SHARCNET General Interest Webinar Series

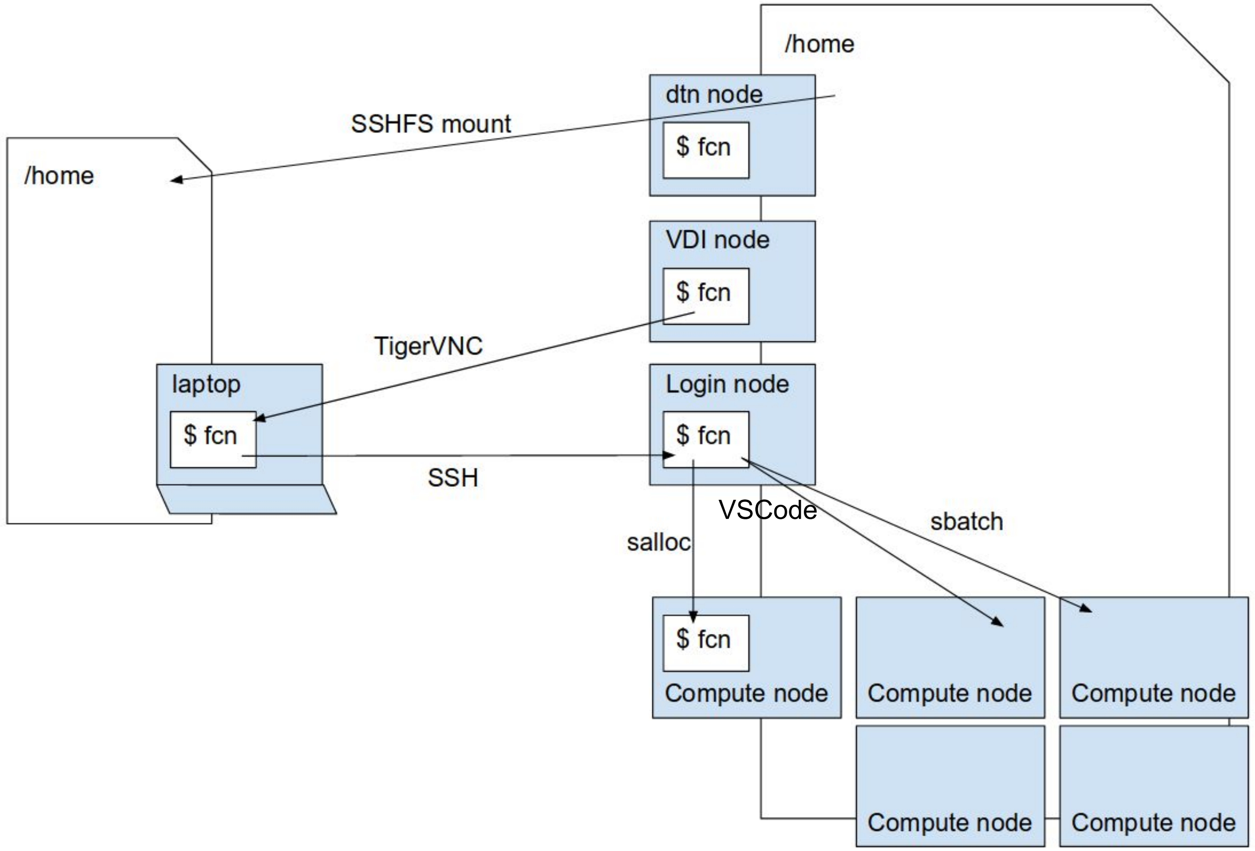
Generating Interactive Visualizations with Plotly on Graham

James Desjardins
High Performance Computing Consultant
SHARCNET, Brock University
February 24th, 2021

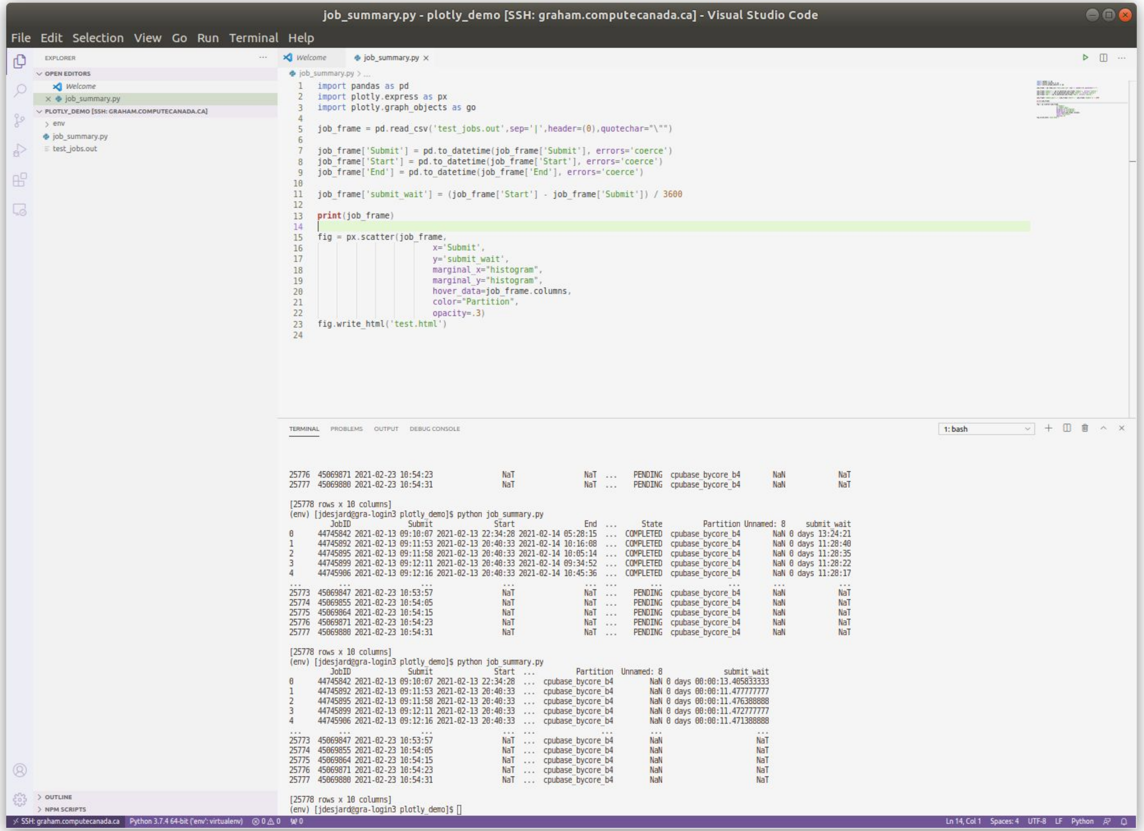
Overview

- Resources for exploring interactive figures generated on Graham
  - Visual Studio Code for code editing and SSH connection to the Graham login node
  - SSHFS file mount
  - TigerVNC for graphical interface connection to the Graham VDI node
  - Slurm salloc for connecting to the Visual Studio Code terminal to a compute node
  - “top” for monitoring resource utilization
- Software tools for building data and visualization
  - Slurm sacct output forwarded to text files for tabular data example
  - Pandas package in Python for reading tabular text files into dataframes
  - Plotly package in Python for generating interactive HTML figures
- Viewing HTML figures
  - Using browser on the local machine access files via SSHFS mount
  - Using browser on gra-vdi node accessed via TigerVNC

Overview of resources and access methods



Visual Studio Code for code editing and SSH

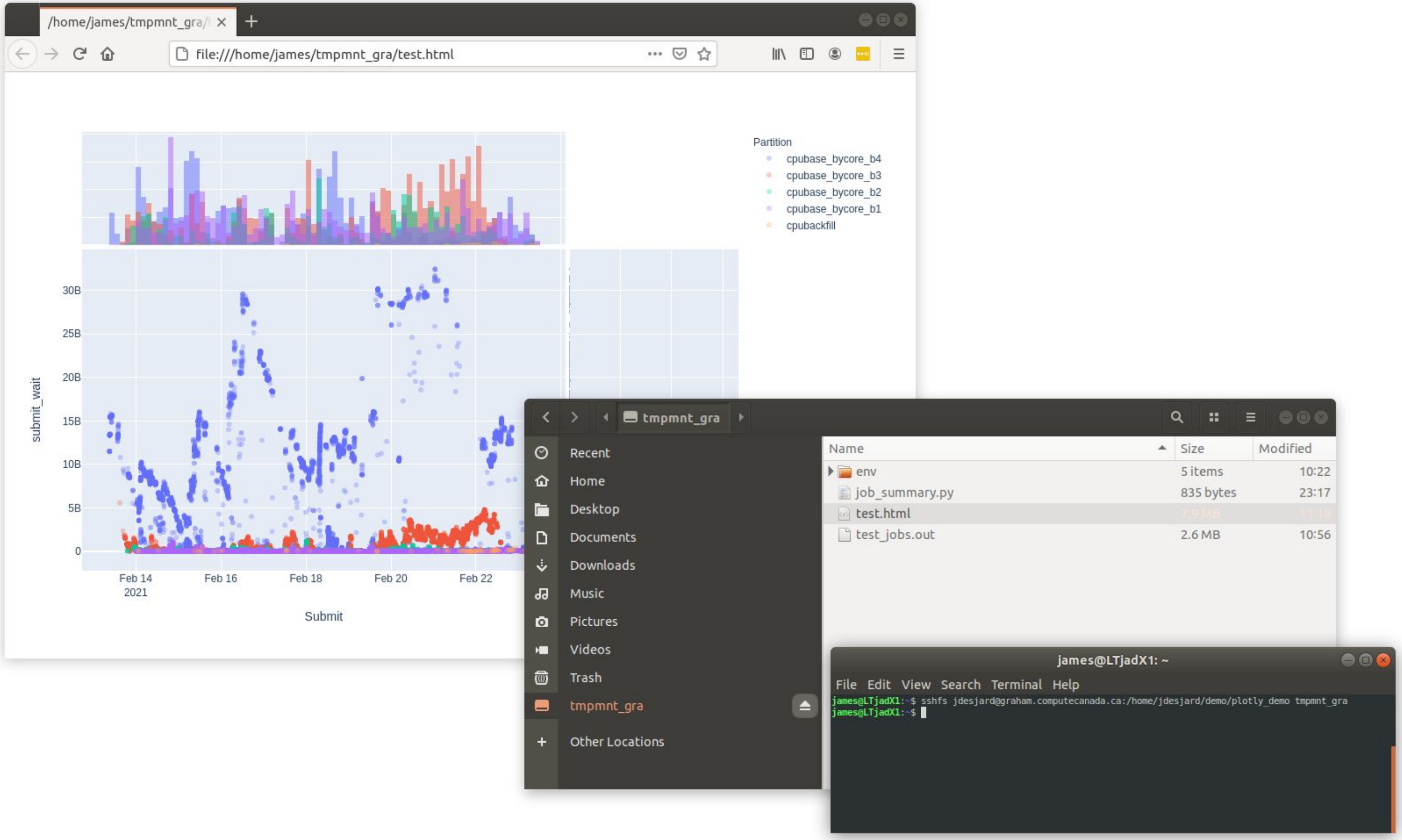


job_summary.py - plotly_demo [SSH: graham.computecanada.ca] - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
OPEN EDITORS
Welcome
job_summary.py
PLOTLY_DEMO [SSH: GRAHAM.COMPUTECANADA.CA]
env
job_summary.py
test_jobs.out
OUTLINE
NPM SCRIPTS
TERMINAL PROBLEMS OUTPUT DEBUG CONSOLE
1: bash
SSH: graham.computecanada.ca
Python 3.7.4 64-bit ('env': virtualenv)
Ln 14, Col 1 Spaces: 4 UTF-8 LF Python

```python
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go

job_frame = pd.read_csv('test_jobs.out',sep='|',header=(0),quotechar="\"")

job_frame['Submit'] = pd.to_datetime(job_frame['Submit'], errors='coerce')
job_frame['Start'] = pd.to_datetime(job_frame['Start'], errors='coerce')
job_frame['End'] = pd.to_datetime(job_frame['End'], errors='coerce')

job_frame['submit_wait'] = (job_frame['Start'] - job_frame['Submit']) / 3600

print(job_frame)

fig = px.scatter(job_frame,
                       x='Submit',
                       y='submit_wait',
                       marginal_x="histogram",
                       marginal_y="histogram",
                       hover_data=job_frame.columns,
                       color="Partition",
                       opacity=.3)
fig.write_html('test.html')

```

```
25776  45069871 2021-02-23 10:54:23                 NaT                 NaT  ...    PENDING  cpubase_bycore_b4       NaN            NaT
25777  45069880 2021-02-23 10:54:31                 NaT                 NaT  ...    PENDING  cpubase_bycore_b4       NaN            NaT

[25778 rows x 10 columns]
(env) [jdesjard@gra-login3 plotly_demo]$ python job_summary.py
          JobID              Submit               Start                 End  ...      State          Partition Unnamed: 8     submit_wait
0      44745842 2021-02-13 09:10:07 2021-02-13 22:34:28 2021-02-14 05:28:15  ...  COMPLETED  cpubase_bycore_b4        NaN 0 days 13:24:21
1      44745892 2021-02-13 09:11:53 2021-02-13 20:40:33 2021-02-14 10:16:08  ...  COMPLETED  cpubase_bycore_b4        NaN 0 days 11:28:40
2      44745895 2021-02-13 09:11:58 2021-02-13 20:40:33 2021-02-14 10:05:14  ...  COMPLETED  cpubase_bycore_b4        NaN 0 days 11:28:35
3      44745899 2021-02-13 09:12:11 2021-02-13 20:40:33 2021-02-14 09:34:52  ...  COMPLETED  cpubase_bycore_b4        NaN 0 days 11:28:22
4      44745906 2021-02-13 09:12:16 2021-02-13 20:40:33 2021-02-14 10:45:36  ...  COMPLETED  cpubase_bycore_b4        NaN 0 days 11:28:17
...         ...                 ...                 ...                 ...  ...        ...                ...        ...             ...
25773  45069847 2021-02-23 10:53:57                 NaT                 NaT  ...    PENDING  cpubase_bycore_b4       NaN            NaT
25774  45069855 2021-02-23 10:54:05                 NaT                 NaT  ...    PENDING  cpubase_bycore_b4       NaN            NaT
25775  45069864 2021-02-23 10:54:15                 NaT                 NaT  ...    PENDING  cpubase_bycore_b4       NaN            NaT
25776  45069871 2021-02-23 10:54:23                 NaT                 NaT  ...    PENDING  cpubase_bycore_b4       NaN            NaT
25777  45069880 2021-02-23 10:54:31                 NaT                 NaT  ...    PENDING  cpubase_bycore_b4       NaN            NaT

[25778 rows x 10 columns]
(env) [jdesjard@gra-login3 plotly_demo]$ python job_summary.py
          JobID              Submit               Start  ...          Partition  Unnamed: 8                submit_wait
0      44745842 2021-02-13 09:10:07 2021-02-13 22:34:28  ...  cpubase_bycore_b4         NaN 0 days 00:00:13.405833333
1      44745892 2021-02-13 09:11:53 2021-02-13 20:40:33  ...  cpubase_bycore_b4         NaN 0 days 00:00:11.477777777
2      44745895 2021-02-13 09:11:58 2021-02-13 20:40:33  ...  cpubase_bycore_b4         NaN 0 days 00:00:11.476388888
3      44745899 2021-02-13 09:12:11 2021-02-13 20:40:33  ...  cpubase_bycore_b4         NaN 0 days 00:00:11.472777777
4      44745906 2021-02-13 09:12:16 2021-02-13 20:40:33  ...  cpubase_bycore_b4         NaN 0 days 00:00:11.471388888
...         ...                 ...                 ...  ...                ...         ...                        ...
25773  45069847 2021-02-23 10:53:57                 NaT  ...  cpubase_bycore_b4         NaN                        NaT
25774  45069855 2021-02-23 10:54:05                 NaT  ...  cpubase_bycore_b4         NaN                        NaT
25775  45069864 2021-02-23 10:54:15                 NaT  ...  cpubase_bycore_b4         NaN                        NaT
25776  45069871 2021-02-23 10:54:23                 NaT  ...  cpubase_bycore_b4         NaN                        NaT
25777  45069880 2021-02-23 10:54:31                 NaT  ...  cpubase_bycore_b4         NaN                        NaT

[25778 rows x 10 columns]
(env) [jdesjard@gra-login3 plotly_demo]$ 
```

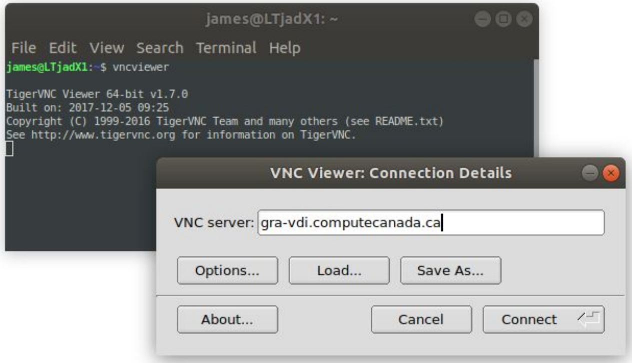
SSHFS mount and local browser viewing



Name	Size	Modified
env	5 items	10:22
job_summary.py	835 bytes	23:17
test.html	7.9 MB	11:10
test_jobs.out	2.6 MB	10:56

```
james@LTjadX1:~$ sshfs jdesjard@graham.computecanada.ca:/home/jdesjard/demo/plotly_demo tmpmnt_gra
james@LTjadX1:~$ 
```

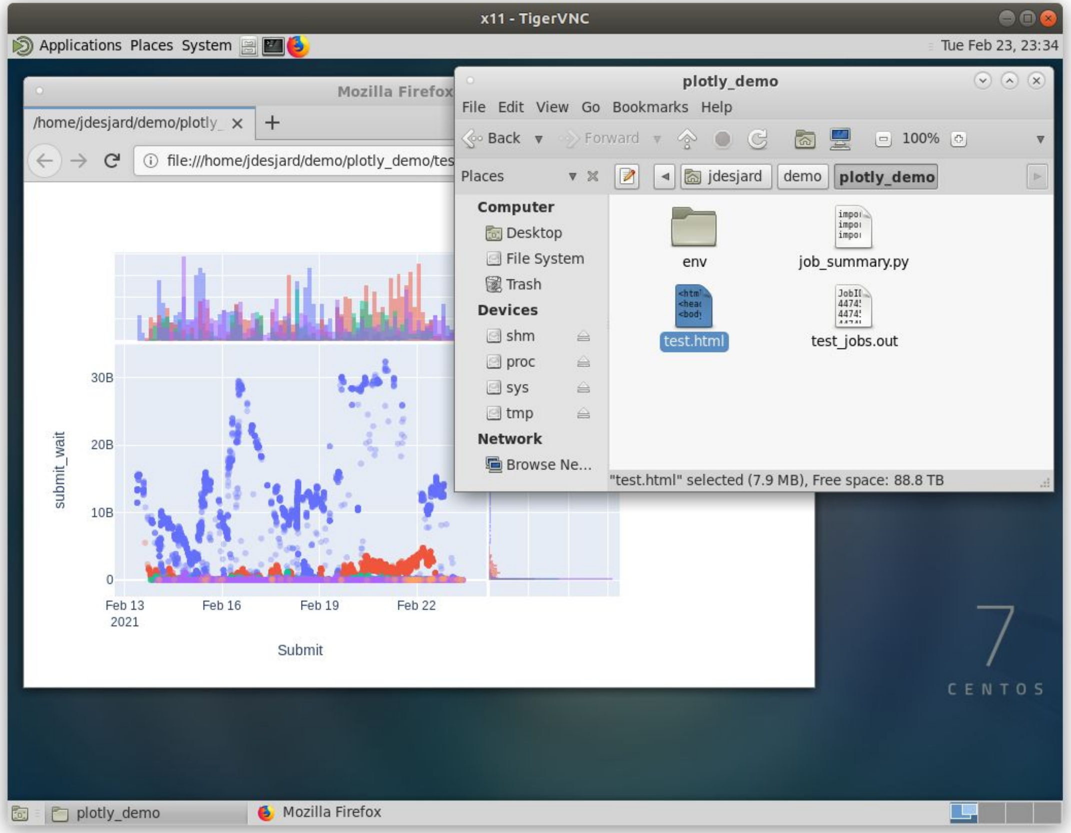
TigerVNC connection to Graham VDI node for browser viewing



```
james@LTjadX1:~$ vncviewer

TigerVNC Viewer 64-bit v1.7.0
Built on: 2017-12-05 09:25
Copyright (C) 1999-2016 TigerVNC Team and many others (see README.txt)
See http://www.tigervnc.org for information on TigerVNC.
```

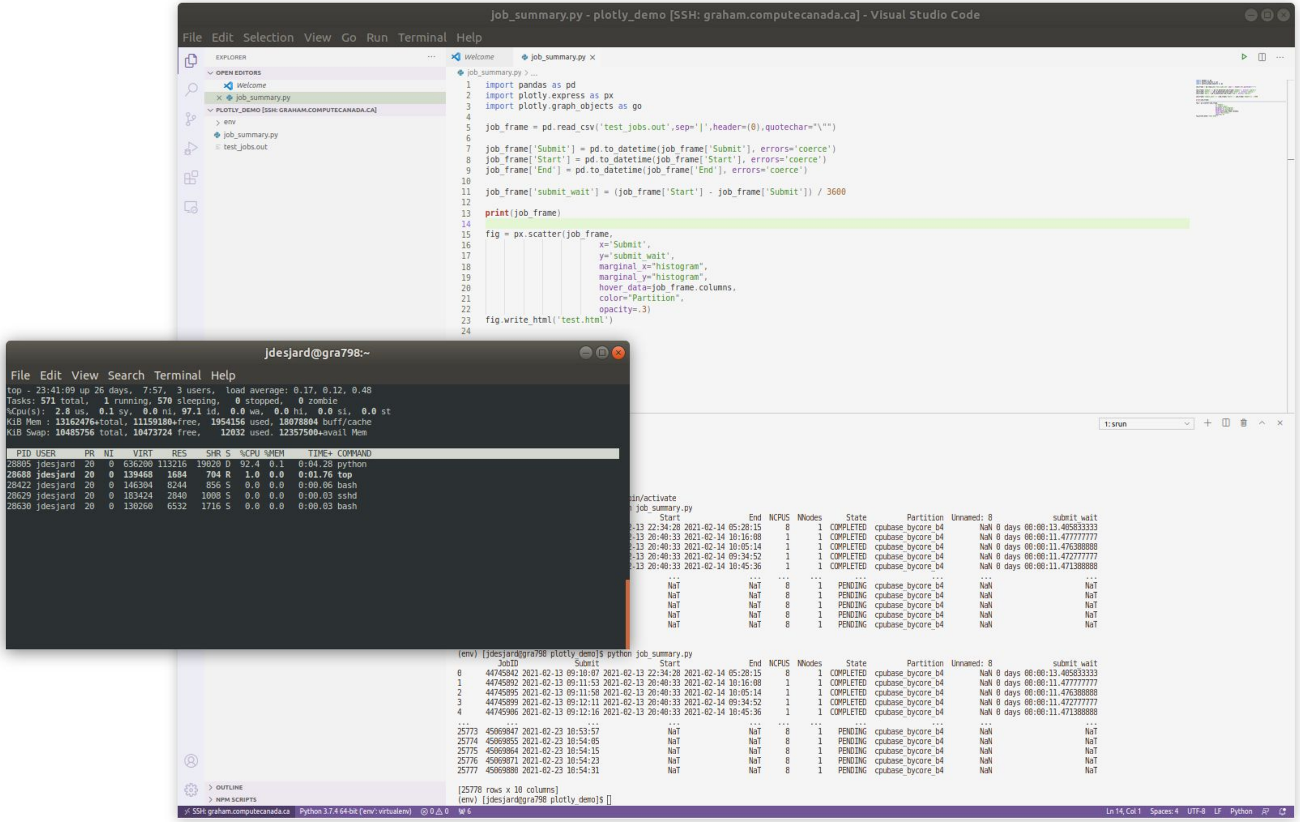
VNC Viewer: Connection Details

VNC server: gra-vdi.computecanada.ca



"test.html" selected (7.9 MB), Free space: 88.8 TB

Slurm salloc compute node access and top resource monitoring

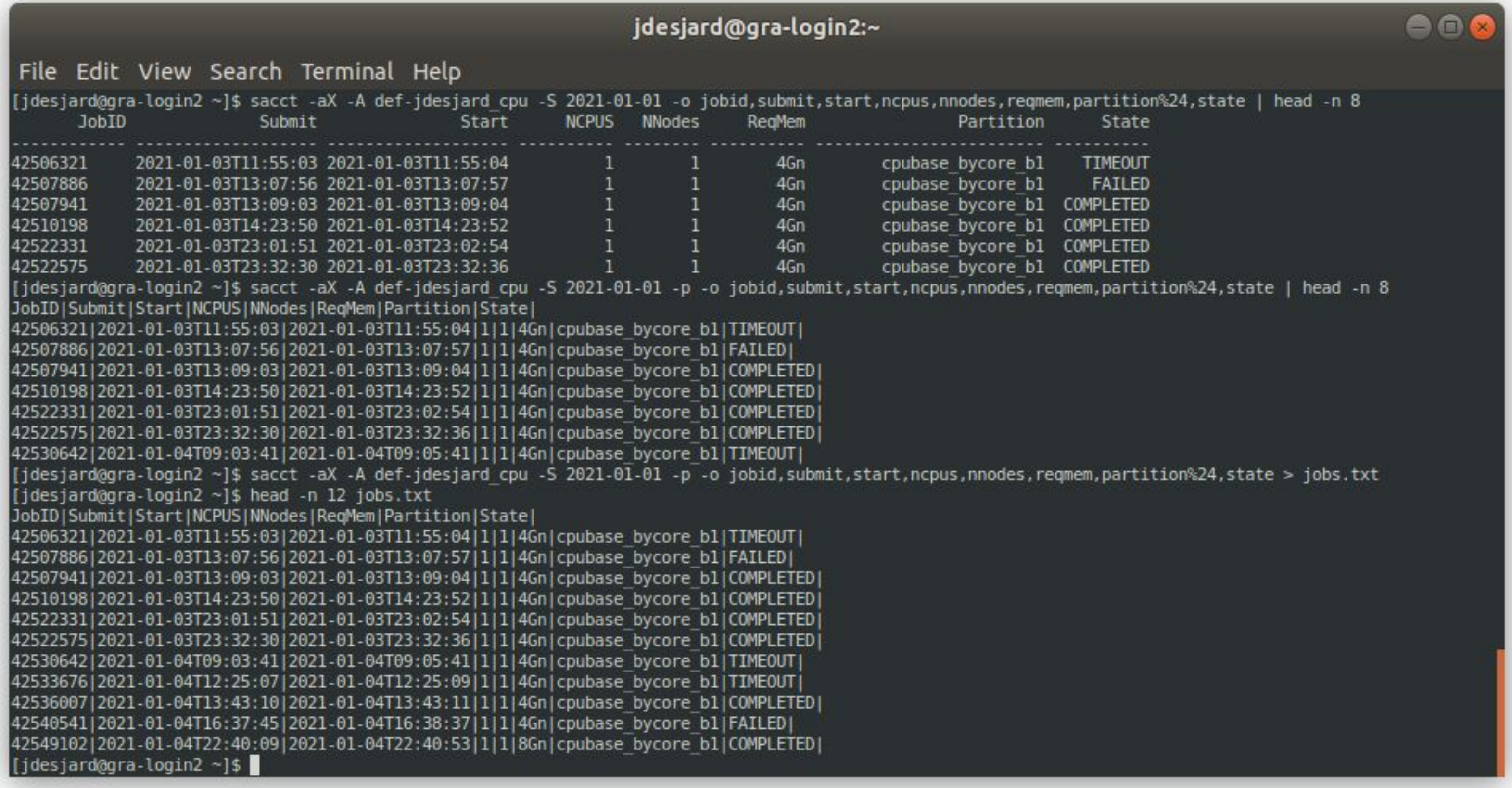

job_summary.py - plotly_demo [SSH: graham.computecanada.ca] - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
OPEN EDITORS
Welcome
job_summary.py
PLOTLY_DEMO [SSH: GRAHAM.COMPUTECANADA.CA]
env
job_summary.py
test_jobs.out
OUTLINE
NPM SCRIPTS
1: srun
SSH: graham.computecanada.ca
Python 3.7.4 64-bit ('env': virtualenv)
Ln 14, Col 1 Spaces: 4 UTF-8 LF Python
jdesjard@gra798:~
File Edit View Search Terminal Help

```python
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go

job_frame = pd.read_csv('test_jobs.out',sep='|',header=(0),quotechar="\"")

job_frame['Submit'] = pd.to_datetime(job_frame['Submit'], errors='coerce')
job_frame['Start'] = pd.to_datetime(job_frame['Start'], errors='coerce')
job_frame['End'] = pd.to_datetime(job_frame['End'], errors='coerce')

job_frame['submit_wait'] = (job_frame['Start'] - job_frame['Submit']) / 3600

print(job_frame)

fig = px.scatter(job_frame,
                 x='Submit',
                 y='submit_wait',
                 marginal_x="histogram",
                 marginal_y="histogram",
                 hover_data=job_frame.columns,
                 color="Partition",
                 opacity=.3)
fig.write_html('test.html')

```

```
top - 23:41:09 up 26 days,  7:57,  3 users,  load average: 0.17, 0.12, 0.48
Tasks: 571 total,   1 running, 570 sleeping,   0 stopped,   0 zombie
%Cpu(s):  2.8 us,  0.1 sy,  0.0 ni, 97.1 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
KiB Mem : 13162476+total, 11159180+free,  1954156 used, 18078804 buff/cache
KiB Swap: 10485756 total, 10473724 free,    12032 used. 12357500+avail Mem

  PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
28805 jdesjard  20   0  636200 113216  19020 D  92.4  0.1   0:04.28 python
28688 jdesjard  20   0  139468   1684    704 R   1.0  0.0   0:01.76 top
28422 jdesjard  20   0  146304   8244    856 S   0.0  0.0   0:00.06 bash
28629 jdesjard  20   0  183424   2840   1008 S   0.0  0.0   0:00.03 sshd
28630 jdesjard  20   0  130260   6532   1716 S   0.0  0.0   0:00.03 bash
```

```
in/activate
 job_summary.py
     Start                    End  NCPUS  NNodes      State         Partition  Unnamed: 8               submit_wait
-13 22:34:28 2021-02-14 05:28:15      8       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:13.405833333
-13 20:40:33 2021-02-14 10:16:08      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.477777777
-13 20:40:33 2021-02-14 10:05:14      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.476388888
-13 20:40:33 2021-02-14 09:34:52      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.472777777
-13 20:40:33 2021-02-14 10:45:36      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.471388888
         ...                  ...    ...     ...        ...               ...         ...                       ...
         NaT                  NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
         NaT                  NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
         NaT                  NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
         NaT                  NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
         NaT                  NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT

(env) [jdesjard@gra798 plotly_demo]$ python job_summary.py
          JobID              Submit               Start                 End  NCPUS  NNodes      State         Partition  Unnamed: 8               submit_wait
0      44745842 2021-02-13 09:10:07 2021-02-13 22:34:28 2021-02-14 05:28:15      8       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:13.405833333
1      44745892 2021-02-13 09:11:53 2021-02-13 20:40:33 2021-02-14 10:16:08      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.477777777
2      44745895 2021-02-13 09:11:58 2021-02-13 20:40:33 2021-02-14 10:05:14      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.476388888
3      44745899 2021-02-13 09:12:11 2021-02-13 20:40:33 2021-02-14 09:34:52      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.472777777
4      44745906 2021-02-13 09:12:16 2021-02-13 20:40:33 2021-02-14 10:45:36      1       1  COMPLETED  cpubase_bycore_b4         NaN 0 days 00:00:11.471388888
...         ...                 ...                 ...                 ...    ...     ...        ...               ...         ...                       ...
25773  45069847 2021-02-23 10:53:57                 NaT                 NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
25774  45069855 2021-02-23 10:54:05                 NaT                 NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
25775  45069864 2021-02-23 10:54:15                 NaT                 NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
25776  45069871 2021-02-23 10:54:23                 NaT                 NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT
25777  45069880 2021-02-23 10:54:31                 NaT                 NaT      8       1    PENDING  cpubase_bycore_b4         NaN                       NaT

[25778 rows x 10 columns]
(env) [jdesjard@gra798 plotly_demo]$ 
```

Slurm sacct job record table generation



```
[jdesjard@gra-login2 ~]$ sacct -aX -A def-jdesjard_cpu -S 2021-01-01 -o jobid,submit,start,ncpus,nnodes,reqmem,partition%24,state | head -n 8
       JobID              Submit               Start      NCPUS   NNodes     ReqMem                Partition      State
------------ ------------------- ------------------- ---------- -------- ---------- ------------------------ ----------
42506321     2021-01-03T11:55:03 2021-01-03T11:55:04          1        1        4Gn        cpubase_bycore_b1    TIMEOUT
42507886     2021-01-03T13:07:56 2021-01-03T13:07:57          1        1        4Gn        cpubase_bycore_b1     FAILED
42507941     2021-01-03T13:09:03 2021-01-03T13:09:04          1        1        4Gn        cpubase_bycore_b1  COMPLETED
42510198     2021-01-03T14:23:50 2021-01-03T14:23:52          1        1        4Gn        cpubase_bycore_b1  COMPLETED
42522331     2021-01-03T23:01:51 2021-01-03T23:02:54          1        1        4Gn        cpubase_bycore_b1  COMPLETED
42522575     2021-01-03T23:32:30 2021-01-03T23:32:36          1        1        4Gn        cpubase_bycore_b1  COMPLETED
[jdesjard@gra-login2 ~]$ sacct -aX -A def-jdesjard_cpu -S 2021-01-01 -p -o jobid,submit,start,ncpus,nnodes,reqmem,partition%24,state | head -n 8
JobID|Submit|Start|NCPUS|NNodes|ReqMem|Partition|State|
42506321|2021-01-03T11:55:03|2021-01-03T11:55:04|1|1|4Gn|cpubase_bycore_b1|TIMEOUT|
42507886|2021-01-03T13:07:56|2021-01-03T13:07:57|1|1|4Gn|cpubase_bycore_b1|FAILED|
42507941|2021-01-03T13:09:03|2021-01-03T13:09:04|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42510198|2021-01-03T14:23:50|2021-01-03T14:23:52|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42522331|2021-01-03T23:01:51|2021-01-03T23:02:54|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42522575|2021-01-03T23:32:30|2021-01-03T23:32:36|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42530642|2021-01-04T09:03:41|2021-01-04T09:05:41|1|1|4Gn|cpubase_bycore_b1|TIMEOUT|
[jdesjard@gra-login2 ~]$ sacct -aX -A def-jdesjard_cpu -S 2021-01-01 -p -o jobid,submit,start,ncpus,nnodes,reqmem,partition%24,state > jobs.txt
[jdesjard@gra-login2 ~]$ head -n 12 jobs.txt
JobID|Submit|Start|NCPUS|NNodes|ReqMem|Partition|State|
42506321|2021-01-03T11:55:03|2021-01-03T11:55:04|1|1|4Gn|cpubase_bycore_b1|TIMEOUT|
42507886|2021-01-03T13:07:56|2021-01-03T13:07:57|1|1|4Gn|cpubase_bycore_b1|FAILED|
42507941|2021-01-03T13:09:03|2021-01-03T13:09:04|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42510198|2021-01-03T14:23:50|2021-01-03T14:23:52|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42522331|2021-01-03T23:01:51|2021-01-03T23:02:54|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42522575|2021-01-03T23:32:30|2021-01-03T23:32:36|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42530642|2021-01-04T09:03:41|2021-01-04T09:05:41|1|1|4Gn|cpubase_bycore_b1|TIMEOUT|
42533676|2021-01-04T12:25:07|2021-01-04T12:25:09|1|1|4Gn|cpubase_bycore_b1|TIMEOUT|
42536007|2021-01-04T13:43:10|2021-01-04T13:43:11|1|1|4Gn|cpubase_bycore_b1|COMPLETED|
42540541|2021-01-04T16:37:45|2021-01-04T16:38:37|1|1|4Gn|cpubase_bycore_b1|FAILED|
42549102|2021-01-04T22:40:09|2021-01-04T22:40:53|1|1|8Gn|cpubase_bycore_b1|COMPLETED|
[jdesjard@gra-login2 ~]$
```

Python virtual environment with Pandas and Plotly

```
$ module load python/3.7.4
$ virtualenv --no-download env
$ source env/bin/activate
$ pip install pandas --no-index
$ pip install plotly --no-index
```

Job_summary.py script for plotting job record scatter plots

```python
import pandas as pd
import plotly.express as px
import plotly.graph_objects as go

job_frame = pd.read_csv('test_jobs.out',sep='|',header=(0),quotechar="\"")

job_frame['Submit'] = pd.to_datetime(job_frame['Submit'], errors='coerce')
job_frame['Start'] = pd.to_datetime(job_frame['Start'], errors='coerce')
job_frame['End'] = pd.to_datetime(job_frame['End'], errors='coerce')

job_frame['submit_wait'] = (job_frame['Start'] - job_frame['Submit']) / 3600
print(job_frame)
fig = px.scatter(job_frame,
                       x='Submit',
                       y='submit_wait',
                       marginal_x="histogram",
                       marginal_y="histogram",
                       hover_data=job_frame.columns,
                       color="Partition",
                       opacity=.3)
fig.write_html('test.html')
```

Demonstration

- Login to Graham using Visual Studio Code
- Create a Python virtual environment in a demo folder
- SSHFS mount the demo folder to the local file system
- SSH to the login node from a terminal and monitor top
- Generate job record tables and redirect them to text files
- Read text files into Python as Pandas dataframe
- Create interactive summary plots using Plotly package
- View HTML figures in the local browser via the SSHFS mount
- View HTML figures in a browser at the Graham VDI node via TigerVNC
- Replicate the process of generating the figures from an interactive job

Documentation

- Using Python on Compute Canada systems
  - https://docs.computecanada.ca/wiki/Python
- Visual Studio Code
  - Home: https://code.visualstudio.com/
  - Remote-SSH: https://code.visualstudio.com/docs/remote/ssh
- Plotly Python graphing library
  - https://plotly.com/python/
- Use local software with remote files systems via SSHFS
  - https://github.com/libfuse/sshfs
  - https://docs.computecanada.ca/wiki/Storage_and_file_management
- Graphical Interface to Graham Virtual Display Infrastructure (VDI) node
  - https://docs.computecanada.ca/wiki/VNC
- Getting help
  - support@computecanada.ca

Things to consider moving forward

- Plotly has a large range of figures:
  - https://plotly.com/python/
- Submitting large scale summaries as batch jobs
  - https://docs.computecanada.ca/wiki/Running_jobs

Thank you for your attention!

SHARCNET™