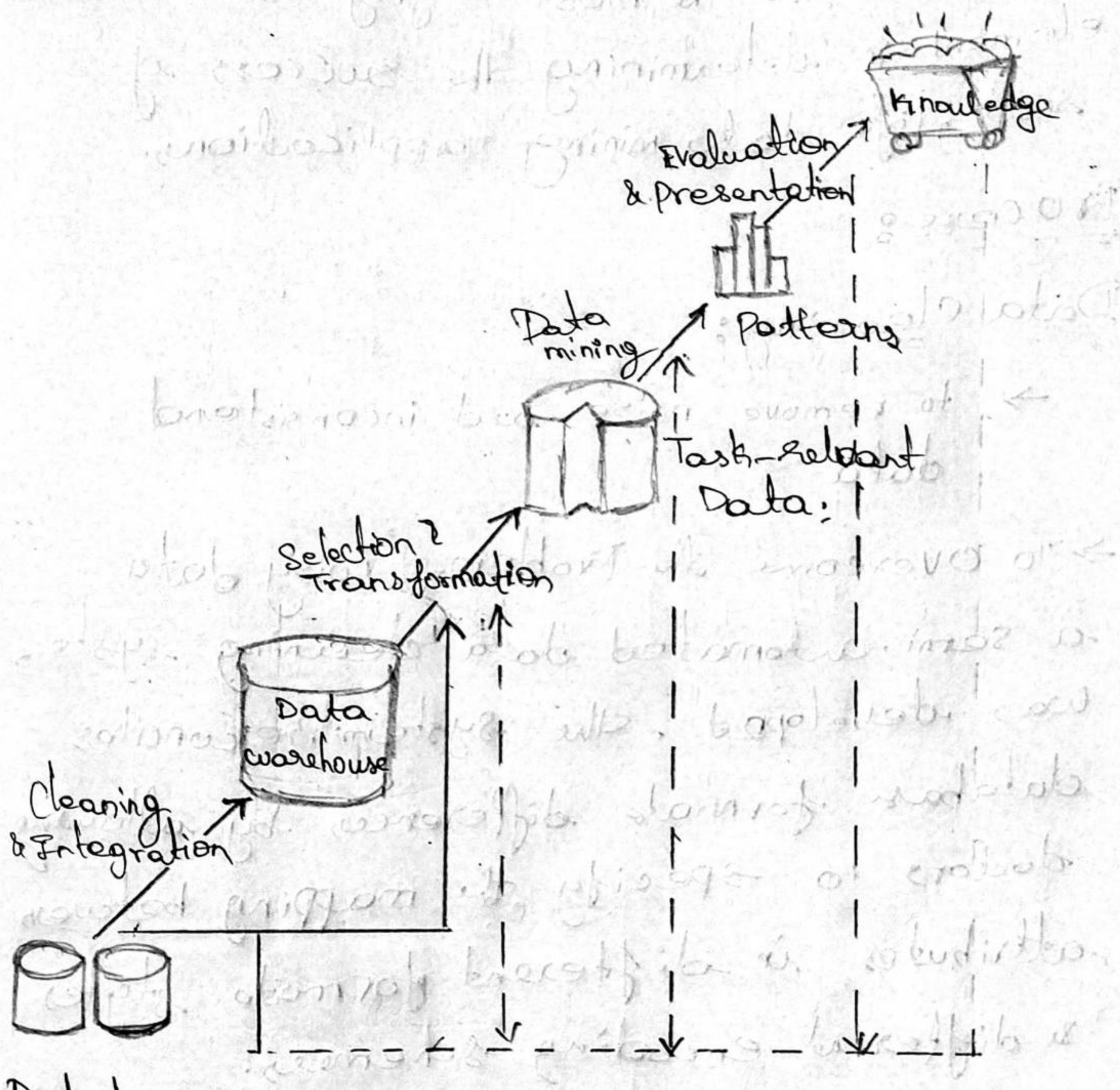
G.NITHISH

19BCS0012

DATA MINING

## Diagram of KDD



Example:. Medical Application: Diabetic Screening:

* Pre-Processing and Post-processing steps can often be the most significant & critical element in determining the success of real-life data-mining applications.

Process:

① Data cleaning:

→ to remove noise and inconsistend data

→ To overcome the problem of noisy data a semi-automated data cleaning system was developed. the system reconciles database formate differences by allowing doctors to specify the mapping between attributes in different formate styles & different encoding schemes.

* The formal difference were reconciled the Problem of identifying & removing duplicated records was addressed.

(2) Data integration :-

→ It defines the heterogeneous data from multiple sources combined in a common source.

* In the hospital diabetic patient data is scattered in hospital Database.

(3) Data selection :-

→ Data selection is defined as the process where data relevant to the analysis is decided and retrived from the data collection.

* Find the diabetic patient info from a specific area or catogory of hospital managment DBMS.

(4) Data Transformation :

* Data Transformation is defined as the Process of transforming data into appropriate from required by mining Procedure :-

It has two step procors :-

- Data Mapping : Assigning elements from Source base to destination to capture transformations
- Code generation : Creation of the actual tranform programe.

5) Data mining:-

Data mining is defined as deliclover techinques that are applied to extract Patterns potentially Usefull.

→ Transforms task relavent data into Patterns.

→ Decides Purpose of model using Classification or characherization.

6.) Patteren Evaluation:-

* To identify the truly interesting Patterns representing knowledge based on some interestingness measures; &

7.) knowledge presentation:-

* When visualization and knowledge representation techniques are used to present the mined knowledge to the user.