

1)

a) Topic name: *Lab7_topic*

```
me17b140@kafka-1-kafka-vm-0: ~ - Google Chrome
ssh.cloud.google.com/projects/lab-7-310219/zones/us-west4-a/instances/kafka-1-kafka-vm-0?authuser=0&hl=
me17b140@kafka-1-kafka-vm-0:~$ ls
producer.py producer.py.save producer.py.save.1 producer.py.save.2 producer.py.save.3
me17b140@kafka-1-kafka-vm-0:~$ python producer.py
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.1,4.0,2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.3,6.1,4.0,2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.3,4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3,1.4,0.1,Iris-setosa
4.3,3,1.1,0.1,Iris-setosa
5.8,4,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
5.4,3.9,1.3,0.4,Iris-setosa
5.1,3.5,1.4,0.3,Iris-setosa
5.7,3.8,1.7,0.3,Iris-setosa
5.1,3.8,1.5,0.3,Iris-setosa
5.4,3.4,1.7,0.2,Iris-setosa
5.1,3.7,1.5,0.4,Iris-setosa
4.6,3.6,1,0.2,Iris-setosa
5.1,3.3,1.7,0.5,Iris-setosa
4.8,3.4,1.9,0.2,Iris-setosa
5,3,1.6,0.2,Iris-setosa
5,3,4,1.6,0.4,Iris-setosa
5.2,3.5,1.5,0.2,Iris-setosa
5.2,3.4,1.4,0.2,Iris-setosa
4.7,3.2,1.6,0.2,Iris-setosa
4.8,3.1,1.6,0.2,Iris-setosa
5.4,3.4,1.5,0.4,Iris-setosa
5.2,4.1,1.5,0.1,Iris-setosa
5.5,4.2,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5,3.2,1.2,0.2,Iris-setosa
5.5,3.5,1.3,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
4.4,3,1.3,0.2,Iris-setosa
5.1,3.4,1.5,0.2,Iris-setosa
5,3.5,1.3,0.3,Iris-setosa
4.5,2.3,1.3,0.3,Iris-setosa
4.4,3.2,1.3,0.2,Iris-setosa
5,3.5,1.6,0.6,Iris-setosa
5.1,3.8,1.9,0.4,Iris-setosa
4.8,3,1.4,0.3,Iris-setosa
5.1,3.8,1.6,0.2,Iris-setosa
```

```
me17b140@kafka-1-kafka-vm-0: ~ - Google Chrome
ssh.cloud.google.com/projects/lab-7-310219/zones/us-west4-a/instances/kafka-1-kafka-vm-0?authuser=0&hl=en_US&projectNumber=1061586762995&useAdminProxy=true
me17b140@kafka-1-kafka-vm-0:~$ kafka-console-consumer.sh --bootstrap-server localhost:9092 --topic lab7_topic --from-beginning
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.1,4.0,2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.3,6.1,4.0,2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.3,4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5.4,3.7,1.5,0.2,Iris-setosa
4.8,3.4,1.6,0.2,Iris-setosa
4.8,3,1.4,0.1,Iris-setosa
4.3,3,1.1,0.1,Iris-setosa
5.8,4,1.2,0.2,Iris-setosa
5.7,4.4,1.5,0.4,Iris-setosa
5.4,3.9,1.3,0.4,Iris-setosa
5.1,3.5,1.4,0.3,Iris-setosa
5.7,3.8,1.7,0.3,Iris-setosa
5.1,3.8,1.5,0.3,Iris-setosa
5.4,3.4,1.7,0.2,Iris-setosa
5.1,3.7,1.5,0.4,Iris-setosa
4.6,3.6,1,0.2,Iris-setosa
5.1,3.3,1.7,0.5,Iris-setosa
4.8,3.4,1.9,0.2,Iris-setosa
5,3,1.6,0.2,Iris-setosa
5,3,4,1.6,0.4,Iris-setosa
5.2,3.5,1.5,0.2,Iris-setosa
5.2,3.4,1.4,0.2,Iris-setosa
4.7,3.2,1.6,0.2,Iris-setosa
4.8,3.1,1.6,0.2,Iris-setosa
5.4,3.4,1.5,0.4,Iris-setosa
5.2,4.1,1.5,0.1,Iris-setosa
5.5,4.2,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
5,3.2,1.2,0.2,Iris-setosa
5.5,3.5,1.3,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
4.4,3,1.3,0.2,Iris-setosa
5.1,3.4,1.5,0.2,Iris-setosa
5,3.5,1.3,0.3,Iris-setosa
4.5,2.3,1.3,0.3,Iris-setosa
4.4,3.2,1.3,0.2,Iris-setosa
5,3.5,1.6,0.6,Iris-setosa
5.1,3.8,1.9,0.4,Iris-setosa
4.8,3,1.4,0.3,Iris-setosa
5.1,3.8,1.6,0.2,Iris-setosa
4.6,3.2,1.4,0.2,Iris-setosa
5.3,3.7,1.5,0.2,Iris-setosa
```

b) Console output while publishing the data:

[illegible]

Console output:

Google Cloud Platformlab 7

Search products and resources

Job details

CLONEDELETESTOPREFRESH

Job output

LINE WRAP: OFF

```
| null|
+-----+
-----+
Batch: 1
-----+
+-----+
|accuracy|
+-----+
| 100.00|
+-----+

-----+
Batch: 1
-----+
|prediction_class|prediction|true_label|correct|
+-----+
|setosa|0.0|0.0|1|
|setosa|0.0|0.0|1|
|setosa|0.0|0.0|1|
|setosa|0.0|0.0|1|
|setosa|0.0|0.0|1|
|setosa|0.0|0.0|1|
|setosa|0.0|0.0|1|
|setosa|0.0|0.0|1|
+-----+

Job output is complete
```

Equivalent: [command line](#)

From lab 5 the Random Forest classifier model is saved and is used for this lab's task. The saved model is in the below screenshot, placed in the bucket.

The screenshot shows the Google Cloud Platform interface for a bucket named 'lab-7-bucket'. The 'OBJECTS' tab is selected, displaying a list of objects. The table includes columns for Name, Size, Type, Created time, Storage class, Last modified, Public access, Encryption, Retention expiration date, and Holds. The objects listed are 'RFC_lab_5_iris_model/' (Folder), 'iris.csv' (4.5 KB), 'iris.py' (2.9 KB), and 'subscriber.py' (2.4 KB).

<input type="checkbox"/>	Name	Size	Type	Created time	Storage class	Last modified	Public access	Encryption	Retention expiration date	Holds
<input type="checkbox"/>	RFC_lab_5_iris_model/	—	Folder	—	—	—	—	—	—	—
<input type="checkbox"/>	iris.csv	4.5 KB	application/vnd.ms-excel	Apr 10, 2021, 1...	Standard	Apr 10, 20...	Not public	Google-managed key	—	None
<input type="checkbox"/>	iris.py	2.9 KB	text/x-python	Apr 13, 2021, 3...	Standard	Apr 13, 20...	Not public	Google-managed key	—	None
<input type="checkbox"/>	subscriber.py	2.4 KB	text/x-python	Apr 13, 2021, 7...	Standard	Apr 13, 20...	Not public	Google-managed key	—	None

The screenshot shows the 'Submit a job' form in the Google Cloud Platform interface. The form includes fields for Job ID, Region, Cluster, Job type, Main python file, and Jar files. The Job ID is 'subscriber_job', Region is 'us-central1', Cluster is 'lab7-cluster', Job type is 'PySpark', and Main python file is 'gs://lab-7-bucket/subscriber.py'. The Jar files section lists several dependencies.

Job ID *
subscriber_job

Region *
us-central1
Specifies the Cloud Dataproc regional service, which determines what clusters are available.

Cluster *
lab7-cluster

Job type *
PySpark

Main python file *
gs://lab-7-bucket/subscriber.py
Can be a GCS file with the gs:// prefix, an HDFS file on the cluster with the hdfs:// prefix, or a local file on the cluster with the file:// prefix"

Additional python files

Jar files

- gs://bdld1/commons-pool2-2.6.2.jar
- gs://bdld1/kafka-clients-2.6.0.jar
- gs://bdld1/lz4-java-1.7.1.jar
- gs://bdld1/scala-library-2.12.10.jar
- gs://bdld1/slf4j-api-1.7.30.jar
- gs://bdld1/snappy-java-1.1.7.3.jar
- gs://bdld1/spark-sql-kafka-0-10_2.12-3.1.1.jar
- gs://bdld1/spark-tags_2.12-3.1.1.jar
- gs://bdld1/spark-token-provider-kafka-0-10_2.12-3.1.1.jar