**Infrastructure Setup: S3, SQS, and Lambda Orchestration**

This Terraform project provisions a serverless AWS data pipeline. It automates daily ingestion of external datasets, stores them in S3, triggers downstream processing through SQS, and runs reporting workflows via Lambda functions. Everything is fully managed with Terraform for reproducibility and easy onboarding.

**Architecture Overview**

* **S3 Bucket** (vm-rearc-data-buckets) – central storage for raw and processed data
* **SQS Queue** (vm-rearc-processing-queue) – receives notifications when new .jsonl files are uploaded under api/
* **IAM Role & Policies** – shared Lambda execution role with permissions for S3, SQS, and CloudWatch Logs
* **Lambdas**
  + Daily Combined Lambda – runs once per day via EventBridge, fetches data from BLS, writes to S3
  + Reporting Lambda – triggered by SQS messages, processes new files, stores results in S3
  + Presign Lambda – generates presigned URLs for secure, temporary file access
* **Event Triggers**
  + S3 → SQS → Reporting Lambda
  + EventBridge → Daily Lambda

**Prerequisites**

* Terraform version 1.5.0 or higher
* AWS provider version ~> 5.0
* AWS CLI configured for region us-east-1
* Lambda deployment packages in the module directory:
  + lambda\_daily\_combined.zip
  + lambda\_reporting.zip
  + generate\_presigned\_url.zip

**Deployment**

1. Initialize Terraform with terraform init
2. Review the plan with terraform plan
3. Apply changes with terraform apply
4. Verify in AWS Console: S3 bucket, SQS queue, and Lambdas are deployed and EventBridge rule is scheduled

**Environmental Variables**

* **Daily Lambda**: DEST\_BUCKET, DEST\_PREFIX, BASE\_URL, CONCURRENCY, TIMESTAMPED
* **Reporting Lambda**: DEST\_BUCKET
* **Presign Lambda**: BUCKET\_NAME

**Cleanup**

Run terraform destroy to remove all resources.

Snap short of S3 bucket upload through Terraform Configurations

A screenshot of a computer

AI-generated content may be incorrect.

Preassigned URL to access the bucket: (link expires after 7 days))

[preassigned URL](https://vm-rearc-data-buckets.s3.us-east-1.amazonaws.com/api/population_data_20251021T135658Z.jsonl?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAZAGFJT2VKX2GVW3P%2F20251021%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20251021T141701Z&X-Amz-Expires=604800&X-Amz-SignedHeaders=host&X-Amz-Signature=9e0b6a82b4737815db50a419bc5ec90f3ec9f19756eb847bb5d1ff7ecb8abba8)

CLI Command to generate this preassigned url:

aws s3 presign s3://vm-rearc-data-buckets/api/population\_data\_20251021T135658Z.jsonl --expires-in 604800