Documentation: Finding Employees with the Highest Salary per Department

Problem Overview:

- We are tasked with finding employees who have the highest salary in each department from two provided tables: the Employee table and the Department table. The goal is to return the department name, employee name, and their salary for employees with the highest salary in each department. If multiple employees have the highest salary in the same department, all such employees should be included in the result.

Input Tables:

1. Employee Table:

- This table contains information about each employee, including their ID, name, salary, and the department they belong to. The departmentId column in the Employee table is a foreign key that references the id column in the Department table.

- ***Schema:***
  - *id:* Integer (Primary Key) – Unique identifier for each employee.
  - *name:* String – Name of the employee.
  - *salary:* Integer – Salary of the employee.
  - *departmentId:* Integer – Foreign key that refers to the department the employee belongs to.

2. **Department Table:**
   - This table contains information about the departments. Each department has a unique id and a name. The id in this table is referenced by the departmentId in the Employee table.

- ***Schema:***
  - *id:* Integer (Primary Key) – Unique identifier for each department.
  - *name:* String – Name of the department.

Problem Requirements:

The task is to determine which employees have the highest salary in each department and return the following fields:

- **Department:** The name of the department.
- **Employee:** The name of the employee(s) with the highest salary in that department.
- **Salary:** The highest salary in that department.

The result should contain one or more rows for each department, depending on how many employees share the highest salary.

Approach:

To solve this problem, we need to follow a structured approach:

1. **Data Joining:**
   - Since the Employee table contains only the departmentId (not the department name), and the Department table contains the id (which matches the departmentId in the Employee table), we first need to join (merge) the Employee and Department tables on their respective departmentId and id columns. This merge will allow us to associate each employee with their respective department's name.

2. Finding the Maximum Salary:

- After merging the two tables, the next step is to group the data by the department and determine the highest salary in each department. This can be done by grouping the data based on the department and then applying a function that retrieves the maximum salary for each group.

3. Filtering Employees:

- Once we know the maximum salary for each department, we need to filter the employees who earn that maximum salary. Since multiple employees may earn the same maximum salary within the same department, we need to ensure all such employees are included in the final result.

4. Result Preparation:

- Finally, after filtering the employees with the highest salary, we will prepare the result to include the department name, employee name, and their salary in the correct format.

Example Walkthrough:

Let's break down an example step-by-step to illustrate the approach:

Input:

Employee table:

id	name	salary	departmentId
1	Joe	70000	1
2	Jim	90000	1
3	Henry	80000	2
4	Sam	60000	2
5	Max	90000	1

Department table:

id	name
1	IT
2	Sales

Step 1: Merge the two tables.

After merging, we will get a DataFrame containing all relevant columns from both tables:

id	name	salary	departmentId	department_name
1	Joe	70000	1	IT
2	Jim	90000	1	IT
3	Henry	80000	2	Sales
4	Sam	60000	2	Sales
5	Max	90000	1	IT

Step 2: Identify the highest salary for each department.

We group by the departmentId or department_name and find the maximum salary in each group:

- For department IT, the highest salary is 90000.
- For department Sales, the highest salary is 80000.

Step 3: Filter employees who have this maximum salary.

For each department, we filter employees whose salary matches the maximum salary identified in Step 2:

- For department IT, both Jim and Max have a salary of 90000.
- For department Sales, Henry has a salary of 80000.

Step 4: Format the result.

The final result should include the department name, employee name(s), and salary:

Department	Employee	Salary
IT	Jim	90000
Sales	Henry	80000
IT	Max	90000

Edge Cases:

- **Multiple Employees with the Same Salary:** If multiple employees have the highest salary in the same department, all such employees should be included in the result.
- **Single Employee per Department:** If a department has only one employee, that employee should automatically be the one with the highest salary.
- **No Employees in a Department:** If a department has no employees, it should not appear in the result.

Time Complexity:

- **Merging the Tables:** The time complexity for merging the Employee and Department tables is (O(n)), where (n) is the total number of rows in the Employee table.
- **Grouping and Finding Maximum:** The time complexity for grouping by department and finding the maximum salary is (O(n)), where (n) is the total number of employees.
- **Filtering the Employees:** The time complexity for filtering employees based on the maximum salary is (O(n)).

Thus, the overall time complexity of the solution is (O(n)), which ensures that the solution will scale efficiently with larger datasets.

Conclusion:

- This approach efficiently solves the problem of finding employees with the highest salary per department by merging the two relevant tables, grouping employees by department, and filtering those who match the highest salary for their respective departments. The final result is a list of department names, employee names, and their salaries, ready to be used for further analysis or reporting.