

What is a Data Warehouse?

ทำให้งานง่ายขึ้น

- 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

ทำให้งานง่ายขึ้น

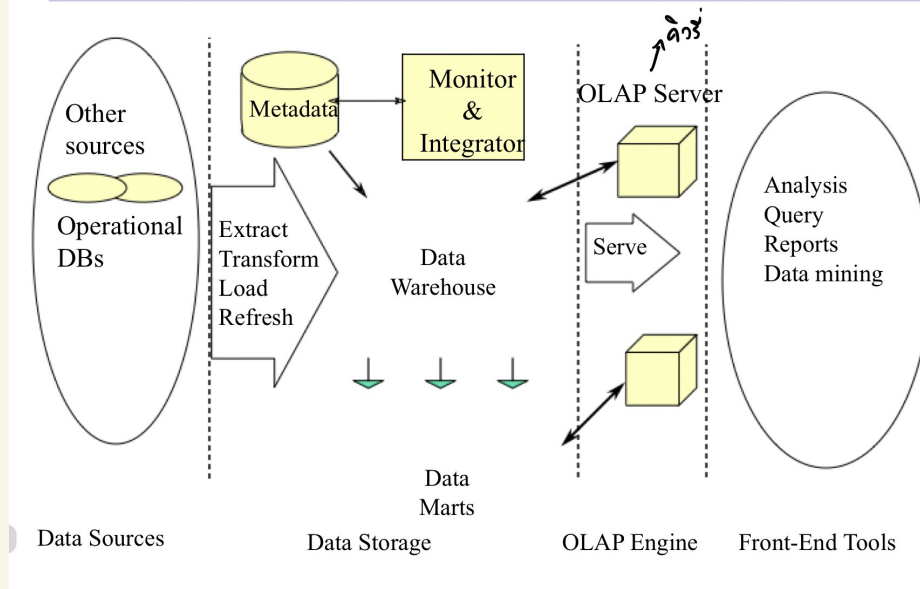
Data Warehouse—Subject-Oriented

- 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**

Data Warehouse—Nonvolatile

- A **physically separate store** of data transformed from the operational environment *Data မှန်ကန်စွာ ထိန်းသိမ်းမှု*
- 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**

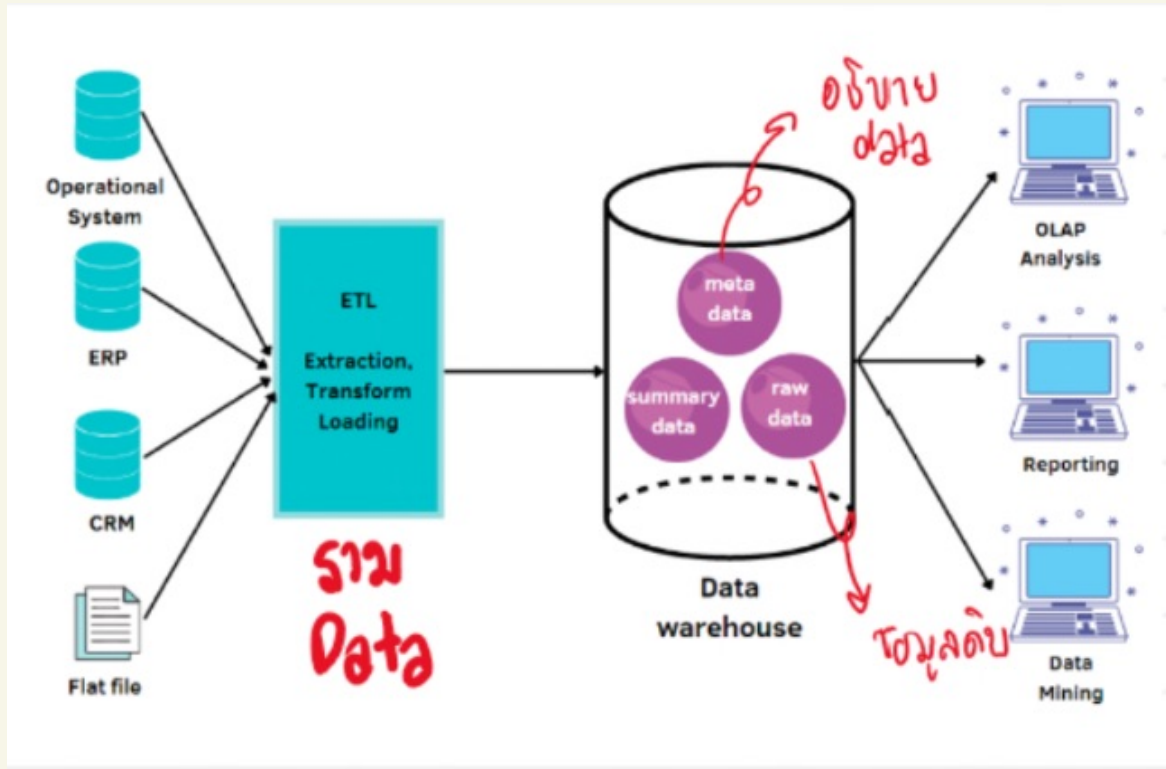
Data Warehouse: A Multi-Tiered Architecture



From Tables and Spreadsheets to Data Cubes

- A **data warehouse** is based on a **multidimensional data model** which views data in the form of a data cube
- A data cube, such as **sales**, allows data to be modeled and viewed in multiple dimensions
- **Dimension tables**, such as **item** (item_name, brand, type), or **time** (day, week, month, quarter, year) *အချက်အလက် မှတ်တမ်း*
- **Fact table** contains **measures** (such as **dollars_sold**) and keys to each of the related dimension tables *→ လက်မှတ်ရေးထိုးမှု*
- In data warehousing literature, an n-D base cube is called a **base cuboid**. The top most 0-D cuboid, which holds the highest-level of summarization, is called the **apex cuboid**. The lattice of cuboids forms a **data cube**.

Data warehouse



Data Lake → เล็ก, น้อย

Data warehouse → ใหญ่ (มีเยอะ)

1) ประเภทของ Data warehouse

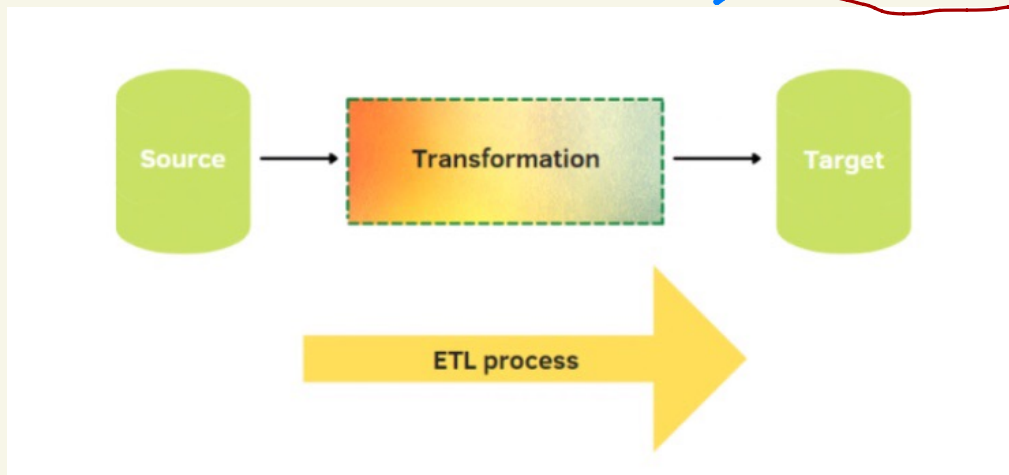
structured → ข้อมูลที่จัดระเบียบ
Unstructured → ข้อมูลที่ไม่สามารถจัดระเบียบ
เช่น รูป, วิดีโอ, ไฟล์เสียง

Data marts → ข้อมูลย่อย (ข้อมูลเฉพาะ)

② OLTP → แล็กในระบบจัดการข้อมูลเพื่อจบงาน (transaction)
เช่น update ข้อมูล, เพิ่มข้อมูล, ลบข้อมูล

OLTP → เป็นระบบจัดการข้อมูลที่มีข้อมูลพร้อมไปพร้อมกัน
เช่น query, filter

③ ETL มสรวม Data



มักเป็น Data Lake → ELT

↓
AI นวัตกรรม

E - Extration

T - Transform

L - Loading → update data → run

① Star Vs Snowflake Schema

⑤ Data Lake

→ เก็บทั้ง structured ด้ย

⑥ mspr=ประมวลผล 2 แบบ

1. Batch Processing → ประมวลผลเป็น 1 ครั้งทำประมวลผล
2. Streaming Processing → ประมวลผลทีละน้อยๆไปเรื่อยๆ