School of Nation Builders

NATIONAL COLLEGE OF SCIENCE & TECHNOLOGY

COMPUTER STUDIES DEPARTMENT

NET 002

INTRODUCTION TO DATAWAREHOUSING

TOPIC 1

PRELIM

SCHEMA

Schema is a logical description of the entire database. It includes the name and description of records of all record types including all associated data-items.

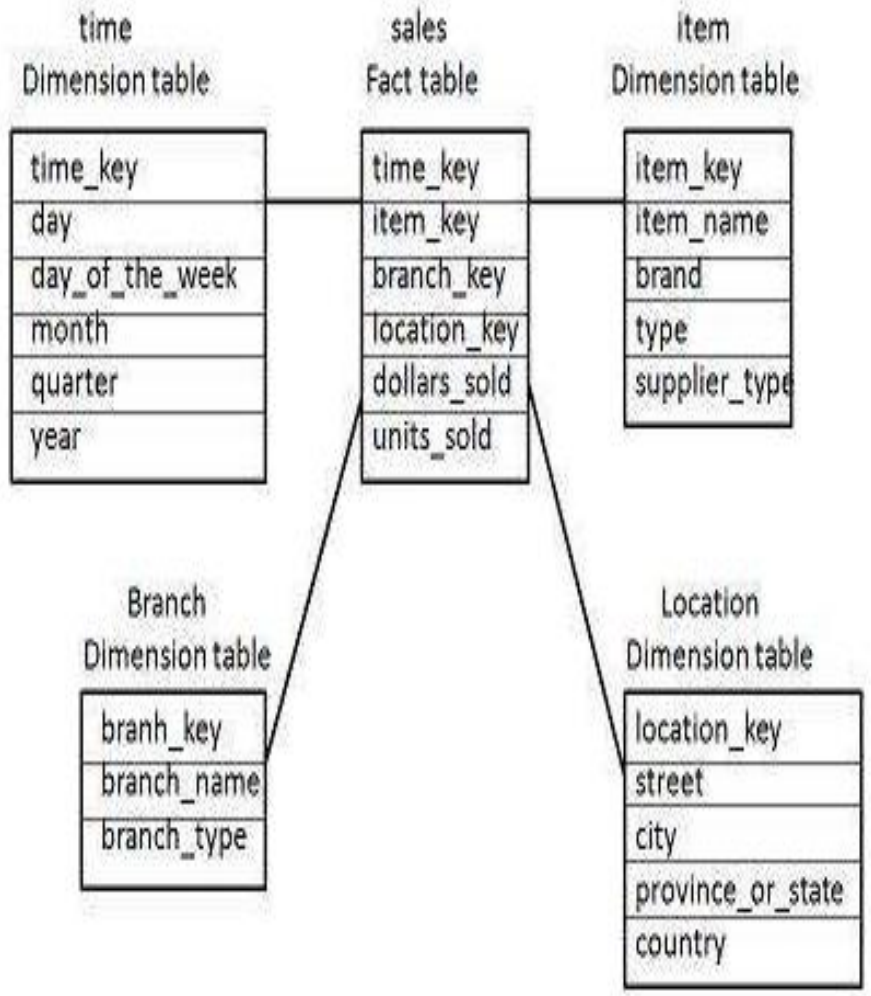
STAR SCHEMA

Each dimension in a star schema is represented with only one-dimension table.

This dimension table contains the set of attributes.

The following diagram shows the sales data of a company with respect to the four dimensions, namely time, item, branch, and location.

There is a fact table at the center. It contains the keys to each of four dimensions.



SNOW FLAKE SCHEMA

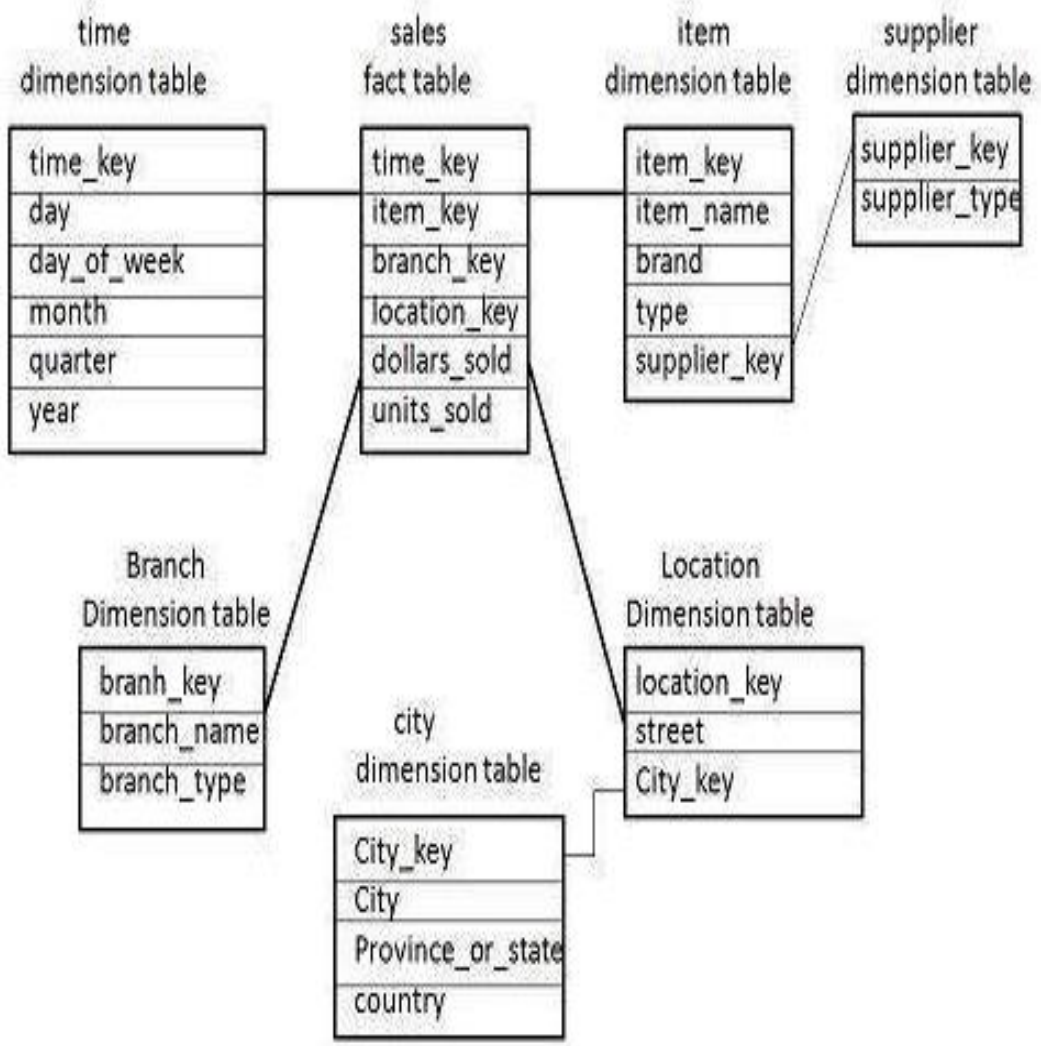
Some dimension tables in the Snowflake schema are normalized.

The normalization splits up the data into additional tables.

Unlike Star schema, the dimensions table in a snowflake schema are normalized.

the item dimension table contains the attributes item_key, item_name, type, brand, and supplier-key.

The supplier key is linked to the supplier dimension table. The supplier dimension table contains the attributes supplier_key and supplier_type.

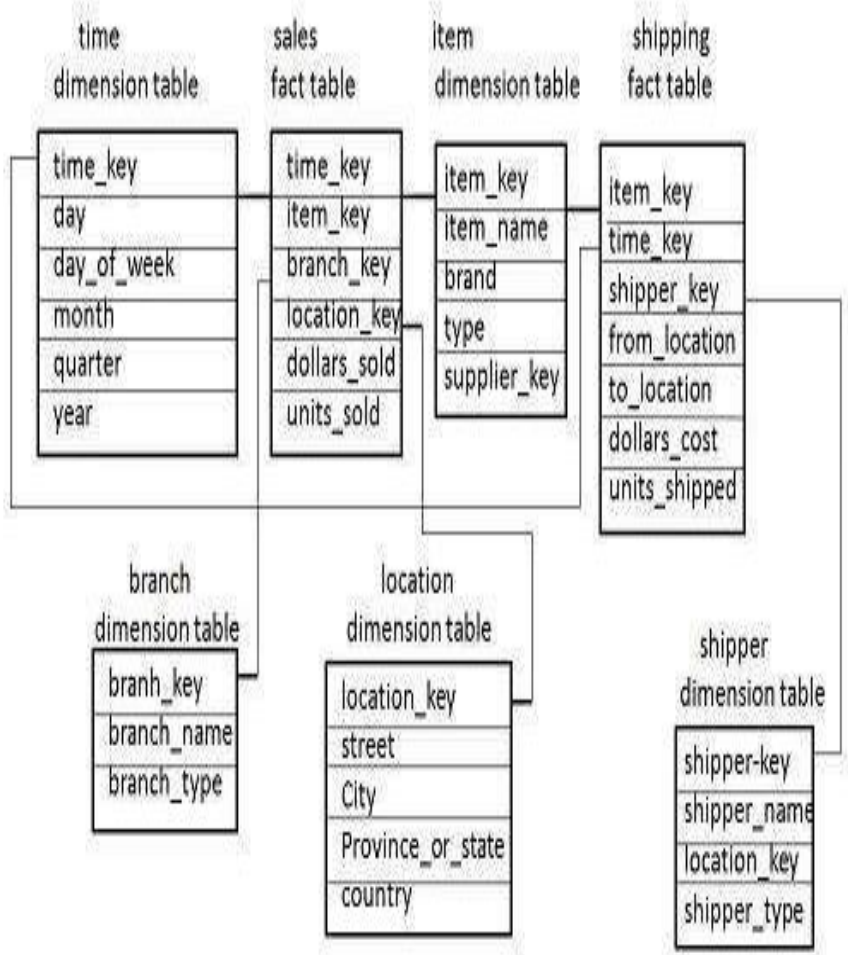


FACT CONSTELLATION SCHEMA

A fact constellation has multiple fact tables. It is also known as galaxy schema.

The following diagram shows two fact tables, namely sales and shipping.

The sales fact table is same as that in the star schema. The shipping fact table has the five dimensions, namely item_key, time_key, shipper_key, from_location, to_location.



ADVANTAGES OF PARTITIONING STRATEGY

MIDTERM WEEK 2

PARTITIONING STRATEGY

Partitioning is done to enhance performance and facilitate easy management of data. Partitioning also helps in balancing the various requirements of the system. It optimizes the hardware performance and simplifies the management of data warehouse by partitioning each fact table into multiple separate partitions.

This section contains the following topics:

- **Bigger Databases**
- **Bigger Individual Tables: More rows in Tables.**
- **More Users Querying the system**
- **More Complex Queries**

Bigger Databases

The ability to split a large database object into smaller pieces transparently simplifies efficient management of very large databases.

Bigger Individual Tables: More Rows in Tables

It takes longer to scan a big table than it takes to scan a small table. Queries against partitioned tables may access one or more partitions that are small in contrast to the total size of the table.

More Users Querying the System

With partitioning, users are more likely to query on isolated and smaller data sets.

More Complex Queries

If smaller data sets are being accessed, then complex calculations are more likely to be processed in memory, which is beneficial from a performance perspective and reduces the application's I/O requirements.

Partitioning for Performance in a Data Warehouse

Analyses run against the database should return within a reasonable amount of time, even if the queries access large amounts of data in tables that are terabytes in size.

Partition Pruning in a Data Warehouse

In partition pruning, the optimizer
analyzes FROM and WHERE clauses in SQL statements to eliminate unneeded partitions when building the partition access list.

Thank You

School of Nation Builders

NATIONAL COLLEGE OF SCIENCE & TECHNOLOGY

COMPUTER STUDIES DEPARTMENT

NET 002

DATA WAREHOUSE OLAP

TOPIC 1

PRELIM

Online Analytical Processing (OLAP)

is based on the multidimensional data model. It allows managers, and analysts to get an insight of the information through fast, consistent, and interactive access to information.

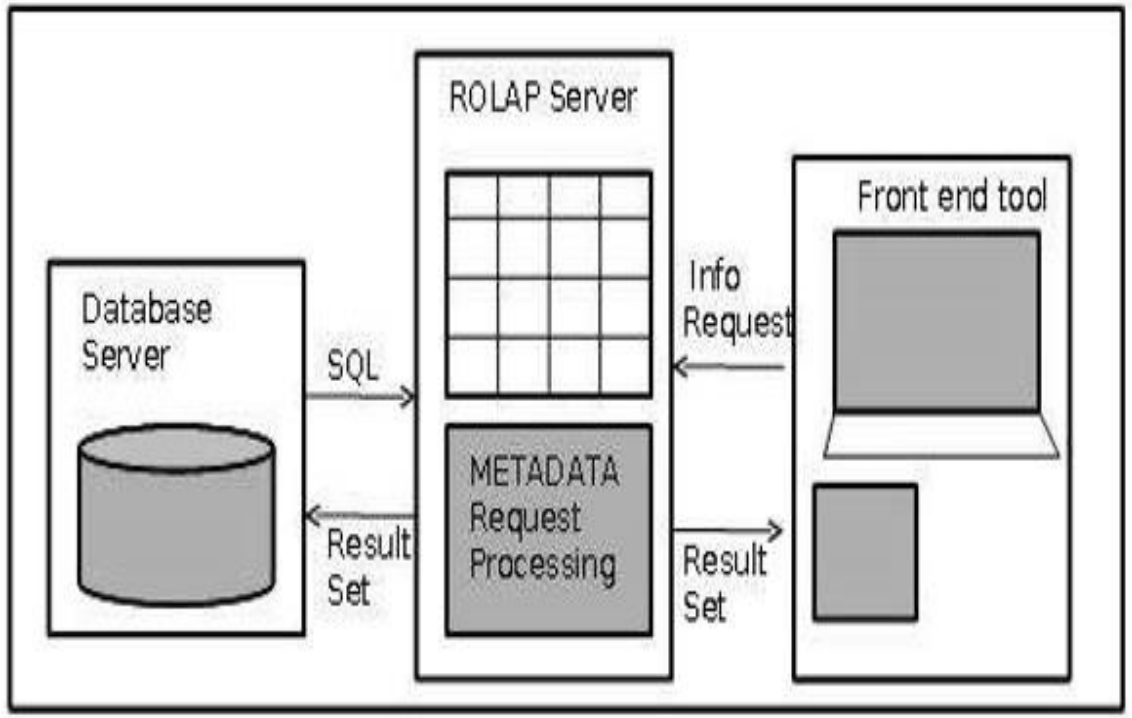
Types of OLAP Servers

We have four types of OLAP servers –

- **Relational OLAP (ROLAP)**
- **Multidimensional OLAP (MOLAP)**
- **Hybrid OLAP (HOLAP)**
- **Specialized SQL Servers**

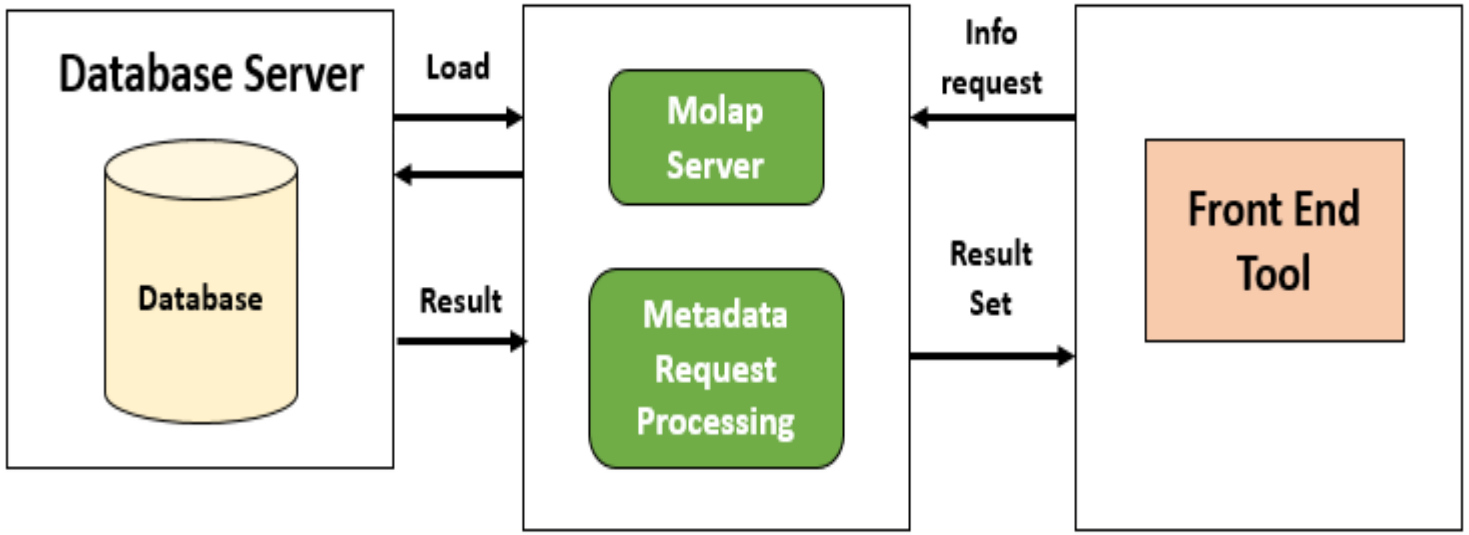
RELATIONAL OLAP

- ROLAP servers are placed between relational back-end server and client front-end tools.
- ROLAP includes the following –
  - Implementation of aggregation navigation logic.
  - Optimization for each DBMS back end.
  - Additional tools and services.



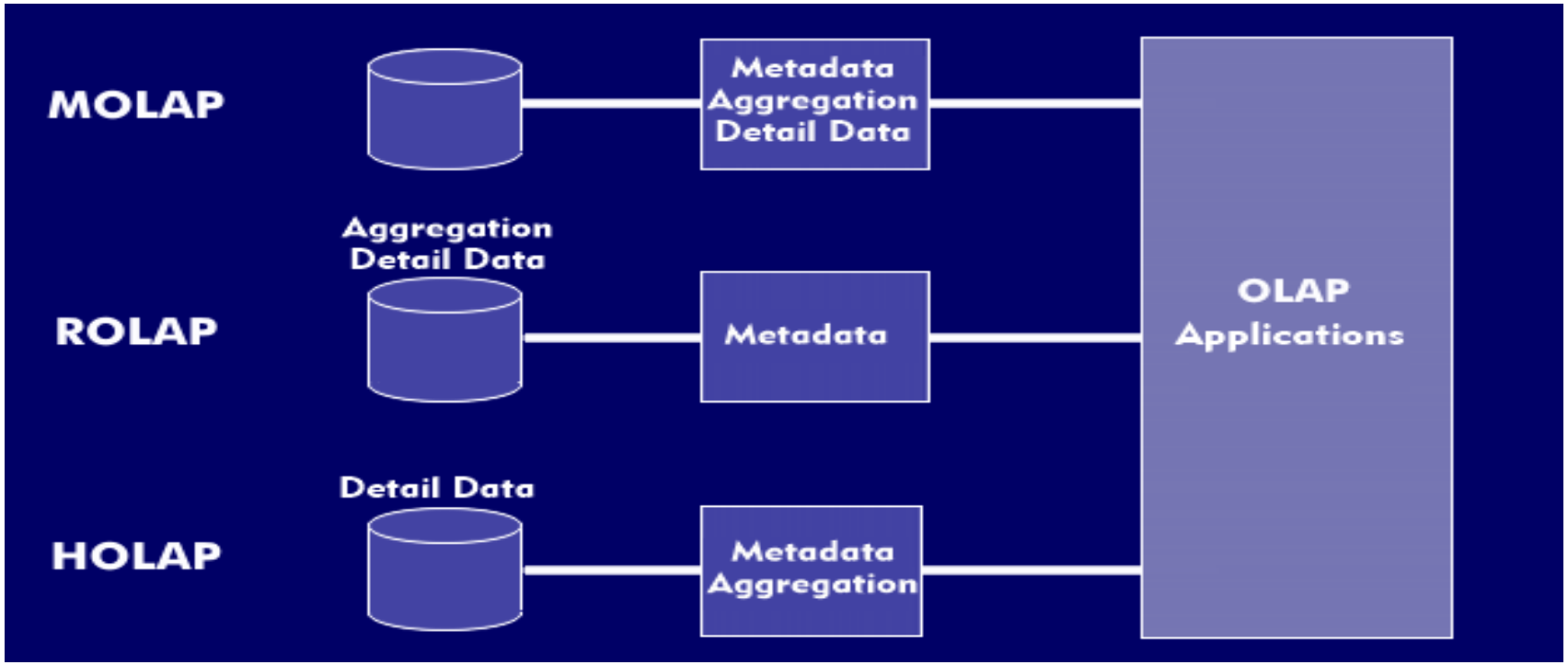
MULTIDIMENTIONAL OLAP

uses array-based multidimensional storage engines for multidimensional views of data. With multidimensional data stores, the storage utilization may be low if the data set is sparse. Therefore, many MOLAP server use two levels of data storage representation to handle dense and sparse data sets.

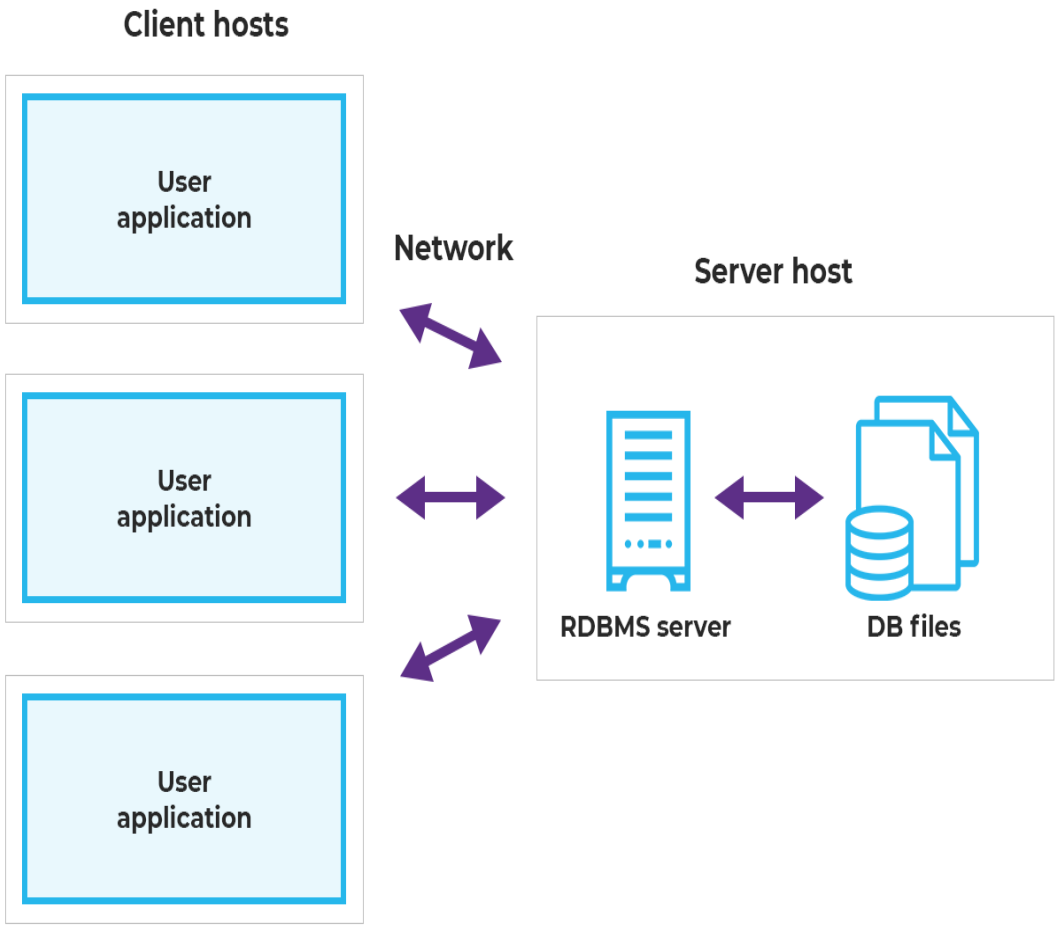


HYBRID OLAP

is a combination of both ROLAP and MOLAP. It offers higher scalability of ROLAP and faster computation of MOLAP. HOLAP servers allows to store the large data volumes of detailed information. The aggregations are stored separately in MOLAP store.



Specialized SQL Servers - Specialized SQL servers provide advanced query language and query processing support for SQL queries over star and snowflake schemas in a read-only environment.



OLAP Operations

OLAP servers are based on multidimensional view of data, we will discuss OLAP operations in multidimensional data.

List of OLAP operations –

- Roll-up
- Drill-down
- Slice and dice
- Pivot (rotate)

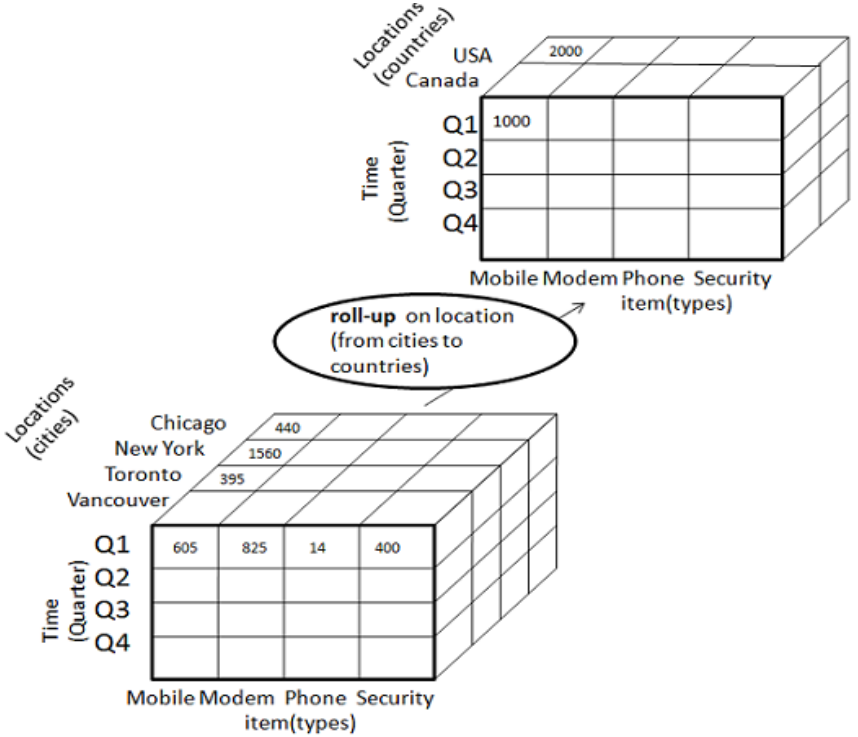
ROLL-UP

This operation aggregation on a data cube in any of the following way:

- **By climbing up a concept hierarchy for a dimension**
- **By dimension reduction**

The following diagram illustrates how roll-up works.

- **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.**



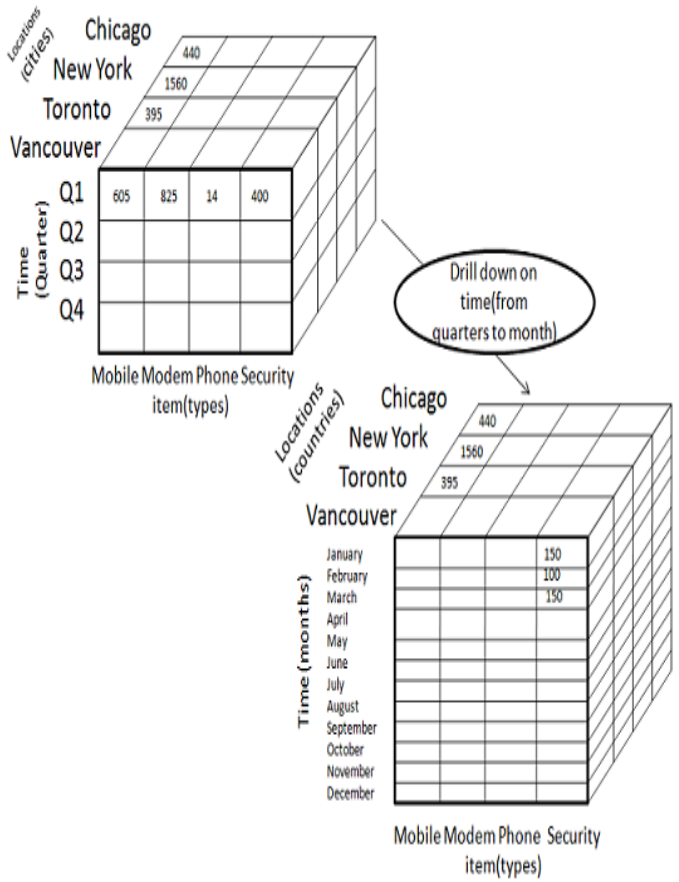
DRILL-DOWN

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.**

The following diagram illustrates how drill-down works –

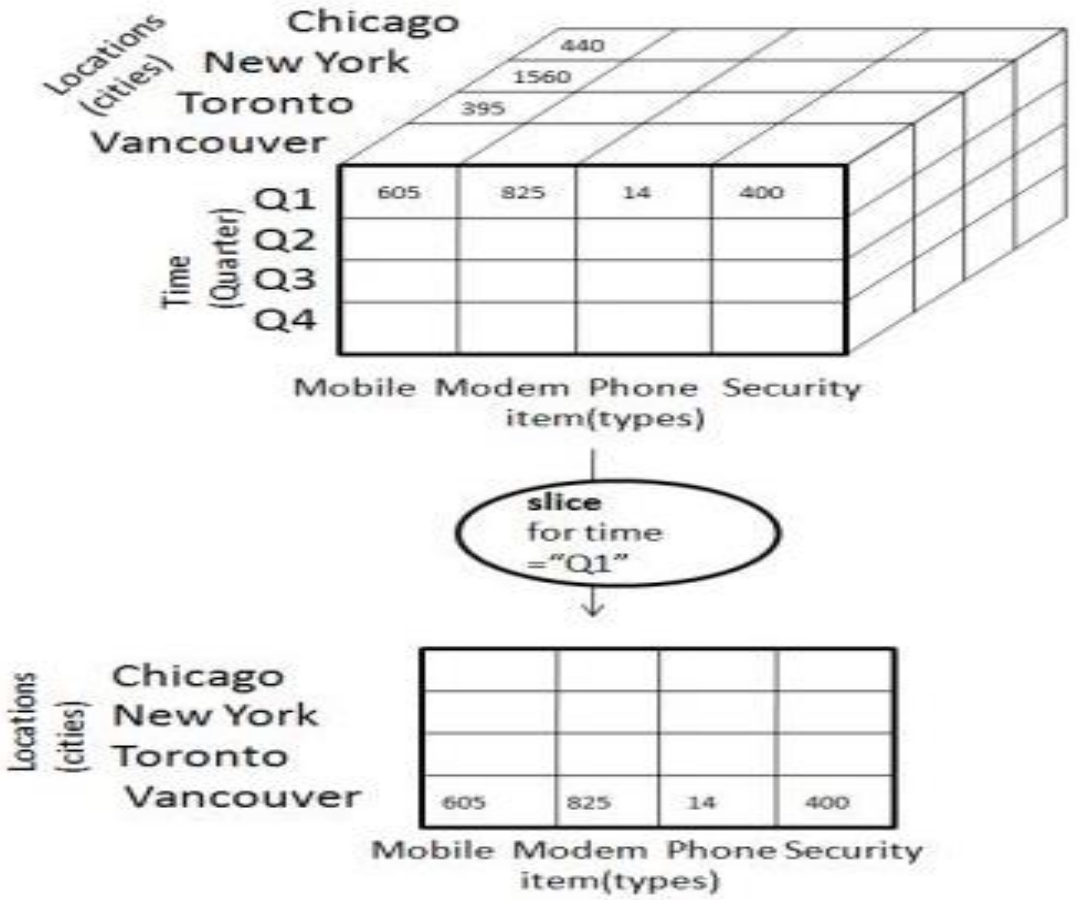
- **Drill-down is performed by stepping down a concept hierarchy for the dimension time.**
- **Initially the concept hierarchy was "day < month < quarter < year."**
- **On drilling down, the time dimension is descended from the level of quarter to the level of month.**
- **When drill-down is performed,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



Slice

The slice operation selects one particular dimension from a given cube and provides a new sub-cube. Consider the following diagram that shows how slice works.

- **Here Slice is performed for the dimension "time" using the criterion time = "Q1".**
- **It will form a new sub-cube by selecting one or more dimensions.**

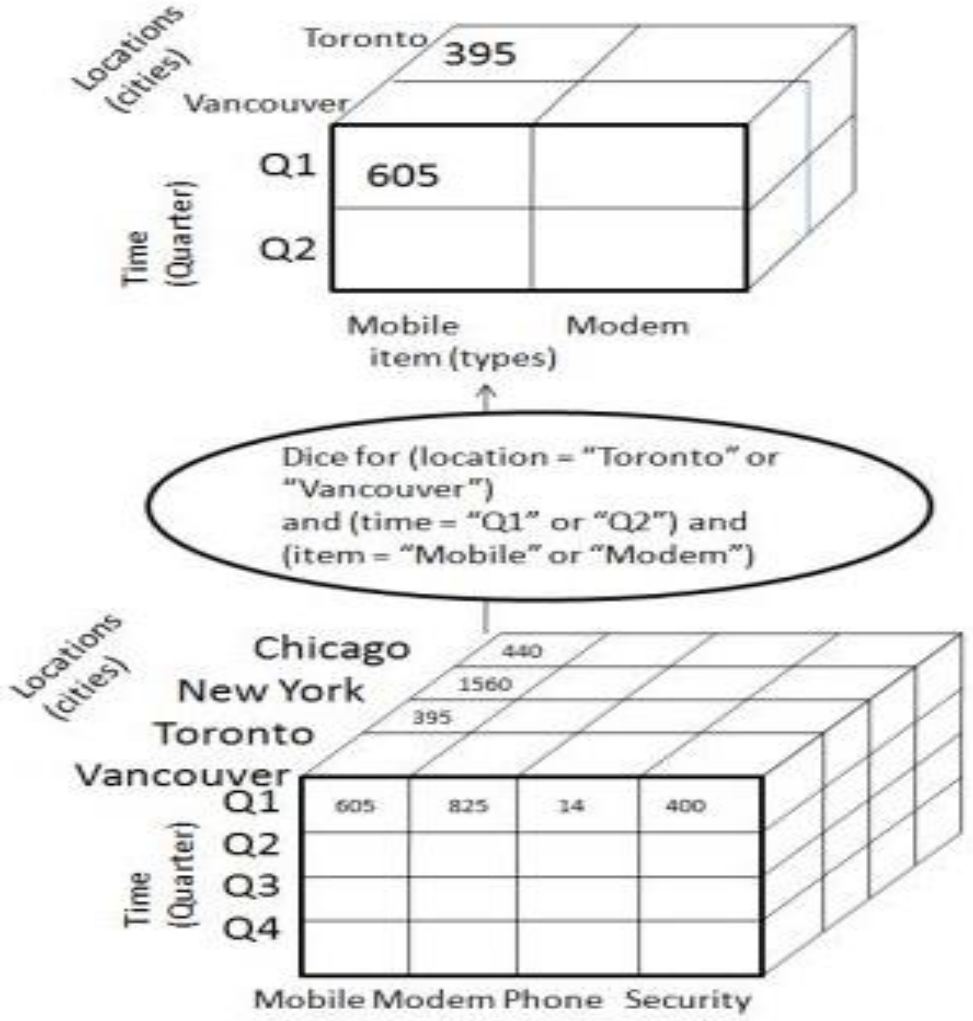


Dice

Dice selects two or more dimensions from a given cube and provides a new sub-cube. Consider the following diagram that shows the dice operation

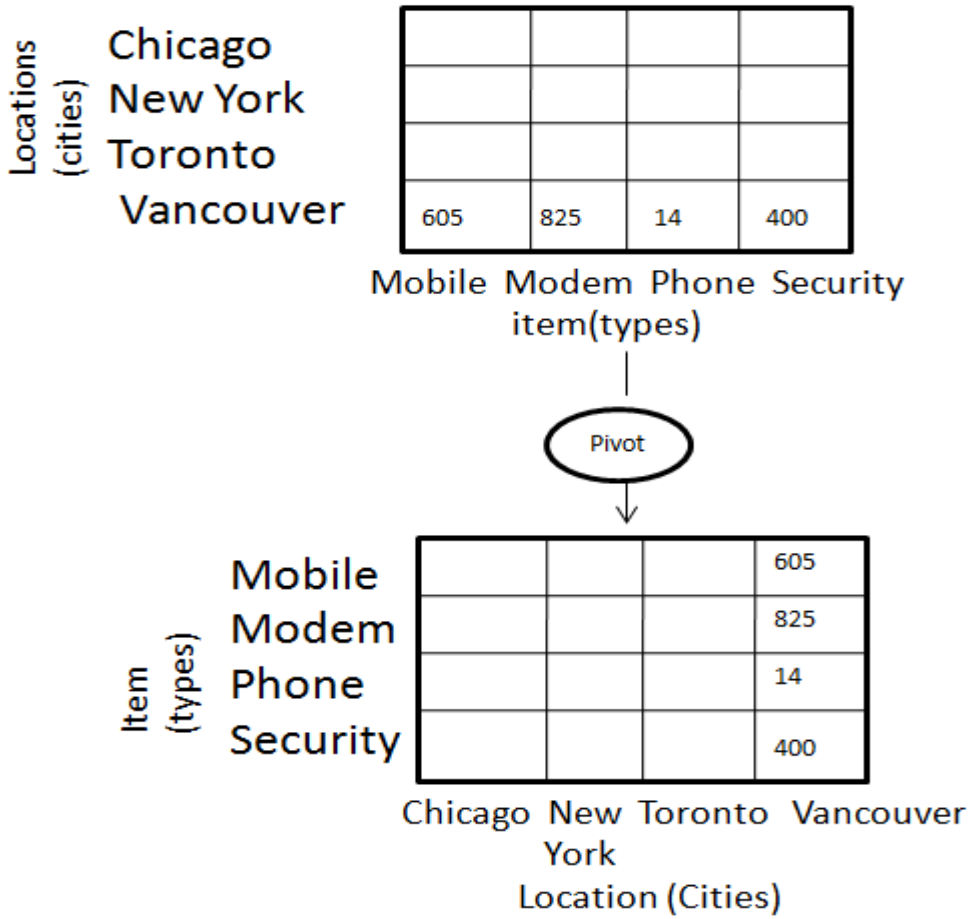
The dice operation on the cube based on the following selection criteria involves three dimensions.

- **(location = "Toronto" or "Vancouver")**
- **(time = "Q1" or "Q2")**
- **(item =" Mobile" or "Modem")**
-



Pivot

The pivot operation is also known as rotation. It rotates the data axes in view in order to provide an alternative presentation of data. Consider the following diagram that shows the pivot operation.



Thank You