A close up of a logo

Description automatically generated

**ALY 6110: Data Management and Big Data**

**CRN:** 81984

**Week 6 Final Assignment:** Airline Market Analysis.

**Submitted By:  Group Beta**

**Ashwini Kumar Pathak [**pathak.as@husky.neu.edu**]**

**Ayush Jain [**jain.ayush@husky.neu.edu**]**

**Megha Ravi [**ravi.m@husky.neu.edu**]**

**Ashishkumar Bidap [**[bidap.a@husky.neu.edu](mailto:bidap.a@husky.neu.edu)**]**

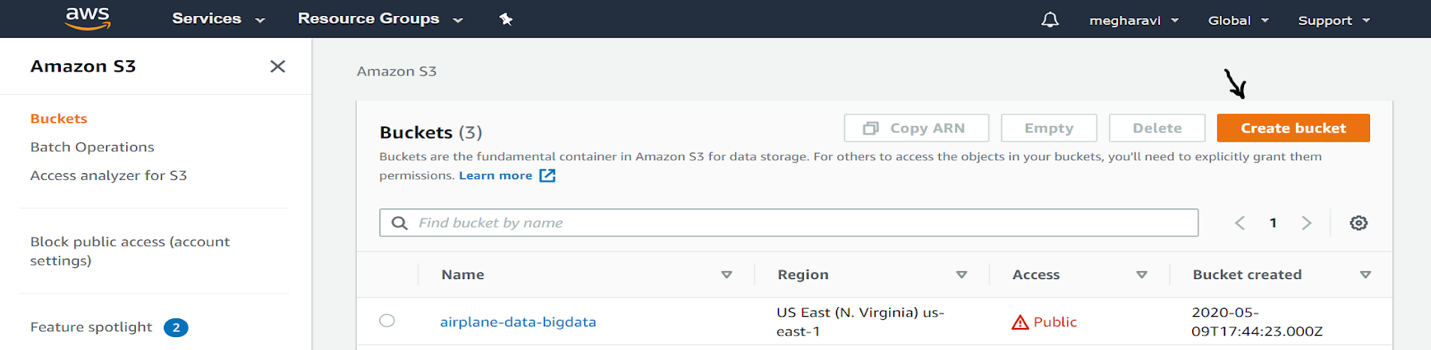
**Working Demo**

Below is the step by step setup of the Airline Market analysis.

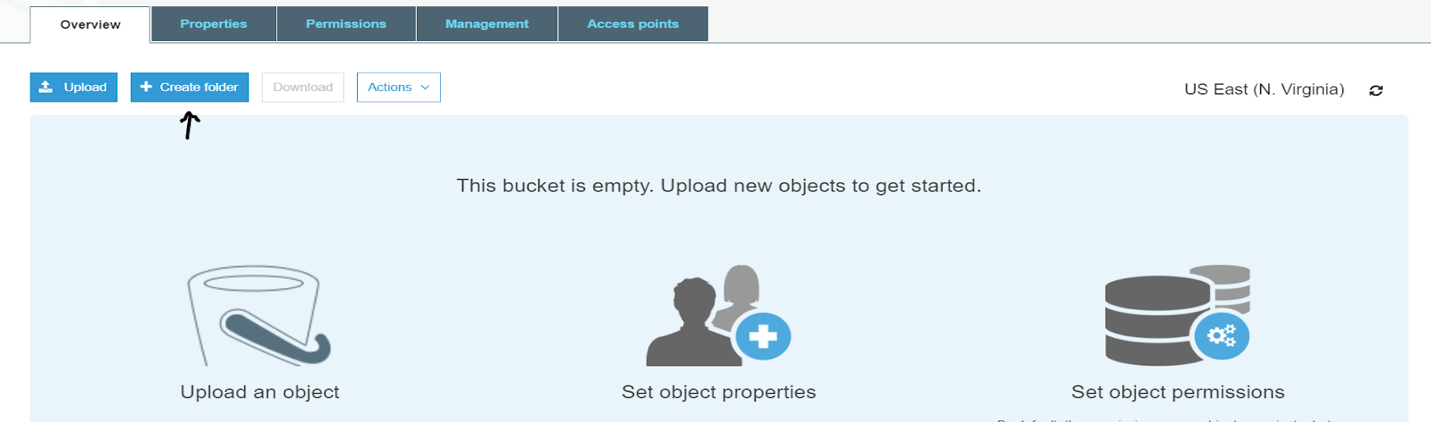
**Step 1** : Download the datasets from the Bureau of Transportation Statistics for the required months and years using the below link. <https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236> .

**Step 2**:Data Storage on the Amazon S3.

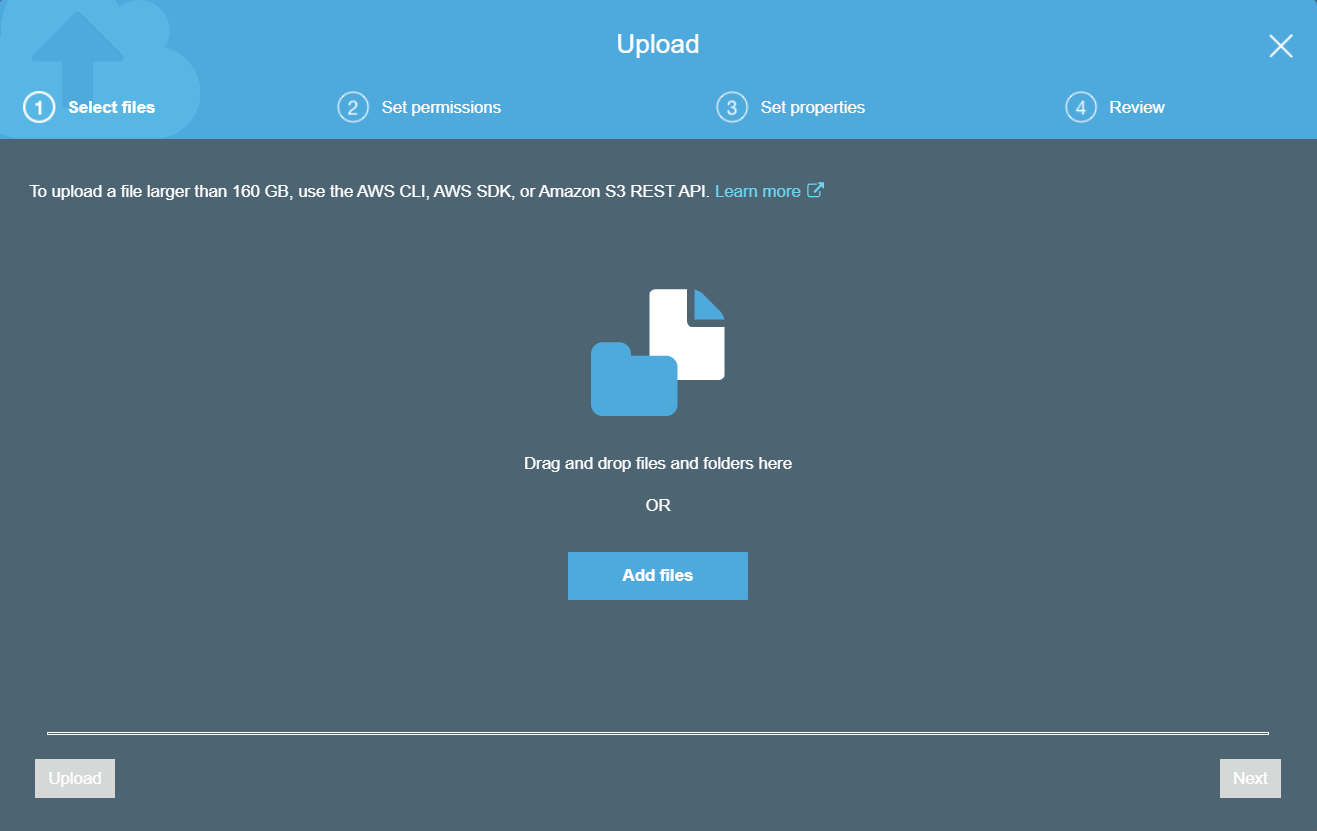
1. Create an account of amazon web services.
2. Once logged in traverse through services on top left corner and search for amazon S3.
3. Create a bucket named “airplane-data-bigdata” and created a folder named “airplane\_data”.



1. Within the “airplane\_data” folder, select the csv files and uploaded them all in one go should be through.



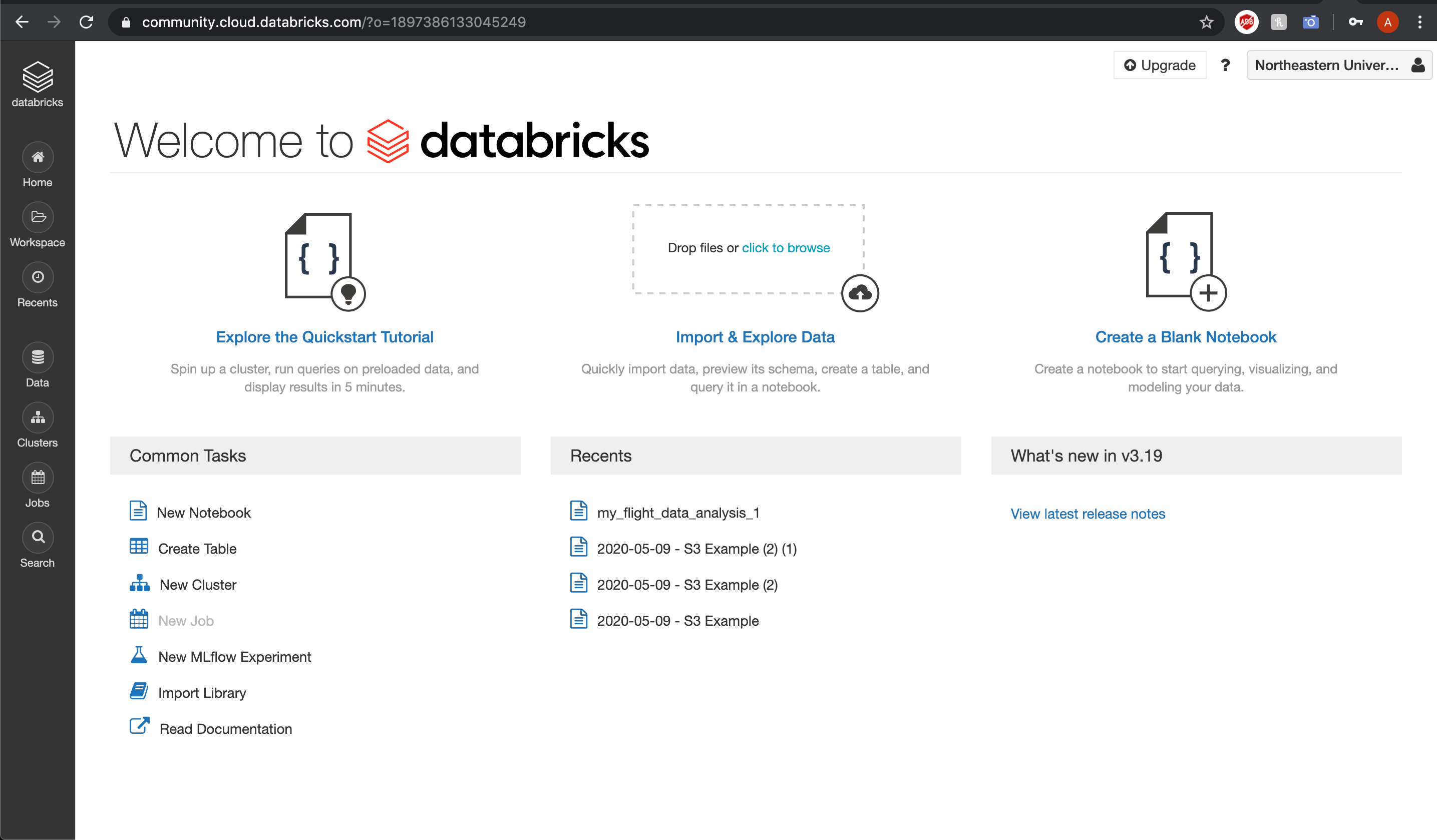
1. While uploading, S3 allows us to set permission – Making the dataset public, allow certain user access or group access, give access to external amazon accounts appropriate assign the permissions.



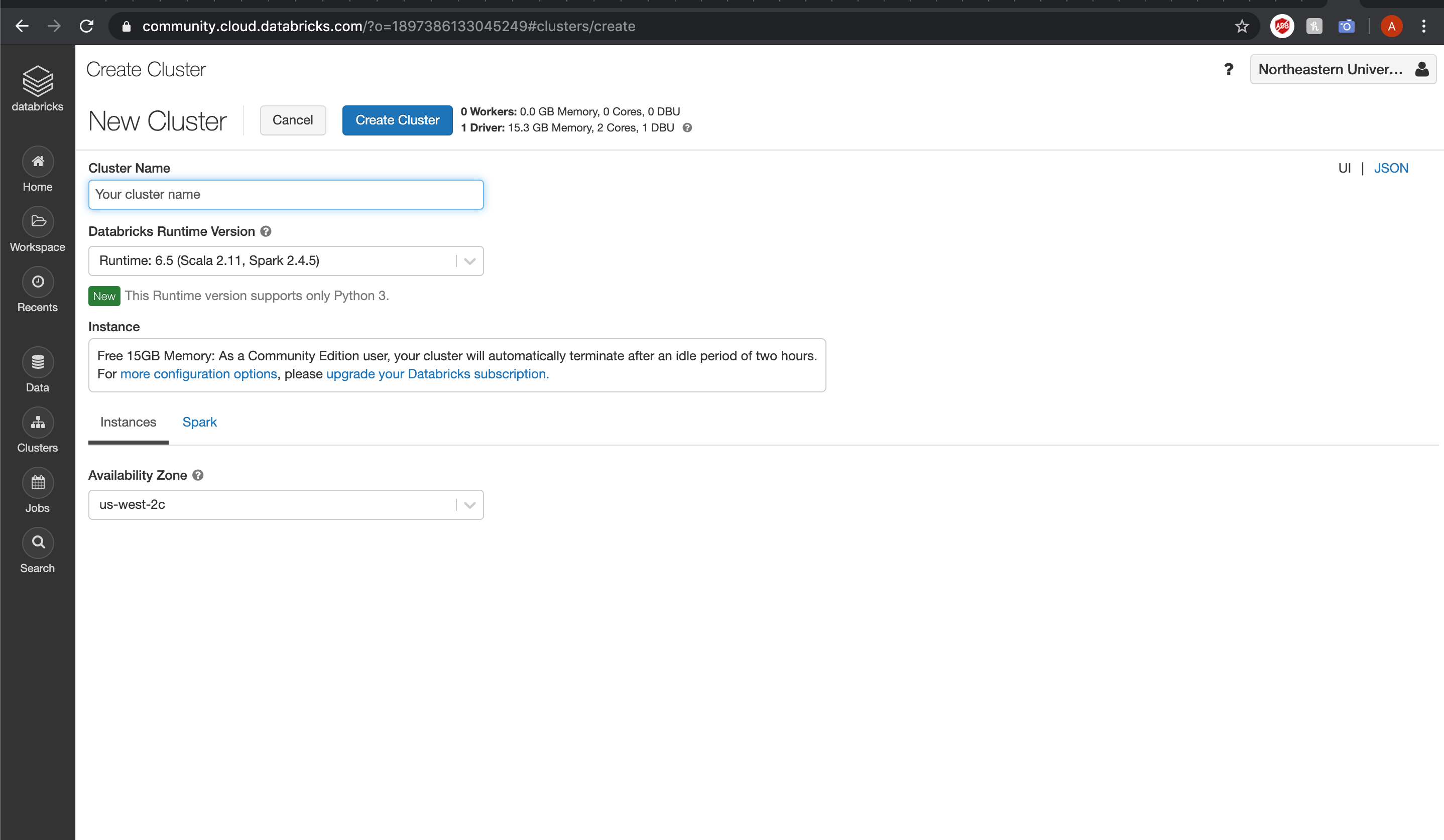
1. With this data storage is ready also note down the access key and the secure key assigned to your account which we would be needing in connecting from Databricks.

**Step 3**: Create an account with Databricks community version using <https://community.cloud.databricks.com/login.html>

**Steps 4**:Once logged in you should view something like this.

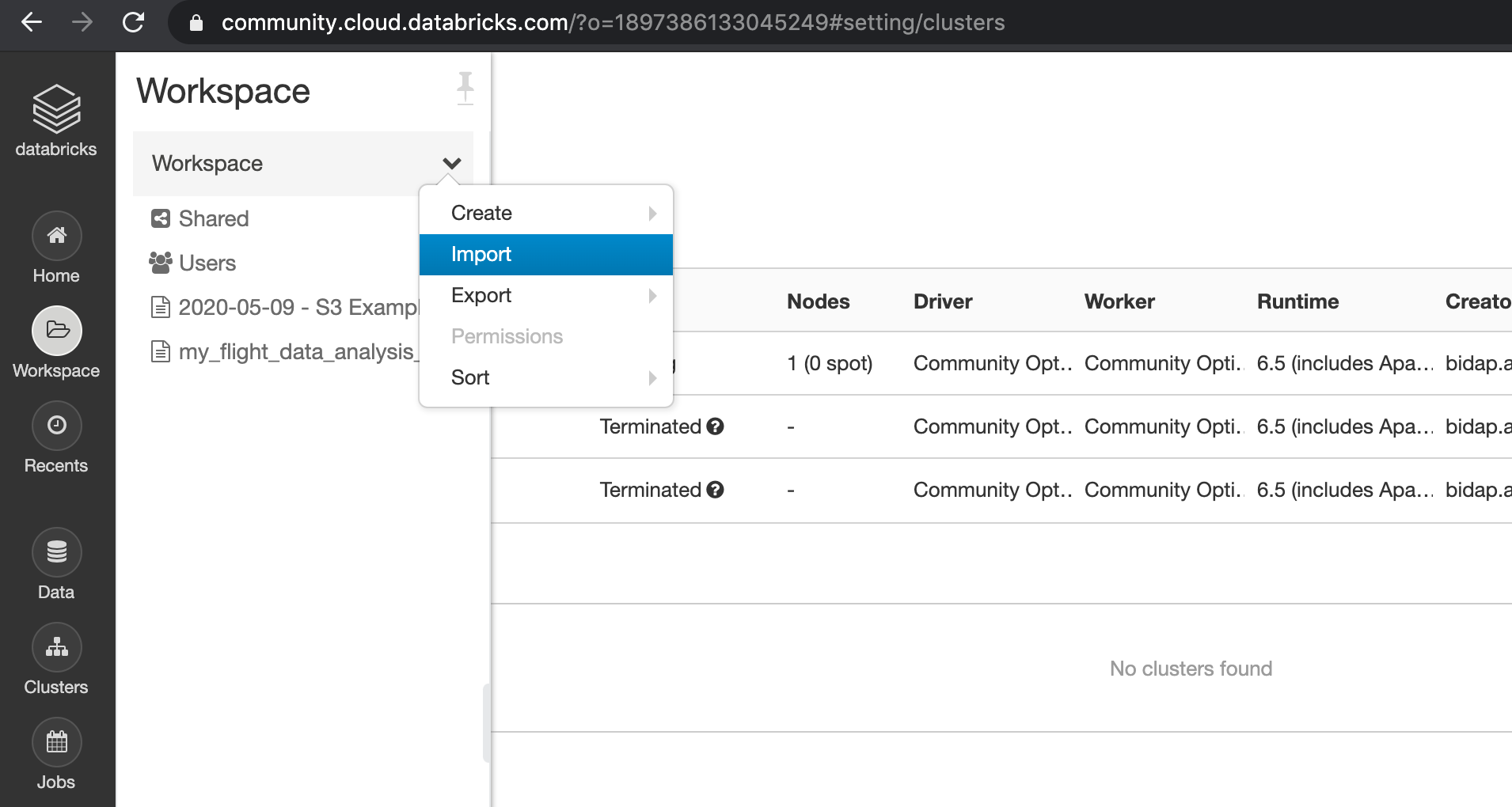


**Step 5:** Traverse through the left pane and look for Clusters and create a new cluster with below shown configuration.



**Step 6:** Wait for the cluster to come into running state.

**Step 7:** Attach the Notebook file of our implementation which is being provided along through the workspace import option.



**Step 8:** Kindly change the access key and security key in the imported notebook file and ignore the first unmount command as no file is mounted as of now.

**Step 9:** Execute the cells and all the steps should be through correctly.