

INFO 7370 38417 Designing Data Architecture Business Intelligence

AWS

Srishti Ashok Mishra

NU ID : 001305178

JOB:

The screenshot displays the AWS CloudFormation console for a job named 'aws_athena_job'. The job is shown in the 'Completed' state. The execution time is 23:42:17 to 23:44:17 on 17/02/2020. The output shows the job ended with exit code 0.

The job consists of four tasks:

- tS3Connection_1
- tS3Get_1
- tS3BucketCreate_1
- tS3Put_1

The job is connected to the AWS S3 service. The output shows the job ended with exit code 0.

The screenshot also shows the 'Basic settings' for the task 'tS3Connection_1'. The settings include:

- Access Key: AKIAZ6OXRKZSP5FOYJ70
- Secret Key: *****
- Inherit credentials from AWS role: ☐
- Assume Role: ☐
- Region and Endpoint: DEFAULT
- Client-side Encrypt: ☐

Using Professor's Access and Secret key and creating a bucket:

The screenshot shows the AWS Step Functions Designer interface. The workflow consists of four components connected sequentially: **tS3Connection_1**, **tS3Get_1**, **tS3BucketCreate_1**, and **tS3Put_1**. Each component is connected to the next via an **OnComponentOk** event. A **Command Prompt** window is open over the workflow, displaying the following text:

```
Microsoft Windows [Version 10.0.18363.628]
(c) 2019 Microsoft Corporation. All rights reserved.
C:\Users\srish>
```

The **ts3Get_1** component is selected, and its **Basic settings** are visible in the left pane:

- Connection:** ☒ Use an existing connection. Component List: **tS3Connection_1**
- Bucket:** **"info7370"**
- Key:** **"vehicle_accidents/Vehicle_Accidents_iowa.csv"**
- File:** **"C:/Users/srish/Documents/data/aws_file_new.csv"**
- Die on error:** ☐

Putting Data from host machine to my AWS S3 bucket :

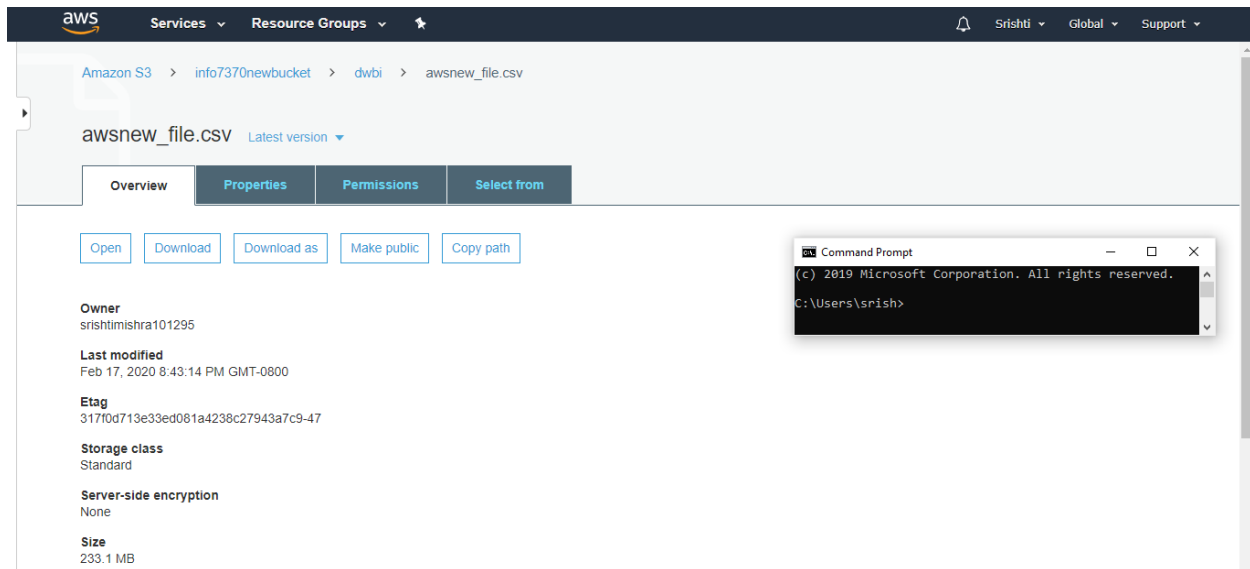
The screenshot shows the AWS Step Functions Designer interface. The workflow consists of four components connected sequentially: **tS3Connection_1**, **tS3Get_1**, **tS3BucketCreate_1**, and **tS3Put_1**. Each component is connected to the next via an **OnComponentOk** event. A **Command Prompt** window is open over the workflow, displaying the following text:

```
Microsoft Windows [Version 10.0.18363.628]
(c) 2019 Microsoft Corporation. All rights reserved.
C:\Users\srish>
```

The **ts3Put_1** component is selected, and its **Basic settings** are visible in the left pane:

- Access Key:** **"AKIAICXQOQ7JISTVJDPA"**
- Secret Key:** **"*****"**
- Inherit credentials from AWS role:** ☐
- Assume Role:** ☐
- Region:** **DEFAULT**
- Client-side Encryption:** ☐
- Bucket:** **"info7370newbucket"**
- Key:** **"dwbi/awsnew_file.csv"**
- File or Stream:** **"C:/Users/srish/Documents/data/aws_file_new.csv"**
- Server-Side Encryption:** ☐
- Die on error:** ☐

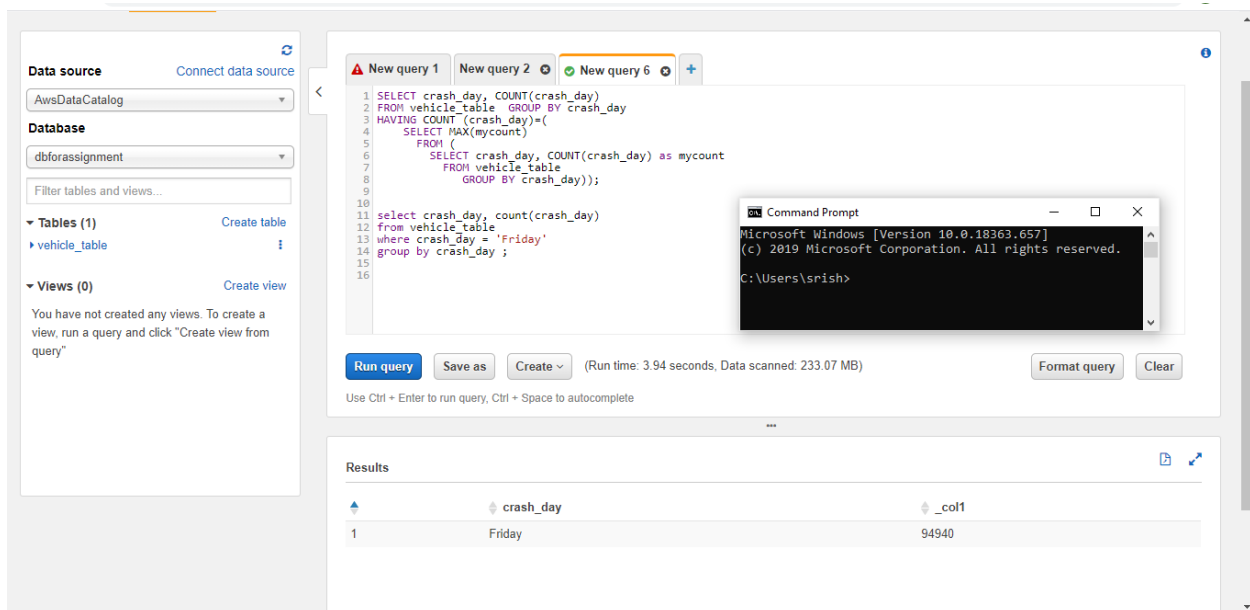
Data in AWS bucket



The screenshot shows the AWS S3 console interface. The breadcrumb navigation indicates the path: Amazon S3 > info7370newbucket > dwbi > awsnew_file.csv. The file 'awsnew_file.csv' is selected, showing its latest version. Below the file name are tabs for Overview, Properties, Permissions, and Select from. The Overview tab is active, displaying metadata: Owner (srishtimishra101295), Last modified (Feb 17, 2020 8:43:14 PM GMT-0800), Etag (317f0d713e33ed081a4238c27943a7c9-47), Storage class (Standard), Server-side encryption (None), and Size (233.1 MB). Action buttons include Open, Download, Download as, Make public, and Copy path. A Command Prompt window is overlaid on the right, showing the command 'C:\Users\srish>'.

SQL:

To find the day on which the maximum crashes happened



The screenshot shows the AWS Glue console interface. On the left, the 'Data source' section shows 'AwsDataCatalog' as the data source and 'dbforassignment' as the database. The 'Tables (1)' section lists 'vehicle_table'. The 'Views (0)' section is empty. The main area shows a SQL query being executed. The query is:

```
1 SELECT crash_day, COUNT(crash_day)
2 FROM vehicle_table GROUP BY crash_day
3 HAVING COUNT(crash_day)=(
4     SELECT MAX(mycount)
5     FROM (
6         SELECT crash_day, COUNT(crash_day) as mycount
7         FROM vehicle_table
8         GROUP BY crash_day));
9
10
11 select crash_day, count(crash_day)
12 from vehicle_table
13 where crash_day = 'Friday'
14 group by crash_day ;
15
16
```

The query is executed, and the results are displayed in a table with two columns: 'crash_day' and '_col1'. The result shows 'Friday' with a count of 94940. A Command Prompt window is overlaid on the right, showing the command 'C:\Users\srish>'.