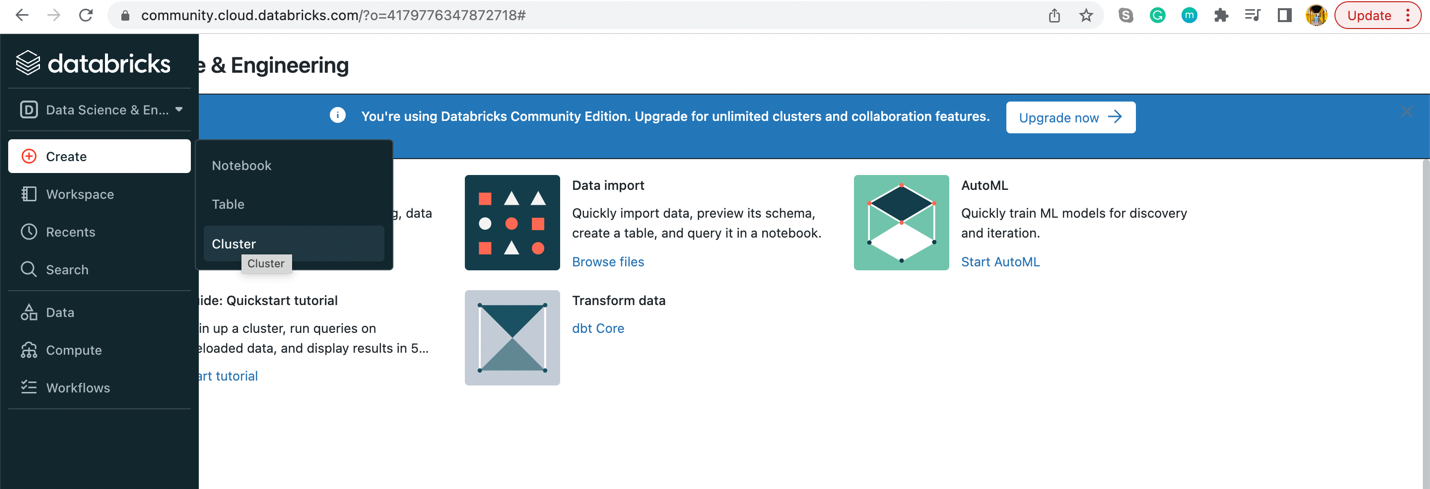
DSCI 5350 SparkSQL Activity-1 Satyasriharsha Buddha

1. Create a cluster



2. Enter the name of your cluster, select DataBricks run time version and click

create cluster

A screenshot of a computer

Description automatically generated

3. Click Event log and verify the cluster status

A screenshot of a computer

Description automatically generated with medium confidence

4. Create new Scala notebook

A screenshot of a computer

Description automatically generated

5. Create an instance of Hive Context:

A screenshot of a computer

Description automatically generated with low confidence

6.Create a hive table person1 with first\_name, last\_name,and age as columns

A screenshot of a computer

Description automatically generated with medium confidence

7.Upload the sample.txt file. Click in the left pane and navigate to create table

or you can navigate to File upload in the Scala notebook screen to upload the

file into tables folder

Note: I did not find the sample.txt.file in the modules,As the file is fairly small, I created a Person1.Csv file with same table contents.

A screen shot of a phone

Description automatically generated with low confidence

A screenshot of a computer

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generated

8. Load the data into person1 table

A screenshot of a computer

Description automatically generated with medium confidence

9. Select the person1 data using HiveQL

A screenshot of a computer

Description automatically generated with medium confidence

10. Create a new table from the output of select query

A screenshot of a computer

Description automatically generated with medium confidence