**Mục lục**

[I. Giới thiệu 2](#_Toc143608987)

[1. Mô tả bài toán 2](#_Toc143608988)

[2. Mục tiêu 2](#_Toc143608989)

[II. Thực hành 3](#_Toc143608990)

[1. Phân tích bài toán và khảo sát nguồn dữ liệu 3](#_Toc143608991)

[2. Mô tả nguồn dataset 3](#_Toc143608992)

[**2.1 Video** 3](#_Toc143608993)

[**2.2 Category video** 4](#_Toc143608994)

[**2.3 Country** 4](#_Toc143608995)

[**2.4 Github link dataset (video,category,country)** 4](#_Toc143608996)

[3. Xác định và mô tả yêu cầu 4](#_Toc143608997)

[4. Phân tích yêu cầu 5](#_Toc143608998)

[5. Entity relationship diagram 5](#_Toc143608999)

[6. Lược đồ quan hệ và mô tả dữ liệu đích 7](#_Toc143609000)

[**6.1 Video** 7](#_Toc143609001)

[**6.2 Category** 8](#_Toc143609002)

[**6.3 Country** 8](#_Toc143609003)

[7. Dataflow architecture 8](#_Toc143609004)

[8. Xây dựng database đích (MS SQL Server) 8](#_Toc143609005)

[9. Xây dựng workflow trong Apache Airflow 10](#_Toc143609006)

[**1.1** **Workflow video\_DAG** 10](#_Toc143609007)

[**1.2** **Workflow category\_and\_country\_DAG** 10](#_Toc143609008)

[**1.3** **Workflow sequence\_YoutubeVideo\_DAG** 12](#_Toc143609009)

[**1.4** **Github link source code** 12](#_Toc143609010)

[10. Trực quan hóa dữ liệu qua PowerBI 12](#_Toc143609011)

[**10.1** **Yêu cầu báo cáo thống kê** 12](#_Toc143609012)

[**10.2** **Xây dựng biểu đồ** 13](#_Toc143609013)

[11. Link github dự án và video 19](#_Toc143609014)

[III. Tài liệu tham khảo 20](#_Toc143609015)

# Giới thiệu

## Mô tả bài toán

**Bài 1:** Một hệ thống ETL dữ liệu cần lấy data từ 3 nguồn: MySQL, MongoDB, DynamoDB và trong tương lai còn lấy thêm từ nhiều cơ sở dữ liệu khác nữa

Yêu cầu: Sử dụng kiến thức Lập trình hướng đối tượng, design pattern, để:

* Thiết kế class diagram cho hệ thống trên
* Viết code implement, tối thiểu phải có chức năng kết nối, lấy dữ liệu. Chụp kết quả cách thực thi code và kết quả dữ liệu lấy được

**Bài 2**: Một doanh nghiệp có một cơ sở dữ liệu quan hệ như sau:

Employee (ID, Name, Nationality, DOB, Phone Number, Employee ID of Manager, Salary, Onboard Date)

Role (ID, Role Name)

Employee\_Role (Employee ID, Role ID, Start Date)

Department (ID, Name)

Employee\_Department (Employee ID, Department ID, Start date, End Date)

Yêu cầu: sử dụng SQL để:

1. Tạo bảng, khởi tạo dữ liệu

2. Lập danh sách nhân viên, phòng ban, vị trí công việc của từng nhân viên ở thời điểm hiện tại

3. Tìm phòng ban có nhiều nhân viên nhất.

4. Tìm nhân viên làm việc cho nhiều phòng ban nhất.

5. Tìm tất cả những nhân viên mà làm chung ít nhất 1 phòng ban với nhân viên tên John Doe

6. Tính mức lương trung bình theo từng phòng ban.

7. Tìm tất cả những nhân viên có làm việc chung với nhân viên John Doe tại tất cả phòng ban.

8. Tìm nhân viên có nhiều cấp dưới nhất

9. Lập danh sách nhân viên và vị trí công việc mà nhân viên đã làm ngay trước khi làm ở vị trí hiện tại (Ví dụ: nhân viên John Doe có quá trình làm việc qua các vị trí RoleA → RoleB → RoleC. Hiện tại đang làm RoleC. Kết quả cần trả về RoleB)

10. Xóa dup dữ liệu bảng Employee\_Department

11. Giả sử doanh nghiệp trên sử dụng cơ sở dữ liệu MySQL. Viết câu lệnh để tạo mới một user và cấp quyền read-all cho user đó, không cấp quyền chỉnh sửa dữ liệu. Với mỗi câu SQL, chụp ảnh kết quả trả về

## Mục tiêu

Với deadline trong vòng 1 ngày, hoàn thành các yêu cầu của 2 bài toán trên

# Thực hành

## Bài 1: Thiết kế class diagram và implement code

### **Class diagram**

…

### **Implement code**

…

### **Github source code**

## Bài 2: Xây dựng database và thực hiện truy vấn theo yêu cầu

## Xác định yêu cầu

1. Tạo bảng, khởi tạo dữ liệu

2. Lập danh sách nhân viên, phòng ban, vị trí công việc của từng nhân viên ở thời điểm hiện tại

3. Tìm phòng ban có nhiều nhân viên nhất.

4. Tìm nhân viên làm việc cho nhiều phòng ban nhất.

5. Tìm tất cả những nhân viên mà làm chung ít nhất 1 phòng ban với nhân viên tên John Doe

6. Tính mức lương trung bình theo từng phòng ban.

7. Tìm tất cả những nhân viên có làm việc chung với nhân viên John Doe tại tất cả phòng ban.

8. Tìm nhân viên có nhiều cấp dưới nhất

9. Lập danh sách nhân viên và vị trí công việc mà nhân viên đã làm ngay trước khi làm ở vị trí hiện tại (Ví dụ: nhân viên John Doe có quá trình làm việc qua các vị trí RoleA → RoleB → RoleC. Hiện tại đang làm RoleC. Kết quả cần trả về RoleB)

10. Xóa dup dữ liệu bảng Employee\_Department

11. Giả sử doanh nghiệp trên sử dụng cơ sở dữ liệu MySQL. Viết câu lệnh để tạo mới một user và cấp quyền read-all cho user đó, không cấp quyền chỉnh sửa dữ liệu. Với mỗi câu SQL, chụp ảnh kết quả trả về

## Phân tích yêu cầu

|  |  |  |
| --- | --- | --- |
| ID | Requirement | Describe |
| 1 | Tạo bảng, khởi tạo dữ liệu | Tạo sample database với tên hrm, các bảng tương ứng , và khởi tạo sample data |
| 2 | Lập danh sách nhân viên, phòng ban, vị trí công việc của từng nhân viên ở thời điểm hiện tại | SQL Query |
| 3 | Tìm phòng ban có nhiều nhân viên nhất | SQL Query |
| 4 | Tìm nhân viên làm việc cho nhiều phòng ban nhất. | SQL Query |
| 5 | Tìm tất cả những nhân viên mà làm chung ít nhất 1 phòng ban với nhân viên tên John Doe | SQL Query |
| 6 | Tính mức lương trung bình theo từng phòng ban. | SQL Query |
| 7 | Tìm tất cả những nhân viên có làm việc chung với nhân viên John Doe tại tất cả phòng ban. | SQL Query |
| 8 | Tìm nhân viên có nhiều cấp dưới nhất | SQL Query |
| 9 | Lập danh sách nhân viên và vị trí công việc mà nhân viên đã làm ngay trước khi làm ở vị trí hiện tại (Ví dụ: nhân viên John Doe có quá trình làm việc qua các vị trí RoleA → RoleB → RoleC. Hiện tại đang làm RoleC. Kết quả cần trả về RoleB) | SQL Query |
| 10 | Xóa dup dữ liệu bảng Employee\_Department | Tạo ngữ cảnh data bị duplicate trong bảng Employee\_Department |
| 11 | Giả sử doanh nghiệp trên sử dụng cơ sở dữ liệu MySQL. Viết câu lệnh để tạo mới một user và cấp quyền read-all cho user đó, không cấp quyền chỉnh sửa dữ liệu. | Database Administration Statements |

## Lược đồ quan hệ và mô tả dữ liệu đích

Employee (**ID**, Name, Nationality, DOB, Phone Number, Employee ID of Manager, Salary, Onboard Date)

Role (**ID**, Role Name)

Employee\_Role (*Employee ID, Role ID*, Start Date)

Department (**ID**, Name)

Employee\_Department (*Employee ID, Department ID,* Start date, End Date)

### **6.1 Employee**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Column\_name | Data\_type | Description |
| 1 | ID | INT | Mã nhân viên (Primary Key) |
| 2 | Name | VARCHAR(100) | Tên nhân viên |
| 3 | Nationality | VARCHAR (100) | Quốc tịch |
| 4 | DOB | DATE | Ngày sinh |
| 5 | PhoneNumber | VARCHAR(15) | Số điện thoại |
| 6 | ManagerID | INT | Mã nhân viên Quản lý |
| 7 | Salary | INT | Lương |
| 8 | OnboardDate | DATE | Ngày vào làm |

### **6.2 Role**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Column\_name | Data\_type | Description |
| 1 | ID | INT | Mã vai trò (Primary Key) |
| 2 | RoleName | VARCHAR (100) | Tên vai trò |

### **6.3 Department**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Column\_name | Data\_type | Description |
| 1 | ID | INT | Mã phòng ban (Primary Key) |
| 2 | Name | VARCHAR (100) | Tên phòng ban |

### **6.4 Employee\_Role**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Column\_name | Data\_type | Description |
| 1 | EmployeeID | INT | Mã nhân viên (Foreign key) |
| 2 | RoleID | INT | Mã vai trò (Foreign key) |
| 3 | StartDate | DATE | Ngày bắt đầu |

### **6.5 Employee\_Department**

|  |  |  |  |
| --- | --- | --- | --- |
| STT | Column\_name | Data\_type | Description |
| 1 | EmployeeID | INT | Mã nhân viên (Foreign Key) |
| 2 | DepartmentID | INT | Mã phòng ban (Foreign Key) |
| 3 | StartDate | DATE | Ngày bắt đầu |
| 4 | EndDate | DATE | Ngày kết thúc |

## Thực hiện yêu cầu (SQL)

1. Tạo bảng, khởi tạo dữ liệu

2. Lập danh sách nhân viên, phòng ban, vị trí công việc của từng nhân viên ở thời điểm hiện tại

3. Tìm phòng ban có nhiều nhân viên nhất.

4. Tìm nhân viên làm việc cho nhiều phòng ban nhất.

5. Tìm tất cả những nhân viên mà làm chung ít nhất 1 phòng ban với nhân viên tên John Doe

6. Tính mức lương trung bình theo từng phòng ban.

7. Tìm tất cả những nhân viên có làm việc chung với nhân viên John Doe tại tất cả phòng ban.

8. Tìm nhân viên có nhiều cấp dưới nhất

9. Lập danh sách nhân viên và vị trí công việc mà nhân viên đã làm ngay trước khi làm ở vị trí hiện tại (Ví dụ: nhân viên John Doe có quá trình làm việc qua các vị trí RoleA → RoleB → RoleC. Hiện tại đang làm RoleC. Kết quả cần trả về RoleB)

10. Xóa dup dữ liệu bảng Employee\_Department

11. Giả sử doanh nghiệp trên sử dụng cơ sở dữ liệu MySQL. Viết câu lệnh để tạo mới một user và cấp quyền read-all cho user đó, không cấp quyền chỉnh sửa dữ liệu. Với mỗi câu SQL, chụp ảnh kết quả trả về

## Link github dự án và video

Github: [CarsonIT/Youtube\_video (github.com)](https://github.com/CarsonIT/Youtube_video)

Link video: <https://drive.google.com/drive/folders/1ko6sgOUCz2dgvCABewtOQ07Rjbyh9PR4?usp=sharing>

# Tài liệu tham khảo

[1]: [Trending YouTube Video Statistics | Kaggle](https://www.kaggle.com/datasets/datasnaek/youtube-new?select=CAvideos.csv)

[2]: [Power BI: Create a YouTube Channel Statistics Dashboard Using YouTube Feeds (No API Key Required) - YouTube](https://www.youtube.com/watch?v=DwzF-3M3tMs)