The globusworld TOUR

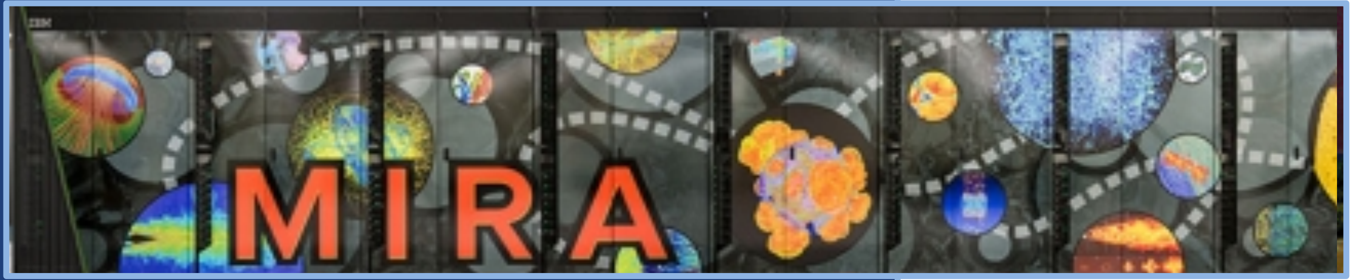
JupyterHub + Globus: A Foundation for Interactive Data Science

Vas Vasiliadis
vas@uchicago.edu

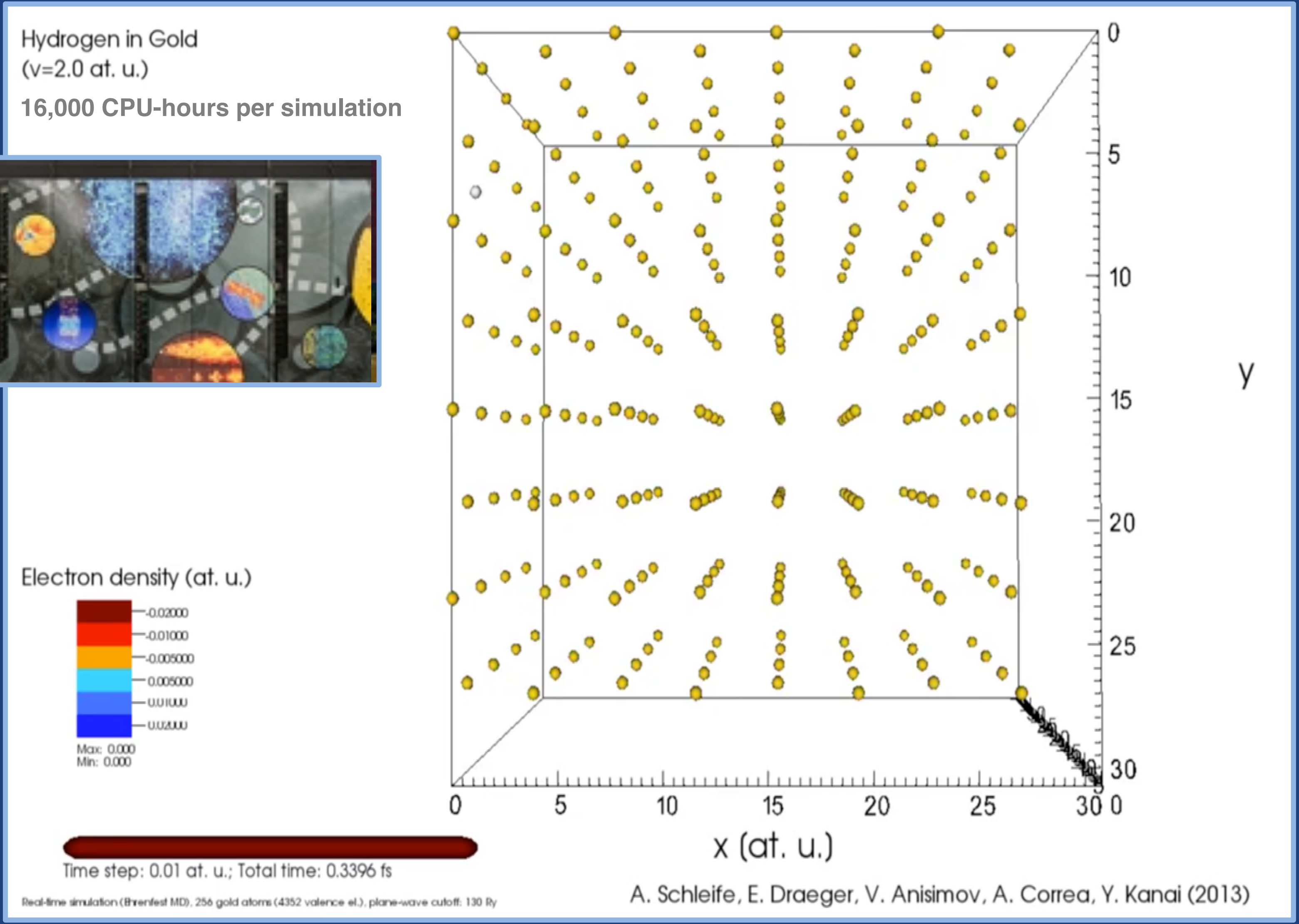
NCAR – September 5, 2018

globus

Modeling stopping power with time-dependent density functional theory



786,432 CPUs, 10 PFLOPS supercomputer
Argonne Leadership Computing Facility



Andre Schleife, UIUC

Jupyter notebooks enable rapid iteration/results

In [35]:
```python
@python_app
def get_stopping_power(lattice_vector, traj_computer):
    return traj_computer.compute_stopping_power([0,0.8,0.85], lattice_vector, 1.0, abserr=0.001,
                                                hit_threshold=2.5, full_output=1)
```

In [37]:
```python
stopping_power_results = []
for d in tqdm(dirs, desc='Submitting'):
    stopping_power_results.append(get_stopping_power(d, traj_computer))
```

Submitting 100% 24/24 [00:00<00:00, 166.06it/s]

In [38]:
```python
stopping_power_results = [s.result() for s in tqdm(stopping_power_results, desc='Waiting')]
```

Waiting 100% 24/24 [18:47:19<00:00, 2818.33s/it]

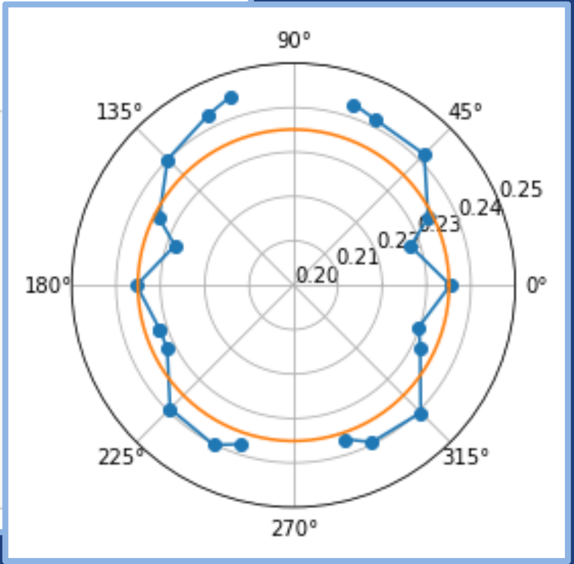
In [62]:
```python
ax = plt.subplot(111, projection='polar')
fig = plt.gcf()

ax.plot(angles + angles[:1], stopping_power + stopping_power[:1], marker='o')

# Plot the 'channel value'
ax.plot(np.linspace(0, 2*np.pi, 100), [ml_stopping_new,]*100)
ax.set_rmax(0.25)
ax.set_rmin(0.2)#min(stopping_power) * 0.99)

fig.set_size_inches(4, 4)
```

Logan Ward



But the data are big, distributed…
…and our science is collaborative

MATERIALS DATA FACILITY

materialsdatafacility.org

3.2M materials data

1 Query

jupyter

4 Share

PETREL

petrel.alcf.anl.gov

globus connect

2PB, 80Gbps store

2 Transfer

3 Learn

globus connect

Cooley: 290 TFLOPS

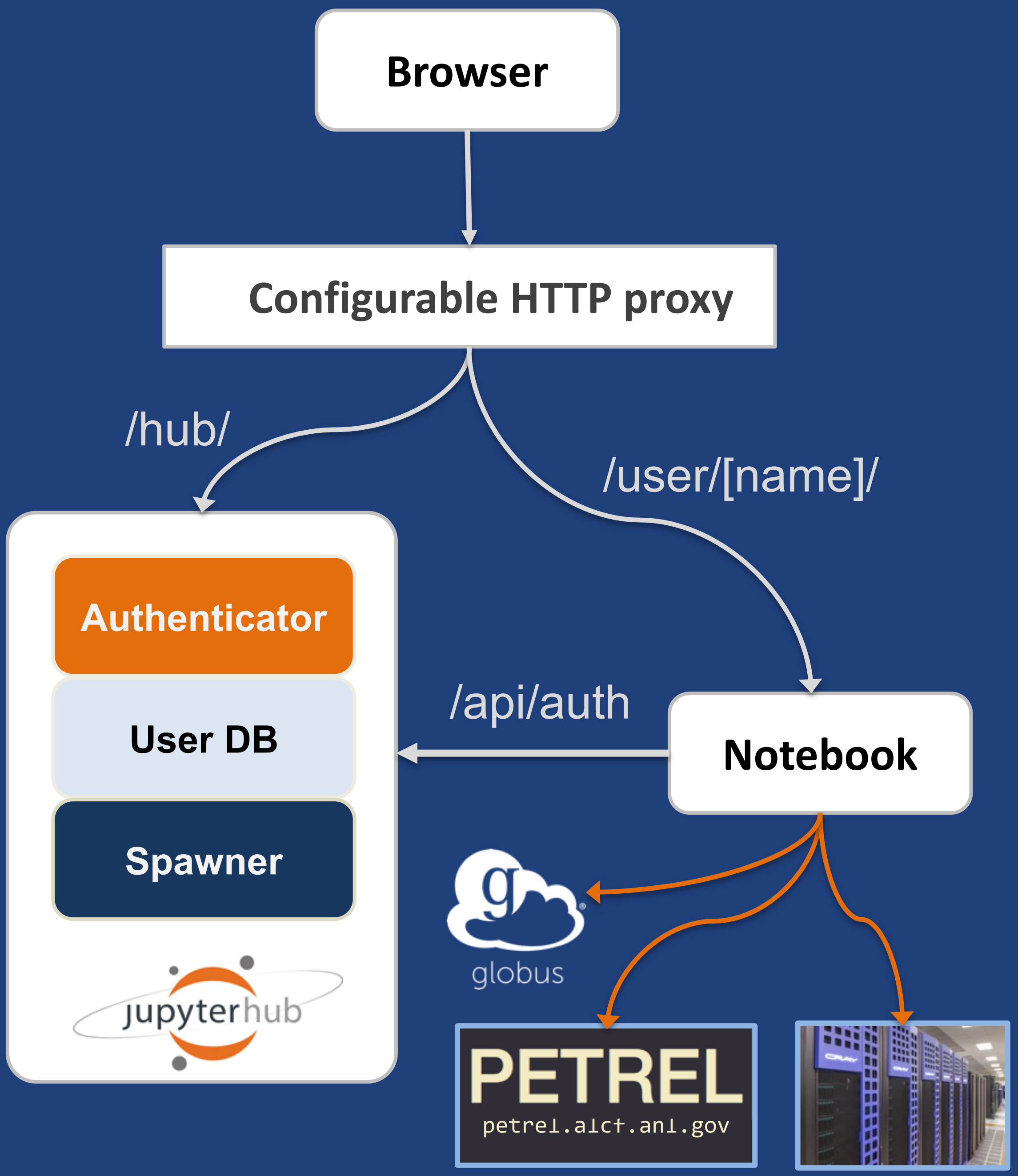
Need multi-credential, multi-service authentication and data management

JupyterHub

- **Multi-user hub**
- **Manages multiple instances of Jupyter notebook server**
- **Configurable HTTP proxy**

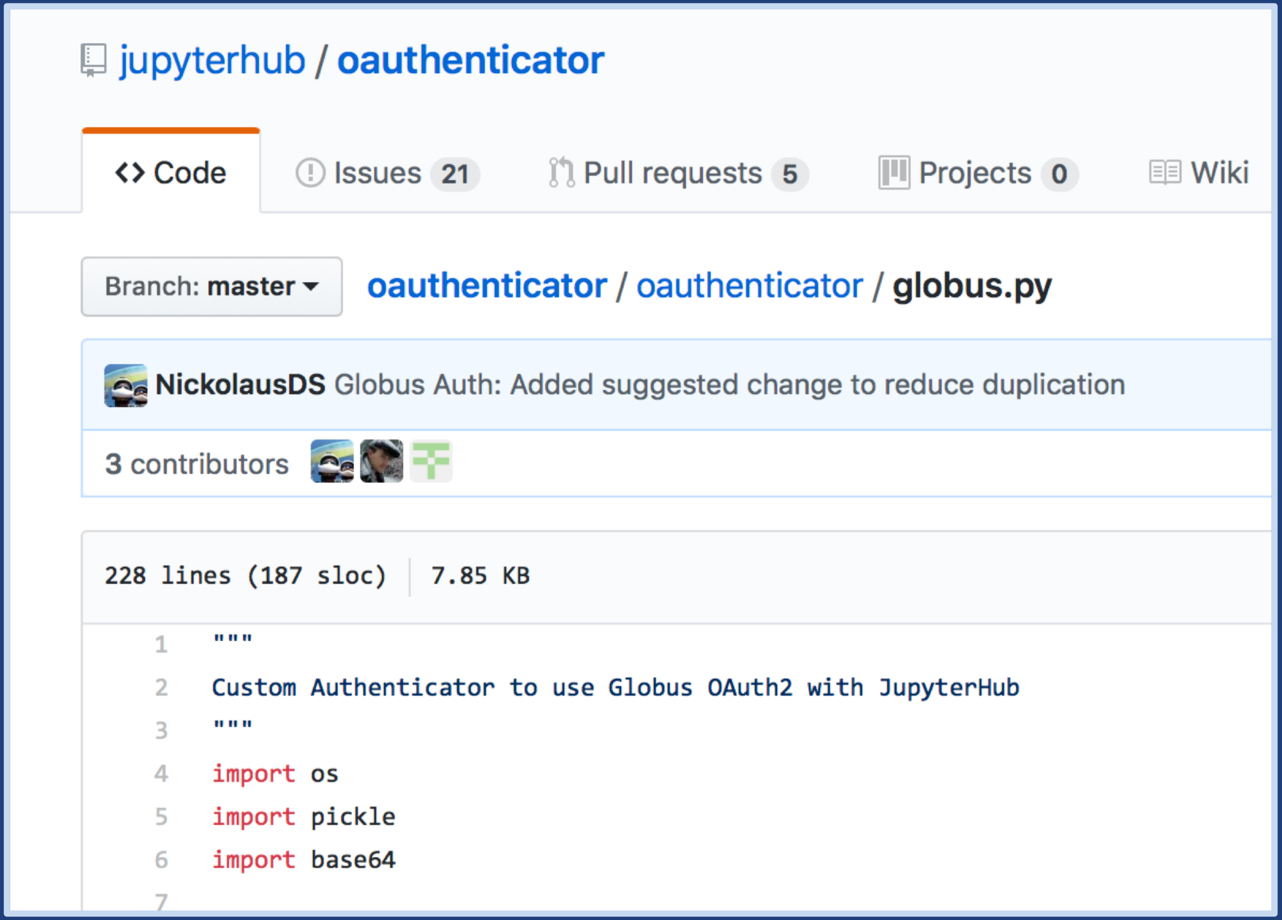
Goal: Liberate the notebook!

- **Tokens for remote services**
- **APIs for remote actions, e.g. data management via Globus service**



Securing JupyterHub with Globus Auth plugin

- Existing OAuth framework
- Can restrict IdP
- Custom scopes
- Tokens passed into notebook environment



github.com/jupyterhub/oauthenticator

Securing JupyterHub with Globus Auth

Visit https://developers.globus.org/ to set up your app. Ensure *Native App* is unchecked and make sure the callback URL looks like:

```
https://[your-host]/hub/oauth_callback
```

Set scopes for authorization and transfer. The defaults include:

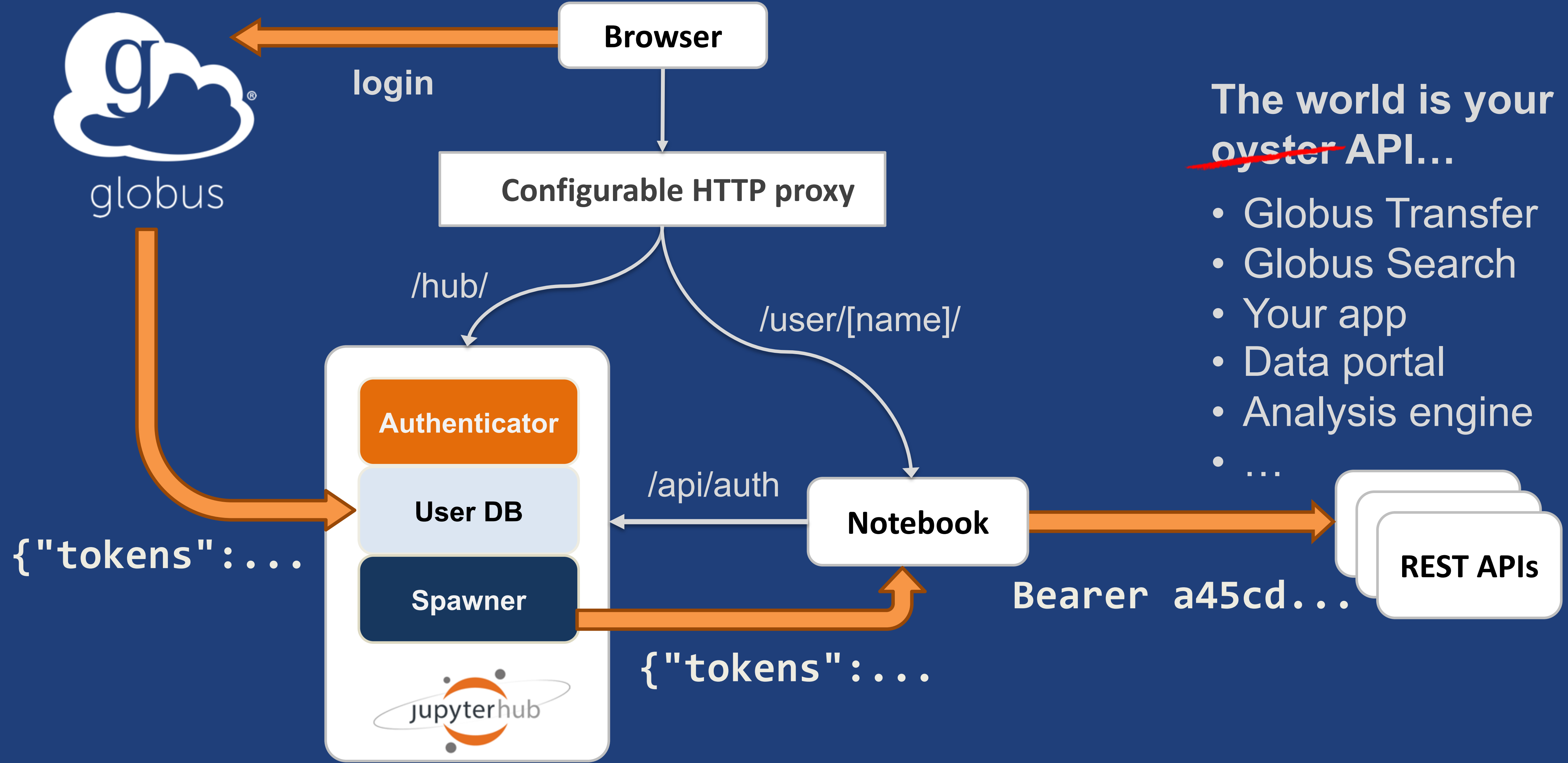
```
openid profile urn:globus:auth:scope:transfer.api.globus.org:all
```

Set the above settings in your `jupyterhub_config`:

```
# Tell JupyterHub to create system accounts
from oauthenticator.globus import LocalGlobusOAuthenticator
c.JupyterHub.authenticator_class = LocalGlobusOAuthenticator
c.LocalGlobusOAuthenticator.enable_auth_state = True
c.LocalGlobusOAuthenticator.oauth_callback_url = 'https://[your-host]/hub/oauth_callback'
c.LocalGlobusOAuthenticator.client_id = '[your app client id]'
c.LocalGlobusOAuthenticator.client_secret = '[your app client secret]'
```

github.com/jupyterhub/oauthenticator#globus-setup

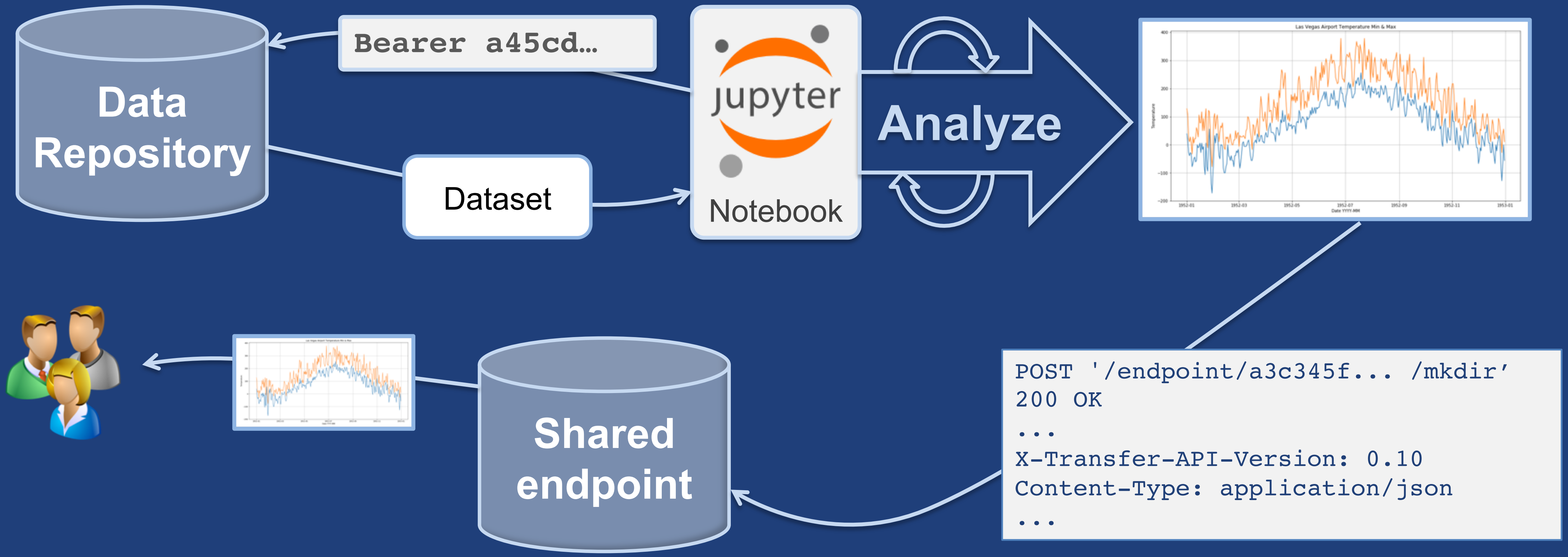
Tokens in Jupyter notebooks



The world is your ~~oyster~~ API…

- Globus Transfer
- Globus Search
- Your app
- Data portal
- Analysis engine
- …

Automated data analysis/results distribution

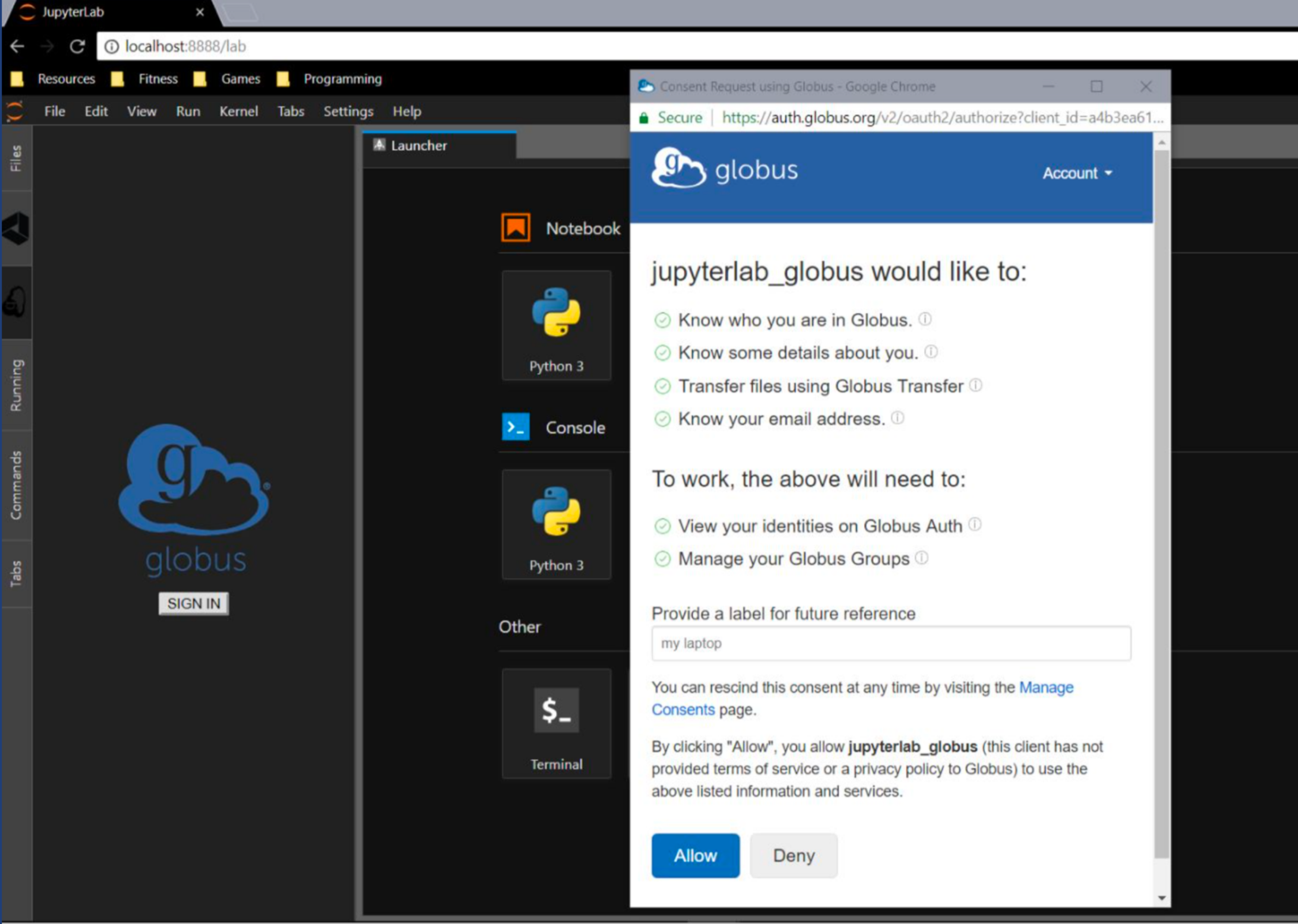


Experiment with the demo notebook

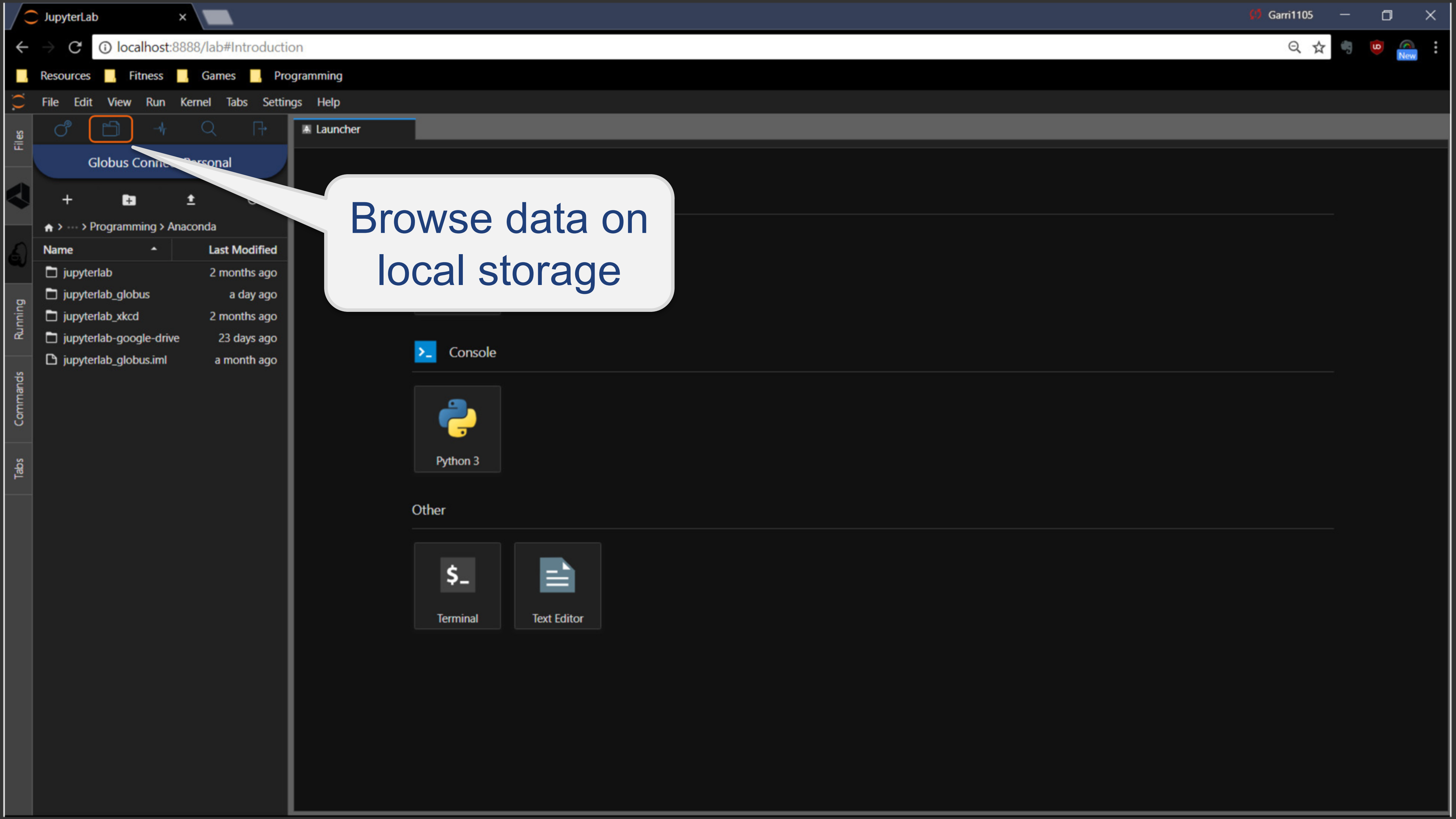
- Login into our JupyterHub*: **jupyter.demo.globus.org**
- Launch (spawn) a notebook server; get tokens
- Access Globus APIs; download some data
- "Analyze" data (generate plot)
- PUT results (graph) on an HTTPS endpoint

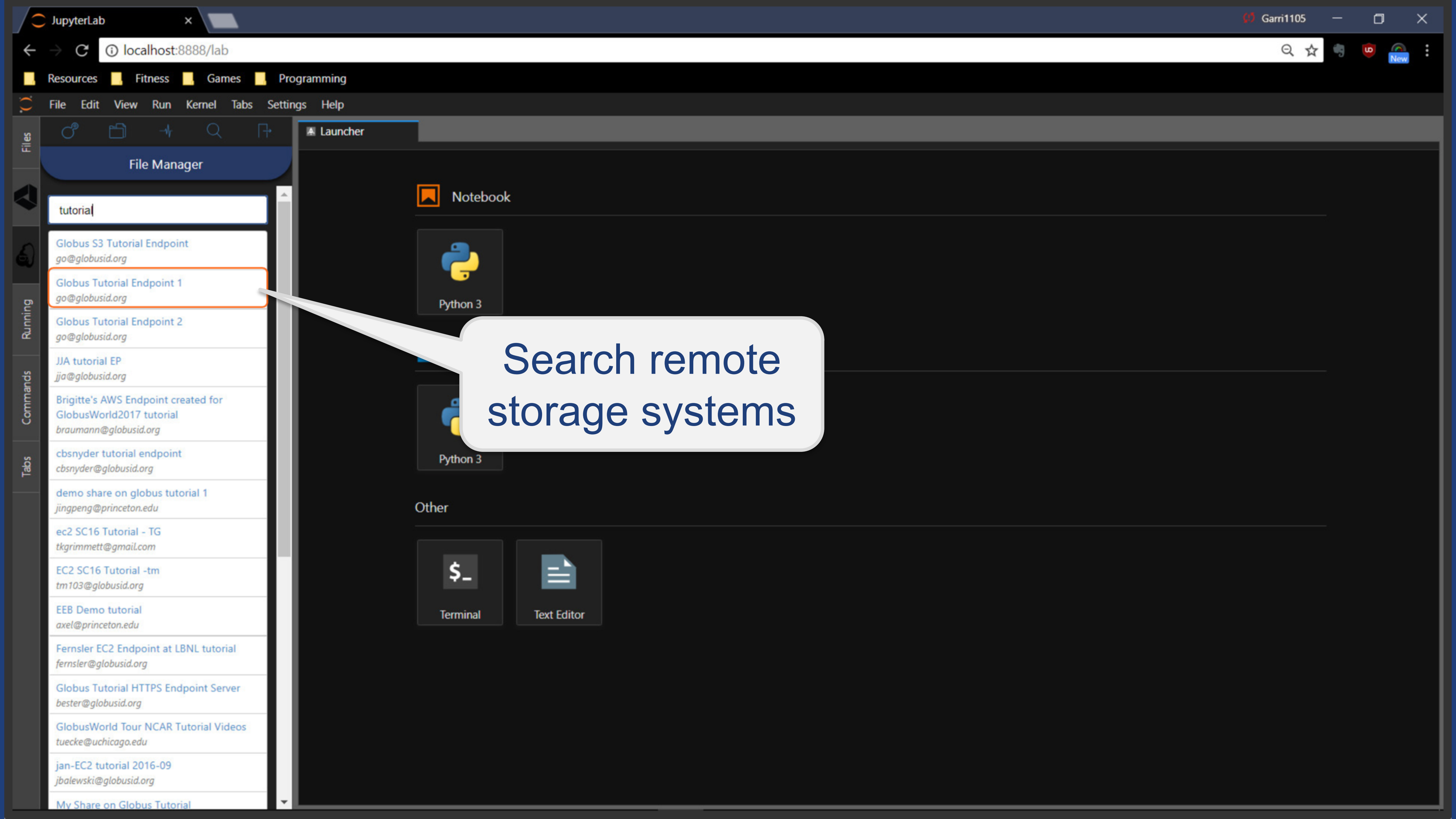
*zero-to-jupyterhub.readthedocs.io

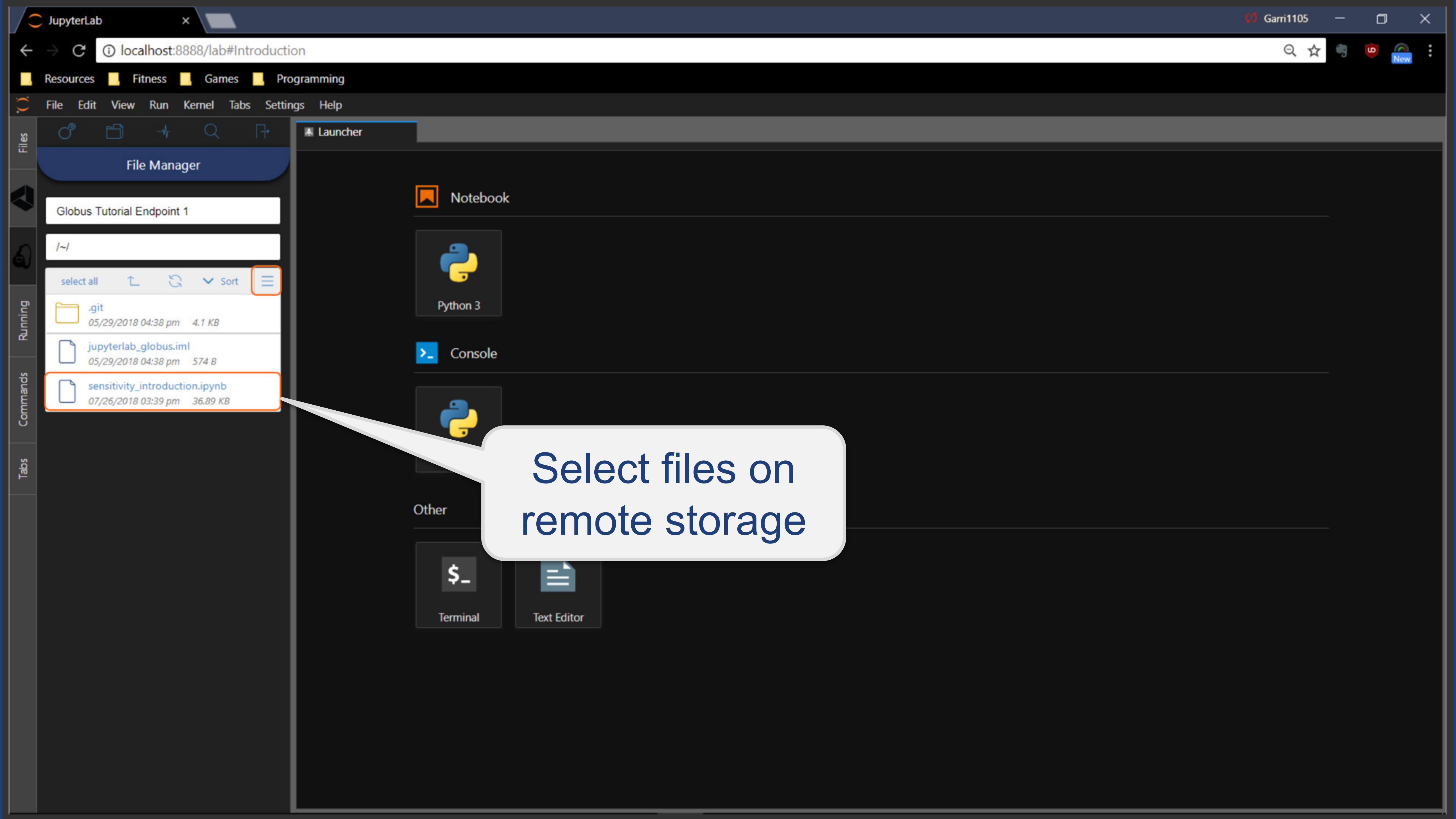
Futures…

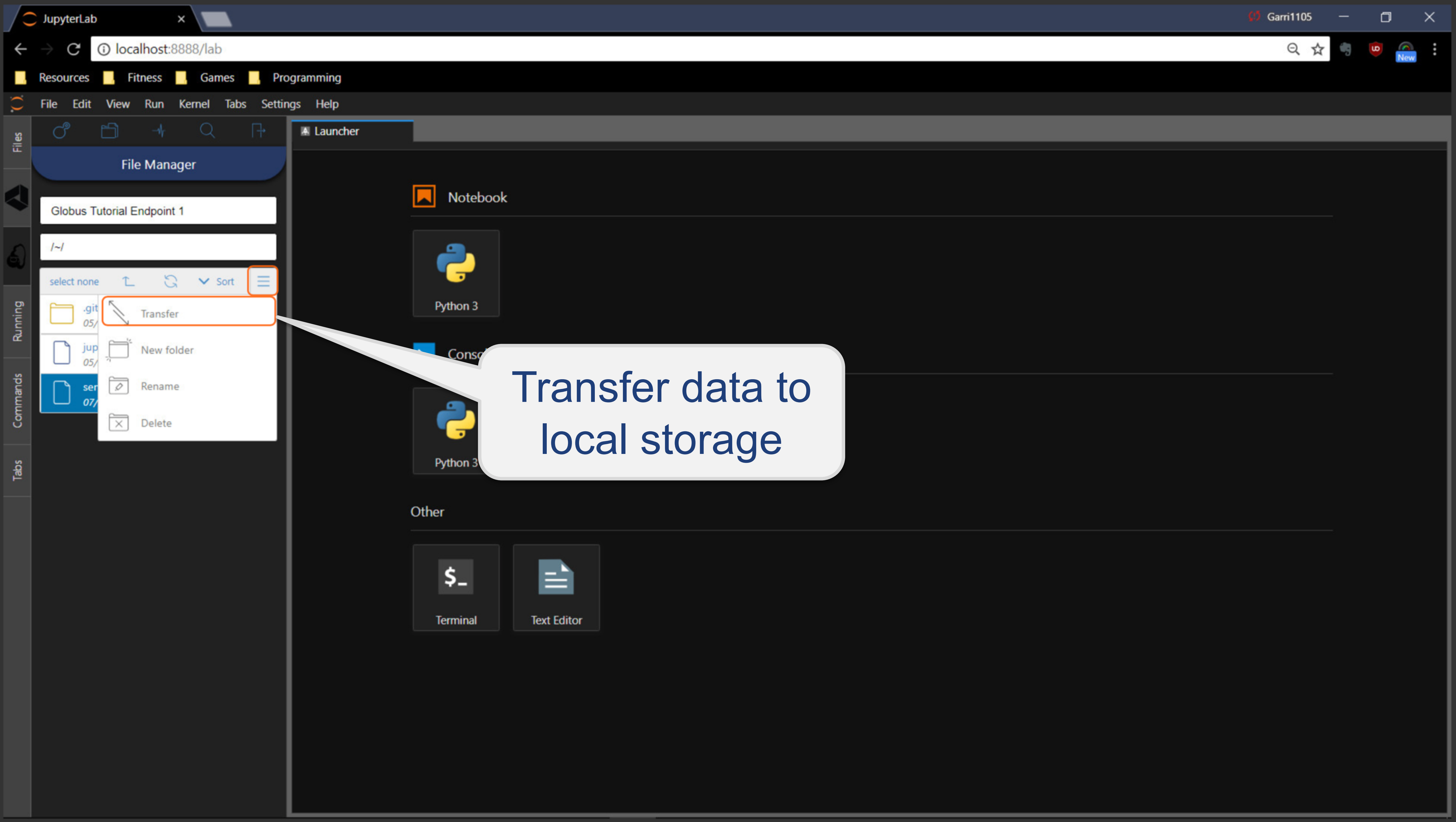


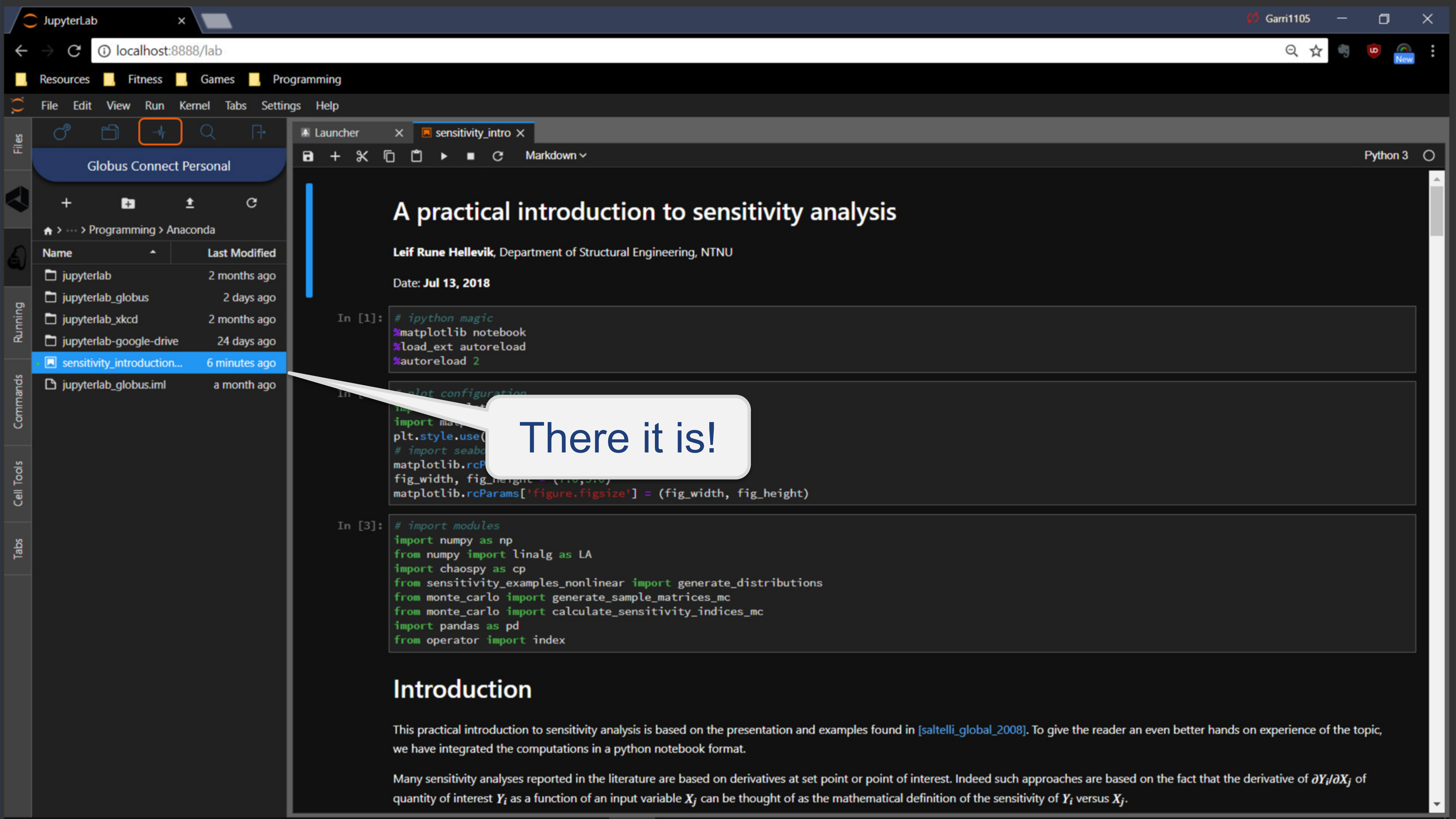
Juan David Garrido

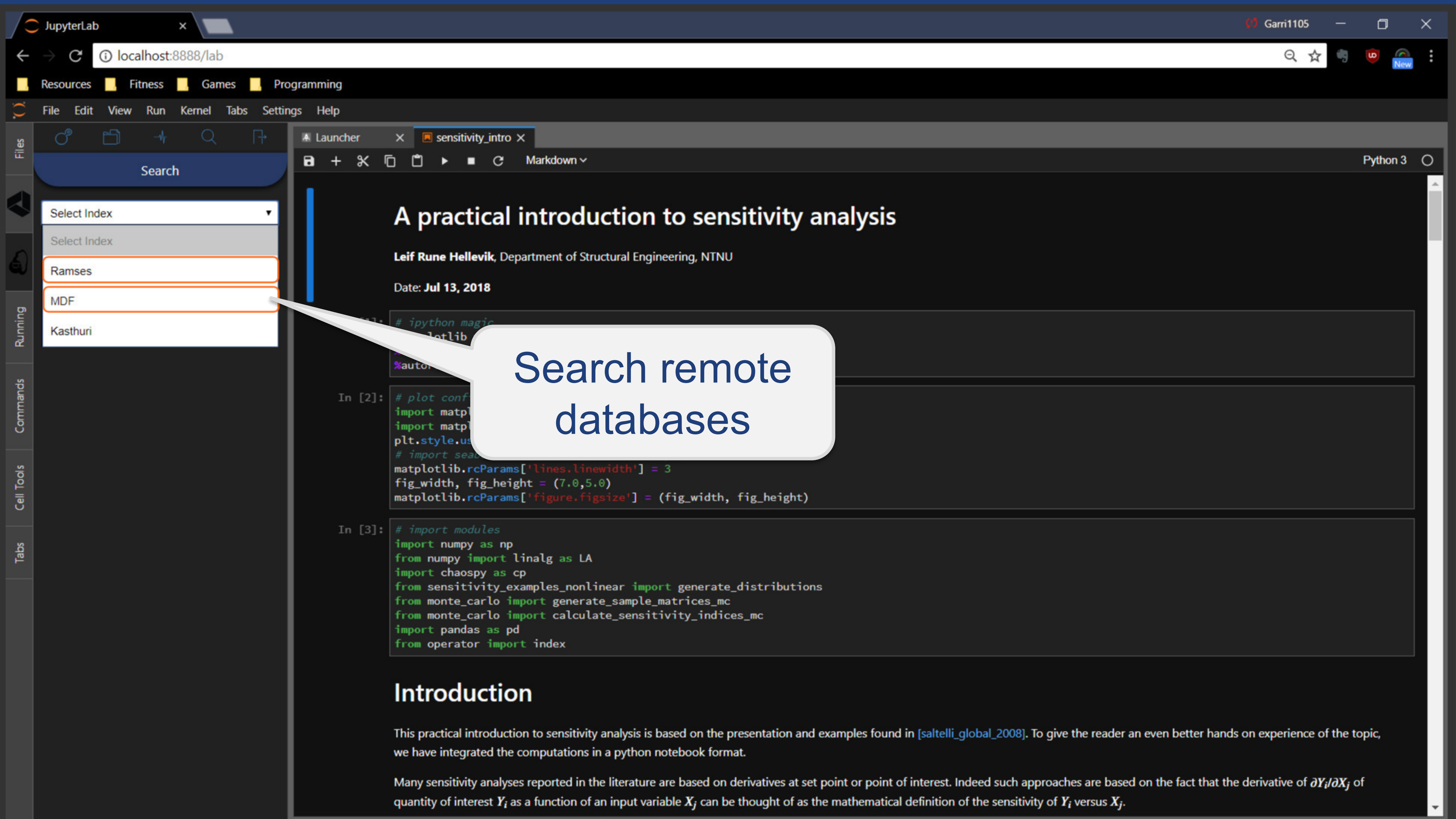

Browse data on local storage
JupyterLab
localhost:8888/lab#Introduction
Resources
Fitness
Games
Programming
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Launcher
Files
Running
Commands
Tabs
Programming › Anaconda
Name
Last Modified
jupyterlab
2 months ago
jupyterlab_globus
a day ago
jupyterlab_xkcd
2 months ago
jupyterlab-google-drive
23 days ago
jupyterlab_globus.iml
a month ago
Console
Python 3
Other
Terminal
Text Editor

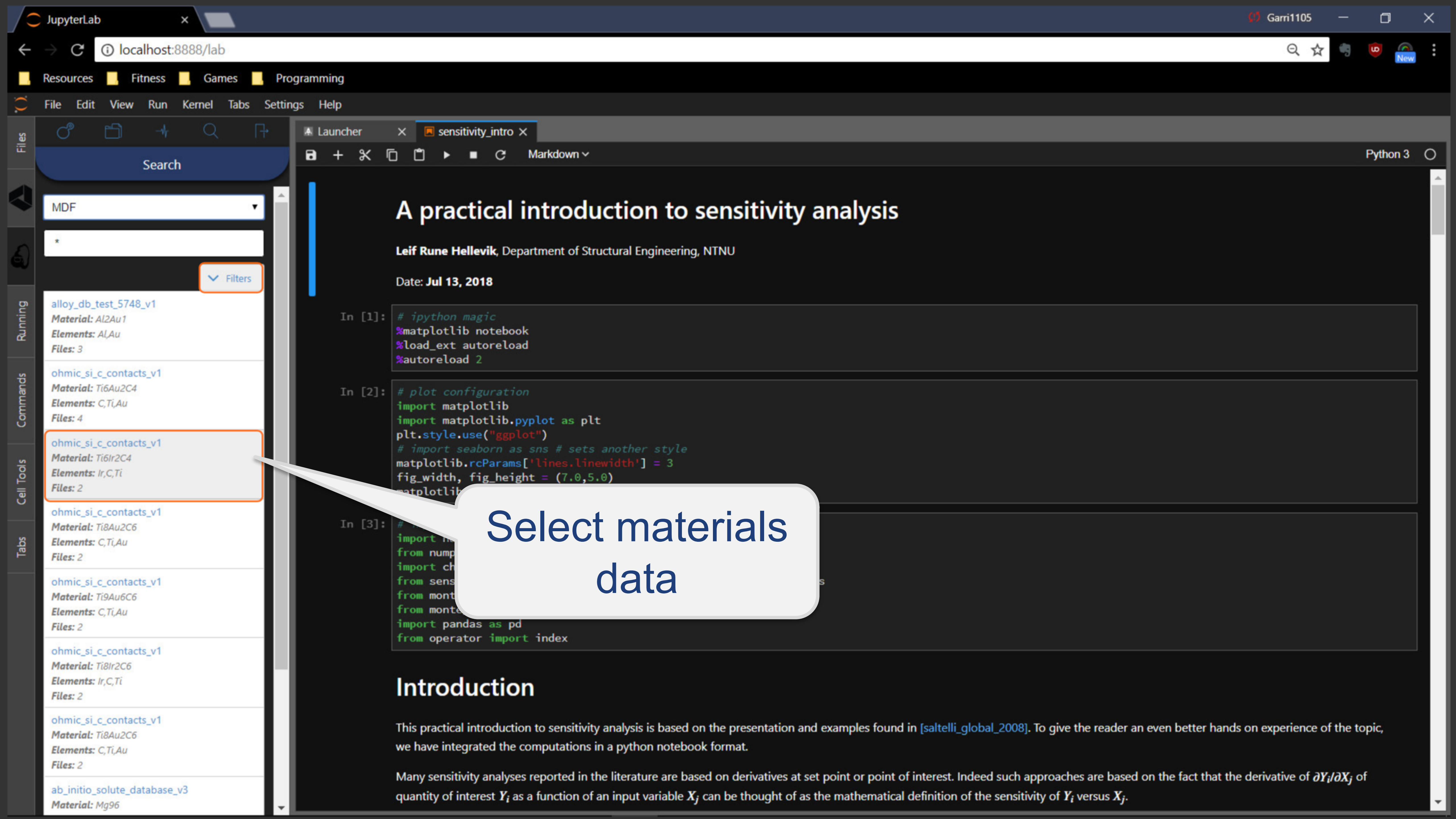

JupyterLab
Garri1105
localhost:8888/lab
Resources
Fitness
Games
Programming
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Files
Running
Commands
Tabs
File Manager
tutorial
Globus S3 Tutorial Endpoint
go@globusid.org
Globus Tutorial Endpoint 1
go@globusid.org
Globus Tutorial Endpoint 2
go@globusid.org
JJA tutorial EP
jja@globusid.org
Brigitte's AWS Endpoint created for GlobusWorld2017 tutorial
braumann@globusid.org
cbsnyder tutorial endpoint
cbsnyder@globusid.org
demo share on globus tutorial 1
jingpeng@princeton.edu
ec2 SC16 Tutorial - TG
tkgrimmett@gmail.com
EC2 SC16 Tutorial -tm
tm103@globusid.org
EEB Demo tutorial
axel@princeton.edu
Fernsler EC2 Endpoint at LBNL tutorial
fernsler@globusid.org
Globus Tutorial HTTPS Endpoint Server
bester@globusid.org
GlobusWorld Tour NCAR Tutorial Videos
tuecke@uchicago.edu
jan-EC2 tutorial 2016-09
jbalewski@globusid.org
My Share on Globus Tutorial
Launcher
Notebook
Python 3
Python 3
Other
Terminal
Text Editor
Search remote storage systems

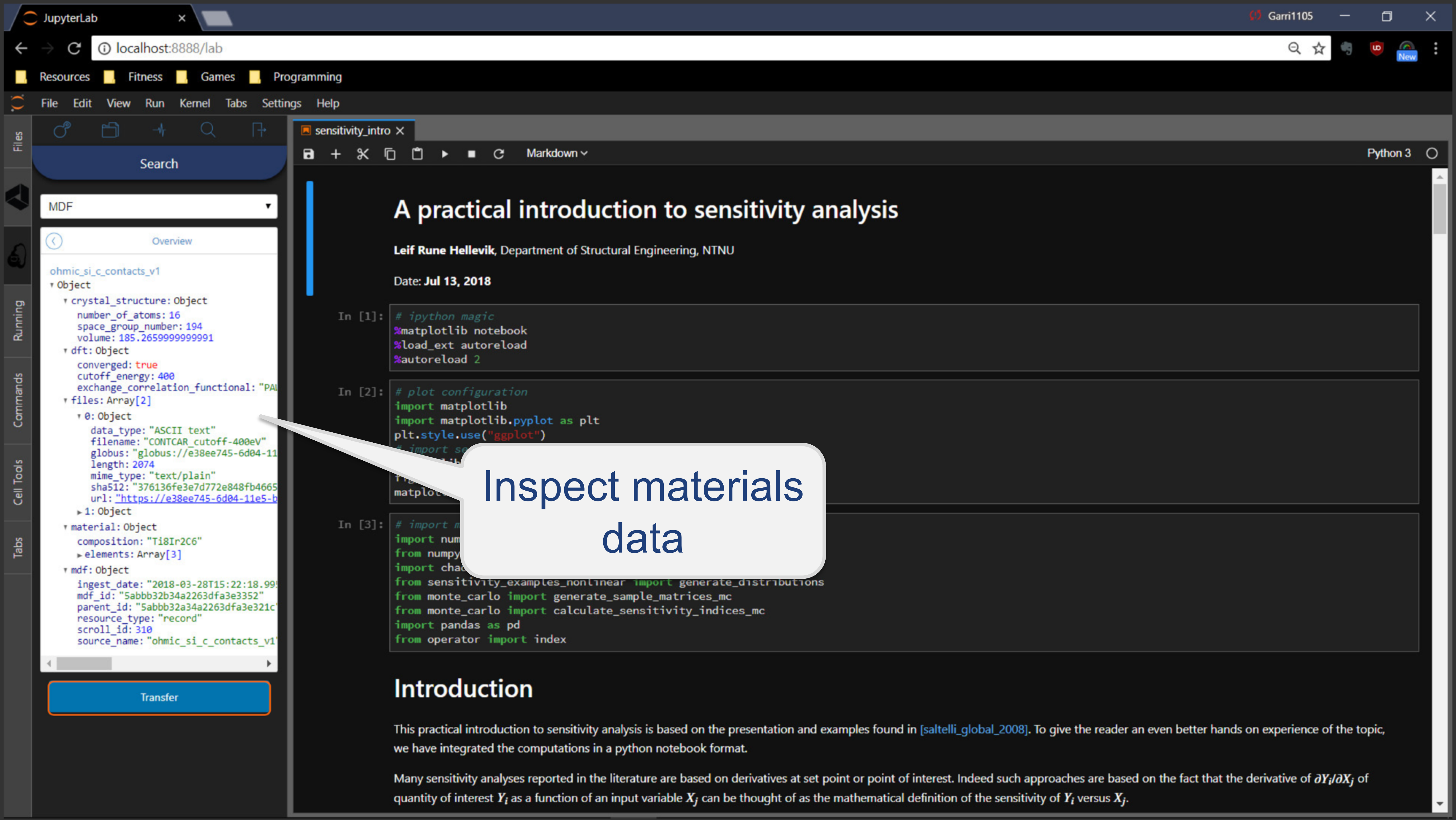

JupyterLab
localhost:8888/lab#Introduction
Resources
Fitness
Games
Programming
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Files
Running
Commands
Tabs
File Manager
Globus Tutorial Endpoint 1
/~/
select all
Sort
.git
05/29/2018 04:38 pm 4.1 KB
jupyterlab_globus.iml
05/29/2018 04:38 pm 574 B
sensitivity_introduction.ipynb
07/26/2018 03:39 pm 36.89 KB
Launcher
Notebook
Python 3
Console
Other
Terminal
Text Editor
Select files on remote storage


JupyterLab
localhost:8888/lab
Resources
Fitness
Games
Programming
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Files
Running
Commands
Tabs
File Manager
Globus Tutorial Endpoint 1
/~/
select none
Sort
Transfer
New folder
Rename
Delete
Launcher
Notebook
Python 3
Python 3
Other
Terminal
Text Editor
Transfer data to local storage


JupyterLab
localhost:8888/lab
Resources
Fitness
Games
Programming
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Files
Running
Commands
Cell Tools
Tabs
Globus Connect Personal
Programming › Anaconda
Name
Last Modified
jupyterlab
2 months ago
jupyterlab_globus
2 days ago
jupyterlab_xkcd
2 months ago
jupyterlab-google-drive
24 days ago
sensitivity_introduction...
6 minutes ago
jupyterlab_globus.iml
a month ago
Launcher
sensitivity_intro
Markdown
Python 3
A practical introduction to sensitivity analysis
Leif Rune Hellevik, Department of Structural Engineering, NTNU
Date: Jul 13, 2018
In [1]:
ipython magic
%matplotlib notebook
%load_ext autoreload
%autoreload 2
plot configuration
import
plt.style.use(
matplotlib.rcParams['figure.figsize'] = (fig_width, fig_height)
In [3]:
# import modules
import numpy as np
from numpy import linalg as LA
import chaospy as cp
from sensitivity_examples_nonlinear import generate_distributions
from monte_carlo import generate_sample_matrices_mc
from monte_carlo import calculate_sensitivity_indices_mc
import pandas as pd
from operator import index
Introduction
This practical introduction to sensitivity analysis is based on the presentation and examples found in [saltelli_global_2008]. To give the reader an even better hands on experience of the topic, we have integrated the computations in a python notebook format.
Many sensitivity analyses reported in the literature are based on derivatives at set point or point of interest. Indeed such approaches are based on the fact that the derivative of ∂Yi/∂Xj of quantity of interest Yi as a function of an input variable Xj can be thought of as the mathematical definition of the sensitivity of Yi versus Xj.
There it is!


Search remote databases
JupyterLab
localhost:8888/lab
Resources
Fitness
Games
Programming
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Search
Select Index
Ramses
MDF
Kasthuri
Files
Running
Commands
Cell Tools
Tabs
Launcher
sensitivity_intro
Markdown
Python 3
A practical introduction to sensitivity analysis
Leif Rune Hellevik, Department of Structural Engineering, NTNU
Date: Jul 13, 2018
In [2]:
matplotlib.rcParams['lines.linewidth'] = 3
fig_width, fig_height = (7.0,5.0)
matplotlib.rcParams['figure.figsize'] = (fig_width, fig_height)
In [3]:
# import modules
import numpy as np
from numpy import linalg as LA
import chaospy as cp
from sensitivity_examples_nonlinear import generate_distributions
from monte_carlo import generate_sample_matrices_mc
from monte_carlo import calculate_sensitivity_indices_mc
import pandas as pd
from operator import index
Introduction
This practical introduction to sensitivity analysis is based on the presentation and examples found in [saltelli_global_2008]. To give the reader an even better hands on experience of the topic, we have integrated the computations in a python notebook format.
Many sensitivity analyses reported in the literature are based on derivatives at set point or point of interest. Indeed such approaches are based on the fact that the derivative of ∂Yi/∂Xj of quantity of interest Yi as a function of an input variable Xj can be thought of as the mathematical definition of the sensitivity of Yi versus Xj.


JupyterLab
localhost:8888/lab
Resources
Fitness
Games
Programming
File
Edit
View
Run
Kernel
Tabs
Settings
Help
Files
Running
Commands
Cell Tools
Tabs
Search
MDF
*
Filters
alloy_db_test_5748_v1
Material: Al2Au1
Elements: Al,Au
Files: 3
ohmic_si_c_contacts_v1
Material: Ti6Au2C4
Elements: C,Ti,Au
Files: 4
ohmic_si_c_contacts_v1
Material: Ti6Ir2C4
Elements: Ir,C,Ti
Files: 2
ohmic_si_c_contacts_v1
Material: Ti8Au2C6
Elements: C,Ti,Au
Files: 2
ohmic_si_c_contacts_v1
Material: Ti9Au6C6
Elements: C,Ti,Au
Files: 2
ohmic_si_c_contacts_v1
Material: Ti8Ir2C6
Elements: Ir,C,Ti
Files: 2
ohmic_si_c_contacts_v1
Material: Ti8Au2C6
Elements: C,Ti,Au
Files: 2
ab_initio_solute_database_v3
Material: Mg96
Launcher
sensitivity_intro
Markdown
Python 3
A practical introduction to sensitivity analysis
Leif Rune Hellevik, Department of Structural Engineering, NTNU
Date: Jul 13, 2018
In [1]:
ipython magic
%matplotlib notebook
%load_ext autoreload
%autoreload 2
In [2]:
# plot configuration
import matplotlib
import matplotlib.pyplot as plt
plt.style.use("ggplot")
# import seaborn as sns # sets another style
matplotlib.rcParams['lines.linewidth'] = 3
fig_width, fig_height = (7.0,5.0)
In [3]:
import pandas as pd
from operator import index
Select materials data
Introduction
This practical introduction to sensitivity analysis is based on the presentation and examples found in [saltelli_global_2008]. To give the reader an even better hands on experience of the topic, we have integrated the computations in a python notebook format.
Many sensitivity analyses reported in the literature are based on derivatives at set point or point of interest. Indeed such approaches are based on the fact that the derivative of ∂Yi/∂Xj of quantity of interest Yi as a function of an input variable Xj can be thought of as the mathematical definition of the sensitivity of Yi versus Xj.



JupyterLab
localhost:8888/lab
Resources
Fitness
Games
Programming
File Edit View Run Kernel Tabs Settings Help
Search
MDF
Overview
Transfer
sensitivity_intro
Markdown
Python 3
A practical introduction to sensitivity analysis
Leif Rune Hellevik, Department of Structural Engineering, NTNU
Date: Jul 13, 2018
Inspect materials data
Introduction
This practical introduction to sensitivity analysis is based on the presentation and examples found in [saltelli_global_2008]. To give the reader an even better hands on experience of the topic, we have integrated the computations in a python notebook format.
Many sensitivity analyses reported in the literature are based on derivatives at set point or point of interest. Indeed such approaches are based on the fact that the derivative of ∂Yi/∂Xj of quantity of interest Yi as a function of an input variable Xj can be thought of as the mathematical definition of the sensitivity of Yi versus Xj.

Incorporate seamless parallel computing

- **(Data) science apps require…**
  - Interactivity
  - Scalability (more than a desktop)
  - Reproducibility (publish code, docs)
- **Solution: JupyterHub + Parsl**
  - Interactive computing environment
  - Notebooks for publication
  - Can run on dedicated hardware

Parsl Python parallel library

- **Tasks exposed as functions (Python, bash)**
- **Python code to glue functions together**
- **Globus for auth and data movement**

parsl-project.org

```
@python_app
def compute_features(chunk):
  for f in featurizers:
    chunk = f.featurize_dataframe(chunk, 'atoms’)
  return chunk

chunks = [compute_features(chunk)
          for chunk in np.array_split(data, chunks)]
```

Containerized data science ecosystem