SECTION: IM01

Project Name: DEI Data Hub: Building a Comprehensive Database for Workplace

Date: 05/09/2023

Team No./Team Name: Team 03/SQL Sultans

Team members: Aditya Aswin Kumar, Pranav Tejasvi Adiraju, Srikanth Parvathala

## Final Project Report

### Introduction:

Diversity, equity, and inclusion (DEI) are crucial to the success of any organization in the modern world. The success of an organization's DEI efforts may be tracked and measured in large part because of HR analytics. A well-designed database is necessary for tracking and measuring DEI parameters effectively. In this project, we will create a database for HR metric analytics based on a DEI Measurements Dataset. This database will give HR professionals the resources they need to monitor and assess the DEI initiatives of their organization and base decisions on the data.

Coming to the final report, we will demonstrate how our objectives and deliverables were achieved during the course of the project phase of developing a database from scratch. We will also be providing a complete overview on the Project Logical Design (Fully normalized tables, Entity-Relationship Diagram (ERD)), Data creation and Query Demonstration which will help answer the questions outlined during our initial project proposal and provide valuable insights related to HR and DEI metrics. Furthermore, we will also be mentioning the assumptions which we have taken into consideration while building this database as well as cite a few challenges and lessons learnt during the journey.

Moreover, we would be providing future recommendations on improving the HR Database to make it an easy-to-use, effective and scalable database for businesses to adopt while fostering a more welcoming and diverse workplace.

**Database Design and Implementation:**

**Logical Design: (The Entity-Relationship Diagram)**
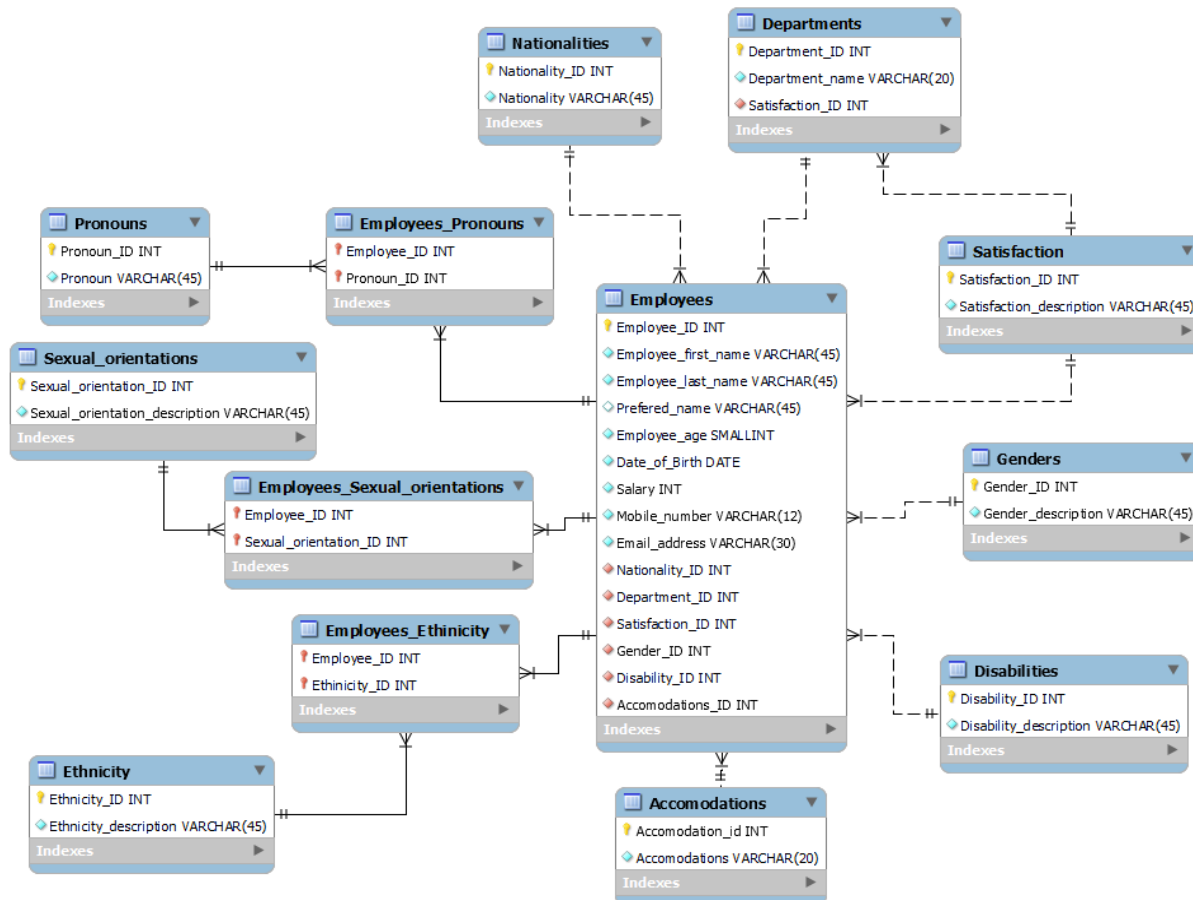


**Physical Database:**

The Physical Design of our database was achieved after forward engineering the ERD to create a schema model and later importing the required sample data and views as needed. We have also created a self-contained backup file in the MySQL Workbench and included the "**CREATE DATABASE**" query in the file. Furthermore, we tested the backup file in other computers to restore the database and to make sure everything is up and running while the required sample data and views are visible to others. Please find the **.mwb** and **.sql** file in the submissions.

## **Sample Data:**

Our project involves merging two datasets, the Human Resources Data Set and the Diversity, Equality and Inclusion Measures Dataset from Kaggle, to create a unified database that will provide insights into DEI metrics and enable workplaces to foster organizational success and employee satisfaction. To ensure that our database is manageable and focused, we have limited the rows from the original 311 rows in the Human Resources dataset and 10,000 rows in the DEI dataset. We will be using columns such as *Nationality, Ethnicity, Employee_Satisfaction* (which has levels from 1 to 5), and *Hobbies* to address diversity and employee engagement within the organization. Additionally, *Sexual_Orientation, Disabilities*, and *Pronouns* columns will be used to address inclusiveness. For the scope of this project, we have discarded columns such as *Employee Retention, Employee Engagement, Employee Survey,* and *DEI training facilitation* from the dataset. We have also normalized our tables using the third normal form (3NF) to ensure that the database is free from anomalies and unwanted dependencies. Also, These datasets are open source and provide license to anyone who wishes to use this dataset for learning or teaching. Here are a few tables and sample data taken from our database for the reference.

## **Assumptions:**

1. While making this database, we have made a few assumptions about the employees and their characteristics.
2. We are making the assumption that each employee has one and only one nationality. This means that all the employees at the company do not have dual citizenship.
3. After consultation with the Professor, It is possible that the employees have more than one sexual orientation.
4. With each employee having a satisfaction rating, the entire department also has a satisfaction in our database.
5. The data is purely factual and does not represent real life scenarios.
6. Each employee belongs to one and only one department.
7. Employee satisfaction is represented by a single satisfaction score, rather than a more complex metric or set of metrics.

**Employees Table:**

| Employee_ID | Employee_firs | Employee_last | Preferred | Employee_age | Date_of_Birth | Salary | Mobile_numbe | Email_address | Nationality | Departmen | Satisfactio | Gender_ID | Disability_I | Accomodations_ID |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Deborah | Addison | Deborah | 29 | 6/4/1993 | 62506 | 363 436 1096 | Deborah.Addisor | 82 | 1 | 5 | 6 | 1 | 1 |
| 2 | Mona | Hill | Mona | 59 | 8/2/1963 | 104437 | 905 980 9989 | Mona.Hill@mail. | 58 | 2 | 5 | 2 | 2 | 2 |
| 3 | Kimberly | Shelton | Kimberly | 43 | 4/25/1979 | 64955 | 788 356 1577 | Kimberly.Sheltor | 17 | 3 | 4 | 1 | 2 | 2 |
| 4 | Denis | Robinson | Denis | 44 | 2/11/1978 | 64991 | 946 855 5822 | Denis.Robinson@ | 17 | 4 | 4 | 2 | 1 | 1 |
| 5 | Carmen | Gunn | Carmen | 50 | 3/11/1972 | 50825 | 537 895 4532 | Carmen.Gunn@r | 56 | 5 | 3 | 2 | 2 | 2 |
| 6 | Nathan | Furnell | Nathan | 37 | 1/11/1985 | 57568 | 317 280 1655 | Nathan.Furnell@ | 17 | 6 | 4 | 2 | 2 | 2 |
| 7 | Luke | Sinclair | Luke | 34 | 9/4/1988 | 95660 | 500 192 9313 | Luke.Sinclair@m | 31 | 2 | 3 | 2 | 2 | 2 |
| 8 | Fred | Callan | Fred | 27 | 11/9/1994 | 59365 | 531 698 4421 | Fred.Callan@ma | 76 | 6 | 1 | 2 | 2 | 2 |

**Views / Queries**

**Requirements table:**

| View Name | JOINS | Filtering | Aggregation | Linking Table | Subquery |
|---|---|---|---|---|---|
| salary_statistics_by_department | ✅ | | ✅ | | |
| highest_salary_employees_satisfaction | ✅ | ✅ | | | ✅ |
| department_employee_summary | ✅ | | ✅ | | |
| avg_salary_range_vs_avg_satisfaction_rate | | ✅ | ✅ | | ✅ |
| ethnicity_avg_salary | ✅ | | ✅ | ✅ | |
| age_sexual_orientation | ✅ | | ✅ | ✅ | |
| disability_satisfaction_score | | | ✅ | | |
| dept_diversity_satisfaction | ✅ | ✅ | ✅ | ✅ | |
| nationality_groups | ✅ | | ✅ | | |
| ethnic_diversity | ✅ | | ✅ | | |

## Changes from Original Design

**Issues encountered:**

While creating the DEI database we encountered many issues and ethical dilemmas throughout the project phase, which we dealt with as a team with the helpful guidance of Professor Duffy and Professor Diker. We couldn't thank them enough for their support throughout the project phase.

Here are the some of the issues highlighted:

1. **Spider Structure:**

   While creating the ERD, we encountered the *"SPIDER"* structure in the design, where all the tables were connected to the Employees table making the design complex.

2. **Ethical Dilemmas:**

   As a DEI Database, our database is supposed to keep in account all the possible DEI factors which could effectively capture all the details that depict and cover the entire user base.

   But for the scope of this project, we could not record all the factors and assumptions as it would add more tables and make our project much more complex.

3. **Varying Opinions:**

   Each of us in our team had a different view on the data and the database design, initially it was a challenge for us to be on the same page and finalize a design.

**Solutions Considered:**

Here are a few solutions that we considered to tackle the issues we faced while implementing the database:

1. To deal with the Spider Structure we had two alternatives, one of which we considered to have satisfaction rating for both individual employee and department as well. So we consider linking

the Satisfaction entity with both Departments and Employees entity. Our Initial approach was to have Satisfaction_ID only for the Employees table.

We also considered a many-to-many relationship between the 'Disabilities' and 'Employees' entities and explored the idea of linking the 'Accommodation' entity with the 'Disabilities_Employees' linking table.

2. Coming to the Ethical Dilemmas issue, we had great confusion on what factors to keep in the database and what to discard to depict the database as accurately as possible. We were struck on whether to have many to many relationships for employees and nationalities and for employees and genders table.

**Solutions Finalized:**

1.To deal with the *"SPIDER"* structure we finalized on having the Satisfaction_ID for both Departments and the Employees table based on Professor's input.

The Department satisfaction rate can be assumed as the average of individual employee ratings in a department or the rating given by the senior leadership.

2.For Ethical Dilemmas, we assumed that in our database, one employee can only have one nationality unlike the real-world case where one employee can have dual citizenship.

Also, for the genders, individuals may identify themselves as more than one gender but for the scope of this project we only considered the gender during birth.

*Note: Please check the assumptions under the Sample Data for reference.*

<u>**Lessons Learnt**</u>

Throughout the course of this project, we learned some valuable lessons at an individual level and as a team. Some of the important lessons were:

1. **Teamwork and Collaboration:**

It has been a great learning experience working as a team for this project and learning about one's strengths and weaknesses while dealing with issues as a team was something that we would take with us moving forward into the corporate sector.

2. **Realizing Expectations:**

During the initial phase of the project, everyone in our team had magnificent ideas and views on how to make this database in its best version possible, not realizing the importance of realistic expectations catering to the needs of the project. Understanding the project requirements and setting up realistic expectations was a valuable lesson learnt in this project journey.

3. **Data Management:**

While managing data is a very important skill for Information Professionals, during this project we pondered upon how huge enterprises manage their data effectively while providing seamless service to their customers and stakeholders.

4. **Scope for Future Learning:**

When our database was up and running, we researched how this database could be further improved and what alternative tools and technologies that can be used to make this possible even better. Our Research revealed that there are promising options that can further make this database more efficient and scalable by latest technology adoption. For example: We can host this database on cloud to manage large data efficiently.

**Potential Future Work**

**Extensions within the current technology and design:**

While the HR DEI Database implemented by our team was able to successfully answer all the questions posed in the project proposal, there are few additional extensions that could be added to this database to make the design more complete and cater to a broader range of user needs while also provide insights needed for the maintaining a more welcoming and diverse workplace.

Firstly, incorporating an additional linking table for Nationality and Gender would have catered to the needs of users who come under the set category.

Moreover, the Satisfaction rating used in our database only stores the most recent rating of that particular review cycle unlike companies which store the ratings for the previous cycles of the performance review as well. One extension could be having individual tables for various review cycles for different years.

Lastly, a dedicated table listing all the disabilities can be implemented through a many to many relationships with the employees table but because of the table constraints for the final project we were not able to implement it in our project.

**Feasibility of alternative implementations**

Exploring new avenues for building HR DEI databases can not only bring fresh perspective to the design but also help improve the performance, scalability, and ease of use. Few options can be:

**1.      Cloud-based Databases:**

Hosting the HR DEI database in AWS cloud can be a great option. By Leveraging RDS (Relational Database Service) from AWS, companies can benefit from improved scalability, performance as well as advanced security and infrastructure provided by AWS. RDS also supports a variety of popular database engines including MySQL, PostgreSQL and Oracle while offering features such as automated backups and software patching also secure encryption at rest and in transit.

**2.      Graph Databases:**

Leveraging Neo4j in the HR DEI database can help businesses improve performance and scalability as well as advanced query capabilities and data visualization tools. Furthermore, Neo4j can provide a more natural and intuitive way of modeling and analyzing complex relationships and patterns in DEI Data.For Example: Companies can use Neo4j to identify patterns of bias or discrimination in their hiring or promotion process as well as analyze DEI initiatives on employee retention rate.

**Reference**

"Diversity, Equity and Inclusion Measures Dataset." *Kaggle*, 2 Nov. 2022,

www.kaggle.com/datasets/keremkurt/diversity-equity-and-inclusion-measures-dataset. Accessed

3 Mar. 2023.

"Human Resources Data Set." *Kaggle*, 19 Oct. 2020,

www.kaggle.com/datasets/rhuebner/human-resources-data-set. Accessed 3 Mar. 2023.

Murach, Joel. *Murachs MySQL*. Mike Murach and Associates, Incorporated, 2019.