

Data warehouse design

based on the book Hovi, Huotari, Lahdenmäki:
Tietokantojen suunnittelu & indeksointi, Docendo
(2003, 2005) chapter 8

© Jouni Huotari & Ari Hovi

Some slides from

[http://www.slideshare.net/idnats/data-](http://www.slideshare.net/idnats/data-warehousing-and-data-mining-presentation-725476)

18.10.2015

[warehousing-and-data-mining-presentation-725476](http://www.slideshare.net/idnats/data-warehousing-and-data-mining-presentation-725476)

jamk.fi

For discussion

- What is a data warehouse?
- How the term Business Intelligence relates to data warehousing?
- How data mart differs from data warehouse?
- Why OLAP (online analytical processing) is important?
- What is the basic idea behind ETL?

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

Basic design principles for data warehouses

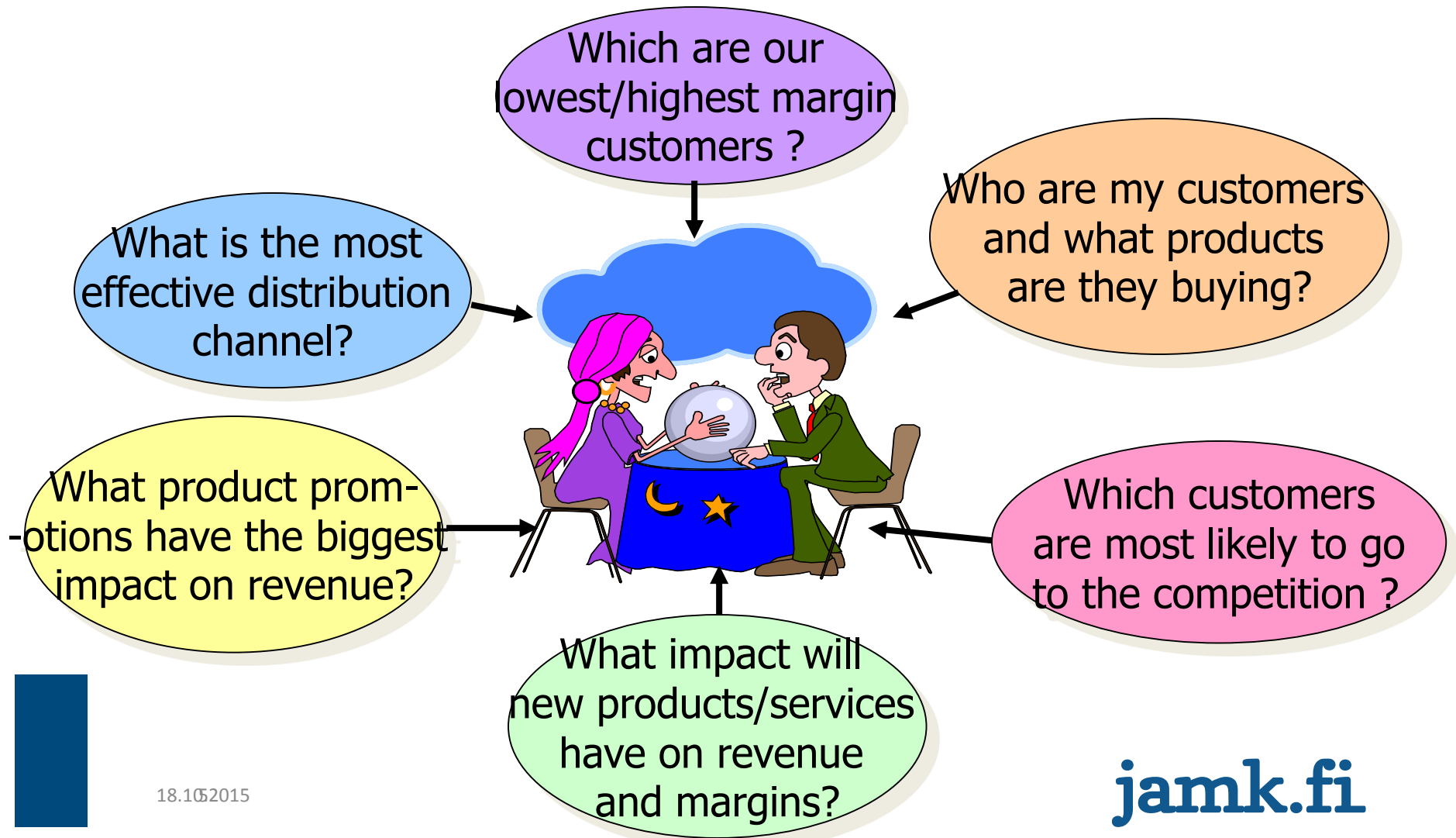
To support fast summary queries, analysis, and reporting

- This is often difficult in operative databases, but some solutions exist
- *can you mention any solutions?*

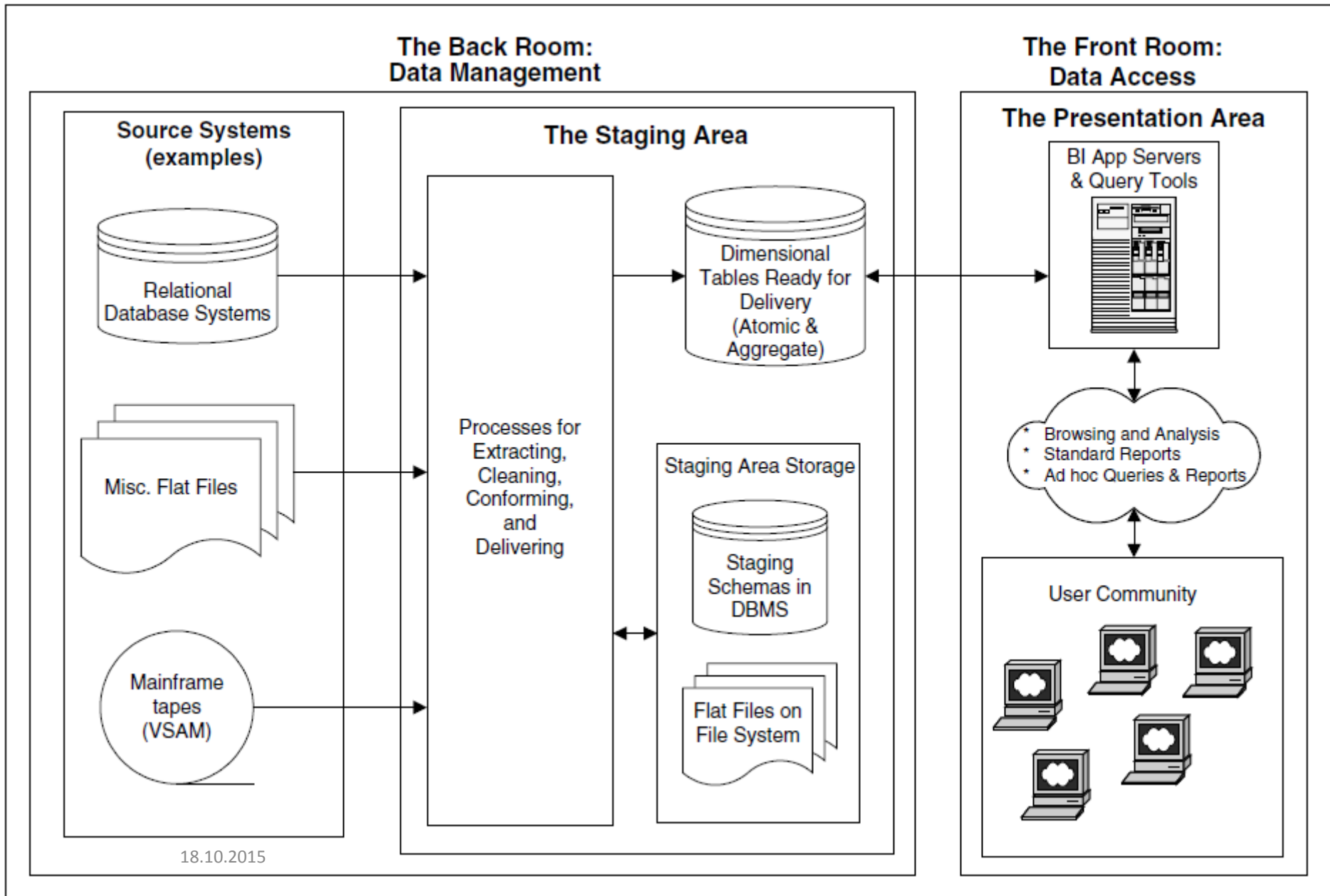
It is important to

- Maintain history for seeing trends etc.
- Clean up and make data consistent from different data sources
- Make the structure clear and understandable

Example: a producer wants to know...



back room and front room of a data warehouse



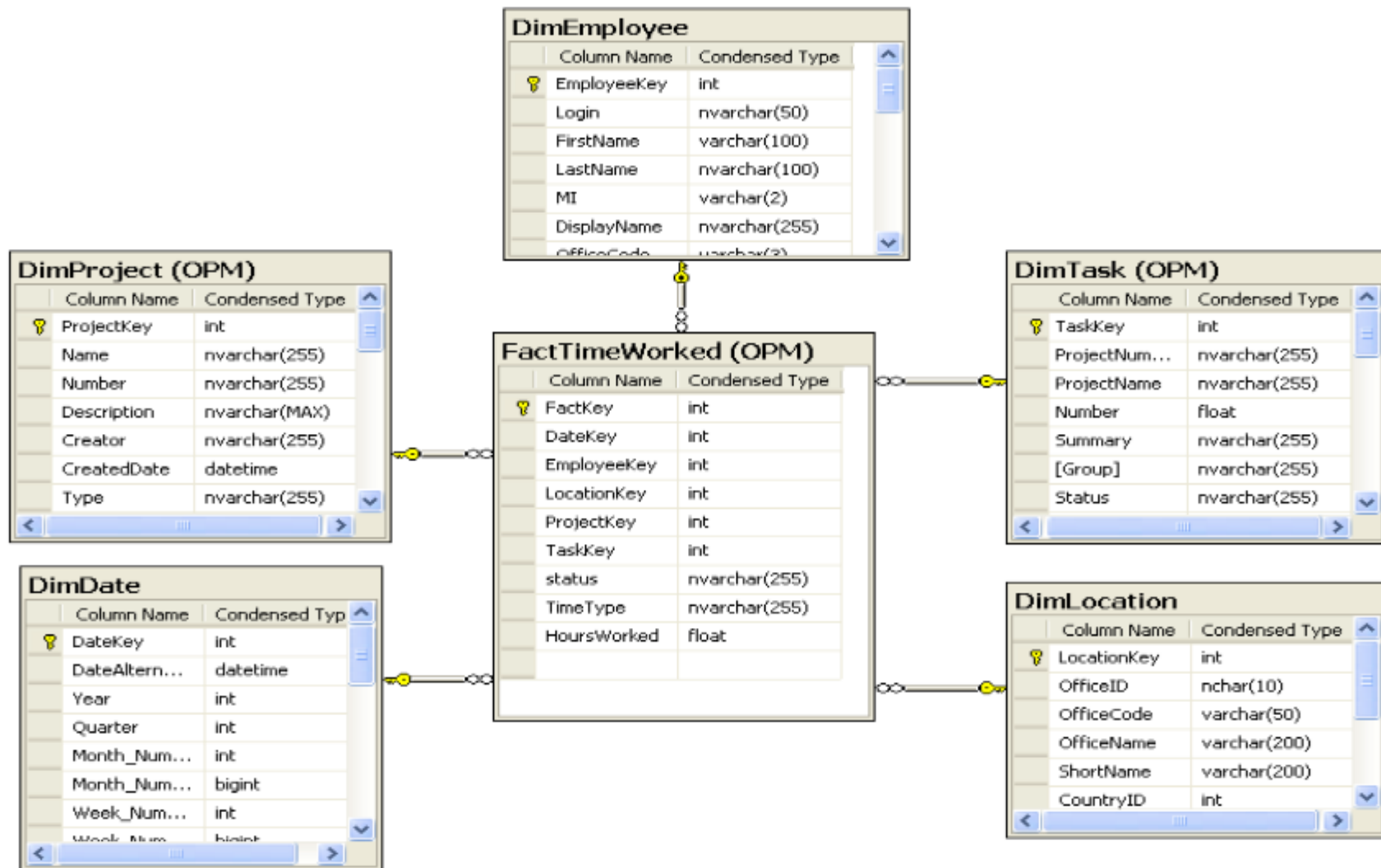
Data Warehouse vs. Operational DBMS

- OLTP (on-line transaction processing)
 - Major task of traditional relational DBMS
 - Day-to-day operations: purchasing, inventory, banking, manufacturing, payroll, registration, accounting, etc.
- OLAP (on-line analytical processing)
 - Major task of data warehouse system
 - Data analysis and decision making
- Distinct features (OLTP vs. OLAP):
 - User and system orientation: customer vs. market
 - Data contents: current, detailed vs. historical, consolidated
 - Database design: ER + application vs. star + subject
 - View: current, local vs. evolutionary, integrated
 - Access patterns: update vs. read-only but complex queries

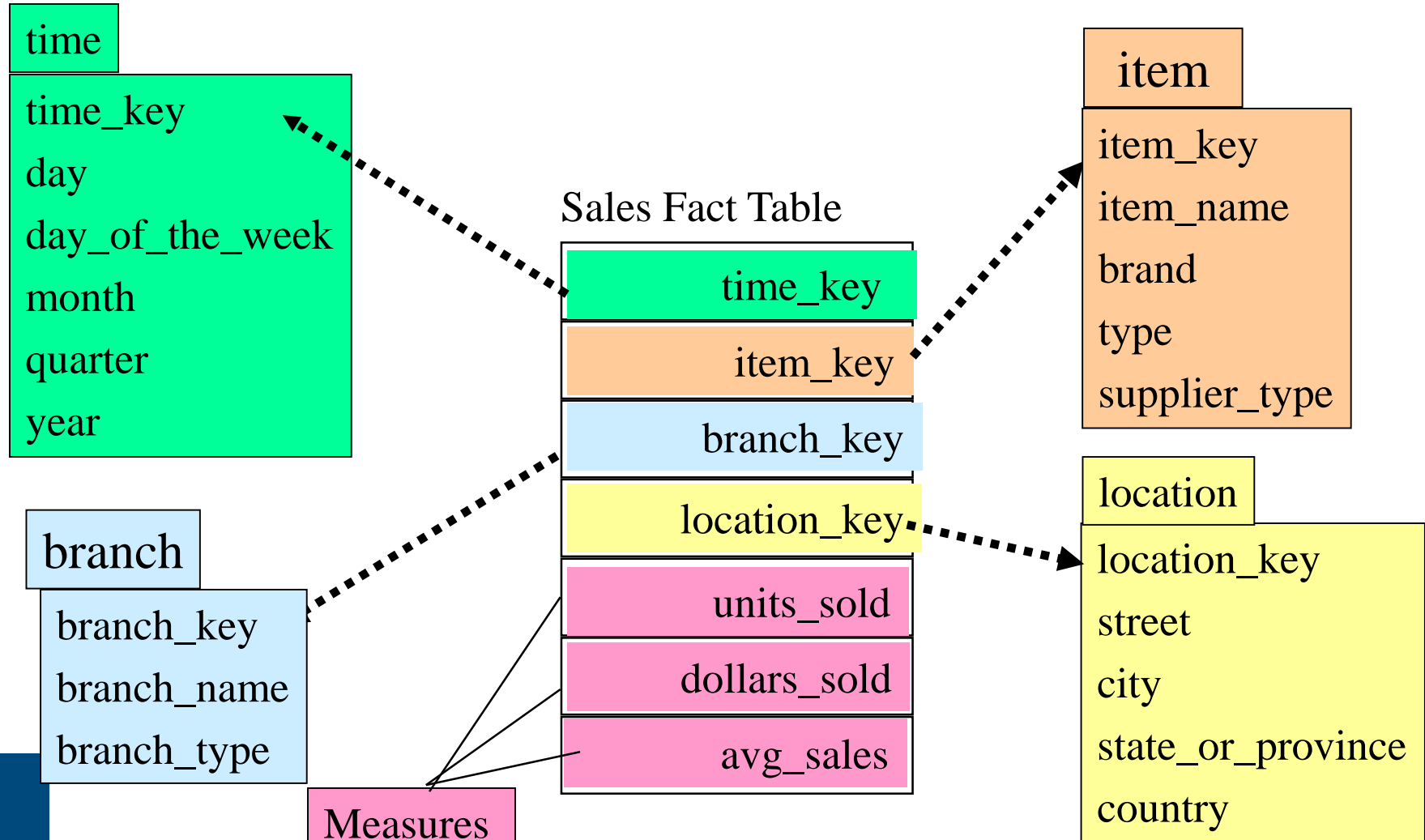
Conceptual Modeling of Data Warehouses

- **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
- **Dimensions** describe who, what, when, where and why for the facts.

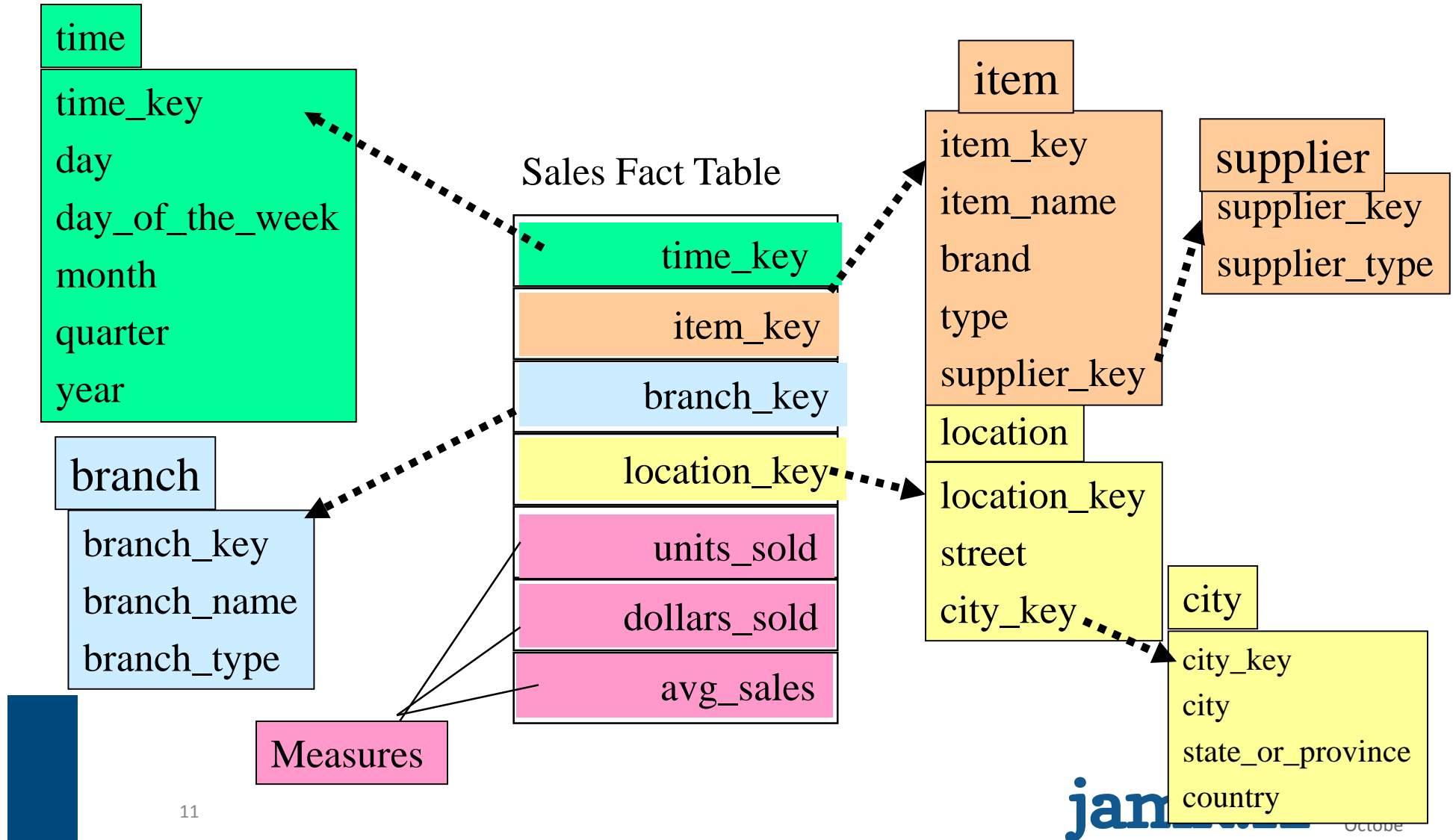
Example of Star Schema



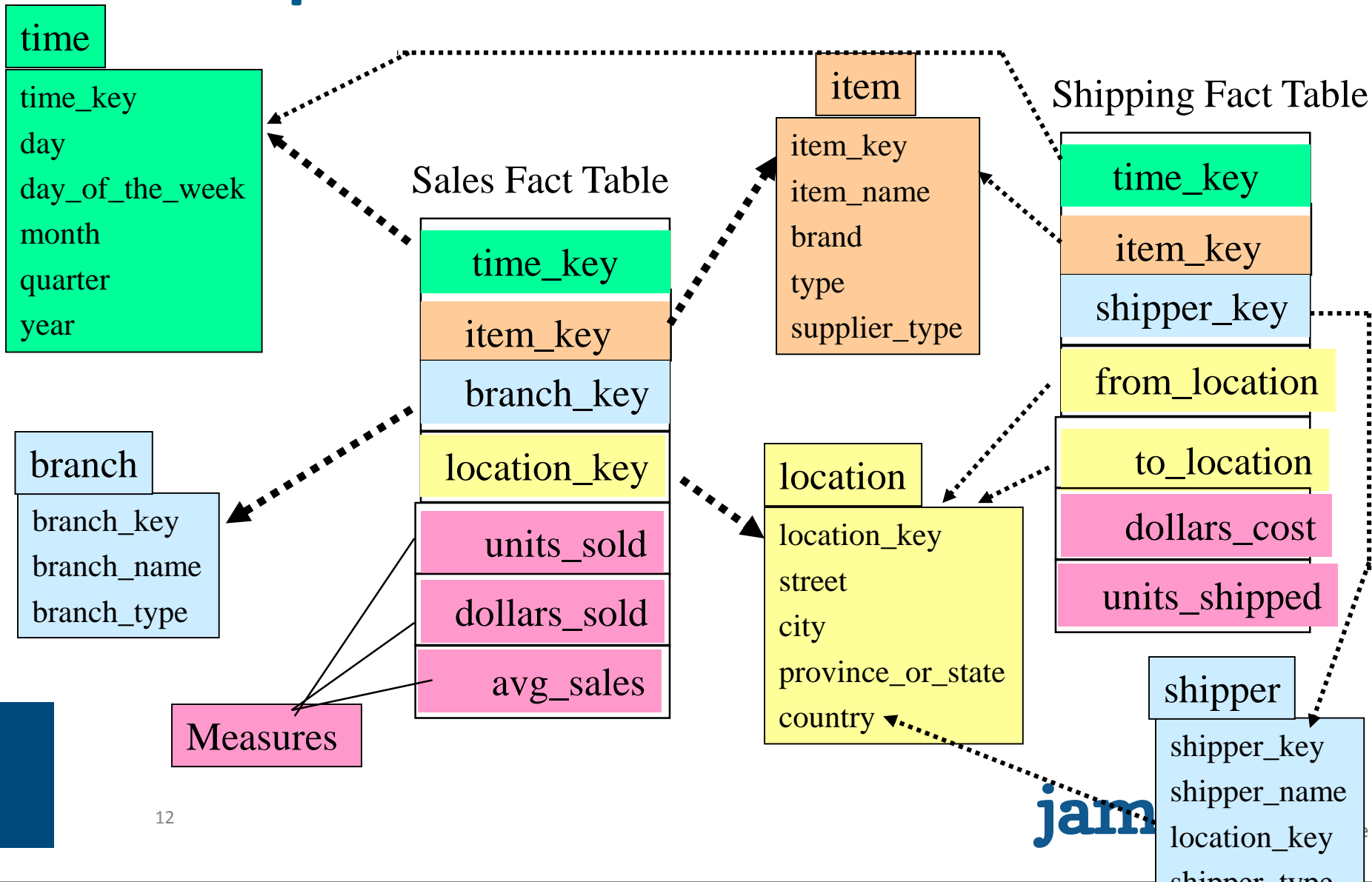
Another example of Star Schema



Example of Snowflake Schema

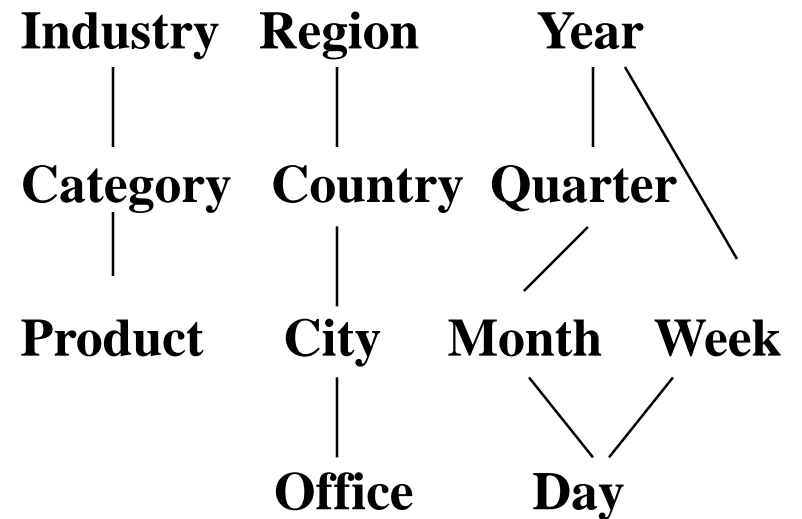
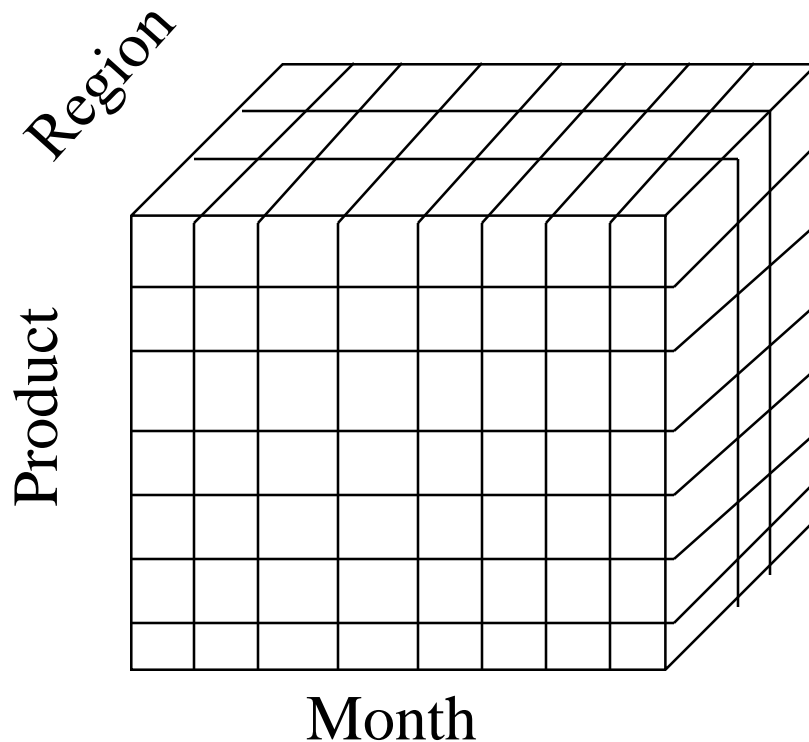


Example of Fact Constellation

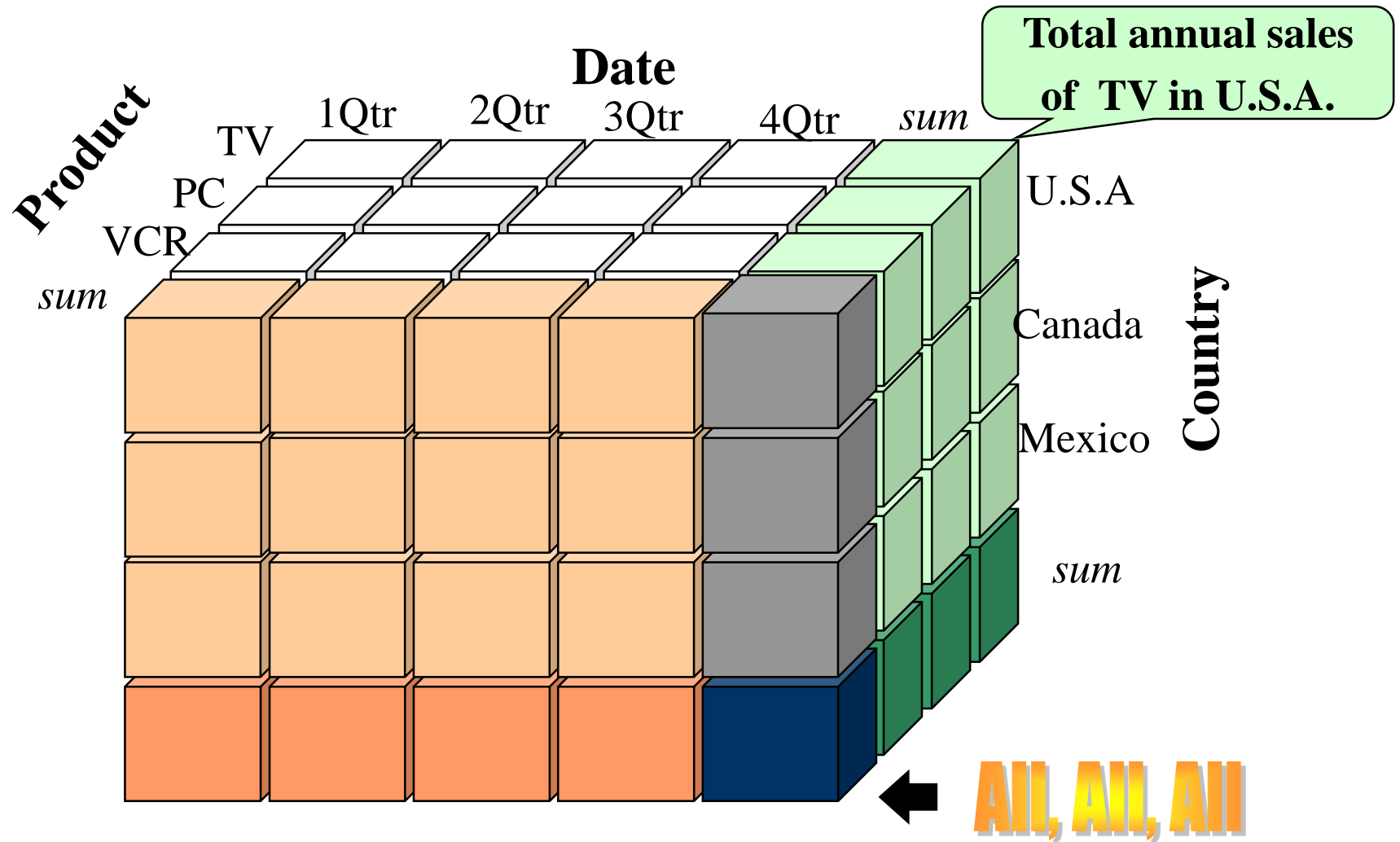


Multidimensional Data

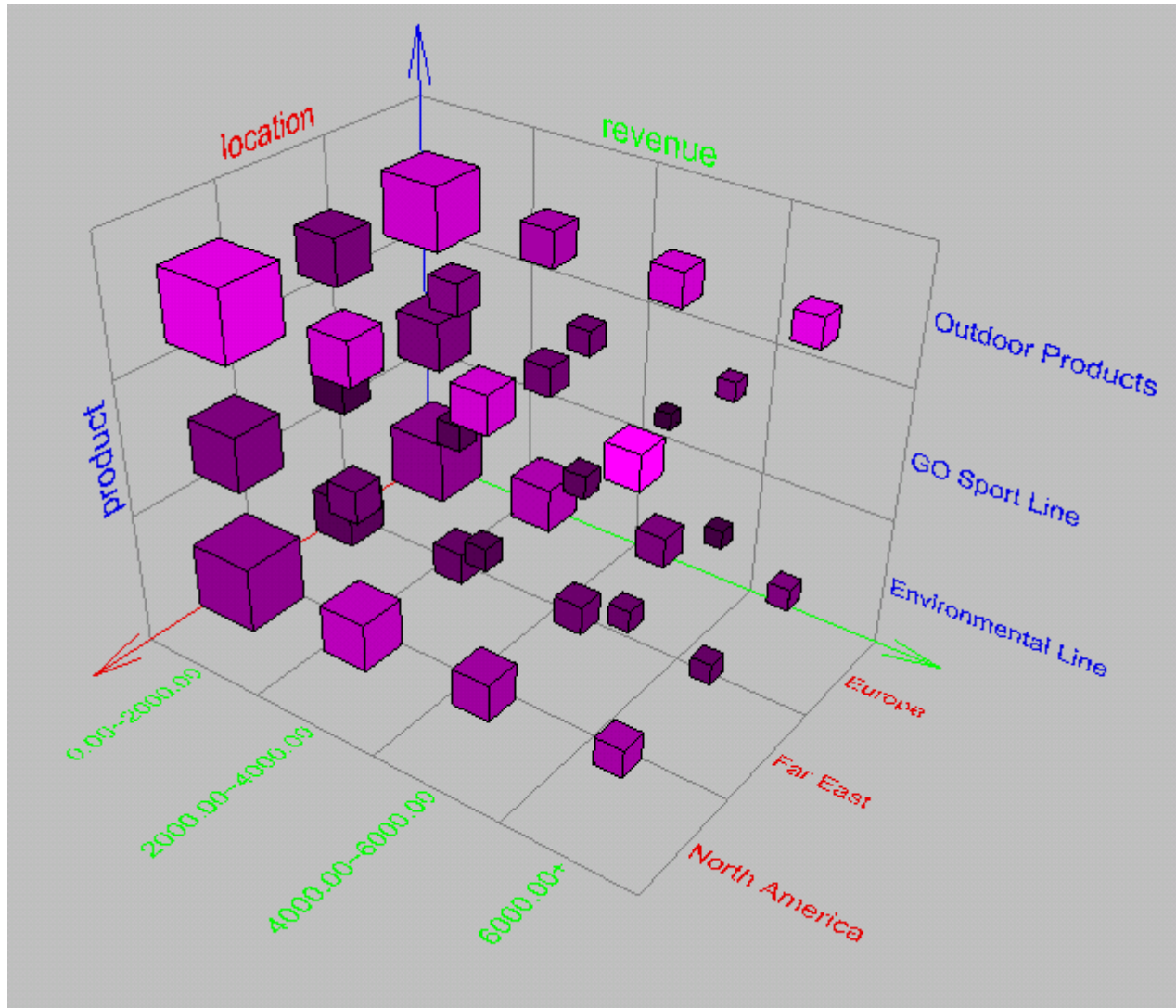
- Sales volume as a function of product, month, and region
- **Dimensions: Product, Location, Time**
- **Hierarchical summarization paths**



A Sample Data Cube

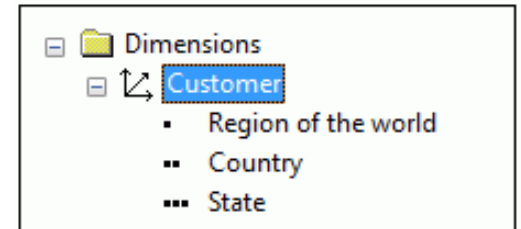


Browsing a Data Cube



- Visualization
- OLAP capabilities
- Interactive manipulation

Exercises



- Install <http://www.olapcube.com/> to your virtual machine and play with the dimensions, create a cube and examine the result (dashboard)
- <http://www.olapcube.com/help/writer/panorama/>
- <http://www.pentaho.com/testdrive>
- [http://www.databaseanswers.org/downloads/Data Warehousing by Example.pdf](http://www.databaseanswers.org/downloads/Data_Warehousing_by_Example.pdf)
- Example (real) data: <http://www.gapminder.org/data/>

What is Business Intelligence (BI)?

- BI refers to skills, technologies, applications and practices used to help a business acquire a better understanding of its commercial context (Wikipedia)
- The goal is to gain insight into the business by bringing together data, formatting it in a way that enables better analysis, and then providing tools that give users power—not just to examine and explore the data, but to quickly understand it. (Business Intelligence with Microsoft Office PerformancePoint Server)

More information from wikipedia:

- http://en.wikipedia.org/wiki/Data_warehouse
- http://en.wikipedia.org/wiki/Data_mart
- http://en.wikipedia.org/wiki/Star_schema
- http://en.wikipedia.org/wiki/Snowflake_schema
- <http://en.wikipedia.org/wiki/OLAP>