Data Mining & warehouse

Experiment – 7

Name :- Akshat Parekh

Sap Id :- 60004200116

Branch :- Computer Engineering

Batch :- B – 2

Aim :- To perform various OLAP operations such as slice, dice, drilldown, rollup, pivot

Theory :-

OLAP is an acronym for On Line Analytical Processing**.** Online Analytical Processing: An OLAP system manages large amount of historical data, provides Facilities for summarization and aggregation, and stores and manages information at different levels of granularity.

OLAP operations:

Slice: A slice is a subset of a multi-dimensional array corresponding to a single value for one or more members of the dimensions not in the subset.

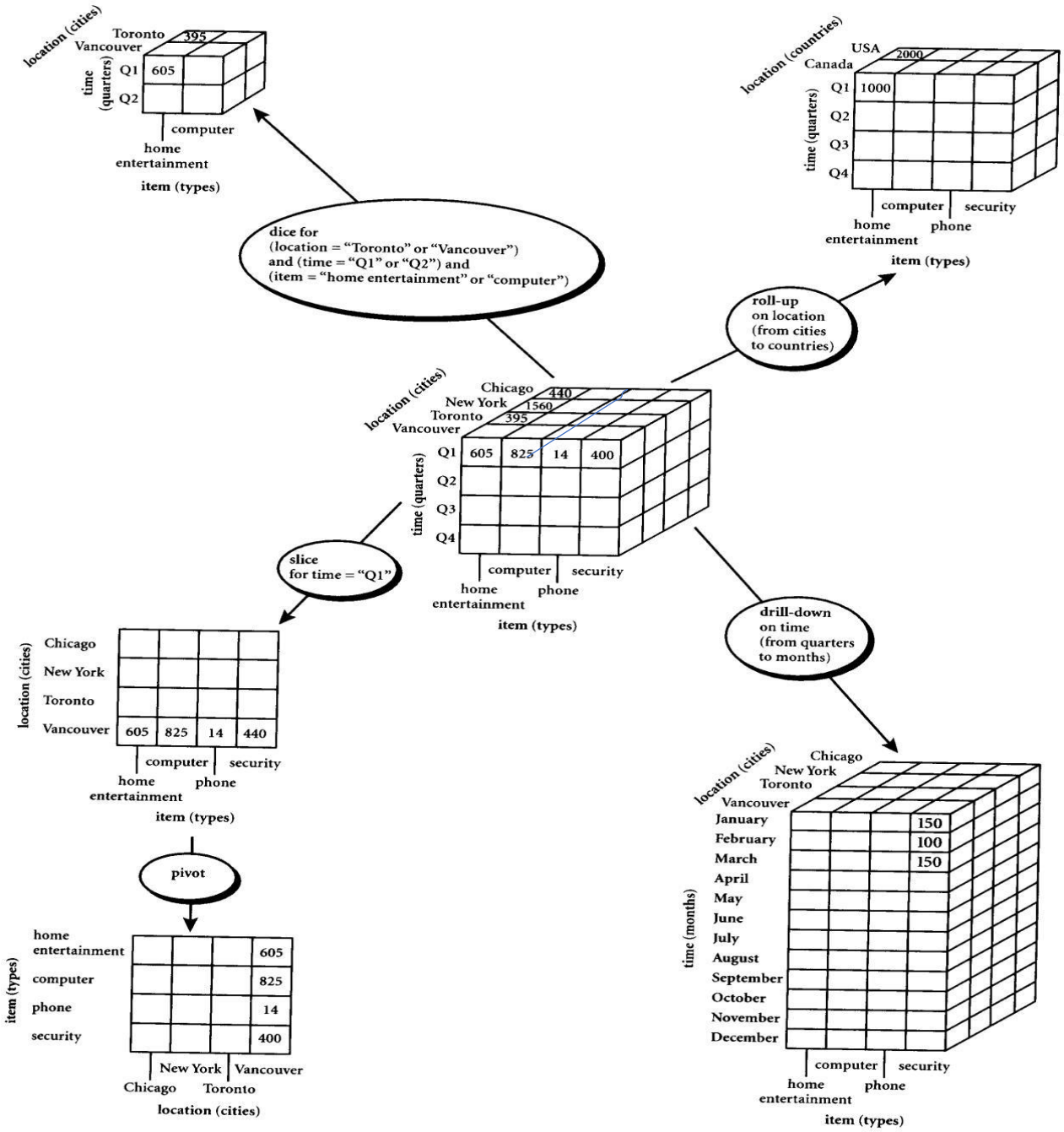
Dice: The dice operation is a slice on more than two dimensions of a data cube (or more than two consecutive slices).

Drill Down/Up: Drilling down or up is a specific analytical technique whereby the user navigates among levels of data ranging from the most summarized (up) to the most detailed (down).

Roll-up: A roll-up involves computing all of the data relationships for one or more dimensions. To do this, a computational relationship or formula might be defined.

Pivot: To change the dimensional orientation of a report or page display.

Example with all the operation performed of OLAP on one dimensional cube.



EXERCISE 1

Consider a data Warehouse for a hospital, where there are three dimensions:

(i) Doctor

(ii) Patient

(iii) Time

With two measures

(a) Count

(b) Charge

Where Charge is the fee that the Doctor charges a patient for a visit.

Using the above example describe the following operations:

(i) Slice

(ii) Dice

(iii) Roll Up

(iv) Drill Down

(v) Pivot

NOTE: Assume data according to the dimensions and measures and explore individual tasks diagrammatically.

(iii) roll up

Doctor	50
Patient	550
Time (monthly)	Monthly count

roll up on time to form monthly count & combine & monthly patient count

Dimension Table

(dimensions)

Doctor	15	500
patient		
Time		
	count	Charge

(measures)

(iv) Drill down

Drill down charge to several level like doctor fees, medicine fees, Hospital utility fees, etc.

[charge]

Doctor	150	200	150
patient			
	doctor fees	medicine cost	hospital utility fees

(i) slice

slice for Doctor

Doctor	15	500
	count	charge

(ii) dice

dice for [dimension = Doctor and patient] and [measure = count]

Doctor	15
patient	
	count

(v) pivot

count	15
charge	500
	Doctor

EXERCISE 2

To create Pivot of Table using MS Excel

Follow these steps ...

1. Start with M.S Excel.
2. In excel sheet create 4 columns PRODUCT, ORIGIN, DAY OF SALE, SOLD UNITS (FACT COLUMN).
3. Insert around fifty rows of data.
4. Save the table data.
5. Go to Insert Tab-> click on Pivot Table-> New work sheet-> Ok.
6. Right side you will find pivot table fields.

   It contains all columns of our table that we created.

   Select product in rows,

   Days in column,

   Unit sold in ∑ values.

   Later apply filter using Origin.

   Also we can flip the rows & columns or combine together as rows only to see different views of same data.

DATA :-

	A	B	C	D	E
1	PRODUCT	ORIGIN	DAY OF SALE	SOLD UNITS	
2	fafda	gujarat	Sunday	150	
3	south indian	banglore	Monday	1000	
4	bengali sweet	bengol	Saturday	100	
5	dabeli	mumbai	Friday	500	
6	frankie	mumbai	Wednesday	1000	
7	oil	dubai	Tuesday	20000	
8	kachori	gujarat	sunday	250	
9	nipat	benglore	monday	1500	
10	vada pav	mumbai	wednesday	5000	
11	gold	dubai	Thursday	300	
12	paratha	panjab	monday	500	
13	chaas	gujarat	sunday	2000	
14	lassi	panjab	sunday	1000	
15					

PIVOT table for above table

	A	B	C	D	E	F	G	H	I	J
1	ORIGIN	(All)								
2										
3	**Sum of SOLD UNITS**	**Column Labels**								
4	**Row Labels**	**Monday**	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**	**Grand Total**	
5	bengali sweet						100		100	
6	chaas							2000	2000	
7	dabeli					500			500	
8	fafda							150	150	
9	frankie			1000					1000	
10	gold				300				300	
11	kachori							250	250	
12	lassi							1000	1000	
13	nipat	1500							1500	
14	oil		20000						20000	
15	paratha	500							500	
16	south indian	1000							1000	
17	vada pav			5000					5000	
18	**Grand Total**	**3000**	**20000**	**6000**	**300**	**500**	**100**	**3400**	**33300**	
19										

Conclusion :-

Thus, we have successfully completed all the operation of OLAP on the dimension cube.