous data sources**
  - relational databases, on-line transaction records
- **Data cleaning and data integration techniques are applied**.
  - Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources
    - E.g., Hotel price: currency, tax, breakfast covered, etc.
  - When data is moved to the warehouse, it is converted.

4.4 Data Warehouse—Time Variant

- **The time horizon for the data warehouse is significantly longer than that of operational systems.**
  - Operational database: current value data.
  - Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)
- **Every key structure in the data warehouse**
  - Contains an element of time, explicitly or implicitly
  - But the key of operational data may or may not contain “time element”.

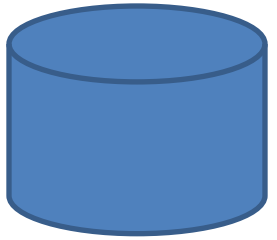
4.4 Data Warehouse—Time Variant (Lanj)

Operational DB



- key may / may not have element of time
- time horizon 60-90 days

Data Warehouse



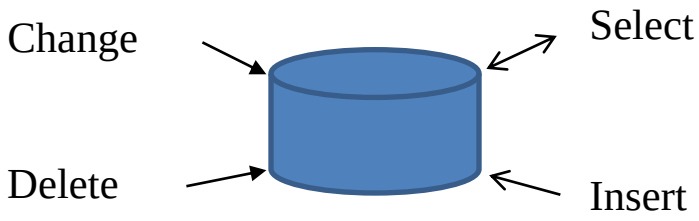
- key contains element of time
- time horizon 5-10 years

4.5 Data Warehouse—Non-Volatile

- A physically separate store of data transformed from the operational environment.
- Operational update of data does not occur in the data warehouse environment.
  - Does not require transaction processing, recovery, and concurrency control mechanisms
  - Requires only two operations in data accessing:
    - *initial loading of data* and *access of data*.

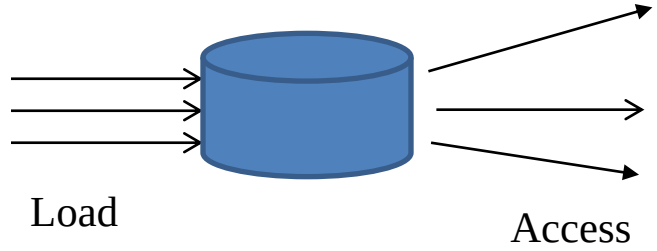
4.5 Data Warehouse Non-Volatile (Lanj..)

Data Warehouse



- Data is not updated, snapshot
- Loaded usually en masse

Operational DB



- Data can be updated
- One record at a time

UNIVERSITAS BUNDA MULIA

4.6 Data Warehouse Architecture

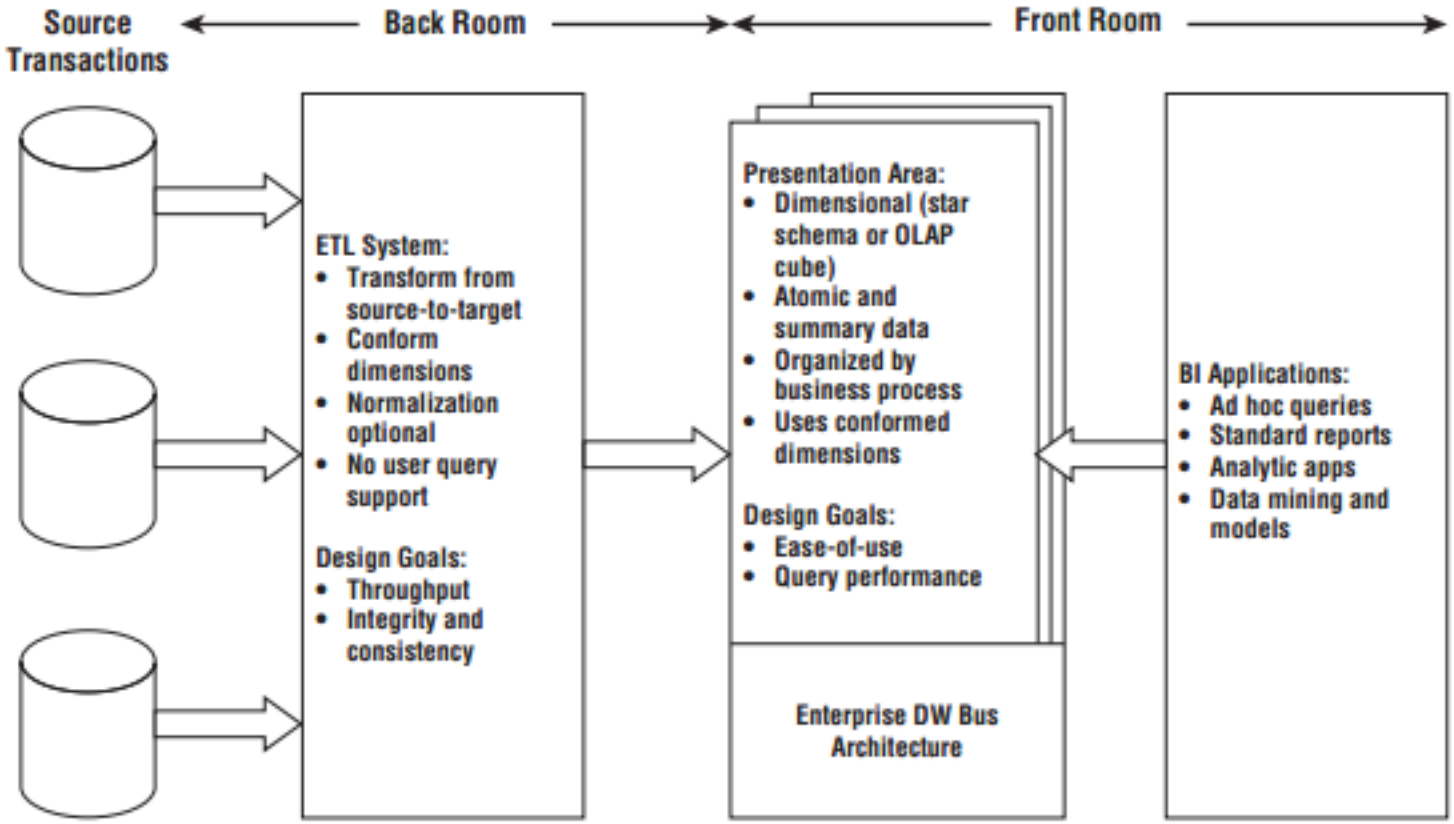


Fig 1. Core elements of the Kimball DW/BI architecture

Sumber: Kimball Ralph., Ross Margy, 2013

4.6 Data Warehouse Architecture

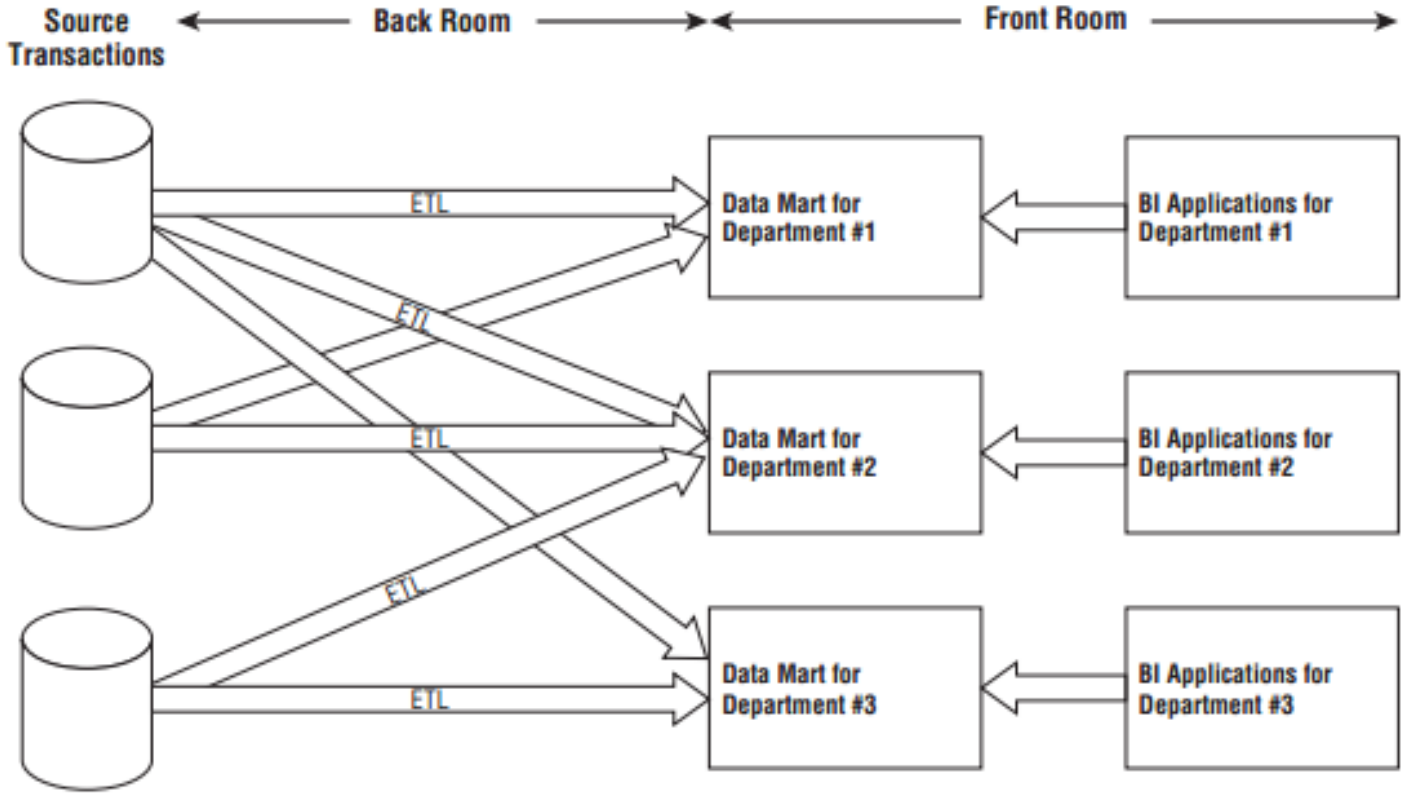


Fig 1. Simplified illustration of the independent data mart "architecture."

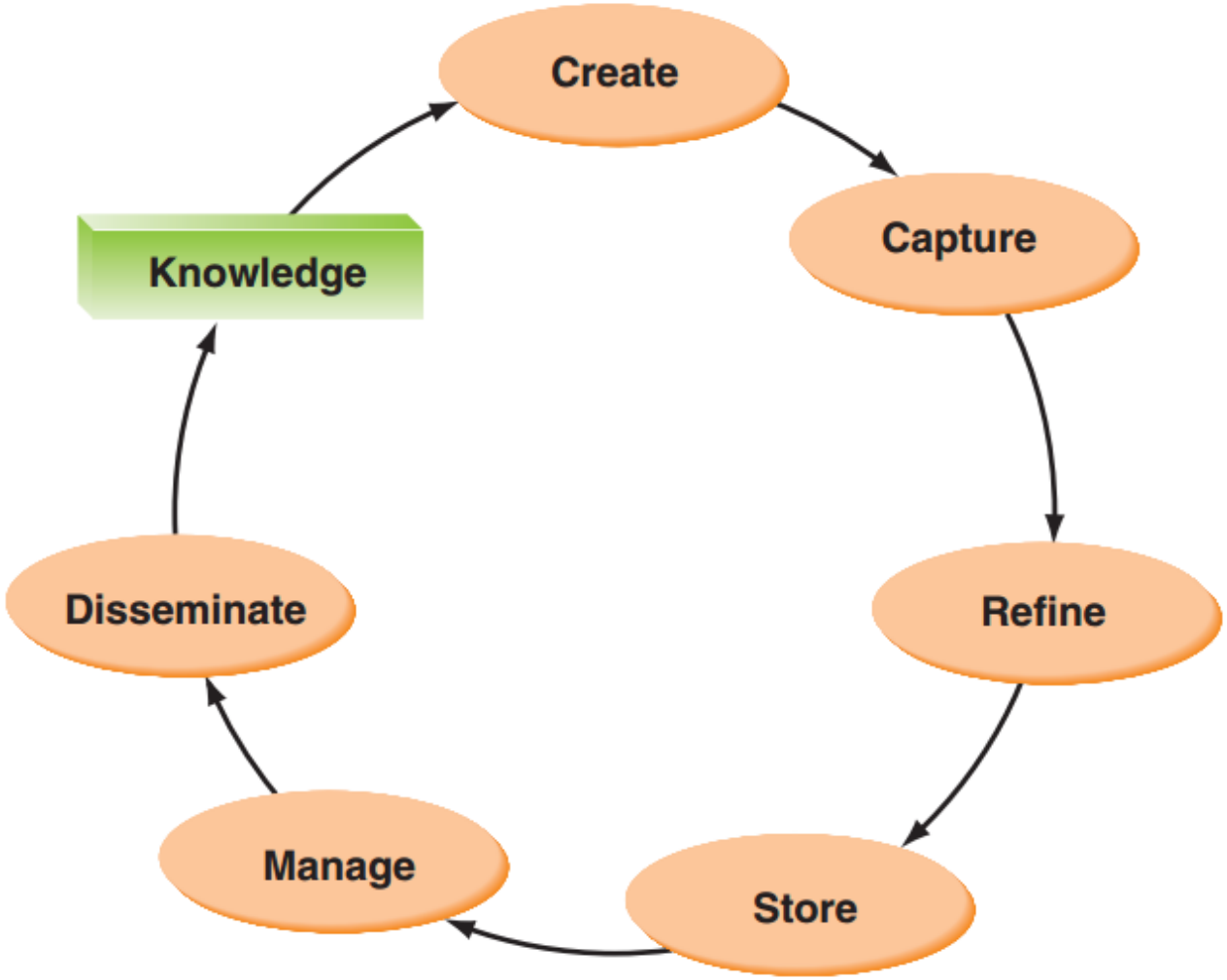
Sumber: Kimball Ralph., Ross Margy, 2013

5. Knowledge Management Systems

5.1 Knowledge Management

- The goal of knowledge management is to help an organization make the most productive use of the knowledge it has accumulated.
- Knowledge management systems (KMSs) refer to the use of modern information technologies—the Internet, intranets, extranets, databases—to systematize, enhance, and expedite intrafirm and interfirm knowledge management.

5.2 The KMS Cycle



Summary

- A data warehouse is a subject-oriented, integrated, time-variant, and nonvolatile collection of data in support of management's decision-making process.
- Key Characteristics of a Data Warehouse are Subject Oriented, integrated, time variant and Non-Volatile.
- Most data warehouses are used for business intelligence to enhance CRM and for data mining.

Summary

- Big Data is composed of high volume, high velocity, and high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery, and process optimization.
- Organizations can realize many benefits with KMSs:
- Best practices are readily available to a wide range of employees.
- Improved customer service.
- More efficient product development.
- Improved employee morale and retention.

TERIMA KASIH