

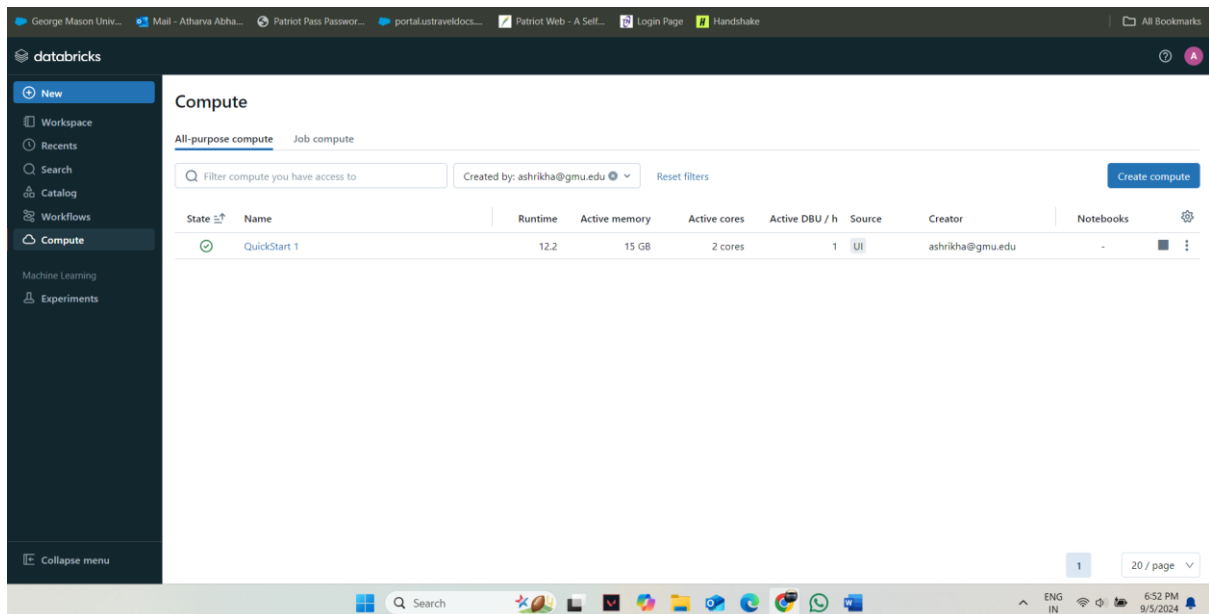
**LAB NO. 1**

**AIT-614-002- FALL-2024**

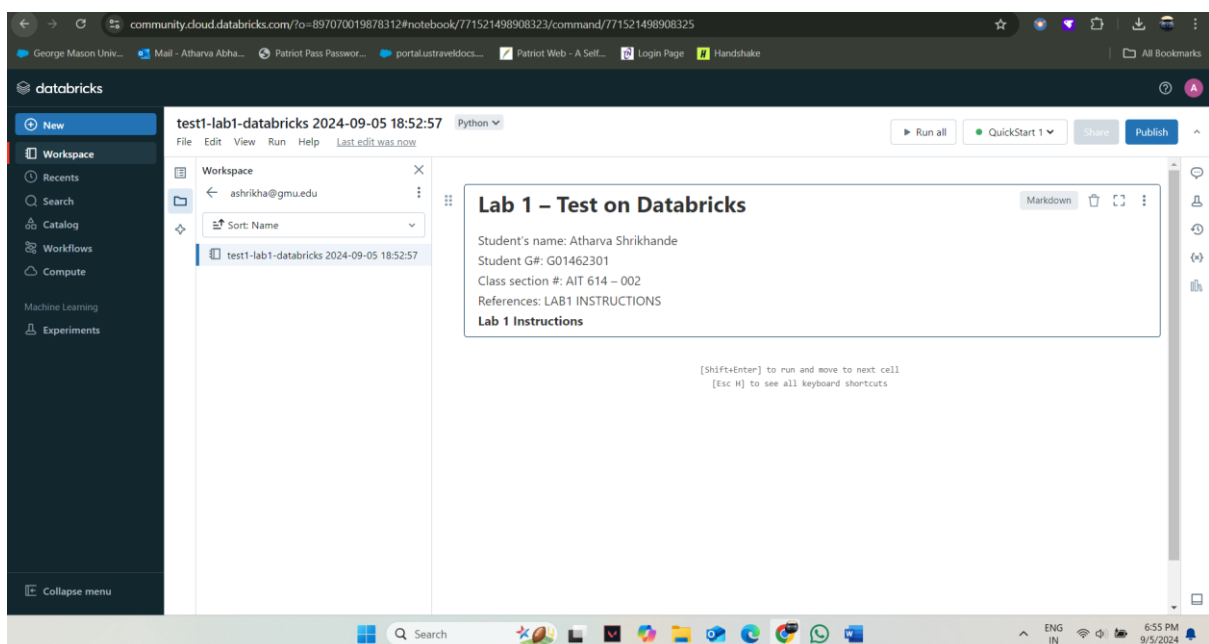
**G01462301**

**ATHARVA ABHAY SHRIKHANDE**

1. (50 points) Save a screenshot of the created Cluster information. It should include your account name.



2. (25 points) Save a screenshot of the created Notebook information. It should include your name, G#, and class section #.



Diamond –

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

default.diamonds

QuickStart 1

Details

History

Description:  
Created at: 2024-09-05 22:34:22  
Last modified: 2024-09-05 22:57:45  
Partition columns:  
Number of files: 1  
Size: 732 kB

Schema:

	col_name	data_type	comment
1	_c0	int	null
2	carat	double	null
3	cut	string	null
4	color	string	null
5	clarity	string	null
6	depth	double	null
7	table	double	null
8	price	int	null
9	x	double	null
10	y	double	null

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

Data

Create Table

Databases

Filter Databases

default

Filter Tables

diamonds

Run all

QuickStart 1

Share

Publish

8

cut	color	clarity	depth	table	price	x	y	z
Very Good	E	SI2	61.5	55	326	3.95	3.98	
Premium	E	SI1	59.8	61	326	3.89	3.84	
Good	E	VSI	56.9	65	327	4.05	4.07	
Premium	I	VSI	62.4	58	334	4.2	4.23	
Good	J	SI2	63.3	58	335	4.34	4.35	
Very Good	J	VVS2	62.8	57	336	3.94	3.96	
Very Good	I	VVS1	62.3	57	336	3.95	3.98	
Very Good	H	SI1	61.9	55	337	4.07	4.11	
Fair	E	VSI	65.1	61	337	3.87	3.78	
Very Good	H	VSI	59.4	61	338	4	4.05	
Good	J	SI1	64	55	339	4.25	4.28	
Fair	J	VSI	62.8	56	340	3.93	3.9	
Premium	F	SI1	60.4	61	342	3.88	3.84	
Fair	J	SI2	62.2	54	344	4.35	4.37	

data due to row limit | 10.73 seconds runtime

Refreshed 4 minutes ago

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

Quickstart Notebook

SQL

File Edit View Run Help

10,000+ rows | Truncated data due to row limit | 4.33 seconds runtime

Refreshed 6 minutes ago

06:56 PM (44s)

6

%python  
diamonds = spark.read.csv("/databricks-datasets/Rdatasets/data-001/csv/ggplot2/diamonds.csv", header="true", inferSchema="true")  
diamonds.write.format("delta").mode("overwrite").save("/delta/diamonds")

(12) Spark Jobs

diamonds: pyspark.sql.dataframe.DataFrame = [c0: integer, carat: double ... 9 more fields]

06:56 PM (4s)

7

DROP TABLE IF EXISTS diamonds;  
  
CREATE TABLE diamonds USING DELTA LOCATION '/delta/diamonds/'

OK

06:56 PM (11s)

8

SELECT \* from diamonds

(2) Spark Jobs

Table +

Collapse menu

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

Quickstart Notebook

SQL

File Edit View Run Help

Run all QuickStart 1 Share Publish

06:56 PM (11s)

8

SELECT \* from diamonds

(2) Spark Jobs

Table +

	c0	1.2 carat	1.2 cut	1.2 color	1.2 clarity	1.2 depth	1.2 table	1.2 price	1.2 x	1.2 y	1.2 z
1	1	0.23	Ideal	E	SI2	61.5	55	326	3.95	3.98	
2	2	0.21	Premium	E	SI1	59.8	61	326	3.89	3.84	
3	3	0.23	Good	E	VS1	56.9	65	327	4.05	4.07	
4	4	0.29	Premium	I	VS2	62.4	58	334	4.2	4.23	
5	5	0.31	Good	J	SI2	63.3	58	335	4.34	4.35	
6	6	0.24	Very Good	J	VVS2	62.8	57	336	3.94	3.96	
7	7	0.24	Very Good	I	VVS1	62.3	57	336	3.95	3.98	
8	8	0.26	Very Good	H	SI1	61.9	55	337	4.07	4.11	
9	9	0.22	Fair	E	VS2	65.1	61	337	3.87	3.78	
10	10	0.23	Very Good	H	VS1	59.4	61	338	4	4.05	
11	11	0.3	Good	J	SI1	64	55	339	4.25	4.28	
12	12	0.23	Ideal	J	VS1	62.8	56	340	3.93	3.9	
13	13	0.22	Premium	F	SI1	60.4	61	342	3.88	3.84	
14	14	0.31	Ideal	J	SI2	62.2	54	344	4.35	4.37	

Collapse menu

3. (25 points) Save a screenshot of the run tutorial notebook and the created table by the tutorial.

The screenshot shows a Databricks notebook interface. The top bar includes the user name 'Aitharva Abha...' and several open tabs: 'Patriot Pass Passwor...', 'portalustraveldocs...', 'Patriot Web - A Self...', 'Login Page', and 'Handshake'. The notebook's left sidebar shows a 'Data' tab. The main area displays a table with 10 columns: 'id', 'name', 'age', 'gender', 'height', 'weight', 'eye\_color', 'hair\_color', 'skin\_color', and 'blood\_type'. The table contains 10 rows of data. Below the table, there is a code cell with the following text: 'The next command manipulates the data and displays the results'. The bottom of the screen shows a Windows taskbar with a search bar, various application icons, and system tray icons including the language 'ENG IN', network status, and the date '9/5/2024'.

The screenshot shows a Databricks notebook interface. The top bar includes the user name 'George Mason Univ...' and several open tabs: 'Mail - Aitharva Abha...', 'Patriot Pass Passwor...', 'portalustraveldocs...', 'Patriot Web - A Self...', 'Login Page', and 'Handshake'. The notebook's left sidebar shows a 'Data' tab. The main area displays a table with 10 columns: 'id', 'name', 'age', 'gender', 'height', 'weight', 'eye\_color', 'hair\_color', 'skin\_color', and 'blood\_type'. The table contains 10 rows of data. Below the table, there is a code cell with the following text: 'The next command creates a table from a Databricks dataset'. The bottom of the screen shows a Windows taskbar with a search bar, various application icons, and system tray icons including the language 'ENG IN', network status, and the date '9/5/2024'.