Unit 6
Advances in Databases

Data Warehousing and OLAP Technology: An Overview

- What is a data warehouse?
- A multi-dimensional data model
- Data warehouse architecture

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

Data Warehouse—Subject-Oriented

- Organized around major subjects, such as customer, product, sales
- Focusing on the modeling and analysis of data for decision makers, not on daily operations or transaction processing
- Provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process

Data Warehouse—Integrated

- Constructed by integrating multiple, heterogeneous data sources
  - relational databases, flat files, on-line transaction records
- Data cleaning and data integration techniques are applied.
  - Ensure consistency in naming conventions, encoding structures, attribute measures, etc. among different data sources
    - E.g., Hotel price: currency, tax, breakfast covered, etc.
  - When data is moved to the warehouse, it is converted.

Data Warehouse—Time Variant

- The time horizon for the data warehouse is significantly longer than that of operational systems
  - Operational database: current value data
  - Data warehouse data: provide information from a historical perspective (e.g., past 5-10 years)
- Every key structure in the data warehouse
  - Contains an element of time, explicitly or implicitly
  - But the key of operational data may or may not contain "time element"

Data Warehouse—Nonvolatile

- A physically separate store of data transformed from the operational environment
- Operational update of data does not occur in the data warehouse environment
  - Does not require transaction processing, recovery, and concurrency control mechanisms
  - Requires only two operations in data accessing:
    - *initial loading of data* and *access of data*

Data Warehouse vs. Heterogeneous DBMS

- Traditional heterogeneous DB integration: A query driven approach
  - Build wrappers/mediators on top of heterogeneous databases
  - When a query is posed to a client site, a meta-dictionary is used to translate the query into queries appropriate for individual heterogeneous sites involved, and the results are integrated into a global answer set
  - Complex information filtering, compete for resources
- Data warehouse: update-driven, high performance
  - Information from heterogeneous sources is integrated in advance and stored in warehouses for direct query and analysis

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

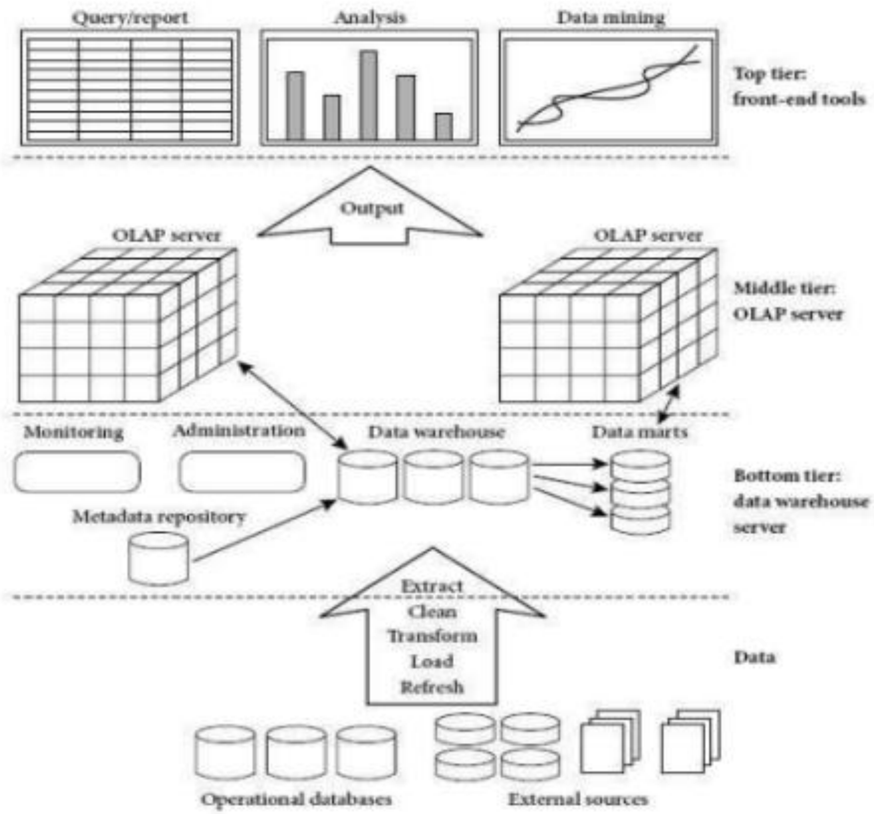
OLTP vs. OLAP

	OLTP	**OLAP**
users	clerk, IT professional	knowledge worker
function	day to day operations	decision support
DB design	application-oriented	subject-oriented
data	current, up-to-date detailed, flat relational isolated	historical, summarized, multidimensional integrated, consolidated
usage	repetitive	ad-hoc
access	read/write index/hash on prim. key	lots of scans
unit of work	short, simple transaction	complex query
# records accessed	tens	millions
#users	thousands	hundreds
DB size	100MB-GB	100GB-TB
metric	transaction throughput	query throughput, response

Why Separate Data Warehouse?

- High performance for both systems
  - DBMS— tuned for OLTP: access methods, indexing, concurrency control, recovery
  - Warehouse—tuned for OLAP: complex OLAP queries, multidimensional view, consolidation
- Different functions and different data:
  - missing data: Decision support requires historical data which operational DBs do not typically maintain
  - data consolidation: DS requires consolidation (aggregation, summarization) of data from heterogeneous sources
  - data quality: different sources typically use inconsistent data representations, codes and formats which have to be reconciled
- Note: There are more and more systems which perform OLAP analysis directly on relational databases

A multitiered architecture



3.12 A three-tier data warehousing architecture.

ETL

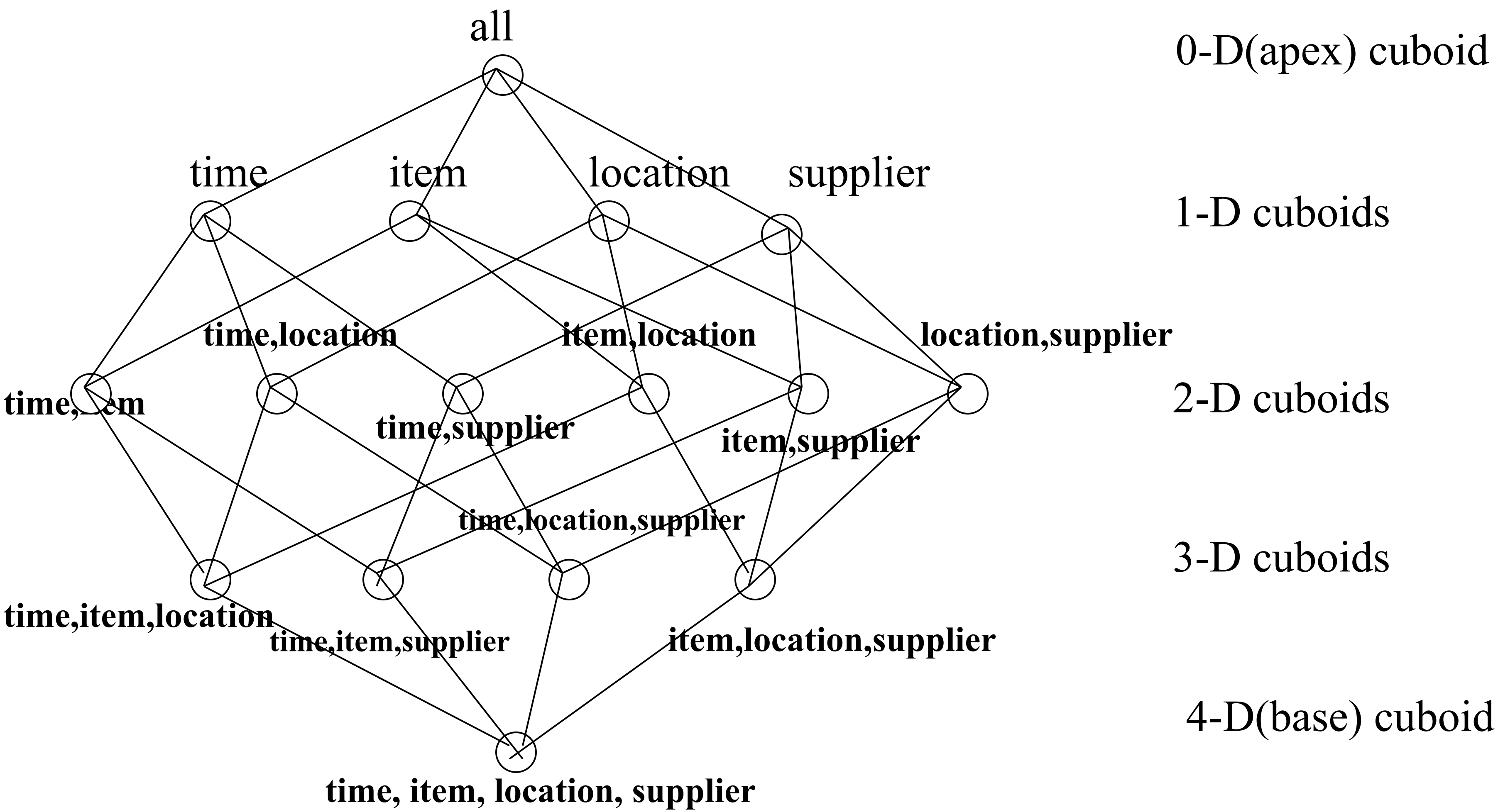
- Data Extraction
- Data Cleaning
- Data Transformation
- Load
- Refresh

A Multidimensional Data Model

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)
  - **Fact table** contains measures (such as dollars_sold) and keys to each of the related dimension tables
- 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.

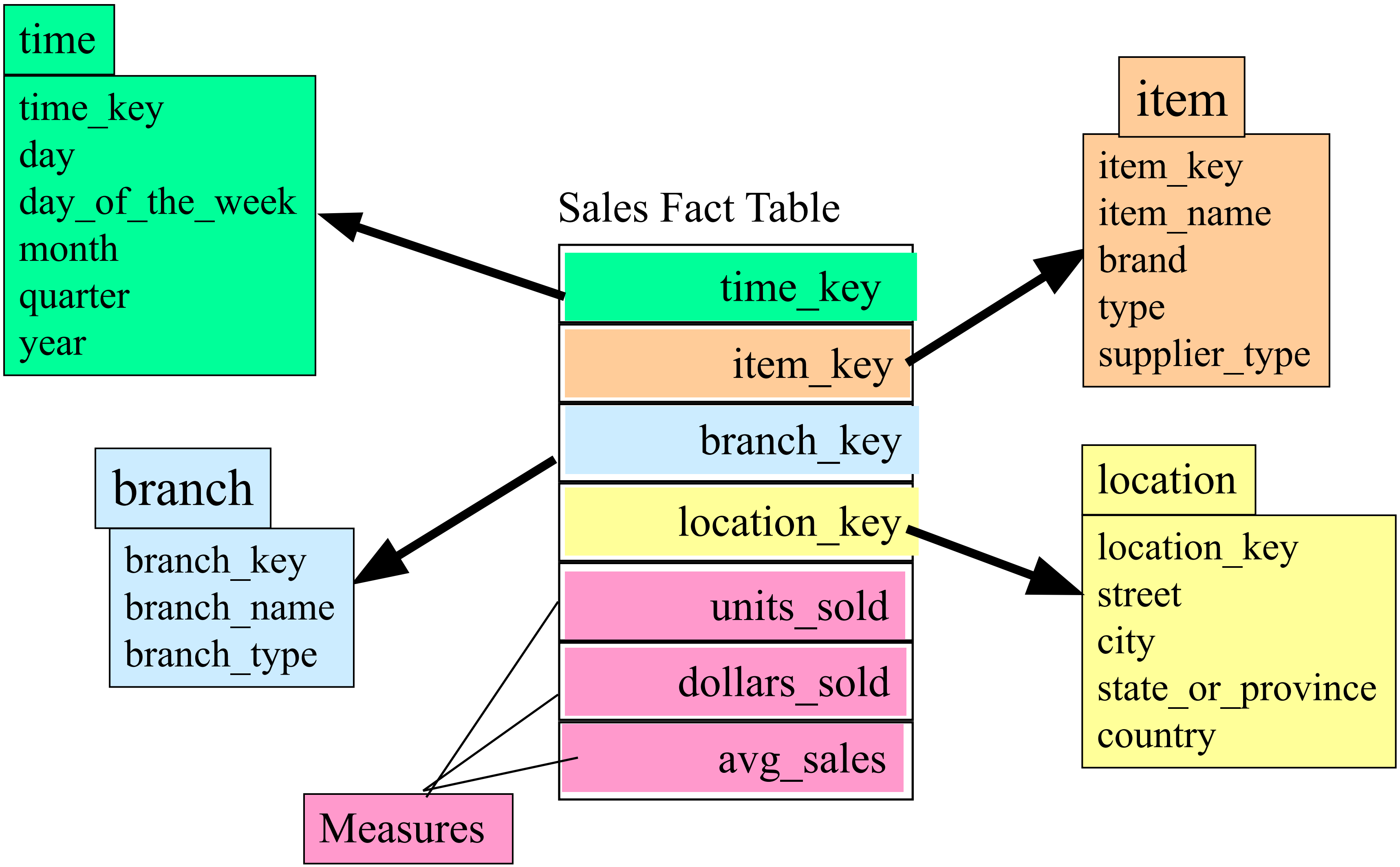
Cube: A Lattice of Cuboids



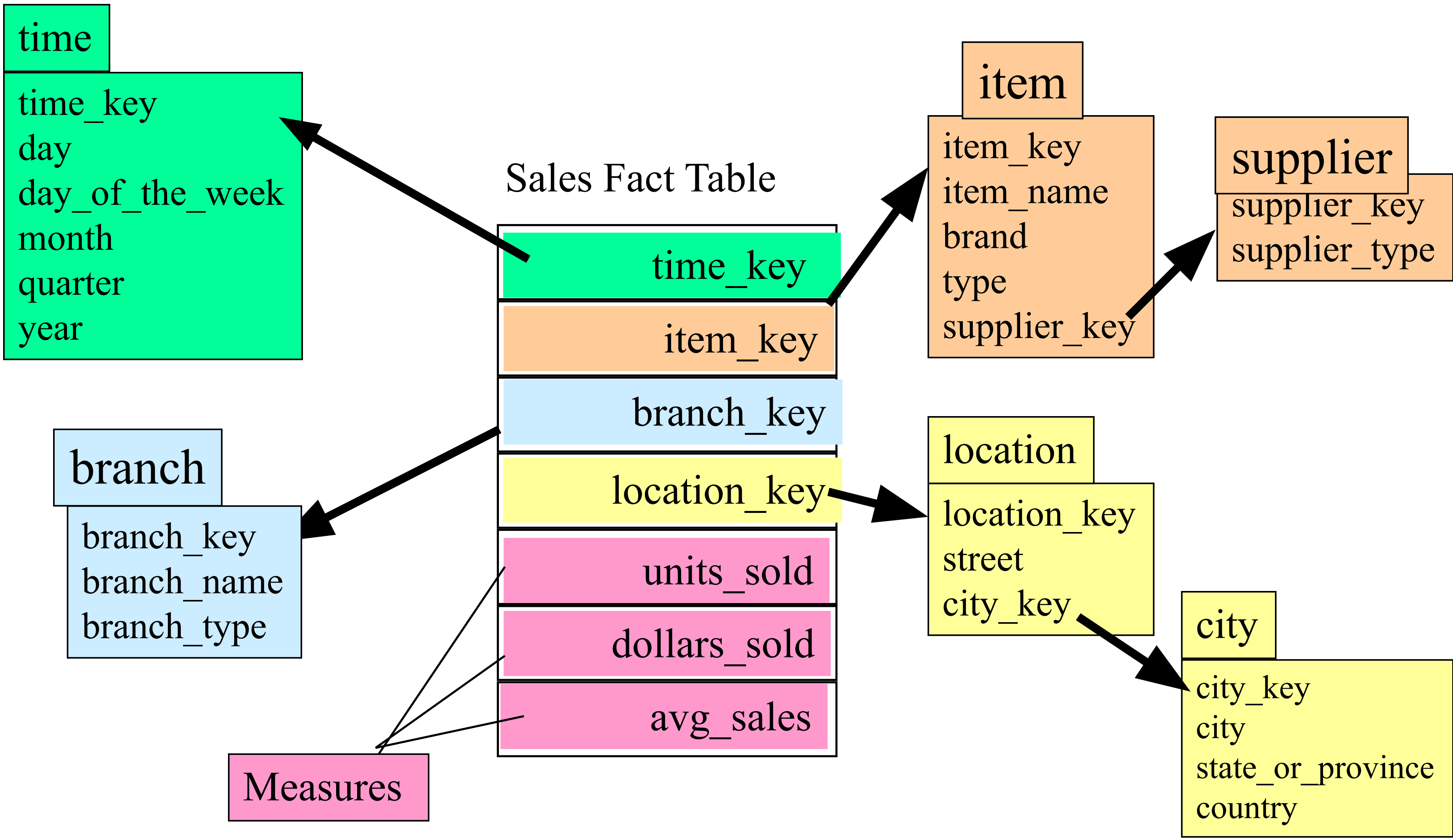
Conceptual Modeling of Data Warehouses

- 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

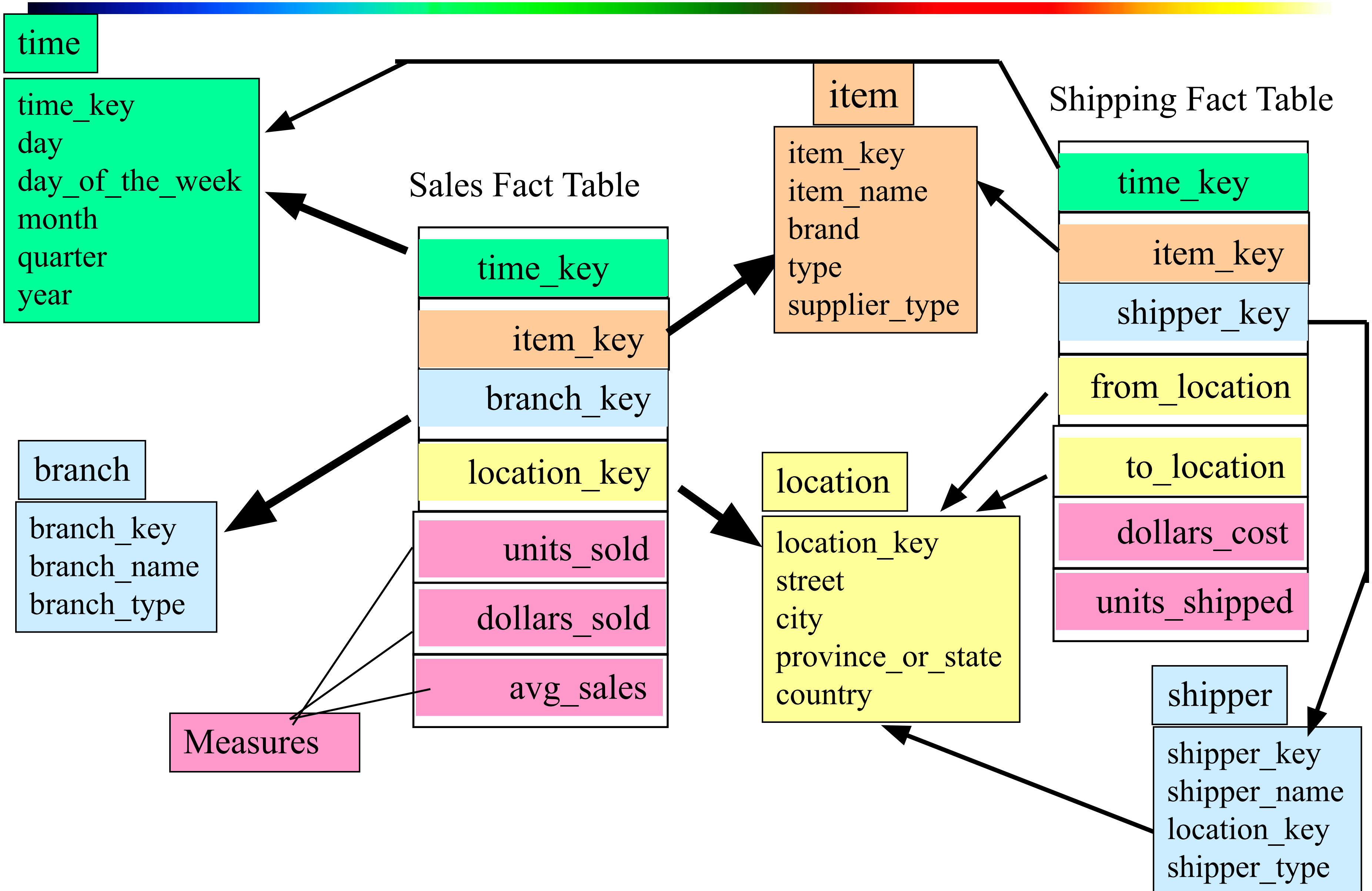
Example of Star Schema



Example of Snowflake Schema

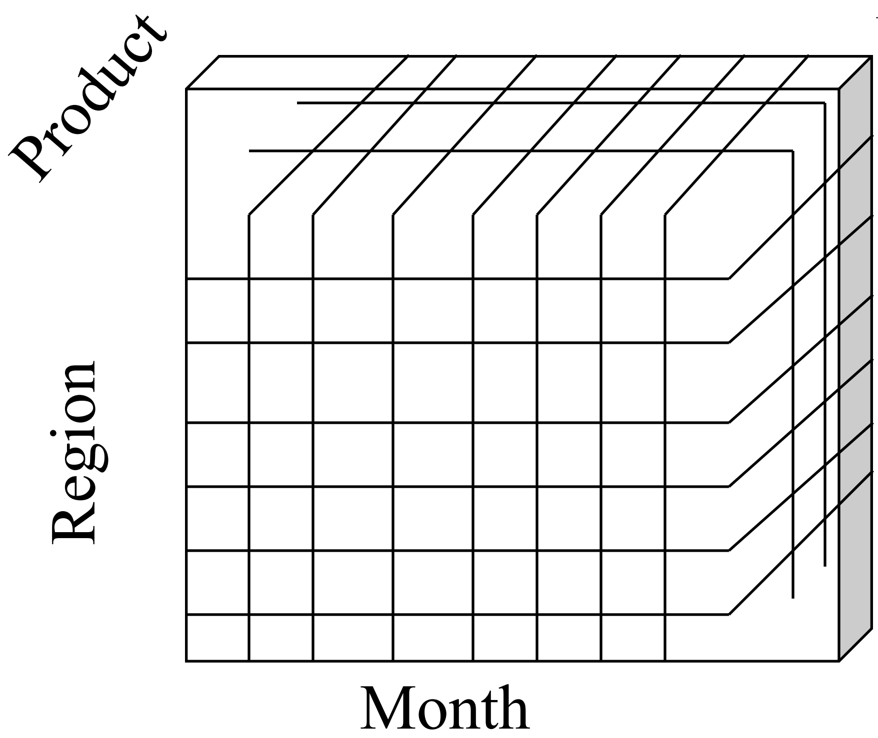


Example of Fact Constellation

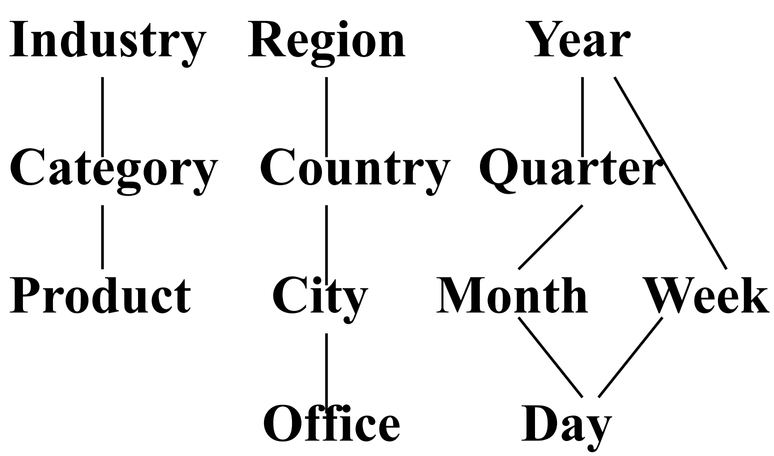


Multidimensional Data

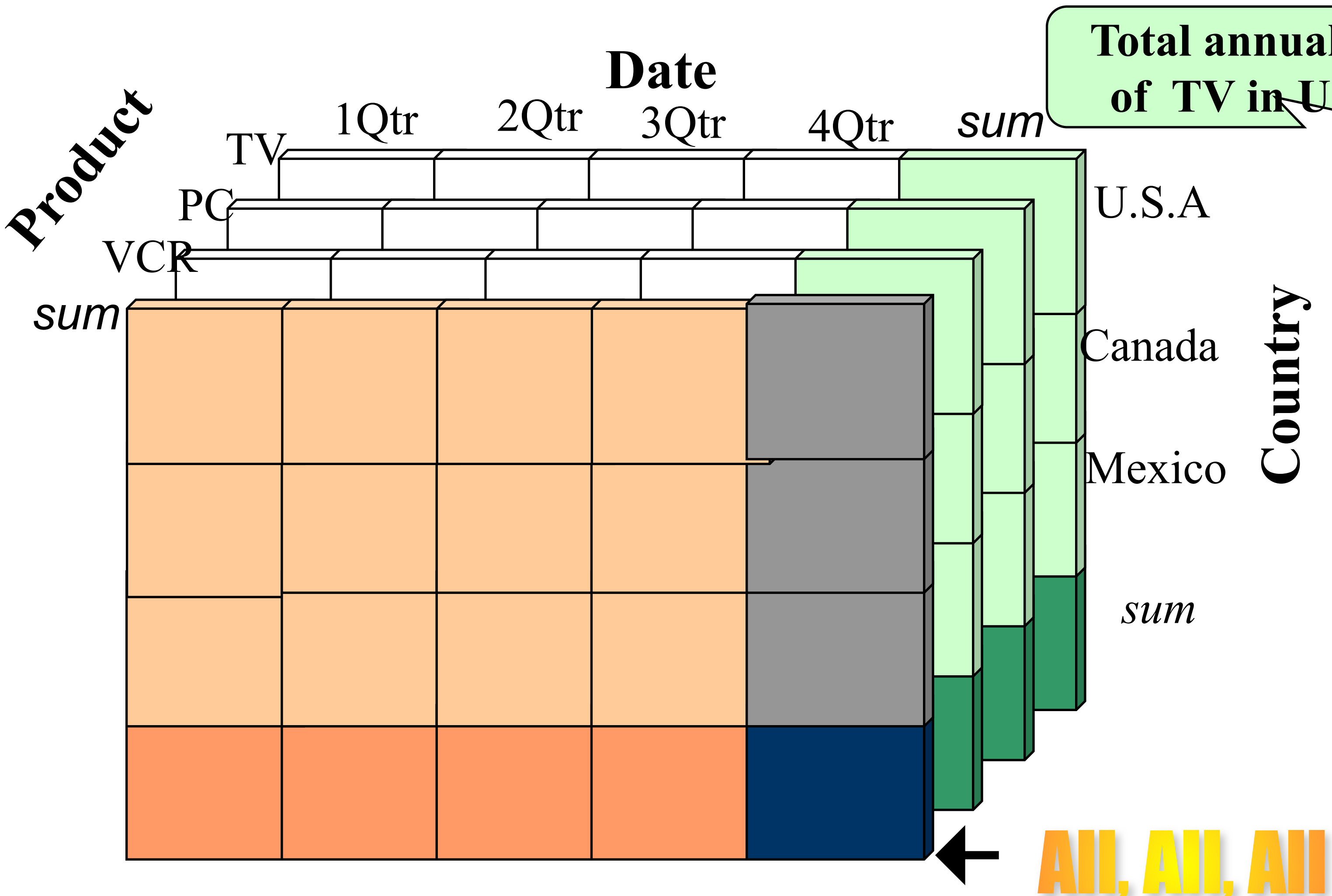
- Sales volume as a function of product, month, and region



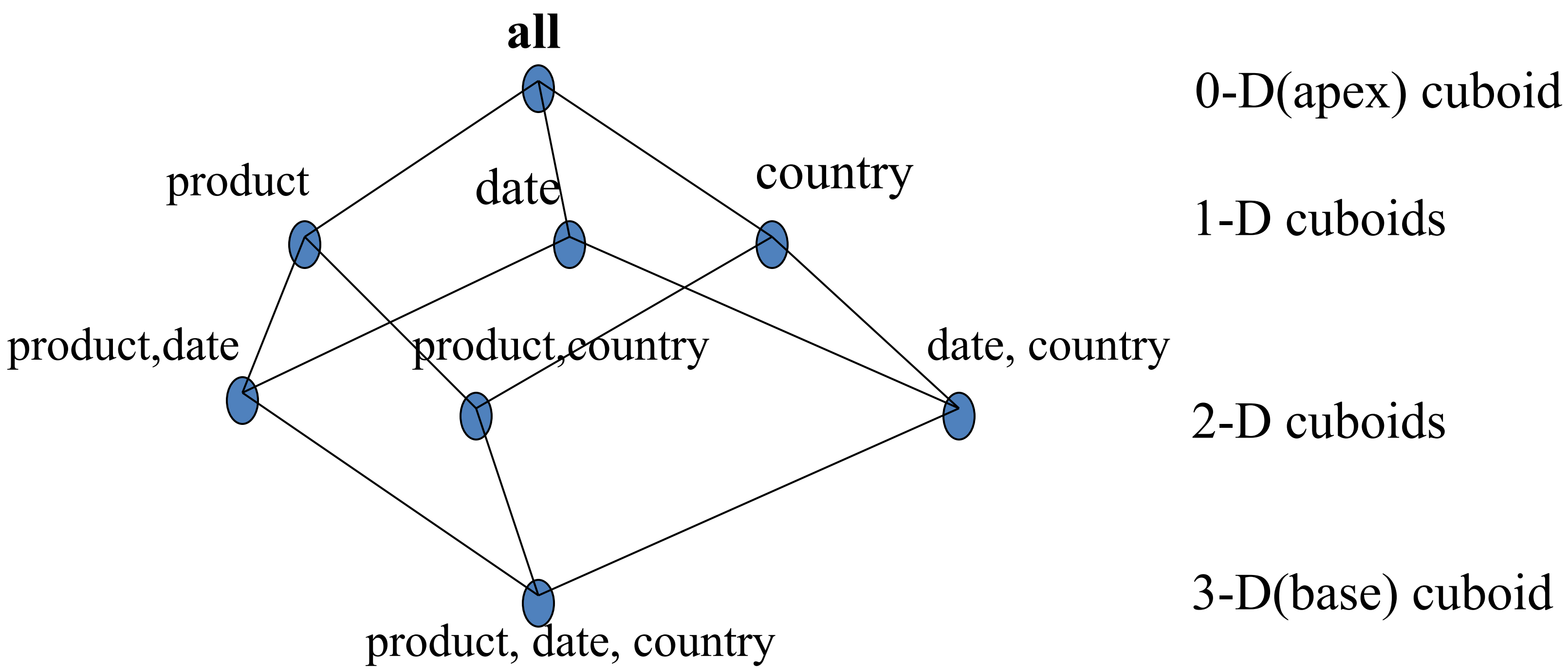
Dimensions: Product, Location, Time
Hierarchical summarization paths



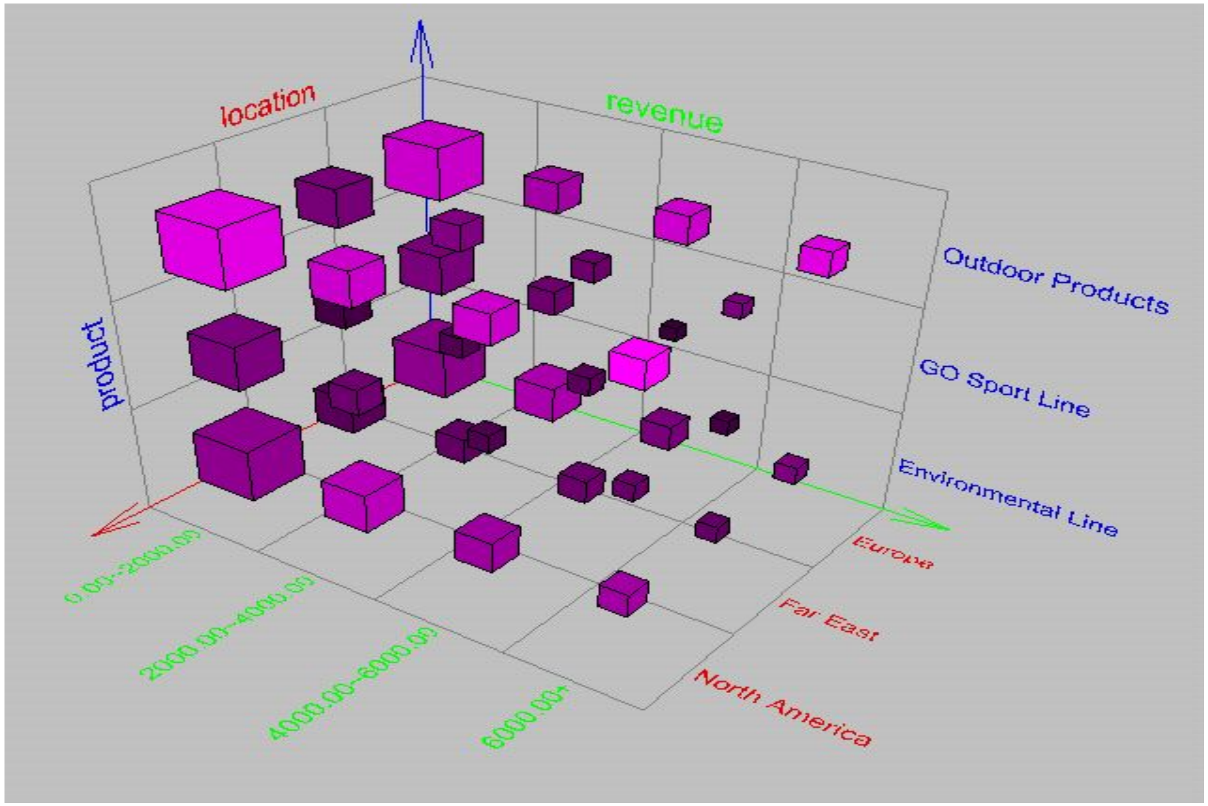
A Sample Data Cube



Cuboids Corresponding to the Cube



Browsing a Data Cube



- Visualization
- OLAP capabilities
- Interactive manipulation

Typical OLAP Operations

- Roll up (drill-up): summarize data
  - *by climbing up hierarchy or by dimension reduction*
- Drill down (roll down): reverse of roll-up
  - *from higher level summary to lower level summary or detailed data, or introducing new dimensions*
- Slice and dice: *project and select*
- Pivot (rotate):
  - *reorient the cube, visualization, 3D to series of 2D planes*
- Other operations
  - *drill across: involving (across) more than one fact table*
  - *drill through: through the bottom level of the cube to its back-end relational tables (using SQL)*

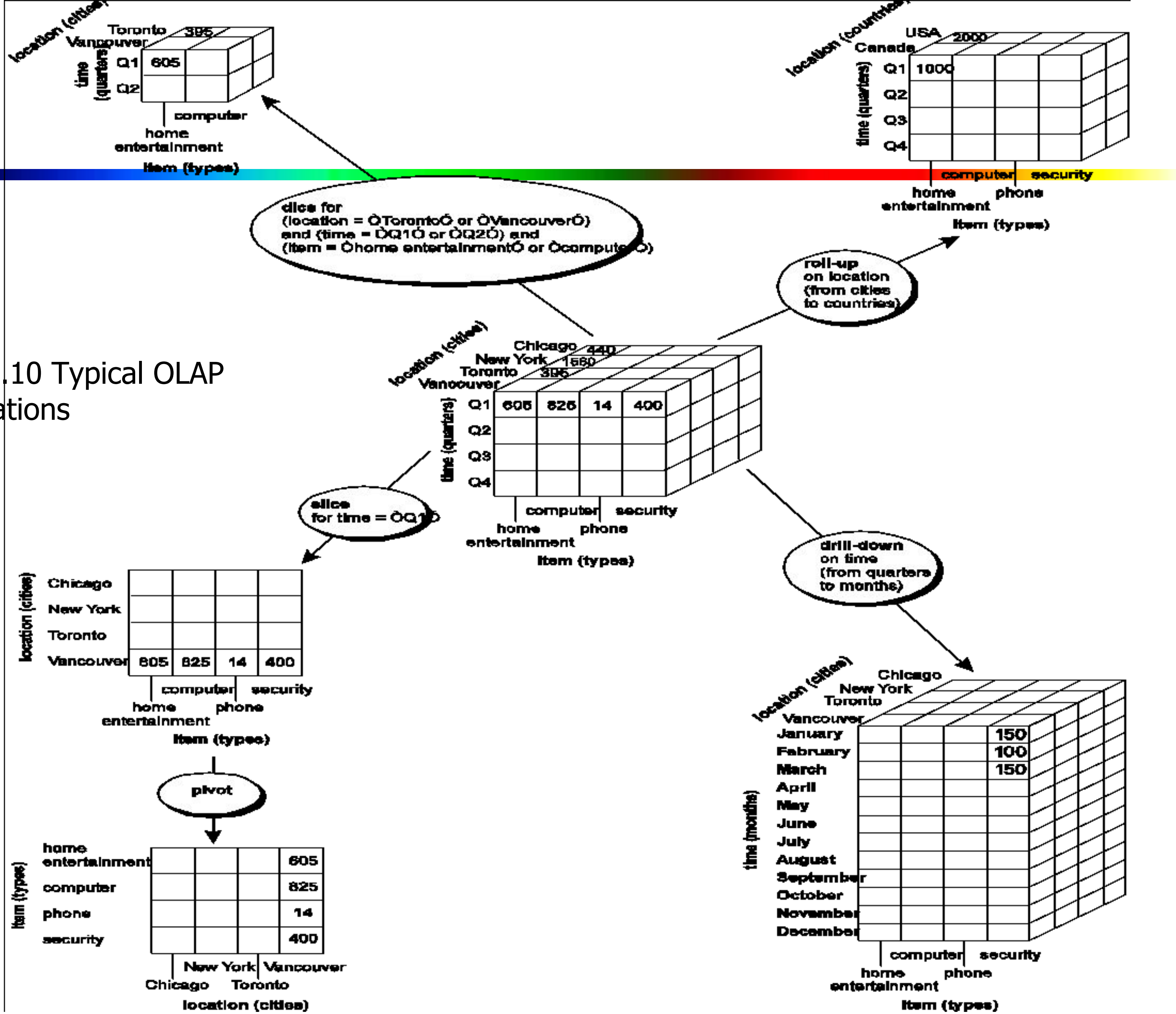


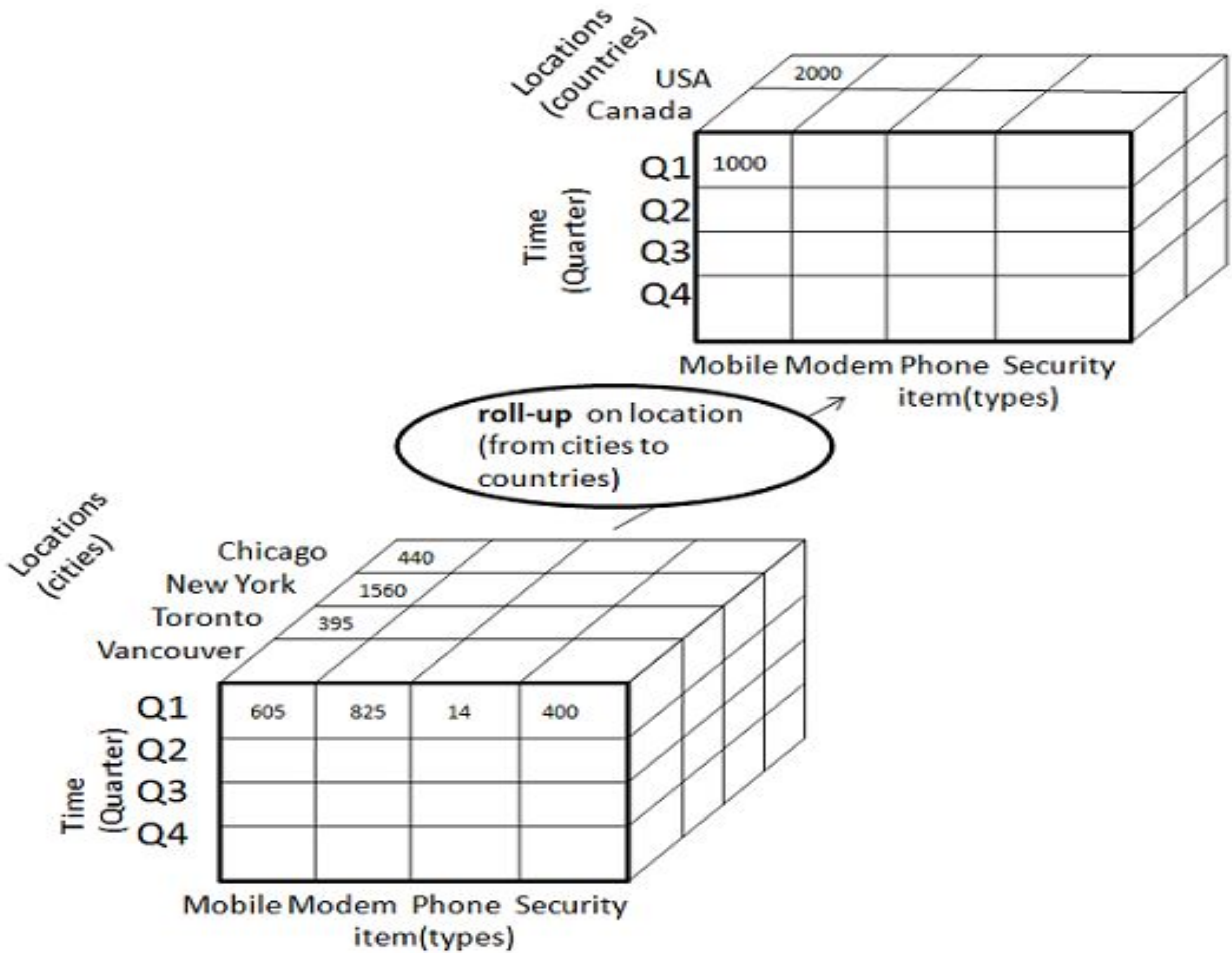
Fig. 3.10 Typical OLAP Operations

CENTRAL CUBE

ROLL UP

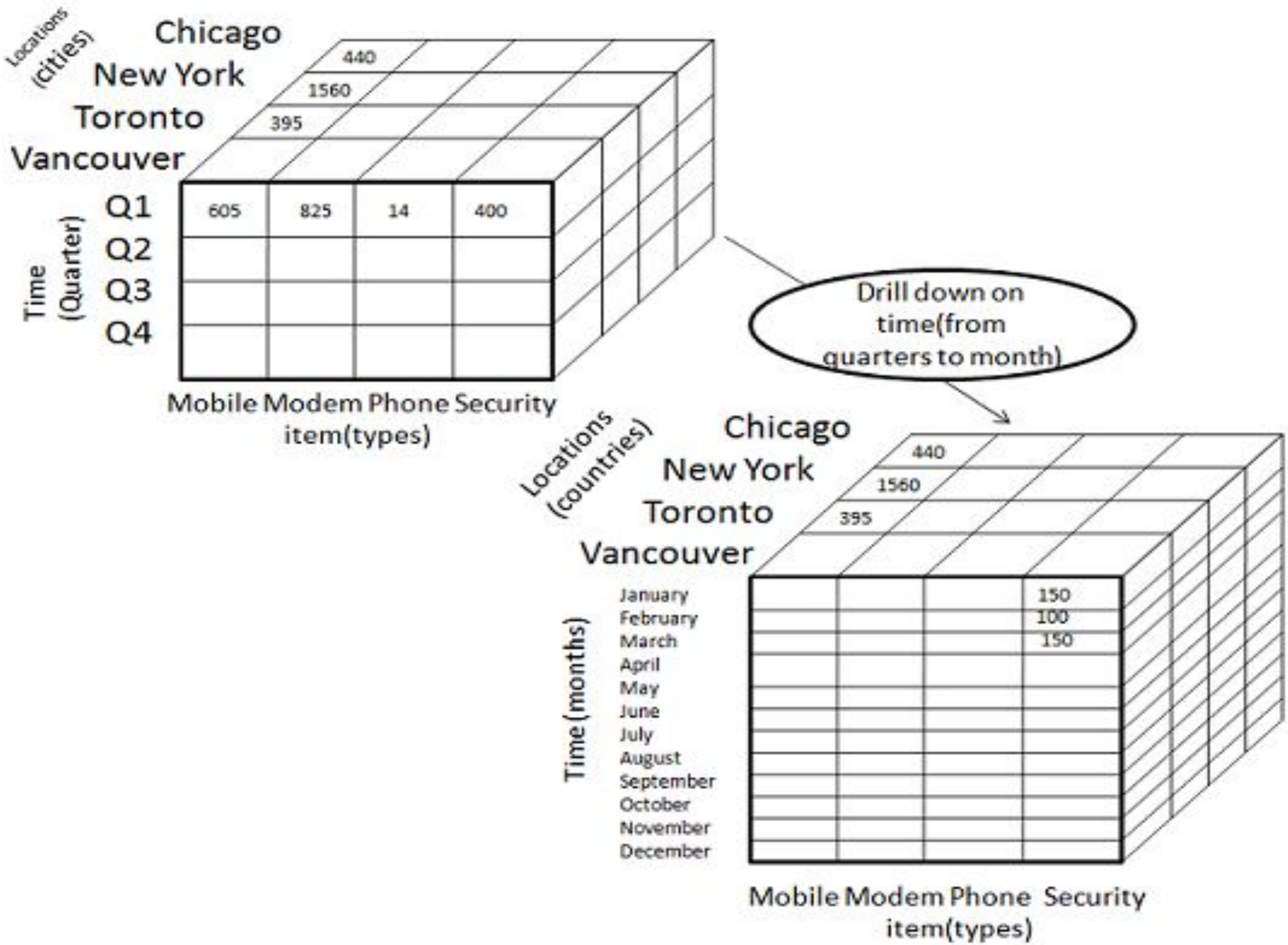
- The roll-up operation (also called the **drill-up** operation by some vendors) performs aggregation on a data cube, either by *climbing up a concept hierarchy* for a dimension or by *dimension reduction*.

- ascending the *location* hierarchy from the level of *city* to the level of *country*.


Locations (countries)
USA
Canada
2000
Q1
Q2
Q3
Q4
Time (Quarter)
1000
Mobile Modem Phone Security
item(types)
roll-up on location (from cities to countries)
Locations (cities)
Chicago
New York
Toronto
Vancouver
440
1560
395
Q1
Q2
Q3
Q4
Time (Quarter)
605
825
14
400
Mobile Modem Phone Security
item(types)

DRILL DOWN

- Drill-down is the ***reverse of roll-up***. It navigates from less detailed data to more detailed data.
- Drill-down can be realized by either *stepping down a concept hierarchy* for a dimension or *introducing additional dimensions*.

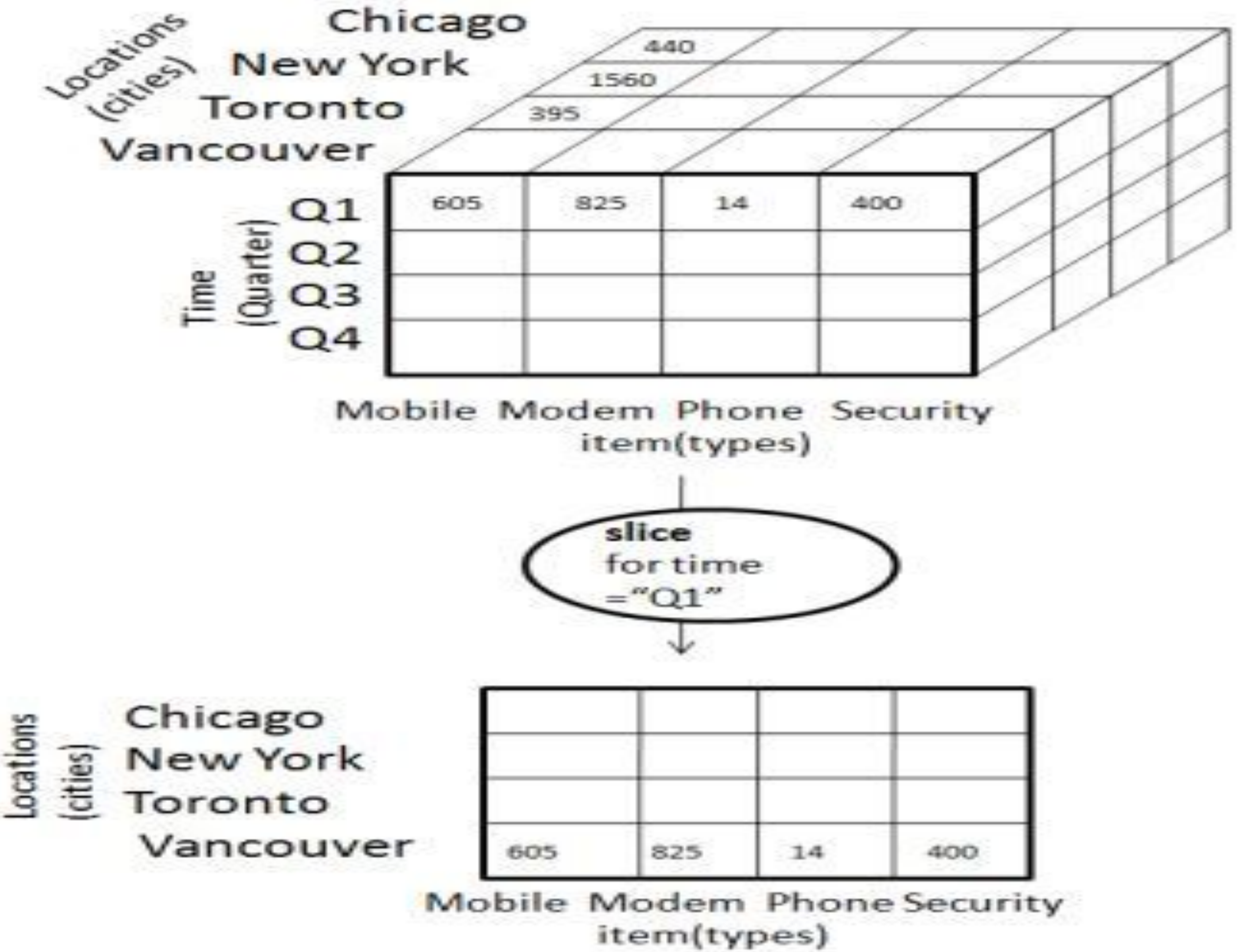


Locations (cities)
Chicago
New York
Toronto
Vancouver
440
1560
395
Time (Quarter)
Q1
Q2
Q3
Q4
605
825
14
400
Mobile Modem Phone Security
item(types)
Drill down on time(from quarters to month)
Locations (countries)
Chicago
New York
Toronto
Vancouver
440
1560
395
Time (months)
January
February
March
April
May
June
July
August
September
October
November
December
150
100
150
Mobile Modem Phone Security
item(types)

SLICING

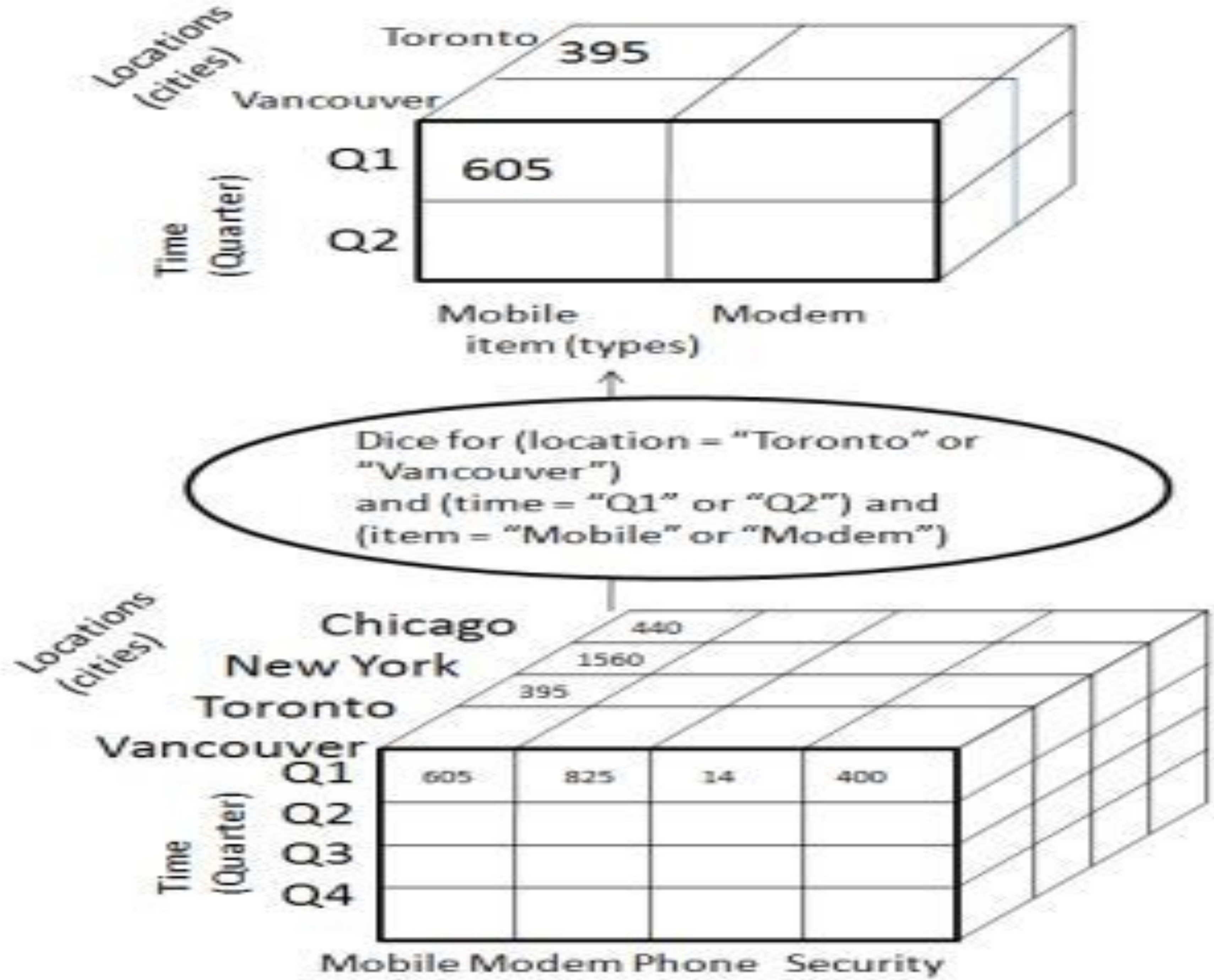
- The *slice* operation performs a selection on one dimension of the given cube, resulting in a sub cube

- the sales data are selected from the central cube for the dimension *time* using the criterion *time =* "Q1."


Locations (cities)
Chicago
New York
Toronto
Vancouver
440
1560
395
Time (Quarter)
Q1
Q2
Q3
Q4
605
825
14
400
Mobile
Modem
Phone
Security
item(types)
slice
for time
="Q1"
Locations (cities)
Chicago
New York
Toronto
Vancouver
605
825
14
400
Mobile
Modem
Phone
Security
item(types)

DICING

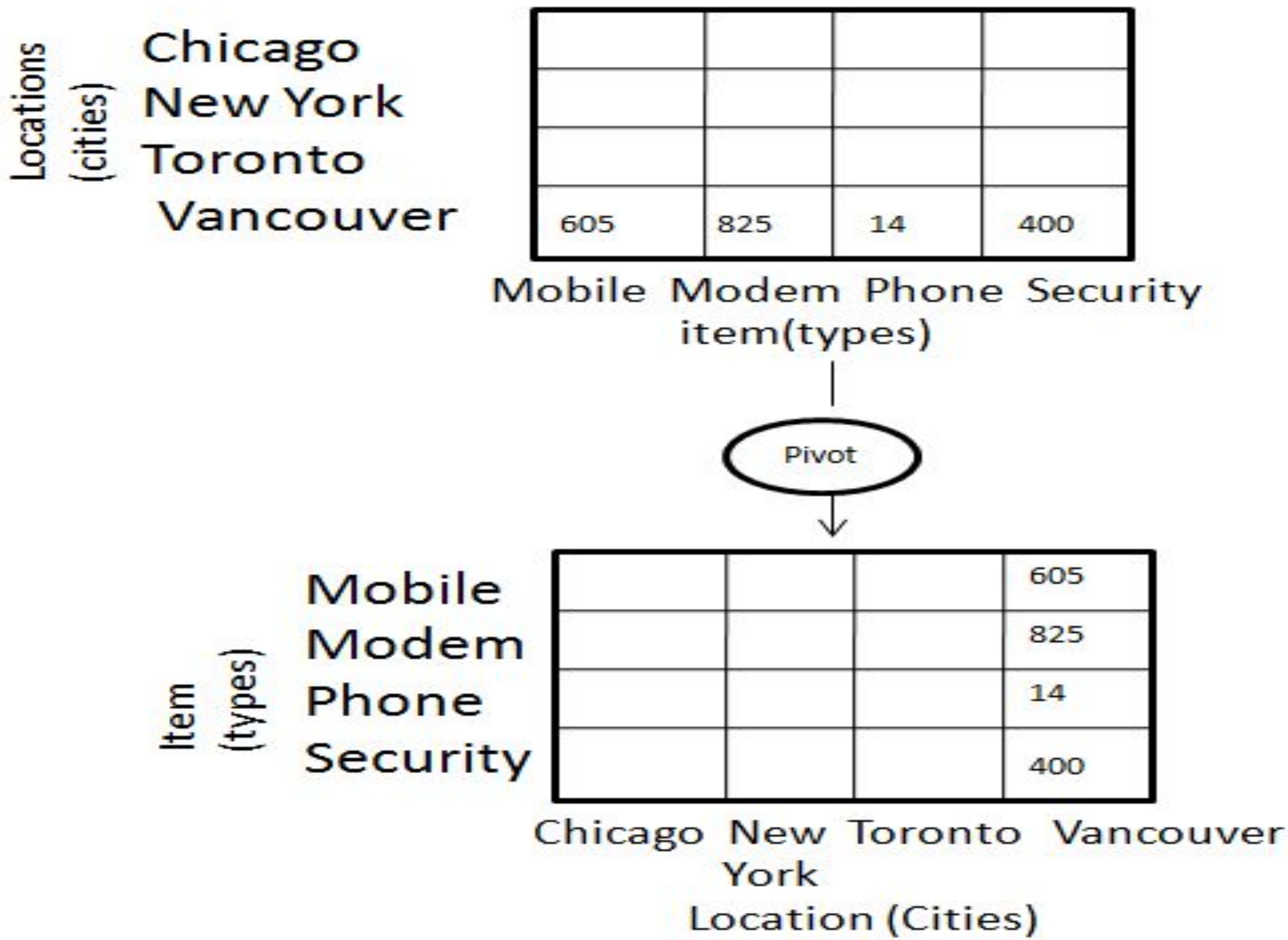
- The *dice* operation defines a sub cube by performing a selection on two or more dimensions.
- On the central cube based on the following selection criteria that involve three dimensions: (*location =* "Toronto" or "Vancouver") and (*time =* "Q1" or "Q2") and (item = "home entertainment" or "computer").


Locations (cities)
Toronto
395
Vancouver
Q1
605
Q2
Time (Quarter)
Mobile
Modem
item (types)
Dice for (location = "Toronto" or "Vancouver")
and (time = "Q1" or "Q2") and
(item = "Mobile" or "Modem")
Locations (cities)
Chicago
440
New York
1560
Toronto
395
Vancouver
Q1
605
825
14
400
Q2
Q3
Q4
Time (Quarter)
Mobile
Modem
Phone
Security

PIVOT

- *Pivot* (also called *rotate*) is a visualization operation that rotates the data axes in view to provide an alternative data presentation.

- a pivot operation where the *item* and *location* axes in a 2-D slice are rotated.

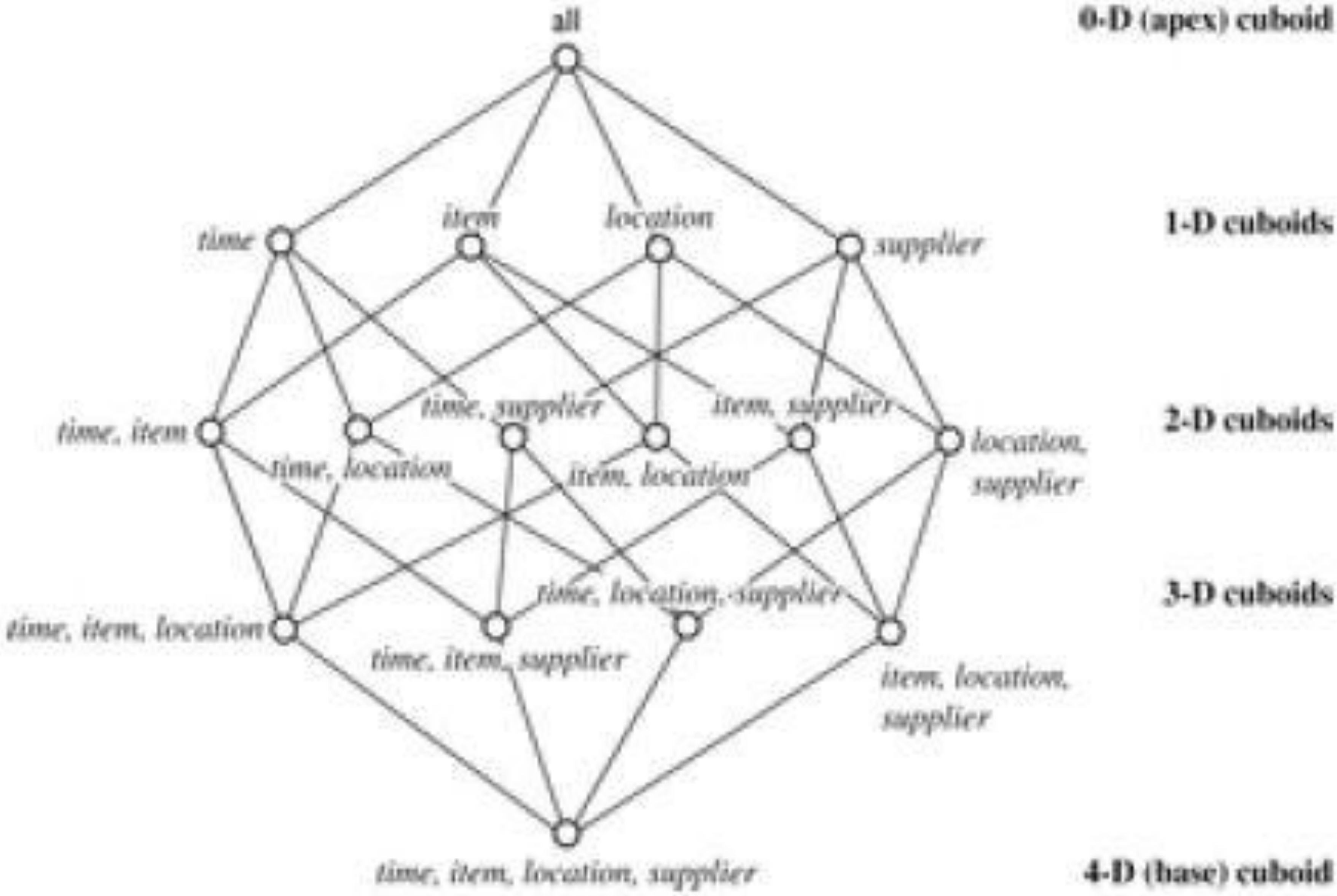

Locations (cities)
Chicago
New York
Toronto
Vancouver
605
825
14
400
Mobile
Modem
Phone
Security
item(types)
Pivot
Item (types)
Mobile
Modem
Phone
Security
605
825
14
400
Chicago
New York
Toronto
Vancouver
Location (Cities)

Other OLAP operations

- **DRILL-ACROSS**
- **DRILL-THROUGH**

- **Drill-across** executes queries involving (i.e., across) more than one fact table.
- **Drill-through** operation uses relational SQL facilities to drill through the bottom level of a data cube down to its back-end relational tables.

4-D DATA CUBE


all
0-D (apex) cuboid
time
item
location
supplier
1-D cuboids
time, item
time, location
time, supplier
item, location
item, supplier
location, supplier
2-D cuboids
time, item, location
time, item, supplier
time, location, supplier
item, location, supplier
3-D cuboids
time, item, location, supplier
4-D (base) cuboid

- The cuboid that holds the lowest level of summarization is called the **base cuboid**.

- The 0-D cuboid, which holds the highest level of summarization, is called the **apex cuboid**

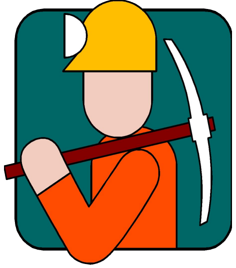
Introduction

- Motivation: Why data mining?
- What is data mining?
- Data Mining: On what kind of data?
- Data mining functionality
- Classification of data mining systems
- Popular data mining algorithms
- Major issues in data mining
- Overview

Why Data Mining?

- The Explosive Growth of Data: from terabytes to petabytes
  - Data collection and data availability
    - Automated data collection tools, database systems, Web, computerized society
  - Major sources of abundant data
    - Business: Web, e-commerce, transactions, stocks, …
    - Science: Remote sensing, bioinformatics, scientific simulation, …
    - Society and everyone: news, digital cameras, YouTube
- <u>We are drowning in data, but starving for knowledge!</u>
- "Necessity is the mother of invention"—Data mining—Automated analysis of massive data sets

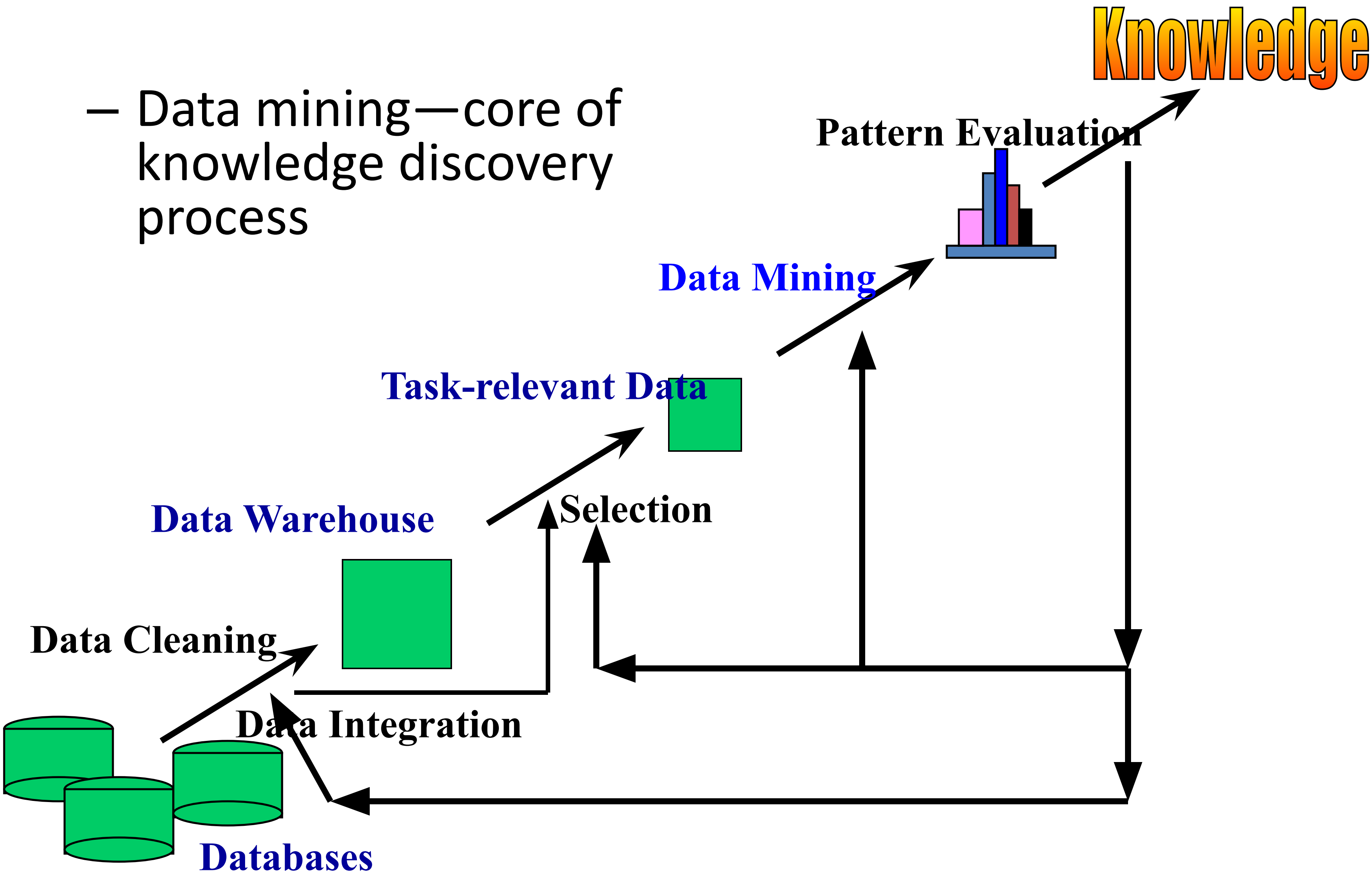
What Is Data Mining?



- Data mining (knowledge discovery from data)
  - Extraction of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data
- Alternative names
  - Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, etc.
- Watch out: Is everything “data mining”?
  - Simple search and query processing
  - (Deductive) expert systems

Knowledge Discovery (KDD) Process

– Data mining—core of knowledge discovery process

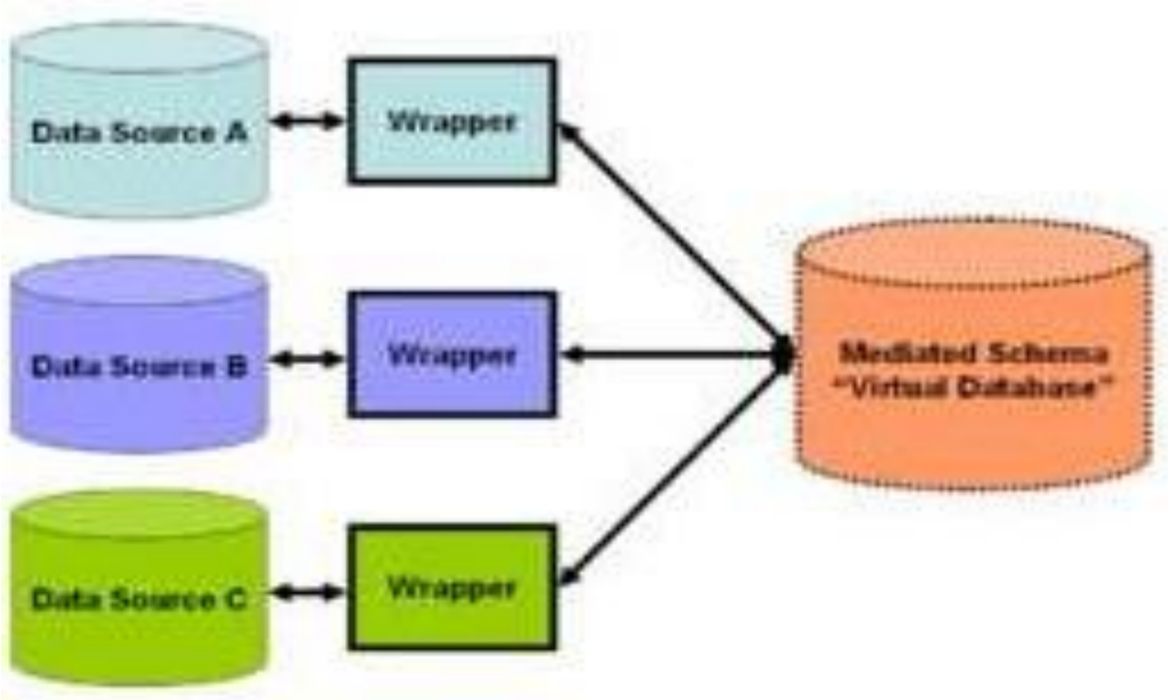


DATA CLEANING

Remove Noise and Inconsistent Data

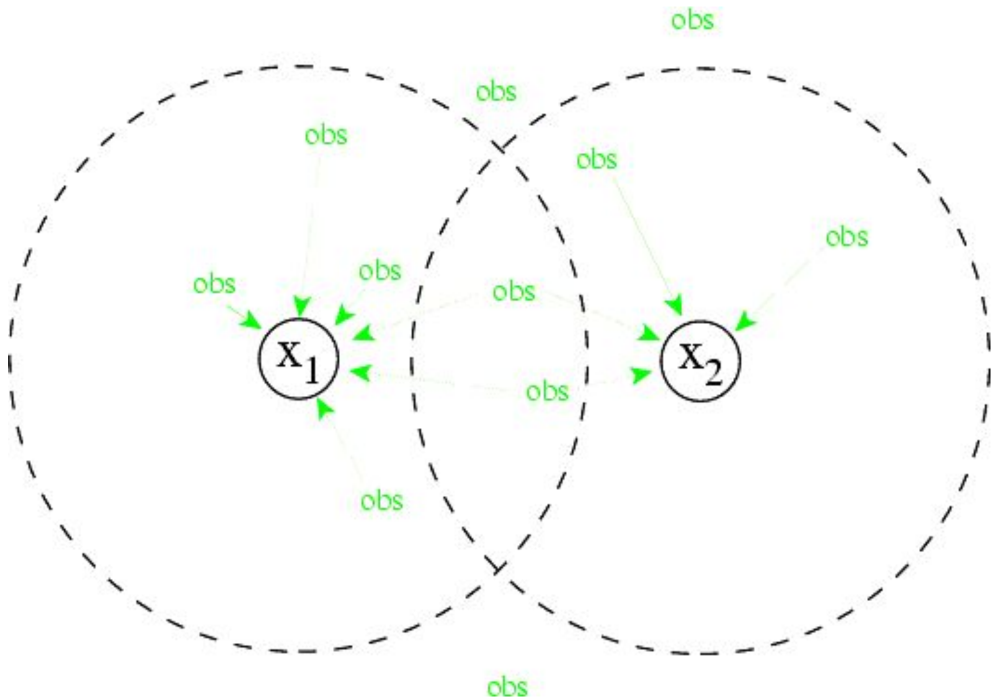
DATA INTEGRATION

- Where multiple data sources may be combined



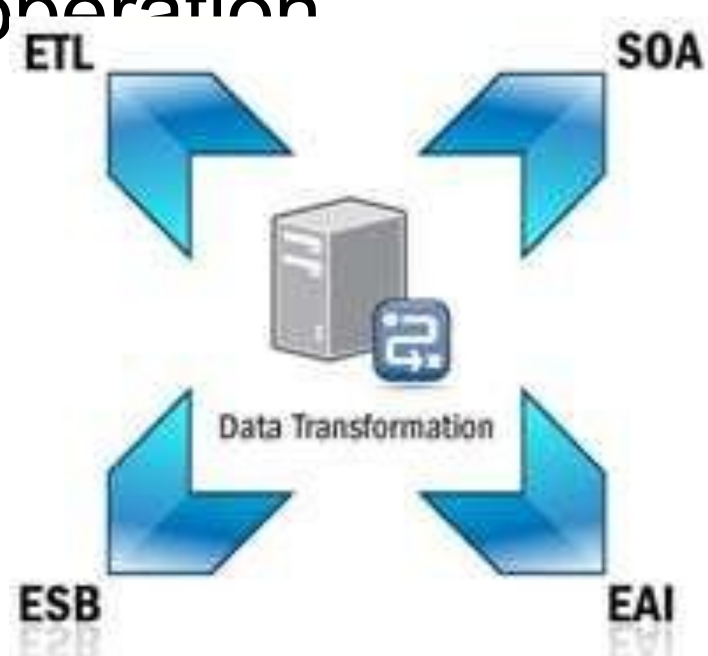
DATA SELECTION

- Where data relevant to the analysis task are retrieved from the data base



DATA TRANSFORMATION

- Where data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operation



Data Mining

- An essential Process where intelligent methods are applied to extract data patterns

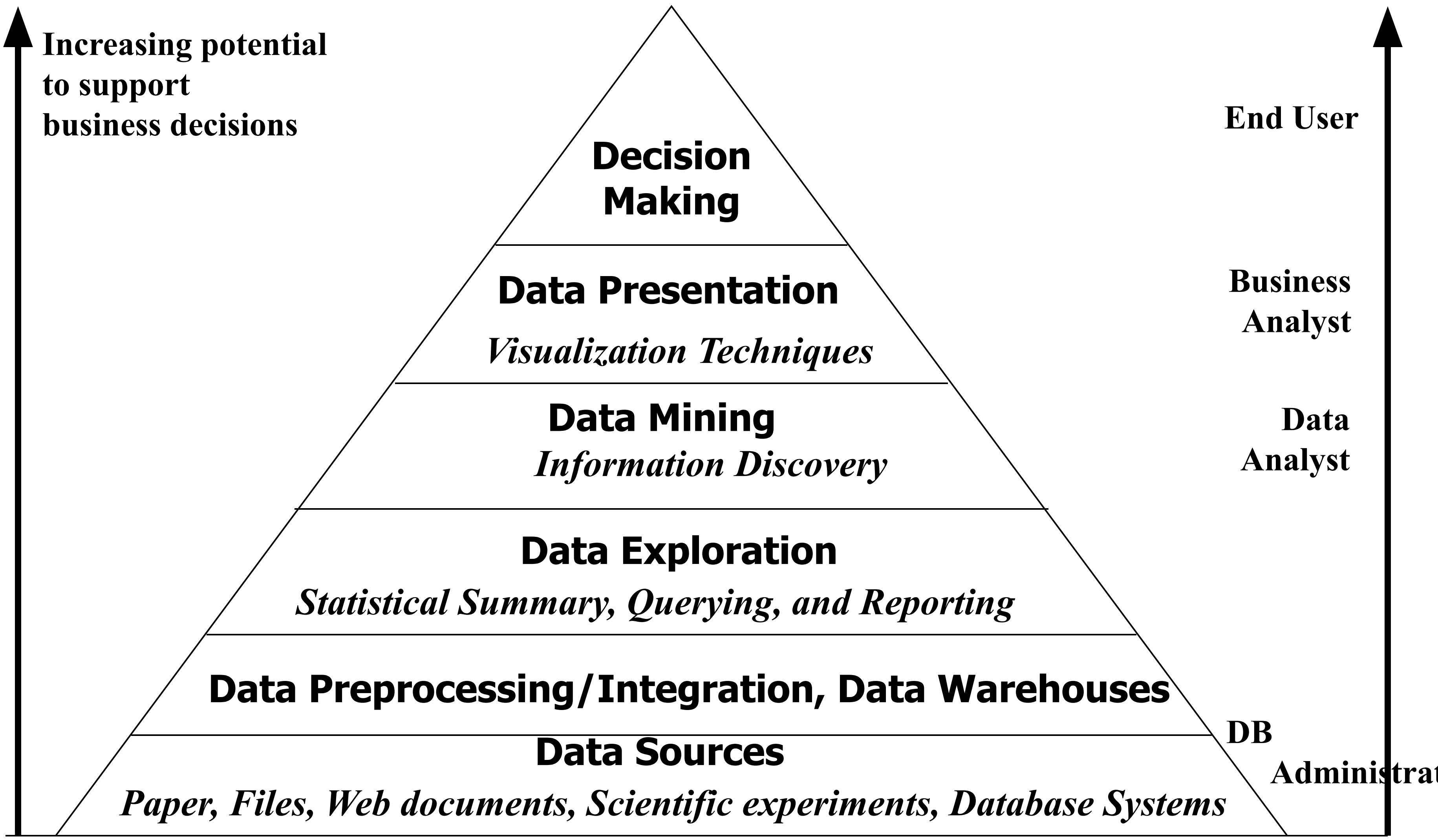
PATTERN EVALUATION

- To identify the truly interesting patterns representing knowledge based on interestingness measures

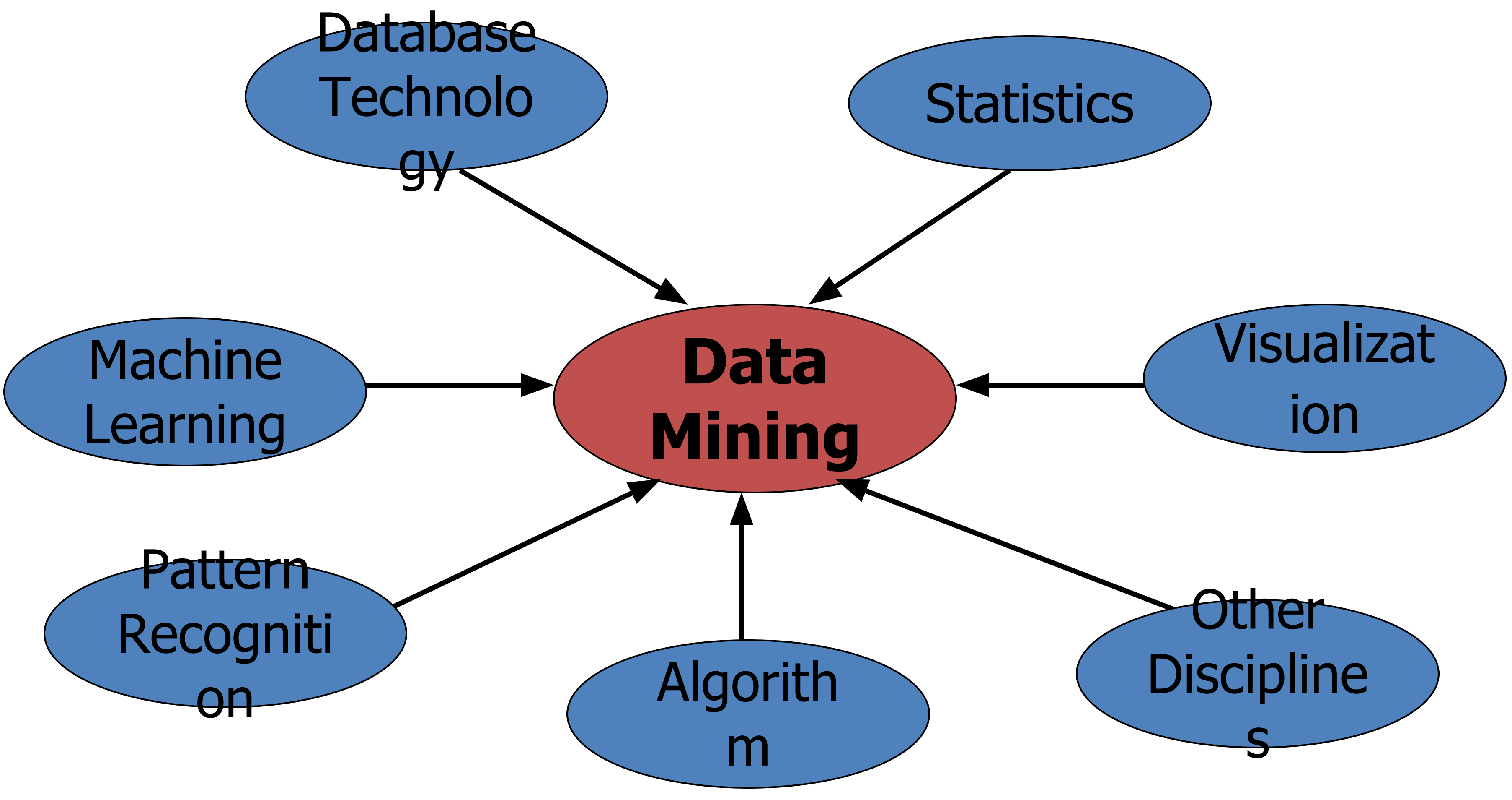
KNOWLEDGE REPRESENTATION

- Where visualization and knowledge representation techniques are used to present mined knowledge to users

Data Mining and Business Intelligence



Data Mining: Confluence of Multiple Disciplines



Why Not Traditional Data Analysis?

- **Tremendous amount of data**
  - Algorithms must be highly scalable to handle such as tera-bytes of data
- **High-dimensionality of data**
  - Micro-array may have tens of thousands of dimensions
- **High complexity of data**
  - Data streams and sensor data
  - Time-series data, temporal data, sequence data
  - Structure data, graphs, social networks and multi-linked data
  - Heterogeneous databases and legacy databases
  - Spatial, spatiotemporal, multimedia, text and Web data
  - Software programs, scientific simulations
- **New and sophisticated applications**

Multi-Dimensional View of Data Mining

- **Data to be mined**
  - Relational, data warehouse, transactional, stream, object-oriented/relational, active, spatial, time-series, text, multi-media, heterogeneous, legacy, WWW
- **Knowledge to be mined**
  - Characterization, discrimination, association, classification, clustering, trend/deviation, outlier analysis, etc.
  - Multiple/integrated functions and mining at multiple levels
- **Techniques utilized**
  - Database-oriented, data warehouse (OLAP), machine learning, statistics, visualization, etc.
- **Applications adapted**
  - Retail, telecommunication, banking, fraud analysis, bio-data mining, stock market analysis, text mining, Web mining, etc.

Data Mining: On What Kinds of Data?

- Database-oriented data sets and applications
  - Relational database, data warehouse, transactional database
- Advanced data sets and advanced applications
  - Data streams and sensor data
  - Time-series data, temporal data, sequence data (incl. bio-sequences)
  - Structure data, graphs, social networks and multi-linked data
  - Object-relational databases
  - Heterogeneous databases and legacy databases
  - Spatial data and spatiotemporal data
  - Multimedia database
  - Text databases
  - The World-Wide Web

Data Mining: Classification Schemes

- General functionality
  - Descriptive data mining
  - Predictive data mining
- Different views lead to different classifications
  - Data view: Kinds of data to be mined
  - Knowledge view: Kinds of knowledge to be discovered
  - Method view: Kinds of techniques utilized
  - Application view: Kinds of applications adapted

Data Mining Functionalities

- Multidimensional concept description: Characterization and discrimination
  - Generalize, summarize, and contrast data characteristics, e.g., dry vs. wet regions
- Frequent patterns, association, correlation vs. causality
  - Beer □ Chips [0.5%, 75%]
- Classification and prediction
  - Construct models (functions) that describe and distinguish classes or concepts for future prediction
    - E.g., classify countries based on (climate), or classify cars based on (gas mileage)
  - Predict some unknown or missing numerical values

Data Mining Functionalities (2)

- **Cluster analysis**
  - Class label is unknown: Group data to form new classes, e.g., cluster houses to find distribution patterns
  - Maximizing intra-class similarity & minimizing interclass similarity
- **Outlier analysis**
  - Outlier: Data object that does not comply with the general behavior of the data
  - Noise or exception? Useful in fraud detection, rare events analysis
- **Trend and evolution analysis**
  - Trend and deviation: e.g., regression analysis
  - Sequential pattern mining: e.g., digital camera large SD memory
  - Periodicity analysis
  - Similarity-based analysis

Supervised vs. Unsupervised Learning

- Supervised learning (classification)
  - Supervision: The training data (observations, measurements, etc.) are accompanied by **labels** indicating the class of the observations
  - New data is classified based on the training set
- Unsupervised learning (clustering)
  - The class labels of training data is unknown
  - Given a set of measurements, observations, etc. with the aim of establishing the existence of classes or clusters in the data

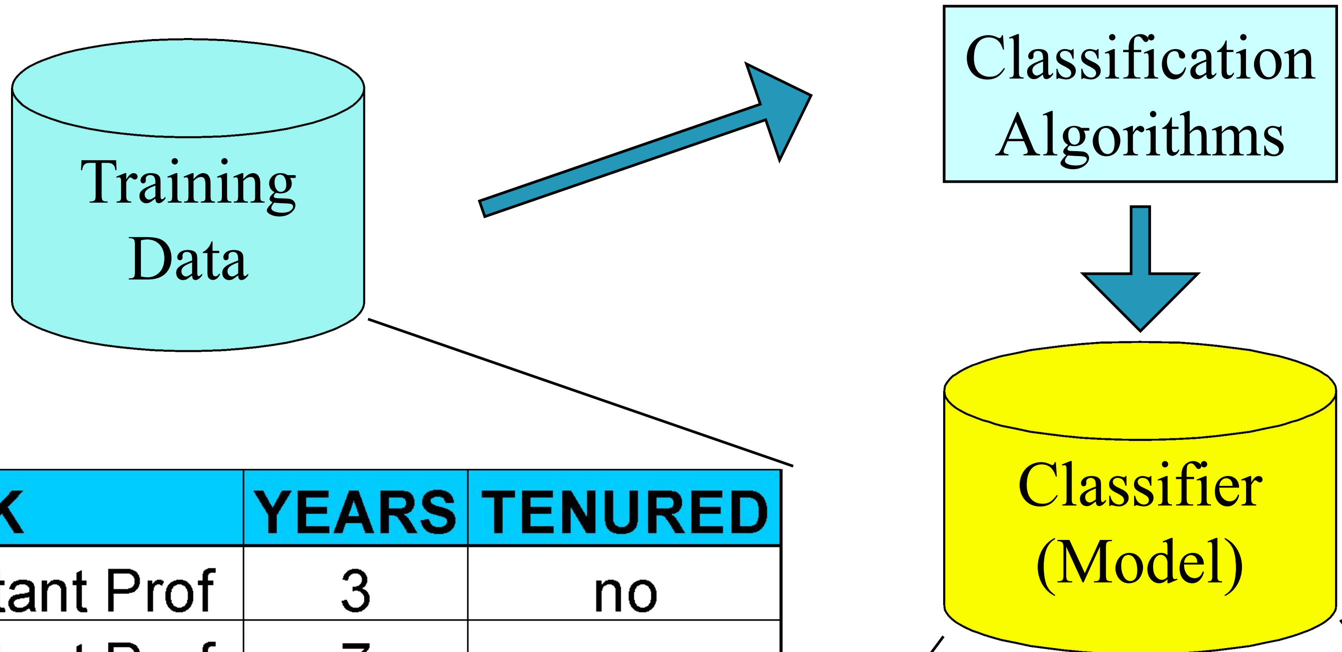
Prediction Problems: Classification vs. Numeric Prediction

- Classification
  - predicts categorical class labels (discrete or nominal)
  - classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data
- Numeric Prediction
  - models continuous-valued functions, i.e., predicts unknown or missing values
- Typical applications
  - Credit/loan approval:
  - Medical diagnosis: if a tumor is cancerous or benign
  - Fraud detection: if a transaction is fraudulent
  - Web page categorization: which category it is

Classification—A Two-Step Process

- Model construction: describing a set of predetermined classes
  - Each tuple/sample is assumed to belong to a predefined class, as determined by the class label attribute
  - The set of tuples used for model construction is training set
  - The model is represented as classification rules, decision trees, or mathematical formulae
- Model usage: for classifying future or unknown objects
  - Estimate accuracy of the model
    - The known label of test sample is compared with the classified result from the model
    - Accuracy rate is the percentage of test set samples that are correctly classified by the model
    - Test set is independent of training set (otherwise overfitting)
  - If the accuracy is acceptable, use the model to classify new data
- Note: If *the test set* is used to select models, it is called validation (test) set

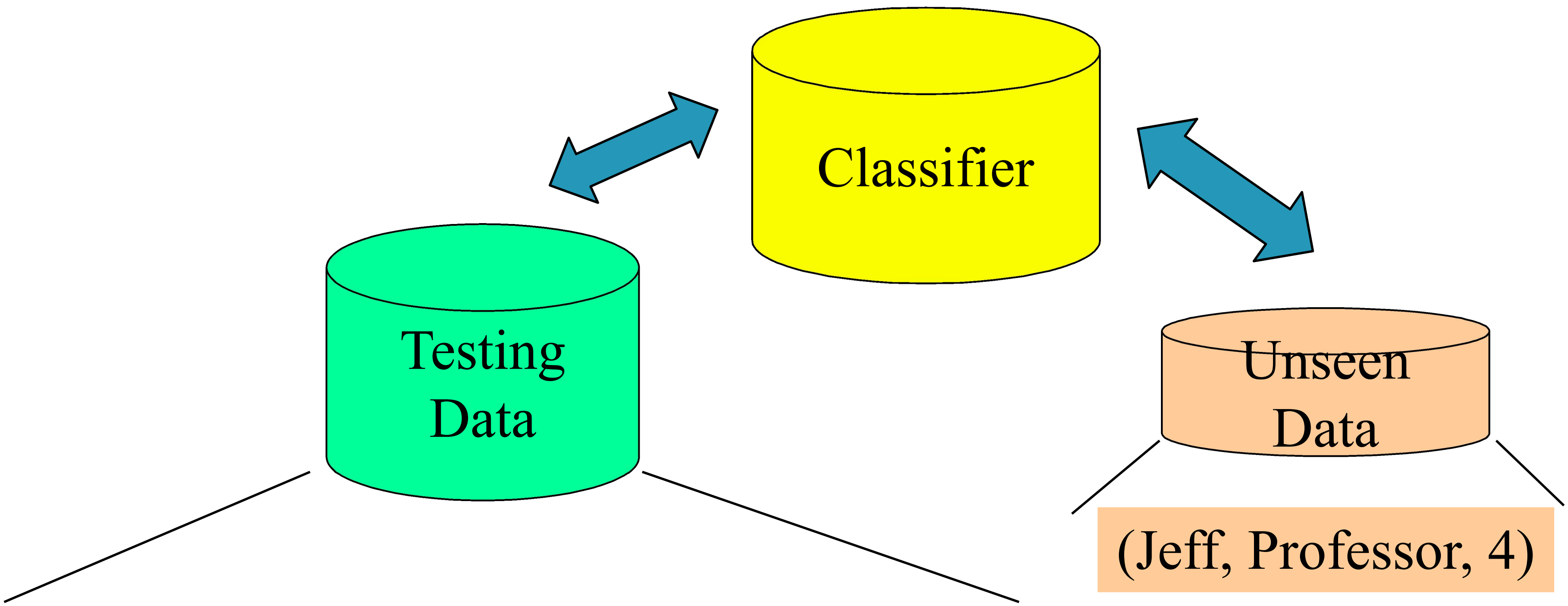
Process (1): Model Construction



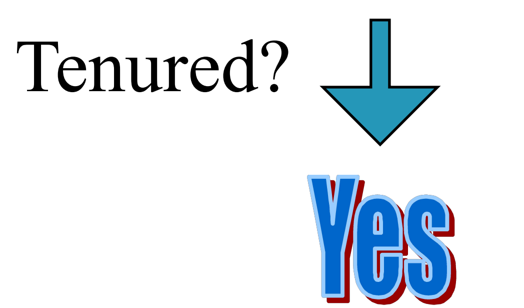
NAME	RANK	YEARS	TENURED
Mike	Assistant Prof	3	no
Mary	Assistant Prof	7	yes
Bill	Professor	2	yes
Jim	Associate Prof	7	yes
Dave	Assistant Prof	6	no
Anne	Associate Prof	3	no

IF rank = ‘professor’
OR years > 6
THEN tenured = ‘yes’

Process (2): Using the Model in Prediction



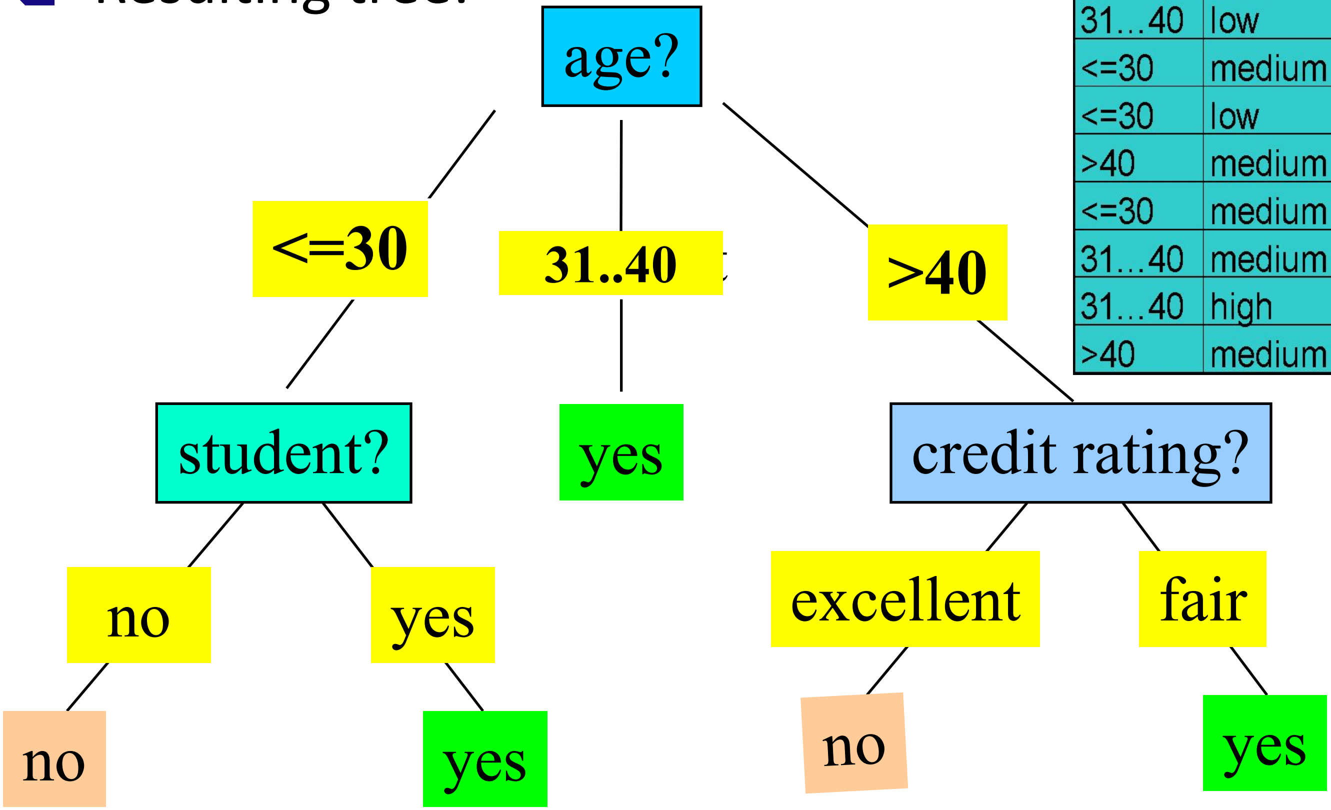
NAME	RANK	YEARS	TENURED
Tom	Assistant Prof	2	no
Merlisa	Associate Prof	7	no
George	Professor	5	yes
Joseph	Assistant Prof	7	yes



Decision Tree Induction: An Example

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

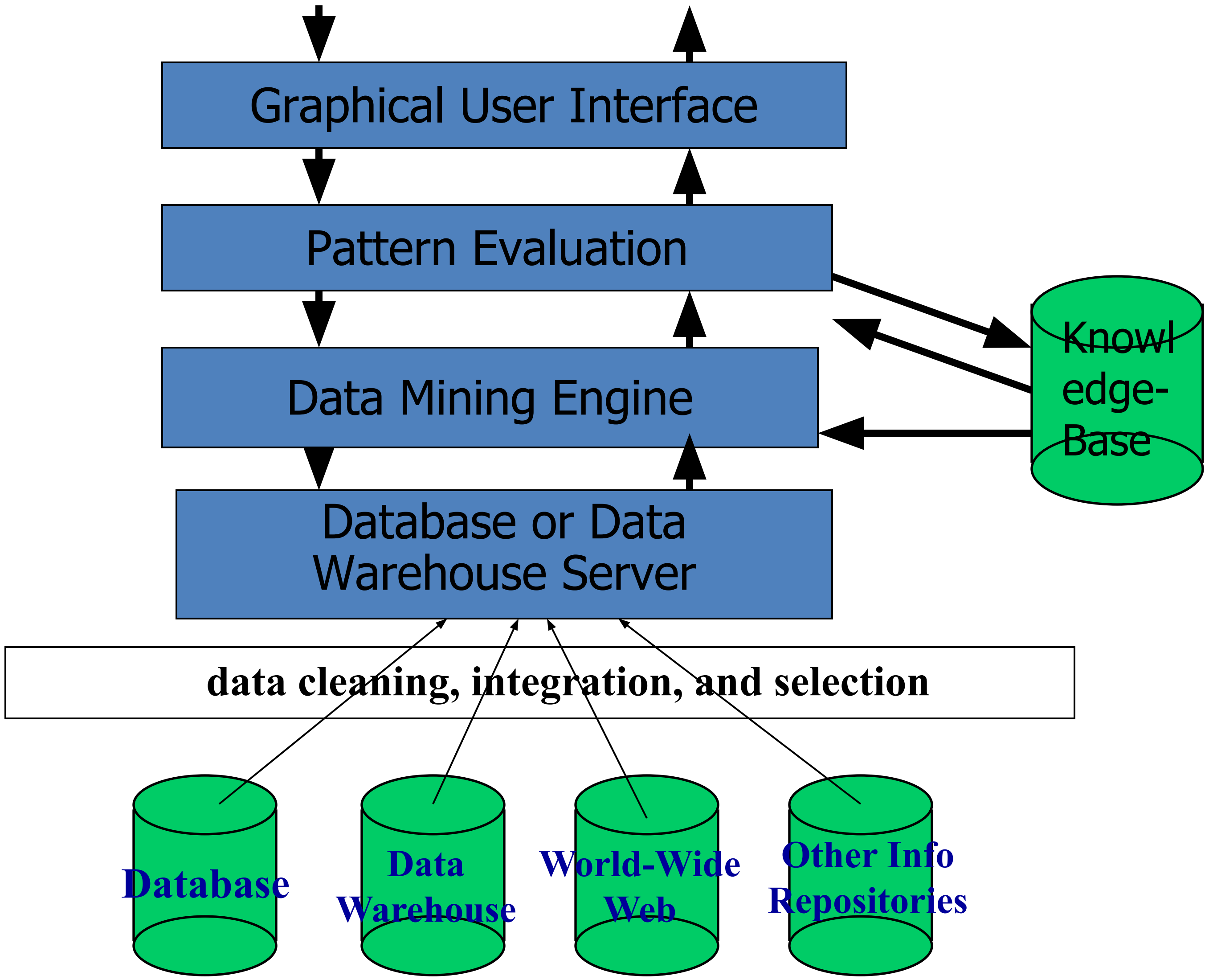
- Training data set: Buys_computer
- The data set follows an example of Quinlan's ID3 (Playing Tennis)
- Resulting tree:



What is Cluster Analysis?

- Cluster: A collection of data objects
  - similar (or related) to one another within the same group
  - dissimilar (or unrelated) to the objects in other groups
- Cluster analysis (or *clustering*, *data segmentation, …*)
  - Finding similarities between data according to the characteristics found in the data and grouping similar data objects into clusters
- Unsupervised learning: no predefined classes (i.e., *learning by observations* vs. learning by examples: supervised)
- Typical applications
  - As a stand-alone tool to get insight into data distribution
  - As a preprocessing step for other algorithms

Architecture: Typical Data Mining System



Major Issues in Data Mining

- Mining methodology
  - Mining different kinds of knowledge from diverse data types, e.g., bio, stream, Web
  - Performance: efficiency, effectiveness, and scalability
  - Pattern evaluation: the interestingness problem
  - Incorporation of background knowledge
  - Handling noise and incomplete data
  - Parallel, distributed and incremental mining methods
  - Integration of the discovered knowledge with existing one: knowledge fusion
- User interaction
  - Data mining query languages and ad-hoc mining
  - Expression and visualization of data mining results
  - Interactive mining of knowledge at multiple levels of abstraction
- Applications and social impacts
  - Protection of data security, integrity, and privacy

Summary

- Data mining: Discovering interesting patterns from large amounts of data
- A natural evolution of database technology, in great demand, with wide applications
- A KDD process includes data cleaning, data integration, data selection, transformation, data mining, pattern evaluation, and knowledge presentation
- Mining can be performed in a variety of information repositories
- Data mining functionalities: characterization, discrimination, association, classification, clustering, outlier and trend analysis, etc.
- Data mining systems and architectures
- Major issues in data mining

Classification vs. Prediction

- **Classification**
  - predicts categorical class labels (discrete or nominal)
  - classifies data (constructs a model) based on the training set and the values (class labels) in a classifying attribute and uses it in classifying new data
- **Prediction**
  - models continuous-valued functions, i.e., predicts unknown or missing values
- **Typical applications**
  - Credit approval
  - Target marketing
  - Medical diagnosis
  - Fraud detection

What Is Prediction?

- (Numerical) prediction is similar to classification
  - construct a model
  - use model to predict continuous or ordered value for a given input
- Prediction is different from classification
  - Classification refers to predict categorical class label
  - Prediction models continuous-valued functions
- Major method for prediction: regression
  - model the relationship between one or more *independent* or **predictor** variables and a *dependent* or **response** variable
- Regression analysis
  - Linear and multiple regression
  - Non-linear regression
  - Other regression methods: generalized linear model, Poisson regression, log-linear models, regression trees

Clustering

Clustering: Rich Applications and Multidisciplinary Efforts

- Pattern Recognition
- Spatial Data Analysis
  - Create thematic maps by clustering feature spaces
  - Detect spatial clusters or for other spatial mining tasks
- Image Processing
- Economic Science (especially market research)
- WWW
  - Document classification
  - Cluster Weblog data to discover groups of similar access patterns

Examples of Clustering Applications

- Marketing: Help marketers discover distinct groups in their customer bases, and then use this knowledge to develop targeted marketing programs
- Land use: Identification of areas of similar land use in an earth observation database
- City-planning: Identifying groups of houses according to their house type, value, and geographical location
- Earth-quake studies: Observed earth quake epicenters should be clustered along continent faults

Quality: What Is Good Clustering?

- A <u>good clustering</u> method will produce high quality clusters with
  - high <u>intra-class</u> similarity
  - low <u>inter-class</u> similarity
- The <u>quality</u> of a clustering result depends on both the similarity measure used by the method and its implementation
- The <u>quality</u> of a clustering method is also measured by its ability to discover some or all of the <u>hidden</u> patterns

Measure the Quality of Clustering

- Dissimilarity/Similarity metric: Similarity is expressed in terms of a distance function, typically metric: $d(i, j)$
- There is a separate "quality" function that measures the "goodness" of a cluster.
- The definitions of distance functions are usually very different for interval-scaled, boolean, categorical, ordinal ratio, and vector variables.
- Weights should be associated with different variables based on applications and data semantics.
- It is hard to define "similar enough" or "good enough"
  - the answer is typically highly subjective.

Major Clustering Approaches (I)

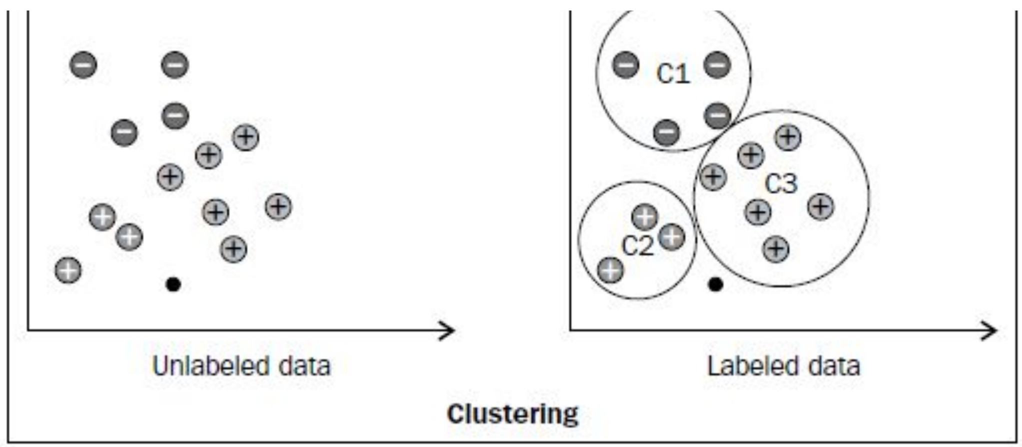
- Partitioning approach:
  - Construct various partitions and then evaluate them by some criterion, e.g., minimizing the sum of square errors
  - Typical methods: k-means, k-medoids, CLARANS
- Hierarchical approach:
  - Create a hierarchical decomposition of the set of data (or objects) using some criterion
  - Typical methods: Diana, Agnes, BIRCH, ROCK, CAMELEON
- Density-based approach:
  - Based on connectivity and density functions
  - Typical methods: DBSACN, OPTICS, DenClue

Major Clustering Approaches (II)

- Model-based:
  - A model is hypothesized for each of the clusters and tries to find the best fit of that model to each other
  - Typical methods: EM, SOM, COBWEB
- Frequent pattern-based:
  - Based on the analysis of frequent patterns
  - Typical methods: pCluster
- User-guided or constraint-based:
  - Clustering by considering user-specified or application-specific constraints
  - Typical methods: COD (obstacles), constrained clustering

Clustering

Clustering is the task of grouping a set of object in such a way that similar objects with similar characteristics are grouped in the same category, but other objects are grouped in other categories. In clustering, the input datasets are not labeled; they need to be labeled based on the similarity of their data structure.



Clustering Algorithms : K-means,k-medoid, hierarchy & density based clustering.

Applications of clustering

1. Market segmentation
2. Social network analysis
3. Organizing computer network
4. Astronomical data analysis

3.Recommendation Algorithms

- A machine-learning technique to predict what new items a user would like based on associations with the user's previous items

- When a customer is looking for a Samsung Galaxy S5 mobile phone on Amazon, the store will also suggest other mobile phones similar to this one, presented in the **Customers Who Bought This Item Also Bought window.**

Types of Recommendations

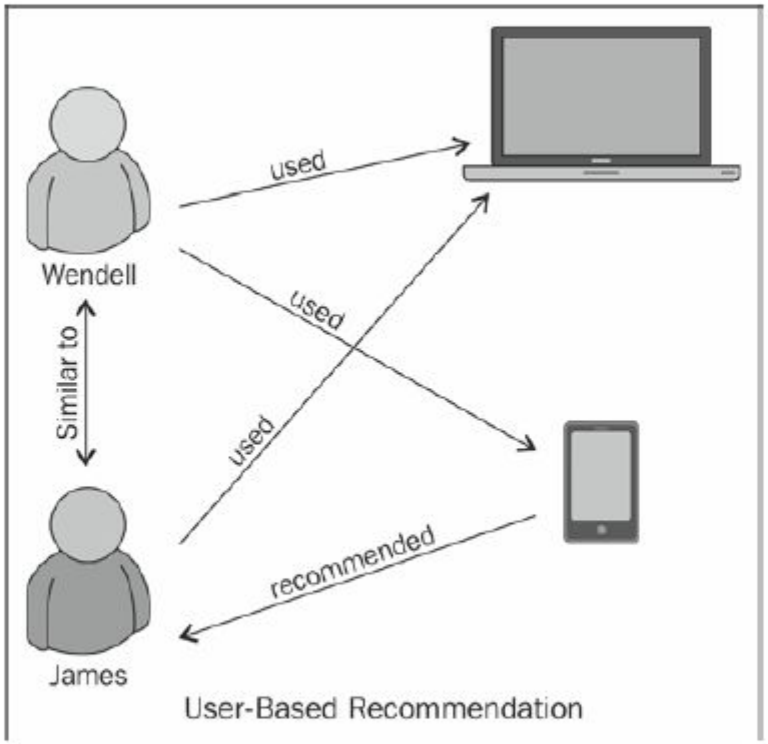
- 1.User Based Recommendation

- 2.Item Based Recommendation

User Based Recommendation

Users similar to the current user are determined

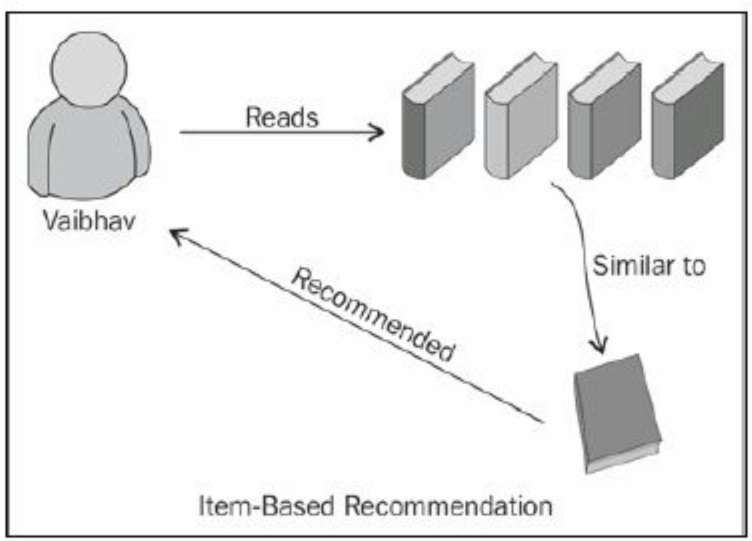
Based on smilarity their liked/used product can be recomended



Item Based Recommendation

- items similar to the items that are being currently used by a user are determined

- Eg:



Steps in R to genearate recommendations

To generate recommendations for users, we need to have datasets in a special format that can be read by the algorithm. Here, we will use the collaborative filtering algorithm for generating the recommendations rather than content-based algorithms.

Recommendations can be derived from the matrix-factorization technique as follows:

```
Co-occurrence matrix * scoring matrix = Recommended Results
```

To generate the recommenders, we will follow the given steps:

1. Computing the co-occurrence matrix.
2. Establishing the user-scoring matrix.
3. Generating recommendations.

Applications /Uses of recommendations

- E- commerce
- Increasing the sales and growing the business
- Customer satisfaction

Bussiness Intelligence

Changing Business Environment

- **The environment in which organizations operate today is becoming more and more complex**
- **The complexity creates opportunities on one hand and problems on the other.**
- **Business environment factors are divided into four major categories:**
  - **markets,**
  - **consumer demands,**
  - **technology,**
  - Societal
- **The intensity of these factors increases with time, hence more pressures, more competition, more management problems**

Business Environment Factors

FACTOR	DESCRIPTION
Markets	Strong competition
	Expanding global markets
	Blooming electronic markets on the Internet
	Innovative marketing methods
	Opportunities for outsourcing with IT support
	Need for real-time, on-demand transactions
Consumer demand	Desire for customization
	Desire for quality, diversity of products, and speed of delivery
	Customers getting powerful and less loyal
Technology	More innovations, new products, and new services
	Increasing obsolescence rate
	Increasing information overload
	Social networking, Web 2.0 and beyond
Societal	Growing government regulations and deregulation
	Workforce more diversified, older, and composed of more women
	Prime concerns of homeland security and terrorist attacks
	Increasing social responsibility of companies
	Greater emphasis on sustainability

Decision Making in Business

- **Management ≅ Decision Making**
- **Decision making means selecting the best solution from two or more alternatives**
- **Management was considered an art because a variety of individual styles could be used in addressing problems**
- Often based on creativity, judgment, intuition, experience rather than on a scientific approach.
- **Studies suggest that managers roles can be classified into 3 major categories:**
  - **Interpersonal – figurehead, leader**
  - **Informational- spokesperson, disseminator**
  - **Decisional- negotiator, resource allocator**

The idea

The right decision = Intelligence + Information
Intelligence = The capacity to acquire and apply knowledge
Information = is used to tell stories, to discover things, to keep track of things, to provide answer and eventually will lead to innovation

Business Intelligence =
The right information + The right time + From the Right Resources

Using information effectively to make better decisions
(Gautner, 1989)

What is Business Intelligence?

- Business Intelligence (**BI**) refers to computer-based techniques used in spotting, digging-out, and analyzing business data, such as sales revenue by products and/or departments or associated costs and incomes

  (Wikipedia,2010)

- Business Intelligence (BI) helps business people make more informed decisions by providing them timely, data-driven answers to their business questions. BI analyzes data stored in data warehouses, operational databases, and/or ERP systems (i.e. SAP®, Oracle, JD Edwards, Peoplesoft) and transforms it into attractive and easy to understand dashboards and reports. BI delivers the insight needed to make strategic planning decisions, improve operational efficiencies, and optimize business processes.

  (Microstrategy.com)

A What is Business Intelligence?

- An umbrella term that combines architectures, tools, databases, applications and methodologies in order to enable interactive access to data, to enable manipulation of data and to give business managers the ability to make more informed and better business decisions

(Turban, 2010)

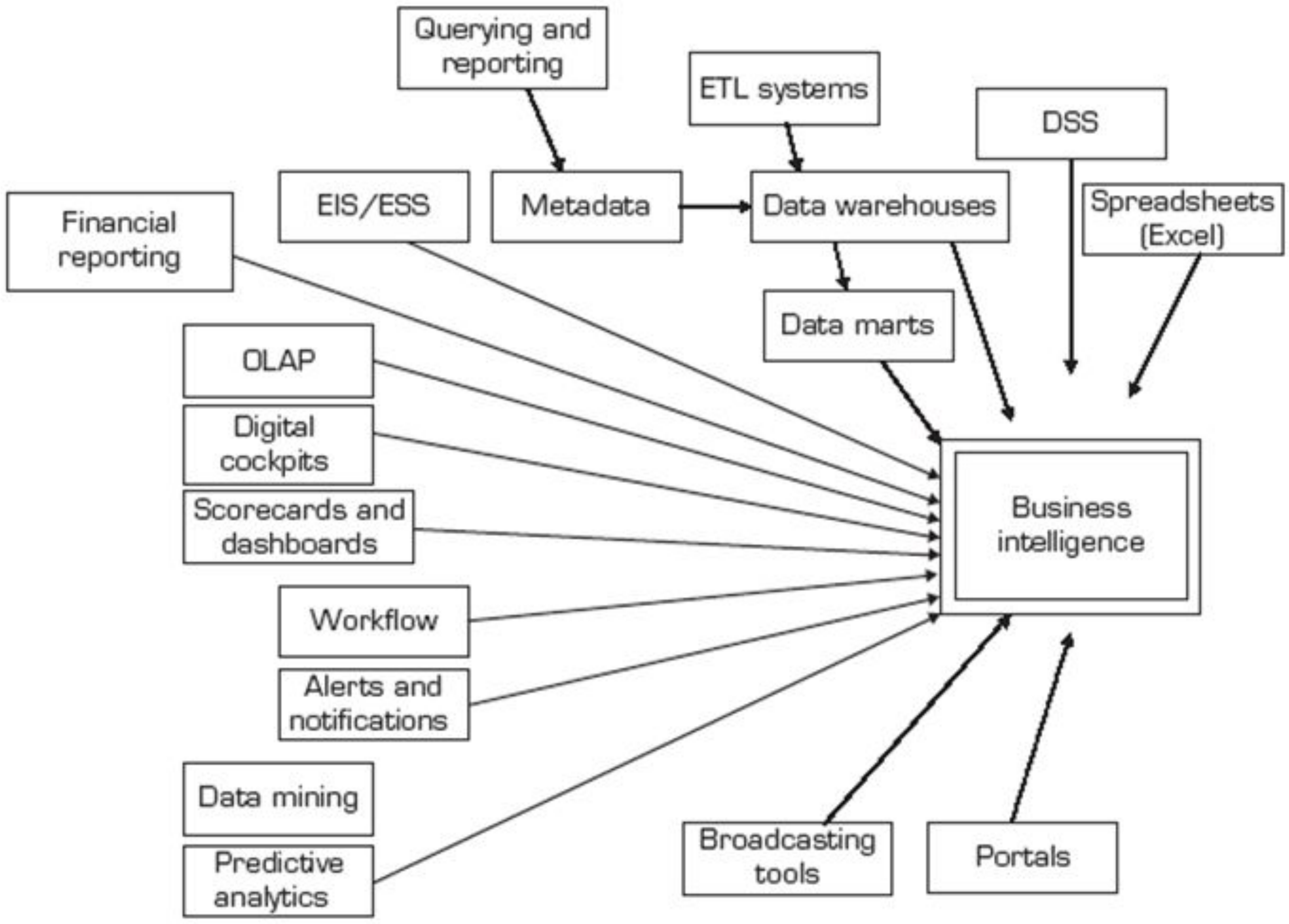
- Business intelligence uses knowledge management, data warehouse[ing], data mining and business analysis to identify, track and improve key processes and data, as well as identify and monitor trends in corporate, competitor and market performance."

(bettermanagement.com)

Business Intelligence main objectives

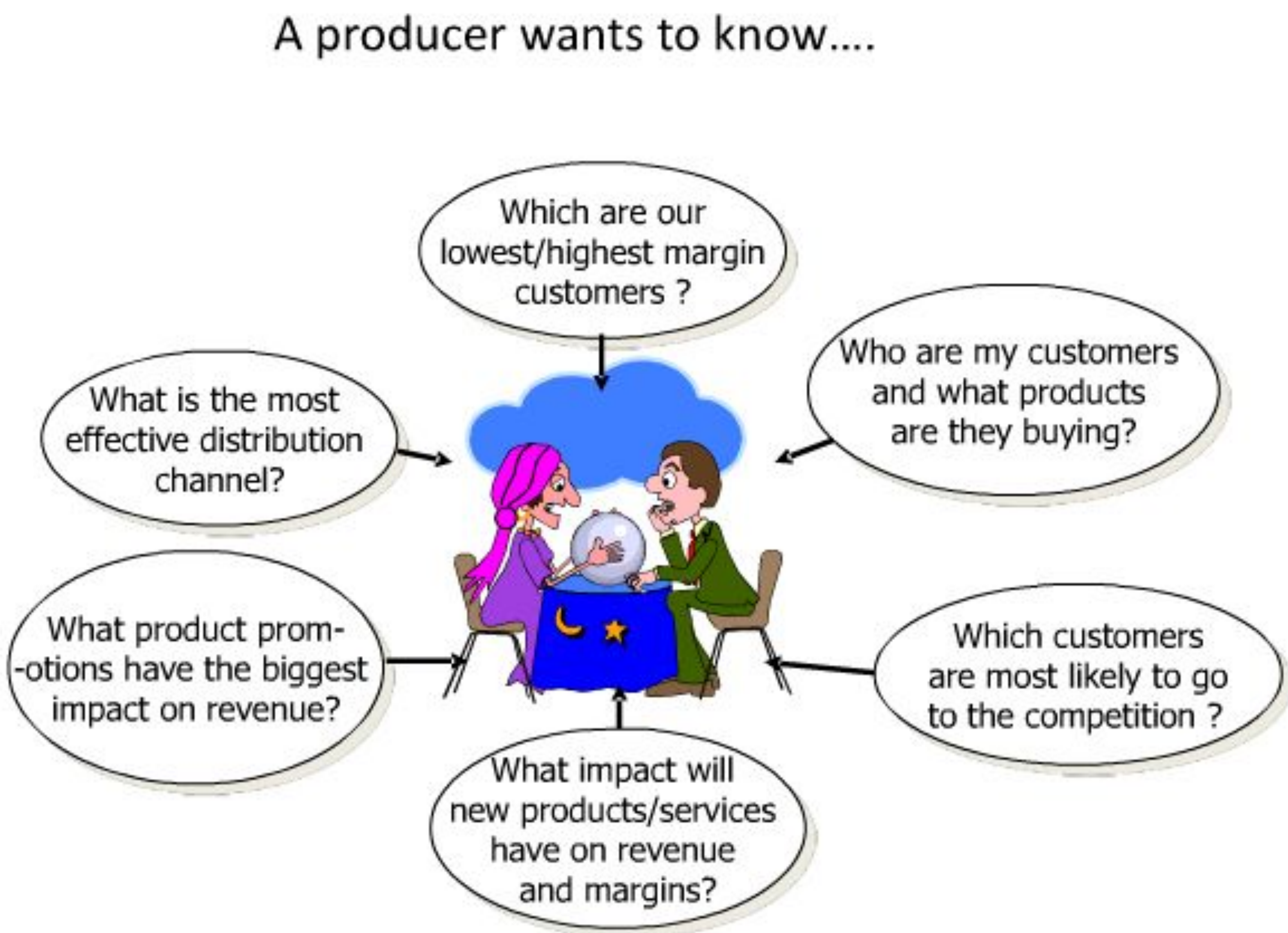
- Enable interactive access to data (sometimes in real time)
- Enable manipulation of data to allow appropriate analysis by managers
- Provide valuable insights to produce informed and better decisions
- The process of BI is based on transformation of data to information, then to decisions and finally to actions
- Facilitate closing the strategy gap of an organization

Various tools and techniques in BI



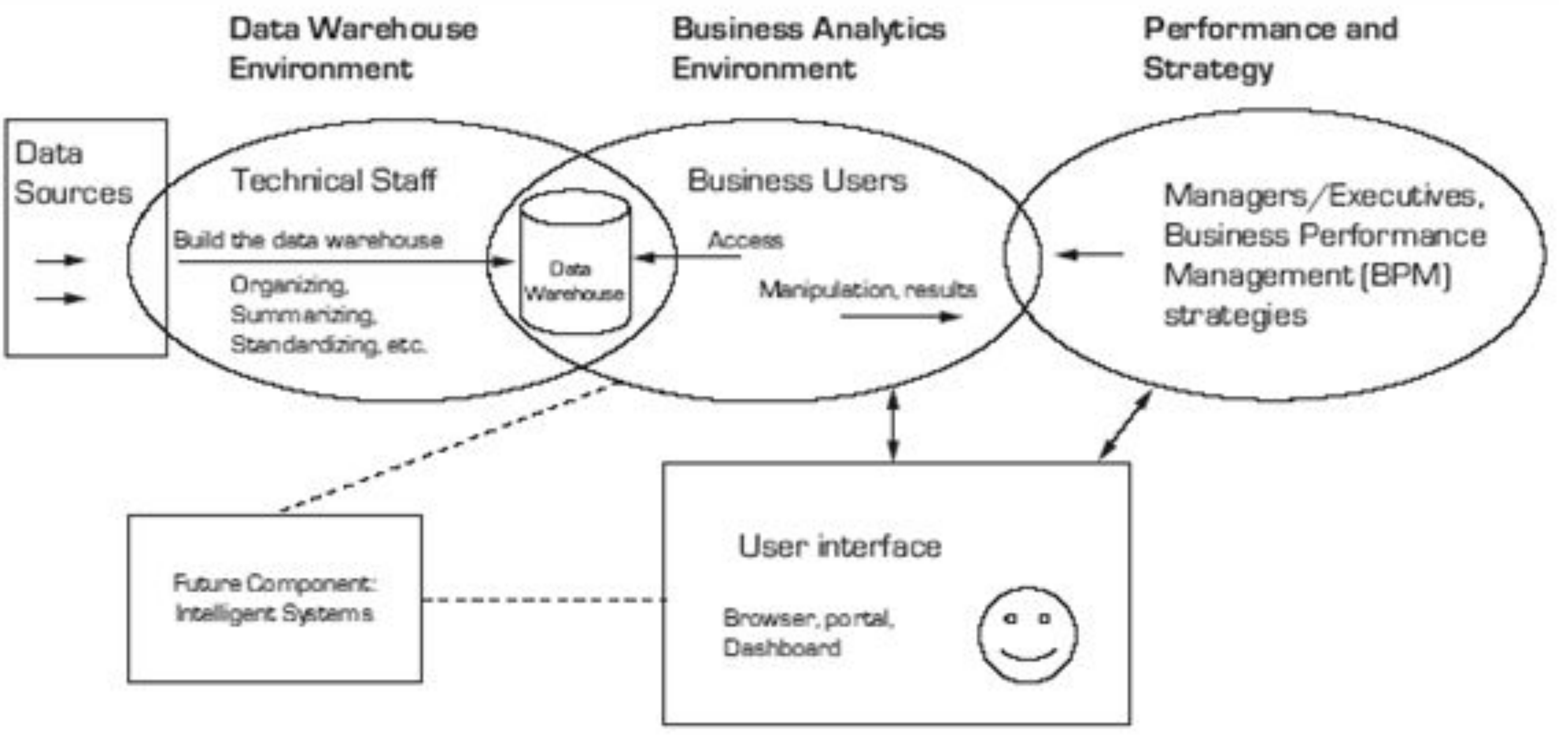
Most sophisticated BI products include most of the above

Decision Making in Business



Will require information

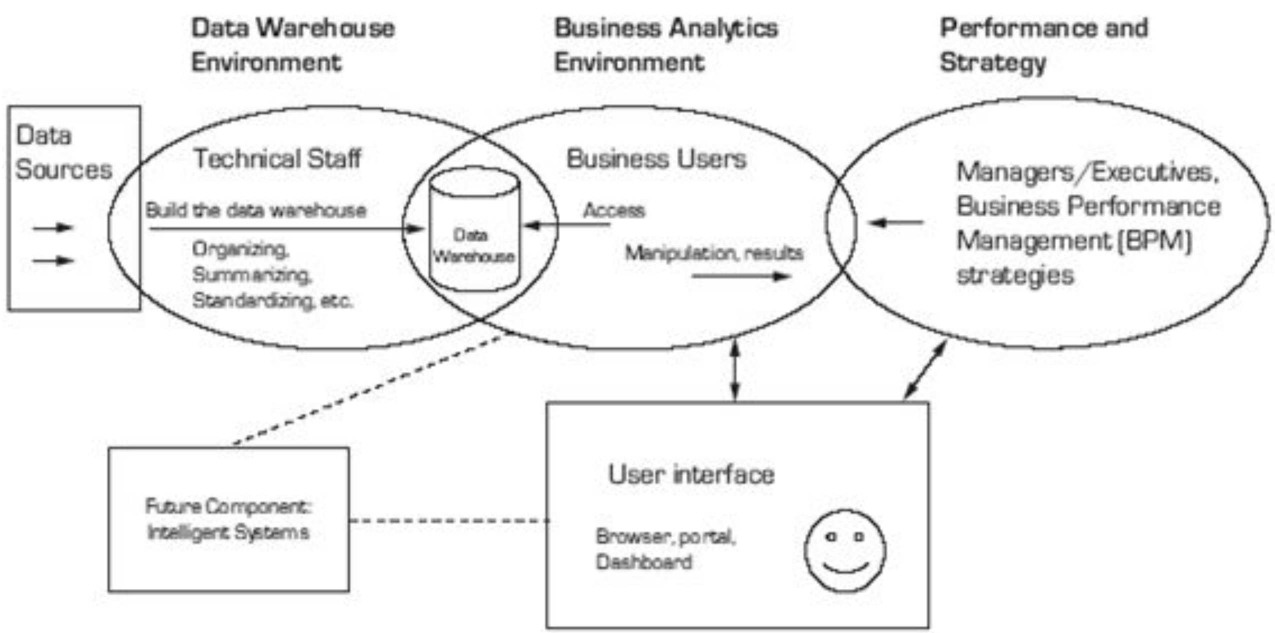
The architecture of Business Intelligence



Four major components

4 major components of Business Intelligence architecture

1. The data warehouse is a special database or repository of data that had been prepared to support decision making applications ranging from simple reporting to complex optimization



4 major components of Business Intelligence architecture

2. Business analytics are the software tools that allow users to create on-demand reports, queries and conduct analysis of data Originally they appear under the name online analytical processing (OLAP)

   Data Mining - A class of information analysis based on databases that looks for hidden patterns in a collection of data which can be used to predict future behavior

   e.g. Amazon.com uses data mining to predict the behaviour of their customers

   Automated Decision Systems - Rule-based system that provide solution usually in one functional area to a specific repetitive managerial problems

4 major components of Business Intelligence architecture

3. Business performance management (BPM) based on balanced scorecard methodology – a framework for defining, implementing, and managing an enterprise's business strategy by linking objectives with factual measures

   Objective is to optimize overall performance of an organization. A real-time system that alert managers to potential opportunities, impending problems, and threats, and then empowers them to react through models and collaboration

The architecture of Business Intelligence

4. User interface allows access and easy manipulation of other BI components

Tools used to broadcast information

Data visualization provides graphical, animation, or video presentation of data and the results of data analysis
The ability to quickly identify important trends in corporate and market data can provide competitive advantage

Business Model

What is a Business Model?

- Model
  - A model is a plan or diagram that is used to make or describe something.
- Business Model
  - A firm's business model is its plan or diagram for how it competes, uses its resources, structures its relationships, interfaces with customers, and creates value to sustain itself on the basis of the profits it generates.
  - The term "business model" is used to include all the activities that define how a firm competes in the marketplace.

Business Models

- Timing of Business Model Development
  - The development of a firm's business model follows the feasibility analysis stage of launching a new venture but comes before writing a business plan.
  - If a firm has conducted a successful feasibility analysis and knows that it has a product or service with potential, the business model stage addresses how to surround it with a core strategy, a partnership network, a customer interface, distinctive resources, and an approach to creating value that represents a viable business.

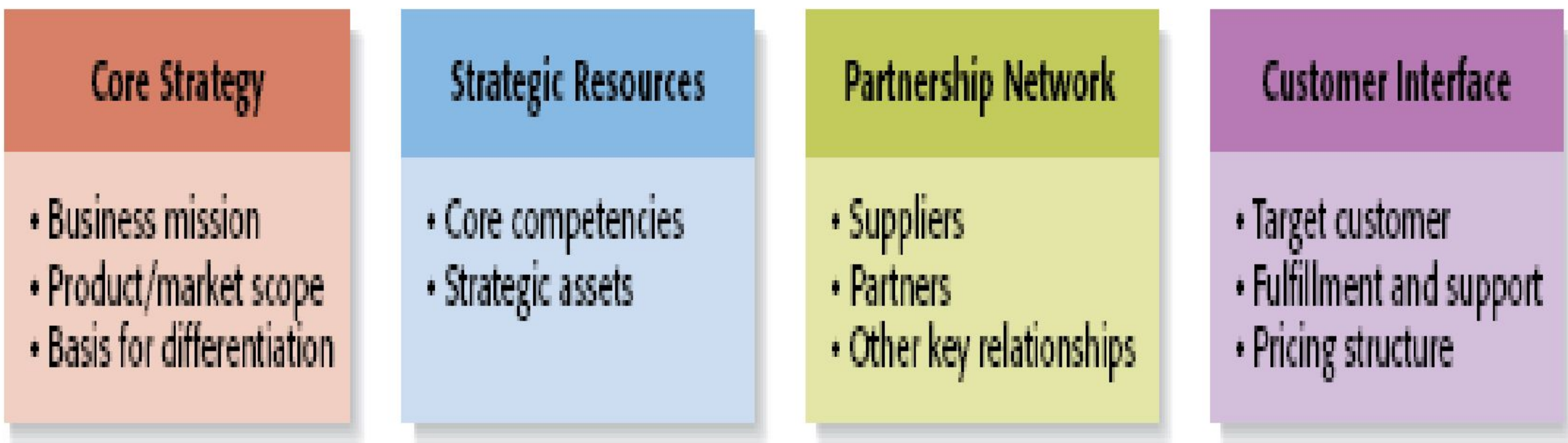
Importance of a Business Model

Having a clearly articulated business model is important because it does the following:

- Serves as an ongoing extension of feasibility analysis. A business model continually asks the question, "**Does this business make sense?"**
- Focuses attention on how all the elements of a business fit together and constitute a working whole.
- Describes why the network of participants needed to make a business idea viable are willing to work together.
- Articulates a company's core logic to all stakeholders, including the firm's employees.

Components of a Business Model

Four Components of a Business Model



Recap: The Importance of Business Models

- Business Models
  - It is very useful for a new venture to look at itself in a holistic manner and understand that it must construct an effective “business model” to be successful.
  - Everyone that does business with a firm, from its customers to its partners, does so on a voluntary basis. As a result, a firm must motivate its customers and its partners to play along.
  - Close attention to each of the primary elements of a firm’s business model is essential for a new venture’s success.