**Document How to use**

P**roblem:** user and decision maker gets the updated data after 24 hours and by that time it’s too late make decision for important updated data which might be risky for airlines /aircraft / engine or business financial health.

**Solution:** this architecture is designed/developed to get the updated data within a given interval like 30 minute or an hour for each flight flying in the sky or grounded at the airport.

Step by Step Process Flow:

1)- API- lambda runs for real time flight, cities, airplanes and aircraft types and get data from Aviation stack .

2)- data get pushed to respective api stream and each stream has only 1 shard.

3)- once data pushed to stream then it gets consumed by kinesis firehose as per stream.

4)-kinesis has 300 seconds buffer time and data gets stored in S3

5)-once data loaded in s3 then we run glue spark every hour for flight data and once in day for Citi, aircraft and airplane data

6)- lambda and glue runs through cloud watch rules

7)- table produced by glue job can be used in Athena and quickshight for analytics.

**Deployment\_Lambda.zip:** this zip file contains all the lambda python scripts and required python libraries to execute the lambda function in aws console. This is uploaded in s3 under bucket **kinesis-producer-lambdas** below are the list of python script used in lambda.

**Name of the python scripts:**

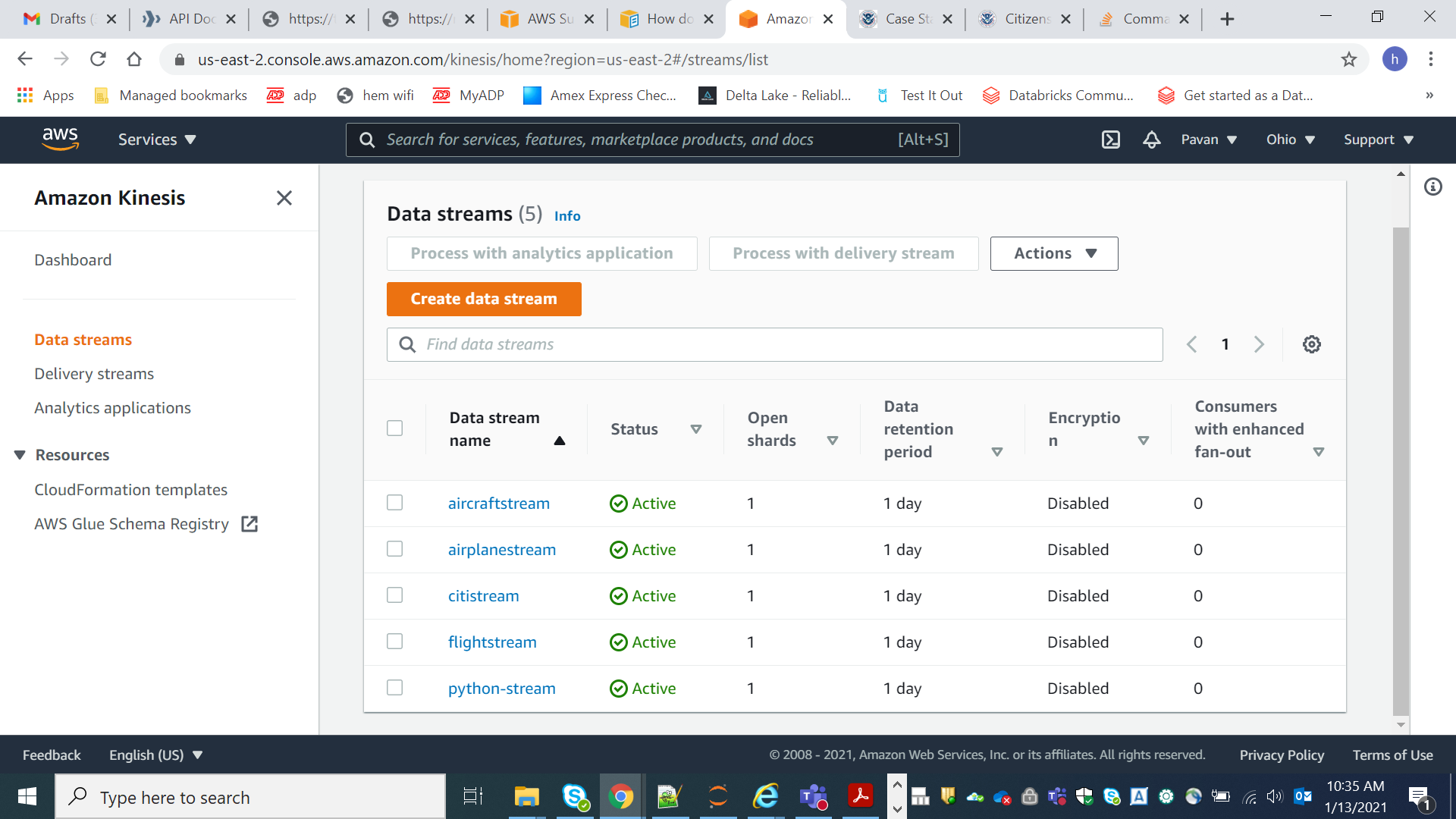
**Aircraftstream.py:** this is being used in aircarftstreamlambda lambda function which will push aircraft data into the stream.

**Airplanestream.py:** this is being used in airplanestreamlambda lambda function, which will push airplane data into the stream

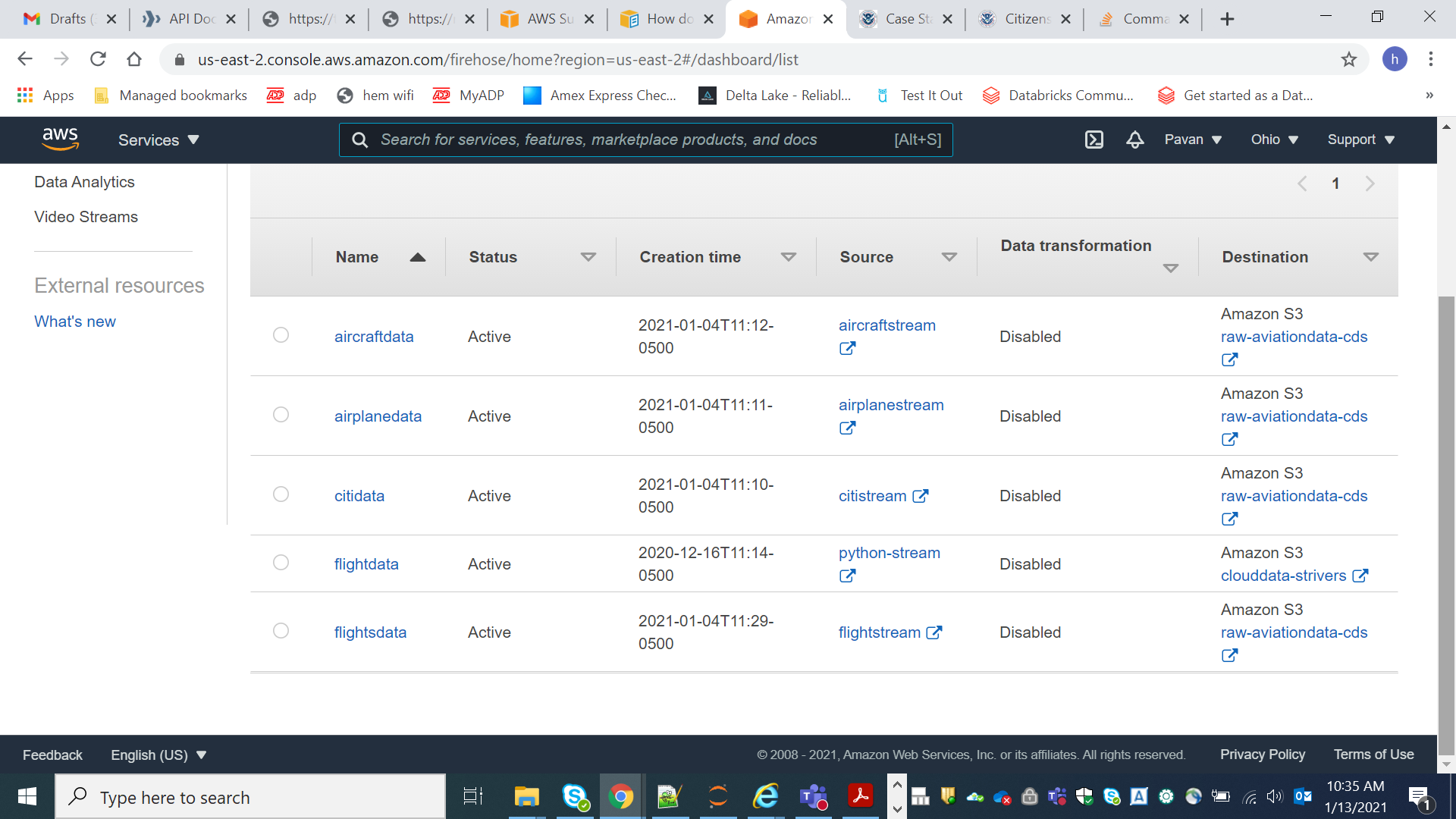
**Citistream.py:** this is being used in citistreamlambda lamda function, which will push city data into the stream

**Flightstream.py:** this is being used in flightstreamlambda lambda function, which will push flight data into the stream

**List of data stream:**

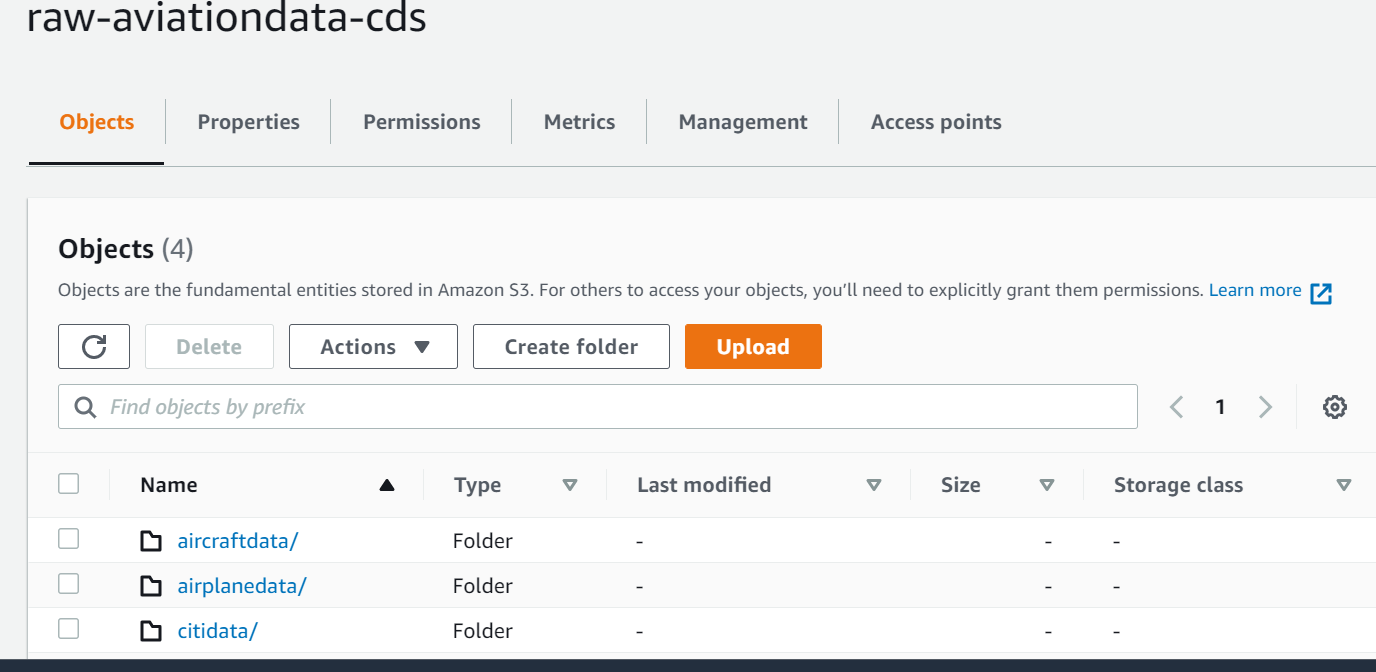


**List of delivery stream:**



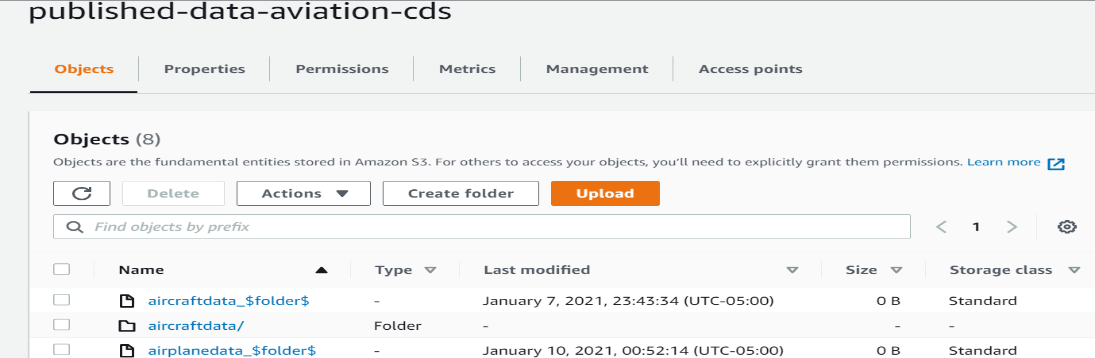
**S3 bucket which stores the raw data loaded by the stream:**

Raw-aviation data-cds bucket contain the respective folder for each api data as seen the the screenshot. Later it is being used by glue to transform.

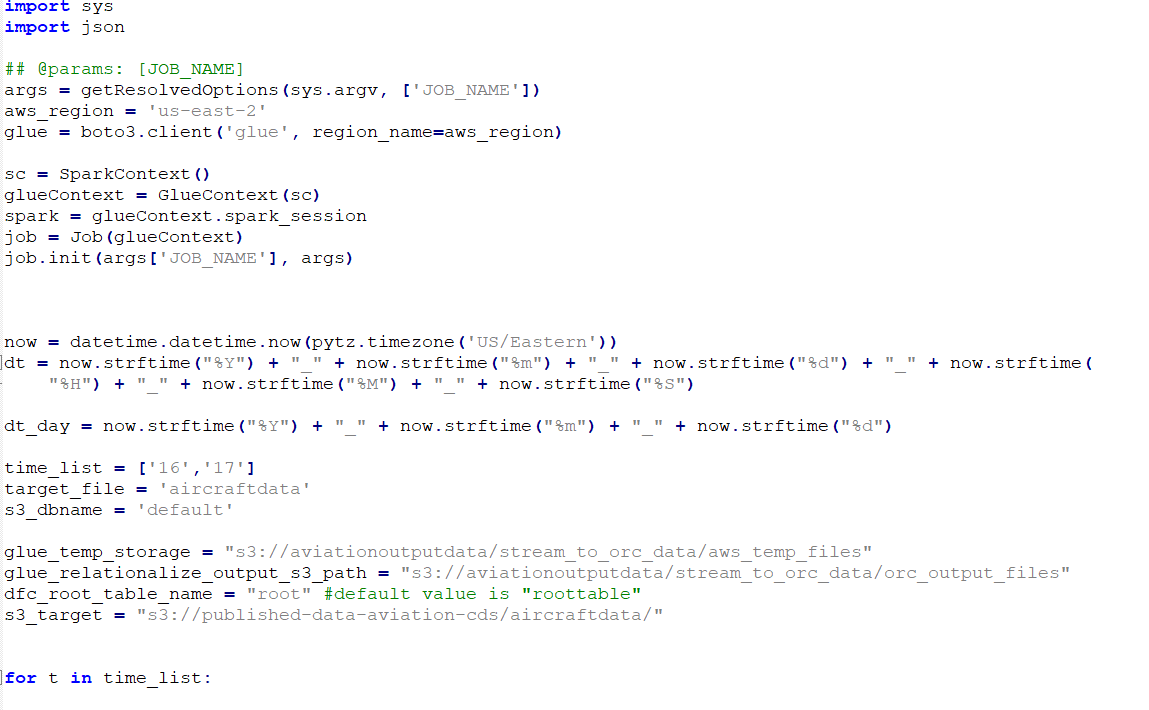
****

**Glue Script name:** aviation\_data\_load.py runs to process raw s3 data and stores transform data in s3 under bucket **published-data-aviation-cds .**

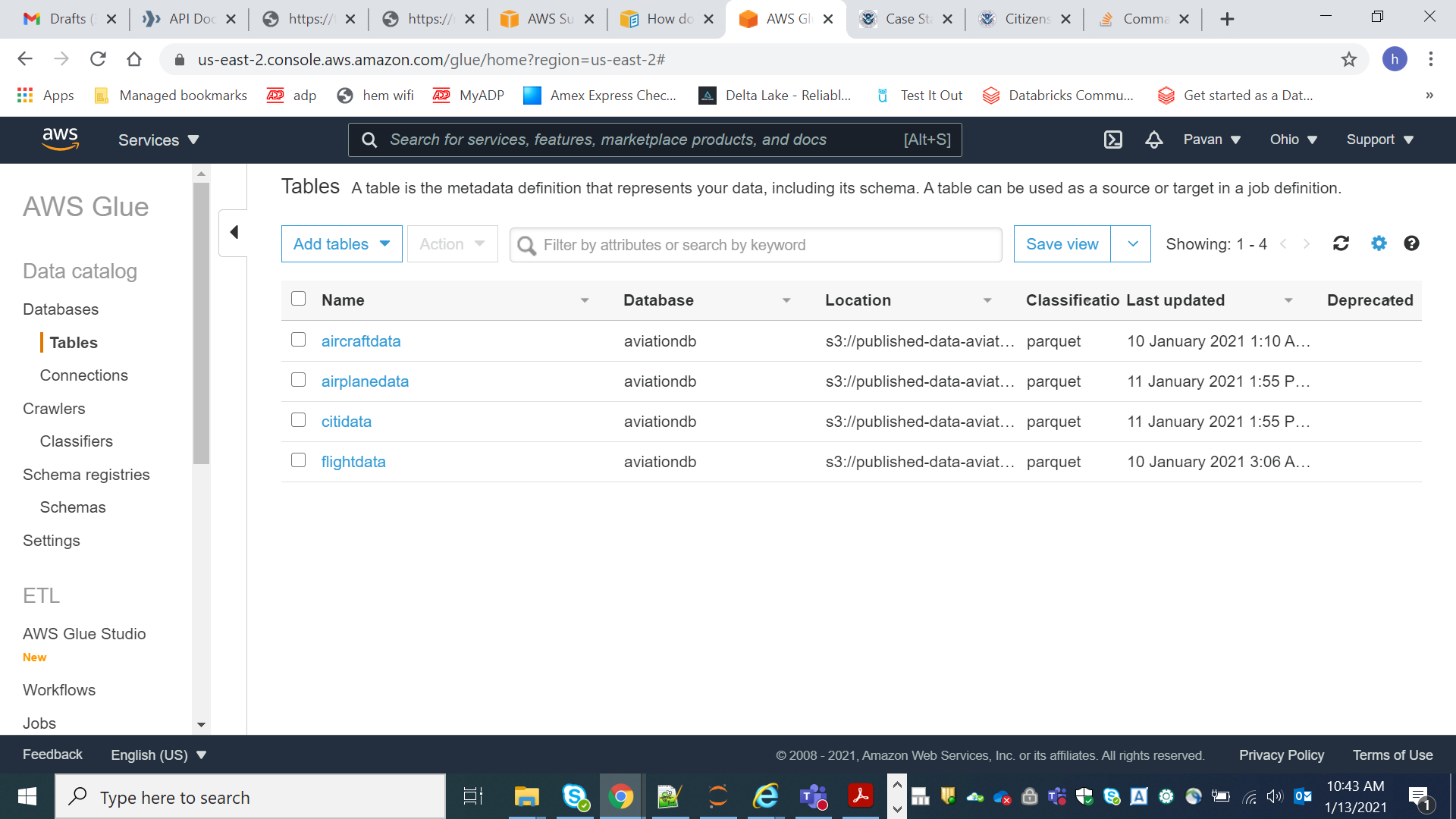
**S3 output bucket:**

****

**aviation\_data\_load.py example:**

****

**Table created by glue job:** table created by glue job can be used in Athena by selection the **aviation dB** database.



**Analytics**: some analytics run top of the data we processed

