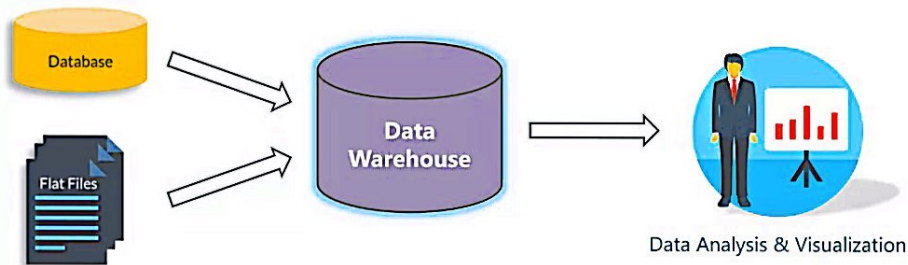


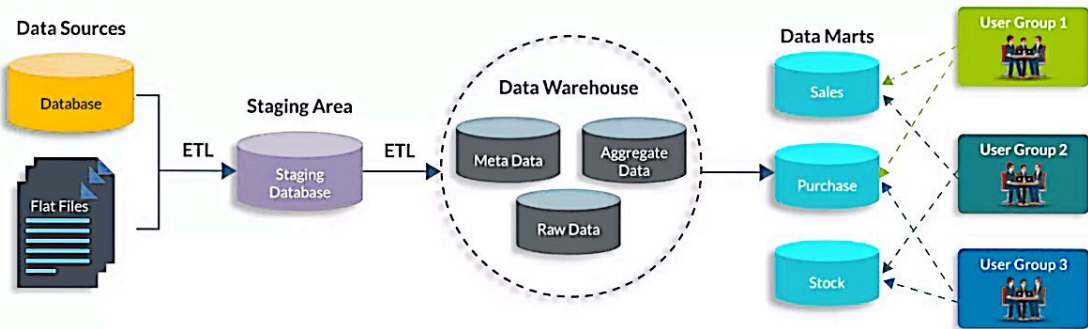
# What Is A Data Warehouse?

- Data Warehouse is like a relational database designed for **analytical needs**.
- It functions on the basis of **OLAP** (Online Analytical Processing).
- It is a central location where consolidated data from multiple locations (databases) are stored.



# What Is Data Warehousing?

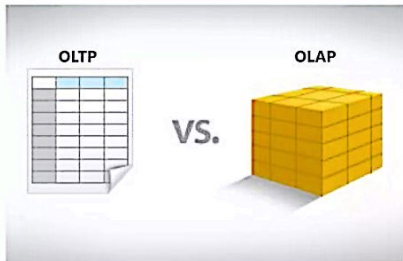
- Data Warehousing is the act of **organizing** & **storing** data in a way so as to make its retrieval efficient and insightful.
- It's also called as the process of transforming **data** into **information**.



Scanned with CamScanner

# OLAP (Online Analytical Processing)

- OLAP is a flexible way for you to make complicated analysis of **multidimensional data**.
- DWH is modeled on the concept of **OLAP**. DBs are modeled on the concept of OLTP (Online Transaction Processing).
- OLTP systems use data stored in the form of two-dimensional tables, with rows and columns.



## Advantages Of OLAP Over OLTP

1. Opens up new views of looking at data.
2. Supports filtering/ sorting of data.
3. Data can be refined.

# Types Of OLAP Cubes

## 1 MOLAP

**MOLAP** is a form of OLAP that processes and stores the data directly into a multidimensional database.

**Advantage:-** Excellent performance; Can perform complex calculations.

**Disadvantage:-** Only limited data can be handled.

## 2 ROLAP

**ROLAP** is a form of OLAP that performs **dynamic multidimensional analysis** of data stored in a relational database rather than in a multidimensional database.

**Advantage:-** Greater amount of data can be processed.

**Disadvantage:-** Requires more processing time/ disk space.

## 3 HOLAP

**HOLAP** (Hybrid OLAP) is a combination of the advantages of MOLAP and ROLAP.

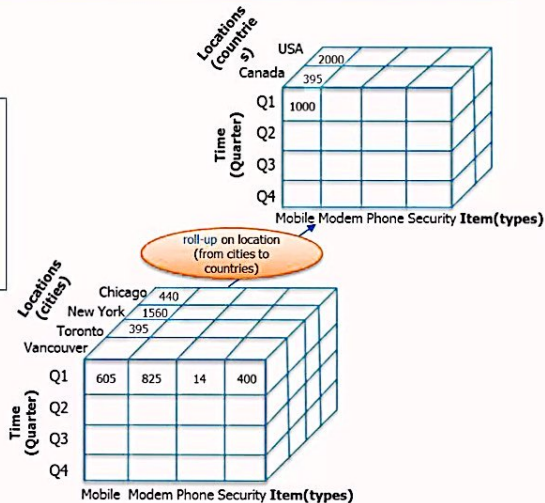
**Advantages:** HOLAP can "drill through" from the cube into underlying relational data.

# OLAP Operations:- Roll-up

Roll-up performs aggregation on a data cube by either:

1. Climbing up a concept hierarchy for a dimension
2. Dimension reduction

The following diagram illustrates how roll-up works.



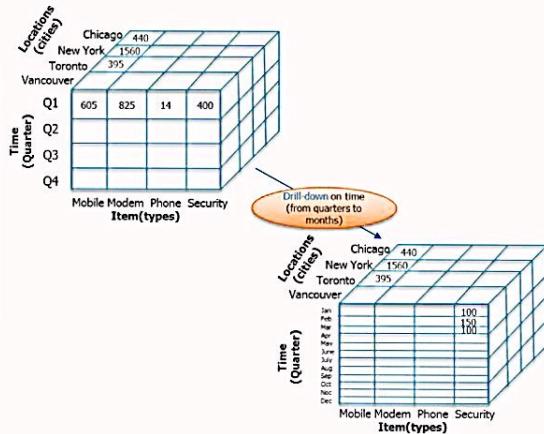
# OLAP Operations:- Drill-down

Drill-down is the reverse operation of roll-up.

It is performed by either:

1. Stepping down a concept hierarchy for a dimension
2. Introducing a new dimension.

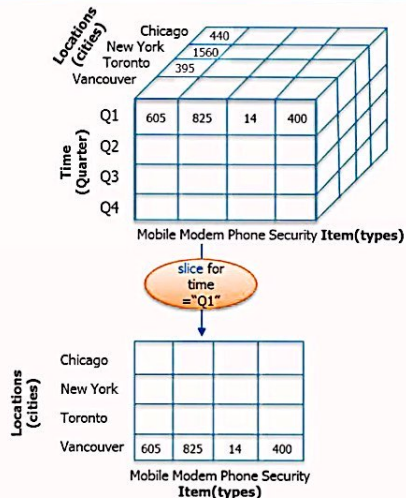
The following diagram illustrates how drill-down works.



# OLAP Operations:- Slice

The slice operation provides a new sub-cube from one particular dimension in a given cube.

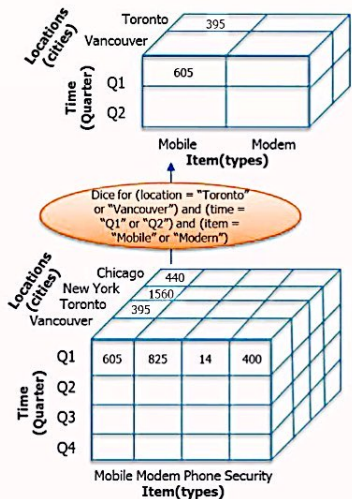
Consider the following diagram that shows how slice works.



# OLAP Operations:- Dice

The Dice operation provides a new sub-cube from two or more dimensions in a given cube.

Consider the following diagram that shows the dice operation.

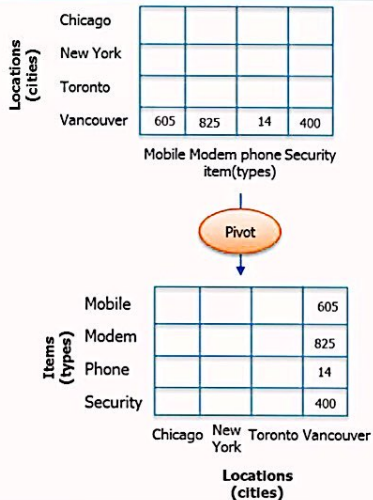




# OLAP Operations:- Pivot

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

Consider the following diagram that shows the pivot operation.



# Dimensions

- The tables that describe the dimensions involved are called **Dimension tables**.
- Dividing a Data Warehouse project into dimensions provides structured information for analysis & reporting.

E-commerce Company								
Customer			Product			Date		
ID	Name	Address	ID	Name	Type	Order date	Shipment date	Delivery date

← Subject

← Dimensions

← Attributes

Scanned with CamScanner

# Dimensions

- End users fire queries on these dimension tables which contain descriptive information.

E-commerce Company								
Customer			Product			Date		
ID	Name	Address	ID	Name	Type	Order date	Shipment date	Delivery date
1	Rita	ABC	001	CD	1A	1/06/14	3/06/14	5/06/14
2	John	XYZ	002	AC	2B	6/06/14	9/06/14	11/06/14
3	Paul	PQR	003	TV	3C	10/06/14	14/06/14	16/06/14

Query  
Result



# Facts & Measures

- A fact is a measure that can be summed, averaged or manipulated.
- A Fact table contains 2 kinds of data – a **dimension key** and a **measure**.
- Every Dimension table is linked to a Fact table.

## Fact Table



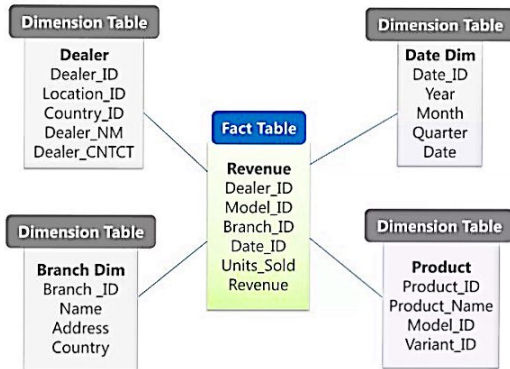
# Schemas

- A schema gives the logical description of the entire data base.
- It gives details about the constraints placed on the tables, key values present & how the key values are linked between the different tables.
- A database uses relational model, while a data warehouse uses **Star**, **Snowflake** and **Fact Constellation** schema.

Employee					Linked	Department	
ID	First Name	Last Name	Age	Dept_ID		Dept_ID	Dept_Name
1234	Rita	Joe	25	0674		0674	Sales
4321	John	Smith	35	0825		0752	HR
5678	Paul	Brady	45	0752		0825	Production
7890	Rose	Michael	65	0825			

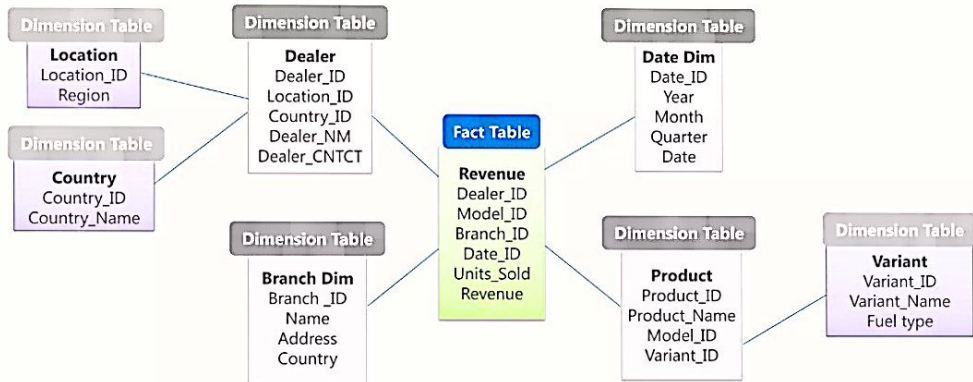
# Types Of Schemas:- Star Schema

- Each dimension in a star schema is represented with a **one-dimension table** which contains a set of attributes.
- **Fact table** is at the center, which contains keys to every dimension table & attributes like: *units sold* and *revenue*.



# Types Of Schemas:- Snowflake Schema

- Dimension tables in the **Snowflake schema** are **normalized**. (Split into additional tables).
- **Dealer** dimension table is split into **Location** & **Country**. Product dimension table is split into **Product** & **Variant**.



# Types Of Schemas:- Galaxy Schema

- Also known as **Fact Constellation** schema. Contains more than **1 Fact table**.
- Below, there are two fact tables: **Revenue** and **Product**.
- Dimensions which are shared are called **Conformed Dimensions**.

