Unidata Science Gateway on the XSEDE Jetstream Cloud

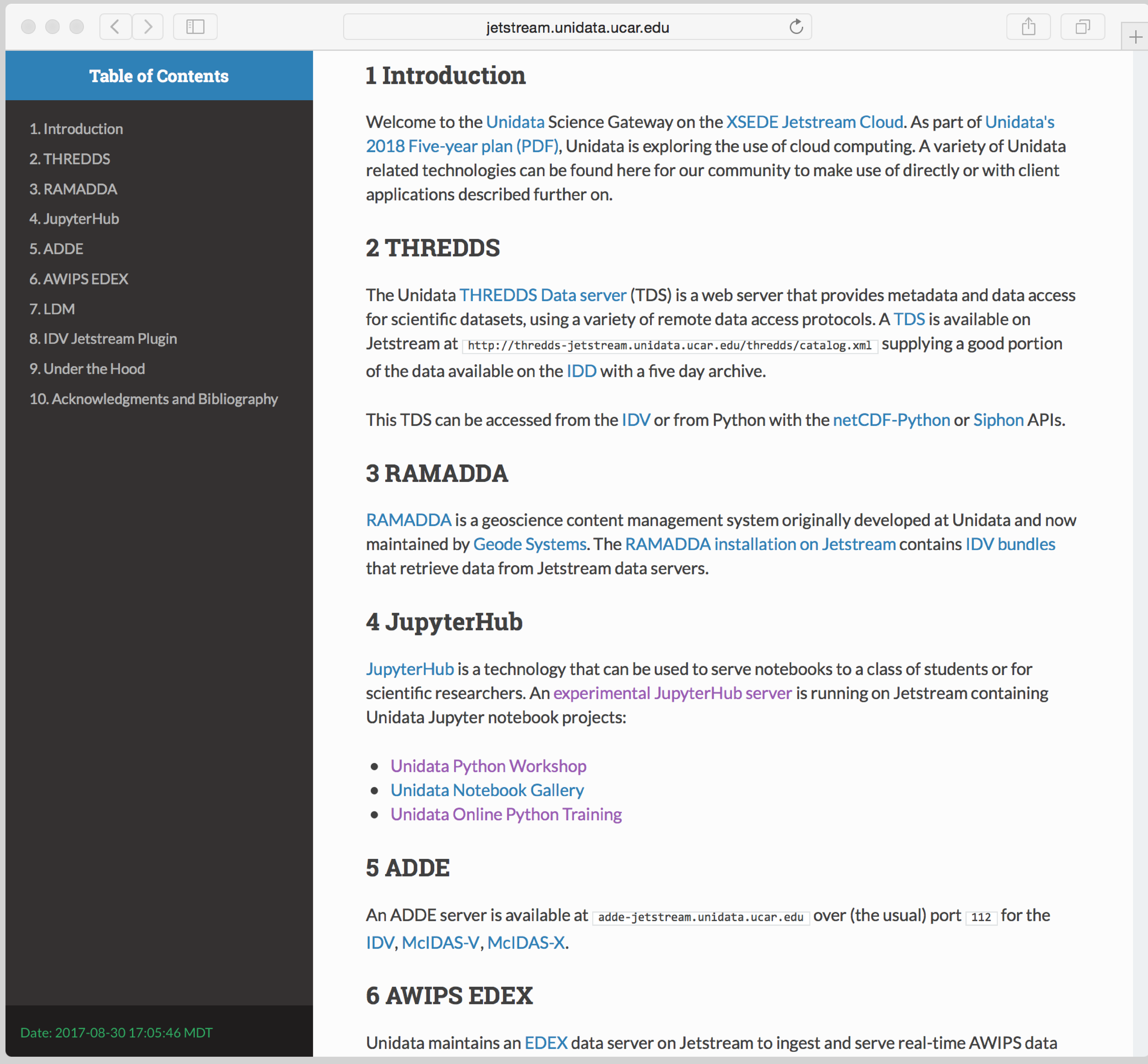
Julien Chastang (chastang@ucar.edu), Mohan Ramamurthy

Unidata Program Center, UCP, University Corporation for Atmospheric Research

Abstract

Cloud computing can accelerate scientific workflows, discoveries, and collaborations by reducing research and data friction. We aim to improve "time to science" by taking advantage of the NSF-funded XSEDE Jetstream cloud. We describe a Unidata science gateway on Jetstream. With the aid of several open-source, cloud computing projects including 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.

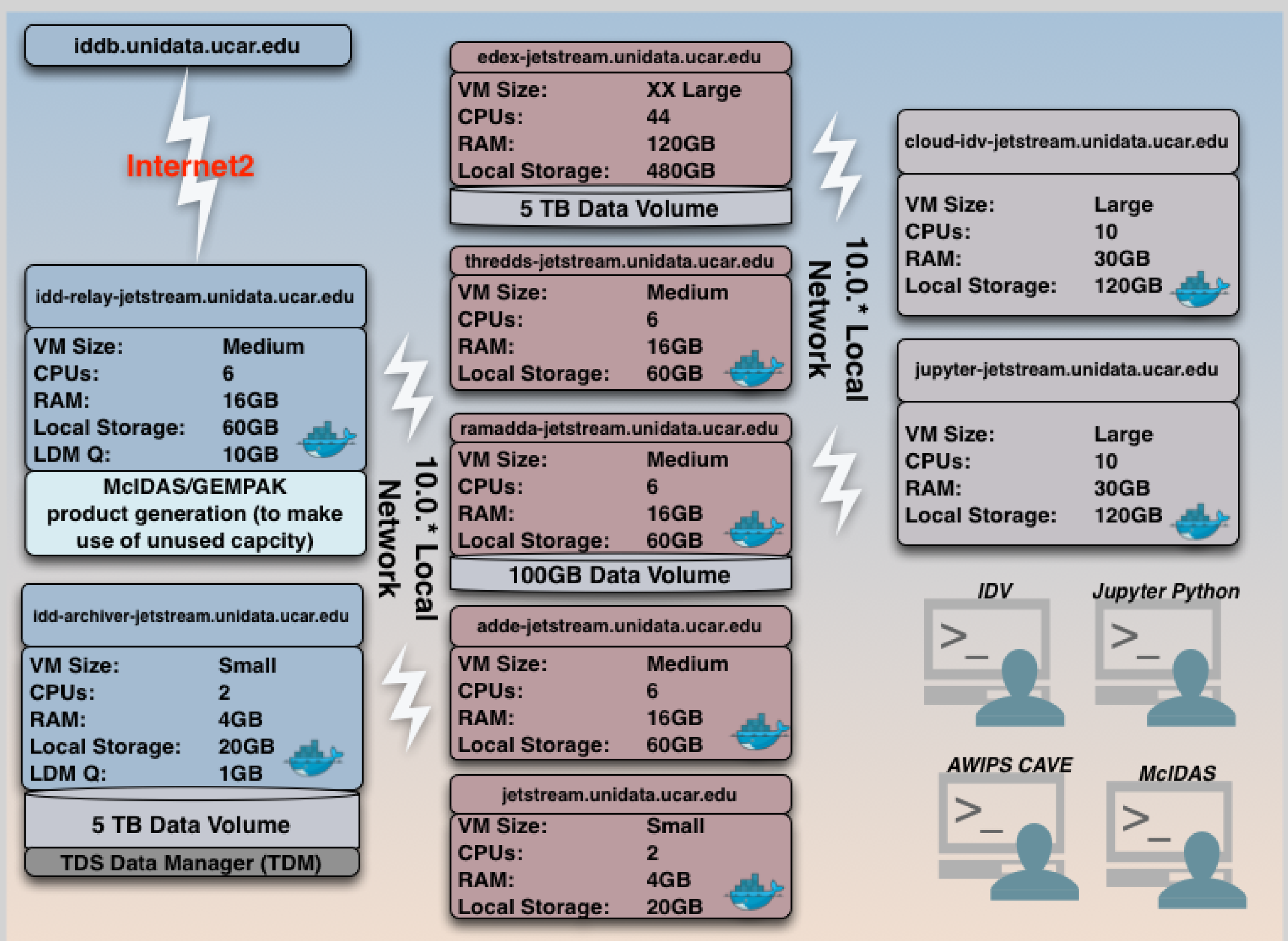
Unidata Science Gateway on Jetstream



Background

With the goal of better serving our community and in fulfillment of objectives articulated in "Unidata 2018: Transforming Geoscience through Innovative Data Services,"[3] Unidata is investigating how its technologies can best take advantage 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, by taking advantage of the cloud's ability to scale and its capacity to store large quantities of data, cloud computing can tackle a class of problems that cannot be approached by traditional, local computing methods.

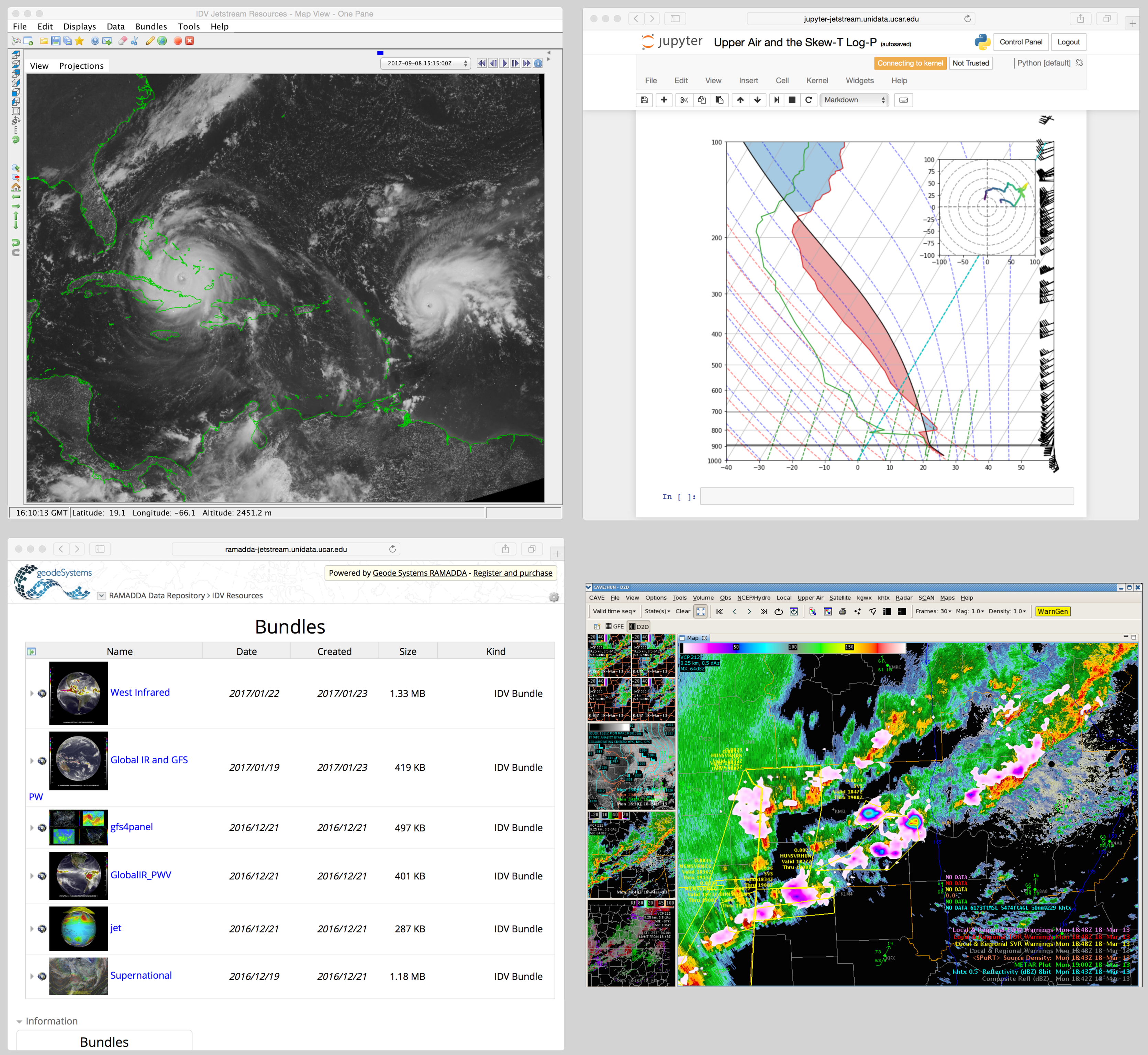
Architecture of VMs, Data Storage and Networking



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[2]. We developed containers for the Unidata LDM and TDS, RAMADDA, and ADDE data distribution and serving technologies. In addition, we make use of 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 make use of Docker, but we are able to 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. All 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 from upper-left, clockwise: IDV remote visualization client, Jupyter Python notebook, RAMADDA geoscience content management system, AWIPS CAVE remote visualization client.

Conclusions

By leveraging the Jetstream cloud and a variety of open-source technologies, we can quickly deploy a fully capable Unidata data center. These resources can be used directly via Jupyter notebooks, or with remote visualization client application such as the IDV and AWIPS CAVE. Future work will involve taking better advantage of the horizontal scalability of the cloud, in a classroom setting for example, where students may be running many Jupyter notebooks at once.

Acknowledgments

We thank Brian Beck, Craig Alan Stewart, George Wm Turner, Jeremy Fischer, Lance Moxley, Marlon Pierce, Maytal Dahan, Nicole Wolter, Peg Lindenlaub, Suresh Marru, and Victor Hazlewood 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*. 2017. URL: https://github.com/Unidata/xsede-jetstream/.

[2] 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://jetstream.unidata.ucar.edu. American Meteorological Society. Seattle, WA USA, 2017. URL: https://ams.confex.com/ams/97Annual/webprogram/Paper315508.html.

[3] 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.

unidata UCAR NSF

Award # NSF-1344155