102046706 – Data Mining & Business Intelligence

Unit-1

Overview and concepts Data Warehousing and Business Intelligence

Outline

- Why Reporting & Analyzing Data?
- Introduction to Business Intelligence
- Introduction to Data Warehousing
- Features of Data Warehousing
- Introduction to Data marts
- Types of Data Marts
- Meta Data

Why Reporting & Analyzing Data?

- The amount of data stored in databases is growing exponentially & databases are now measured in gigabytes(GBs) and terabytes(TBs).
- However row data does not provide useful information.
- In today's highly competitive business environment, companies need to turn these terabytes of raw data into some **useful information**.
- The general methods of analysis/reporting can be broadly classified into two categories: **non-parametric** analysis & **parametric** analysis
- **Example**
  - Managers will generally be more interested in **actual data** and non-parametric analysis results, while engineers will be more concerned with parametric analysis.

What is Business Intelligence?

- BI technologies provide **historical, current and predictive views** of business operations.
- Common functions of business intelligence technologies include reporting, online analytical processing, analytics, data mining, process mining, business performance management, text mining, predictive analytics and prescriptive analytics.
- BI technologies can handle **large amounts of structured** and sometimes **unstructured data** to **help business & also identify, develop** new strategic **business opportunities**.
- Identifying new opportunities and implementing an effective strategy based on insights can provide businesses with a **competitive market advantage** and **long-term stability**.

Business Intelligence (Cont..)

- Business intelligence (BI) make up the **strategies and technologies used by enterprises for the data analysis** of business information.
- BI tools access and analyze data sets and present analytical findings in **reports, summaries, dashboards, graphs, charts and maps** to provide users with **detailed intelligence** about the **state of the business.**
- Typical BI infrastructure components are as follows:
  - Software solution for gathering, cleansing, integrating, analyzing and sharing data.
- It produces analysis and provides **believable information** to help **making effective and high quality business decisions**.

Business Intelligence (Cont..)

- The most common kinds of **business intelligence systems** are:
  - **MIS** - Management Information Systems
  - **CRM** - Customer Relationship Management
  - **EIS** - Executive Information Systems
  - **DSS** - Decision Support Systems
  - **GIS** - Geographic Information Systems
  - **OLAP** - Online Analytical Processing

Types of Users of Business Intelligence:

- **Analyst (Data Analyst or Business Analyst):** They are the statistician of the company, they used BI on the basis of historical data priorly stored in the system.
- **Head or Manager of the Company:** Head of the company uses Business Intelligence used to increase the profitability of their company by increasing the efficiency in their decisions on the basis of all the knowledge they discovered.
- **IT Engineer:** For his company.
- **Small Business Owners:** Can be used by a small businessman because it is quite affordable too.
- **Government Officials:** In the decision-making of the government.

Applications of Business Intelligence:

- In Decision Making of the company by decision-makers of the organizations.
- In Data Mining while extracting knowledge.
- In Operational Analytics and operational management.
- In Predictive Analytics.
- In Prescriptive Analytics.
- Making Structured data from unstructured data.
- In Decision Support System.
- In Executive Information System (EIS).

Advantages of Business Intelligence

1. Decision-making is improved because users have access to real-time data and insights through business intelligence tools. This enables users to base their decisions on correct and current information.
2. **Efficiency gain:** Many manual data analysis operations are automated by business intelligence systems, freeing up time and resources for other tasks.
3. **Better data management:** Business intelligence technologies aid in the administration and organization of data, making it simpler to locate the facts required for decision-making.
4. **Greater visibility:** Business intelligence solutions give users a comprehensive picture of the functioning of the firm, enabling them to spot areas that could use improvement.
5. **A better understanding of customers:** Business intelligence technologies helps firms understand their customers better, enabling them to customize products and services to suit their needs.
6. **Cost savings:** Business intelligence technologies assist firms in locating inefficiencies and cost savings, which boosts revenue.
7. **Better forecasting:** Organizations may evaluate past data and predict future patterns using business intelligence technologies, which enables them to plan more successfully for the future.
8. **Competitive advantage:** By granting access to important data and insights that can guide them in making better decisions, business intelligence technologies provide firms a leg up on their rivals.
9. Collaboration is improved as a result of using business intelligence technologies to disseminate information between teams and departments. This promotes better decision-making and collaboration.
10. **Better Monitoring:** Business intelligence technologies assist firms in tracking important metrics like revenue, customer happiness, and staff performance and in monitoring performance.

Disadvantages of Business Intelligence

1. **Complexity:** The implementation and upkeep of business intelligence systems can be extremely difficult and complicated. This may be a drawback for companies with constrained IT resources.
2. **High costs:** Some businesses find it prohibitively expensive to implement and purchase business intelligence technologies.
3. Business intelligence strongly depends on accurate and current data. The insights produced by business intelligence technologies could not be accurate if the data is inconsistent, erroneous, or incomplete.
4. **Data Security:** Business intelligence systems handle and store a lot of sensitive data, which, if not adequately protected, is susceptible to security breaches.
5. **Dependence on IT:** Because business intelligence solutions frequently rely largely on IT assistance, it may be challenging for enterprises to quickly get the data they require.
6. **Limited scalability:** For firms with huge data volumes, business intelligence solutions may not be able to handle enormous amounts of data.

Business Intelligence	Data Warehouse
It is a set of tools and methods to analyze data and discover, extract and formulate actionable information that would be useful for business decisions.	It is a system for storage of data from various sources in an orderly manner as to facilitate business-minded reads and writes.
It is a Decision Support System (DSS).	It is a data storage system.
Serves at the front end.	Serves at the back end.
The aim of business intelligence is to enable users to make informed, data-driven decisions.	A data warehouse's main aim is to provide the users of business intelligence; a structured and comprehensive view of available data of an organization.
Collects data from the data warehouse for analysis.	Collects data from various disparate sources and organizes it for efficient BI analysis.
Comprises business reports, charts, graphs, etc.	Comprises of data held in “fact tables” and “dimensions” with business meaning incorporated into them.
BI as such doesn't have much use without a data warehouse as large amounts of various and useful data is required for analysis.	BI is one of many use-cases for data warehouses, there are more applications for this system.

Handled by executives and analysts relatively higher up in the hierarchy.	Handled and maintained by data engineers and system administrators who report to/work for the executives and analysts.
The role of Business Intelligence lies in improving the performance of business by utilizing tools and approaches that focus on counts, statistics, and visualization.	The reflection of actual database development and integration process is given by Data Warehouse and in addition, Data Profiling and Company validation standards.
It deals with- • OLAP (Online Analytical Processing) • Data Visualization • Data Mining • Query/Reporting Tools	It deals with- • Acquiring/gathering of data • Metadata management • Cleaning of data • Transforming data • Data dissemination • Data recovery/backup planning
Examples of BI software: SAP, Sisense, Datapine, Looker, etc.	Examples of Data warehouse software: BigQuery, Snowflake, Amazon, Redshift, Panoply, etc.

Introduction to Data Warehouse

- Collections of databases that work together are called **data warehouses.**
- This makes it possible to **integrate data from multiple databases** & it is used to help individuals and organizations **make better decisions.**
- A database consists of one or more files that need to be stored on a computer.
- In large organizations, databases are typically not stored on the individual computers of employees but in a **central system (server).**

Need for Data Warehouse

An ordinary Database can store MBs to GBs of data and that too for a specific purpose. For storing data of TB size, the storage shifted to Data Warehouse. Besides this, a transactional database doesn't offer itself to analytics. To effectively perform analytics, an organization keeps a central Data Warehouse to closely study its business by organizing, understanding, and using its historic data for taking strategic decisions and analyzing trends.

Data Warehouse (Cont..)

- A **server** is a computer system that provides a **service over a network**. The server is often located in a specific place with controlled access, so only authorized person can get physical access for it.
- In a typical setting, the database files reside on the server, but it can be accessed from many different computers in the organization.
- As the number and complexity of databases grows, we start referring to them together as a data warehouse.
- **The ultimate goal of a database is not just to store data, but to help businesses make decisions based on that data.**
- A data warehouse supports this goal by providing an architecture and tools to systematically organize and understand data from multiple databases.

Data Warehouse (Cont..)

- According to William H. Inmon, a leading architect in the construction of data warehouse systems, "A data warehouse is a **subject-oriented**, **integrated**, **time-variant**, and **nonvolatile** collection of data in support of management's decision making process".

- **Features of Data Warehousing**
  - Subject-oriented
  - Integrated
  - Time-variant
  - Nonvolatile

Features of Data Warehouse

- **Subject-oriented:**
  - A data warehouse is organized around major subjects, such as **customer, supplier, product, and sales.**
  - Rather than concentrating on the day-to-day operations and transaction processing of an organization, a **data warehouse focuses on the modeling and analysis of data for decision makers.**
  - Data warehouses **typically provide a simple and concise view around particular subject** issues by excluding data that are not useful in the decision support process.

Features of Data Warehouse (Cont..)

- **Integrated:**
  - A data warehouse is usually constructed by **integrating multiple heterogeneous sources**, such as **relational databases, flat files, and on-line transaction records.**
  - Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures, and so on.
- **Time-variant:**
  - Data are stored to provide information from a historical perspective (e.g., the past 5–10 years).
  - Every key structure in the data warehouse contains, either implicitly or explicitly, an element of time.

Features of Data Warehouse (Cont..)

- **Nonvolatile:**
  - A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment.
  - Due to this separation, a **data warehouse does not require transaction processing, recovery, and concurrency control mechanisms**.
  - It usually requires only two operations in data accessing: **initial loading of data and access of data**.

Data Warehouse Design Process

- A data warehouse can be built using a **top-down** approach, a **bottom-up** approach, or a **combination of both**.
- **Top Down Approach**
  - The top-down approach starts with the overall design and planning.
  - It is useful in cases where the technology is mature and well known, and where the business problems that must be solved are clear and well understood.
- **Bottom up Approach**
  - The bottom-up approach starts with experiments and prototypes.
  - This is useful in the early stage of business modeling and technology development.
  - It allows an organization to move forward at considerably less expense and to evaluate the benefits of the technology before making significant commitments.
- **Combined Approach**
  - In the combined approach, an organization can exploit the planned and strategic nature of the top-down approach while retaining the rapid implementation and opportunistic application of the bottom-up approach.

Types of Data Warehouse

- The three main types of data warehouses are:
  - **Enterprise Data Warehouse**
  - **Operational Data Store**
  - **Data Mart**

Data Warehouse Types (Cont..)

- **Enterprise Data Warehouse**:
  - Enterprise Data Warehouse is a **centralized warehouse**, which provides **decision support service across the enterprise**.
  - It offers a **unified approach to organizing and representing data**.
  - It also provides the **ability to classify data according to the subject** and give access according to those divisions.
- **Operational Data Store**:
  - Operational Data Store, also called ODS, is data store required when neither data warehouse nor OLTP systems support organizations reporting needs.
  - It is widely preferred for **routine activities like storing records.**.
  - In ODS, Data warehouse is refreshed in real time.
- **Data Mart**:
  - A Data Mart is a subset of the data warehouse.
  - It specially designed for specific segments like sales, finance, sales, or finance.
  - In an independent data mart, data can collect directly from sources.

Introduction to Data Marts

- A data mart is a **simple form of a data warehouse** that is focused on a single subject (or functional area), such as **Sales or Finance or Marketing**.
- Data marts are often built and controlled by a **single department within an organization,** given their single-subject focus, data marts usually draw data from only a few sources.
- The sources could be internal operational systems, a central data warehouse, or external data.

Introduction to Data Marts (Cont..)

- A data mart is a repository of data that is **designed to serve a particular community of knowledge workers**.
- The difference between a data warehouse and a data mart can be confusing because the two terms are sometimes used incorrectly as synonyms.
- A data warehouse is a **central repository for all an organization's data**.
- The goal of a data mart, however, is to meet the particular demands of a specific group of users within the organization, such as human resource management (HRM).
- Generally, an organization's data marts are subsets of the organization's data warehouse.

Introduction to Data Marts

- A data mart contains a subset of corporate-wide data that is of value to a specific group of users.
- The scope is confined to specific selected subjects.
- For example, a marketing data mart may confine its subjects to customer, item, and sales.
- The data contained in data marts tend to be summarized.

Introduction to Data Marts (Cont..)

- Depending on the source of data, data marts can be categorized as **independent** or **dependent**.
- **Independent data** marts are sourced from data captured from one or more operational systems or external information providers, or from data generated locally within a particular department or geographic area.
- Dependent data marts are sourced directly from enterprise data warehouses.

Virtual warehouse

- A virtual warehouse is a set of views over operational databases.
- For efficient query processing, only some of the possible summary views may be materialized.
- A virtual warehouse is easy to build but requires excess capacity on operational database servers.

Reasons for Creating a Data Marts

- Easy access to frequently needed data
- Creates collective view by a group of users
- Improves end-user response time
- Ease of creation
- Lower cost than implementing a full data warehouse
- Potential users are more clearly defined than in a full data warehouse
- Contains only business essential data and is less cluttered

Data Warehouse v/s Data Mart

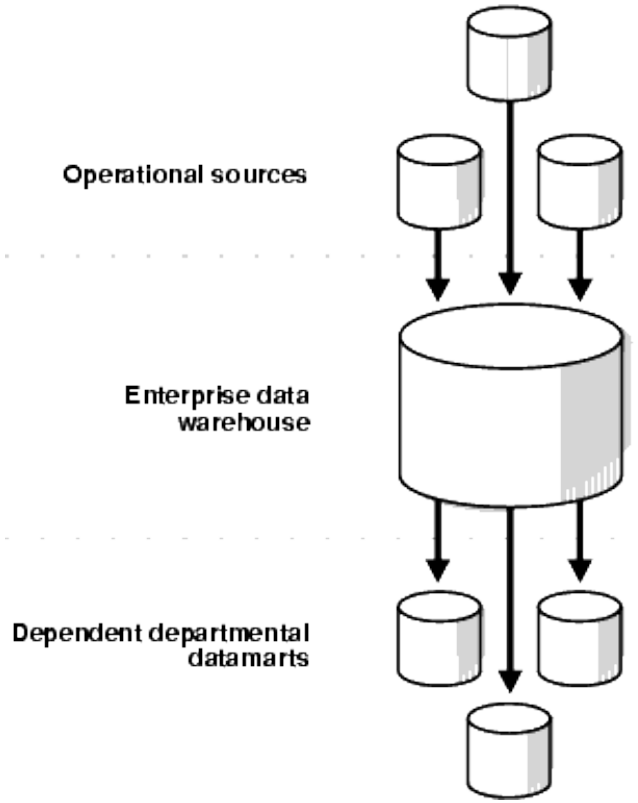
- **Data warehouse**:
  - Holds multiple subject areas Holds very
  - detailed information
  - Works to integrate all data sources Size
  - (typical) 100 GB-TB+
  - Implementation Time : Months to Years
- **Data mart**:
  - Often holds only one subject area- for example, Finance, or Sales
  - May hold more summarized data
  - Concentrates on integrating information from a given subject area or set of source systems
  - Size (typical) < 100GB
  - Implementation Time : Months

Types of Data Marts

- There are three kinds of Data-Marts (DMs), which are as follows:
  1) **Dependent DM**: Created from a data warehouse to a separate physical data-store. (build over data warehouse physically)
  2) **Independent DM**: Created from operational systems and have separate physical data-store.
  3) **Logical or Hybrid DM**: Exists as a subset of data warehouse. (build over data warehouse logically)

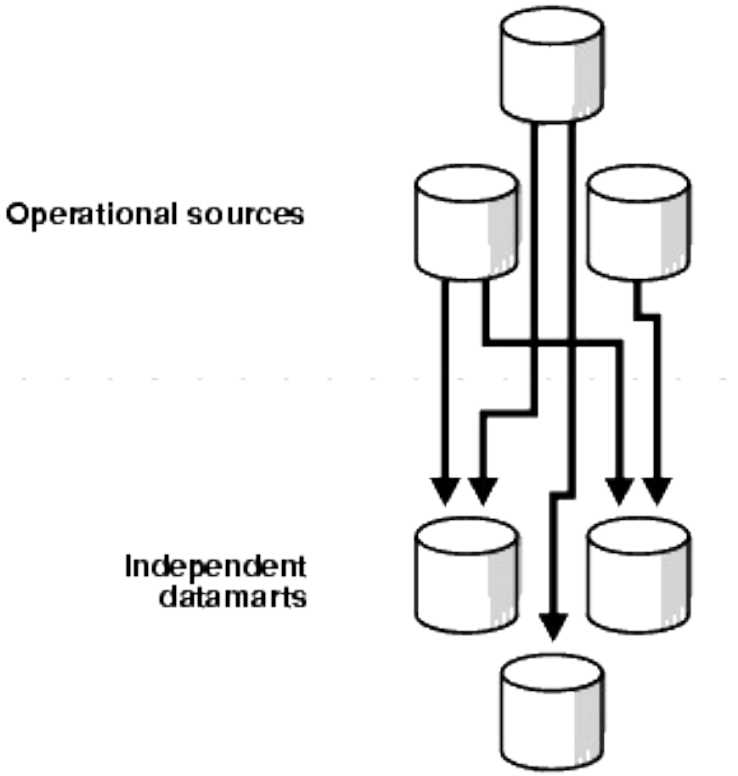
1) Dependent Data Marts

- A dependent data mart allows you to unite your organization's data in one data warehouse.
- This gives you the usual advantages of centralization.



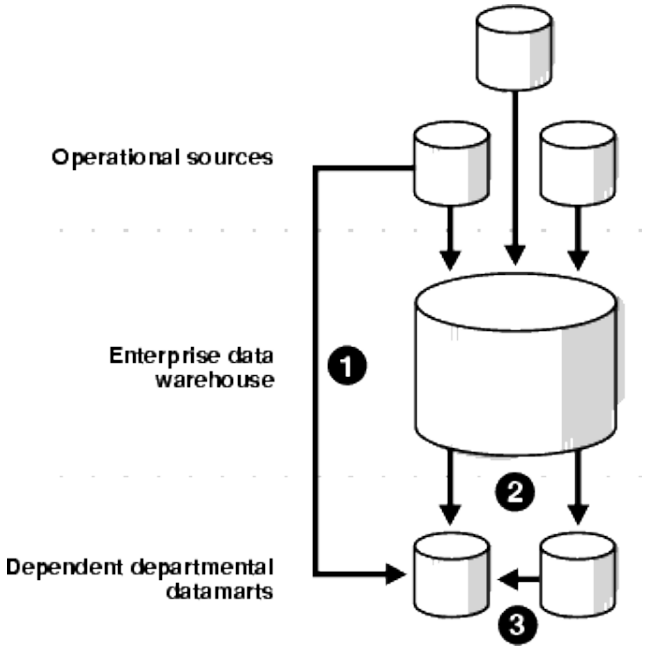
2) Independent Data Marts

- An independent data mart is created without the use of a central data warehouse.
- This could be desirable for smaller groups within an organization.



3) Hybrid Data Mart

- A hybrid data mart allows you to combine input from sources other than a data warehouse.
- This could be useful for many situations, especially when you need ad hoc integration, such as after a new group or product is added to the organization.



Meta data

- Metadata are **data about data**.
- When meta data is used in a data warehouse, that defines warehouse objects.
- Metadata are created for the data names and definitions of the given warehouse.
- Additional metadata are created and captured for time stamping any extracted data, the source of the extracted data, and missing fields that have been added by data cleaning or integration processes.

Meta data

- A metadata repository should contain the following:
- A description of the **data warehouse structure**, which includes the warehouse schema, view, dimensions, hierarchies, and derived data definitions, as well as data mart locations and contents.
- **Operational metadata**, which include data lineage (history of migrated data and the sequence of transformations applied to it), currency of data (active, archived, or purged), and monitoring information (warehouse usage statistics, error reports, and audit trails).
- The **algorithms used for summarization**, which include measure and dimension definition algorithms, data on granularity, partitions, subject areas, aggregation, summarization, and predefined queries and reports.

Meta data

- **Mapping from the operational environment to the data warehouse**, which includes databases and their contents, gateway descriptions, data partitions, data extraction, cleaning, transformation rules and defaults, data refresh and purging rules, and security (user authorization and access control).
- **Data related to system performance**, which include indices and profiles that improve data access and retrieval performance, in addition to rules for the timing and scheduling of refresh, update, and replication cycles.
- **Business metadata**, which include business terms and definitions, data ownership information, and charging policies.

Metadata – Example

- **To Describe Meta Data of a Book Store:**
  - Name of Book
  - Summary of the Book
  - The Date of publication
  - High level description of what it contains
  - How you can find the book
  - Author of the book
  - Whether the book is available OR not
- **The information helps you to:**
  - Search for the book
  - Access the book
  - Understand the book before you access OR buy it.

102046706 – Data Mining & Business Intelligence

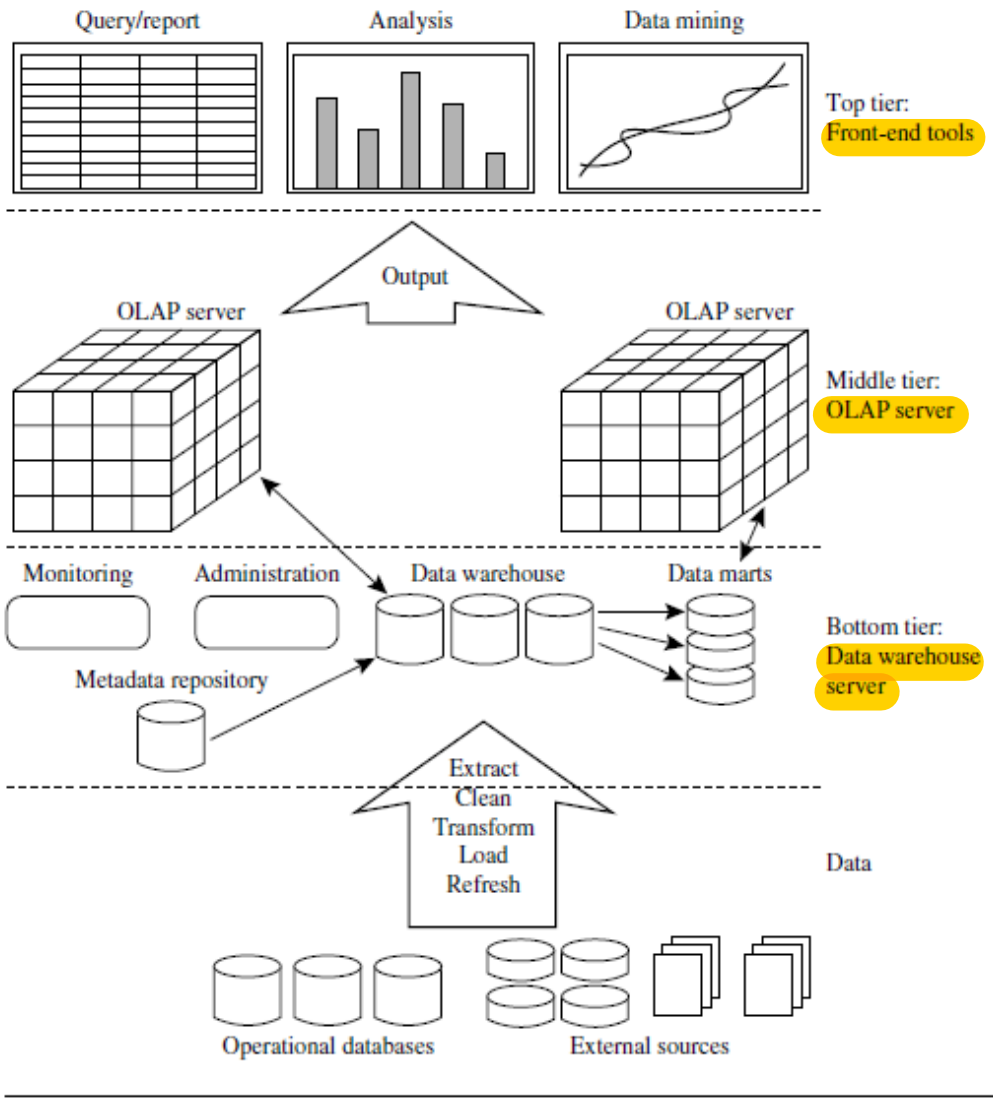
Unit-2

The Architecture of BI and DW

Outline

- Data Warehouse Architecture
- OLTP v/s OLAP
- Data Warehouse Schema Architecture
- OLAP Operations
- OLAP Servers

Three-tier Data Warehouse architecture



Data Warehouse Architecture

- **Bottom tier:**
  - The **bottom tier** is a warehouse **database server** that is almost always a relational database system.
  - **Back-end tools and utilities are used to feed data** into the bottom tier from operational databases or other external sources.
  - These tools and utilities **perform data extraction, cleaning, and transformation**, as well as load and refresh functions to update the data warehouse.
  - The data are extracted using application program interfaces known as **gateways**.
  - A gateway is supported by the underlying DBMS and allows client programs to generate SQL code to be executed at a server.
  - Examples of gateways include ODBC (Open Database Connection) and OLEDB (Open Linking and Embedding for Databases) by Microsoft and JDBC (Java Database Connection).
  - This tier also **contains a metadata repository**, which stores information about the data warehouse and its contents.

Data Warehouse Architecture

- **Middle tier:**
  - The middle tier is an OLAP (Online Analytical Processing Server) that is typically implemented using either
    - A **relational OLAP (ROLAP)** model, that is, an extended relational DBMS that maps operations on multidimensional data to standard relational operations or,
    - A **multidimensional OLAP (MOLAP)** model, that is, a special-purpose server that directly implements multidimensional data and operations.
- **Top tier:**
  - The top tier is a front-end client layer, which contains **query and reporting tools, analysis tools, and/or data mining tools**.

OLAP (**O**n-**L**ine **A**nalytical **P**rocessing)

- Data warehouse systems, serve users or knowledge workers in the role of data analysis and decision-making. Such systems can organize and present data in various formats in order to accommodate the diverse needs of different users. These systems are known as online analytical processing (OLAP) systems.
- OLAP is characterized by relatively **low volume of transactions**.
- Queries are often **very complex and involve aggregations**.
- For OLAP systems a **response time is an effectiveness measure**.
- OLAP applications are widely used by Data Mining techniques.
- In OLAP database there is **aggregated, historical data, stored in multi-dimensional** schemas (usually star schema).

OLTP (**O**n-**L**ine **T**ransaction **P**rocessing)

- The major task of online operational database systems is to perform online transaction and query processing. These systems are called online transaction processing (OLTP) systems.
- They cover most of the day-to-day operations of an organization such as purchasing, inventory, manufacturing, banking, payroll, registration, and accounting.
- It is characterized by a **large number of short on-line transactions** (INSERT, UPDATE, DELETE).
- The main emphasis for OLTP systems is put on very fast query processing, maintaining data integrity in multi-access environments and an effectiveness measured by number of transactions per second.
- In OLTP database, **there is detailed and current data**, and schema used to store transactional databases is the entity model (usually 3NF).

OLTP v/s OLAP (Understanding)

OLTP	OLAP
Many Short Transactions (Queries + Updates)	Long Transactions (Complex Queries)
Examples • Update account balance • Enroll in course • Add book to shopping cart	**Examples** • Report total sales for each department in each month • Identify top-selling books • Count classes with fewer than 10 students
Queries touch small amount of data (one record or few records)	Queries touch large amount of data
Updates are frequent	Updates are infrequent

OLTP v/s OLAP

Functionality	OLTP	OLAP
Characteristic	Operational processing informational processing	Transaction Analysis
Orientation	Transaction	Analysis
User	Clerk, DBA, database professional	Knowledge worker (e.g., manager, executive, analyst)
Function	day-to-day operations	long-term informational requirements, decision support
DB design	ER based, **application-oriented**	Star/snowflake, **subject-oriented**
Data	Current; guaranteed up-to-date	Historical; accuracy maintained over time
Summarization	Primitive, highly detailed	Summarized, consolidated
View	Detailed, flat relational	Summarized, multidimensional
Unit of work	Short, simple transaction	Complex query
Access	Read/write	Mostly read

Data Warehouse Modeling: Data Cube and OLAP

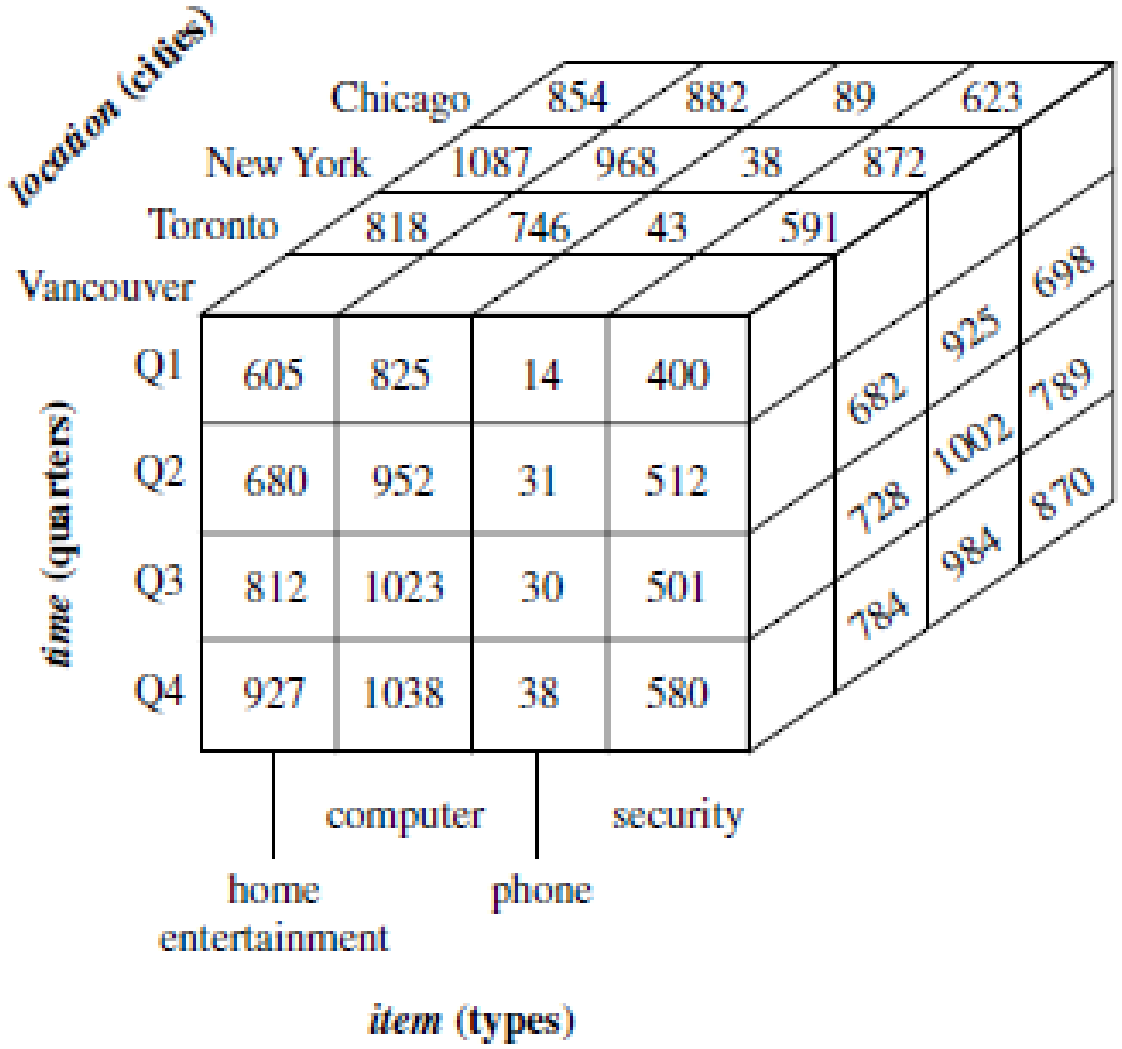
- Data warehouses and OLAP tools are based on a multidimensional data model.
- This model views data in the form of a data cube.
- A data cube allows data to be modeled and viewed in multiple dimensions.
- It is defined by dimensions and facts.
- **Dimensions** are the perspectives or entities with respect to which an organization wants to keep records.

Example of Data Cube

3-D View of Sales Data for *AllElectronics* According to *time*, *item*, and *location*

	location = "Chicago"				***location*** = "New York"				***location*** = "Toronto"				***location*** = "Vancouver"			
	item				***item***				***item***				***item***			
	home				*home*				*home*				*home*			
time	*ent.*	*comp.*	*phone*	*sec.*	*ent.*	*comp.*	*phone*	*sec.*	*ent.*	*comp.*	*phone*	*sec.*	*ent.*	*comp.*	*phone*	*sec.*
Q1	854	882	89	623	1087	968	38	872	818	746	43	591	605	825	14	400
Q2	943	890	64	698	1130	1024	41	925	894	769	52	682	680	952	31	512
Q3	1032	924	59	789	1034	1048	45	1002	940	795	58	728	812	1023	30	501
Q4	1129	992	63	870	1142	1091	54	984	978	864	59	784	927	1038	38	580

Example of Data Cube



Data Warehouse Schema Architecture

- Data Warehouse environment usually transforms the relational data model into some special architectures.
- There are many schema models designed for data warehousing but the most commonly used are:
  - **Star Schema**
  - **Snowflake Schema**
  - **Fact constellation(Group of star, Collection of fact tables) Schema**
- The determination of which schema model should be used for a data warehouse based upon the analysis of project requirements, accessible tools and project team preferences.

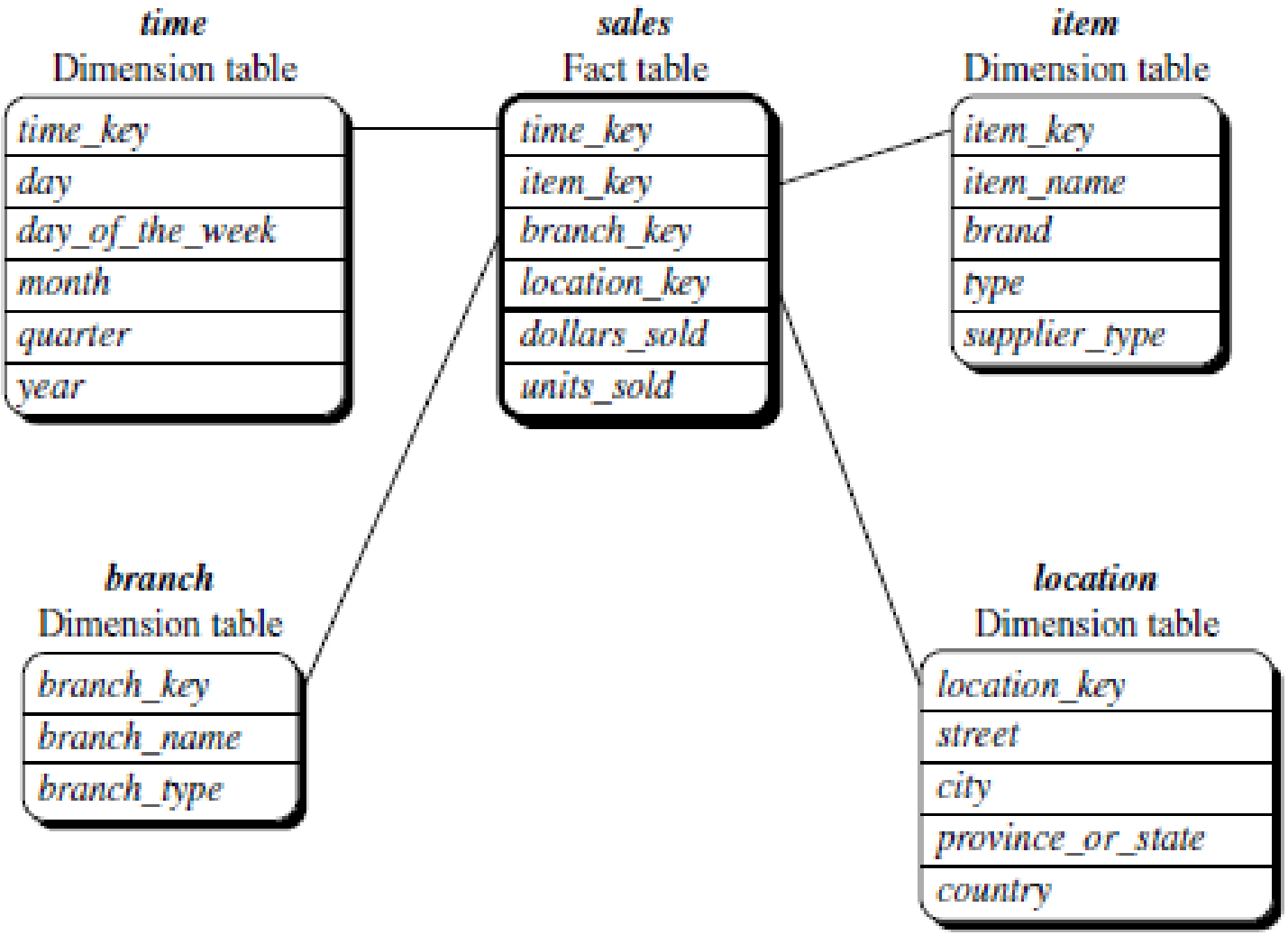
Star Schema

- **Star schema**: The most common modeling paradigm is the star schema, in which the data warehouse contains (1) a large central table **(fact table)** containing the bulk of the data, with no redundancy, and (2) a set of smaller attendant tables **(dimension tables),** one for each dimension.
- The schema graph resembles a starburst, with the dimension tables displayed in a radial pattern around the central fact table.

Star Schema

- The star schema architecture is the **simplest data warehouse schema.**
- It is called a star schema because the diagram resembles a **star**, with points radiating from a center.
- The center of the star consists of **fact table** and the **points of the star are the dimension tables**.
- Usually the fact tables in a star schema are in third normal form (3NF) whereas dimensional tables are de-normalized.
- Despite the fact that the star schema is the simplest architecture, it is **most commonly used nowadays** and is recommended by Oracle.

Star Schema - Example



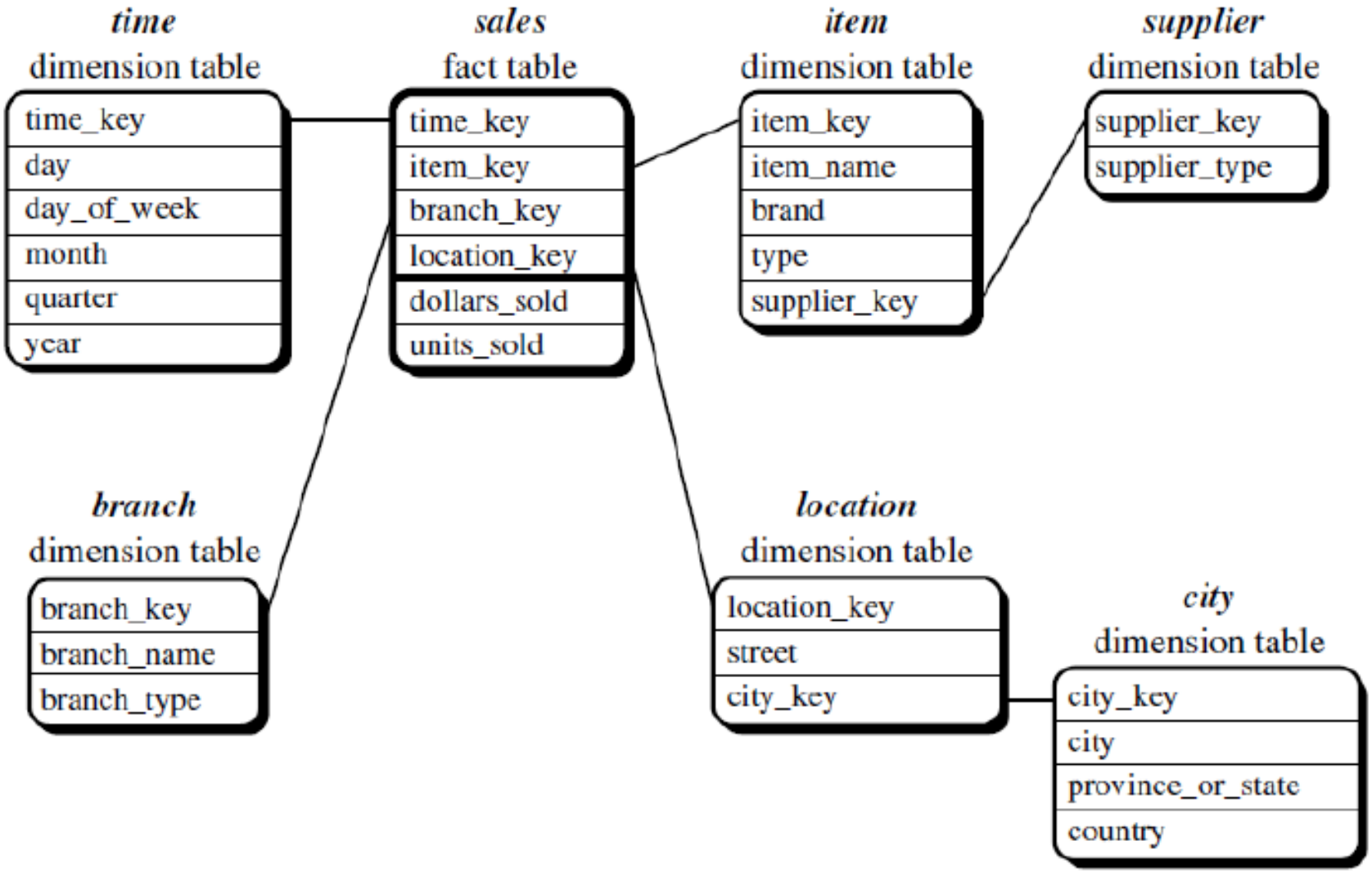
Snowflake Schema

- Snowflake schema: The snowflake schema is a variant of the star schema model, where some dimension tables are normalized, thereby further splitting the data into additional tables.
- The resulting schema graph forms a shape similar to a snowflake.

Snowflake Schema

- The snowflake schema architecture is a **more complex variation of the star schema** used in a data warehouse, because the tables which describe the dimensions are normalized.
- This table is easy to maintain and saves storage space.
- However, this saving of space is negligible in comparison to the typical size of the fact table.
- Furthermore, the snowflake structure can reduce the effectiveness of browsing, since **more joins** will be needed to execute a query.
- Hence, although the **snowflake schema reduces redundancy**, it is not as popular as the star schema in data warehouse design.

Snowflake Schema - Example



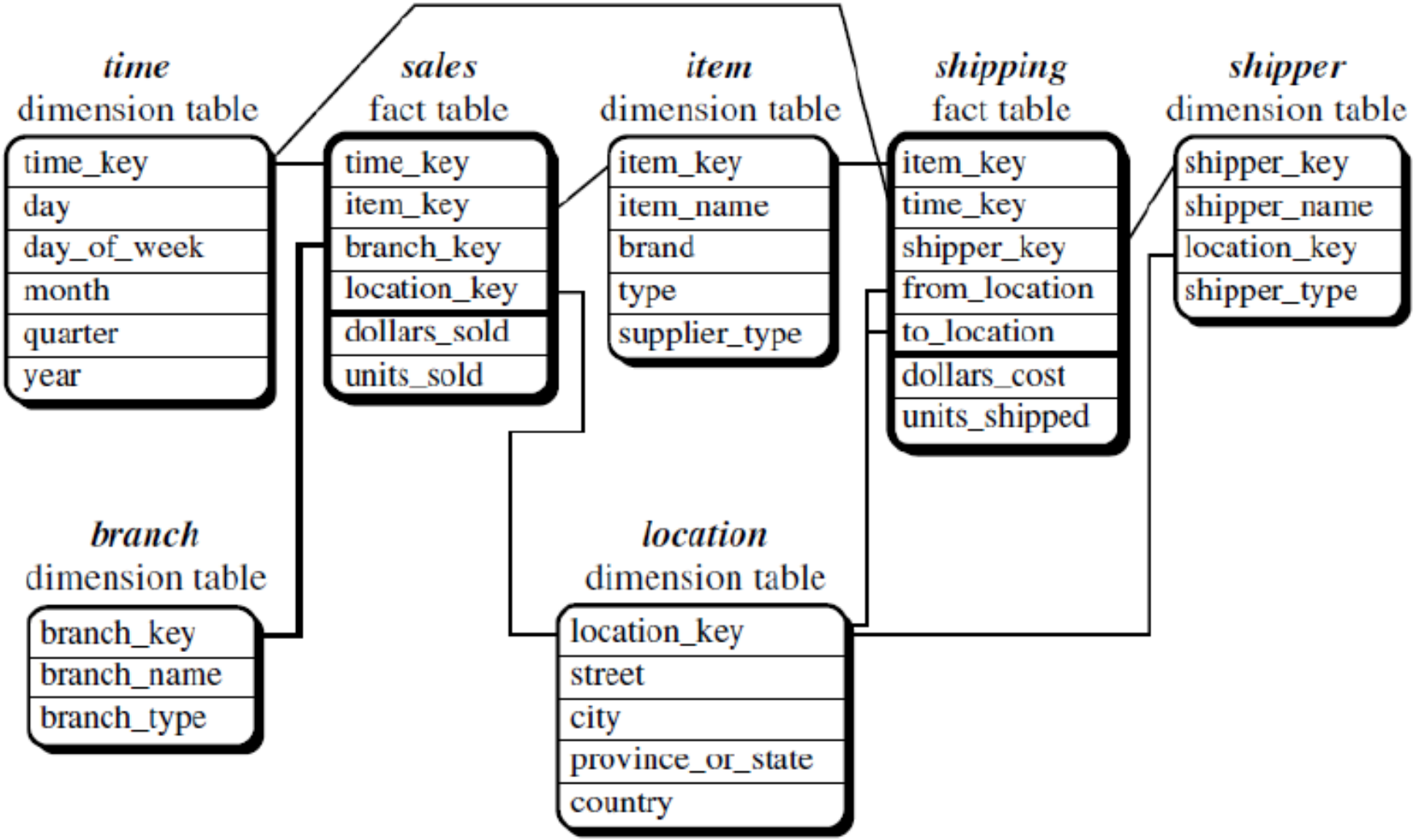
Fact Constellation Schema

- **Fact constellation or Galaxy Schema:** Sophisticated applications may require multiple fact tables to share dimension tables. This kind of schema can be viewed as a collection of stars, and hence is called a galaxy schema or a fact constellation.

Fact Constellation Schema

- Sophisticated applications may require **multiple fact tables** to share dimension tables.
- This kind of schema can be viewed as a **collection of stars**, and hence is called a **galaxy schema** or a **fact constellation**.
- A fact constellation schema allows dimension tables to be shared between fact tables.
- For example, the dimensions tables for ***time***, ***item***, and ***location*** are shared between both the **sales and shipping** fact tables.
- The main shortcoming of the fact constellation schema is a more complicated design because many variants for particular kinds of aggregation must be considered and selected.

Fact Constellation Schema



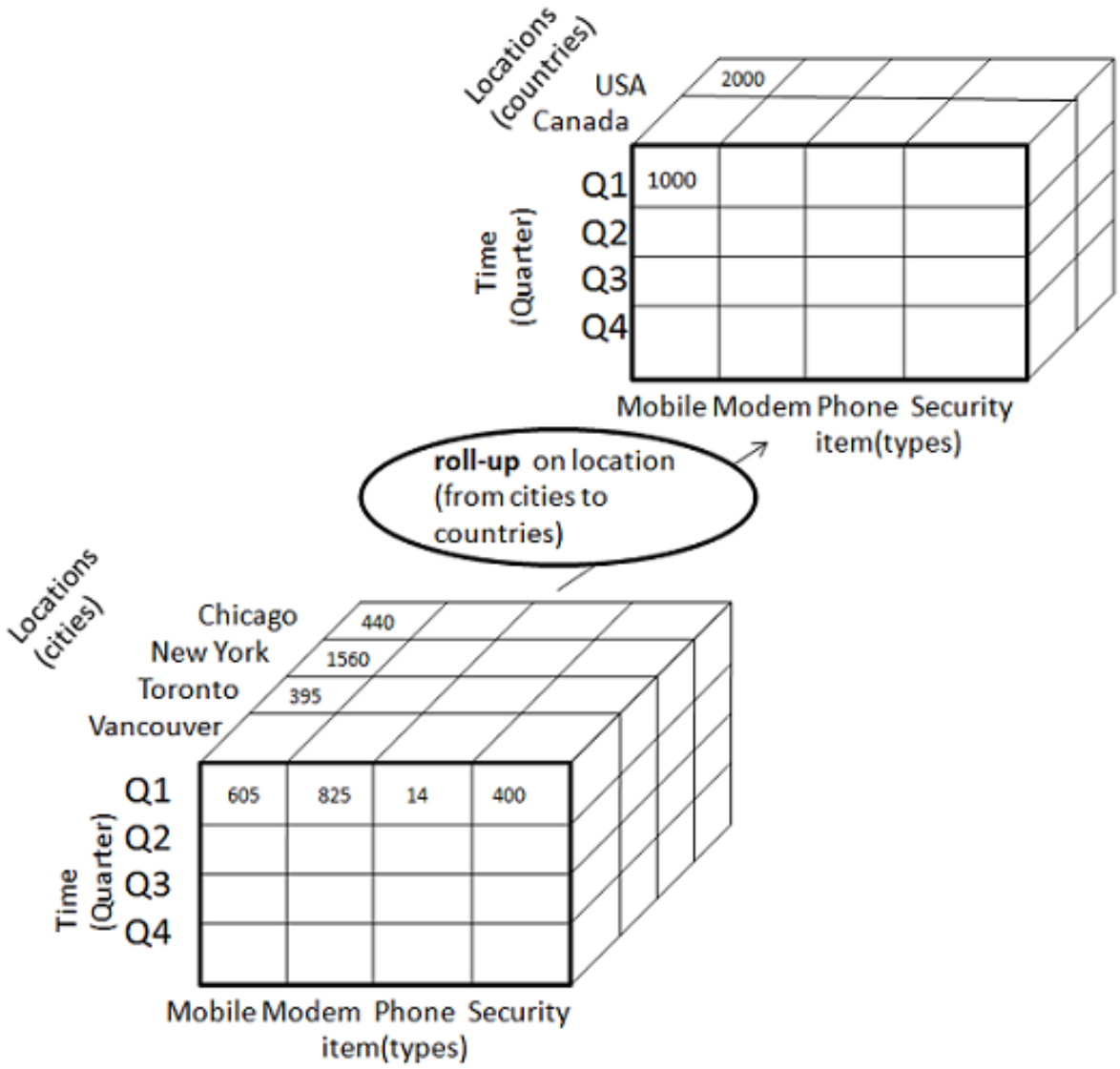
OLAP Operations

- Roll up
- Drill Down
- Slice
- Dice
- Pivot (Rotate)

Roll up – OLAP Operation

- The roll-up operation (also called drill-up or aggregation operation) **performs aggregation on a data cube** by following ways:
  - By climbing up a concept hierarchy for a dimension
  - By dimension reduction
- Roll-up is performed by **climbing up** a concept hierarchy for the dimension location.
- Initially the concept hierarchy was "street < city < province < country".
- On rolling up, the **data is aggregated by ascending the location hierarchy from the level of city to the level of country**.
- The data is grouped into cities rather than countries.
- When roll-up is performed, one or more dimensions from the data cube are removed.

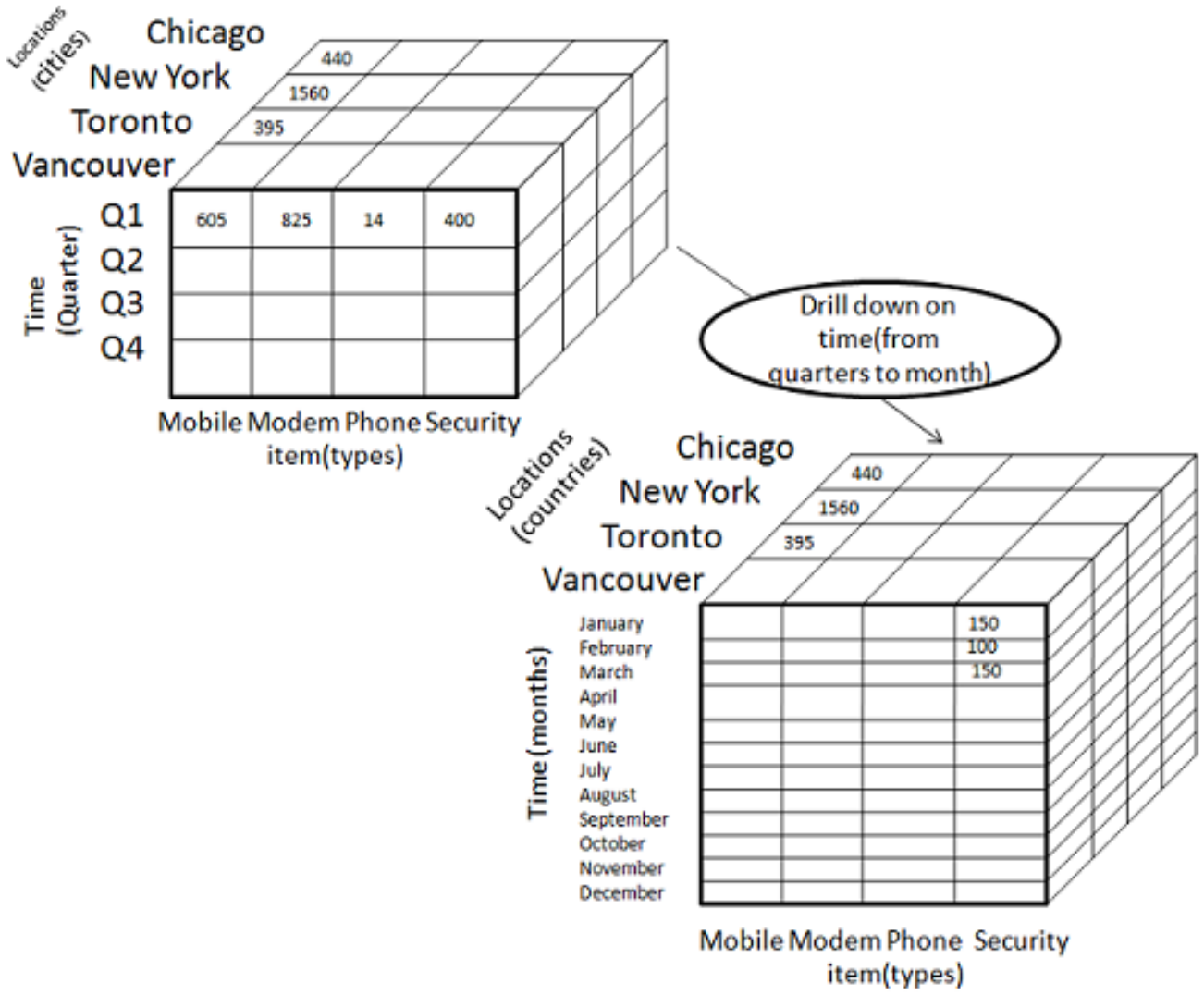
Roll up – OLAP Operation



Drill Down – OLAP Operation

- Drill-down is the reverse operation of roll-up. It is performed by either of the following ways:
  - By stepping down a concept hierarchy for a dimension
  - By introducing a new dimension
- Drill-down is performed by **stepping down** a concept hierarchy for the dimension time.
- Initially the concept hierarchy was "day < month < quarter < year."
- On drilling down, **the time dimension is descended from the level of quarter to the level of month**.
- When drill-down is performed, **one or more dimensions from the data cube are added**.
- It navigates the data from less detailed data to highly detailed data.

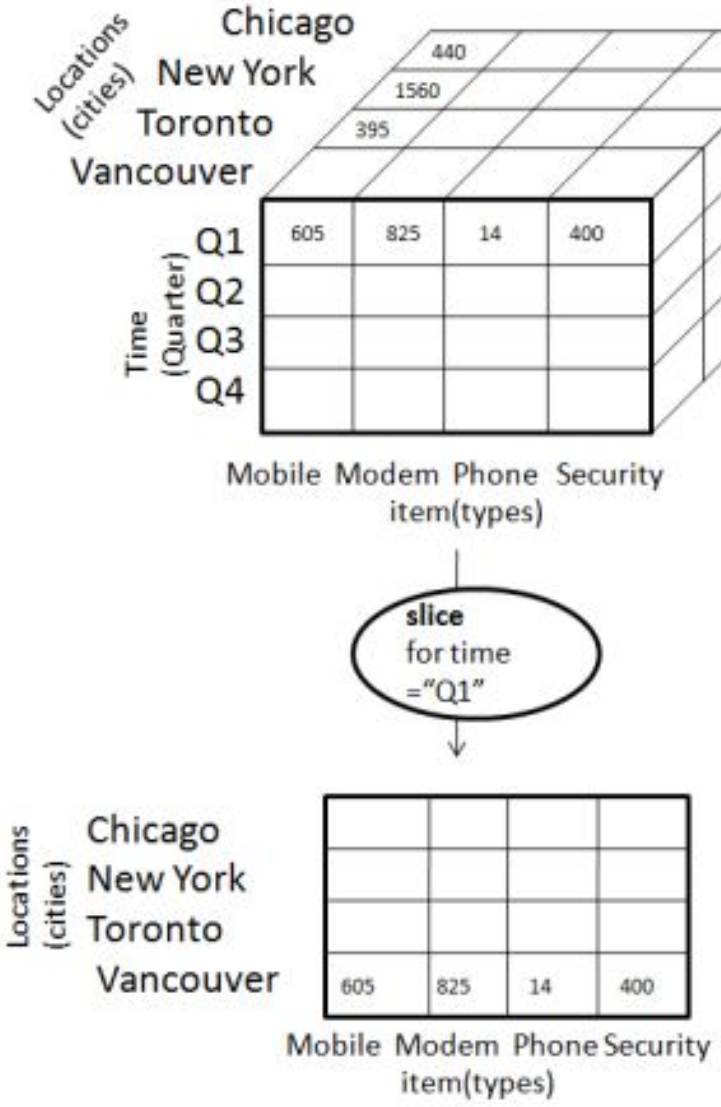
Drill Down – OLAP Operation



Slice – OLAP Operation

- The slice operation **selects one particular dimension from a given cube and provides a new sub cube**.
- Here Slice is performed for the dimension "time" using the criterion time = "Q1“, time = "Q2“, time = "Q3“ etc.
- It will form a new sub-cube by selecting one or more dimensions.

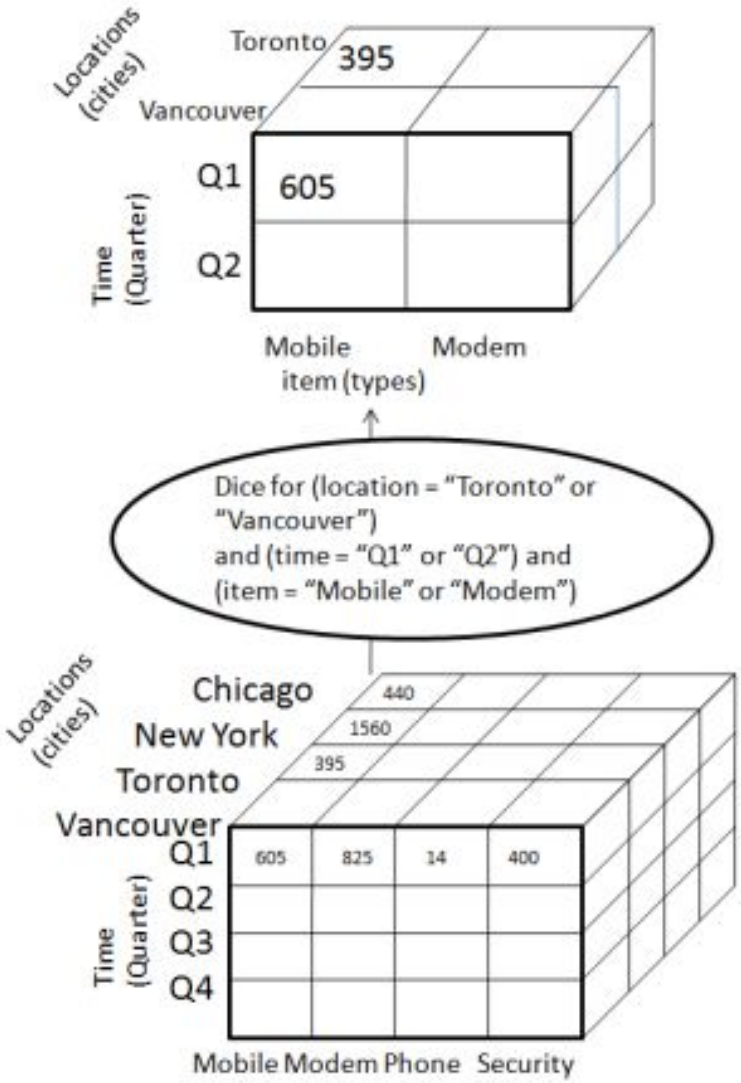
Slice – OLAP Operation



Dice – OLAP Operation

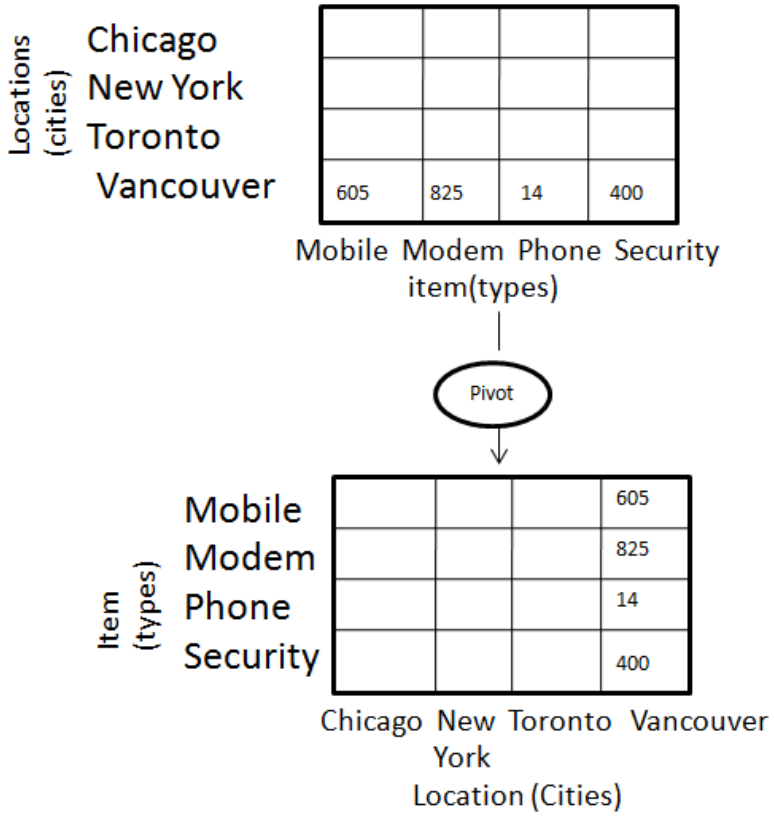
- Dice **selects two or more dimensions** from a given cube and provides a new sub cube.
- The dice operation on the cube based on the following selection criteria involves three dimensions.
  - (location = "Toronto" or "Vancouver")
  - (time = "Q1" or "Q2")
  - (item =" Mobile" or "Modem")

Dice – OLAP Operation



Pivot – OLAP Operation

- The pivot operation is also known as **rotation**.
- It rotates the data axes in view in order to provide an alternative presentation of data.



OLAP Servers

- Relational OLAP (ROLAP)
- Multidimensional OLAP (MOLAP)
- Hybrid OLAP (HOLAP)

Relational OLAP (ROLAP)

- Relational On-Line Analytical Processing (ROLAP) work mainly for the data that resides in a **relational database**, where the base data and dimension tables are **stored as relational tables**.
- 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.
  - **Advantages of ROLAP**
    - ROLAP can handle large amounts of data.
    - Can be used with data warehouse and OLTP systems.
  - **Disadvantages of ROLAP**
    - Limited by SQL functionalities.
    - Hard to maintain aggregate tables.

Multidimensional OLAP (MOLAP)

- Multidimensional On-Line Analytical Processing (MOLAP) support **multidimensional views of data** through array-based multidimensional storage engines.
- With multidimensional data stores, the storage utilization may be low if the data set is sparse.
  - **Advantages of MOLAP**
    - Optimal for slice and dice operations.
    - Performs better than ROLAP when data is dense(heavy).
    - Can perform complex calculations.
  - **Disadvantages of MOLAP**
    - Difficult to change dimension without re-aggregation.
    - MOLAP can handle limited amount of data.

Hybrid OLAP (HOLAP)

- 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.
  - **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.

Basis	ROLAP	MOLAP	HOLAP
Storage location for summary aggregation	Relational Database is used as storage location for summary aggregation.	Multidimensional Database is used as storage location for summary aggregation.	Multidimensional Database is used as storage location for summary aggregation.
Processing time	Processing time of ROLAP is very slow.	Processing time of MOLAP is fast.	Processing time of HOLAP is fast.
Storage space requirement	Large storage space requirement in ROLAP as compare to MOLAP and HOLAP.	Medium storage space requirement in MOLAP as compare to ROLAP and HOLAP.	Small storage space requirement in HOLAP as compare to MOLAP and ROLAP.
Storage location for detail data	Relational database is used as storage location for detail data.	Multidimensional database is used as storage location for detail data.	Relational database is used as storage location for detail data.
Latency	Low latency in ROLAP as compare to MOLAP and HOLAP.	High latency in MOLAP as compare to ROLAP and HOLAP.	Medium latency in HOLAP as compare to MOLAP and ROLAP.
Query response time	Slow query response time in ROLAP as compare to MOLAP and HOLAP.	Fast query response time in MOLAP as compare to ROLAP and HOLAP.	Medium query response time in HOLAP as compare to MOLAP and ROLAP.