

Databases

Lab sheet 1: Introduction to Database Management and MS Excel

OBJECTIVES:


The objectives of this lab are to demonstrate the limitations of using Excel for database management, including data redundancy, integrity, and security issues. Students will learn how to organise data in tables, create relationships manually, and perform basic queries. The lab highlights Excel's inefficiencies for large datasets and introduces the advantages of a DBMS for managing complex data.

Task 1: Dealing with Unstructured Data

Scenario:

You are a new data engineer at a company, and HR has provided you with the following unstructured employee data. They need you to calculate the **average duration** of all projects assigned to employees.

HR Data:

 employee_data.txt - Notepad

File Edit Format View Help

Employee number 1059 is assigned to project 122, starting on 07/03/2013 and ending on 13/11/2014. No assigned supervisor, and the employee's function is Designer, Manager.

Employee number 1001 is assigned to project 122, starting on 08/03/2011 and ending on 28/06/2011. The supervisor number is 1059, with the name MERABAT, and the employee's function is Manager.

Employee number 1022 is assigned to project 122, starting on 15/06/2011 and ending on 04/10/2011. The supervisor number is 1059, with the name MERABAT, and the employee's function is Developer.

Employee number 1059 is assigned to project 103, starting on 12/09/2010 and ending on 01/11/2010. No assigned supervisor, and the employee's function is Designer, Manager.

Employee number 1001 is assigned to project 208, starting on 15/06/2011 and ending on 12/10/2011. The supervisor number is 1059, with the name MERABAT, and the employee's function is Manager.

Employee number 1059 is assigned to project 208, starting on 15/06/2011 and ending on 06/03/2012. No assigned supervisor, and the employee's function is Designer, Manager.

Employee number 1022 is assigned to project 208, starting on 01/09/2011 and ending on 17/12/2011. The supervisor number is 1059, with the name MERABAT, and the employee's function is Developer.

Employee number 1059 is assigned to project 133, starting on 06/11/2011 and ending on 19/02/2012. No assigned supervisor, and the employee's function is Designer, Manager.

Employee number 1053 is assigned to project 208, starting on 01/09/2011 and ending on 06/03/2012. The supervisor number is 1026, with the name BOURAS, and the employee's function is Administrator.

Employee number 1026 is assigned to project 208, starting on 19/08/2011 and ending on 06/03/2012. The supervisor number is 1059, with the name MERABAT, and the employee's function is Analyste, Manager.

Question:

Before calculating the average project duration, **what should you do** to make this data easier to work with?

Task 2: Querying and Analyzing the Data

1. Employee Role Analysis:

What is the **function** of employee **1022** in both project **122** and project **208**? What do you notice? How would you solve this problem?

2. Identifying Designers:

List all employees who work as **Designers** in the company.

Discuss the encountered issue? How to solve it ?

3. Supervisor Query:

Who is the supervisor of **employee 1053**? What is the name of the **supervisor number 1026**? What do you notice? What do you propose ?

4. Data Integrity Check:

Change the **Start_Affect** date for employee **1059** in project **122** from a date to the text "March". What happens?

5. Data Security in Excel:

Suppose you have sensitive employee data, like salaries, that you need to protect. How could you secure this information in Excel? What are the limitations of Excel's security features?

Task 3: Manual Data Queries in Excel

1. Create a new Excel workbook, then organize the data from Task 2 into separate sheets, ensuring that each sheet represents a distinct table.
2. You have been asked by the company to find the function of employee 1001, and you're considering using Excel functions like VLOOKUP() to accomplish this. What challenges did you face when manually querying the data in Excel?

Task 4: Reflection on Excel's Limitations

1. Excel's Limitations:

Reflect on the limitations of using Excel as a database management tool. Think about data redundancy, manual errors, and difficulty in maintaining relationships between tables.

- **Discussion:** Why would a company need to move from Excel to a more advanced database management system (DBMS)? What advantages does a DBMS provide over Excel?