AWS Architecture Design for E-commerce Financial Data Pipeline

Introduction:

This document outlines a production-ready AWS architecture to support the financial data processing needs of an e-commerce platform. The pipeline ingests, validates, transforms, and analyzes transaction, order, and chargeback data to enable financial reporting and fraud detection.

Architecture Diagram:

[See next page]

Component Descriptions:

1. Amazon S3
   - Acts as the central data lake.
   - Stores raw data from e-commerce systems (transactions.json, orders.json, chargebacks.csv).
   - Also used for storing processed/enriched output and logs.

2. Amazon EventBridge
   - Triggers on new file uploads in S3.
   - Starts pipeline jobs through Lambda or Step Functions automatically.

3. AWS Lambda
   - Handles data validation (Pydantic), light transformations, and schema enforcement.
   - Scales automatically with file arrival rate.

4. AWS Step Functions (or AWS Glue)
   - Coordinates pipeline stages: validation -> enrichment -> aggregation.
   - AWS Glue is preferred for large-scale batch processing; Step Functions for flexibility.

5. Amazon Athena
   - Performs SQL queries directly on S3-based Parquet/CSV outputs.
   - Enables rapid business metric extraction without provisioning compute.

6. Amazon Redshift (optional)

- Serves as a data warehouse for structured financial analytics or BI integration.

7. Amazon CloudWatch

- Aggregates logs and metrics from Lambda, Glue, and Step Functions.
- Provides centralized observability and alerting.

8. Amazon SNS

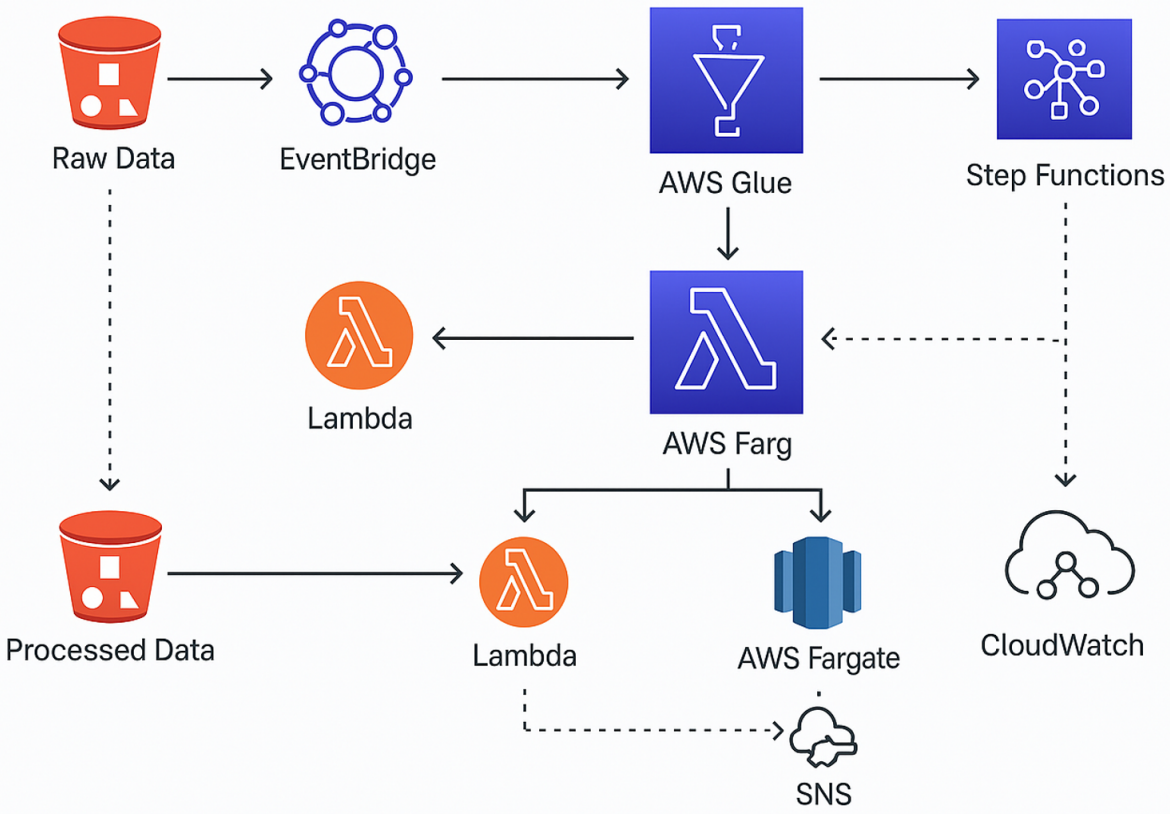
- Delivers alerts to developers or operations teams on job failures or key events.

Design Highlights:

- Fully serverless and auto-scaling.
- Cost-optimized via pay-per-use services and S3 storage lifecycle policies.
- Modular pipeline with decoupled stages for flexibility and fault isolation.
- Easily extended to real-time streaming with Amazon Kinesis if needed.

Use Case Support:

- Daily transaction volume/value reporting
- Chargeback rate and fraud detection
- Payment method performance analysis
- Failed transaction diagnostics



AWS Architecture for E-Commerce Data Pipeline