

BÁO CÁO TUẦN

Tên nhân viên: Nguyễn Đức Anh

Chức vụ: Thực tập sinh

Phòng/ban: Phòng công nghệ thông tin

Tuần: Từ ngày 3/6/2024 đến 7/6/2024

1. Mục tiêu

- Mô phỏng dữ liệu liên tục (continuous data)
- Hiển thị dữ liệu qua giao diện quản trị (Streamlit)
- Đẩy dữ liệu lên Google Cloud Storage

2. Công việc đã thực hiện

- Triển khai Cloud Function Producer để đẩy dữ liệu lên Google Cloud Storage:

Để bắt đầu quá trình này, cloud function producer đã được lập trình để tự động tạo dữ liệu giao dịch giả và tải lên google cloud storage. Mỗi dữ liệu giao dịch bao gồm các thông tin như mã giao dịch, loại giao dịch, tên và email người gửi, trạng thái giao dịch (success,ongoing,error) và thời gian giao dịch:

```
[
  {
    "transaction_id": "GXAbYE6QypI2vJW2W8B4VeHIkzf8Au1WL0hdCvCfHIQ6nQAf9e",
    "name": "Charlotte",
    "type": "Withdrawals",
    "status": "error",
    "email": "charlotte@gmail.com",
    "timestamp": "2024-06-06 12:16:00.589525+07:00"
  },
  {
    "transaction_id": "8IKkduExxGVPjUkXtM5jIcMOVWYecYG8i54BFQJcQ5n9MitV20",
    "name": "Daniel",
    "type": "Standing Orders",
    "status": "success",
    "email": "daniel@gmail.com",
    "timestamp": "2024-06-06 12:16:00.589698+07:00"
  }
],
```

- Tạo dữ liệu giao dịch:** Dữ liệu giao dịch được tạo ra một cách ngẫu nhiên với số lượng giao dịch trong khoảng từ 1 đến 10 cho mỗi lần chạy.
- Tải lên Google Cloud Storage:** Dữ liệu giao dịch được chuyển thành định dạng JSON và được tải lên một bucket xác định trên Google Cloud Storage. Mỗi tệp JSON được đặt tên dựa trên thời gian hiện tại để đảm bảo tính duy nhất và dễ dàng truy xuất.

b. Tạo job cho Cloud Scheduler để gọi Cloud Function Producer:

Nhiệm vụ của cloud scheduler là gửi HTTP request (GET, POST) đến Cloud Function Producer mỗi phút một lần. Các bước thực hiện bao gồm:

- i. **Thiết lập Cloud Scheduler:** Cloud Scheduler được thiết lập để gửi yêu cầu HTTP đến URL của Cloud Function Producer. Phương thức GET đã được sử dụng để kích hoạt cloud function producer
- ii. **Lịch trình:** Cloud Scheduler được cài đặt để chạy mỗi phút, đảm bảo rằng dữ liệu giao dịch được tạo và tải lên Google Cloud Storage một cách liên tục.

c. Triển khai Cloud Function để xử lý dữ liệu từ Google Cloud Storage và lưu vào MongoDB:

Cloud Function này được thiết kế để lắng nghe các sự kiện từ Google Cloud Storage. Khi một tệp dữ liệu mới được tải lên bucket, Cloud Function sẽ tự động kích hoạt và thực hiện các bước sau:

- i. **Đọc tệp dữ liệu từ Google Cloud Storage:** Cloud Function sẽ đọc tệp JSON mới nhất từ bucket.
- ii. **Kết nối đến MongoDB:** Sử dụng URI kết nối đến MongoDB Atlas, Cloud Function sẽ mở kết nối đến cơ sở dữ liệu MongoDB.
- iii. **Chuyển dữ liệu vào MongoDB:** Dữ liệu từ tệp JSON sẽ được chuyển vào cơ sở dữ liệu MongoDB, mỗi giao dịch sẽ được lưu thành một bản ghi riêng biệt.

Quá trình này đảm bảo rằng tất cả các giao dịch có thể truy xuất dễ dàng khi cần thiết.

d. Sử dụng Streamlit để hiển thị dữ liệu (admin view):

Để quan sát được lượng dữ liệu ra vào mongodb, Streamlit đã được sử dụng để đơn giản hóa quá trình:

- i. **Kết nối tới MongoDB:** Ứng dụng Streamlit kết nối tới cơ sở dữ liệu MongoDB để lấy dữ liệu giao dịch.
- ii. **Hiển thị dữ liệu:** Sử dụng các công cụ trực quan hóa của Streamlit, dữ liệu giao dịch được hiển thị theo từng phút. Admin có thể thấy số lượng giao dịch đã diễn ra trong mỗi phút, và kiểm tra cloud function có gửi hụt lượng dữ liệu trên mỗi phút.

3. Kế hoạch tuần tới

- a. Điều chỉnh lượng dữ liệu tồn tại trong GCS qua admin view (Streamlit)
- b. Xuất được các log metadata của Log Explorer, và chuyển hóa nó thành metric
- c. Quan sát được các metric hoạt động của cloud function qua Streamlit
- d. Sửa các cloud function không deploy được