**-** **USE CASE -  
FETCH FILES FROM STAGING FOLDER, PROCESSING AND PUT PROCESSED FILE IN DESTINATION FOLDER**

Thực hiện : Nguyễn Văn Tuấn - TuanNV90

Quản lý : Ứng Việt Anh - AnhUV

Chủ đề : Use Cases for Apache Nifi

**THUẬT NGỮ VÀ TỪ VIẾT TẮT**

S3 : Simple Storage Service

AWS : Amazon Web Services

**USE CASE : FETCH FILES FROM STAGING FOLDER, PROCESSING AND PUT PROCESSED FILE IN DESTINATION FOLDER**

1. **Mô tả usecase**

* Lấy tệp từ thư mục tạm thời (FETCH FILES FROM STAGING FOLDER):
* Mô tả: Trong bước này, hệ thống sẽ lấy các tệp từ một thư mục tạm thời được gọi là "staging folder". Đây thường là nơi lưu trữ các tệp ban đầu trước khi được xử lý.
* Xử lý tệp (PROCESSING):
* Mô tả: Sau khi các tệp đã được lấy từ thư mục tạm thời, bước tiếp theo là xử lý chúng. Quá trình này có thể bao gồm nhiều tác vụ khác nhau như chuyển đổi định dạng, làm sạch dữ liệu, phân tích, hoặc bất kỳ xử lý nào khác cần thiết.
* Đưa tệp đã xử lý vào thư mục đích (PUT PROCESSED FILE IN DESTINATION FOLDER):
* Mô tả: Sau khi tệp đã được xử lý, bước cuối cùng là lưu trữ các tệp đã xử lý vào một thư mục đích. Thư mục đích này có thể là nơi lưu trữ lâu dài hoặc là điểm trung chuyển đến hệ thống khác.

1. **Quy trình**

* Fetch files from staging (S3)
* Do some processing (Apache Nifi): remove special chars from phone numbers
* Put processed file into destination folder (S3)

1. **Chuẩn bị**

* Tài khoản Amazon Web Services (AWS)
* Apache Nifi đã dựng sẵn

1. **Thiết lập và cài đặt:**

Để Nifi có quyền đọc ghi trên S3, chúng ta cần phải thiết lập một cặp khoá truy cập (access key và secret access key). Để thiết lập cặp khoá truy cập, chúng ta cần tạo IAM user trong Amazon Web Services để lấy nó.

**Bước 1: Tạo IAM user**

Bước 1.1: Đăng nhập AWS

Bước 1.2: Truy cập vào trang quản lí các IAM user

* Lên thanh tìm kiếm của AWS console, gõ IAM và chọn “IAM”
* Chọn mục Users

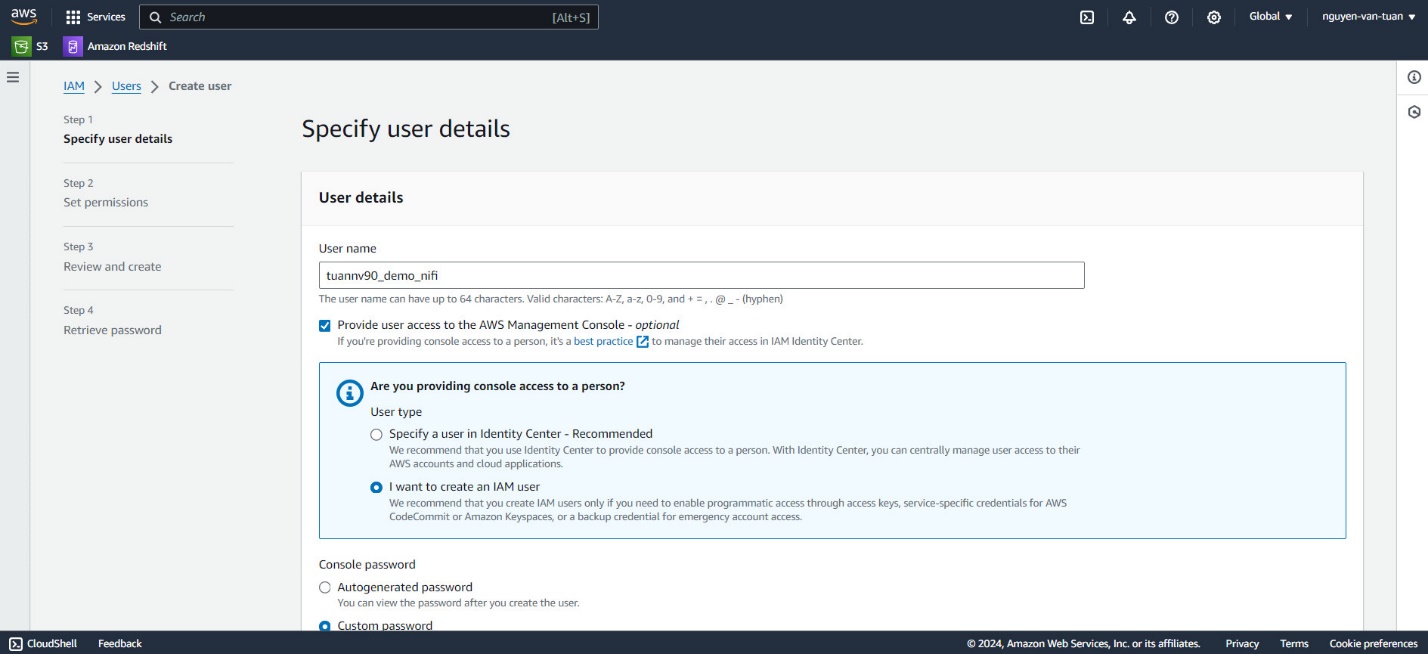
A screenshot of a computer

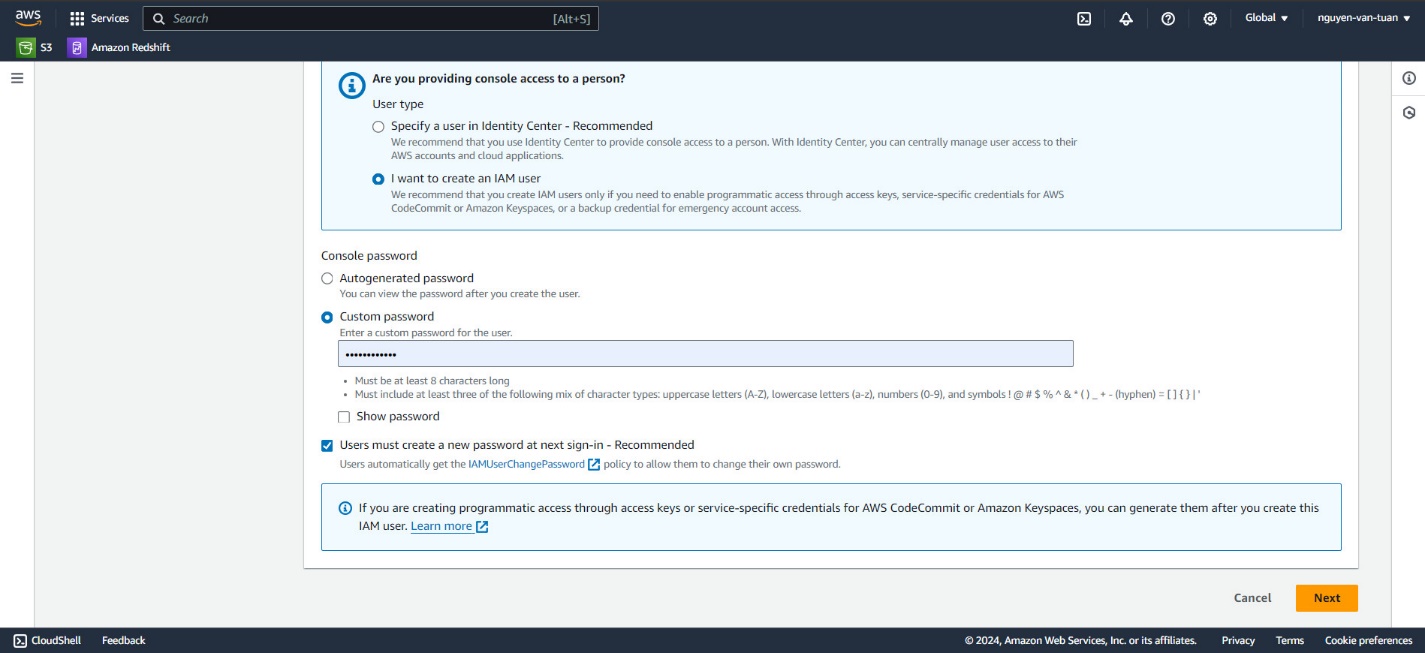
Description automatically generated

Bước 1.3: Chọn “Create user”

Bước 1.4: Thiết lập thông tin về IAM user

* User name: tên của User name
* Thiết lập thông tin như mẫu và chọn “Next”





Bước 1.5: Phân quyền cho IAM user

* Ấn Create group

A screenshot of a computer

Description automatically generated

* Đặt tên User group
* ở phần Permissions policies, tìm kiếm “s3” và chọn “AmazonS3FullAcess”
* Ấn Create user group

A screenshot of a computer

Description automatically generated

* Chọn User group vừa mới tạo và ấn “Next”

A screenshot of a computer

Description automatically generated

Bước 1.6: Chọn Create user

A screenshot of a computer

Description automatically generated

Bước 1.7: Chọn Download .csv file để tải file .csv chứa thông tin đăng nhập và chọn “Return to users list” để hoàn thành tạo IAM user.

A screenshot of a computer

Description automatically generated

**Bước 2: Tạo cặp khoá truy cập**

Bước 2.1: Chọn user vừa tạo trong danh sách các IAM user

A screenshot of a computer

Description automatically generated

Bước 2.2: Chọn Create access key

A screenshot of a computer

Description automatically generated

Bước 2.3: Chọn Third-party service và Next

A screenshot of a computer

Description automatically generated

Bước 2.4: Thêm Description tag và chọn Create access key

A screenshot of a computer

Description automatically generated

Bước 2.5: Chọn Download .csv file để lưu file csv chứa thông tin cặp khoá vì cặp khoá này chỉ có thể thấy được 1 lần duy nhất. Sau đó chọn Done để hoàn thành việc tạo cặp khoá truy cập

A screenshot of a computer

Description automatically generated

**Bước 3: Tạo nơi lưu trữ dữ liệu trước xử lí và nơi lưu trữ dữ liệu sau xử lí trong S3**

Chúng ta tạo 2 bucket trên S3 để làm nơi lưu dữ liệu , 1 cho staging, 1 cho destination storage

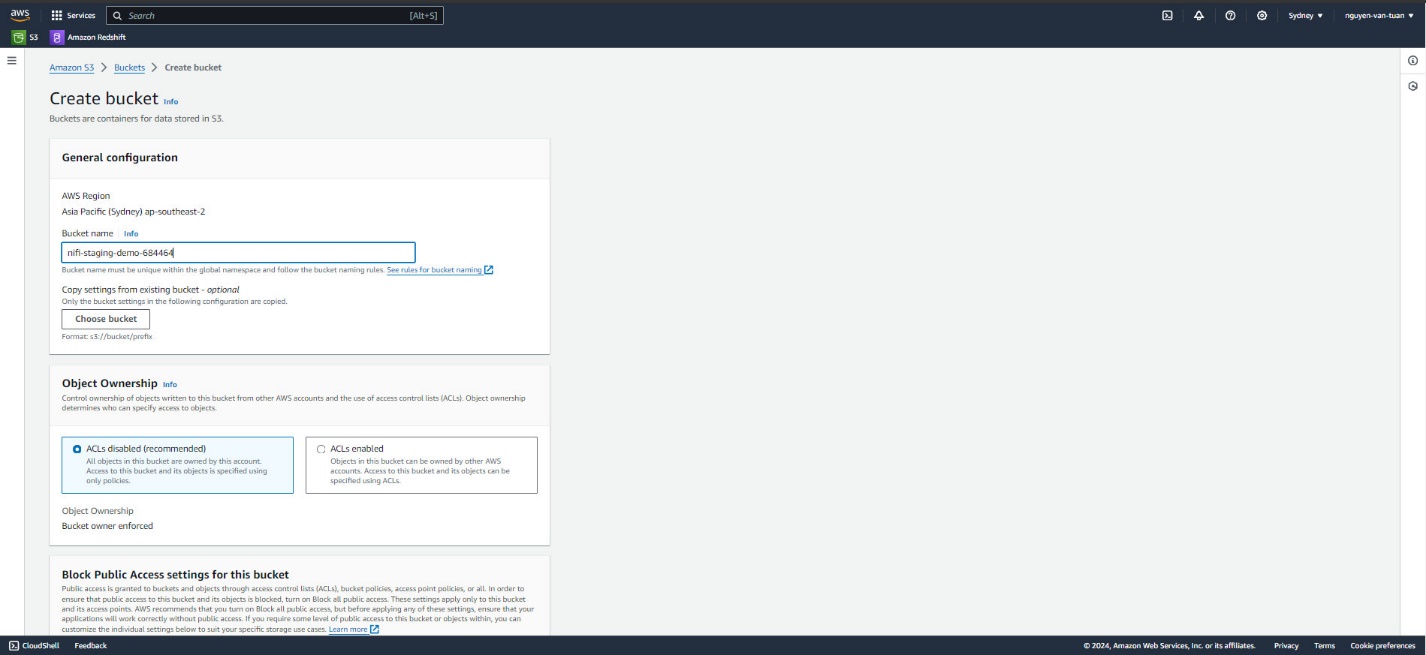
Bước 3.1: Lên thanh tìm kiếm gõ S3 và chọn “S3”

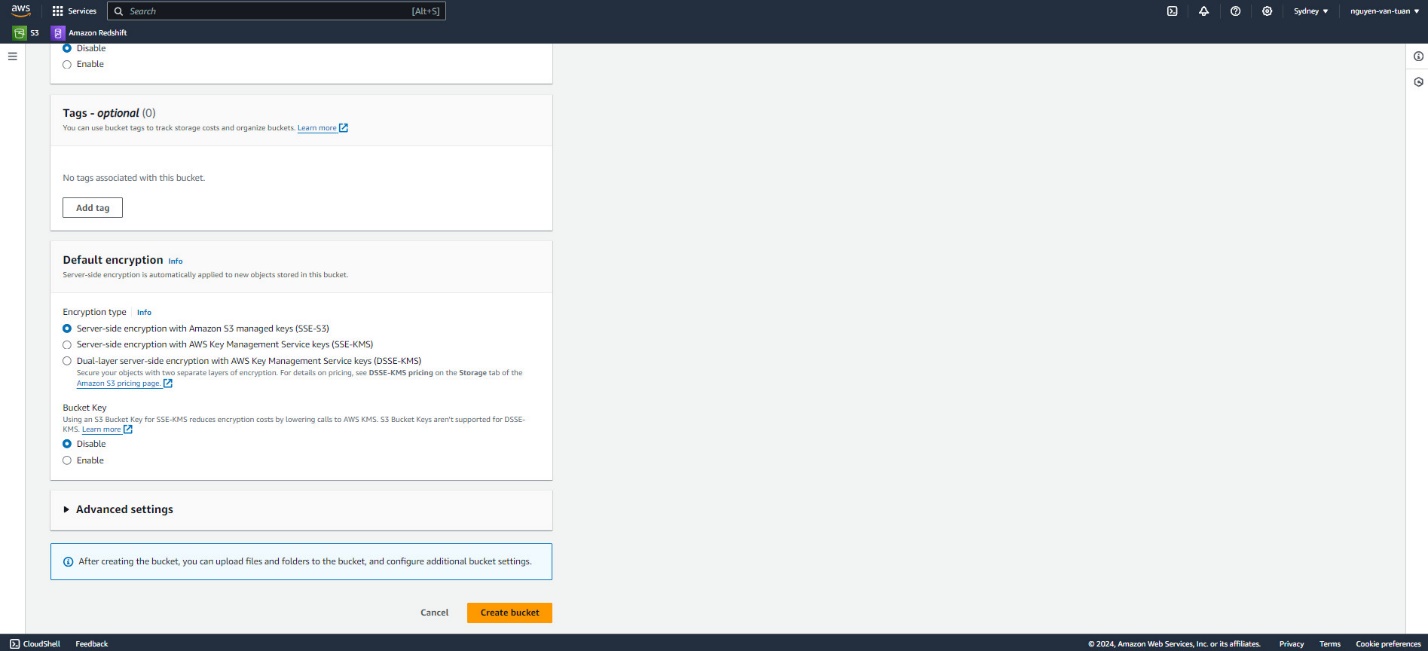
Bước 3.2: Trong giao diện S3 chọn create bucket

A screenshot of a computer

Description automatically generated

Bước 3.3: Điền “Bucket name” và thiết lập như mẫu, sau đó chọn create bucket để hoàn thành việc tạo bucket





Chúng ta tạo 2 bucket tương tự như sau



**Bước 4: Triển khai luồng trên Apache Nifi**

Bước 4.1: Đầu tiên tạo một Process Group bằng việc kéo và thả vào giao diện, sau đó hãy đặt tên cho nó. Process Group này sẽ chứa các Processor và các kết nối giữa Processor của chúng ta, việc tạo Process Group giúp chúng ta quản lí dễ dàng hơn

A screenshot of a computer

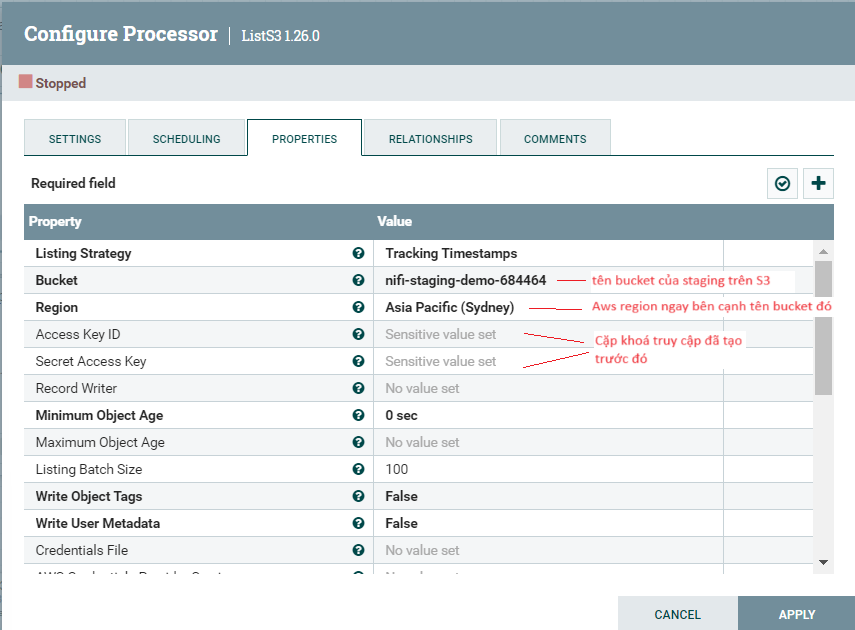
Description automatically generated

Bước 4.2: Double-click vào Process Group vừa tạo, chúng ta sẽ kéo mục Processor để thêm các Processor và cấu hình nó như sau:

* ListS3:
* Processor này được sử dụng để liệt kê các đối tượng (objects) trong một bucket S3 cụ thể. Nó tạo ra các FlowFile đại diện cho từng đối tượng được liệt kê.

A screenshot of a computer

Description automatically generated



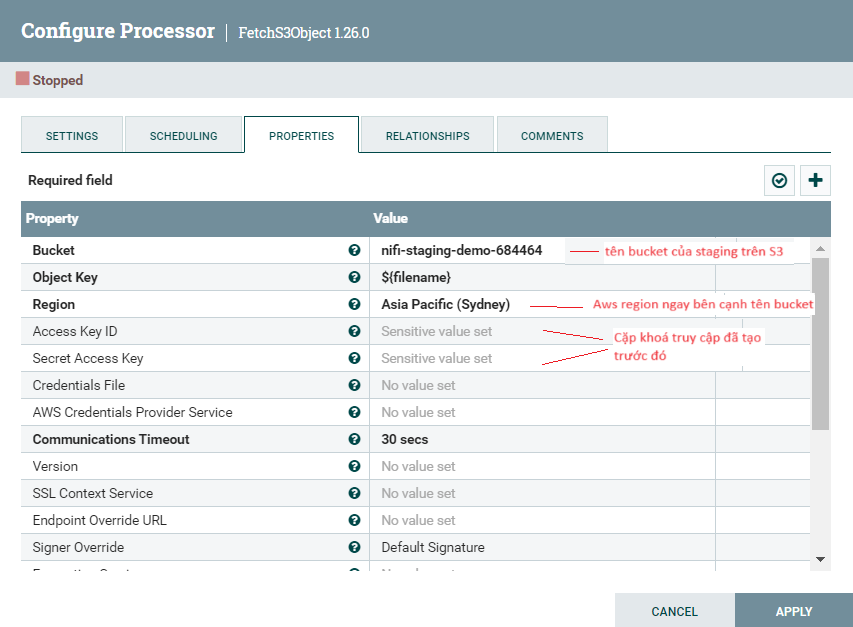
A screenshot of a computer

Description automatically generated

* Chọn “Apply” để hoàn tất thiết lập
* FetchS3Object:
* Processor này dùng để tải về (fetch) một đối tượng cụ thể từ S3 dựa trên thông tin được cung cấp (như tên bucket và key của đối tượng).

A screenshot of a computer

Description automatically generated



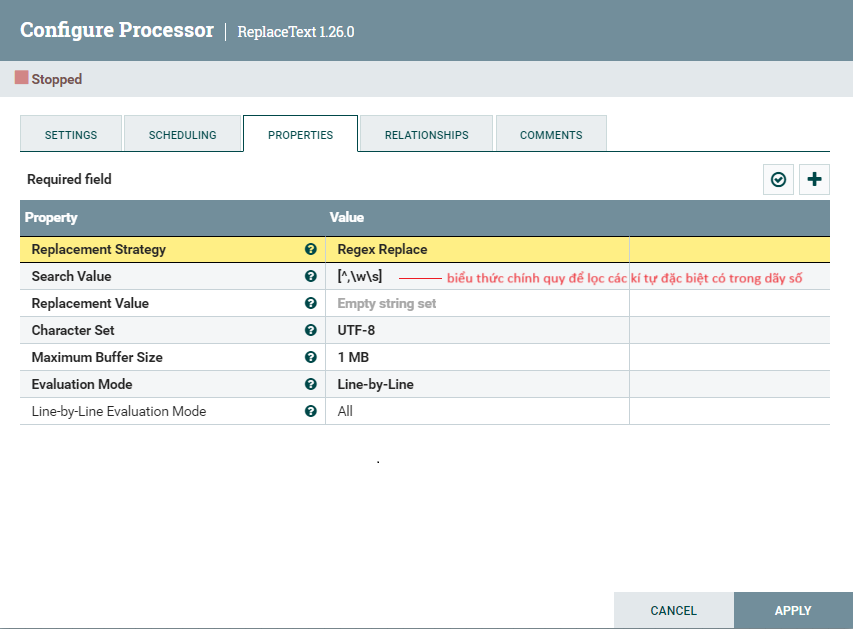
A screenshot of a computer

Description automatically generated

* Chọn “Apply” để hoàn tất thiết lập
* ReplaceText :
* Processor này dùng để thay thế văn bản trong nội dung của FlowFile dựa trên các quy tắc hoặc mẫu (patterns) được định nghĩa

A screenshot of a computer

Description automatically generated



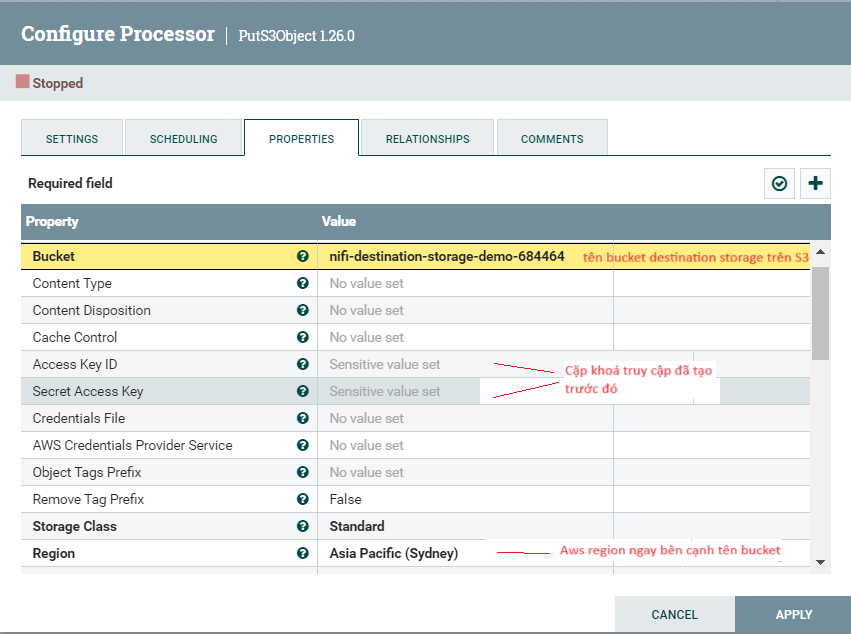
A screenshot of a computer

Description automatically generated

* Chọn “Apply” để hoàn tất thiết lập
* PutS3Project:
* Processor này dùng để tải lên (put) một FlowFile thành một đối tượng trong S3.

A screenshot of a computer

Description automatically generated



A screenshot of a computer

Description automatically generated

* Chọn “Apply” để hoàn tất thiết lập

Bước 4.3: Kết nối các Processor để triển khai luồng

Chúng ta di chuột và nhấn kéo biểu tượng mũi tên từ processor này sang processor khác để tạo kết nối giữa chúng. Khi giao diện connection hiện lên, hãy chọn hết các mục ở “For Relationships”, ví dụ:

A screenshot of a computer

Description automatically generated

Đây là luồng chúng ta sẽ triển khai:

A screenshot of a computer

Description automatically generated

=> hoàn thành dựng luồng

**Bước 5: Khởi chạy luồng**

Bước 5.1: Chuột phải vào khoảng trống trong giao diện và chọn “Start” để luồng bắt đầu hoạt động

A screenshot of a computer

Description automatically generated

Bước 5.2: Chúng ta vào lại bucket staging trong S3, thêm vào file .csv gồm id và số điện thoại có chứa kí tự đặc biệt

A screenshot of a computer

Description automatically generated

File to test: <https://drive.google.com/drive/folders/1hYNYvo8SYSGFJLgvRlYjY8-qo9rxBbE_?usp=sharing>

* Chọn vào bucket staging
* Chọn “Upload”
* Kéo thả file vào hoặc chọn “Add files” để thêm file

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

Sau khi thêm , chúng ta sẽ thấy luồng trong Nifi đang hoạt động

Vào bucket destination storage và chúng ta thấy dữ liệu sau xử lí đã được đẩy vào

A screenshot of a computer

Description automatically generated

Ấn vào file trong destination storage và ấn download để tải nó về. Chúng ta thấy dữ liệu trong file .csv đã được xử lí

A screenshot of a computer

Description automatically generated

Như vậy chúng đã triển khai thành công Usecase 1