Unidata Science Gateway on the XSEDE Jetstream Cloud

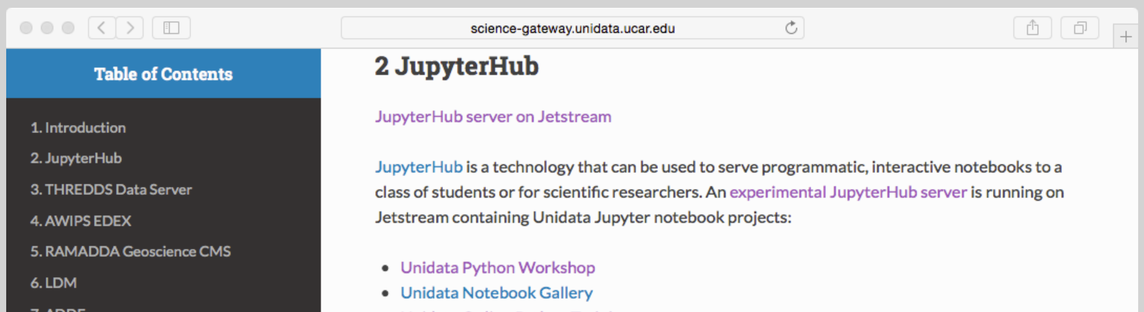
Julien Chastang (chastang@ucar.edu), Mohan Ramamurthy

Unidata Program Center, UCP, University Corporation for Atmospheric Research

Abstract

Cloud computing accelerates scientific workflows, discoveries, and collaborations by reducing research and data friction. We aim to improve "time to science" with the NSF-funded XSEDE Jetstream cloud[5, 6]. We describe a Unidata science gateway on Jetstream. With the aid of open-source cloud computing projects such as OpenStack and Docker on Linux VMs, we deploy a variety of scientific computing resources on Jetstream for our scientific community. These systems can be leveraged with data-proximate Jupyter notebooks, and remote visualization clients such as the Unidata Integrated Data Viewer (IDV) and AWIPS CAVE. This gateway will enable students and scientists to spend less time managing their software and more time doing science.

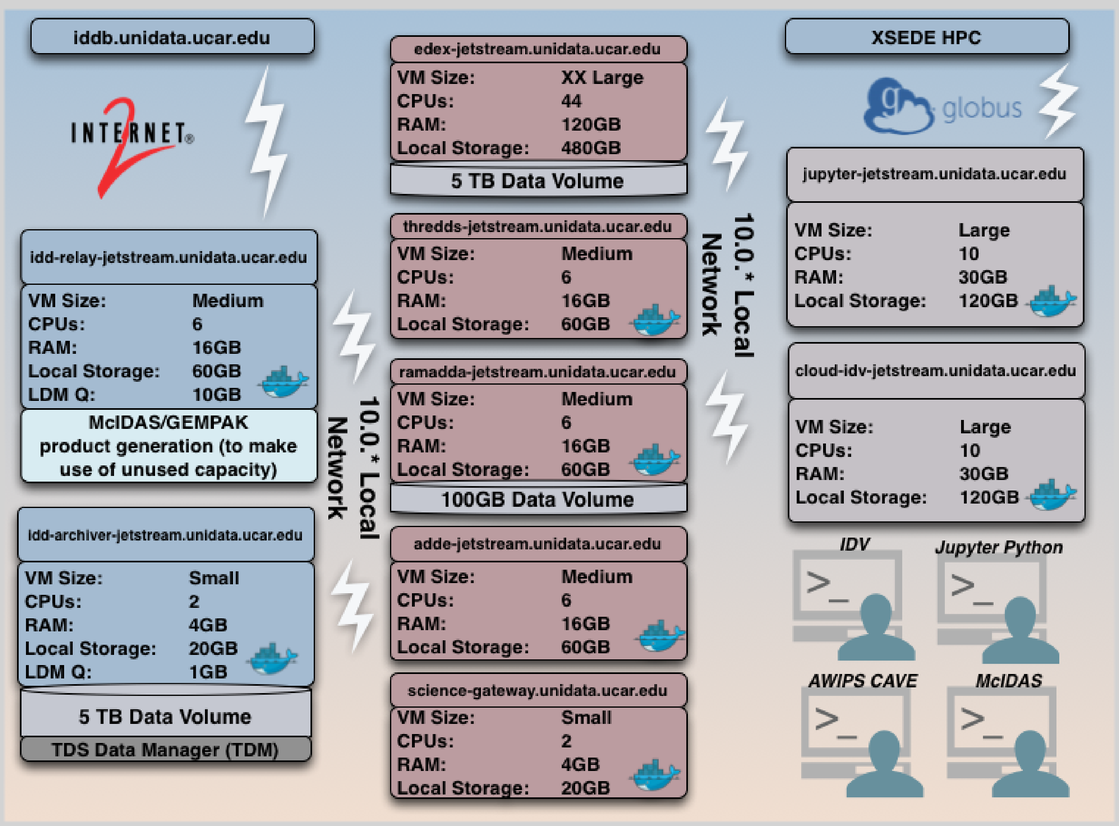
http://science-gateway.unidata.ucar.edu



Background

With the goal of better serving our community and in fulfillment of objectives articulated in "Unidata 2018: Transforming Geoscience through Innovative Data Services,"[4] Unidata is investigating how its technologies can best make use of cloud computing. The observation that science students and professionals are spending too much time distracted by software that is difficult to access, install, and use, motivates Unidata's investigation. In addition, cloud computing can tackle a class of problems that cannot be approached by traditional, local computing methods because of its ability to scale and its capacity to store large quantities of data.

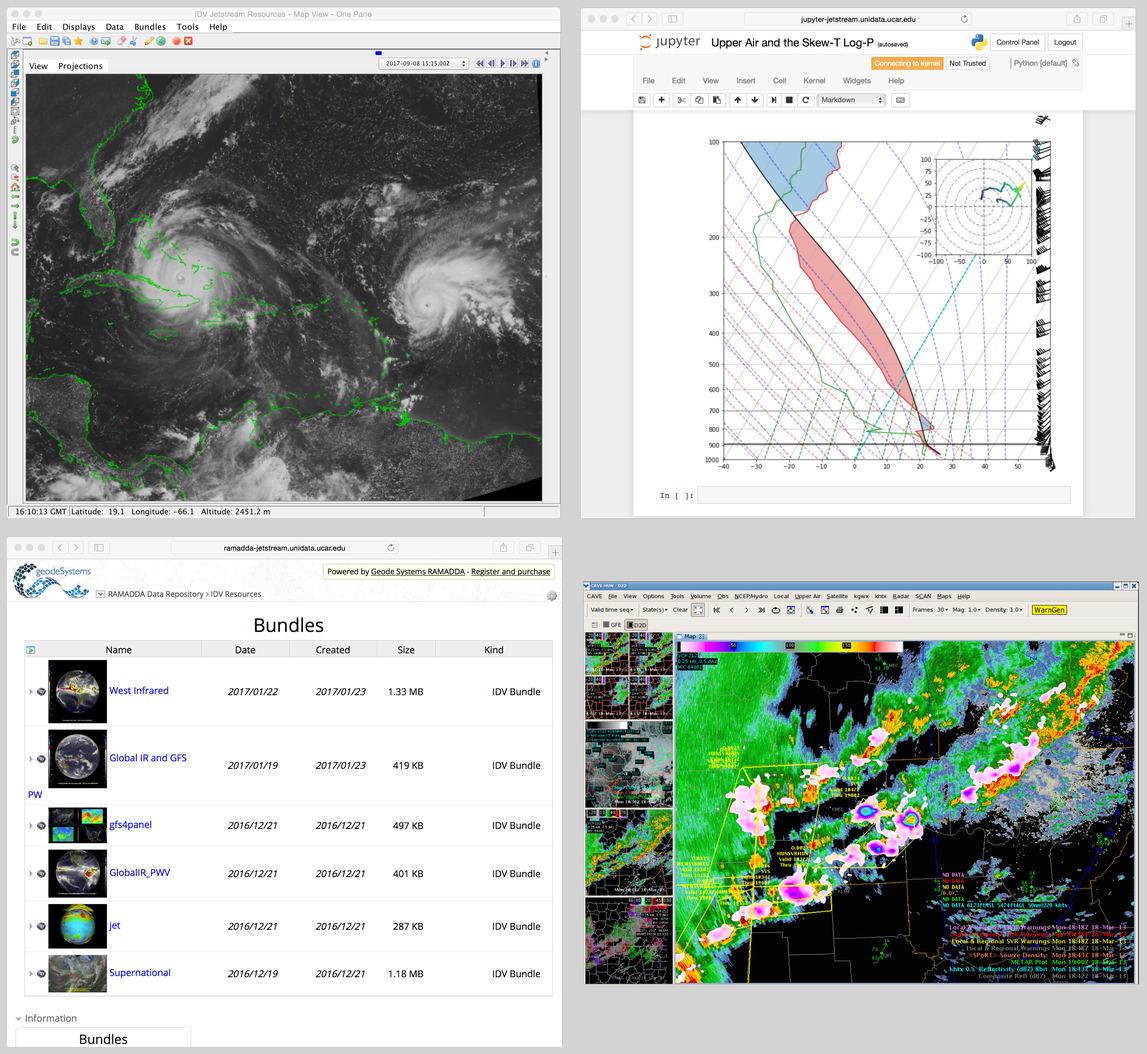
Architecture of VMs, Data Storage and Networking



Methods

To build the Unidata Science Gateway, we employed open-source and cloud computing technologies. We created several Docker containers for Unidata software offerings and reused other open-source containers[3]. We developed containers for the Unidata LDM and TDS, RAMADDA, and ADDE data distribution and serving technologies. We employ Apache Tomcat and JupyterHub containers maintained by open-source groups. With the Jetstream OpenStack API, we deployed a collection of Linux virtual machines (VMs) attached with disk storage to run these containers. Containers are orchestrated with `docker-compose`. The AWIPS EDEX server does not use Docker, but we can allocate a very large VM as this server requires significant computing resources. In addition, we setup an internal subnetwork with OpenStack for fast inter-VM communication via TCP ports and NFS mounts. With the LDM and Unidata Internet Data Distribution (IDD) network, we can deliver large quantities of geoscience data to Jetstream in a timely manner because of the Internet2® network accessible on Jetstream. The work presented here is developed in an open-source manner using git and github version control technology[1] and employing software carpentry best-practices.

Data-Proximate and Remote Analysis and Visualization



Starting at upper-left, clockwise: IDV, Jupyter Python notebook, AWIPS CAVE, RAMADDA.

Conclusions

We can quickly deploy a fully capable Unidata data center by leveraging the Jetstream cloud and a variety of open-source technologies. Science students and professionals can use the Unidata gateway directly via Jupyter notebooks, or with remote visualization client applications such as the IDV and AWIPS CAVE. Gateway users, coupled with XSEDE HPC resources, can achieve complete end-to-end scientific computing workflows[2]. 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 on the XSEDE Jetstream Cloud*. https://github.com/Unidata/xsede-jetstream/. 2017. URL: https://github.com/Unidata/xsede-jetstream/.

[2] J. Chastang and R. Signell. "Met/Ocean Modeling Workflows on XSEDE via HPC & Cloud". In: See also http://science-gateway.unidata.ucar.edu. Earth Science Information Partners. Bloomington, IN USA, 2017. DOI: 10.6084/m9.figshare.5249845.v1. URL: https://figshare.com/articles/Met_Ocean_Modeling_Workflows_on_XSEDE_via_HPC_Cloud/5249845.

[3] J. Chastang, T. Yoksas, and M. K. Ramamurthy. "Geoscientific Data Distribution in the XSEDE Jetstream Cloud". In: *Proceedings, 33nd Conference on Environmental Information Processing Technologies, 97th AMS Annual Meeting*. See also http://science-gateway.unidata.ucar.edu. American Meteorological Society. Seattle, WA USA, 2017. URL: https://ams.confex.com/ams/97Annual/webprogram/Paper315508.html.

[4] M. K. Ramamurthy. *Unidata 2018: Transforming Geoscience through Innovative Data Services*. Tech. rep. 3300 Mitchell Ln, Boulder, CO 80301 USA: Unidata Program Center, UCAR Community Programs, University Corporation for Atmospheric Research, 2013. URL: https://www.unidata.ucar.edu/publications/Unidata_2018.pdf.

[5] 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.

[6] 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