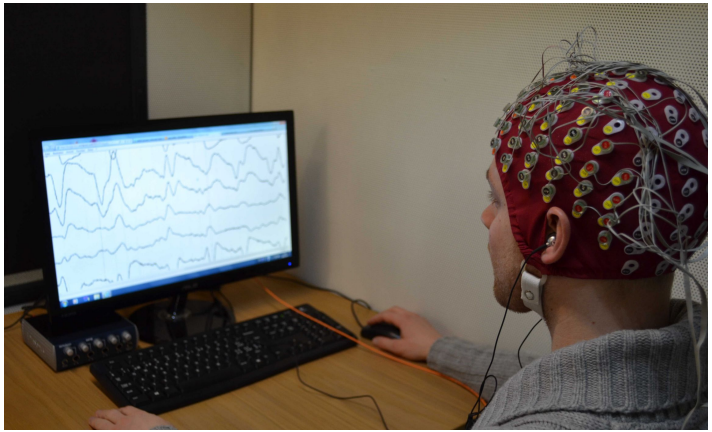
Using Python and Fabric for analyzing brain signals on OSG connect

Scott Cole
Neurosciences Graduate Program
UC San Diego

Outline

- Brain rhythms
- Running python on OSG

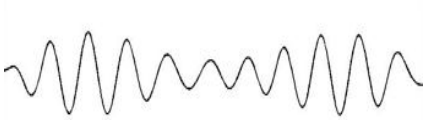
Large-scale brain recordings

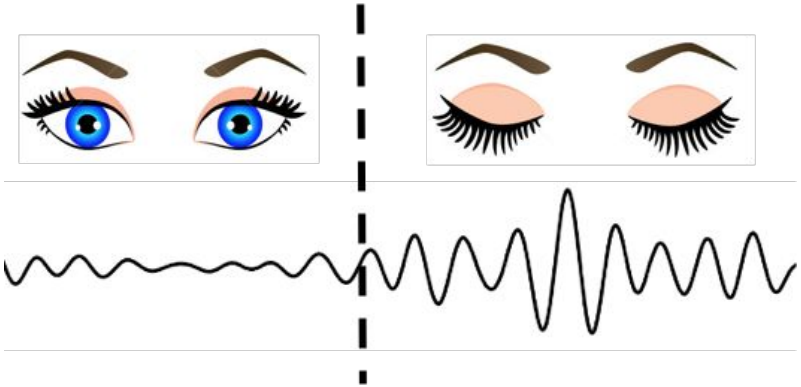


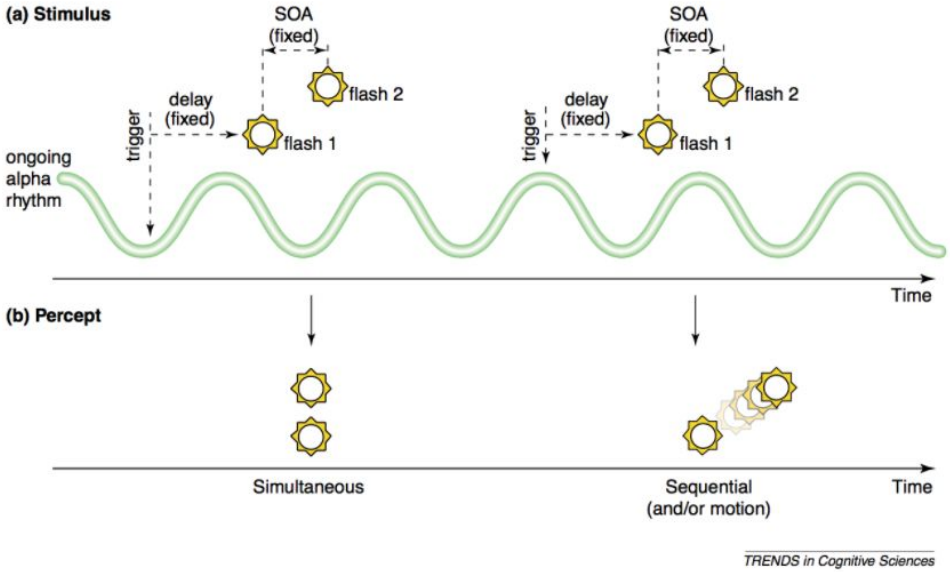
- Healthy humans
- Patients
- Animals

- Sampling: 500-30000 Hz
- Duration: 1 hour - 1 week
- Channels: 1 - 250+
- Several GB per subject

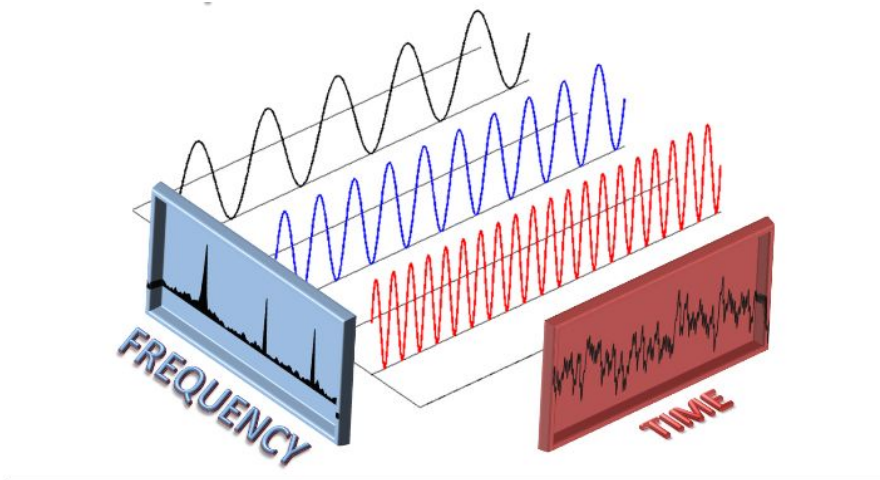
Neural oscillations (a.k.a. Brain rhythms)

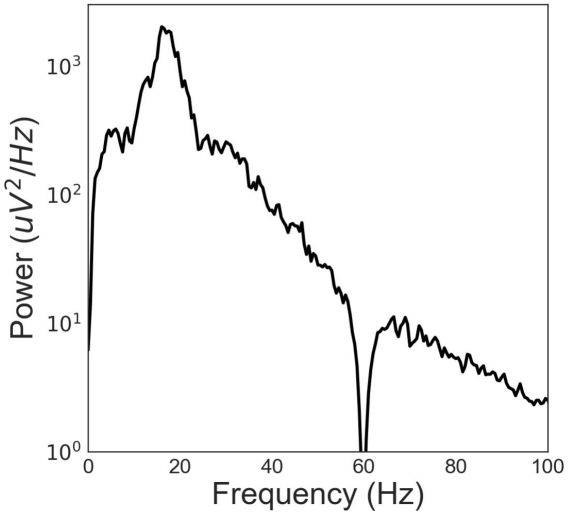




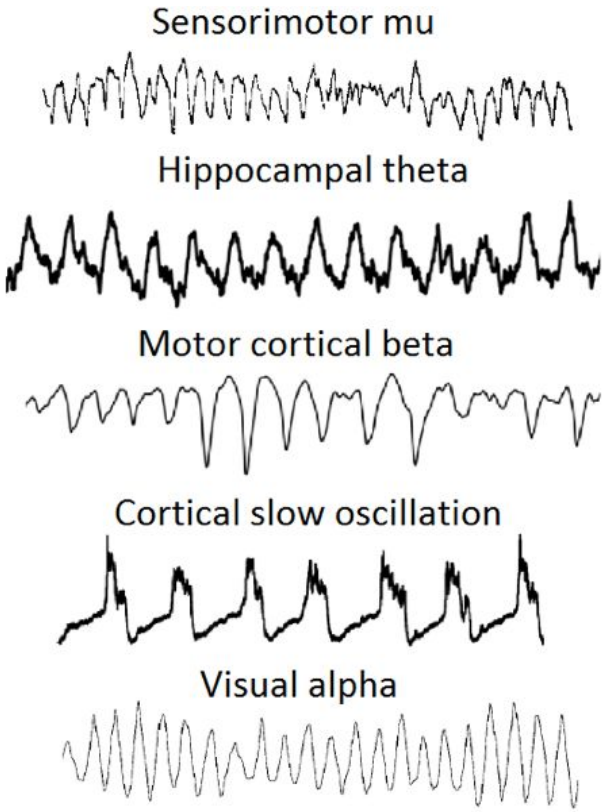


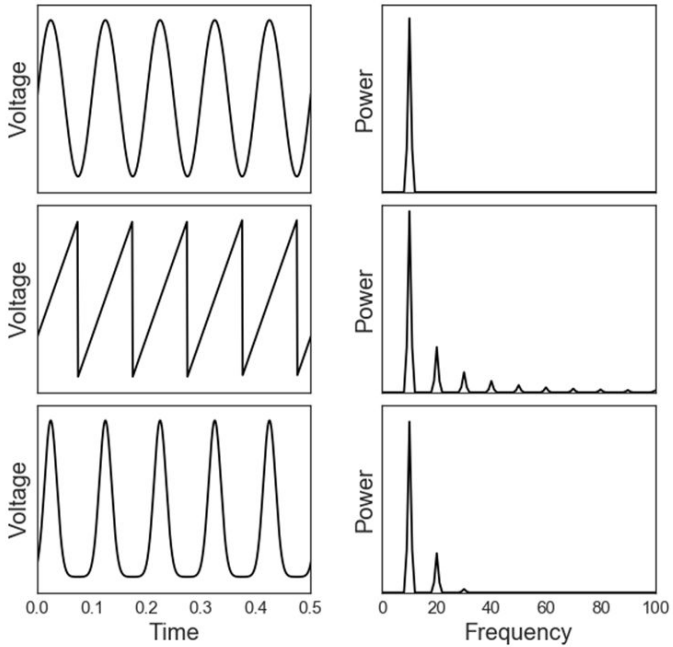
Fourier Transform-based analysis





Fourier Transform-based analysis

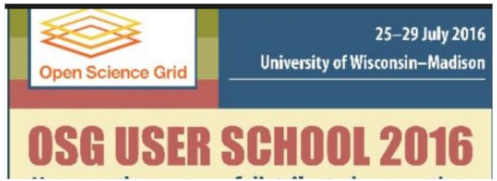




Custom analysis in Python

shape.py

```
def pt_sharp(x, Ps, Ts, window_half, method='diff'):
    """
    Calculate the sharpness of extrema

    Parameters
    ----------
    x : array-like 1d
        voltage time series
    Ps : array-like 1d
        time points of oscillatory peaks
    Ts : array-like 1d
        time points of oscillatory troughs
    window_half : int
        Number of samples in each direction around extrema to use for sharpness estimation

    Returns
    -------
    Psharps : array-like 1d
        sharpness of peaks
    Tsharps : array-like 1d
        sharpness of troughs

    """

    # Assure input has the same number of peaks and troughs
    if len(Ts) != len(Ps):
        raise ValueError('Length of peaks and troughs arrays must be equal')

    # Calculate the sharpness of each peak
    P = len(Ps)
    Psharps = np.zeros(P)
    for e in range(P):
        if method == 'deriv':
            Edata = x[Ps[e]-window_half: Ps[e]+window_half+1]
            Psharps[e] = np.mean(np.abs(np.diff(Edata)))
        elif method == 'diff':
            Psharps[e] = np.mean((x[Ps[e]]-x[Ps[e]-window_half],x[Ps[e]]-x[Ps[e]+window_half]))
```



osg connect

Free supercomputing for research: A tutorial on using Python on the Open Science Grid

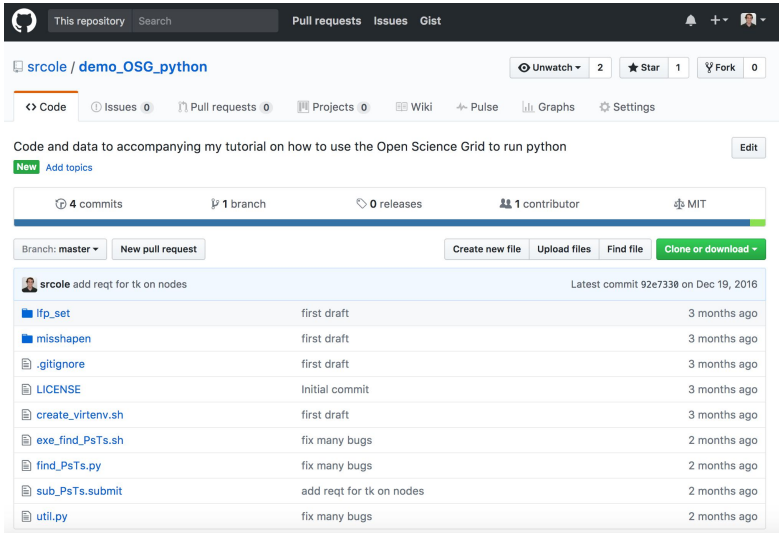
Jan 3, 2017

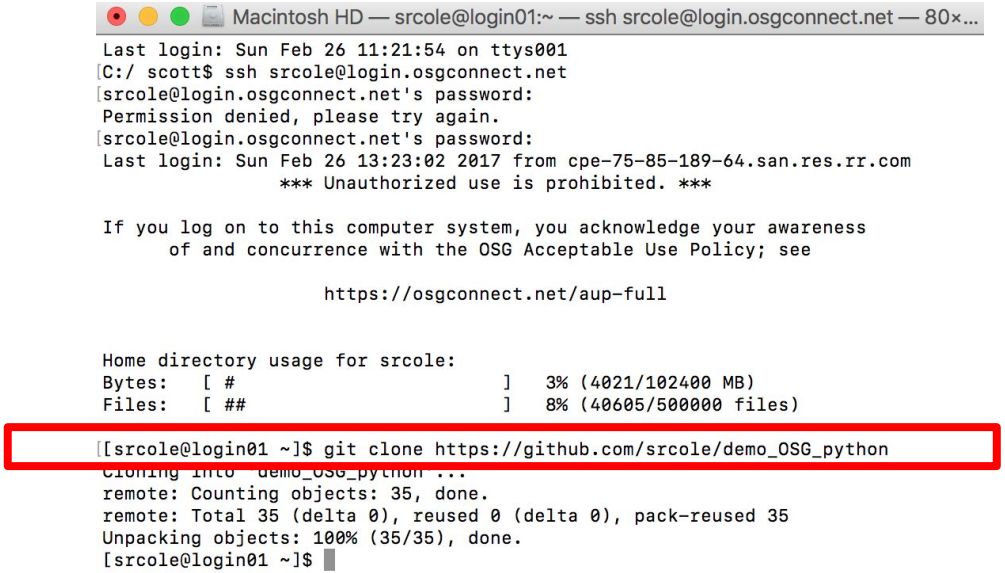
Supercomputing resources typically cost money, but the Open Science Grid (OSG) provides high-throughput computing to any researcher in the US **for free**. Briefly, OSG users can run jobs on servers owned by dozens of academic institutions, whenever those servers are not actively running a job for its owners.

Besides cost, a second major barrier to entry for those who are new to supercomputing (or, specifically, Condor) is the necessary troubleshooting before we can actually run our jobs. The purpose of this tutorial is to provide a complete example for running Python jobs on the OSG. This example is nontrivial, in that it includes multiple data sets, public libraries (e.g. scipy), private libraries, and analyzing output. In complement to this tutorial, the OSG has tutorials, a structured class, and extremely helpful online support when you get stuck.

This tutorial goes through the steps of manually connecting to and running commands on the remote server, but see the **Fabfile** section at the bottom for how this can be automated on your local machine.

1. Clone tutorial (.py, .sh, & .submit) from GitHub





2. Create python virtual environment

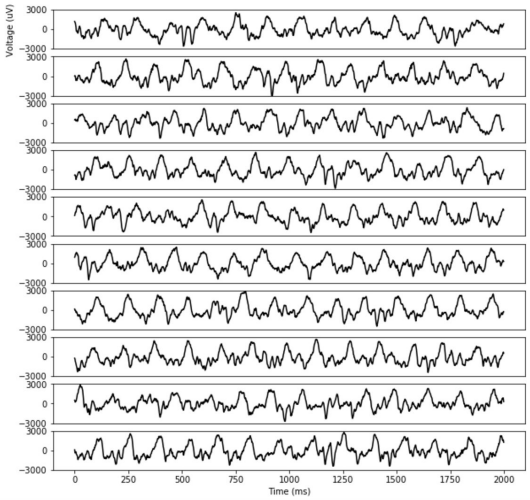
create_virtenv.sh

```
#!/bin/bash

# Load python 2.7
module load python/2.7

# Activate virtual environment
virtualenv-2.7 python_virtenv_demo
source python_virtenv_demo/bin/activate

# Install python packages (some are unused in this demo)
pip install numpy
pip install scipy
pip install scikit-learn
pip install matplotlib
pip install h5py
pip install pandas

# Deactivate virtual environment
deactivate

# Compress virtual environment
tar -cvzf python_virtenv_demo.tar.gz python_virtenv_demo
```

Macintosh HD — srcole@login01:~/demo_OSG_python — ssh srcole@login.osg...

```
Home directory usage for srcole:
Bytes:   [ #                         ]   3% (4021/102400 MB)
Files:   [ ##                        ]   8% (40605/500000 files)

[srcole@login01 ~]$ git clone https://github.com/srcole/demo_OSG_python
Cloning into 'demo_OSG_python'...
remote: Counting objects: 35, done.
remote: Total 35 (delta 0), reused 0 (delta 0), pack-reused 35
Unpacking objects: 100% (35/35), done.
[srcole@login01 ~]$ cd demo_OSG_python/
[srcole@login01 demo_OSG_python]$ chmod +x create_virtenv.sh
[srcole@login01 demo_OSG_python]$ ./create_virtenv.sh
New python executable in python_virtenv_demo/bin/python
Installing setuptools, pip...done.
You are using pip version 6.0.8, however version 9.0.1 is available.
You should consider upgrading via the 'pip install --upgrade pip' command.
Collecting numpy
  Using cached numpy-1.12.0.zip
    Running from numpy source directory.
Installing collected packages: numpy
  Running setup.py install for numpy
```

3. Transfer input data to Stash



Macintosh HD — srcole@login01:~/demo_OSG_python — ssh srcole@login.osg...

```
python_virtenv_demo/lib/python2.7/fnmatch.pyc
python_virtenv_demo/include/
python_virtenv_demo/include/python2.7
python_virtenv_demo/bin/
python_virtenv_demo/bin/python
python_virtenv_demo/bin/python2
python_virtenv_demo/bin/python2.7
python_virtenv_demo/bin/pip
python_virtenv_demo/bin/pip2
python_virtenv_demo/bin/pip2.7
python_virtenv_demo/bin/easy_install
python_virtenv_demo/bin/easy_install-2.7
python_virtenv_demo/bin/activate
python_virtenv_demo/bin/activate.fish
python_virtenv_demo/bin/activate_this.py
python_virtenv_demo/bin/activate.csh
python_virtenv_demo/bin/f2py
python_virtenv_demo/pip-selfcheck.json
[srcole@login01 demo_OSG_python]$ rm -R python_virtenv_demo
[srcole@login01 demo_OSG_python]$ mv lfp_set/ /stash/user/srcole/lfp_set/
mv: inter-device move failed: `lfp_set/' to `/stash/user/srcole/lfp_set/lfp_set'
; unable to remove target: Directory not empty
[srcole@login01 demo_OSG_python]$ mv lfp_set/ /stash/user/srcole/lfp_set/
[srcole@login01 demo_OSG_python]$
```

4. Compress personal libraries

shape.py

```
def pt_sharp(x, Ps, Ts, window_half, method='diff'):
    """
    Calculate the sharpness of extrema

    Parameters
    ----------
    x : array-like 1d
        voltage time series
    Ps : array-like 1d
        time points of oscillatory peaks
    Ts : array-like 1d
        time points of oscillatory troughs
    window_half : int
        Number of samples in each direction around extrema to use for sharpness estimation

    Returns
    -------
    Psharps : array-like 1d
        sharpness of peaks
    Tsharps : array-like 1d
        sharpness of troughs

    """

    # Assure input has the same number of peaks and troughs
    if len(Ts) != len(Ps):
        raise ValueError('Length of peaks and troughs arrays must be equal')

    # Calculate the sharpness of each peak
    P = len(Ps)
    Psharps = np.zeros(P)
    for e in range(P):
        if method == 'deriv':
            Edata = x[Ps[e]-window_half: Ps[e]+window_half+1]
            Psharps[e] = np.mean(np.abs(np.diff(Edata)))
        elif method == 'diff':
            Psharps[e] = np.mean((x[Ps[e]]-x[Ps[e]-window_half],x[Ps[e]]-x[Ps[e]+window_half]))
```

Macintosh HD — srcole@login01:~/demo_OSG_python — ssh srcole@login.osg...

```
python_virtenv_demo/bin/python2.7
python_virtenv_demo/bin/pip
python_virtenv_demo/bin/pip2
python_virtenv_demo/bin/pip2.7
python_virtenv_demo/bin/easy_install
python_virtenv_demo/bin/easy_install-2.7
python_virtenv_demo/bin/activate
python_virtenv_demo/bin/activate.fish
python_virtenv_demo/bin/activate_this.py
python_virtenv_demo/bin/activate.csh
python_virtenv_demo/bin/f2py
python_virtenv_demo/pip-selfcheck.json
[srcole@login01 demo_OSG_python]$ rm -R python_virtenv_demo
[srcole@login01 demo_OSG_python]$ mv lfp_set/ /stash/user/srcole/lfp_set/
mv: inter-device move failed: `lfp_set/' to `/stash/user/srcole/lfp_set/lfp_set'
; unable to remove target: Directory not empty
[srcole@login01 demo_OSG_python]$ mv lfp_set/ /stash/user/srcole/lfp_set/
[srcole@login01 demo_OSG_python]$ tar -cvzf misshapen.tar.gz misshapen
misshapen/
misshapen/__init__.py
misshapen/nonshape.py
misshapen/shape.py
[srcole@login01 demo_OSG_python]$ rm -R misshapen
[srcole@login01 demo_OSG_python]$
```

5. Submit job

exe_find_PsTs.sh

```
#!/bin/bash
# load necessary modules
module load python/2.7
module load lapack
module load gcc
module load libgfortran
module load atlas
module load hdf5
module load stashcp

# transfer data from stashcache
stashcp /user/srcole/lfp_set/$2.npy data.npy

# untar and activate virtual environment
tar -xzf python_virtenv_demo.tar.gz
source ./python_virtenv_demo/bin/activate

# untar local library
tar -xzf misshapen.tar.gz

# Make directory for output files
mkdir out

# Run python script
./python_virtenv_demo/bin/python2.7 find_PsTs.py

# tar output file
tar -czf out.$1.$2.tar.gz out

# Remove loaded data so not copied back
rm data.npy

# deactivate virtual environment
deactivate
```

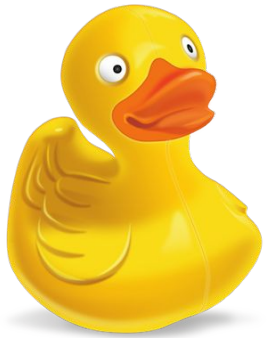
sub_PsTs.submit

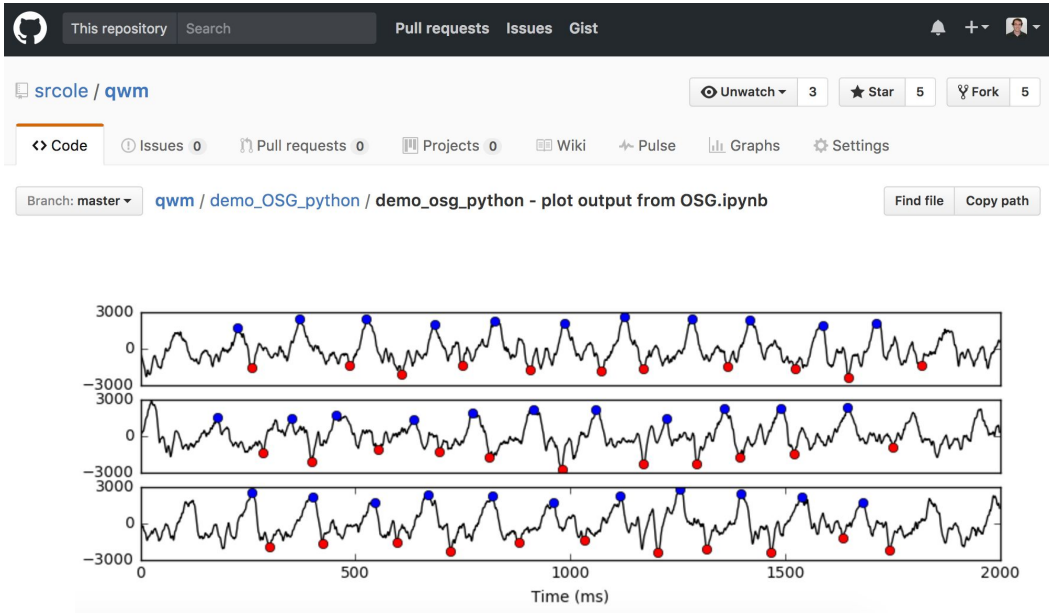
```
# Standard universe
Universe = vanilla

# Shell file we want to run
Executable = exe_find_PsTs.sh

# Files to transfer and trasnfer settings
should_transfer_files = YES
transfer_input_files = python_virtenv_demo.tar.gz,
misshapen.tar.gz, util.py, find_PsTs.py
when_to_transfer_output = ON_EXIT

# Log file save location
output        = Log/job_hc2_v1.$(Cluster).$(Process).out
error         = Log/job_hc2_v1.$(Cluster).$(Process).error
log           = Log/job_hc2_v1.$(Cluster).$(Process).log

# Specs needed on nodes
request_cpus = 1
request_memory = 1GB
request_disk = 1GB

# Arguments for the shell file we run
Arguments = $(Cluster) $(Process)

# Node requirements
+WantsStashCache = true
requirements = (HAS_MODULES == True &&
HAS_FILE_usr_lib64_atlas_libptf77blas_so_3 == True &&
HAS_FILE_usr_lib64_libtk8_5_so == True )

# Do 10 processes (0-9)
queue 10
```

Macintosh HD — srcole@login01:~/demo_OSG_python — ssh srcole@login.osg...

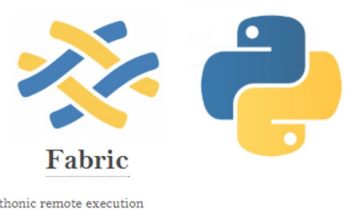
```
python_virtenv_demo/bin/easy_install
python_virtenv_demo/bin/easy_install-2.7
python_virtenv_demo/bin/activate
python_virtenv_demo/bin/activate.fish
python_virtenv_demo/bin/activate_this.py
python_virtenv_demo/bin/activate.csh
python_virtenv_demo/bin/f2py
python_virtenv_demo/pip-selfcheck.json
[srcole@login01 demo_OSG_python]$ rm -R python_virtenv_demo
[srcole@login01 demo_OSG_python]$ mv lfp_set/ /stash/user/srcole/lfp_set/
mv: inter-device move failed: `lfp_set/' to `/stash/user/srcole/lfp_set/lfp_set'
; unable to remove target: Directory not empty
[srcole@login01 demo_OSG_python]$ mv lfp_set/ /stash/user/srcole/lfp_set/
[srcole@login01 demo_OSG_python]$ tar -cvzf misshapen.tar.gz misshapen
misshapen/
misshapen/__init__.py
misshapen/nonshape.py
misshapen/shape.py
[srcole@login01 demo_OSG_python]$ rm -R misshapen
[srcole@login01 demo_OSG_python]$ mkdir Log
[srcole@login01 demo_OSG_python]$ condor_submit sub_PsTs.submit
Submitting job(s)..........
10 job(s) submitted to cluster 30296317.
[srcole@login01 demo_OSG_python]$
```

6. Transfer output and plot results





Steps 1-6 with Fabric



1. Prepare files (clone repo)
2. Create virtual environment
3. Transfer input data
4. Prepare (compress) local libraries
5. Submit job
6. Transfer output data

NOTE:
Not Python 3 compatible

fabfile.py

```python
# -*- coding: utf-8 -*-
# Load fabfile and other necessary modules
import time
from fabric.api import env, run
from fabric.context_managers import cd
from fabric.operations import get

# Declare remote host
env.hosts = ['login.osgconnect.net']

# Declare remote username and key info (optional)
with open('C:/gh/data2/username.txt','r') as myfile:
    env.user = myfile.read()
with open('C:/gh/data2/pw.txt','r') as myfile:
    env.password = myfile.read()

# Commands to execute on the remote server
def run_demo():
    run("git clone https://github.com/srcole/demo_OSG_python")
    with cd('demo_OSG_python'):
        run("chmod +x create_virtenv.sh")
        run("./create_virtenv.sh")
        run("rm -R python_virtenv_demo")
        run("mv lfp_set/ /stash/user/"+env.user+"/lfp_set/")
        run("tar -cvzf misshapen.tar.gz misshapen")
        run("rm -R misshapen")
        run("mkdir Log")
        run("condor_submit sub_PsTs.submit")
        # Need to wait until done running; should be less than 5 minutes
        time.sleep(300)
        get("./out*")
```

Summary

Fabric

Pythonic remote execution

Open Science Grid

San Diego Union-Tribune

Comme

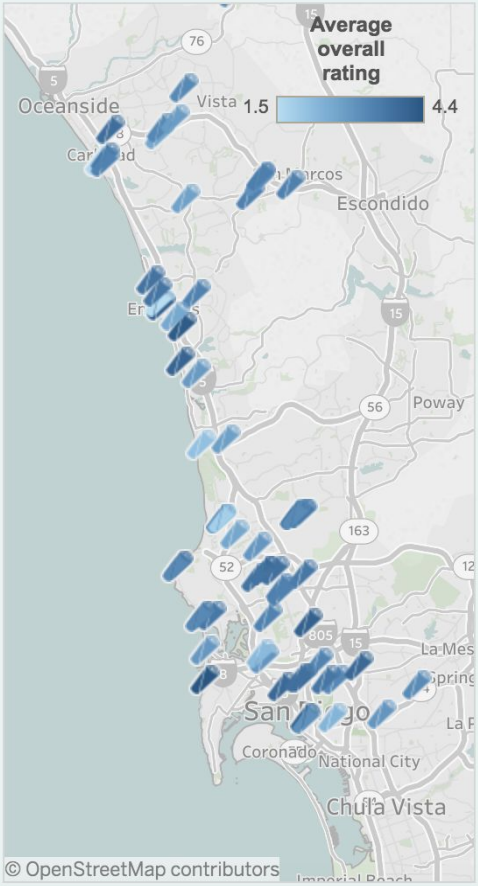
										J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
										Cost	Hunger	Length	Circum	Volume	Tortilla	Temp	Meat	Fillings	Meat:filling	Uniformity	Salsa	Synergy	Wrap	overall
										6	2	23	23.5	1.01	5	4.5	5	5	5	5		5	4	5
										7.99	3	25.5	23	1.07	4	4	4	3.5	2.5	3.5	4	3.5	2.5	3.7
										6.99	2.5	21.5	22.125	0.84	3.5	4	3.5	3.5	3	1	4	4	4.5	3.8
										5.55	3.5	17.5	23	0.74	4	4.5	3	3	5	3	4	3	5	3.5
										5.55	2.5	20.5	22.5	0.83	4.5	4	3.5	4	3.5	4	3	4	5	3.9
										7.25	3.5	18.5	22.5	0.75	1.5	1	2	2	2	2	2	2	1	1.8
										7.14	3.5	20	23	0.84	4	2	4	5	4	2	4	4.5	5	4.4
										7.14	5	20.5	24	0.94	4	2	4.5	4	3.5	2	3	4	4	4.1
										6.02	3.2	19.5	20.5	0.65	4.5	2	3.3	3.4	4.5	4.5	3.8	3.7	5	3.7
										6.5	2	18.5	21.5	0.68	1.5	4	2	2	4.5	1	1	2	5	1.5
										5.4	3	18	20	0.57	2	3.5	4	2.5	3.5	1.5	1.5	3	5	2.7
										5.99	4	20.5	23.5	0.9	4.5	5	3	4	2	2	2	3.75	4	3.2
										5.5	3	20.75	23.5	0.91	3.5	4	3.5	3	3.5	3.5	2.5	3	2.5	3
										6.25	4	15	20.5	0.5	1.5	2.5	3	4.5	4	2	2.75	3.75	3	2.6
										8.95	3.5	18.5	25	0.92	3	3	4	4	3.5	3.5	2.5	4	0.5	3
										3.5	1	16	20	0.51	3.5	4.5		3		4	3	2.5	4	2.7
										11.95	4	18	23.5	0.79	3.5	4	4	4.5	2.5	2.5	3	4	2	4.2
										11.95	3	20	22.75	0.82	3	4	4	4	3	2.5	4.5	4	1.5	4
										7.5	4				3	5	3.5	3.5	4	4	2.5	3	5	3.5
										4.87	2	19	23.5	0.83	3	4	1.5	2	3	5	3	2.5	5	2.6
										4.87	2.5	19.5	20.5	0.65	3.5	3	2.5	2.4	3.5	4.5	3	3	1.5	2.6
										9.25	3.5	25.5	24	1.17	2	1	3	3	1		3	2	4	2.3
										7	3	20	22	0.77	4.5	4	3.5	4	1	1	2	2	4	2.5
										6.25	2.5	18.5	22.5	0.75	3	4	3	2	1.5	1	3	1.5	3	2
142	Los Primos Mexican Food	Baja monster	6/3/2016							6	4				3	3	1	4	2	2	3	2	1	2
143	Los Primos Mexican Food	Baja monster	6/3/2016							6	3				4	3.5	4	4	2	4	4	4	3	3.8
144	Los Primos Mexican Food	Carnitas	6/3/2016							6.35	4.5	18	20.5	0.6	4	4	2.5	2	3	2	2	2	3	2
145	Los tacos	Local	6/5/2016	Oceanside	2183 Vista Way	https://ww	4	4.2		7.9	3.5	20.5	21.5	0.75	3.5	4.5	4.5	4	1.5	1.5	3	4	5	3.5
146	Los tacos	Fusion	6/5/2016							8.5	3.5	18.5	21	0.65	4	4	4	3.5	4	4	4	4.5	4.5	4.2
147	Taco Stand	Al pastor	6/6/2016							6.99	3.5	18.5	21	0.65	4	3.5	4	3.5	4	5	2.5	4	4.5	3.9
148	Taco Stand	Al pastor	6/6/2016							6.99	5				4	4	4	4	2	3	3	4	3	3.5
149	Albertacos	California	6/8/2016	San Marcos	500 W San Marco	https://ww	3.5	3.9		5.7	3.5	20.5	21.5	0.75	3.5	3	3	4	1.5	4	4	4	4	3.7
150	Albertacos	Carne asada	6/8/2016							5.25	4	21	21	0.74	3.5	3	2.8	3.5	4	4	4	3.5	5	3.2
151	El Rey Moro	California Surf	6/9/2016	Clairemont	4471 Clairemont M	https://ww	4	3.8		9.19	2.8	20.5	23.5	0.9	3.5	4.5	4	3.5	5	4.5	3	3.5	5	4
152	Tacos por favor	Carne asada	6/11/2016	Santa Monica	1408 Olympic Blv	https://ww	4	4.1	x	7.85	3	16.5	20.5	0.55	3	4.5	4.5	3.5	3.5	3.5	1	4	1	3.5
153	Tacos por favor	Al pastor	6/11/2016							6.85	4	15.5	21	0.54	3	5	4	3	5	4	3	4	3	4
154	Tacos por favor	Carne asada	6/11/2016							6.85	4	18			4	5	3.5	3	4	4	3	3	3	3.5
155	El dorado Mexican food	California	6/16/2016	Temecula	29000 Old Town F	https://ww	4.5	4		6.94	4				4	4	4	4.5	3.2	3	4.5	4.5	2	4.5
156	El dorado Mexican food	Super	6/16/2016							6.94	3.75				3.5	3.5	3.5	3.5	4	4	4.5	3.5	4.5	3.8



San Diego Burrito Ratings



Over 250 burritos were rated across San Diego (for more information, go to https://srcole.github.io/100burritos). For this map, the 10 burrito dimensions were averaged across all burritos eaten at each restaurant. Use the sliders on the right to filter the map and locate the best burritos in each dimension.



Click a burrito on the map to see the details of that restaurant below.

Restaurant information

Taco Stand

La Jolla

621 Pearl St

Yelp rating: **4.5**
burritos rated: **10**
Average cost: **$7.44**

Average ratings:
Overall burrito: 4.2
Volume (liters): **0.69**
Flavor synergy: **4.3**
Salsa: **3.5**
Tortilla: **3.8**
Temperature: **3.8**
Meat: **4.3**
Nonmeat fillings: **3.9**
Meat-to-filling ratio: **3.9**
Filling distribution: **3.8**
Wrap integrity: **4.1**

Neighborhood

(All)

This visualization was based on a template by Will Griffiths

Overall rating
1.5 — 4.4

Tortilla quality
1.0 — 5.0

Yelp rating
2.5 — 4.5

Temperature
1.0 — 5.0

Cost
$4.00 — $11.00

Meat quality
1.0 — 5.0

 # burritos rated
1 — 28

Nonmeat quality
1.5 — 5.0

Volume (liters)
0.50 — 1.20

Meat-to-filling ratio
1.0 — 5.0

Flavor Synergy
1.0 — 5.0

Filling distribution
1.0 — 5.0

Salsa quality
1.0 — 4.6

Wrap integrity
1.0 — 5.0

scott.cole0@gmail.com

Thank you

VOYTEKlab

UC San Diego

GRFP
NSF GRADUATE RESEARCH FELLOWSHIP PROGRAM

osg connect

Open Science Grid