



CS 412 Intro. to Data Mining

Chapter 4. Data Warehousing and On-line Analytical Processing

Jiawei Han, Computer Science, Univ. Illinois at Urbana-Champaign, 2017



Chapter 4: Data Warehousing and On-line Analytical Processing

❑ Data Warehouse: Basic Concepts



❑ Data Warehouse Modeling: Data Cube and OLAP

❑ Data Warehouse Design and Usage

Model ของ Data Warehouse คือ

1. Data Cube คือ การผสมของ Dimension

มี 3 แบบ คือ star Schema , Snowflake Schema , Fact Constellation

2.OLAP (On-Line Analytical Processing)

❑ Data Warehouse Implementation

❑ Summary

What is a 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

รู้ว่าสร้างเพื่อ ? มาจากหลายแหล่ง แต่ละตารางต้องมีช่วงเวลาเดียวกัน

เป็น Data ที่ไม่เคลื่อนไหวเป็น Batch

ต้องมีเป้าหมายในการจะสร้าง Data Warehouse ว่าจะสร้างมาเพื่ออะไร
สร้างมาเพื่อตอบคำถาม , สร้างมาเพื่อ support การตัดสินใจของผู้บริหาร

- ❑ Data warehousing:
 - ❑ The process of constructing and using data warehouses

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
 - ตัวที่อธิบายว่าแต่ละ Dimension ของข้อมูลจำแนกเป็นอะไรบ้าง
❑ **Dimension tables**, such as item (item_name, brand, type), or time(day, week, month, quarter, year)
 - ตารางที่เก็บตัวเลขต่างๆ
❑ **Fact table** contains **measures** (such as dollars_sold) and keys to each of the related dimension tables
- ❑ **Data cube**: A lattice of cuboids
 - ❑ 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**.