Business Intelligence

Unit I

https://www.scribd.com/document/390973082/Business-Intelligence-concepts

Content

- Introduction to Business Intelligence
- Introduction to Digital Data and Its Types
  - Structured,
  - Semi-Structured and
  - Unstructured,
- Introduction to OLTP and OLAP (MOLAP, ROLAP, HOLAP),
- BI Definitions & Concepts,
  - BI Framework,
- Data Warehousing Concepts and
  - Its Role in BI;
- BI Infrastructure Components –
  - BI Process,
  - BI Technology,
- BI Roles & Responsibilities,
- Business Applications of BI,
- BI Best Practices.

Business Intelligence

- The term Business Intelligence (BI) is relatively new but the it is synonymous with a range of applications that have been around for years;
  - Decision support systems
  - Executive Information Systems
  - On-line Analytical Processing (E.F Codd early 90's) or multi-dimensional modelling
- It is the conversion of data into information in such a way that the business is able to analyse the information to gain insight and take action

History of BI

- Its roots go back to the late 1960s
- In the 1970s, there were decision support systems (DSS)
- In the 1980s, there were EIS, OLAP, GIS, and more
- Data warehousing and dashboards/scorecards became popular in the 1990s
- Howard Dresner, a Gartner analyst, coined the BI term in the early 1990s
- Today there is much discussion of analytics
- There are many BI definitions, but the following is useful

Functional Areas of BI

- ETL Tools
- Data Warehouse
- Query Tools and Reporting
- OLAP technologies
- Data Mining
- Process Mining
- Complex event processing
- BPM – Business Performance Management
- Bench Marking
- Text Mining
- Predictive Analysis
- Prescriptive Analysis
- SaaS – Software as a Service
- Cloud Computing or Cloud

Where is Business Intelligence applied?

Operational Efficiency

- ERP Reporting
- KPI Tracking
- Product Profitability
- Risk Management
- Balanced Scorecard
- Activity Based Costing
- Global Sourcing
- Logistics

Customer Interaction

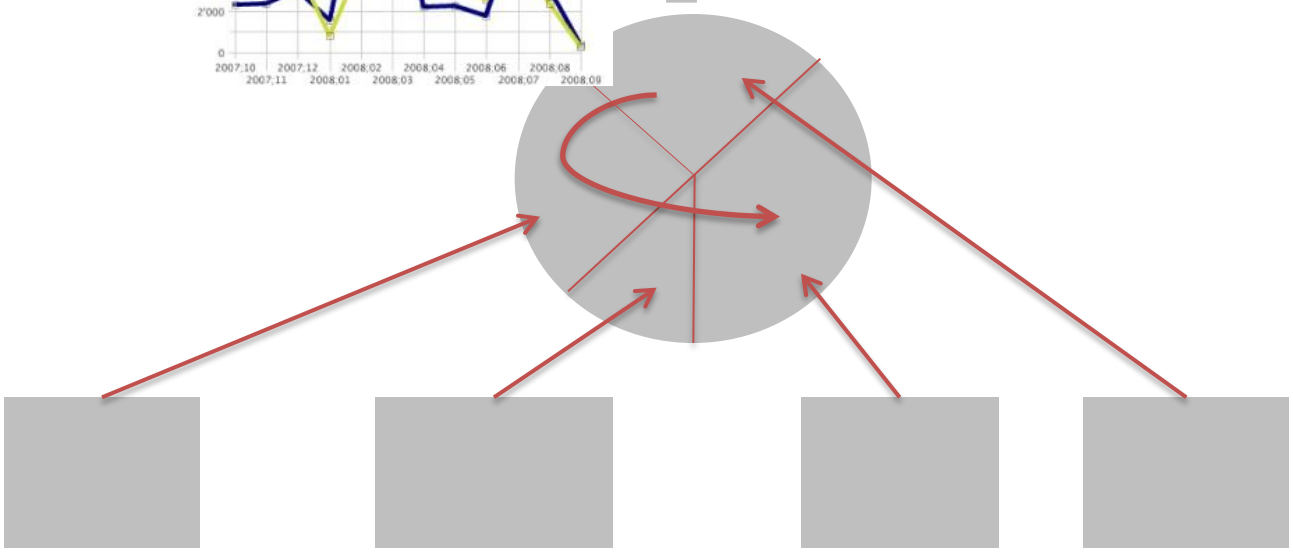
- Sales Analysis
- Sales Forecasting
- Segmentation
- Cross-selling
- CRM Analytics
- Campaign Planning
- Customer Profitability

Business Intelligence

Reports

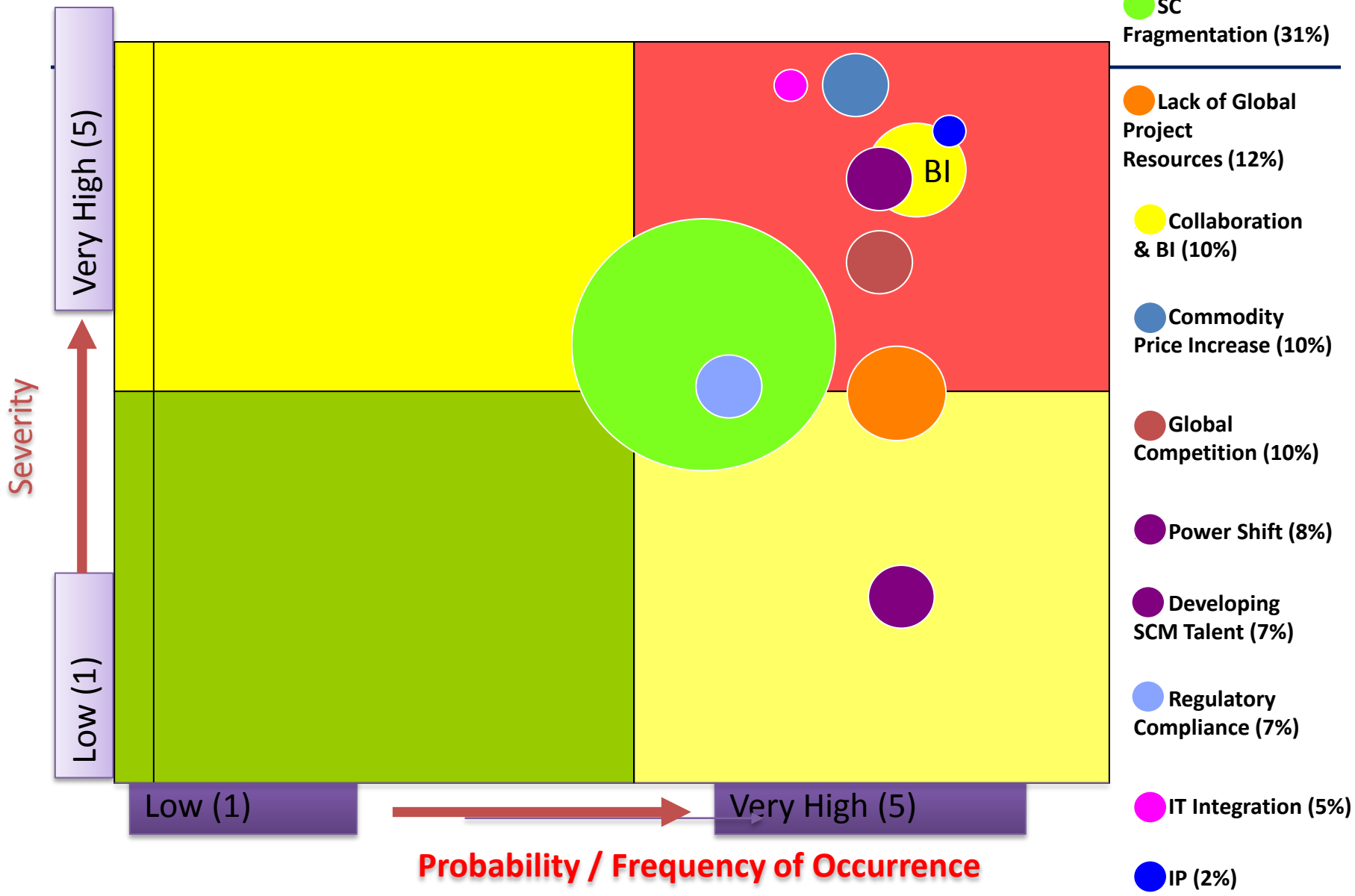
Data repository

Different systems



7

Exec survey: What Keeps You Up at Night?



Sample BI Architecture

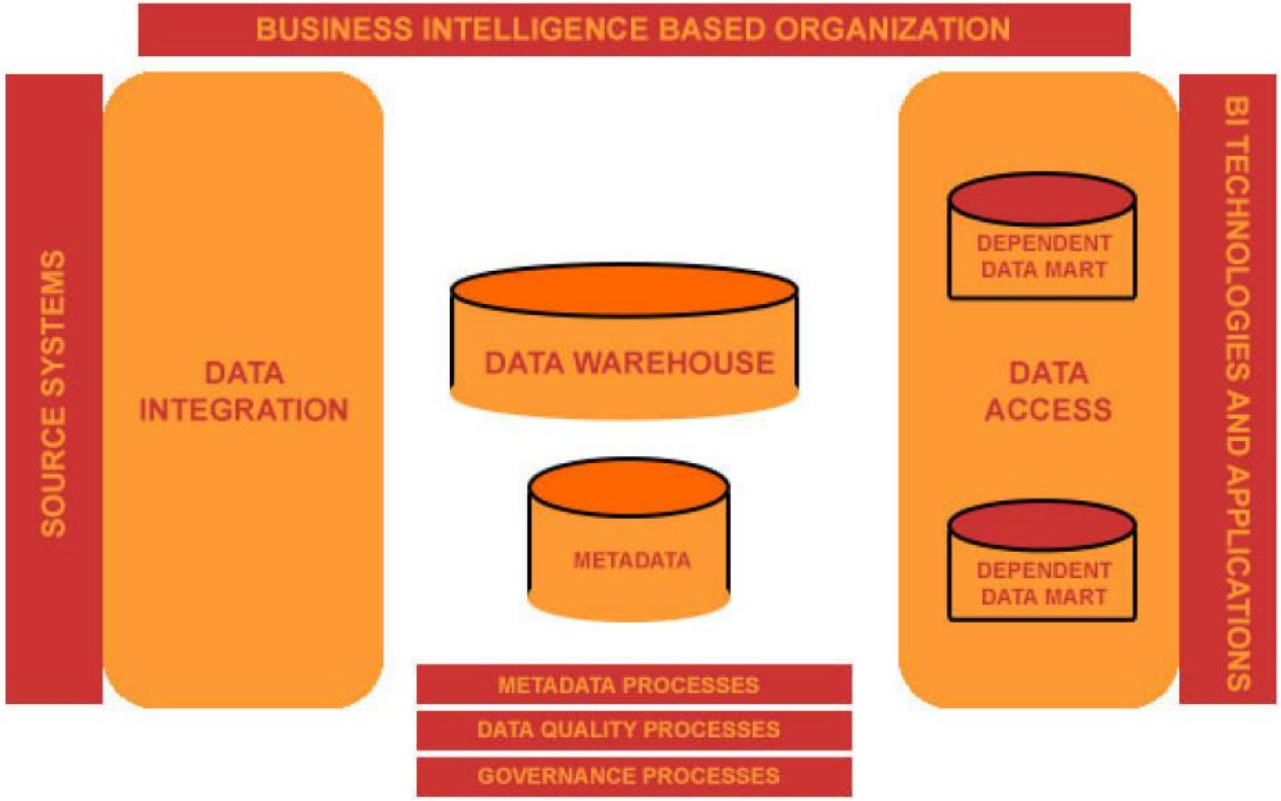


Figure: Sample BI Architecture

Detailed BI Architecture

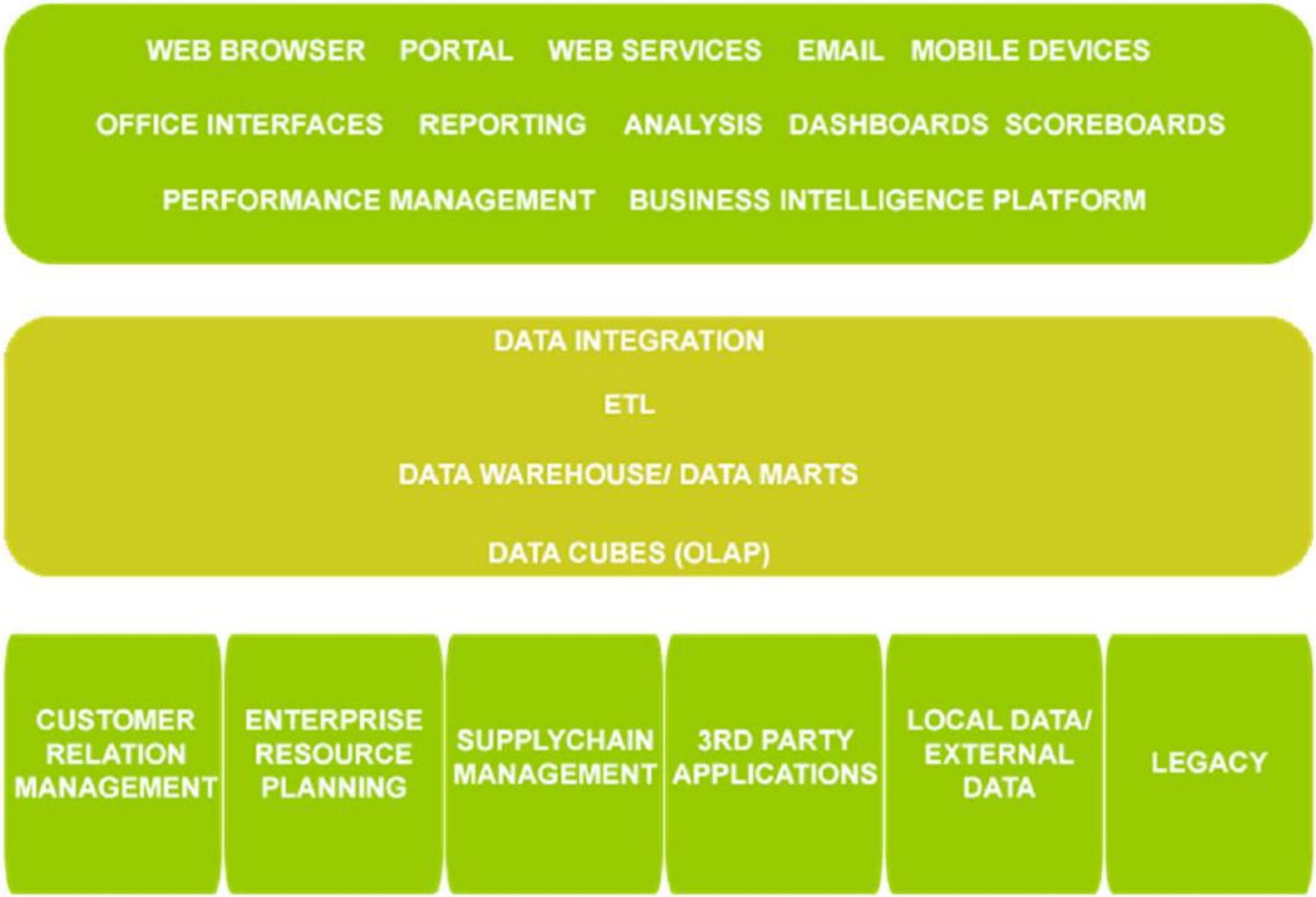
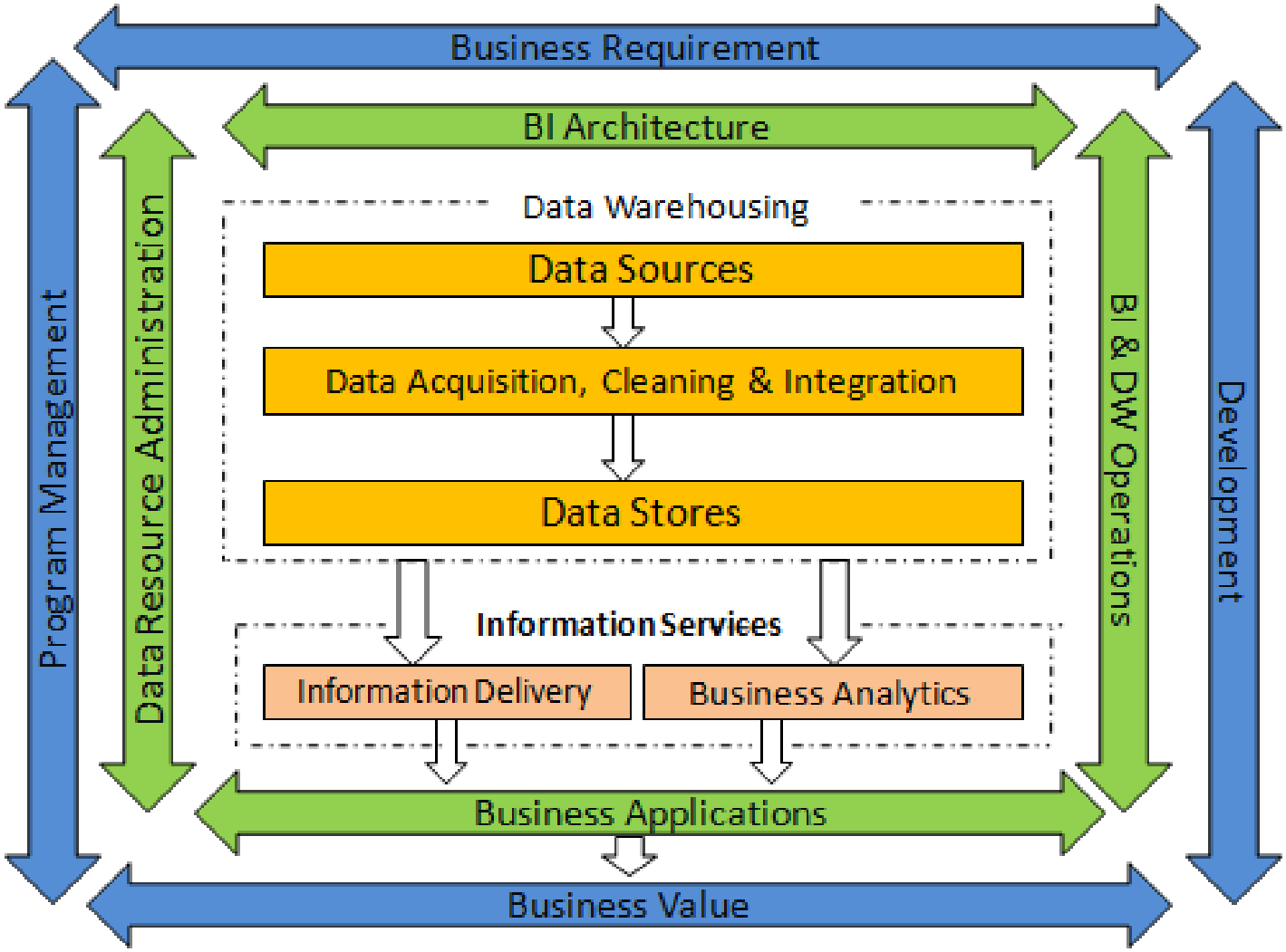
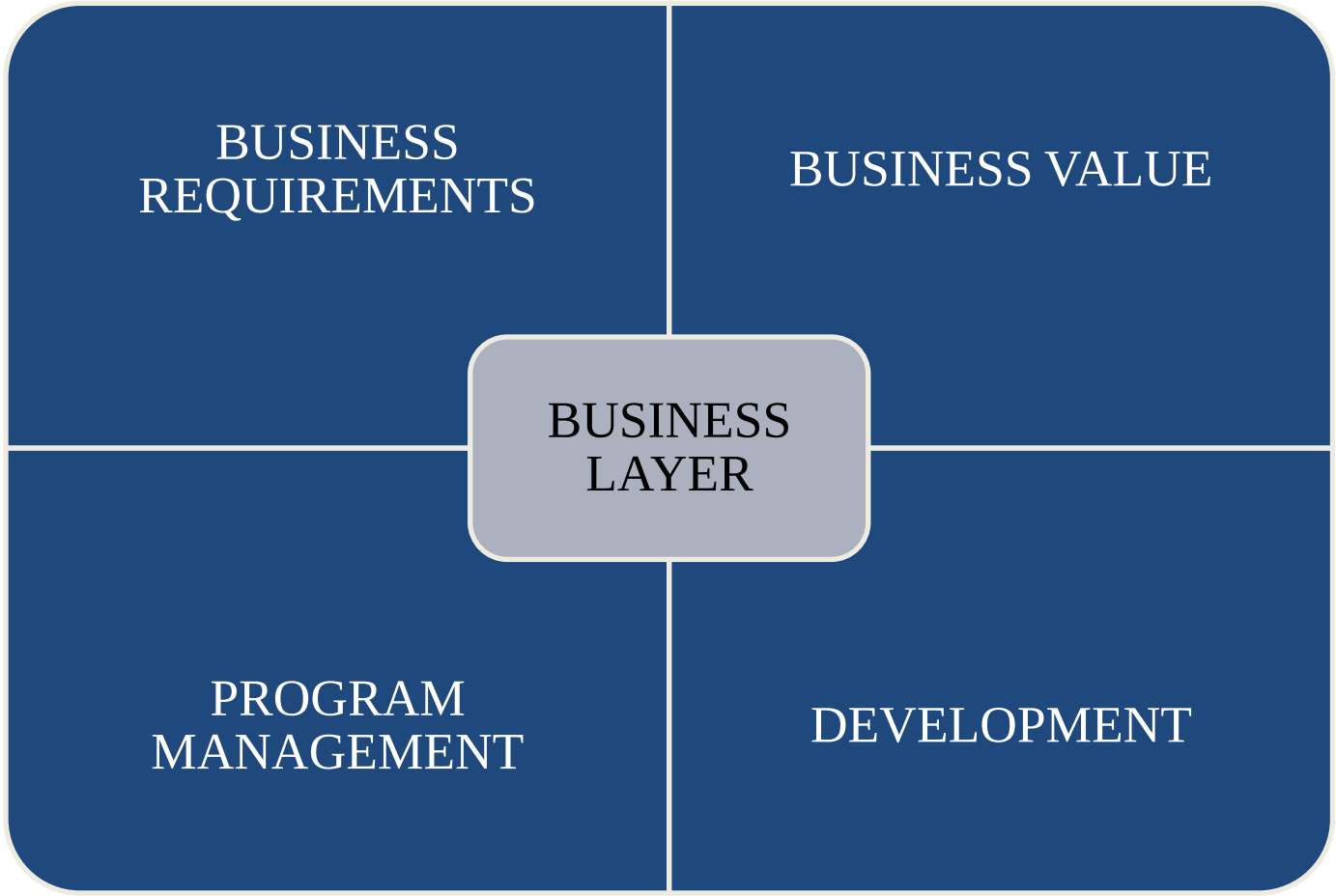


Figure: BI Components and Architecture

BI Framework



Business Layer



Business Layer

Business requirements: The requirements are a product of three steps of a process that includes:

- ***Business drivers*** (the impulses that initiate the need to act).

  Examples: changing workforce, changing labor laws, changing economy, changing technology, etc.
- ***Business goals*** (the targets to be achieved in response to the business drivers).

  Examples: increased productivity, improved market share, improved profit margins, improved customer satisfaction, cost reduction, etc.
- ***Business strategies*** (the planned course of action that will help achieve the set goals).

  Examples: outsourcing, global delivery model, partnerships, customer retention programs, employee retention programs, competitive pricing, etc.

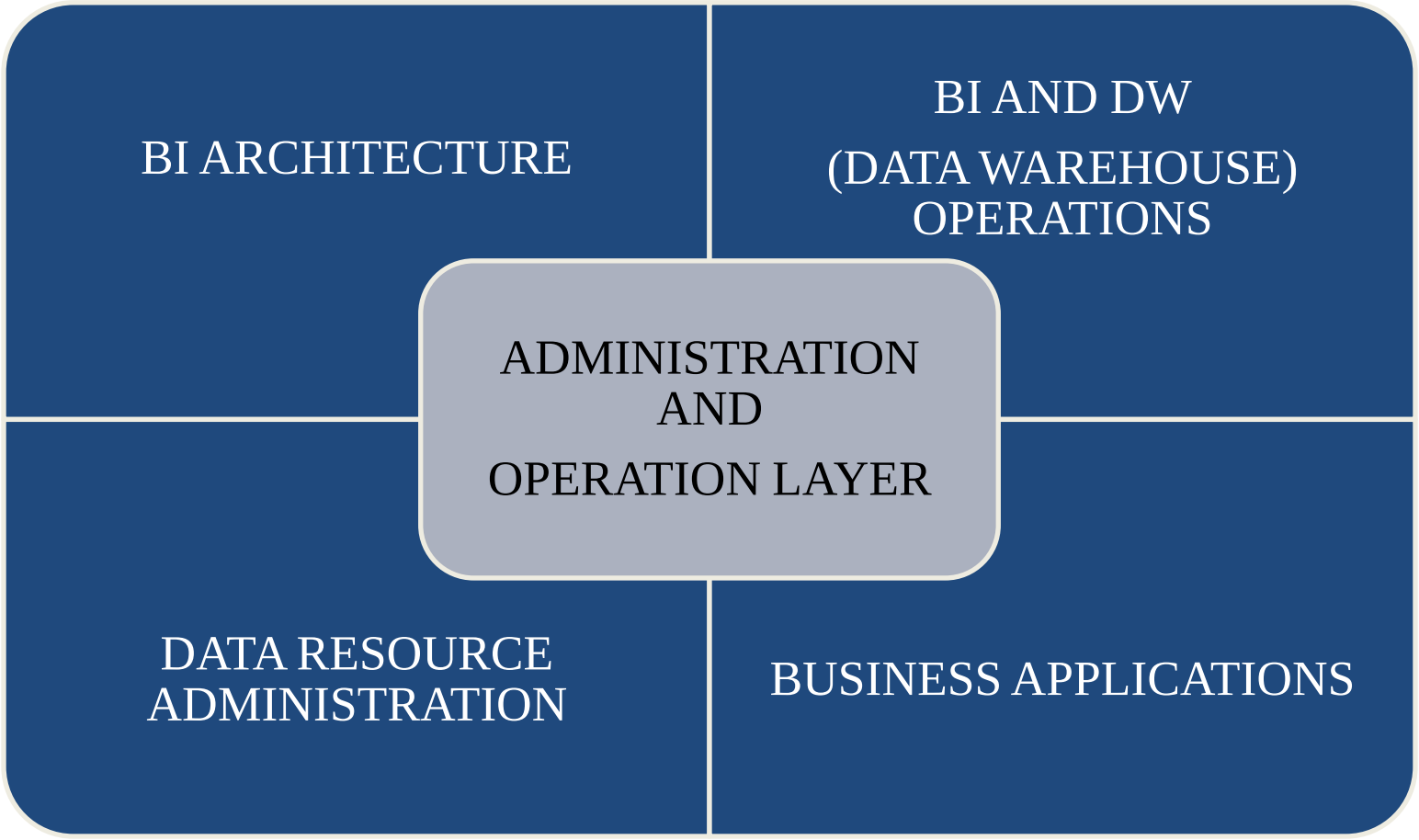
Business Layer

Business Value: Business value can be measured in terms of ROI (Return on Investment), ROA (Return on Assets), TCO (Total Cost of Ownership), TVO (Total Value of Ownership), etc.

Program management: It is the component that ensures people, projects and priorities work in a manner in which individual processes are compatible with each other; so as to ensure seamless integration and smooth functioning of the entire program.

Development: The process of development consists of *database/data-warehouse development* (consisting of ETL, data profiling, data cleansing and database tools), *data integration system development* (consists of data integration tools and data quality tools) and *business analytics development* (about processes and various technologies used).

Administration and Operation Layer



Administration and Operations Layer

BI Architecture

DATA	• Should follow design standards • Must have a logically apt data model • Metadata should be of high standard
INTEGRATION	• Performed according to business semantics and rules • During integration, certain processing standards have to be followed • Data must be consistent
INFORMATION	• Information derived from data that has been integrated should be usable, findable and as per the requirements
TECHNOLOGY	• Technology used for deriving information must be accessible • Also, it should have a good user-interface • Should support analysis, decision support, data and storage management
ORGANIZATION	• Consists of different roles and responsibilities, like management, development, support and usage roles

Administration and Operations Layer

BI and DW operations: Data warehouse administration requires the usage of various tools to monitor the performance and usage of the warehouse, and perform administrative tasks on it. Some of these tools would be:

- Backup and restore
- Security
- Configuration management
- Database management

Data resource administration: Involves *data governance* and *metadata management*.

Data governance is a technique for controlling data quality, which is used to assess, improve, manage and maintain information. It helps to define standards that are required to maintain data quality. The distribution of roles for governance of data is as follows:

- Data ownership
- Data stewardship
- Data custodianship

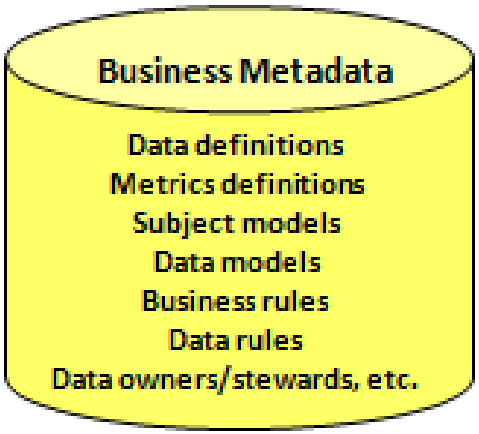
Administration and Operations Layer

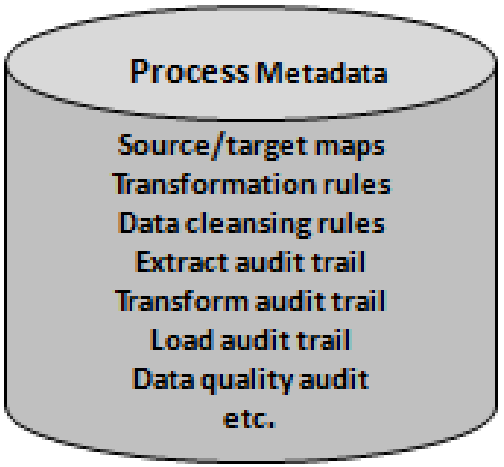
Metadata management: Metadata is data about data.

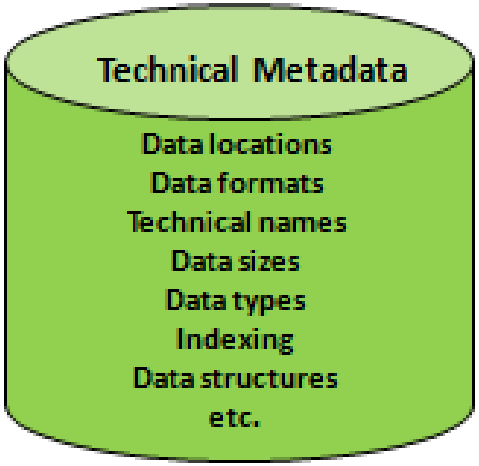
Metadata can be divided into four groups:

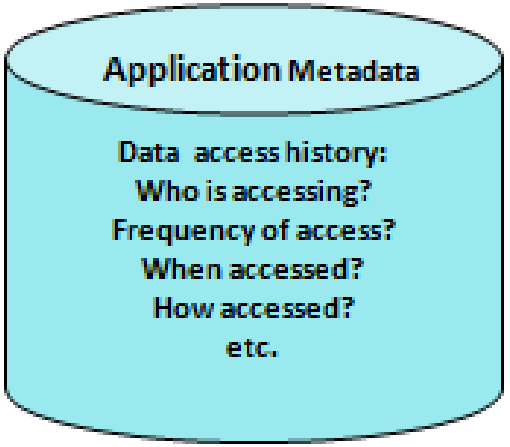
- **Business metadata:** business definition, structure, hierarchy of the data, ownership characteristics.
- **Process metadata:** transformation rules, data cleansing rules, data quality audit, etc.
- **Technical metadata:** data location, data format, data sizes, data structures, etc.
- **Application metadata:** who is accessing?, who has accessed? Frequency of access? When was it accessed? How was it accessed? etc

Administration and Operations Layer

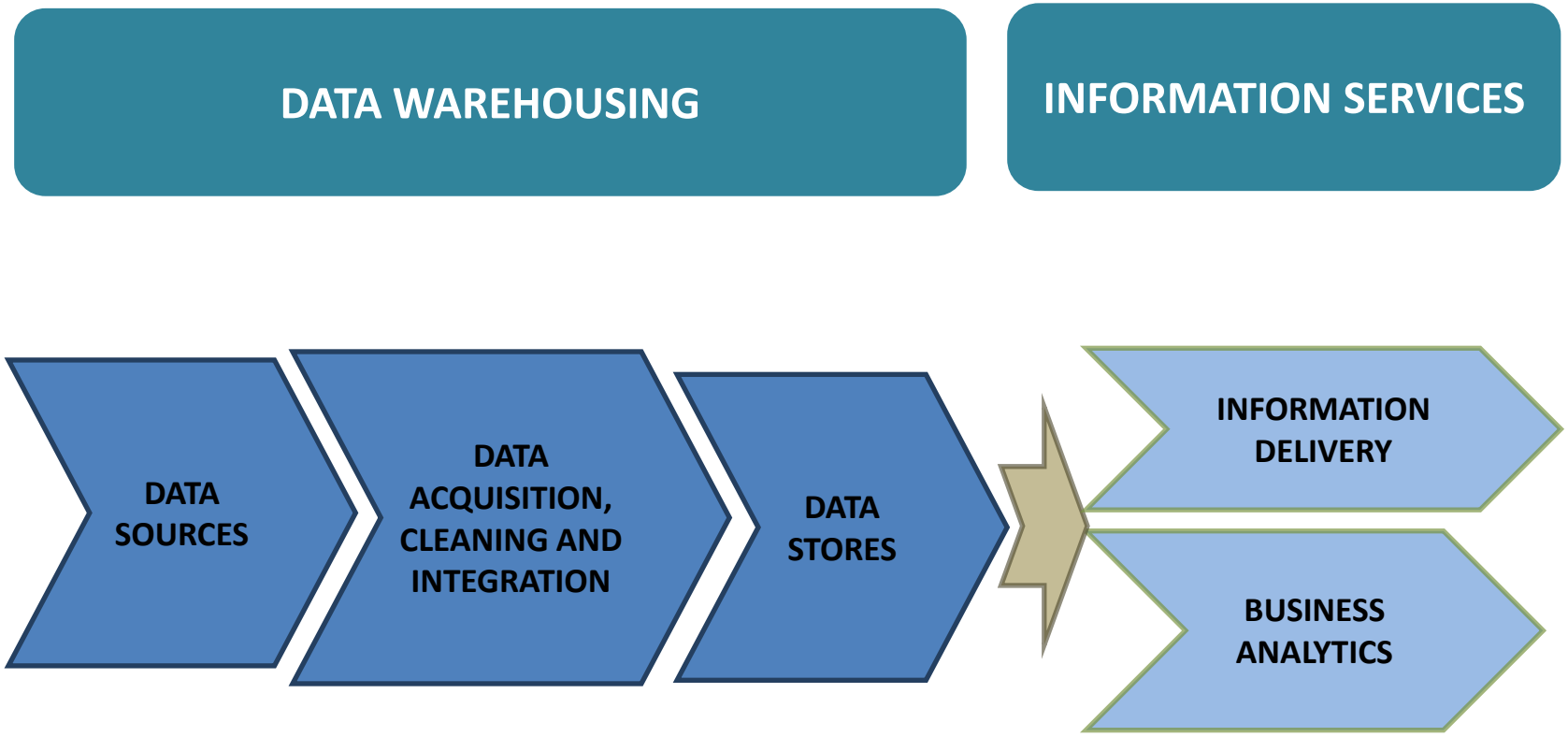




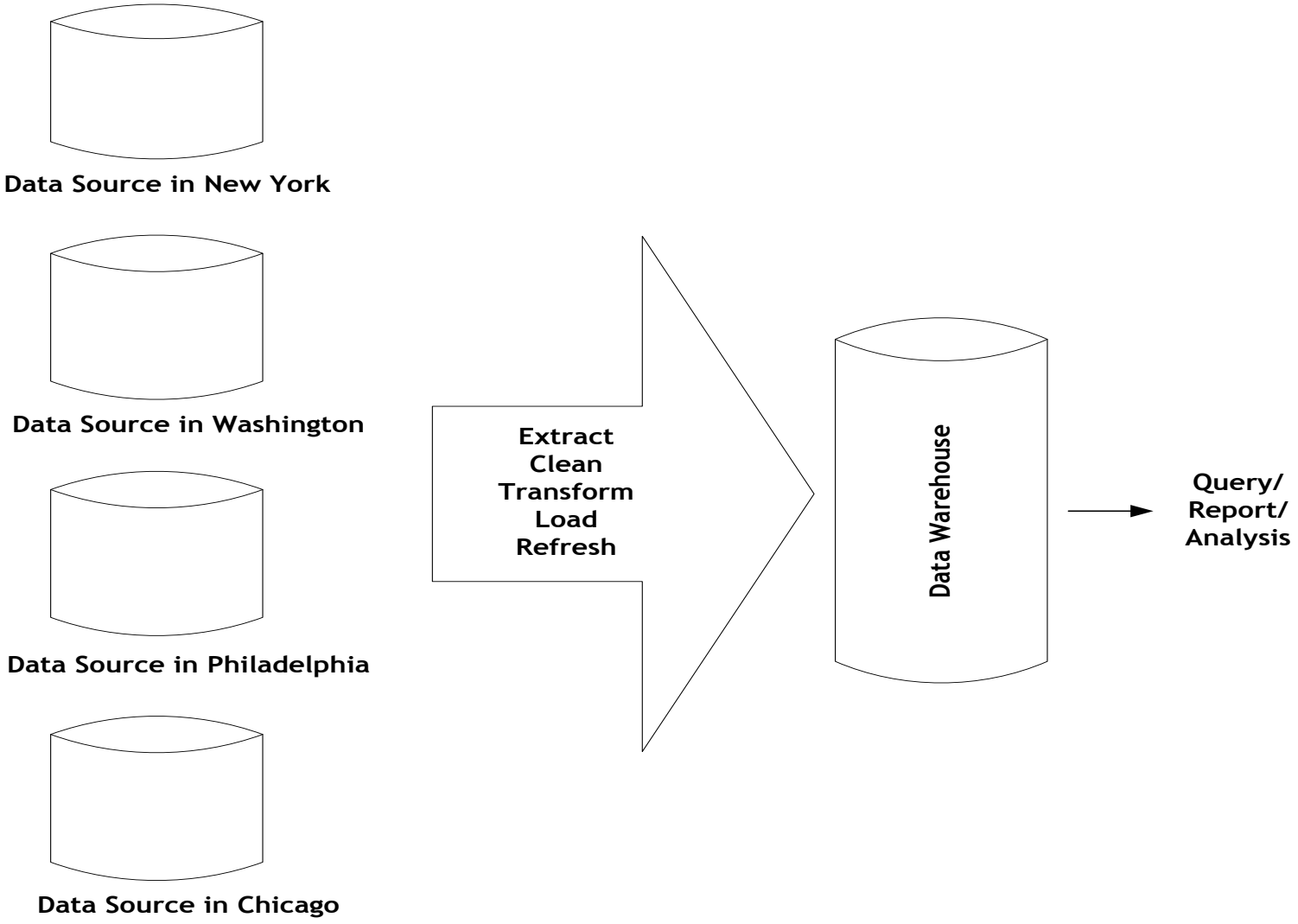




Implementation Layer



Implementation Layer

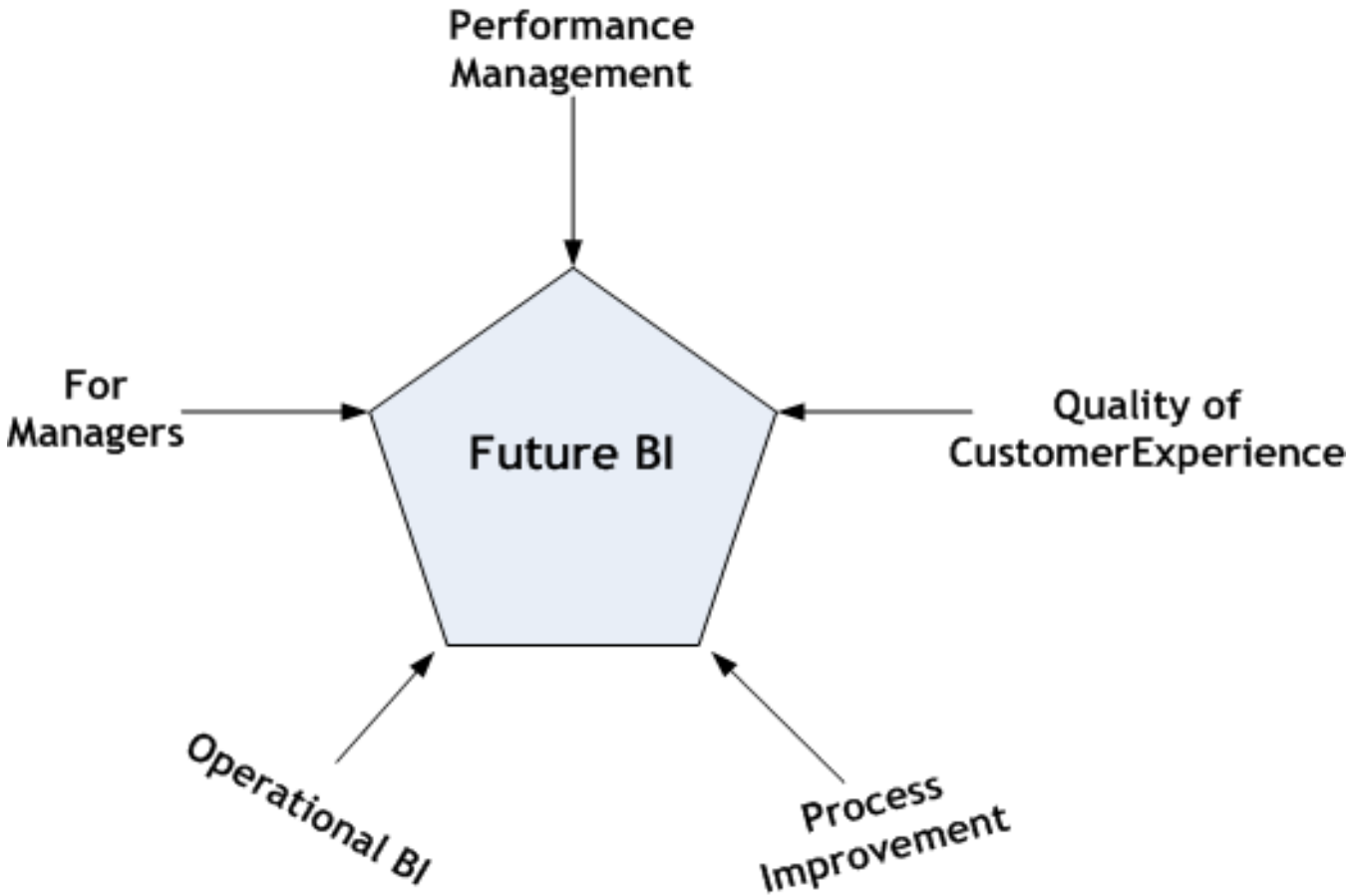


Implementation Layer

Information services:

- It is not only the process of producing information; rather, it involves ensuring that the information produced is aligned with business requirements and can be acted upon to produce value for the company.
- Information is delivered in the form of KPI's, reports, charts, dashboards or scorecards, etc., or in the form of analytics.
- Data mining is a practice used to increase the body of knowledge.
- Applied analytics is generally used to drive action and produce outcomes.

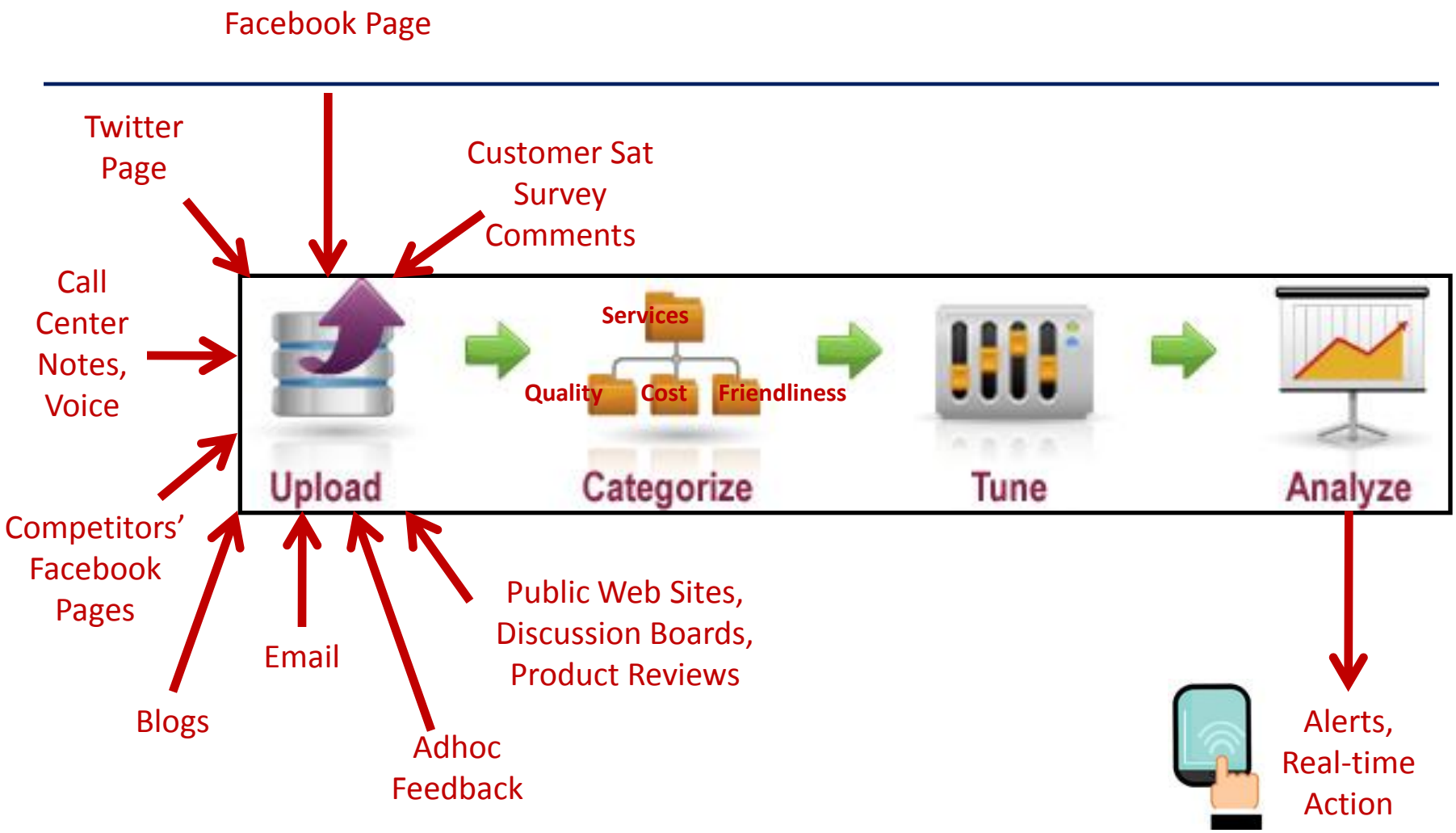
Who is BI for?


Performance Management
For Managers
Future BI
Quality of CustomerExperience
Operational BI
Process Improvement

Types of BI Users

Type of user	Casual users/ Information consumers	Power users/Information producers
Example of such users	Executives, managers, customers, suppliers, field/operation workers, etc.	SAS, SPSS developers, administrators, business analysts, analytical modelers, IT professionals, etc.
Usage	Information consumers	Information producers
Data Access	Tailor made to suit the needs of their respective role	Ad hoc/exploratory
Tools	Pre-defined reports/dashboards	Advanced Analytical/ Authoring tools
Sources	Data warehouse/Data Marts	Data Warehouse/Data Marts (both internal and external)

Unstructured Text Processing



Types of Digital Data

Digital Data

- Digital data can be
  - Unstructured
  - Semi-structured
  - Structured

- According to Merrill Lynch 80–90% of business data is either unstructured or semi-structured

- Data is usually in a format which makes it difficult to extract information from it

Unstructured data

- Data which does not conform to a data model or is not in a form which can be used easily by a computer program.
- 80-90% of the data of an organization is in this format.
- E.g. memos, chat rooms, ppts, images, videos, letters, researches, white papers, body of an e-mail etc.
- It can be classified into 2 broad categories:
  - Bitmap Objects: E.g.: image, video or audio files.
  - Textual Objects: E.g.: Microsoft Word documents, e-mails or Microsoft Excel spreadsheets.

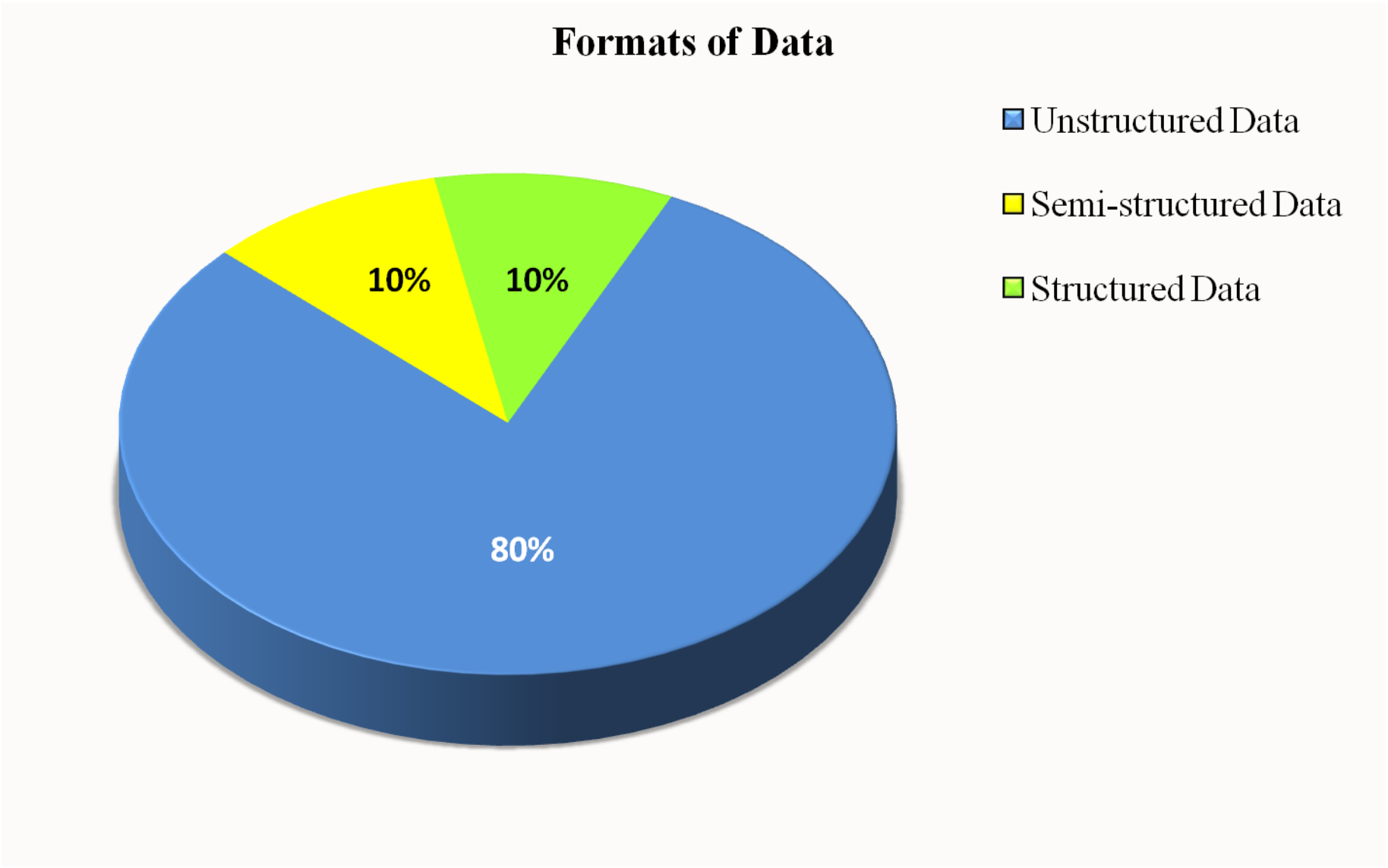
Semi- Structured data-

- Semi-structured data does not conform to any data model i.e. it is difficult to determine the meaning of data
- It is not in a form which can be used easily by a computer program. Metadata is available but it is not sufficient.
- E.g. e-mails, XML, mark- up languages like HTML

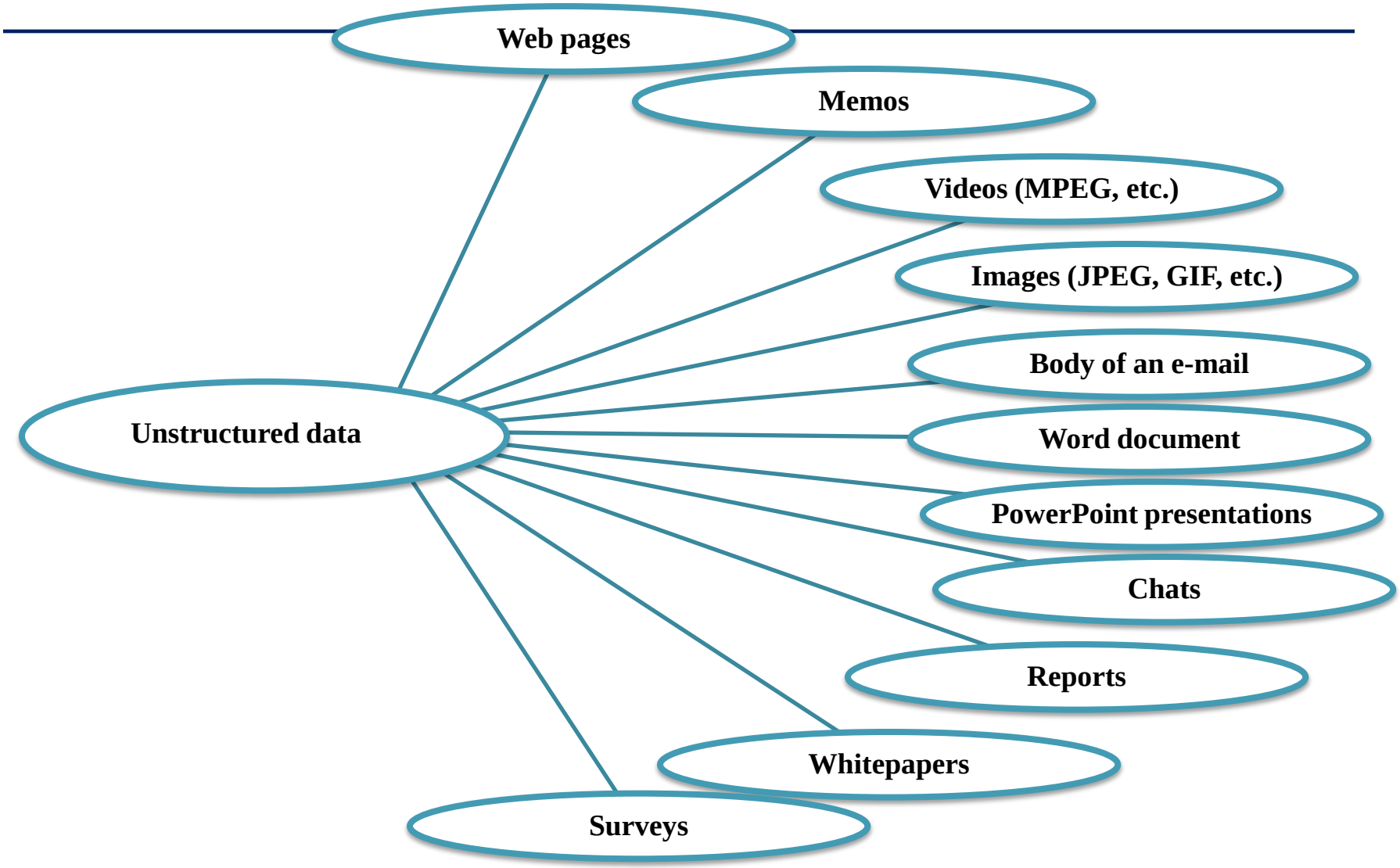
Structured Data

- Data which is in an organized form e.g. in rows and columns;
- which can be easily used by a computer program.
- Relationships exist between entities such as classes and their objects.
- E.g. data stored in databases

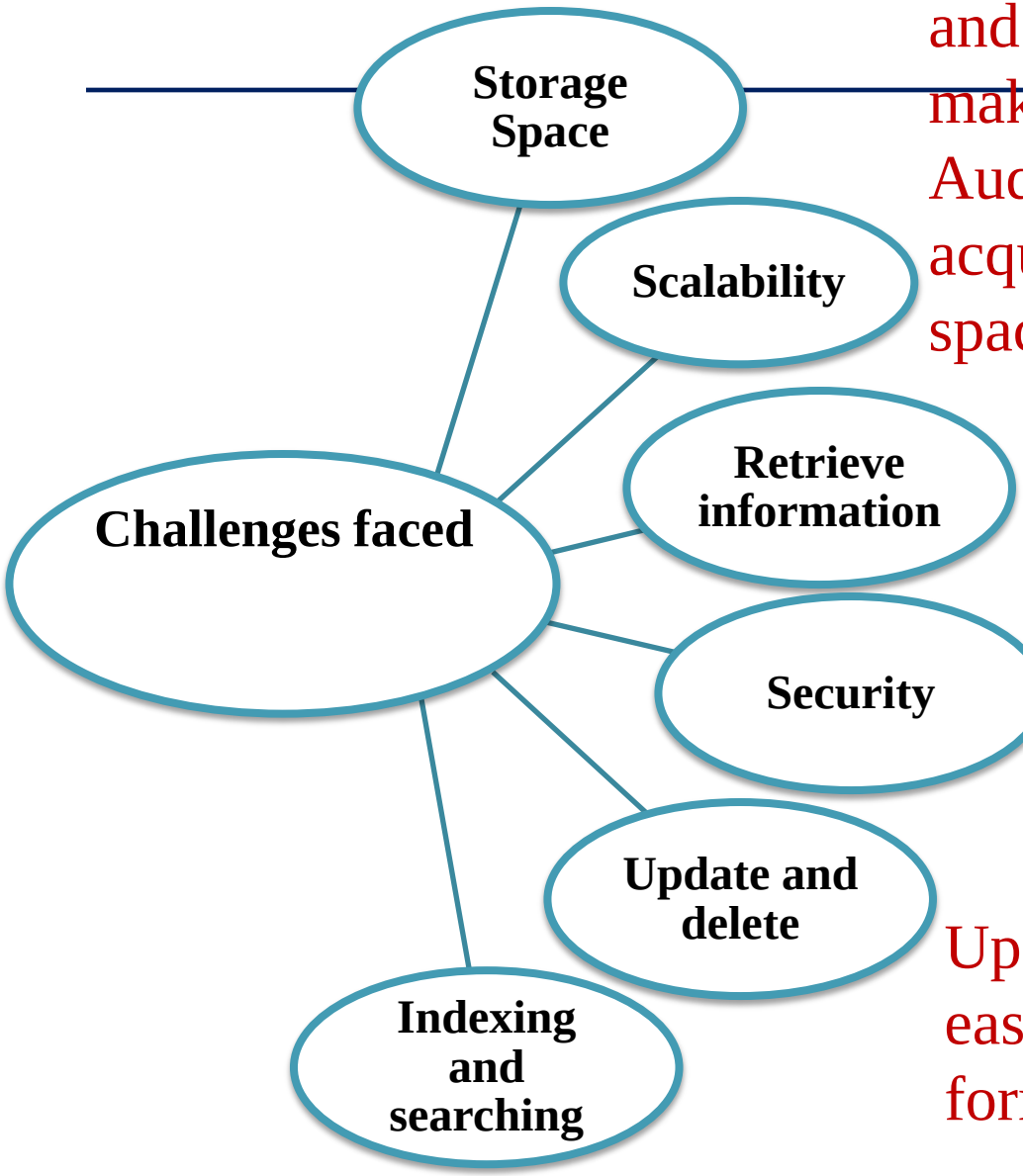
Formats of Digital Data



Where does Unstructured Data Come from?



How to Store Unstructured Data?



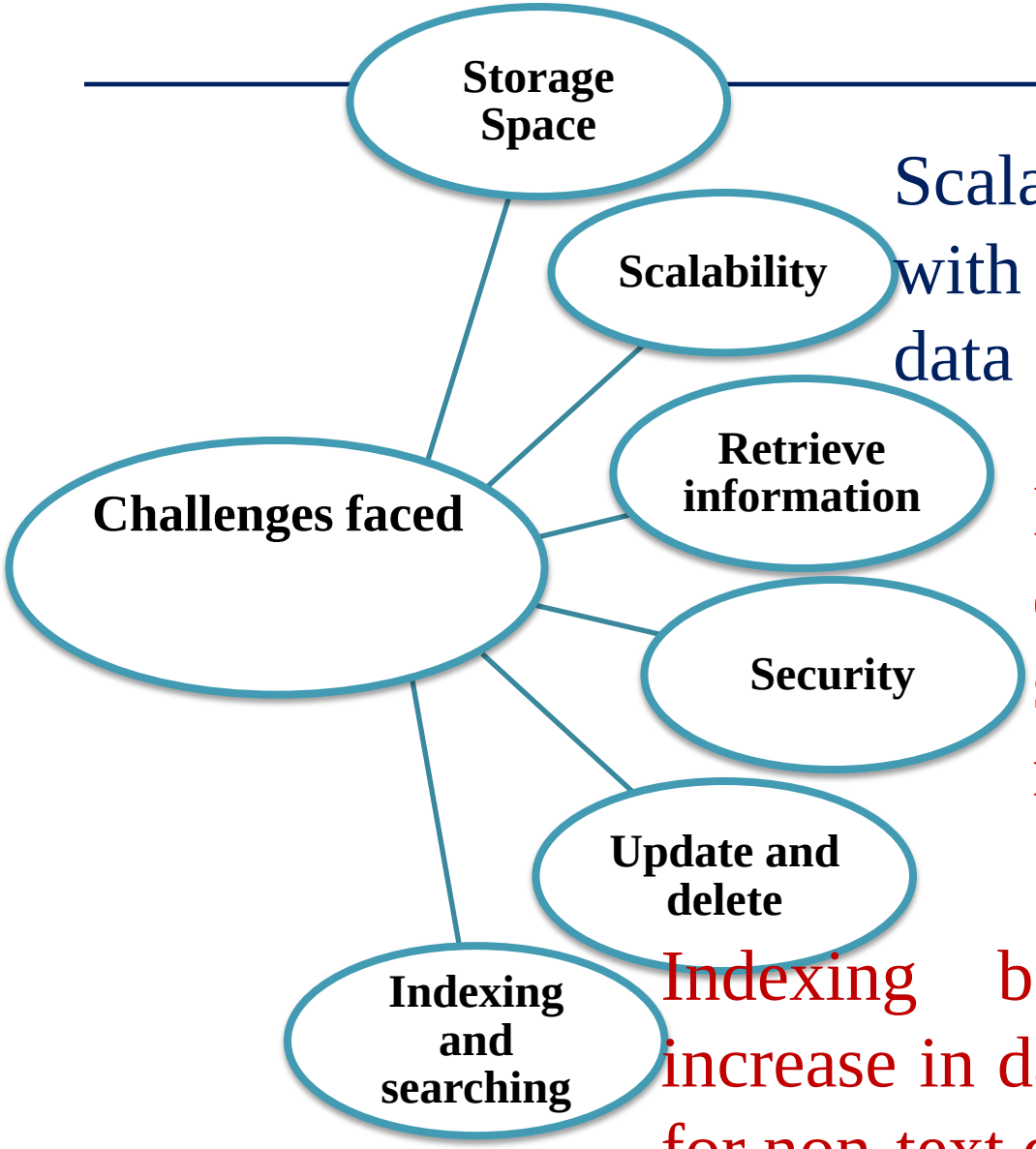
Sheer volume of unstructured data and its unprecedented growth makes it difficult to store.

Audios, videos, images, etc. acquire huge amount of storage space

Retrieving and recovering unstructured data are cumbersome

Updating, deleting, etc. are not easy due to the unstructured form

How to Store Unstructured Data?

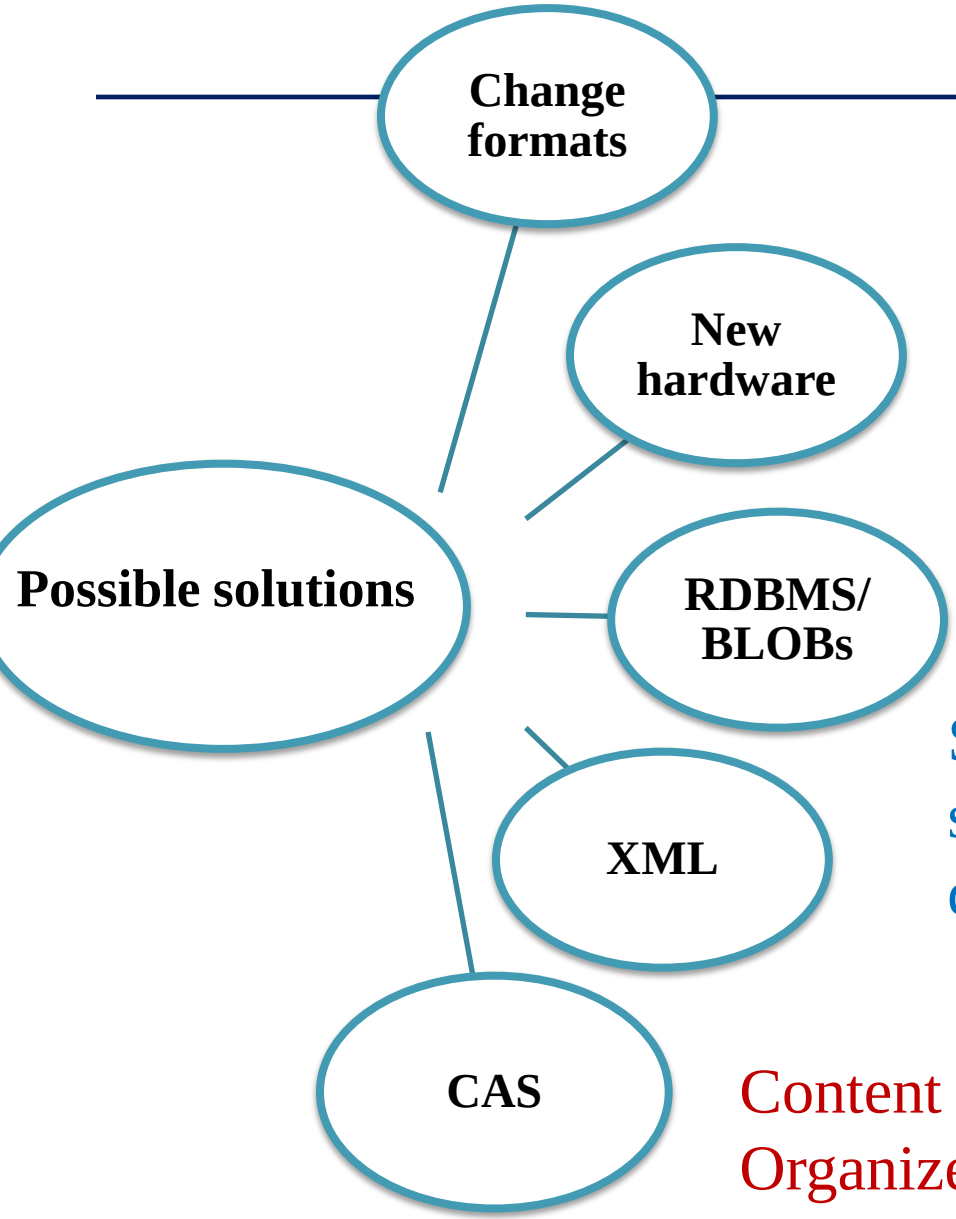


Scalability becomes an issue with increase in unstructured data

Ensuring security is difficult due to varied sources of data (e.g. e-mail, web pages)

Indexing becomes difficult with increase in data. Searching is difficult for non-text data

How to Store Unstructured Data?



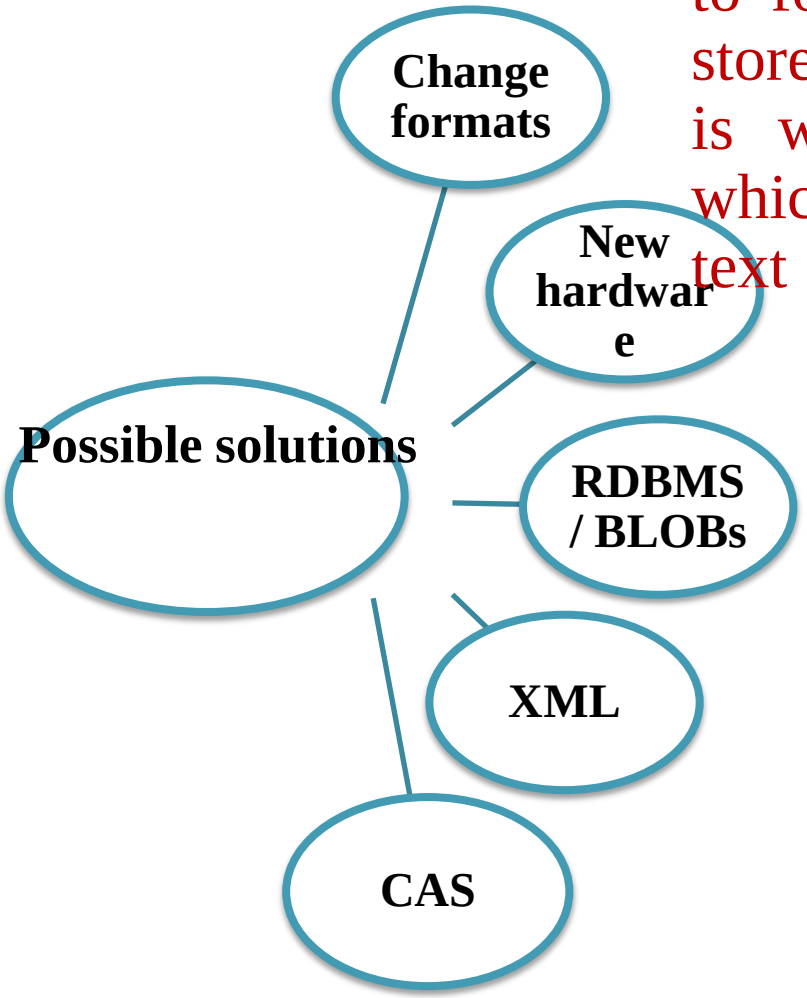
Create hardware which support unstructured data either compliment the existing storage devices or be a stand alone for unstructured data

Store in XML which tries to give some structure to unstructured data by using tags and elements

Content Addressable Storage (CAS) Organize files based on their metadata

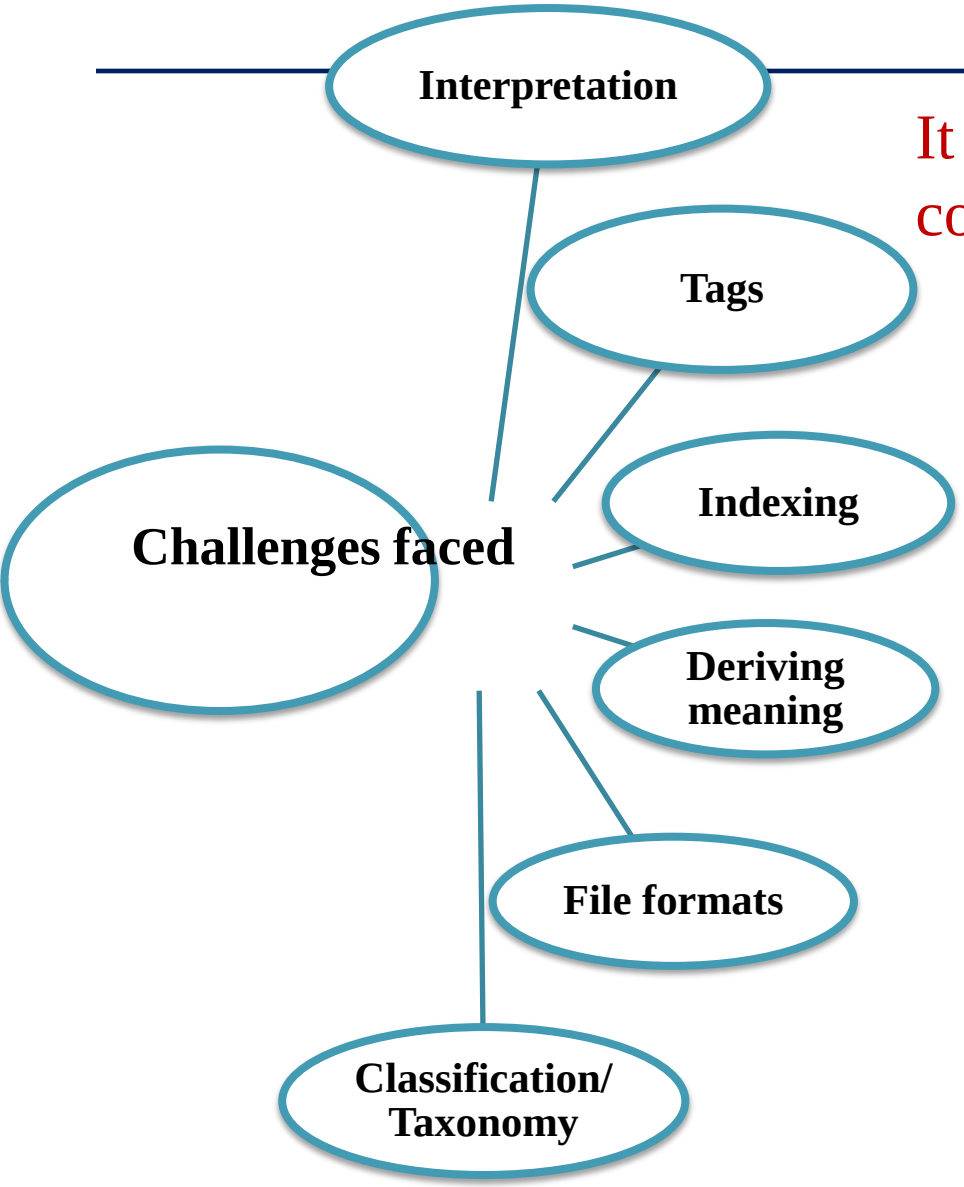
How to Store Unstructured Data?

Unstructured data may be be converted to formats which are easily managed, stored and searched. For example, IBM is working on providing a solution which converts audio , video, etc. to text



Store in relational databases which support BLOBs which is Binary Large Objects

How to Extract Information from Unstructured Data?

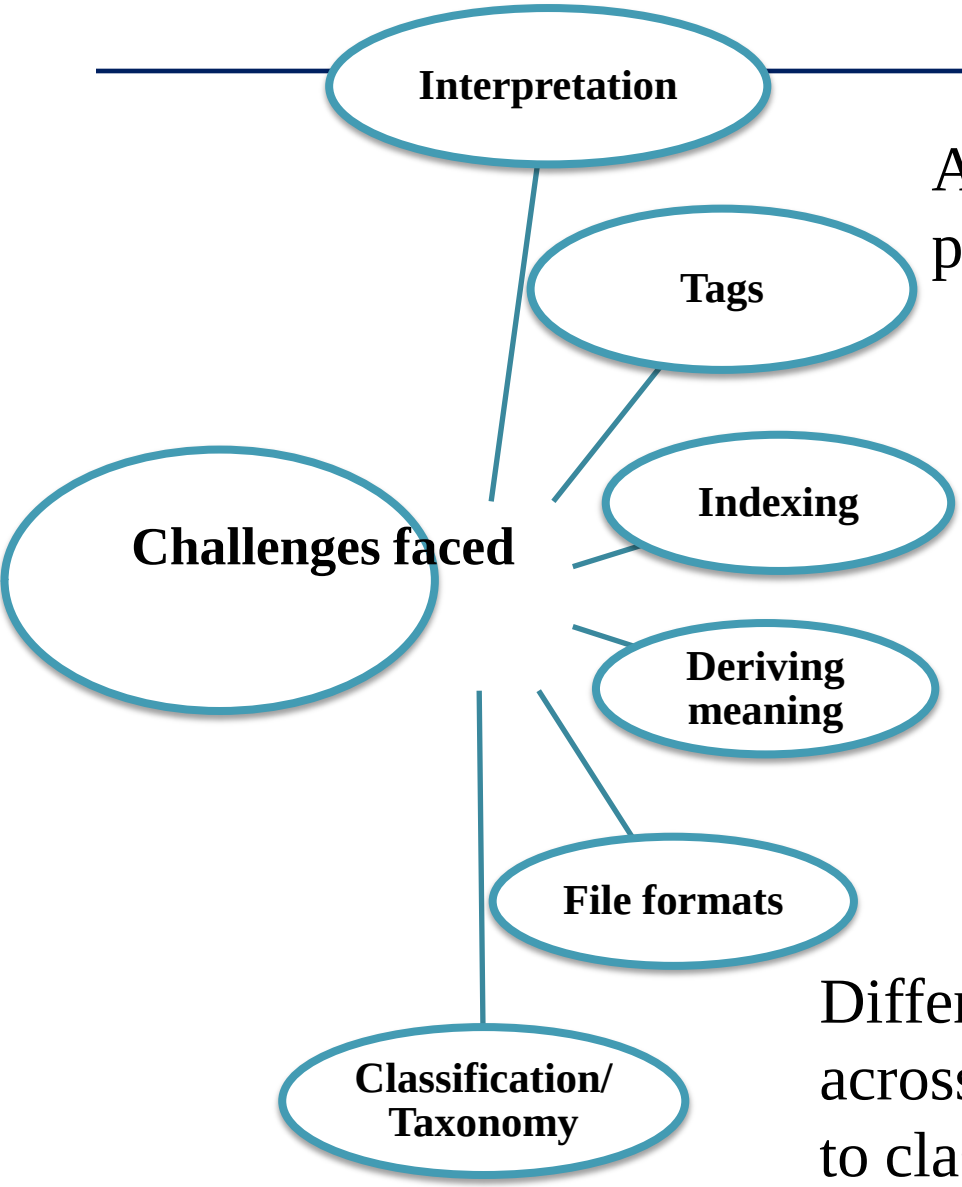


It is not easily interpreted by conventional search algorithms

Designing algorithms to understand the meaning of the document and then tag or index them accordingly is difficult

Increasing number of file formats make it difficult to interpret data

How to Extract Information from Unstructured Data?

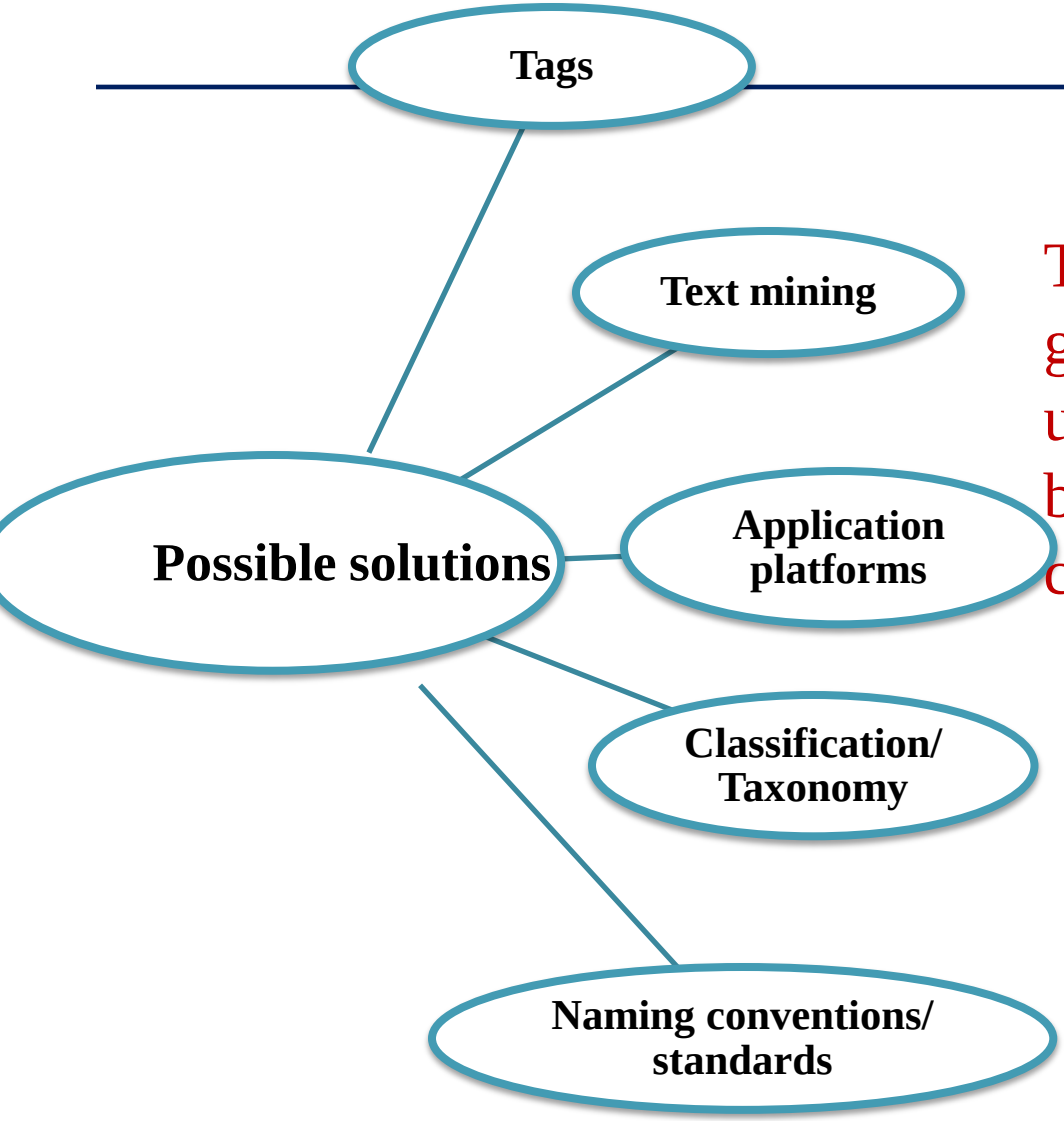


As the data grows it is not possible to put tags manually

Computer programs cannot automatically derive meaning/structure from unstructured data

Different naming conventions followed across the organization make it difficult to classify data.

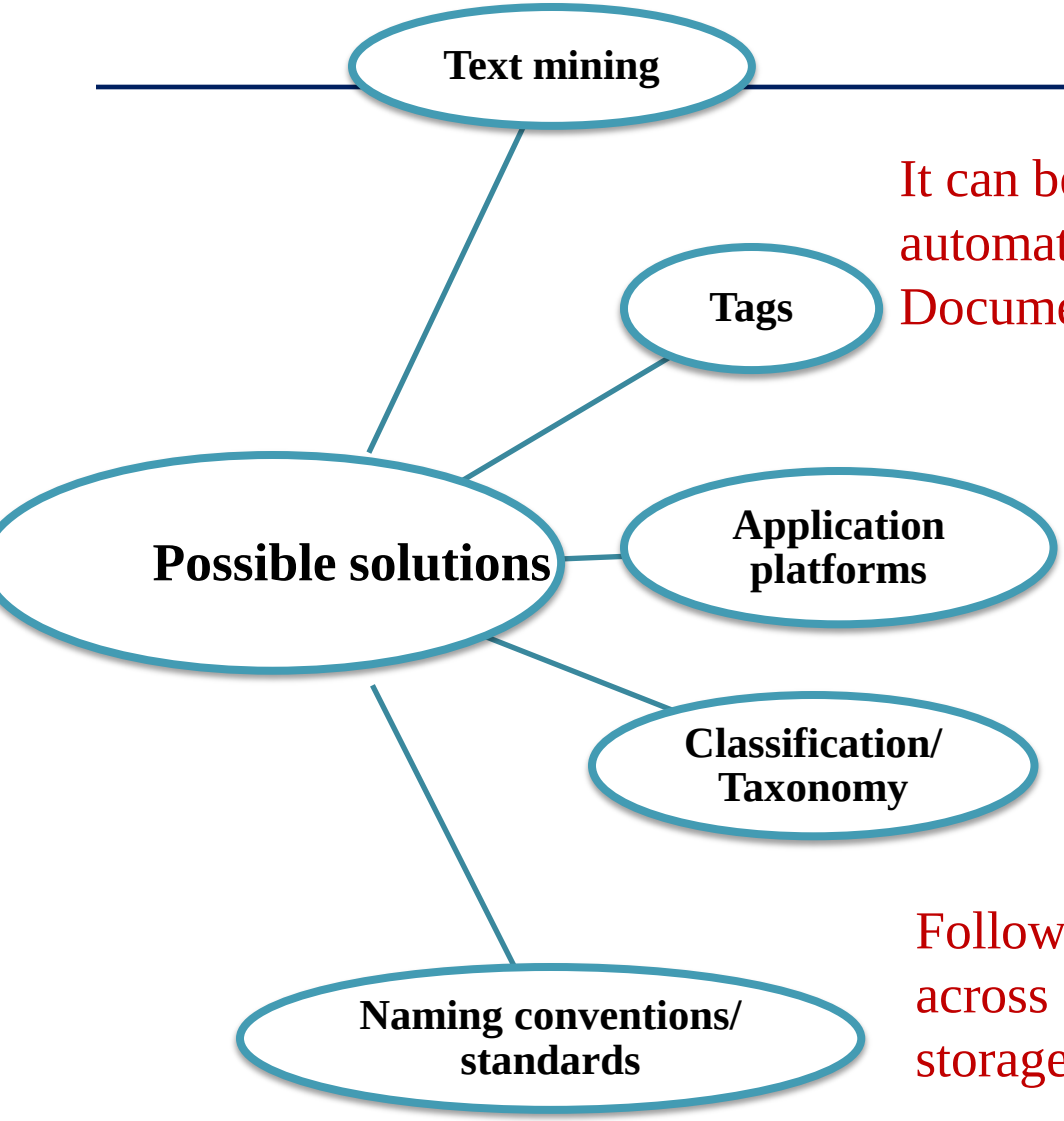
How to Extract Information from Unstructured Data?



Text mining tools help in grouping and classifying unstructured data and analyze by considering grammar, context, synonyms ,etc.

Taxonomies within the organization can be managed automatically to organize data in hierarchical structures

How to Extract Information from Unstructured Data?

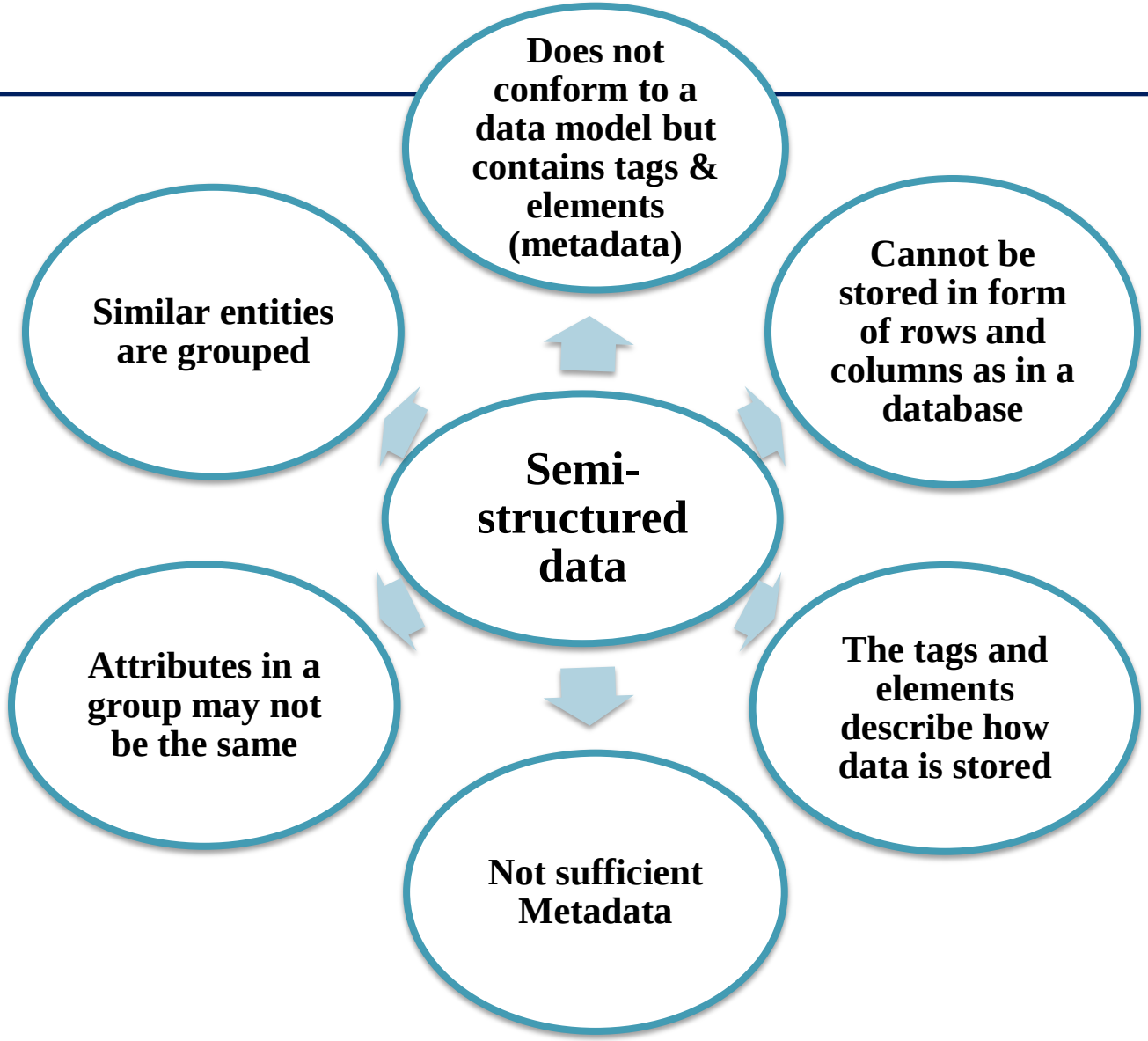


It can be stored in a virtual repository and be automatically tagged. For example, Documentum provides this type of solution

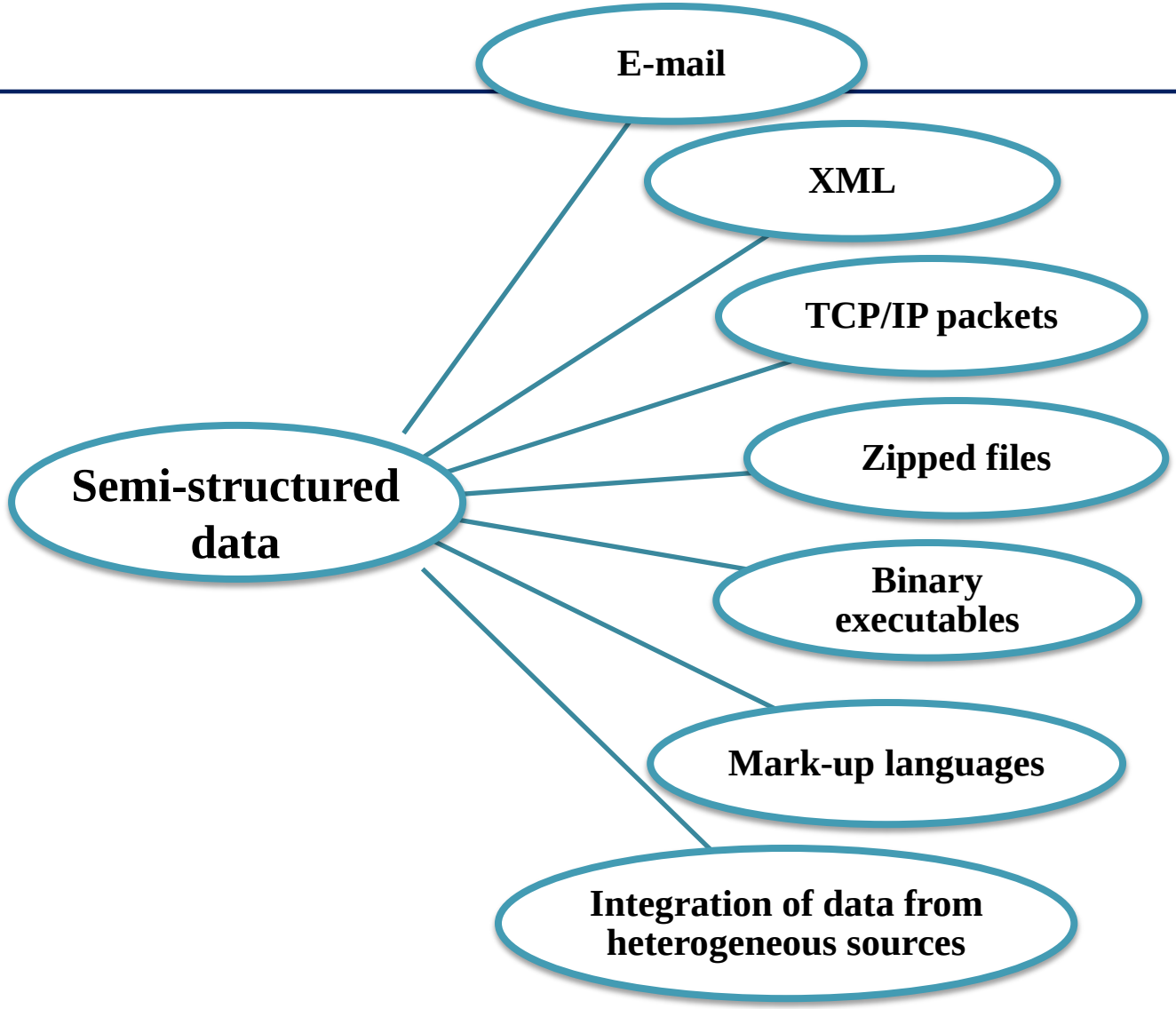
Application platforms like XOLAP help extract information from e-mail and XML based documents

Following naming conventions or standards across an organization can greatly improve storage and retrieval

What is Semi-structured Data?



Where does Semi-structured Data Come from?

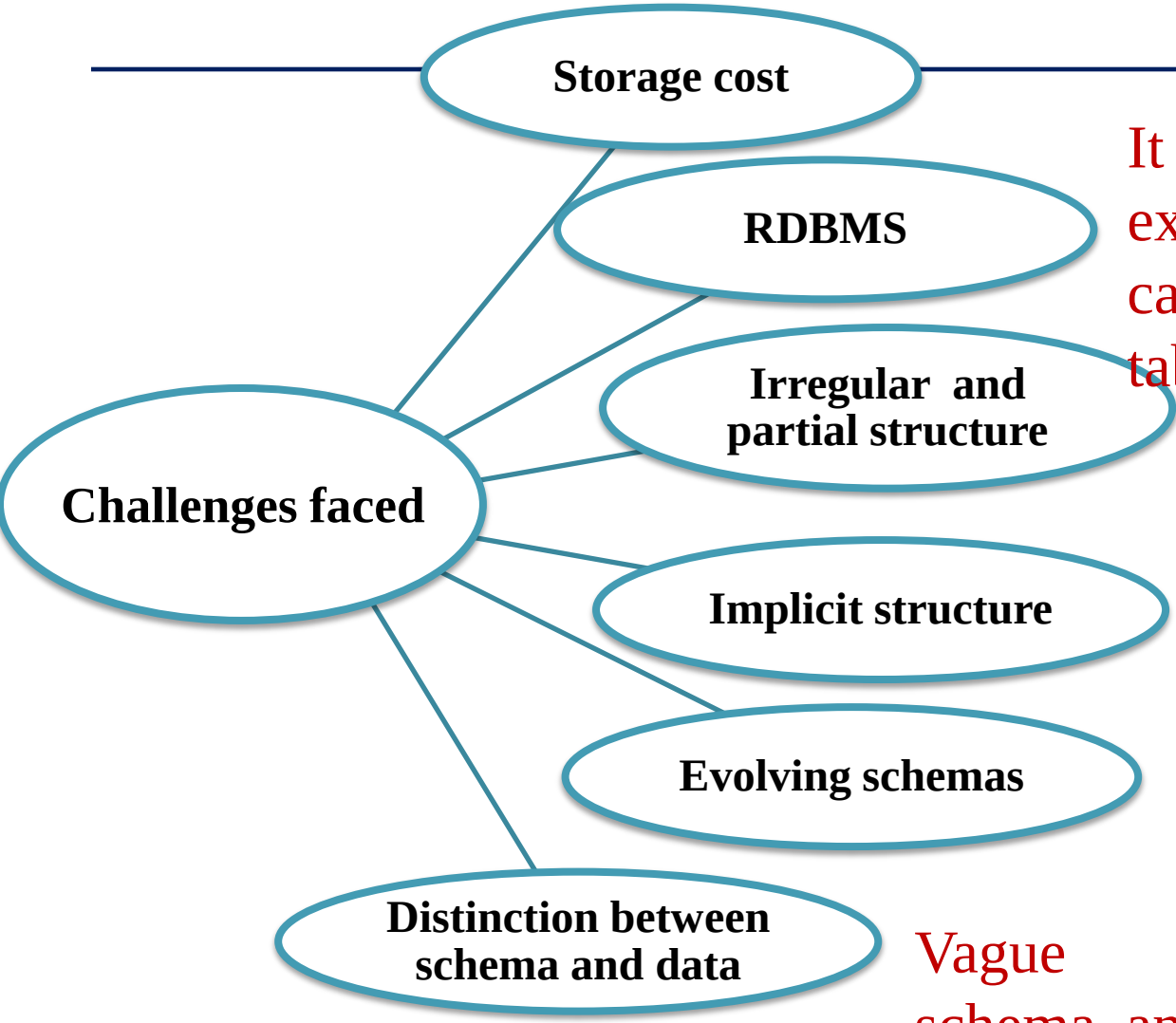


How to Manage Semi-structured Data?

Some ways in which semi-structured data is managed and stored

Schemas	Graph-based data models	XML
• Describe the structure and content of data to some extent • Assign meaning to data hence allowing automatic search and indexing	• Contain data on the leaves of the graph. Also known as ‘schema less’ • Used for data exchange among heterogeneous sources	• Models the data using tags and elements • Schemas are not tightly coupled to data

How to Store Semi-structured Data?

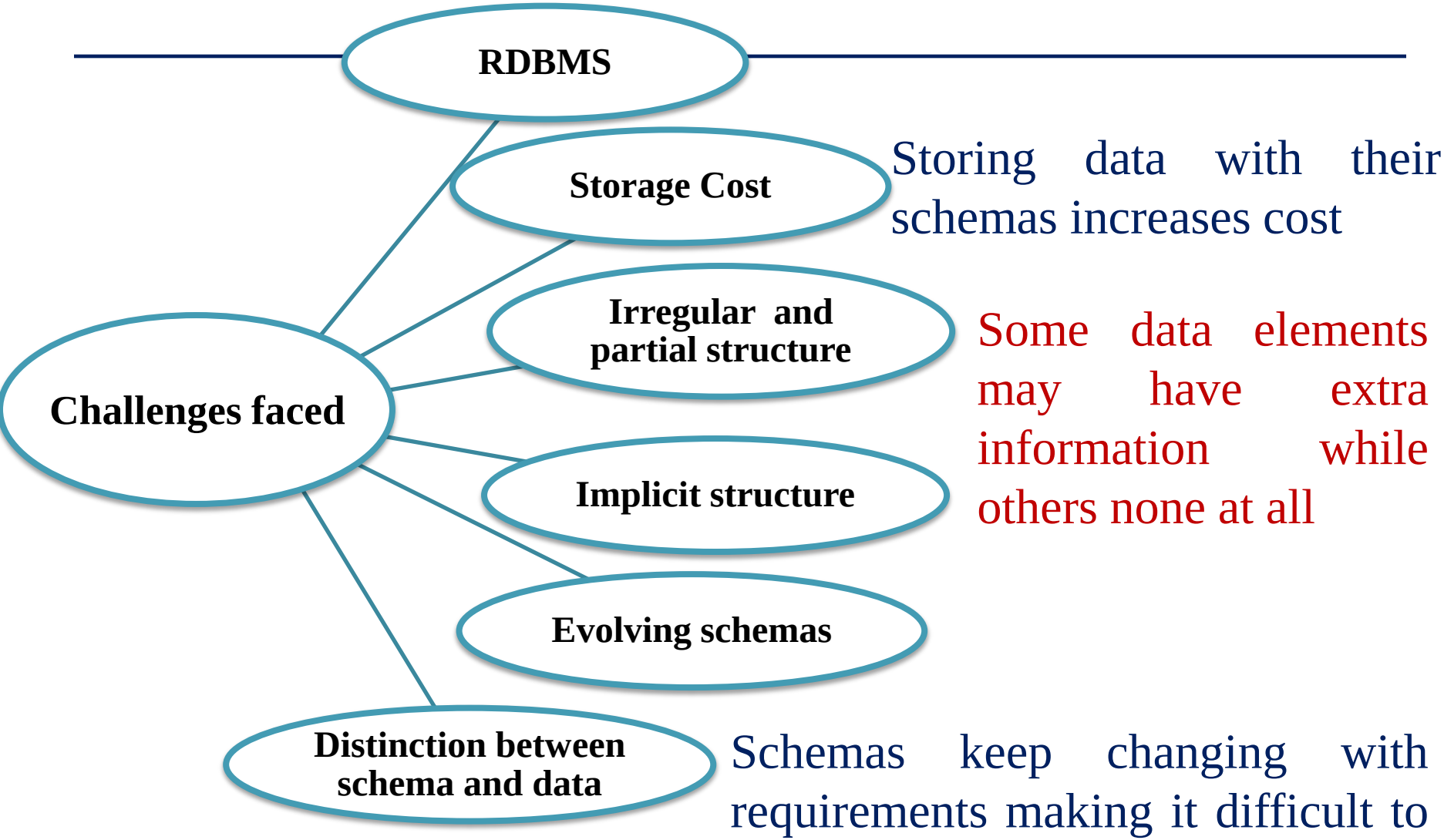


It cannot be stored in existing RDBMS as data cannot be mapped into tables directly

In many cases the structure is implicit. Interpreting relationships and correlations is very difficult

Vague distinction between schema and data exists at times making it difficult to capture data

How to Store Semi-structured Data?

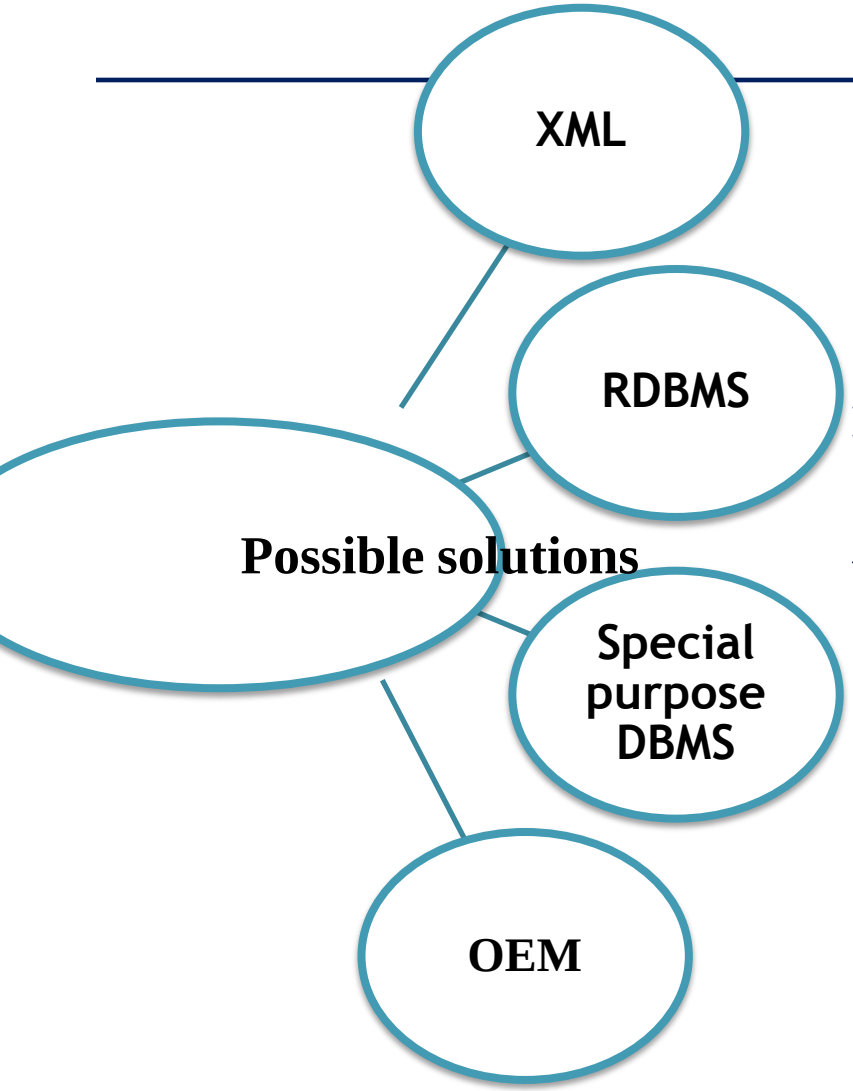


Storing data with their schemas increases cost

Some data elements may have extra information while others none at all

Schemas keep changing with requirements making it difficult to capture it in a database

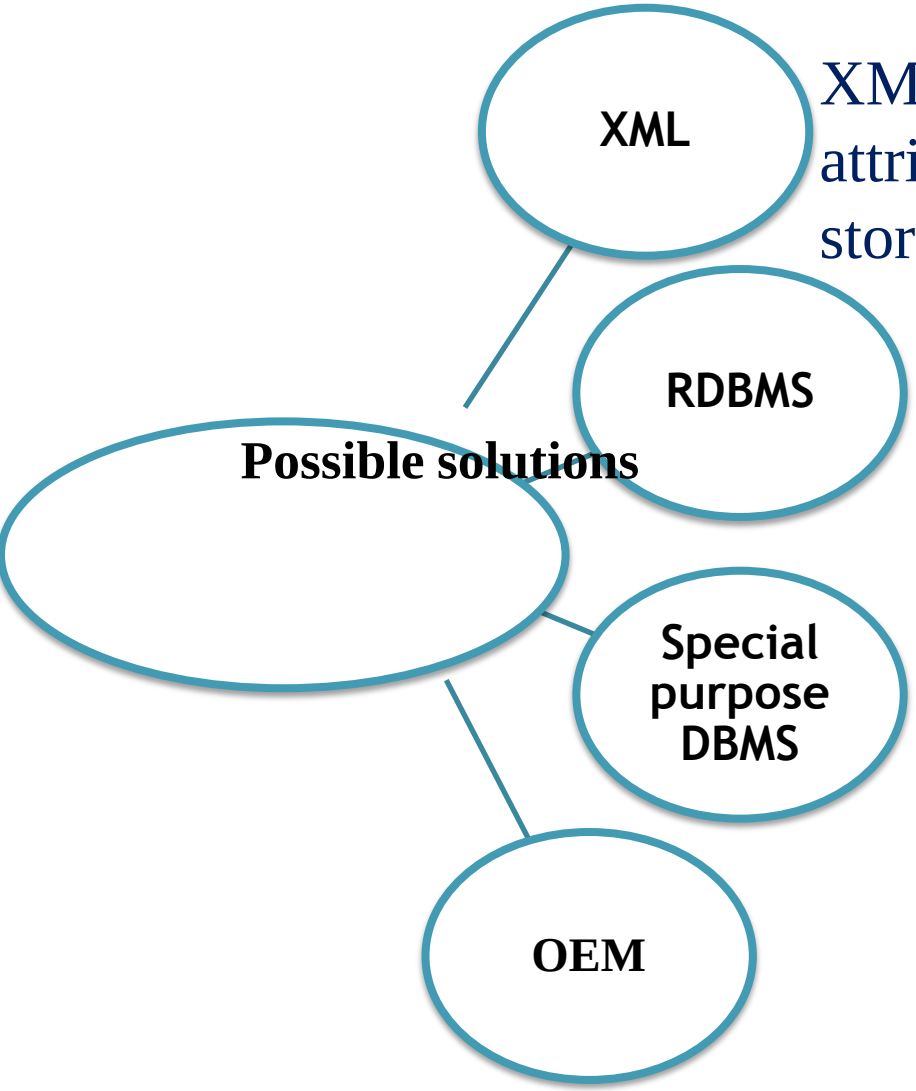
How to Store Semi-structured Data?



Semi-structured data can be stored in a relational database by mapping the data to a relational schema which is then mapped to a table

Object Exchange Model.
Data can be stored and exchanged in the form of graph where entities are represented as objects which are the vertices in a graph

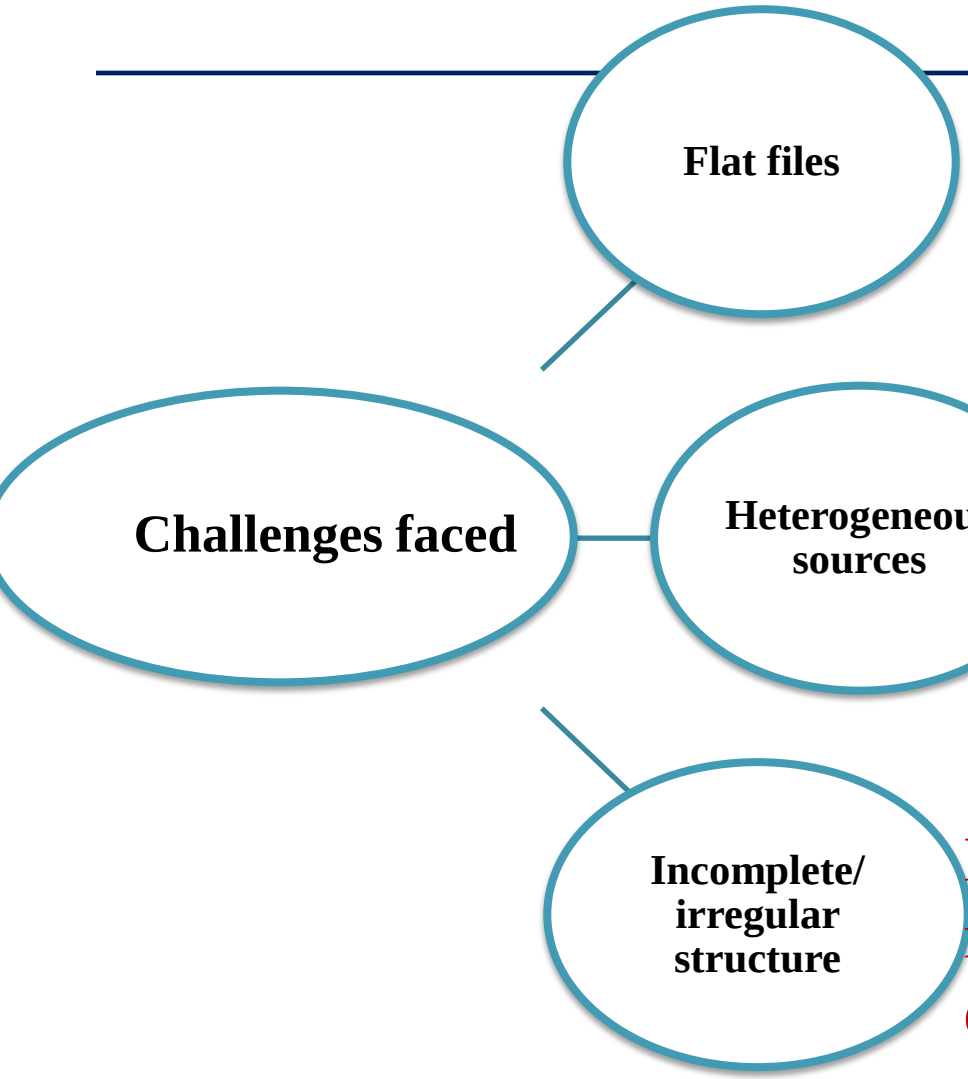
How to Store Semi-structured Data?



XML allows to define tags and attributes to store data. Data can be stored in a hierarchical/nested structure

Databases which are specifically designed to store semi-structured data

How to Extract Information from Semi-structured Data?

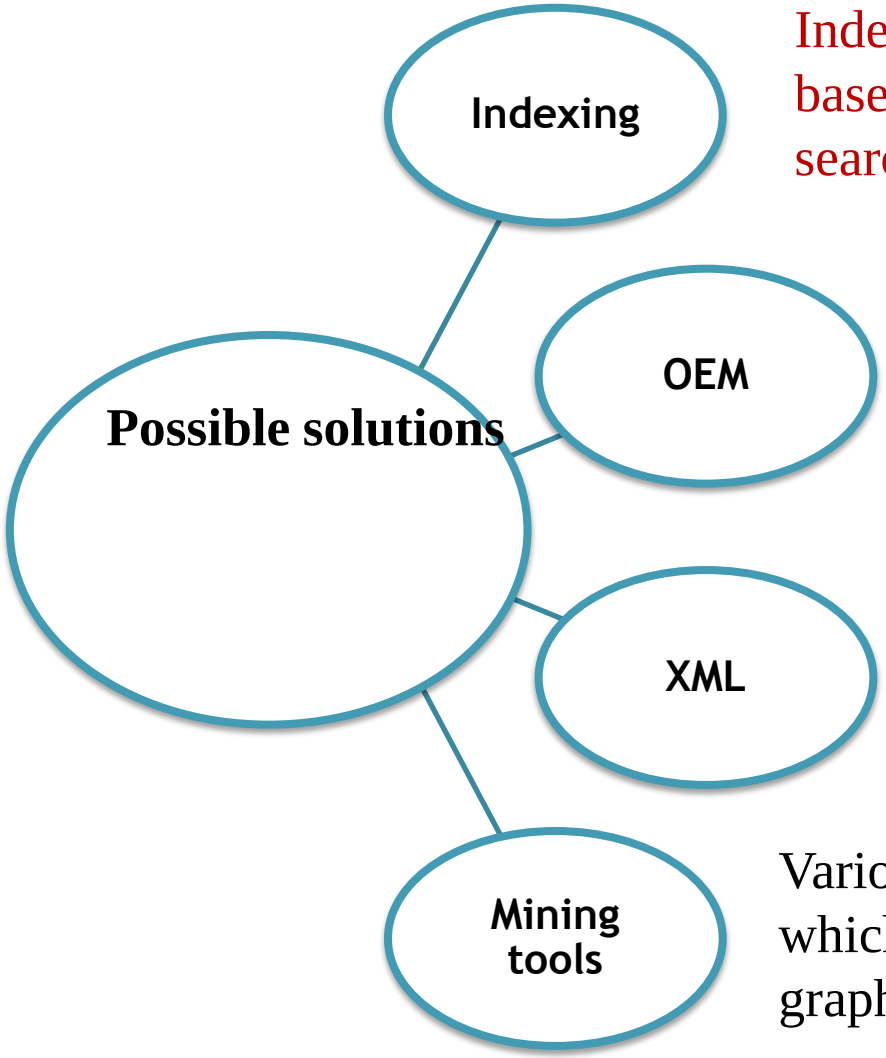


Semi-structured is usually stored in flat files which are difficult to index and search

Data comes from varied sources which is difficult to tag and search

Extracting structure when there is none and interpreting the relations existing in the structure which is present is a difficult task

How to Extract Information from Semi-structured Data?



Indexing data in a graph-based model enables quick search

Allows data to be stored in a graph-based data model which is easier to index and search

Allows data to be arranged in a hierarchical or tree-like structure which enables indexing and searching

Various mining tools are available which search data based on graphs, schemas, structure, etc.

XML – A Solution for Semi-structured Data Management

XML	Extensible MarkUp Language
What is XML?	Open-source mark up language written in plain text. It is hardware and software independent
Does what?	Designed to store and transport data over the Internet
How?	It allows data to be stored in a hierarchical/nested structure. It allows user to define tags to store the data

XML – A Solution for Semi-structured Data Management

XML has no predefined tags

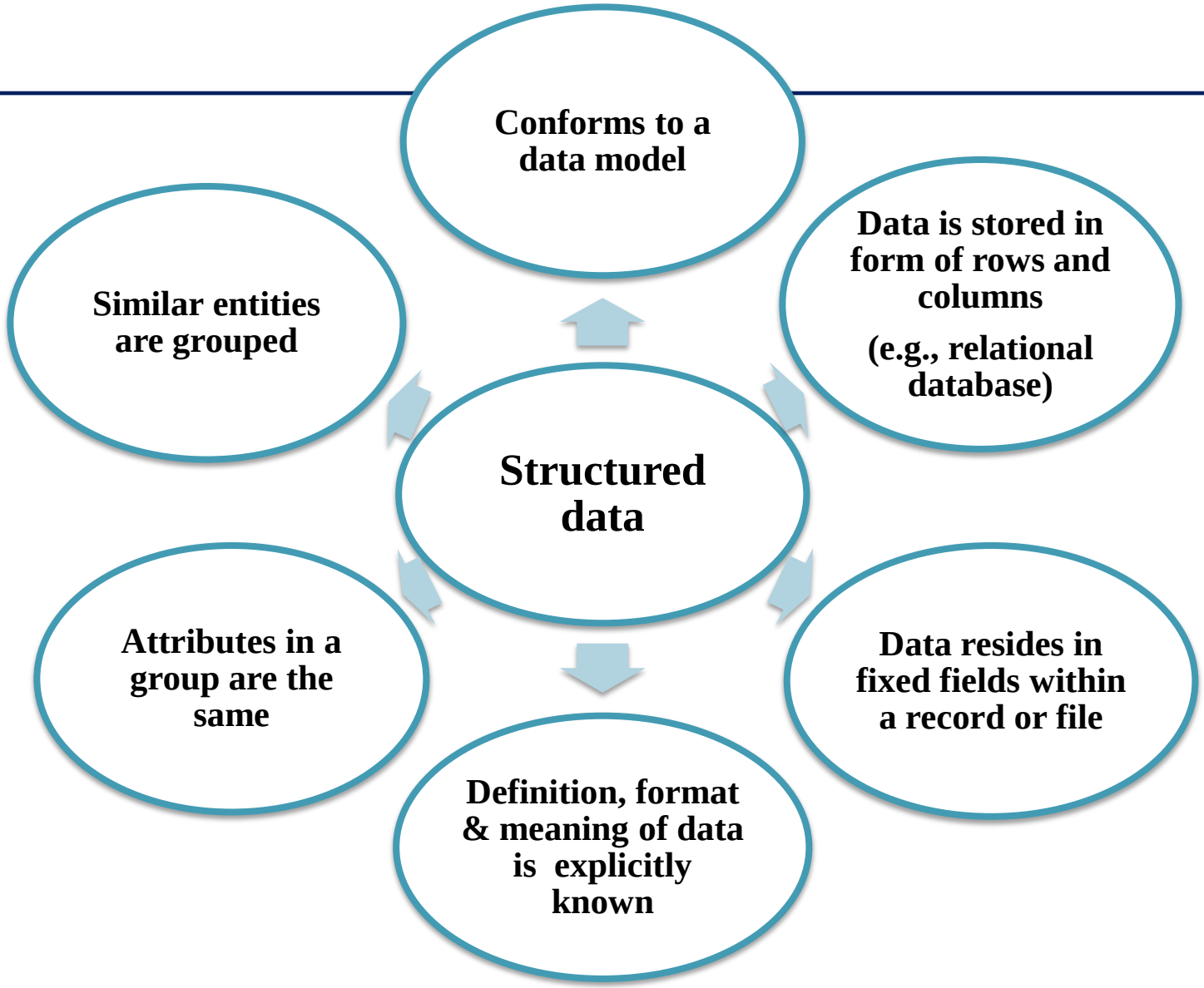
```
<message>
<to> XYZ </to>
<from> ABC </from>
<subject> Greetings </subject>
<body> Hello! How are you? </body>
</message>
```

The words in the <> (angular brackets) are user-defined tags

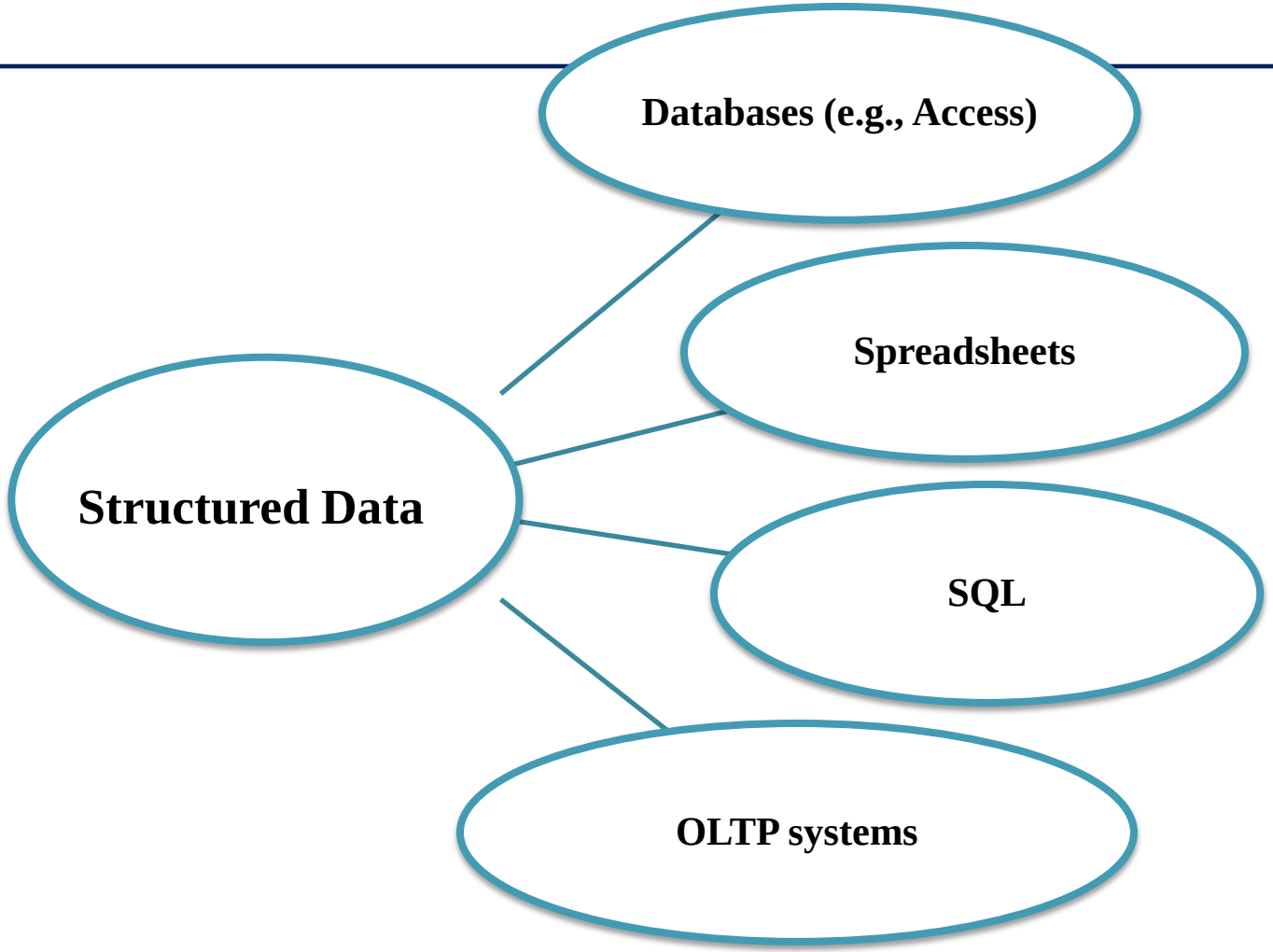
XML is known as self-describing as data can exist without a schema and schema can be added later

Schema can be described in XSLT or XML schema

What Is Structured Data?



Where does Structured Data Come from?



Structured Data: Everything in its Place

- **Fully described datasets**
- **Clearly defined categories and sub-categories**
- **Data neatly placed in rows and columns**
- **Data that goes into the records is regulated by a well-defined structure**
- **Indexing can be easily done either by the DBMS itself or manually**

Structured Data

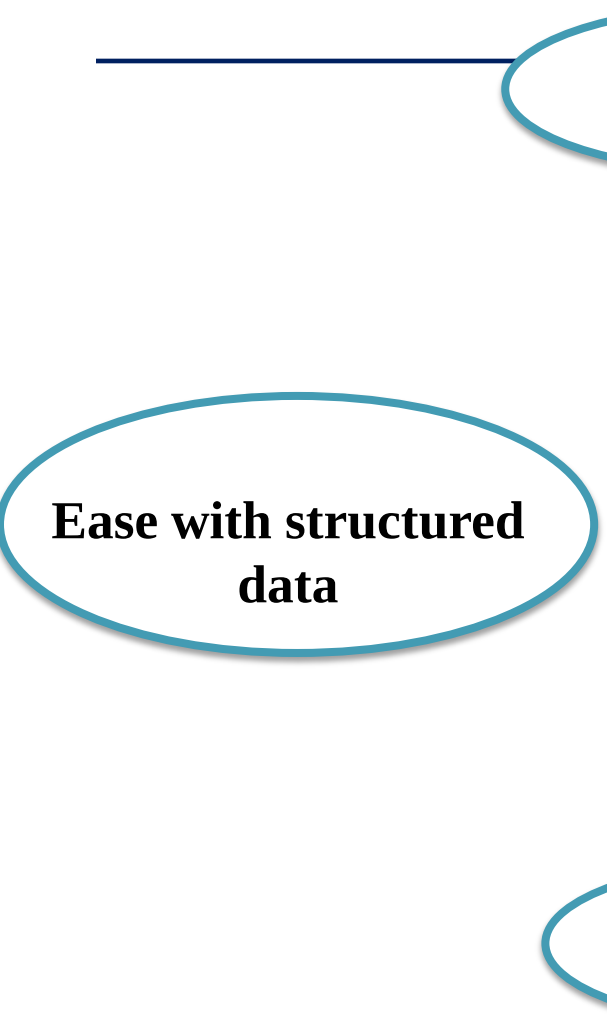
Semi-structured

Name	E-mail
Patrick Wood	ptw@dcs.abc.ac.uk, p.wood@ymail.uk
First name: Mark Last name: Taylor	MarkT@dcs.ymail.ac.uk
Alex Bourdoo	AlexBourdoo@dcs.ymail.ac.uk

Structured

First Name	Last Name	E-mail Id	Alternate E-mail Id
Patrick	Wood	ptw@dcs.abc.ac.uk	p.wood@ymail.uk
Mark	Taylor	MarkT@dcs.ymail.ac.uk	
Alex	Bourdoo	AlexBourdoo@dcs.ymail.ac.uk	

Ease with Structured Data-Storage

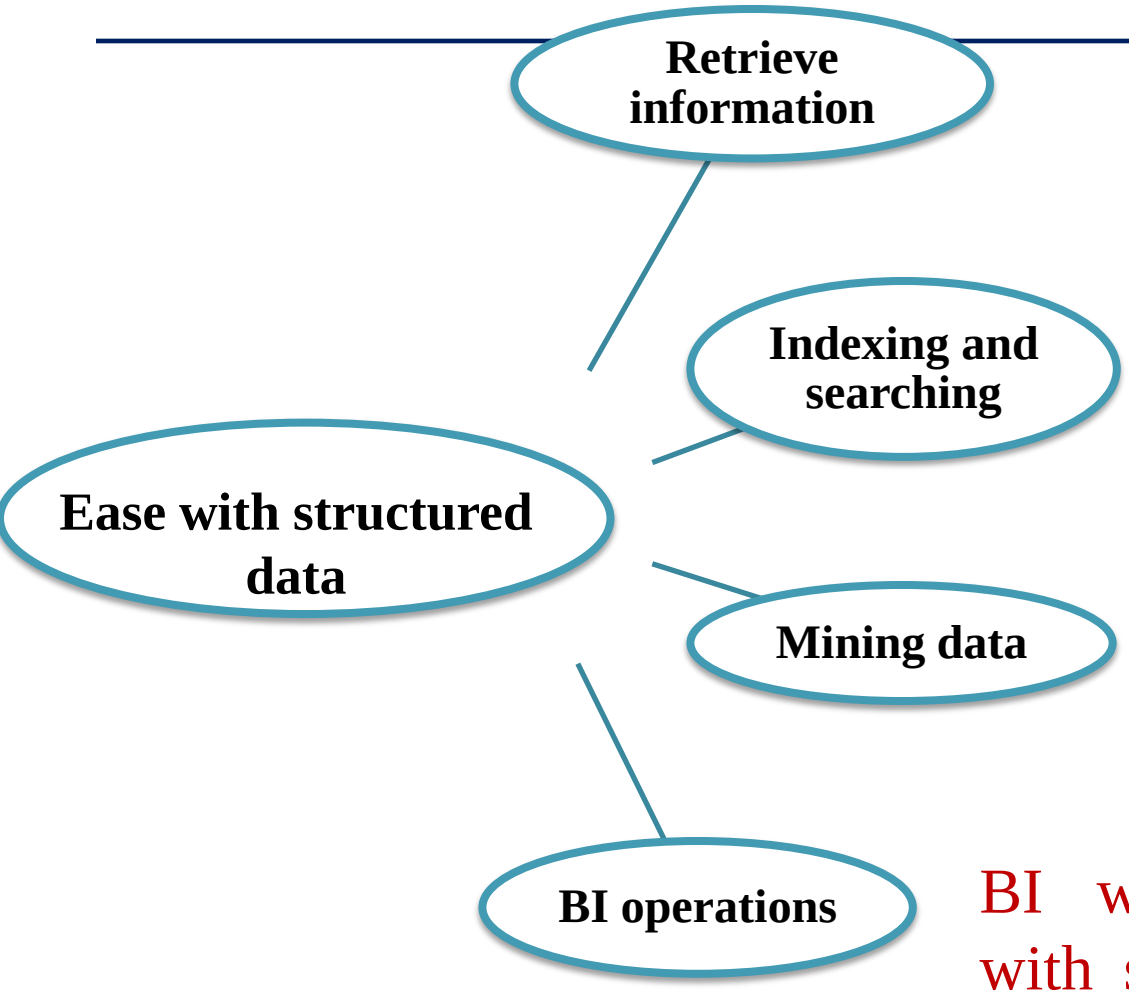


Data types – both defined and user defined help with the storage of structured data

Scalability is not generally an issue with increase in data

Updating, deleting, etc. is easy due to structured form

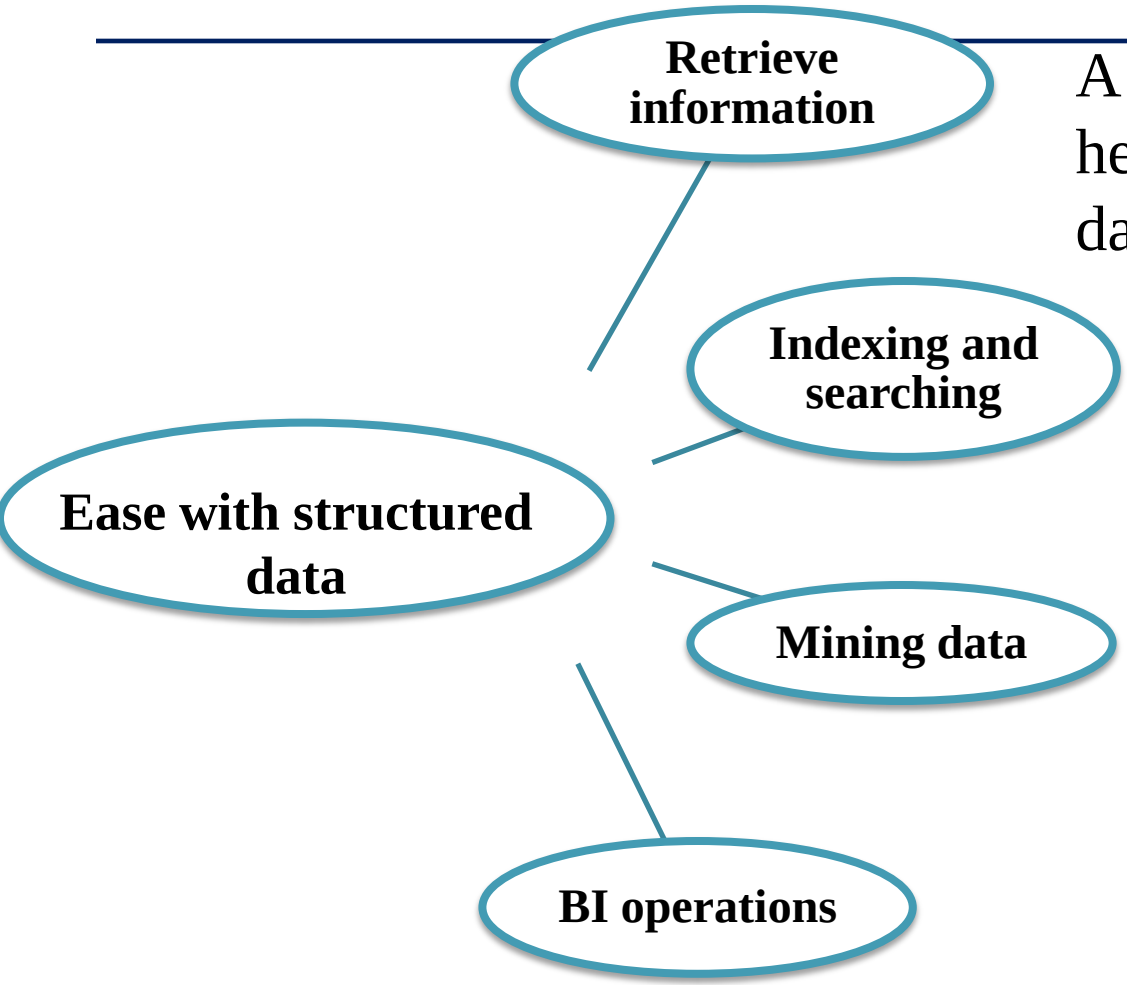
Ease with Structured Data-Retrieval



Data can be indexed based not only on a text string but other attributes as well. This enables streamlined search

BI works extremely well with structured data. Hence data mining, warehousing, etc. can be easily undertaken

Ease with Structured Data-Retrieval



A well-defined structure helps in easy retrieval of data

Structured data can be easily mined and knowledge can be extracted from it

Data Warehouse

Architecture and Role in BI

Definitions

- 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

Features of data Warehouse

- Following are the major features of data warehouse:
  - Subject Oriented
  - Integrated
  - Time variant
  - Non volatile

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*

Conceptual Modeling of Data Warehouses

- Modeling data warehouses: dimensions & measures instead of relational model
- Subject, facilitates on-line data analysis oriented
- Most popular model is the multidimensional model
- Most common modeling paradigm:
  - **Star schema**
  - Data warehouse contains a **large** central table (**fact table**)
    - Contains the data without redundancy
  - A set of **dimension tables** (each for each dimension)

Fact Table

- A fact table is the central table in a star schema of a data warehouse.
- A fact table stores quantitative information for analysis and is often de-normalized.
- A fact table holds the data to be analyzed.
- A dimension table stores data about the ways in which the data in the fact table can be analyzed.

Example of Fact Table

- Suppose that a company sells products to customers. Every sale is a fact that happens, and the fact table is used to record these facts. For example:

Time ID	Product ID	Customer ID	Unit Sold
4	17	2	1
8	21	3	2
8	4	1	1

Dimension Table

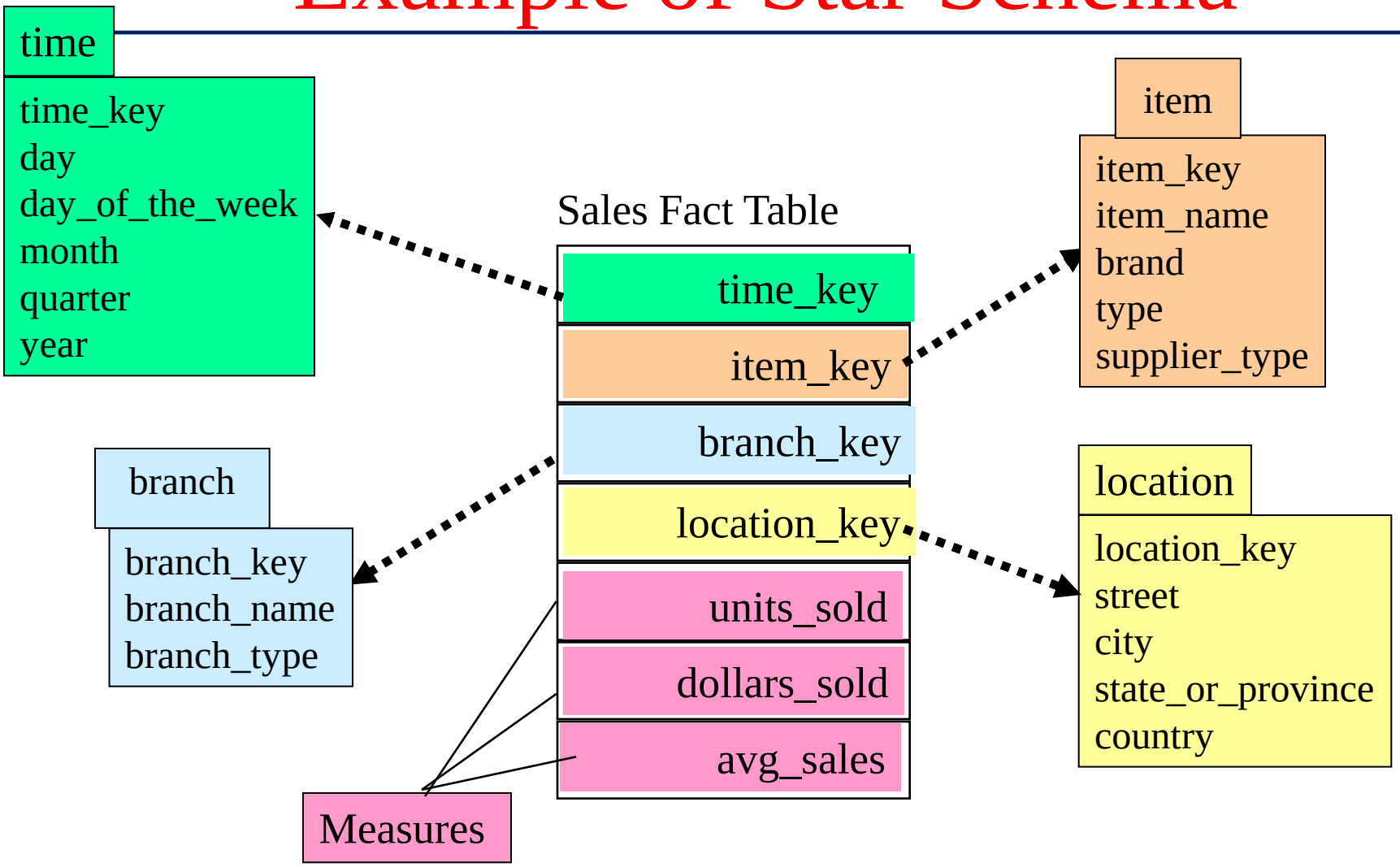
- A dimension table stores data about the ways in which the data in the fact table can be analyzed.
- A dimension table is analogous to relations of an RDBMS.

Example of Dimension Table

- Now we can add a dimension table about customers:

Customer ID	Name	Gender	Income	Education	Region
1	Brian Edge	M	2	3	4
2	Fred Smith	M	3	5	1
3	Sally Jones	F	1	7	3

Example of Star Schema



Advantages of Star Schema

- **Simpler Queries:**
  - Join logic of star schema is quite cinch in compare to other join logic which are needed to fetch data from a transactional schema that is highly normalized.
- **Simplified Business Reporting Logic:**
  - In compared to a transactional schema that is highly normalized, the star schema makes simpler common business reporting logic, such as as-of reporting and period-over-period.
- **Feeding Cubes:**
  - Star schema is widely used by all OLAP systems to design OLAP cubes efficiently.
  - In fact, major OLAP systems deliver a ROLAP mode of operation which can use a star schema as a source without designing a cube structure.

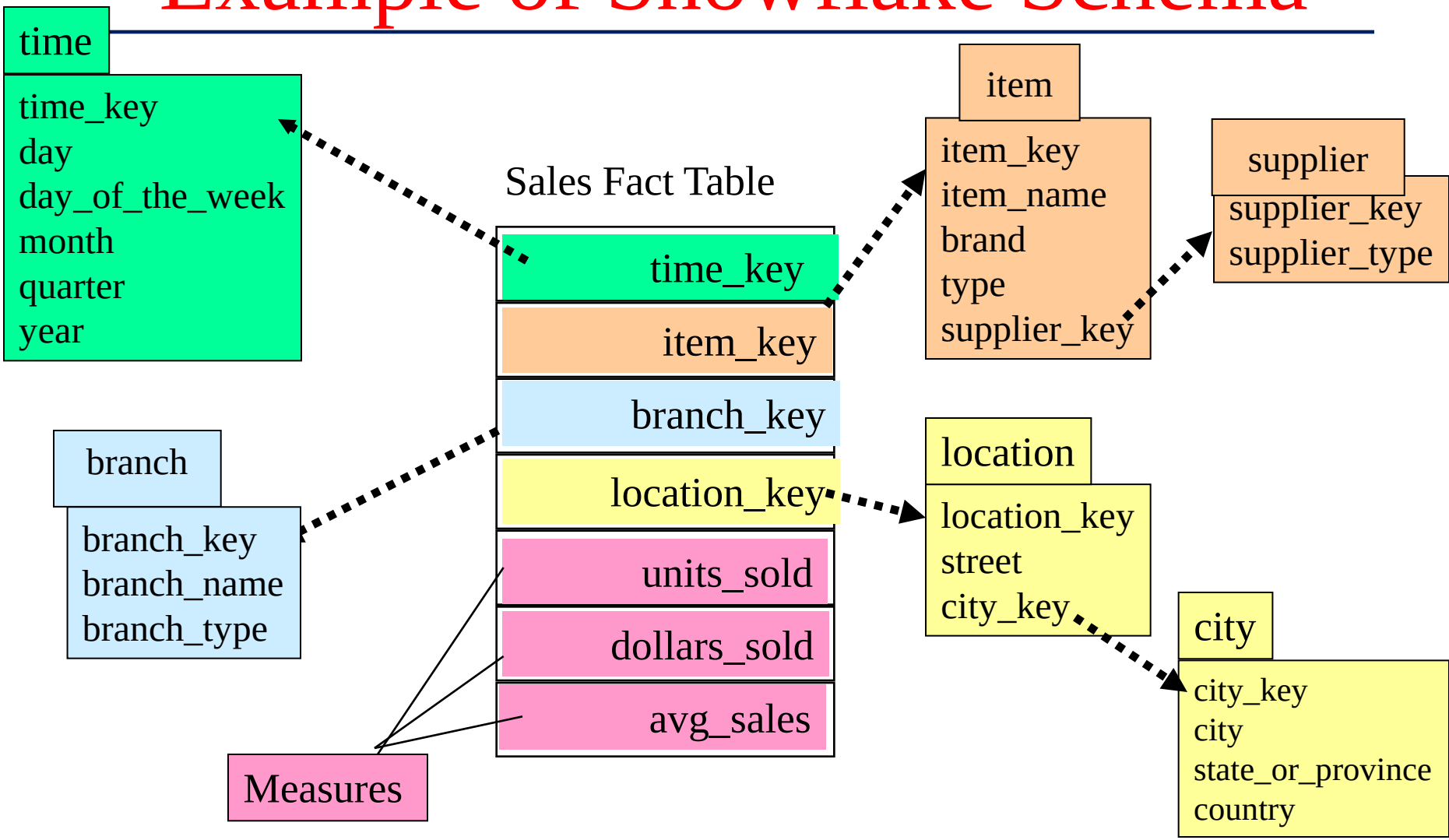
Disadvantages of Star Schema

- Data integrity is not enforced well since in a highly de-normalized schema state.
- Not flexible in terms if analytical needs as a normalized data model.
- Star schemas don’t reinforce many-to-many relationships within business entities – at least not frequently.

Snowflake schema

- 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

Example of Snowflake Schema



Characteristics of snowflake schema

- The snowflake schema uses small disk space.
- It is easy to implement dimension is added to schema.
- There are multiple tables, so performance is reduced.
- The dimension table consist of two or more sets of attributes which define information at different grains.
- The sets of attributes of the same dimension table are being populate by different source systems.

Advantages

- There are two main advantages of snowflake schema given below:
- It provides structured data which reduces the the problem of data integrity.
- It uses small disk space because data are highly structured.

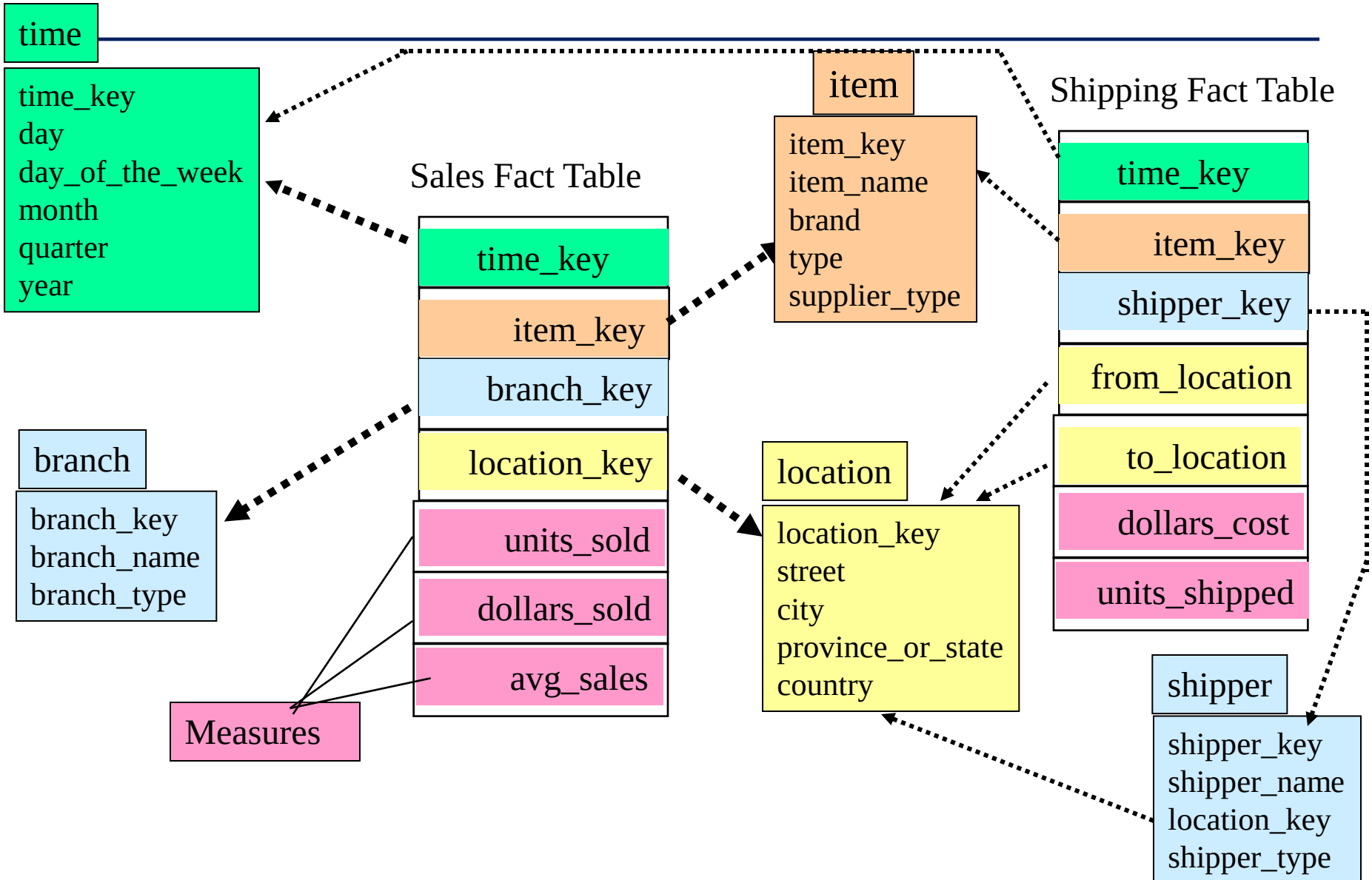
Disadvantages

- Snowflaking reduces space consumed by dimension tables, but compared with the entire data warehouse the saving is usually insignificant.
- Avoid snowflaking or normalization of a dimension table, unless required and appropriate.
- Do not snowflake hierarchies of one dimension table into separate tables. Hierarchies should belong to the dimension table only and should never be snowfalked.
- Multiple hierarchies can belong to the same dimension has been designed at the lowest possible detail.

Fact constellations

- Fact constellations: **Multiple** fact tables share dimension tables, viewed as a collection of stars, therefore called galaxy schema or fact constellation

Example of Fact Constellation



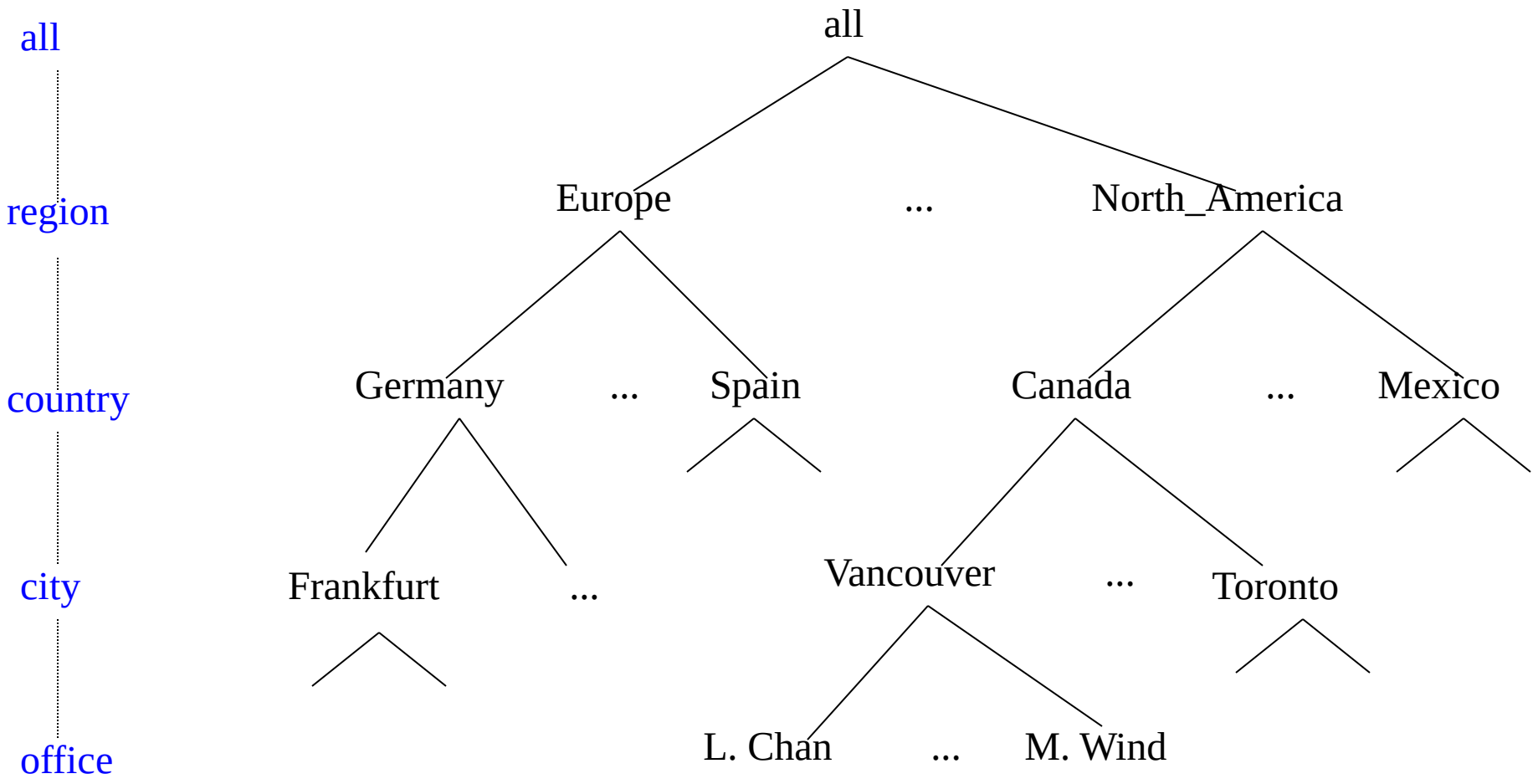
Difference Between Star & Snowflake

S.NO	STAR SCHEMA	SNOWFLAKE SCHEMA
1.	In star schema, The fact tables and the dimension tables are contained.	While in snowflake schema, The fact tables, dimension tables as well as sub dimension tables are contained.
2.	Star schema is a top-down model.	While it is a bottom-up model.
3.	Star schema uses more space.	While it uses less space.
4.	It takes less time for the execution of queries.	While it takes more time than star schema for the execution of queries.
5.	In star schema, Normalization is not used.	While in this, Both normalization and denormalization are used.
6.	It’s design is very simple.	While it’s design is complex.
7.	The query complexity of star schema is low.	While the query complexity of snowflake schema is higher than star schema.
8.	It’s understanding is very simple.	While it’s understanding is difficult.

Hierarchies

- Independent variables are often related in hierarchies (taxonomy)
  - Determine ways in which dependent data can be aggregated
- Temporal hierarchy
  - Seconds, minutes, hours, days, weeks, months, years
- Same data can be aggregated in many different ways
  - Same independent variable can belong to different hierarchies

Hierarchy - Location



Exercise

Design the data warehouse for a wholesale furniture company. The data warehouse has to allow to analyze the company's situation at least with respect to the Furniture, Customers and Time.
Moreover, the company needs to analyze:

- the furniture with respect to its type (chair, table, wardrobe, cabinet. . .), category (kitchen, living room, bedroom, bathroom, office. . .) and material (wood, marble. . .)
- the customers with respect to their spatial location, by considering at least cities, regions and states

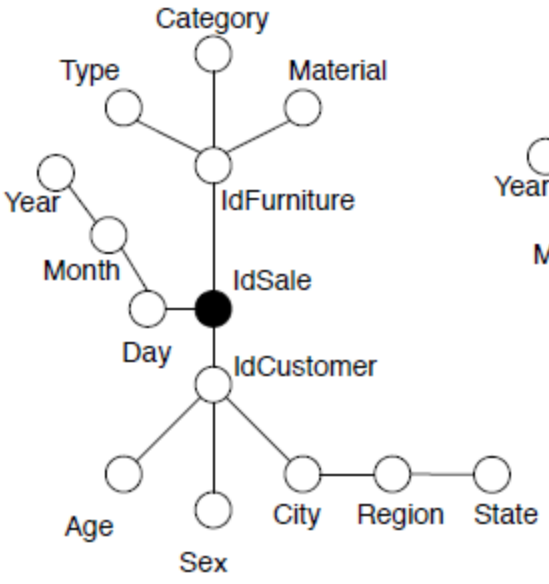
The company is interested in learning at least the quantity, income and discount of its sales.

Exercise

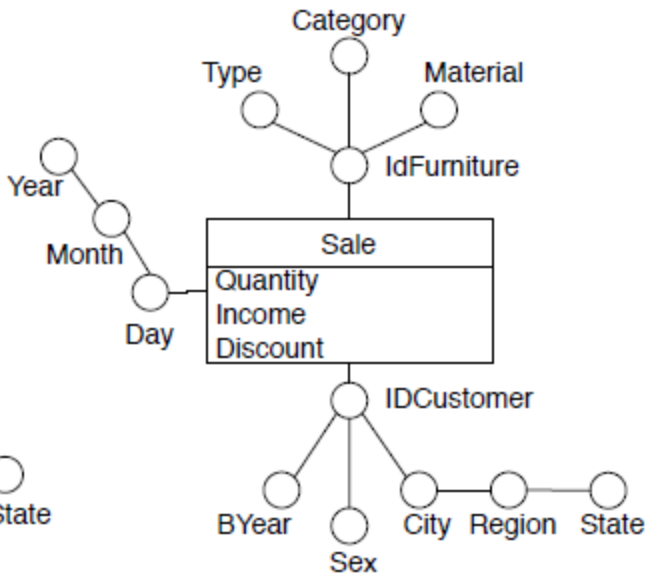
- 1. Identify facts, dimensions and measures
- 2. For each fact:
  - produce the attribute tree and fact schema
  - design the star or snowflake schema

Possible Solution

FACT Sales
MEASURES Quantity, Income, Discount
DIMENSIONS Furniture (Type, Category, Material)
Customer (Age, Sex, City → Region → State)
Time (Day → Month → Year)

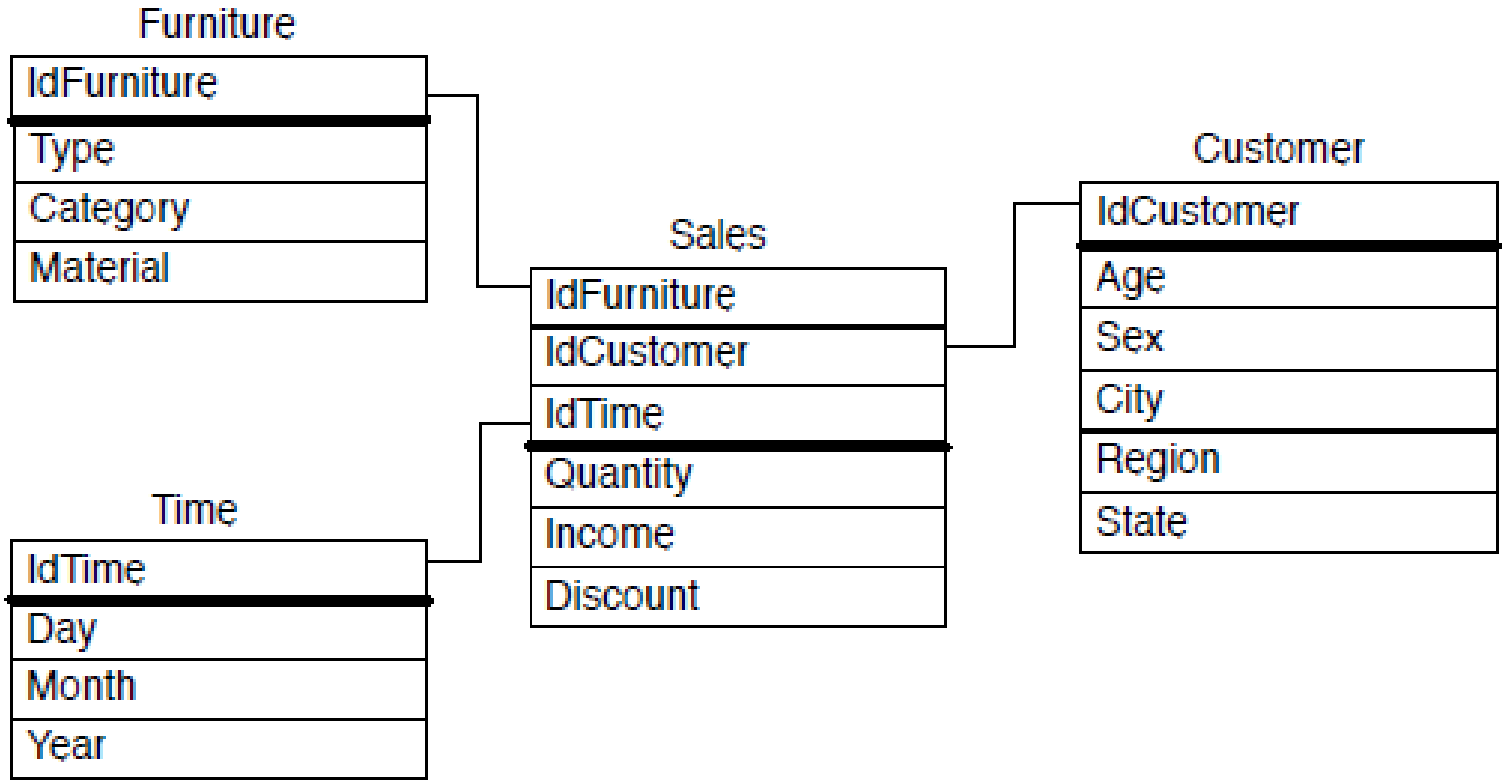


Attribute tree



Fact schema

Possible Solution- Star Schema

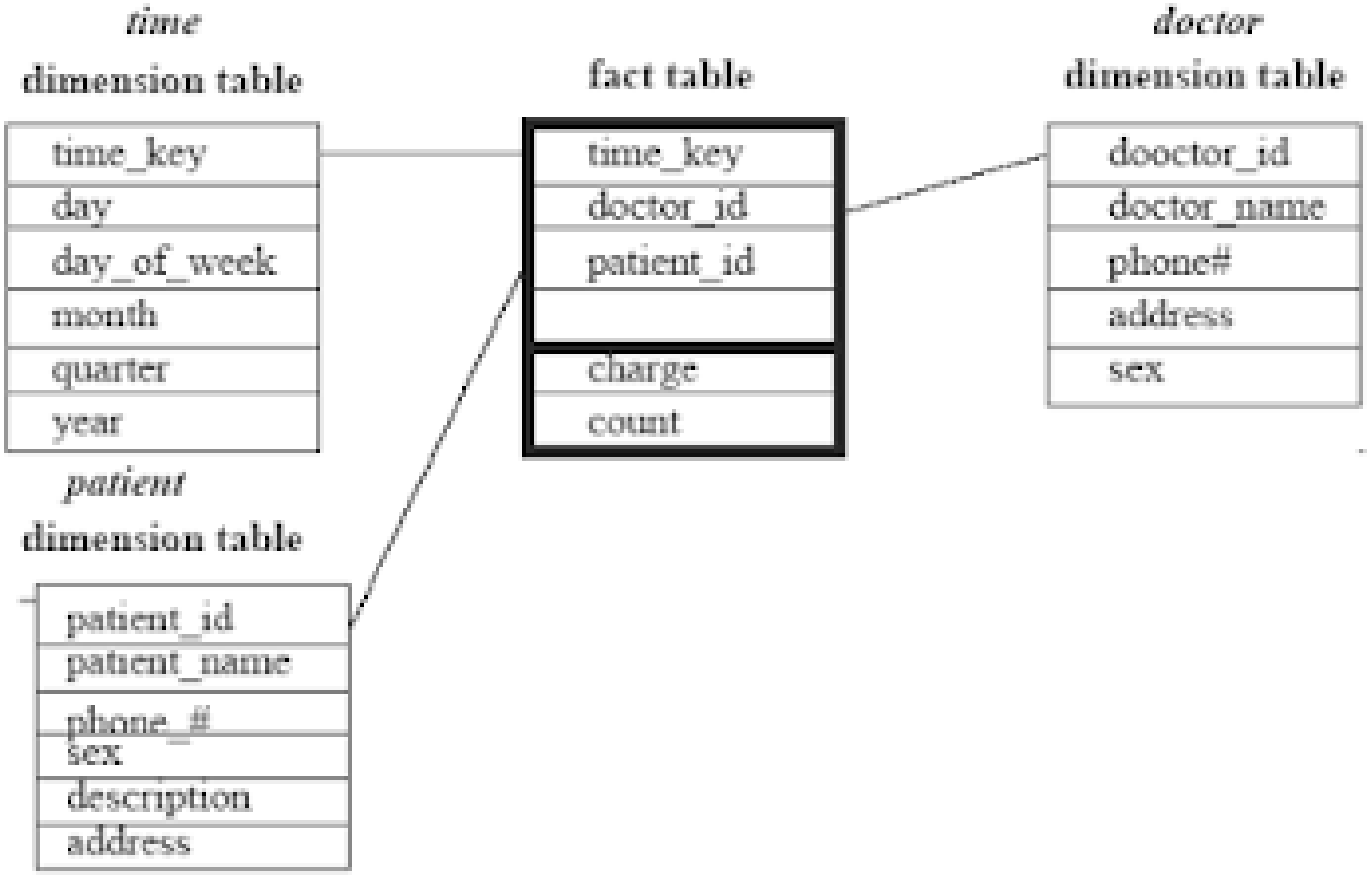


Exercise

Suppose that you are working as a Software Engineer on a project of developing a data warehouse for a hospital. It should consist of at least three dimensions Doctor, Patient and Time, and the two measures count and charge (fee charged by the doctor from patients for a single visit).

Draw the star schema diagram for the above Data Warehouse.

Possible Solution

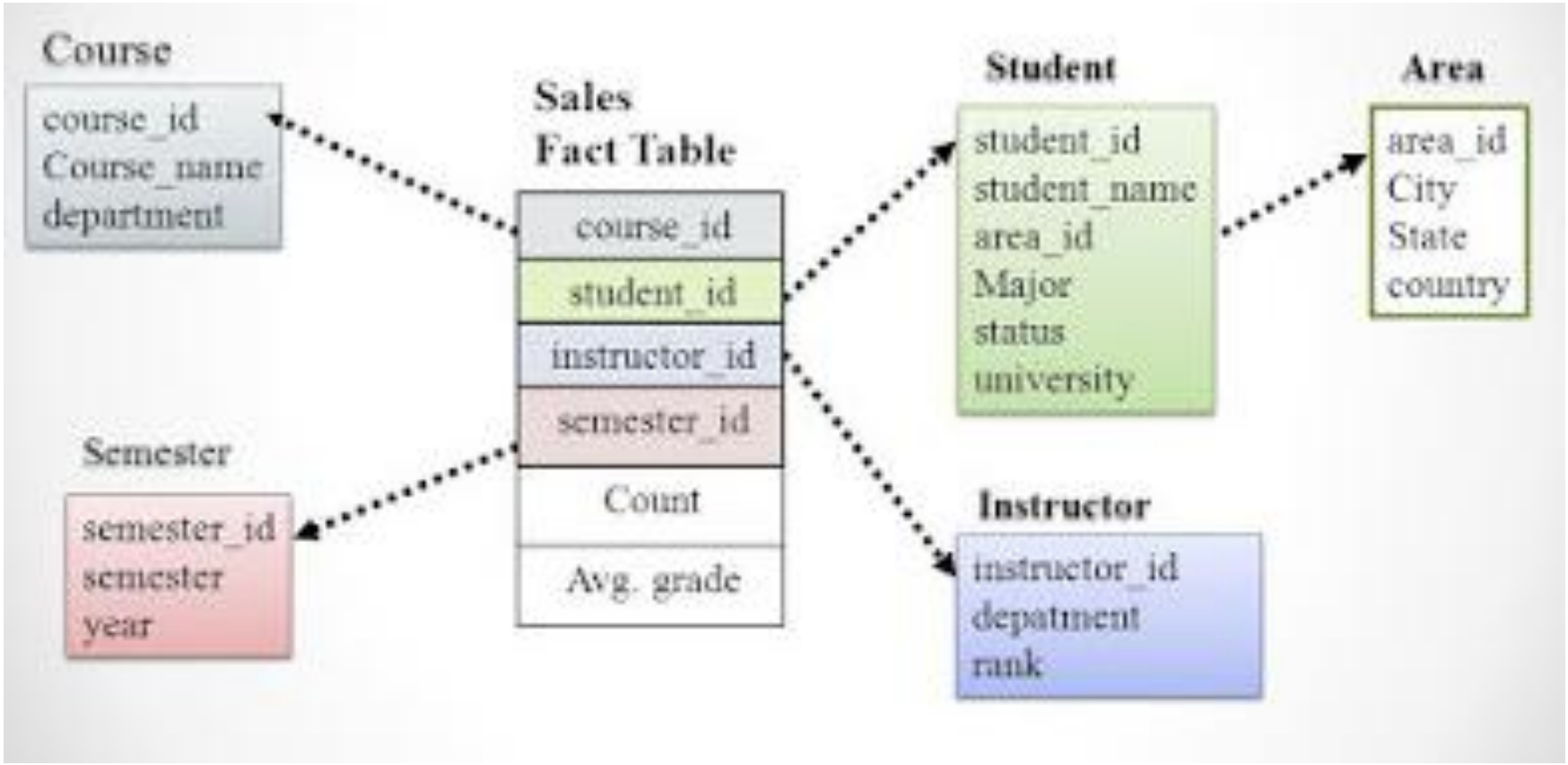


Exercise

Suppose that you are working on a project of developing a Data Warehouse for a University, which should consist of at least four dimensions: student, semester, course, and instructor, and two measures count and avg_grade.

Draw a snow-flake schema diagram for the above Data Warehouse.

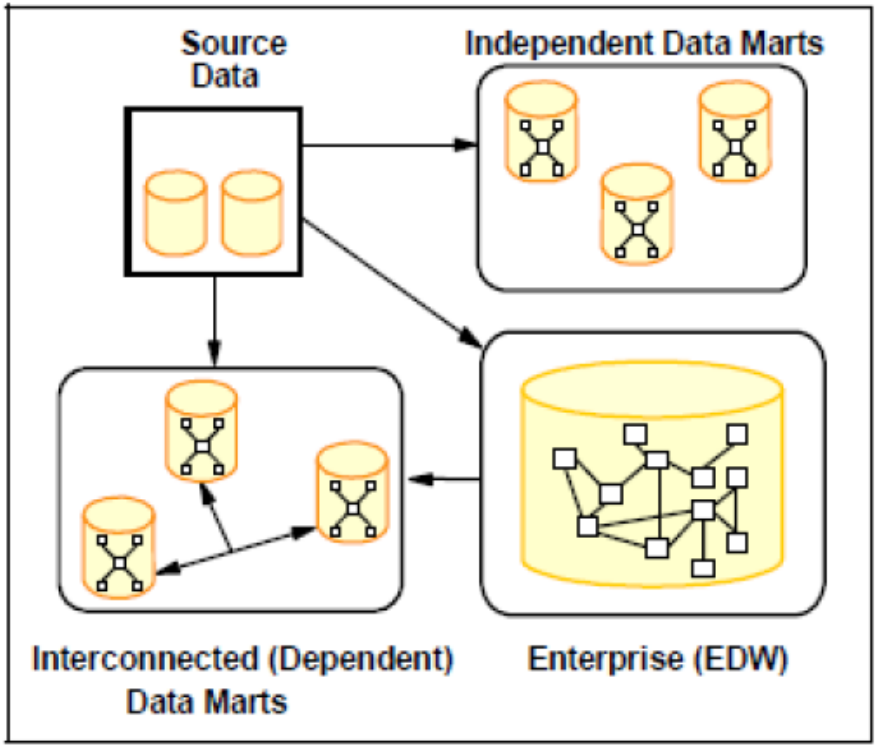
Possible Solution



DW Architecture

The three architectural approaches, listed below:

- Enterprise data warehouse (EDW)
- Independent data marts
- Dependent data marts



Various data warehouse architectures

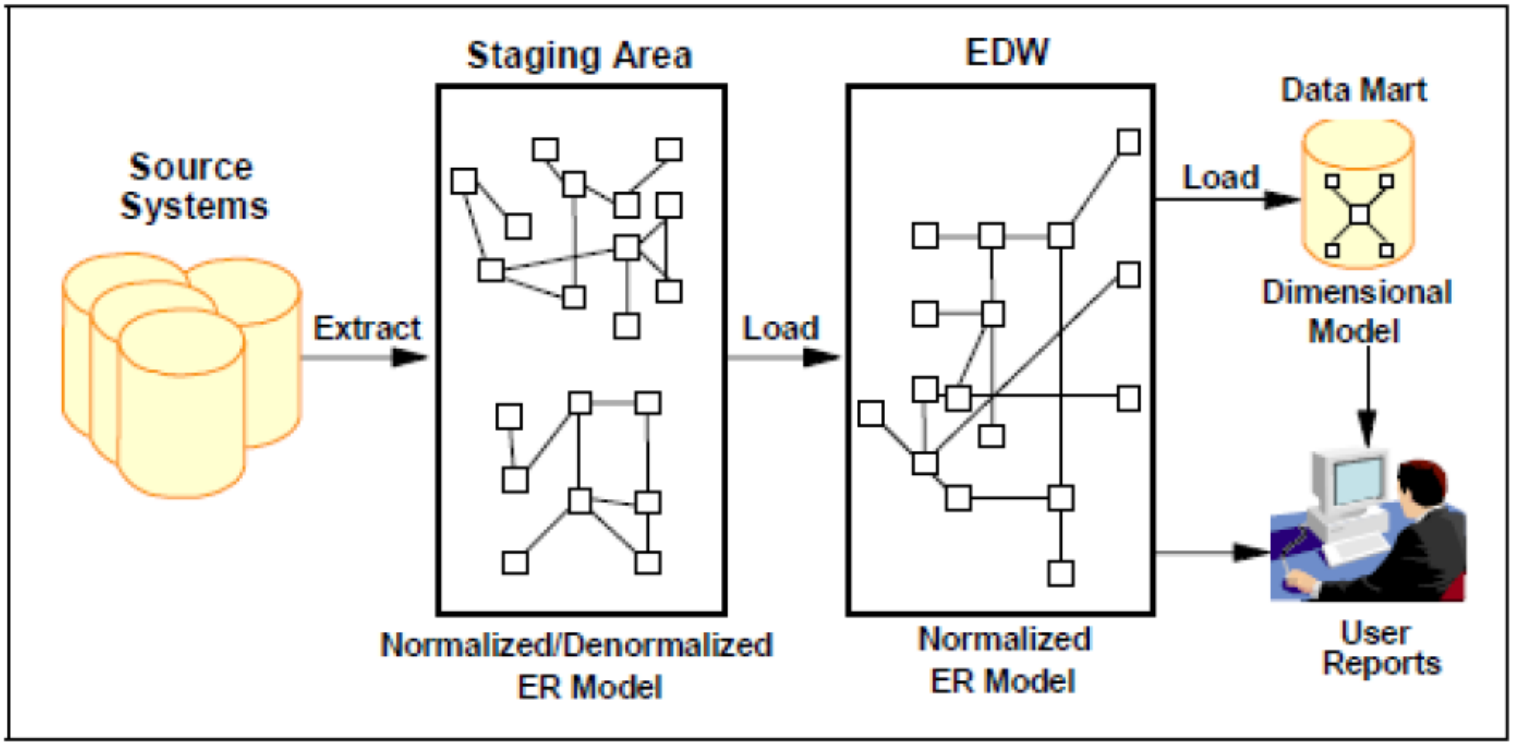
Data Warehouse vs. Data Marts

- *Enterprise warehouse*: collects all information about subjects (`customers,products,sales,assets, personnel`) that span the entire organization
  - Requires extensive business modeling (may take years to design and build)
- *Data Marts*: Departmental subsets that focus on selected subjects
  - Marketing data mart: customer, product, sales
  - Faster roll out, but complex integration in the long run
- *Virtual warehouse*: views over operational dbs
  - Materialize sel. summary views for efficient query processing
  - Easy to build but require excess capability on operat. db servers

Difference Between Data Warehouse and Data Mart

S.NO	DATA WAREHOUSE	DATA MART
1.	Data warehouse is a Centralised system.	While it is a decentralised system.
2.	In data warehouse, lightly denormalization takes place.	While in Data mart, highly denormalization takes place.
3.	Data warehouse is top-down model.	While it is a bottom-up model.
4.	To built a warehouse is difficult.	While to build a mart is easy.
5.	In data warehouse, Fact constellation schema is used.	While in this, Star schema and snowflake schema are used.
6.	Data Warehouse is flexible.	While it is not flexible.
7.	Data Warehouse is the data-oriented in nature.	While it is the project-oriented in nature.
8.	Data Ware house is long life.	While data-mart is short life than warehouse.
9.	In Data Warehouse, Data are contained in detail form.	While in this, data are contained in summarized form.

Enterprise-DW (EDW)



Enterprise data warehouse architecture

Basic Elements of the Data Warehouse

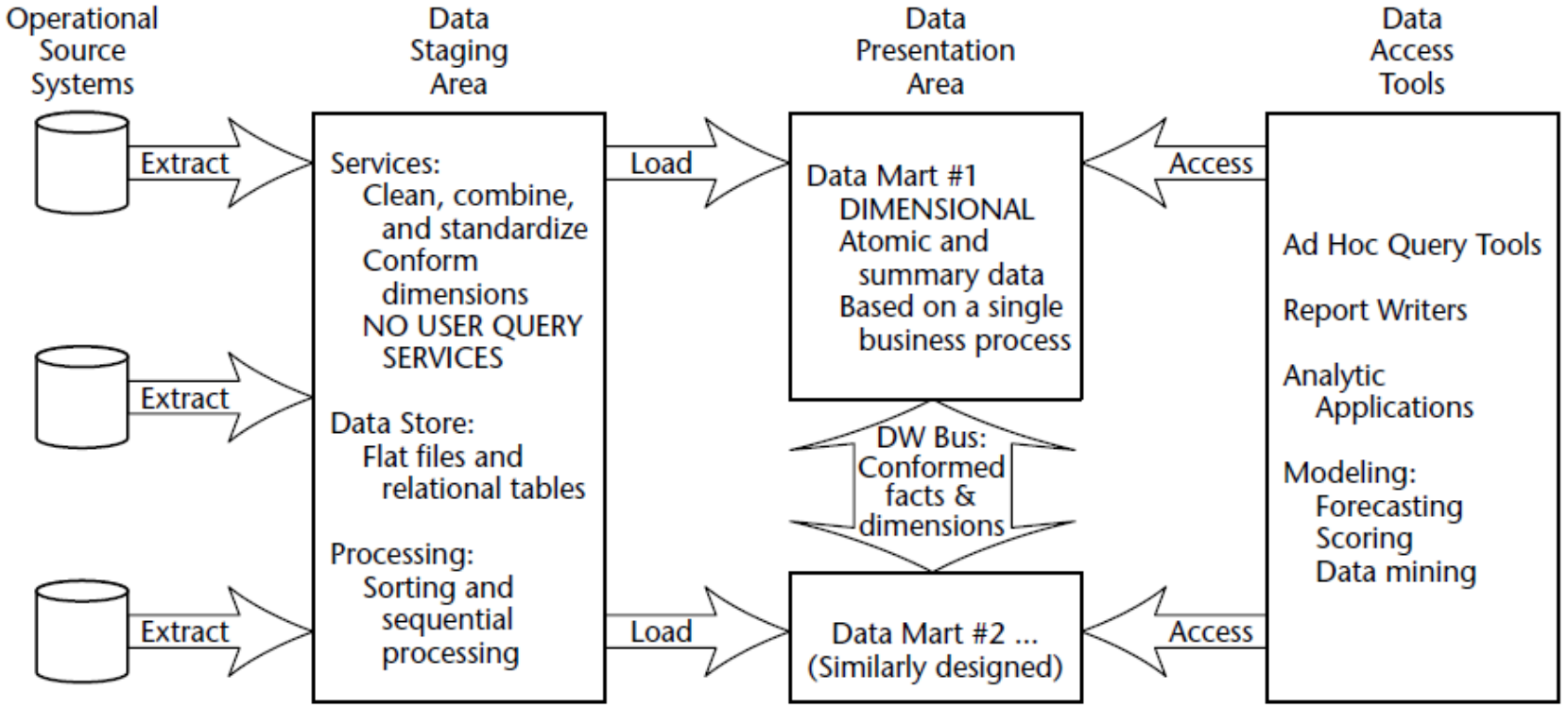


Figure 1.1 Basic elements of the data warehouse.

Ralph Kimball, Margy Ross, **The Data Warehouse Toolkit, 2nd Edition, *2002***

Operational Source Systems

- capture the transactions of the business
- queries against source systems are narrow stovepipe application

Data Staging Area

- a storage area

  AND

- a set of ETL processes

  (*extract-transform-load)*

- It is off-limits to business users and does *not* provide query and presentation services.

Data Staging Area - ETL

- EXTRACTION
  - reading and understanding the source data and copying the data needed for the data warehouse into the staging area for further manipulation.
- TRANSFORMATION
  - cleansing, combining data from multiple sources, deduplicating data, and assigning warehouse keys
- LOADING
  - loading the data into the data warehouse presentation area

Data Presentation Area

- where data is organized, stored and made available for direct querying by users, report writers, and other analytical applications
- it is all the business community sees and touches via data access tools
- dimensional data modeling
  - user understandability
  - query performance
  - resilience to change
- detailed, atomic data

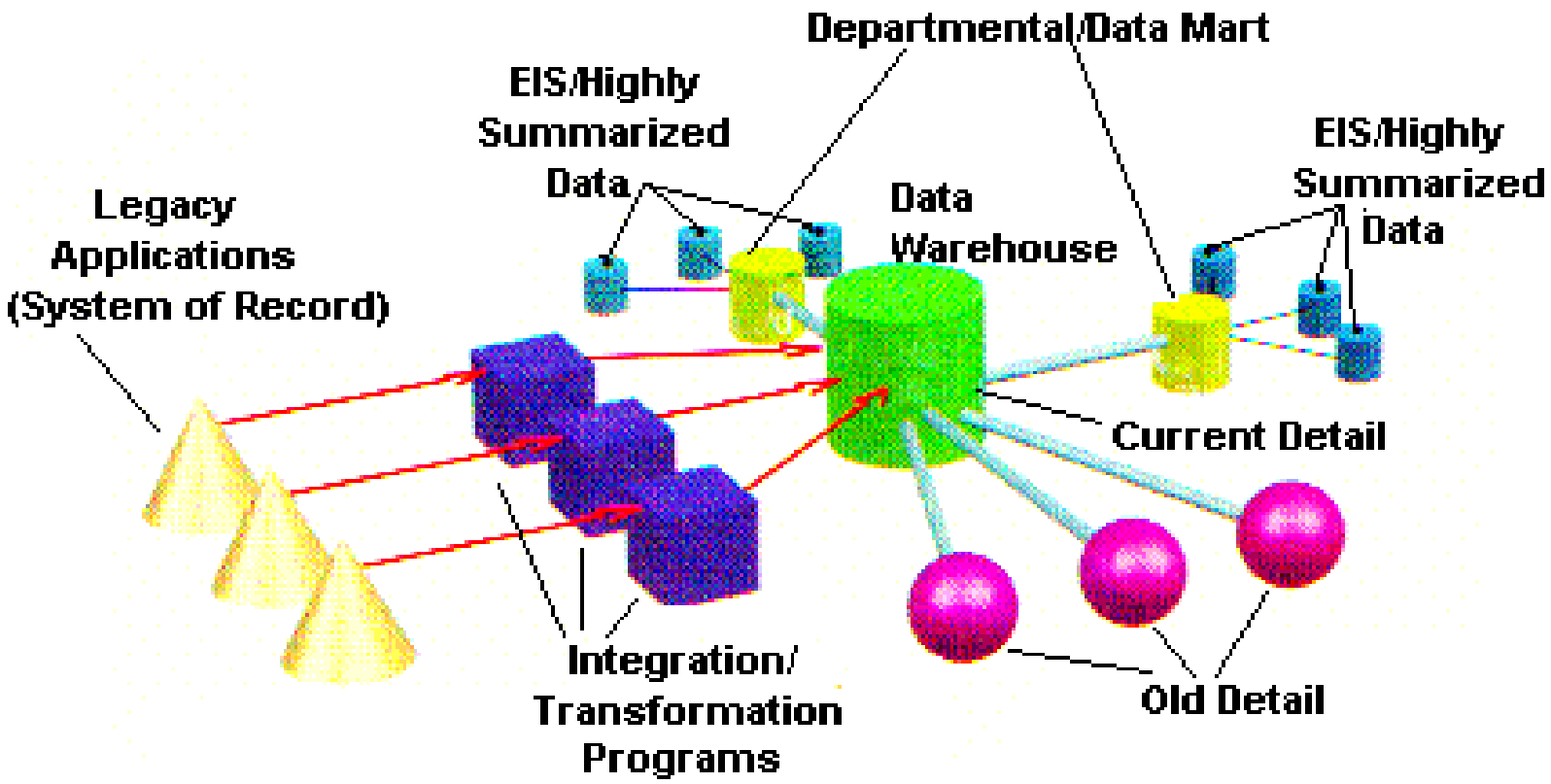
Data Access Tools

- tools that query the data in the data warehouse's presentation area
- the variety of capabilities that can be provided to business users to leverage the presentation area for analytic decision making.
  - prebuilt parameter-driven analytic applications
  - ad hoc query tools
  - data mining, modeling, forecasting

Microsoft SQL Server

- SQL Server Integration Services (SSIS)
  - tool for the ETL process
- SQL Server Analysis Services (SSAS)
  - tool for multidimensional modeling
- SQL Server Reporting Services (SSRS)
  - tool for reporting

Data Warehouse Architecture



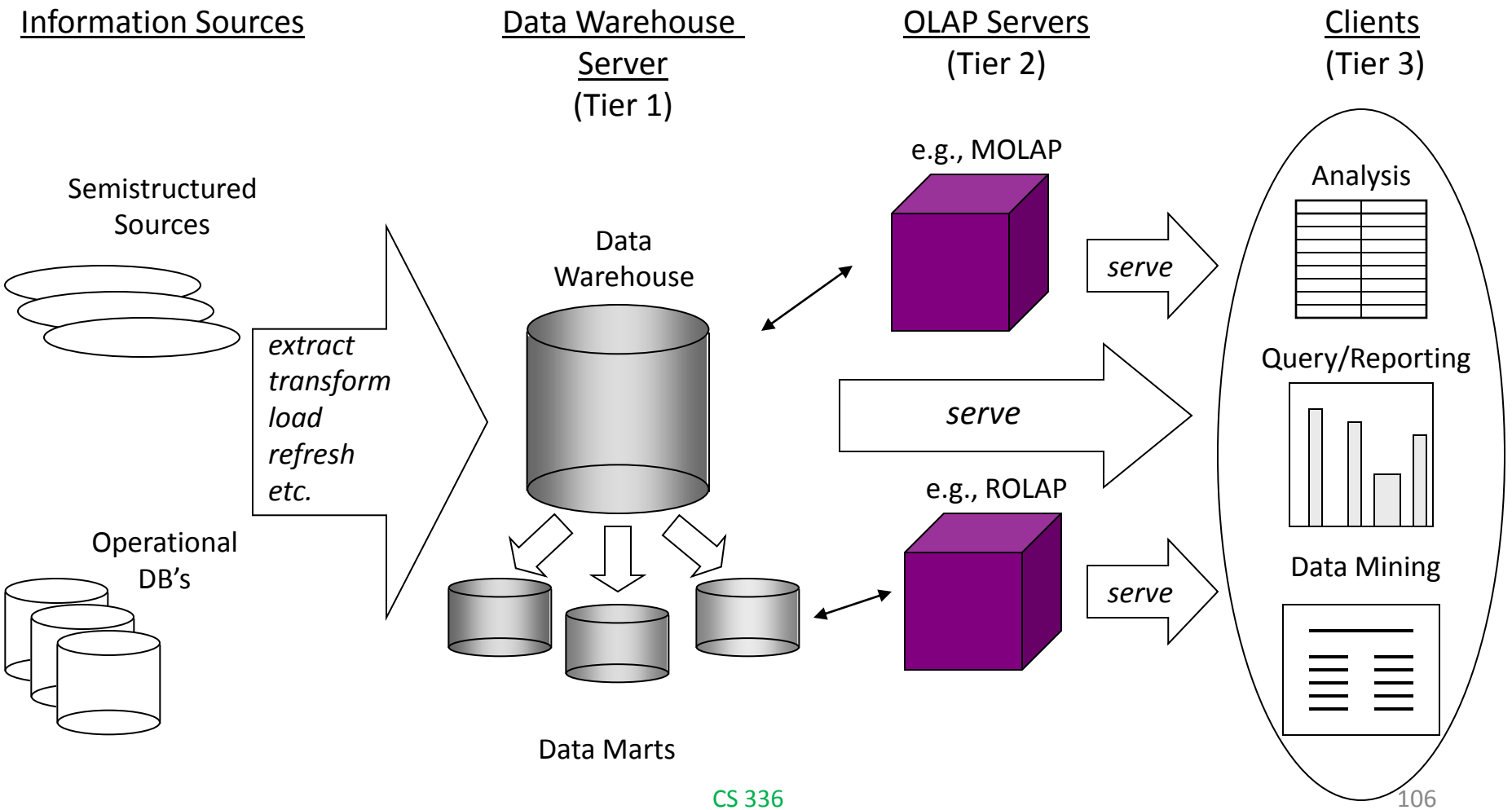
Decision Support

- Information technology to help the knowledge worker (executive, manager, analyst) make faster & better decisions
  - *"What were the sales volumes by region and product category for the last year?"*
  - *"How did the share price of comp. manufacturers correlate with quarterly profits over the past 10 years?"*
  - *"Which orders should we fill to maximize revenues?"*
- On-line analytical processing (OLAP) is an element of decision support systems (DSS)

Three-Tier Decision Support Systems

- Warehouse database server
  - Almost always a relational DBMS, rarely flat files
- OLAP servers
  - Relational OLAP (ROLAP): extended relational DBMS that maps operations on multidimensional data to standard relational operators
  - Multidimensional OLAP (MOLAP): special-purpose server that directly implements multidimensional data and operations
- Clients
  - Query and reporting tools
  - Analysis tools
  - Data mining tools

The Complete Decision Support System



OLAP

- OLAP databases contain two basic types of data:
  - **measures,** which are numeric data, the quantities and averages that you use to make informed business decisions, and
  - **dimensions,** which are the categories that you use to organize these measures.
- OLAP databases help organize data by many levels of detail, using the same categories that you are familiar with to analyze the data.

Advantages and Challenges of an OLTP System

Advantages of an OLTP System

- Simplicity – It is designed typically for use by clerks, cashiers, clients, etc.
- Efficiency – It allows its users to read, write and delete data quickly.
- Fast query processing – It responds to user actions immediately and also supports transaction processing on demand.

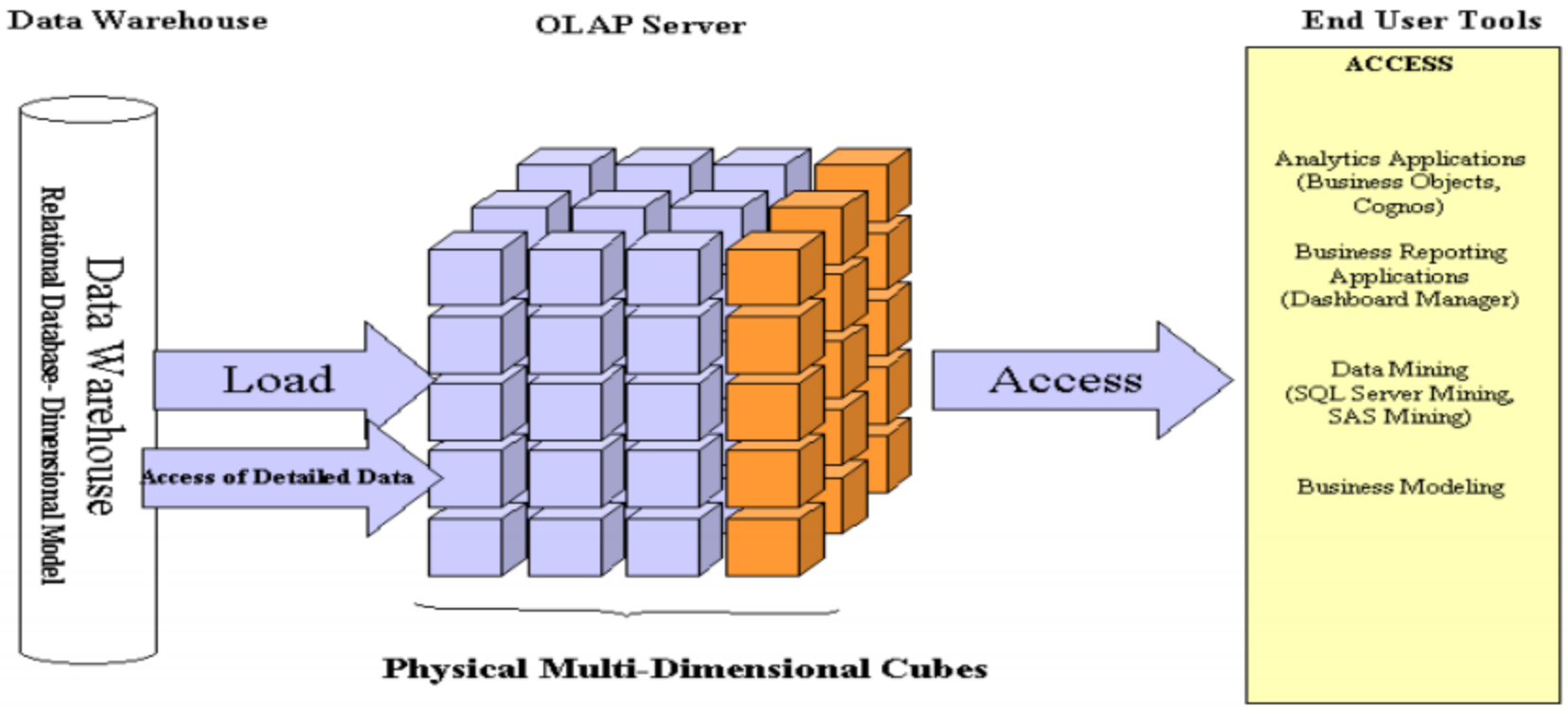
Challenges of an OLTP System

- Security – An OLTP system requires concurrency control (locking) and recovery mechanisms (logging).
- OLTP system data content not suitable for decision making – A typical OLTP system manages the current data within an enterprise/organization. This current data is far too detailed to be easily used for decision making.

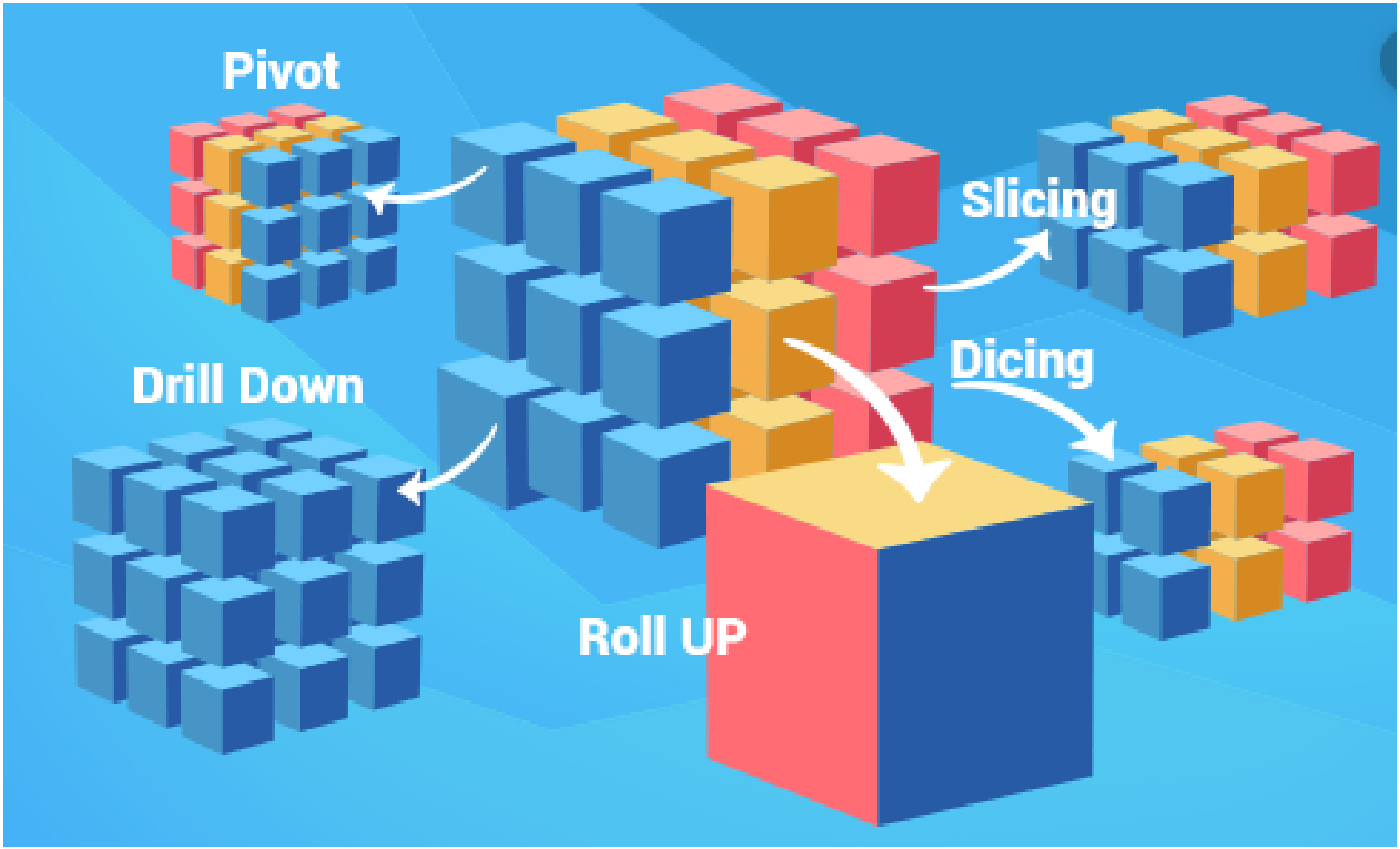
OLAP Cube

- An OLAP Cube is a data structure that allows fast analysis of data.
- The arrangement of data into cubes overcomes a limitation of relational databases.
- The OLAP cube consists of numeric facts called measures which are categorized by dimensions.

OLAP Cube


Data Warehouse
OLAP Server
End User Tools
Data Warehouse
Relational Database- Dimensional Model
Load
Access of Detailed Data
Access
ACCESS
Analytics Applications (Business Objects, Cognos)
Business Reporting Applications (Dashboard Manager)
Data Mining (SQL Server Mining, SAS Mining)
Business Modeling
Physical Multi-Dimensional Cubes

OLAP Operations



OLAP Operations

- There are different kind of operations which we can perform in OLAP
  - Roll up
  - Drill Down
  - Slice
  - Dice
  - Pivot
  - Drill-across
  - Drill-through

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)*

Roll-up

- Takes the current aggregation level of fact values and does a further aggregation on one or more of the dimensions.
- Equivalent to doing GROUP BY to this dimension by using attribute hierarchy.
- SELECT [attribute list], SUM [attribute names] FROM [table list] WHERE [condition list] GROUP BY [grouping list]

Example

- Roll upon Location from cities to countries.

Location	Medal
Delhi	5
New York	2
Patiala	3
Los Angeles	5

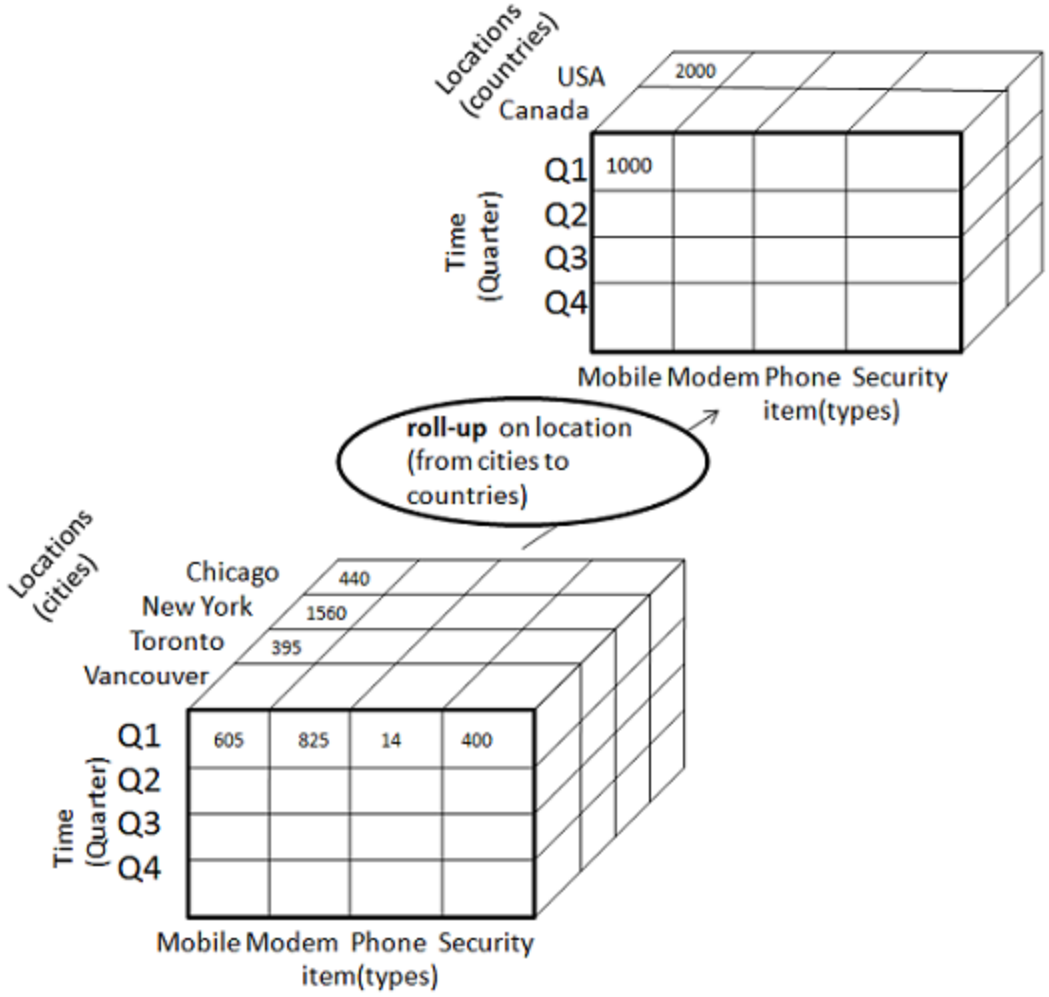
Before Roll up

- More detailed data to less detailed data.

Location	Medal
India	8
America	7

After Roll up

Example of Roll up



Drill- down

- Drill-down is the reverse of roll-up.
- That means lower level summary to higher level summary.
- Increases a number of dimensions - adds new headers
- **Drill-down can be performed either by**
  - Stepping down a concept hierarchy for a dimension
  - By introducing a new dimension.

Example

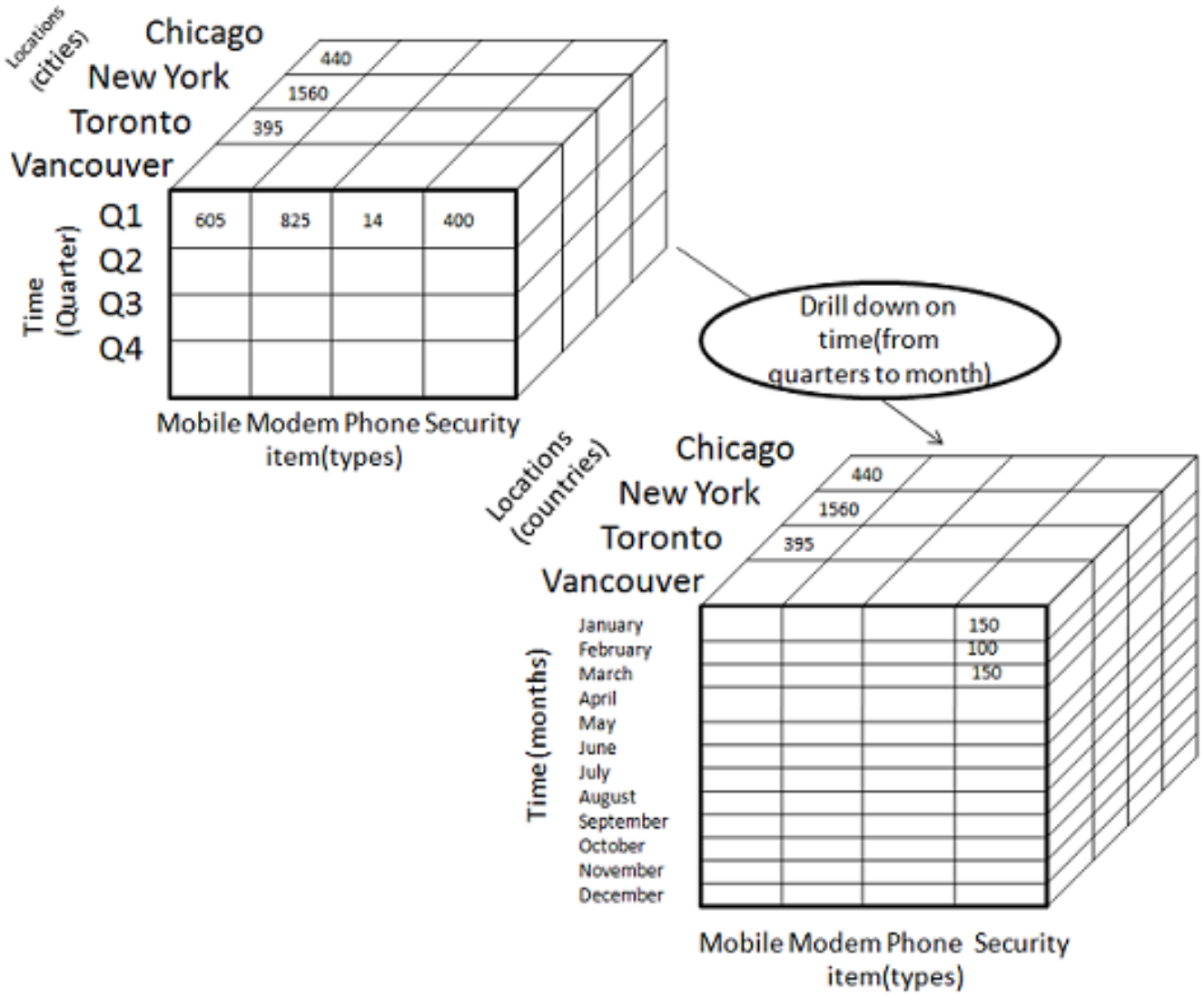
Location	Medal
India	8
America	7

After Drill Down

Location	Medal
Delhi	5
New York	2
Patiala	3
Los Angeles	5

Before Drill down

Example of Drill down

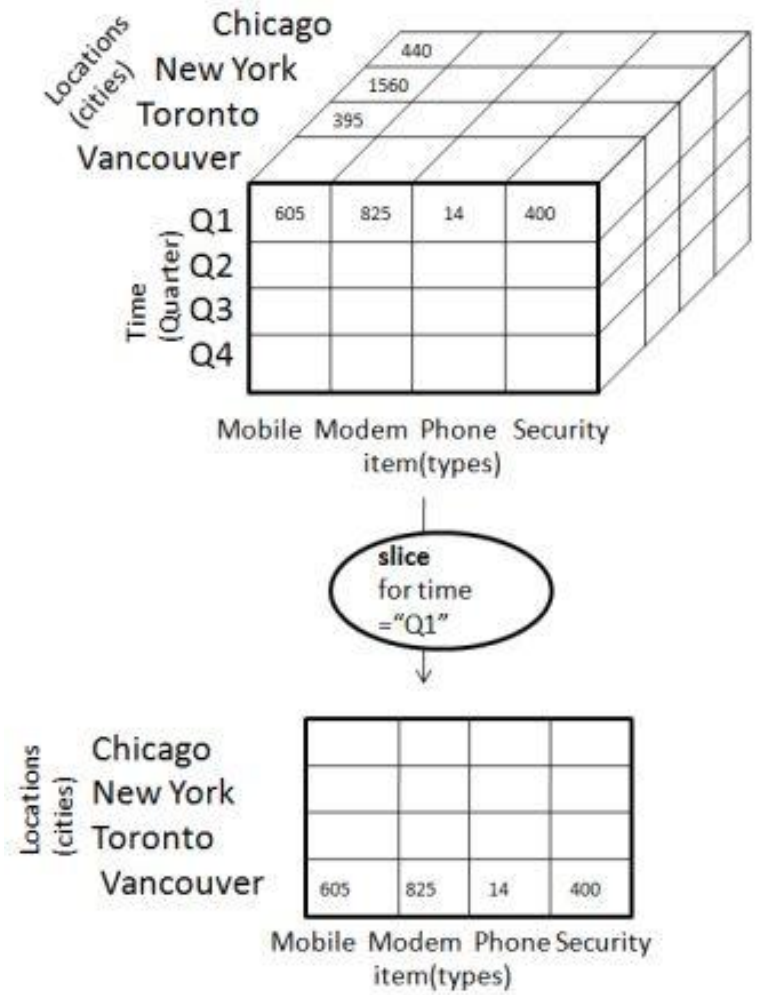


Slice

- Performs a selection on one dimension of the given cube.
- Sets one or more dimensions to specific values and keeps a subset of dimensions for selected values.

Example

- Here Slice is performed for the dimension "time" using the criterion time = "Q1".

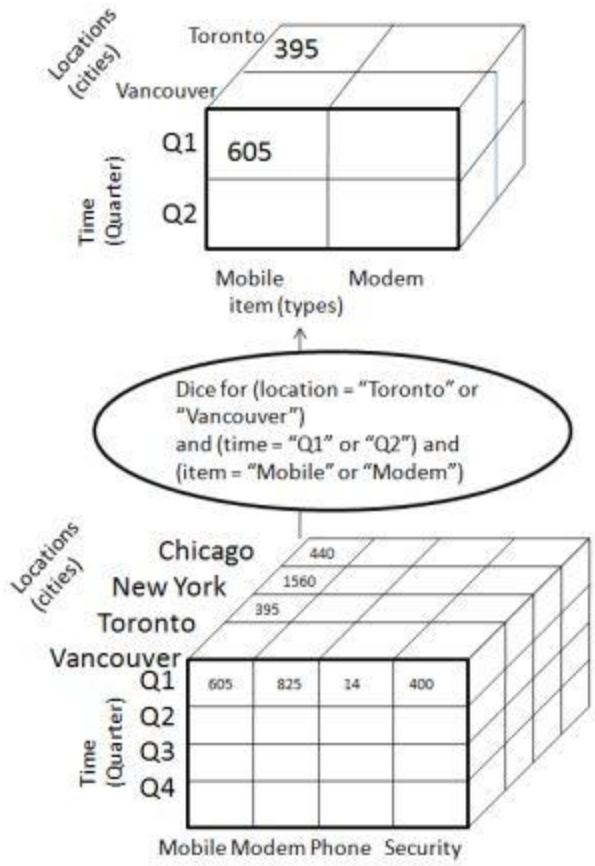


Dice

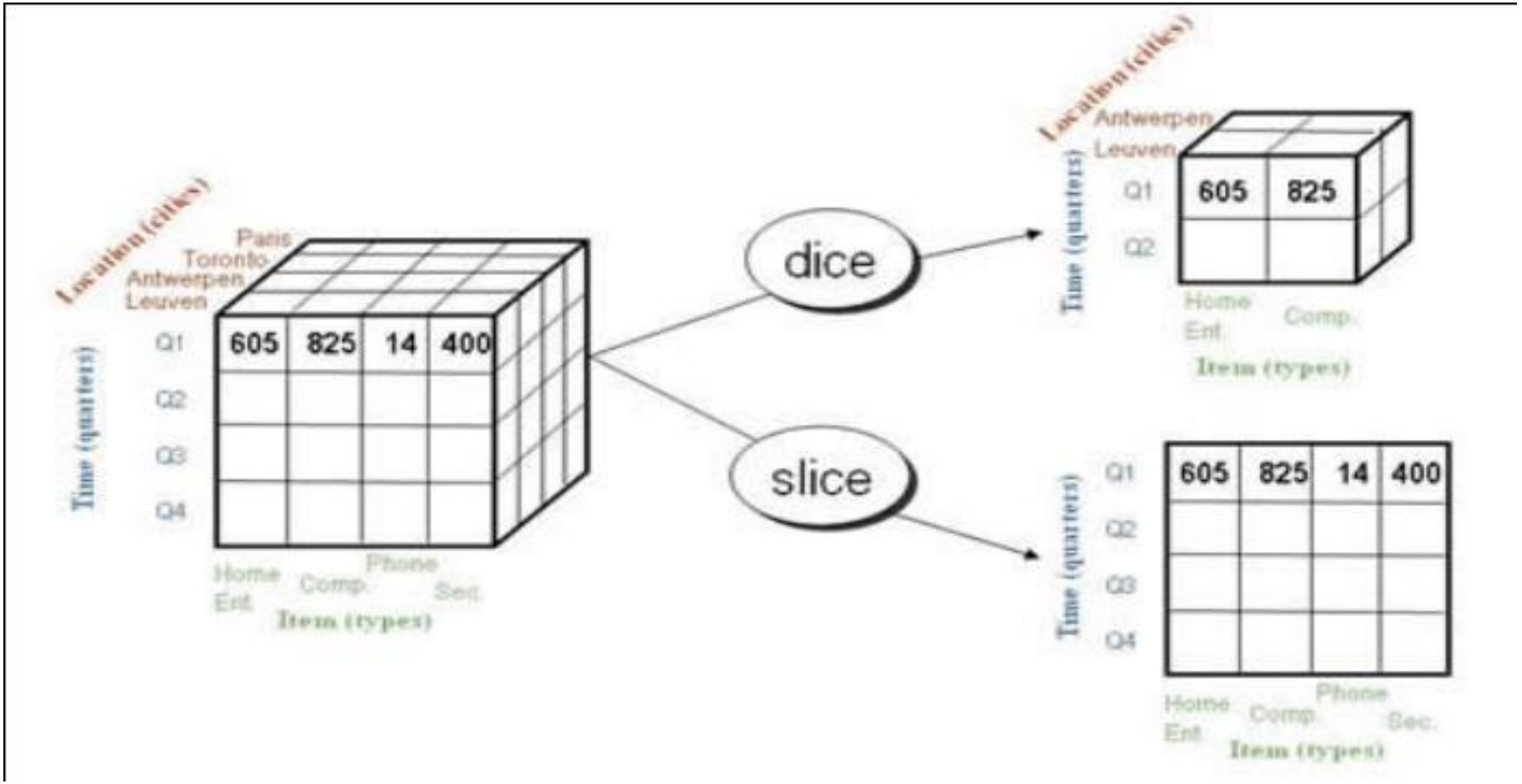
- Define a sub-cube by performing a selection of one or more dimensions.
- Refers to range select condition on one dimension, or to select condition on more than one dimension.
- Reduces the number of member values of one or more dimensions.

Example

- The dice operation on the cube based on the following selection criteria involves three dimensions.
  - (location = "Toronto" or "Vancouver")
  - (time = "Q1" or "Q2")
  - (item =" Mobile" or "Modem")

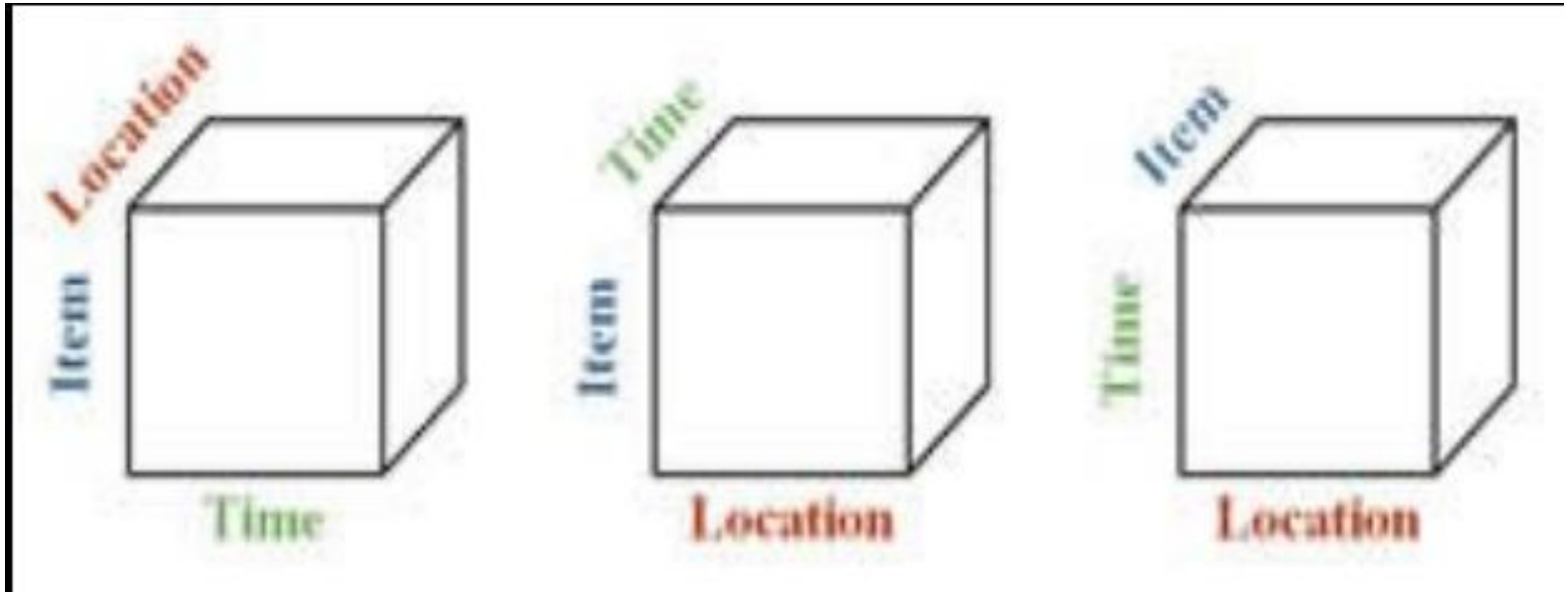


Example of Slice & dice



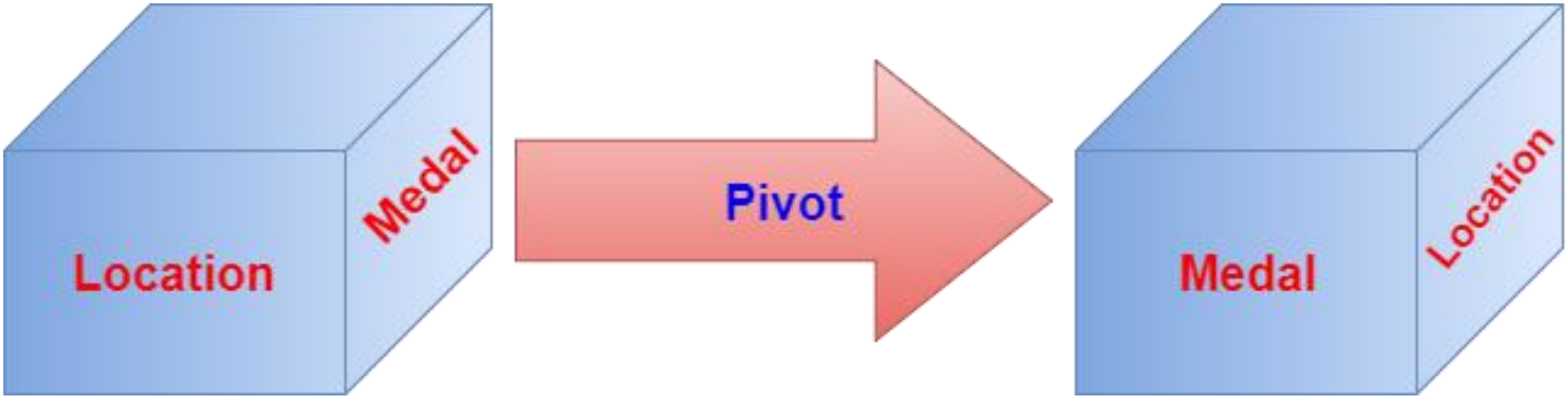
Pivot

- Rotates the data axis to view the data from different perspectives.
- Groups data with different dimensions

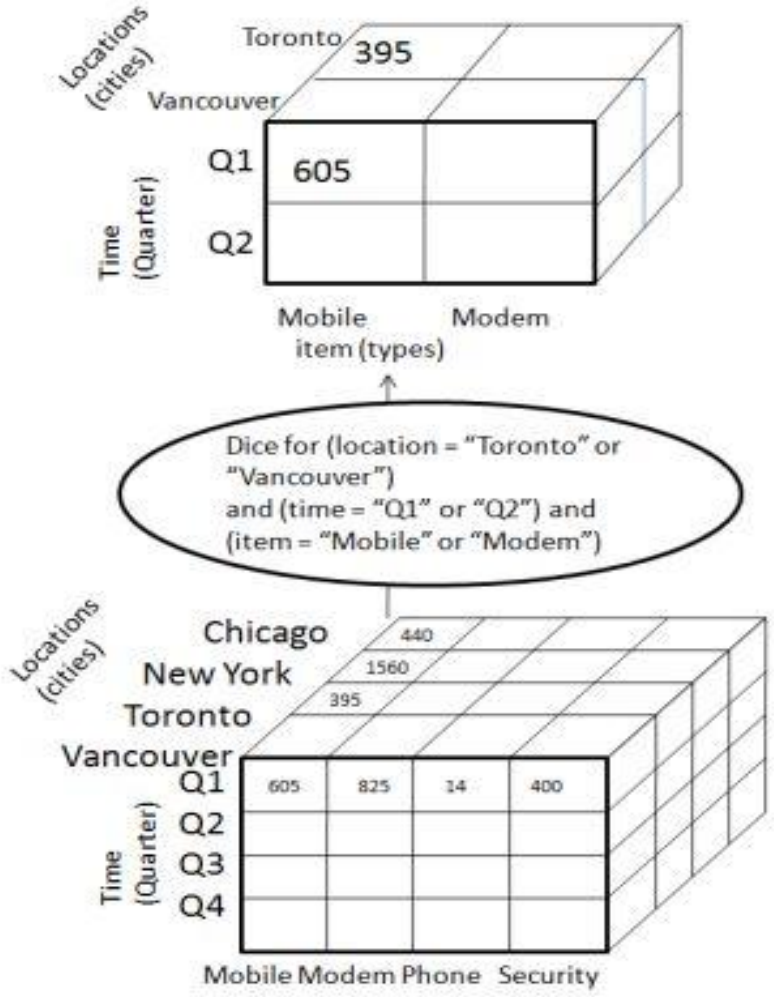


Pivot

- Pivot is also known as rotate.
- It Rotates the data axis to view the data from different perspectives.
-



Example

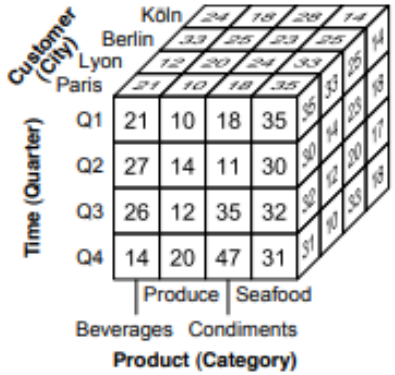

Locations (cities)
Toronto
Vancouver
395
Q1
605
Q2
Time (Quarter)
Mobile
Modem
item (types)
Dice for (location = "Toronto" or "Vancouver")
and (time = "Q1" or "Q2") and
(item = "Mobile" or "Modem")
Locations (cities)
Chicago
New York
Toronto
Vancouver
440
1560
395
Q1
Q2
Q3
Q4
605
825
14
400
Time (Quarter)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

Drill across & Drill through

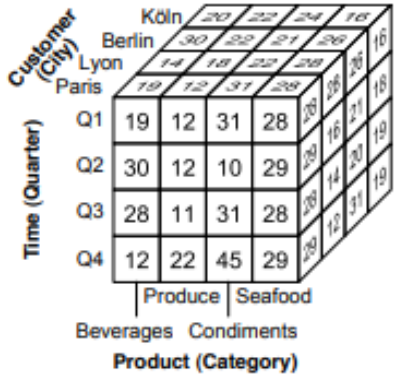
- Drill-across : Accesses more than one fact table that is linked by common dimensions. Combines cubes that share one or more dimensions.
- Drill-through: Drill down to the bottom level of a data cube down to its back-end relational tables.

Drill Across

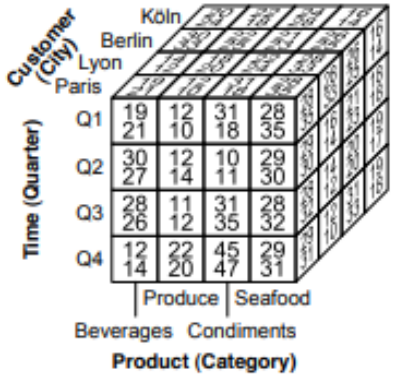
Sales2012



Sales2011



Q: Compare the sales quantities in 2012 with those in 2011



Sales2011-2012 ← DRILLACROSS(Sales2011, Sales2012)

- **Drill-across**: combines cells from two data cubes that have the same schema

Drill Through

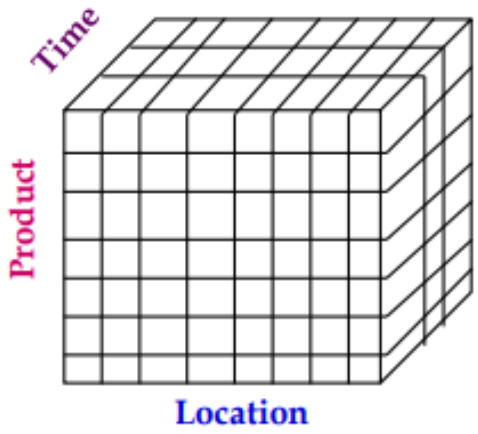
All customers

Time	Thailand		Japan		Total
	Food	NonFood	Food	NonFood	
2006	2400	2200	11000	5000	20600
2007	4500	3200	12000	6000	25700
2008	5600	2900	10000	5500	24000
Total	12500	8300	33000	16500	70300

In order to see the detail at the relational table, we can perform 'drill through'.

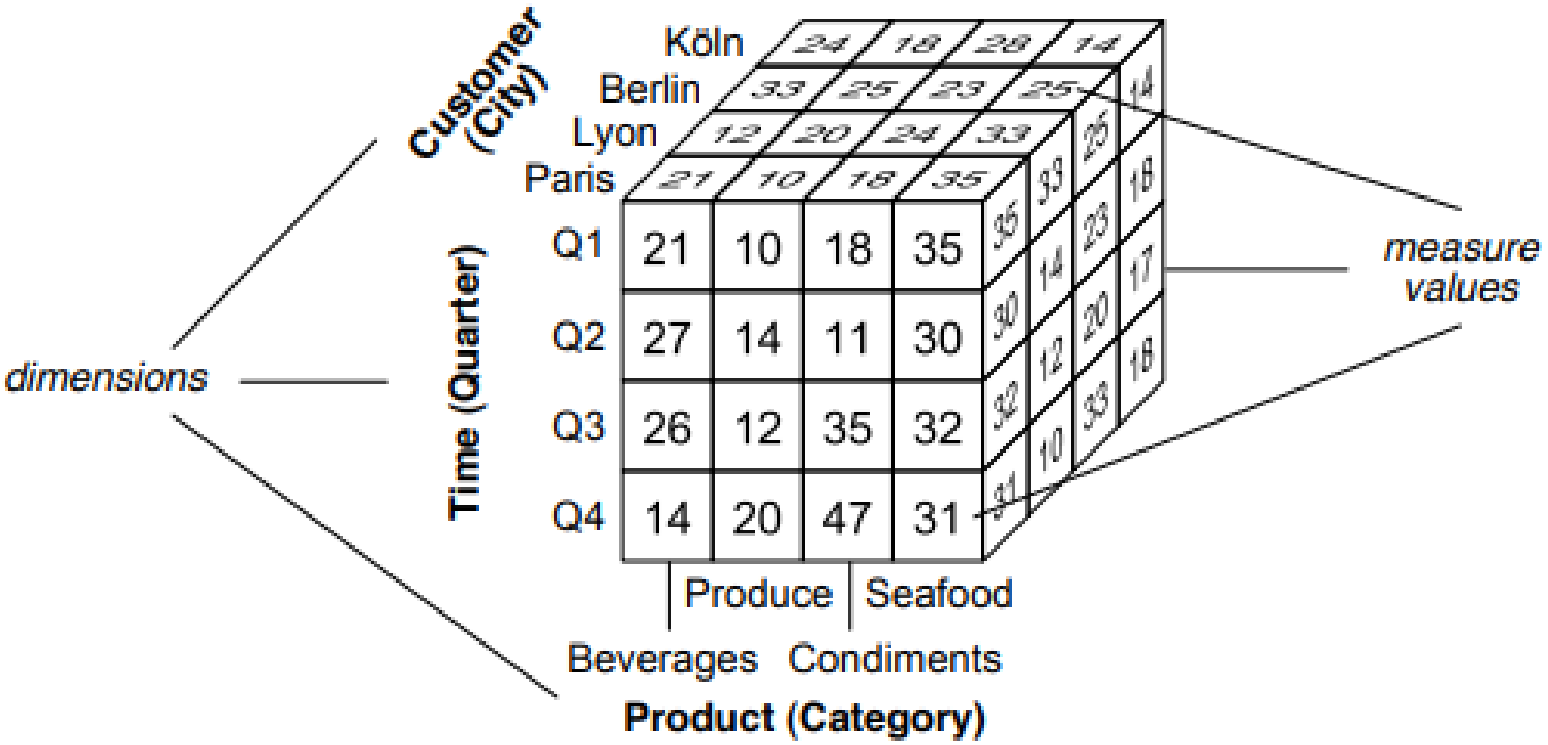
Drill Through

Sales



Date	Branch	Item	Buyer	Units sold	Dollars sold
1/1/2008	Phuket	VCD	First Company	10	250
1/1/2008	Bangkok	TV	First Company	30	900
10/1/2008	Phuket	TV	First Company	10	300
4/2/2008	Phuket	Stereo	First Company	40	200
15/2/2008	Bangkok	VCD	Best Company	30	750
2/5/2008	Bangkok	Computer	Best Company	20	600

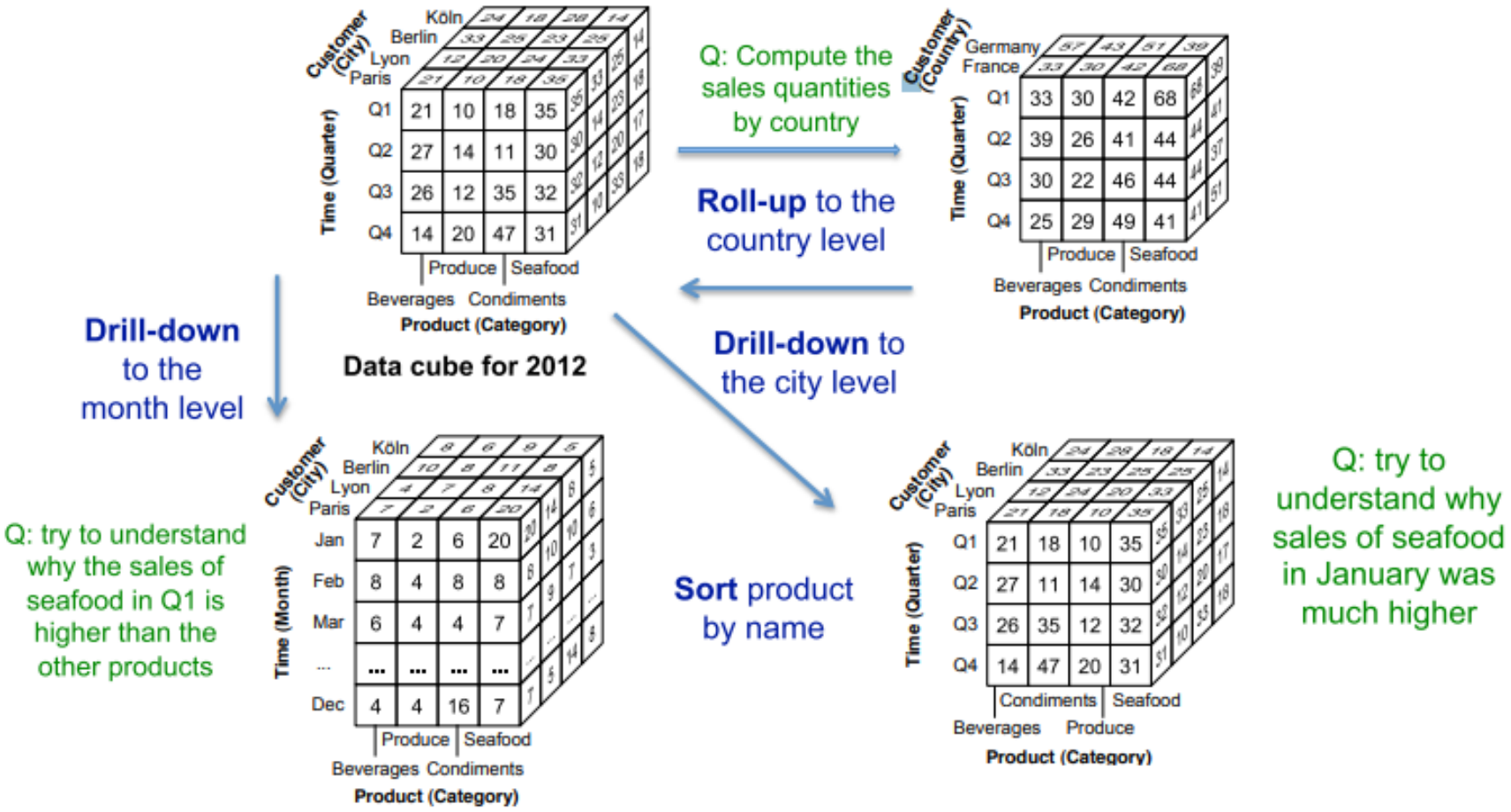
Exercise


Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1
Q2
Q3
Q4
21 10 18 35
27 14 11 30
26 12 35 32
14 20 47 31
Produce
Seafood
Beverages
Condiments
Product (Category)
dimensions
measure values

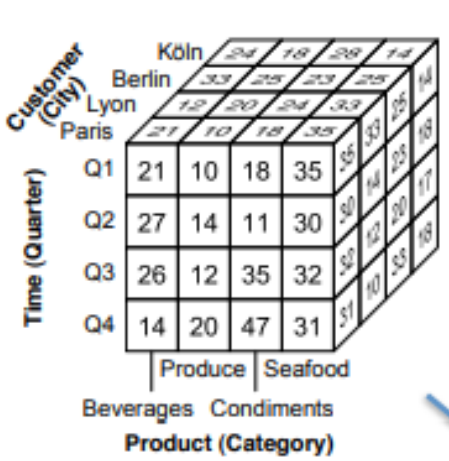
Exercise

1. Compute the sales Quantity by Country
2. Try to understand why sales of seafood in Q1 is higher than the other products.
3. Try to understand why sales of seafood in January was higher much
4. Visualize the cube with time dimension in X axis
5. Visualize data only for paris
6. Visualize data only for paris or Lyon and Quarters Q1 or Q2

Solution


Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1 21 10 18 35
Q2 27 14 11 30
Q3 26 12 35 32
Q4 14 20 47 31
Produce
Seafood
Beverages
Condiments
Product (Category)
Data cube for 2012
Q: Compute the sales quantities by country
Roll-up to the country level
Customer (Country)
Germany
France
Q1 33 30 42 68
Q2 39 26 41 44
Q3 30 22 46 44
Q4 25 29 49 41
Drill-down to the month level
Drill-down to the city level
Time (Month)
Jan 7 2 6 20
Feb 8 4 8 8
Mar 6 4 4 7
... ...
Dec 4 4 16 7
Q: try to understand why the sales of seafood in Q1 is higher than the other products
Sort product by name
Q1 21 18 10 35
Q2 27 11 14 30
Q3 26 35 12 32
Q4 14 47 20 31
Q: try to understand why sales of seafood in January was much higher

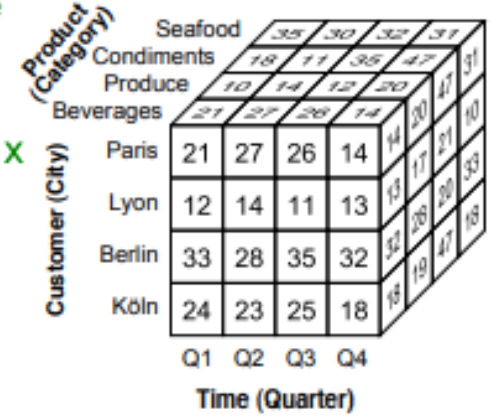
Solution



Data cube for 2012

Q: Visualize the cube with the Time dimension in the x axis

Pivot

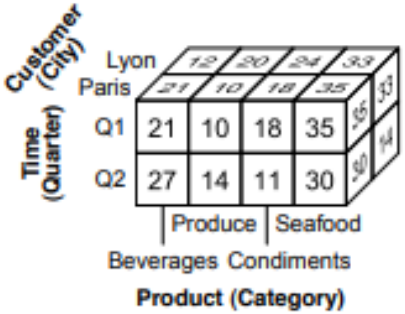


Q: Visualize data only for Paris

Slice on city = 'Paris'

Dice on city = 'Paris' or 'Lyon' and Quarter = 'Q1' or 'Q2'

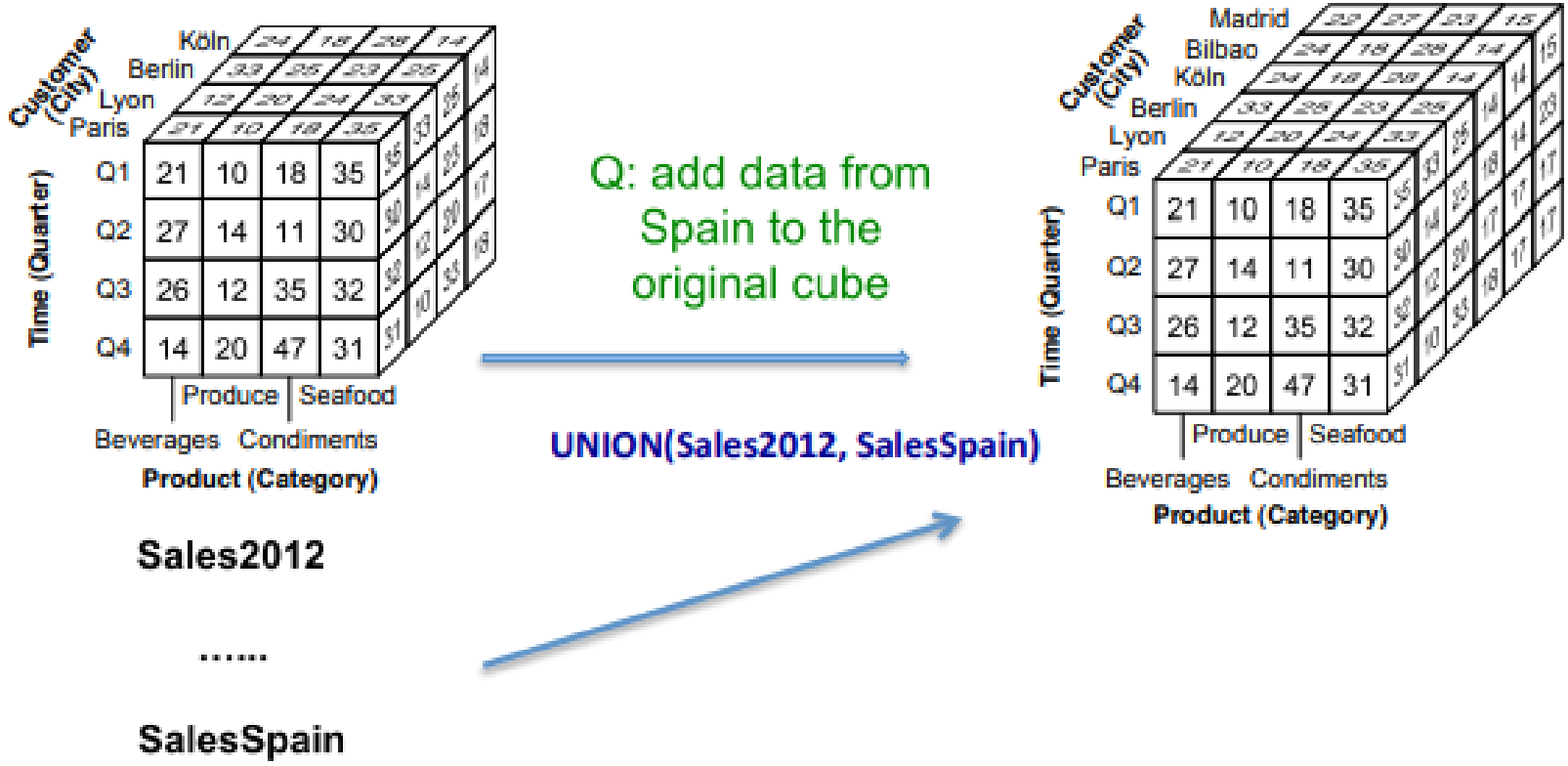
Q: Visualize data only for Paris or Lyon and quarters Q1or Q2



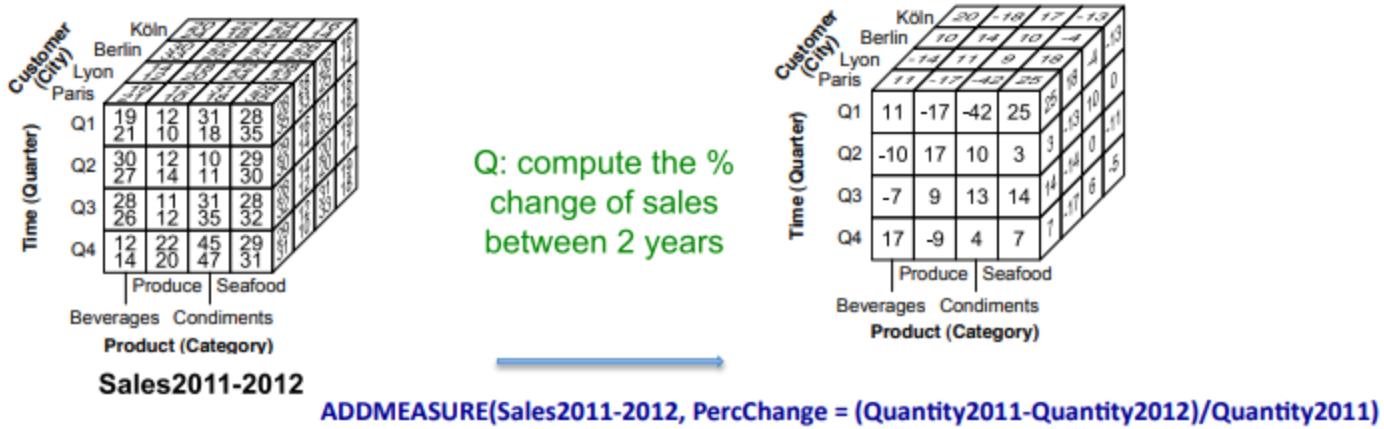
Other OLAP Operations

- Sort
  - Sort brings the cube back where the members of a dimension were sorted.
- Add Measure
  - This OLAP operation one is able to add new measures to a cube.
- Drop Measure
  - In contrast to Add Measure, it’s also possible to get rid of a measure from a data cube if it's not necessary.
- Union
  - Due to an opportunity of Union, you can unite a number of cubes which have the same scheme but separate instances.
- Difference
  - Difference eliminates the cells in a cube which are owned by another one. These two cubes must possess the same scheme.

Union

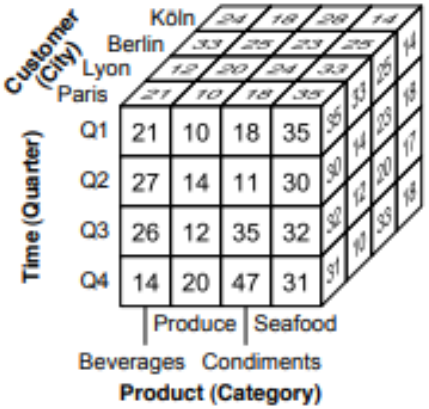

Customer (City)
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1 21 10 18 35
Q2 27 14 11 30
Q3 26 12 35 32
Q4 14 20 47 31
Produce
Seafood
Beverages
Condiments
Product (Category)
Sales2012
.....
SalesSpain
Q: add data from Spain to the original cube
UNION(Sales2012, SalesSpain)
Customer (City)
Madrid
Bilbao
Köln
Berlin
Lyon
Paris
Time (Quarter)
Q1 21 10 18 35
Q2 27 14 11 30
Q3 26 12 35 32
Q4 14 20 47 31
Produce
Seafood
Beverages
Condiments
Product (Category)

Add and Drop Measure



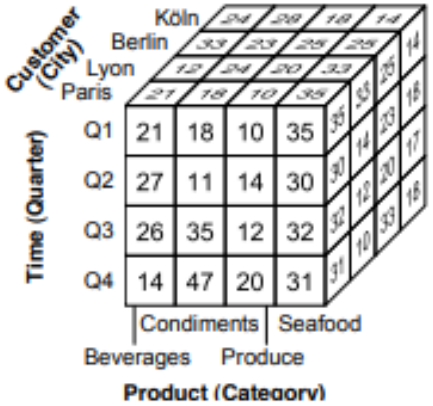
- **Add Measure**: adds new measures to a cube
  - **ADDMEASURE(CubeName, (NewMeasure = Expression)*)**
- **Drop measure**: Deletes a measure from a cube schema
  - **DROPMEASURE(CubeName, Measure*)**

Sort



Sales2012

Q: try to understand why sales of seafood in January was much higher



SORT(Sales2012, Product, Category)

- **Sort** returns a cube where the members of a dimension have been sorted

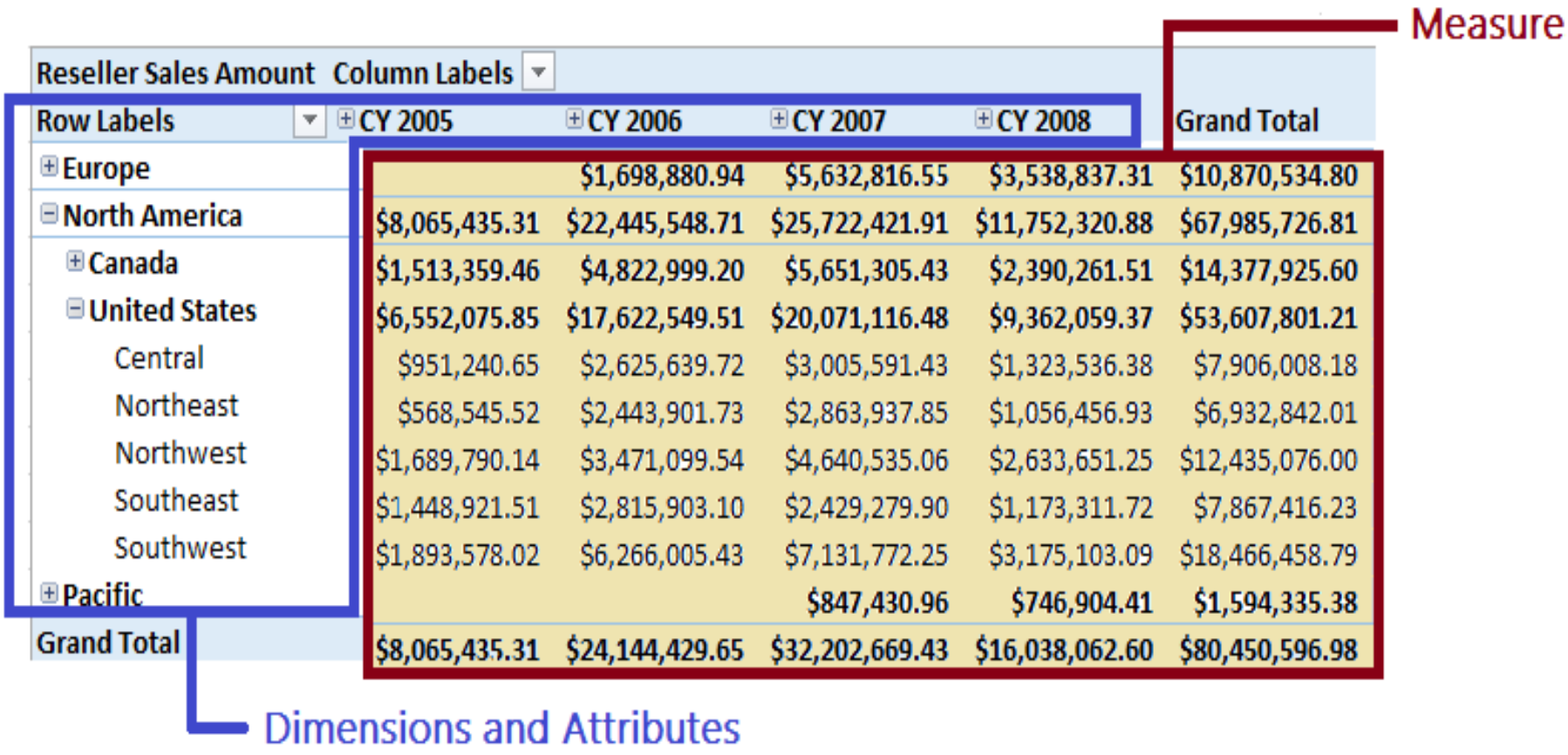
Multi-Dimensional Data

- Dimensions - business parameters that define a transaction
- Example: Analyst may want to view ***sales*** data (measure) by *geography*, by *time*, and by *product* (dimensions)
- Dimensional modeling is a technique for structuring data around the business concepts
- ER models describe “entities” and “relationships”
- Dimensional models describe “measures” and “dimensions”

Measure

- A set of values in a cube that are based on a column in the cube's fact table and that are usually numeric values.
- Measures are the central values in the cube that are preprocessed, aggregated, and analyzed. Common examples include sales, profits, revenues, and costs.

Example


Measure
Reseller Sales Amount
Column Labels
Row Labels
CY 2005
CY 2006
CY 2007
CY 2008
Grand Total
Europe
North America
Canada
United States
Central
Northeast
Northwest
Southeast
Southwest
Pacific
Grand Total
$1,698,880.94
$5,632,816.55
$3,538,837.31
$10,870,534.80
$8,065,435.31
$22,445,548.71
$25,722,421.91
$11,752,320.88
$67,985,726.81
$1,513,359.46
$4,822,999.20
$5,651,305.43
$2,390,261.51
$14,377,925.60
$6,552,075.85
$17,622,549.51
$20,071,116.48
$9,362,059.37
$53,607,801.21
$951,240.65
$2,625,639.72
$3,005,591.43
$1,323,536.38
$7,906,008.18
$568,545.52
$2,443,901.73
$2,863,937.85
$1,056,456.93
$6,932,842.01
$1,689,790.14
$3,471,099.54
$4,640,535.06
$2,633,651.25
$12,435,076.00
$1,448,921.51
$2,815,903.10
$2,429,279.90
$1,173,311.72
$7,867,416.23
$1,893,578.02
$6,266,005.43
$7,131,772.25
$3,175,103.09
$18,466,458.79
$847,430.96
$746,904.41
$1,594,335.38
$8,065,435.31
$24,144,429.65
$32,202,669.43
$16,038,062.60
$80,450,596.98
Dimensions and Attributes

Member

- An item in a hierarchy representing one or more occurrences of data.
- A member can be either unique or nonunique.
- For example, 2007 and 2008 represent unique members in the year level of a time dimension,
- whereas January represents nonunique members in the month level because there can be more than one January in the time dimension if it contains data for more than one year.

KPIs

- A KPI is a special calculated measure that is defined on the server that allows you
  - to track "key performance indicators" including status (Does the current value meet a specific number?) and
  - trend (what is the value over time?).
- These calculations are a combination of Multidimensional Expressions (MDX) expressions or calculated members.
- KPIs also have additional metadata that provides information about how client applications should display the results of the KPI's calculations.

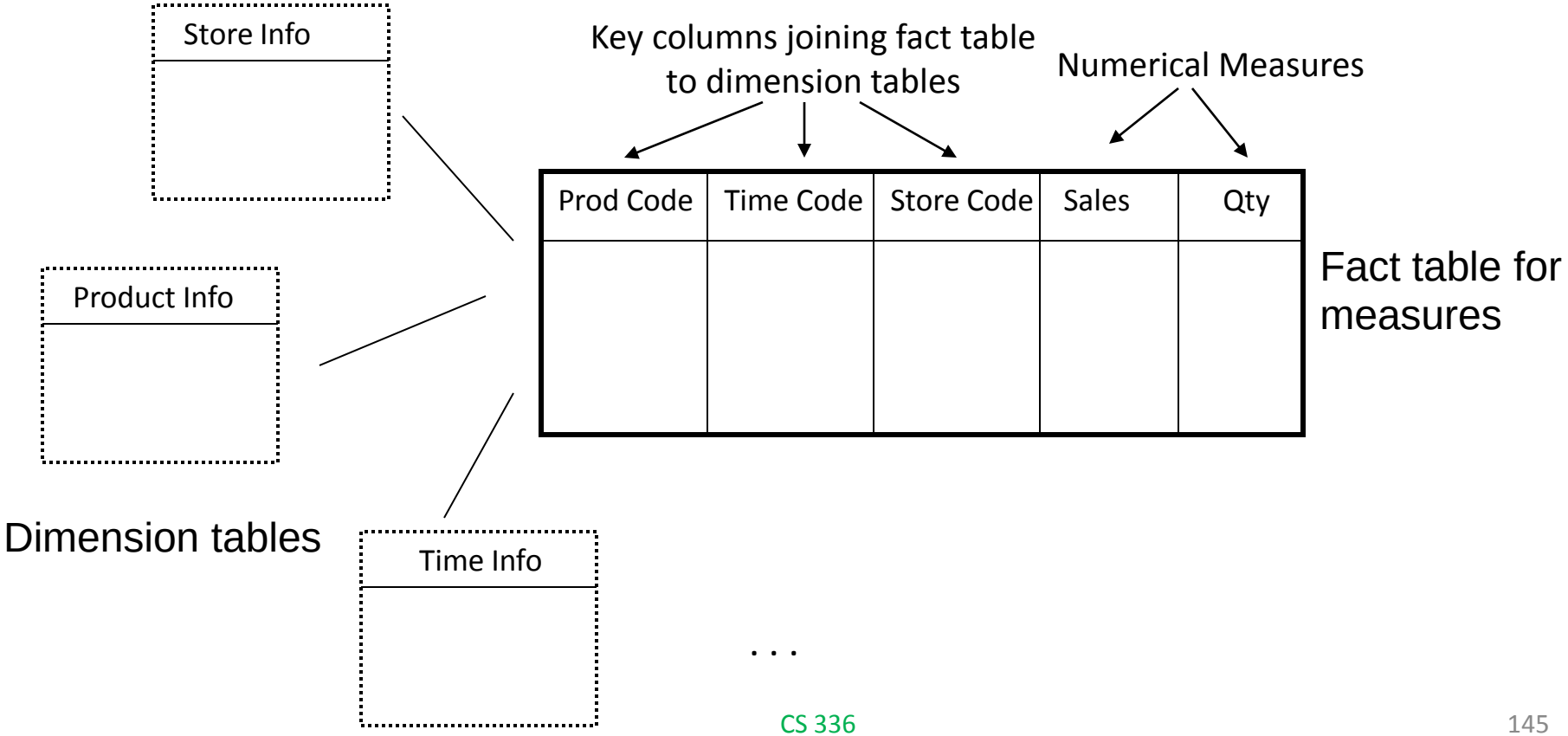
Examples

- The sales department of an organization may use monthly gross profit as a KPI,
- But, the human resources department of the same organization may use quarterly employee turnover.

The Multi-Dimensional Model

"Sales by product line over the past six months"

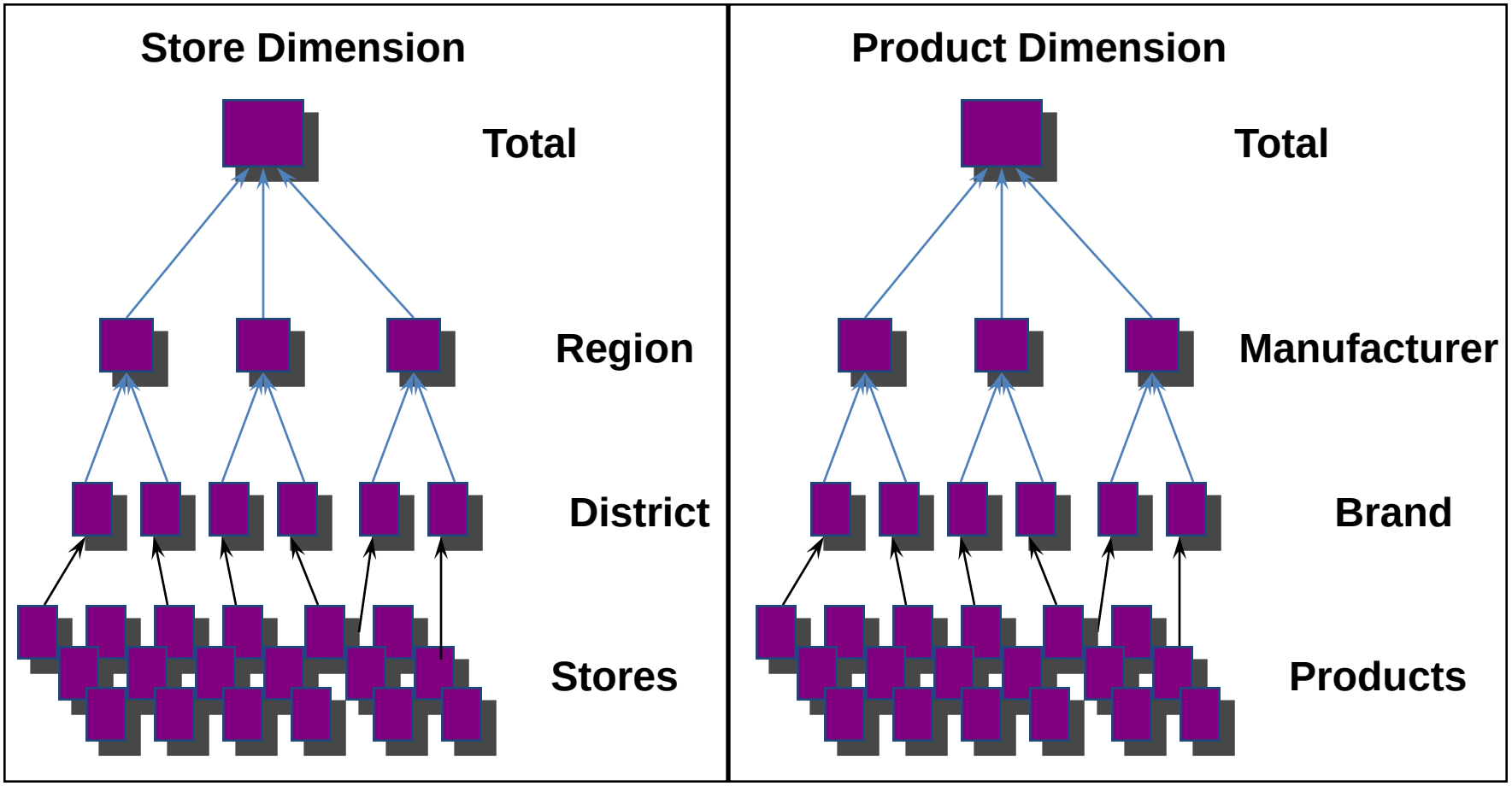
"Sales by store between 1990 and 1995"



Dimensional Modeling

- Dimensions are organized into hierarchies
  - E.g., Time dimension: days → weeks → quarters
  - E.g., Product dimension: product → product line → brand
- Dimensions have attributes

Dimension Hierarchies



Types of OLAP servers

- **Three types of OLAP servers are:-**
  - 1 Relational OLAP (ROLAP)
  - 2 Multidimensional OLAP (MOLAP)
  - 3 Hybrid OLAP (HOLAP)

ROLAP

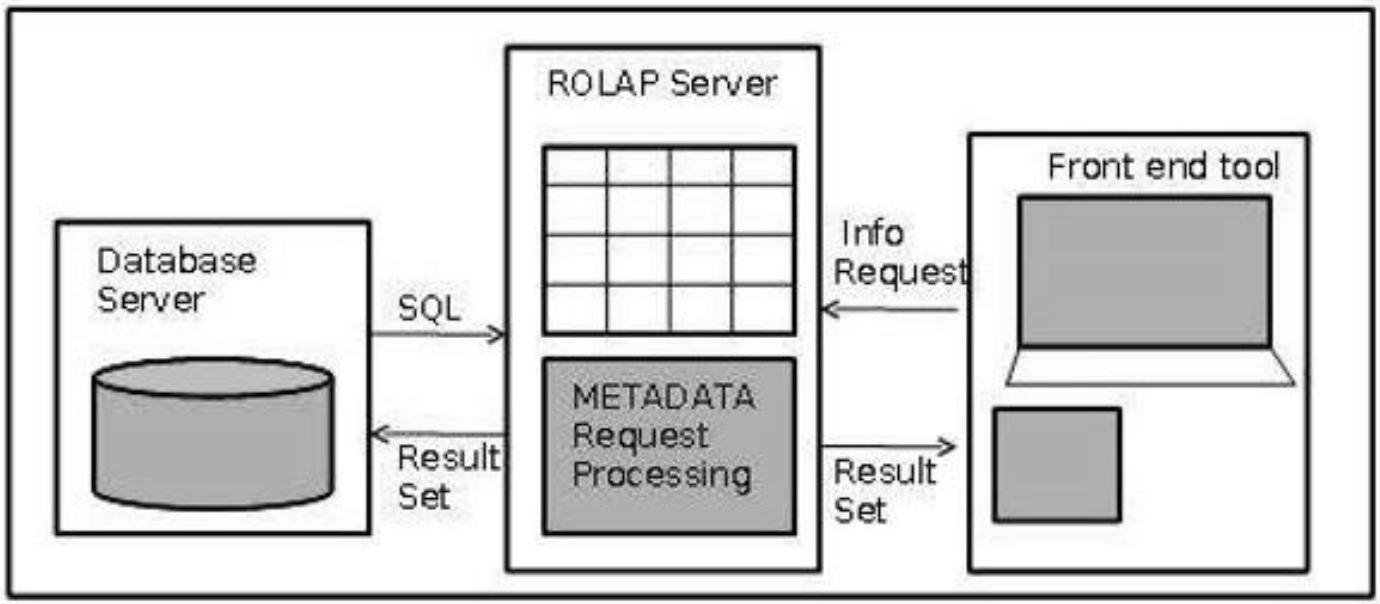
- ROLAP servers are placed between the relational back-end server and client front-end tools.
- ROLAP servers use RDBMS to store and manage warehouse data, and OLAP middleware to support missing pieces.

ROLAP: Dimensional Modeling Using Relational DBMS

- Special schema design: *star, snowflake*
- Special indexes: bitmap, multi-table join
- Special tuning: maximize query throughput
- Proven technology (relational model, DBMS), tend to outperform specialized MDDB especially on large data sets
- Products
  - IBM DB2, Oracle, Sybase IQ, RedBrick, Informix

Relational OLAP Architecture

- ROLAP includes the following components –
  - Database server
  - ROLAP server
  - Front-end tool.



Advantages of ROLAP Dimensional Modeling

- Define complex, multi-dimensional data with simple model
- Reduces the number of joins a query has to process
- ROLAP can handle large amounts of data.
- Can be used with data warehouse and OLTP systems.
- Allows the data warehouse to evolve with rel. low maintenance

Disadvantages of ROLAP

- Limited by SQL functionalities.
- Hard to maintain aggregate tables.

MOLAP: Dimensional Modeling Using the Multi Dimensional Model

- MDDB: a special-purpose data model
- Facts stored in multi-dimensional arrays
- Dimensions used to index array
- Sometimes on top of relational DB
- Products
  - Pilot, Arbor Essbase, Gentia

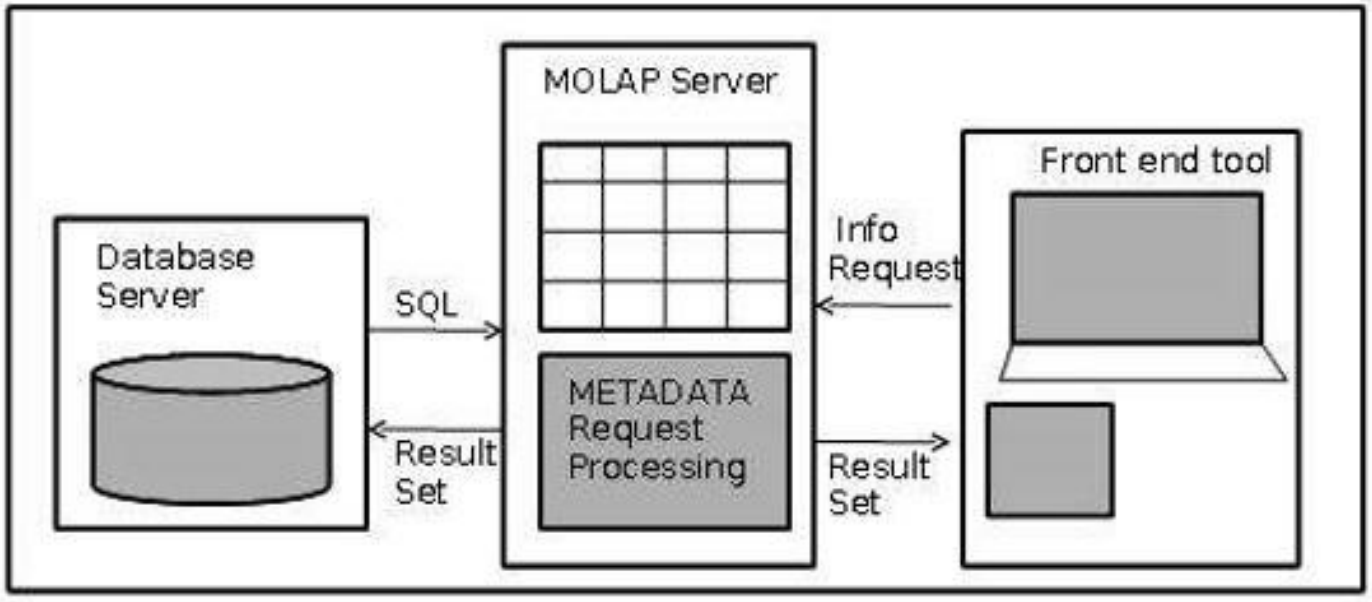
MOLAP

- The first generation of server-based multidimensional OLAP (MOLAP) solutions use multidimensional databases (MDDBs).

- The main advantage of an MDDB over an RDBMS is that an MDDB can provide information quickly since it is calculated and stored at the appropriate hierarchy level in advance.

MOLAP Architecture

- MOLAP includes the following components −
  - Database server.
  - MOLAP server.
  - Front-end tool.



MOLAP

- **Advantages of MOLAP**
  - Optimal for slice and dice operations.
  - Performs better than ROLAP when data is dense.
  - Can perform complex calculations.
- **Disadvantages of MOLAP**
  - Difficult to change dimension without re-aggregation.
  - MOLAP can handle limited amount of data.

HOLAP

- The hybrid OLAP system combines the performance and functionality of the MDDB with the ability to access detail data, which provides greater value to some categories of users.
- Hybrid On-Line Analytical Processing (HOLAP) is a combination of ROLAP and MOLAP.
- HOLAP provide greater scalability of ROLAP and the faster computation of MOLAP.

HOLAP

- **Advantages of HOLAP**
  - HOLAP provide advantages of both MOLAP and ROLAP.
  - Provide fast access at all levels of aggregation.
- **Disadvantages of HOLAP**
  - HOLAP architecture is very complex because it support both MOLAP and ROLAP servers.

ROLAP vs. MOLAP

Sl. No.	MOLAP	ROLAP
1	Information retrieval is fast.	Information retrieval is comparatively slow.
2	Uses sparse array to store data-sets.	Uses relational table.
3	MOLAP is best suited for inexperienced users, since it is very easy to use.	ROLAP is best suited for experienced users.
4	Maintains a separate database for data cubes.	It may not require space other than available in the Data warehouse.
5	DBMS facility is weak.	DBMS facility is strong.

Advantages of an OLAP System

- Multi-dimensional data representation.
- Consistency of information.
- “What if ” analysis.
- Provides a single platform for all information and business needs – planning, budgeting, forecasting, reporting and analysis.
- Fast and interactive ad hoc exploration.

OLAP Products

- MOLAP
  - SAS CFO Vision
  - Comshare Decision
  - Hyperion Essbase
  - PowerPlay Enterprise Server
- ROLAP
  - Cartesis Carat
  - MicroStrategy
- HOLAP
  - Oracle Express
  - Seagate Holos
  - Speedware Media/M
  - Microsoft OLAP Services

OLAP Server Architecture

Relational OLAP (ROLAP)

- Use relational or extended-relational DBMS to store and manage warehouse data and OLAP middle ware to support missing pieces
- Include optimization of DBMS backend, implementation of aggregation navigation logic, and additional tools and services
- Greater scalability

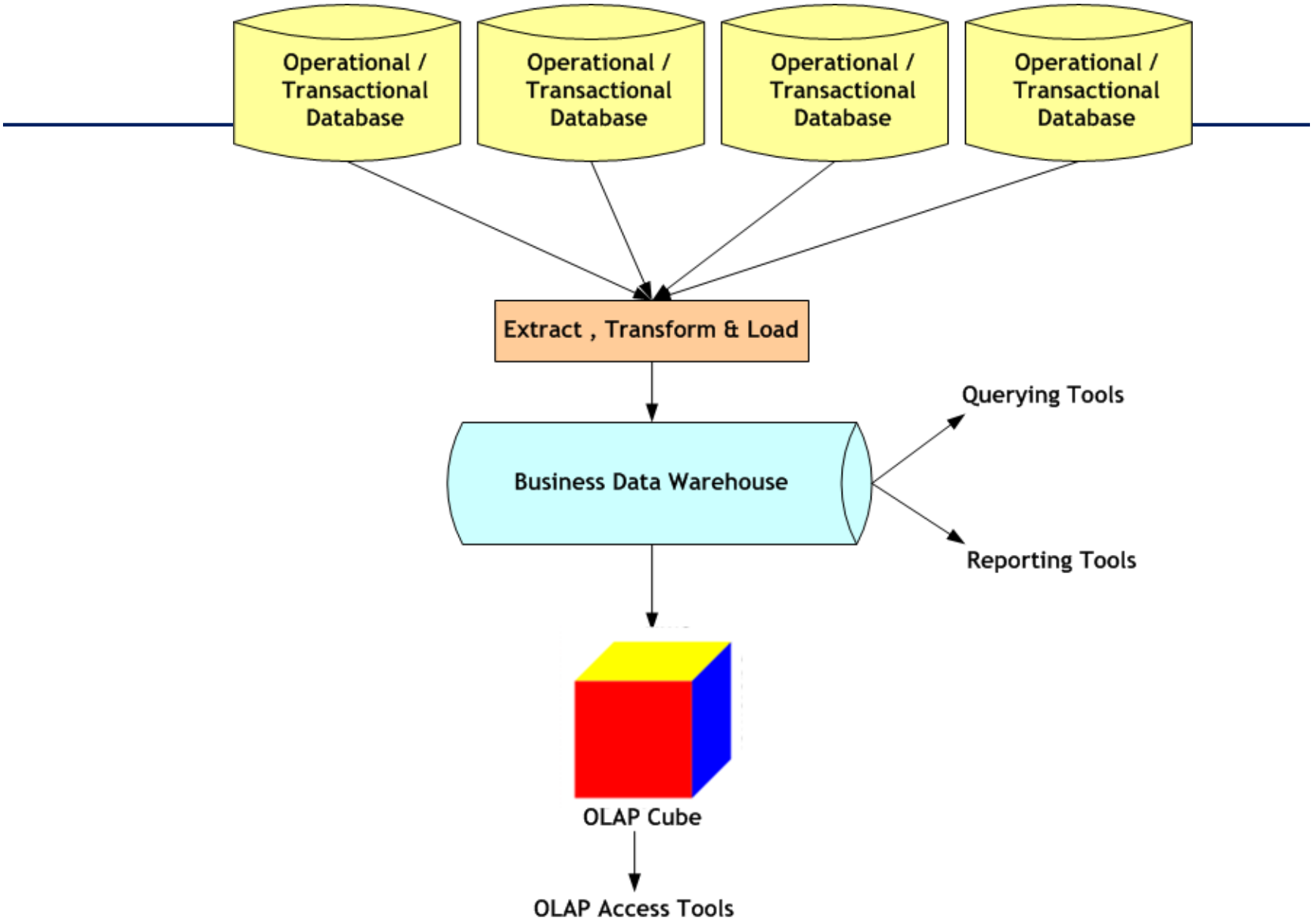
Multidimensional OLAP (MOLAP)

- Array-based multidimensional storage engine (sparse matrix techniques)
- Fast indexing to pre-computed summarized data

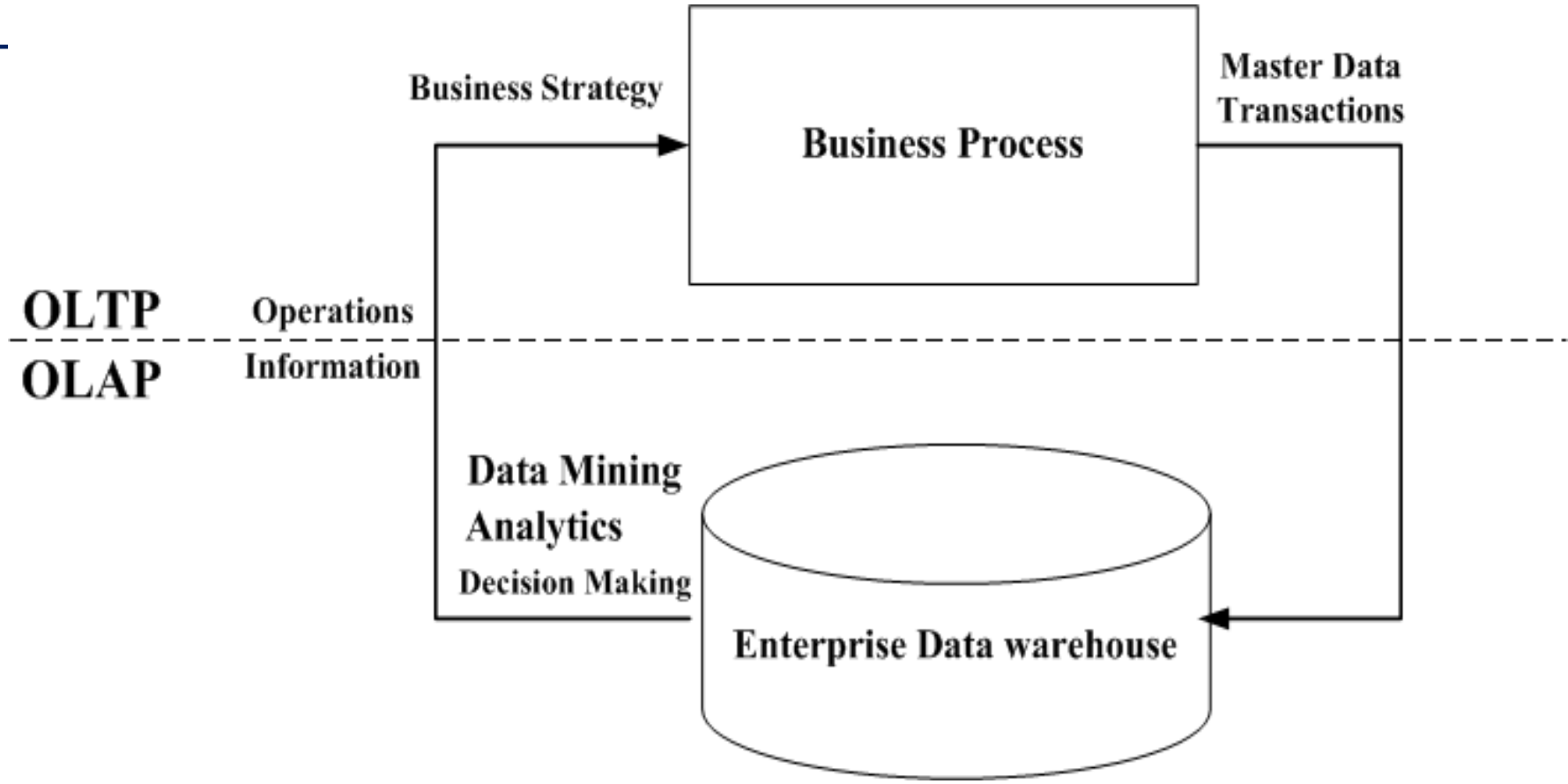
Hybrid OLAP (HOLAP)

- User flexibility, e.g. low level: relational, high-level: array
- Specialized SQL servers
- Specialized support for SQL queries over star/snowflake schemas

OLAP in BI


Operational / Transactional Database
Operational / Transactional Database
Operational / Transactional Database
Operational / Transactional Database
Extract , Transform & Load
Business Data Warehouse
Querying Tools
Reporting Tools
OLAP Cube
OLAP Access Tools

OLTP and OLAP



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	100 MB-GB	100 GB-TB
Metric	Transaction Throughput	Query throughput, response

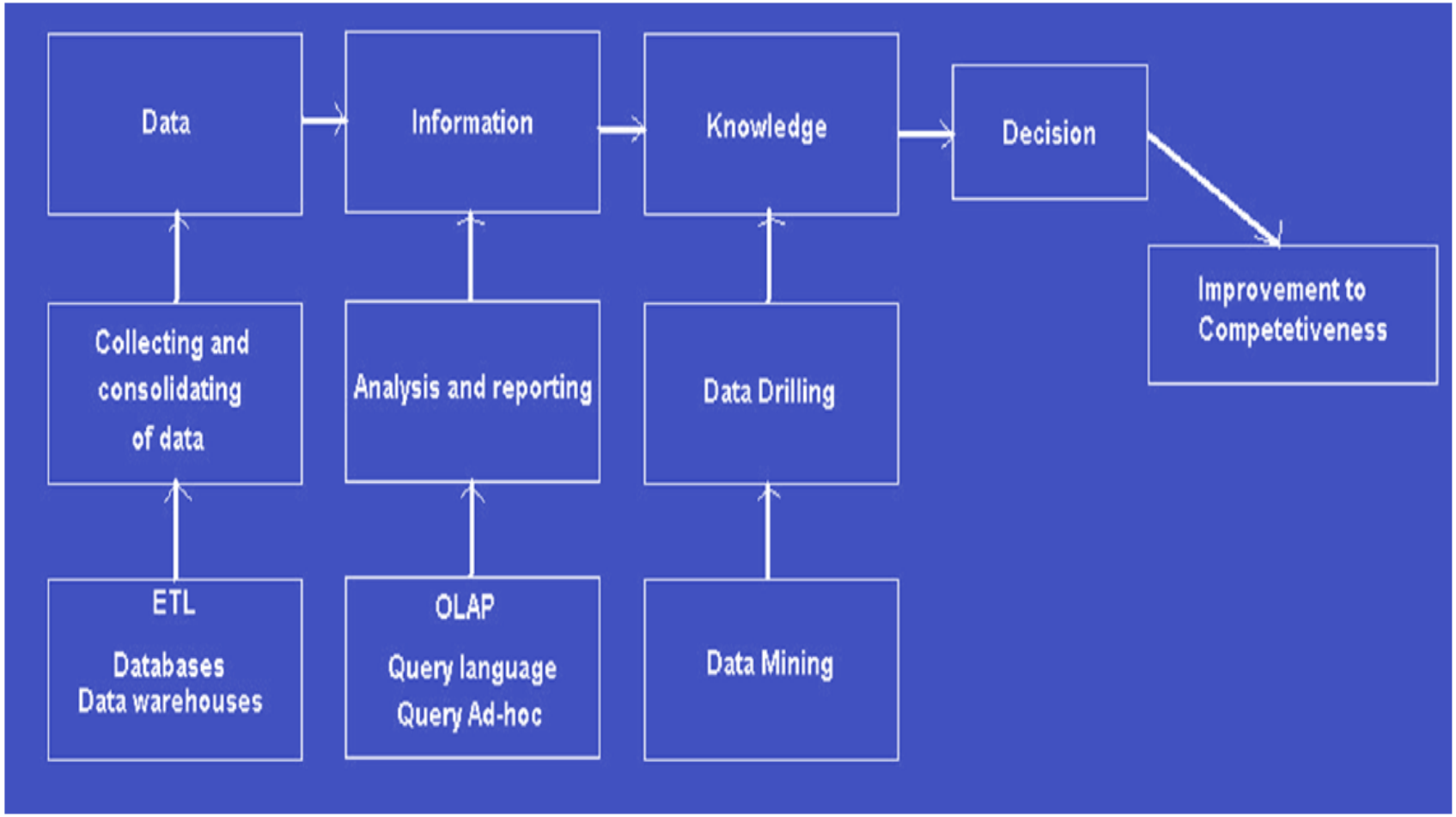
Fit of BI in existing infrastructure

- A BI that includes effective data warehouse and also a reactive component capable of monitoring the time critical operational processes to allow tactical and operational decision makers to tune their actions according to the company strategy.

- BI provides an in-depth analysis of detailed business data, including database and application technologies as well as analysis practices.

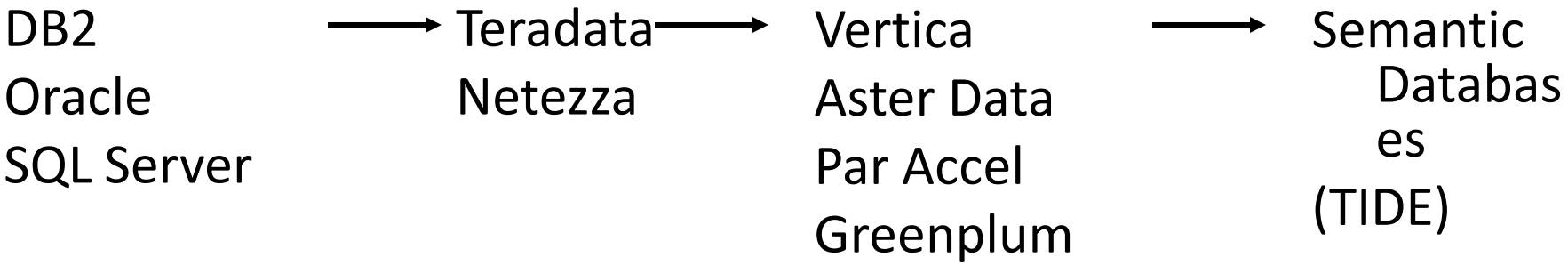
- Applications
  - Strategic management of business
  - Performance measurement of business
  - Collaboration of processes within and external to the organization

BI Process


Data
Information
Knowledge
Decision
Improvement to Competetiveness
Collecting and consolidating of data
Analysis and reporting
Data Drilling
ETL
Databases
Data warehouses
OLAP
Query language
Query Ad-hoc
Data Mining

BI Technologies

•Analytic Databases



•BI is a consolidating industry

- Oracle: Siebel, Hyperion, Brio, Sun
- SAP: Business Objects, Sybase
- IBM: Cognos, SPSS, Coremetrics, Unica, **Netezza**
- EMC: **Greenplum**
- HP: **Vertica**
- Teradata: **Aster Data**

•Independent vendors: MicroStrategy, Informatica, SAS

•Reporting standards determined mainly by Microsoft, Apple and Adobe

BI Technologies (cont'd)

- If you want to learn more about Analytic Databases:

http://hosted.mediasite.com/mediasite/Viewer/?peid=120d6b7ba227498b96a8c0cd01349a791d

- If you want to learn more about BI in the Cloud:

http://hosted.mediasite.com/mediasite/Viewer/?peid=e6d91148a71a47969824c22b3b20d6221d

BI Roles

- Data Warehouse/Business Intelligence Manager
- Data Modeler/Data Architect
- Business Analyst
- Extract, Transform, and Load (ETL) Developer
- Report Developer
- Data Analyst
- Data Scientist

Data Warehouse/Business Intelligence Manager

- **Description**
  - The Data Warehouse/Business Intelligence Manager is responsible for managing and directing the project.
  - This person is the tactical leader and not the executive sponsor.
  - He or she needs to be able to communicate well with both IT and business staff for project success.
- **Skills Needed**
  - Data Architecting Background
  - Project Management
  - Communication and Presentation Skills
  - At minimum, working knowledge of all aspects of the BI stack
- **Common Tools Used**
  - SQL Server Management Studio
  - Microsoft Excel, Microsoft Project, or some other Project Management Tool
  - Data Modeling Software

Data Modeler/Data Architect

- **Description**
  - The Data Modeler/Architect is responsible for developing the marts and warehouse structures based off of the business and technology requirements that have been gathered.
  - They tend to perform quite a bit of data profiling and define the final data warehouse/data mart structure.
- **Skills Needed**
  - Data Architecting Background
  - Data Profiling Skills
  - Advanced Knowledge of SQL and Data Structures
- **Common Tools Used**
  - SQL Server Management Studio
  - Microsoft Excel
  - Data Modeling Software
  - SQL Server Integration Services (SSIS)

Business Analyst

- **Description**
  - The Business Analyst is responsible for defining the business requirements for the business intelligence solution.
  - A business intelligence solution helps solve business problems, therefore this role is critical to project success.
- **Skills Needed**
  - Industry-specific Business Background is preferred
  - General Data Warehouse/Business Intelligence Background
  - Communication and Presentation Skills
- **Common Tools Used**
  - Microsoft Excel
  - SQL Server Management Studio (depending on the technical skills of the analyst)

Extract, Transform, and Load (ETL) Developer

- **Description**
  - The ETL Developer is the programmer who develops the packages that move data from its source location to the ODS (operational data store) and the marts or warehouse.
  - In Microsoft SQL Server implementations, SSIS (SQL Server Integration Services) is the tool most often used.
- **Skills Needed**
  - SQL Programming
  - Some C#/VB.net Background Preferred
  - Knowledge of Data Structures
  - ETL/SSIS Development Experience
- **Common Tools Used**
  - SQL Server Integration Services (SSIS)
  - SQL Server Management Studio
  - Microsoft Excel

Report Developer

- **Description**
  - The Report Developer is the developer responsible for creating reports, dashboards, and other visualizations that provide clarity to end users.
  - Proper data visualization is a key component of a successful business intelligence implementation, so a capable Report Developer is a must.
- **Skills Needed**
  - SQL Programming
  - Some C#/VB.net Background Preferred
  - SQL Server Reporting Services (SSRS), Crystal Reports,
  - Strong Industry-Specific Business Background Preferred
- **Common Tools Used**
  - SQL Server Reporting Services (SSRS), Tableau, Crystal Reports or Other Data Visualization Software
  - SQL Server Management Studio
  - Microsoft Excel
  - Microsoft SharePoint (if SharePoint is the BI Portal in your deployment)
  - R, SAS, SPSS or some other statistical software package

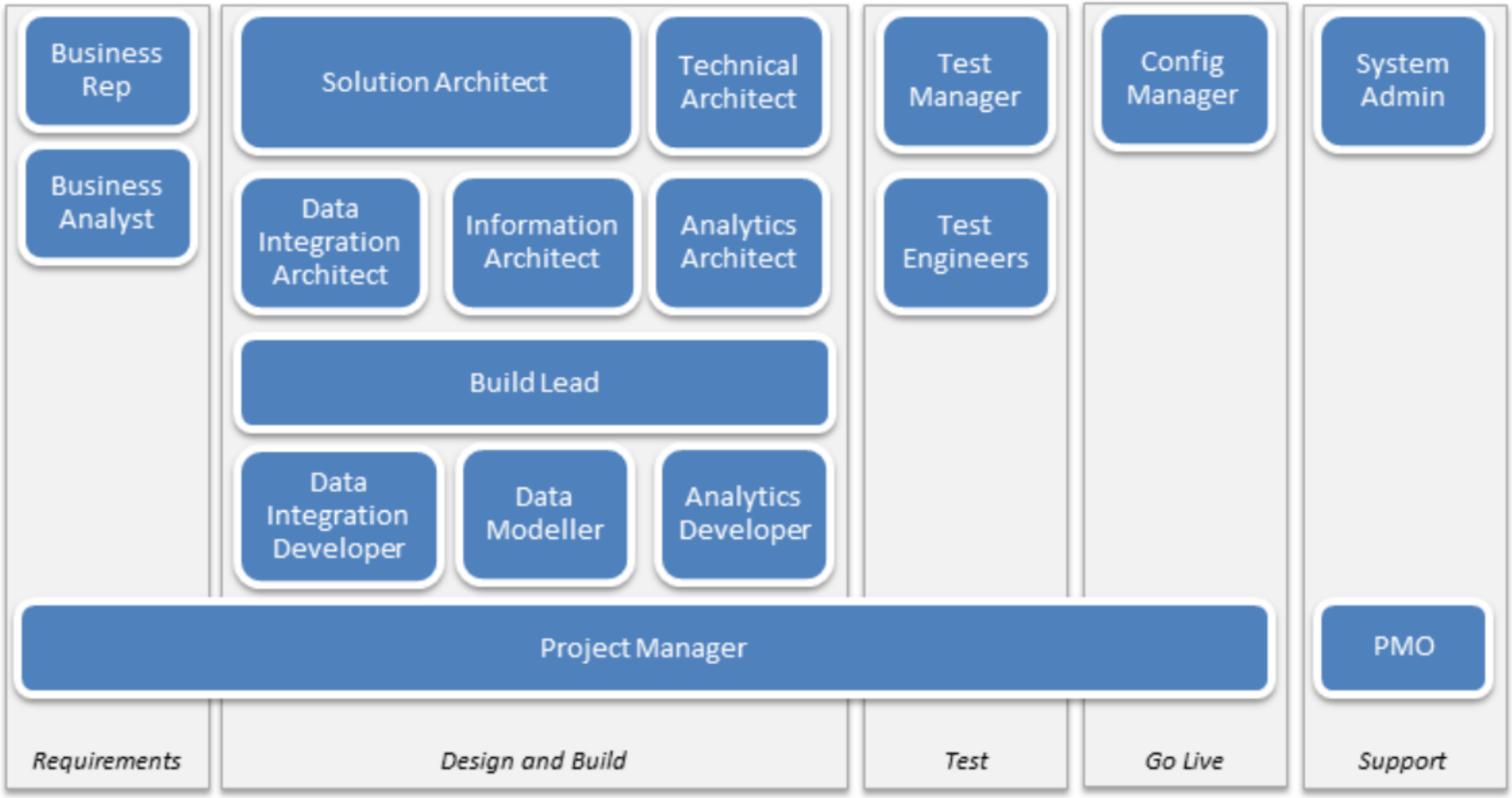
Data Analyst

- **Description**
  - The Data Analyst role is responsible for assisting business users to analyze and interpret the data contained in your data warehouse/data mart environment.
  - This person often works directly with business users to help them solve specific business problems using the data contained in the business intelligence solution.
- **Skills Needed**
  - SQL Programming
  - SQL Server Reporting Services (SSRS), Crystal Reports, Tableau or other data visualization tool
  - Industry-Specific Business Background
  - Statistics or Math Background Preferred
- **Common Tools Used**
  - SQL Server Reporting Services (SSRS), Tableau, Crystal Reports or Other Data Visualization Software
  - SQL Server Management Studio
  - Microsoft Excel
  - Microsoft SharePoint (if SharePoint is the BI Portal in your deployment)
  - Statistical Software Package (R, SAS, SPSS, etc.)

Data Scientist

- **Description**
  - They must be well-versed in statistics and mathematics while also having a strong background in programming or scripting.
  - They perform advanced data-driven statistical modeling to develop predictive models, clustering analysis & other advanced data mining & statistical models.
- **Skills Needed**
  - Advanced Mathematics/Statistics Background
  - SQL Programming
  - R, SAS, SPSS, or Other Statistical Software Package Skills
  - Data Visualization Development
- **Common Tools Used**
  - SQL Server Reporting Services (SSRS), Tableau, Crystal Reports or Other Data Visualization Software
  - SQL Server Management Studio
  - Microsoft Excel
  - R, SAS, SPSS, Python, or other statistical software package.

BI Roles and Responsibilities



BI Team Roles

- **BI Project Team Roles**
- **Business Manager**
- **BI Business Specialist**
- **BI Project Manager**
- **Business Requirements Analyst**
- **Decision Support Analyst**
- **BI Designer**
- **ETL Specialist**
- **Database Administrator**
- **More on http://www.b-eye-network.com/view/8360**

Application of BI in all industries

- A few examples
  - Pro Sports
    - Oakland A's, New England Patriots – recruiting players
    - Dallas Cowboys – merchandising
  - Gambling
    - Harrah's → Caesar's
- Quote from Intel manager:
  - "In God we trust, all others bring data"
    - Demming

A sample of Oracle BI Customers

BI Best Practices |1

1. Get organization wide buy-in

- Disjointed BI practices and failed universal adoption is a quick path to BI failure. To maximize BI success it is essential to get organization wide buy-in. It is in everyone's best interest; in the end every department especially Sales, Marketing, Finance, and Management will benefit from BI. To that end the right parties should be involved from the get go. This best practice includes getting CIOs and CMOs to collaborate from the beginning!
- IT is another key stakeholder you should involve from the beginning. Yes, the right SaaS BI tool won't require IT heavy lifting. This doesn't mean IT shouldn't be included in analytics procurement, implementation, and management processes. IT and Dev teams offer valuable knowledge bases that should be utilized. They can ensure that appropriate security and governance measures are adhered to. IT is a great resource for knowledge retention and knowledge transfer. In the end, IT doesn't necessarily have to manage the BI but they should be kept in the loop.

BI Best Practices |2

2. Have a plan, from the beginning

- Approaching analysis without a strategy can cost you a lot of money, time and stress. Before you even choose a BI tool you need to identify the top business needs and develop a set of business requirements and goals. Don't develop these in a vacuum or just at the executive level. Since you already have organization wide buy-in (right?!) consult with all key stakeholders, including finance, marketing, sales and operations. In the end setting clear objectives and having pre-determined key performance indicators(KPIs) will help guide a successful BI adoption.

BI Best Practices |3

3. Start Small

- Odds are there are a lot of business questions you want answered and insights you need to gain. When you start to compile these needs and begin pondering the overall size of your data it is easy to get overwhelmed. This is only compounded when every employee starts clamoring for data immediately. This leads us to the next business intelligence best practice: start small. Start with a small list of crucial questions. Try to answer them with easily available data. If you end up with answers, then great. Odds are this process will lead to more questions that you can add to your roadmap. This process will also help you discover what data sources you need to compile, or what data you are missing in general. When it comes to BI rollouts, we are fans of Agile methodology best practices which promote an iterative approach that constantly involves stakeholders. Gather requirements and input, build, test, and repeat!

BI Best Practices |4

4. Identify necessary data sources

- Odds are your organization, no matter the size, is collecting data surrounding most business operations. The problem is this data is often spread across a variety of different systems and software. Data may be stored in various ERP systems, CRMs, databases and Excel spreadsheets. With data spread across multiple systems, getting the information you need can be an arduous task. This is where you once again need to start small and identify the necessary data sources to get started. View your data not as an overwhelming mass but as a collection of answers to specific questions. Approach it with a question or a hypothesis in mind and check if the data gathered confirms your assumption. In the end you may take advantage of a data warehouse to improve performance. With the right tool you can easily compile and blend various data sources.

BI Best Practices |5

5. Foster a data culture

- To get your organization on board with a new BI plan you are going to have to evangelize the importance of data. For your BI to succeed your organization needs to foster a data culture. When all departments of a company are working together towards a common goal, the resulting data insights and subsequent actions will be more meaningful and valuable.
- There may be push back. Departments may be discouraged by a lack of time, data acumen and resources and shy away from encouraging enterprise adoption of BI. They may not see that the adoption costs outweigh the benefits. They need to see that the right tool will benefit all teams.
- Some users may just be overwhelmed by new technologies and data. This can lead to change resistance. Work to show that while a little training may be necessary the right tool will easily empower everyone to be their own analyst. Look for a tool that allows for a wide range of users to easily connect to, explore, visualize and communicate their data. Easy drag and drop interfaces that require little training and no prior data analysis or SQL skills will also drive adoption. Empowering staff with business dashboards they can immediately use encourages them to make fast and confident decisions, and benefits the company as a whole.