DATA WAREHOUSING AND MINING

T.E. CSE-DS, Sem V
Academic Year: 2023-24

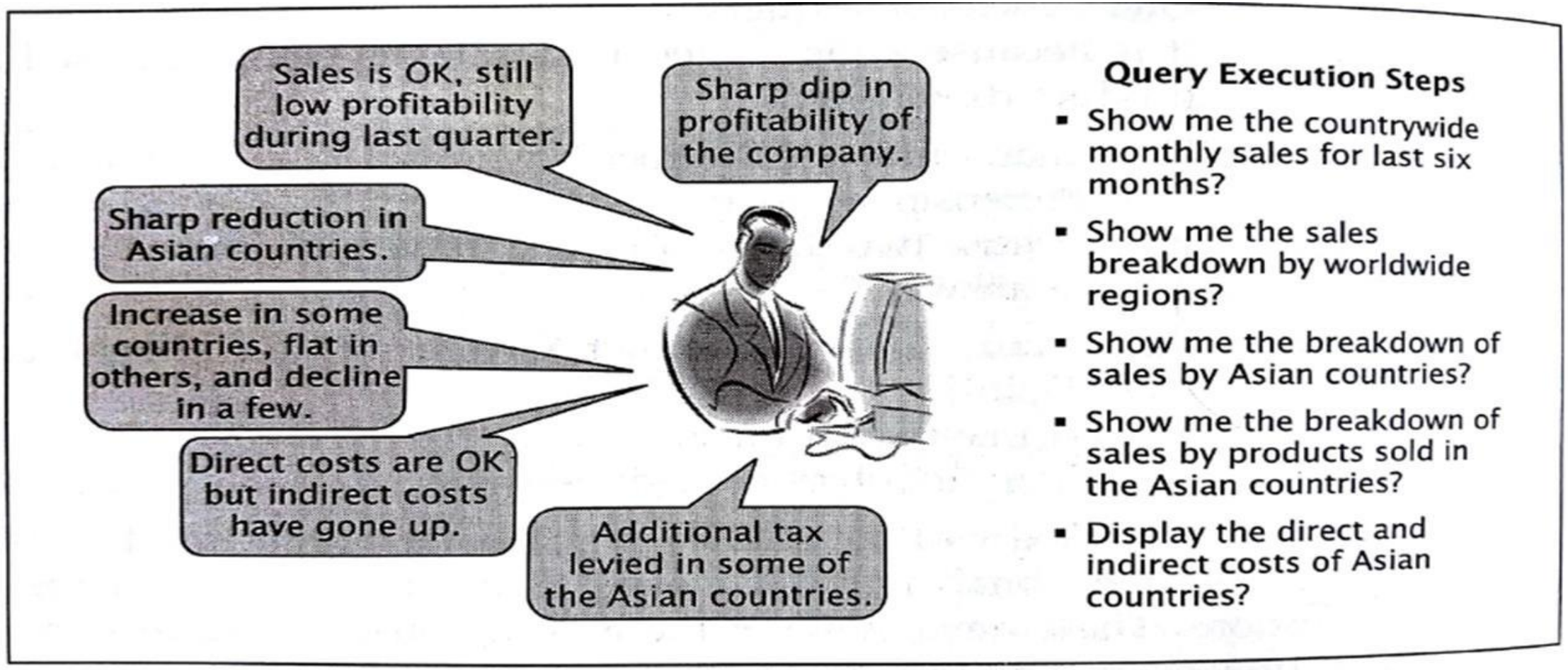
Data Warehousing Fundamentals: OLAP, Multidimensional Analysis

Online Analytical Processing (OLAP)

- Data warehouse is designed to perform strategic analysis using the data that is stored in it.
- The data must be stored and presented in a way to facilitate analysis of key indicators over time along business dimensions.
- Any data structure designed using the dimensional modelling technique fits the framework of carrying out such analysis.

OLAP is a category of s/w, tool or technology that enables DW users to gain an insight into the data through fast, consistent interactive access to provide superior performance in executing business queries that perform complex strategic analysis

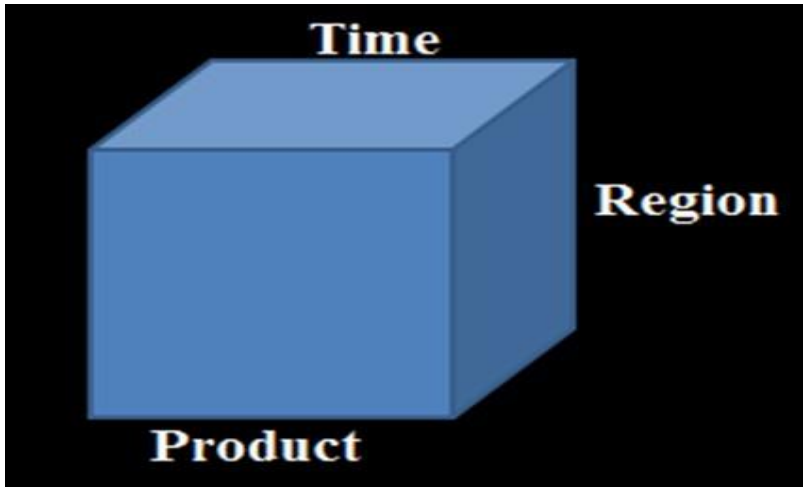
Typical Analysis Session



A typical analysis session

OLAP – Multidimensional Analysis

Online Analytical Processing (OLAP)



One dimensional query:

How many units of product_x were sold?

Two dimensional query:

How many units of product_x were sold on 17 Jan 2018?

Three dimensional query:

How many units of product_x were sold on 17 Jan 2018 in new delhi outlet?

Multidimensional Analysis

- Multidimensional analysis means analysing the facts across multiple dimensions
- DW end users are no longer satisfied with one and two dimensional queries and simple analysis
- The user needs an environment that offers a multidimensional views of data
- The users must be able to analyse the data and finally be able to see/view the results in a Varity of formats and different perspectives
- Irrespective of the query (whether it is simple or complex/whether it is single dimensional or multidimensional) the amount of time to receive the result set must be consistent.

OLAP CUBE

At the core of the OLAP concept, is an OLAP Cube.

The OLAP cube is a data structure optimized for very quick data analysis.

OLAP vs OLTP

Features	OLTP Systems	OLAP System
Characteristic	Operational processing	Informational processing
Orientation	Transaction	Analysis
User	Clark, DBA, Programmers and Database professionals	Business analysts, knowledge workers, decision makers
Focus	Data in	Information out
DB design	ER model based and application oriented	Star, snowflake and subject oriented
Age of data	Current Guaranteed and up to date data	Historical and consolidated
Unit of work	Short and simple transaction	Complex query

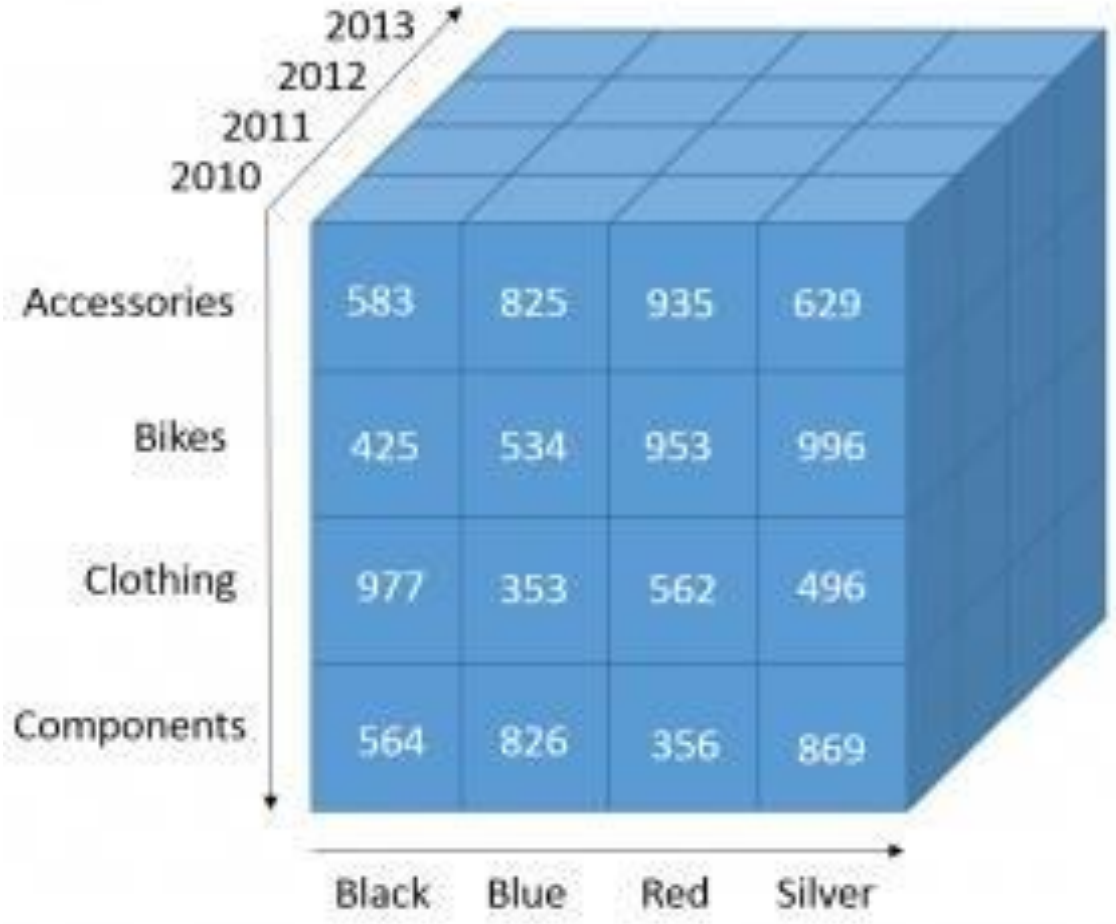
OLAP vs OLTP

Features	OLTP Systems	OLAP System
Access	Read, write, update	Mostly read Periodic data insertion Can not delete
Priority	High performance High availability	High flexibility End user autonomy
Usage	Predefined and repetitive	Not predefined Adhoc
Response time	Fast	Moderate
DB size	100 mb to few GB	>100 GB,TB or PB

Characteristics of OLAP

- It allows the users to have multidimensional and logical view of the data in the data warehouse.
- It provides the ability for interactive query and complex analysis
- Enables data warehouse users to perform drill down and roll up operation
- Displays results in a variety of formats including charts and graphs.

OLAP Cube



summarized sales data by product, time-period and color.

OLAP should allow :

- ✓Display the total sales in every individual product of Red color for the May 2010.
- ✓Show the average sales of individual color for clothing products for every year.

Star Schema with its corresponding data cube

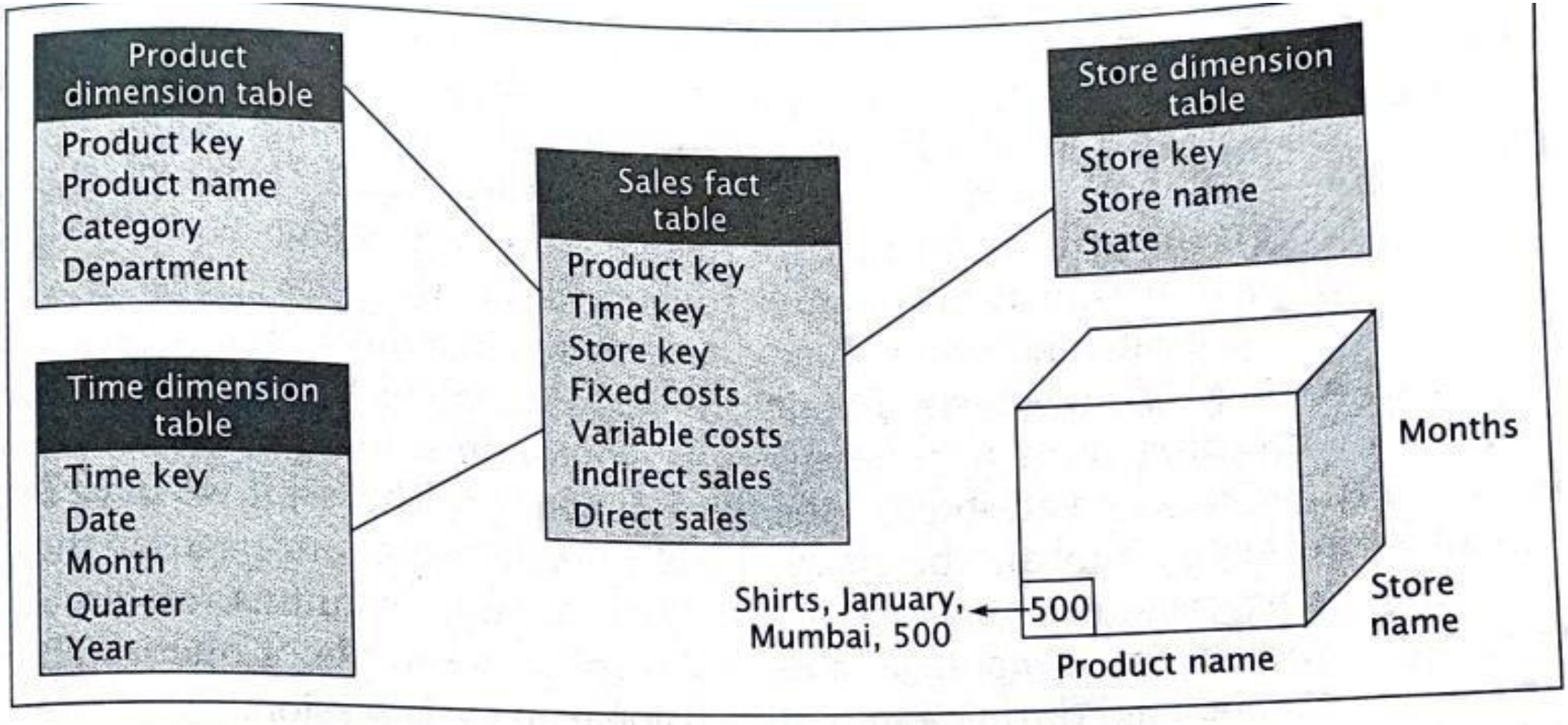


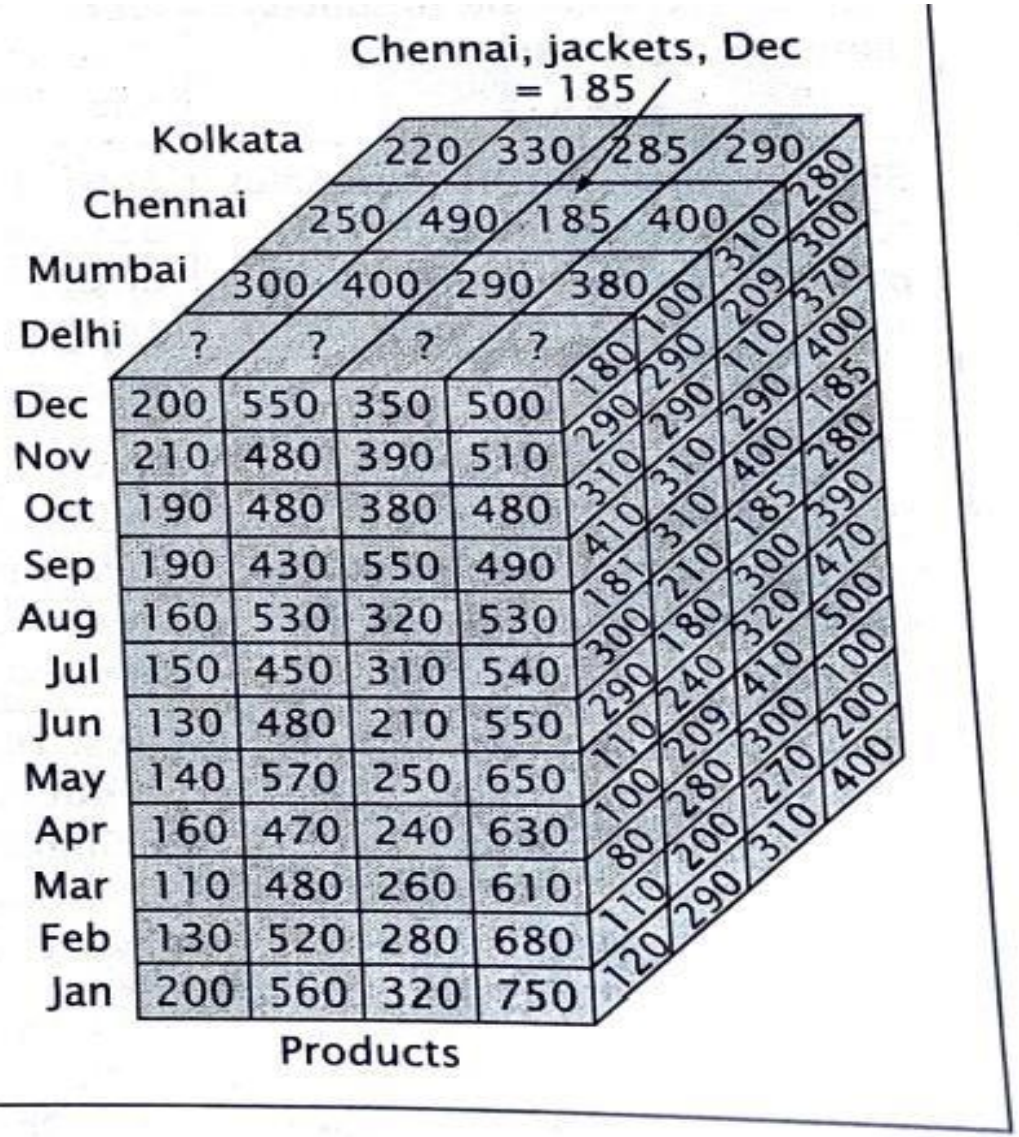
Table Storing data along three dimensions

Pages:
Store dimension, ex-Delhi

Columns:
Product dimension

Rows: Time dimension

	Shirts	T- shirts	Jackets	Trousers
Jan	200	550	350	500
Feb	210	480	390	510
Mar	190	480	380	480
Apr	190	430	350	490
May	160	530	320	530
Jun	150	450	310	540
Jul	130	480	270	550
Aug	140	570	250	650
Sep	160	470	240	630
Oct	170	480	260	610
Nov	180	520	280	680
Dec	200	560	320	750



OLAP Operations in multidimensional Data Model

- In multidimensional model data is organized into multiple dimensions and each dimension contains multiple levels of abstraction.
- This organization of data provides users with flexibility to view data from different perspectives.

OLAP operations help in materializing these views

Roll-up

Drill-down

Slice-Dice

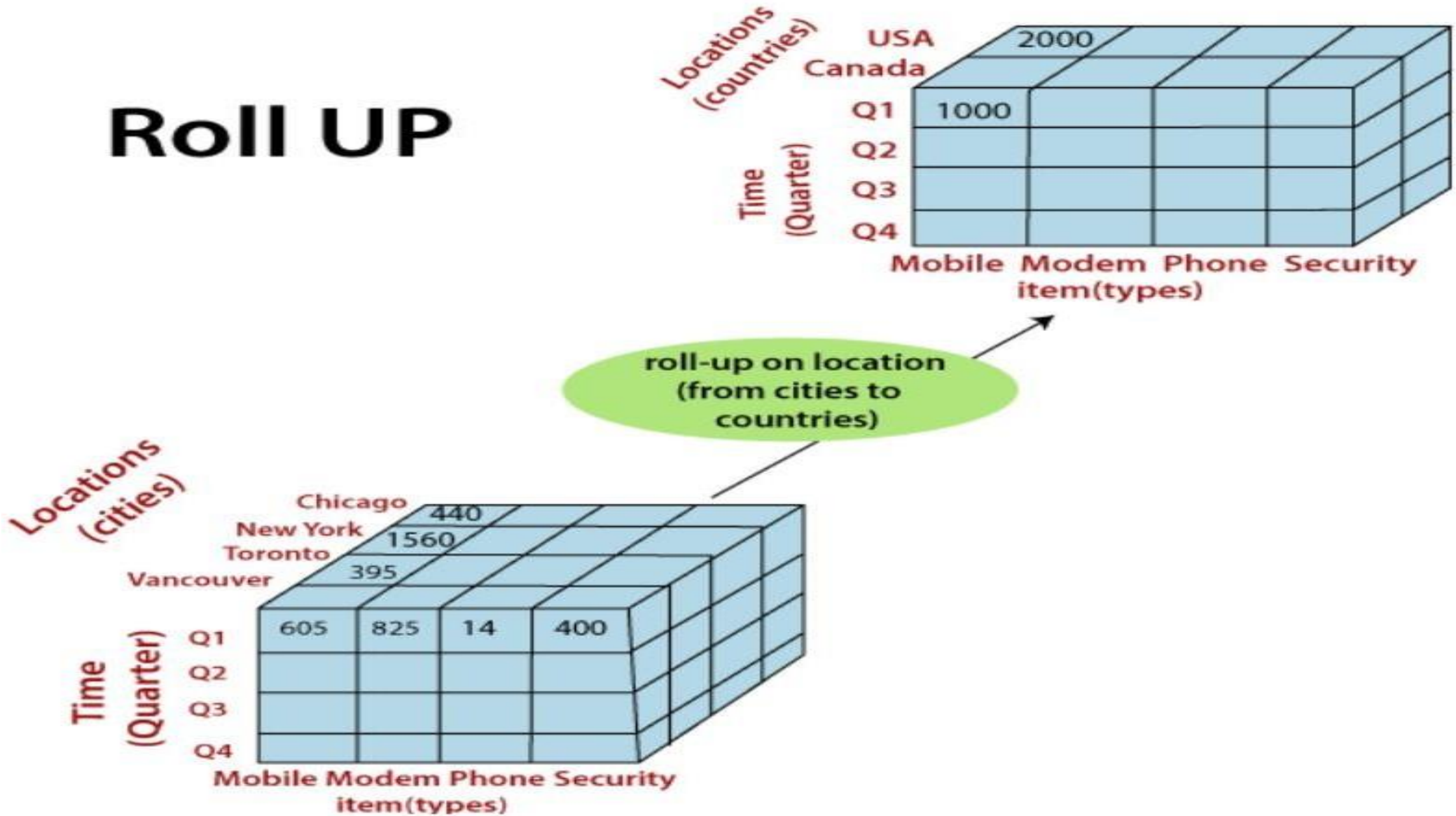
Pivot/ Rotate.

Roll -Up

Roll-up is also known as "consolidation" or "aggregation." The Roll-up operation can be performed in 2 ways

1. Reducing dimensions
2. Climbing up concept hierarchy. Concept hierarchy is a system of grouping things based on their order or level.

Roll UP


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

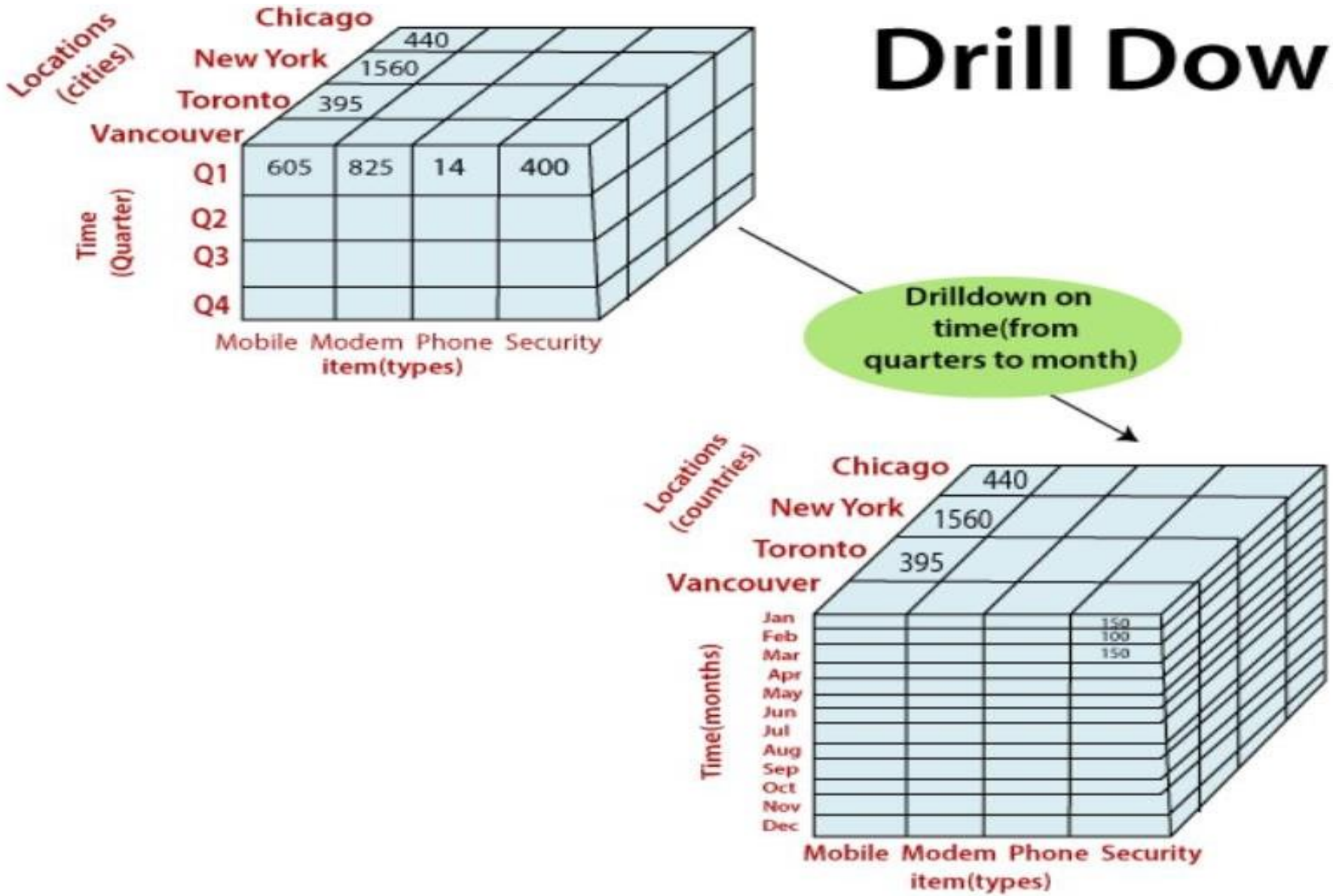
Roll Up

- In this example, cities New York and Chicago and rolled up into country USA
- The sales figure of New Jersey and Los Angeles are 440 and 1560 respectively. They become 2000 after roll-up
- In this aggregation process, data is location hierarchy moves up from city to the country.
- In the roll-up process at least one or more dimensions need to be removed. In this example, Cities dimension is removed.

Drill-down

- In drill-down data is fragmented into smaller parts. It is the opposite of the rollup process.
- It navigates from less detailed record to more detailed data.
- It can be done by Moving down the concept hierarchy
- Increasing a dimension

Drill Down

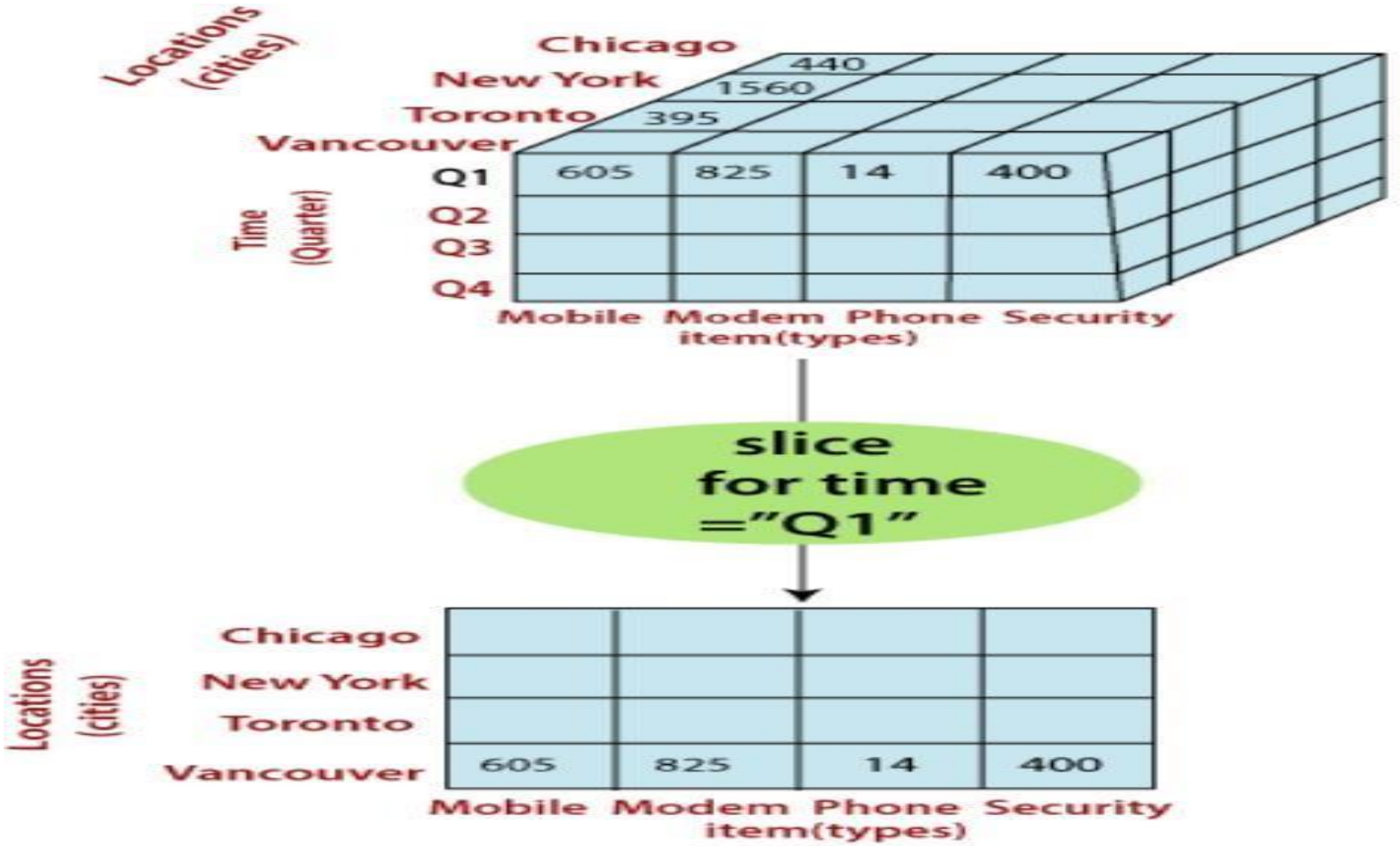

Locations (cities)
Chicago
New York
Toronto
Vancouver
440
1560
395
Time (Quarter)
Q1
Q2
Q3
Q4
605
825
14
400
Mobile Modem Phone Security
item(types)
Drilldown on time(from quarters to month)
Locations (countries)
Chicago
New York
Toronto
Vancouver
440
1560
395
Time(months)
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
150
100
150
Mobile Modem Phone Security
item(types)

Drill Down

- Quarter Q1 is drilled down to months January, February, and March. Corresponding sales are also registers.
- In this example, dimension months are added.

Slice

Here, one dimension is selected, and a new sub-cube is created.

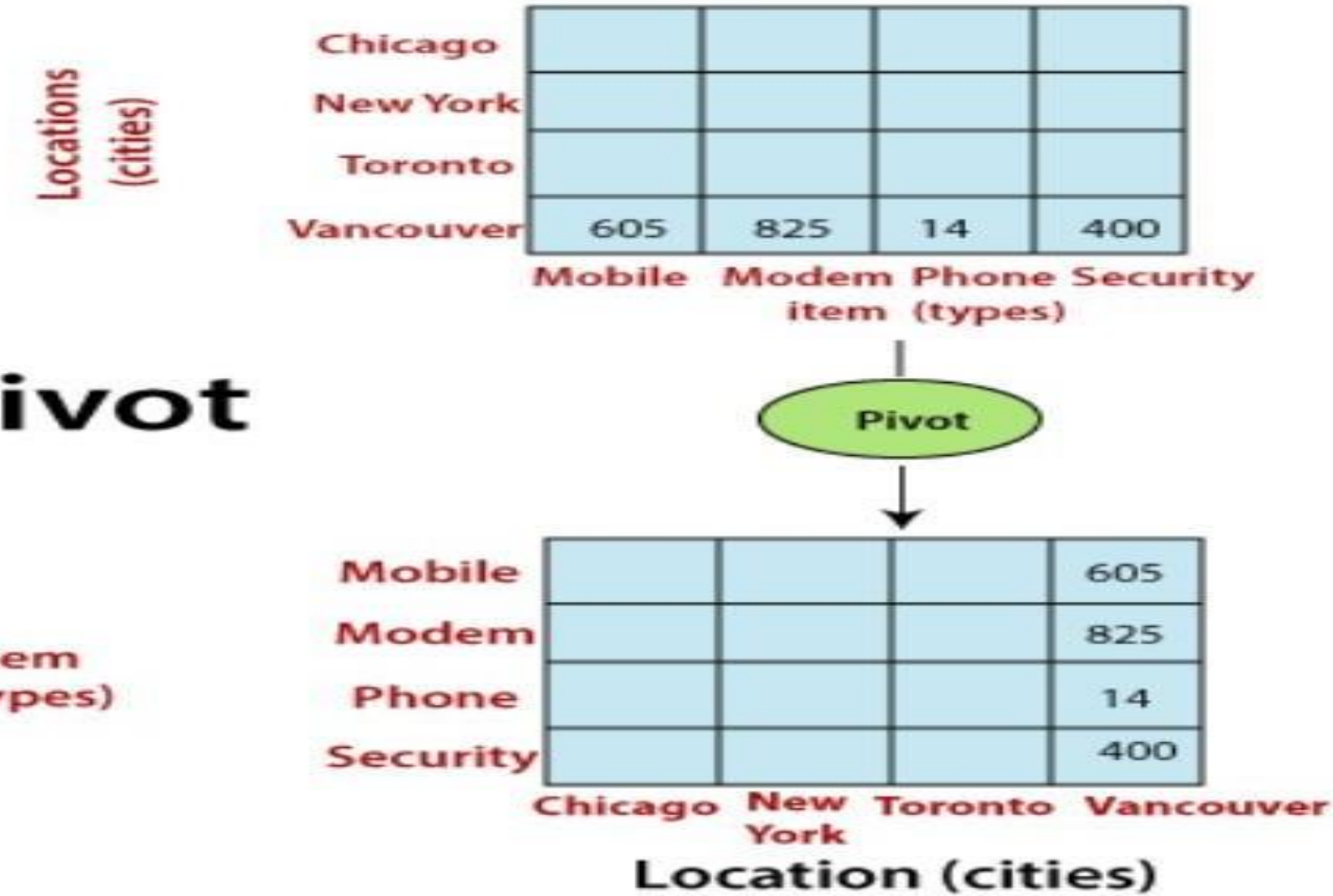


PIVOT

The pivot operation is also called a rotation.

Pivot is a visualization operations which rotates the data axes in view to provide an alternative presentation of the data.

Pivot


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

Sample Schema

Table customer

cid	cname	age	gender
cust1	Steve	20	M
cust2	Emma	25	F
cust3	Peter	23	M
cust4	Alice	22	F

Table item

itid	category	color
item1	Tshirt	blue
item2	jacket	blue
item3	Tshirt	red
item4	Jacket	red

Table store

sid	scity	scountry
store1	new york	USA
store2	chicago	USA
store3	delhi	India
store4	mumbai	India

sales_id	sid	cid	itid	price
1	store1	cust1	item1	30
2	store1	cust1	item2	35
3	store1	cust1	item3	25
4	store1	cust1	item4	40
5	store1	cust2	item4	40
6	store1	cust2	item3	25
7	store1	cust2	item2	35
8	store1	cust2	item1	30
9	store1	cust3	item1	30
10	store1	cust3	item2	35
11	store1	cust3	item3	25
12	store1	cust3	item4	40
13	store1	cust4	item4	40
14	store1	cust4	item3	25
15	store1	cust4	item2	35
16	store1	cust4	item1	30
17	store2	cust1	item1	30
18	store2	cust1	item2	35
19	store2	cust1	item3	25
20	store2	cust1	item4	40
21	store2	cust2	item4	40
22	store2	cust2	item3	25
23	store2	cust2	item2	35
24	store2	cust2	item1	30
25	store2	cust3	item1	30
26	store2	cust3	item2	35
27	store2	cust3	item3	25
28	store2	cust3	item4	40
29	store2	cust4	item4	40
30	store2	cust4	item3	25
31	store2	cust4	item2	35
32	store2	cust4	item1	30

Roll UP

Show total sale for each country

```
SELECT  scountry , category ,sum(price)
FROM sales_fact Sl, store s, item I
WHERE Sl.sid=s.sid and Sl.itid=i.itid
GROUP BY scountry ,category;
```

Drill Down

```
SELECT scountry , scity, gender, category ,sum(price)
FROM sales_fact Sl, store s, item i, customer c
WHERE Sl.sid=s.sid and Sl.itid=i.itid and Sl.cid=c.cid
GROUP BY scountry ,scity ,gender,category;
```

SLICE

Total sale for Mumbai store

```
SELECT Sl.sid, itid, sum(price)
FROM sales_fact Sl, store s
WHERE Sl.sid = s.sid and  scity='mumbai’
GROUP BY  Sl.sid, itid ;
```

DICE

Get total sale for Mumbai store for Item color RED

```
SELECT    Sl.sid, Sl.itid ,sum(price)
FROM      sales_fact Sl,  store s, item i
WHERE     Sl.sid = s.sid  and  Sl.itid=i.itid  and  scity='mumbai' and
           color='red’
GROUP BY Sl.sid, Sl.itid;
```

OLAP SERVERS

Relational OLAP (ROLAP)

Multidimensional OLAP (MOLAP)