**CHALLENGE 9**

**REPORT**

**Introduction**:

This study is regarding data of a fictitious company, Pewlett Hackard, regarding its employees, departments, employee-salaries, and employee-titles from the 1980s and the 1990s.

**Data**:

The 6 CSV files provided are:

1. departments:

Number of records: 9

Attributes: department number, department name

1. dept\_emp:

Number of records: 331,603

Attributes: department number, employee number

1. dept\_manager:

Number of records: 24

Attributes: department number, employee number for each manager

1. employees:

Number of records: 300,024

Attributes: employee number, title number, date of birth, first name, last name, sex, starting date of employment

1. salaries:

Number of records: 300,024

Attributes: employee number, salary

1. titles:

Number of records: 7

Attributes: title number, title

**Analysis**:

Part I

In PostgreSQL 12, using pgAdmin, a database called **company\_db** is created.

Six tables are created, followed by importing data from each corresponding CSV file into each table.

In order to determine a primary key (attribute) for each table, a check for uniqueness was conducted.

Both the ‘**employees**’ table and ‘**salaries**’ table show that there are 300,024 employees. But the ‘**dept\_emp**’ table has 331,603 records, indicating that employees may belong to more than one department.

Table ‘dept\_emp’ has duplicates in all/both of its columns, ‘dept\_no’ and ‘emp\_no’. Therefore, for this table, a composite key was created.

Foreign keys were also determined for each table to establish inter-table relationships.

Note: The primary and foreign keys were added **after** creating and importing data, to avoid encountering error due to potential erroneous choices of primary keys.

Part II

Information from the tables created in Part I, including the primary and the foreign keys for each table was input into quickDBD, to create an Entity Relationship Diagram (ERD) at the following link:

<https://app.quickdatabasediagrams.com/#/d/rGDkj2>

A screenshot of a computer

Description automatically generated

Part III

The following are the queries created and run and the results for each query, respectively. A sample of 15 rows from each query result is shown in each table.

1. List the employee number, last name, first name, sex, and salary of each employee.

Table 1.

A screenshot of a computer

Description automatically generated

1. List the first name, last name, and hire date for the employees who were hired in 1986.

Table 2.

A screenshot of a computer

Description automatically generated

1. List the manager of each department along with their department number, department name, employee number, last name, and first name.

Table 3.

A screenshot of a computer

Description automatically generated

1. List the department number for each employee along with that employee’s employee number, last name, first name, and department name.

Table 4.

A screenshot of a computer

Description automatically generated

1. List first name, last name, and sex of each employee whose first name is Hercules and whose last name begins with the letter B.

Table 5.

A screenshot of a computer

Description automatically generated

1. List each employee in the Sales department, including their employee number, last name, and first name.

Table 6.

A screenshot of a computer

Description automatically generated

1. List each employee in the Sales and Development departments, including their employee number, last name, first name, and department name.

Table 7.

A screenshot of a computer

Description automatically generated

To check if any employee is in Sales as well as in Development, a query was run to count the number of such employees. Table 7a. shows that no such employee exists in the company.

Table 7a.

A close up of a website

Description automatically generated

Table 7b. shows that the sum of employees of each department is the same as the number of employees from the original query (all employees in departments, Sales and Development combined).

This indicates that there is no ‘common’ employee of both departments.

Table 7b.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

1. List the frequency counts, in descending order, of all the employee last names (that is, how many employees share each last name).

Table 8.

A screenshot of a table

Description automatically generated

**Conclusion**:

* From Part I, it can be observed that the data imported into the created tables can have duplicate values in each column. The data can be checked for uniqueness prior to determining primary keys.
* Part II provides a visual aid to understanding relationships between tables, via foreign keys for each table.
* Part III enables us to understand the data quantitatively through queries.