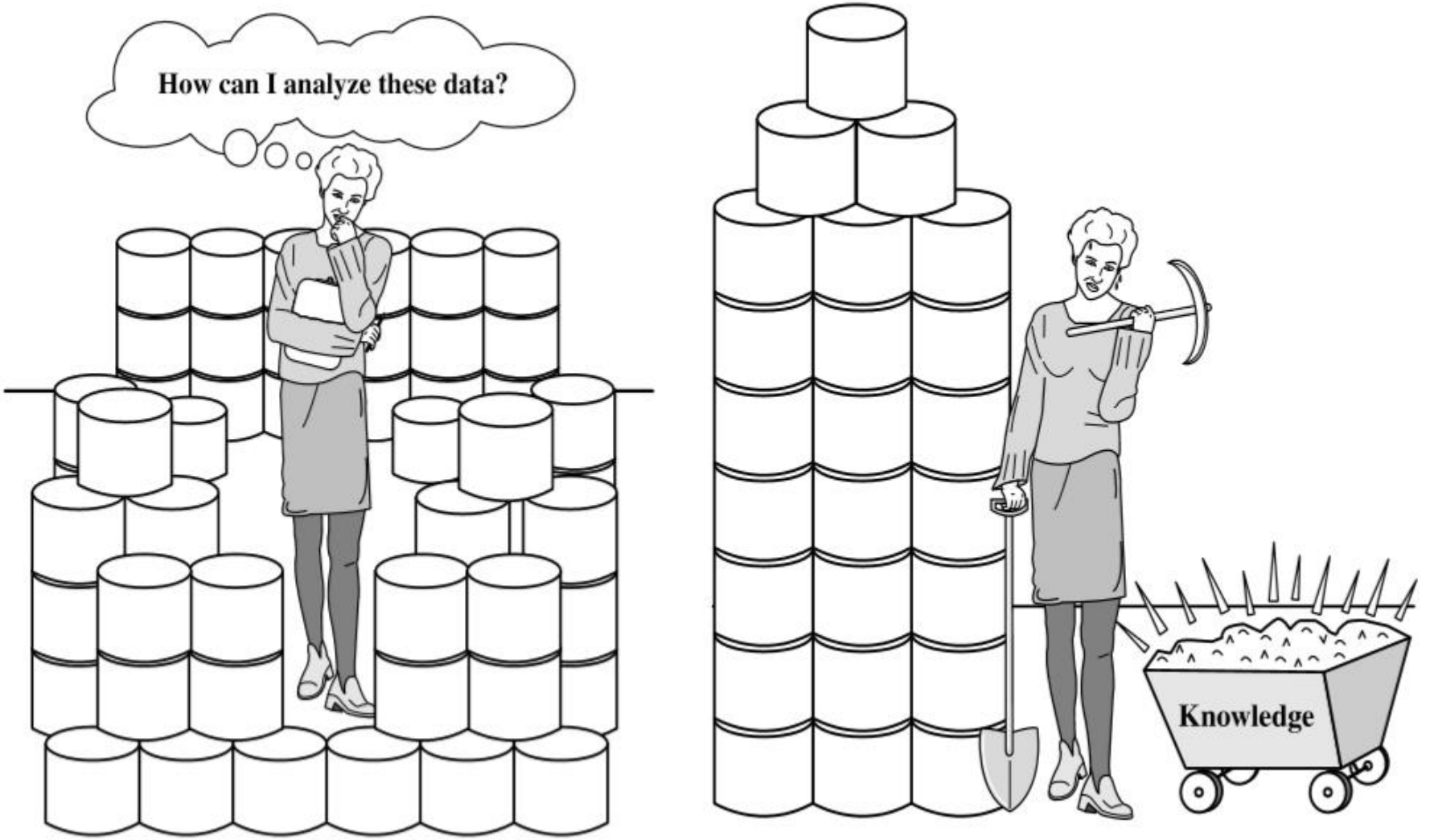
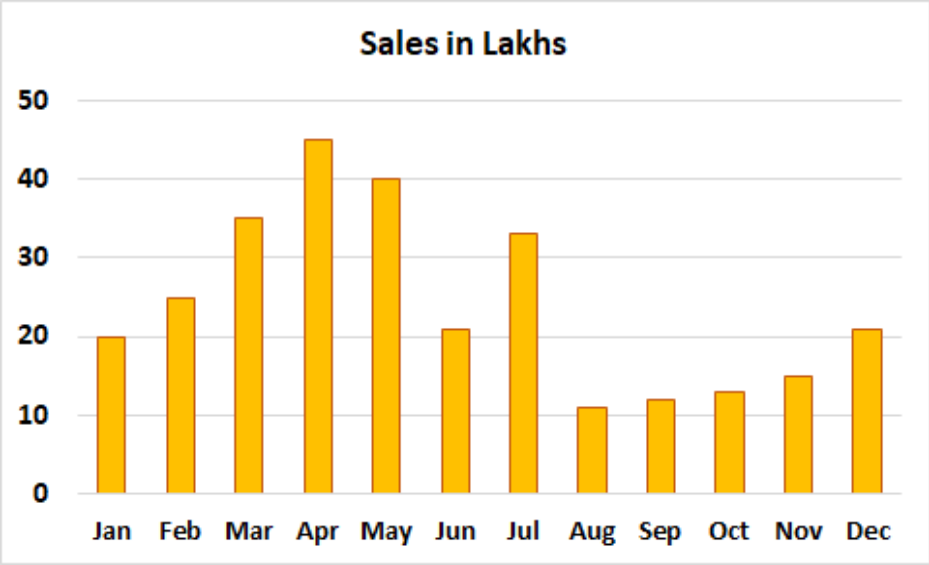
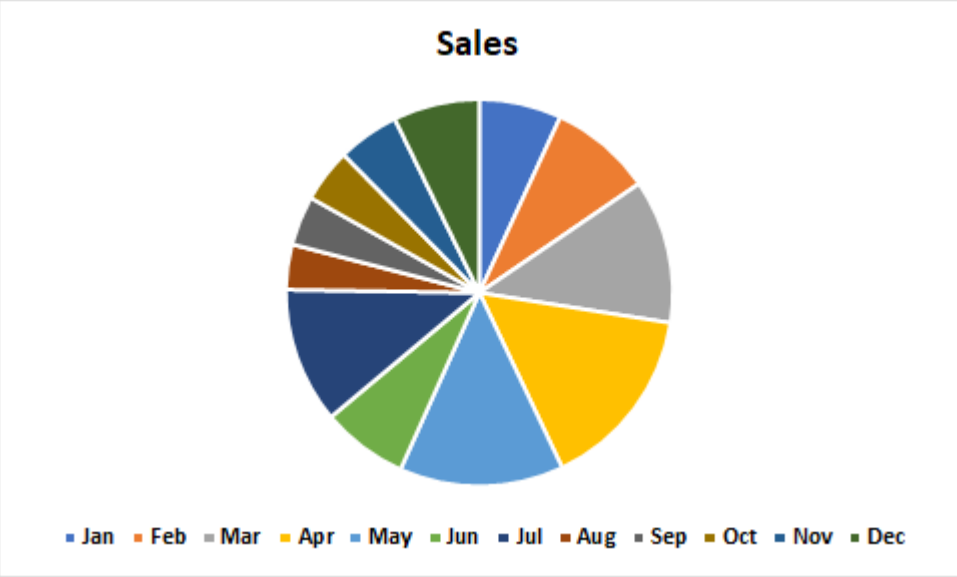
# Knowledge Discovery from Data (KDD)



# Knowledge Discovery from Data (KDD)



# Knowledge Discovery from Data (KDD)

**Step 1: Data Cleaning – to remove noise and inconsistent data**

**Step 2:Data Integration – multiple data sources are combined**

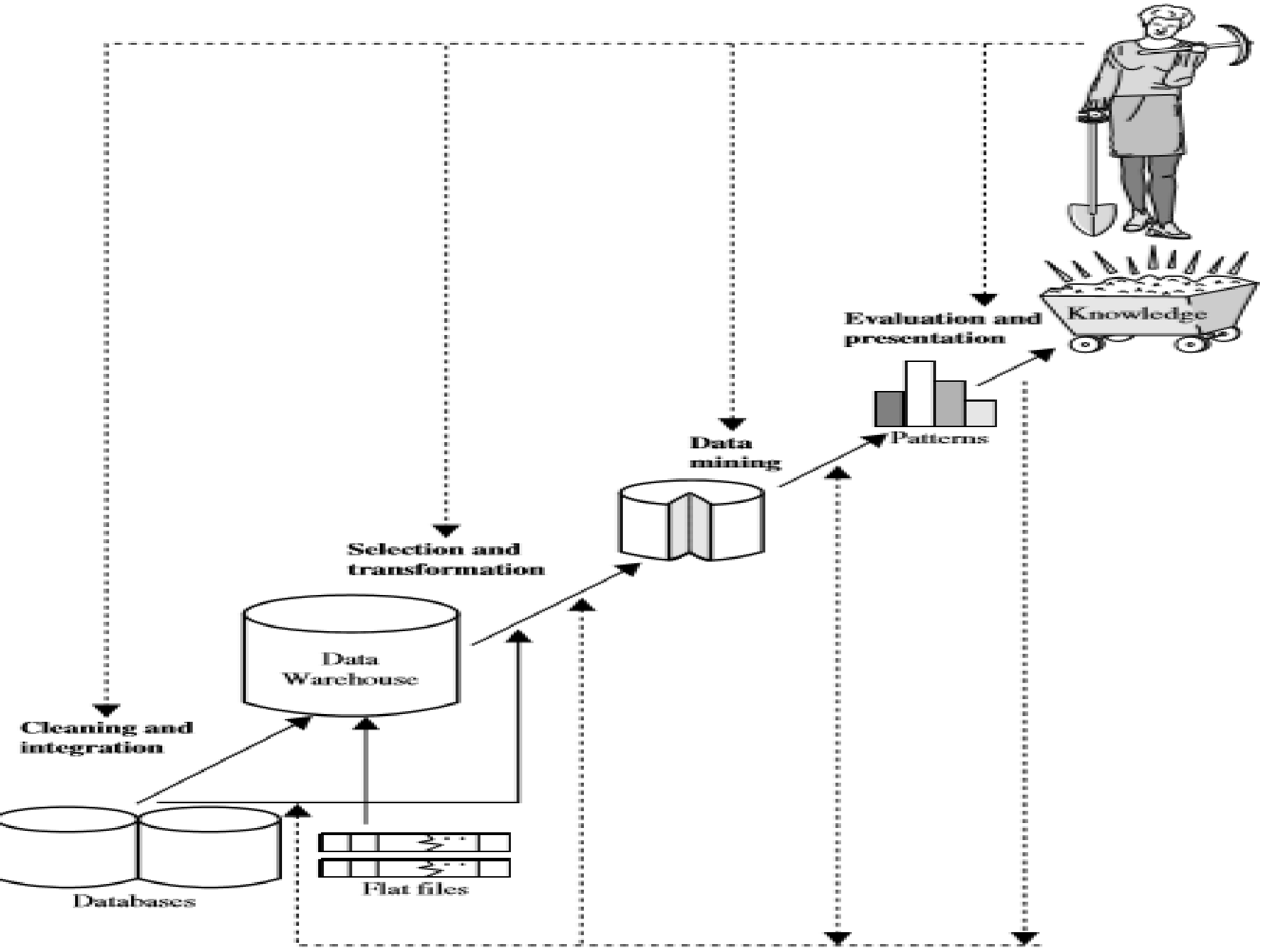
**Step 3: Data Selection – Relevant data of analyst task is retrieved from the database**

**Step 4: Data Transformation – data are transformed and consolidated into forms appropriate for mining by performing summary or aggregation operations**

**Step 5: Data Mining – Intelligent methods are applied to extract data patterns**

**Step 6 : Pattern Evaluation**

**Step 7: Knowledge Presentation**




Data mining as a step in the process of knowledge discovery.

# Kinds of Data


## 1. Data Base Data-

**DBMS**
**Relational Database**
**ER Datamodel**

**Examples:**

| | |
|---|---|
| *customer* | (*cust_ID, name, address, age, occupation, annual_income, credit_information, category, . . .*) |
| *item* | (*item_ID, brand, category, type, price, place_made, supplier, cost, . . .*) |
| *employee* | (*empl_ID, name, category, group, salary, commission, . . .*) |
| *branch* | (*branch_ID, name, address, . . .*) |
| *purchases* | (*trans_ID, cust_ID, empl_ID, date, time, method_paid, amount*) |
| *items_sold* | (*trans_ID, item_ID, qty*) |
| *works_at* | (*empl_ID, branch_ID*) |

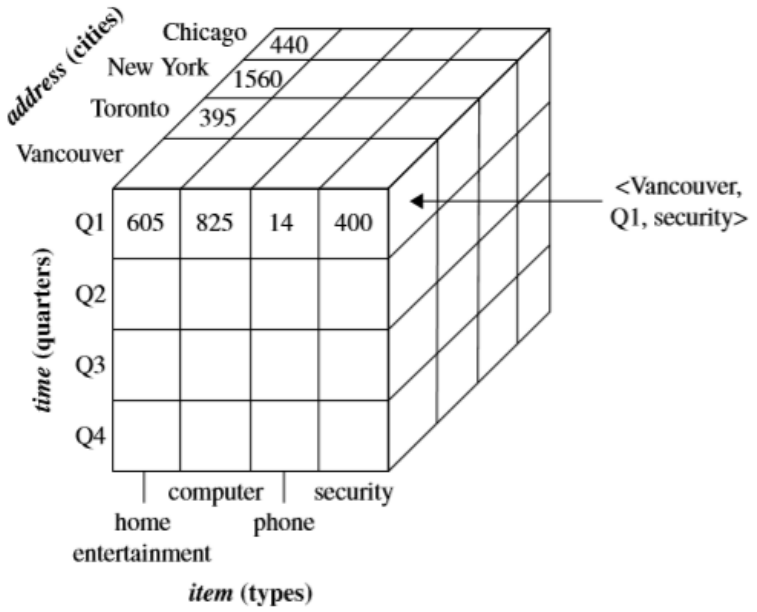
# Kinds of Data (Contd..)

## 2. Data Warehouses-

**is a repository information**
**collected from multiple heterogeneous databases**
**stored under a unified schema, at a single reside**

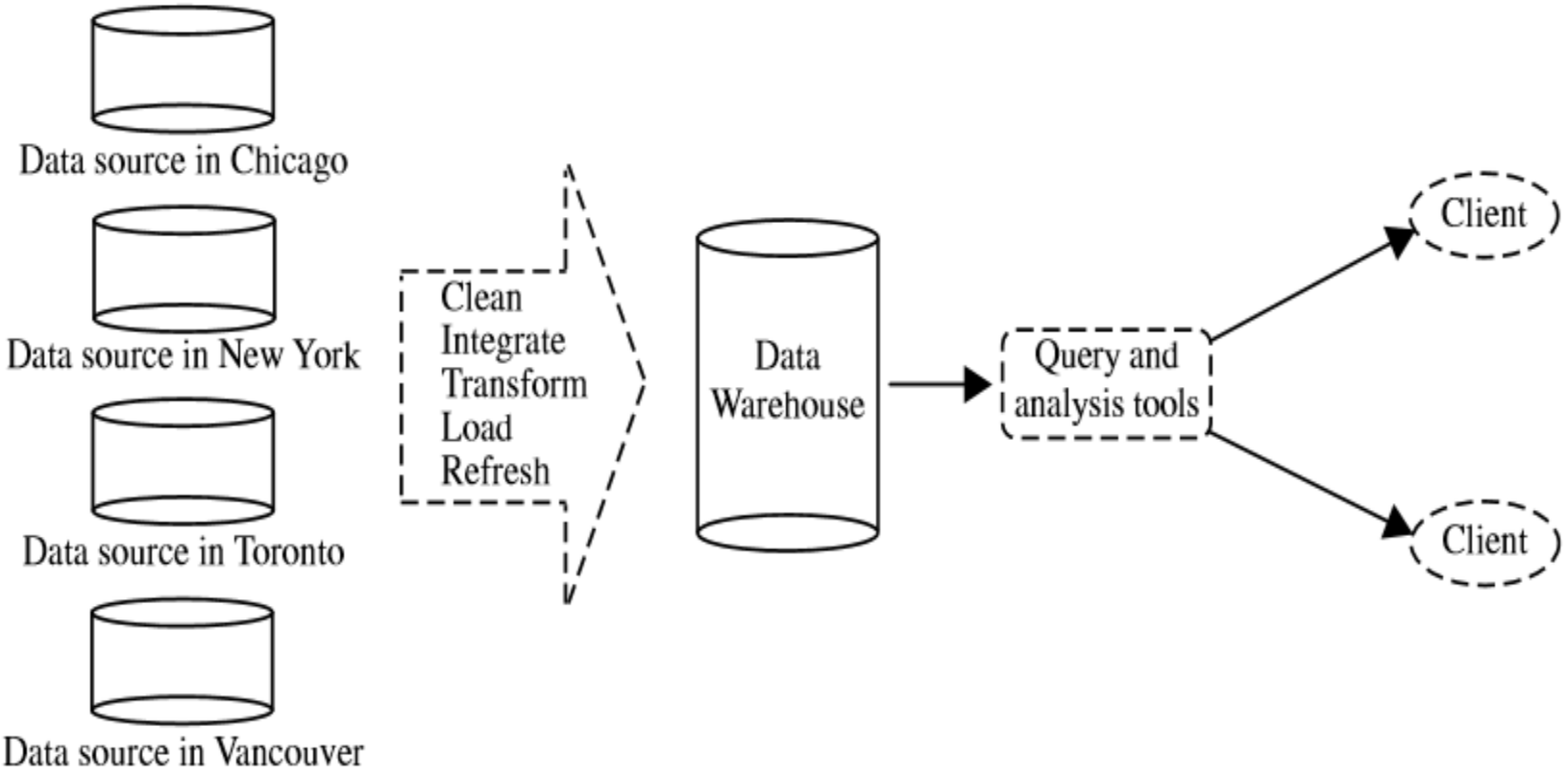
**Example:** **Address Vs Item (2-dimensional data) – relational data**
**Customer Vs Item Vs Supplier (3-dimensional data)- data cube**

| | Itemno | Price | Qty | GST | Invoice no |
|---|---|---|---|---|---|
| Hyd | | | | | |
| Hyd | | | | | |
| BAN | | | | | |
| ..... | | | | | |



# Kinds of Data (Contd..)

## Typical Architecture of Data Warehouse



# Kinds of Data (Contd..)

**Data Cube –** **“ a data warehouse is usually modeled by a multi-dimensional data structure, called a data cube”**

**dimension ?**

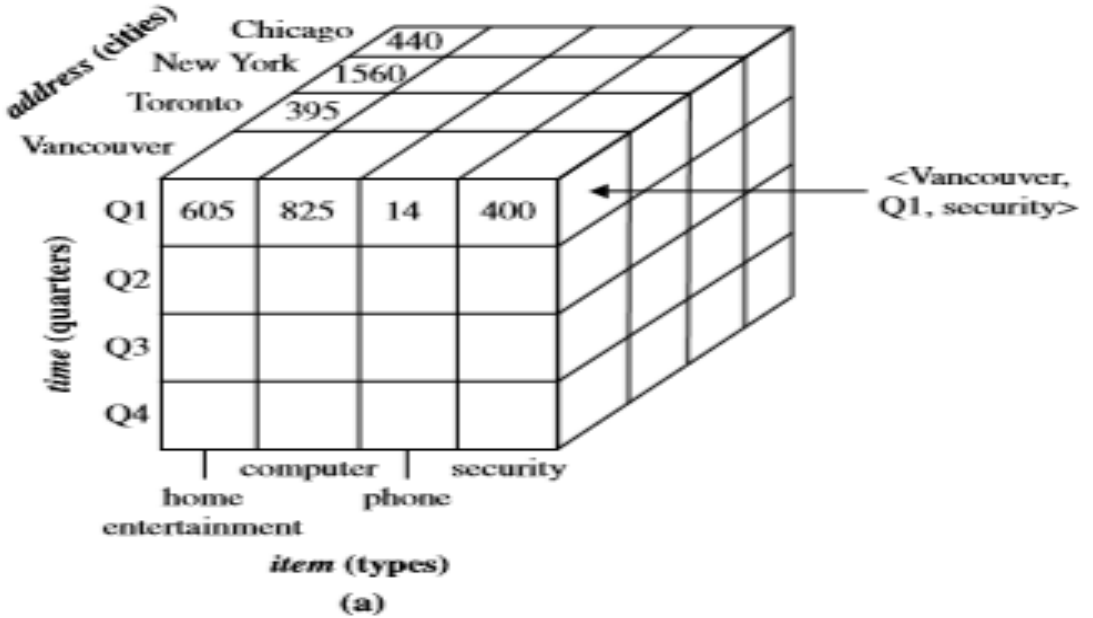
**cell ?**

## Data Warehouses- Query and Analysis Tools

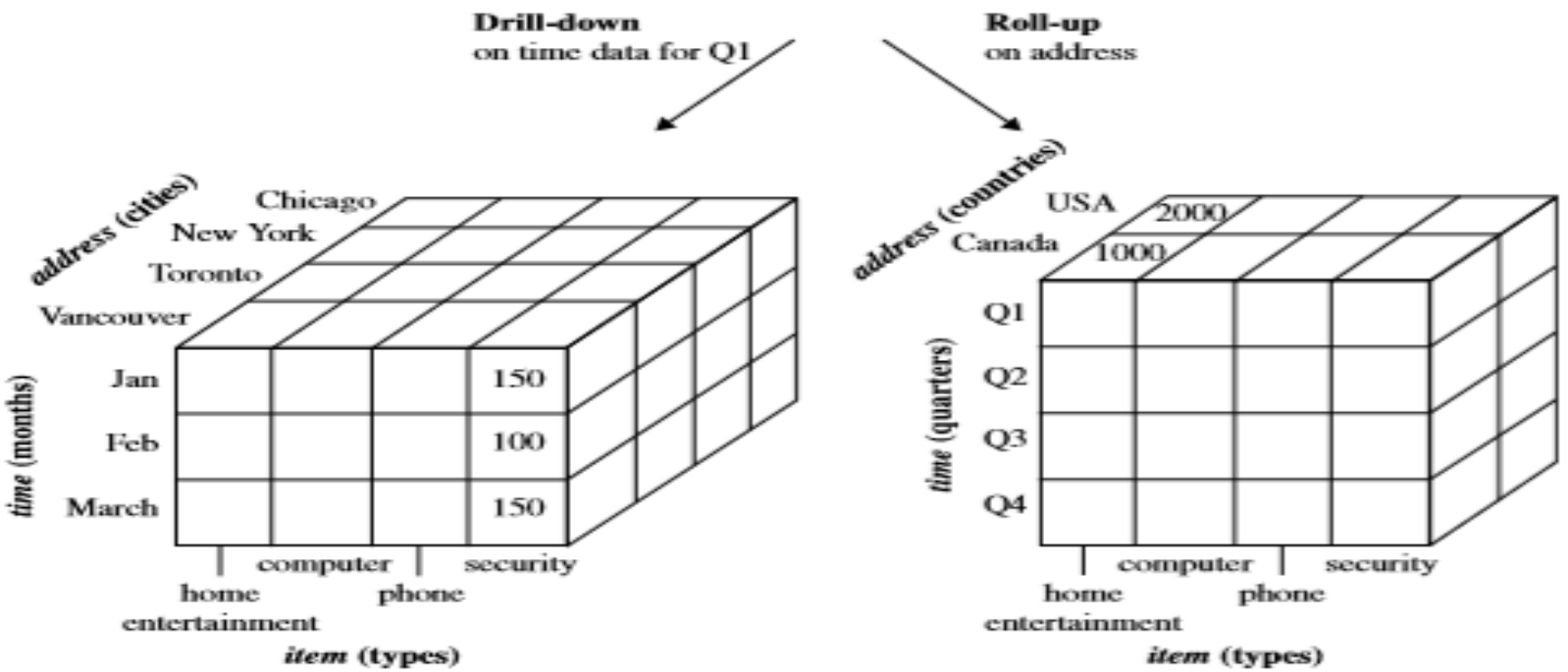
**Online Analytical Processing Operations (OLAP)**

- **Drill down**
- **Roll up**

**(Which allow the user to view the data at different degrees of summarization)**


address (cities)
Chicago
New York
Toronto
Vancouver
440
1560
395
Q1
Q2
Q3
Q4
605
825
14
400
time (quarters)
<Vancouver, Q1, security>
home entertainment
computer
phone
security
item (types)

(a)


Drill-down
on time data for Q1
Roll-up
on address
address (cities)
Chicago
New York
Toronto
Vancouver
Jan
Feb
March
150
100
150
time (months)
home entertainment
computer
phone
security
item (types)
address (countries)
USA
Canada
2000
1000
Q1
Q2
Q3
Q4
time (quarters)
home entertainment
computer
phone
security
item (types)

(b)

# Kinds of Data (Contd..)

**3. Transactional database – "Each record in the transactional database captures a transaction"**

**Transaction – transacation ID and list of items for the transactions**

**Example:**

| *trans_ID* | *list_of_item_IDs* |
|---|---|
| T100 | I1, I3, I8, I16 |
| T200 | I2, I8 |
| ... | ... |

# Possible Questions

1. Define data warehouse?
2. Data cube is a ---------- and it having----------, and,---------
3. Data modeling of dbms is ----------
4. OLAP and OLTP refers to ----------, and ----------
5. What is the use of drill down and roll up operations
6. What is transactional database?

# Thank You