Reducing Time to Science: Unidata and JupyterHub Technology Using the XSEDE Jetstream Cloud

Julien Chastang (chastang@ucar.edu), Rich Signell (USGS Woods Hole), Jeremy Fischer (XSEDE, Indiana University)

Unidata Program Center, UCP, University Corporation for Atmospheric Research

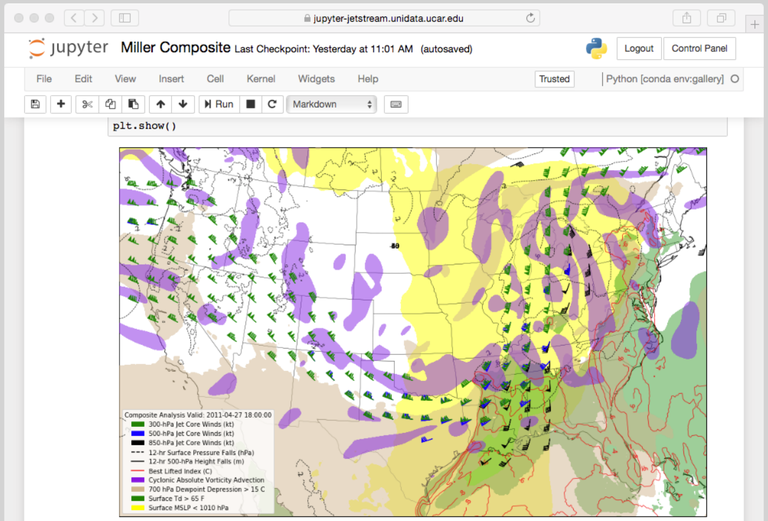
Abstract

The ascendancy of Jupyter notebooks and cloud computing coupled with established Unidata technology provides a significant opportunity to improve "time to science" by reducing research and data impediments. Scientists and students no longer need to transfer unwieldy geoscientific datasets to local computing environments and can rely upon data center reliability and horsepower. Moreover, complex software environments required to analyze and visualize these data can be preconfigured by experts on the scientists' behalf. We describe the deployment of Unidata and JupyterHub technologies on the NSF-funded XSEDE Jetstream cloud[4, 5].

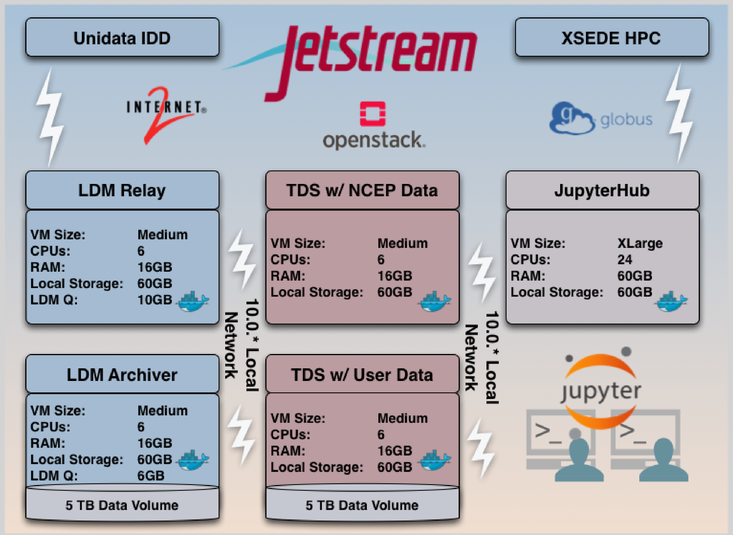
Methods

- Cloud Technologies
  - XSEDE Jetstream Cloud
  - OpenStack
  - Linux Virtual Machines
- Docker Containers[2]
  - Unidata Local Data Manager[1] (LDM)
  - Unidata THREDDS Data Server[3] (TDS)
  - JupyterHub
- Data Transfer
  - Internet2®
  - LDM
  - Internet Data Distribution (IDD) network with NCEP data
  - Globus CL
- JupyterHub with Unidata Notebooks Projects
  - Unidata Python Workshop
  - Notebook Gallery
  - Online Python Training
- Unidata Python APIs
  - netCDF4-python
  - MetPy
  - Siphon
- Globus OAuth

JupyterHub with Unidata Notebooks



Cloud Architecture



Conclusions

Deploying a multi-user JupyterHub server on the Jetstream cloud with easily accessible geoscientific data holdings has many benefits. Cloud computing environments are fast, reliable and scalable. Scientists and students analyze, visualize, and share data using only browser-based Jupyter notebook and JupyterHub technology. No local specialized desktop software or fast Internet connection are required. These Jupyter notebooks are preconfigured by specialists for the geoscience community. In addition, we can leverage decades of Unidata expertise in delivering and serving data with the LDM and TDS co-located on Jetstream delivering large quantities of geoscience data. This environment will enable scientists to spend less time managing their software and more time doing science. Future work will explore cloud elasticity, in a classroom setting for example, where students may be running many Jupyter notebooks at once.

Acknowledgments

We thank B. Beck, M. Dahan, J. Fischer, V. Hazlewood, P. Lindenlaub, S. Marru, L. Moxley, M. Pierce, C. A. Stewart, G. W. Turner, N. Wilkins-Diehr and N. Wolter for their assistance with this effort, which was made possible through the XSEDE Extended Collaborative Support Service (ECSS) program.

References

[1] J. Chastang. *Unidata LDM Docker Container.* https://github.com/Unidata/ldm-docker/. 2016. URL: https://github.com/Unidata/ldm-docker/.

[2] J. Chastang, W. Fisher, and S. Arms. "UniCloud: Docker Use at Unidata". In: *Proceedings, 32nd Conference on Environmental Information Processing Technologies, 96th AMS Annual Meeting.* See also http://unidata.github.io/Unidata-Dockerfiles/index.html. American Meteorological Society. New Orleans, LA, 2016. URL: https://ams.confex.com/ams/96Annual/webprogram/Paper287336.html.

[3] Wilcox K. and J. Chastang. *Unidata THREDDS Docker Container.* https://github.com/Unidata/thredds-docker/. 2015. URL: https://github.com/Unidata/thredds-docker/.

[4] Craig A. Stewart et al. "Jetstream: A Self-provisioned, Scalable Science and Engineering Cloud Environment". In: *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure.* XSEDE '15. St. Louis, Missouri, USA: ACM, 2015, 29:1–29:8. ISBN: 978-1-4503-3720-5. DOI: 10.1145/2792745.2792774. URL: http://doi.acm.org/10.1145/2792745.2792774.

[5] John Towns et al. "XSEDE: Accelerating Scientific Discovery". In: *Computing in Science & Engineering* 16.5 (2014), pp. 62–74. ISSN: 1521-9615. DOI: 10.1109/MCSE.2014.80.

unidata UCAR NSF

Award # NSF-1344155