

Cloud Computing Assignment 3

Report

Crime Rates

Table of Contents

Signed Contribution Sheet	3
Introduction	
Implementation	
1. Creating buckets and IAM role:	6
(i) Set up s3 Buckets	6
(ii) Create IAM roles	6
2. Uploading files to s3 bucket:	7
(i) Obtain the dataset and extract csv	7
(ii) Upload the csv to bucket	7
3. AWS Data lake setup:	8
4. Database creation and AWS Glue Setup	9
(i) Prerequisite	9
(ii) Creating database	10
(ii) Crawler Setup	11
(iii) Loading additional dataset using Glue ETL Job	12
(iv) Running Crawler to add tables in database	15
5. Verifying database content using AWS Athena	15
6. AWS RedShift setup	16
(i) Creating cluster in AWS Redshift	16
(ii) Creating a new database inside Redshift	17
(iii) Creating security group	18
(iv) Adding VPC endpoints	19
(v) Creating connection to Redshift inside AWS Glue	20
(vi) Creating jobs for transferring tables	20
(vii) Creating view	24
7. AWS QuickSight Setup	26
(i)Creating view in AWS Athena	26
(ii)Loading dataset from view created	27
8. Redshift vs Athena	30
9. Elastic Beanstalk environment setup:	30
10. Route 53 setup:	32

Cloud Computing Assignment 3 Report

• • •

11. AWS Certification Manager:	33
12. AWS Cloud Formation:	35
13. AWS Cognito:	36
14. AWS API Gateway:	36
15. AWS Lambda:	37
16. Developer Manual	38
References:	40

Signed Contribution Sheet

Project title:

Group name here:

Student Name: Prabhat Kumar Singh	Student Name: Karthi Narendrababu Geetha
Student ID: s3833321	Student ID:s3835901
Contributions: 1. AWS Glue: Creating and editing jobs, database and crawler. 2. Athena: Creating and editing views and queries. 3. RedShift: Creating clusters, VPC connection and endpoints. 4. Quick sight: Creating and editing dashboard and performing analysis on the dataset.	Contributions: AWS Elastic beanstalk AWS Cloud formation AWS Cognito AWS API-Gateway AWS Lambda AWS Route 53
Contribution Percentage: 50%	Contribution Percentage: 50%
By signing below, I certify all information is true and correct to the best of my knowledge. Signature: prabhat Date: 14/06/2021	By signing below, I certify all information is true and correct to the best of my knowledge. Signature: karthi Date: 14/06/2021

Project Link

The project has been deployed and the link is: viccrimnal-dashboard.click

Introduction

The Australian government provides the Crime statistics dataset with all the offenses against persons and property based on the suburbs.

The objective is to educate people with the insights obtained by analyzing crime statistics. It is important to know their respective neighborhoods in order to be safe. When a person is planning to move to a new place and does not have sufficient easily available information, this application will come in handy.

The dataset is extracted from the Australian government website and it constitutes 3,17,504 records. This project utilizes the history data from 2011 to 2020. Another dataset is collected from the World Cities Database to extract the cities, Longitudes, and Latitudes as the main dataset contained only the suburbs.

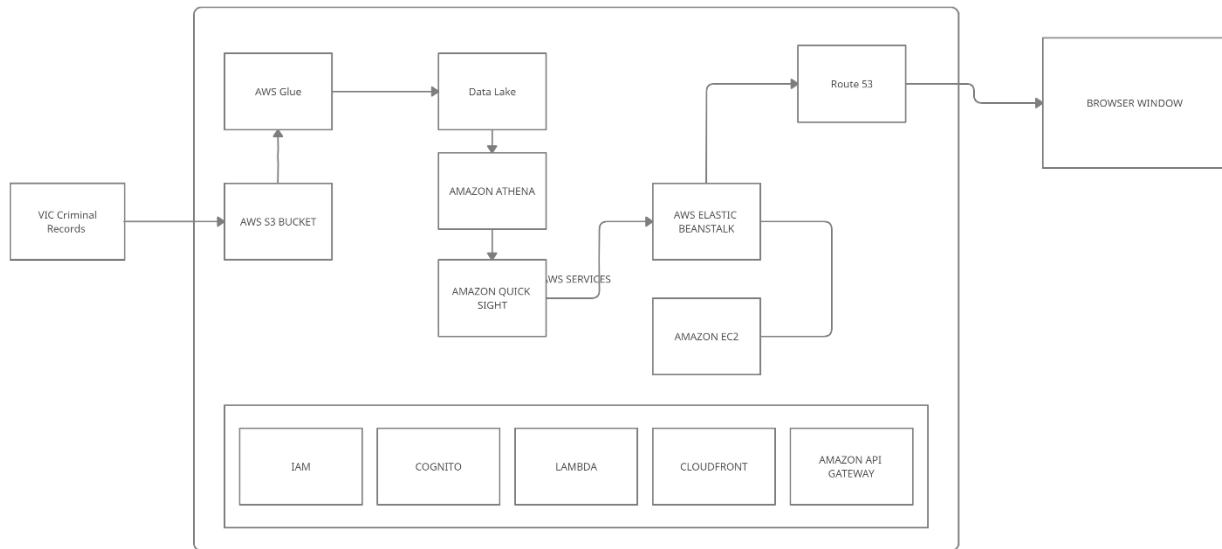
For data analysis, have used AWS services such as AWS Glue, Athena, Quick Sight RedShift, S3 Buckets etc. And for deployment, we have used services such Elastic Beanstalk, Cognito, Lambda, Route 53 etc.

Considering the increasing crime rate and the access to modern technology, it is easy to be updated about the safety of every other place in Australia.

Summary

The goal of this application to provide an analysis of the crime statistics dataset. This application uses AWS services to perform all the activities from uploading the dataset to deploying the application. AWS services like Quick Sight, Athena, RedShift, Elastic Beanstalk, Cognito etc., are used to provide a userfriendly interface with interactive dashboard. This project can benefit Australians to be safe by alerting them about their surroundings.

Architecture



Implementation

The following step by step guide shows how to recreate the CrimeRate project.

1. Creating buckets and IAM role:

Set up s3 Buckets

Bucket Name	Purpose
dataset-crime	To upload all dataset required for the project, scripts used for the jobs.

Create IAM roles

IAM Roles	Purpose
AWSGlueServiceRole-crawler-iam	Used for access s3 bucket, datalake, redshift and glue services

2. Uploading files to s3 bucket:

(i) Obtain the dataset and extract csv

The main dataset can be obtained from [Criminal Incident by LGA – Year ending Dec 2020](#) page.

The dataset is in xlsx format and contains multiple sheets. We require only third table from it. To extract it, use the following script in python. It can be run locally in your machine.

```
import csv
import pandas as pd
import requests
from urllib.request import urlopen
import openpyxl

# loading data from crime dataset xlsx file
crime_dataset_link = "https://www.crimestatistics.vic.gov.au/sites/default/files/embridge_cache/emshare/original/public/users/202103/f1/092cf1840/Data_Table_s_LGA_Criminal_Incidents_Year_Ending_December_2020.xlsx"

xlsx_link_open = urlopen(crime_dataset_link)
crime_xlsx_file = xlsx_link_open.read()

# only table 3 is required
crime_excel1 = pd.read_excel(crime_xlsx_file, 'Table 03', engine='openpyxl')
crime_excel1.to_csv('data_3.csv', encoding='utf-8', index=False)
```

(ii) Upload the csv to bucket

The above script will generate data_3.csv. Create a new folder inside s3 bucket **dataset-crime**, name the folder **dataset** and create one more folder **crime_records**. Upload this csv file inside **crime_records**.

Cloud Computing Assignment 3 Report

• • •

The resulting path will be: **dataset-crime/dataset/crime_records/data_3.csv**

Amazon S3 > dataset-crime > dataset/ > crime_records/

crime_records/

Copy S3 URI

Objects Properties

Objects (1)

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](#)

C Copy S3 URI Copy URL Download Open Delete Actions ▾

Create folder Upload

Find objects by prefix

Name	Type	Last modified	Size	Storage class
data_3.csv	csv	June 9, 2021, 22:46:04 (UTC+10:00)	39.8 MB	Standard

3. AWS Data lake setup:

(a) Open datalake in your AWS account and click **Get Started** and then click **Register Location** to register you s3 bucket location.

AWS Lake Formation

Dashboard

Data catalog

Databases

Tables

Settings

Register and ingest

Data lake locations

Blueprints

Crawlers

Jobs

Permissions

Administrative roles and tasks

Policy tags

Tag permissions

Data permissions

▼ Data lake setup

Quickly set up your data lake in Lake Formation.

Stage 1	Stage 2	Stage 3
Register your Amazon S3 storage Lake Formation manages access to designated storage locations within Amazon S3. Register the storage locations that you want to be part of the data lake. Register location	Create a database Lake Formation organizes data into a catalog of logical databases and tables. Create one or more databases and then automatically generate tables during data ingestion for common workflows. Create database	Grant permissions Lake Formation manages access for IAM users, roles, and Active Directory users and groups via flexible database, table, and column permissions. Grant permissions to one or more resources for your selected users. Grant permissions

Recent access activity (0/50)

Recent access activity for your data lake in AWS CloudTrail. Events can take several minutes to appear in CloudTrail and are limited to the last 90 days.

Find events

Cloud Computing Assignment 3 Report

• • •

The screenshot shows the 'Amazon S3 location' registration step in the AWS Lake Formation console. It includes fields for the Amazon S3 path (set to 's3://dataset-crime'), IAM role selection (set to 'AWSServiceRoleForLakeFormationDataAccess'), and a note about selecting the service-linked role for EMR. At the bottom are 'Cancel' and 'Register location' buttons.

(b) After registration you can see your s3 bucket inside **Data lake locations**

The screenshot shows the 'Data lake locations' page in the AWS Lake Formation console. It lists a single registered location: 's3://dataset-crime' with an IAM role of 'AWSServiceRoleForLakeFormationDataAccess'. The page includes a search bar, actions menu, and navigation controls.

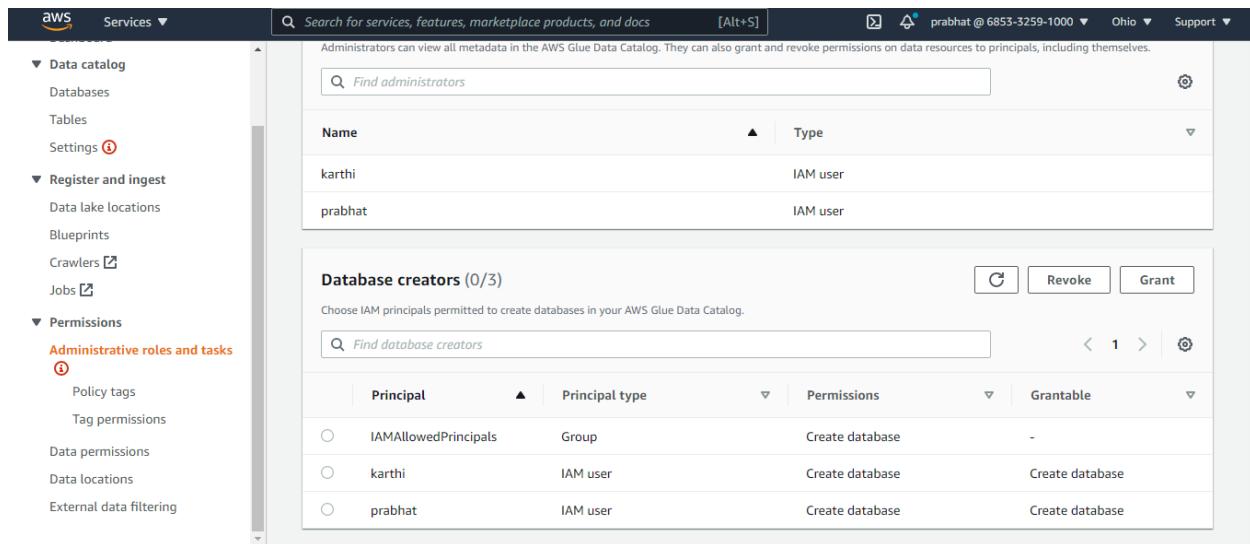
4. Database creation and AWS Glue Setup

Prerequisite

If you are sharing account, the under **Permissions/Administrative roles and tasks**, check whether you have access to Create database.

Cloud Computing Assignment 3 Report

• • •

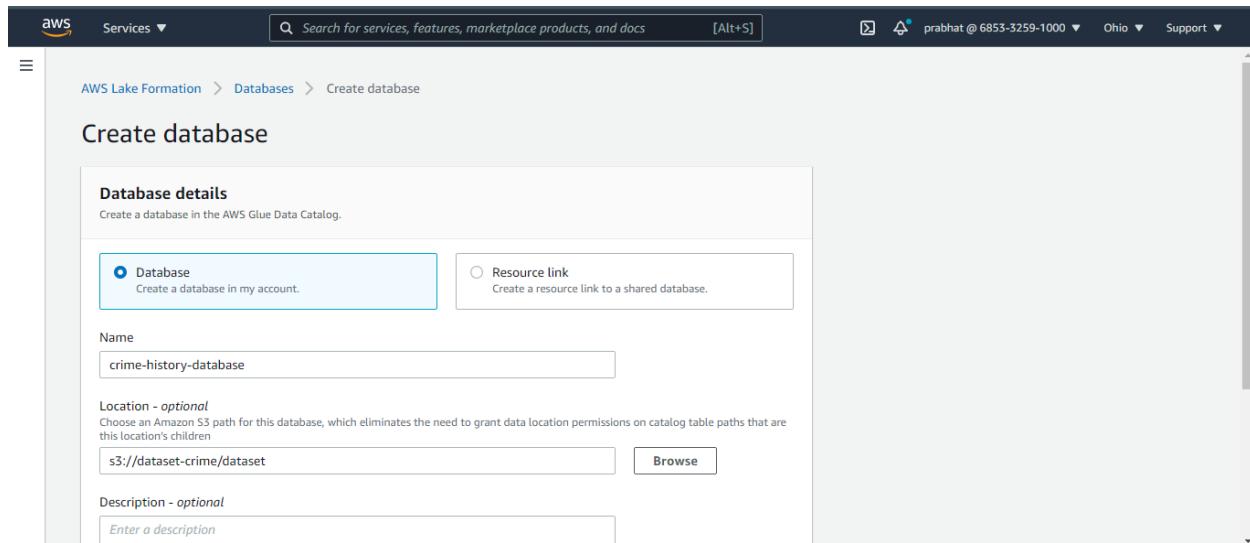


The screenshot shows the AWS Glue Data Catalog interface. On the left, a sidebar lists categories: Data catalog, Register and ingest, and Permissions. Under Permissions, there's a sub-section for Administrative roles and tasks. The main area displays two tables. The first table, titled "Administrators", lists users "karti" and "prabhat" both categorized as "IAM user". The second table, titled "Database creators (0/3)", shows no entries. At the bottom right of the main area are three buttons: "C" (Create), "Revoke", and "Grant".

(i) Creating database

(a) Open AWS Glue, go to Databases inside Data Catalog and click **Add database**.

(b) Provide a name for the database and location to dataset. We have named it **crime-history-database** and the location is provided as “dataset-crime/dataset/”.



The screenshot shows the "Create database" dialog from AWS Lake Formation. It has a header "AWS Lake Formation > Databases > Create database". The main section is titled "Create database" and contains a "Database details" form. The "Name" field is filled with "crime-history-database". The "Location - optional" field contains "s3://dataset-crime/dataset/" and has a "Browse" button next to it. There are two radio button options: "Database" (selected) and "Resource link".

(c) Verify the created database.

Cloud Computing Assignment 3 Report

• • •

The screenshot shows the AWS Lake Formation console. On the left, there's a sidebar with options like Dashboard, Data catalog (selected), Databases (highlighted in orange), Tables, Settings, Register and ingest, and Permissions. The main area shows a database named 'crime-history-database'. It has a 'Database details' section with fields: Name (crime-history-database), Amazon S3 path (s3://dataset-crime/dataset), Description (-), and two checkboxes: 'Default permissions for newly created tables' and 'Use only IAM access control for new tables in this database'. Below this is a 'Policy tags (0)' section with an 'Edit tags' button.

(ii) Crawler Setup

Go to Crawler under the Data Catalog section and click **Add crawler**. Fill in the details for crawler, as per the following images. We have used **crime-database-crawler** as the crawler's name.

(Note) The IAM **AWSGlueServiceRole-crawler-iam** will be created here.

The screenshot shows the 'Add crawler' wizard. On the left, there's a sidebar with checkboxes for: Crawler info (checked), Crawler source type (checked), Data store (checked), IAM Role (checked), Schedule (checked), Output (checked), and Review all steps (unchecked). The main area has three tabs: 'Crawler info' (Name: crime-database-crawler, Tags: -), 'Data stores' (Data store: S3: s3://dataset-crime/dataset, Include path: s3://dataset-crime/dataset, Connection: , Exclude patterns:), and 'IAM role' (IAM role: arn:aws:iam::68532591000:role/service-role/AWSGlueServiceRole-crawler-iam). At the bottom, there are buttons for 'Next Step' and 'Cancel'.

Cloud Computing Assignment 3 Report

Add crawler

- Crawler info: crime-database-crawler
- Crawler source type: Data stores
- Data store: S3: s3://dataset-cr...
- IAM Role: arn:aws:iam:686532591000:role/service-role/AWSGlueServiceRole-crawler-iam
- Schedule: Run on demand
- Output: Database: crime-history-database, Prefix added to tables (optional): false

Finish

(iii) Loading additional dataset using Glue ETL Job

- (a) To the IAM role, AWSGlueServiceRole-crawler-iam, provide access to s3 bucket for read and write operation.

Identity and Access Management (IAM)

Dashboard

Access management

User groups

Users

Roles

Policies

Identity providers

Account settings

Access reports

Access analyzer

Archive rules

Analyzers

Settings

Credential report

AWSGlueServiceRole-crawler-iam

Service	Access level	Resource	Request condition
S3	Full: Read, Write	Multiple	None

Policy summary { JSON Edit policy Simulate policy

Filter

Allow (1 of 284 services) Show remaining 283

Permissions boundary (not set)

Generate policy based on CloudTrail events

You can generate a new policy based on the access activity for this role, then customize, create, and attach it to this role. AWS uses your CloudTrail events to...

Cloud Computing Assignment 3 Report

• • •

- (b) Create a python script **load_dataset.py**, with the following content. Create a new folder inside s3 bucket dataset-crime, name it dataset-loading and upload the script there. The dataset is from [World Cities Database](#) page.

```
import csv
import pandas as pd
import requests
from urllib.request import urlopen
import boto3

# accessing the s3 bucket
bucket_name = "dataset-crime"
s3_resources = boto3.resource('s3')
s3_bucket = s3_resources.Bucket(bucket_name)

# loading city dataset in csv format
city_dataset_link = "https://simplemaps.com/static/data/country-
cities/au/au.csv"
city_csv_open = requests.get(city_dataset_link, allow_redirects=False)

# saving the file csv locally
open('austrlian_city.csv', 'wb').write(city_csv_open.content)

# uploading the csv file to s3 bucket
s3_bucket.meta.client.upload_file('austrlian_city.csv', bucket_name, 'austrlian_
_city/austrlian_city.csv')
```

The screenshot shows the AWS S3 console interface. At the top, the navigation bar includes the AWS logo, 'Services ▾', a search bar ('Search for services, features, marketplace products, and docs [Alt+S]'), user information ('prabhat @ 6853-3259-1000 ▾ Global ▾ Support ▾'), and a 'Copy S3 URI' button. Below the navigation, the path 'Amazon S3 > dataset-crime > dataset_loading/' is displayed. The main area shows a table titled 'Objects (1)' with one item: 'load_dataset.py'. The table has columns for Name, Type, Last modified, Size, and Storage class. The object 'load_dataset.py' is listed with a size of 1.0 KB and a storage class of Standard. At the bottom of the page, there are links for 'Feedback', 'English (US) ▾', '© 2008 - 2021, Amazon Internet Services Private Ltd. or its affiliates. All rights reserved.', 'Privacy Policy', 'Terms of Use', and 'Cookie preferences'.

Name	Type	Last modified	Size	Storage class
load_dataset.py	py	June 12, 2021, 00:16:44 (UTC+10:00)	1.0 KB	Standard

Cloud Computing Assignment 3 Report

• • •

(c) Create a new job inside AWS glue. The name of the job is **job_loadDataset** and configure it as per the following image.

Jobs A job is your business logic required to perform extract, transform and load (ETL) work. Job runs are initiated by triggers which can be scheduled or driven by events.

User preferences

Add job Action ▾ Filter by tags and attributes Showing: 1 - 3 < > 🔍 ?

Name	Type	ETL	Script	Last modified	Job
History	Details	Script			
Name job_loadDataset	Type Python shell	Python lib path -			
IAM role AWSGlueServiceRole-crawler-iam		Jar lib path -			
Glue version 1.0		Other lib path -			
Python version 3		Job parameters -			
Script location s3://dataset-crime/dataset_loading/load_dataset.py		Non-overridable -			
Job bookmark Disable		Job parameters -			
Job metrics Disable		Connections -			
Continuous logging Disable		Maximum capacity 0.0625			
Server-side encryption Disabled		Job timeout (minutes) 2880			
		Delay notification threshold (minutes) -			
		Tags -			

(d) Run the job and it will create a new csv file named **australian_city.csv** inside the bucket.

AWS Services ▾ Search for services, features, marketplace products, and docs [Alt+S] prabhat @ 6853-3259-1000 ▾ Ohio ▾ Support ▾

AWS Glue User preferences

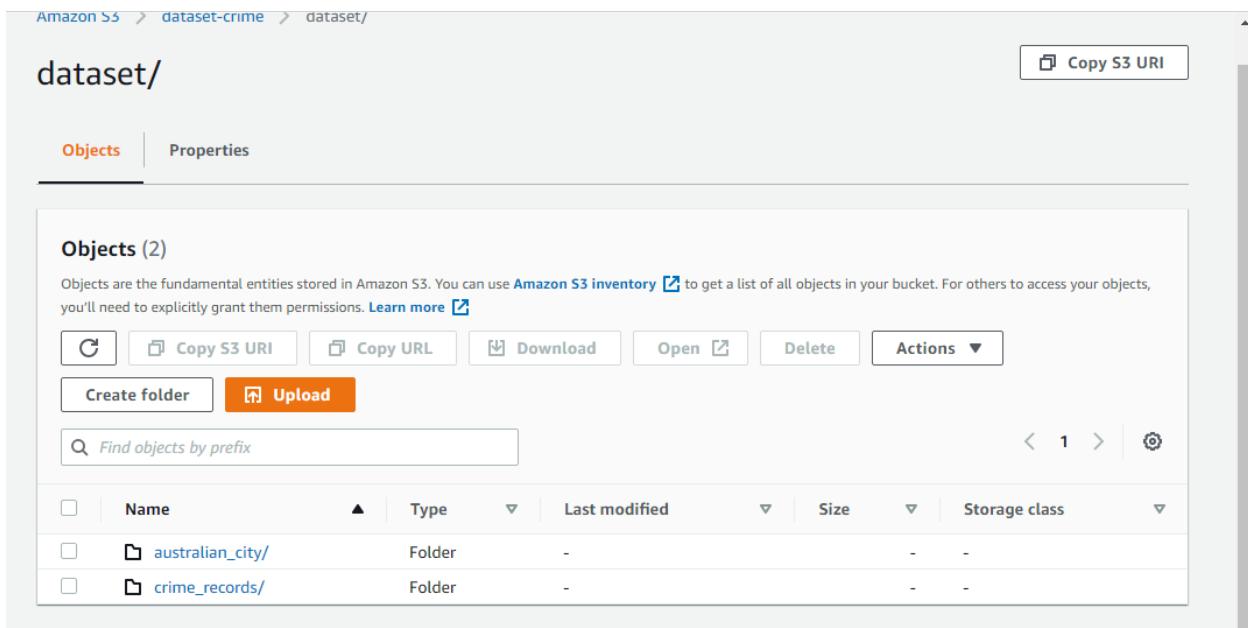
Add job Action ▾ Filter by tags and attributes Showing: 1 - 1 < > 🔍 ?

Name	Type	ETL language	Script location	Last modified	Job bookmark
job_loadDataset	Python shell	s3://dataset-...	12 June 2021 12:18 A...	Disable	

(e) Verify that the csv file is created.

Cloud Computing Assignment 3 Report

• • •

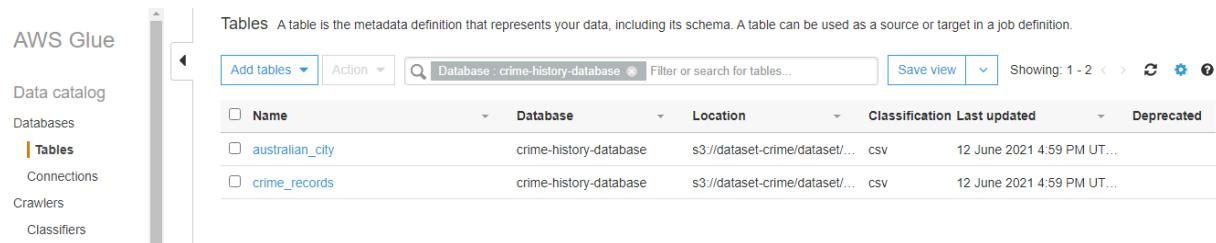


The screenshot shows the Amazon S3 console interface. The path is `Amazon S3 > dataset-crime > dataset/`. The bucket name is `dataset/`. There are two objects listed: `australian_city/` and `crime_records/`. The `Objects` tab is selected. At the top right, there is a `Copy S3 URI` button. Below the objects, there are buttons for `Create folder`, `Upload`, and search. A table below lists the objects with columns for Name, Type, Last modified, Size, and Storage class.

Name	Type	Last modified	Size	Storage class
<code>australian_city/</code>	Folder	-	-	-
<code>crime_records/</code>	Folder	-	-	-

(iv) Running Crawler to add tables in database

Run the crawler “crime-database-crawler”, and it will add tables to the crime-history-database.



The screenshot shows the AWS Glue Data Catalog interface. The sidebar shows options like AWS Glue, Data catalog, Databases, Tables (selected), Connections, Crawlers, and Classifiers. The main area is titled "Tables" and describes them as metadata definitions. It shows a list of tables with columns for Name, Database, Location, Classification, Last updated, and Deprecated. Two tables are listed: `australian_city` and `crime_records`.

Name	Database	Location	Classification	Last updated	Deprecated
<code>australian_city</code>	crime-history-database	s3://dataset-crime/dataset/...	csv	12 June 2021 4:59 PM UT...	
<code>crime_records</code>	crime-history-database	s3://dataset-crime/dataset/...	csv	12 June 2021 4:59 PM UT...	

5. Verifying database content using AWS Athena

Load the database in AWS Athena and verify its content. At this point we should have two tables, `crime_records` and `australian_city`.

Cloud Computing Assignment 3 Report

• • •

The screenshot shows the AWS Athena console interface. On the left, the sidebar displays the Data source (AwsDataCatalog), Database (crime-history-database), and Tables (australian_city, crime_records). The Views section is empty. The main area shows a query editor with the following content:

```
1 |SELECT * FROM "crime-history-database"."crime_records" limit 10;
```

Below the query editor, the status bar indicates: (Run time: 0.99 seconds, Data scanned: 1.58 MB). The results table contains 6 rows of data:

	year	year ending	local government area	postcode	suburb/town name	offence division	offence subdivision
1	2016	December	Moonee Valley	3040	Essendon	B Property and deception offences	B30 Burglary/Break and ent
2	2016	December	Moonee Valley	3040	Essendon	B Property and deception offences	B30 Burglary/Break and ent
3	2016	December	Moonee Valley	3040	Essendon	B Property and deception offences	B30 Burglary/Break and ent
4	2016	December	Moonee Valley	3040	Essendon	B Property and deception offences	B40 Theft
5	2016	December	Moonee Valley	3040	Essendon	B Property and deception offences	B40 Theft
6	2016	December	Moonee Valley	3040	Essendon	B Property and deception offences	B40 Theft

The screenshot shows the AWS Athena console interface, identical to the one above but with a different query. The sidebar shows the same database and tables. The main area shows a query editor with the following content:

```
1 |SELECT * FROM "crime-history-database"."australian_city" limit 10;
```

Below the query editor, the status bar indicates: (Run time: 1.28 seconds, Data scanned: 67.07 KB). The results table contains 6 rows of data:

	city	lat	long	country	iso2	admin_name	capital	population	population_proper
1	Sydney	-33.865	151.2094	Australia	AU	New South Wales	admin	5312163	4840600
2	Melbourne	-37.8136	144.9631	Australia	AU	Victoria	admin	5078193	4529500
3	Brisbane	-27.4678	153.0281	Australia	AU	Queensland	admin	2514184	2360241
4	Perth	-31.9522	115.8589	Australia	AU	Western Australia	admin	2059484	2039200
5	Adelaide	-34.9289	138.6011	Australia	AU	South Australia	admin	1345777	1295714
6	Gold Coast	-28.0167	153.4	Australia	AU	Queensland		679127	638090

6. AWS RedShift setup

(i) Creating cluster in AWS Redshift

Create a new cluster in redshift, name it **redshift-cluster-1**. The node type dc2.large is sufficient for our database.

Cloud Computing Assignment 3 Report

• • •

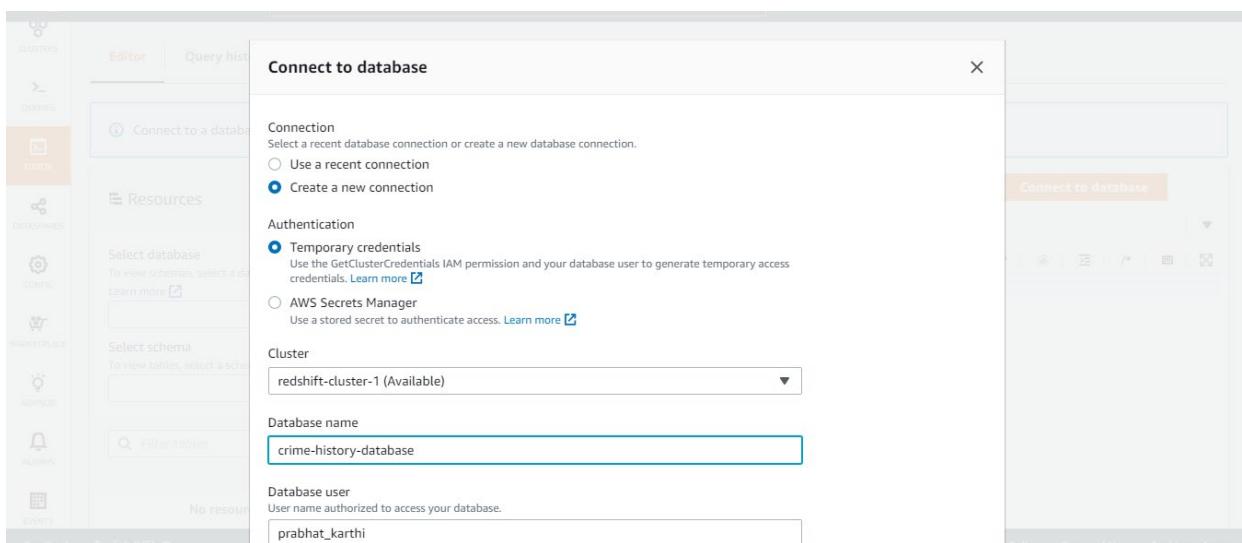
The screenshot shows the Amazon Redshift console interface. On the left, there's a sidebar with icons for Dashboard, Clusters (selected), Editor, DataShares, Config, Marketplace, Advisor, and Events. The main area shows the 'Clusters' section for 'redshift-cluster-1'. The cluster identifier is 'redshift-cluster-1', status is 'Available', node type is 'dc2.large', number of nodes is 1, storage used is 'AQUA', and endpoint is 'redshift-cluster-1.chp9nvtwk4ha.us-east-1.redshift.amazonaws.com'. JDBC URL is 'jdbc:redshift://redshift-cluster-1.chp9nvtwk4ha.us-east-1.redshift.amazonaws.com:5439', and ODBC URL is 'Driver={Amazon Redshift (x64)}; Server=redshift-cluster-1.chp9nvtwk4ha.us-east-1.redshift.amazonaws.com; Port=5439; Database=public; User=prabhat; Password=...'. Below the cluster details, there are tabs for Cluster performance, Query monitoring, Schedules, Maintenance, and Properties (selected).

(ii) Creating a new database inside Redshift

We have to create a new database inside redshift so transfer our tables from existing database (crime-history-database). As hyphens are not allowed for nomenclature in redshift query, name of new database will be **crime_history_database**.

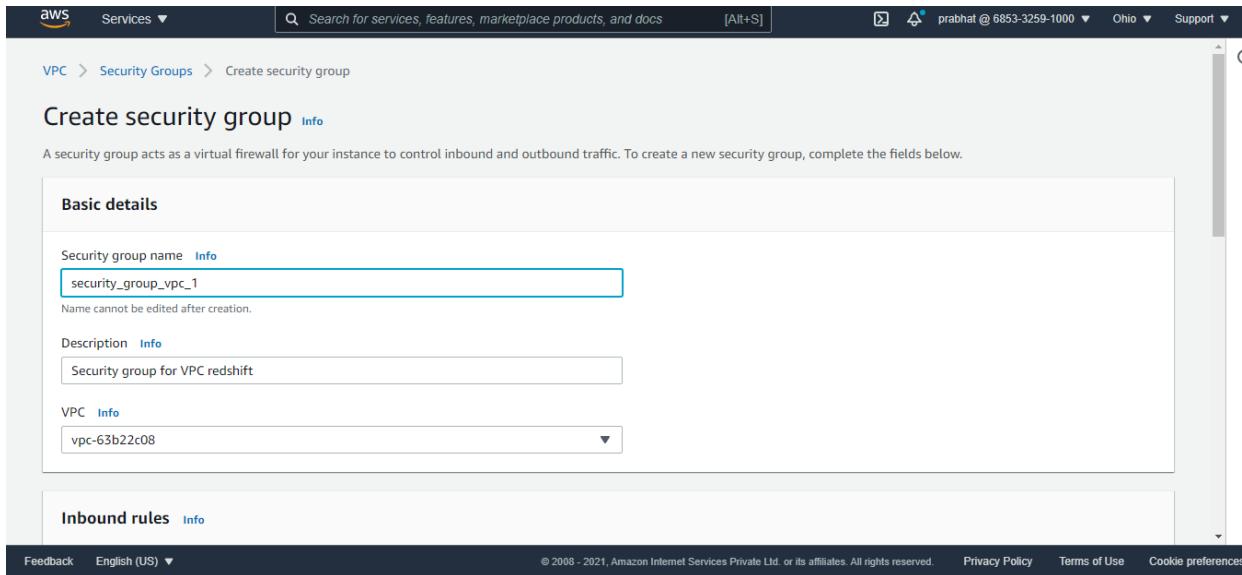
The screenshot shows the AWS Redshift Query Editor. The sidebar includes icons for Editor (selected), DataShares, Config, Marketplace, Advisor, and Events. The main area has a search bar and a 'Query 1' tab. Under 'Select database', it says 'To view schemas, select a database.' and 'dev' is selected. Under 'Select schema', it says 'To view tables, select a schema.' and 'public' is selected. A 'Filter tables' dropdown shows '1'. Below the schema dropdown, it says 'No resources' and 'No resources to display'. At the bottom, there are buttons for 'Run', 'Save', 'Schedule', and 'Clear', and a 'Send feedback' link.

Cloud Computing Assignment 3 Report



(iii) Creating security group

Creating a new security group for the VCP, with the following inbound and outbound rules.



Cloud Computing Assignment 3 Report

• • •

Inbound rules | Outbound rules | Tags

Inbound rules (2)

Type	Protocol	Port range	Source	Description - optional
All TCP	TCP	0 - 65535	52.15.247.160/27	US East (Ohio) (us-east-2)
All traffic	All	All	sg-09204c7f / default	Default

Inbound rules | **Outbound rules** | Tags

Outbound rules (3)

Type	Protocol	Port range	Destination	Description - optional
All TCP	TCP	0 - 65535	52.15.247.160/27	US East (Ohio) (us-east-2)
All TCP	TCP	0 - 65535	sg-09204c7f / default	Default
All traffic	All	All	0.0.0.0/0	-

(iv) Adding VPC endpoints

The VPC needs to access S3 bucket, thus creating new endpoint.

AWS Services ▾ [Alt+S] prabhat @ 6853-3259-1000 ▾ Ohio ▾ Support ▾

A VPC endpoint enables you to securely connect your VPC to another service. There are three types of VPC endpoints – Interface endpoints, Gateway Load Balancer endpoints, and gateway endpoints. Interface endpoints and Gateway Load Balancer endpoints are powered by AWS PrivateLink, and use an elastic network interface (ENI) as an entry point for traffic destined to the service. Interface endpoints are typically accessed using the public or private DNS name associated with the service, while gateway endpoints and Gateway Load Balancer endpoints serve as a target for a route in your route table for traffic destined for the service.

Service category AWS services Find service by name Your AWS Marketplace services

Service Name com.amazonaws.us-east-2.s3

Service Name	Owner	Type
com.amazonaws.us-east-2.s3	amazon	Gateway
com.amazonaws.us-east-2.s3	amazon	Interface

VPC* vpc-63b22c08

Cloud Computing Assignment 3 Report

The screenshot shows the AWS IAM Policy creation interface. At the top, there are three radio buttons: 'Full Access' (selected), 'Custom', and 'AWS Lambda'. Below this, a note says 'Use the [policy creation tool](#) to generate a policy, then paste the generated policy below.' A large text area contains a JSON template:

```
{ "Statement": [ { "Action": "sts:AssumeRole", "Effect": "Allow", "Resource": "*" } ] }
```

Below the JSON area are two input fields: 'Key' (128 characters maximum) and 'Value' (256 characters maximum).

(v) Creating connection to Redshift inside AWS Glue

Create a new connection inside the AWS Glue to newly created redshift cluster with the same VCP for which we have added security group and endpoint.

The screenshot shows the 'Add connection' wizard in AWS Glue. On the left, a sidebar lists steps: 'Connection properties' (checked), 'redshift_connection' (Type: Amazon Redshift), 'Connection access' (checked), 'redshift-cluster-1' (VPC Id: vpc-63b22c08), and 'Review all steps' (unchecked). The main area is divided into two sections: 'Connection properties' and 'Connection access'.

Connection properties

Name	redshift_connection
Type	Amazon Redshift
Require SSL connection	false
Description (optional)	redshift connection for crime_history_database inside redshift

Connection access

Cluster	redshift-cluster-1
Username	prabhat_karthi
VPC Id	vpc-63b22c08
Subnet	subnet-6900fc14
Security groups	sg-09204c7f

At the bottom are 'Back' and 'Finish' buttons.

(vi) Creating jobs for transferring tables

Cloud Computing Assignment 3 Report

• • •

(a) To transfer the tables from local database to redshift database, first the IAM roles required access to Lake Formation. Thus, use Attach Policy and add **AWSLakeFormationDataAdmin** policy to the IAM role **AWSGlueServiceRole-crawler-iam**.

The screenshot shows the AWS IAM Permissions policies page. On the left, there's a sidebar with navigation links like Dashboard, Access management, Roles, Policies, and Access reports. The 'Roles' section is currently selected. The main area displays a table of policies:

Policy name	Policy type	X
AWSGlueServiceRole	AWS managed policy	X
AWSGlueServiceRole-crawler-iam	Managed policy	X
AWSLakeFormationDataAdmin	AWS managed policy	X

Below the table, there are tabs for 'Policy summary' and 'JSON'. A 'Simulate policy' button is also present. A 'Filter' search bar is available. At the bottom, there's a table for 'Explicit deny (1 of 284 services)' with columns for Service, Access level, Resource, and Request condition.

(b) Now, add a new job with the following configuration to upload the table *australian_city* to redshift database *crime_history_database*. A new python script is created for this purpose.

The screenshot shows the 'Add job' configuration page. On the left, there's a sidebar with tabs: Job properties (selected), Data source, Transform type, Data target, and Schema. The main area is titled 'Configure the job properties'.

Job properties settings:

- Name: job_addCityDataToRedShift
- IAM role: AWSGlueServiceRole-crawler-iam
- Type: Spark
- Glue version: Spark 2.4, Python 3 with improved job startup times (Glue Version 2.0)

This job runs settings:

- A proposed script generated by AWS Glue (selected)
- An existing script that you provide
- A new script to be authored by you

Cloud Computing Assignment 3 Report

• • •

Add job

Job properties
 Data source
 Transform type
 Data target
 Schema

job_addCityDataToRedShift
S3 path where the script is stored
s3://dataset-crime/script_redshift
Temporary directory
s3://aws-glue-temporary-685332591000-us-east-2/prabhat

► Advanced properties
► Monitoring options
► Tags (optional)
► Security configuration, script libraries, and job parameters (optional)
► Catalog options (optional)

Add job

Job properties
job_addCityDataToRedShift
 Data source
australian_city
 Transform type
Change schema
 Data target
 Schema

Choose a data source

Filter by attributes or search by keyword

Name	Database	Location	Classification
viccriminalrec	viccriminalrec	s3://assignment3-glue/vicCriminalRec/	csv
australian_city	crime-history-database	s3://dataset-crime/dataset/australian...	csv
elb_logs	sampledb	s3://athena-examples-us-east-2/elb/p...	Unknown
crime_records	crime-history-database	s3://dataset-crime/dataset/crime_rec...	csv

Showing: 1 - 4 < >

Back Next

Add job

Job properties
job_addCityDataToRedShift
 Data source
australian_city
 Transform type
Change schema
 Data target
 Schema

Choose a transform type

Machine learning transforms are currently not supported for Glue 2.0.

Change schema
Change schema of your source data and create a new target dataset.

Find matching records
Use machine learning to find matching records within your source data

Back Next

Cloud Computing Assignment 3 Report

● ● ●

The screenshot shows the AWS Glue 'Add job' configuration interface. On the left, there is a sidebar with several checked configuration items: 'Job properties', 'Data source', 'Transform type', 'Data target', and 'Schema'. The main area displays a schema mapping table comparing columns from a source table ('australian_city') and a target table ('redshift_connection'). The columns listed are lat, long, country, iso2, admin_name, capital, population, and population_proper. Each column has its data type (double, string, bigint) and a mapping arrow pointing from the source column to the target column. To the right of each target column are four icons: a delete icon (X), a down arrow, an up arrow, and a refresh icon.

Source Column	Type	Target Column	Type
lat	double	lat	double
long	double	long	double
country	string	country	string
iso2	string	iso2	string
admin_name	string	admin_name	string
capital	string	capital	string
population	bigint	population	long
population_proper	bigint	population_proper	long

[Back](#) [Save job and edit script](#)

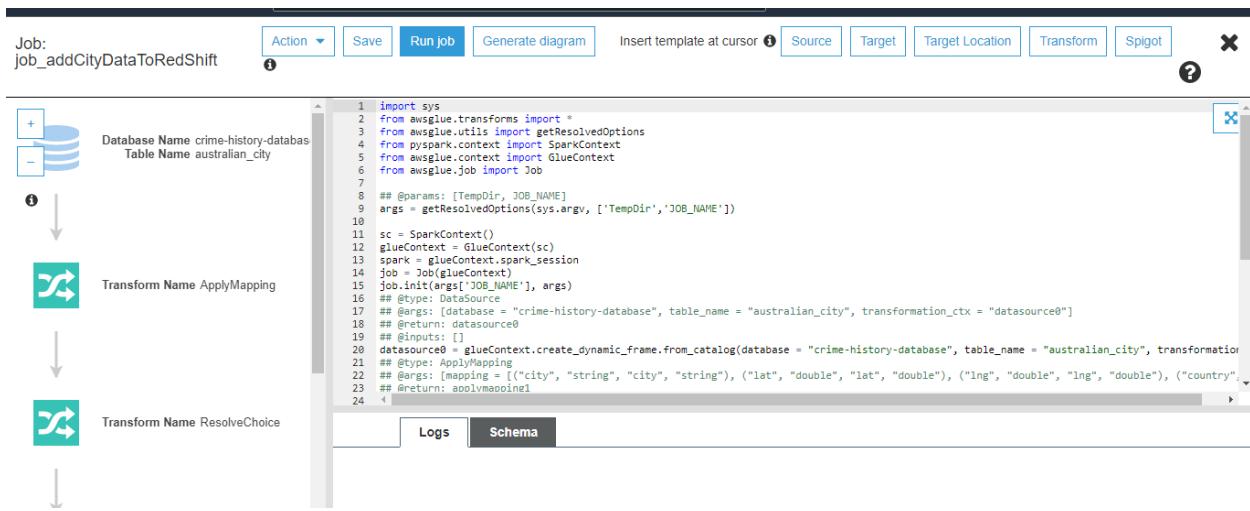
The screenshot shows the 'Choose a data target' step of the AWS Glue 'Add job' configuration. On the left, the same checked configuration items as the previous screen are listed. The main area is titled 'Choose a data target' and contains two radio button options: 'Create tables in your data target' (selected) and 'Use tables in the data catalog and update your data target'. Below these options are three dropdown menus: 'Data store' (set to 'JDBC'), 'Connection' (set to 'redshift_connection'), and 'Database name' (set to 'crime_history_database'). At the bottom of the screen, there are 'Back' and 'Next' buttons.

[Back](#) [Next](#)

(c) The following script is created automatically by the AWS based on the above configuration setting. Run the job to transfer tables to the redshift database.

Cloud Computing Assignment 3 Report

● ● ●



(d) After running the job, we can see that tables are loaded inside redshift database.

The screenshot shows the AWS Redshift Query Editor. On the left, a sidebar menu includes options like QUERIES (selected), EDITOR, DATASHARES, CONFIG, MARKETPLACE, ADVISOR, ALARMS, EVENTS, and WHAT'S NEW. The main area has a "Change connection" button and a "Query 2" tab. The "Select database" dropdown is set to "crime_history_database" and the "Select schema" dropdown is set to "public". A "Filter tables" search bar shows "australian_city" as the current filter. Below the schema list, "crime_records" is expanded, showing its attributes: "year", "year ending", "local government area", "postcode", "suburb/town name", and "offence division". The central pane contains a query editor with the SQL command "SELECT TOP 10 * FROM crime_records". Below the editor are buttons for Run, Save, Schedule, and Clear. To the right, there are icons for "Send feedback" and "Query results" (which is currently selected). At the bottom, there is a "Table details" tab.

(vii) Creating view

Views are required to create database inside AWS Quick Sight. Thus, creating a view for the same. Any changes done in the view will be reflected inside QuickSight as well. Following query is used to select required attributes, from both the tables.

Cloud Computing Assignment 3 Report

• • •

```
CREATE OR REPLACE VIEW crime_record_view AS

SELECT

"crime_rec"."year", "crime_rec"."year ending", "crime_rec"."local government area",
"crime_rec"."postcode", "crime_rec"."suburb/town name", "crime_rec"."offence division",
"crime_rec"."offence subdivision", "crime_rec"."offence subgroup", "crime_rec"."incidents recorded",

"aus_city"."city", "aus_city"."lat", "aus_city"."lng", "aus_city"."country", "aus_city"."population",
"aus_city"."population_proper", "aus_city"."admin_name"

FROM crime_records crime_rec, australian_city aus_city

WHERE

"crime_rec"."local government area" = "aus_city"."city" OR "crime_rec"."suburb/town name" =
"aus_city"."city";
```

The screenshot shows the AWS Lambda SQL Editor interface. On the left, there's a sidebar with various AWS services like Editor, DataShares, Config, Marketplace, Advisor, Alarms, Events, and What's New. The main area has tabs for Resources, Services, and a search bar. A connection named 'crime_rec...' is selected. The code editor contains the SQL query provided in the previous text block. Below the code editor are buttons for Run, Save, Schedule, and Clear. At the bottom, there are tabs for Query results and Table details, along with links for Feedback, English (US), Privacy Policy, Terms of Use, and Cookie preferences.



7. RedShift vs Athena

For this project, we are using Athena. The view which has been created for RedShift can be used with in Athena as well with modification in database name.

8. AWS QuickSight Setup

Creating view in AWS Athena

We can use the same query as used inside RedShift, with changes in database name. The view name is **crime_record_view**.

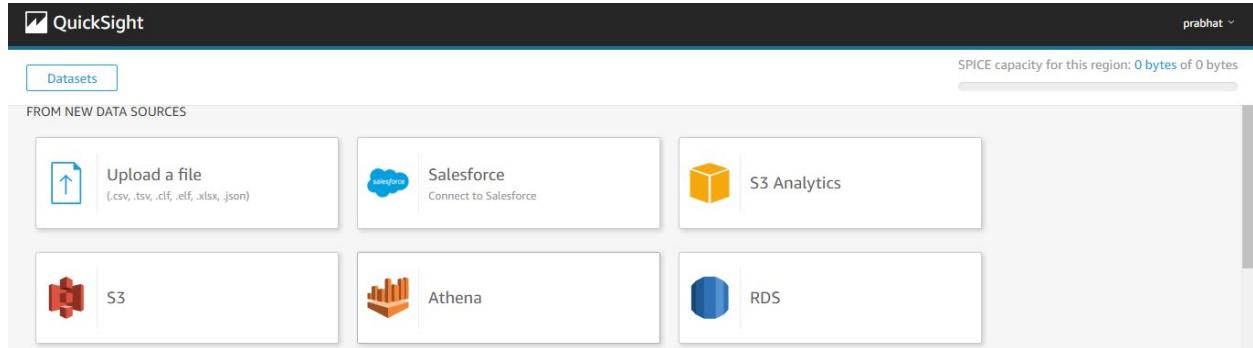
```
CREATE OR REPLACE VIEW crime_record_view AS  
  
SELECT  
  
"crime_rec"."year", "crime_rec"."year ending", "crime_rec"."local government area",  
"crime_rec"."postcode", "crime_rec"."suburb/town name", "crime_rec"."offence division",  
"crime_rec"."offence subdivision", "crime_rec"."offence subgroup", "crime_rec"."incidents recorded",  
  
"aus_city"."city", "aus_city"."lat", "aus_city"."lng", "aus_city"."country", "aus_city"."population",  
"aus_city"."population_proper", "aus_city"."admin_name"  
  
FROM "crime-history-database"."crime_records" crime_rec, "crime-history-database"."australian_city"  
aus_city  
  
WHERE  
  
"crime_rec"."local government area" = "aus_city"."city" OR  
"crime_rec"."suburb/town name" = "aus_city"."city";
```

Cloud Computing Assignment 3 Report

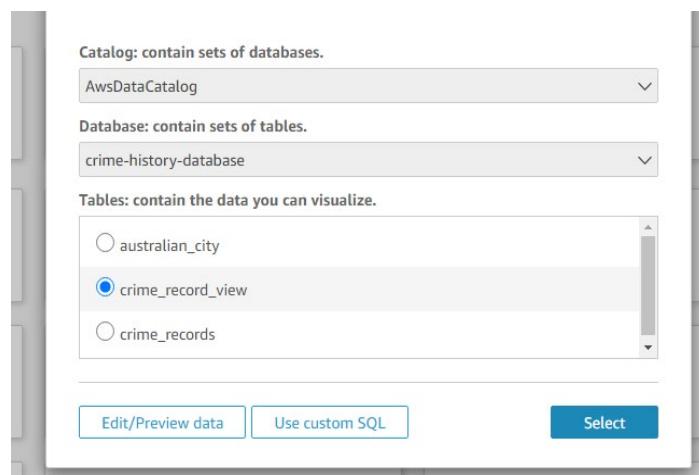
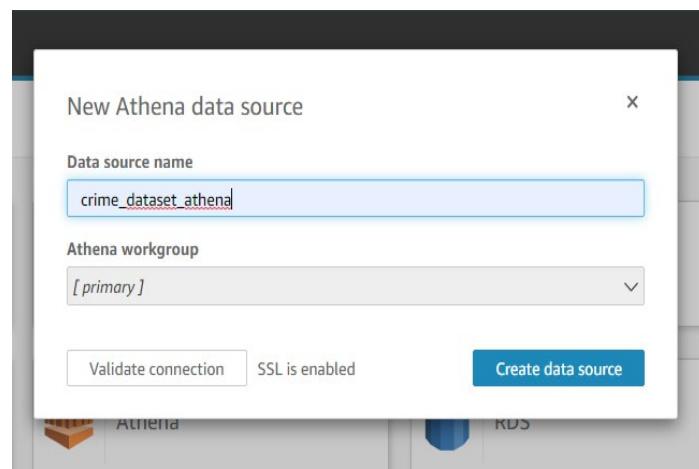
• • •

Loading dataset from view created

- (a) Go to AWS QuickSight and select *Datasets > New dataset > Athena*.



- (b) Name the data source as **crime_dataset_athena**, select the view created from the last step *crime_record_view* and fill the details according to the given images.



Cloud Computing Assignment 3 Report

• • •

(c) Verify the loaded data and start the analysis.

Dataset Name: crime_record_view

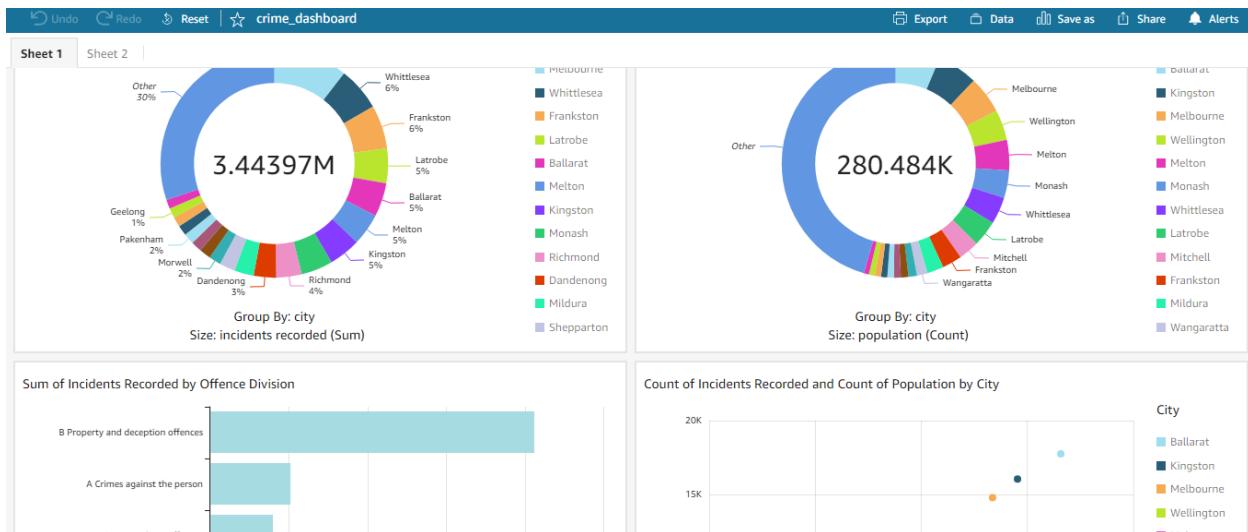
Query mode: SPICE

Fields: All fields included

Focus: All fields

Excluded fields: No fields excluded

year	year ...	local...	post...	subu...	offe...	offe...	offe...	incid...	city
2020	December	Ballarat	3350	Alfredton	A Crimes ag...	A20 Assault ...	A211 FV Ser...	7	Ball...
2020	December	Ballarat	3350	Alfredton	A Crimes ag...	A20 Assault ...	A212 Non-F...	7	Ball...
2020	December	Ballarat	3350	Alfredton	A Crimes ag...	A20 Assault ...	"A22 Assault...		Ball...
2020	December	Ballarat	3350	Alfredton	A Crimes ag...	A20 Assault ...	A231 FV Co...	5	Ball...
2020	December	Ballarat	3350	Alfredton	A Crimes ag...	A20 Assault ...	A232 Non-F...	8	Ball...
2020	December	Ballarat	3350	Alfredton	A Crimes ag...	A50 Robbery	A51 Aggrav...	1	Ball...
2020	December	Ballarat	3350	Alfredton	A Crimes ag...	"A70 Stalkin...	harassment		Ball...





9. Elastic Beanstalk environment setup:

Creating an Elastic beanstalk application which works as a front-end python flask application.

Environment information

Choose the name, subdomain, and description for your environment. These cannot be changed later.

Application name: vicriminal

Environment name: Vicriminal-env

Domain: vicriminal.us-east-2.elasticbeanstalk.com

Description: (empty)

Platform

Managed platform Platforms published and maintained by AWS Elastic Beanstalk. Learn more [\[?\]](#)

Custom platform Platforms created and owned by you.

Platform: Python

Platform branch: Python 3.8 running on 64bit Amazon Linux 2

Platform version: 3.3.1 (Recommended)

Platform

Managed platform Platforms published and maintained by AWS Elastic Beanstalk. Learn more [\[?\]](#)

Custom platform Platforms created and owned by you.

Platform: Python

Platform branch: Python 3.8 running on 64bit Amazon Linux 2

Platform version: 3.3.1 (Recommended)

Application code

Sample application Get started right away with sample code.

Existing version Application versions that you have uploaded for vicriminal.

app-Ofc1-210614_035446

Upload your code Upload a source bundle from your computer or copy one from Amazon S3.

Create environment

Cloud Computing Assignment 3 Report

• • •

Setting up the listeners, processes in order to access https port.

Port	Protocol	SSL certificate	Default process	Enabled
80	HTTP	--	default	<input checked="" type="checkbox"/>
443	Pending create	www.viccrimnal-dashboard.click - f569fdad-1d80-40e5-a5b9-5ab941091410	default	<input checked="" type="checkbox"/>

Name	Port	Protocol	HTTP code	Health check path	Stickiness
default	80	HTTP	/	/	disabled
https	Pending create	443	200	/	disabled

Name	Listener port	Priority	Host headers	Path patterns	Process
http	Pending create	80	1	/*	default
https	Pending create	443	2	/*	default

Access log files
Configure Elastic Load Balancing to capture logs with detailed information about requests sent to your Load Balancer. Logs are stored in Amazon S3. [Learn more](#)

Store logs
(Standard Amazon S3 charges apply.)
 Enabled

S3 bucket
(You must first configure bucket permissions. [Learn more](#))
-- Choose an Amazon S3 bucket --

Prefix
Logical hierarchy in the bucket. If you don't specify a prefix, Elastic Load Balancing stores access logs at the bucket's root.

Cancel Continue Apply

Cloud Computing Assignment 3 Report

• • •

10. Route 53 setup:

Registering a domain in AWS Route 53 console.

The screenshot shows the AWS Route 53 Registered domains page for the domain `viccrimnal-dashboard.click`. The domain was registered on 2021-06-13 and expires on 2022-06-13. It has an auto-renew setting enabled. The name servers listed are ns-1849.awsdns-39.co.uk, ns-546.awsdns-04.net, ns-1393.awsdns-46.org, and ns-463.awsdns-57.com. The DNSSEC status is disabled. The registrant contact is Karthi Narendrababu Geetha, and the administrative contact is also Karthi Narendrababu Geetha. Both contacts have email addresses `ngkarthi71@gmail.com` and phone numbers +61 0477213120. The technical contact is listed as Karthi Narendrababu Geetha, ngkarthi71@gmail.com, +61 0477213120, Unit 209, Unit 209, MELBOURNE VIC 3000 AU.

Creating a CNAME record in order to route all the traffics from the elastic beanstalk URL via the Route 53 domain.

The screenshot shows the AWS Route 53 Quick create record interface. A message at the top says "Introducing the new Route 53 console" and provides a link to the old console. The "Record 1" section is active, showing a CNAME record being created. The record name is `viccrimnal`, the record type is `CNAME`, and the value is `viccrimnal.us-east-2.elasticbeanstalk.com`. The TTL is set to 300 seconds, and the routing policy is `Simple routing`. Below the form, there's a link to "View existing records".

11. AWS Certification Manager:

Requesting for SSL certificate for our domain.

The screenshot shows the 'Request a certificate' page in the AWS Management Console. At the top, there's a navigation bar with the AWS logo, services dropdown, search bar ('Search for services, features, marketplace products, and docs'), and user information ('ngkarthi71', 'Ohio', 'Support'). Below the navigation is a message: 'Choose Import a certificate to import an existing certificate instead of requesting a new one. Learn more.' followed by a blue 'Import a certificate' button. The main section is titled 'Request a certificate' with the sub-instruction 'Choose the type of certificate for ACM to provide.' It contains two options: ' Request a public certificate' (selected) and ' Request a private certificate'. Each option has a detailed description and a 'Learn more' link. At the bottom right of this section are 'Cancel' and 'Request a certificate' buttons. The footer of the page includes links for 'Feedback', 'English (US)', 'Privacy Policy', 'Terms of Use', and 'Cookie preferences'.

Cloud Computing Assignment 3 Report

• • •

12. AWS Cloud Formation:

Creating a static s3 application by following the steps given in the URL. The end goal of this stack formation is to provide users the public access to the Quick sight dashboard via the cloud front distribution URL.

The screenshot shows the AWS CloudFormation console with the 'Stacks' page selected. A search bar at the top right contains the placeholder 'Search for services, features, marketplace products, and docs [Alt+S]'. The main area displays three stacks:

- awsbe-e-cdyyyjum6-stack**: Status: UPDATE_COMPLETE, Created: 2021-06-14 12:47:02 UTC+1000.
- CognitoQuickSight**: Status: CREATE_COMPLETE, Created: 2021-06-14 01:24:09 UTC+1000.
- vpc-publicprivate**: Status: CREATE_COMPLETE, Created: 2021-06-09 05:47:58 UTC+1000.

The right panel shows the detailed configuration for the **CognitoQuickSight** stack, including its Stack ID, Status, Root stack, Created time, Updated time, Drift status, Termination protection, and various status reasons and IAM roles.

After the stack formation task, we need to upload the .js and .html files to the s3 bucket, which is dedicatedly created for the public cloud formation.

The screenshot shows the AWS S3 console with the 'Upload' page selected. A search bar at the top right contains the placeholder 'Search for services, features, marketplace products, and docs [Alt+S]'. The main area shows a file upload interface:

- A message: "Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDK or Amazon S3 REST API. Learn more [?]."
- A large input field with the placeholder "Drag and drop files and folders you want to upload here, or choose Add files, or Add folders."
- A table titled "Files and folders (4 Total, 547.8 KB)" showing the uploaded files:

Name	Folder	Type	Size
amazon-cognito-auth.min.js	-	text/javascript	23.4 KB
auth.js	-	text/javascript	722.0 B
aws-cognito-sdk.js	-	text/javascript	516.7 KB
index.html	-	text/html	6.9 KB
- A "Destination" section with the URL "s3://cognitoquicksight-s3website-1osvhhek3mfj1".
- A "Destination details" section with the note "Bucket settings that impact new objects stored in the specified destination."
- A "Permissions" section with the note "Grant public access and access to other AWS accounts."

Cloud Computing Assignment 3 Report

• • •

13. AWS Cognito:

Once the cloud formation stack is completed, a static s3 website is created. In order to provide authentication for the website, we're embedding AWS Cognito. Cognito will authenticate the users and provide access to specific content.

For the Cognito App client, the callback URL should be the cloud-front domain.

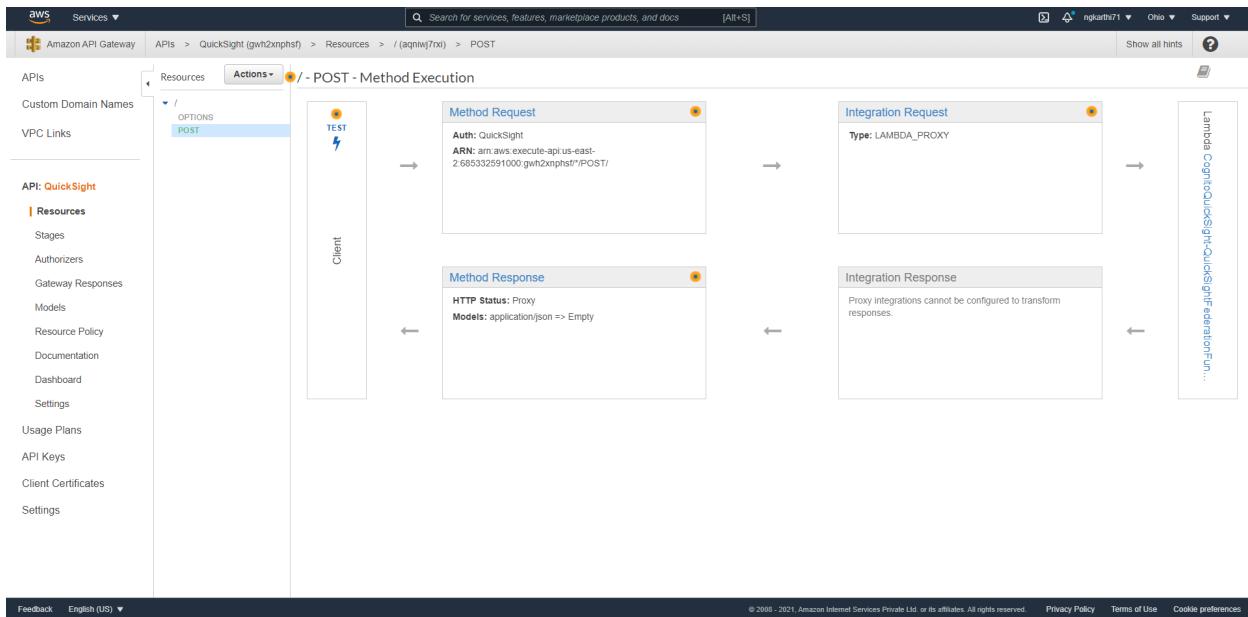
The screenshot shows the AWS Cognito User Pools console. On the left, there's a navigation sidebar with various options like General settings, Users and groups, Attributes, Policies, MFA and verifications, Advanced security, Message customizations, Tags, Devices, App clients, Triggers, Analytics, App integration, Federation, Identity providers, and Attribute mapping. The 'App clients' section is currently selected. The main content area is titled 'What identity providers and OAuth 2.0 settings should be used for your app clients?'. It shows a configuration for an app client named 'QuickSight'. Under 'Enabled Identity Providers', 'Cognito User Pool' is selected. Under 'Sign in and sign out URLs', the 'Callback URL(s)' field contains 'https://d43p0bun2dy6r.cloudfront.net'. Under 'OAuth 2.0', the 'Allowed OAuth Flows' section includes 'Implicit grant' and 'OpenID Connect'. The 'Allowed OAuth Scopes' section includes 'email', 'openid', and 'aws.cognito.signin.user.admin'. At the bottom, there's a note about the 'Hosted UI'.

The screenshot shows the continuation of the AWS Cognito User Pools console. The left sidebar remains the same. The main content area is titled 'What domain would you like to use?'. It has two sections: 'Amazon Cognito domain' and 'Your own domain'. In the 'Amazon Cognito domain' section, a domain prefix 'viccriminal-qs' is entered in the 'Domain prefix' input field, followed by '.auth.us-east-2.amazoncognito.com' and a 'Delete domain' button. In the 'Your own domain' section, there's a note about needing an ACM certificate and an alias record. A 'Use your domain' button is present. At the bottom, there are 'Go to summary' and 'Customize UI' buttons, along with standard footer links for Feedback, English (US), Privacy Policy, Terms of Use, and Cookie preferences.



14. AWS API Gateway:

The cloud formation stack will also create an API-Gateway in order to make a secure transaction between Cognito and the s3 static application. Here we're using the Cognito user pool as an Authorizer.



The screenshot shows the AWS API Gateway interface for creating a new Authorizer. The configuration is as follows:

- Name:** QuickSight
- Type:** Cognito
- Cognito User Pool:** us-east-2 - QuicksightUsers
- Token Source:** Token Validation
- Authorization:** (empty field)

Once the API model is deployed, a staging URL is generated that will be used by the application as an endpoint.

Cloud Computing Assignment 3 Report

The screenshot shows the AWS API Gateway Stage Editor for the 'prod' stage of the 'QuickSight' API. The left sidebar lists various API resources like APIs, Stages, Authorizers, and Models. The main panel displays the 'prod Stage Editor' with tabs for Settings, Logs/Tracing, Stage Variables, SDK Generation, Export, Deployment History, Documentation History, and Canary. Under Settings, there are sections for Cache Settings (Enable API cache), Default Method Throttling (Enable throttling checked, Rate: 10000 requests per second, Burst: 5000 requests), and Web Application Firewall (WAF) (Select Web ACL: None). There is also a Client Certificate section (Certificate: None). At the bottom right is a 'Save Changes' button.

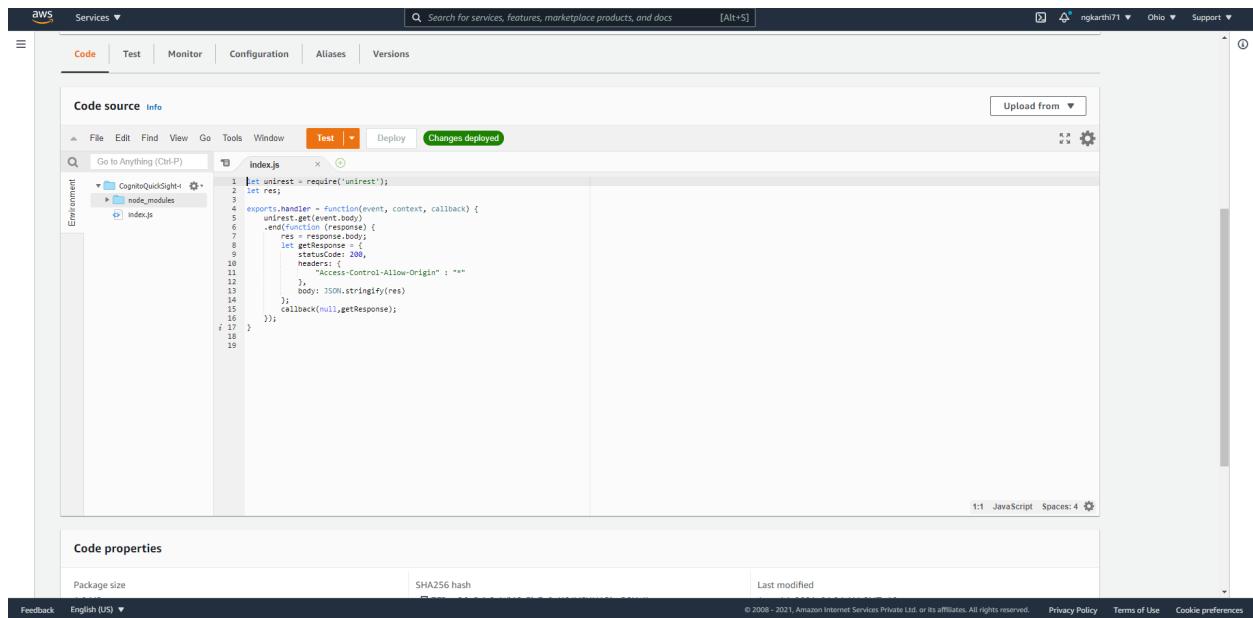
15. AWS Lambda:

The cloud stack formation creates a lambda function, which will be triggered by the AWS API-Gateway and it will provide response to the s3 application.

The screenshot shows the AWS Lambda Function Overview page for the 'CognitoQuickSight-QuickSightFederationFunction' function. The top navigation bar includes services, search, and user information. The main area shows the function name and a message indicating it belongs to an application. Below this is the 'Function overview' section, which includes a thumbnail of the function, its ARN, layers (0), and triggers (API Gateway). A 'Code' tab is selected at the bottom.

Cloud Computing Assignment 3 Report

• • •



The screenshot shows the AWS Lambda Code source interface. At the top, there's a navigation bar with 'Services' dropdown, search bar ('Search for services, features, marketplace products, and docs'), and user info ('ngkarthi71 Ohio Support'). Below the navigation is a tabs bar with 'Code' (selected), 'Test', 'Monitor', 'Configuration', 'Aliases', and 'Versions'. The main area is titled 'Code source' with tabs 'Info' and 'Code properties'. A code editor window displays 'index.js' with the following code:

```
1  let unirest = require('unirest');
2  let res;
3
4  exports.handler = function(event, context, callback) {
5    unirest.get(event.body)
6      .end(function (response) {
7        res = response.body;
8        let headers = {
9          statusCode: 200,
10            headers: {
11              "Access-Control-Allow-Origin": "*"
12            },
13            body: JSON.stringify(res)
14        };
15      });
16    callback(null,getResponse);
17  }
18
19
```

Below the code editor are sections for 'Code properties' (Package size, SHA256 hash, Last modified), 'Feedback' (English (US)), and footer links (Privacy Policy, Terms of Use, Cookie preferences).

16. Developer Manual:

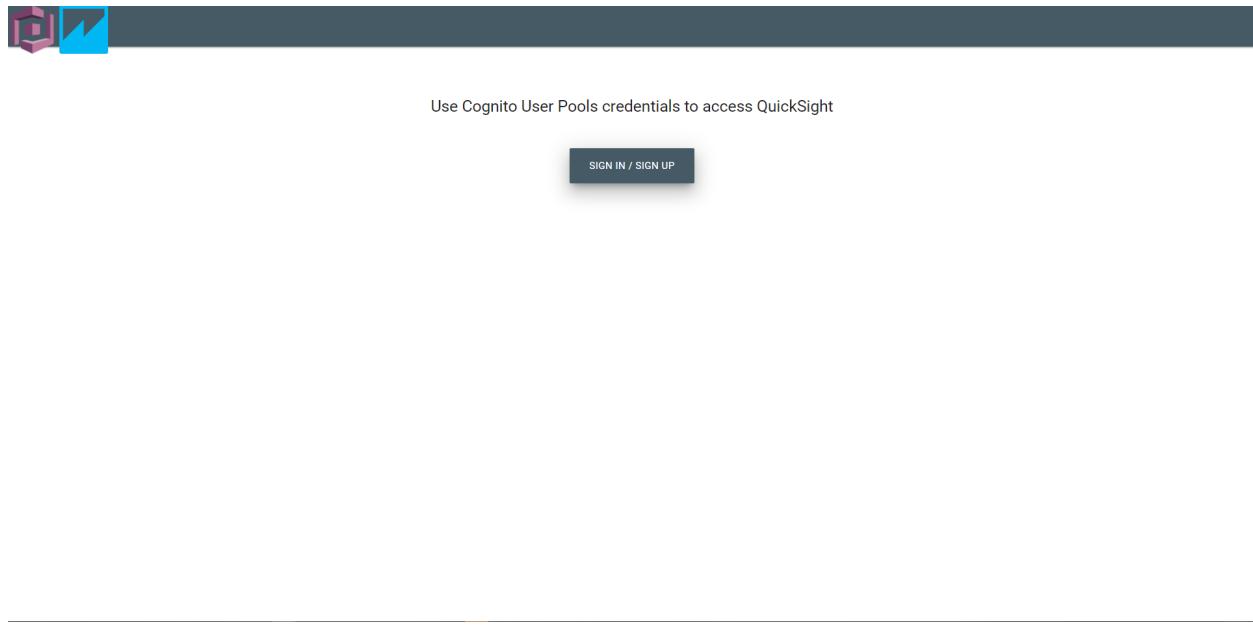
User will click on the view dashboard button, which will move to the login/register page.



Cloud Computing Assignment 3 Report

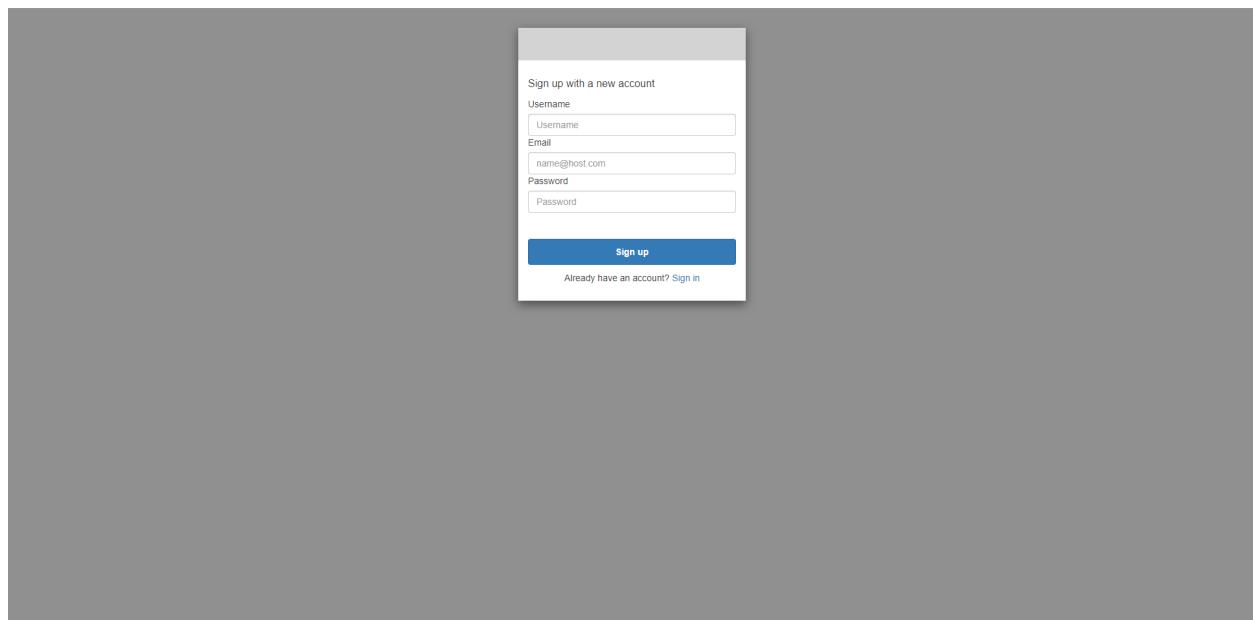
• • •

Click on the signin/signup to proceed with the login.



Register:

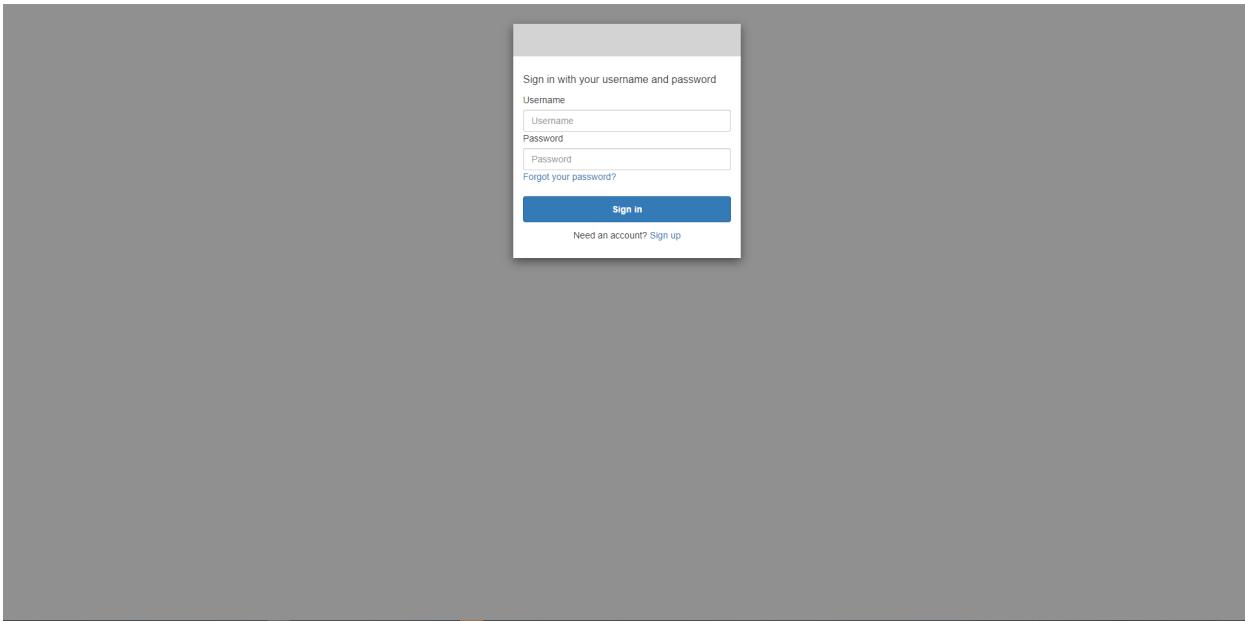
User can register into the app by providing the details available in the screenshot.



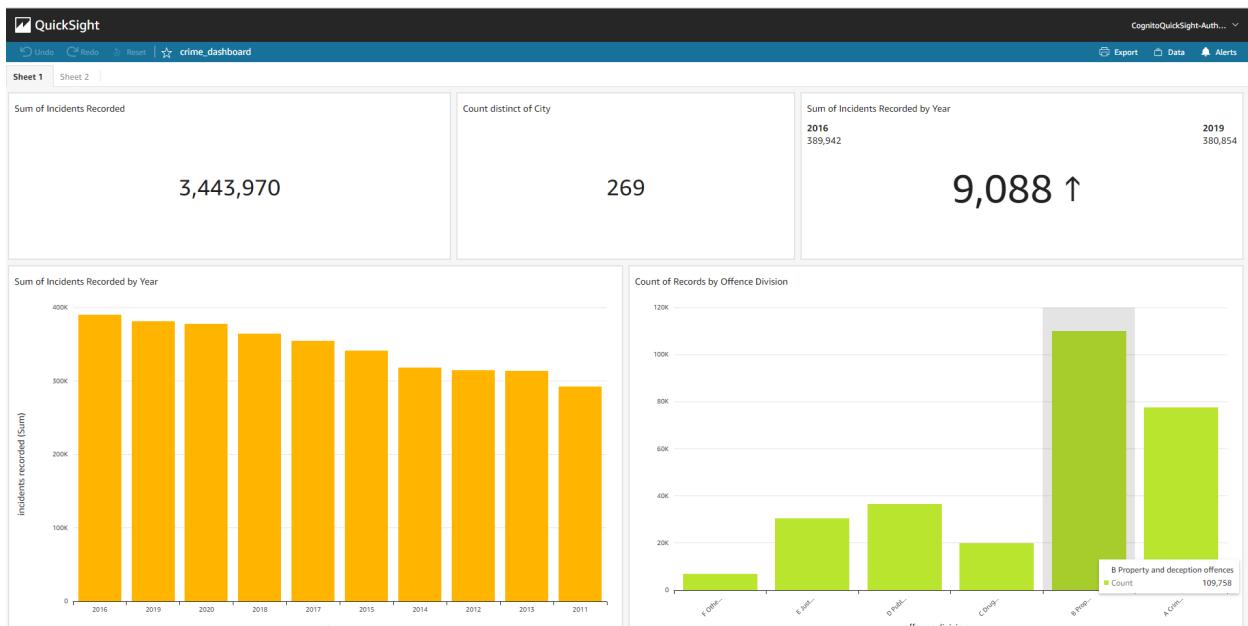
Cloud Computing Assignment 3 Report

• • •

Already a registered user, he/she can login directly with the credentials.



After login user will be moved to the quicksight dashboard.





References:

Dataset:

[1]. Data VIC, *Criminal incident by LGA - Year ending Dec 2020*, Data VIC, viewed 13 July 2021, <<https://discover.data.vic.gov.au/dataset/crime-by-location-data-table/resource/c9b3c3e3-4a71-4546-989b-6ff96ae6ad97>>

[2]. Simple Maps, *Australia Cities Database*, Data VIC, viewed 13 July 2021, <<https://simplemaps.com/data/au-cities>>

AWS Documentation:

[3]. AWS, *How can I make a private Amazon Redshift cluster publicly accessible?*, AWS, viewed 13 July 2021, <<https://aws.amazon.com/premiumsupport/knowledge-center/redshift-cluster-private-public/>>