

Business Intelligence

Unit 2

* What is business intelligence?

→ an umbrella term that covers the processes and methods of collecting, storing and analyzing data from business operations to optimize performance.

→ Business Intelligence combines:-

(i) business analytics

(ii) data mining

(iii) data visualization

(iv) data tools and infrastructure

(v) best practices to help orgs make better decisions.



* BI Methods

→ Data mining - uncover trends in large datasets

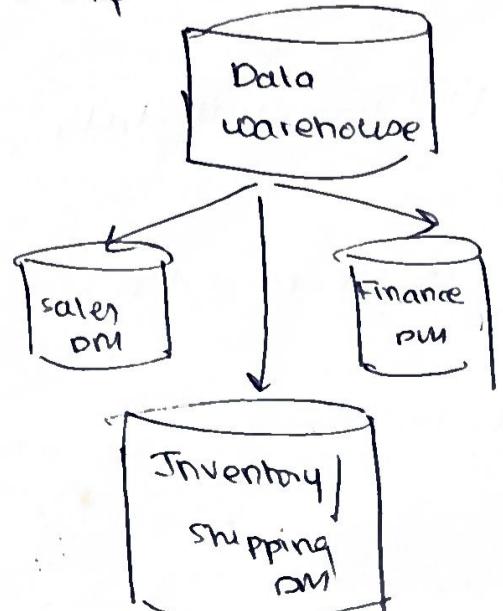
use co/
data warehousing

→ Descriptive analytics

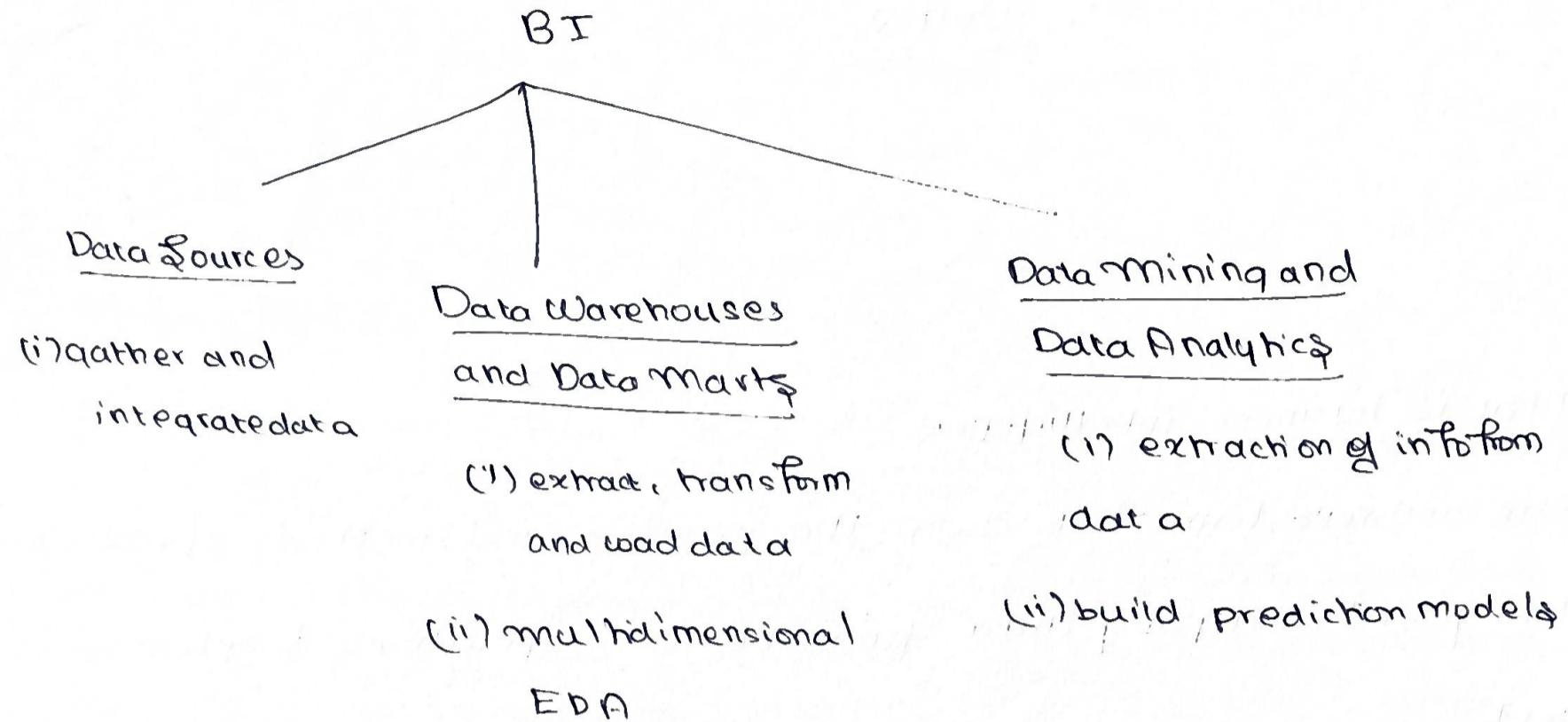
→ Statistical analytics

→ Data visualization

→ Data preparation



* BI Architecture



eg Building a telecom customer retention model — given a customer's telecom behavior, predict if the customer will stay or leave.

* Data Warehouse

→ repository of the data available for BI and decision support systems

→ can be:

(i) internal data
→ back office : transactional records, orders, invoices
→ front office : call centers, sales office
→ web-based : sales transactions on e-commerce web sites

(ii) external data

→ market surveys, GIS systems

(iii) personal data

→ data about individuals

(iv) meta data

→ data about dataset

* Data Marts

: a subset of a data warehouse for one function - eg. marketing

BI vs. BA

BI : uses historic data - descriptive analysis

BA : predictive analysis

BI : how processes affect KPI

BA : why processes affected KPI

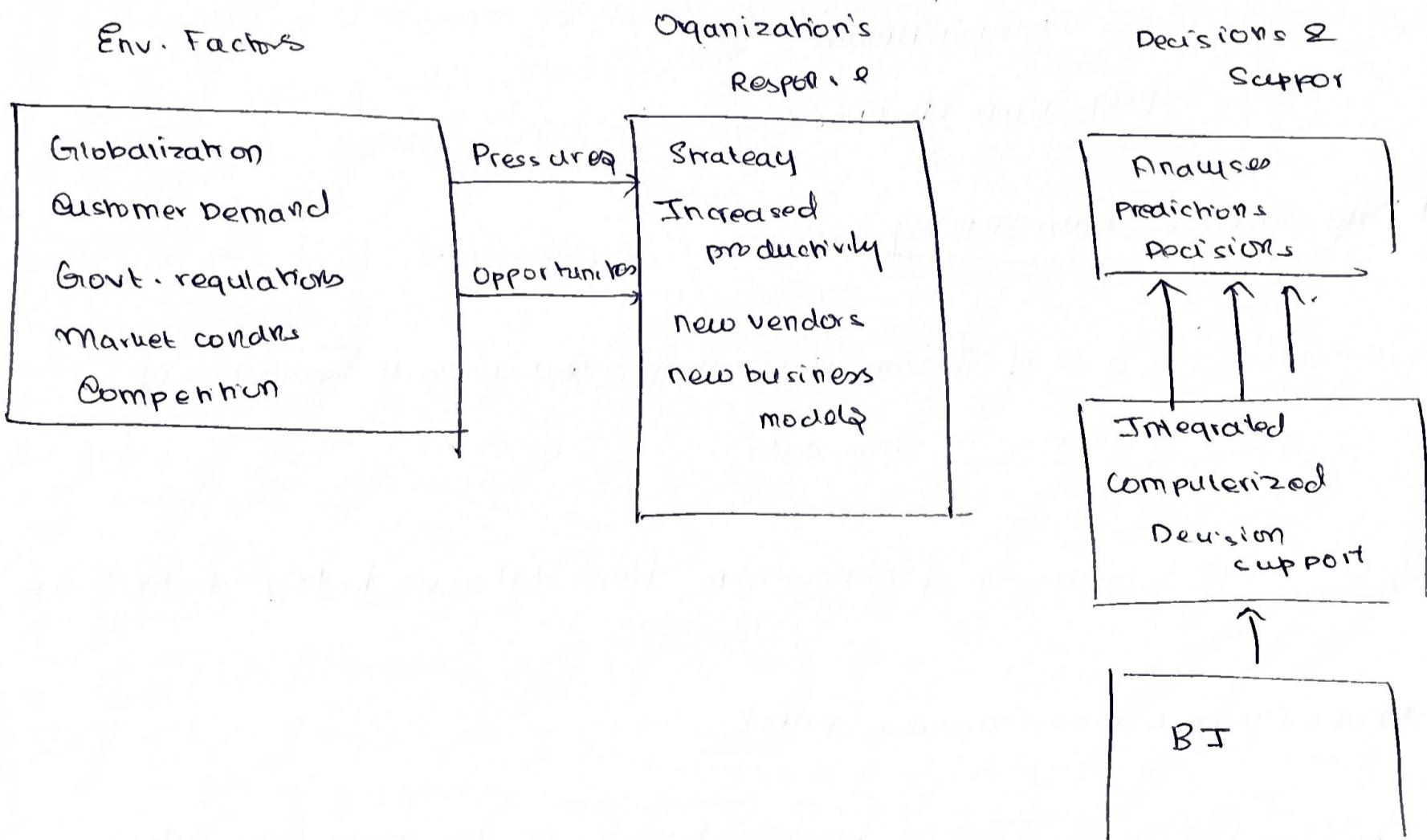
BI : less mathematical

BA : more mathematically advanced

OLAP
online analytic processing

Slicing & dicing data
stored in a dimensional
format, drilling down in
greater detail & aggregating
data

Business Pressures - Responses - Support Model



* Benefits of BI

- (i) Increased profitability
- (ii) Decreased costs
- (iii) Improved customer - relationship mngt
- (iv) Decreased risk

Some fields that BI can improve are:

- (i) Customer Analysis
- (ii) Behavior Analysis
- (iii) Human Capital Productivity Analysis
- (iv) Business Productivity Analysis
- (v) Sales Channel Analysis
- (vi) Supply Chain Analysis

* Data Science vs. Data mining

Data Science : a set of fundamental principles that guide the extraction of knowledge from data

Data mining : extraction of knowledge from data via tools / technologies

* Data Driven Decision - Making (DDD)

DDD - practice of making decisions based on the analysis of data rather than intuition

Type 1 decision - discover something new in data

Type 2 decision - repeat decisions at massive scale (automatic decision making)

* Decision Support Systems

software - eg. Check Box

Yonex

XLStat

QlikView

5

- DSS = an interactive information system that analyzes large volumes of decisions.
- DSS leverages a combination of raw data, documents, personal knowledge, business models.
- can include relational datasources, cubes, data warehouses, sales projections etc.
- Decision Intelligence updates / reinvents decision support systems

Purpose of DSS

- analyzes data rather than collect data
- integrate and synthesize multiple variables to project the likelihood of various outcomes

Examples

Route optimization

Crop Planning

Clinical DSS

ERP Dashboards

DSS vs BI

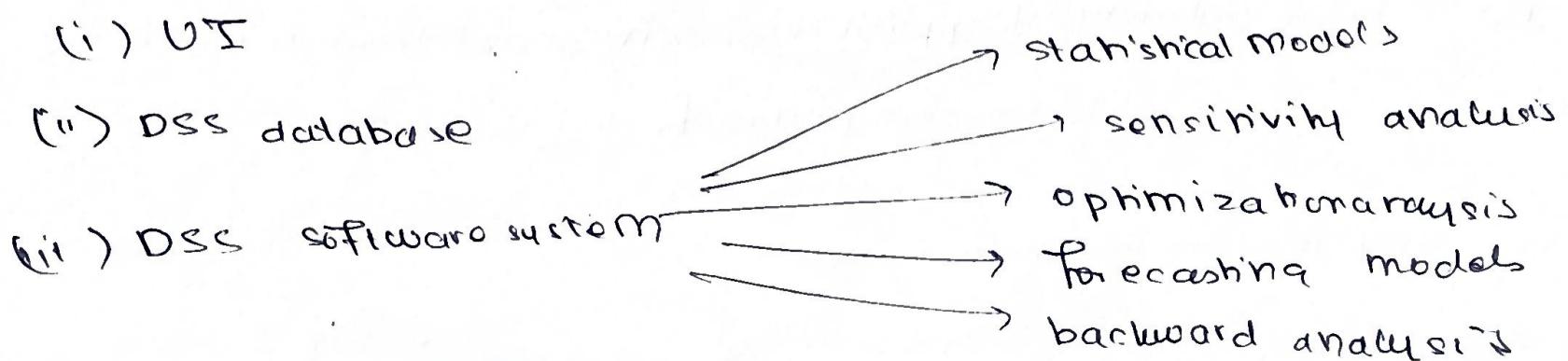
BI - broad category of applications, services and technologies for gathering, storing and accessing data for decision-making

DSS - only for decision-making

* Types of DSS

- (i) Data driven DSS - manage & manipulate large DBs of structured data
- (ii) Model driven DSS - use accounting & financial models
access & manipulate a model
- (iii) Knowledge driven DSS : suggest / recommend actions to managers
- called advisory system
- provides specialized problem-solving expertise
- used for classification, configuration & diagnosis
- (iv) Document-driven DSS - integrate storage & processing technologies for document retrieval and analysis
- e.g. search engine
- (v) Communication-driven DSS -
& group DSS
Focus on communication, collaboration & coordination
- GIDSS - support groups of decision-makers

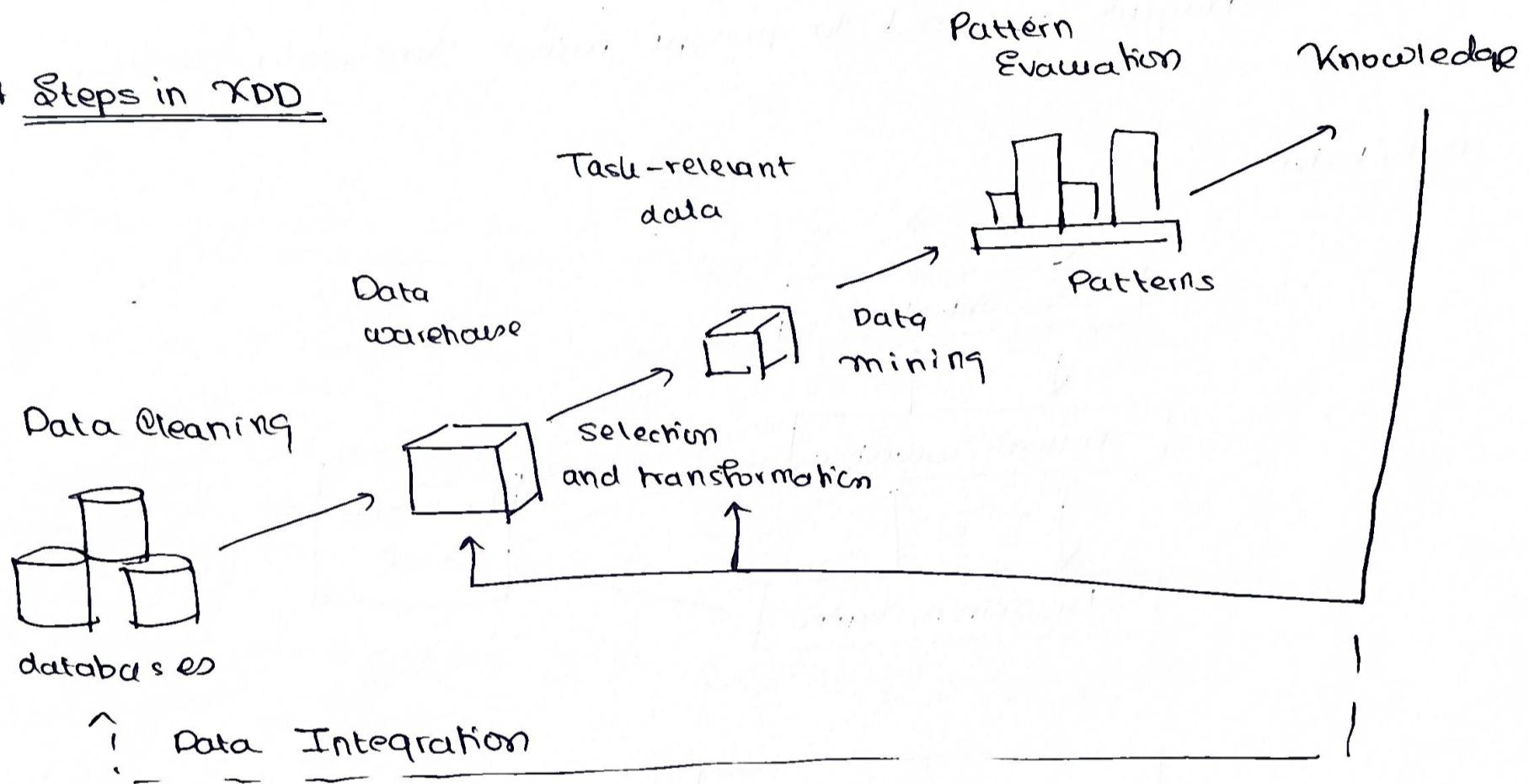
* Components of a DSS



* Data Mining / KDD - Knowledge Discovery from Data

→ extraction of non-trivial, implicit, previously unknown and potentially useful patterns or knowledge from a huge amount of data.

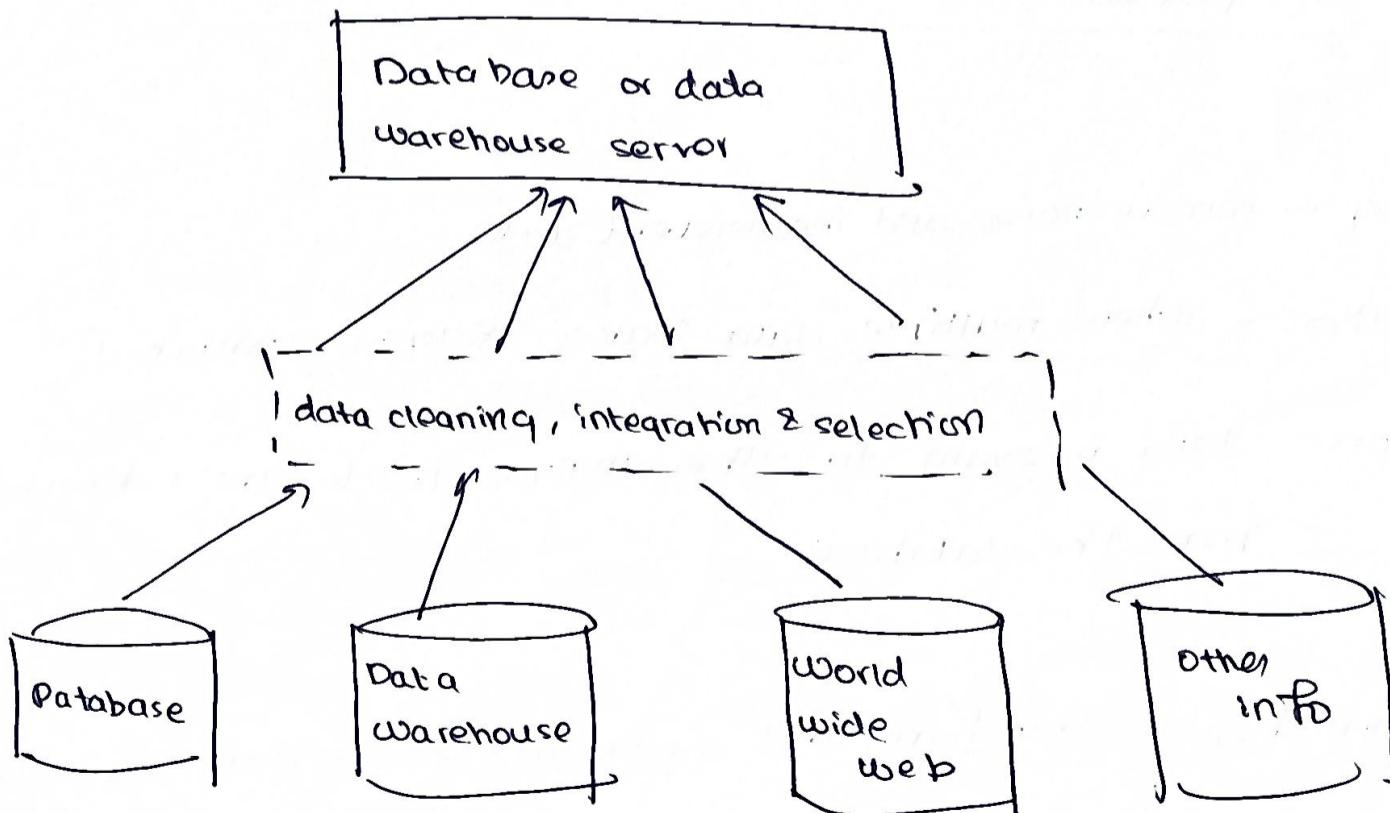
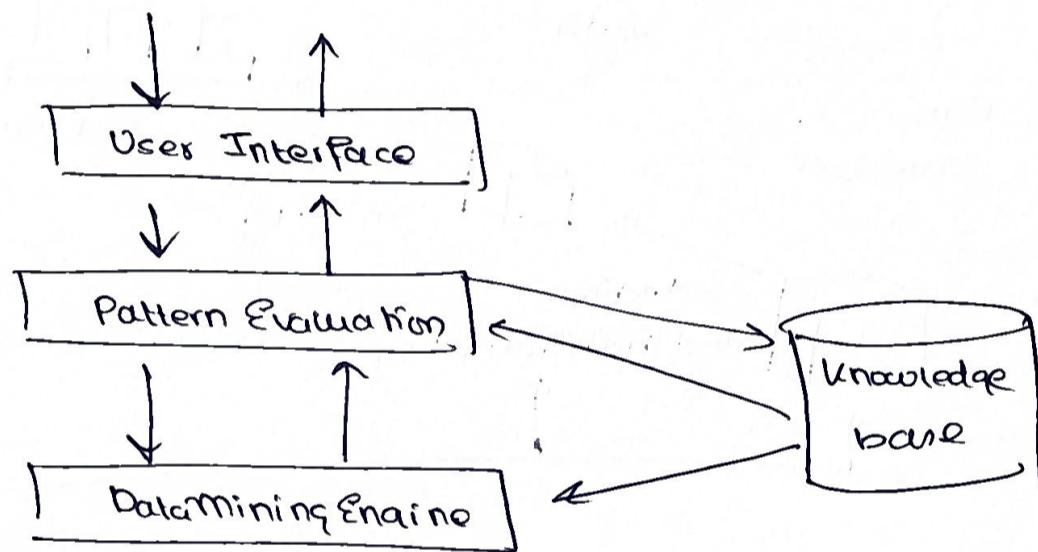
* Steps in KDD



- (i) Data Cleaning - remove noise and inconsistent data
- (ii) Data Integration - where multiple data sources may be combined
- (iii) Data selection - data relevant to the analysis task are retrieved from the database
- (iv) Data Transformation - Transform and consolidate into forms appropriate for mining by performing summary or aggregation operations

- (v) Data Mining - intelligent methods are applied to extract data patterns
- (vi) Pattern Evaluation - identify interesting patterns
- (vii) Knowledge Presentation - visualization and knowledge representation techniques are used to present mined knowledge to users

Architecture



* KDD from an ML and Statistics POV

Data Preprocessing

- (i) data integration
- (ii) normalization
- (iii) feature selection
- (iv) dimension reduction

Data Mining

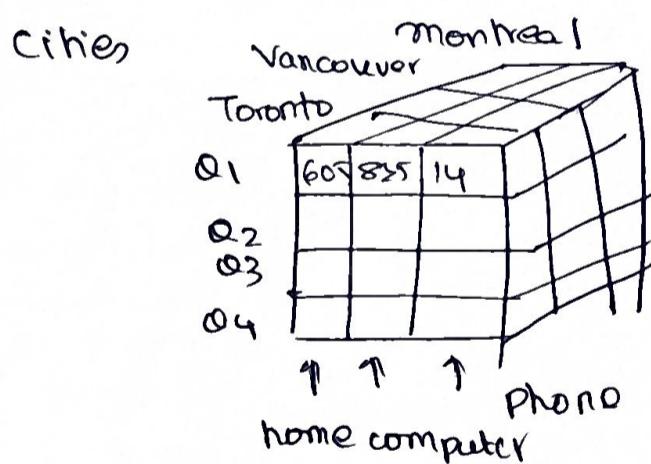
- (i) clustering
- (ii) Outlier Analysis
- (iii) Pattern Discovery

Post-processing

- (i) pattern evaluation
- (ii) pattern selection
- (iii) pattern interpretation
- (iv) pattern visualization

* Multidimensional View of KDD

- modelled using a data cube
- helps in fast access of summarized data
- each dimension corresponds to a set of attributes or an attr.
- each cell stores value of an aggregated measure



* Tasks in Data mining

Descriptive → derive patterns which summarize the underlying relationship

Predictive Tasks - Perform induction on the current data in order to make predictions