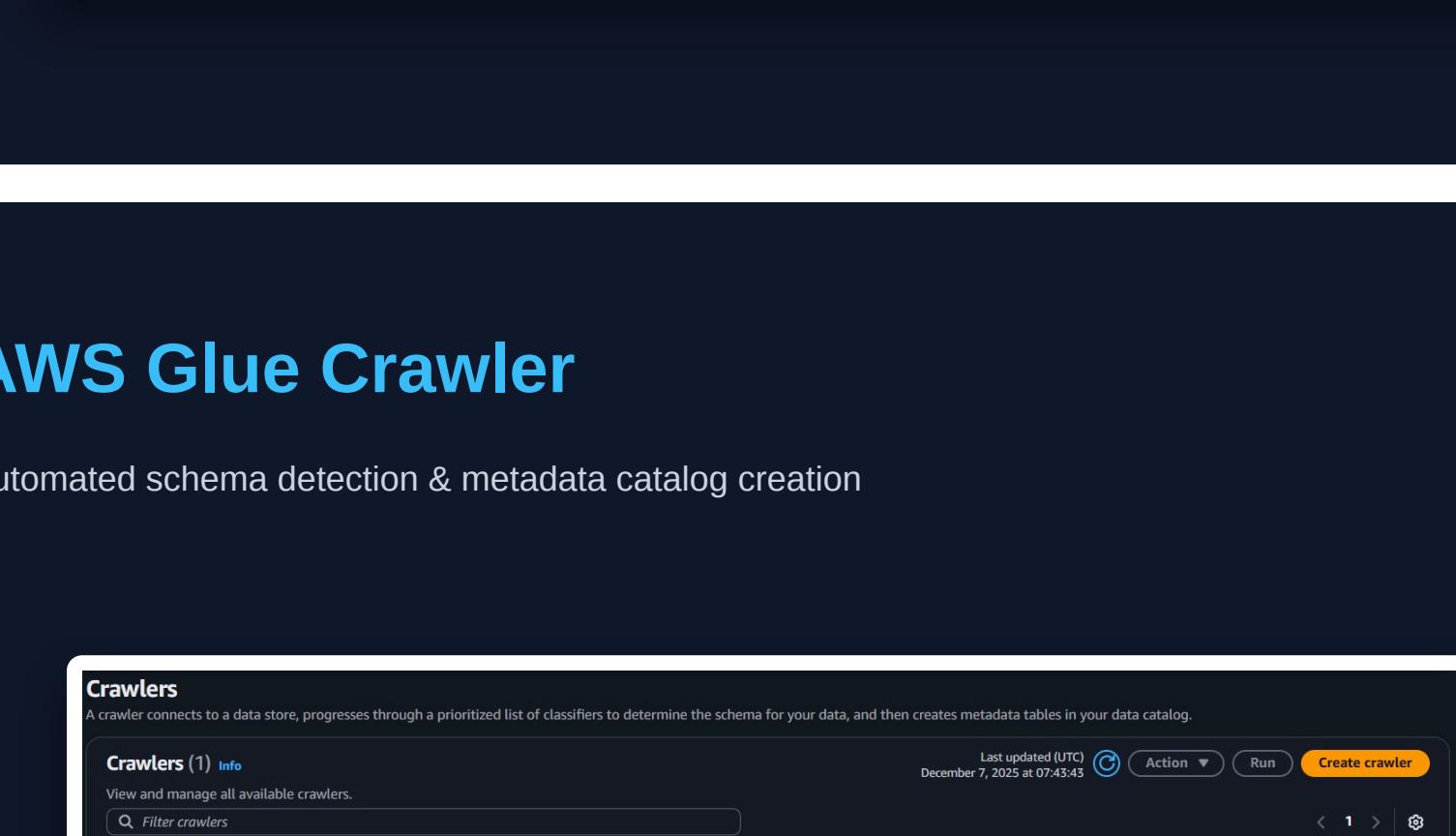


# AWS Serverless ETL Pipeline

Data Catalog • Lambda • Athena

Serverless Data Engineering Architecture

## Complete Event Driven Data Pipeline



## AWS Glue Crawler

Automated schema detection & metadata catalog creation

Crawlers

A crawler connects to a data store, progresses through a prioritized list of classifiers to determine the schema for your data, and then creates metadata tables in your data catalog.

Last updated (UTC) December 7, 2020 at 09:45:48 Action Run Create crawler

View and manage all available crawlers.

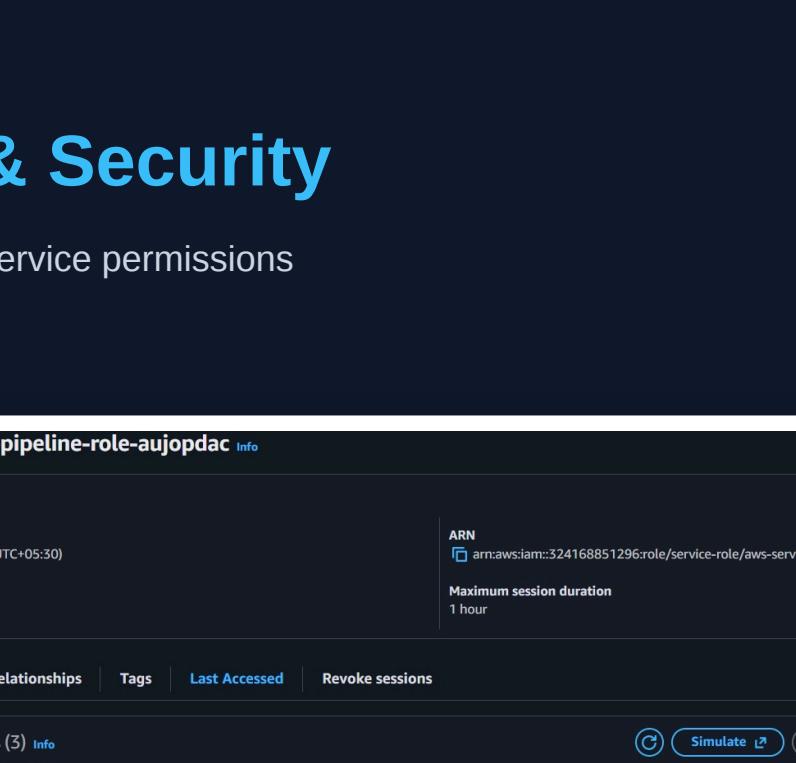
Filter crawlers

Name State Schedule Last run times... Log Table changes from last run

etl\_pipeline\_crawler Ready Succeeded December 7, 2020... View log 1 created

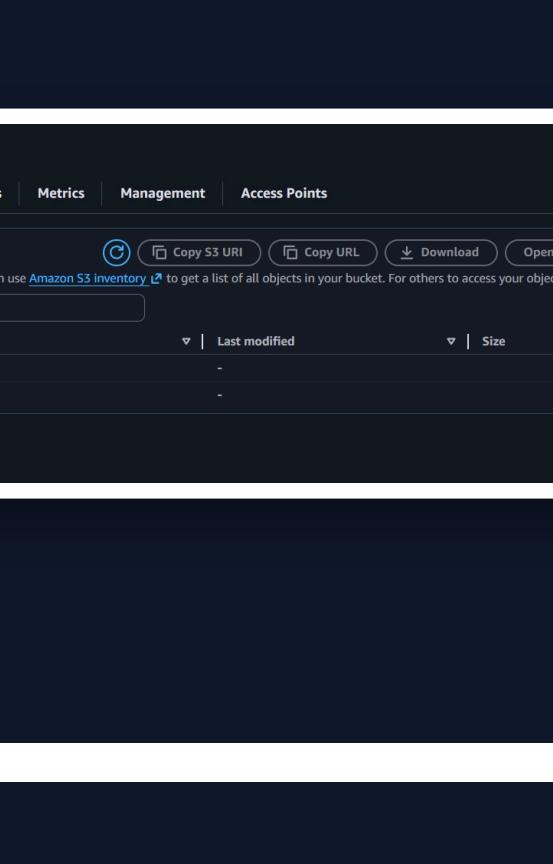
## Lambda Functions - Code

Python-based transformation logic with boto3, pandas



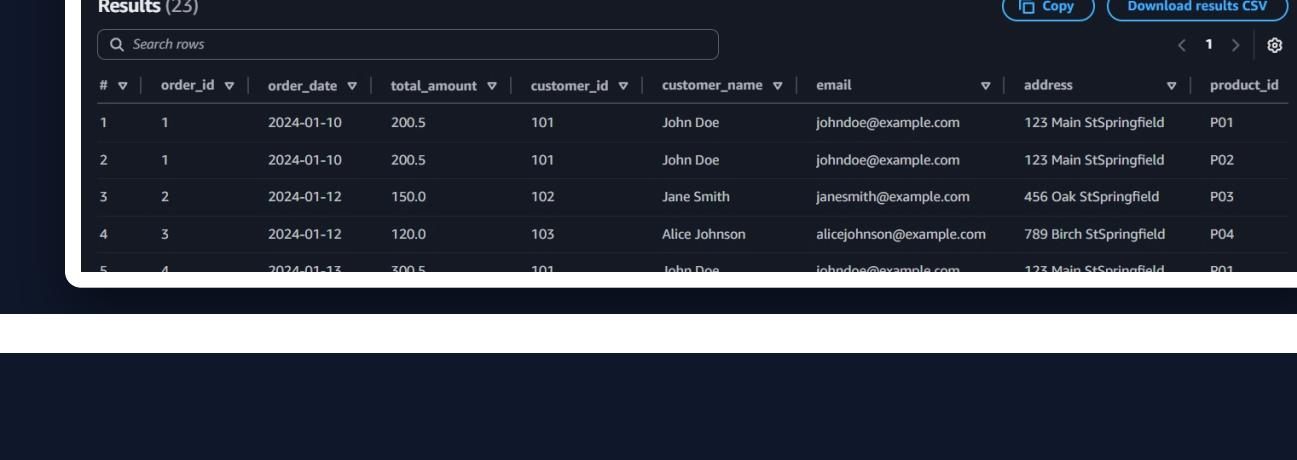
## Lambda Functions - Configuration

Runtime settings, layers, and managed configurations



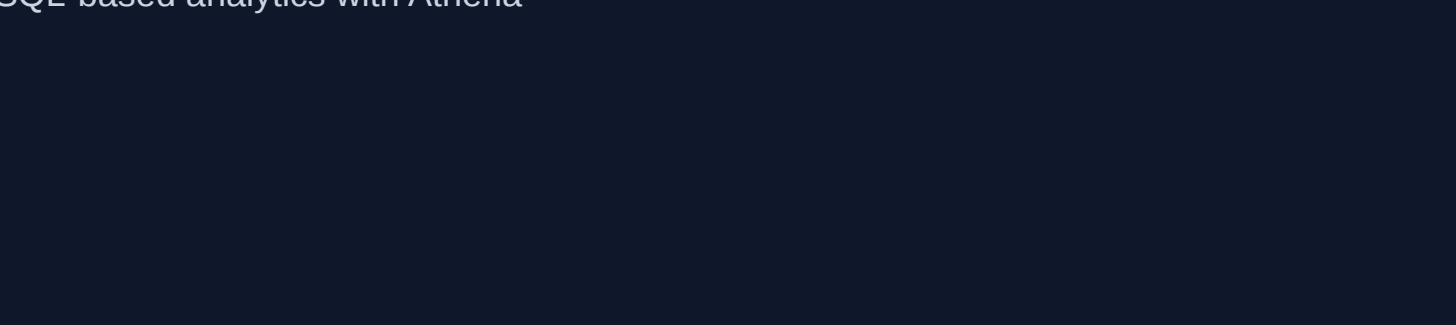
## IAM Roles & Security

S3, Lambda, and Glue service permissions



## S3 Data Lake

Organized storage: raw and processed data



## Amazon Athena

Serverless SQL queries on S3 data



## Key Benefits

- Fully serverless - no infrastructure management
- Auto-scaling with pay-per-use pricing
- Integrated data catalog and metadata
- SQL-based analytics with Athena

AWS Serverless ETL in Action

Production-Ready Data Pipeline

Thank You