

Term project – METCS777 (gk@bu.edu)

Spark History screenshots – (main_task1.py)

- App Name: Project-task1
- App ID: [application_1588417284511_0002](#)

https://hjoeazet7vgghhzevjkrnmju7m-dot-us-east1.dataproc.googleusercontent.com/sparkhistory/ - Google Chrome

project3 - My Project - Go... Clusters - Dataproc - My... https://hjoeazet7vgghhzevjkrnmju7m-dot-us-east1.dataproc.googleusercontent.com/sparkhistory/

evident-hexagon-113810 > project2 Sign out

History Server

Event log directory: hdfs://project2-m/user/spark/eventlog
Last updated: 2020-05-02 15:05:59
Client local time zone: America/New_York

Search:

App ID	App Name	Started	Completed	Duration	Spark User	Last Updated	Event Log
application_1588417284511_0002	Project-task1	2020-05-02 12:58:24	2020-05-02 15:10:35	2.2 h	root	2020-05-02 15:10:35	Download
application_1588417284511_0001	Assignment5-task3	2020-05-02 07:02:36	2020-05-02 11:33:11	4.5 h	root	2020-05-02 11:33:11	Download

Showing 1 to 2 of 2 entries
[Show incomplete applications](#)

https://hjoeazet7vgghhzevjkrnmju7m-dot-us-east1.dataproc.googleusercontent.com/gateway/default/sparkhistory/history/application_1588417284511_0002/jobs/ - Google Chrome

project3 - My Project - Go... Clusters - Dataproc - My... https://hjoeazet7vgghhzevjkrnmju7m-dot-us-east1.dataproc.googleusercontent.com/gateway/default/sparkhistory/history/application_1588417284511_0002/jobs/

evident-hexagon-113810 > project2 Sign out

Spark Jobs (7)

User: root
Total Uptime: 2.2 h
Scheduling Mode: FAIR
Completed Jobs: 462

▼ Event Timeline
Enable zooming

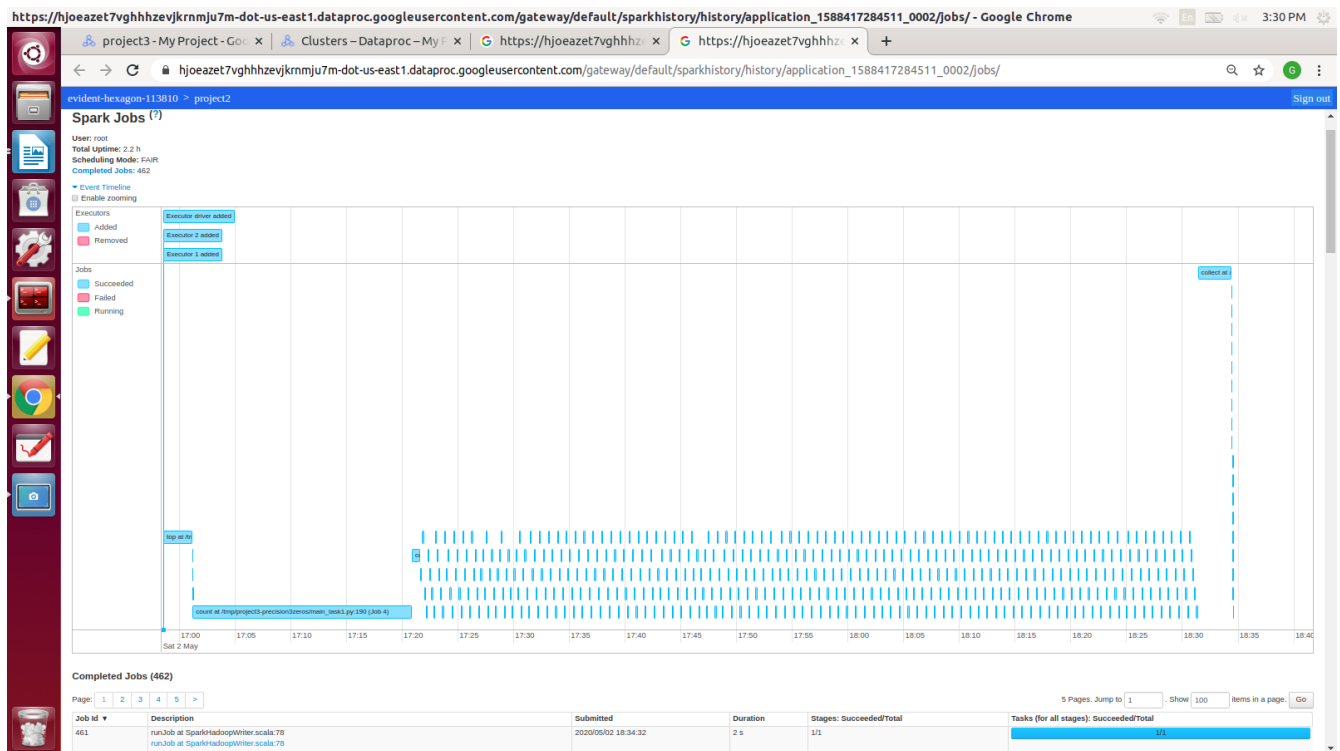
Executors
Added
Removed

Jobs
Succeeded
Failed
Running

count at /tmp/project3-precision3overmean_task1.py (Job 5)

Completed Jobs (462)
Page: 1 2 3 4 5 >

5 Pages. Jump to 1 - Show 100 items in a page. Go



Google Storage screenshots – output folder (Output_task1) generated by main_task1.py

Bucket details - My Project - Google Cloud Platform - Google Chrome

project3 - My Project - Google Cloud Platform | Overview - BigQuery | Clusters - Dataproc - My Project | Clusters - Dataproc - My Project | https://hjoeazet7vghhzevjkrnmju7m-dot-us-east1.dataproc.googleusercontent.com/gateway/default/sparkhistory/history/application_1588417284511_0002/jobs/ | https://5kfa1... | met-cs-777-as... | Bucket details |

console.cloud.google.com/storage/browser/gagankaushal.com/Project?forceOnBucketsSortingFiltering=false&project=eventid-hexagon-113810

Your free trial is waiting: activate now to get \$300 credit to explore Google Cloud products. [Learn more](#) DISMISS ACTIVATE

Google Cloud Platform

My Project

Search resources and products

Storage

Browser
Transfer
Transfer for on-premises
Transfer Appliance
Settings

Bucket details

EDIT BUCKET REFRESH BUCKET

gagankaushal.com

Public

Objects Overview Permissions Bucket Lock

Upload files Upload folder Create folder Manage holds Delete

Filter by prefix...

Buckets / gagankaushal.com / Project

Name	Size	Type	Storage class	Last modified	Public access	Encryption	Retention expiration date	Holds
Assignment_4_Final_Output_Task2_part-00000	525.86 KB	application/octet-stream	Standard	5/2/20, 6:26:18 AM UTC-4	Public to internet CD	Google-managed key	-	None
Output_task1/	-	Folder	-	-	Public to internet	-	-	-
main_task1.py	26.19 KB	text/x-python	Standard	5/2/20, 6:20:04 AM UTC-4	Public to internet CD	Google-managed key	-	None

Bucket details - My Project - Google Cloud Platform - Google Chrome

project3 - My Project - Google Cloud Platform x | https://hjoeazet7vghhzh x | https://hjoeazet7vghhzh x | Bucket details - My Project x +

console.cloud.google.com/storage/browser/gagankaushal.com/Project/Output_task1/?forceOnBucketsSortingFiltering=false&project=evident-hexagon-113810

Your free trial is waiting: activate now to get \$300 credit to explore Google Cloud products. [Learn more](#) DISMISS ACTIVATE

Google Cloud Platform My Project Search resources and products

Storage

Bucket details EDIT BUCKET REFRESH BUCKET

gagankaushal.com

Objects Overview Permissions Bucket Lock

Upload files Upload folder Create folder Manage holds Delete

Filter by prefix...

Buckets / gagankaushal.com / Project / Output_task1

Name	Size	Type	Storage class	Last modified	Public access	Encryption	Retention expiration date	Holds
_SUCCESS	0 B	application/octet-stream	Standard	5/2/20, 11:33:11 AM UTC-4	Not public	Google-managed key	-	None
part-00000	1.36 KB	application/octet-stream	Standard	5/2/20, 11:33:10 AM UTC-4	Not public	Google-managed key	-	None

Project_Output_task1_part-00000 (~/Desktop/BU/Spring2020/BigDataAnalytics/Project/project/FINAL) - gedit

Open ~/Desktop/BU/Spring2020/BigDataAnalytics/Project/project/FINAL Project_Output_task1_part-00000 Save

*Untitled Document 1 Things_to_keep_in_m...d_SparkLibraries.txt Project_Output_task1_part-00000

```
(#####, ' PERCEPTRON - RESULTS ', '#####')
('F1 score for classifier =', 98.39, '%')

('Time taken to read Testing and Training Data and preprocess them (days:seconds:microsecond):', datetime.timedelta(0, 2400, 746215))
('Time taken to train the Perceptron model (days:seconds:microsecond):', datetime.timedelta(0, 1009, 124341))
('Time taken to test the Perceptron model (days:seconds:microsecond):', datetime.timedelta(0, 204, 948016))
('Total Time taken by Perceptron:', datetime.timedelta(0, 3614, 818572))

('Number of True Positives', 365)
('Number of False Positives', 0)
('Number of False Negatives', 12)
('Number of True Negatives', 18347)

(#####, ' SUPPORT VECTOR - RESULTS ', '#####')
('F1 score for classifier =', 98.79, '%')

('Time taken to read Testing and Training Data and preprocess them (days:seconds:microsecond):', datetime.timedelta(0, 2400, 746215))
('Time taken to train the Support Vector Machine model (days:seconds:microsecond):', datetime.timedelta(0, 1715, 655587))
('Time taken to test using the Support Vector Machine model (days:seconds:microsecond):', datetime.timedelta(0, 138, 874497))
('Total Time taken by Support Vector Machine (days:seconds:microsecond):', datetime.timedelta(0, 4255, 276299))

('Number of True Positives', 368)
('Number of False Positives', 0)
('Number of False Negatives', 9)
('Number of True Negatives', 18347)]
```

Plain Text Tab Width: 8 Ln 25, Col 36 INS