Accessing Distributed Jupyter/Spark in OnDemand

Jeremy W. Nicklas, Eric Franz, Alan Chalker, Doug Johnson, Morgan Rodgers, David E. Hudak

Ohio Supercomputer Center

This work is supported by the National Science Foundation of the United States under the awards NSF SI2-SSE-1534949 and NSF CSSI-1835725

Come see my poster tonight for more info!

OPEN OnDemand

Overview

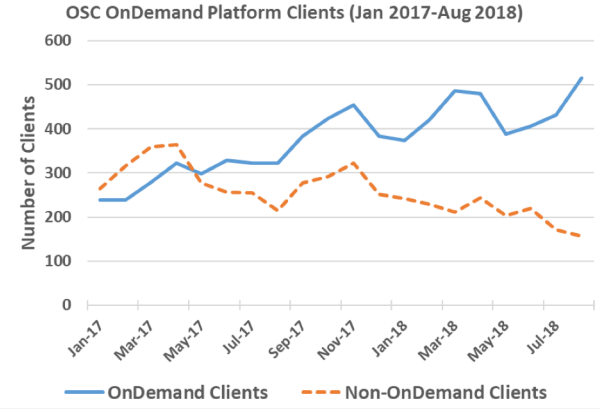
- Easy to install and use
- Web-based access to supercomputers
- Support for interactive supercomputing

Features include

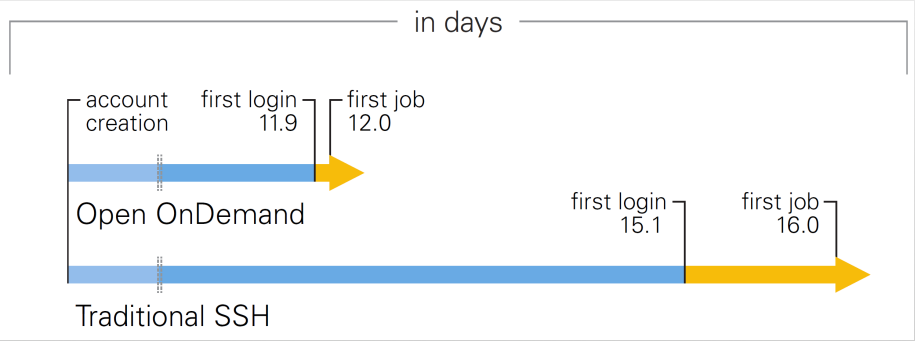
- Plugin-free web experience
- Easy file management
- Command-line shell access
- Job management and monitoring
- Graphical desktops and applications

OSC Install Details and Impact

- Launched Sep. 2016, serving OSC clients globally
- % of users has steadily increased since launch



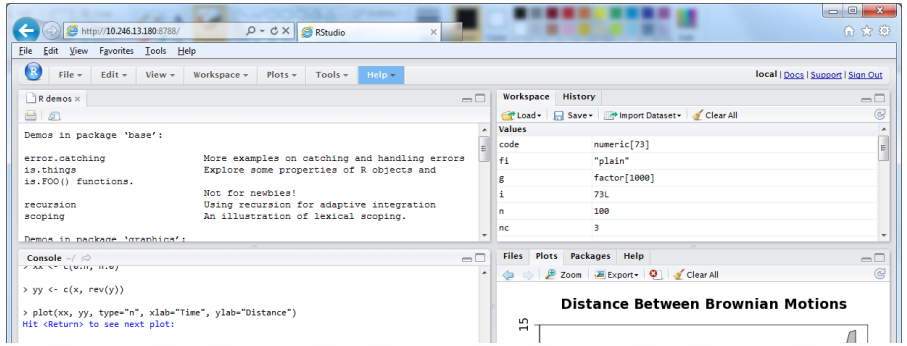
- **Improving Time to Science**: new OnDemand users start faster than ssh users: first login and first job



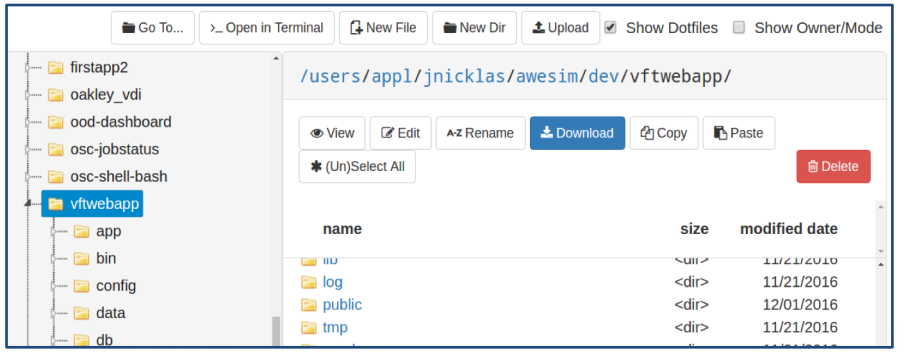
Ohio Supercomputer Center
An OH·TECH Consortium Member

Interactive Apps & Cluster Access

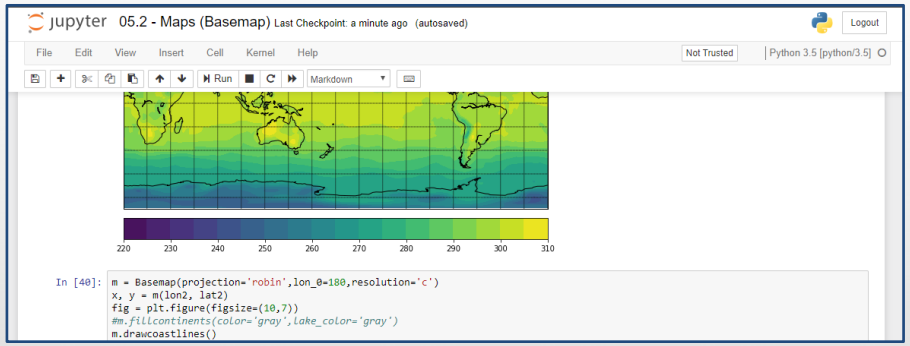
RStudio Server – R IDE



File Access (browse, edit, etc)

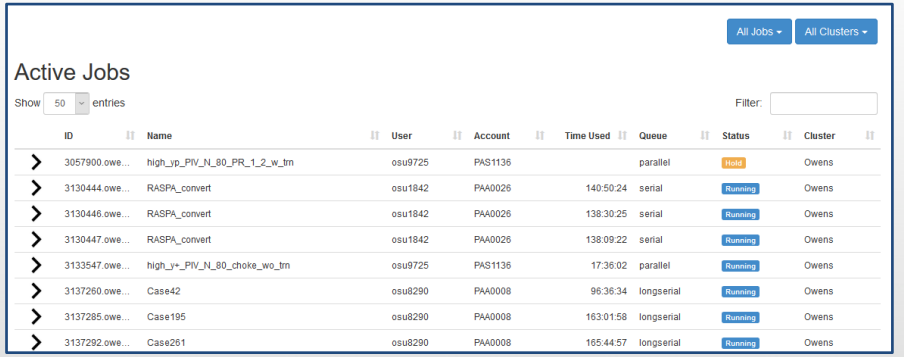


Jupyter Notebook – Python IDE



Come see my poster tonight for more info!

Manage Jobs (view, submit, etc)



And many more, such as ANSYS Workbench, Abaqus/CAE, MATLAB, Paraview, COMSOL Multiphysics

And many more, such as in-browser terminal, job apps, noVNC desktops and apps

Example Current Engagements and Deployments

Production Deployments

In Process of Installing

Come see my poster tonight for more info!

Get Started!

- SGCI Affiliate; Listed in SGCI Catlog
- Documentation and code repository available at: http://openondemand.org/
- Send email to ood-users-request@lists.osc.edu with the subject “subscribe” to join the mailing list
- Webinars and conference publications available on the website

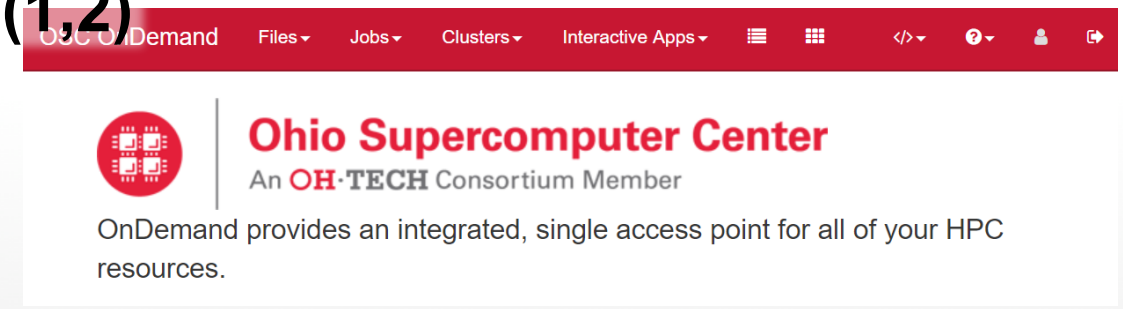
Open OnDemand website
QR code

Based upon work supported by the National Science Foundation under grant numbers 1534949 and 1835725.

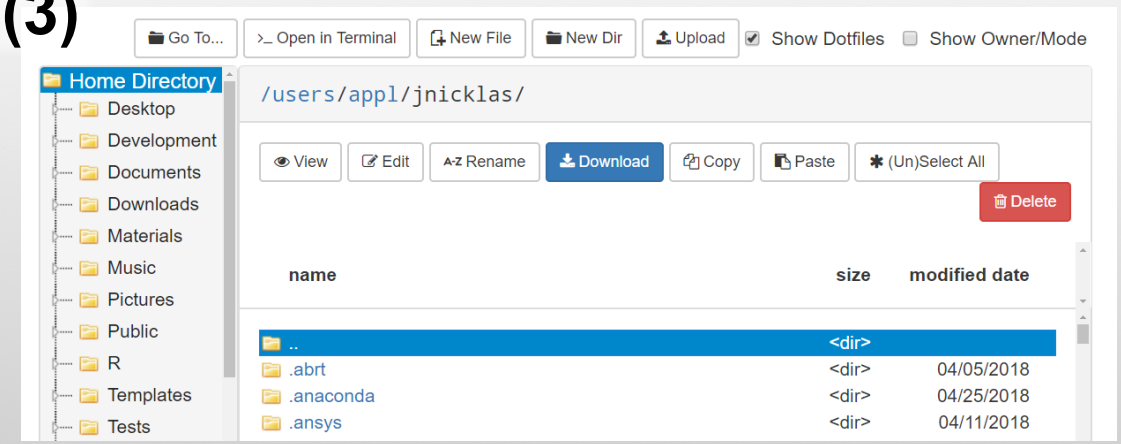
OSC OnDemand Overview

1. Browse to OSC OnDemand (Dashboard)
2. Show navigational elements
3. Launch File Explorer in home directory
4. Launch Active Jobs to show jobs on cluster
5. Launch Shell App
6. Go to Interactive Sessions for Jupyter demo

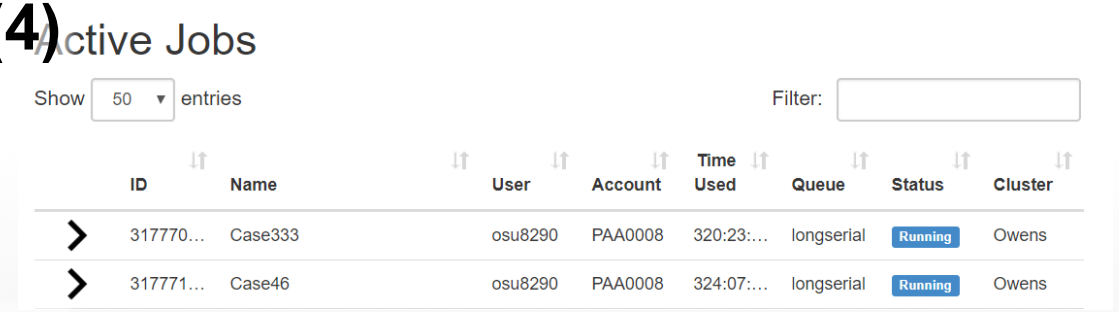
(1,2)



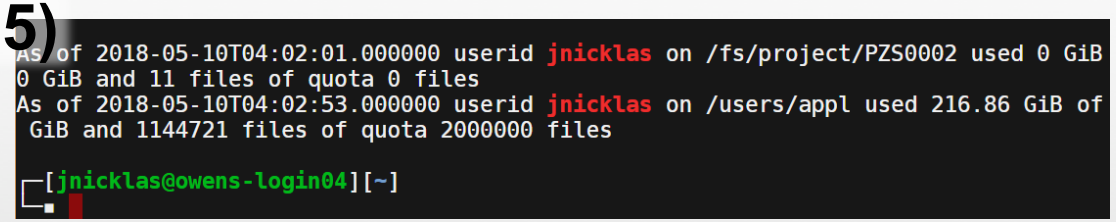
(3)



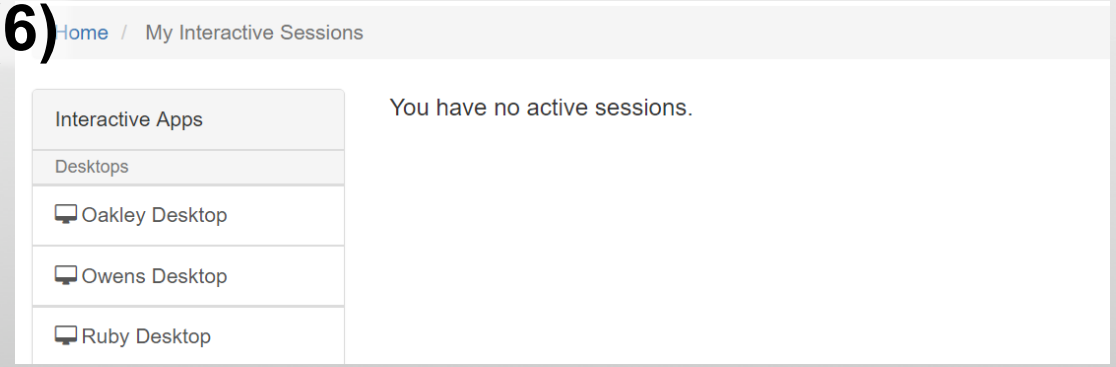
(4)



(5)



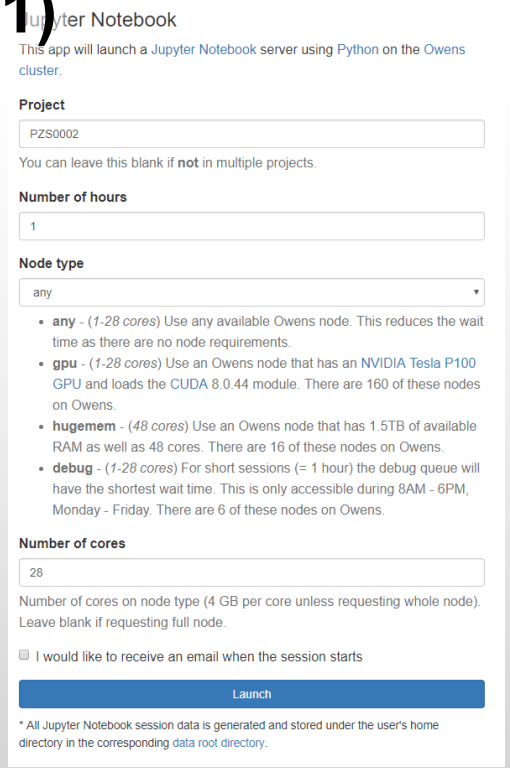
(6)



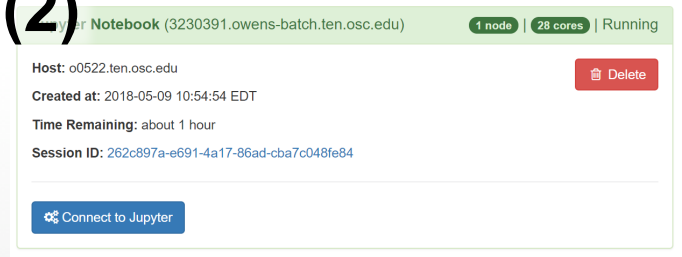
Jupyter through OnDemand (1 node | 28 cores)

1. Launch Jupyter from OnDemand
2. Connect to Jupyter when it starts
3. Open a terminal in Jupyter with `htop`
4. Open a Notebook using one of the cluster-installed Anaconda modules
5. Benchmark a Pi calculation while observing resource utilization in realtime (uses single core)

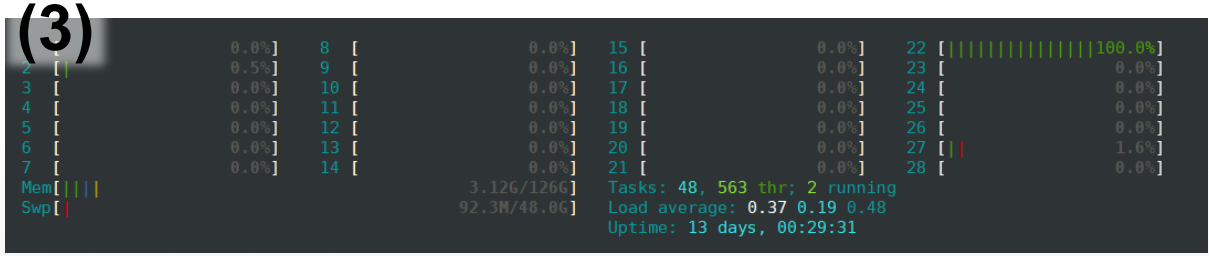
(1)



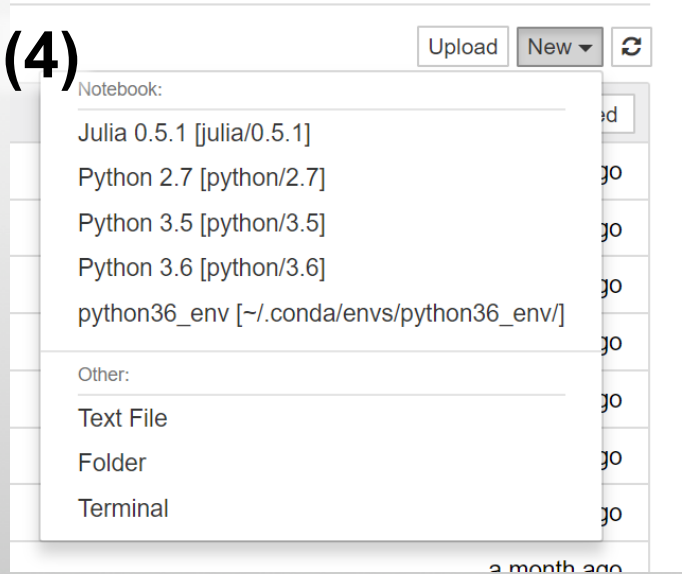
(2)



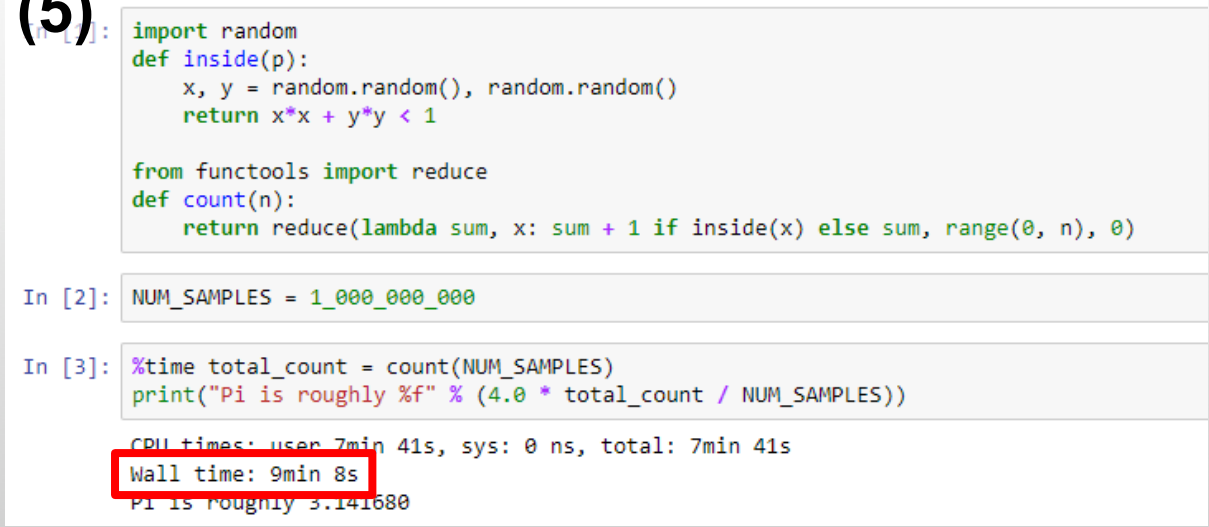
(3)



(4)



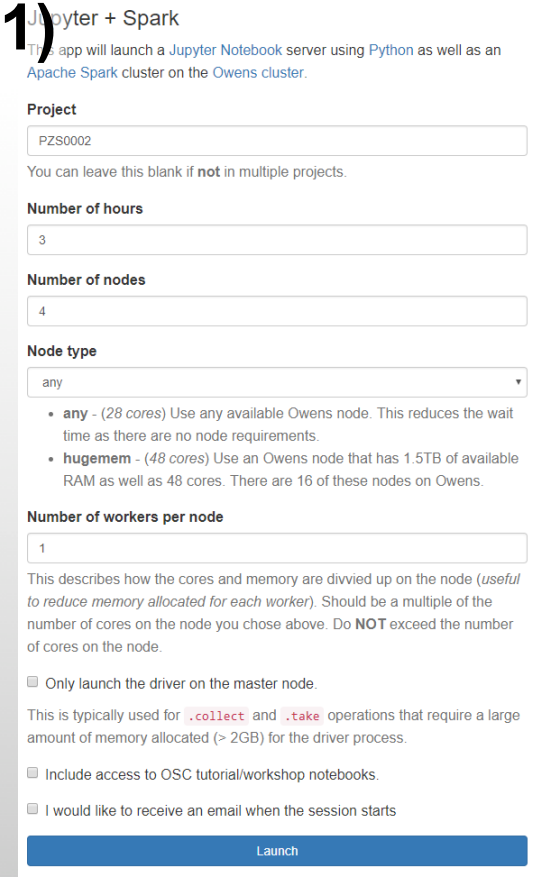
(5)



Scaling Science — Jupyter / Spark (4 nodes | 112 cores)

1. Launch Jupyter / Spark from OnDemand requesting 4 nodes
2. Connect to Jupyter when it starts
3. Open a `pyspark` notebook
4. Show number of cores Spark is using
5. Launch same Pi calculation using Spark

(1)

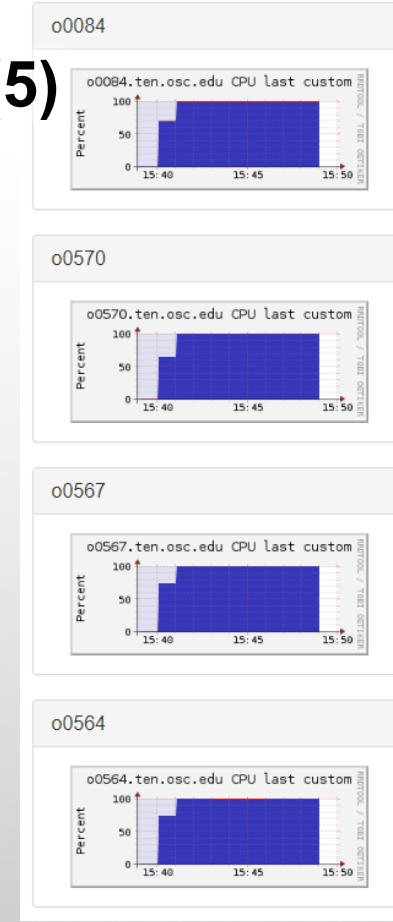


(4)

```
In [1]: print("Number of cores: %d" % sc.defaultParallelism)
        Number of cores: 112

In [2]: import random
        def inside(p):
            x, y = random.random(), random.random()
            return x*x + y*y < 1

        def count(n):
            return sc.parallelize(range(0, n)).filter(inside).count()

In [3]: NUM_SAMPLES = 1_000_000_000

(5)
In [4]: %time total_count = count(NUM_SAMPLES)
        print("Pi is roughly %f" % (4.0 * total_count / NUM_SAMPLES))

        CPU times: user 19.2 ms, sys: 8.49 ms, total: 27.7 ms
        Wall time: 6.31 s
        Pi is roughly 3.141834
```

87x speedup

(5)



Summary

- Jupyter / Spark is a powerful distributed environment
- OnDemand makes it easy to use on an HPC cluster
- JupyterHub + BatchSpawner is a nice Jupyter-only alternative
  - Both rely on reverse-proxy based solution
- OnDemand also makes it easy to use COMSOL Server, RStudio Server, and X11 applications running in a VNC/websockify stack

Visit our website to get started:
http://openondemand.org

Join our mailing list to keep in touch:
https://lists.osu.edu/mailman/listinfo/ood-users