University of South-Eastern Norway

Lecture 3: Multidimensional Data Representation and Manipulation

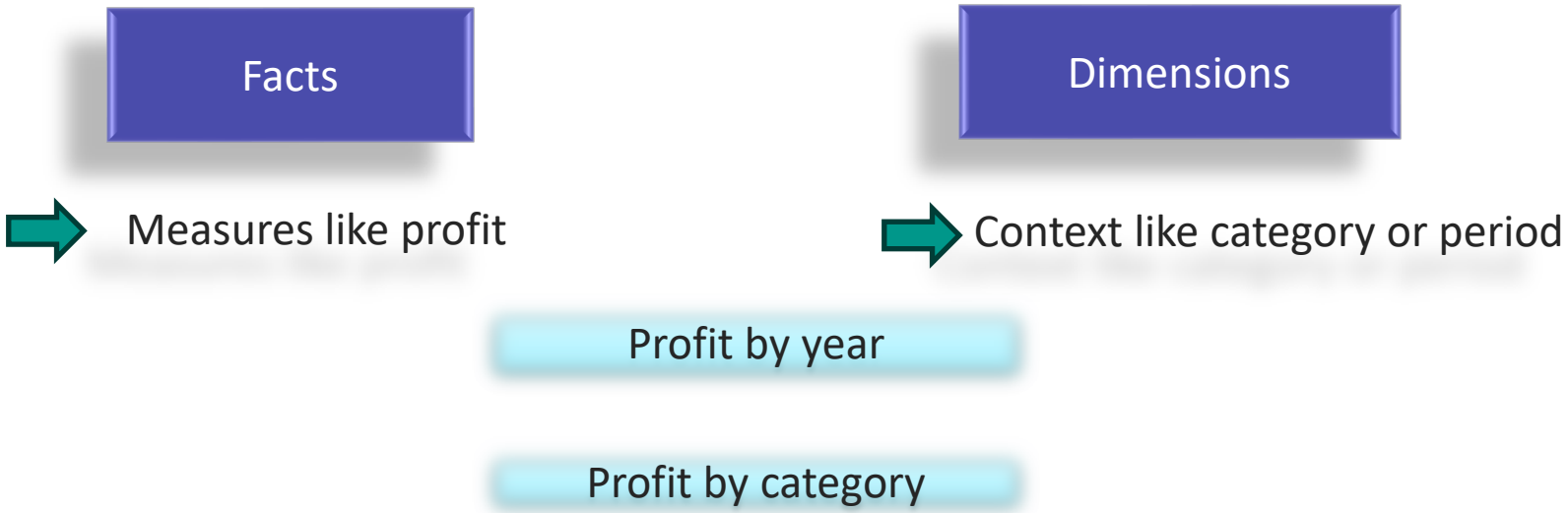
Veralia Sánchez
Associate Professor

Outline

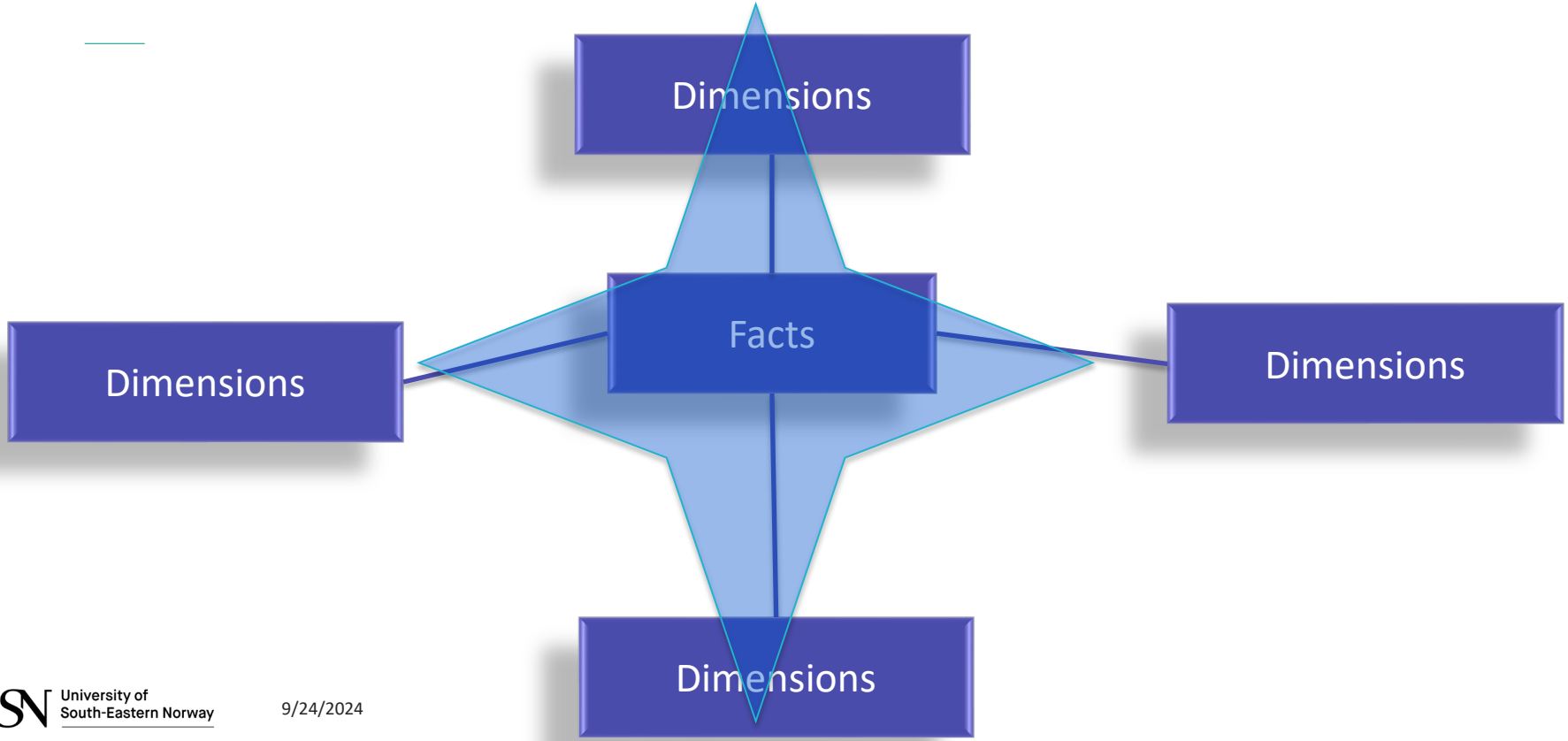
- Dimensional modelling
- Data Cube Concepts
- Data Cube Operators

Dimensional modelling

- Methods of organising data (in a data warehouse)



Dimensional modelling



Dimensional modelling

- Unique technique of structuring data
- Commonly used in DWH
- Optimized for faster data retrieval
- Oriented around performance and usability
- Designed for reporting /OLAP

Facts

Facts

- ✓ Foundation of DWH
- ✓ Key mearuments
- ✓ Aggregated and analzed
- ✓ Fact table: PK, FK, Facts

Dimensions

- ✓ Categorizes facts
- ✓ Supportive and descriptive
- ✓ Filtering, grouping and labeling
- ✓ Non aggregatable
- ✓ (More) static

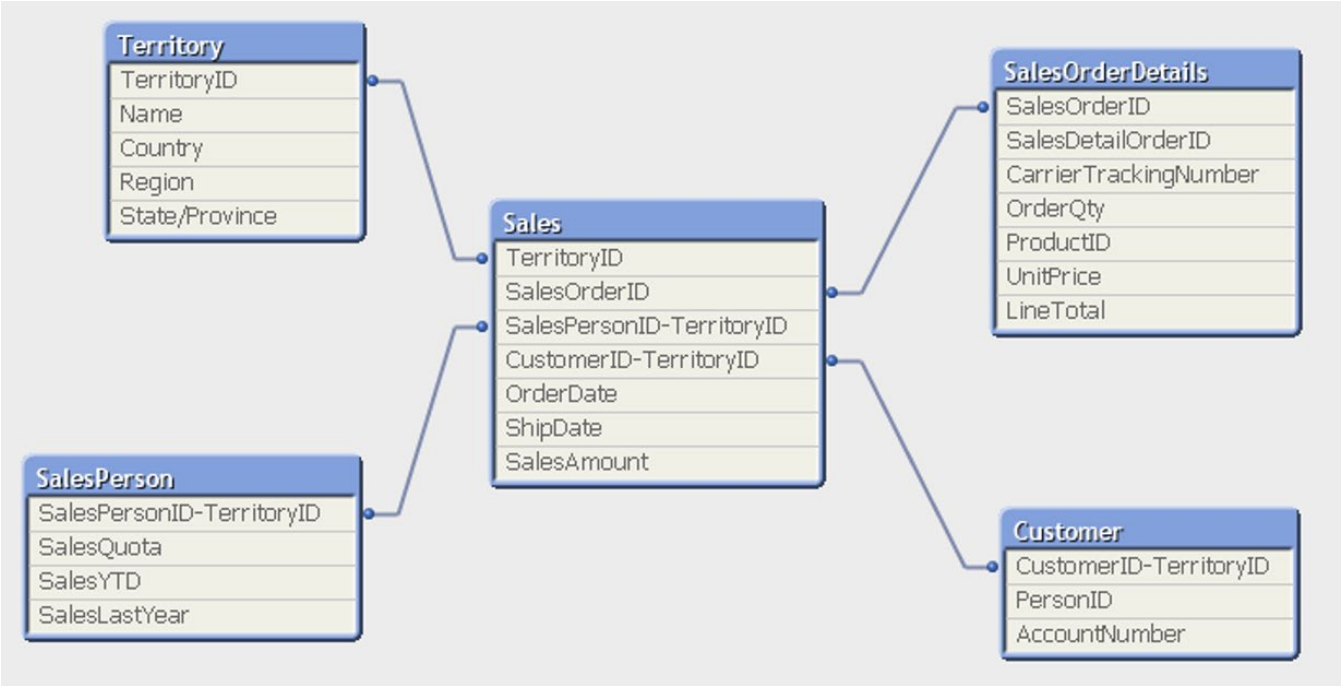
In-class tutorial

- Product and category dimensions in PostgreSQL.

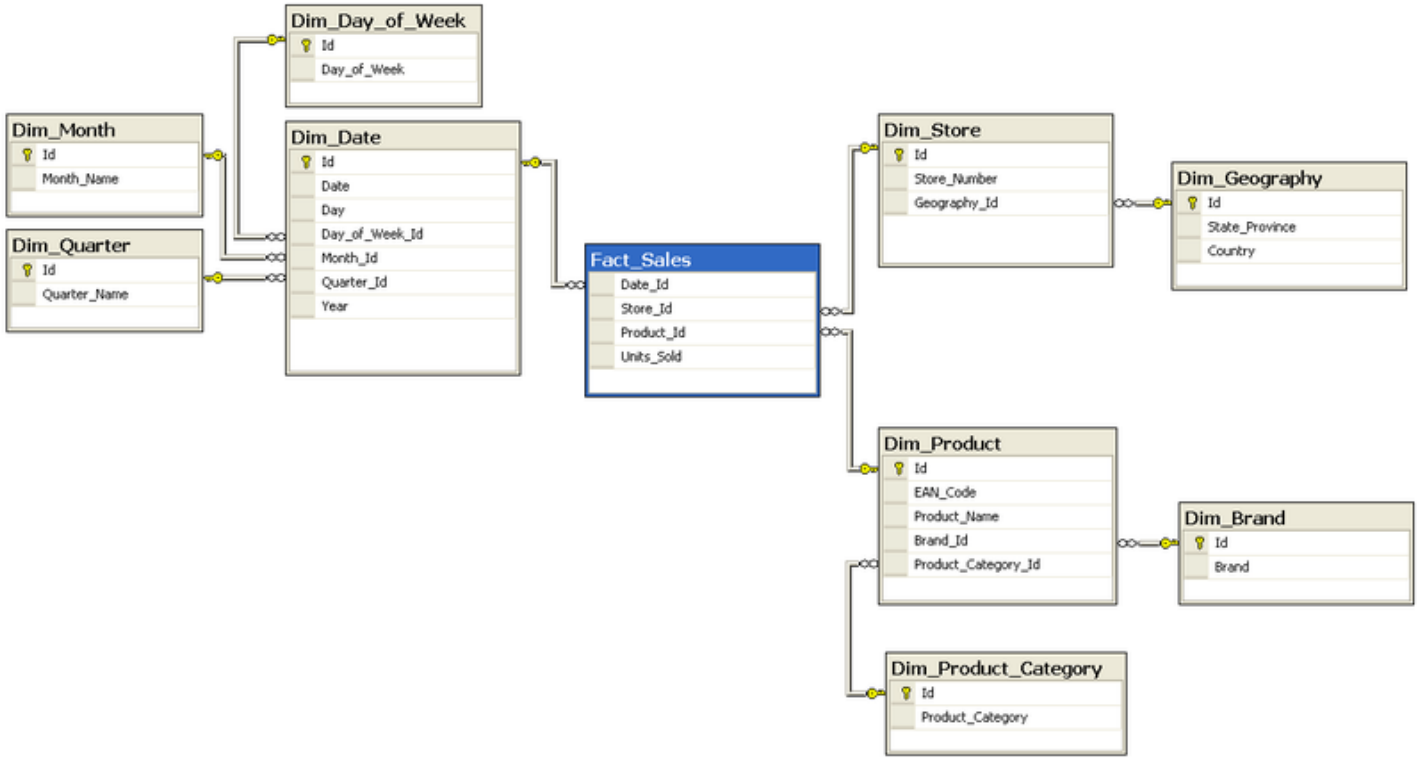
Representation of Data in DW

- Dimensional Modeling
  - A retrieval-based system that supports high-volume query access
- Star schema
  - The most commonly used and the simplest style of dimensional modeling
  - Contain a fact table surrounded by and connected to several dimension tables
- Snowflakes schema
  - An extension of star schema where the diagram resembles a snowflake in shape

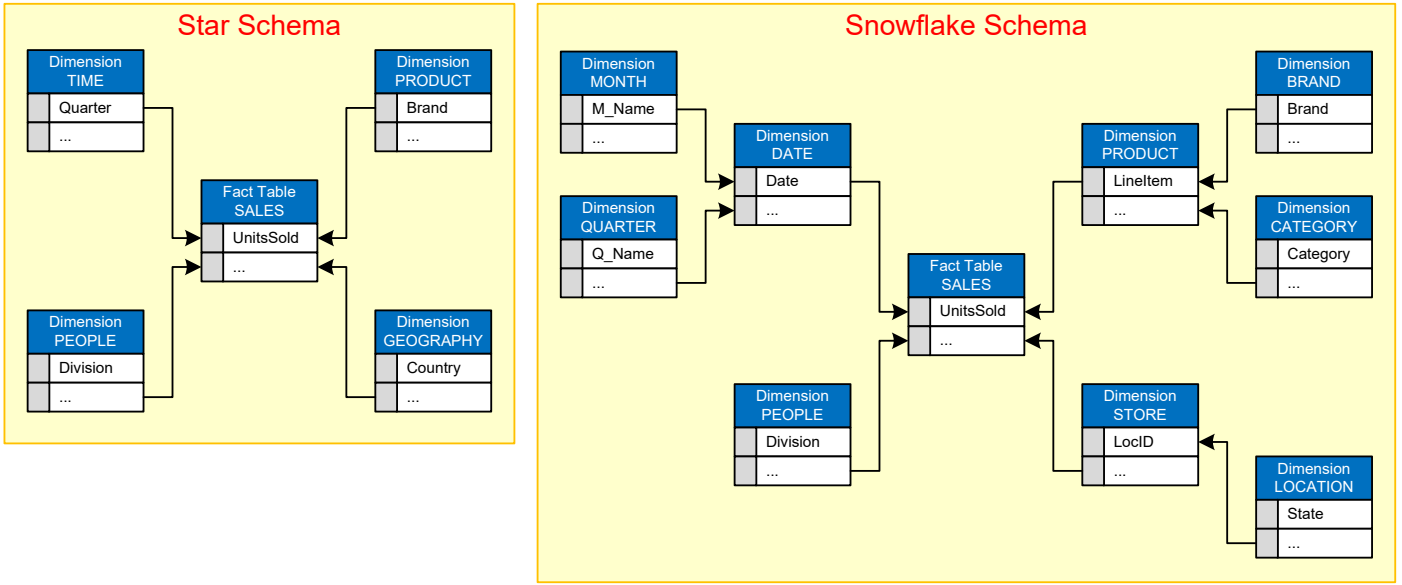
Example of a Star Schema:



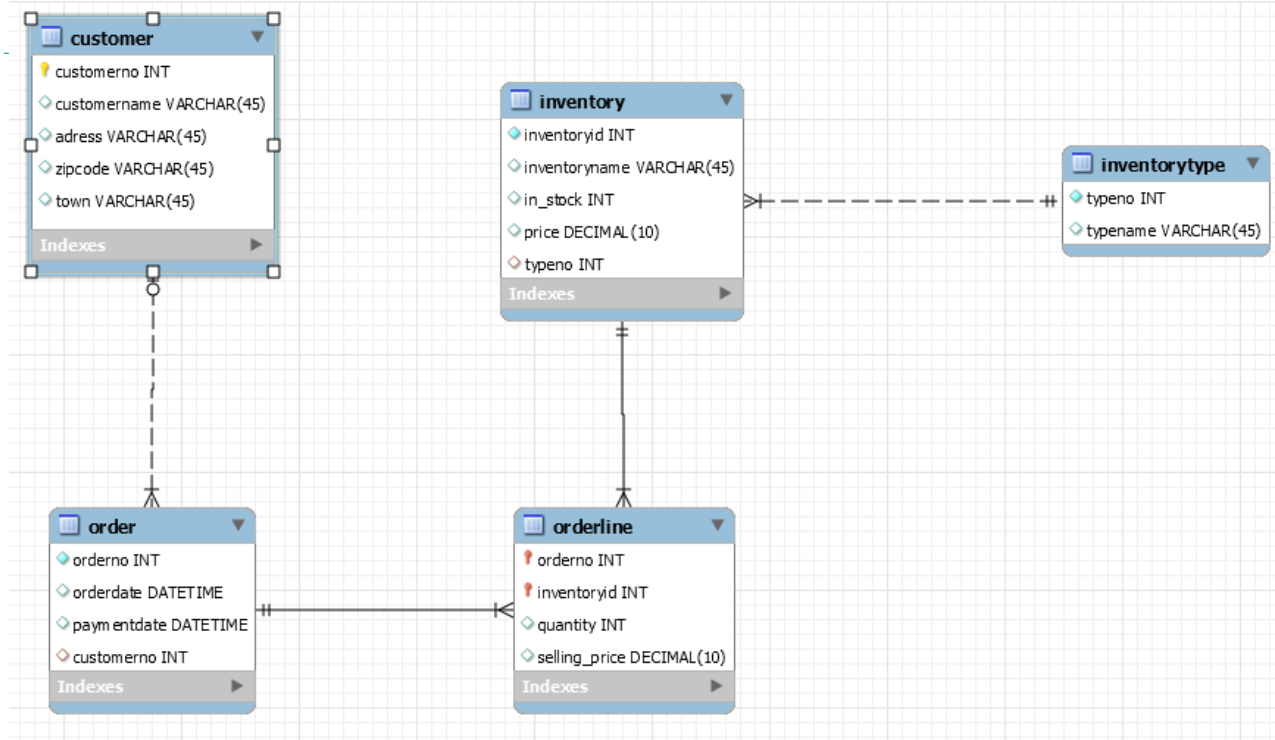
Example of a Snowflake Schema:



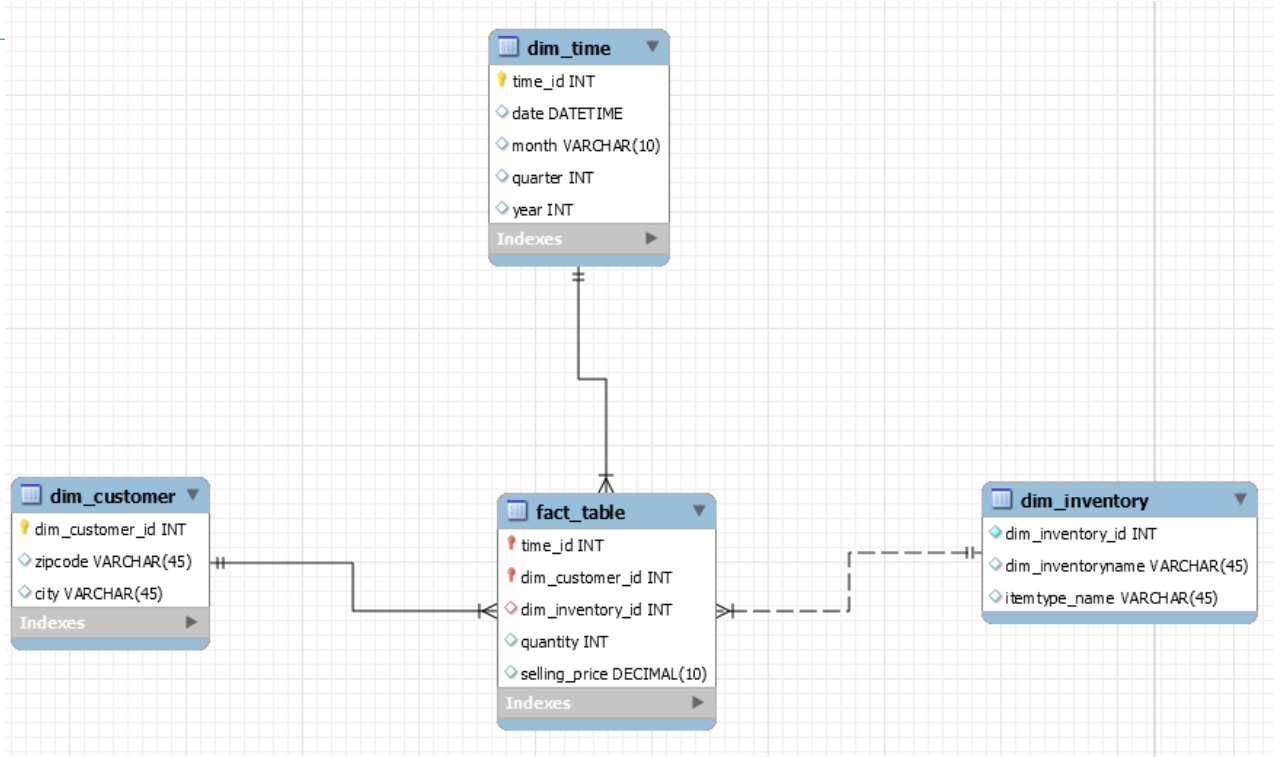
Star versus Snowflake Schema



Example of a Transaction System Database:



The Resulting Star Schema in the Data Warehouse:



Dimensional Modeling

- https://www.youtube.com/watch?v=lWPiSZf7-uQ

In-class assignment

- You are assigned to create a data warehouse as the responsible BI Consulatant.
- The source data is coming directly from the project management tool that is used in your company.
- The structure of the output table on the right

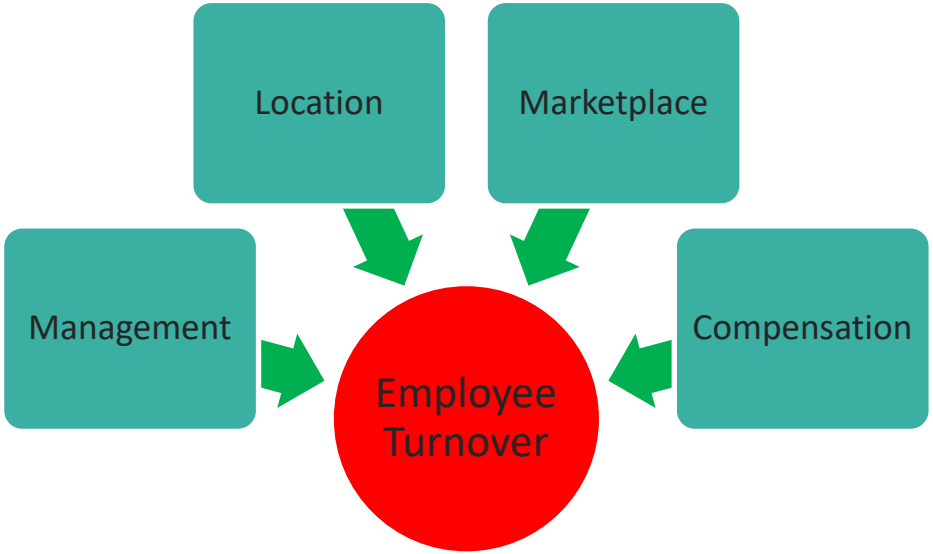
Requirements:

- The project managers and the division managers need to be able to analyze how many hours where logged by different attributes. They would also like to be able to analyze the hours by Month, Quarter and Year. You are free to create an additional table for that if necessary

ProjectLogs
Log_id
Update_Date
Hours_logged
Project_ID
Project_Name
Project_priority
Employee_ID
Employee_name
Division
Head_of_division

Data Cube Concepts

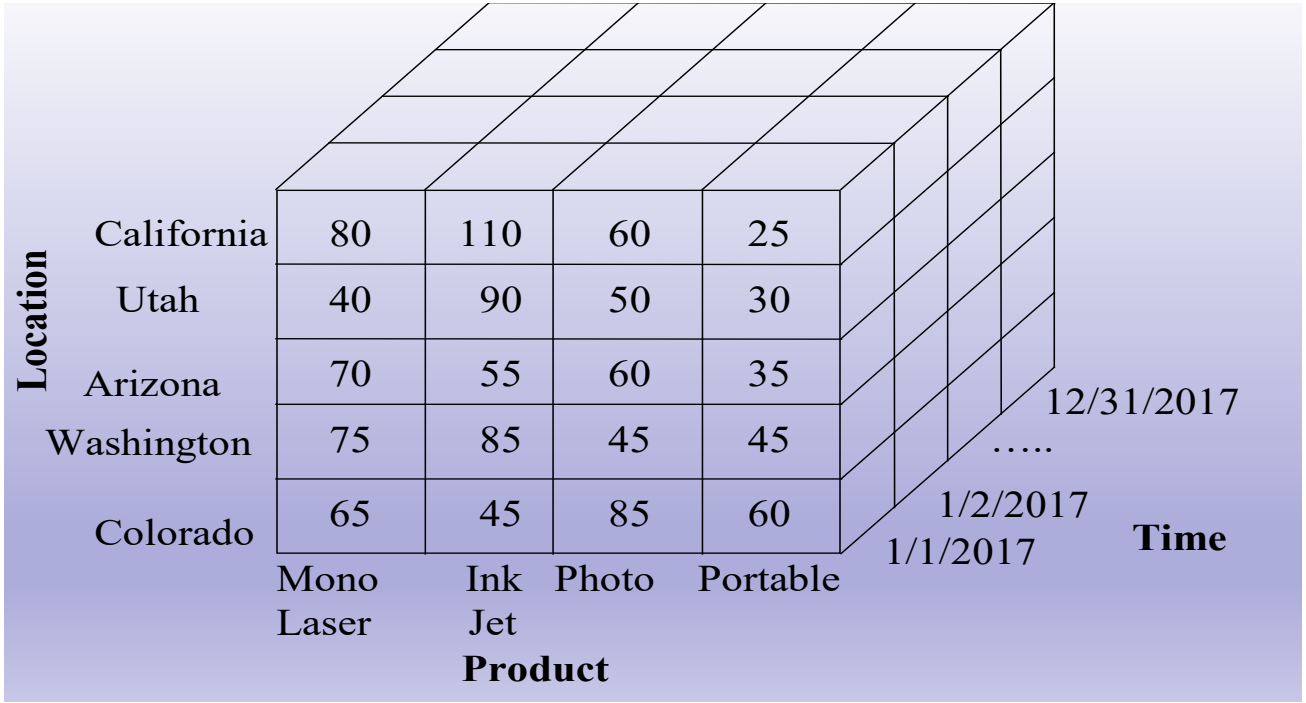
Business Analyst Perspective



Data Cube Basics

- Business analyst model
  - Factors or influencing variables of interest
  - Quantitative variables
  - Multidimensional arrangement
- Terminology
  - Dimension: subject label for a row or column
  - Member: value of dimension
  - Measure: quantitative variables stored in cells

Sales Data Cube Example



Notes on Dimensions and Measures

- Hierarchical dimensions with sub members
- Sparsity
  - Many cells do not have values
  - Increases with dimension detail and number of dimensions
- Measures
  - Derived measures
  - Multiple measures in cells

Measure Aggregation Properties

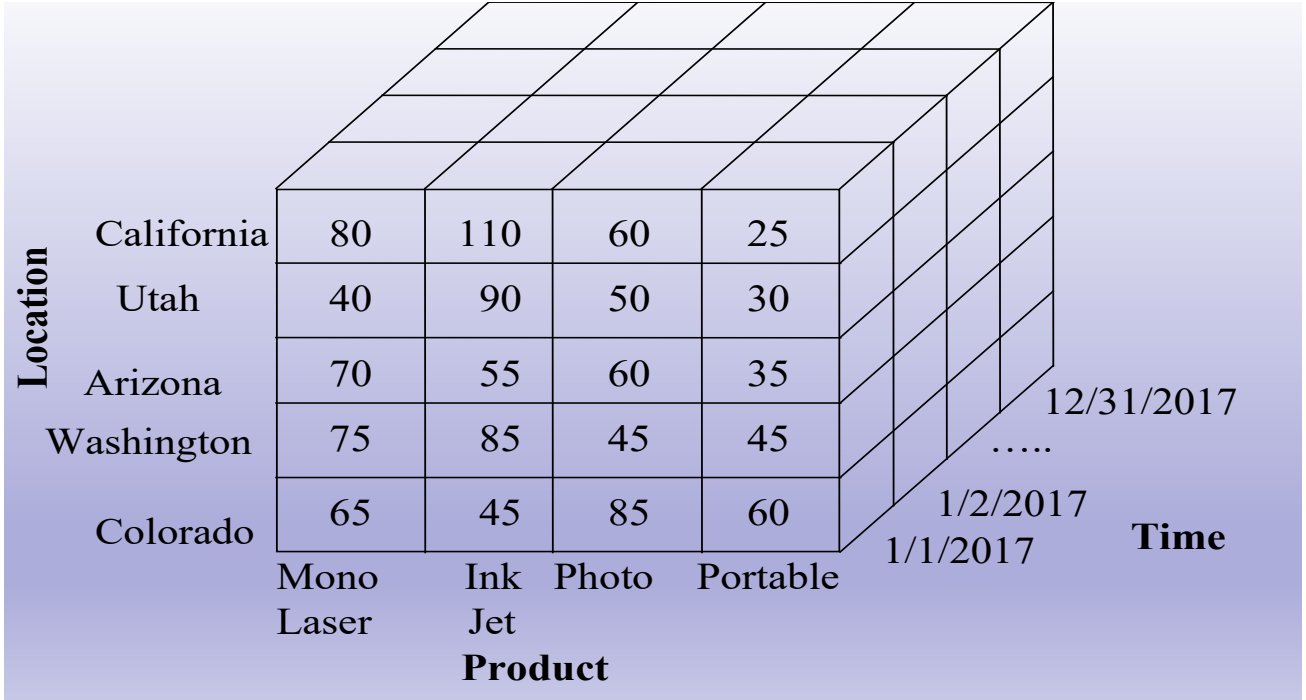
- Additive
  - Summarized by addition across all dimensions
  - Common measures such as sales, cost, and profit
- Semi-Additive
  - Summarized by addition in some but not all dimensions such as time
  - Periodic measurements such as account balances and inventory levels
- Non-Additive
  - Cannot be summarized by addition through any dimension
  - Historical facts such as unit price for a sale

Measure Aggregation Example

- Dimensions
  - Course: course id, degree, department, and college
  - Student: student id, major, department, and college
  - Time: semester, academic year, academic decade
- Measures:
  - Credit hours
  - Grade
  - Unit tuition
  - Tuition
- Aggregation properties for measures: ?

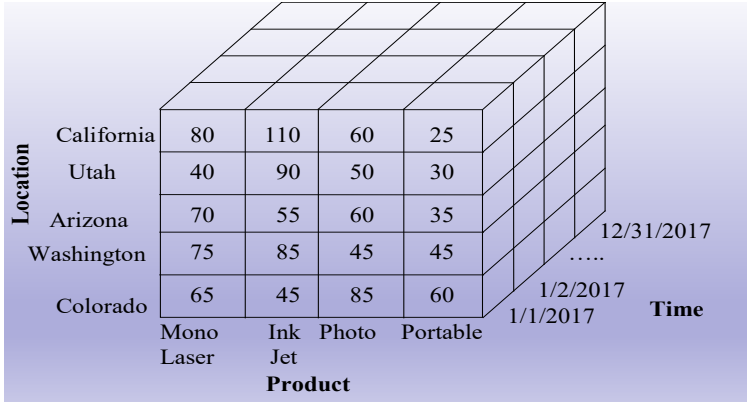
Data Cube Operators

Sales Data Cube Example



Slice Operator

- Subset of dimensions
- Set dimension to specific value

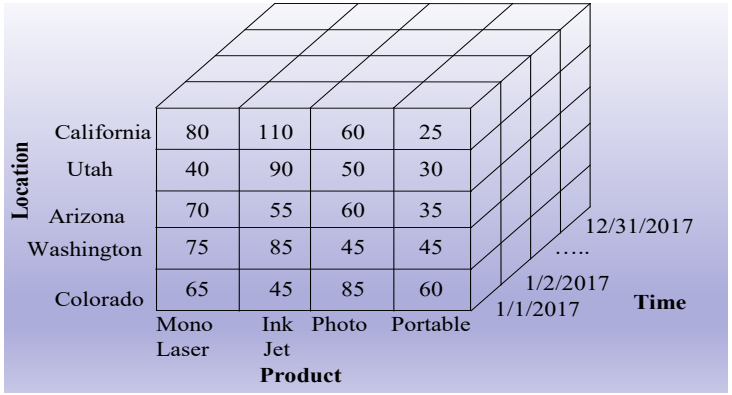


(Location × Product Slice for Time = 1/1/2017)

Location	Product			
	Mono Laser	*Ink Jet*	*Photo*	*Portable*
California	80	110	60	25
Utah	40	90	50	30
Arizona	70	55	60	35
Washington	75	85	45	45
Colorado	65	45	85	60

Slice Summarize Variation

- Replace a dimension with a summary of its values across all members



(Location × Time Slice SUM Product Sales)

Location	Time			
	1/1/2017	*1/2/2017*	*…*	*Total Sales*
California	275	670	…	16,250
Utah	210	190	…	11,107
Arizona	220	255	…	21,500
Washington	250	285	…	20,900
Colorado	255	245	…	21,336

Dice Operator

- Replace a dimension with a subset of values
- Dice operation often follows a slice operation

Location	Product			
	Mono Laser	*Ink Jet*	*Photo*	*Portable*
California	80	110	60	25
Utah	40	90	50	30
Arizona	70	55	60	35
Washington	75	85	45	45
Colorado	65	45	85	60

(Utah, Colorado, Arizona Dice)

Location	Product			
	Mono Laser	*Ink Jet*	*Photo*	*Portable*
Utah	40	90	50	30
Arizona	70	55	60	35
Colorado	65	45	85	60

Navigation Operators

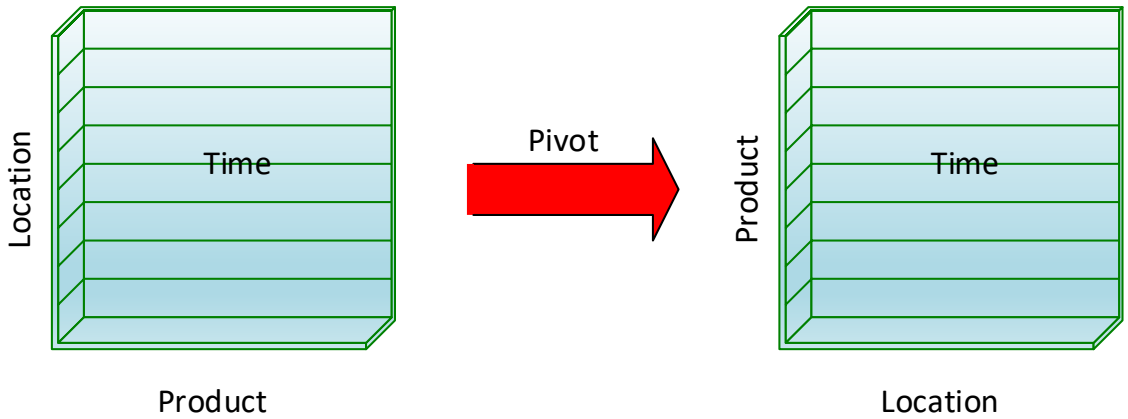
- Operators for hierarchical dimensions
- Drill-down: add detail to a dimension
- Roll-up: remove detail from a dimension
- Distribute or recalculate measure values

Drill-down Example

Location	Product			
	Mono Laser	Ink Jet	Photo	Portable
California	80	110	60	25
- Utah				
Salt Lake	20	20	10	15
Park City	5	30	10	5
Ogden	15	40	30	10
Arizona	70	55	60	35
Washington	75	85	45	45
Colorado	65	45	85	60

Pivot Operator

- Rotate or rearrange dimensions



Operator Summary

Operator	Purpose	Description
Slice	Focus attention on a subset of dimensions	Replace a dimension with a single member value or with a summary of its measure values
Dice	Focus attention on a subset of member values	Replace a dimension with a subset of members
Drill-down	Obtain more detail about a dimension	Navigate from a more general level to a more specific level
Roll-up	Summarize details about a dimension	Navigate from a more specific level to a more general level
Pivot	Present data in a different order	Rearrange the dimensions in a data cube