




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
- ❑ Data Warehouse Implementation
- ❑ Summary

Model was Data Warehouse

- Data Cube = new Dimension

is a new Star Schema,

Snowflake Schema, Fact Constellation

- OLAP

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
 - Handwritten notes in Thai: “คือ = ว่าเป็น” (is = to be), “ข้อมูลที่รวบรวมไว้” (collected data), “ข้อมูลที่มีลักษณะเฉพาะ” (data with specific characteristics), “ข้อมูลที่เก็บไว้” (data stored), “ข้อมูลที่มีลักษณะเฉพาะ” (data with specific characteristics), “ข้อมูลที่เก็บไว้” (data stored), “ข้อมูลที่มีลักษณะเฉพาะ” (data with specific characteristics), “ข้อมูลที่เก็บไว้” (data stored).
- ❑ Data warehousing:
 - ❑ The process of constructing and using data warehouses

From Tables and Spreadsheets to Data Cubes

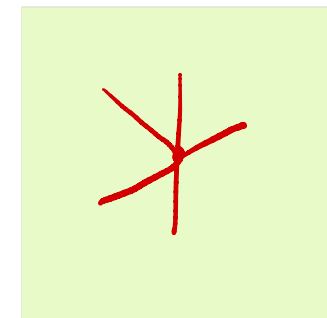
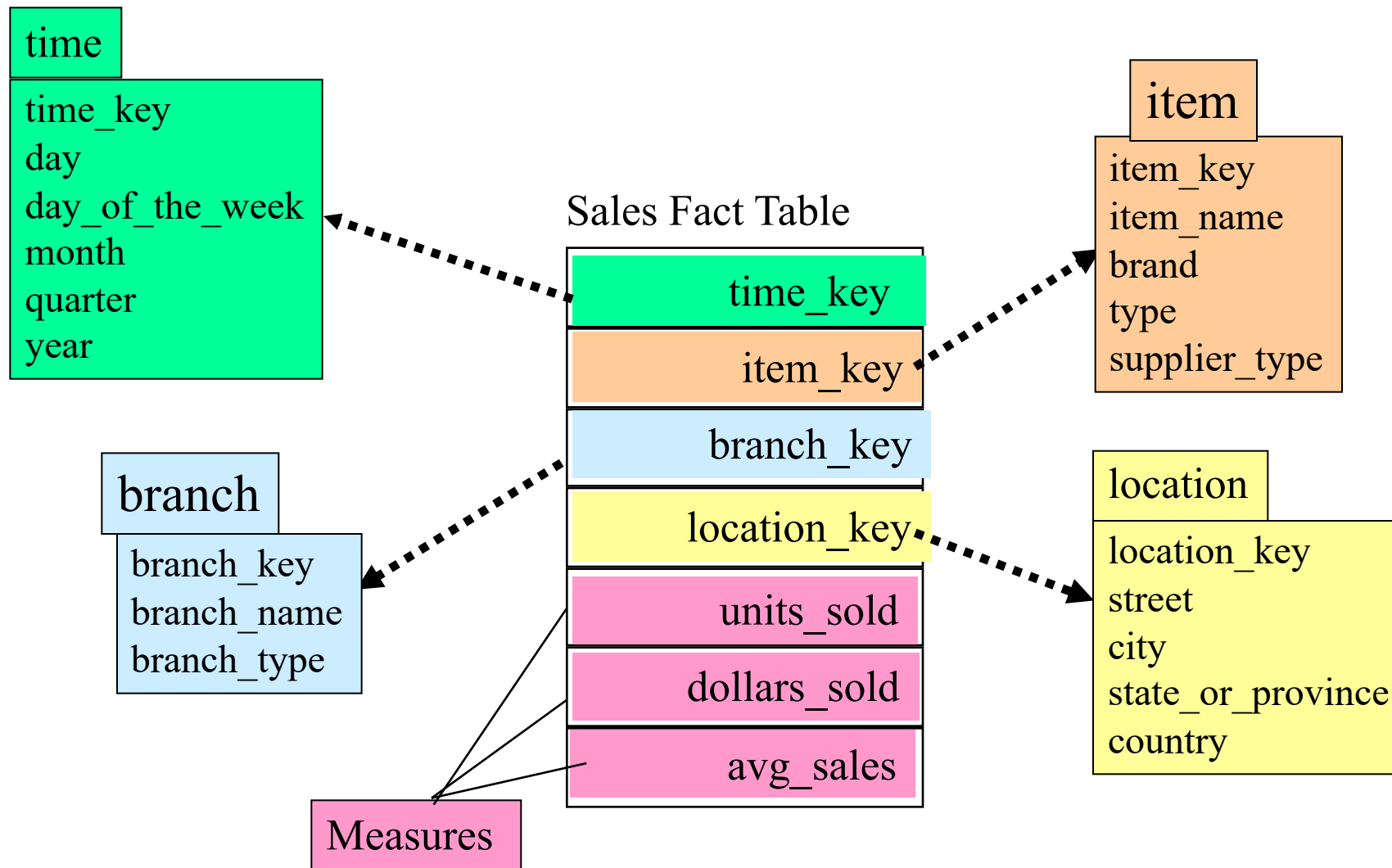
- ❑ A **data warehouse** is based on a multidimensional data model which views data in the form of a data cube
- ❑ A data cube, such as sales, allows data to be modeled and viewed in multiple dimensions
 - ❑ **Dimension tables**, such as item (item_name, brand, type), or time(day, week, month, quarter, year)
 - ❑ ^{measuring} **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**.

Conceptual Modeling of Data Warehouses

- ❑ Modeling data warehouses: dimensions & measures
 - ❑ Star schema: A fact table in the middle connected to a set of dimension tables
 - ❑ Snowflake schema: A refinement of star schema where some dimensional hierarchy is normalized into a set of smaller dimension tables, forming a shape similar to snowflake
 - ❑ Fact constellations: Multiple fact tables share dimension tables, viewed as a collection of stars, therefore called **galaxy schema** or fact constellation

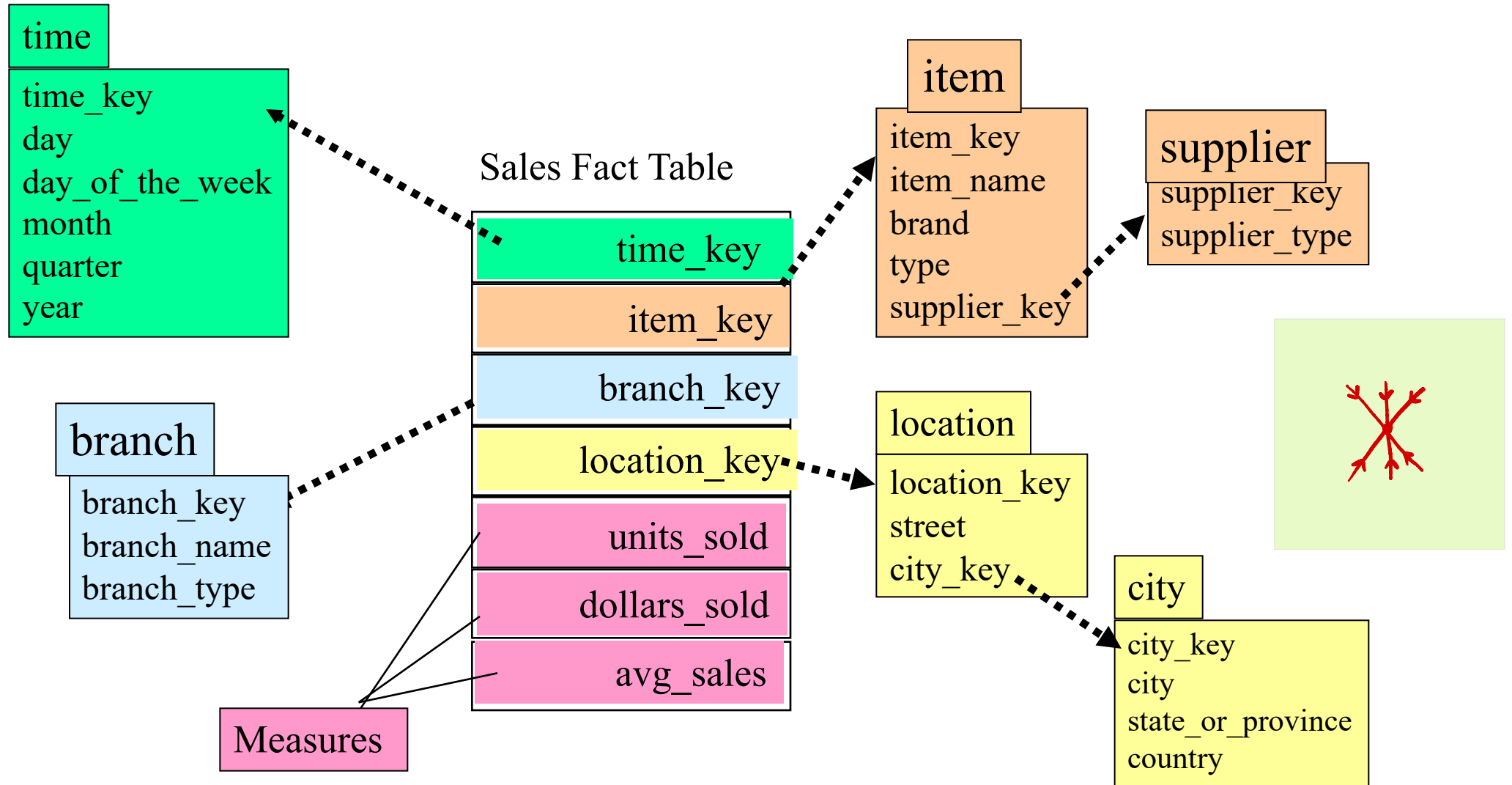
Star Schema: An Example

Star Schema: An Example

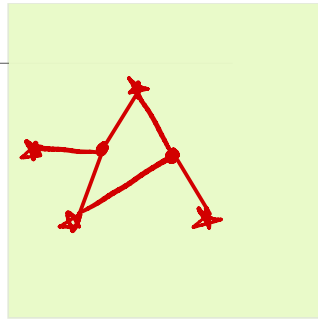
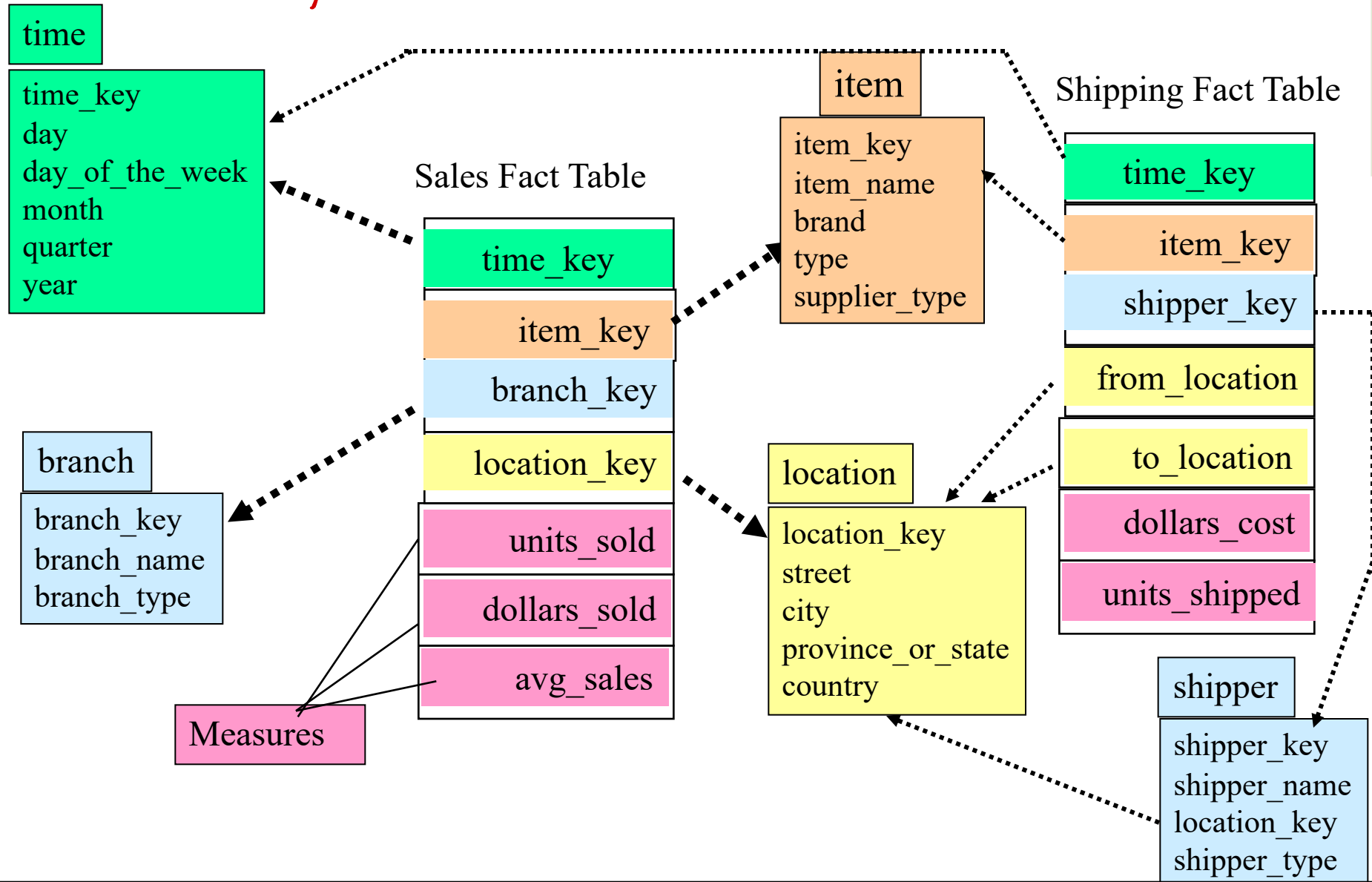


Snowflake Schema: An Example

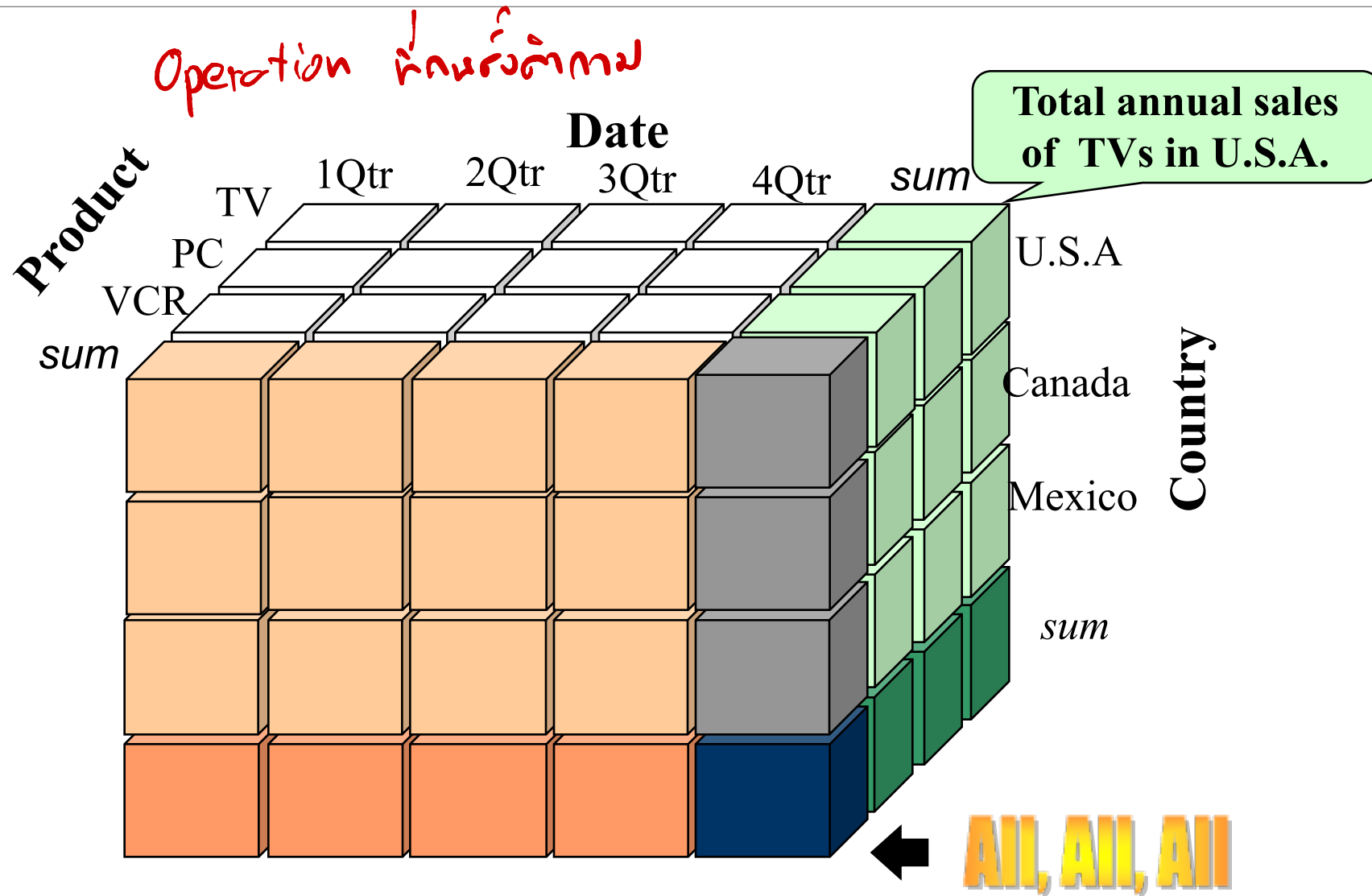
اسمیتا کمار



Fact Constellation: An Example



A Sample Data Cube



Typical OLAP Operations

