JNIESTR's
Smt. Indira Gandhi College of Engineering
Plot No -1, Sector- 8, Ghansoli, Navi Mumbai-400701
Accredited by NAAC with A Grade
Contact:- +91-8080949300, +91-7738921077 Email: contactus@sigce.edu.in Website: www.sigce.edu.in
Department: CSE-IOT

Experiment No-3

Title: Implement OLAP operations

Objective:

• To learn fundamental of data warehousing

• To learn concepts of dimensional modeling

• To learn star, snowflake & Galaxy schema

Description:

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

Since OLAP servers are based on multidimensional view of data, we will discuss OLAP operations in multidimensional data. Here is the list of OLAP operations:

Roll-up

· Drill-down

· Slice and dice

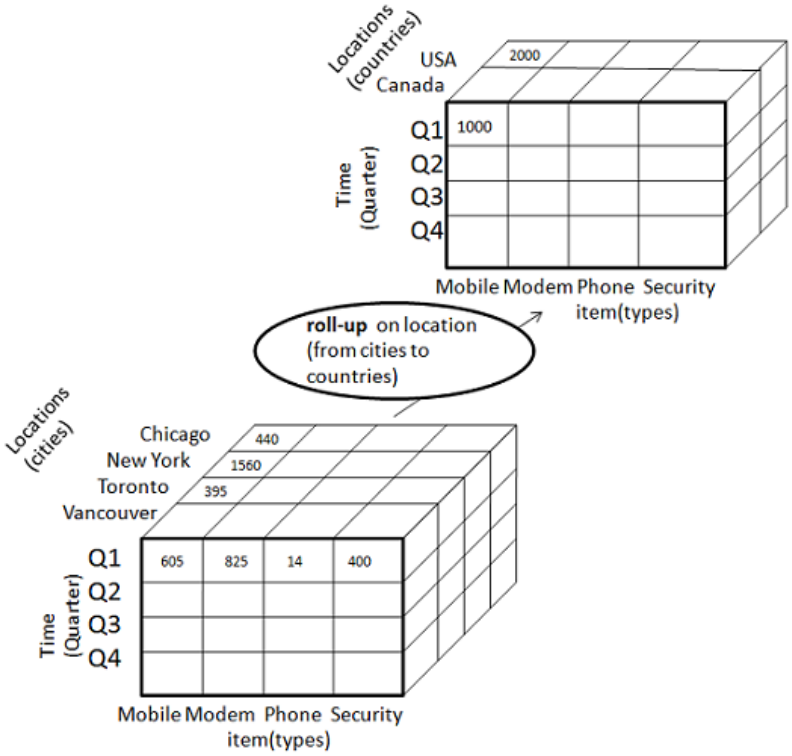
· Pivot (rotate)

Roll-up

Roll-up performs aggregation on a data cube in any of the following ways −

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

The following diagram illustrates how roll-up works.



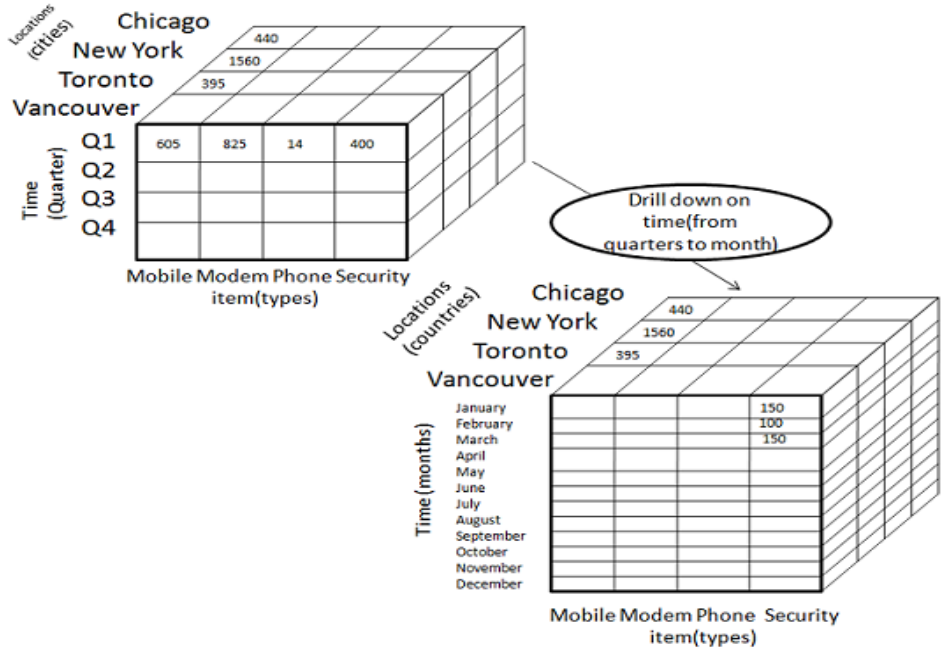
- Roll-up is performed by climbing up a concept hierarchy for the dimension location.
- Initially the concept hierarchy was "street < city < province < country".
- On rolling up, the data is aggregated by ascending the location hierarchy from the level of city to the level of country.
- The data is grouped 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.

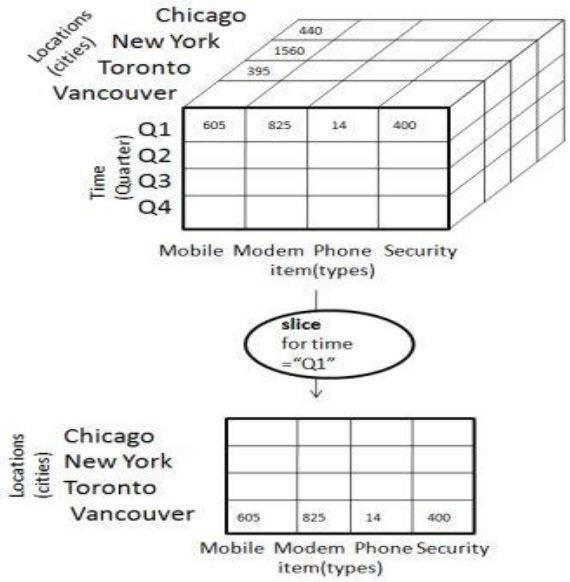
The following diagram illustrates how drill-down works −



- 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

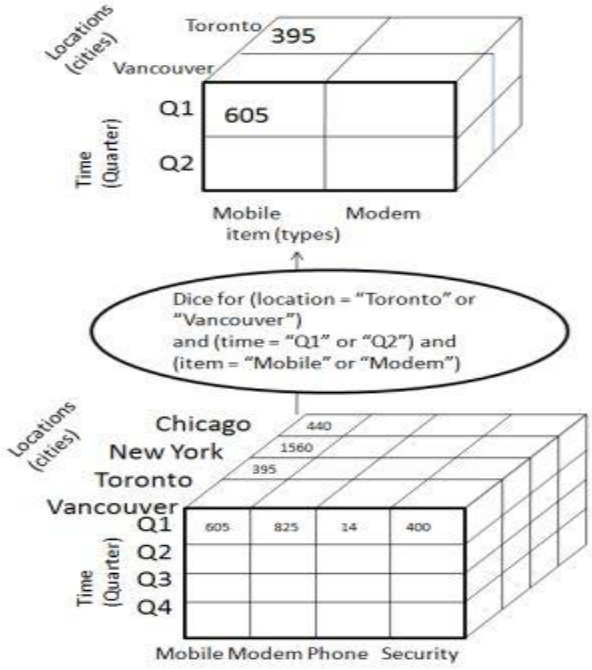
The slice operation selects one particular dimension from a given cube and provides a new sub-cube. Consider the following diagram that shows how slice works.



- 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.



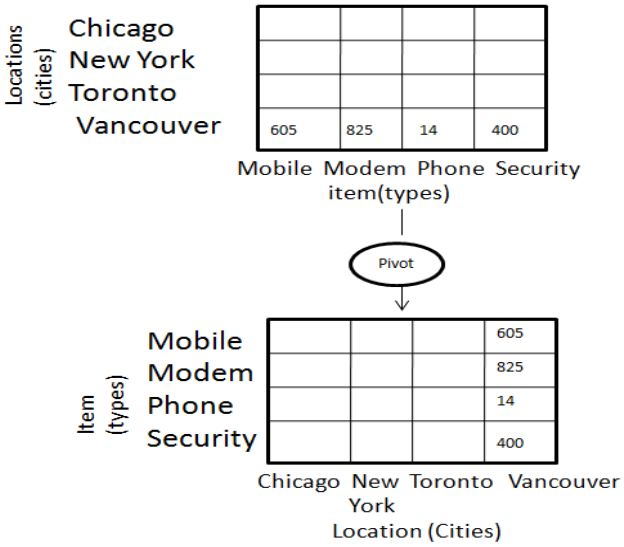
The dice operation on the cube based on the following selection criteria involves three dimensions.

- (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.

JNIESTR's
Smt. Indira Gandhi College of Engineering
Plot No -1, Sector- 8, Ghansoli, Navi Mumbai-400701
Accredited by NAAC with A Grade
Contact:- +91-8080949300, +91-7738921077 **Email:** contactus@sigce.edu.in **Website:** www.sigce.edu.in
Department: CSE-IOT



JNIESTR's
Smt. Indira Gandhi College of Engineering
Plot No -1, Sector- 8, Ghansoli, Navi Mumbai-400701
Accredited by NAAC with A Grade
Contact:- +91-8080949300, +91-7738921077 Email: contactus@sigce.edu.in Website: www.sigce.edu.in
Department: CSE-IOT

Experiment no 4:

AIM: Implement data exploration

- **Solving exercise in data exploration**
  Ques: Explore the following data with help of exploration technique:
  Compute mean median, mode, also compute five number summary .
  13,15,16,19,20,20,21,22,22,25,25,25,25,30, 35,35,35
- **Solving exercise in data preprocessing**
  Ques: Smooth the following data by using
  1. BIN Median
  2. BIN Boundaries
  3. BIN Mean

  8 16, 9, 15, 21, 21, 24, 30, 26, 27, 30, 34

THEORY:

Ques: Explore the following data with help of exploration technique:
Compute mean, median, mode, also compute five number summaries.
13,15,16,19,20,20,21,22,22,25,25,25,25,30, 35,35,35

Soln 1:

Arithmetic Mean $=1/n \sum x_i$

- Mean x=23.705882352941
- Median x=22
- Mode=25

Range=22
Minimum=13
Maximum=35
Count n=17
Sum=403
Quartiles
Q_1 -->19.5
Q_2 -->22
Q_3 --> 27.5
Interquartile Range IQR=8
Outliers: none

Ques: Smooth the following data by using

1. BIN Median
2. BIN Boundaries
3. BIN Mean

8 16, 9, 15, 21, 21, 24, 30, 26, 27, 30, 34

Soln2:

1. **BIN Mean:**

- **Bin 1:** 8, 9, 15, 16
- **Bin 2:** 21, 21, 24, 26,
- **Bin 3:** 27, 30, 30, 34

Bin mean:

For Bin 1:

(8+ 9 + 15 +16 / 4) = 12

Bin 1 = 12, 12, 12, 12

For Bin 2:

(21 + 21 + 24 + 26 / 4) = 23

Bin 2 = 23, 23, 23, 23

For Bin 3:

(27 + 30 + 30 + 34 / 4) = 30

Bin 3 = 30, 30, 30, 30

2.BIN Boundaries

- **Bin 1:** 8, 9, 15, 16
- **Bin 2:** 21, 21, 24, 26,
- **Bin 3:** 27, 30, 30, 34

BIN Boundaries

Bin 1: 8, 8, 8, 15

Bin 2: 21, 21, 25, 25

Bin 3: 26, 26, 26, 34

2. **BIN Median**

- **Bin 1:** 8, 9, 15, 16
- **Bin 2:** 21, 21, 24, 26,
- **Bin 3:** 27, 30, 30, 34

BIN Median

Bin 1: 12, 12,12, 12

Bin 2: 22.5,22.5,22.5,22.5

Bin 3: 30,30,30,30