

Phần 1: Hệ thống nền tảng

1.1. Cơ sở dữ liệu siêu dữ liệu & Quản lý tài sản

Cơ sở dữ liệu: Triển khai một cơ sở dữ liệu (ví dụ: Firestore, DynamoDB) để lưu trữ siêu dữ liệu (metadata) của tài sản.

Lược đồ siêu dữ liệu (Metadata Schema): Mỗi mục tài sản phải bao gồm:

s3_path: Đường dẫn S3 đầy đủ.

asset_type: video, audio, image, hoặc narration (lời dẫn).

category: Danh mục do người dùng định nghĩa.

tags: Một danh sách (tùy chọn) các thẻ văn bản.

duration: Thời lượng của tài sản tính bằng giây (phải được tự động tính toán khi tải lên).

Giao diện người dùng (UI) tải lên tài sản: Tạo một giao diện người dùng trên bảng điều khiển để:

Tải tệp lên S3.

Gán một danh mục.

Thêm một hoặc nhiều thẻ.

Logic liên kết tài sản:

Tài sản có thẻ (Tagged Assets): Chỉ có thể được ghép nối với các tài sản có cùng thẻ chính xác.

Tài sản không có thẻ (Untagged Assets): Có thể được ghép nối với bất kỳ cảnh quay nào trong công việc đó.

1.2. Công cụ tạo lời dẫn tự động

Công cụ UI: Xây dựng một công cụ trên bảng điều khiển để tải lên tệp văn bản có cấu trúc (ví dụ: CSV).

Định dạng tệp đầu vào: Tệp phải có các cột cho text (văn bản), campaign_part (phần của chiến dịch), và tags (thẻ - tùy chọn).

Lưu ý xử lý backend: Tạo một script thực hiện các việc sau:

Đọc tệp đã tải lên.

Gọi API của ElevenLabs để tạo âm thanh cho văn bản mới.

Lưu tệp MP3 vào S3.

Tạo một mục tài sản mới trong cơ sở dữ liệu với siêu dữ liệu đầy đủ.

Phần 2: UI & API cấu hình công việc

Tất cả cài đặt phải được gửi dưới dạng một JSON duy nhất gọi là "Bộ chỉ thị công việc" (Job Instruction Set) đến worker.

2.1. Cài đặt công việc chính

Lựa chọn video nguồn: Trường để chọn một hoặc nhiều video nguồn.

Số lượng biến thể: Ô nhập liệu để chỉ định số lượng video độc nhất cần tạo cho mỗi video nguồn.

2.2. Định nghĩa cấu trúc động

Số lượng phần: UI phải yêu cầu số lượng phần cho phân đoạn video tùy chỉnh.

Lựa chọn chế độ:

Nếu là 1-Phần (1-Part), hiển thị các tùy chọn xử lý đơn giản.

Nếu là Nhiều-Phần (Multi-Part), tạo một khu vực cấu hình cho mỗi phần.

2.3. Tùy chọn cấu hình

Đối với chế độ 1-Phần:

Âm thanh nền: Cho phép chọn một hoặc nhiều danh mục âm thanh.

Lớp phủ (Overlay): Hộp kiểm để bật lớp phủ hình ảnh/logo ngẫu nhiên.

Văn bản trên màn hình:

Chỉ thêm văn bản vào video nếu thẻ của tài sản khớp với thẻ của tệp văn bản.

Cung cấp 10 tùy chọn phông chữ và kích thước văn bản ngẫu nhiên.

Công cụ tạo tính độc nhất (Uniqueness Engine):

Hộp kiểm để bật.

Nếu được bật, hiển thị các tùy chọn để ngẫu nhiên hóa độ tương phản, màu sắc, âm thanh, v.v.

Đối với chế độ Nhiều-Phần (Cài đặt cho mỗi phần):

Lựa chọn tài sản: Cho mỗi phần, cho phép chọn một hoặc nhiều danh mục tài sản.

Nối cảnh quay:

Chỉ định số lượng clip video sẽ sử dụng cho mỗi phần.

Đặt thời lượng tối thiểu/tối đa cho mỗi clip.

Lời dẫn: Cho phép chọn tài sản lời dẫn.

Nhạc nền (Soundtrack):

Chọn một danh mục nhạc nền.

Chỉ định các phần mà nhạc nền sẽ phát qua (ví dụ: từ Phần 1 đến Phần 2).

Lớp phủ văn bản:

Hộp kiểm để bật.

Trường nhập văn bản.

Sử dụng một danh sách phông chữ được định sẵn, chọn ngẫu nhiên.

Lớp phủ hình ảnh/logo: Hộp kiểm để bật.

Cài đặt chung cuối cùng:

Lật ngang (Horizontal Flip): Hộp kiểm chính để chỉ lật video nguồn ban đầu.

Công cụ tạo tính độc nhất: Hộp kiểm chính để bật ngẫu nhiên hóa cuối cùng cho độ tương phản, độ sáng, tốc độ và cao độ (pitch).

Phần 3: Logic của Worker Engine

3.1. Quy trình xử lý cốt lõi của Worker

Phân tích chỉ thị: Đọc cấu hình công việc dạng JSON từ tin nhắn SQS.

Vòng lặp biến thể: Chạy vòng lặp theo "Số lượng biến thể" đã chỉ định.

Tập hợp "Danh sách cảnh quay": Bên trong vòng lặp, truy vấn cơ sở dữ liệu để lấy một bộ tài sản ngẫu nhiên mới khớp với các quy tắc của công việc.

Tạo video:

Nếu chế độ 1-Phần: Áp dụng lớp phủ, âm thanh và hiệu ứng độc nhất đã chọn vào video nguồn.

Nếu chế độ Nhiều-Phần:

Xác định thời lượng: Thời lượng của phân đoạn video được quyết định bởi thời lượng của âm thanh lời dẫn.

Tạo các phần: Tạo mỗi phần thành một tệp video tạm thời.

Nối cảnh quay: Nối các clip video lại với nhau bằng cách cắt trực tiếp (hard cuts), lặp lại hoặc cắt bớt để khớp với thời lượng lời dẫn.

Trộn âm thanh: Kết hợp lời dẫn và nhạc nền.

Kết hợp các phần: Sử dụng bộ lọc concat của FFmpeg để nối tất cả các tệp video tạm thời lại.

Nối cuối cùng: Kết hợp video nguồn ban đầu với video nhiều phần vừa tạo.

Áp dụng Công cụ tạo tính độc nhất (Nếu được bật): Thực hiện một lượt xử lý cuối cùng bằng FFmpeg để áp dụng các thay đổi ngẫu nhiên (độ tương phản, tốc độ, cao độ, v.v.).

Tải lên: Tải video cuối cùng lên S3.

Lặp lại vòng lặp: Tiếp tục cho đến khi tất cả các biến thể được tạo xong.