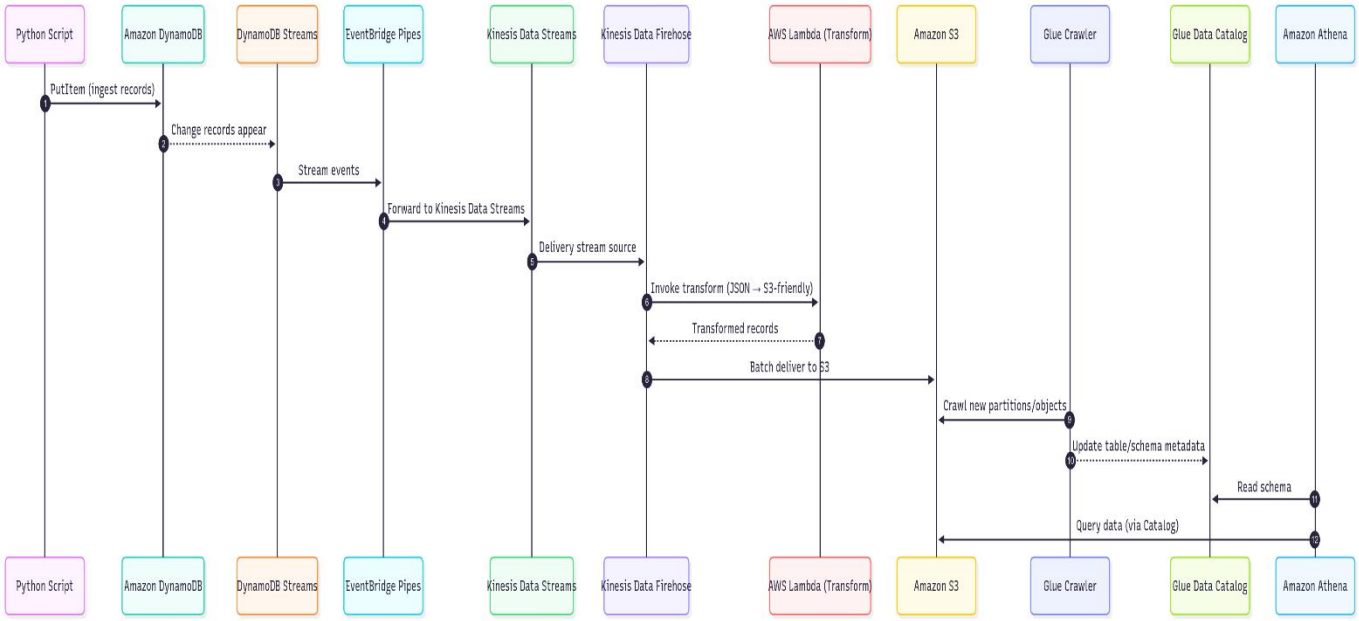
CDC Data Pipeline for Sales Data Analysis

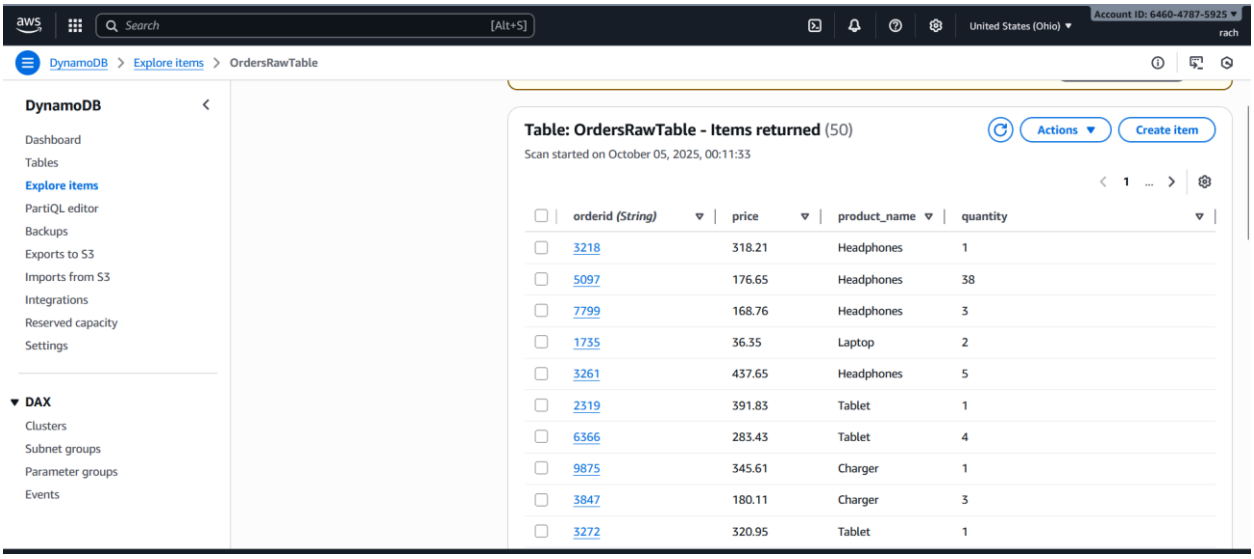
This real time streaming pipeline aims to capture data changes in Dynamodb using Kinesis Data Streams,Kinesis Firehose,Glue, Lambda,S3 and Eventbridge. Finally, Athena is used to run queries for data analysis.

Architecture Diagram



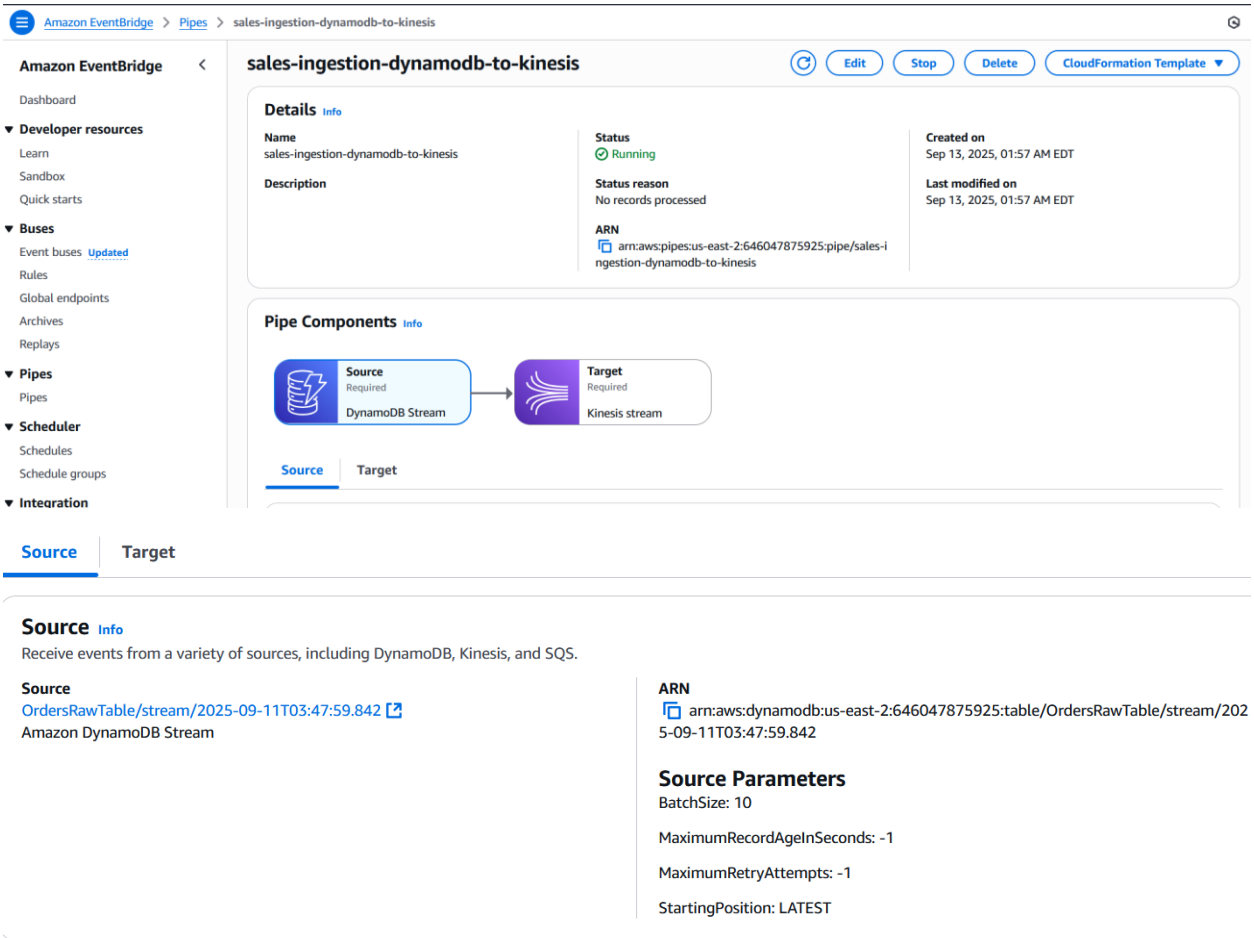
Step 1: Create records in Dynamodb using Python script.

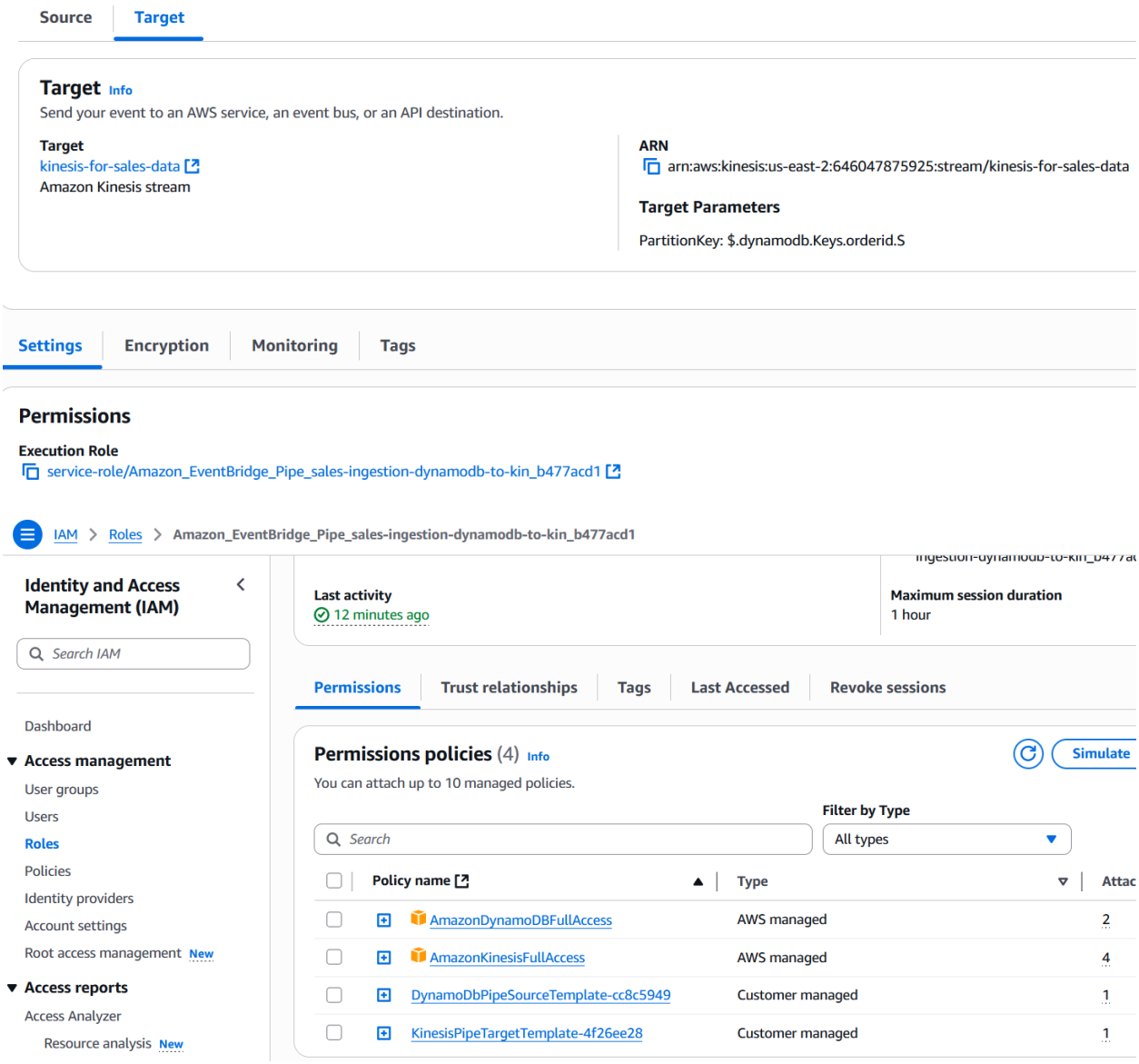
Python with boto3 pushes data to dynamodb.



Step 2: Deliver stream of data to Kinesis Data Stream

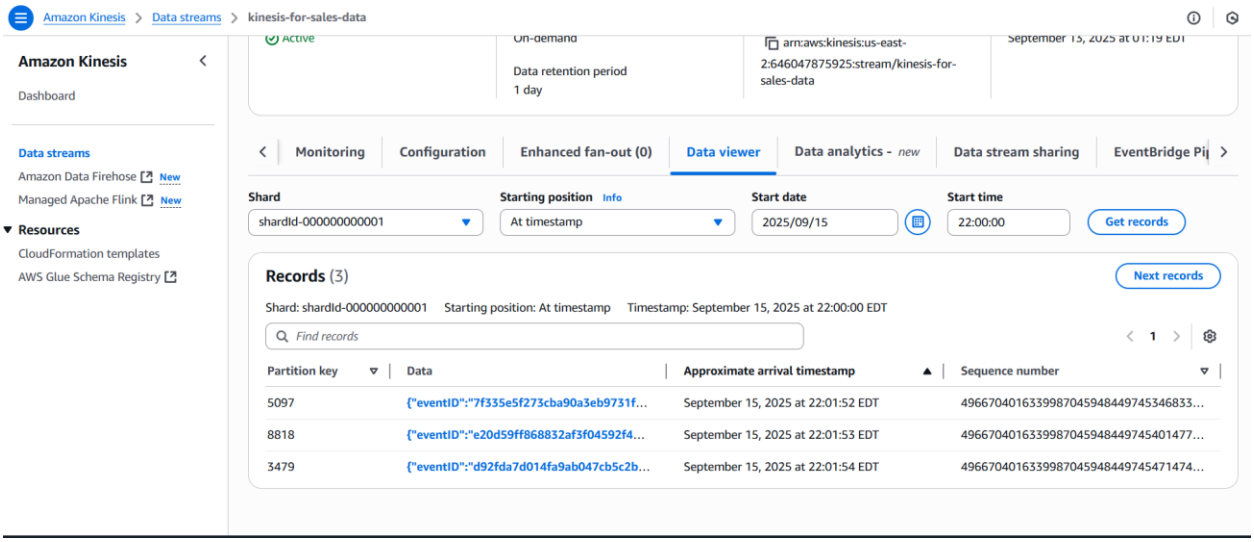
The stream is then delivered by EventBridge pipe to Kinesis Data stream. Setup an IAM role for eventbridge with access to DynamoDB and Kinesis.



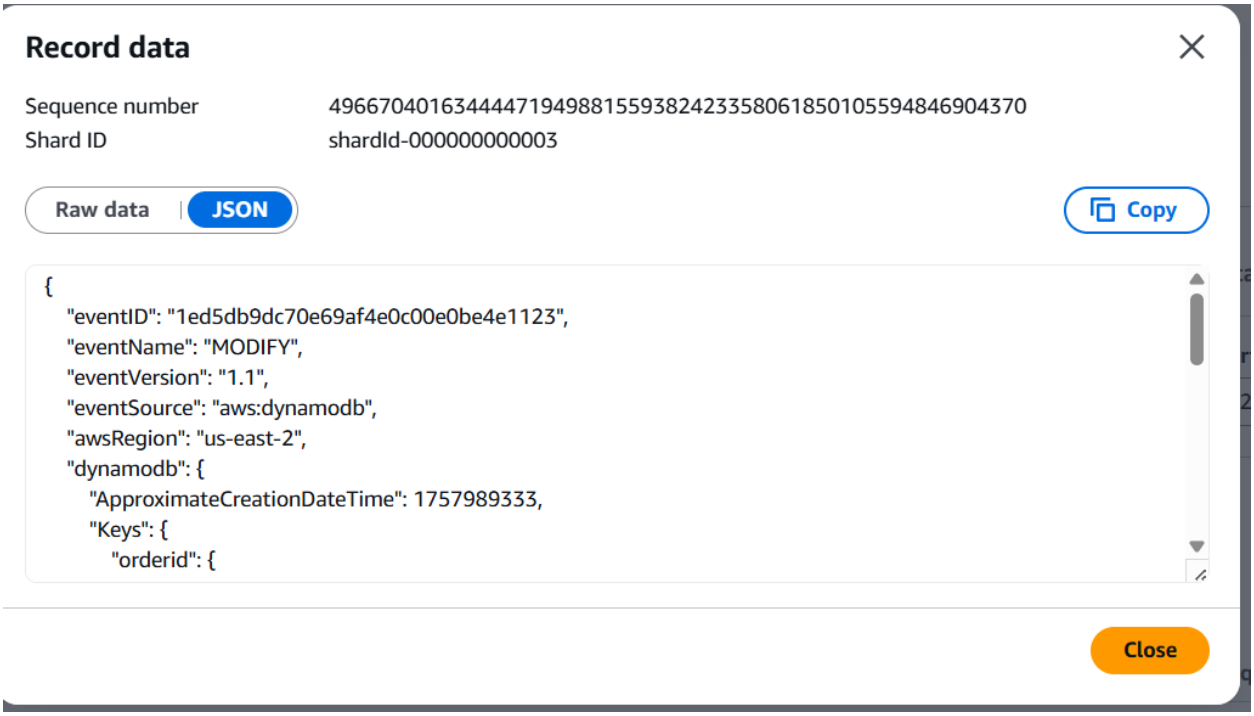


Step 3: Create a Kinesis data stream to view the change data capture performed in DynamoDB

Choose the shard and accurate timestamp to view streamed data. Event name “Modify” implies the change made in data .

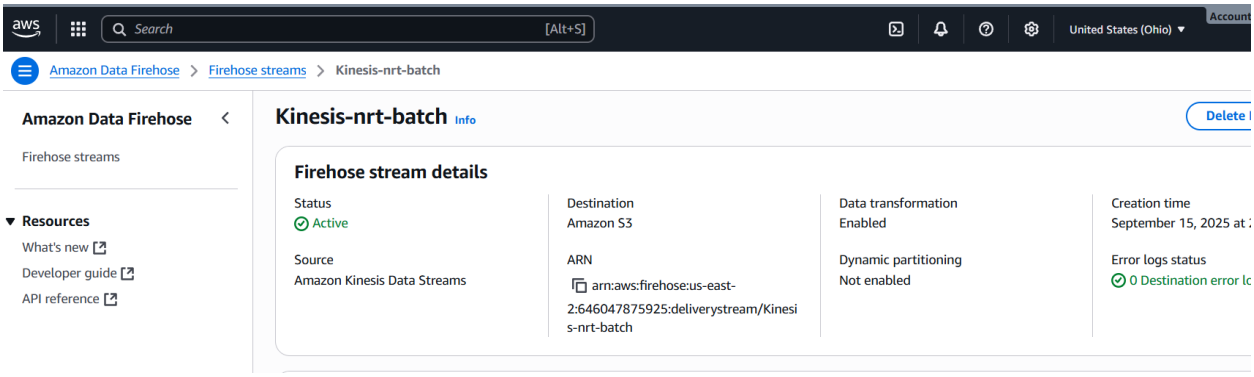


Change data Capture



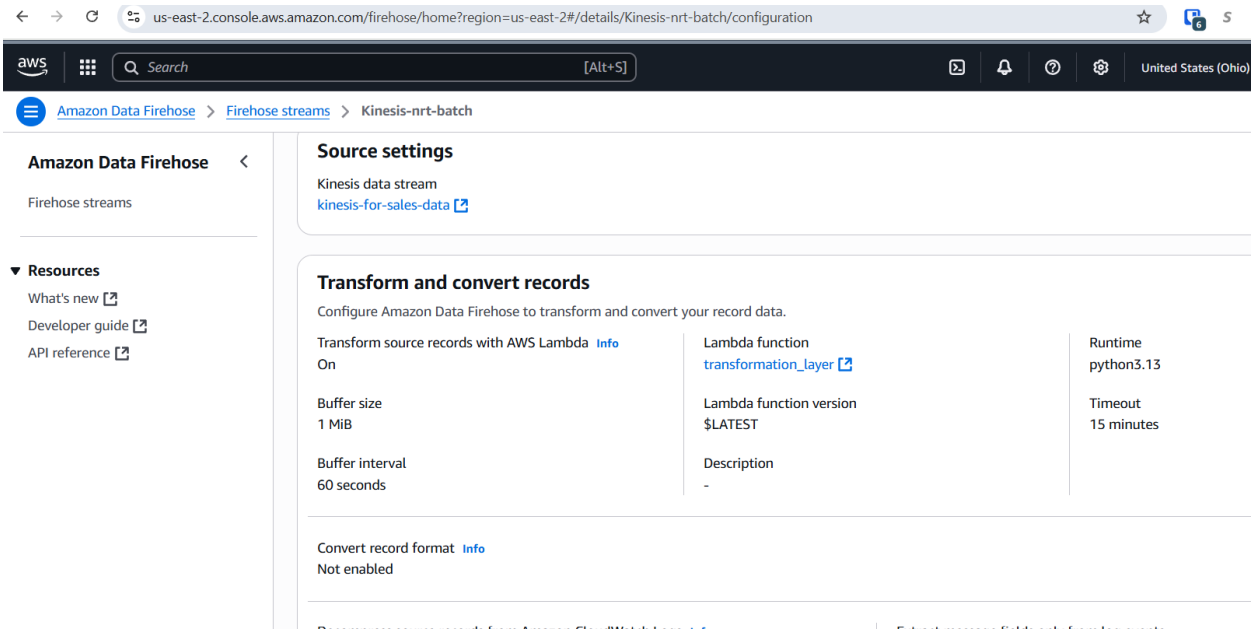
Step 4: Create Kinesis Data Firehose

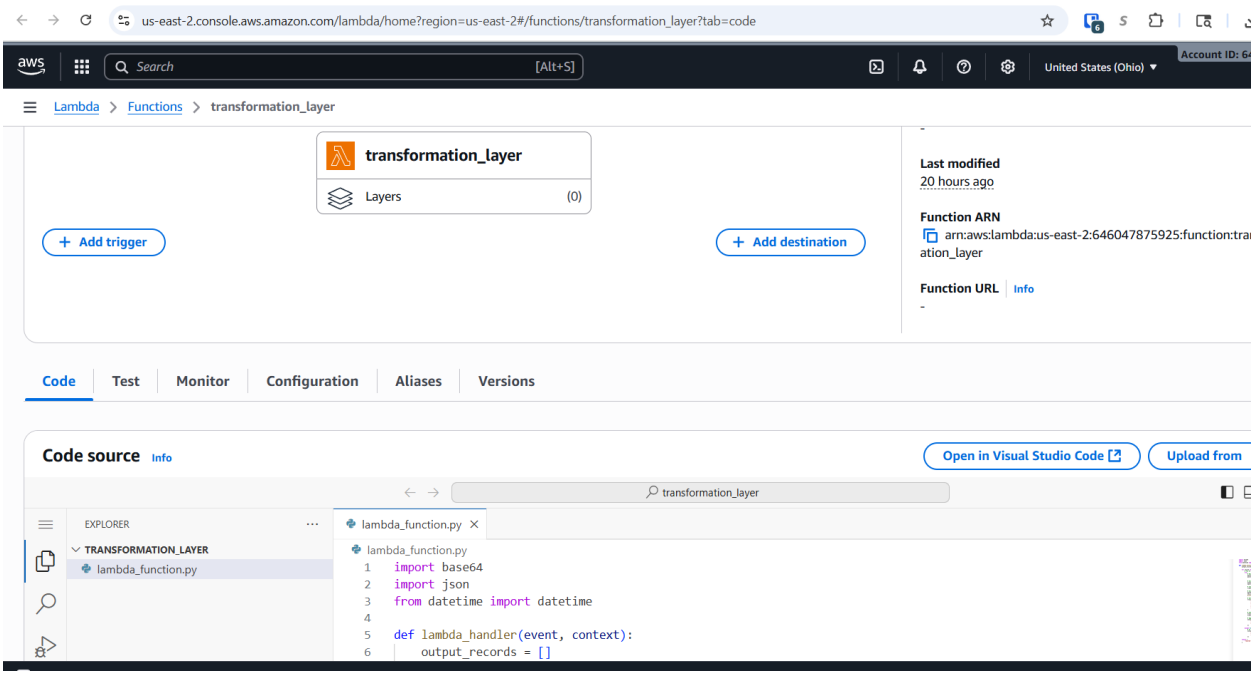
The purpose of kinesis firehose is to Batch data coming from kinesis stream and dump into target. Here source is kinesis data stream and target is S3.



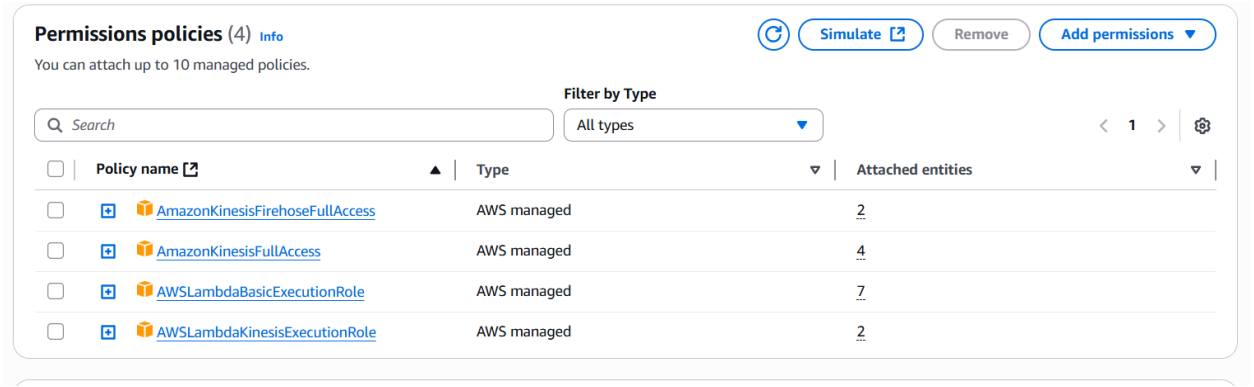
Step 5: Transformation using Lambda

Lambda function is using to transform the data from source kinesis data stream and then push it back to firehose. We use lambda transformation since the data coming from kinesis data stream is in json format and we want to push it to S3 in tabular format to write athena queries. All kind of transformations can be done using lambda

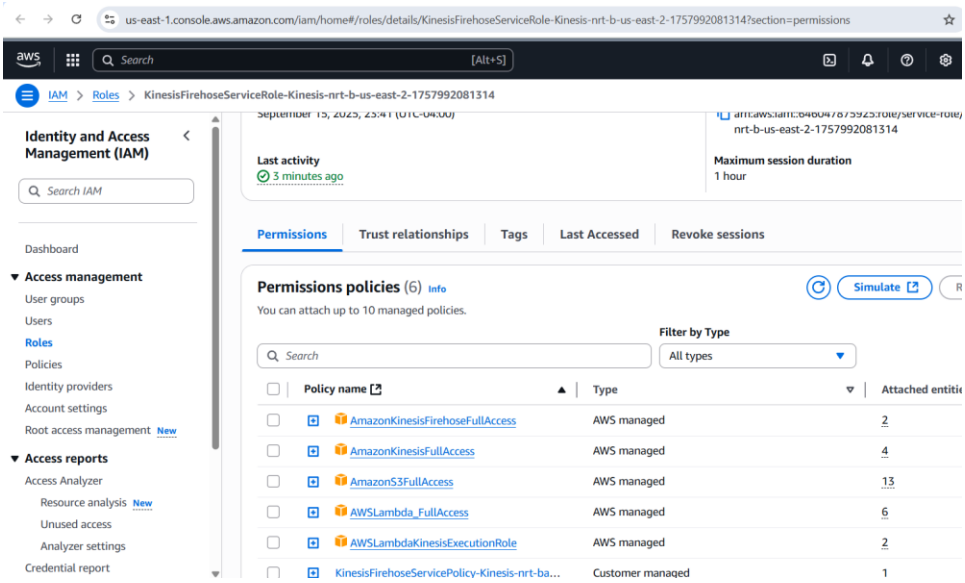




Lambda code selects order id and other values from New image of kinesis data stream creates a new json and then pushes that back firehose which then dumps to S3. IAM roles for lambda function

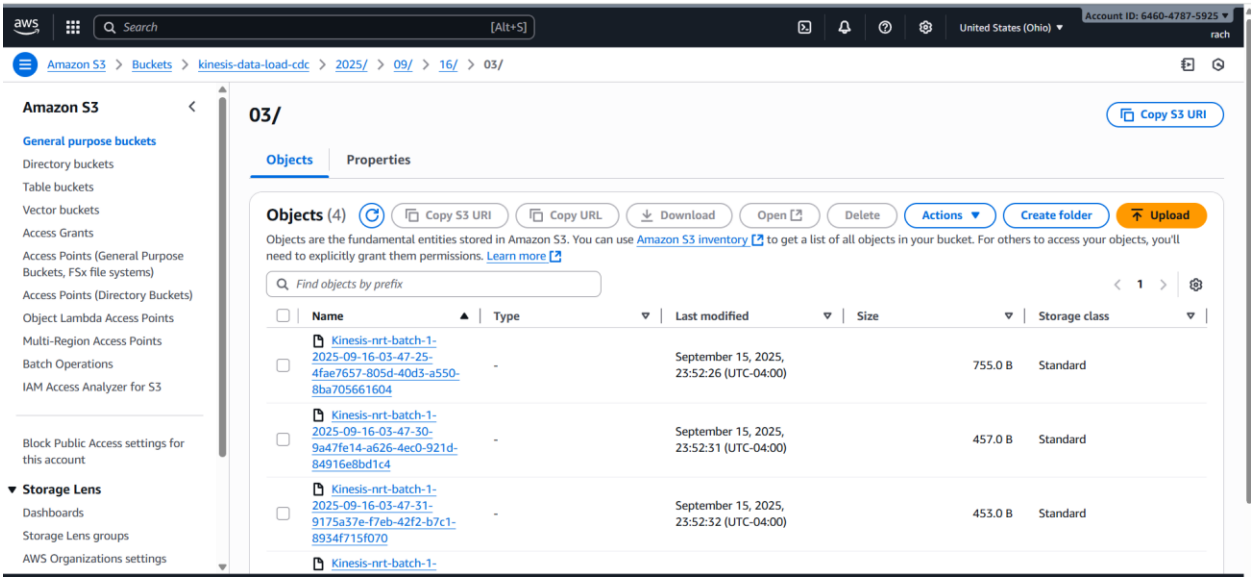


IAM role for kinesis firehose



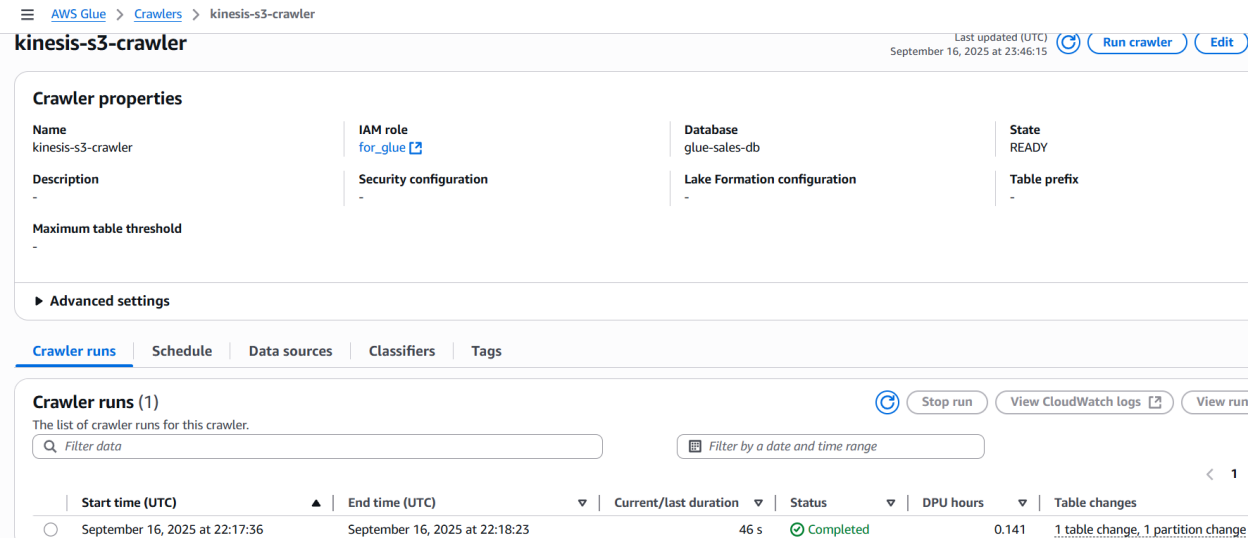
Step 6: Target S3 bucket is created where data gets dumped in real time

S3 bucket contents in near real time

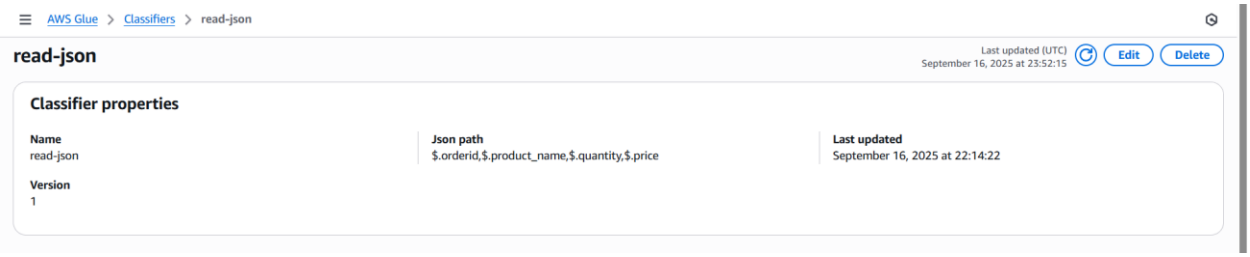


Step 7: Create a Glue crawler and catalog to crawl the contents in S3

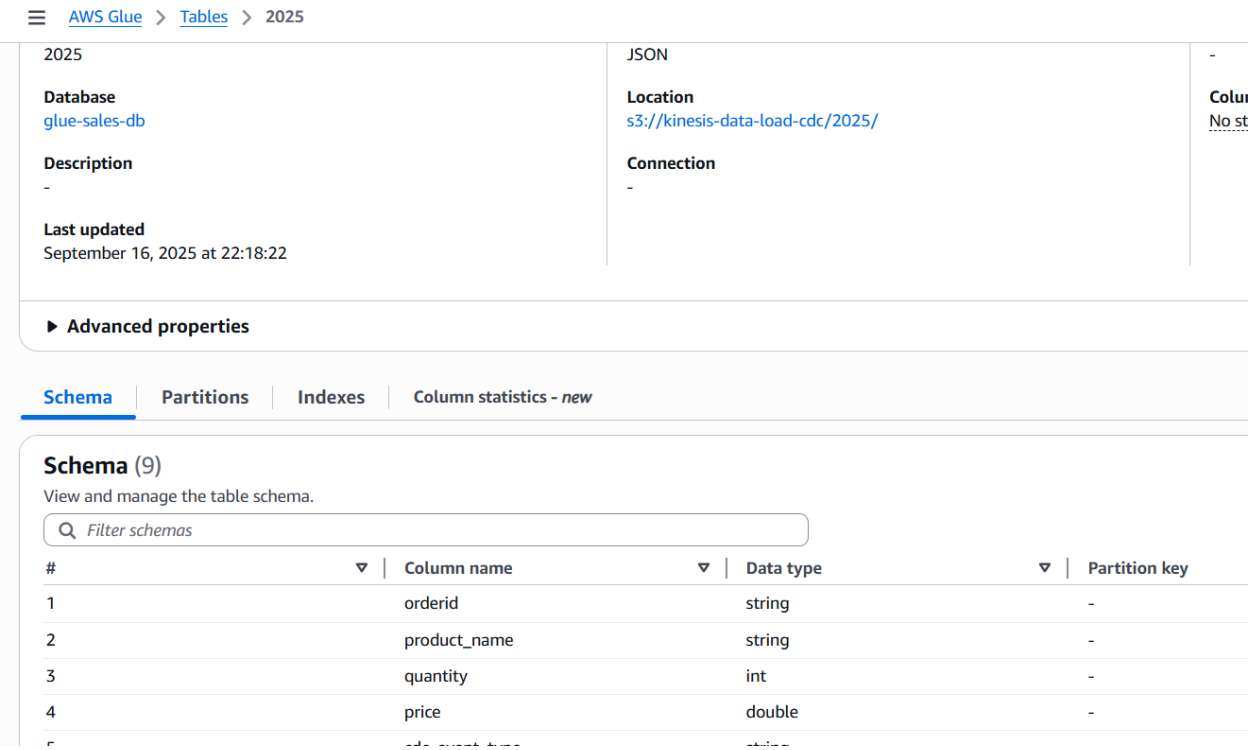
Glue crawler to crawl data in S3 so that its ready for Athena querying



Custom classifier for Glue crawler



Glue table created



Step 8: Open Athena query editor to run queries against glue catalog table.

This data is used for analysis

Athena query