**Project Requirements (Deliverable 1 & 2)**

**Goal:** Demonstrate proficiency in data handling in SQL (using MySQL).

**Synopsis:** VivaK is a successful retail chain in the fashion industry, the head office of which is located in Southlake, Texas, USA. The HR Manager of the VivaK head office has decided a hire a group of analysts (you) to support their HR analytics by devising an Online Analytical Processing (OLAP) Database. Unfortunately, the information management of the HR department has not been robust so far, and it maintains its data in various data/file formats. The data is also known to have anomalies. The HR manager narrates the following story:

*We are in a tough situation with this data, and our analysts find it highly challenging to perform their job without a well-designed OLAP database. Allow me to clarify how our organizational structure works. We operate in several regions and countries, but we manage our operations via seven office locations in different parts of the world. We have three offices in the US, one in Canada, two in the UK, and one in Denmark. The President sits at the head office in Southlake, Texas, USA. All the other locations have the same well-structured management team reporting to two Vice Presidents at each location: VP Administration and VP Operations. Each location has 11 departments, 10 of which are managed by a dedicated Manager; the other employees report to these managers. We pay our employees on the USD salary scale, and you may find the monthly salary range per job title and each employee’s monthly salary in the sample dataset we have provided to you. We also keep track of each employee’s dependents. Our data designer made an error and called the “location\_id” column in employees data “department\_id”.*

**Your Tasks:**

**Task 1: Analyse the provided data and design a complete (and effective) MySQL Data Model** (File → New Model) call VivaK\_Data\_Model You must include the diagram in your presentation. You must identify the following:

1. Entities

2. Attributes

3. Constraints (Including Primary Keys, Not Null columns, and Unique Keys)

4. Relationships and Foreign Keys

**Task 2: Create a Schema based on VivaK\_Data\_Model and call it VivaKHR.** You must include all the statements in your SQL file and only the important statements/ outputs in your presentation.

1. All salary-related columns must be double. Data must be recorded in 2 decimals.

2. All ID columns and their related foreign keys must be Integer.

3. The employee table MUST contain the report\_to column to record the ID of the employee’s manager.

4. The data type of All the date columns MUST be DATE.

5. Include a column called experience\_at\_VivaK in the appropriate table to include the number of months that each employee has worked at VivaK.

6. Include a column called last\_performance\_rating in the appropriate table to include the performance rating (0-10) of each employee after the annual performance appraisal.

7. Include a column called salary\_after\_increment in the appropriate table to record the salary anticipated after the annual performance appraisal.

8. Include a column called annual\_dependent\_benefit in the appropriate table to record the dependent bonus that each employee receives per dependent.

**Task 3: Import and clean the data.** Import the sample data to test the integrity and efficiency of your OLAP schema. You must include all the statements/outputs in your SQL file and only the important statements in your presentation. Consider the following:

1. Handle Duplicates: All tables must be in the full-normalized form, meaning that there should not be any data redundancy. Check for duplicates in each table using qualifying candidate keys.

* 1. 2. Format the data according to the designated data types. a. Ensure the floating-point data is represented as double.
  2. b. Ensure that the phone numbers are all recorded in the format: ‘+000-000-000-0000’, where the first four characters refer to the country code.
  3. c. Ensure the dates are recorded in the format: ‘yyyy-mm-dd’.
  4. **3. Treat Missing Values:**
  5. a. Fill up the report\_to column by analyzing the available data.
  6. b. Devise a strategy to fill in the missing entries in the salary column. Justify your answers and state your assumptions.

**Task 4 Perform the following calculations and updates.** You must include all the statements in your SQL file and only the important statements/outputs in your presentation.

1. experience\_at\_VivaK: calculate the time difference (in months) between the hire date and the current date for each employee and update the column.

2. last\_performance\_rating: to test the system, generate a random performance rating figure (a decimal number with two decimal points between 0 and 10) for each employee and update the column.

3. salary\_after\_increment: calculate the salary after the performance appraisal and update the column by using the following formulas:

***salary\_after\_increment*** = salary \* increment

|  |  |
| --- | --- |
| ***increment*** = 1 + (0.01 \* experience\_at\_VivaK) + rating\_increment last\_performance\_rating | rating\_increment |
| >= 0.9 | 0.15 |
| >= 0.8 | 0.12 |
| >= 0.7 | 0.10 |
| >= 0.6 | 0.08 |
| >= 0.5 | 0.05 |
| < 0.5 | 0.02 |

4. annual\_dependent\_benefit: Calculate the annual dependent benefit per dependent (in USD) and update the column as per the table below:

|  |  |
| --- | --- |
| Title | Dependent beneift per dependent |
| Executives | 0.2 \* annual salary |
| Managers | 0.15 \* annual salary |
| Other Employees | 0.05 \* annual salary |

5. email***:*** Until recently, the employees were using their private email addresses, and the company has recently bought the domain VivaK.com. Replace employee email addressed to ‘<emailID>@vivaK.com’. emailID is the part of the current employee email before the @ sign.