# Assignment A1 #1403

- Title: Design star/snowflake schemas for analyzing the process

- problem statement: for an organization of your choice, choose a set of business processes. Design star/snowflake schemas for analyzing these processes. Create a fact constellation schema by combining them. extract data from different data sources apply suitable transformation & load into destination tables using an ETL tool. for e.g. business organizations: Sales order.

- Learning objective: To understand star/snowflake schema
  To understand ETL tools
- Learning outcomes: Students will be able to use ETL tool to design and also be able to analyze different schemas (snow, star, etc)

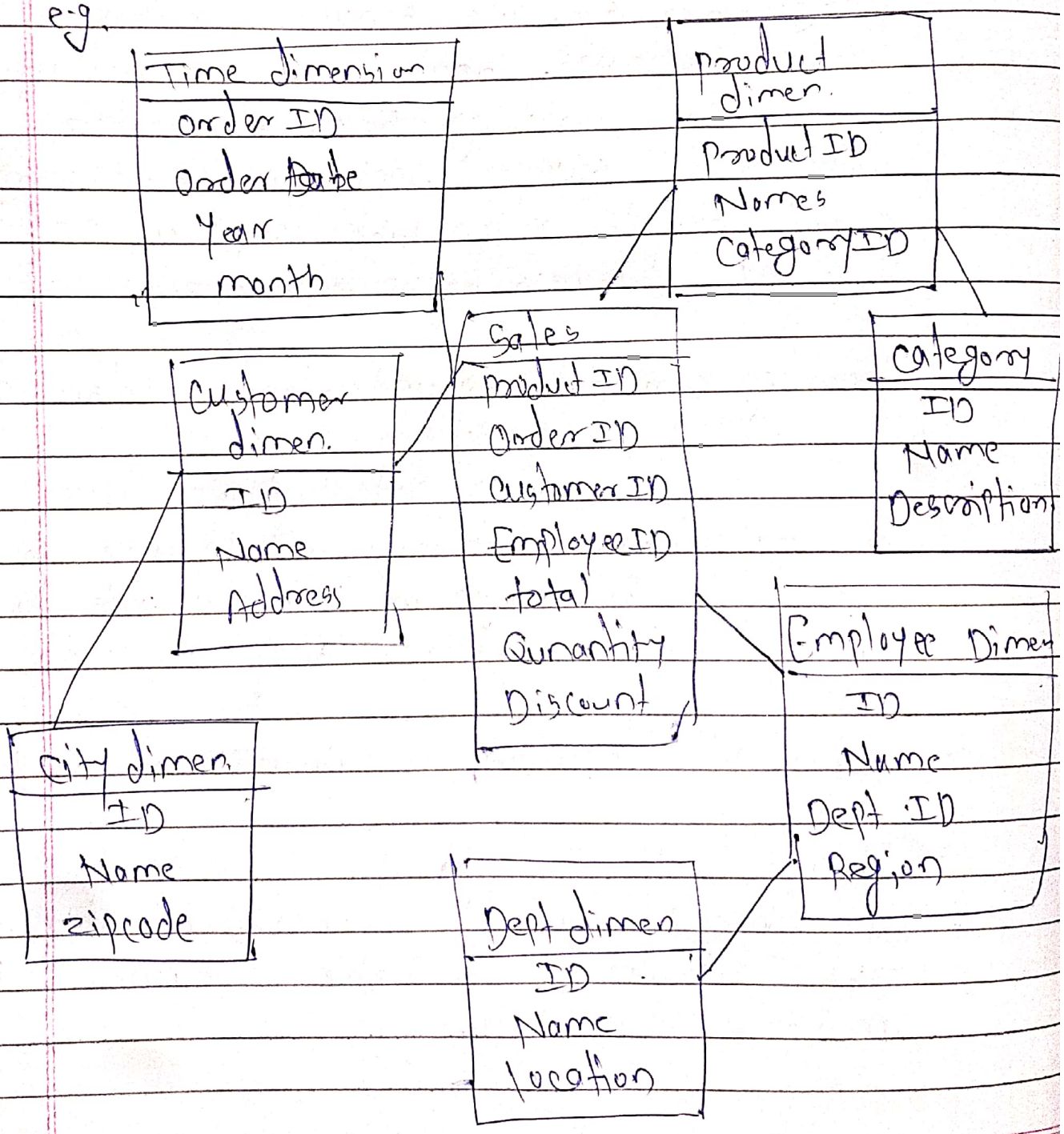
- S/w H/w requirements: An ETL tool eg Pentaho

- Theory:
- Snowflake Schema: It is variant of star schema. The centralized fact table is connected to multiple dimensions.

41403

In the snowflake schema, dimensions are present in normalized form in multiple ~~related~~ tables. This effect affects only the dimension tables and not the fact tables

e.g.



CA1403

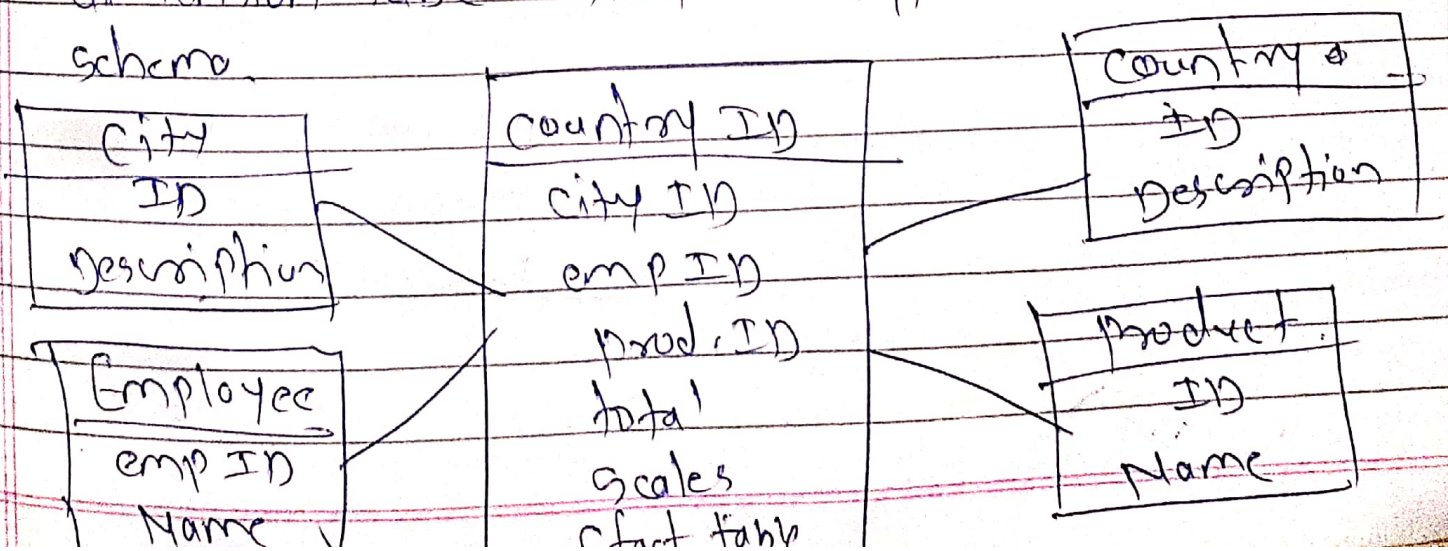
Dimension are maintained in normalized form to reduce dependency. Tables are easy to maintain & some storage space

- Characteristics
  - use small disk space
  - easy to implement
  - performance reduces due to multiple tables

- Advantages
  - provides structur data which reduces the problem of data integrity
  - highly structured data reduces disk space

- Disadvantages
  Hierarchy should belong to the dimension table only & should never be snowflaked

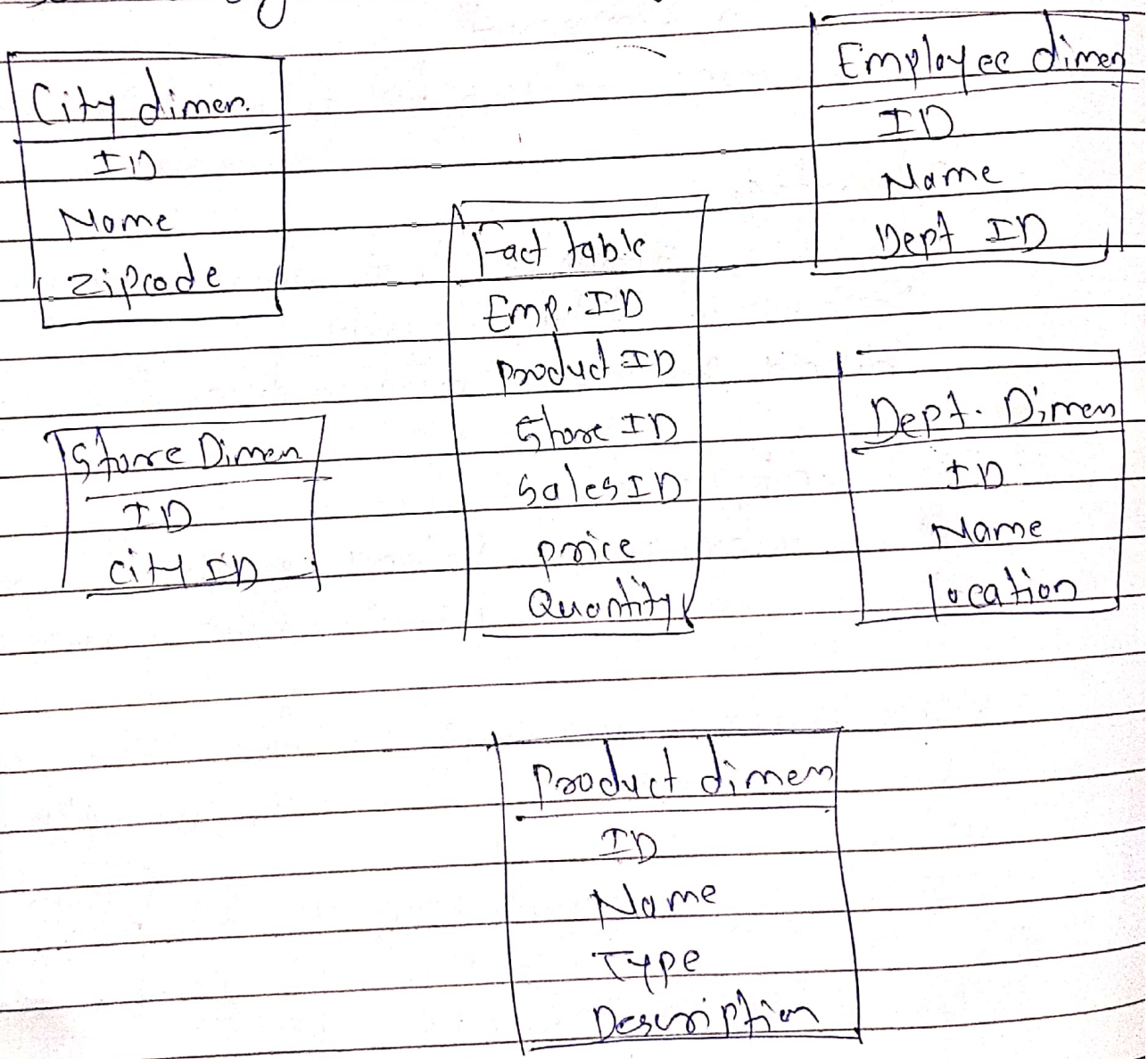
- Star schema : can have one fact table and many dimension table. Simplest type of data warehouse schema.



4403

- characteristics:
  - only one fact table
  - Dimension table are not joined to each other
  - easy to understand

- Selected organization : college snow flake schema



41403

- ETL tools:
  - ETL stands for extract transformation tool
  - Extract means extracting the data from heterogenous or homogenous sources into our environment for integration & generate insight from it.
  - ETL tools extract data from different sources (tables, flat files, etc) & process this data.
  - Transformation phase data is cleaned according to need. Data can be trimmed, appended, filtered, joins to be generated, etc.
  - In the bad phase, final data is loaded into the targed DB. or into flat files or in the form of web service.

- conclusion: Thus, I understood the different schema used (star, snowflake, etc) used Pentaho tool for performing translation