Management Information Systems

Larisa Cherkasov, MBA

Martin V. Smith School of Business and Economics

CSU Channel Islands

Email: Larisa.Cherkasov@csuci.edu

Channel Islands
CALIFORNIA STATE UNIVERSITY

Data Warehouse, Business Intelligence and Data Visualization

Channel Islands
CALIFORNIA STATE UNIVERSITY

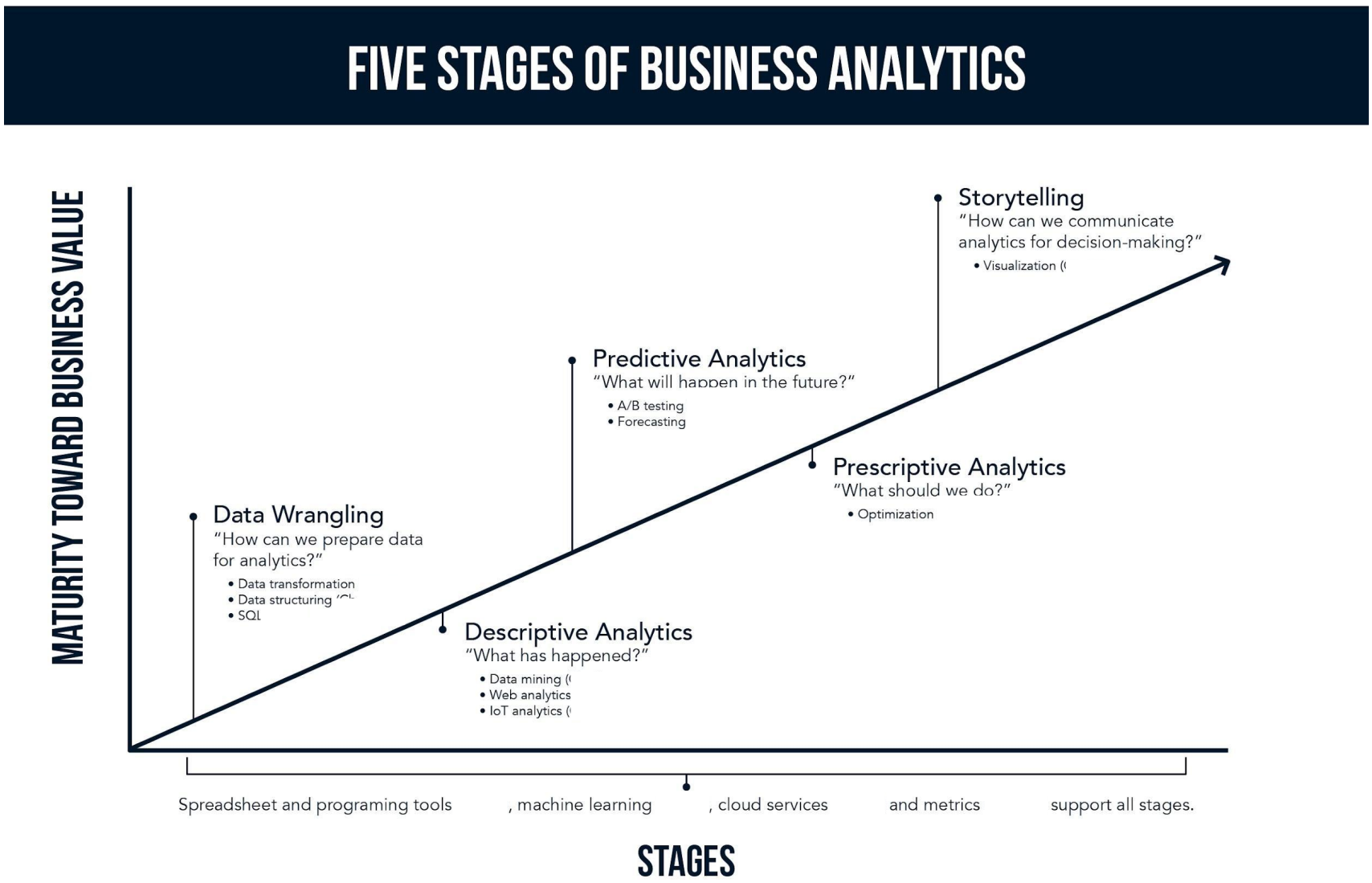
Business Analytics Definitions

- **Business Intelligence (BI):**
  - The process of gathering meaningful information to answer questions and identify significant trends or patterns by analyzing and visualizing past data
- **Advanced analytics:**
  - Tools and techniques used to understand why something happened, predict future outcomes, or discover hidden patterns in large data sets
- **Business Analytics**
  - Umbrella term for all these concepts

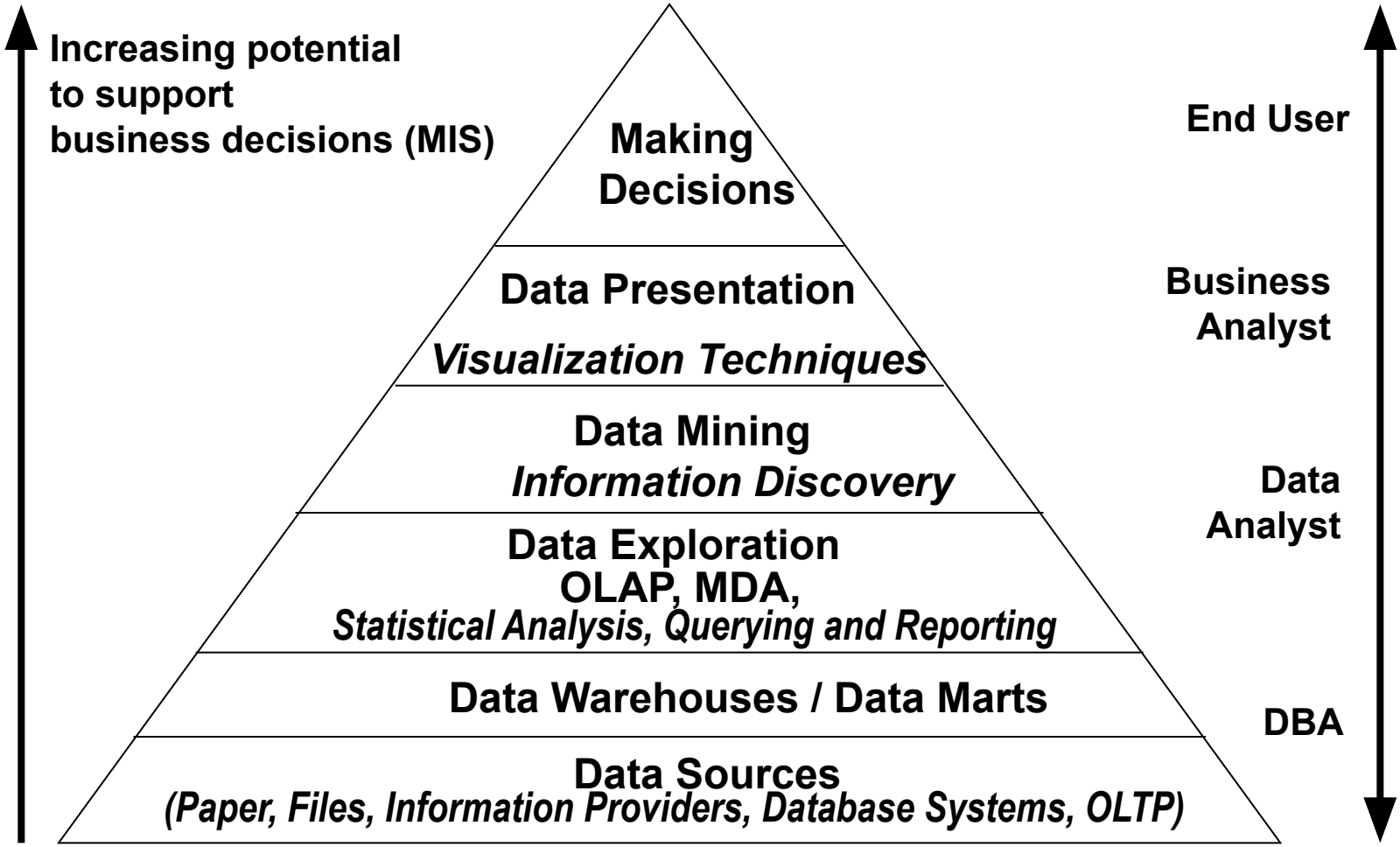
Business Analytics Questions

- What happened?
- What were our total sales this month?
- What's happening?
- Are our sales going up or down?
- Trend analysis
- Why have sales gone down?
- What will happen?
- Forecasting & "What If" Analysis
- What do I want to happen?
- Planning & Targets

Five Stages of Business Analytics

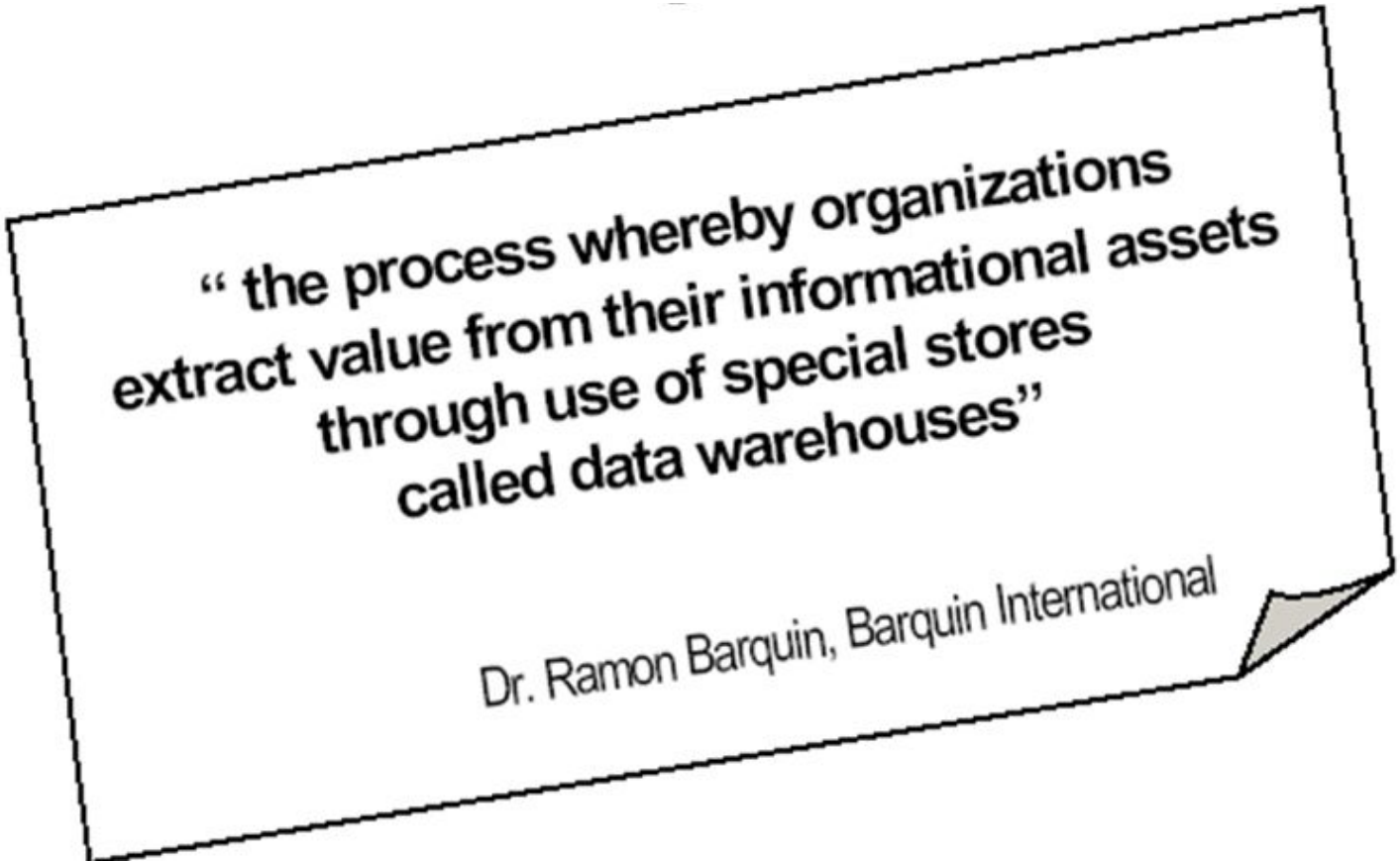


Business Analytics Flow



Data Warehouse

- Data warehouse is a database designed for centralizing data for optimized analysis
  - **What do Data Warehouse and Business Intelligence do**



DW: Inmon's Definition

What is Data Warehouse?

- A warehouse is a
  - subject-oriented,
  - integrated,
  - time-variant and
  - non-volatile

  collection of data in support of management's decision making process.

Bill Inmon in 1990

https://www.1keydata.com/datawarehousing/data-warehouse-definition.html

DW: Inmon's Definition

- **Subject-oriented**

Organized around major subjects such as customer, supplier, product, and sales.

  - Focus on modeling and analysis vs. operations and transaction processing.

- **Integrated**
  - Sources
    - operational systems data
      - relational databases, flat files, online transaction records.
  - Processes
    - data cleansing, data scrubbing

- **Time-variant**

Data contained in the warehouse provide information from an historical perspective.

- **Nonvolatile**

Data contained in the warehouse are physically separate from data present in the operational environment.

DW: Definitions & Purpose

- Purpose of DW
  - Support decision making
  - Collects data from OLTP systems and supports OLAP analysis for decision-making.
- OPTP: Online Transaction Processing
  - Processing at operational site
- OLAP: Online Analytical Processing
  - Processing and computing at warehouse
- DW is kept separate from operational databases (OD)

Operational Databases

- Constructed for well-known tasks and workloads.
- Query allows to read and modify operations.
- Maintains current data.

DW

- Queries are complex and present a general form of data
- OLAP query needs only **read only** access of data
- Maintains historical data

DW: OLTP vs. OLAP

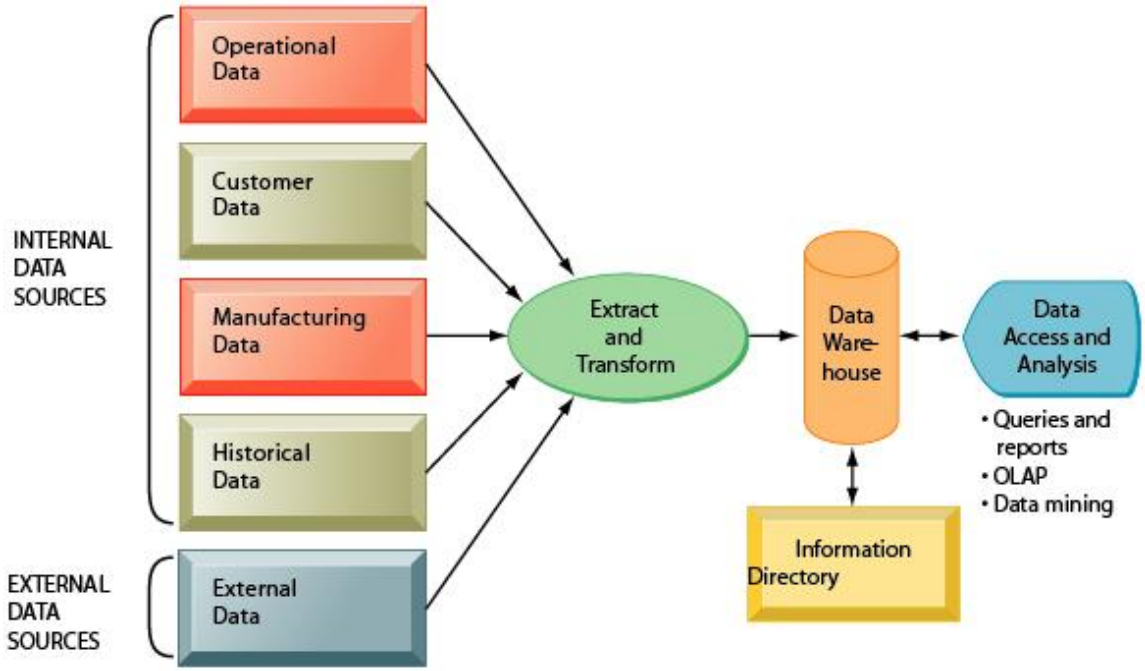
- OLTP Questions:
  - When did that order ship?
  - How many units are in inventory?
  - Does this customer have unpaid bills?
  - Are any of customer X's line items on backorder?
- Analysis Questions
  - What factors affect order processing time?
  - How did each product line (or product) contribute to profit last quarter?
  - Which products have the lowest Gross Margin?
  - What is the value of items on backorder, and is it trending up or down over time?

DW: OLTP vs. OLAP

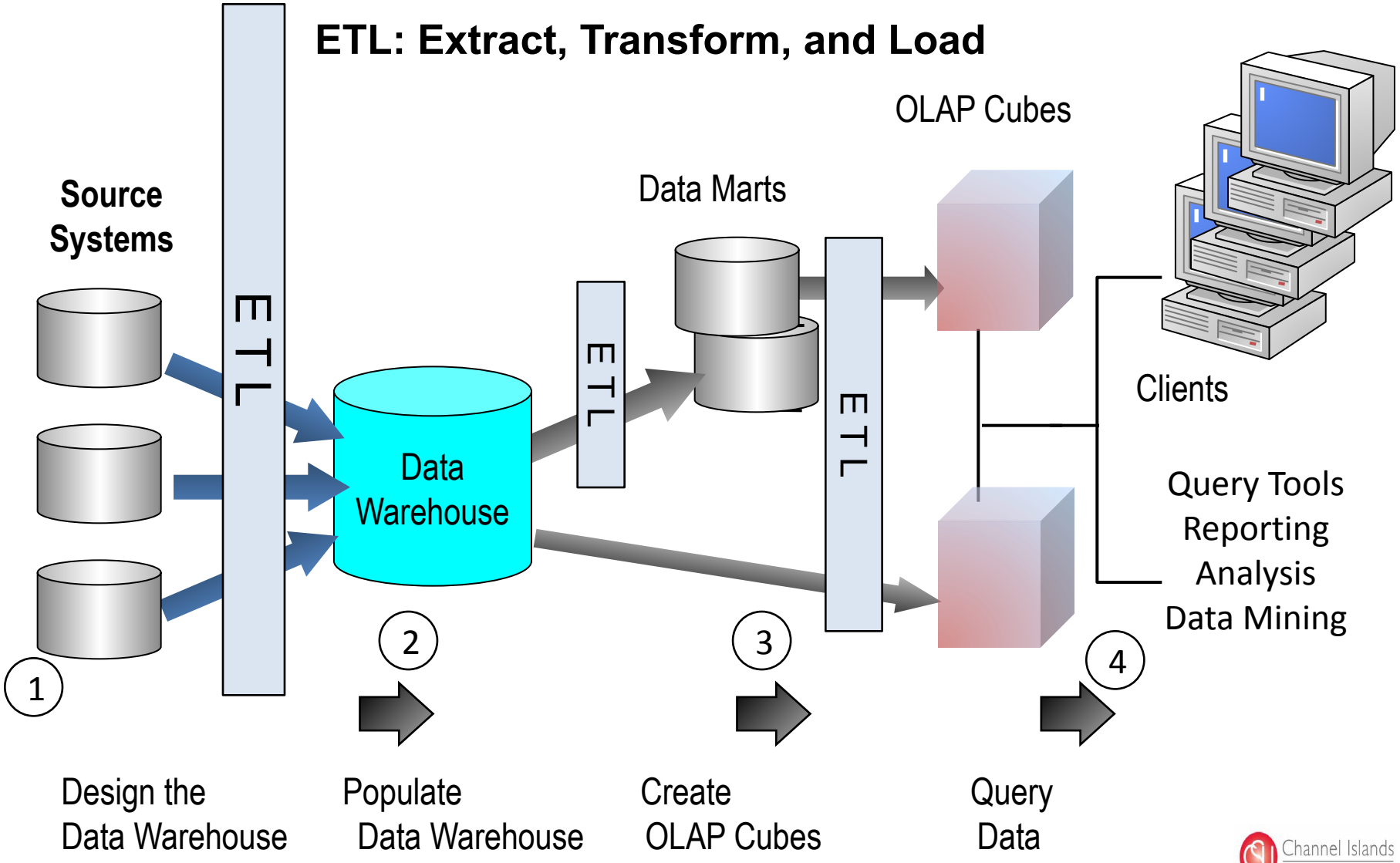
	OLTP System	OLAP System
Source of data	Operational data; OLTPs are the original source of the data.	Consolidation data; OLAP data comes from the various OLTP Databases
Purpose of data	To control and run fundamental business tasks	To help with planning, problem solving, and decision support
What the data	Reveals a snapshot of ongoing business processes	Multi-dimensional views of various kinds of business activities
Inserts and Updates	Short and fast inserts and updates initiated by end users	Periodic long-running batch jobs refresh the data
Queries	Relatively standardized and simple queries Returning relatively few records	Often complex queries involving aggregations
Processing Speed	Typically very fast	Depends on the amount of data involved; batch data refreshes and complex queries may take many hours; query speed can be improved by creating indexes
Space Requirements	Can be relatively small if historical data is archived	Larger due to the existence of aggregation structures and history data; requires more indexes than OLTP
Database Design	Highly normalized with many tables	Typically de-normalized with fewer tables; use of star and/or snowflake schemas
Backup and Recovery	Backup religiously; operational data is critical to run the business, data loss is likely to entail significant monetary loss and legal liability	Instead of regular backups, some environments may consider simply reloading the OLTP data as a recovery method

DW: Definitions & Purpose

- Data from the operational systems are
  - Extracted
  - Cleansed
  - Transformed
  - Aggregated
  - Loaded into the DW

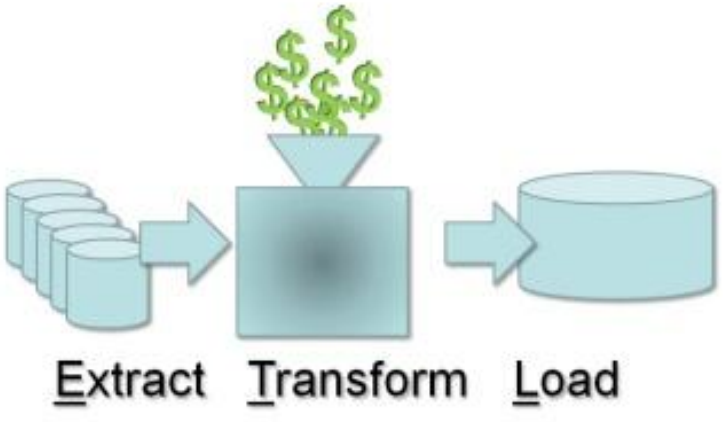


DW: Architecture & Process



DW: Extract, Transform, Load

- ETL = Extract, Transform, Load.
- ETL cycle includes
  - Build reference data
  - Extract
    - From sources
  - Validate
  - Transform
    - Clean, apply business rules, check for data integrity, create aggregates
  - Stage
    - Load into staging tables
  - Audit reports on compliance with business rules.
  - Publish/load
  - Clean up



DW: Multidimensional Modeling

- BI requires Multidimensional Modeling
- Data is divided into:
  - Facts
    - Have measures that can be aggregated.
    - Each sales record is a fact, and its sales value is a measure

Product	Type	Category	Store	City	County	Date	Sales
Top	Beer	Beverage	Trøjborg	Århus	Århus	25 May, 2009	5.75

Product Store Time

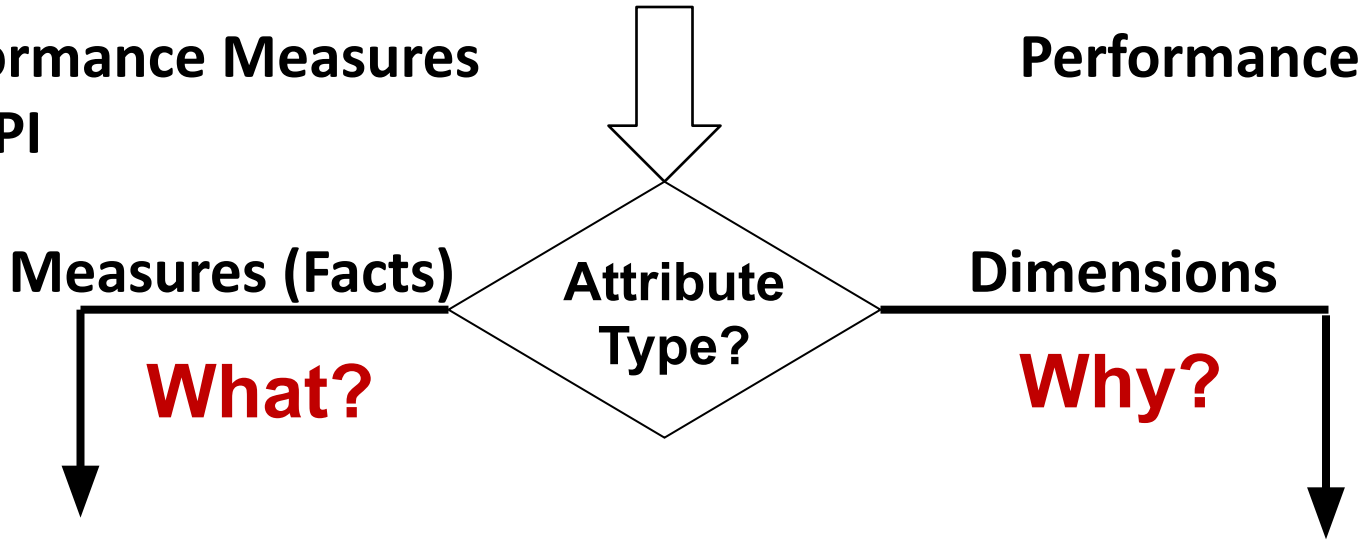
  - Dimensions
    - Group correlated attributes into the same dimension
    - Dimensions describe facts
- Facts "live" in a multidimensional cube
- Goal for dimensional modeling:
  - Surround facts with as much context (dimensions) as possible

DW: Multidimensional Modeling

Information for Decision Making

Performance Measures for KPI

Performance Drivers



The attribute (column) varies continuously:

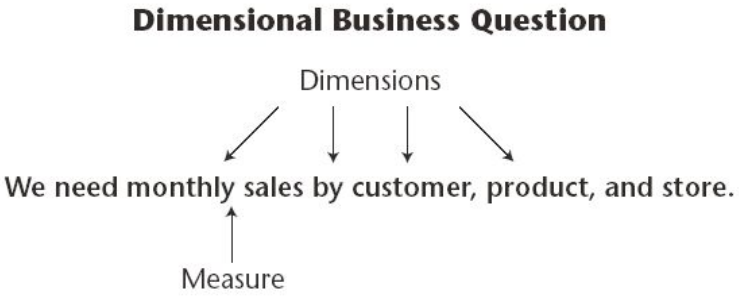
- Unit Sold
- Cost
- Sales
- Balance

The attribute is perceived as a constant or discrete value:

- Name/Description
- Location
- Color
- Size

DW: Multidimensional Modeling

- Dimensions are the core of multidimensional databases
- Dimensions are used for
  - Selection of data
  - Grouping of data at the right level of detail
- Dimension values
- May have an ordering
- Hierarchies with levels
  - Typically 3-5 levels (of detail)
- Values are organized in a tree structure
  - Product: Product->Type->Category
  - Store: Store->Area->City->County
  - Time: Day->Month->Quarter->Year



DW: Multidimensional Modeling

- Facts represent the subject of the desired analysis
- A fact is identified via its dimension values
- Fact should be attached to exactly one dimension value in each dimension
- Measures represent the fact property to study and optimize
- A measure has two components
  - Numerical value
  - Aggregation formula

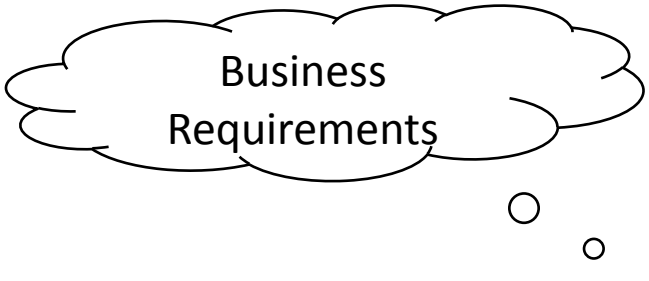
Review Questions

- What is the purpose of a Data Warehouse?
- What are four main properties of a Data Warehouse?
- How Data Warehouse is different from the operational database?
- What are the steps in ETL process?
- What are two attributes types of Multidimensional Modeling?
- What are the differences between dimensions and facts? Provide examples of both.

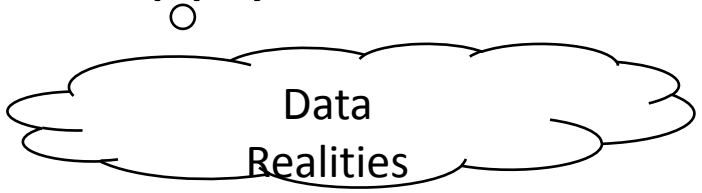
DW: Models & Operators

- Data Models
  - Relations
  - Stars & Snowflakes
  - Cubes
- Operators
  - Slice & Dice
  - Roll-up, drill down
  - Pivoting
  - Other

DW: Dimensional Design Process



- Select the business process to model
- Declare the grain of the business process/data in the fact table
  - The grain represents the most atomic level by which the facts may be defined.
- Identify the numeric facts/measures that will populate each fact table row
- Choose the dimensions that apply to each fact table row



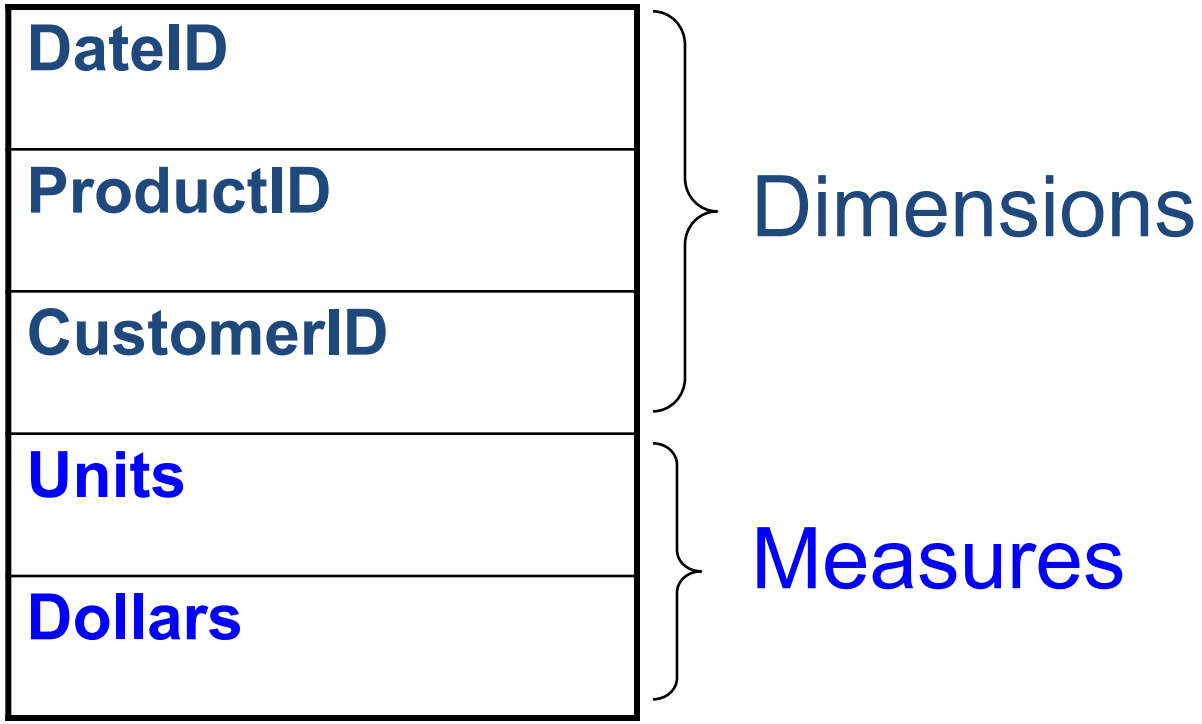
Ref: http://en.wikipedia.org/wiki/Fact_table

DW: Dimensional Design Process

- Selection of the business process to model
  - Not business departments or business functions
  - Cross-functional business processes
  - Business events
  - Examples:
    - Raw materials purchasing
    - Order fulfillment process
    - Shipments
    - Invoicing
    - Inventory
    - General ledger
    - Insurance claims
    - Class enrollment
    - Airline ticket sales

DW: Dimensional Design Process

- The Fact Table contains keys and units of measure
  - Measurements of business events



DW: Dimensional Design Process

- Fact Tables Characteristics:
  - It contains numeric measures of the business.
  - It may contain aggregated data.
  - It almost always contains date-stamped data.
  - Measures are additive.
  - Have key value
    - Concatenated key composed of the primary keys of the dimensions.
  - Joined to dimension tables through foreign keys
  - Fact tables are narrow (few attributes) but many records.

DW: Dimensional Design Process

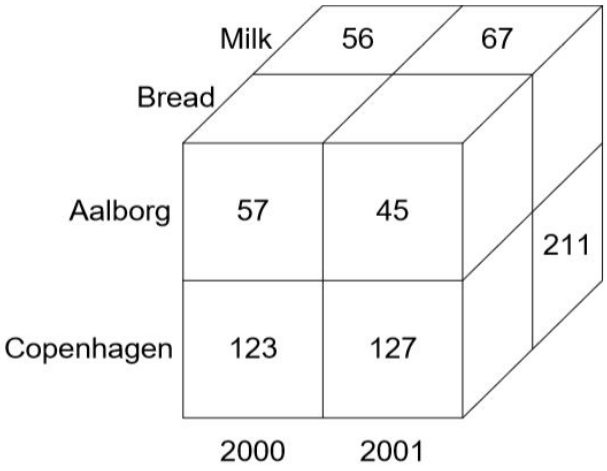
- Dimension Table Characteristics:
  - Key: Uniquely identify each row.
    - Use surrogate key.
  - Table is wide
    - A table have many attributes .
  - Textual attributes.
    - Descriptive attributes in string format
  - Attributes not directly related
  - Often not normalized (star schema).
  - Drilling down and rolling up along a dimension.
  - One or more hierarchy within a dimension.
  - Fewer number of records.

DW: Dimensional Design Process

- Identify fact tables
  - Translate business measures into fact tables
  - Analyze information from source systems for additional measures
  - Identify base and derived measures
  - Document additivity of measures
    - non-additive (price),
    - semi-additive (quantity-on-hand is not additive over time)
    - additive (quantity))
- Identify dimension tables
- Link fact tables to the dimension tables
- Create views for users

DW: Data Cube

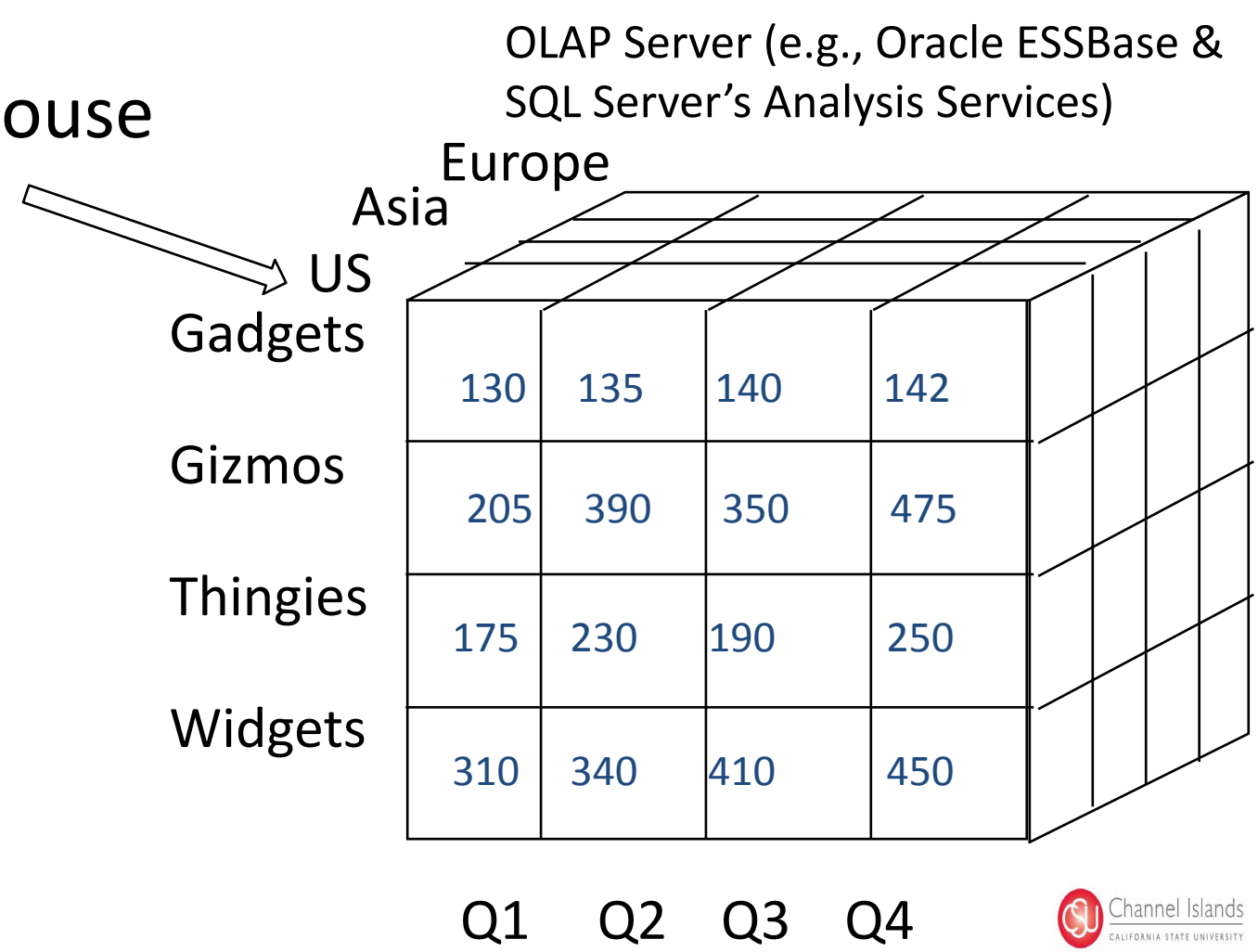
- A data cube is a collection of data that's been aggregated to allow queries to return data quickly
  - A "cube" may have many dimensions!
  - Dimensionality reduced by queries via projection/aggregation
- A cube consists of cells
  - A given combination of dimension values
  - A cell can be empty (no data for this combination)
  - A sparse cube has few non-empty cells
  - A dense cube has many non-empty cells



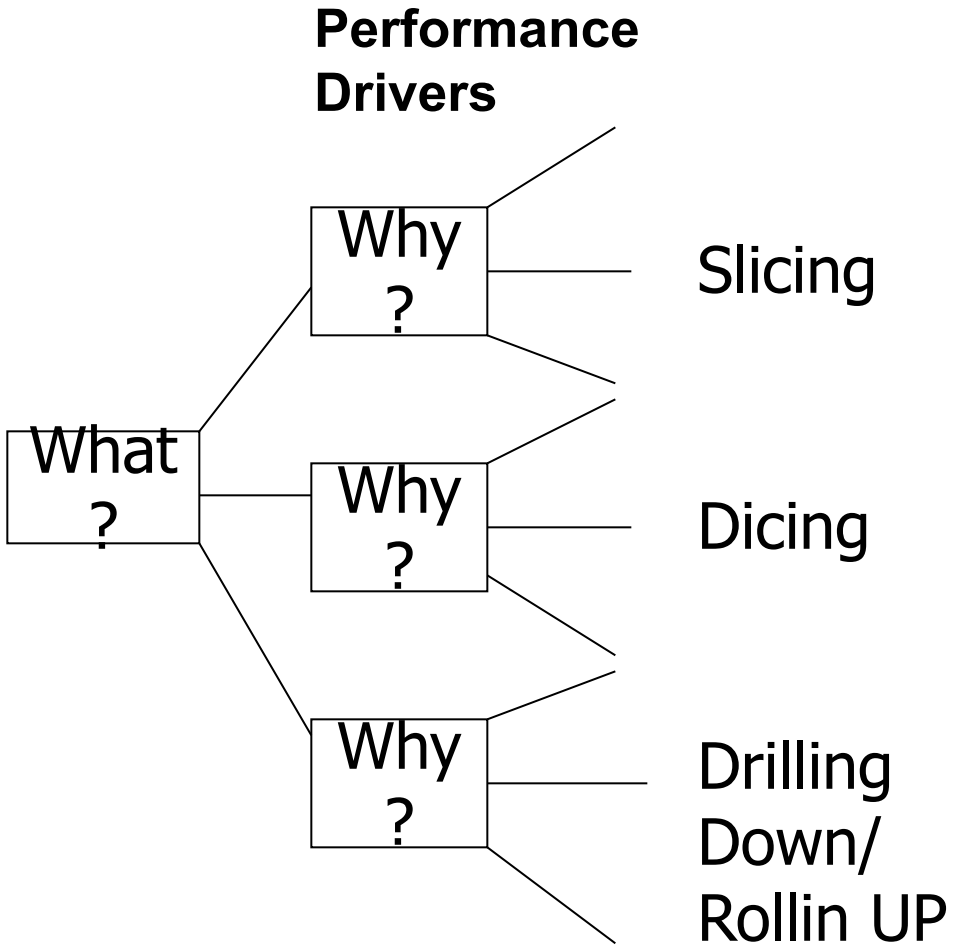
DW: OLAP Solutions

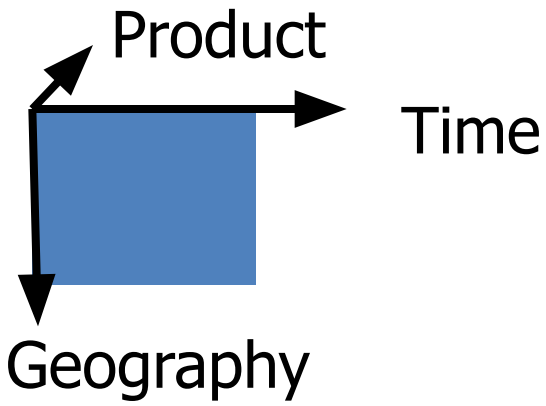
- An OLAP cube is a multidimensional database structure to support slicing, dicing, and drill-down

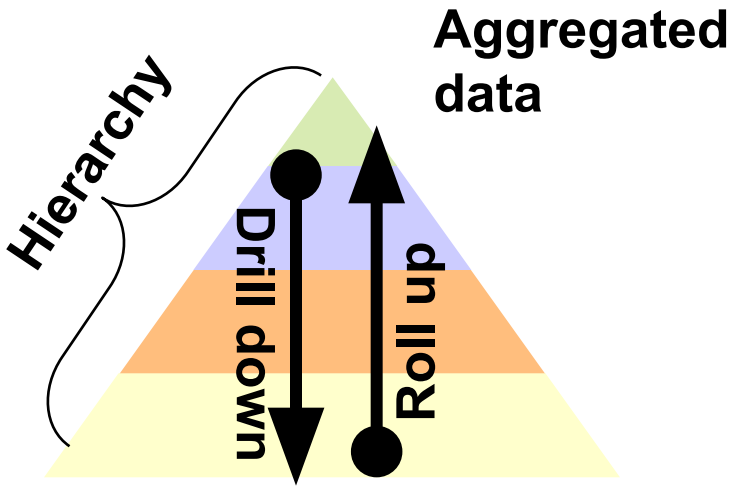
- Data Warehouse
- Data Mart
- Cubes
- Dimensions
- Measures
- Cells



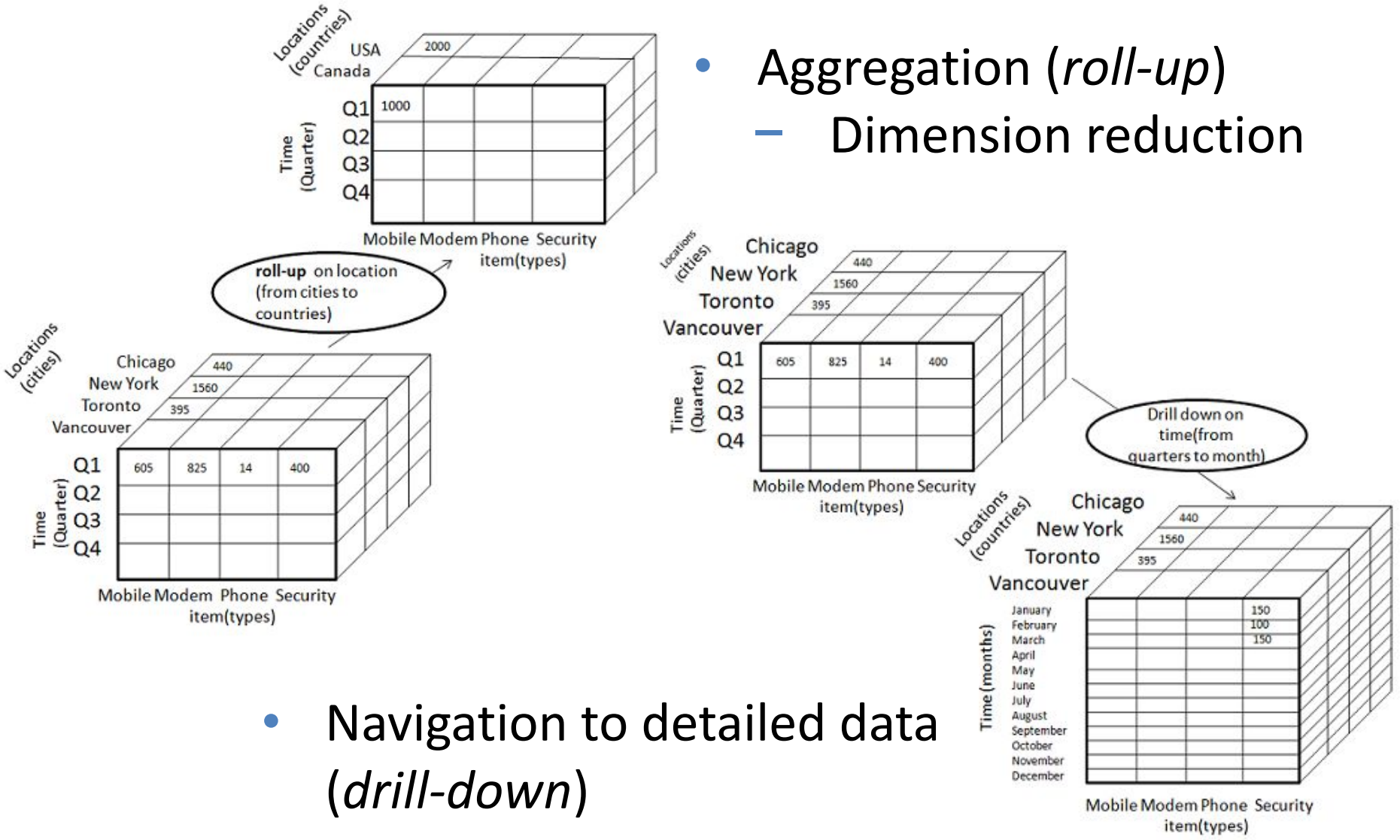
DW: Multidimensional Query Techniques







DW Operations: Roll-Up and Drill-Down

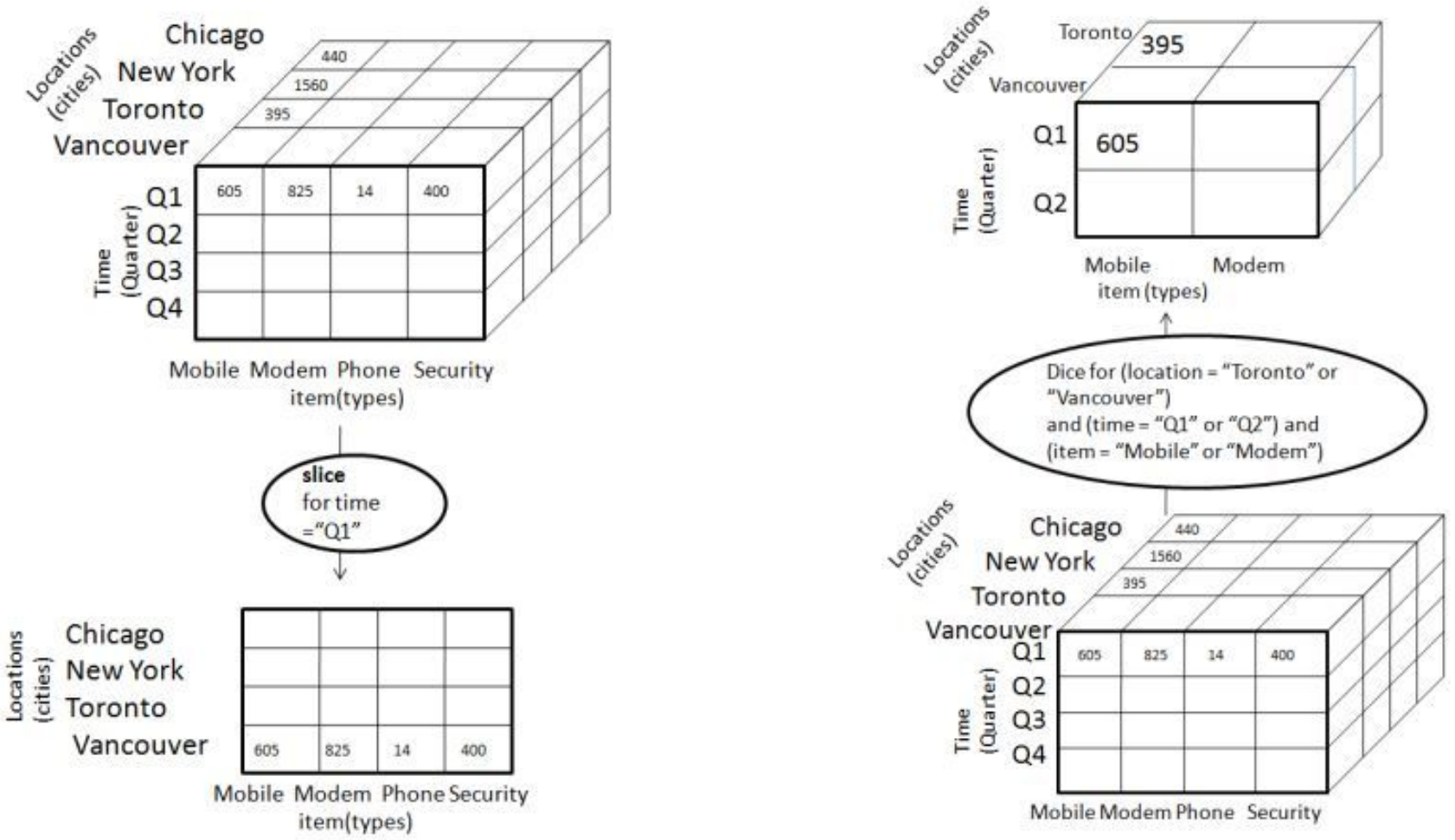


- Aggregation (*roll-up*)
  - Dimension reduction
- Navigation to detailed data (*drill-down*)

Source: http://www.tutorialspoint.com/dwh/dwh_olap.htm

DW Operations : Slice and Dice

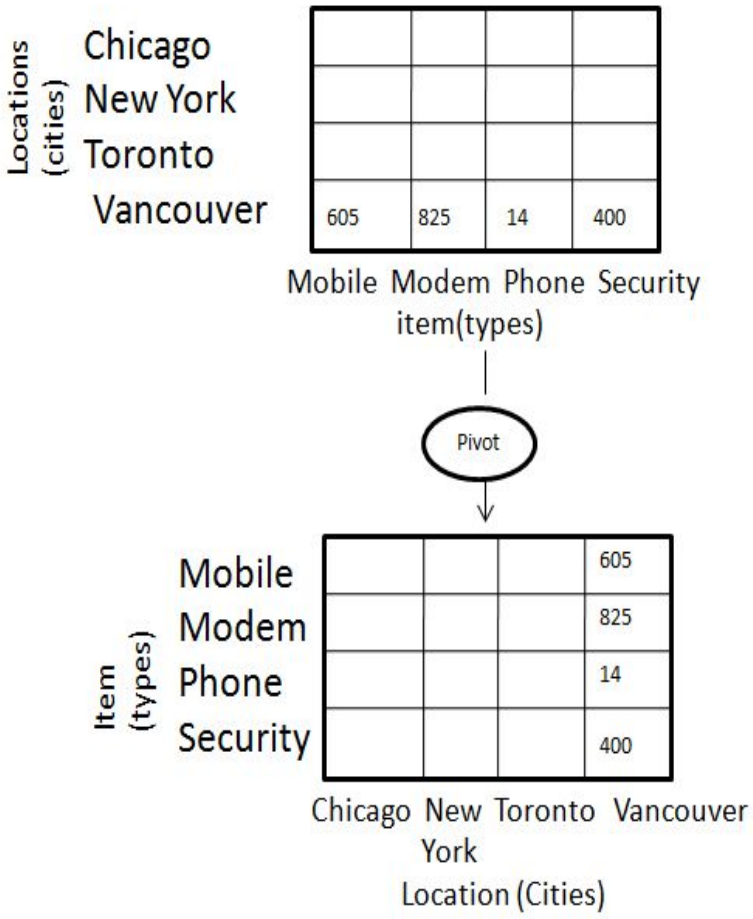
- Selection (*slice or dice*) defines a subcube

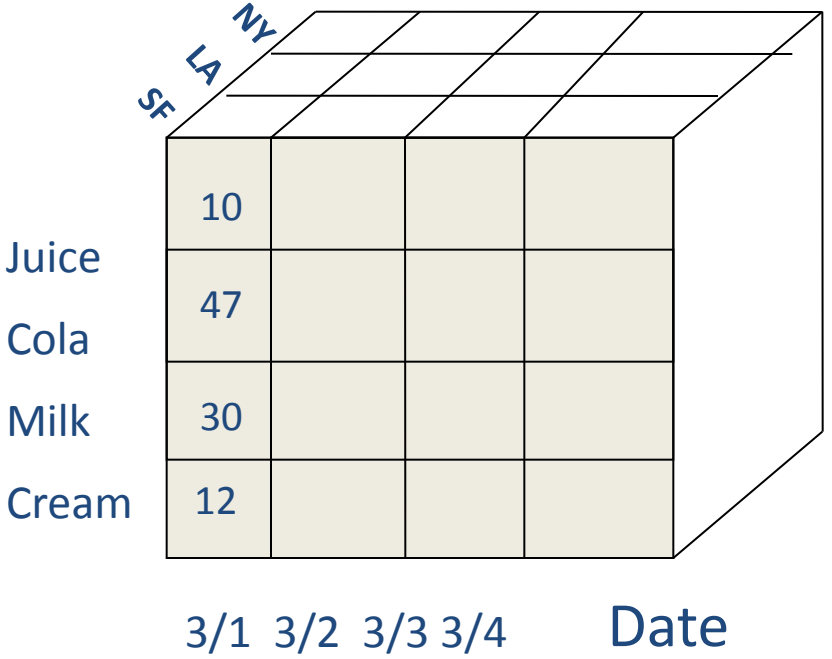


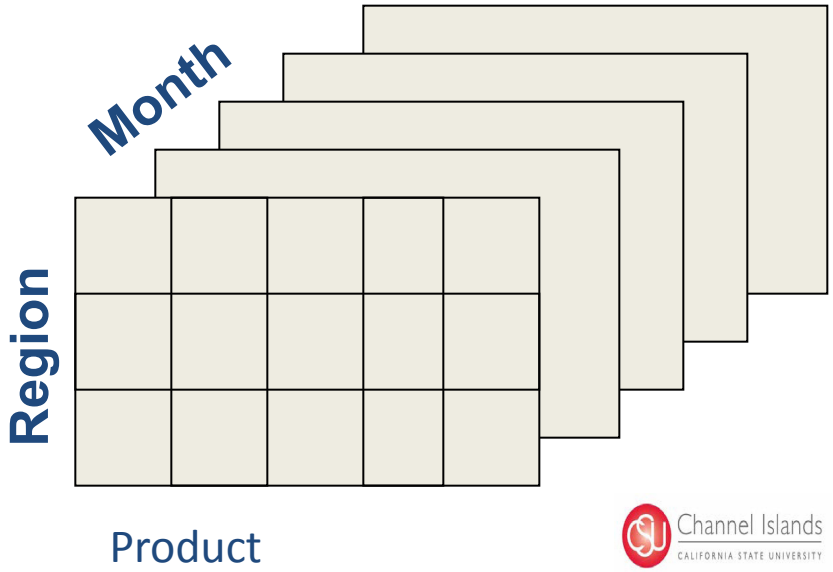
Source: http://www.tutorialspoint.com/dwh/dwh_olap.htm

DW Operations : Pivot

- Visualization operations (e.g., Pivot)



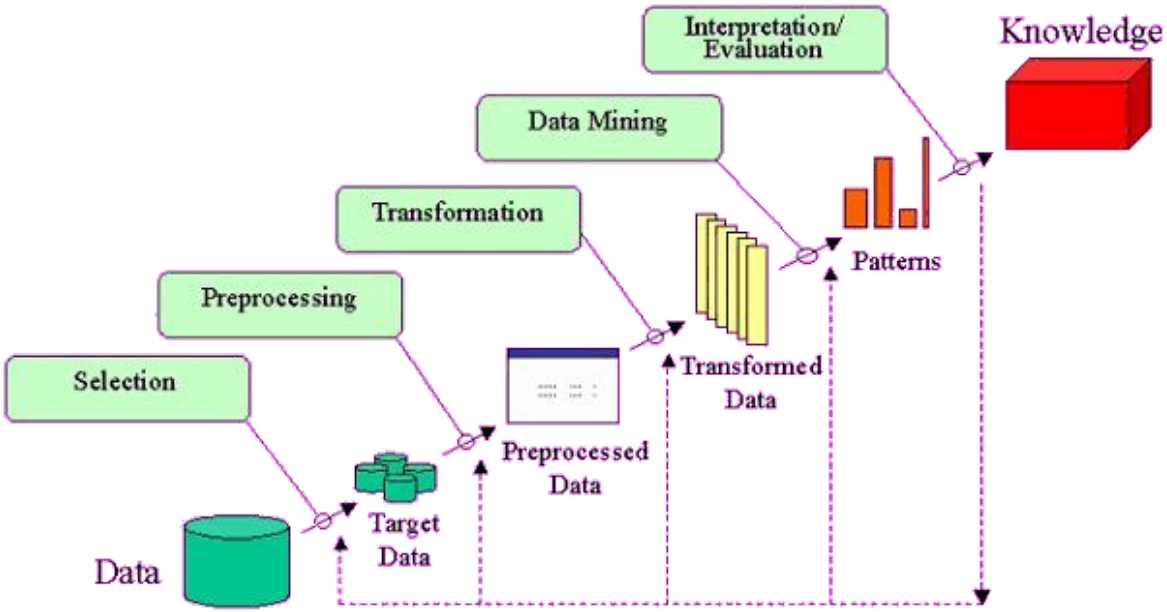




Source: http://www.tutorialspoint.com/dwh/dwh_olap.htm

Data Mining

- Data Mining is the practice of searching through large amounts of computerized data to find useful patterns or trends
  - Analyzing data with qualitative methods to extract actionable information



Data Mining vs. OLAP

- OLAP and Data Mining address different types of questions
  - Reporting and OLAP are informative about past facts
  - Data Mining can help to predict the future of the business.

OLAP	Data Mining
What was the response rate to our mailing?	What is the profile of people who are likely to respond to future mailings?
How many units of our new product did we sell to our existing customers?	Which existing customers are likely to buy our next new product?
Who were my 10 best customers last year?	Which 10 customers offer me the greatest profit potential?
Which customers didn't renew their policies last month?	Which customers are likely to switch to the competition in the next six months?
Which customers defaulted on their loans?	Is this customer likely to be a good credit risk?
What were sales by region last quarter?	What are expected sales by region next year?
What percentage of the parts we produced yesterday are defective?	What can I do to improve throughput and reduce scrap?

Source: http://www.dmreview.com/editorial/dmreview/print_action.cfm?articleId=2367

Examples: What is (not) Data Mining?

- **What is not Data Mining?**
  - Look up phone number in phone directory
  - Query a Web search engine for information about "Amazon"

- **What is Data Mining?**
  - Certain names are more prevalent in certain US locations (O'Brien, O'Rurke, O'Reilly… in Boston area)
  - Group together similar documents returned by search engine according to their context (e.g. Amazon rainforest, Amazon.com,)

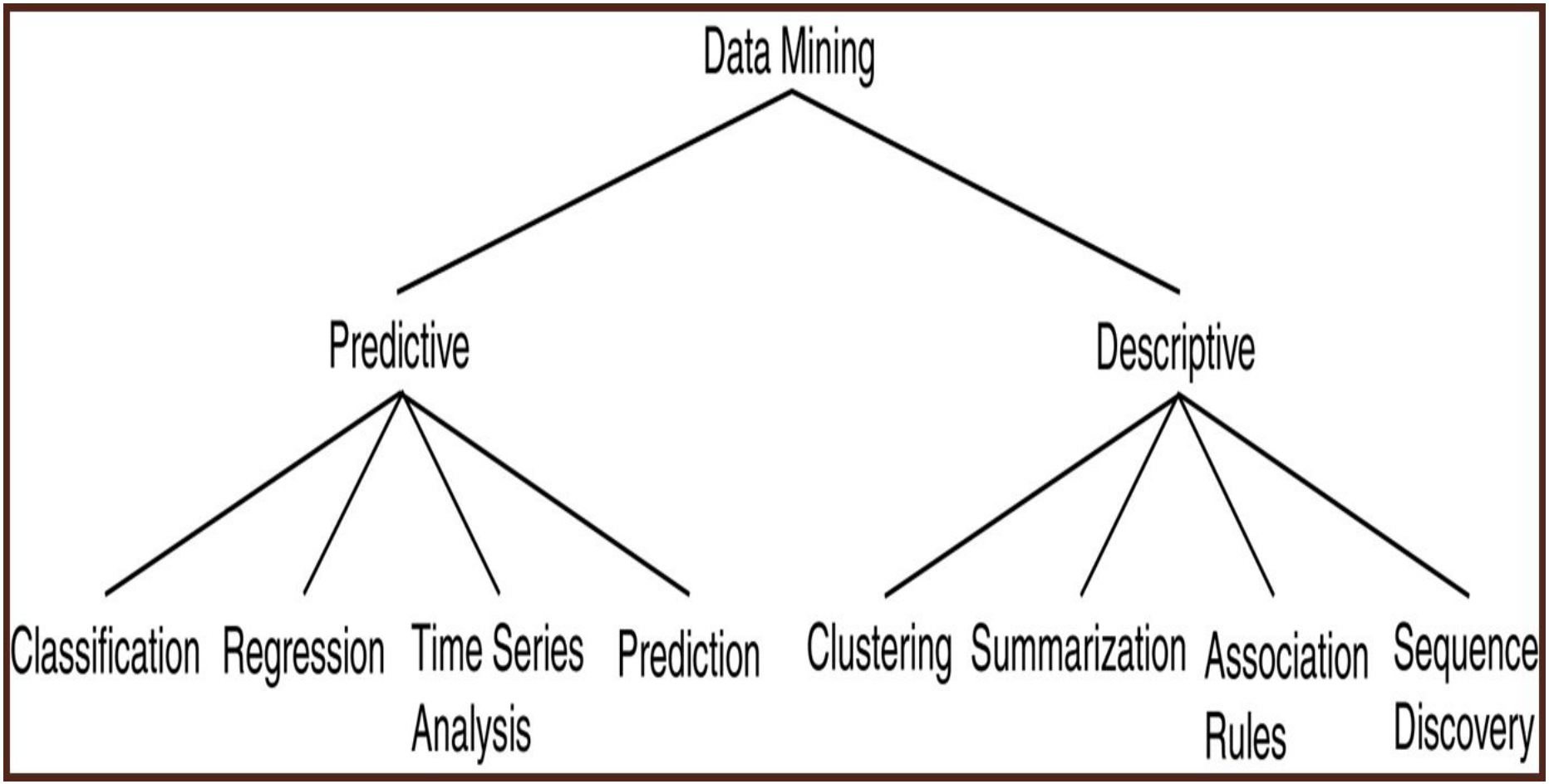
Data Mining: How You're Revealing More Than You Think

Data Mining: Methods

- Classification
  - Assign data into predefined categories
    - Predicting if an email is spam or not spam
- Regression
  - Predict a numeric or continuous value
    - Estimating a house price based on size, location, and age.
- Clustering
  - Finding “natural” groupings of data points based on similarity metrics
    - Grouping customers by similar purchasing behavior
- Summarization
    - Creating a sales summary report showing average revenue per region
- Association learning
    - Finding that people who buy bread often also buy butter
- Anomaly detection
    - Identifying fraudulent credit card transactions.
- Text and Web content mining
  - Draws patterns from unstructured data
  - Uses clickstream data and Web crawlers
  - Associations
    - Analyzing customer reviews to find common opinions or sentiments.

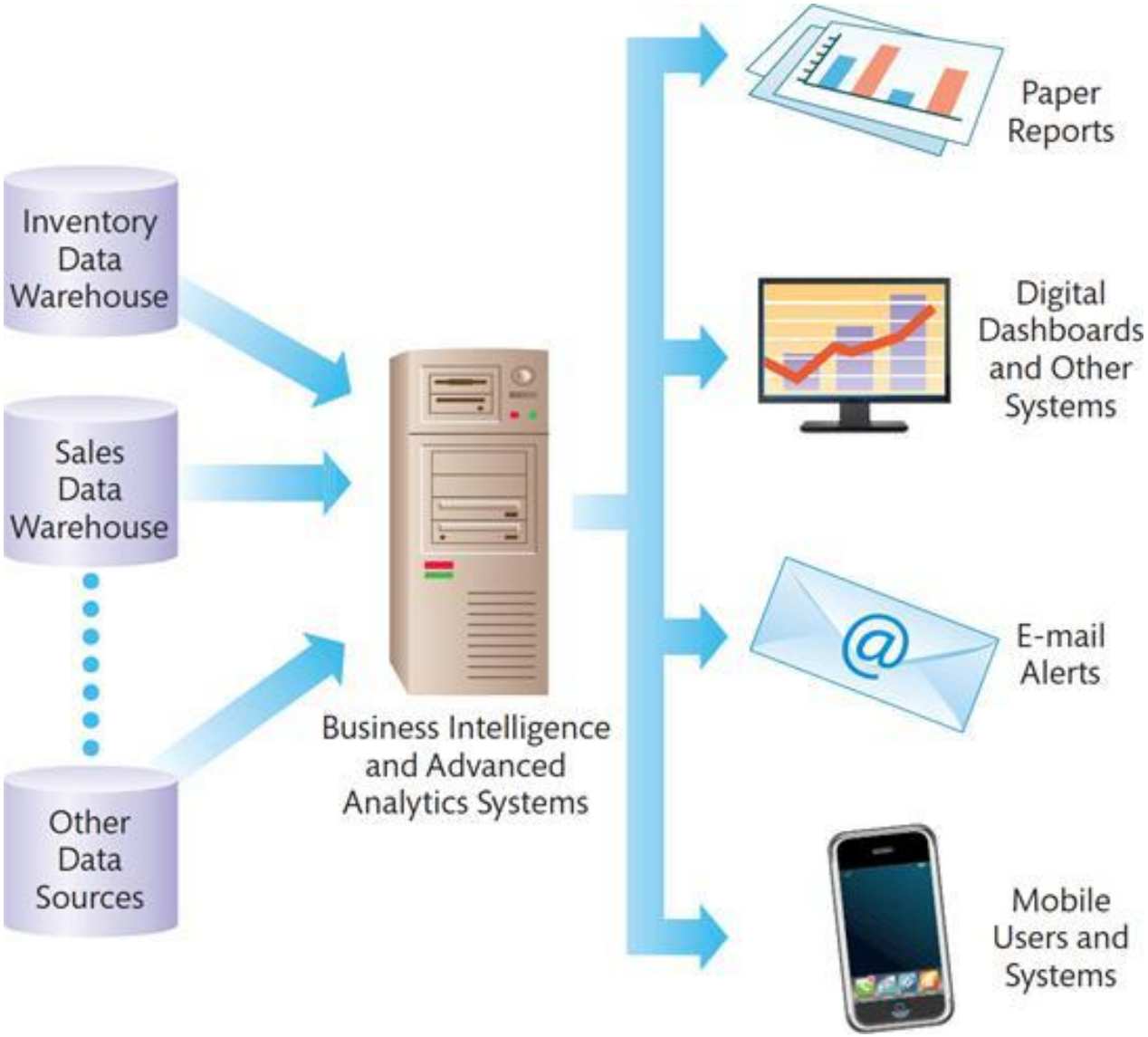
Data Mining: How You're Revealing More Than You Think

Data Mining Models and Tasks



Data Mining: How You're Revealing More Than You Think

Data Mining: Results



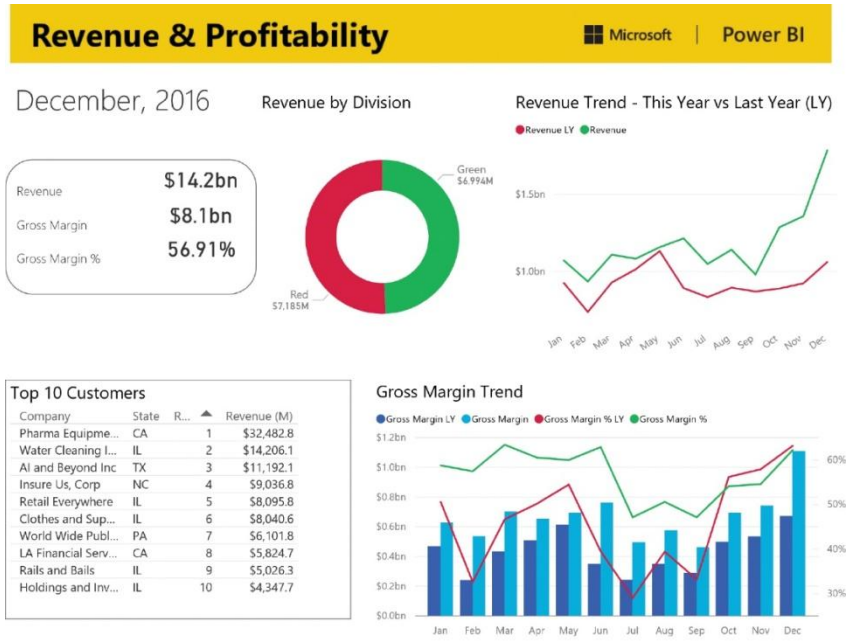
Information Visualization

- Data visualization is the graphical representation of information and data
- Common general types of data visualization:
  - Charts
  - Tables
  - Graphs
  - Maps
  - Infographics
  - Dashboards

Ice Bucket Challenge Data Visualization

Information Visualization

- Digital dashboards visually present key performance indicators
  - Data highly aggregate
  - Use a variety of design elements to present data in a user friendly way

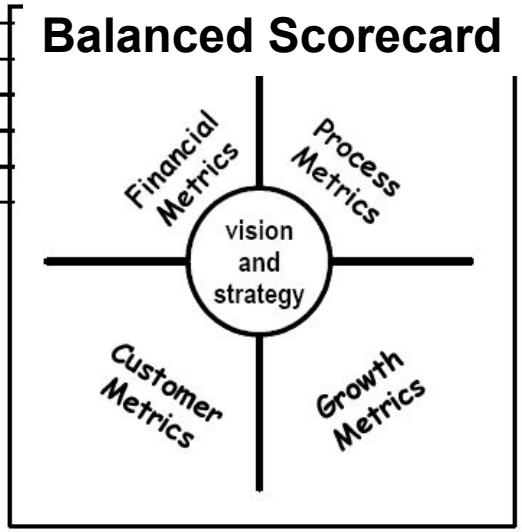


Source: Power B I 2016, Windows 10, Microsoft Corporation.

Information Visualization

- Balanced Scorecard
  - Financial measures
    - The results of actions already taken
  - Operational measures
    - The drivers of future financial performance.

trend	metric	actual	target	variance
↑	same store sales	$108.0m	$120.0m	- 10%
→	customer retention	96%	95%	+0.9%
→	new customers	3.8k	5.0k	-24.0%
↓	charge cards issued			
↑	30 day past-due accounts			
↑	60 day past-due accounts			
↑	90 day past-due accounts			
↓	merchandise return rate			
→	inventory turnover rate			



The Balanced Scorecard—Measures that Drive Performance

Information Visualization: Solutions

- Tableau is a business intelligence system that helps companies visualize and understand their data.
  - Benefits:
    - Connection to various data sources
    - Intuitive and easy interface
    - Advanced collaboration
    - Several methods for data investigation
    - Flexible pricing & deployment
    - Reliable support
    - Free one-year licenses to students at accredited academic institutions through Tableau for Students program
      - Tutorial: Get Started with Tableau Desktop

Information Visualization: Solutions

- Microsoft Power BI is an interactive data visualization software product with a primary focus on business intelligence
  - A Windows desktop application called Power BI Desktop.
  - An online software as a service (SaaS) service called the Power BI service.
  - Power BI Mobile apps for Windows, iOS, and Android devices.
  - Benefits:
    - Customized dashboards and interactive reports
    - Enables Big Data analysis and sharing
    - Artificial Intelligence
      - image recognition, and text analytics, build machine learning models, and interface with Azure Machine Learning
    - Integration for multiple data sources
    - Economical and cost-effective
      - Power BI Desktop is free to download and use
      - Get started with Power BI Desktop
    - User-friendly: No specialized tech support is necessary

Information Visualization: Solutions

- Features Needed in Data Visualization Software:
  - Ability to choose different visuals and graphs
  - Trend tracking capability
  - High level of security
  - Simplified software interfacing
- Data Visualization Tools
- Best Data Visualization Software Solutions of 2025

Tool	Monthly Cost	Complexity	Data Prep	Best For
Power BI	$10-20	Low	Basic	Microsoft users
Tableau	$70+	Medium	Limited	Advanced viz
Looker Studio	Free-$9	Low	Basic	Google users
Qlik Sense	$30+	Medium	Good	Self-service
Apache Superset	Free	High	Medium	Open source
D3.js	Free	High	None	Developers
Grafana	Free-$8.50	Medium	Limited	Monitoring
Sisense	Custom	Low	Good	Big data
Metabase	Free-$85	Low	Basic	Self-hosted
Domo	$83+	Medium	Good	Cloud-native
Chartio	Atlassian	Low	Basic	Collaboration

AI for Business Intelligence and Data Visualization

- AI enhances BI through predictive analytics, natural language queries, and automated insights.
  - Process large volumes of data at unparalleled speeds
  - Automate data mining
  - Use advanced algorithms to provide real-time, data-driven insights
  - Detect patterns, trends, and correlations unnoticed by conventional BI techniques by utilizing machine learning algorithms.
- AI-driven visualization tools
  - Help users interpret large datasets quickly
    - Power BI with Copilot
    - Tableau with AI features
- AI can be seen as an evolution of BI
  - Move from descriptive dashboards toward predictive and prescriptive analytics.

Data Mining: How You're Revealing More Than You Think

Sources:
Joseph Valacich, Christoph Schneider, *Information Systems Today: Managing in the Digital World*, 8th Edition
John Gallaugher, Information Systems: A Manager's Guide to Harnessing Technology, v. 7.0
Minder Chen,Ph.D., Management Information Systems Lectures