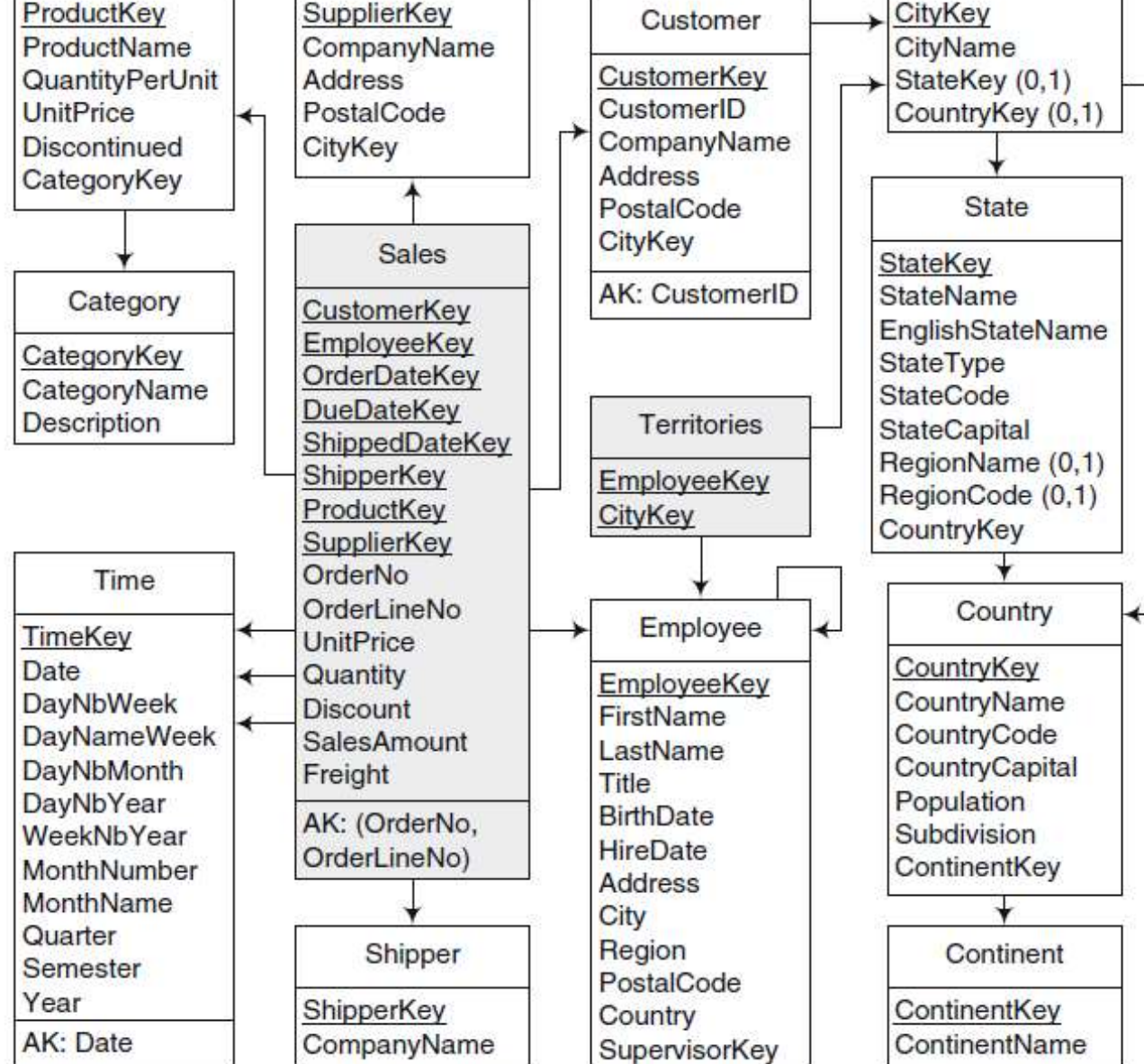
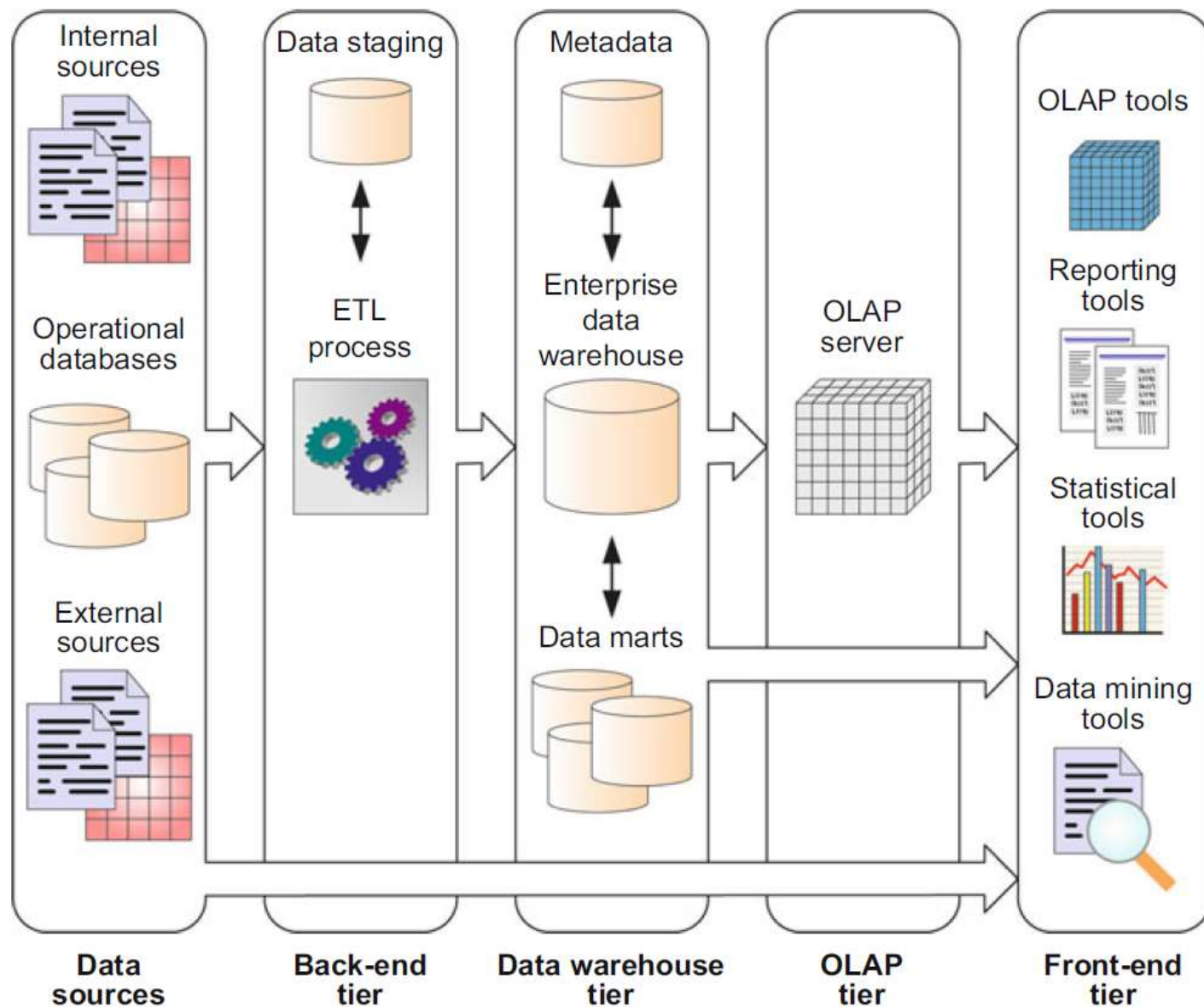

DS-306 Data Warehousing and Business Intelligence

Topic 6: OLAP and Its Operations

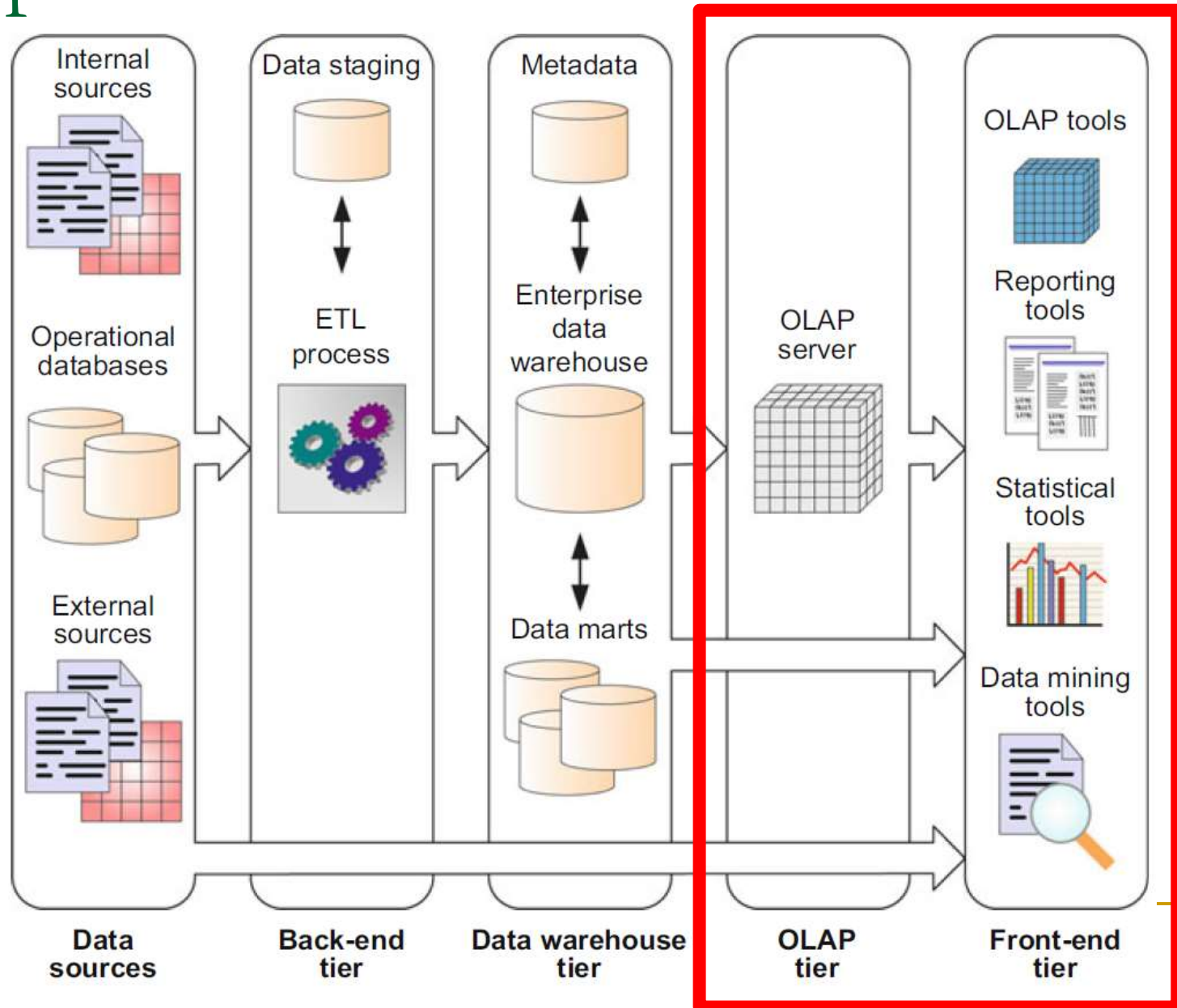
Dr. Khurram Shahzad



Typical DW architecture

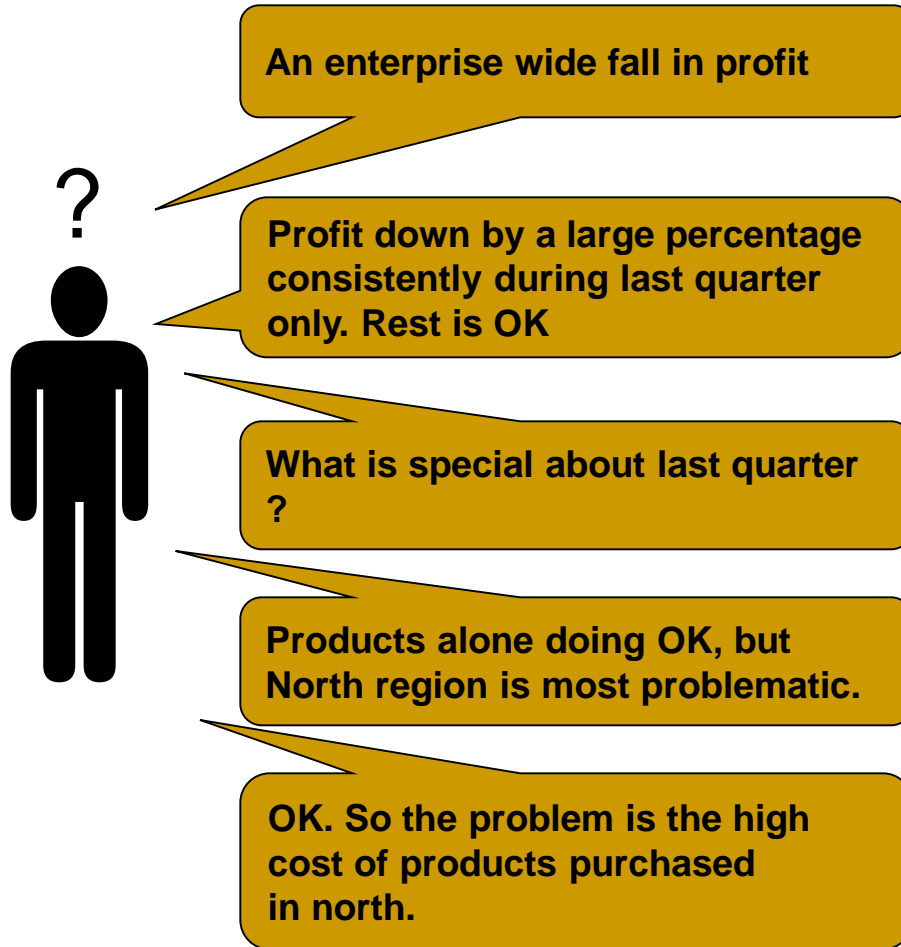


Typical DW architecture

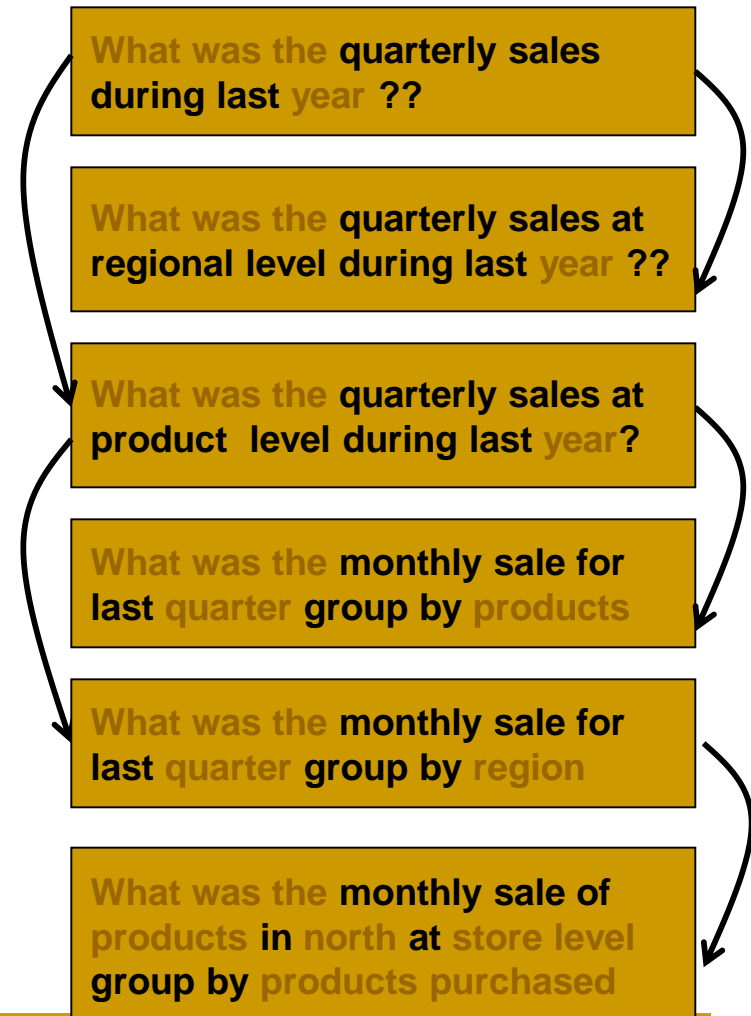


Supporting the human thought process

THOUGHT PROCESS



QUERY SEQUENCE



How many such query sequences can be programmed in advance?

Observations on Example

- Analysis is **Ad-hoc**
- Analysis is **interactive (User-Driven)**
- Analysis is **iterative**
 - Answer to one question leads to a dozen more
- Analysis is directional
 - Aggregating to lower level in dimensional hierarchy
 - Aggregating to higher level in dimensional hierarchy
 - Including multiple dimensions

Observations on Example

- Writing **pre-defined queries will not** work
 - Does not remain user-driven
 - Does not remain Ad-hoc and hence is not interactive
- **Enable Ad-hoc query** support
 - User can build his/her own queries (No SQL)
 - Do not afford on-the-fly computation
- **Problem**
 - Want to **Pre-compute the answers but do not know the questions**
- **Solution**
 - **Compute answers to all possible questions**
 - Remember, questions are always multidimensional aggregations at some level

OLAP Solution

- Online Analytical processing system
 - Online = no waiting for answer
- DW **stores** data and OLAP is **for analysis of data** and facilitates analysis
- OLAP: *“A category of applications and technologies for collecting, managing, processing and presenting **multidimensional data** for analysis and management purposes”*

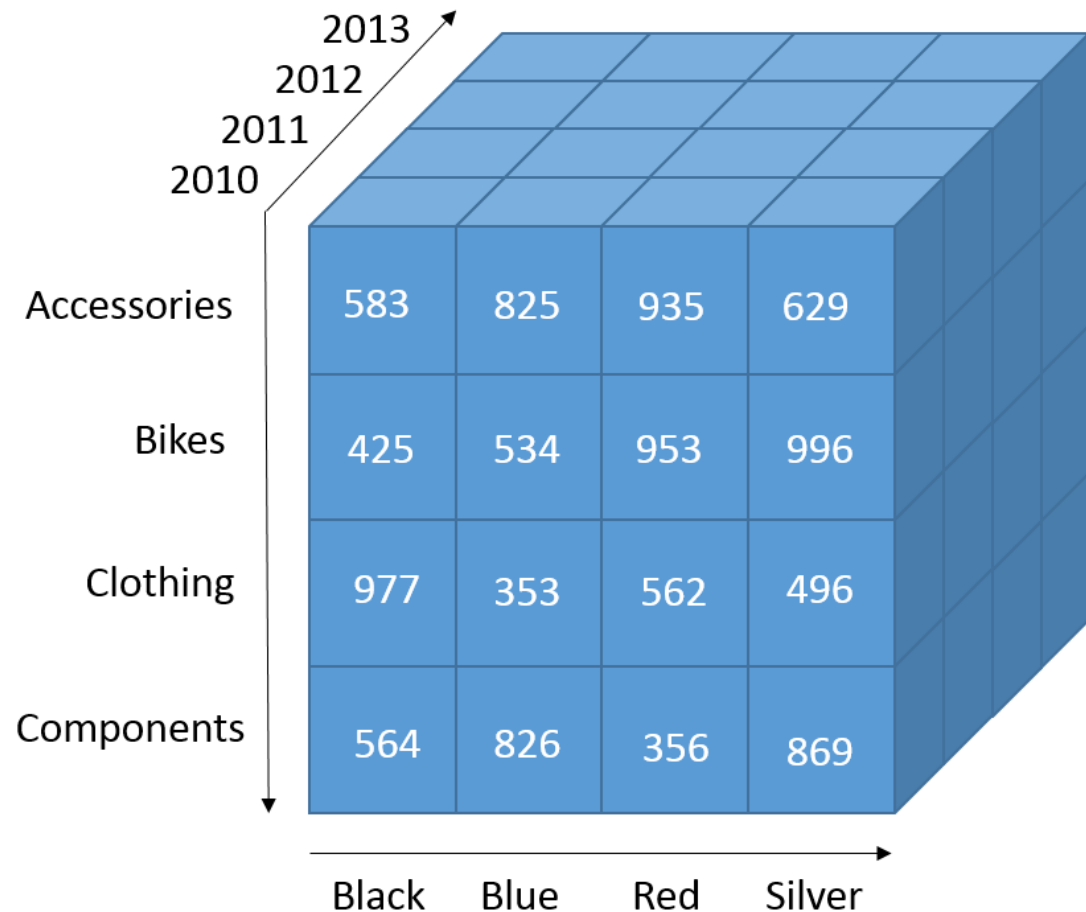
Two dimensions/Traditional Analysis

- Example of two dimensions can be represented by a table

Month	Sales	Direct Costs	Indirect Costs	Total Costs	Margin
January	790	480	110	590	200
February	850	520	130	650	200
March	900	530	140	670	230
April	910	590	150	740	170
May	860	600	120	720	140
June	830	490	100	590	240
July	880	500	110	610	270
August	900	620	130	750	150
September	790	300	90	390	400
October	820	540	100	640	180
November	840	570	150	720	120
December	810	600	120	720	90
Total	10,180	6,340	1,450	7,790	2,390

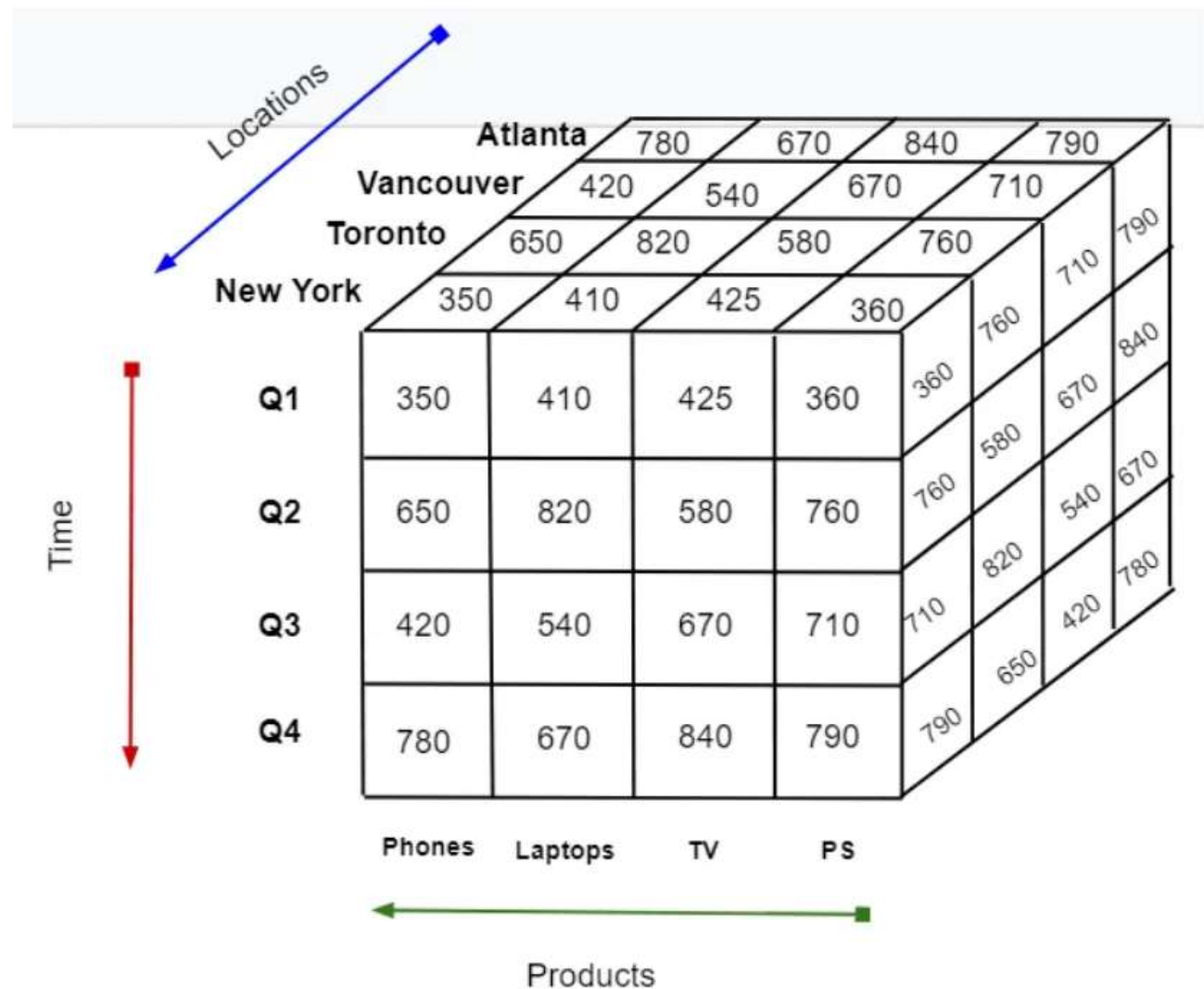
Three-dimensional Cube

- Three dimensions visualization



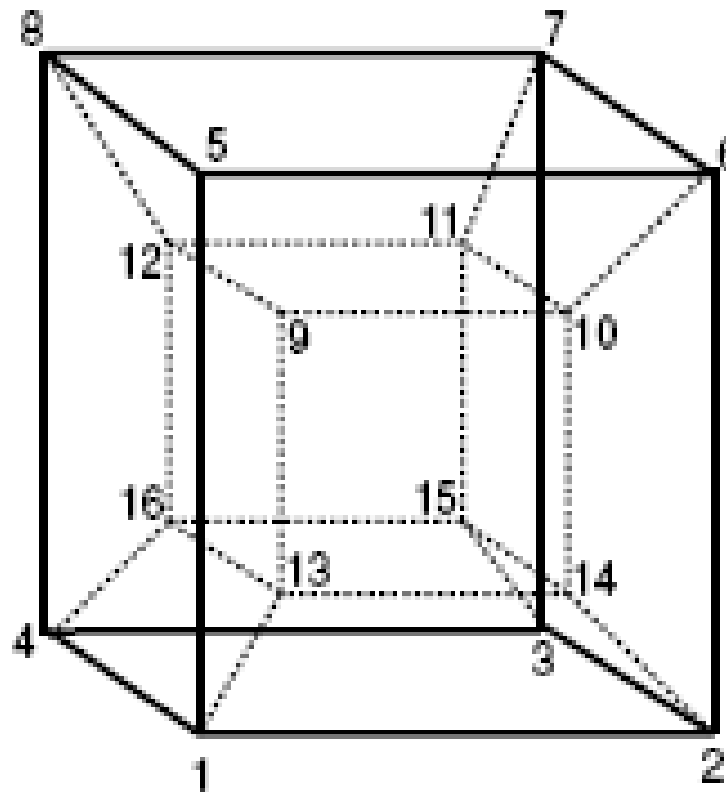
OLAP Intro.

■ Cube



Beyond three dimensions

- If we have more than three dimensions
- Hyper-cubes, is a cube with more than three dimensions



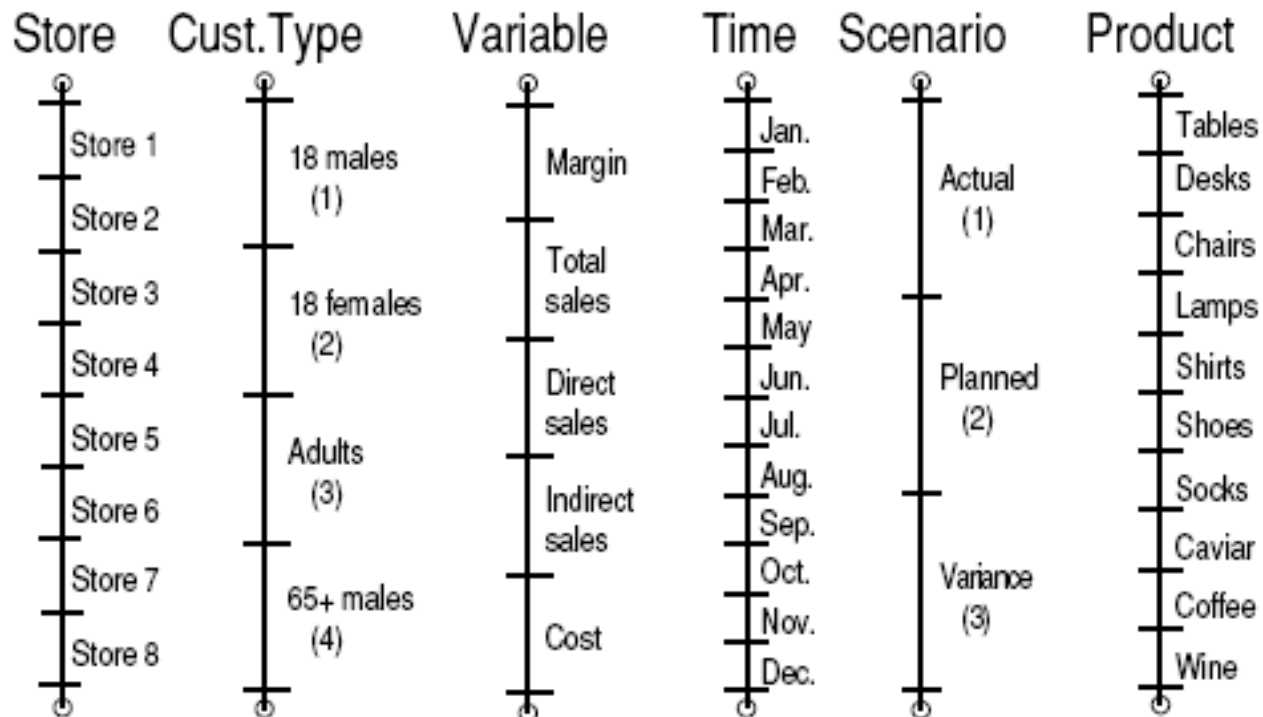
Another form for 4 dimensions

- Four dimensions can be represented by four lines

Time	Product	Variable	Store
Jan.	Tables	Margin	Store 1
Feb.	Desks		Store 2
Mar.	Chairs	Total sales	Store 3
Apr.	Lamps		Store 4
May	Shirts	Direct sales	Store 5
Jun.	Shoes		Store 6
Jul.	Socks	Indirect sales	Store 7
Aug.	Caviar		Store 8
Sep.	Coffee		
Oct.			
Nov.		Cost	
Dec.	Wine		

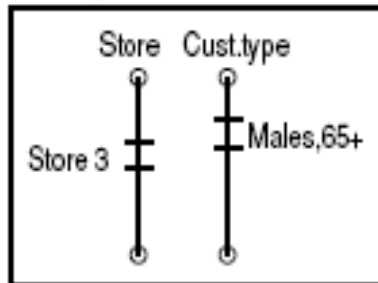
Hyper-cubes on computer screen

- Six dimensions can be represented on computer screen

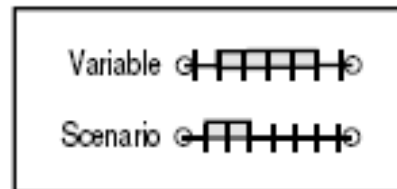


Diff. Six-dimension display

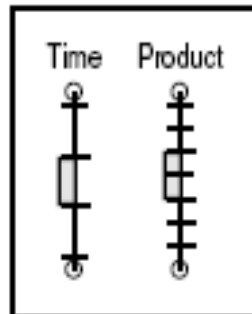
Page order dimension



Column order dimension

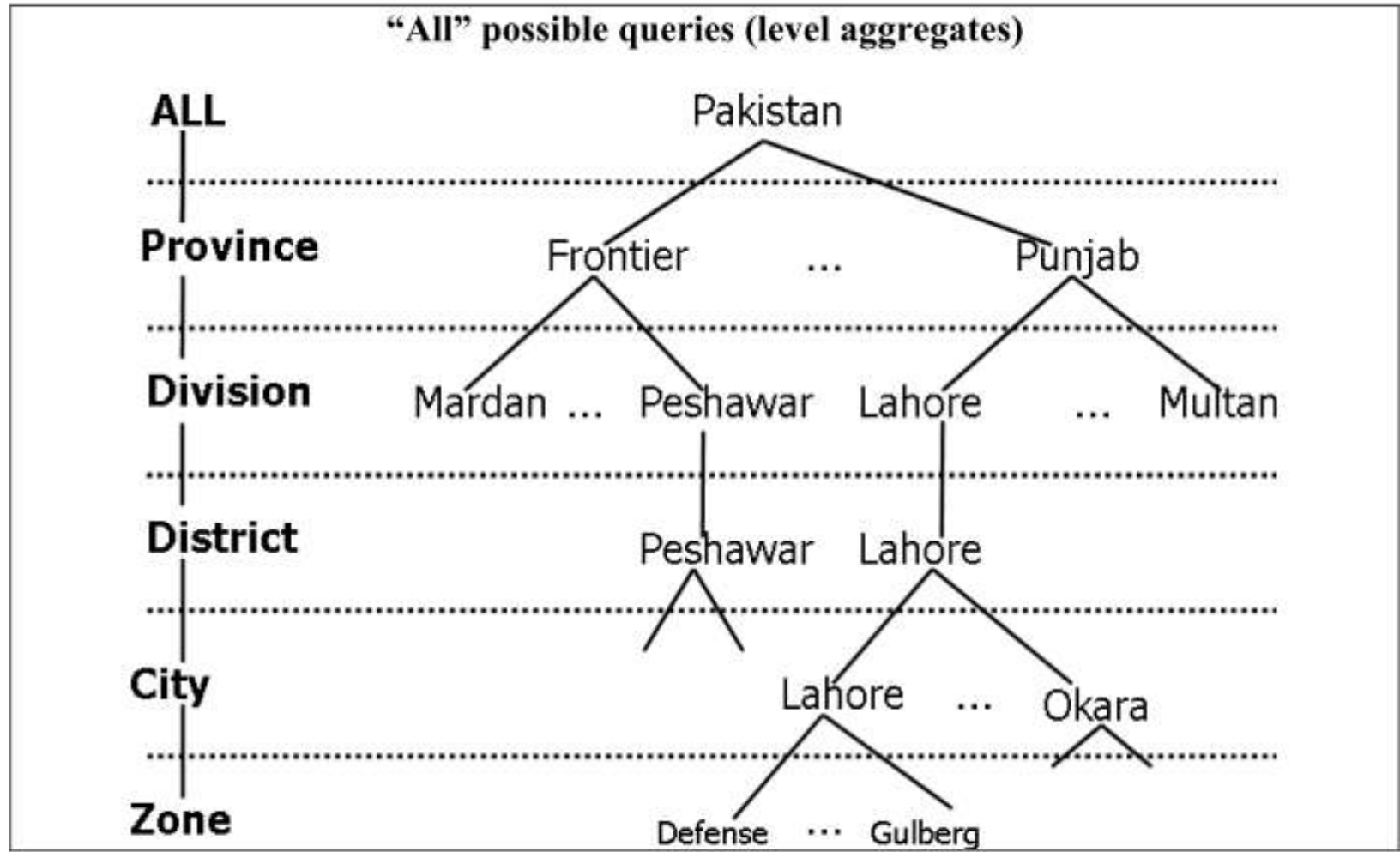


Row order dimension



Store 3	Males,65+	Direct sales		Indirect sales		Total sales	
		Actual	Planned	Actual	Planned	Actual	Planned
January	Desks	250	300	125	150	375	450
	Lamps	267	320	133	160	400	480
February	Desks	333	400	167	200	500	600
	Lamps	283	340	142	170	425	510
March	Desks	350	420	175	210	525	630
	Lamps	250	300	125	150	375	450

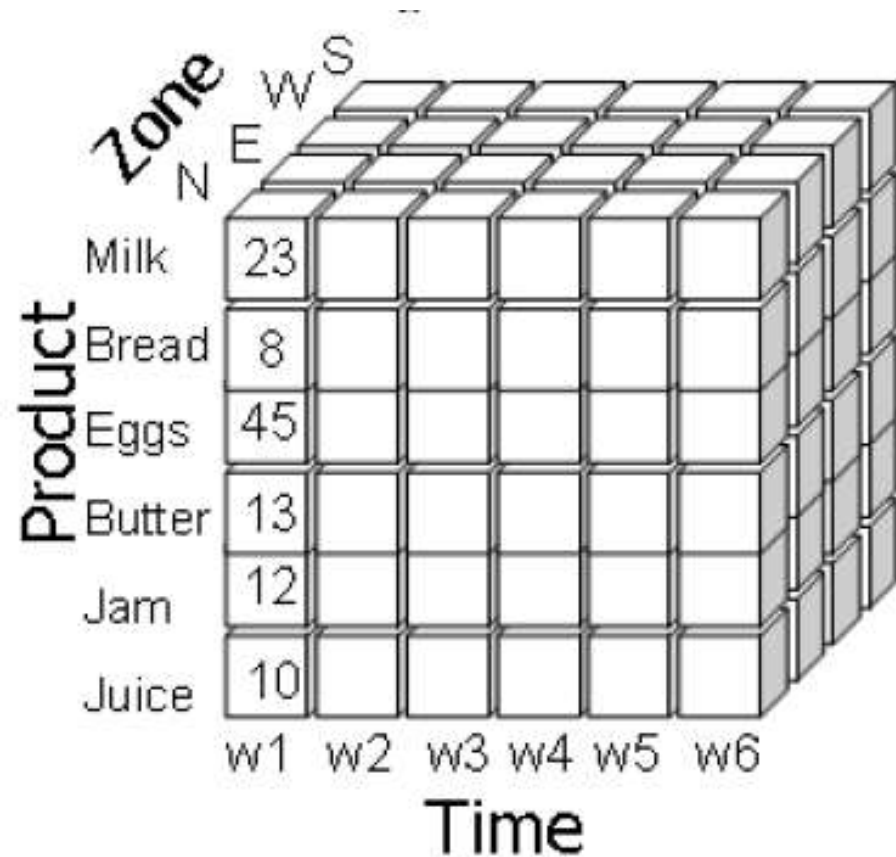
Levels in Dimension



Operations in OLAP

Cube Navigation Operations

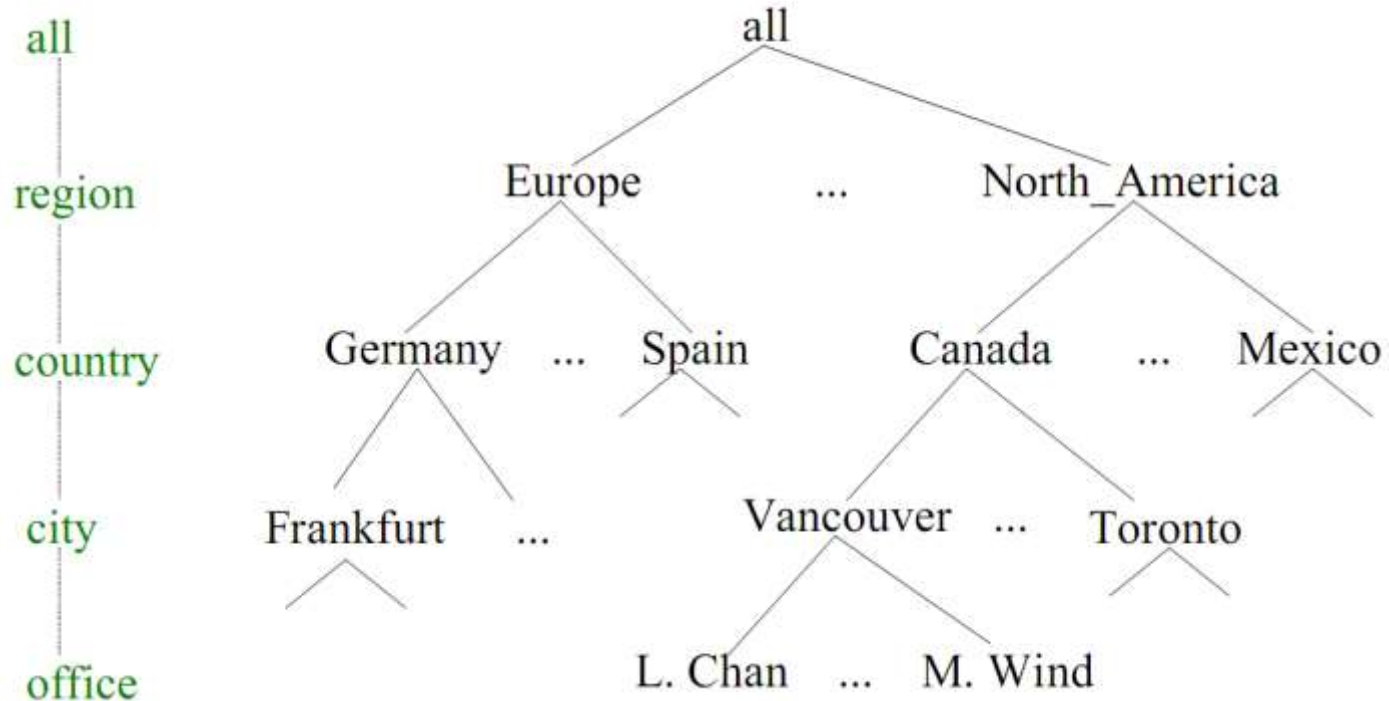
- Roll up
- Drill down
- Pivot
- Slice
- Dice



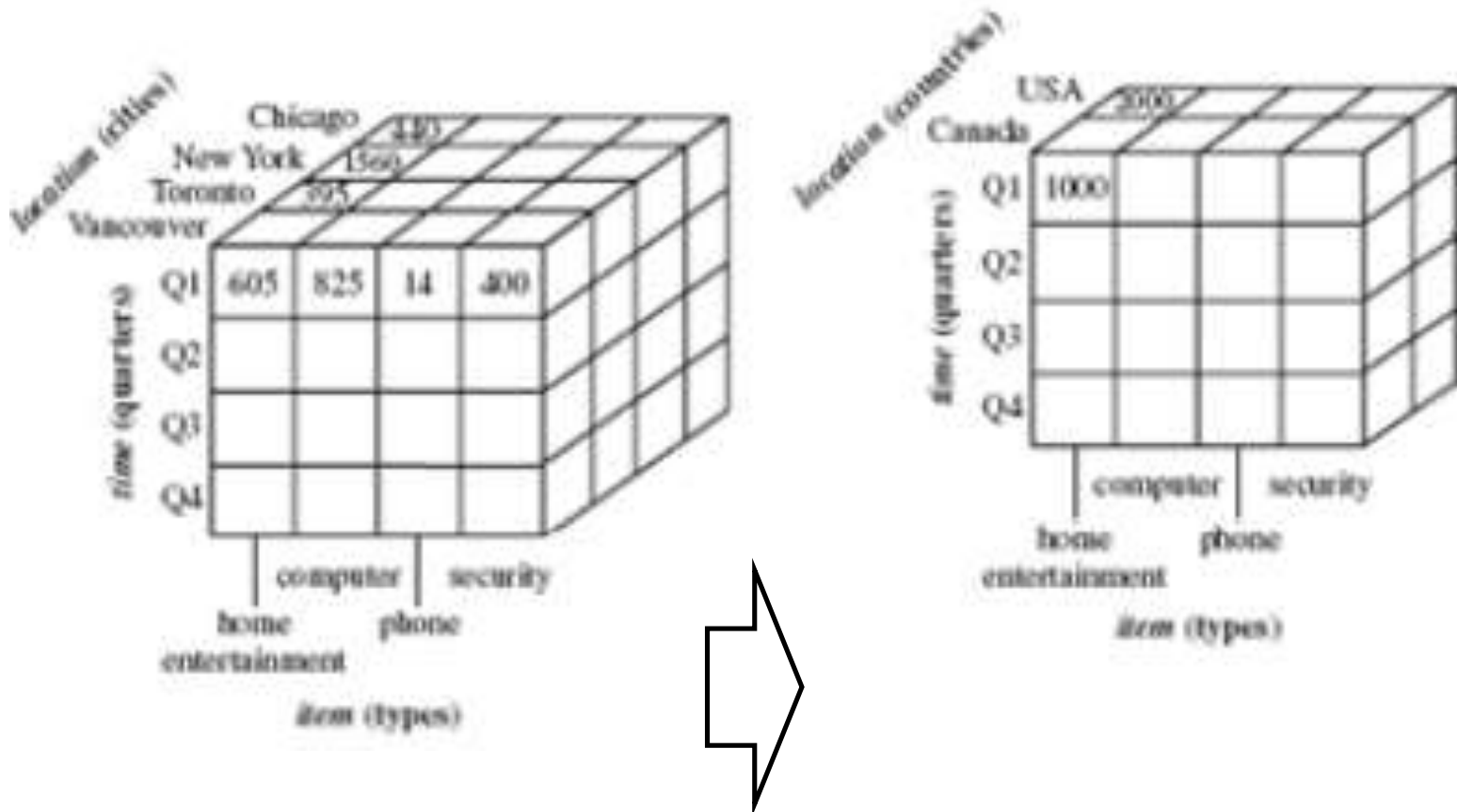
Navigation Operations: Roll-up

■ Roll Up

- Perform aggregation on a data cube, involves summarizing the data along **one dimension**
- Summarize data e.g. given sales



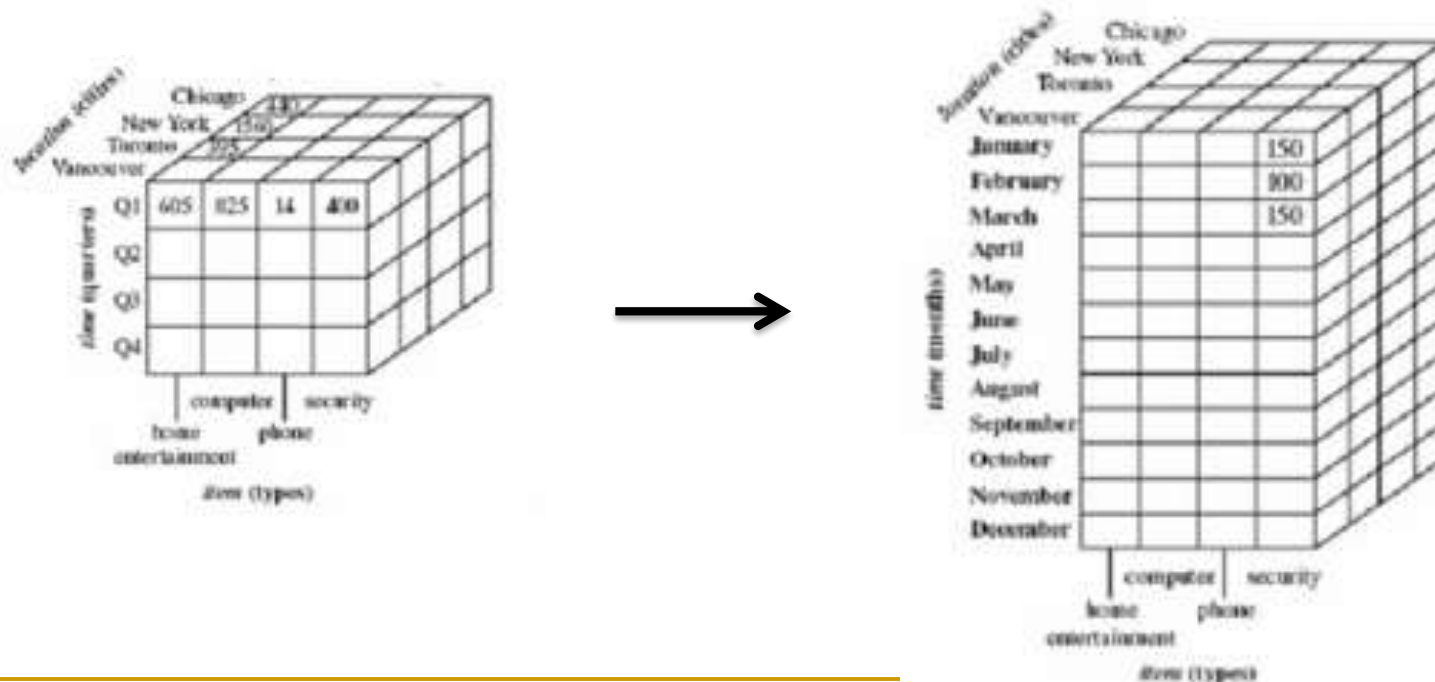
Navigation Operations: Roll-up



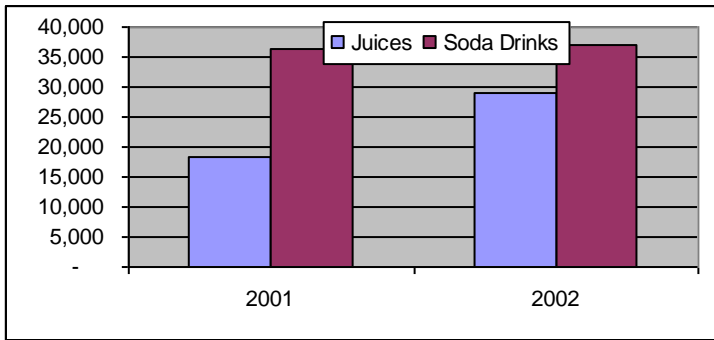
Navigation Operations: Drill-down

■ Drill down

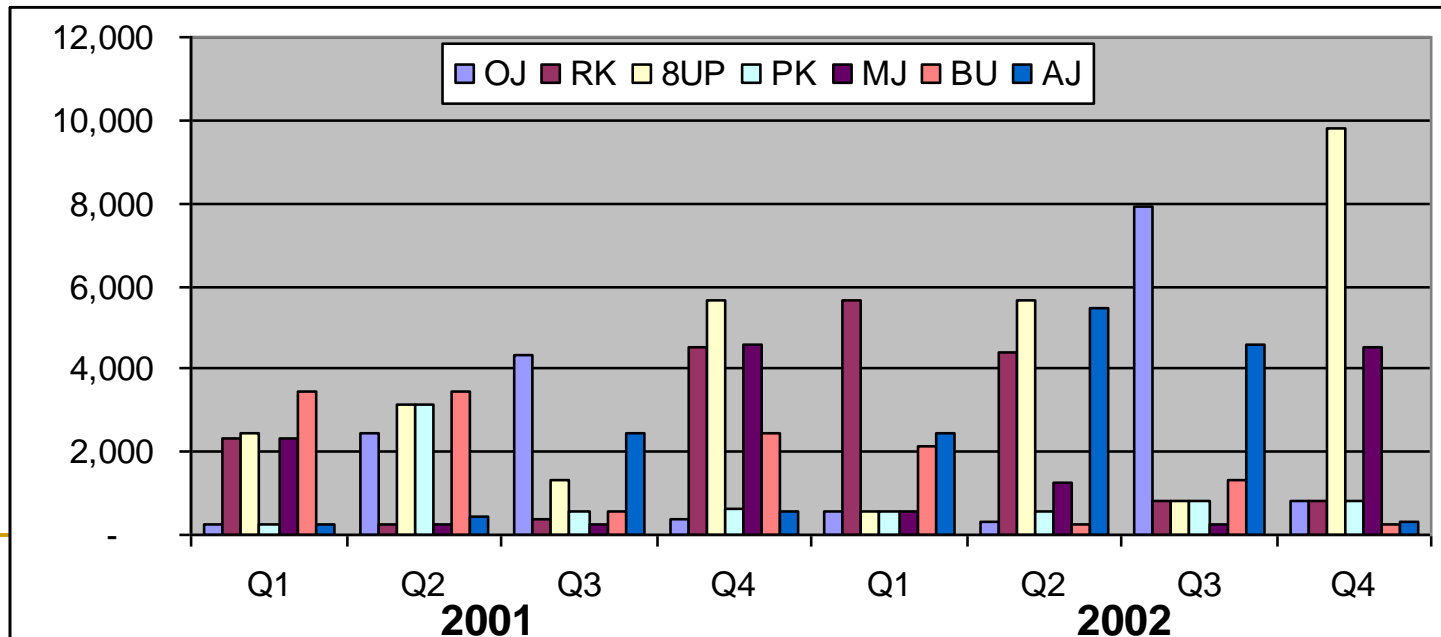
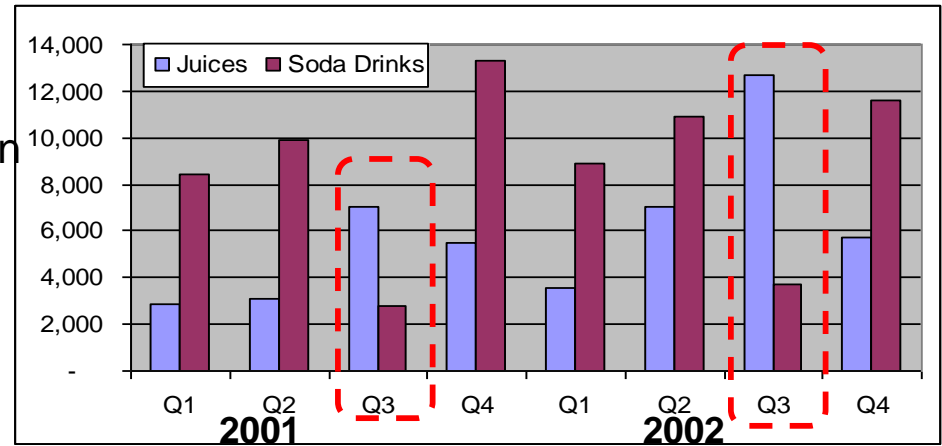
- Get more details, find breakup of sales from Quarter to Months
- Allows users to navigate among levels of data ranging from the most summarized (up) the most detailed (down)



Graphical form



Drill-Down
Roll-Up



Drill-down

Navigation Operations: Drill-down

LINE	TOTAL SALES
Clothing	\$12,836,450
Electronics	\$16,068,300
Video	\$21,262,190
Kitchen	\$17,704,400
Appliances	\$19,600,800
Total	\$87,472,140

High level summary
by product line

Navigation Operations: Drill-down

LINE	TOTAL SALES
Clothing	\$12,836,450
Electronics	\$16,068,300
Video	\$21,262,190
Kitchen	\$17,704,400
Appliances	\$19,600,800
Total	\$87,472,140

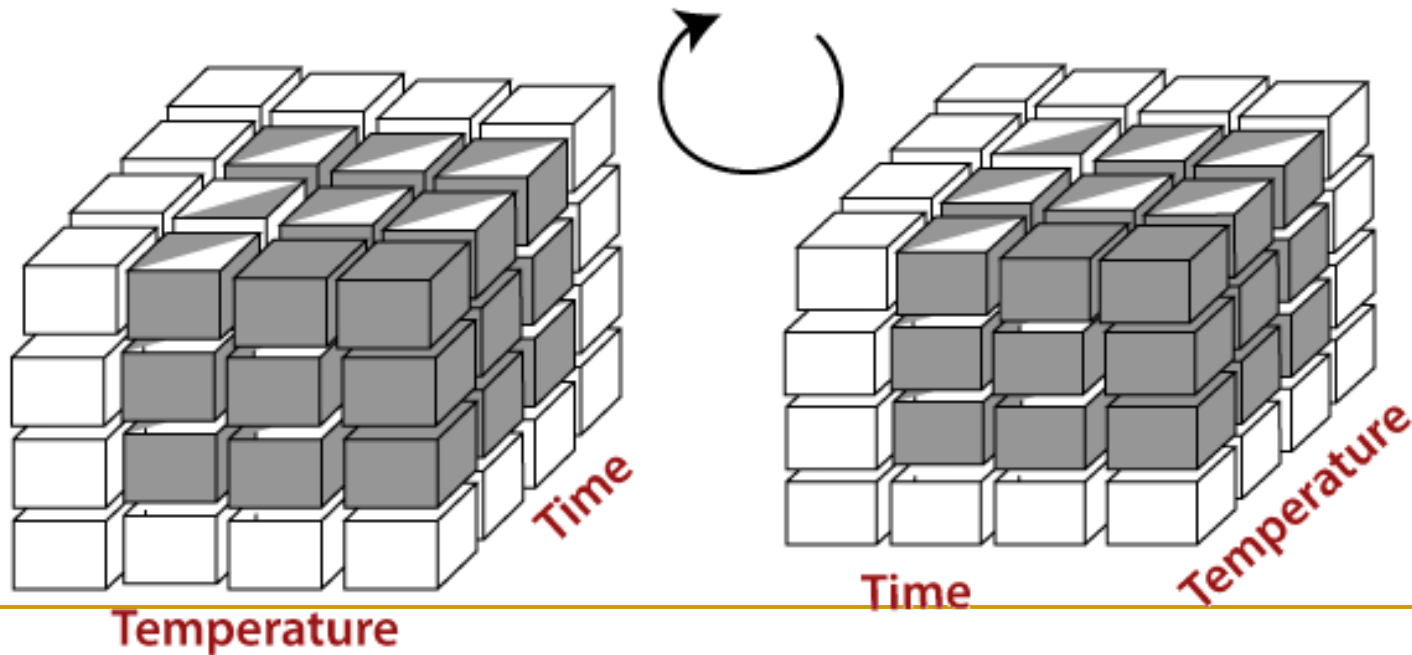
High level summary
by product line

LINE	1998	1999	2000	TOTAL
Clothing	\$3,457,000	\$3,590,050	\$5,789,400	\$12,836,450
Electronics	\$5,894,800	\$4,078,900	\$6,094,600	\$16,068,300
Video	\$7,198,700	\$6,057,890	\$8,005,600	\$21,262,190
Kitchen	\$4,875,400	\$5,894,500	\$6,934,500	\$17,704,400
Appliances	\$5,947,300	\$6,104,500	\$7,549,000	\$19,600,800
Total	\$27,373,200	\$25,725,840	\$34,373,100	\$87,472,140

Drill down
by year

Navigation Operations: Pivot

- Pivot operation is also called rotation
- Rotates the data axes in view to provide alternative presentation



Navigation Operations: Pivot

LINE	1998	1999	2000	TOTAL
Clothing	\$3,457,000	\$3,590,050	\$5,789,400	\$12,836,450
Electronics	\$5,894,800	\$4,078,900	\$6,094,600	\$16,068,300
Video	\$7,198,700	\$6,057,890	\$8,005,600	\$21,262,190
Kitchen	\$4,875,400	\$5,894,500	\$6,934,500	\$17,704,400
Appliances	\$5,947,300	\$6,104,500	\$7,549,000	\$19,600,800
Total	\$27,373,200	\$25,725,840	\$34,373,100	\$87,472,140

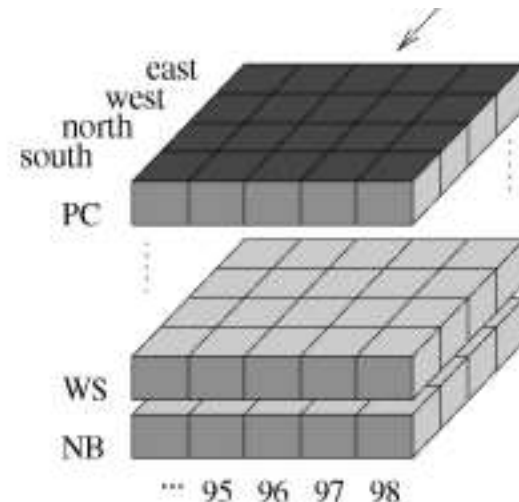
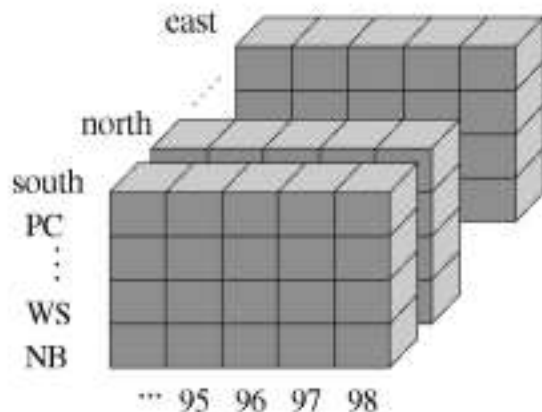
Rotate columns
to rows

YEAR	Clothing	Electronics	Video	Kitchen	Appliances	TOTAL
1998	\$3,457,000	\$5,894,800	\$7,198,700	\$4,875,400	\$5,947,300	\$27,373,200
1999	\$3,590,050	\$4,078,900	\$6,057,890	\$5,894,500	\$6,104,500	\$25,725,840
2000	\$5,789,400	\$6,094,600	\$8,005,600	\$6,934,500	\$7,549,000	\$34,373,100
Total	\$12,836,450	\$16,068,300	\$21,262,190	\$17,704,400	\$19,600,800	\$87,472,140

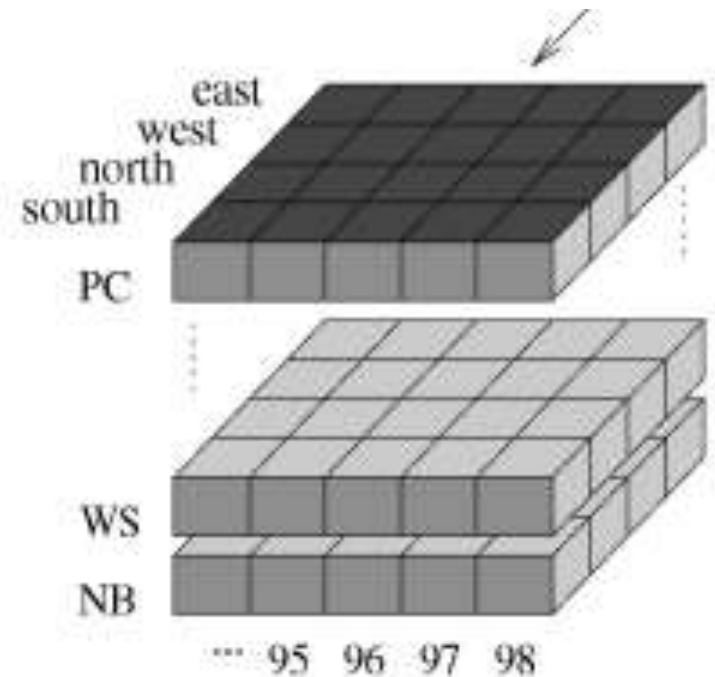
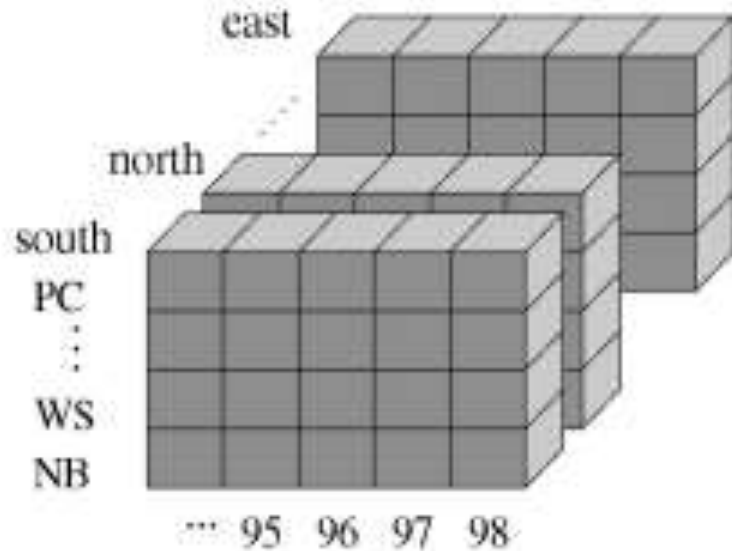
Navigation Operations: Slice

■ Slice Operation

- ❑ Correspond to **reducing** the number of dimensions by taking a project of data in cube on **a dimension**
- ❑ Slice operation will reduce the number of dimensions



Navigation Operations: Slice



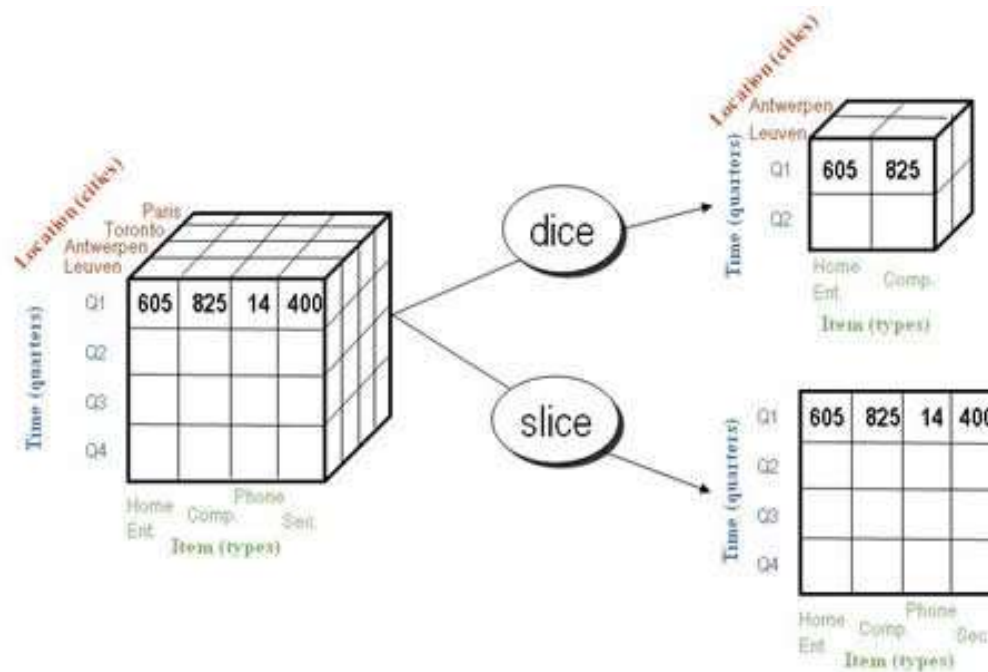
Navigation Operations: Dice

■ Dice Operation

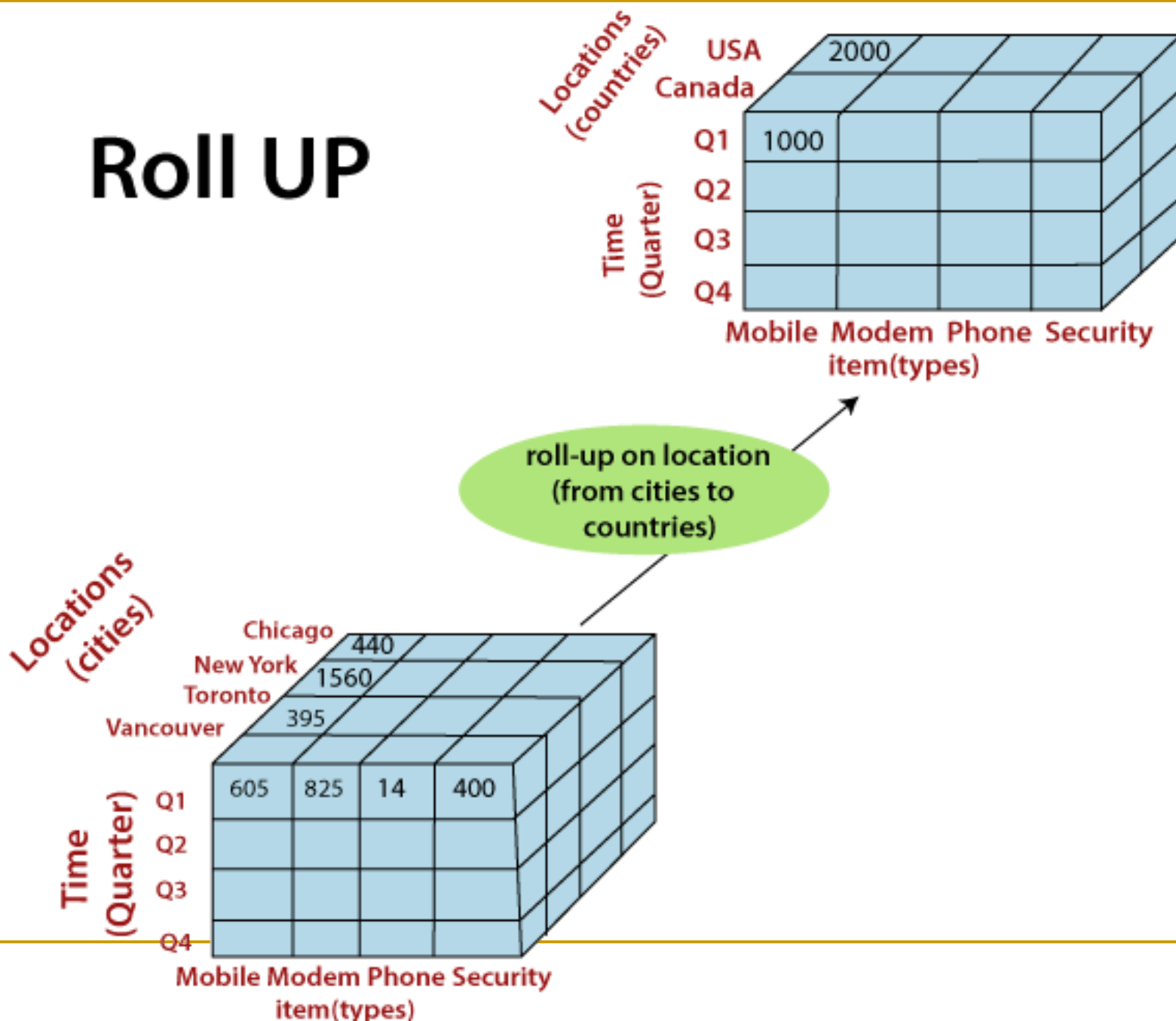
- ❑ Produces a sub-cube based on specific values of **multiple dimensions** (two or more dimensions)
- ❑ Select data that satisfies a list of predicates connected by logical operators AND, OR, or NOT

Navigation Operations: Slice vs. Dice

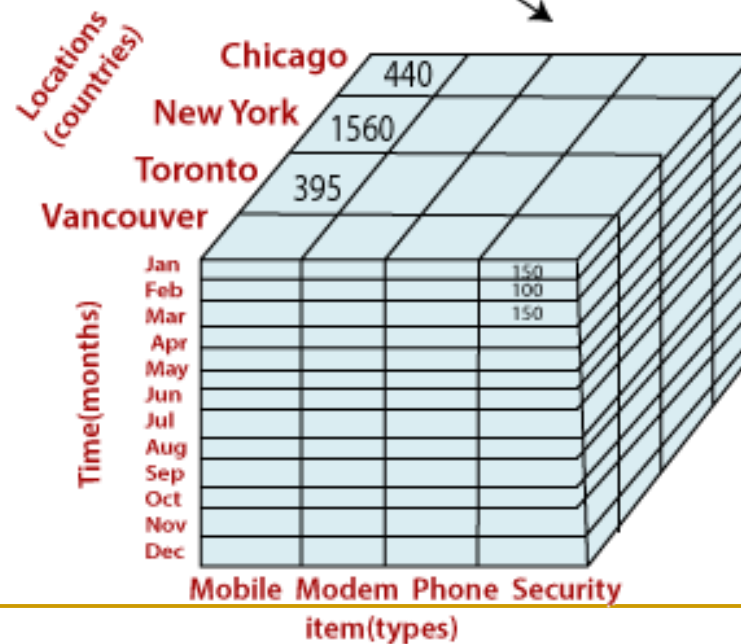
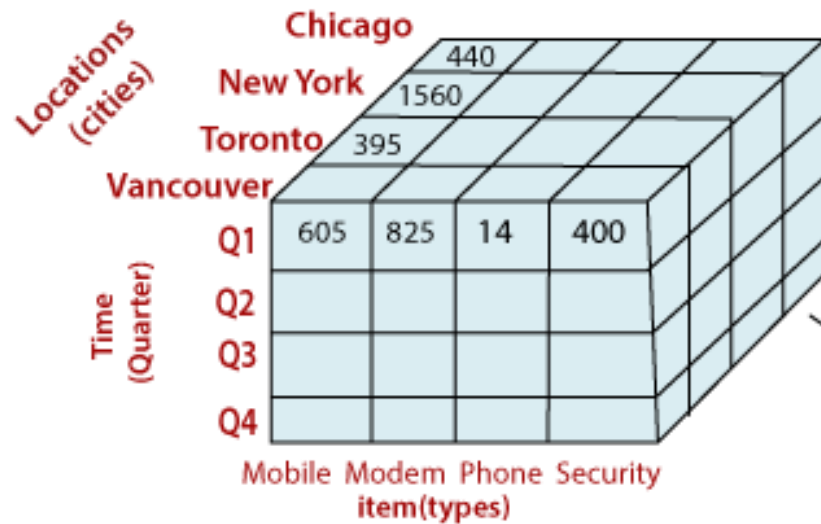
- Slice performs a selection on **one dimension** of a cube
- Dice operation defines a **sub-cube** by performance a selection on two or more dimensions



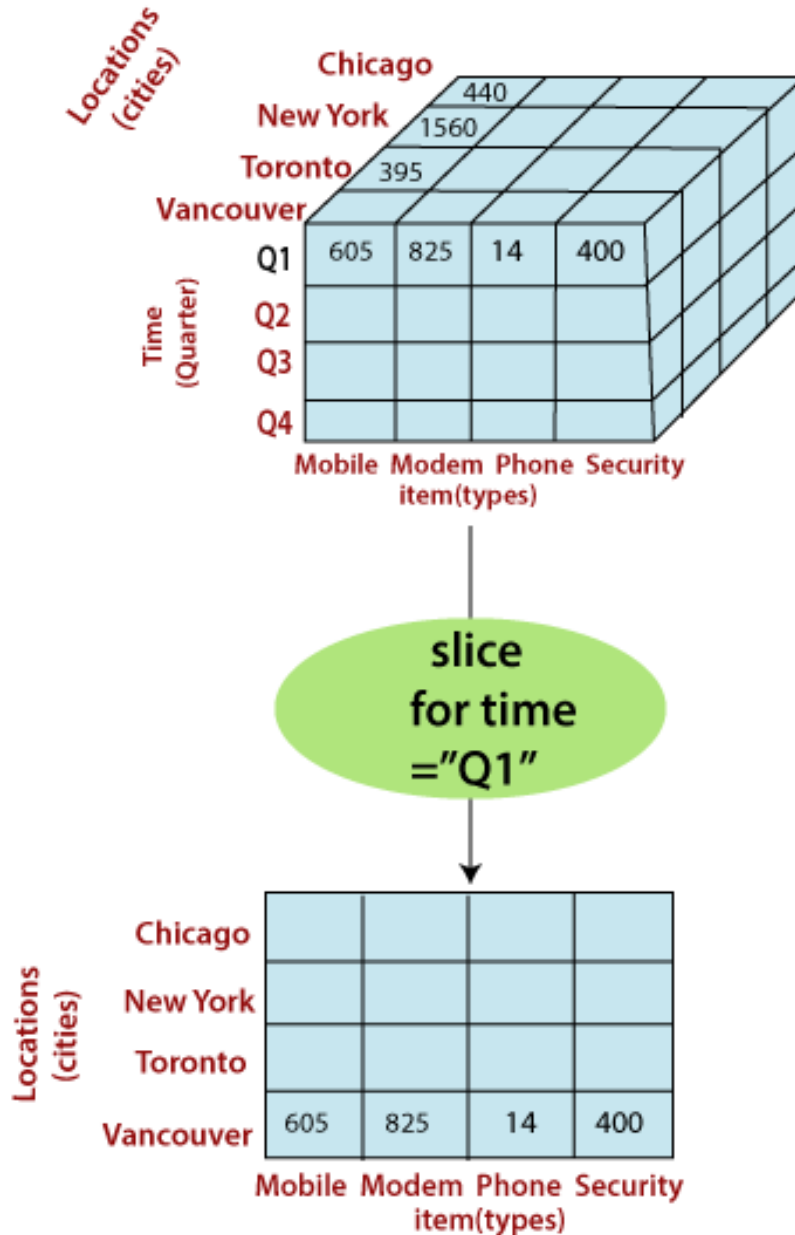
Roll UP



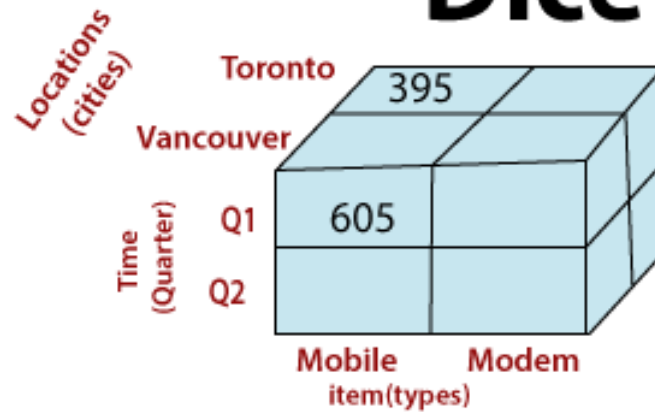
Drill Down



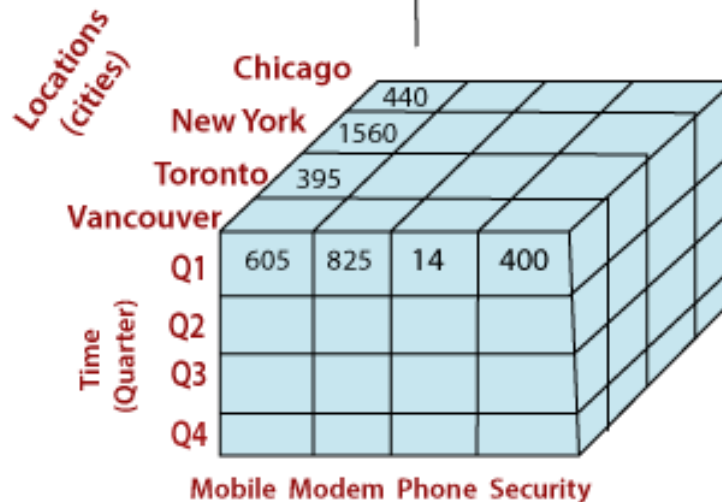
Slice



Dice



Dice for (location="Toronto" or "Vancouver")
and(time="Q1" or "Q2") and
(item="Mobile" or "Modem")



Sample Cube (visualization)

	Sales	Count	Sales	Count
+ Washington	\$520	120	\$700	32
+ California	\$440	129	\$400	20
+ Washington	\$850	120	\$850	44
	+ Redmond		\$250	9
	+ Seattle		\$320	15
+ California	+ Berkeley		\$820	41
	+ Beverly Hills	129	\$910	50
	+ Los Angeles		\$680	38

Sample Cube

Cube Browser - Sales

Product: All Product Promotion: All Promotion Name
 Store: All Store Time: All Time
 Yearly Income: All Yearly Income

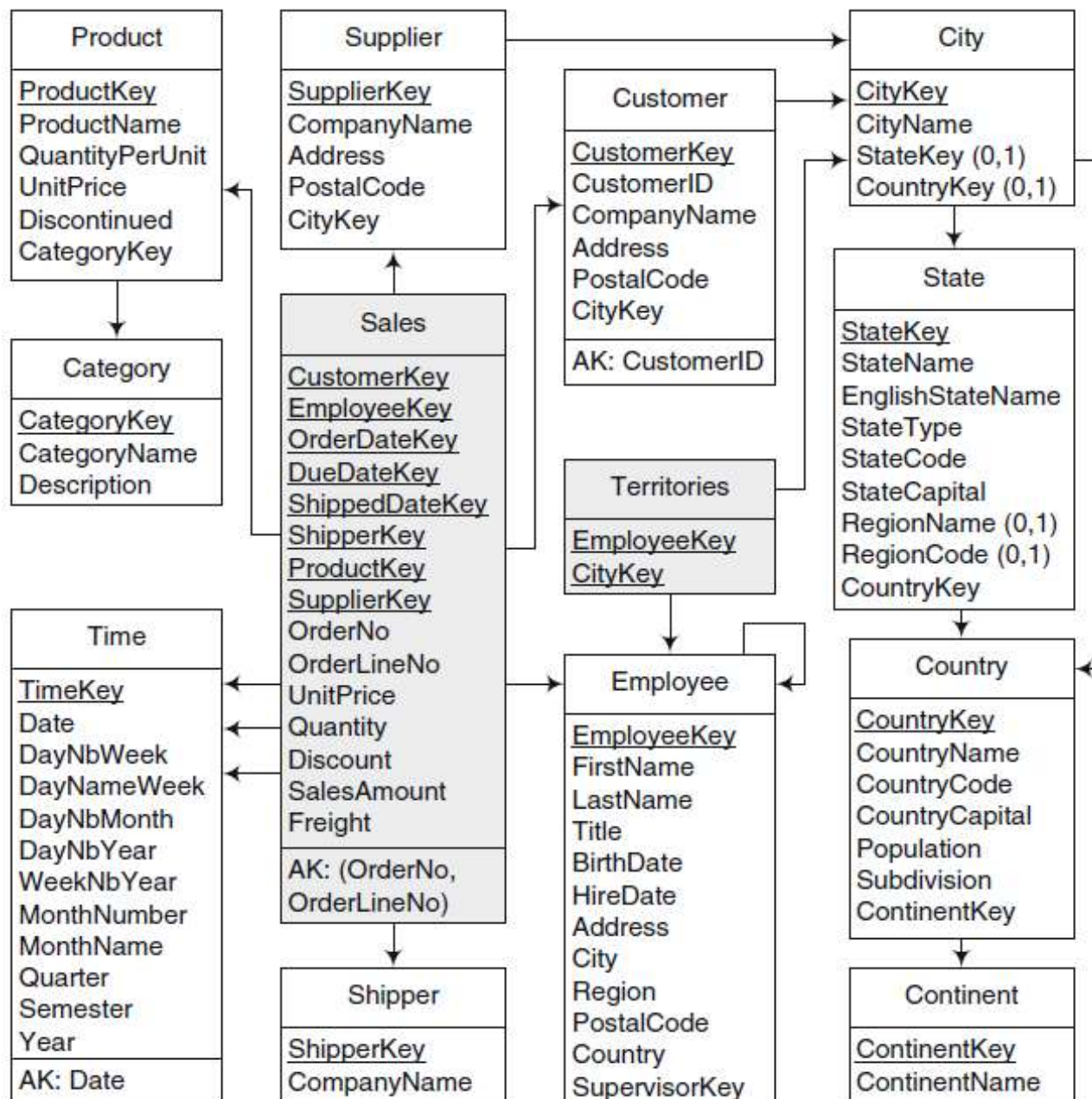
- Country	- State Province	- City	Lname	MeasuresLevel	Store Sales	Store
All Customer	All Customer Total				1,079,147.47	
+ Canada	Canada Total				98,045.46	
+ Mexico	Mexico Total				430,293.59	
	USA Total				550,808.42	
	+ CA	CA Total			154,513.49	
	+ OR	OR Total			128,598.50	
		WA Total			267,696.43	
		+ Anacortes	Anacortes Total		1,338.23	
		+ Ballard	Ballard Total		5,301.58	
		+ Bellingham	Bellingham Total		1,679.21	
		+ Bremerton	Bremerton Total		25,927.72	
		+ Burien	Burien Total		5,091.41	
		+ Edmonds	Edmonds Total		4,583.23	
		+ Everett	Everett Total		5,427.29	
		+ Issaquah	Issaquah Total		4,583.63	
		+ Kirkland	Kirkland Total		6,013.32	
		+ Lynnwood	Lynnwood Total		5,199.78	
		+ Marysville	Marysville Total		4,851.97	
		+ Olympia	Olympia Total		27,800.70	
		+ Port Orchard	Port Orchard Total		25,207.47	
		+ Puyallup	Puyallup Total		23,123.39	
			Redmond Total		5,158.29	
			Abbey		30.33	
			Alstorn		104.98	
			Autobee			
			Bagwell		108.43	
			Banks		8.04	
			Bateman		35.12	
			Bates		169.90	
			Beerbaum		85.24	
			Berner		31.57	

Double-click a member to drill up or down.

Close Help

Northwind Example

OLAP Operations



Implementation of OLAP

OLAP Implementation Architectures

- MOLAP
 - ROLAP
 - HOLAP
-

OLAP Implementation

- **MOLAP**: OLAP implemented with a multi-dimensional data structure.
- **ROLAP**: OLAP implemented with a relational database.
- **HOLAP**: OLAP implemented as a hybrid of MOLAP and ROLAP.

OLAP Implementation: MOLAP

- Data is stored in the form of multi-dimensional arrays (cubes).
 - Analytical tools to allow analysis of data
 - It requires the pre-computation and storage of information in cube
-

OLAP Implementation: MOLAP

■ Advantages

- ❑ Instance response (pre-calculated aggregates)
- ❑ Excellent performance

■ Disadvantages

- ❑ Limited in the amount of data it can handle
 - ❑ Long load time (due to pre-calculation)
 - ❑ Very sparse cube (space wastage) for high cardinality
 - ❑ Actually, the cubes are less than 1% full
-

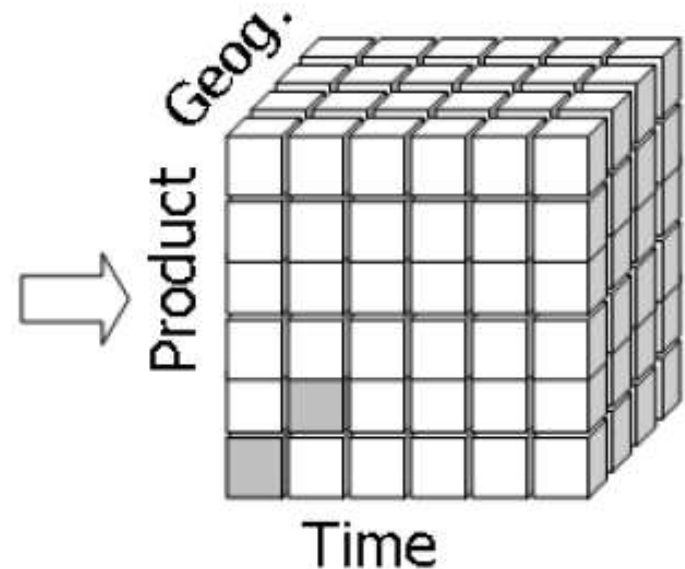
Data Sparsity and Dense Value

- The term sparsity has been used for missing, inapplicable, zero value
- Notice all sparse or blank cell
 - Identification of sparse or blank cell is very important
 - Very sparse cube (wastage of space)
- Opposite value is DENSE value
- Some values are meaningless (Identify them)
- Need to distinguish missing & meaningless value

Data Sparsity and Dense Value

Fact Table

Month	Product	Zone	Sale K <u>Rs.</u>
M1	P1	Z1	250
M2	P2	Z1	500



OLAP Implementation: ROLAP

- Data is stored as relational tables
- ROLAP and space requirements
 - Increase in number of dimensions increases number of summary tables
 - Consider two dimensions
 - Time: Day, Week, Month, Quarter, Year, All days
 - Product: Item, subcategory, Category, all products

②	2001				2002			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Orange juice	232	2,432	4,353	354	535	345	7,897	789
Rola-Kola	2,342	243	353	4,535	5,655	4,424	789	798
8-UP	2,424	3,131	1,313	5,675	567	5,675	789	9,797
Pola-Kola	242	3,112	567	646	567	567	789	798
Mango juice	2,342	243	243	4,564	564	1,232	242	4,553
Bubbly-UP	3,453	3,453	535	2,422	2,131	242	1,321	245
Apple juice	253	456	2,433	567	2,442	5,453	4,566	345

③	2001				2002			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Soda Drinks	8,461	9,939	2,768	13,278	8,920	10,908	3,688	11,638
Juices	2,827	3,131	7,029	5,485	3,541	7,030	12,705	5,687

ROLAP as a “Cube”

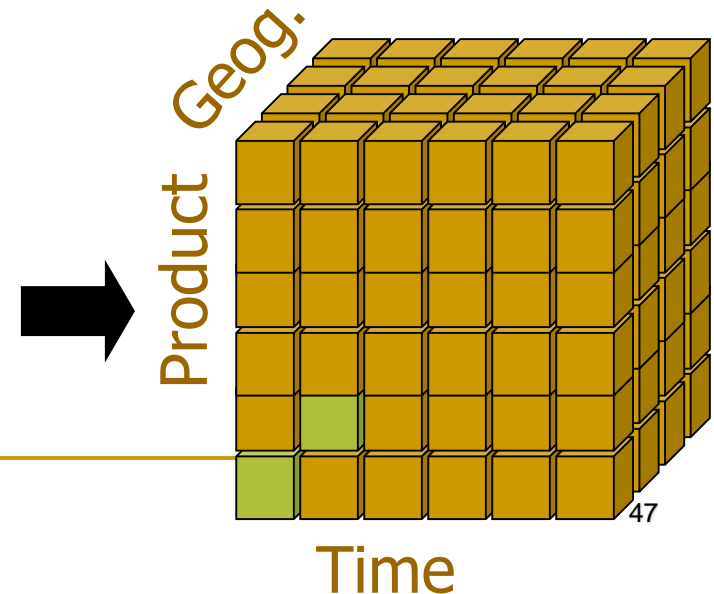
- OLAP data is stored in a relational database (e.g. a star schema)

The fact table is a way of *visualizing* as a “un-rolled” cube.

- So where is the **cube**?
 - It's a matter of perception

Fact Table

Month	Product	Zone	Sale K Rs.
M1	P1	Z1	250
M2	P2	Z1	500



OLAP Implementation: ROLAP

■ Advantages

- ❑ Can handle large amount of data
- ❑ Can leverage functionalities inherent in the relational DB

■ Disadvantages

- ❑ Performance can be slow
 - ❑ Limited by SQL functionalities
-

OLAP Implementation: HOLAP

- Hybrid OLAP
 - Target is to get the best of both worlds
 - HOLAP allow co-existence of pre-build MOLAP cubes alongside relational OLAP or ROLAP structure
-

OLAP Implementation: Summary

- Multidimensional OLAP (MOLAP)
 - Best Query Performance
 - Relational OLAP (ROLAP)
 - Ideal for large databases
 - Hybrid OLAP (HOLAP)
 - Best & Worst of both worlds!
-