JupyterHub for Atmospheric Science Research and Education on the NSF Jetstream Cloud

AMS 2019

Julien Chastang

Thursday, 10 January 2019

Outline

Background

Jupyter and Related Technologies

NSF Jetstream Cloud

A JupyterHub for Atmospheric Science

Background

Unidata 2024 Proposal: Science as a Service

The Science as a Service concept draws together Unidata's ongoing work to provide geoscience data and software for analysis and visualization with access to workflows designed to take advantage of cloud computing resources.

Award # NSF-1344155

Current Trends

- Web browser is the de facto cross-platform 'app'.
- Cloud-computing enables easy deployments of complex software at scale
  - Virtual Machines
  - Containerization
  - Software orchestration across data center
- Data are becoming unwieldy and expensive to move

Award # NSF-1344155

Opportunity for Research and Education

- Difficult to install and configure scientific software can be installed on cloud on user's behalf by experts
- Allowing researchers and student to focus on science instead of installing and configuring software

Award # NSF-1344155

Jupyter and Related Technologies

Award # NSF-1344155

What is a Jupyter Notebook?

A narrative of:

- Explanatory and expository text
- Software code (Python, R, etc.) and output
- Equations (MathJax, LaTeX)
- Figures and multimedia

Lorenz System

The Lorenz system is a series of Ordinary Differential equation studied by Edward Lorenz.

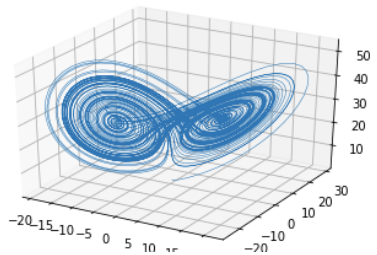
$$\dot{x} = \sigma(y - x)$$
$$\dot{y} = \rho x - y - xz$$
$$\dot{z} = -\beta z + xy$$

In [10]:
```
def lorenz(x, y, z, s=10, r=28, b=2.667):
    x_dot = s*(y - x)
    y_dot = r*x - y - x*z
    z_dot = x*y - b*z
    return x_dot, y_dot, z_dot

dt = 0.01; stepCnt = 10000
xs = np.empty((stepCnt + 1,))
ys = np.empty((stepCnt + 1,))
zs = np.empty((stepCnt + 1,))
xs[0], ys[0], zs[0] = (0., 1., 1.05)

for i in range(stepCnt):
    x_dot, y_dot, z_dot = lorenz(xs[i], ys[i], zs[i])
    xs[i + 1] = xs[i] + (x_dot * dt)
    ys[i + 1] = ys[i] + (y_dot * dt)
    zs[i + 1] = zs[i] + (z_dot * dt)

fig = plt.figure()
ax = fig.gca(projection='3d')
ax.plot(xs, ys, zs, lw=0.5)
plt.show()
```



Award # NSF-1344155

Success of Jupyter in Research and Education

nature
International journal of science

TOOLBOX · 30 OCTOBER 2018

Why Jupyter is data scientists' computational notebook of choice

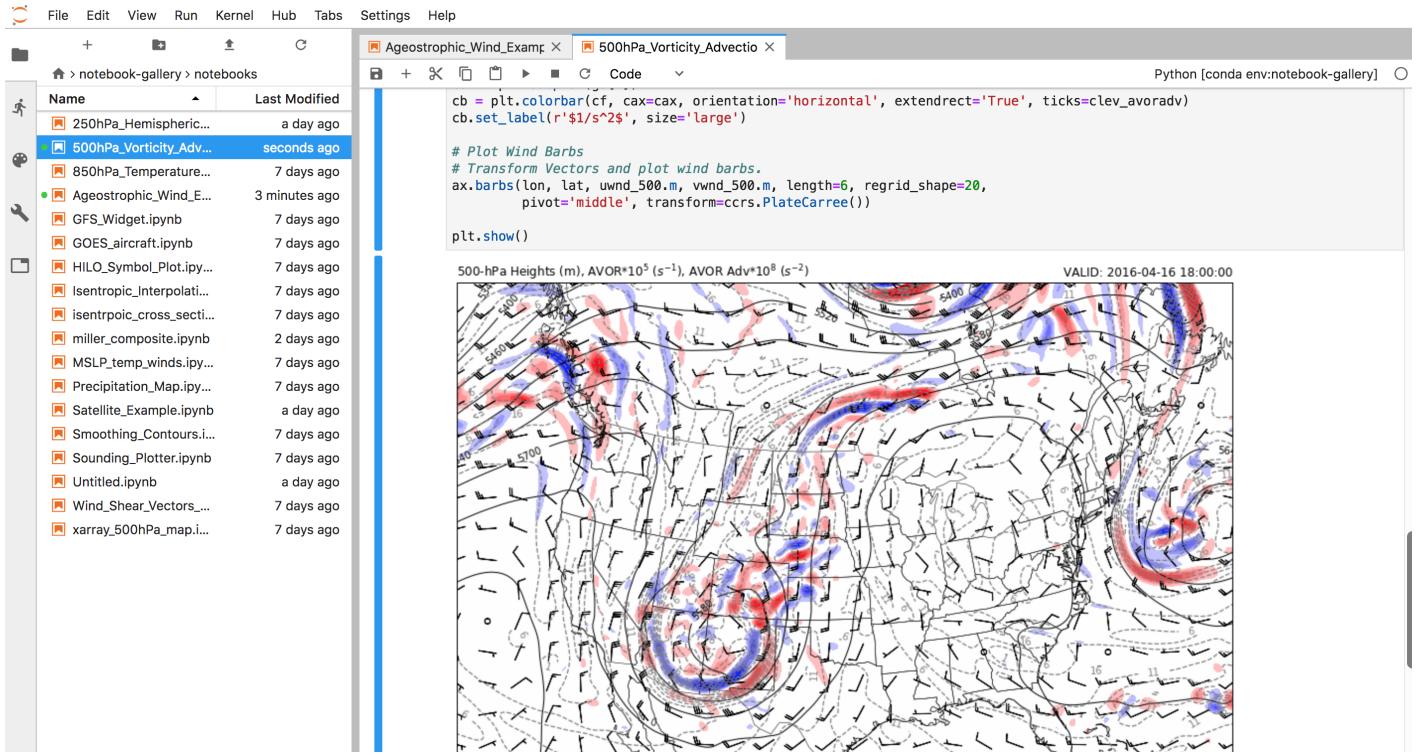
An improved architecture and enthusiastic user base are driving uptake of the open-source web tool.

Jeffrey M. Perkel

[Jupyter] notebooks are really a killer app for teaching computing in science and engineering - Lorena Barba, Engineering Professor, GWU

Award # NSF-1344155

JupyterLab: Next Generation UI

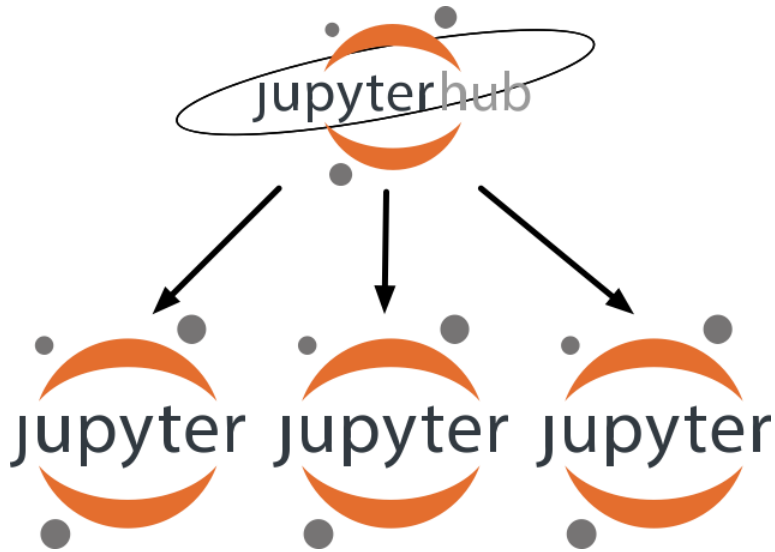


- Terminal (`git`, `conda`, etc.)
- Text Editor

JupyterHub: Multi-user Jupyter Notebooks

Fernando Pérez: It is infeasible for IT support to assist 800 students install complex software on their laptops.

- Users log in to a JupyterHub server
- Users have their own work space
- Excellent for workshops or in the classroom
- Administrator can configure ahead of time on behalf of user



Award # NSF-1344155

Unidata JupyterHubs

- Notre Dame of Maryland University
- Southern Arkansas University
- 2018 IS-GEO Workshop
- 2018 Fall Unidata Python Workshop

Award # NSF-1344155

Zero to JupyterHub Project

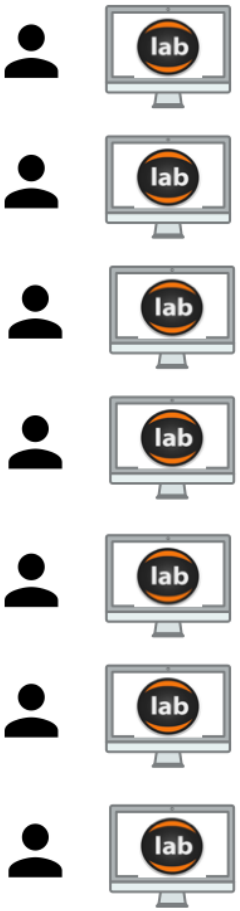
Problem: A single JupyterHub server running on a large VM can only serve a small number of students (< 10).

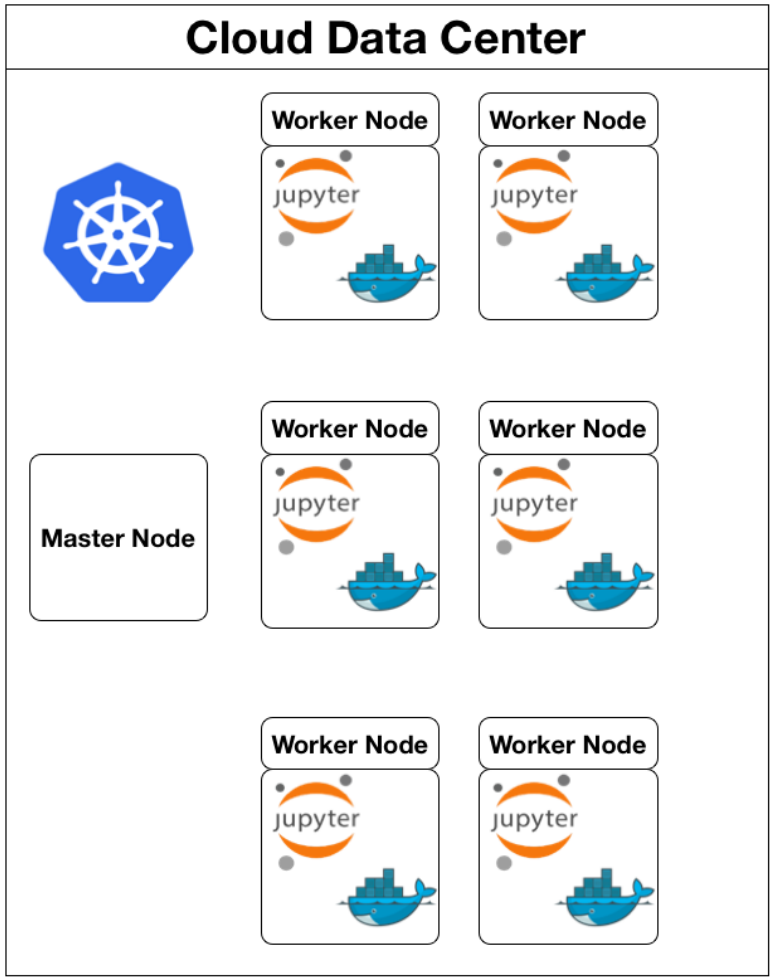
Solution: Zero to JupyterHub project aims to take advantage of the elastic computational capacity of the cloud.

- Virtual Machine
- Software Containers (i.e., Docker)
- Data center software orchestration (i.e., Kubernetes)

Zero to JupyterHub allows for many more users.

Zero to JupyterHub





NSF Jetstream Cloud

Award # NSF-1344155

Cloud Computing

CIO JOURNAL.

Harvard-MIT's Broad Institute Powers Genomic Research in the Cloud

'You need to build a scale-out infrastructure, and it doesn't make sense to do it yourself,' CIO says

By Steven Norton

Mar 12, 2018 6:41 pm ET

2 COMMENTS

a more scalable and accessible computing infrastructure better serves researchers and spurs advances in the field.

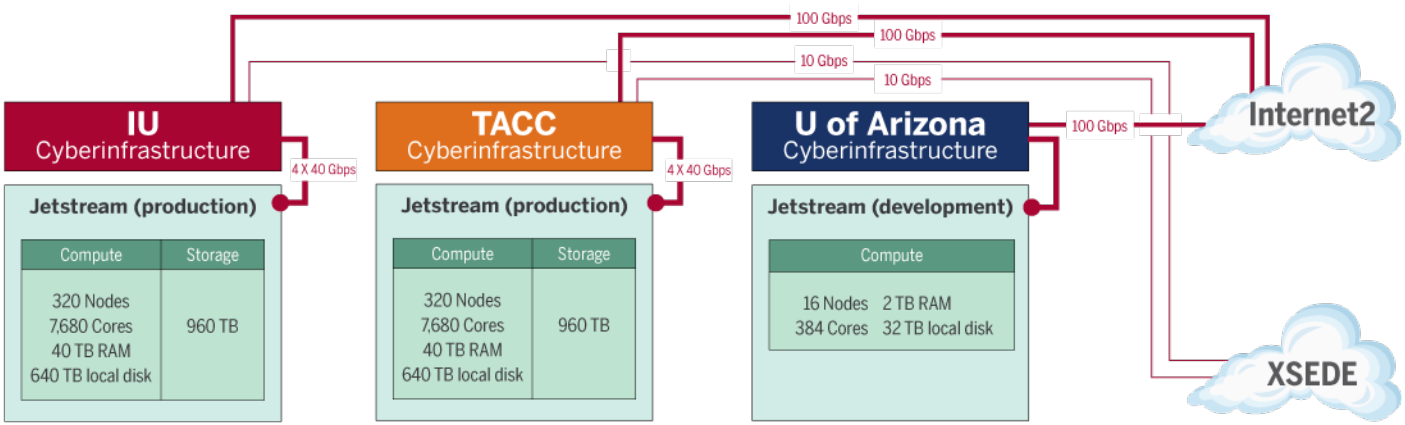
Commercial Cloud Computing Is $$$

- GEMPAK product generation server running on Amazon cloud
- m1.large: 2 vCPU, 7.5GB memory, 2x 420GB disk
- Costs Unidata $3,000/year!

NSF Jetstream Cloud

- What is Jetstream?
  - A National Science and Engineering Cloud funded by an $11 million NSF grant.
  - Data centers at IU and TACC.
- Attached to fast Internet2 capability.
- Cloud based on OpenStack for creation of VMs, networks etc.
- Unidata has been operating on Jetstream for 3 years through research grants

NSF Jetstream Cloud



Award # NSF-1344155

A JupyterHub for Atmospheric Science

Award # NSF-1344155

Unidata JupyterHub on Jetstream

- Three Unidata Notebook projects with environments pre-configured
  - Notebook Gallery, example atmos science notebooks
  - Python Workshop, netcdf4-python, metpy, siphon training
  - Online Python Training Teaching Python with an atmos focus
- Zero to JupyterHub
- Login with GitHub credentials
- JupyterLab

Award # NSF-1344155

Technical Specifications

- Two Kubernetes clusters running at Jetstream TACC data center
- Each cluster has:
  - 50 vCPUs
  - 150 GB RAM
  - spread over 5 VMs
- Each user has 1GB of persistent disk storage
- Can accommodate 40 users per cluster

Award # NSF-1344155

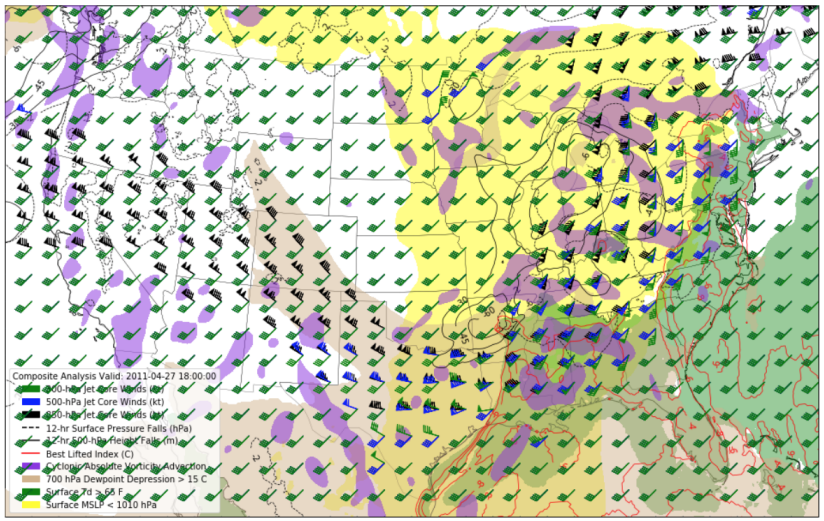
Example Notebook: Miller Composite

Satellite_Exam | 500hPa_Vortic | miller_composi

Code | Python [conda env:notebook-gallery]

```
Legend
rple = mpatches.Patch(color='BlueViolet', label='Cyclonic Absolute Vorticity
llow = mpatches.Patch(color='yellow', label='Surface MSLP < 1010 hPa')
een = mpatches.Patch(color='green', label='Surface Td > 65 F')
n = mpatches.Patch(color='tan', label='700 hPa Dewpoint Depression > 15 C')
d_line = lines.Line2D([], [], color='red', label='Best Lifted Index (C)')
shed_black_line = lines.Line2D([], [], linestyle='dashed', color='k',
                               label='12-hr Surface Pressure Falls (hPa)')
ack_line = lines.Line2D([], [], linestyle='solid', color='k',
                        label='12-hr 500-hPa Height Falls (m)')
t.legend(handles=[jet300, jet500, jet850, dashed_black_line, black_line, red
                  purple, tan, green, yellow], loc=3,
         title='Composite Analysis Valid: {:s}'.format(str(vtimes[0])))

t.show()
```



Example Notebook: Satellite + GFS Model

Satellite_Exam | 500hPa_Vortic | miller_composi

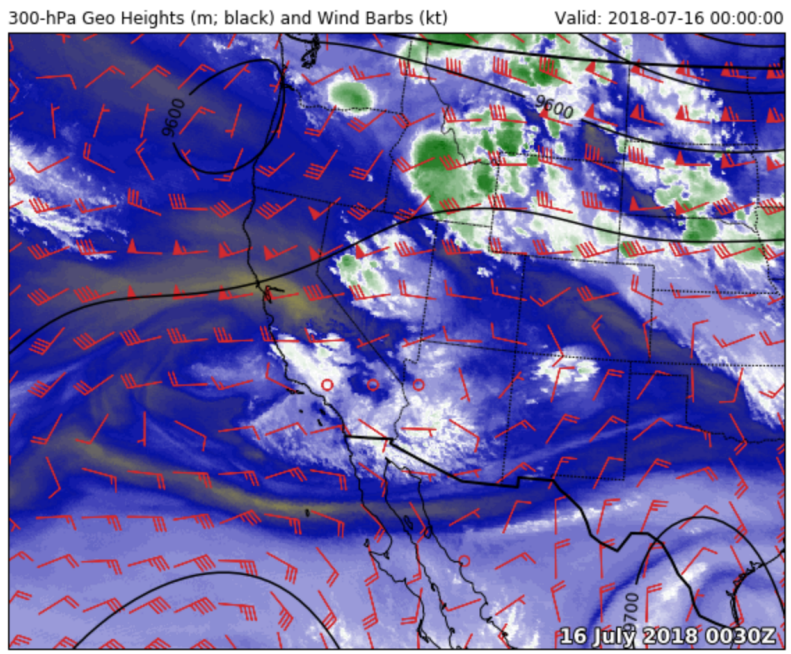
Code | Python [conda env:notebook-gallery]

```
ax.set_extent([-132, -95, 25, 47], ccrs.Geodetic())
cs = ax.contour(lon, lat, Z_300, colors='black', transform=ccrs.PlateCarree(
ax.clabel(cs, fontsize=12, colors='k', inline=1, inline_spacing=8,
          fmt='%i', rightside_up=True, use_clabeltext=True)

ax.barbs(lon, lat, uwnd_300.to('knots').m, vwnd_300.to('knots').m, color='ta
         length=7, regrid_shape=15, pivot='middle', transform=ccrs.PlateCar

ax.set_title('300-hPa Geo Heights (m; black) and Wind Barbs (kt)', loc='left
ax.set_title('Valid: {}'.format(vtime[0]), loc='right')

plt.show()
```

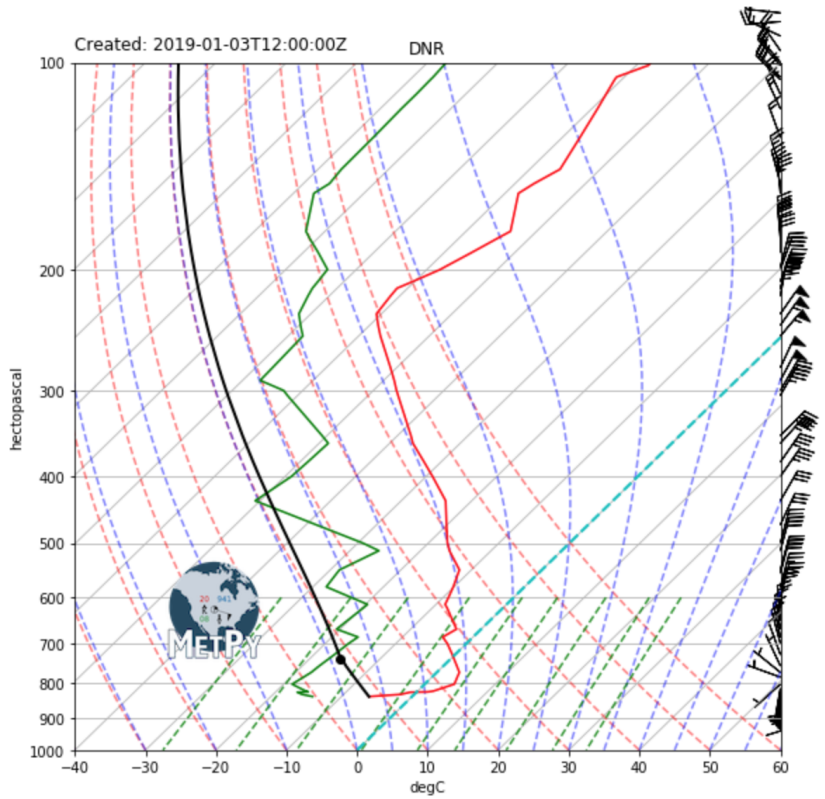


Example Notebook: Upper Air SkewT

Sounding_Plotter.ipynb

Code | Python [conda env:notebook-gallery]

```
        with tempfile.NamedTemporaryFile(suffix='.png') as f:
            skewt.ax.figure.savefig(f.name)
            uploader.upload_to(f.name, posixpath.join(args.gdrive, fname))
    else:
        skewt.ax.figure.savefig(make_name(args.site, request_time))
```



JupyterHub URLs

https://www.unidata.ucar.edu/projects/#jupyterlab

https://js-104-95.jetstream-cloud.org

https://js-16-87.jetstream-cloud.org

Future Plans

- Better URL name – jupyterhub.unidata.ucar.edu
- Supporting two classes at Southern Arkansas University – roughly 50 students
- Community outreach (i.e., more users)
- You?

Award # NSF-1344155

Acknowledgments

We thank Brian Beck, Maytal Dahan, Jeremy Fischer, Victor Hazlewood, Peg Lindenlaub, Suresh Marru, Lance Moxley, Marlon Pierce, Semir Sarajlic, Craig Alan Stewart, George Wm Turner, Nancy Wilkins-Diehr, Nicole Wolter and Andrea Zonca for their assistance with this effort, which was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program.

Special thanks to Andrea

Questions

https://www.unidata.ucar.edu/projects/#jupyterlab

https://js-104-95.jetstream-cloud.org

https://js-16-87.jetstream-cloud.org

Award # NSF-1344155