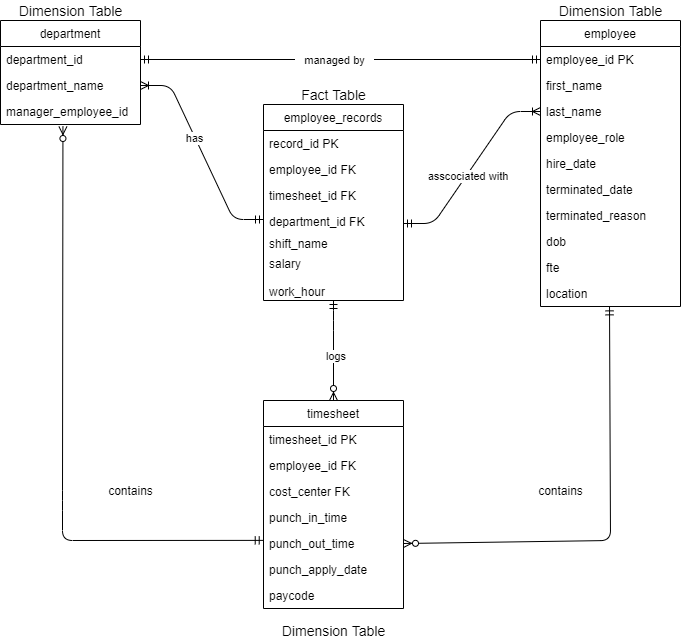
Data Warehouse

Assignment #1

Prepared by Nischal Badal

**Identifying fact table and dimension table from given dataset and creating logical and physical model of the data warehouse**

In our given data set, there are data of employee and timesheet in weeks. To create the data warehouse, we need to classify them into fact table and dimensions table. The proposed data warehouse logical structure for the mentioned requirements could be:



*Fig: Logical ER Model of Data Warehouse*

Now let us see how this ER model satisfies our requirements:

1. **Clients should be able to know if an employee was working on a particular day or not.**
2. **If they worked,**
3. **What time did they start and left?**
4. **How many hours?**
5. **Were they charge on the day?**
6. **If they didn’t,**
7. **Were they on call?**

**Answer:**

In our data warehouse, we have timesheet as dimension table where we can filter the punch\_in\_time, punch\_out\_time and work\_hour. Also, there is a column called paycode which determines if they were charge or not.

Similarly, the CALL paycode on the timesheet table determines of an employee was on call or not in a particular day.

1. **Clients should be able to know if the employee had a Morning (Starting between 5: 00 AM - 11:00 AM) or Evening (Starting after 12:00 PM) shift.**

**Answer:**

Similarly, there is a shift column in the fact table which can be calculated as Morning (Starting between 5: 00 AM - 11:00 AM) or Evening (Starting after 12:00 PM) shift based on the maximum punch in time of any specific date.

This can help to identify if the employee had a morning or evening shift.

1. **Clients should be able to know if the employees are working regularly on a weekend (SUN, SAT)**

**Answer:**

This can be done by converting punch\_apply\_date to char in postgresql which gives the day name and filtering whether the employee has worked on SUN or SAT. If yes, we can sum the total number of worked days in weekend.

1. **Clients want to analyze if any employee has to cover for other team members regularly.**

**Answer:**

For this we can filter out the employees in same department by department id and we have the CHARGE column in paycode, which

1. **Clients want to analyze the data on a biweekly basis starting from 2021-01-01**

**Answer:**

The records in the timesheet dimension table are stored as per day. It can be grouped as two weeks to obtained the data on a biweekly basis.

1. **Clients want to analyze the data based on the employee role.**

**Answer:**

There is a column called employee\_role in employee dimension table. This can be joined with employee\_id of the fact table and can be analyzed based upon the employee role.

1. **Clients want to analyze the salary distribution by department.**

**Answer:**

The salary is dependent up on the department id and employee\_role. The salary and department\_id both are in the facts table and hence it can be analyzed.