



# **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





# 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.



## JC board

 Q EI C MA 👤 ⚙️

TO DO

+ Create issue

IN PROGRESS

DONE 6 ✓

Connect to API

☒ JC-3 ✓ EI

Create RDS Database

☒ JC-4 ✓ MA

Assign IAM roles and permissions

☒ JC-2 ✓ EI

Create Lambda function to pull from API and load data into RDS ⋮

☒ JC-5 ✓ C

Data cleaning

☒ JC-6 ✓ C

Complete PRD Document

☒ JC-1 ✓ EI



**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.

# 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.

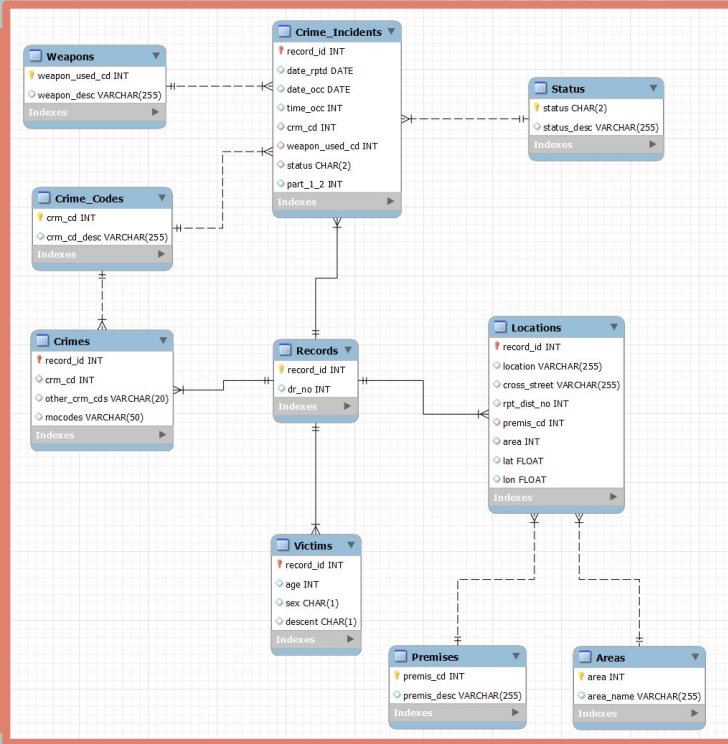
Data catalog schema (optional)

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

Cancel Create database



# 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"/>		 <a href="#">AmazonEC2FullAccess</a>	AWS mar
<input type="checkbox"/>		 <a href="#">AmazonQFullAccess</a>	AWS mar
<input type="checkbox"/>		 <a href="#">AmazonRedshiftFullAcc...</a>	AWS mar
<input type="checkbox"/>		 <a href="#">AmazonRedshiftQueryE...</a>	AWS mar
<input type="checkbox"/>		 <a href="#">AmazonVPCFullAccess</a>	AWS mar
<input type="checkbox"/>		 <a href="#">AWSGlueConsoleFullAcc...</a>	AWS mar
<input type="checkbox"/>		 <a href="#">AWSMigrationHubFullA...</a>	AWS mar
<input type="checkbox"/>		<a href="#">capstone-permissions</a>	Custome
<input type="checkbox"/>		 <a href="#">CloudWatchEventsFullA...</a>	AWS mar
<input type="checkbox"/>		 <a href="#">EC2InstanceConnect</a>	AWS mar
<input type="checkbox"/>		 <a href="#">IAMUserChangePassword</a>	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](#)

< 1 > [Settings](#)

<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

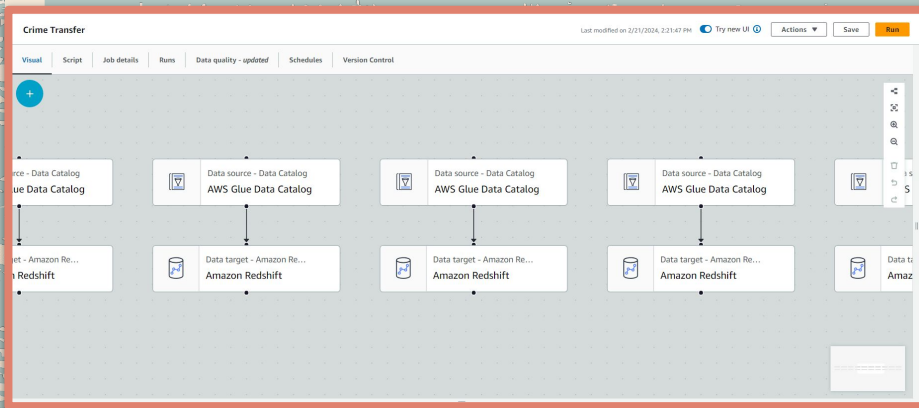
To query AWS Lake Formation managed data, map an AWS Glue database to the the Amazon Redshift database you create.

Data catalog schema (optional)

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

Cancel Create database





# AWS GLUE TO AWS REDSHIFT

**Redshift query editor v2**

Untitled 1 x Limit 100 Explains Isolated session Serverless de... dev Schedule Export Chart

Filter resources

Serverless default-workgroup

awsdatacatalog

public

Tables

ases

crime\_codes

crime\_incidents

crimes

locations

premises

records

status

victims

weapons

Views

Functions

Stored procedures

jump-captone

sample\_data\_dev

Result 1 (100)

id	location	crime_street	offense_id	premise_id	area	lat	lon
175	3700 S CENTRAL AV	NULL	1024	501	18	33.9402	-118.1
770	3300 SAN FERNANDO RD	NULL	1124	726	11	34.118	-118.1
886	8500 COLUMBUS AV	NULL	1963	502	19	34.2246	-118.1
964	1600 N FULLER AV	NULL	643	203	6	34.0966	-118.1
700	1400 W 24TH ST	NULL	506	101	5	33.9335	-118.1
456	6400 DIX ST	NULL	626	501	6	34.1058	-118.1
47	15200 LA MANA ST	NULL	971	606	9	34.1065	-118.1
686	7600 DALTON AV	NULL	1250	501	12	33.9709	-118.1
600	23000 E LEONARD DR	NULL	2172	501	21	34.4658	-118.1
46	2700 S CENTRAL AV	NULL	1333	750	13	34.0174	-118.1
49	1700 ORCHARD AV	NULL	645	203	6	34.103	-118.1
700	5700 W 6TH ST	NULL	733	501	7	34.0649	-118.1
78	10700 PACIFIC BL	NULL	1465	502	14	33.9993	-118.1
882	300 E 59TH PL	NULL	1365	501	13	33.9661	-118.1
517	200 W 99TH ST	NULL	1822	501	18	33.9465	-118.1
908	20000 PRAIRIE ST	NULL	1772	502	17	34.2392	-118.1
690	700 E 100TH ST	NULL	1541	501	15	33.9893	-118.1
587	3100 CALVERT ST	NULL	1025	104	15	34.1812	-118.1
785	3800 NORTON AV	NULL	373	501	3	34.0171	-118.1
582	10100 WILMINGTON AV	NULL	1827	501	18	33.9446	-118.1
297	PERCY	SPENCE	464	102	4	34.0445	-118.1
172	4400 BURNET AV	NULL	971	501	9	34.1559	-118.1
766	200 N CARONDELLE ST	NULL	235	502	2	34.0702	-118.1
966	14600 TUNDRRA DR	NULL	1925	501	19	34.2944	-118.1

Elapsed time: 35 ms Total rows: 100

**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

