

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

The screenshot shows the Unidata Science Gateway on Jetstream website. The left sidebar contains a Table of Contents with links to sections like Introduction, THREDDS, RAMADDA, JupyterHub, ADDE, AWIPS EDEX, LDM, IDV Jetstream Plugin, Under the Hood, Acknowledgments and Bibliography. The main content area includes:

- 1 Introduction**: Welcome to the Unidata Science Gateway on the XSEDE Jetstream Cloud. As part of Unidata's 2018 Five-year plan (PDF), Unidata is exploring the use of cloud computing. A variety of Unidata related technologies can be found here for our community to make use of directly or with client applications described further on.
- 2 THREDDS**: The Unidata THREDDS Data server (TDS) is a web server that provides metadata and data access for scientific datasets, using a variety of remote data access protocols. A TDS is available on Jetstream at <http://thredds-jetstream.unidata.ucar.edu/thredds/catalog.xml>, supplying a good portion of the data available on the IDD with a five day archive.
- 3 RAMADDA**: RAMADDA is a geoscience content management system originally developed at Unidata and now maintained by Geode Systems. The RAMADDA installation on Jetstream contains IDV bundles that retrieve data from Jetstream data servers.
- 4 JupyterHub**: JupyterHub is a technology that can be used to serve notebooks to a class of students or for scientific researchers. An experimental JupyterHub server is running on Jetstream containing Unidata Jupyter notebook projects:

 - Unidata Python Workshop
 - Unidata Notebook Gallery
 - Unidata Online Python Training

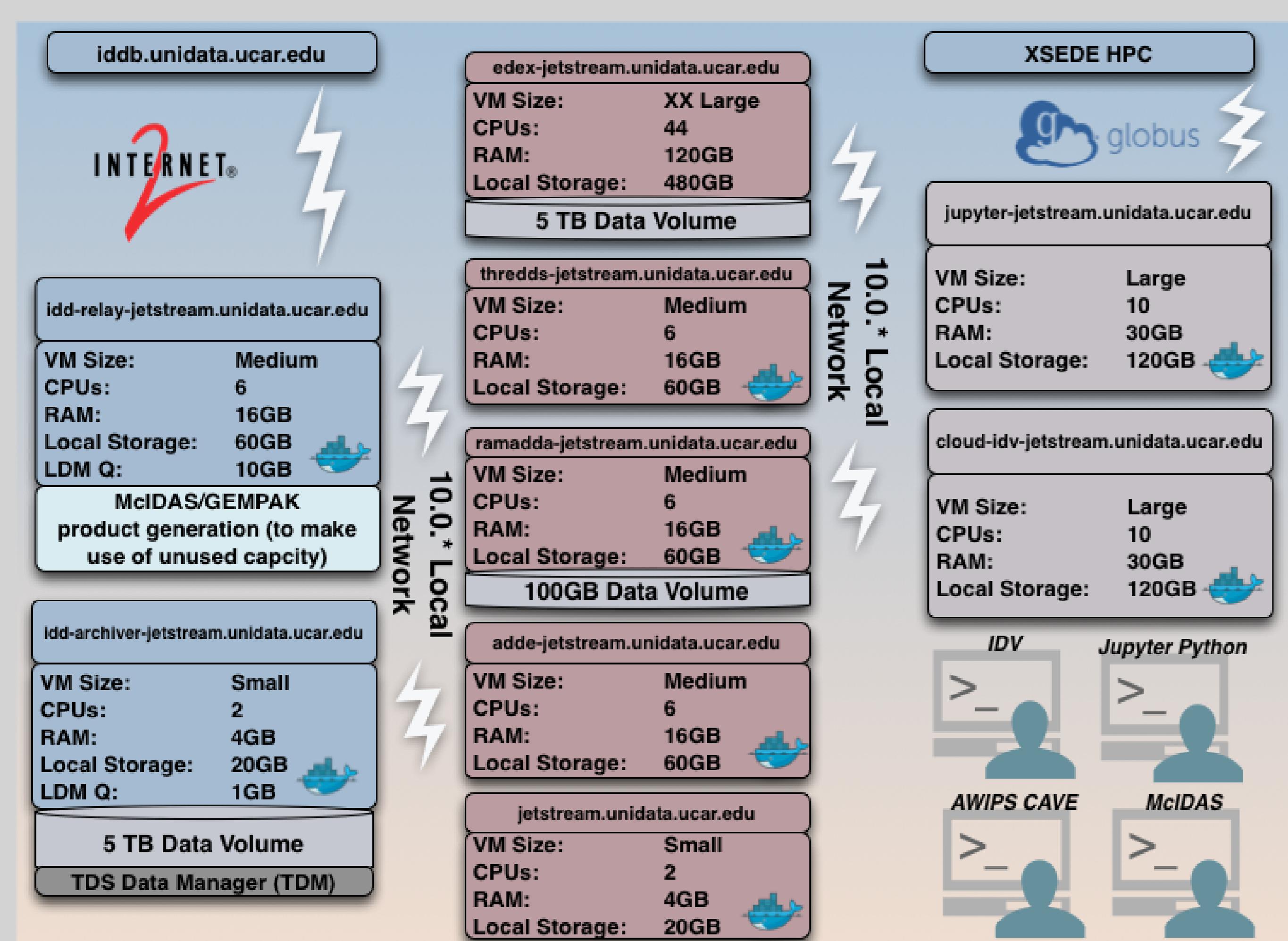
- 5 ADDE**: An ADDE server is available at adde-jetstream.unidata.ucar.edu over (the usual) port 112 for the IDV, McIDAS-V, McIDAS-X.
- 6 AWIPS EDEX**: Unidata maintains an EDEX data server on Jetstream to ingest and serve real-time AWIPS data

Date: 2017-08-30 17:05:46 MDT

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 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.

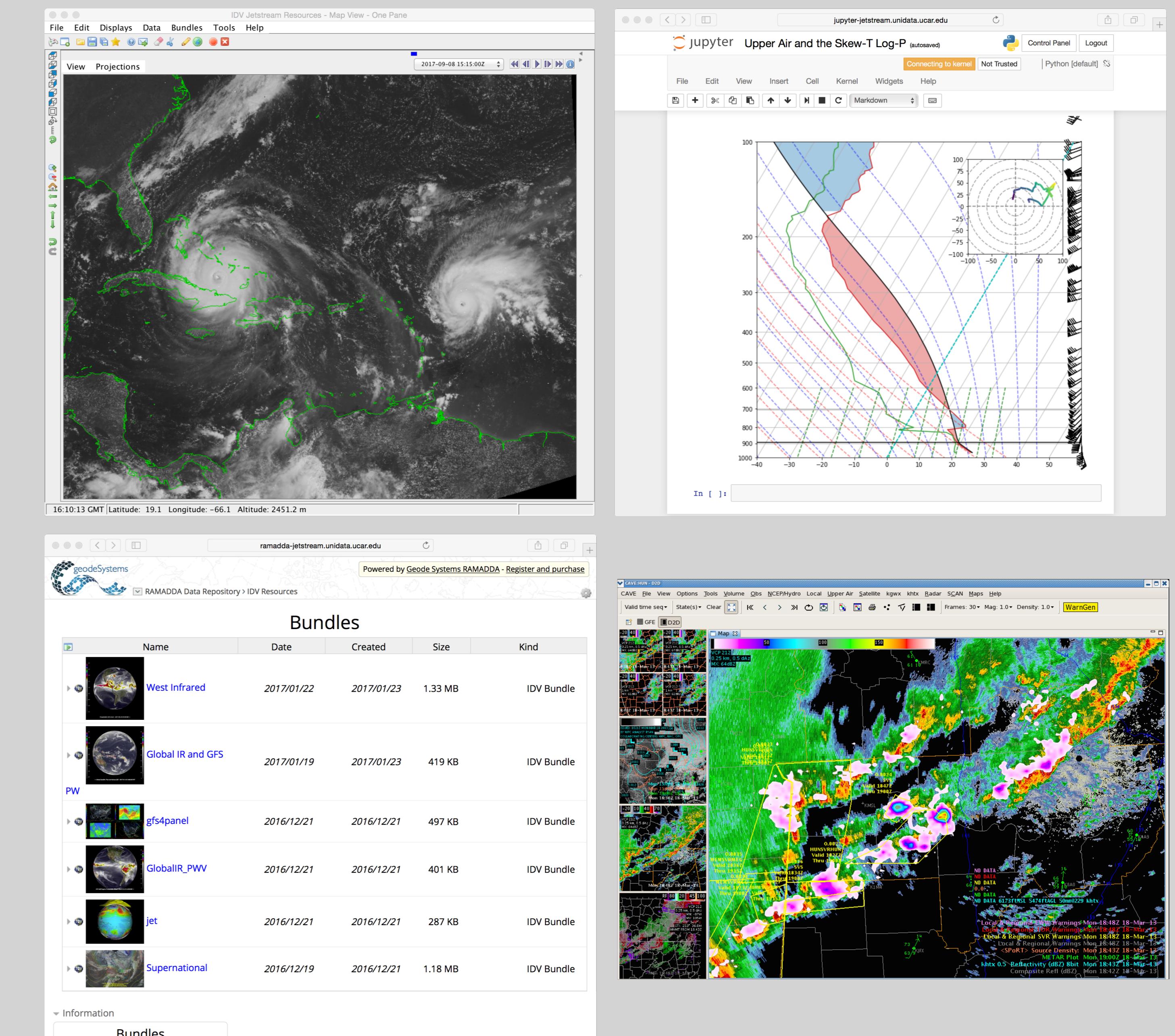
Architecture of VMs, Data Storage and Networking



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[3]. We developed containers for the Unidata LDM and TDS, RAMADDA, and ADDE data distribution and serving technologies. 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 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. 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

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 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*. <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://jetstream.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. Yokas, and M. K. Ramamurthy. "Geoscientific Data Distribution in the XSEDE Jetstream Cloud". In: *Proceedings, 33rd 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>.
- [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.