Unidata and data-proximate analysis and visualization in the cloud

Mohan Ramamurthy and Many
Unidata Staff

1 June 2017

Modeling in the Cloud Workshop

Unidata: A program of the community, by the community, and for the community

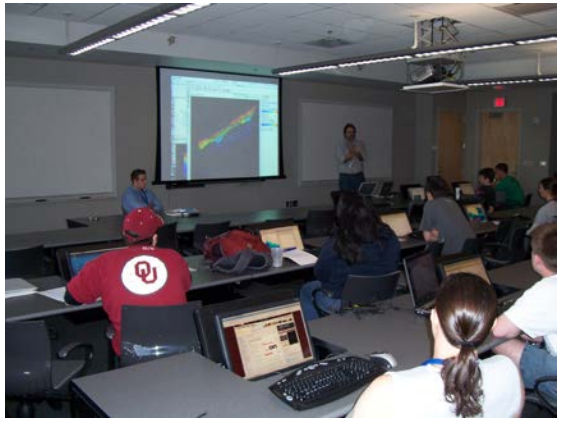
Established in 1984; Primarily funded by NSF

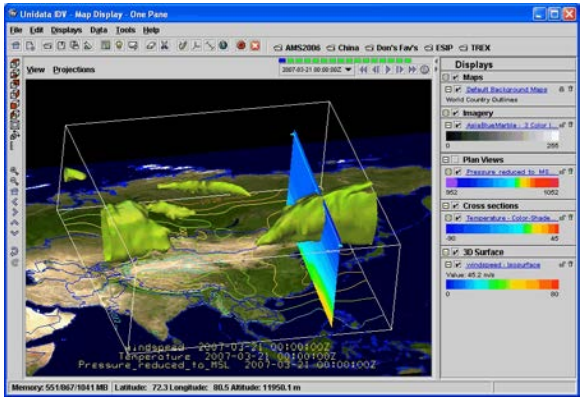
- Acquire and distribute real-time meteorological data;
- Develop software for accessing, managing, analyzing, visualizing geoscience data;
- Provide training and support to users;
- Negotiate data & software agreements on behalf of universities;
- Facilitate advancement of standards and conventions;
- Provide Equipment Grants

Niché and Mission

Niché: **Providing data services to advance Earth System science research and education.**

- **Reduce *"data friction"*, lower the barriers for accessing and using data, and shrink the "time to science."**





A Snapshot of Products & Services

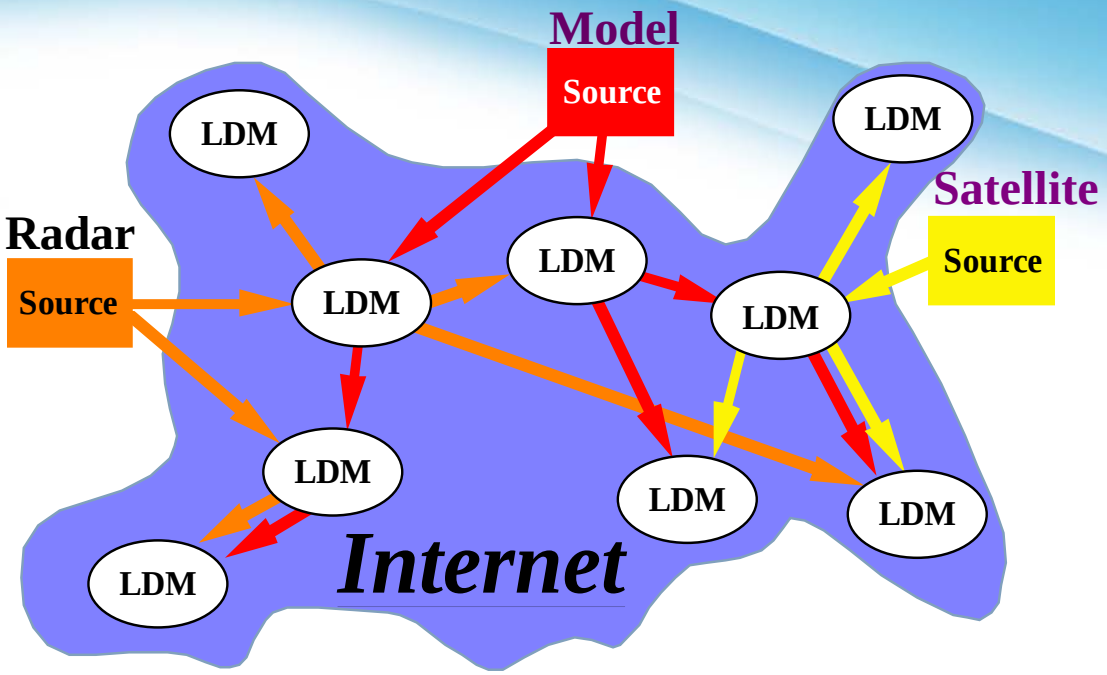
Data:
Over 30 data streams provided in real-time
Data collection, cataloging, and distribution
Both push and pull technologies are used

Software:
Data Distribution: LDM
Remote Data Access: THREDDS Data Server, ADDE, and RAMADDA
Data Management: netCDF, UDUNITS, and Rosetta
Analysis and Visualization: GEMPAK, McIDAS, IDV, and AWIPS II
GIS support via TDS (WCS, WMS) and KML and Shapefiles

User Support & Training:
Direct email support
Community mailing lists (~60)
Annual Training Workshops, Triennial Users Workshops, and Regional Workshops as needed.

Community:
Community Engagement; Equipment Awards to universities; Seminars; Advocacy;

Real-time Data Distribution



About 30 different streams of real-time weather data from diverse sources are provided to ~1250 computers worldwide.

Unidata's outbound traffic out of UCAR network is about 31 Terabytes/day. In fact, we move more data via Internet 2 than any other advanced application.

Remote Data Access

- Complements the IDD/LDM push data delivery system
- Available via THREDDS Data Server, RAMADDA, and ADDE data servers that support several protocols and APIs:
  - OPeNDAP
  - ADDE
  - HTTP
  - FTP
  - WCS and WMS

The Unidata Program Center operates a data server, that provides the above services. Nearly one terabyte of data are downloaded each day from our servers.

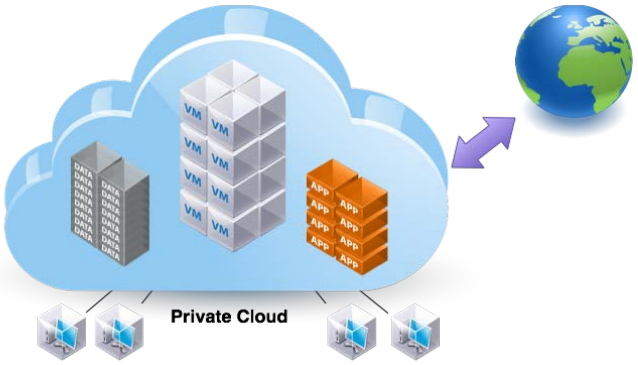
What is our motivation?

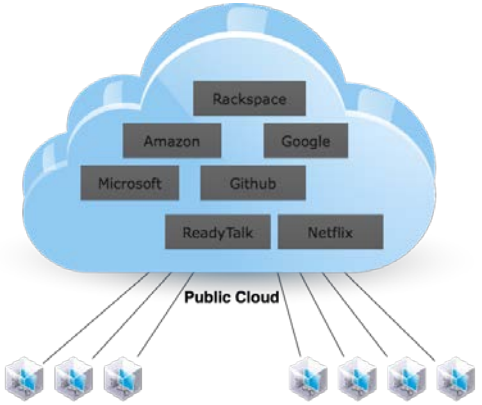
- **Data volumes** are getting to be **too large** to bring all of the data to your **local environment**.
- Need to keep data close to the point of origin or dissemination and provide the requisite tools and services and create a “playground” and workbench in the cloud.
- Bottom line: **We need to move from** “bringing the data to the scientist” **to** “bringing the science to the data”.
- **We would like to exploit the elasticity and easy virtualization aspects of the cloud.**

For these and other reasons, Unidata made a decision to transition data services to the cloud about 4 years ago.

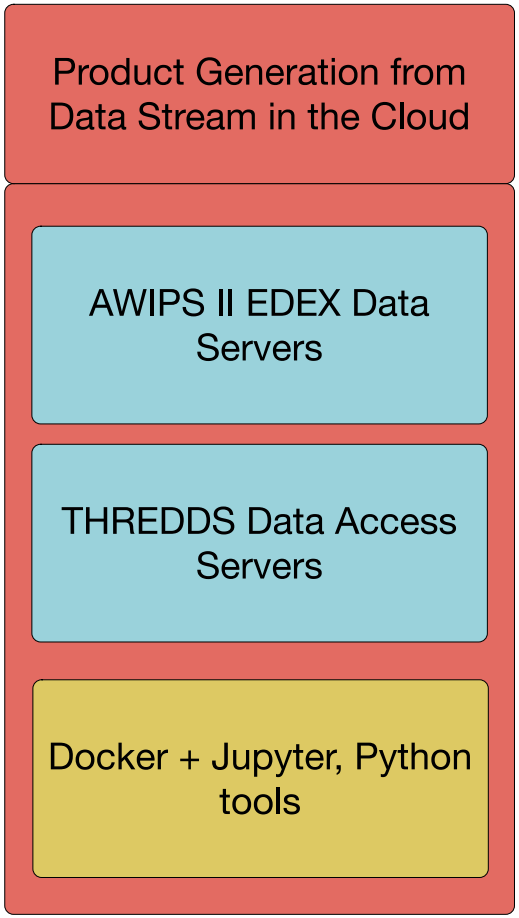
Goals for our Cloud work

- Along with providing data access, **develop and provide data-proximate processing, analysis and visualization services that are portable.**
- Provide portable, cloud-compatible software (i.e., Docker containers) that **users can run on their own cloud, private or public**.

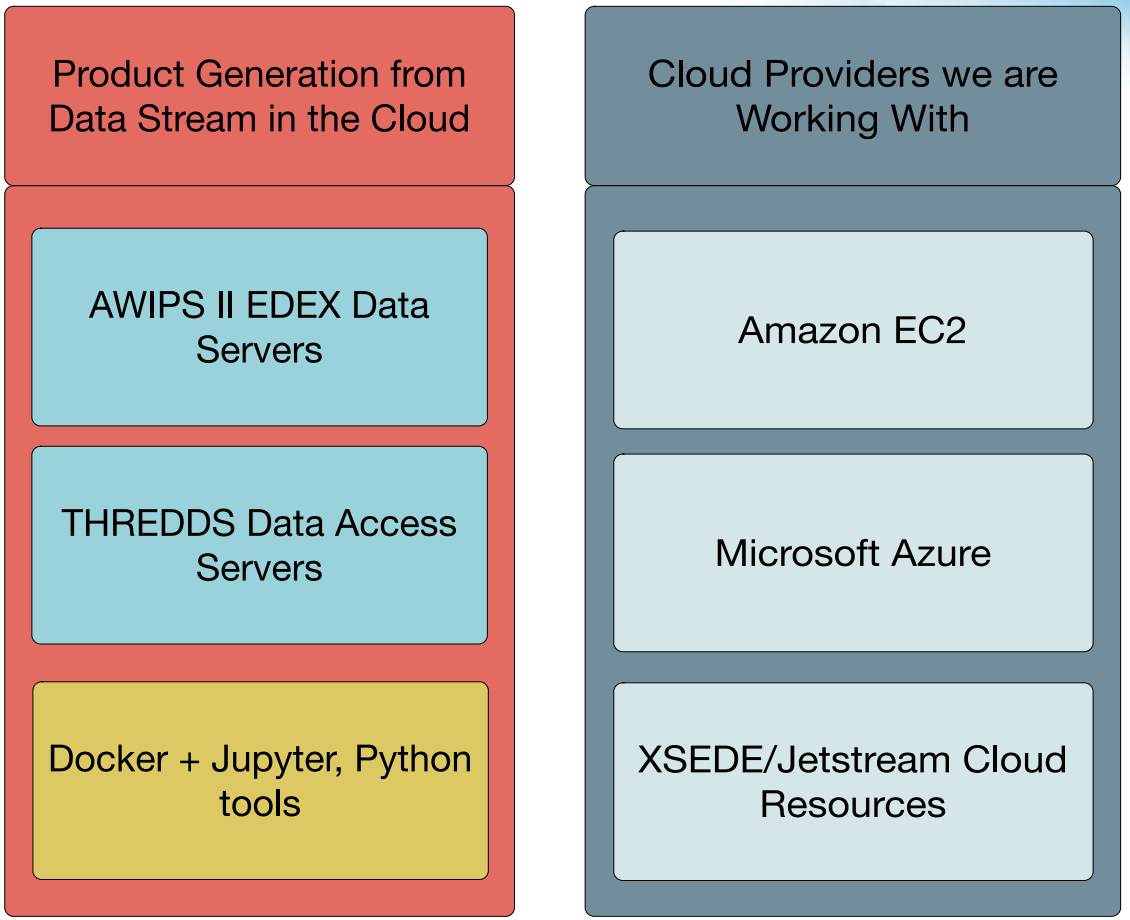




Unidata Cloud Projects

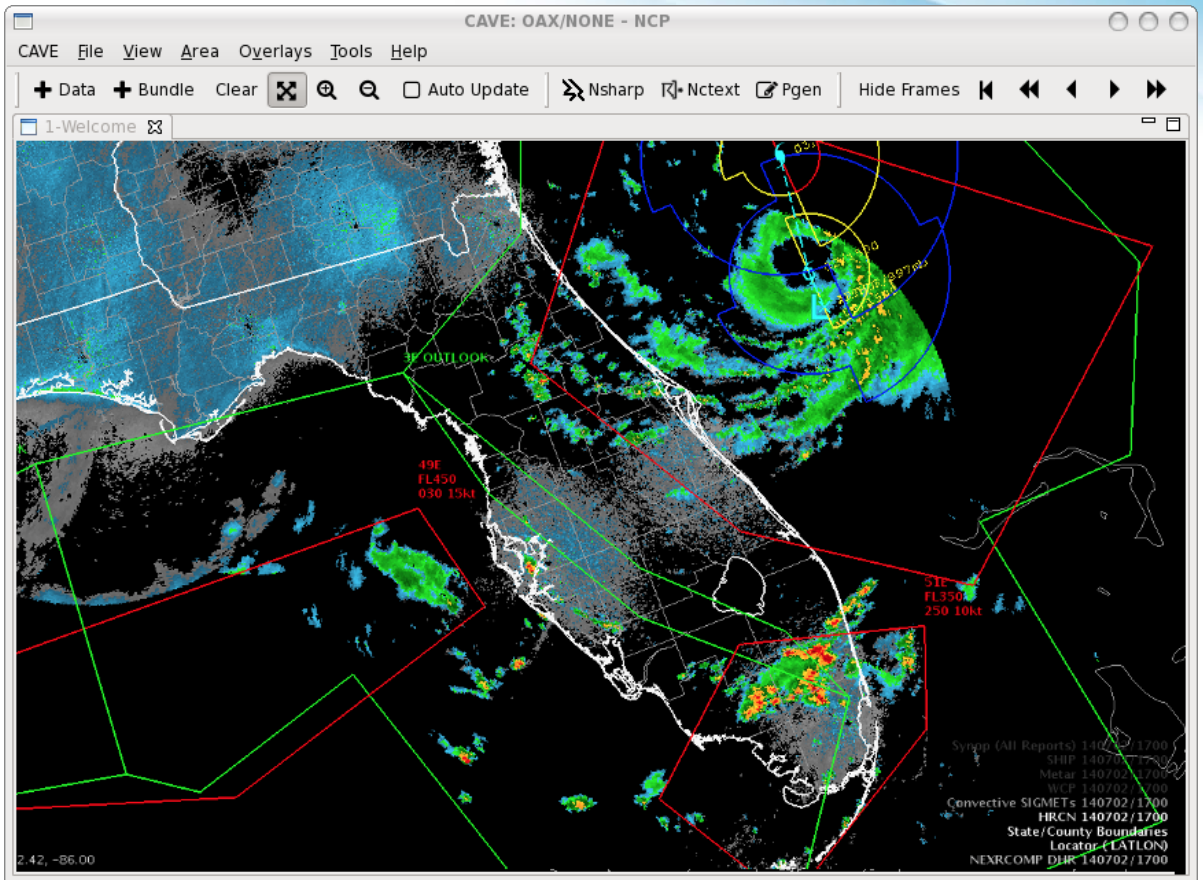


Unidata Cloud Partners



AWIPS Data Servers in the Cloud

Unidata is running AWIPS-EDEX data server in the Microsoft Azure cloud and exploring use in the Jetstream Cloud.



AWIPS Data Servers in the Cloud

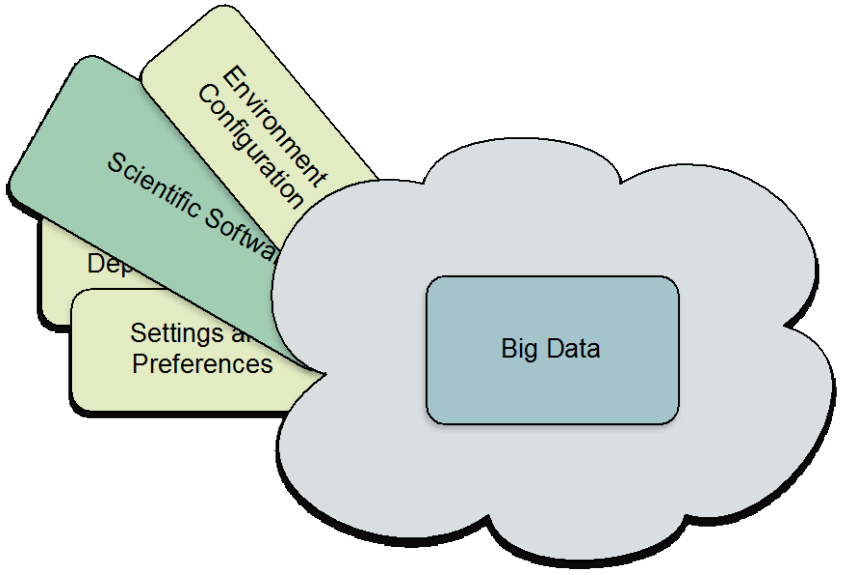
Unidata is running AWIPS-EDEX data server in the Microsoft Azure cloud and exploring use in the Jetstream Cloud.

alaska.edu
albany.edu
bu.edu
calu.edu
cmich.edu
cod.edu
colostate.edu
erau.edu
iastate.edu
illinois.edu
ku.edu
lyndonstate.edu
millersville.edu
missouri.edu
msudenver.edu
ncsu.edu
osu.edu
ou.edu
plymouth.edu
psu.edu
sdsmt.edu
sjsu.edu
slu.edu
sunysb.edu
tamu.edu
uah.edu
ucf.edu
ucla.edu
ucsd.edu
uga.edu
uic.edu
uiuc.edu
umd.edu
umich.edu
unca.edu
uncc.edu
unl.edu
uprm.edu
uwm.edu
uwyo.edu
valpo.edu
wisc.edu
wku.edu
yorku.ca

44 universities are using Unidata's Azure-hosted EDEX.

Easing the Community Burden when Deploying Software in the Cloud

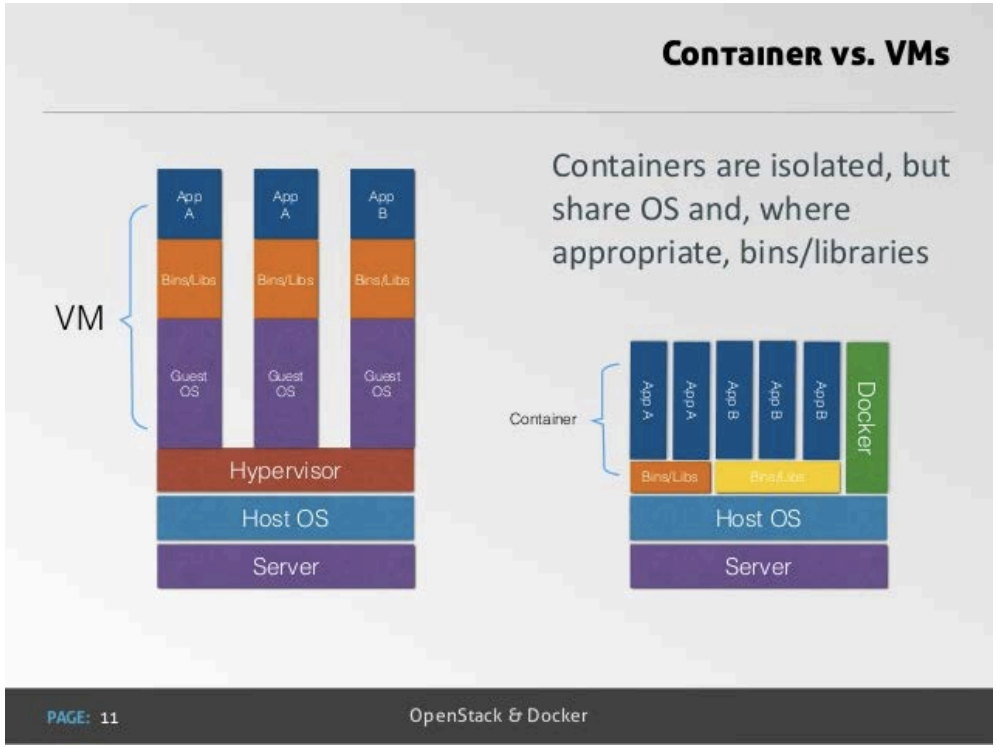
- Deploying services **to** the cloud/maintaining services **in** the cloud can be **complicated and time consuming.**



Easing the Community Burden by Deploying Portable Software in the Cloud

- Solution: Containerization, e.g. **Docker**

Virtual Machines vs. Containers



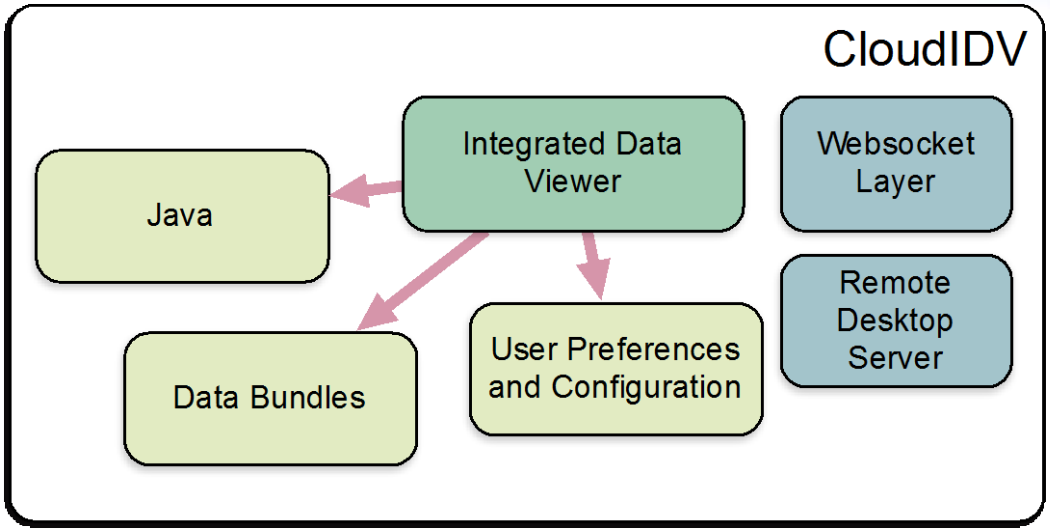
Docker Benefits

- **Small Footprint**
- **Rapid and Lightweight Deployment**
- **Portability**
- **Reuse**

Containerizing Applications

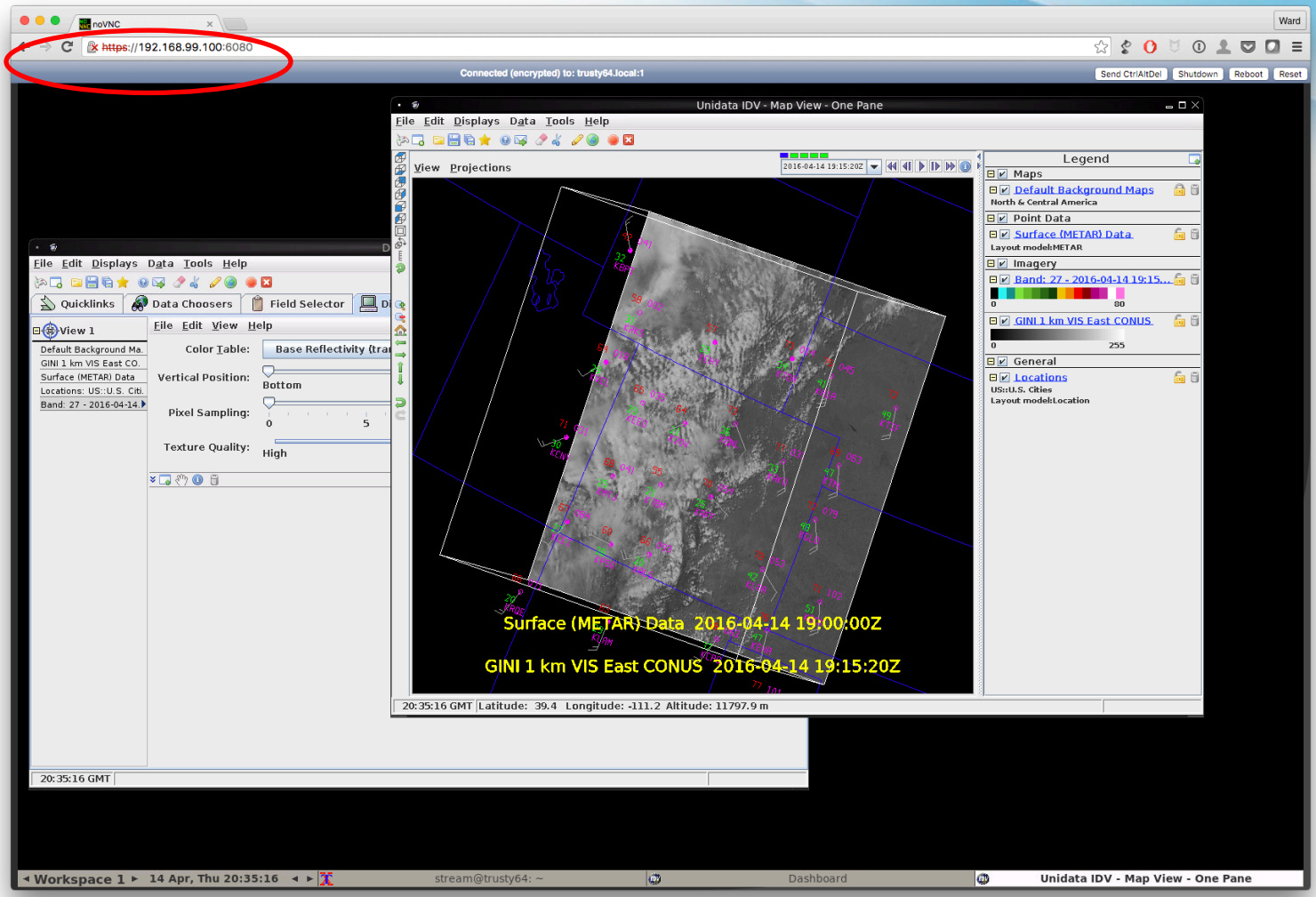
- We have created **Docker container image**s for several Unidata applications, including the **Integrated Data Viewer (IDV)**, **THREDDS Data Server**, **Local Data Manager (LDM)**, and many Python tools.
- We have been deploying these applications in our **own cloud instances** and also making them available as downloadable software to our users.
- We have released a technology stack (dubbed **CloudStream)** to make it easy to deploy **desktop software** (as opposed to server software) in **the cloud**.

CloudIDV



The CloudIDV Docker image contains the **standard IDV** as well as **all of the technology required** to run it in the cloud, **accessed via browser.**

CloudIDV


noVNC
https://192.168.99.100:6080
Connected (encrypted) to: trusty64.local:1
Send CtrlAltDel
Shutdown
Reboot
Reset
Unidata IDV - Map View - One Pane
File Edit Displays Data Tools Help
View Projections
Legend
Maps
Default Background Maps
North & Central America
Point Data
Surface (METAR) Data
Layout model:METAR
Imagery
Band: 27 - 2016-04-14 19:15...
GINI 1 km VIS East CONUS
General
Locations
US::U.S. Cities
Layout model:Location
Surface (METAR) Data 2016-04-14 19:00:00Z
GINI 1 km VIS East CONUS 2016-04-14 19:15:20Z
20:35:16 GMT Latitude: 39.4 Longitude: -111.2 Altitude: 11797.9 m
File Edit Displays Data Tools Help
Quicklinks
Data Choosers
Field Selector
View 1
Default Background Ma.
GINI 1 km VIS East CO.
Surface (METAR) Data
Locations: US::U.S. Citi.
Band: 27 - 2016-04-14...
File Edit View Help
Color Table:
Vertical Position:
Bottom
Pixel Sampling:
Texture Quality:
High
20:35:16 GMT
Workspace 1
14 Apr, Thu 20:35:16
stream@trusty64: ~
Dashboard
Unidata IDV - Map View - One Pane

Remote Data Analysis & Visualization

- In addition to enabling cloud-hosted data access, Unidata is leveraging cloud technologies to enable data proximate analysis and visualization capabilities.
- Specifically, Unidata is integrating the capabilities of THREDDS Data Server and AWIPS II EDEX Server, Jupyter Notebook platform, Siphon Python data access tool, and MetPy/CartoPy/Matplotlib, IDV and GEMPAK analysis and visualization applications.

Siphon

MetPy

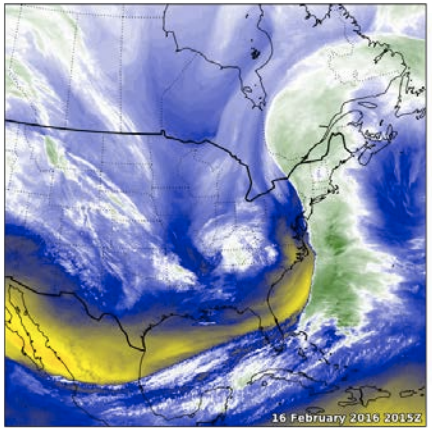
TDS+Siphon+Python Plotting

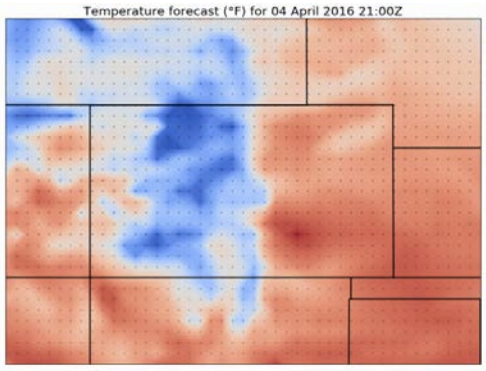
Siphon

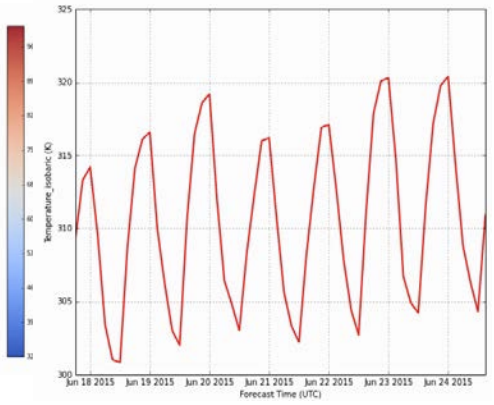
Siphon is a collection of Python utilities for downloading data from Unidata data technologies. Siphon's current functionality focuses on access to data hosted on a THREDDS Data Server.

Using Siphon to query the NetCDF Subset Service and plotting it to a map

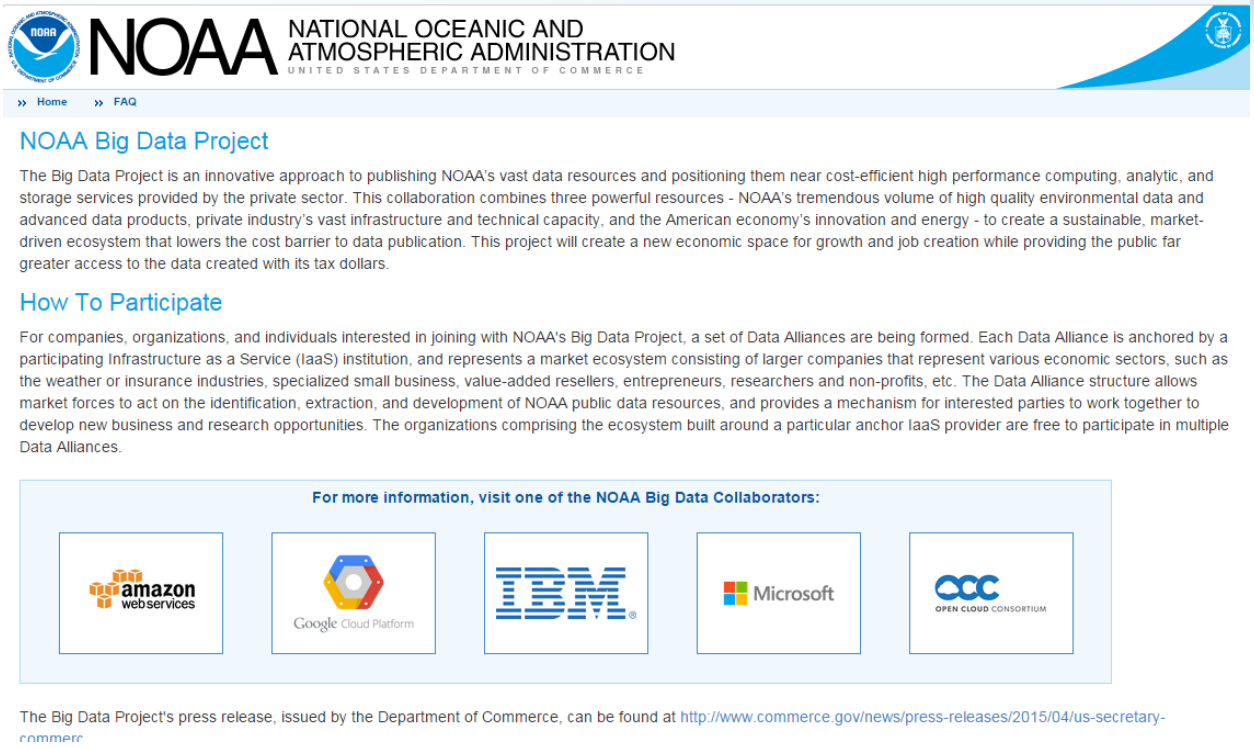
```
from siphon.catalog import TDSCatalog
best_gfs = TDSCatalog('http://thredds.ucar.edu/thredds/catalog/grib/NCEP/GFS/'
                      'Global_0p25deg/catalog.xml?dataset=grib/NCEP/GFS/Global_0p25deg/Best')
print(list(best_gfs.datasets))
```







NOAA Big Data Project and Unidata Cloud Activities



NOAA NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
UNITED STATES DEPARTMENT OF COMMERCE

» Home » FAQ

NOAA Big Data Project

The Big Data Project is an innovative approach to publishing NOAA's vast data resources and positioning them near cost-efficient high performance computing, analytic, and storage services provided by the private sector. This collaboration combines three powerful resources - NOAA's tremendous volume of high quality environmental data and advanced data products, private industry's vast infrastructure and technical capacity, and the American economy's innovation and energy - to create a sustainable, market-driven ecosystem that lowers the cost barrier to data publication. This project will create a new economic space for growth and job creation while providing the public far greater access to the data created with its tax dollars.

How To Participate

For companies, organizations, and individuals interested in joining with NOAA's Big Data Project, a set of Data Alliances are being formed. Each Data Alliance is anchored by a participating Infrastructure as a Service (IaaS) institution, and represents a market ecosystem consisting of larger companies that represent various economic sectors, such as the weather or insurance industries, specialized small business, value-added resellers, entrepreneurs, researchers and non-profits, etc. The Data Alliance structure allows market forces to act on the identification, extraction, and development of NOAA public data resources, and provides a mechanism for interested parties to work together to develop new business and research opportunities. The organizations comprising the ecosystem built around a particular anchor IaaS provider are free to participate in multiple Data Alliances.

For more information, visit one of the NOAA Big Data Collaborators:

amazon web services | Google Cloud Platform | IBM | Microsoft | OCC Open Cloud Consortium

The Big Data Project's press release, issued by the Department of Commerce, can be found at http://www.commerce.gov/news/press-releases/2015/04/us-secretary-commerc

Unidata is collaborating with Amazon Web services and Open Commons Consortium CRADA Partners.

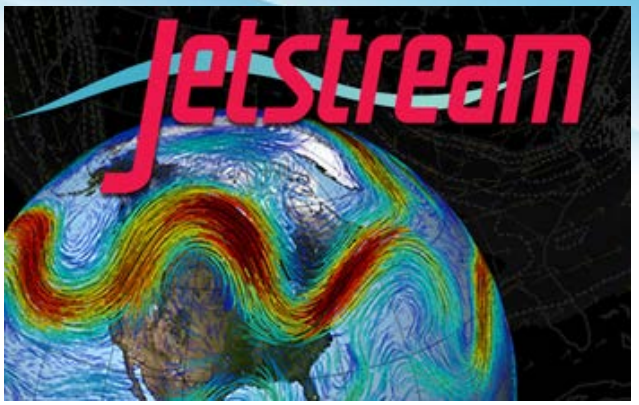
Collaborative Activities with AWS

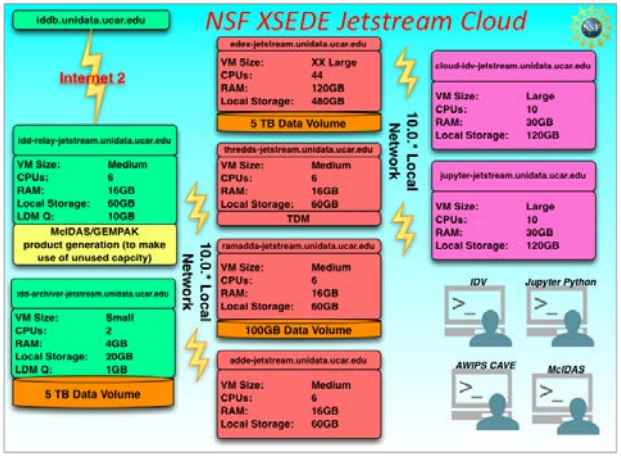
- Streaming real-time NEXRAD radar data to AWS/S3 operationally using the Unidata LDM software. We are continuing our partnership and now moving GOES-16 data. We will next start moving NCEP model output (including the National Water Model Output) to AWS.
- **Running Docker-containerized THREDDS Data Server to serve radar data from AWS/S3.**
- Providing JupyterHub multi-user Python environment, including plotting tools.
- **Providing individual Docker containers.**
- We are continuing our partnership with AWS on the NOAA Big Data Project and on other Unidata efforts, including the provision of GOES-16 data

For the first time, users have seamless access to both historical and real-time WSR-88D Radar Data from the same location and interface.

Cloud Partnerships

- Unidata received a sizeable allocation on the NSF XSEDE Jetstream cloud.
- We are currently deploying an array of Unidata services in that environment.





Thank You

Unidata is one of the University Corporation for Atmospheric Research (UCAR)'s Community Programs (UCP), and is funded primarily by the National Science Foundation (Grant NSF-1344155).