

CS4830 - LAB 7 ASSIGNMENT

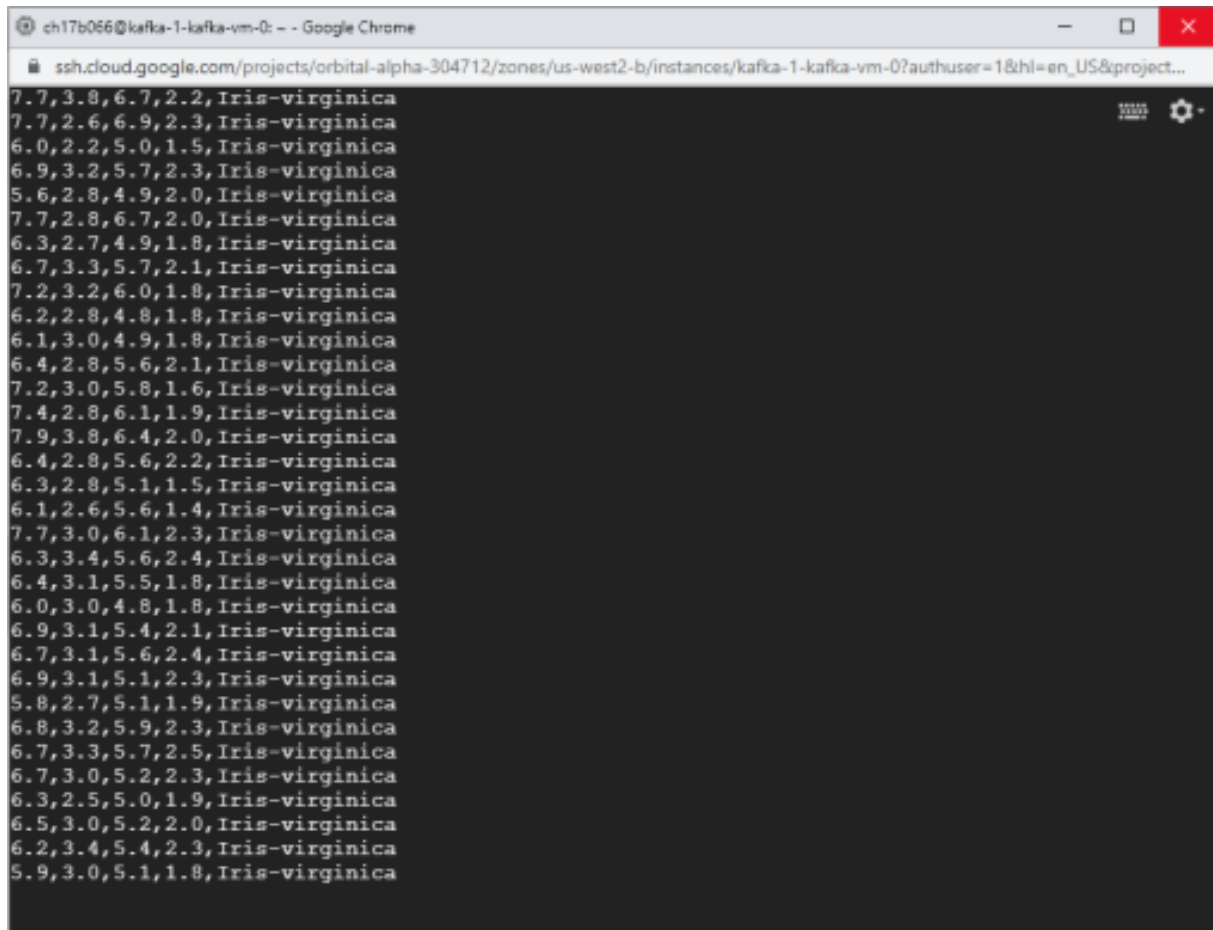
BIG DATA LABORATORY: JAN-MAY 2021

Raj Jain | CH17B066

April 13, 2021

A

Rows written by producer.py are received by the consumer



```
ch17b066@kafka-1-kafka-vm-0: ~ - Google Chrome
ssh.cloud.google.com/projects/orbital-alpha-304712/zones/us-west2-b/instances/kafka-1-kafka-vm-0?authuser=1&hl=en_US&project...
7.7,3.8,6.7,2.2,Iris-virginica
7.7,2.6,6.9,2.3,Iris-virginica
6.0,2.2,5.0,1.5,Iris-virginica
6.9,3.2,5.7,2.3,Iris-virginica
5.6,2.8,4.9,2.0,Iris-virginica
7.7,2.8,6.7,2.0,Iris-virginica
6.3,2.7,4.9,1.8,Iris-virginica
6.7,3.3,5.7,2.1,Iris-virginica
7.2,3.2,6.0,1.8,Iris-virginica
6.2,2.8,4.8,1.8,Iris-virginica
6.1,3.0,4.9,1.8,Iris-virginica
6.4,2.8,5.6,2.1,Iris-virginica
7.2,3.0,5.8,1.6,Iris-virginica
7.4,2.8,6.1,1.9,Iris-virginica
7.9,3.8,6.4,2.0,Iris-virginica
6.4,2.8,5.6,2.2,Iris-virginica
6.3,2.8,5.1,1.5,Iris-virginica
6.1,2.6,5.6,1.4,Iris-virginica
7.7,3.0,6.1,2.3,Iris-virginica
6.3,3.4,5.6,2.4,Iris-virginica
6.4,3.1,5.5,1.8,Iris-virginica
6.0,3.0,4.8,1.8,Iris-virginica
6.9,3.1,5.4,2.1,Iris-virginica
6.7,3.1,5.6,2.4,Iris-virginica
6.9,3.1,5.1,2.3,Iris-virginica
5.8,2.7,5.1,1.9,Iris-virginica
6.8,3.2,5.9,2.3,Iris-virginica
6.7,3.3,5.7,2.5,Iris-virginica
6.7,3.0,5.2,2.3,Iris-virginica
6.3,2.5,5.0,1.9,Iris-virginica
6.5,3.0,5.2,2.0,Iris-virginica
6.2,3.4,5.4,2.3,Iris-virginica
5.9,3.0,5.1,1.8,Iris-virginica
```

B

Training & console output

Job details

CLONEDELETESTOPREFRESH

Job output

LINE WRAP: OFF

21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Step Size: 1.000
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Val and Grad Norm: 0.00222308 (rel: 2.09e-06) 0.08180e-06
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Step Size: 1.000
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Val and Grad Norm: 0.00222308 (rel: 7.86e-07) 1.05270e-05
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Step Size: 1.000
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Val and Grad Norm: 0.00222307 (rel: 3.64e-06) 1.73636e-05
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Step Size: 1.000
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Val and Grad Norm: 0.00222305 (rel: 9.44e-06) 2.97010e-05
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Step Size: 1.000
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Val and Grad Norm: 0.00222308 (rel: 2.38e-05) 4.79217e-05
21/04/13 19:02:40 INFO breeze.optimize.LBFGS: Converged because max iterations reached

Best_model valid performance (LR): 0.9523809523809523

21/04/13 19:02:43 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/metadata/_temporary/0/task_21/04/13 19:02:44 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/metadata/' directory.
21/04/13 19:02:46 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/stages/0_StringIndexer_92ce
21/04/13 19:02:46 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/stages/0_StringIndexer_92ce
21/04/13 19:02:46 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/stages/0_StringIndexer_92ce
21/04/13 19:02:50 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/stages/0_StringIndexer_92ce
21/04/13 19:02:53 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/stages/1_VectorAssembler_3c
21/04/13 19:02:53 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageFileSystem: Successfully repaired 'gs://rajlab7/model/stages/1_VectorAssembler_3c

Equivalent:

command line

rajlab7

OBJECTS

CONFIGURATION

PERMISSIONS

RETENTION

LIFECYCLE

Buckets

rajlab7

UPLOAD FILES

UPLOAD FOLDER

CREATE FOLDER

MANAGE HOLDS

DOWNLOAD

DELETE

Filter by name prefix only

Filter

Filter objects and folders

<input type="checkbox"/>	Name	Size	Type	Created time	Storage class	Last modified	Public access	Encryption	Retention e	
<input type="checkbox"/>	iris.csv	4.4 KB	application/vnd.ms-excel	Apr 14, 2021, 1...	Standard	Apr 14, 20...	Not public	Google-managed key	—	<div></div>
<input type="checkbox"/>	model/	—	Folder	—	—	—	—	—	—	<div></div>
<input type="checkbox"/>	subscriber.py	2.8 KB	application/octet-stream	Apr 14, 2021, 1...	Standard	Apr 14, 20...	Not public	Google-managed key	—	<div></div>
<input type="checkbox"/>	train.py	2.8 KB	application/octet-stream	Apr 14, 2021, 1...	Standard	Apr 14, 20...	Not public	Google-managed key	—	<div></div>

Job details

CLONEDELETESTOPREFRESH

Job output

LINE WRAP: OFF

Batch: 5

| Class|label|prediction|accuracy|

Iris-versicolor	0.0	0.0	1
Iris-versicolor	0.0	0.0	1
Iris-versicolor	0.0	0.0	1
Iris-versicolor	0.0	0.0	1
Iris-virginica	1.0	1.0	1
Iris-virginica	1.0	1.0	1
Iris-virginica	1.0	1.0	1
Iris-setosa	2.0	2.0	1
Iris-setosa	2.0	2.0	1
Iris-setosa	2.0	2.0	1

END