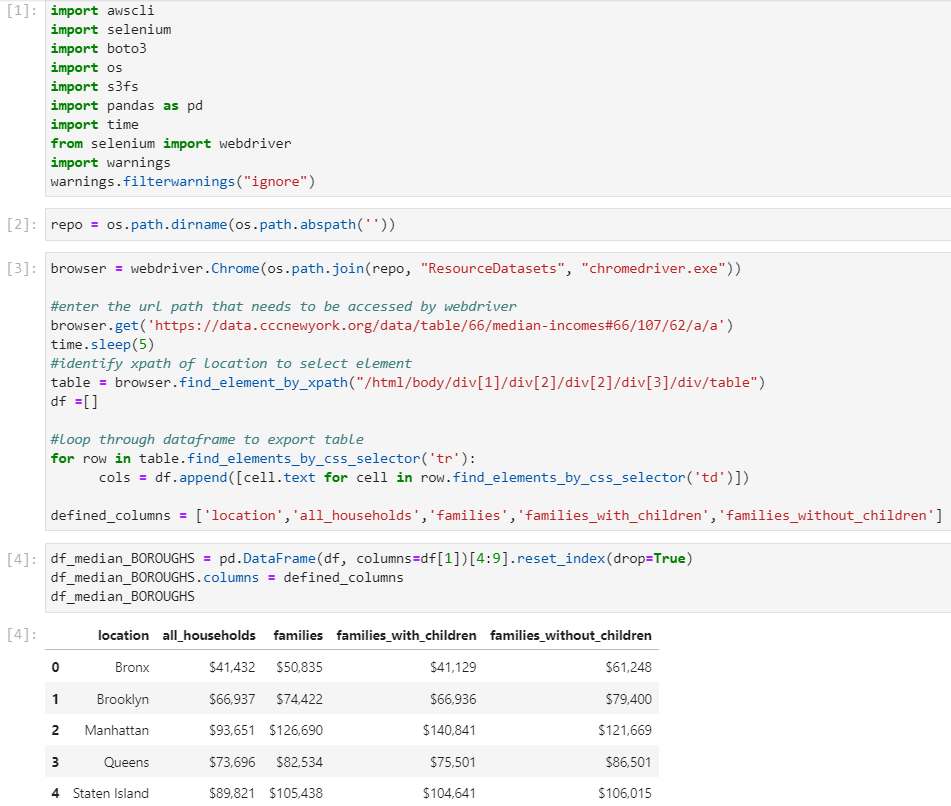
**6100 Information Architecture Summer 2021 Final Project Report**

Student: Xiaolan Li, Bernard Cooper

Professor: Brandon Chiazza

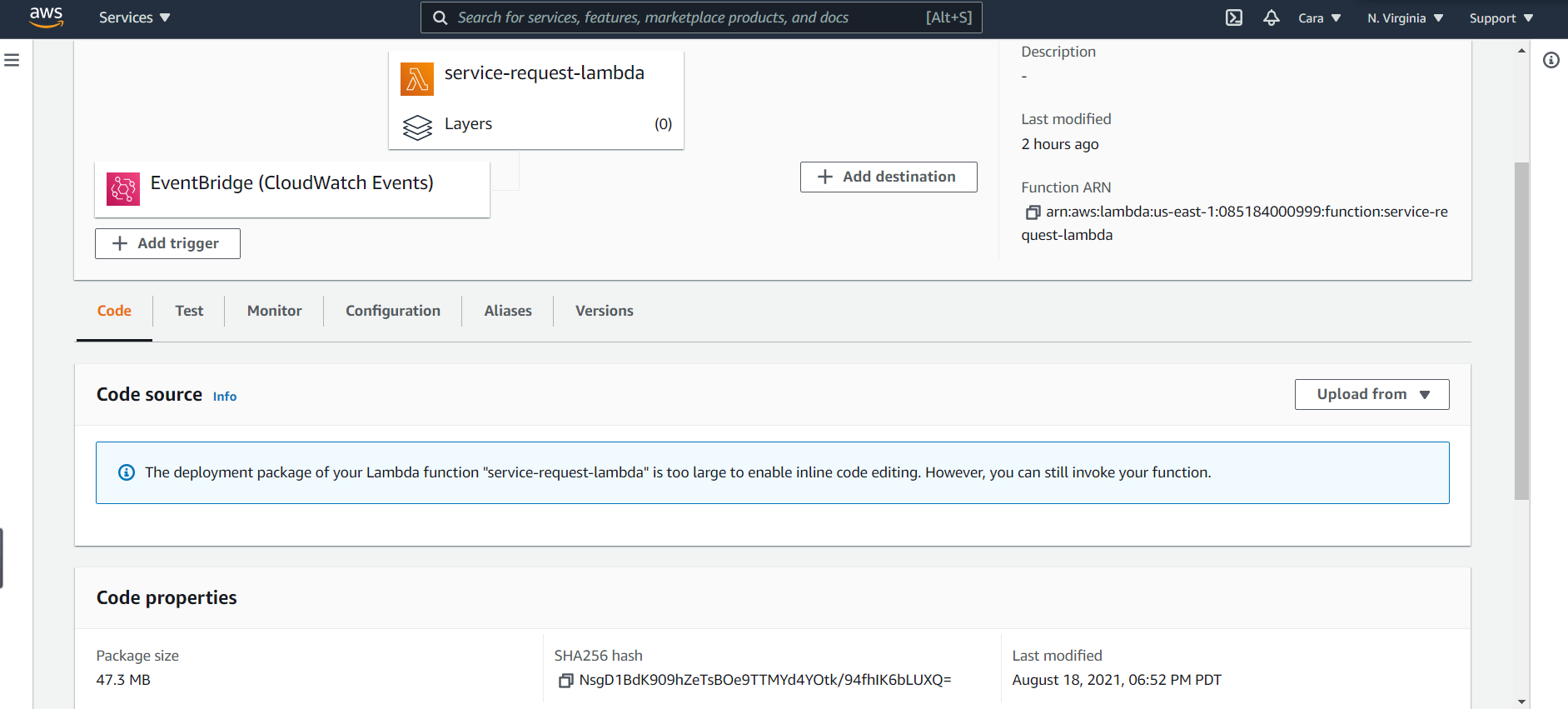
1. **Using Chrome Driver Selenium method to obtain the unstructured NYC median income dataset.**

<https://data.cccnewyork.org/data/table/66/median-incomes#66/107/62/a/a>

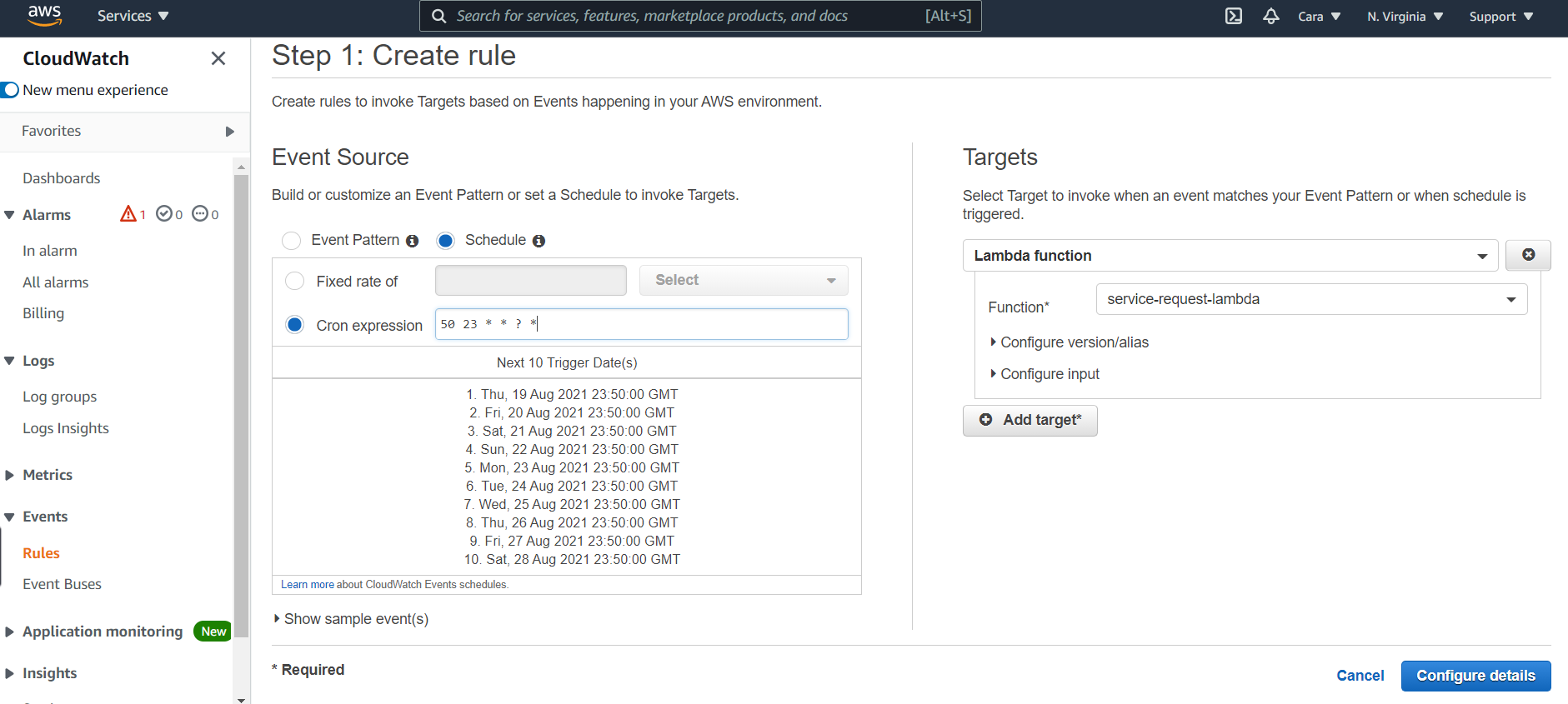


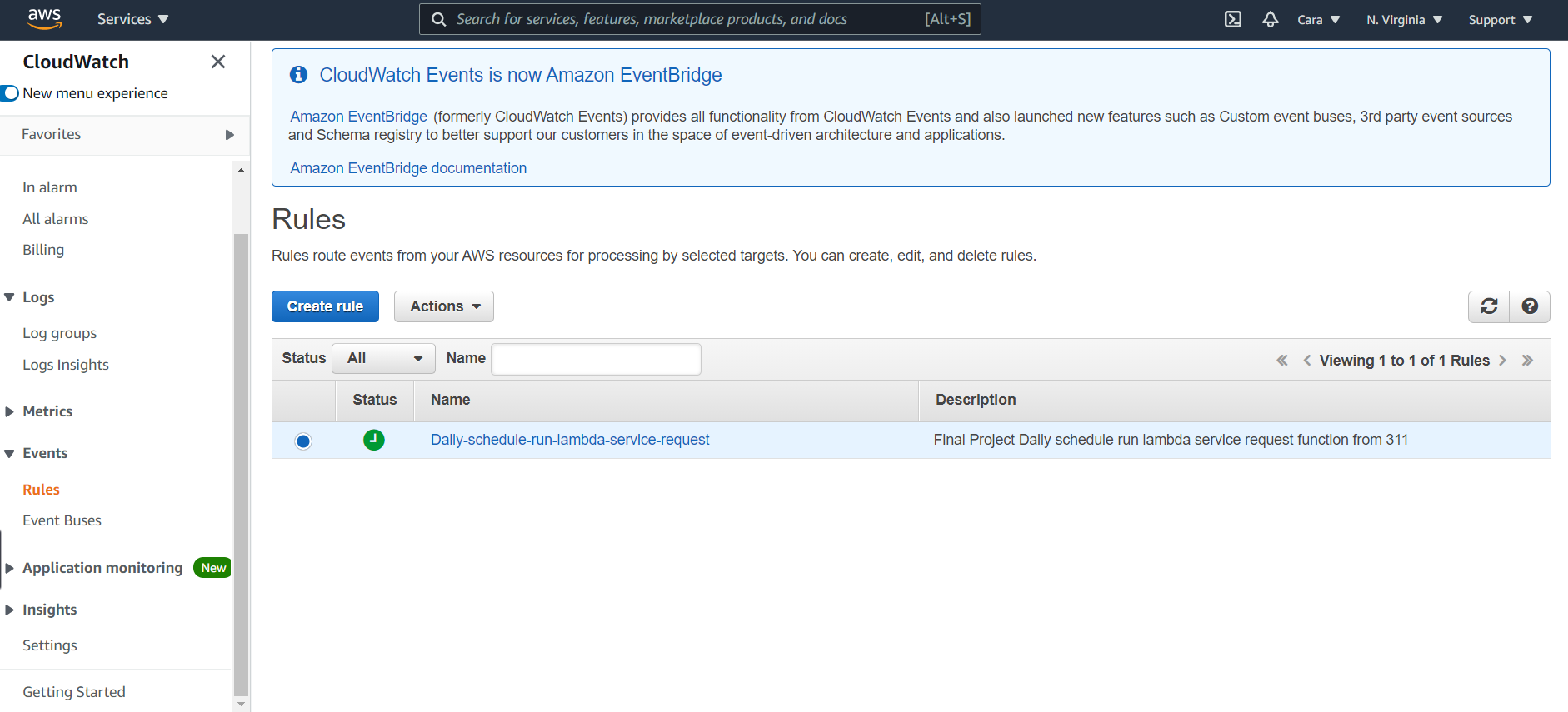
1. **Using Lambda Service with REST API method to daily obtain the last day incident resource from 311 service request structured database**

<https://data.cityofnewyork.us/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9>

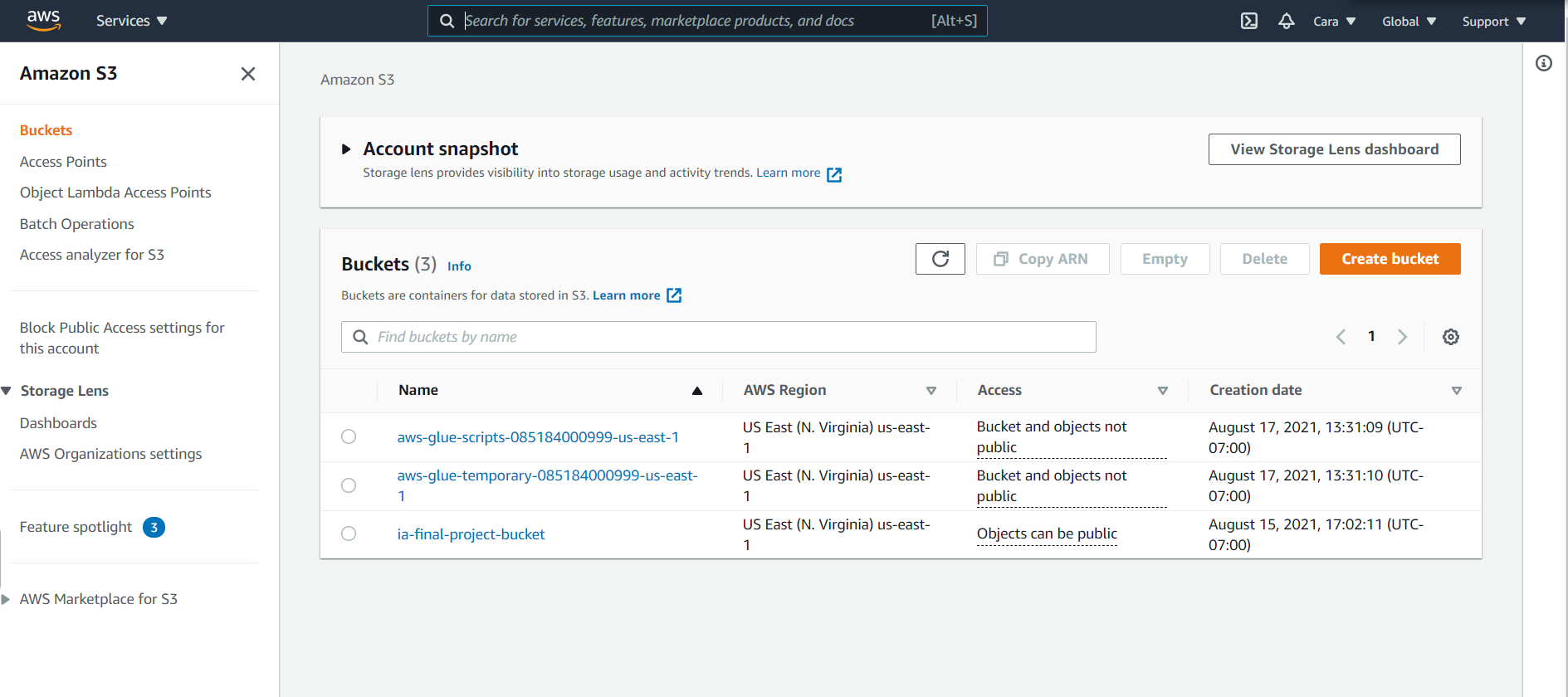


1. **Using CloudWatch Serice to schedule the time to run lambda function**

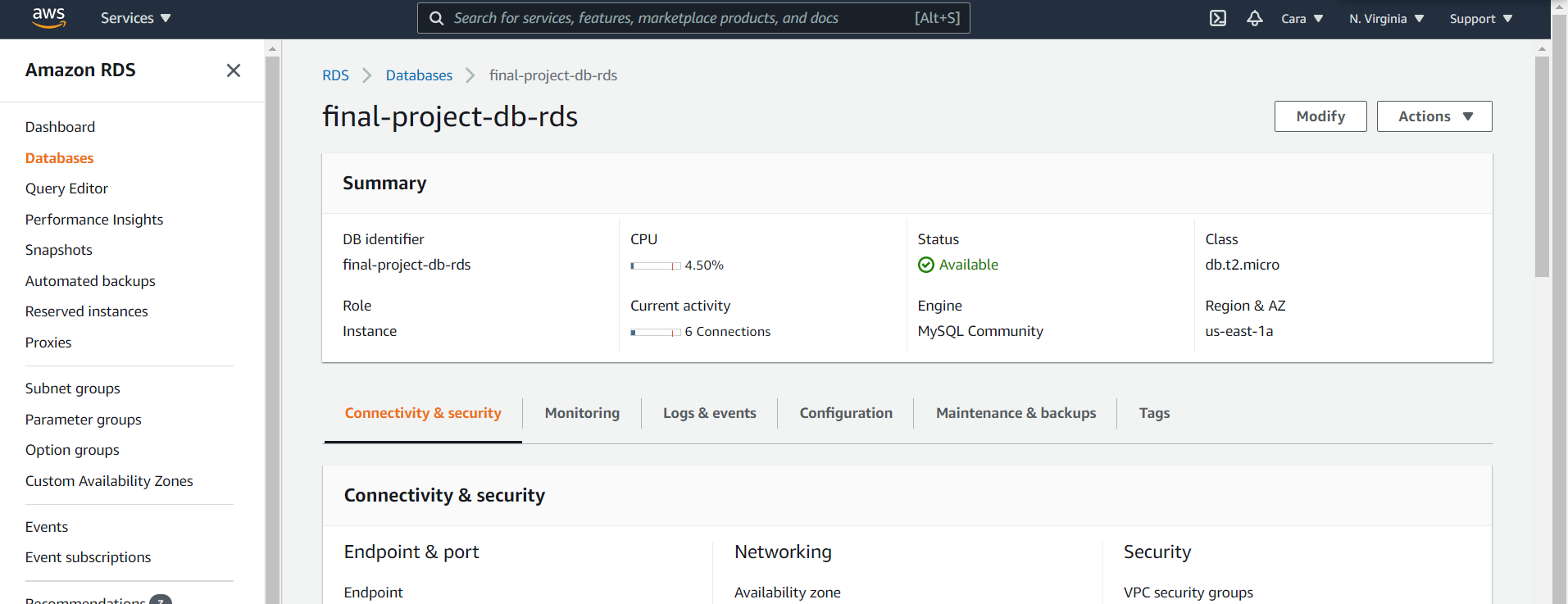




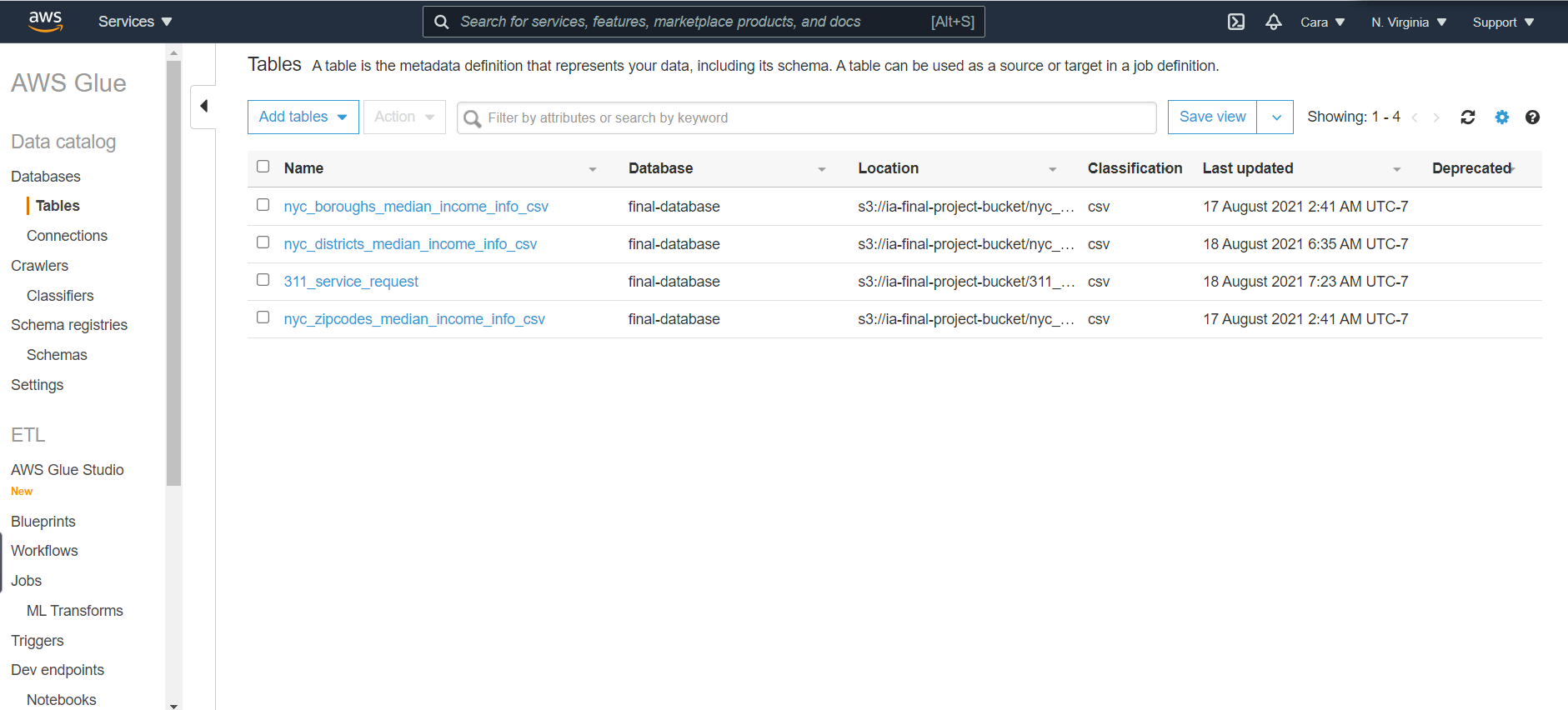
1. **Using S3 Service to store the two data sources in `ia-final-project-bucket` bucket**

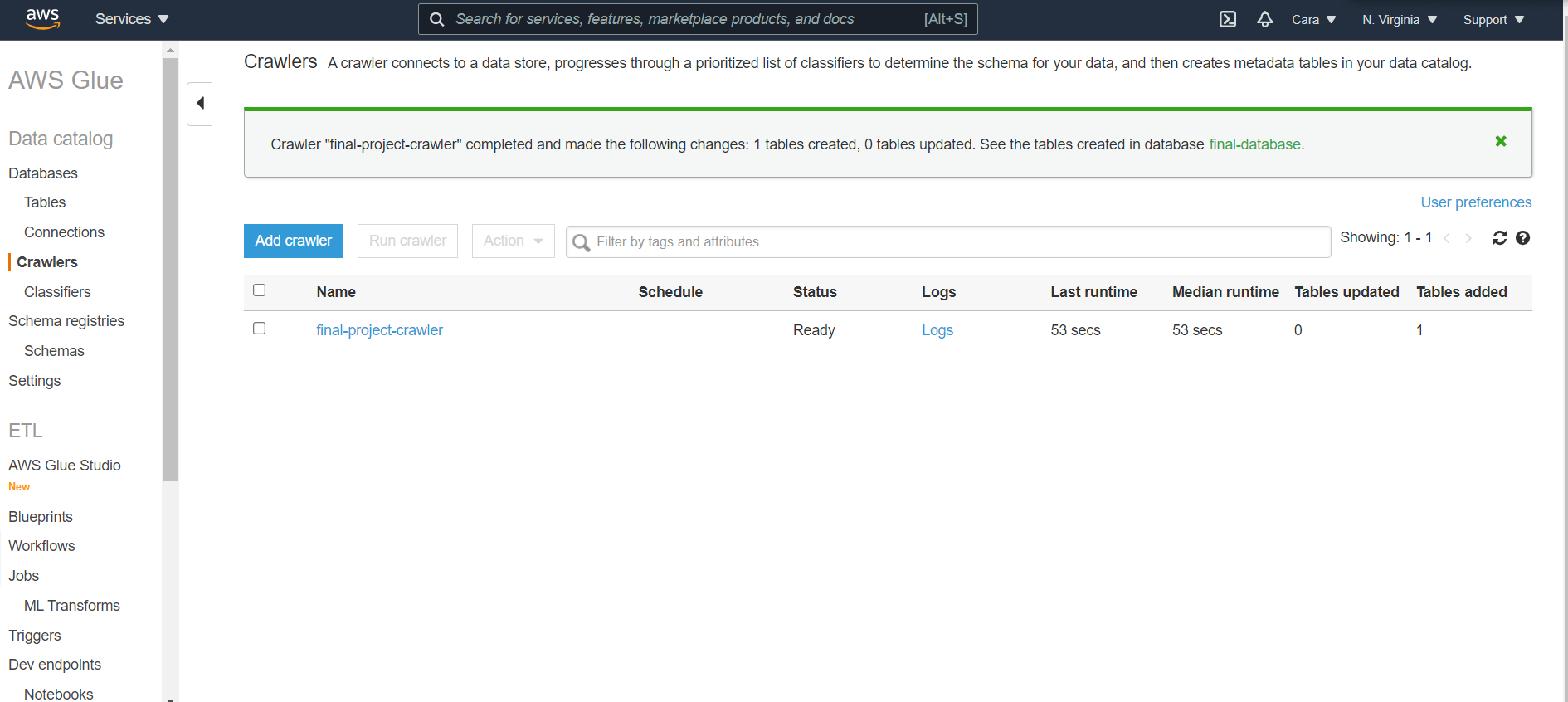


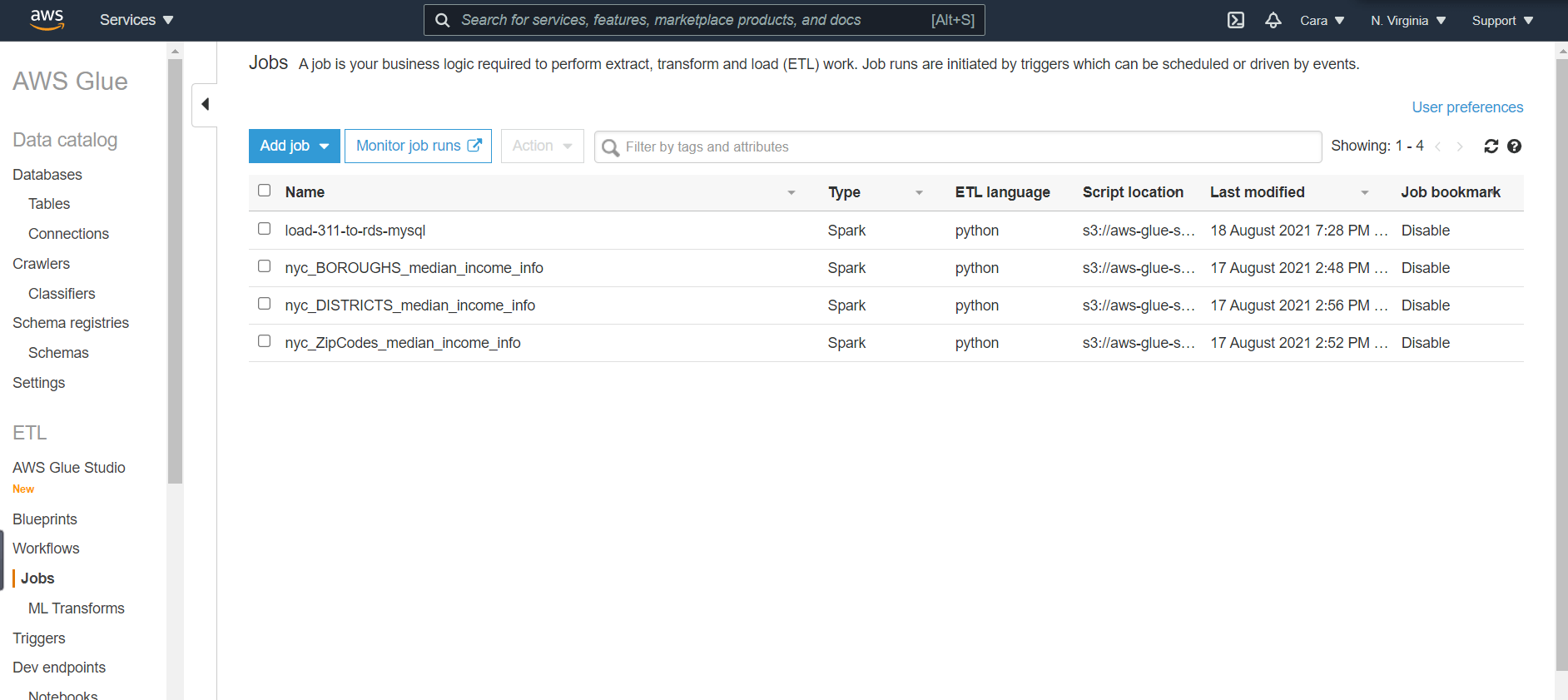
1. **Using RDS to store the sources schema and data warehouse**

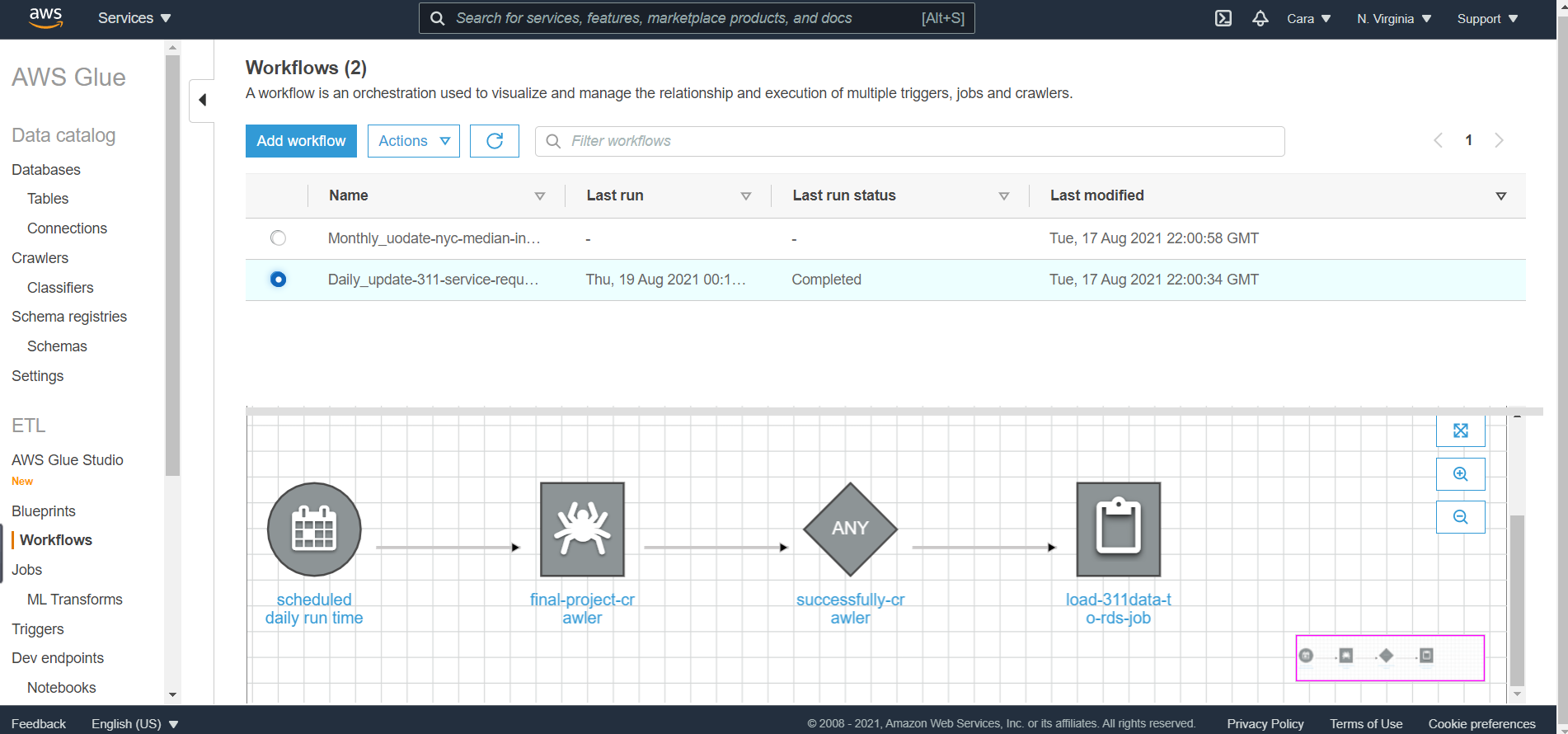


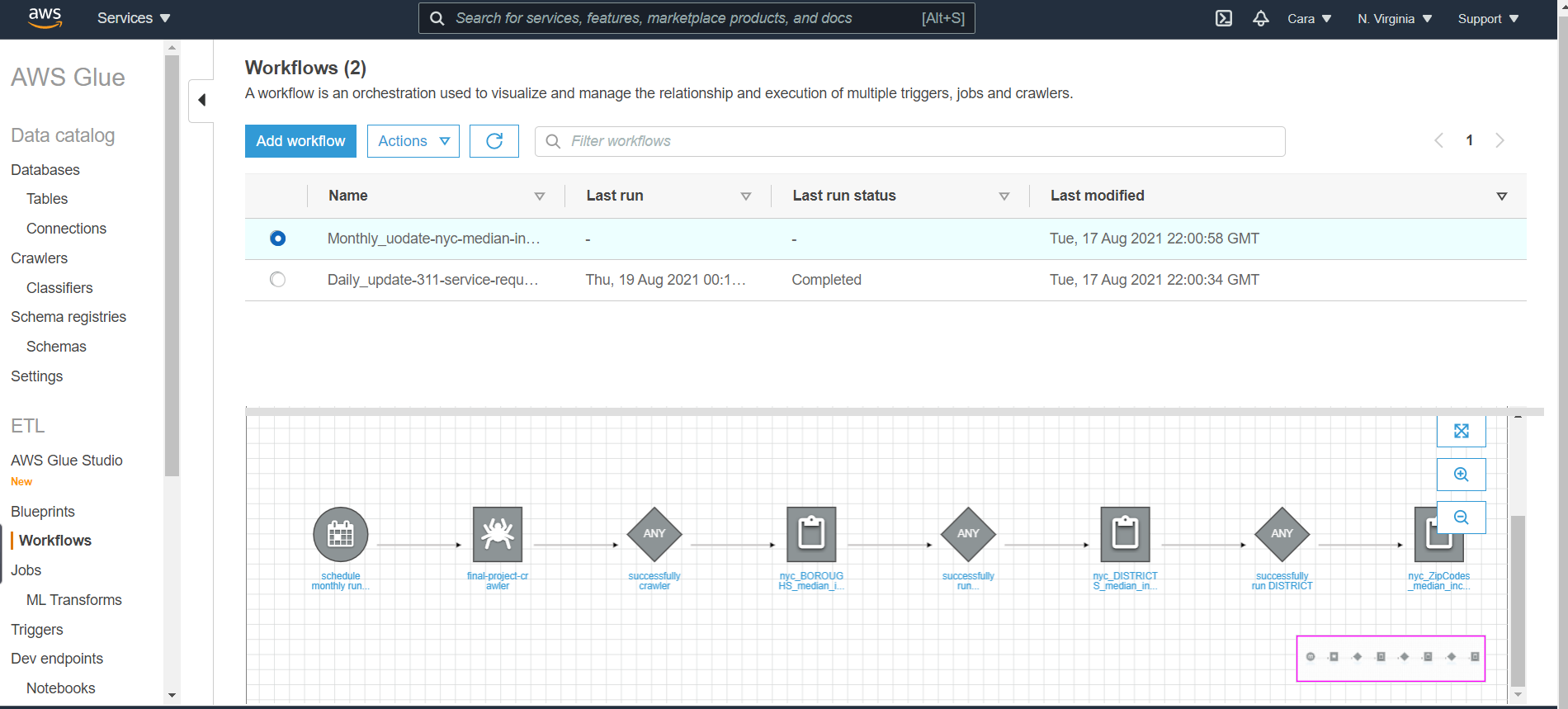
1. **Using GLUE service to load the data into RDS with MySQL Work Bench in daily schedule with workflow steps crawler table from s3 and running jobs for each table.**







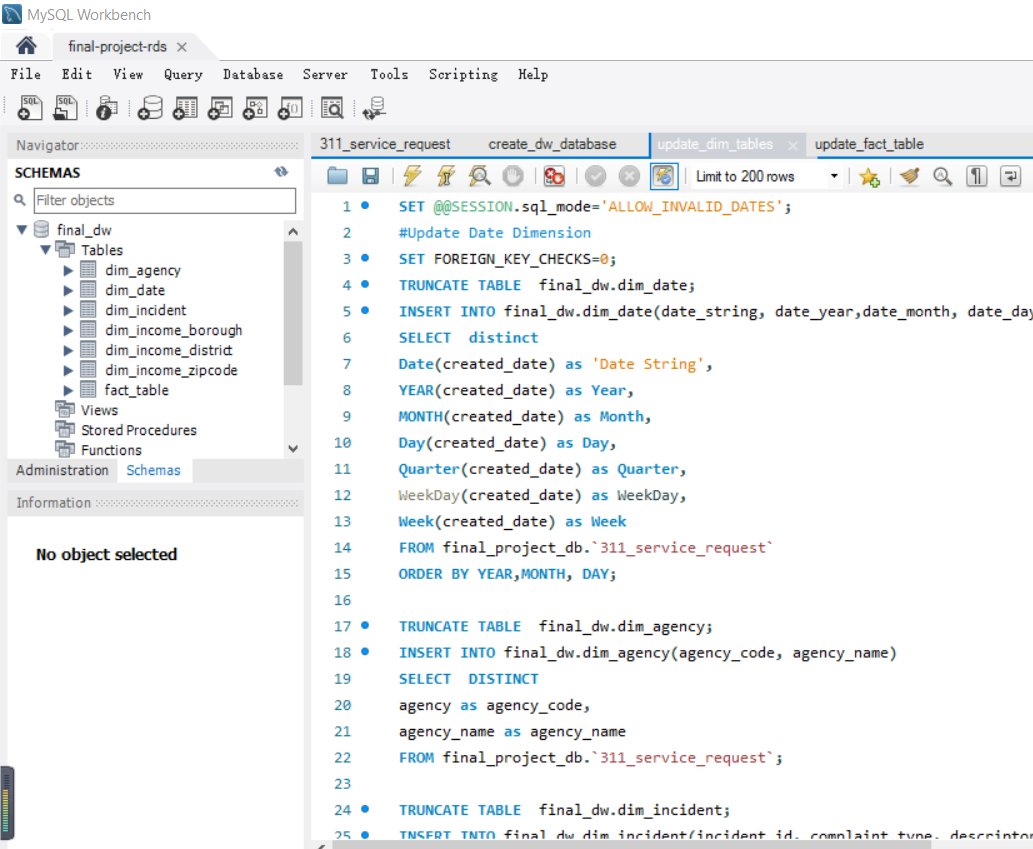




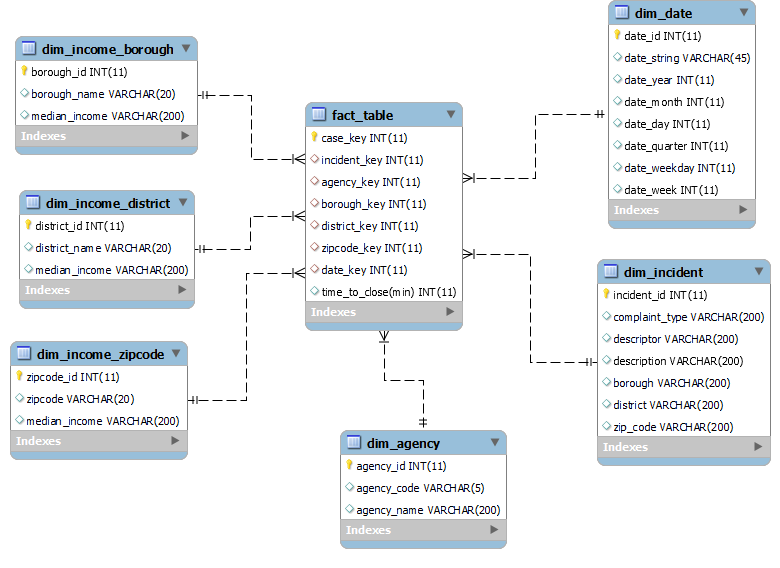
1. **Database Reverse Engineer ER diagram for Resource database `final\_project\_db`**



1. **Create Star schema Data warehouse DDL in sql, create update the dimensional table DDL, create update fact table DDL.**

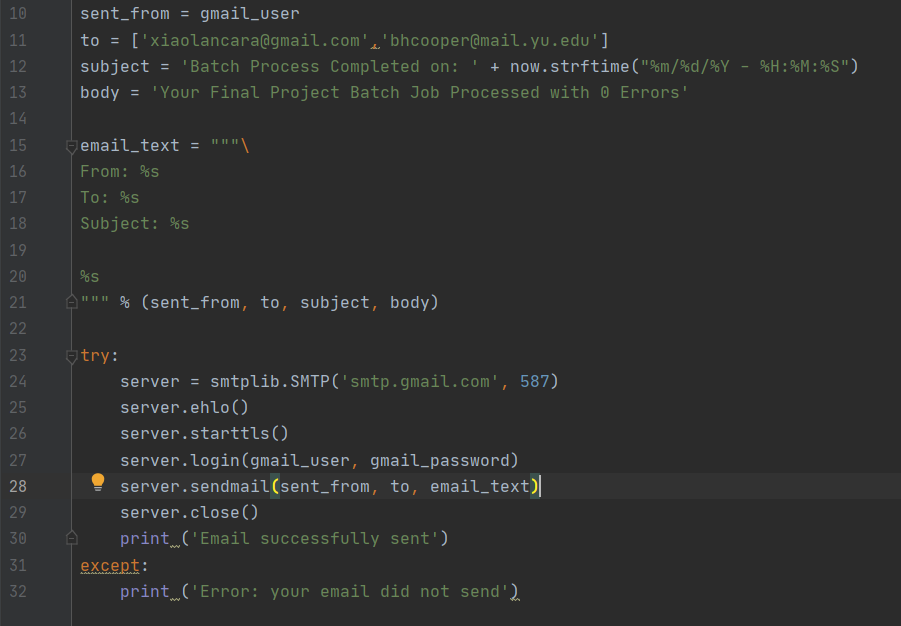


1. **Database Reverse Engineer ER diagram for star schema Data Warehousedatabase `final\_dw`**



1. **Using python connect to the RDS MySql and schedule to run the updated data warehouse sql script daily time. Connecting to the gmail module and send the notification log when it successfully updated data warehouse.**





1. **Using Tableau to Analyze the data warehouse**