L.A Criminal Activity Analysis

PRESENTED BY LAW AND DISORDER

ETTIONE C. STUCKEY II | MARIO AVILA | CHI ASANGWE

Table of contents

01 Overview

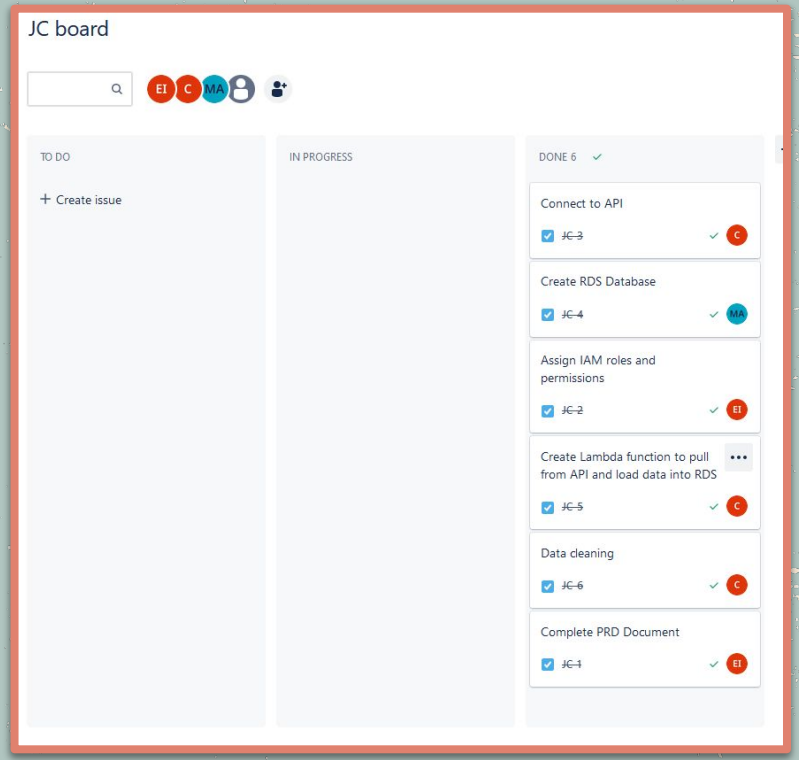
Introduction

CHIEF OF POLICE, JAY L. THYME HAS EMPLOYED OUR GROUP OF DATA SCIENTIST TO HELP MAKE MORE EFFICIENT USE OF HIS L.A POLICE FORCE.

HE'S ASKED US TO ANALYZE POLICE REPORTS IN LA AND PROVIDE DETAILED INSIGHTS ON THE TYPES OF CRIMES BEING COMMITTED ACROSS THE CITY.

Team Communication

Our team chose to use a Jira board to allocate tasks, and held daily stand up meetings to ensure that we are hitting our checkpoints on time with Ettione as our scrum master.

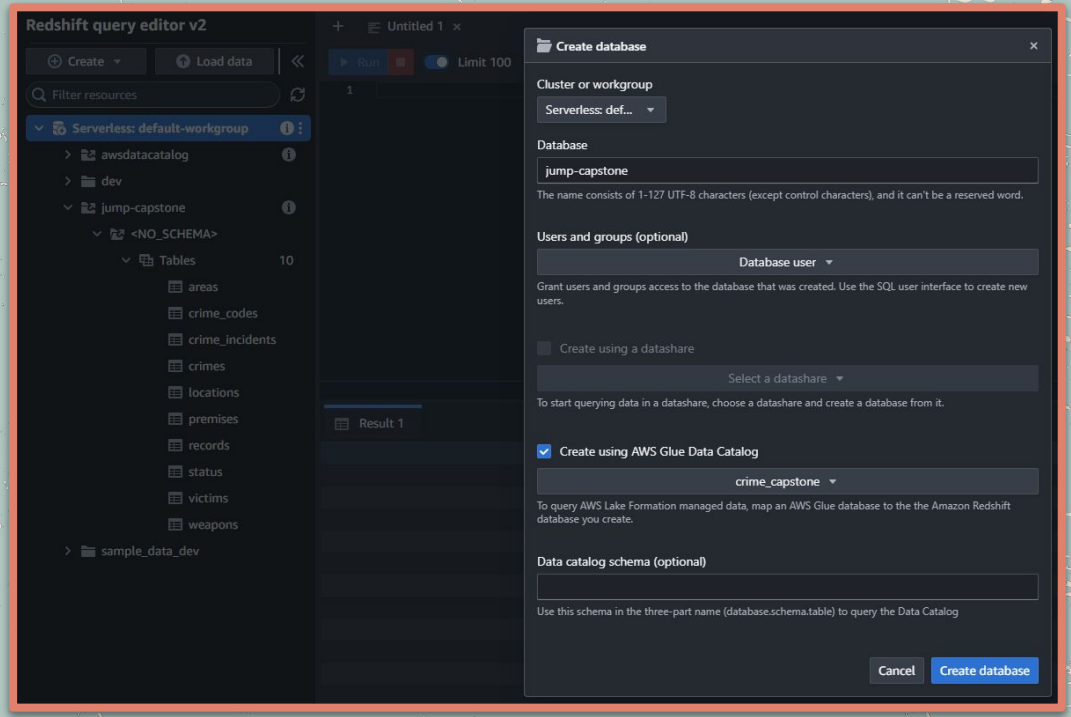



Where are we getting the data from?

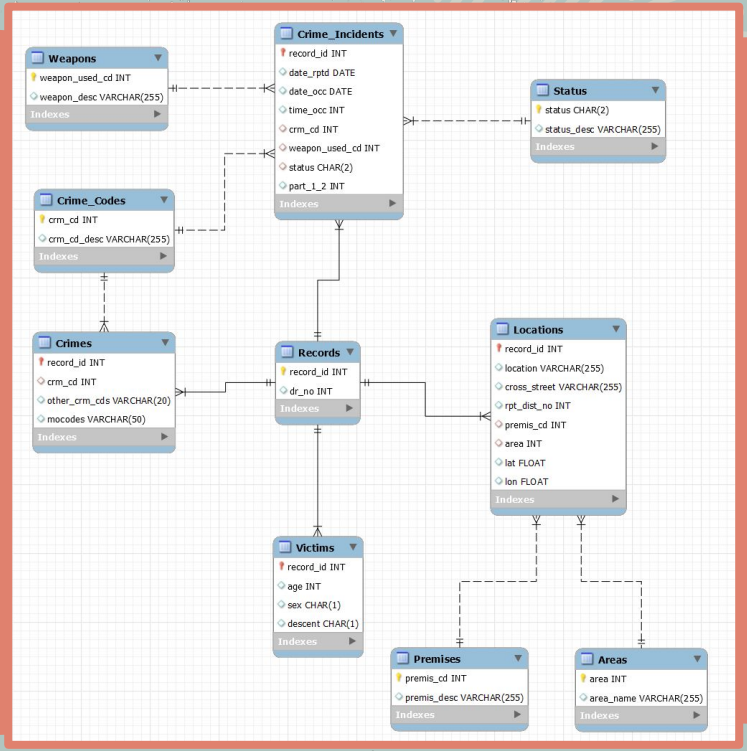
L.A Crime API

BY USING THE L.A CRIME DATA API, WE WERE ABLE TO PULL LIVE DATA IN THE FOR OF A JSON FILE IN ORDER TO PERFORM OUR ANALYTICS AND BUILD OUR DATABASE.

REDSHIFT CONFIGURATION



DATABASE SCHEMA



SETTING UP IAM

BEFORE WE CAN GET STARTED USING OUR AWS RESOURCES, WE NEED TO ASSIGN USERS, GROUPS AND PERMISSIONS.

Chi Asangwe

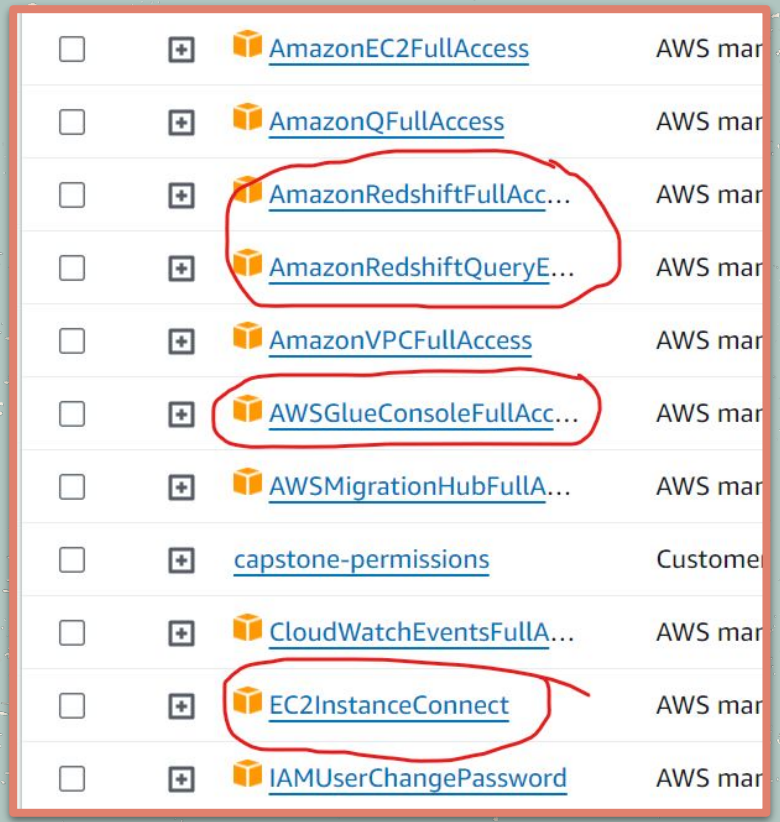
NEEDS ACCESS TO AWS GLUE AND REDSHIFT

Ettione Stuckey II

NEEDS ACCESS TO EC2 AND S3

Mario Avila

NEEDS ACCESS TO AWS GLUE, REDSHIFT, AND LAMBDA



Key IAM Policies

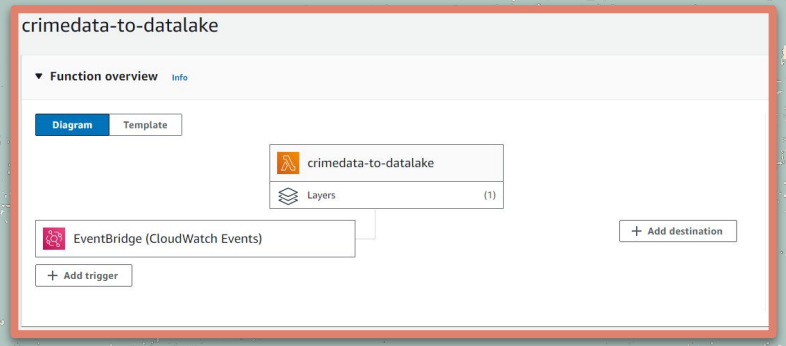
- **AWS Glue Full Access**
- **AWS Redshift Full Access**
- **EC2 Instance connect**

Hosting the Data Lake

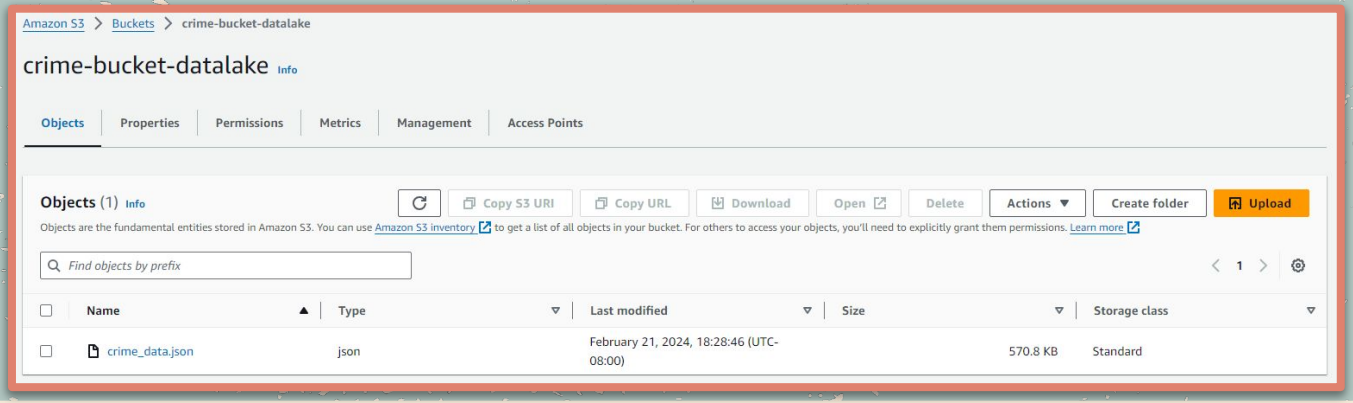
02

WITH AWS S3

USING LAMBDA AND CLOUDWATCH



TO UPLOAD DATA TO S3 BUCKET



03 Setting up AWS Glue

FOR CATALOGING DATA AND ETL JOBS

```
# Write DataFrames to the Glue Data Catalog
records_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Records").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Records")
victims_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Victims").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Victims")
status_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Status").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Status")
premises_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Premises").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Premises")
areas_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Areas").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Areas")
weapons_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Weapons").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Weapons")
locations_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Locations").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Locations")
crime_codes_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Crime_Codes").option("createTableColumnTypes", "snappy").saveAsTable(database_name + "
    .Crime_Codes")
crimes_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Crimes").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Crimes")
crime_incidents_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Crime_Incidents").option("createTableColumnTypes", "snappy").saveAsTable(database_name + "
    .Crime_Incidents")
```

AWS Glue > Databases > crime_capstone

crime_capstone

Last updated (UTC) February 22, 2024 at 04:58:14 | Edit | Delete

Database properties

Name	Description	Location	Created on (UTC)
crime_capstone	-	-	February 20, 2024 at 20:54:56

Tables (10)

View and manage all available tables.

Last updated (UTC) February 22, 2024 at 04:58:16 | Delete | Add tables using crawler | Add table

Filter tables

Name	Database	Location	Classification	Deprecated	View data	Data quality
areas	crime_capstone	s3://crime-bucket-datalake/Areas	-	-	Table data	View data quali
crime_codes	crime_capstone	s3://crime-bucket-datalake/Crime_Codes	-	-	Table data	View data quali
crime_incidents	crime_capstone	s3://crime-bucket-datalake/Crime_Incidents	-	-	Table data	View data quali
crimes	crime_capstone	s3://crime-bucket-datalake/Crimes	-	-	Table data	View data quali
locations	crime_capstone	s3://crime-bucket-datalake/Locations	-	-	Table data	View data quali
premises	crime_capstone	s3://crime-bucket-datalake/Premises	-	-	Table data	View data quali
records	crime_capstone	s3://crime-bucket-datalake/Records	-	-	Table data	View data quali
status	crime_capstone	s3://crime-bucket-datalake/Status	-	-	Table data	View data quali
victims	crime_capstone	s3://crime-bucket-datalake/Victims	-	-	Table data	View data quali
weapons	crime_capstone	s3://crime-bucket-datalake/Weapons	-	-	Table data	View data quali

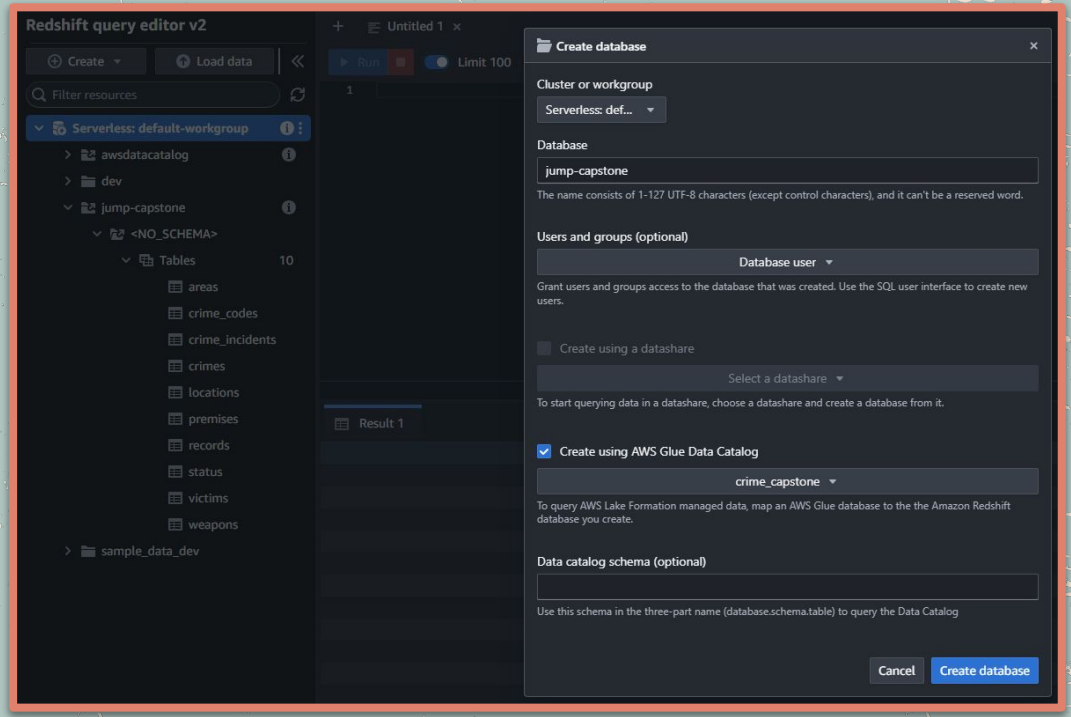
CONNECTING TO AWS GLUE

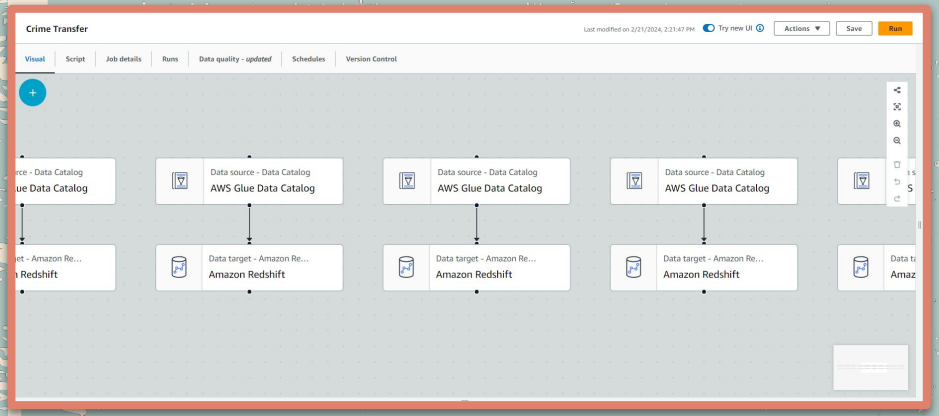
04

Setting up AWS Redshift

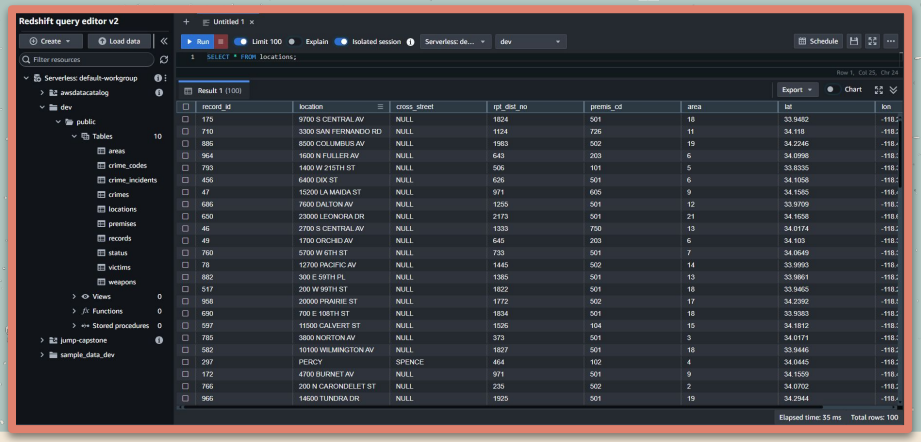
CLOUD-BASED DATA WAREHOUSING

REDSHIFT CONFIGURATION





AWS GLUE TO AWS REDSHIFT



05

Challenges

A FEW OBSTACLES WE HAD TO OVERCOME

Challenge

AWS GLUE AND AWS REDSHIFT WAS NEW TO US ALL.

Solution

WE HAD TO DO OUR RESEARCH ON WHAT EACH SERVICE DOES AND CONFIGURE THEM TO OUR NEEDS.

Challenge

WHEN IMPORTING TABLES FROM AWS GLUE TO AWS REDSHIFT, ALL OF THEM BUT ONE WAS TRANSFERRED SUCCESSFULLY

Solution

WE DID THE DATA CLEANING ON THE INITIAL UNSTRUCTURED DATA, SO WE FORGOT TO TAKE CARE OF NULLS AND DUPLICATES WHEN STRUCTURING THE DATA. YOU CAN'T HAVE A NULL PRIMARY KEY!

Challenge

While writing our scripts for visualizations and analytics, we noticed that the attributes for the time that a crime occurred were all in 2020. This is an issue considering this data should be from 2020 to 2023 onward.

Solution

We saw this as a data entry error, and decided to move forward with our analytics using the date that the crime was reported. This way, we could get a better grasp of when crimes were actually happening.

Questions?

06

CLICK HERE TO VIEW OUR ANALYTICS