



L.A Criminal Activity Analysis

PRESENTED BY LAW AND DISORDER

ETTIONE C. STUCKEY II | MARIO AVILA | CHI ASANGWE





Table of contents

01

Overview

02

Hosting Data Lake

03

Setting up AWS Glue

04

Setting up AWS Redshift

05

Challenges

06

Demo/Questions

01

Overview





**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.**



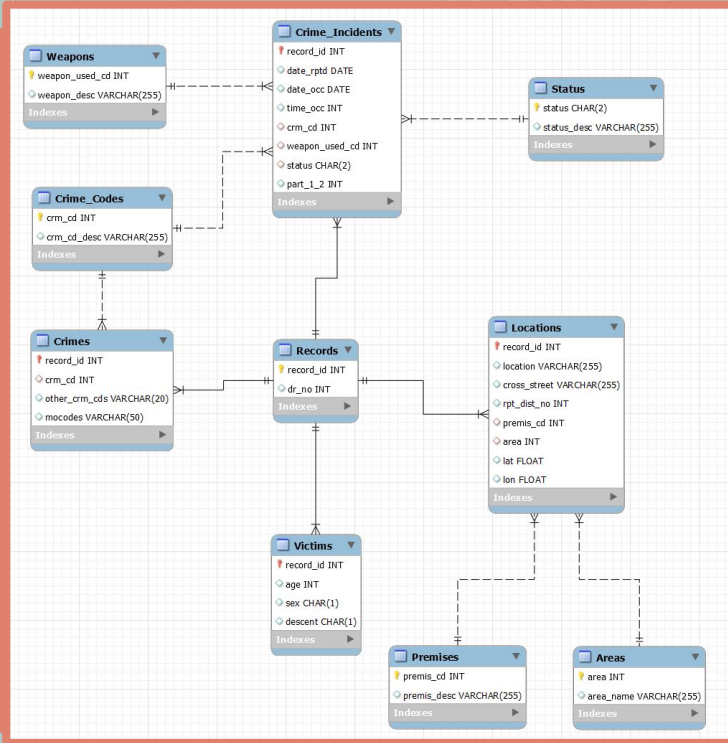
Where are we getting the data from?



L.A Crime API

BY USING THE [L.A CRIME DATA API](#), WE WERE ABLE TO PULL LIVE DATA IN THE FORM OF A JSON FILE IN ORDER TO PERFORM OUR ANALYTICS AND BUILD OUR DATABASE.

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

| | | | |
|--------------------------|---|--|---------|
| <input type="checkbox"/> |  |  AmazonEC2FullAccess | AWS mar |
| <input type="checkbox"/> |  |  AmazonQFullAccess | AWS mar |
| <input type="checkbox"/> |  |  AmazonRedshiftFullAcc... | AWS mar |
| <input type="checkbox"/> |  |  AmazonRedshiftQueryE... | AWS mar |
| <input type="checkbox"/> |  |  AmazonVPCFullAccess | AWS mar |
| <input type="checkbox"/> |  |  AWSGlueConsoleFullAcc... | AWS mar |
| <input type="checkbox"/> |  |  AWSMigrationHubFullA... | AWS mar |
| <input type="checkbox"/> |  | capstone-permissions | Custome |
| <input type="checkbox"/> |  |  CloudWatchEventsFullA... | AWS mar |
| <input type="checkbox"/> |  |  EC2InstanceConnect | AWS mar |
| <input type="checkbox"/> |  |  IAMUserChangePassword | AWS mar |

Key IAM Policies

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



Hosting the Data Lake

WITH AWS S3

02



USING LAMBDA AND CLOUDWATCH

crimedata-to-datalake

Function overview [Info](#)

Diagram [Template](#)



EventBridge (CloudWatch Events)

+ Add destination

+ Add trigger

TO UPLOAD DATA TO S3 BUCKET

[Amazon S3](#) > [Buckets](#) > crime-bucket-datalake

crime-bucket-datalake [Info](#)

[Objects](#) [Properties](#) [Permissions](#) [Metrics](#) [Management](#) [Access Points](#)

Objects (1) [Info](#)



Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|-----------------|------|---|----------|---------------|
| <input type="checkbox"/> | crime_data.json | json | February 21, 2024, 18:28:46 (UTC-08:00) | 570.8 KB | Standard |



03

Setting up AWS Glue

FOR CATALOGING DATA AND ETL JOBS



```

82 # Write DataFrames to the Glue Data Catalog
83 records_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Records").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Records")
84 victims_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Victims").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Victims")
85 status_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Status").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Status")
86 premises_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Premises").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Premises")
87 areas_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Areas").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Areas")
88 weapons_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Weapons").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Weapons")
89 locations_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Locations").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Locations")
90 crime_codes_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Crime_Codes").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Crime_Codes")
91 crimes_df.write.format("parquet").mode("overwrite").option("path", "s3://crime-bucket-datalake/Crimes").option("createTableColumnTypes", "snappy").saveAsTable(database_name + ".Crimes")
92 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 04:38:14 [Edit](#) [Delete](#)

Database properties

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

Tables (10)

Last updated (UTC) February 22, 2024 at 04:38:16 [Delete](#) [Add tables using crawler](#) [Add table](#)

View and manage all available tables.

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

CONNECTING TO AWS GLUE



04



Setting up AWS Redshift

CLOUD-BASED DATA WAREHOUSING



REDSHIFT CONFIGURATION



Redshift query editor v2

Create Load data

Filter resources

Serverless: default-workgroup

- awsdatacatalog
- dev
- jump-capstone
 - <NO_SCHEMA>
 - Tables 10
 - areas
 - crime_codes
 - crime_incidents
 - crimes
 - locations
 - premises
 - records
 - status
 - victims
 - weapons
 - sample_data_dev

Untitled 1

Run Limit 100

1

Result 1

Create database

Cluster or workgroup

Serverless: def...

Database

jump-capstone

The name consists of 1-127 UTF-8 characters (except control characters), and it can't be a reserved word.

Users and groups (optional)

Database user

Grant users and groups access to the database that was created. Use the SQL user interface to create new users.

Create using a datashare

Select a datashare

To start querying data in a datashare, choose a datashare and create a database from it.

Create using AWS Glue Data Catalog

crime_capstone

To query AWS Lake Formation managed data, map an AWS Glue database to the the Amazon Redshift database you create.

Data catalog schema (optional)

Use this schema in the three-part name (database.schema.table) to query the Data Catalog

Cancel Create database

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?

[CLICK HERE TO VIEW OUR ANALYTICS](#)

06

