Business Intelligence 3

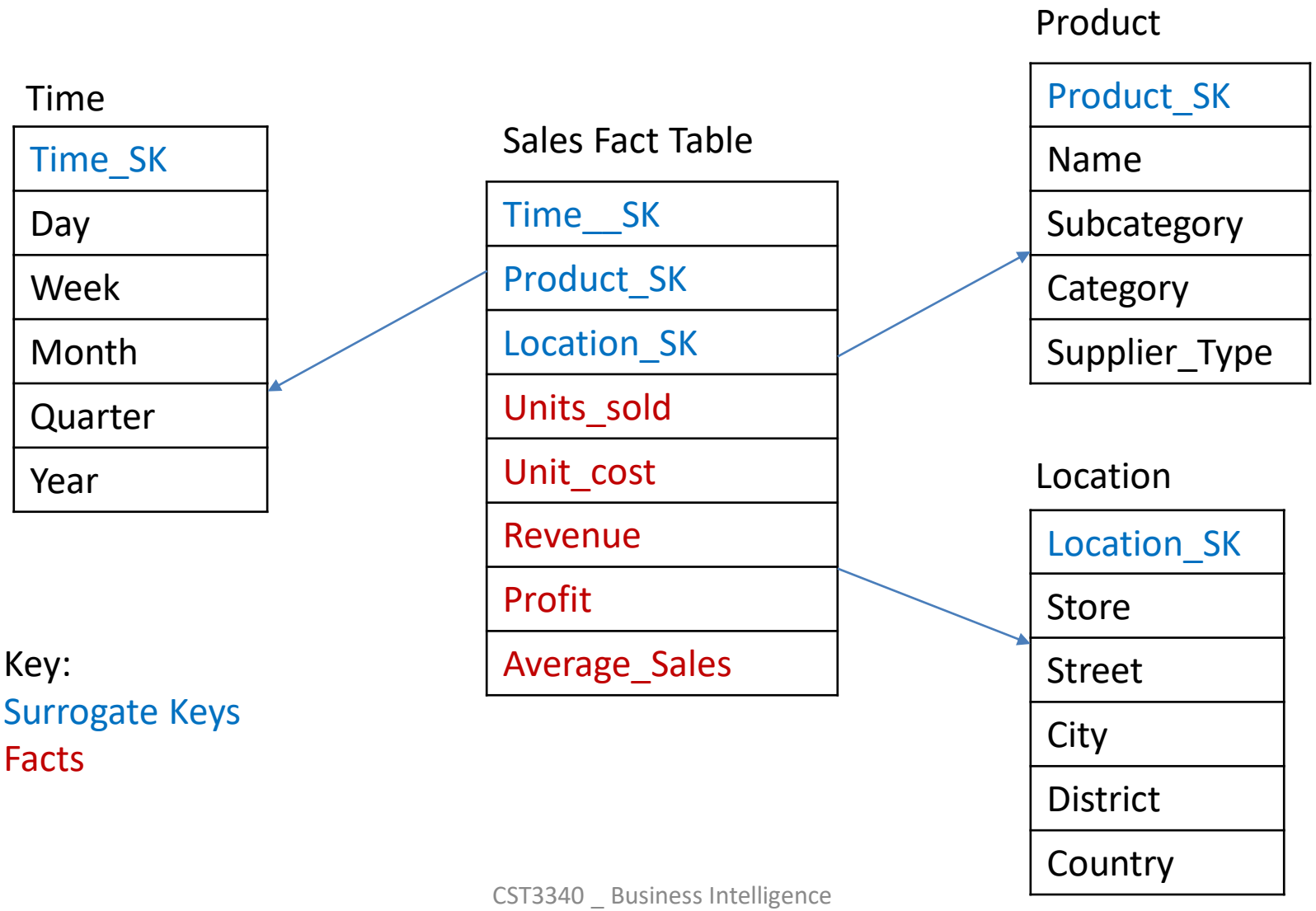
Multidimensional Data Modeling and OLAP

Slides adapted from Jiawei Han, Micheline Kamber, Jian Pei, (2011), Data Mining: Concepts and Techniques, Third Edition, The Morgan Kaufmann Series in Data Management System.

From Tables and Spreadsheets to Data Cubes

- The multidimensional data model allows a **data warehouse** to be viewed 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 product (produt_name, category, sub_category), or time(day, week, month, quarter, year)..
  - **Fact table** contains **measures** (such as units_sold, Revenue, Profit, Cost) and foreign keys to each of the related dimension tables.

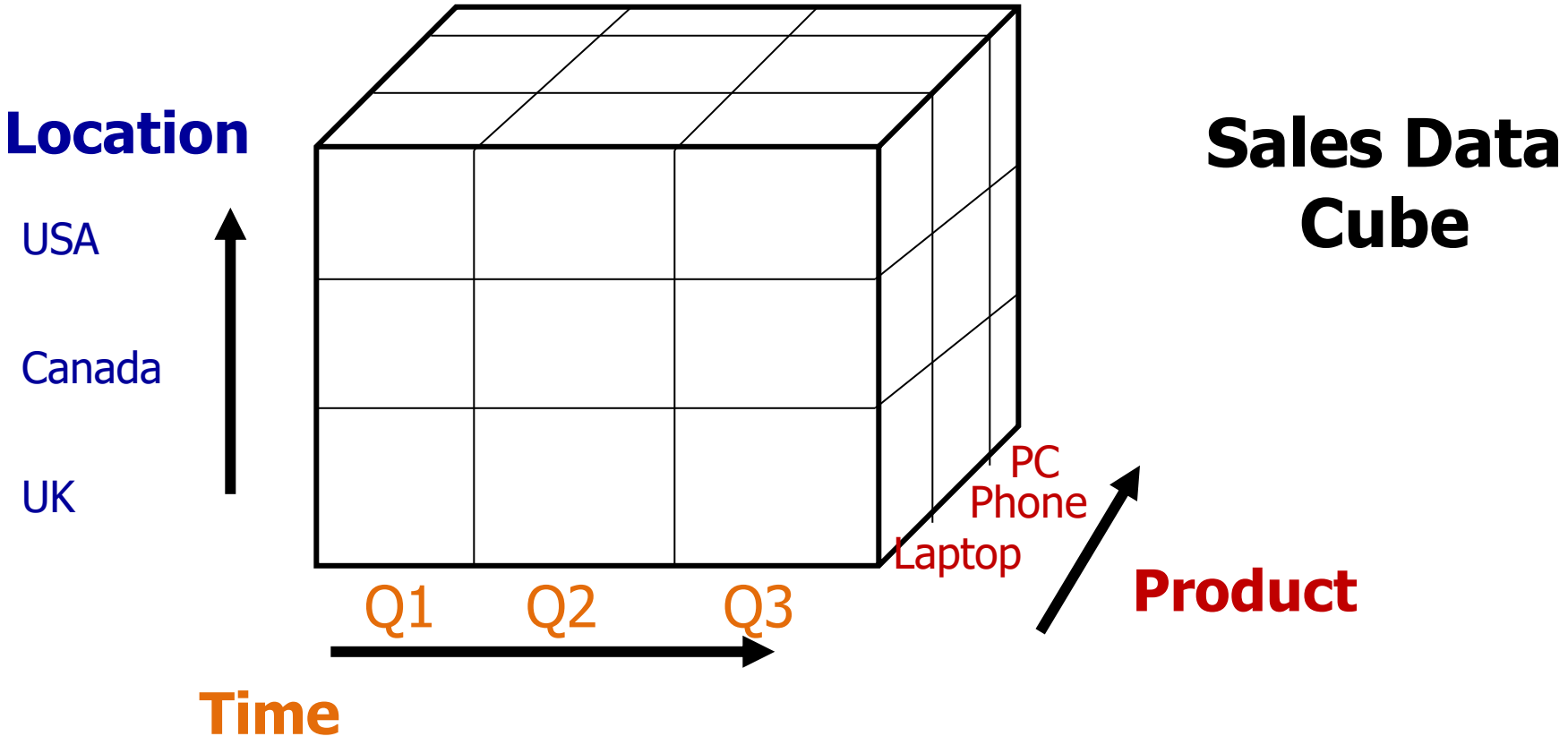
Example of a Star Schema



Data Cube 1

- A star schema with 3 dimensions can be represented by a cube.
- Each side of the cube represents a dimension.
- The cells of the cube store the measures.
- An n-dimensional star schema can be represented by an n-sides shape but cannot be visualised.

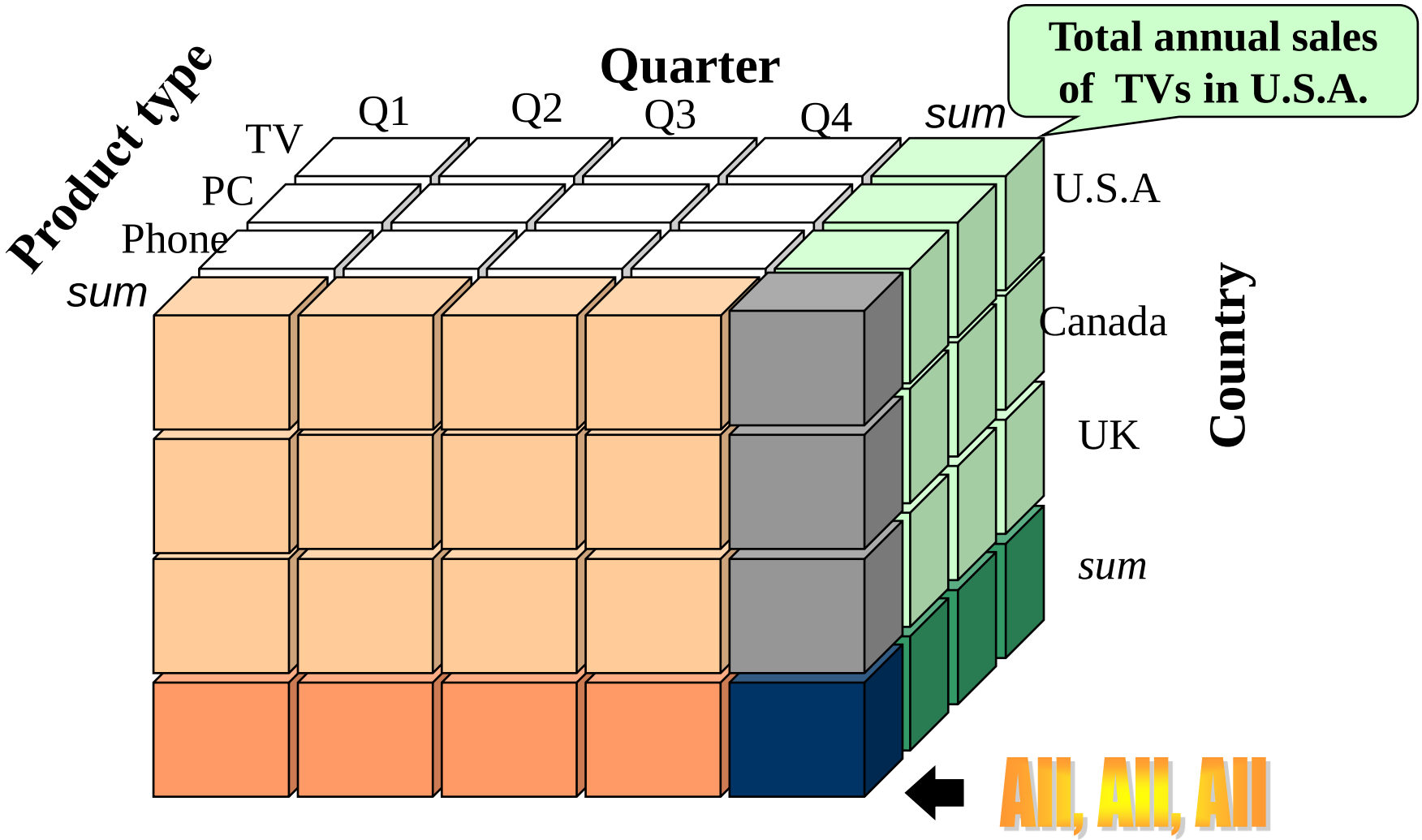
Visualisation of Multi-dimensional Modelling



Data Cube 2

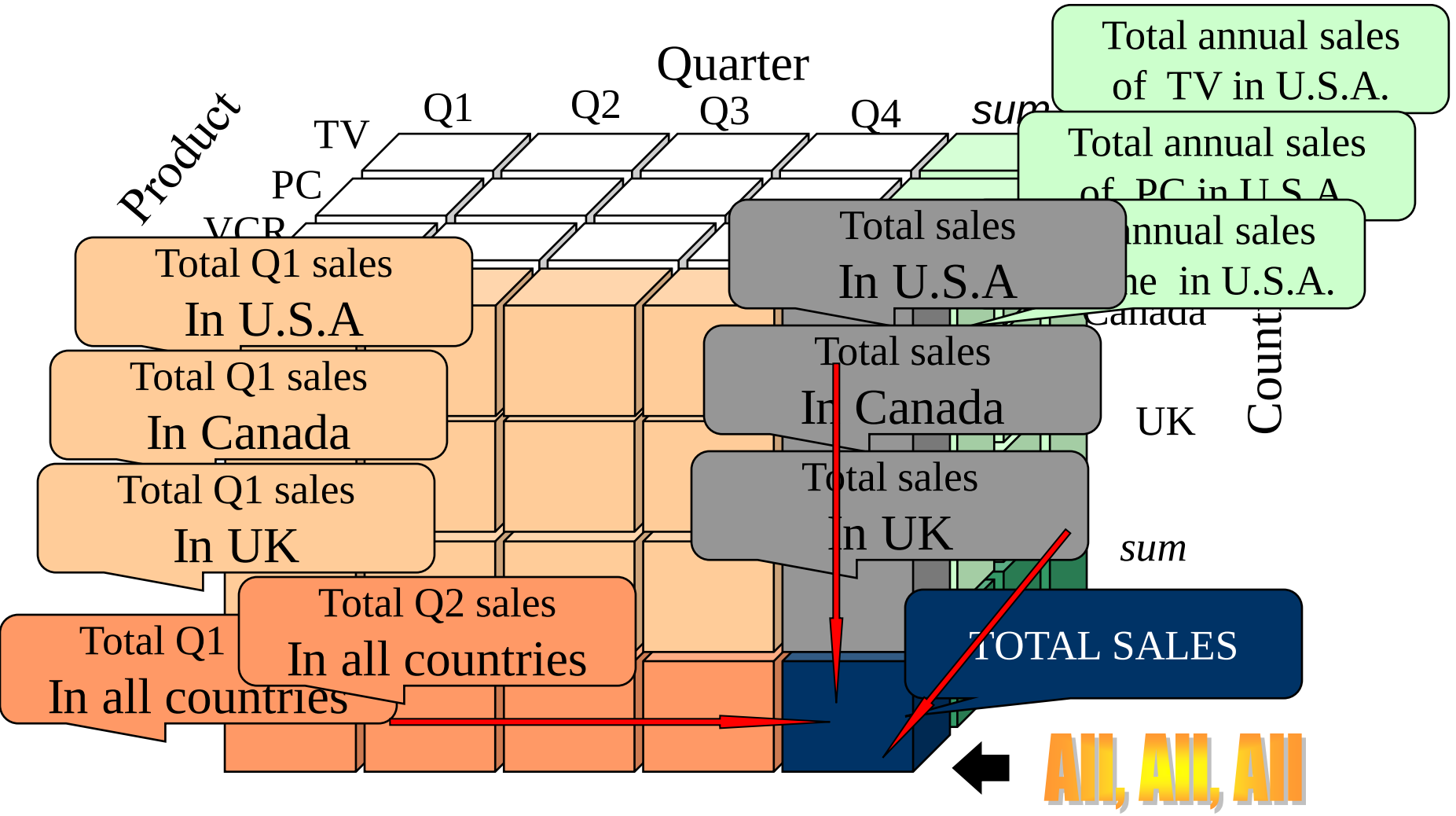
- Different types of data stored in the cube:
  - Data values such as unit sales, profit, costs.
  - Aggregated values such as sum, average etc.
  - Column and row totals such as
    - Sum of sales in Q1.
    - Total sales in Canada

A Sample Data Cube



Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

SAMPLE CUBE with aggregates (totals)



Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

What is OLAP?

- OLAP (On-Line Analytical Processing) is used to process data in a multidimensional cube.
- Data are stored in data cubes, which are defined over multiple dimensions.
- Used to discover trends and patterns in the data using
  - Business Intelligence Report
  - Complex calculations
  - Forecasting – What if analysis.
  - Ad-hoc queries
- Used to power many Business applications such as
  - Business Performance Management
  - Knowledge Discovery
  - Simulation Models

OLAP Operations 1

- Roll up (drill-up): Data is summarised over dimensions – aggregated data
  - *Moving up the hierarchy of a dimension.*
  - *Reducing the number dimensions.*
- Drill down (roll down): Opposite of roll-up – more detailed data
  - *Moving down the hierarchy of a dimension*
  - *Increasing the number of dimensions.*

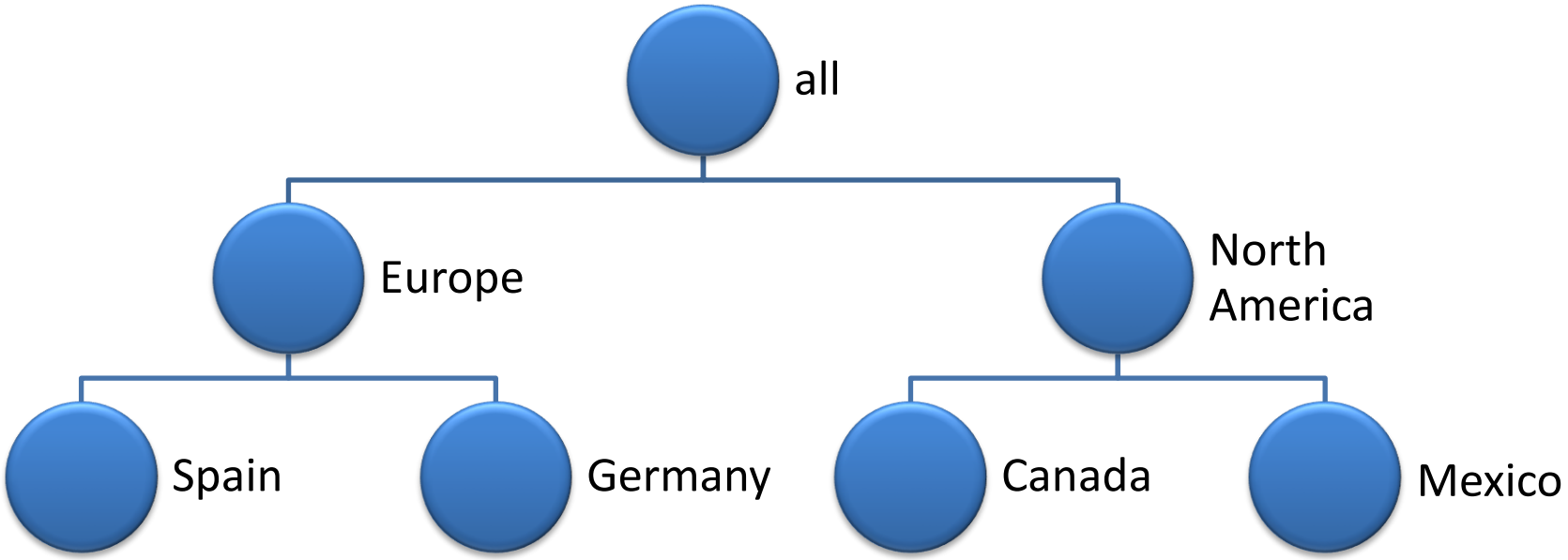
OLAP Operations 2

- Slice:
  - *Using a slice of cube,*
  - *Reducing the number of dimensions*
- Dice:
  - *Using a sub-cube*
  - *Reducing the number of values considered on each dimension*

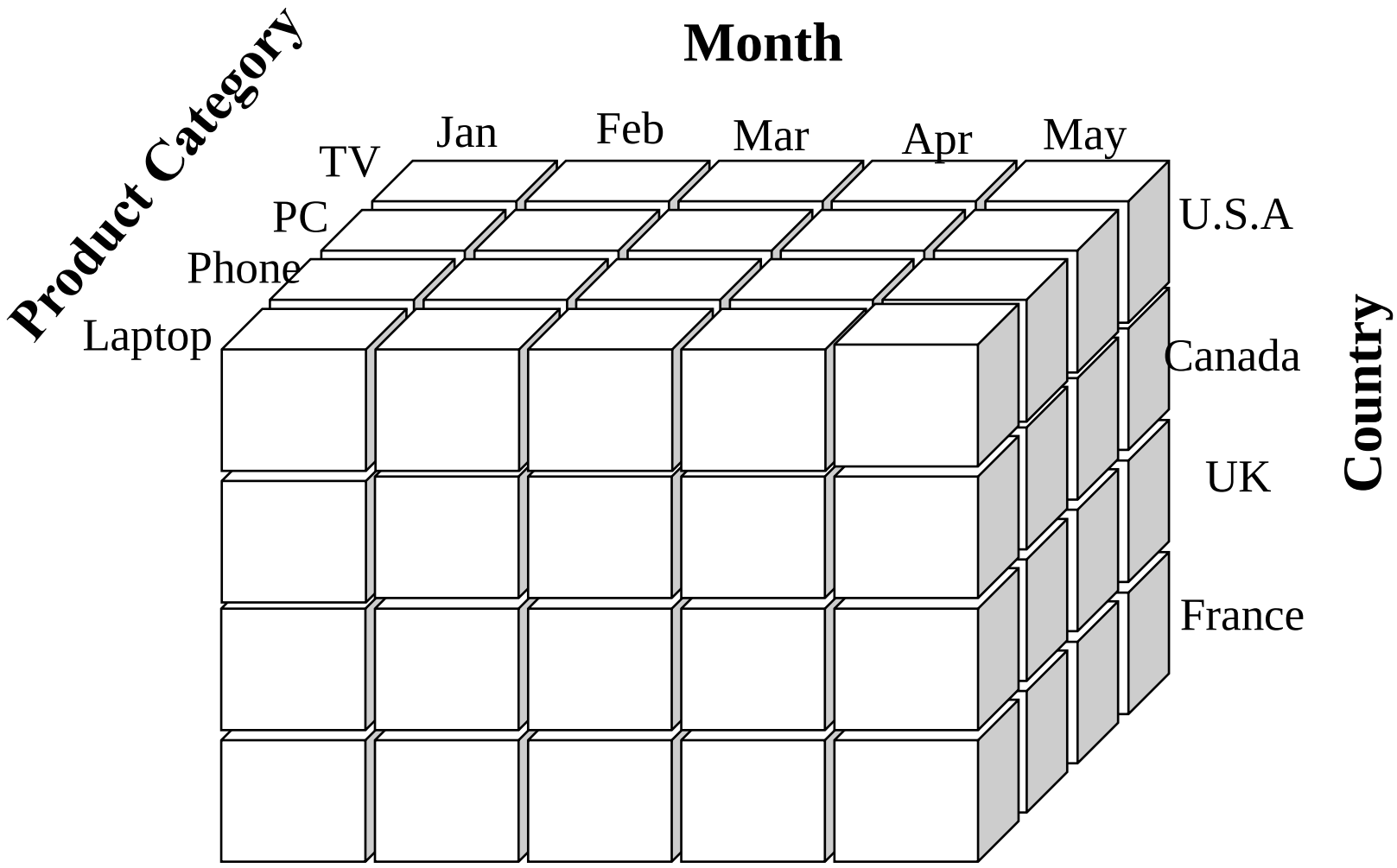
OLAP Operations 3

- Pivot (rotate):
  - *Considering different faces of the cube*
  - *Visualising the 3D cube as a series of 2D planes.*
- Other operations
  - *drill across: uses the fact constellation – several fact table.*
  - *drill through: uses SQL to access the data in the original relational tables.*

Location Dimension Hierarchy

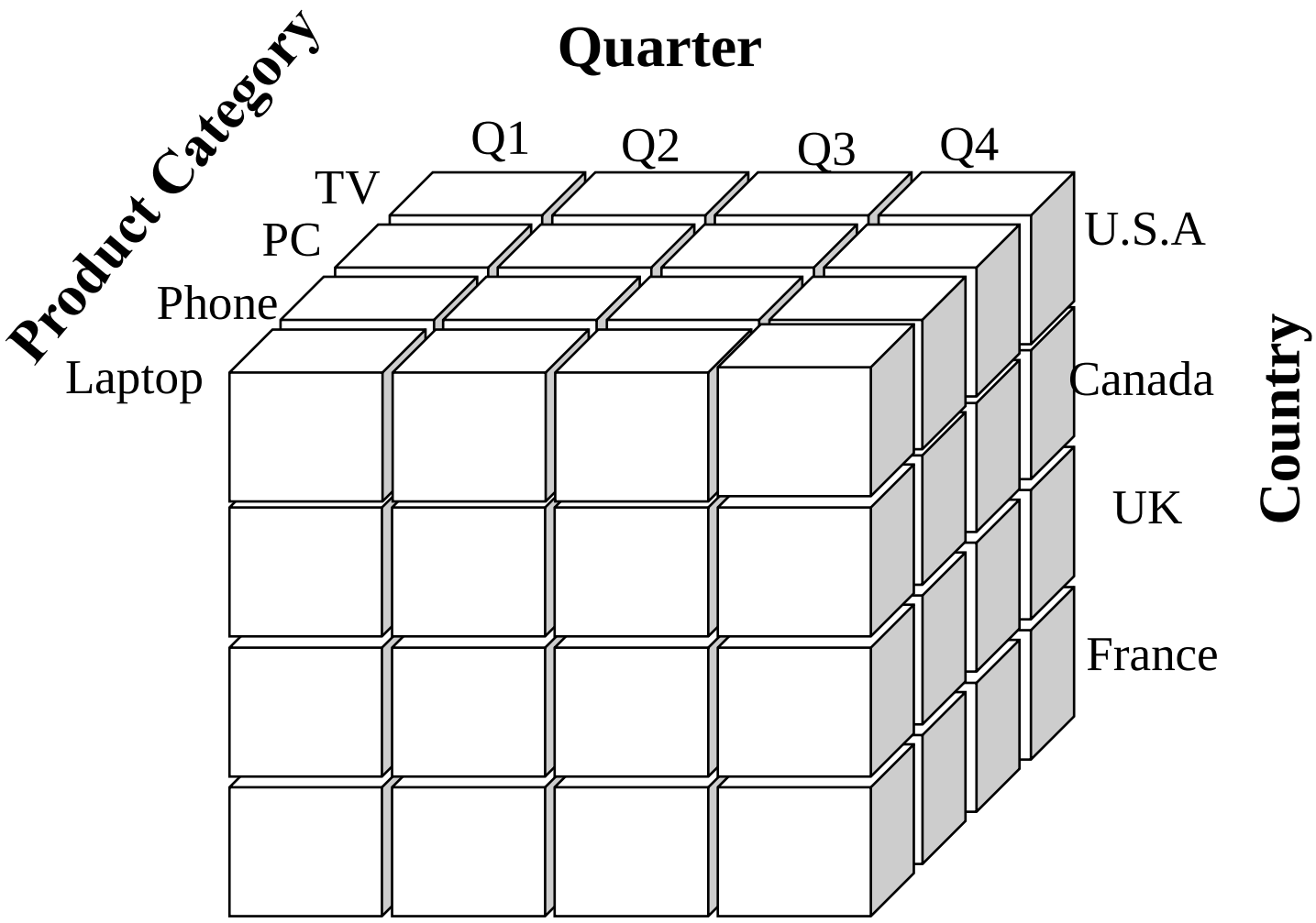


Original Data Cube



Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

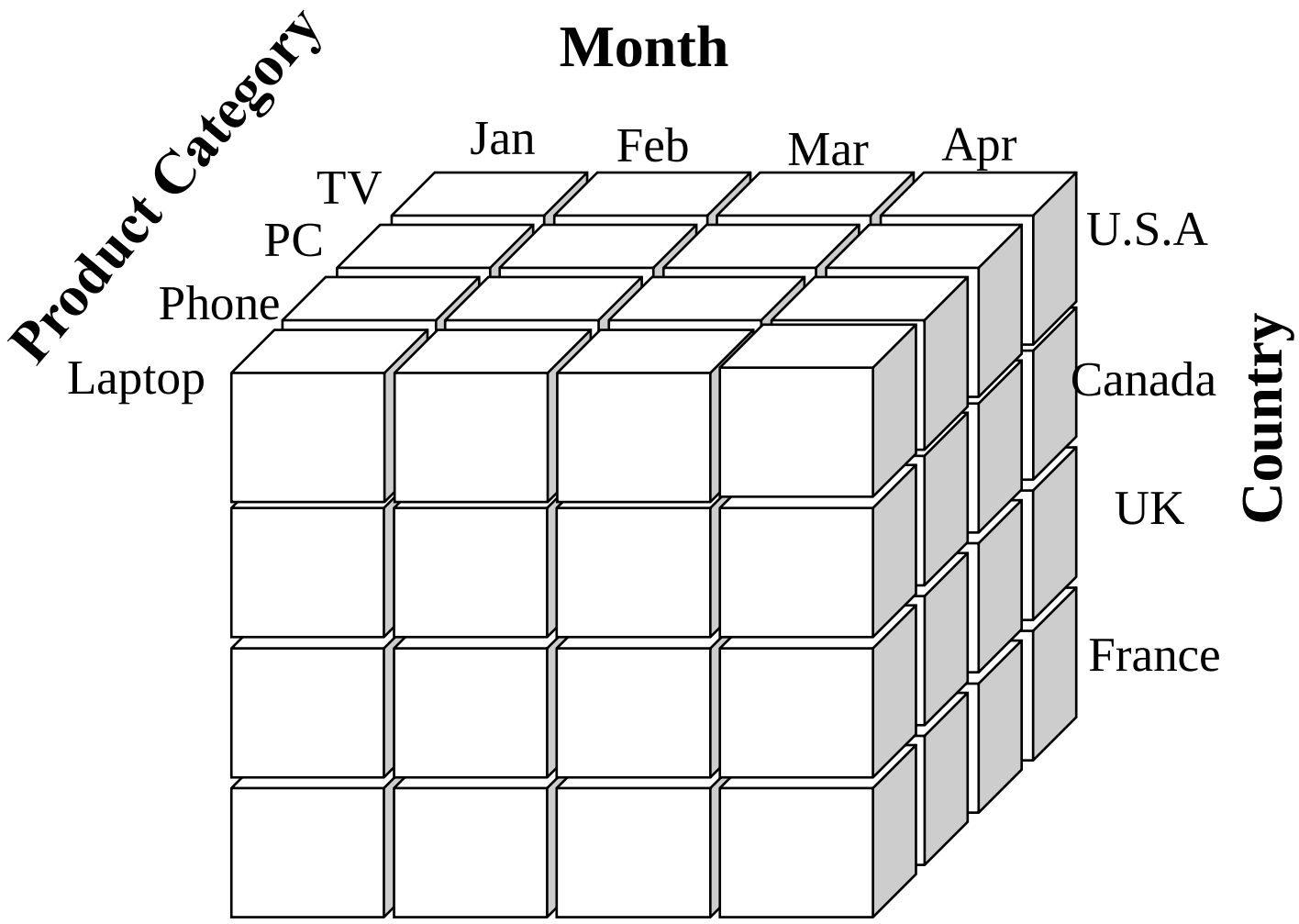
Example of Roll Up



The data on the Time dimension is rolled up from months to quarters. The data is aggregated

Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

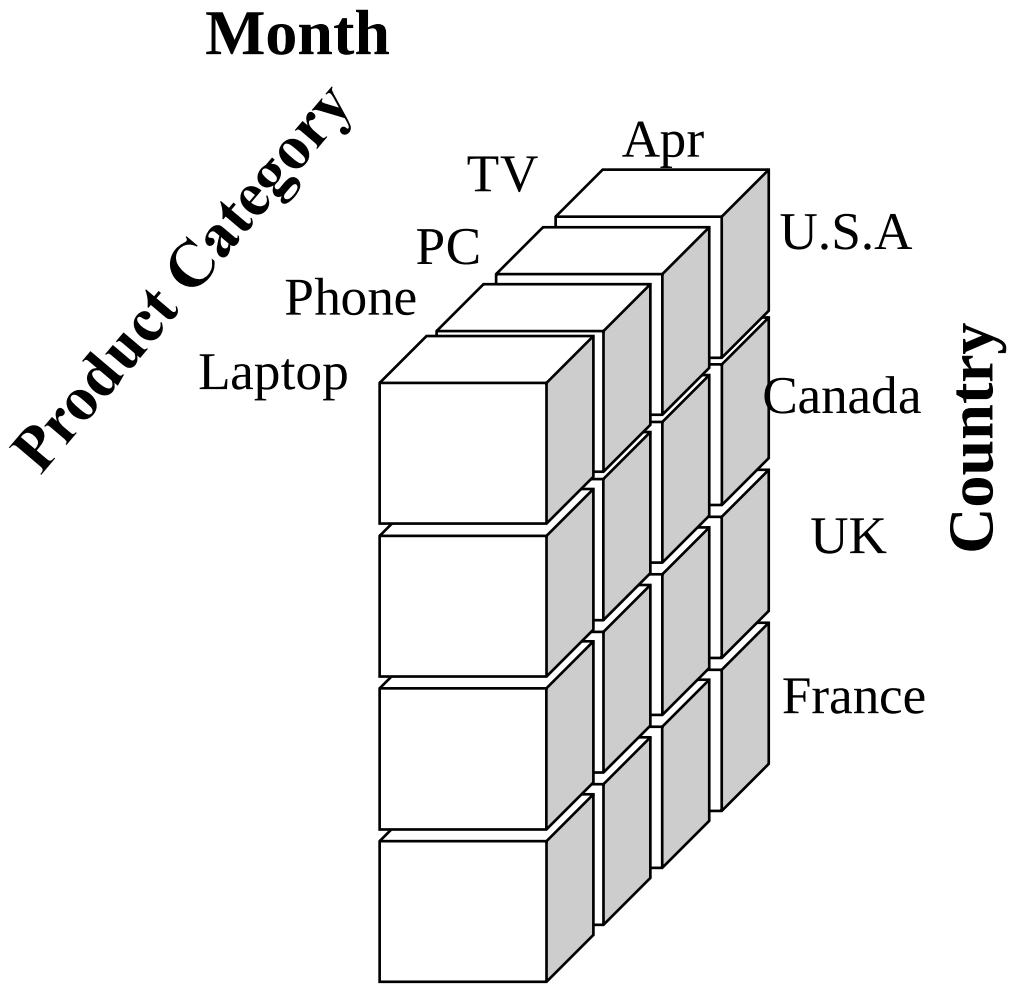
Example of Roll Down



The data on the Time dimension is rolled down from quarters to Months. The data is more detailed

Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

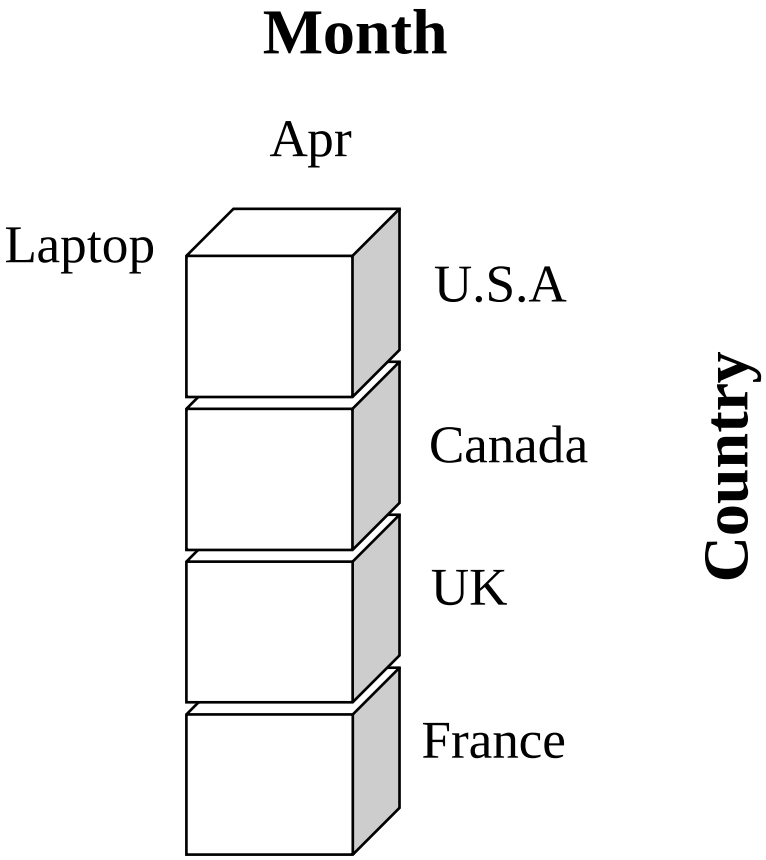
Example of Slice on 1 dimension



A slice of the cube is extracted. For example consider the sales in Apr. This produces a sub cube.

Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

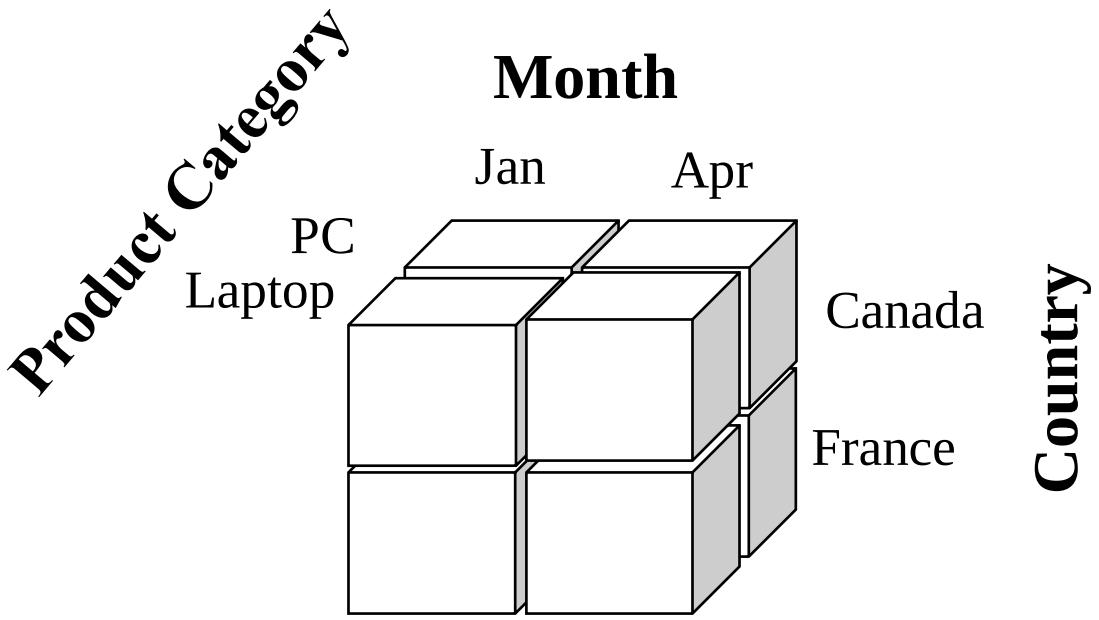
Example of Slice on 2 dimensions



A slice of the cube is extracted. For example consider the sales of laptops in Apr. This produces a sub cube.

Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

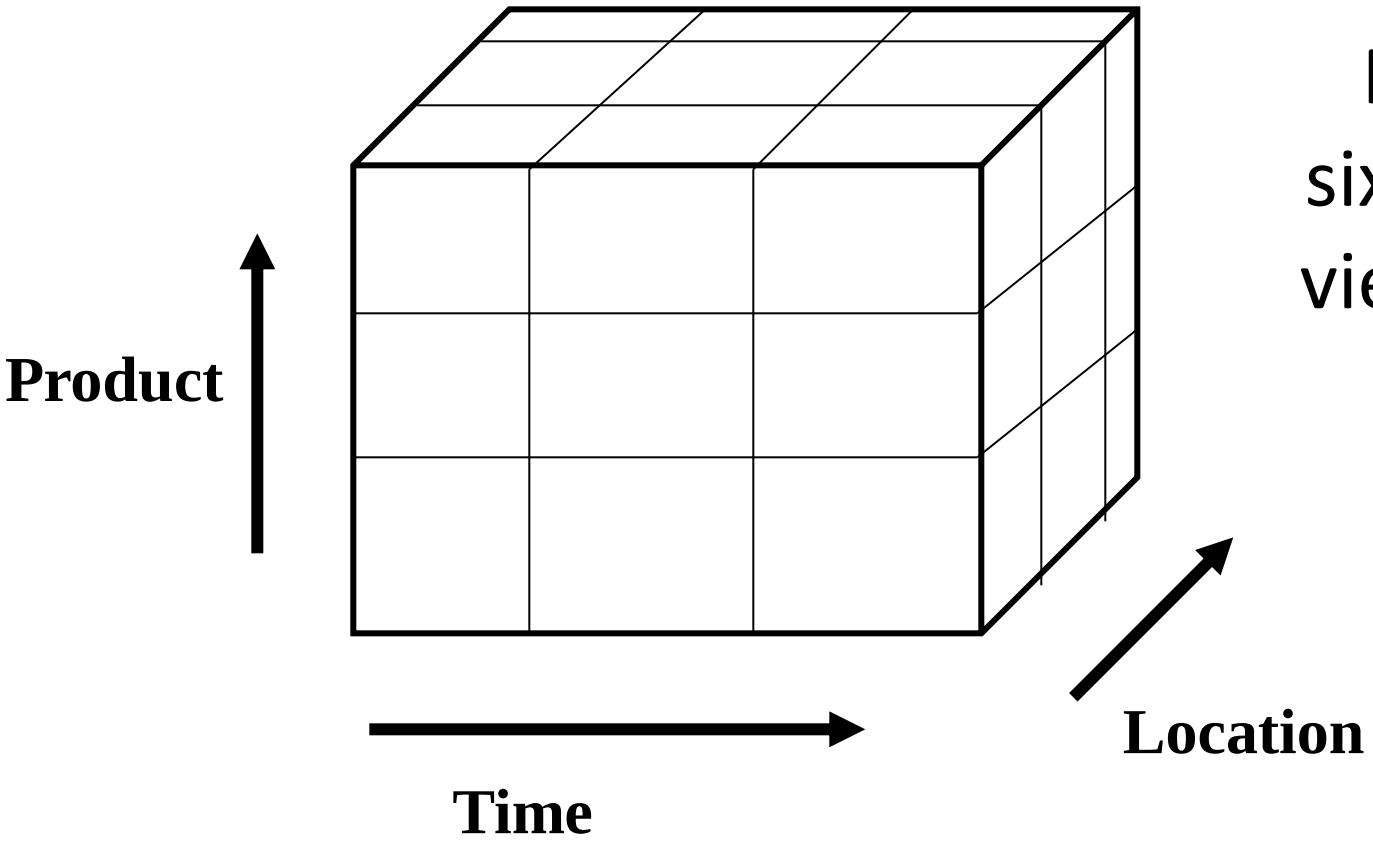
Example of Dice



A sub cube is extracted. For Example, consider the sale of Laptops and PCs in January and April in Canada and France

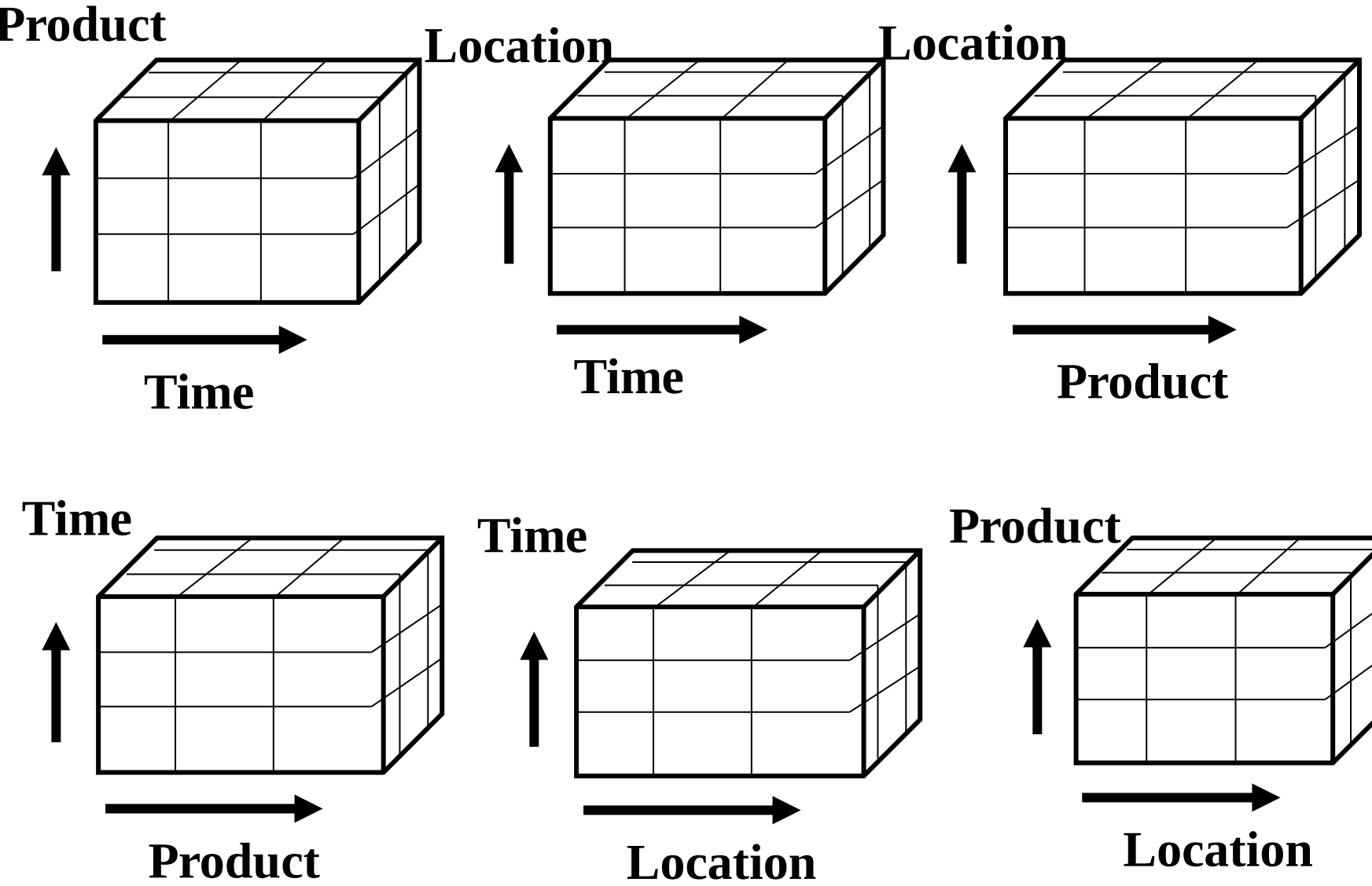
Source:Jiawei Han, Micheline Kamber, Jian Pei (2011)

Example of Rotation



Rotation six different views - Data slicing

Data Rotation/Slicing – Different views of the data



Different way data is stored

- The data cube is a good way to visualise the data in an OLAP environment.
- However, Data is not stored on disc as a data cube
- There are three different logical data models
  - ROLAP: Relational Online Analytical Processing
  - MOLAP: Multi-dimensional Online Analytical Processing.
  - HOLAP: Hybrid Online Analytical Processing.

ROLAP: Relational Online Analytical Processing 1

- Data is stored and manipulate by a relational DBMS.
- Gives the appearance of traditional OLAP functionality.
- Advantages:
  - Handles large amounts of data.
  - Has access to relational database functionality.
  - Easy to modify and maintenance is low.
  - Uses star and snowflake schemas.
  - Proved technology (DBMS and relational Model).
  - Can out perform Multi Dimensional Database for large data sets.

ROLAP: Relational Online Analytical Processing 2

- Disadvantages:
  - Performance can be slow
    - ROLAP can require multiple complex SQL queries which can be slow for large data sets.
  - Limited by functionality of SQL
    - SQL does not meet all the requirement of ROLAP, such as complex calculations.
    - Many vendors mitigate this risk by building complex functions or allowing user defined functions.

MOLAP: Multi-dimensional Online Analytical Processing 1.

- Data (Facts) stored in multidimensional arrays, separate from the Data warehouse.
- Known as a Multidimensional Database
- Users have direct access to the arrays.
- Dimensions used as index for arrays.

MOLAP: Multi-dimensional Online Analytical Processing 2.

- Advantages:
  - Excellent performance: MOLAP cubes are designed for fast and efficient retrieval.
  - Optimal efficiency for slicing and dicing operations.
  - Performance of complex calculations is optimised as many calculations are generated as the cube is formed.

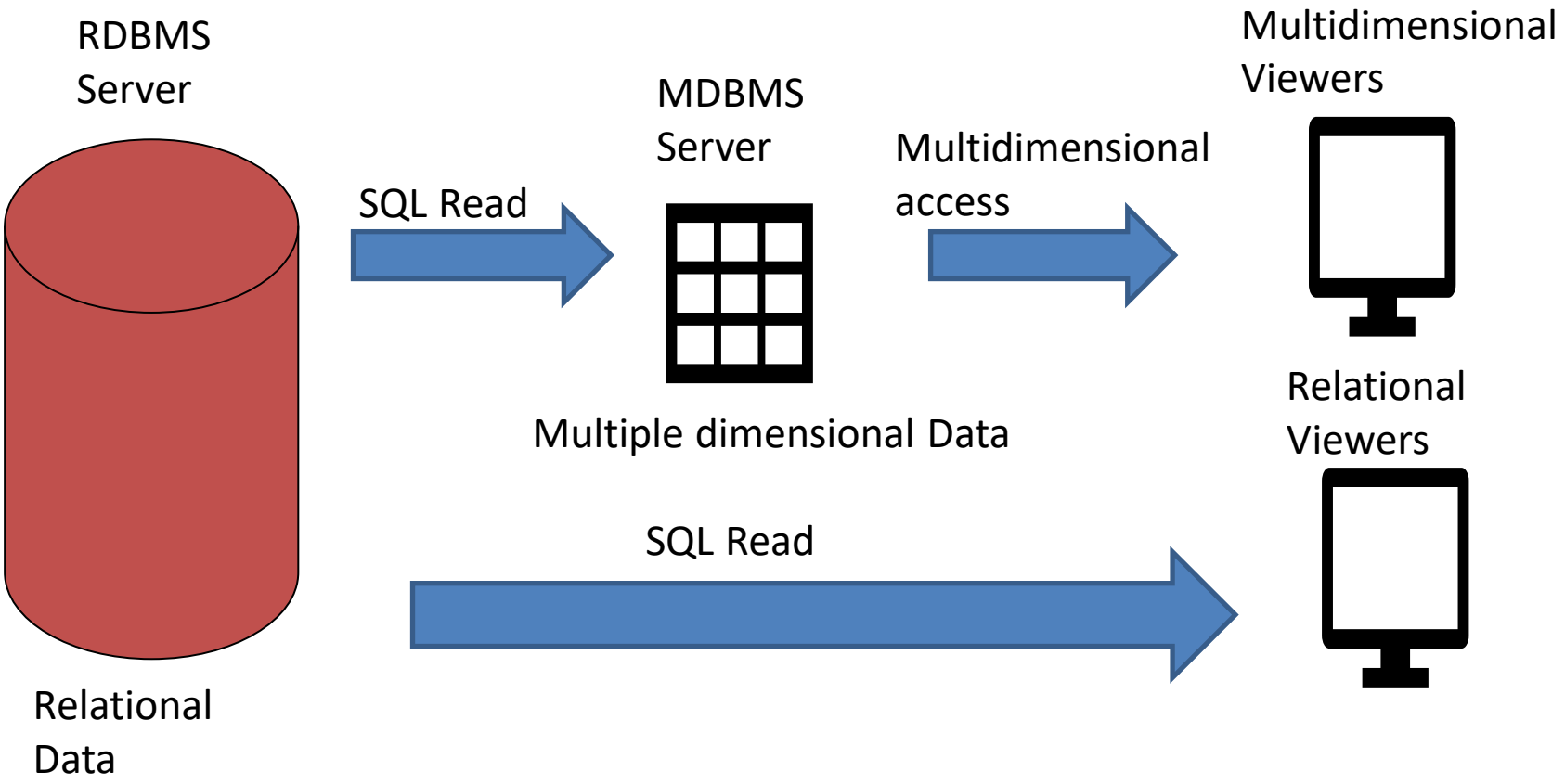
MOLAP: Multi-dimensional Online Analytical Processing 3.

- Disadvantages:
  - Limited data as calculations are performed as the cube is created, not all data can be included.
  - Requires additional investment in human and capital resources.
    - Extra technology required to store and manipulate cubes.
    - Extra technicians required to support the MDDB

HOLAP: Hybrid Online Analytical Processing 1.

- ROALP and MOLAP combined
- Detailed data stored on RDBMS
- Aggregated data stored in MDBMS
- User access MDBMS via MOLAP tools.

HOLAP: Hybrid Online Analytical Processing 2.



ROLAP vs MOLAP vs HOLAP

	ROLAP	MOLAP	HOLAP
Storage Method	Relational DB	Multidimensional DB	Both
Data Arrangement & schemas	Star Schemas. Tables with rows and columns	Data Cubes	Multidimensional
Volume	Enormous amount of data processed	Limited data can be processed	Large amounts of data processed
Technique	SQL	Sparce Matrix technology	SQL and Sparse Matrix Technology
Design View	Dynamic	Static	Dynamic

Reading

- Chapter 4, Section 4.2:
- Jiawei Han, Micheline Kamber, Jian Pei, (2011), Data Mining: Concepts and Techniques, Third Edition, The Morgan Kaufmann Series in Data Management System.

- Chapter 3, Section 3.6:
- Sharda, Delen, Turban (2018), Business Intelligence Analytics and Data Science: A management Perspective. Pearson.