

AirBnB Stream Data Ingestion

Objective

Build a simulated data pipeline for Airbnb booking data that integrates various AWS services, demonstrating real-time data processing, filtering, and storage.

Prerequisites

- AWS Account
- Basic understanding of AWS Lambda, Amazon SQS, Amazon S3, AWS CodeBuild, and Amazon EventBridge.
- Familiarity with AWS CLI or AWS Management Console.
- Knowledge of programming (preferably Python for Lambda functions).

Assignment Steps

Part 1: Create an Amazon SQS Queue with DLQ

- Create an SQS Standard Queue named **AirbnbBookingQueue**.
- Setup a Dead Letter Queue (DLQ): Create another SQS queue named **AirbnbBookingDLQ**. Configure the **AirbnbBookingQueue** to send messages to **AirbnbBookingDLQ** after 3 unsuccessful delivery attempts.

Part 2: Create Producer Lambda Function

- Lambda Function - Producer: Create a Lambda function named **ProduceAirbnbBookingData**. This function will generate mock Airbnb booking data and publish it to **AirbnbBookingQueue**.
- Suggested Mock Data Structure:

```
{
  "bookingId": "UUID",
  "userId": "UserID",
  "propertyId": "PropertyID",
  "location": "City, Country",
  "startDate": "YYYY-MM-DD",
  "endDate": "YYYY-MM-DD",
  "price": "Price in USD"
}
```

Part 3: Setup EventBridge Pipe

- EventBridge Pipe: Create an EventBridge Pipe to consume messages from **AirbnbBookingQueue**. Filter messages where the booking duration is more than 1 day.

- Filtering Logic: Use the **startDate** and **endDate** to calculate the booking duration.

Part 4: Create Destination Lambda Function

- Lambda Function - Consumer: Create a Lambda function named **ProcessFilteredBookings**. This function will be triggered by the EventBridge Pipe and will write the filtered records to an S3 bucket.
- S3 Bucket: Ensure the S3 bucket is created beforehand to store the booking records. The bucket can be named **airbnb-booking-records**.

Part 5: Implement CI/CD using AWS CodeBuild

- AWS CodeBuild Project: Setup an AWS CodeBuild project to automate the deployment of Lambda function code changes. Use a source repository of your choice (e.g., GitHub) for hosting the Lambda function code.
- Configure build specifications to install dependencies, package the Lambda function, and deploy it to AWS.
- Ensure the IAM role used by CodeBuild has the necessary permissions for Lambda deployment.

Deliverables

- A link to the source code repository for the Lambda functions.
- Screenshots or logs demonstrating successful data flow from the producer Lambda through SQS, EventBridge Pipe, to the consumer Lambda, and finally to the S3 bucket.
- A reflection on the assignment, including challenges faced and how they were overcome.