Name	Shubham Golwal
UID no.	2020300**15**
Experiment No.	01

AIM:	Data Models (Dimensional Modelling)
Program 1	
PROBLEM STATEMENT:	The goal of this experiment is to design and develop a data warehouse solution that provides HR departments with a centralized, transparent, and comprehensive view of HR data. This solution should support the collection and storage of HR data from various sources and allow HR departments to perform complex data analysis to gain insights into key HR metrics. The solution should also allow HR departments to easily query the data and generate reports to support decision-making processes. Prepare a Dimension Modelling for Human Resources Management System. Generate Star and snowflake schema for the one.
Theory:	**What is Dimensional modeling?** Dimensional modeling is a data warehousing technique used to design a database to support business intelligence (BI) activities. The basic premise of dimensional modeling is to represent data in a way that allows for easy querying and analysis, while minimizing data redundancy and complexity. The key components of dimensional modeling are: • Fact table: contains the central data for analysis and is often the largest table in the data warehouse. • Dimension tables: provide context to the fact table by describing the characteristics of the objects being analyzed. • Attributes: categorical information about the dimensions that is used to filter and aggregate the data. The goal of dimensional modeling is to provide a flexible, scalable, and user-friendly data structure that can be used to support a wide range of BI activities, including reporting, data analysis, and predictive modeling. It is widely used by data warehousing practitioners due to its simplicity,

	scalability, and ease of use. **Fact Table:** A fact table is the central table in a dimensional model and contains the data that is used for analysis. It records business transactions or events and provides the context for analyzing the data. A fact table typically includes foreign keys to dimension tables and contains measures, which are the numerical values that describe the facts being recorded. Example: Consider a retail store that sells products. The fact table for the retail store could record the sales of each product on a daily basis. The columns in the fact table could include: • Date: the date of the sale • Product ID: a foreign key to the dimension table that describes the product being sold • Store ID: a foreign key to the dimension table that describes the store where the sale took place • Quantity sold: the number of units sold • Total Sales: the total value of the sale In this example, the fact table is recording the sale of each product at each store on a daily basis. The foreign keys to the dimension tables provide the context for the sales data, such as the product and store being sold. The measures in the fact table, such as the quantity sold and total sales, describe the facts being recorded and provide the basis for analysis. **Dimension Tables** A dimension table is a table in a dimensional model that provides context to the data in a fact table. It describes the characteristics of the objects being analyzed and allows for filtering and grouping of the data. Dimension tables are typically linked to a fact table through foreign keys, which act as pointers to the relevant rows in the dimension table. Each dimension table is composed of one or more attributes, which are categorical information that can be used to further describe the dimension. For example, consider a retail store that sells products. The dimension tables for the retail store could include: • Product: a table that describes each product that is sold, including information such as product name, product type, manufacturer, and price. • Store: a table that describes each store that is part of the retail chain, including information such as store name, location, and size. • Time: a table that describes the time periods being analyzed, such as

year, quarter, month, and day.

In this example, the dimension tables provide context to the data in the fact table. For example, when analyzing sales data, you could use the product dimension to filter sales data by product type or manufacturer or use the time dimension to filter sales data by year or quarter. By providing this context, dimension tables make it easier to understand and analyze the data in the fact table.

Characteristic of dimensional modelling:

1. Clear and well-defined structure
2. Fact and dimension tables
3. De-normalized form
4. Easy to understand
5. Supports query performance
6. Helps in data analysis and decision making
7. Facilitates incremental data loads
8. Handling of multi-valued and derived attributes.

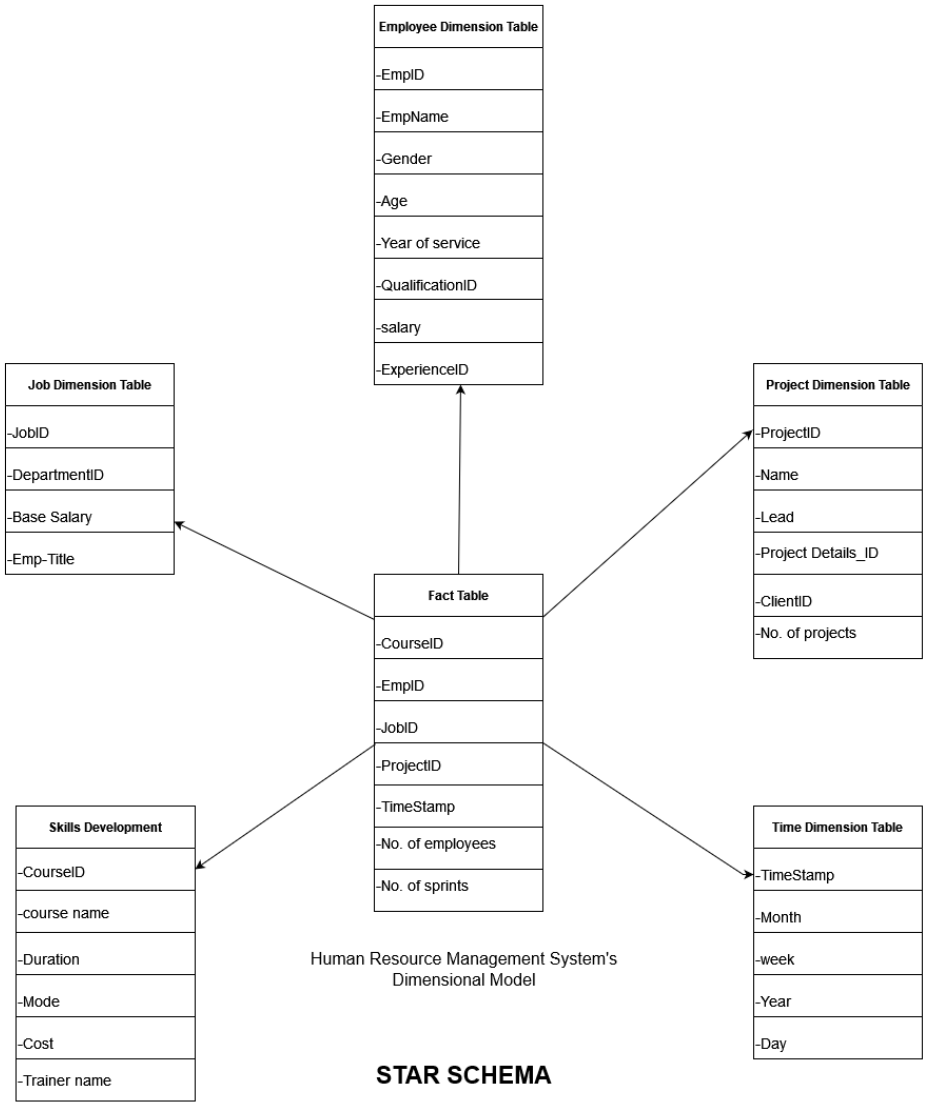
Advantages of Dimension Models:

1. Improved data analysis
2. Better data understanding
3. Improved data integrity
4. Flexibility
5. Scalability

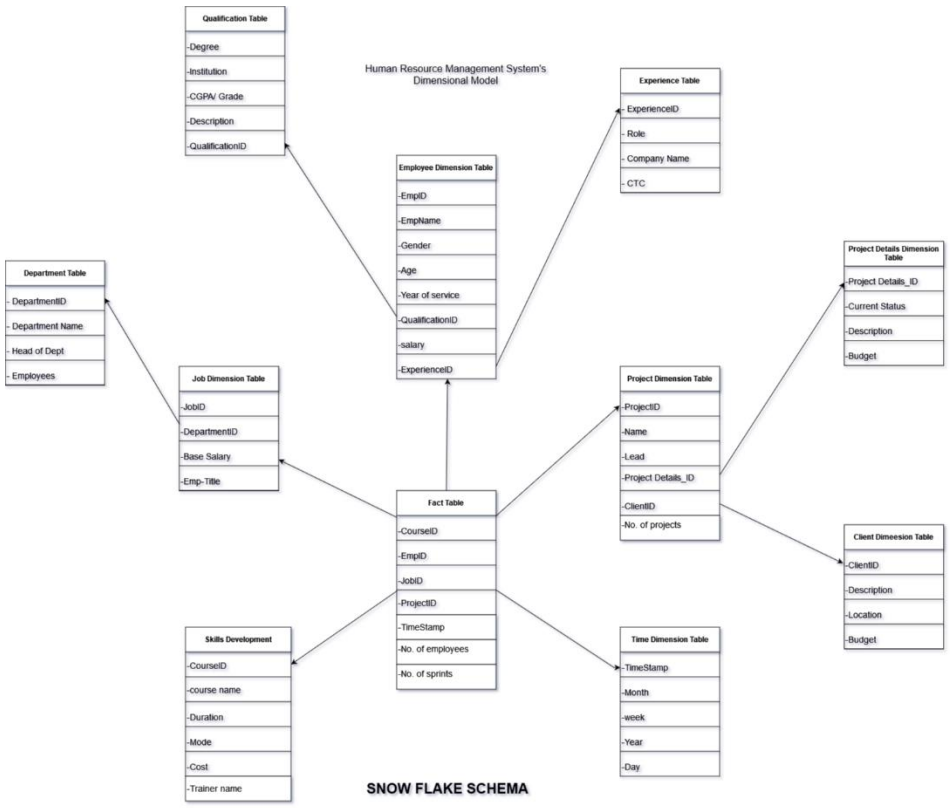
Disadvantages of Dimension Models:

1. Complexity
2. Maintenance
3. Performance
4. Limited data quality
5. Training

Output	Star schema:



Snowflake schema:



Conclusion: After the designing the models throughout this experiment:

- I now have clear understanding of the concepts of dimension modelling, including star and snowflake schemas, data redundancy, scalability.
- I'm able to apply the knowledge and skills acquired through the experiment to create dimension models for real-world data sets.
- I have improved their problem-solving skills through the process of defining the requirements, transforming, and cleansing data, and creating the dimension model.
- I'm aware about advantages and disadvantages of dimensional model and have got analogy between ER diagram and dimensional models.