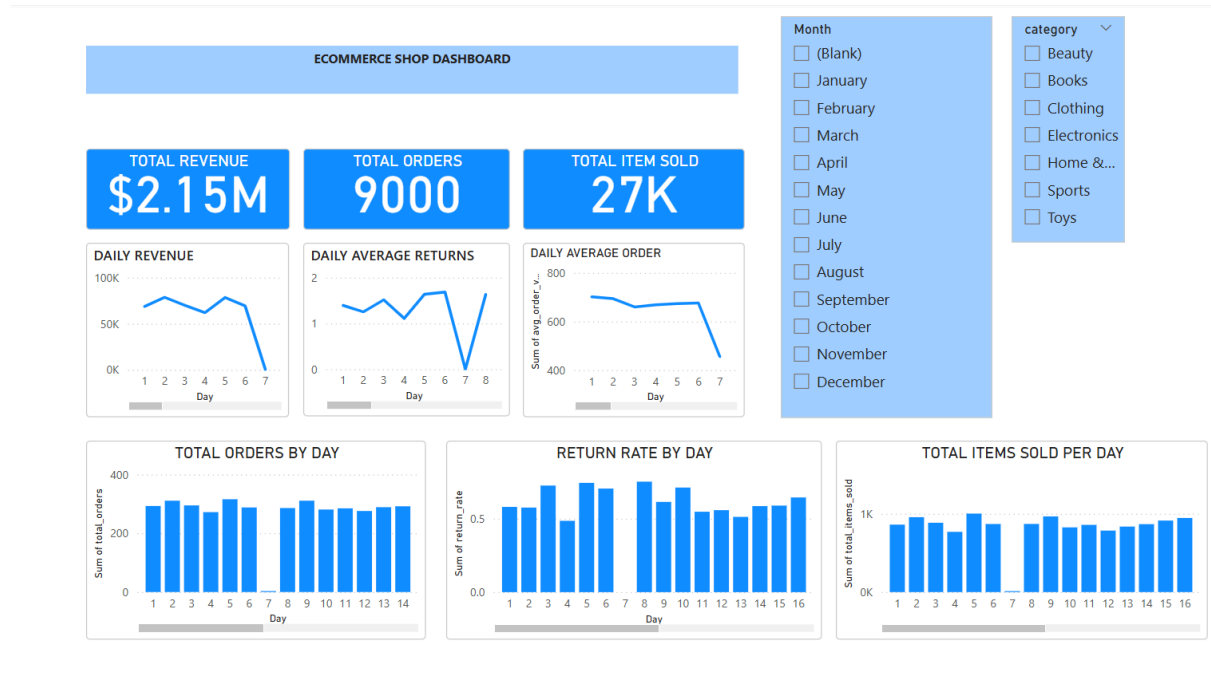


EVENT-DRIVEN PIPELINE FOR AN E-COMMERCE SHOP

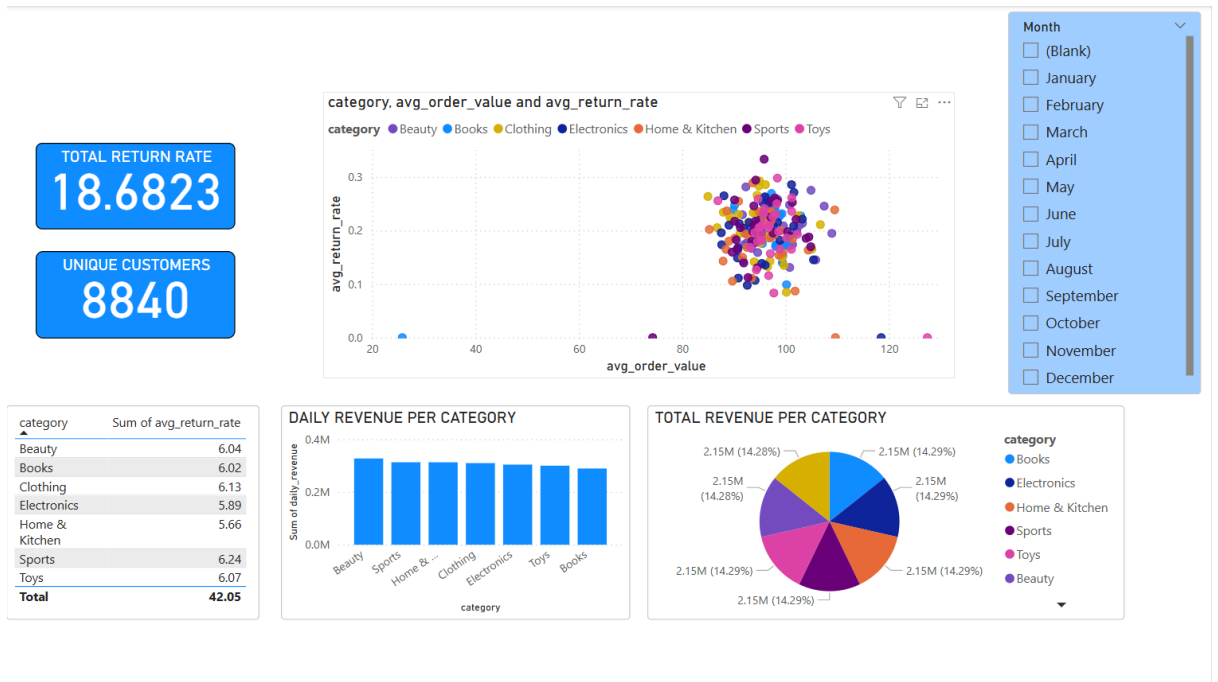
A company wants to deal with a continuous inflow of transactional data (orders, products, order_items) in flat forms. These files arrive randomly in an **Amazon S3** bucket. The goal is to validate, transform, and compute business KPIs and store results for real-time querying via **Amazon DynamoDB**.



BUSINESS KPIs

1. CATEGORY-LEVEL KPIs (Per Category, Per Day)

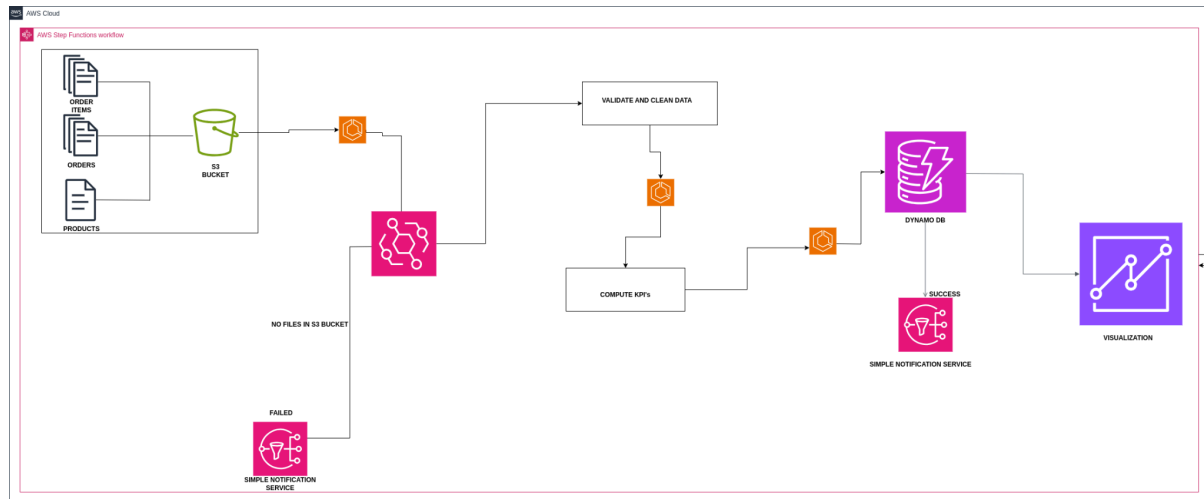
| Field | Description |
|-----------------|----------------------------------------------------|
| category | Product category |
| order_date | Date of the summarized orders |
| daily_revenue | Total revenue from that category for the day |
| avg_order_value | Average value of individual orders in the category |
| avg_return_rate | Percentage of returned orders in the category |



2. ORDER-LEVEL KPIs(Per Day)

| Field | Description |
|------------------|------------------------------------------------|
| order_date | Date of the summarized orders |
| total_orders | Count of unique orders |
| total_revenue | Total revenue from all orders |
| total_items_sold | Total number of items sold |
| return_rate | Percentage of orders that were returned |
| unique_customers | Number of distinct customers who placed orders |

FLOW DIAGRAM

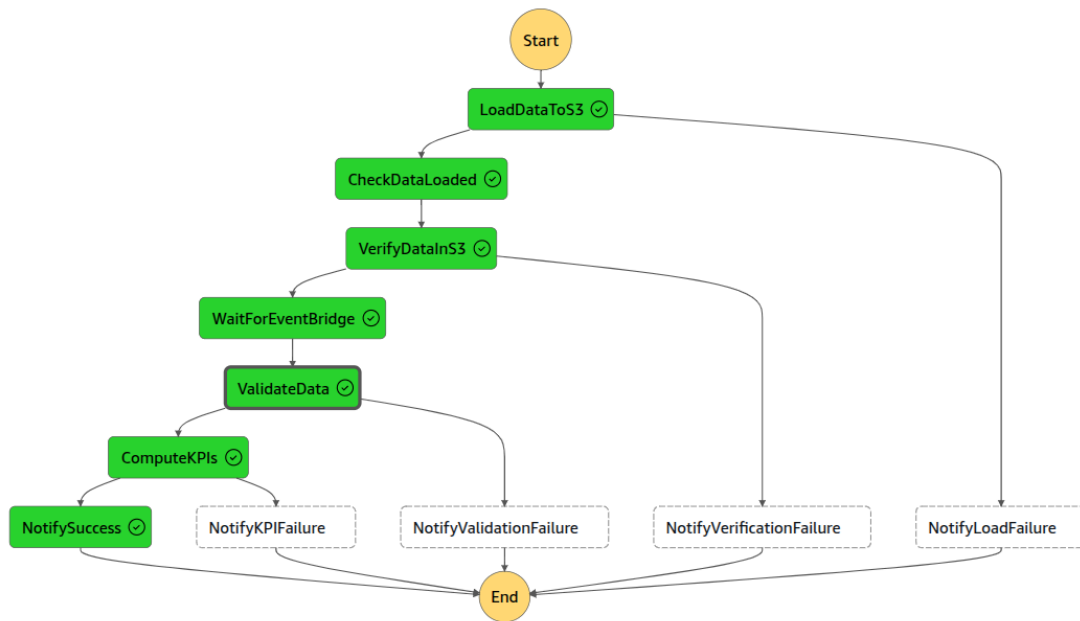


ORCHESTRATION

Overview

This Step Function orchestrates an E-Commerce data processing pipeline using AWS ECS Fargate tasks. The workflow loads data to S3, verifies its presence, validates the data, computes KPIs, and handles success/failure notifications through SNS.

FLOW DIAGRAM



State Definitions

1. LoadDataToS3

- Type: Task (ECS RunTask)
- Purpose: Initiates data loading to S3 bucket

2. CheckDataLoaded

- Type: Wait
- Purpose: Pauses for 30 seconds after data load

3. VerifyDataInS3

- Type: Task (EventBridge PutEvents)
- Purpose: Triggers EventBridge rule to verify S3 data

4. Wait For EventBridge

- Type: Wait
- Purpose: Pause for 40 seconds for EventBridge verification

5. Validate Data

- Type: Task (ECS RunTask)
- Purpose: Runs data validation process

6. Compute KPIs

- Type: Task (ECS RunTask)
- Purpose: Calculate key performance indicators

Notification System

Success Notification

SNS Topic:

Trigger: After successful completion of all tasks

Message: Pipeline completion confirmation

Failure Notifications

All failure notifications use topic