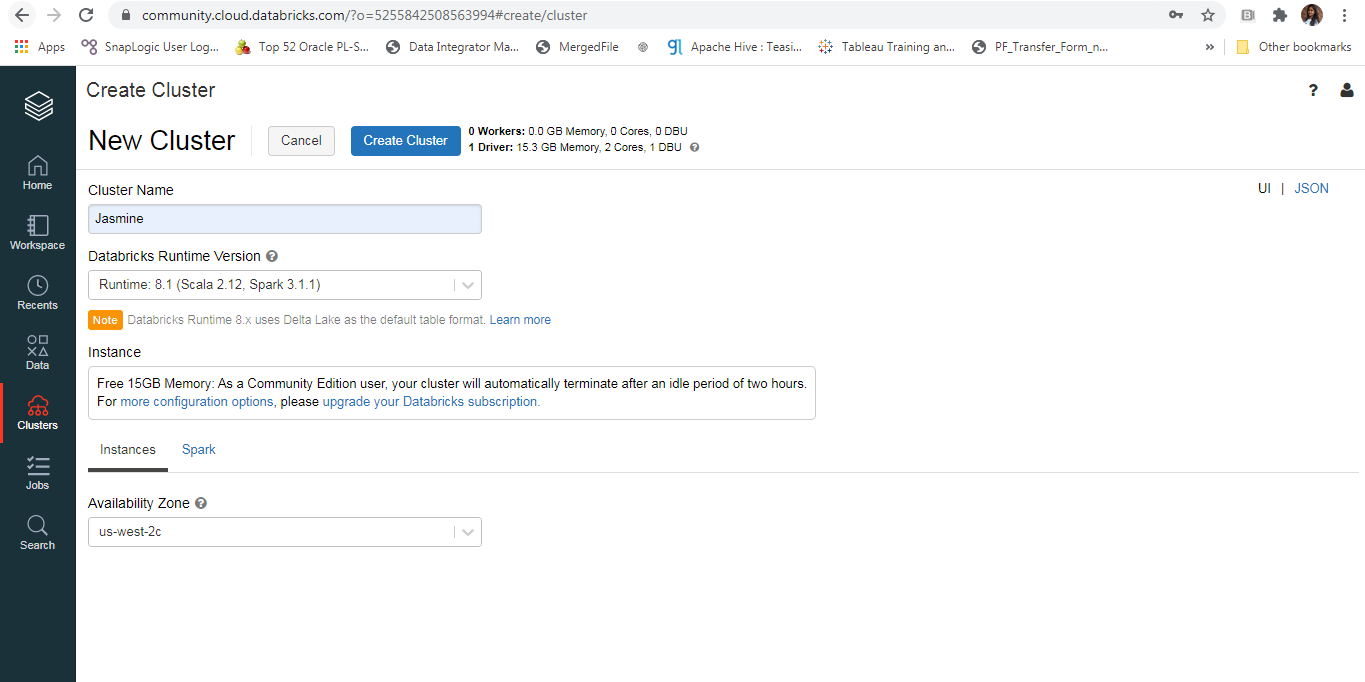
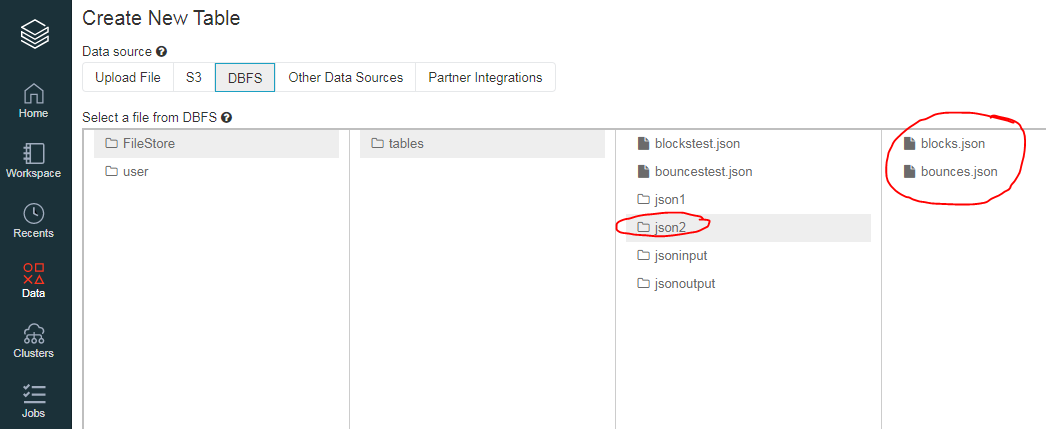
**Steps involved in loading the data through data bricks cluster**

1. Cluster ***‘Jasmine’*** is created in databricks through **Databricks community edition** provided by Databricks.



1. The Cluster is created when given above details.
2. We have workspace to work with our requirement
3. When cluster is up and running folder is created using below command in the workspace

***dbutils.fs.mkdirs("/FileStore/tables/json2")***

In File store the folder is created as json2 as shown below

1. In this folder json2, blocks and bounces json files that needs to be converted to single csv are uploaded.

**Merging of two Json files:**

1. ***The two json files needs to be merged and converted to csv, and also the column ‘source’ is to be flagged to know from which json(bounces/blocks), it is generated***.
2. Read each json file to dataframe and added the required column to the dataset with code below

***from pyspark.sql.functions import lit***

***filepath="/FileStore/tables/json2/blocks.json"***

***data=spark.read.option("multiline",True).json(filepath)***

***df1=data.withColumn("source", lit('blocks'))***

***df1.display()***

***Note: Here lit is a function used to take constant value as input when creating a new column***

1. Similarly the second json file is also read into json and column source is added with the value

***filepath="/FileStore/tables/json2/bounces.json"***

***data=spark.read.option("multiline",True).json(filepath)***

***df2=data.withColumn("source", lit('bounces'))***

***df2.display()***

1. As the two json files are read into data frames then the two data frames can be merged using the below code.

***df = df1.union(df2)***

**To CSV:**

1. The merged data frame is then read to csv file through overwrite mode as used in below code

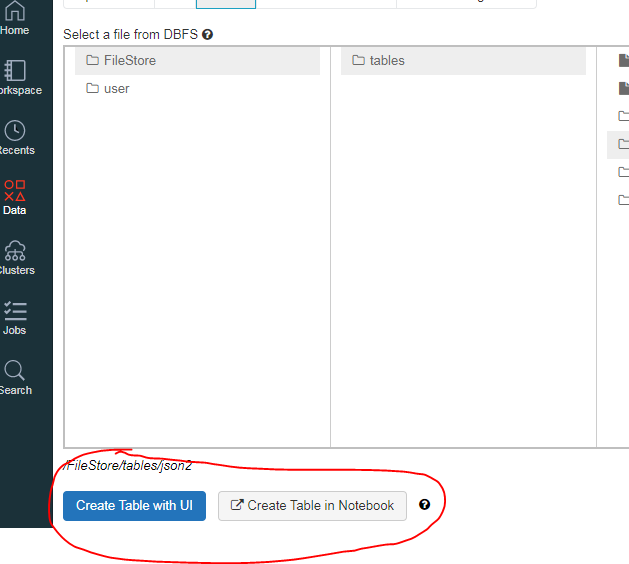
***output="/FileStore/tables/jsonoutput/output.csv"***

***df.write.mode("overwrite").csv(output,sep=',',header=True)***

***display(df)***

**To Table:**

1. The csv file data which is uploaded to DBFS is then loaded into a table using UI or through notebook as given in the databricks cluster.



1. Created table is stored in */user/hive/warehouse/bounce\_block* this path.

Two tables temporary and permanent are created.

