

# CSE606 - DATA WAREHOUSE DECISION TOOL

## INFORMATION MODEL TO STAR SCHEMA

---

**PROBLEM STATEMENT** - Given an Information model we have to convert it to Star Schema

**METHODOLOGY** - We take input from a CSV file which contains multiple rows, each row represents one Information. Each row consists of Facts and Dimensions and their attributes and measures. We read those respective values and form a star schema connecting Facts with their dimensions. This is repeated for each Information and when our algorithm terminates it gives the desired Star Schema as the output.

**TECHNOLOGY USED** - We used Python language to code the algorithm and convert our input ( Information Model ) to output the Star Schema. For GUI we have used Flask and HTML.

**INPUTS** - Information Model

### Informational Model to Star Schema

---

This decisional tool converts an informational model consisting of information such as facts, categories, attributes, computed from data to a well-structured star schema with links between facts, dimensions and their attributes. The star schema is the simplest type of Data Warehouse Schema. It is known as a star schema because its structure resembles the shape of a star. In the star schema, the centre contains a fact table and the pointed corners denote the dimensions of the informational model.

Input the informational model in CSV file format as specified in this [sample file](#)

No file chosen

**Group Details:** Ayushi Shrivastava (2016025), Meghna Gupta (2016056), Tanish Gupta (2016106)

## OUTPUTS - The desired Star Schema

### Star Schema

---

#### Facts

**Fact:** Time Spent for Consultation (Time of registration, Time of consultation, Department\_key, Unit\_key, Patient\_key, date\_key, )  
**Fact:** Distribution of Personnel (Number of patients, Doctors, Nurses, Record Clerks, Department\_key, Unit\_key, date\_key, )

---

#### Dimensions

**Dimension:** Date (date\_key, Daily, 5 years, inserted\_timestamp, updated\_timestamp, inserted\_timestamp, updated\_timestamp, )  
**Dimension:** Unit (name, Unit\_key, inserted\_timestamp, updated\_timestamp, inserted\_timestamp, updated\_timestamp, )  
**Dimension:** Department (name, Department\_key, inserted\_timestamp, updated\_timestamp, inserted\_timestamp, updated\_timestamp, )  
**Dimension:** Patient (rn, name, age, gender, Patient\_key, inserted\_timestamp, updated\_timestamp, inserted\_timestamp, updated\_timestamp, )

---

#### Mappings

Department --> Time Spent for Consultation  
Unit --> Time Spent for Consultation  
Patient --> Time Spent for Consultation  
Date --> Time Spent for Consultation  
Department --> Distribution of Personnel  
Unit --> Distribution of Personnel  
Date --> Distribution of Personnel  
Unit --> Department

## Group Members -

1. Ayushi Srivastava
2. Meghna Gupta
3. Tanish Gupta