UNIT I

DATA MINING
Data-Types of Data

TYPES OF DATA

- Relational database

- Data warehouse

- Transactional Database

Database Data

- DB system or DBMS – consists of
  - interrelated data, known as database &
  - Set of software programs to manage and access the data

- Software programs provide mechanisms
  - for defining DB structures and data storage
  - For specifying and managing concurrent, shared, or distributed data access
  - For ensuring consistency and security of information stored despite system crashes or attempts at unauthorized access

Database Data

- Relational DB is a collection of tables, each with unique name.

- Each table consists of set of attributes (columns or fields) and large set of tuples (records or rows)

- Each tuple represents object identified by unique key and set of attribute values.

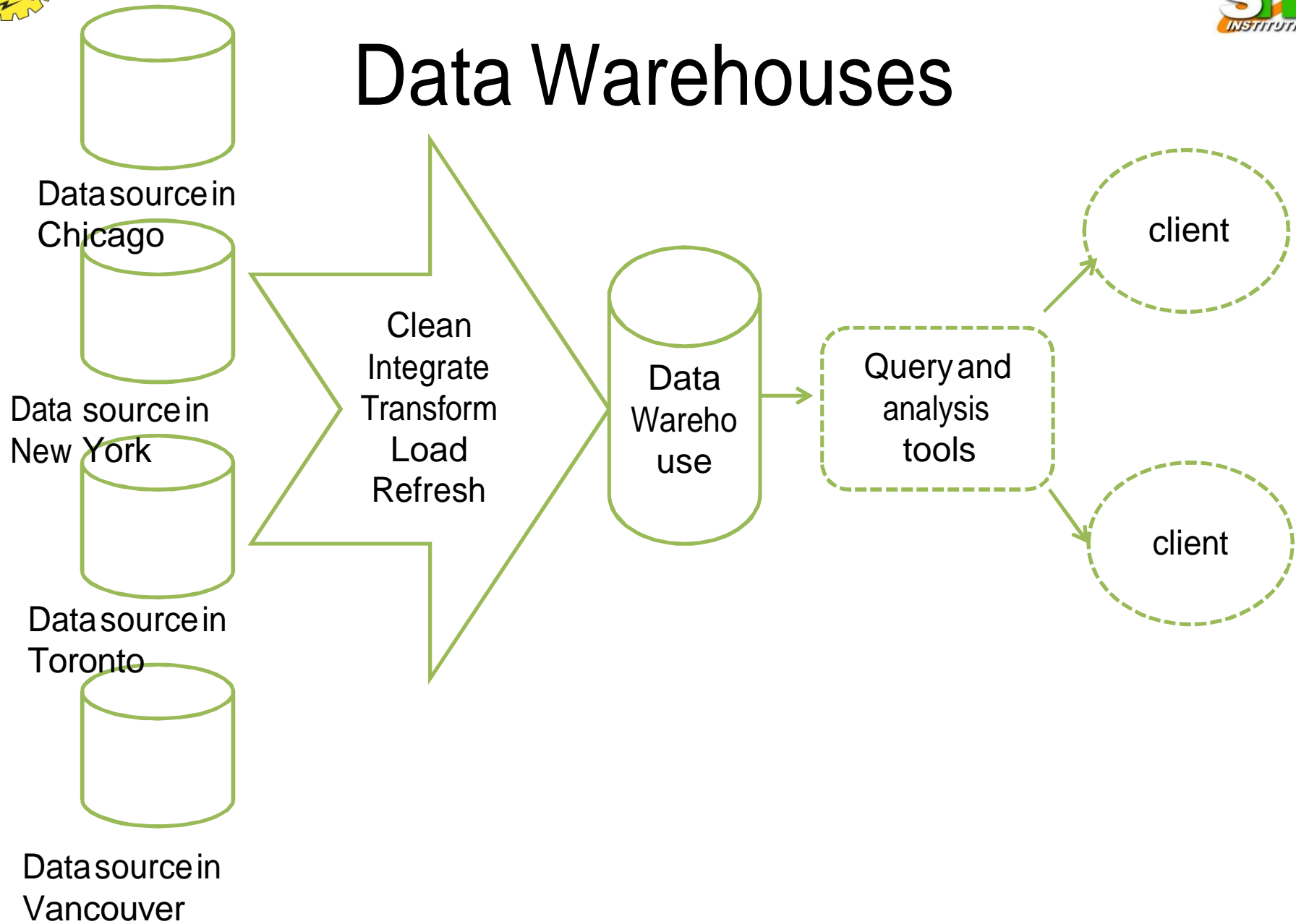- Example: Entity-Relational (ER) Data Model

Data Warehouses

- Repository of information collected from multiple sources, stored under a unified schema, reside at a single site

- Constructed via

- process of data cleaning

- Data Integration

- Data Transformation

- Data loading

- Periodic data refreshing

Data Warehouses

Data source in
Chicago

Data source in
New York

Data source in
Toronto

Data source in
Vancouver

Clean
Integrate
Transform
Load
Refresh

Data
Wareho
use

Query and
analysis
tools

client

client

Data Warehouses

- For decision making,
- Data in warehouse are organized around major subjects
- Eg: customer, item, supplier
- Historical perspective
- Summarized data
- Data cube – each dimension corresponds to attribute

Transactional Data

- Each record in transactional DB captures transaction like customer's purchase flight booking, user's click on a web page)

- It includes unique transaction identity number (trans_ID) and a list of items making up the transaction I.e itemspurchased

- Transactional DB may have additional tables, which contain other information related to transactions like item description, information about salesperson orbranch

Other kinds of data

- Time-related or sequence data (eg: historical records, stock exchange data)
- Data streams (video surveillance, sensor data)
- Spatial data (maps)
- Engineering design data(design of buildings)
- Hypertext and multimedia data (text, image, audio, video)
- Graph and networked data(social & information networks)
- Web (internet)

Thank you