A Cloud-based Science Gateway for Enabling Science as a Service to Facilitate Open Science and Reproducible Research

Mohan Ramamurthy and Julien Chastang
Unidata, University Corporation for Atmospheric Research

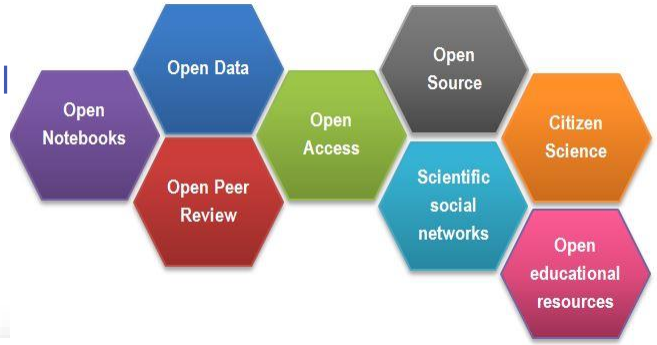
Unidata

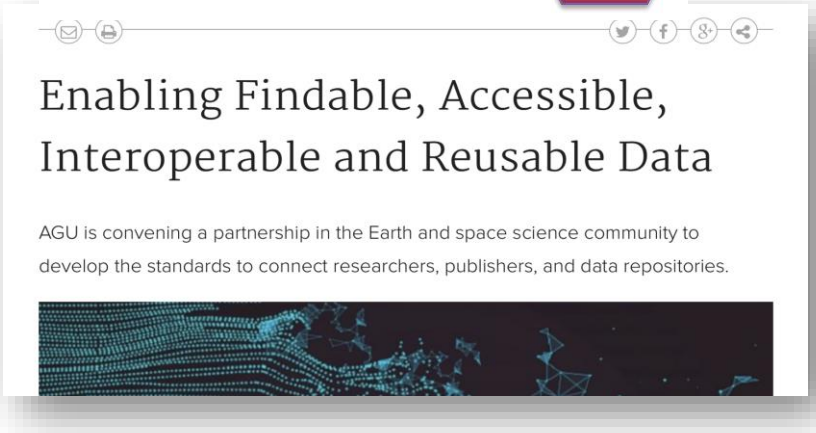
Unidata is a geoscience data and software facility funded primarily by the U. S. National Science Foundation

- Acquire and distribute real-time meteorological data;
- Acquire, develop, and provide software for accessing, managing, analyzing, visualizing geoscience data;
- Provide training and support to users;

Open and Reproducible Science

Tabl





Enabling Findable, Accessible, Interoperable and Reusable Data

AGU is convening a partnership in the Earth and space science community to develop the standards to connect researchers, publishers, and data repositories.

Motivation

- **Data volumes** are getting to be **too large** to bring all of the data to your **local environment**.
- Instead of “bringing data to scientists” **we need to** “bring their science to the data”. Users should not only be able to use data and applications provided to them, but also upload their own data and software to their own workspace and use them collectively.
- **In doing so, we should exploit the on-demand resource provision, elasticity, scalability and virtualization aspects of the "cloud."**

Approach

- The cloud-enabled deployment of **Infrastructure as a Service, Data as a Service,** and **Software as a Service** is key to the delivery of **Science as a Service**.
- Transform “client-side” applications to become “server-side” tools that reside close to the location of data. (Similar to Office 365 and Adobe Photoshop)
- Connect those server-side applications and data via workflows by providing an ecosystem of tools and services for scientists.

Science Gateway Attributes

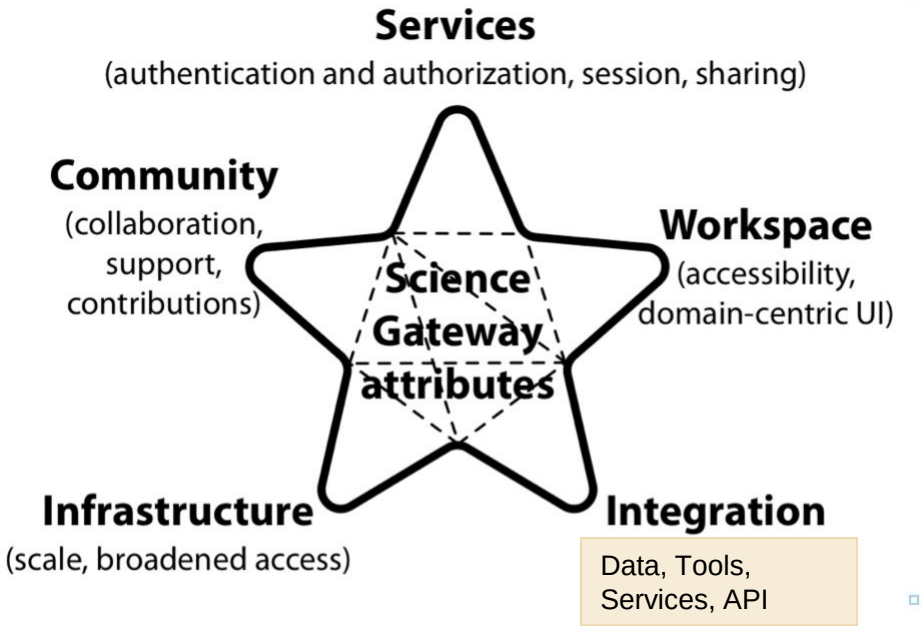
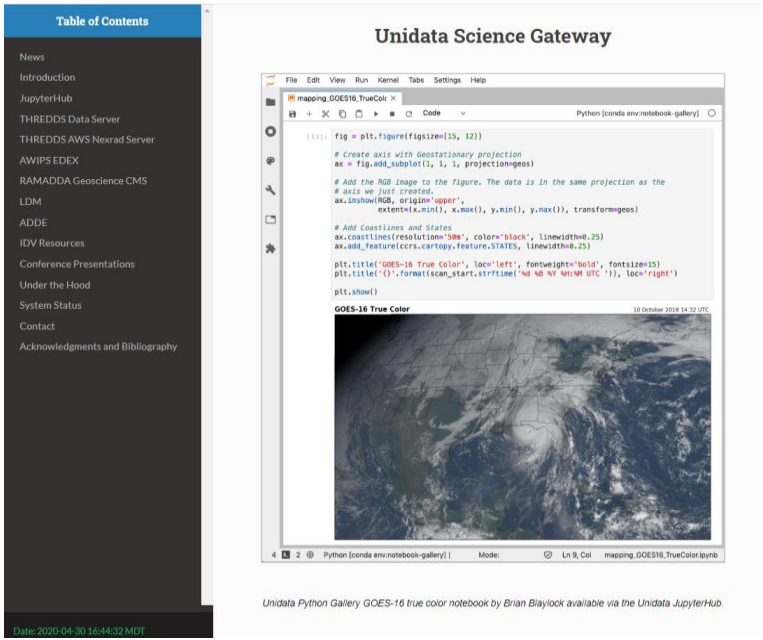
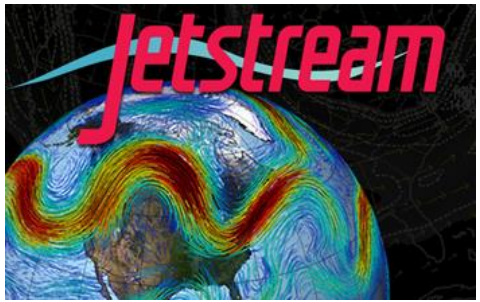


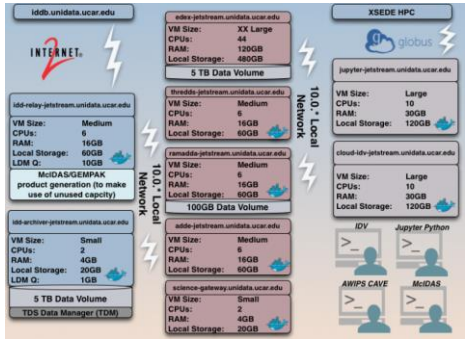
Fig. 2. Interconnected attributes of a science gateway.

Pierce et al., 2018

Unidata Science Gateway





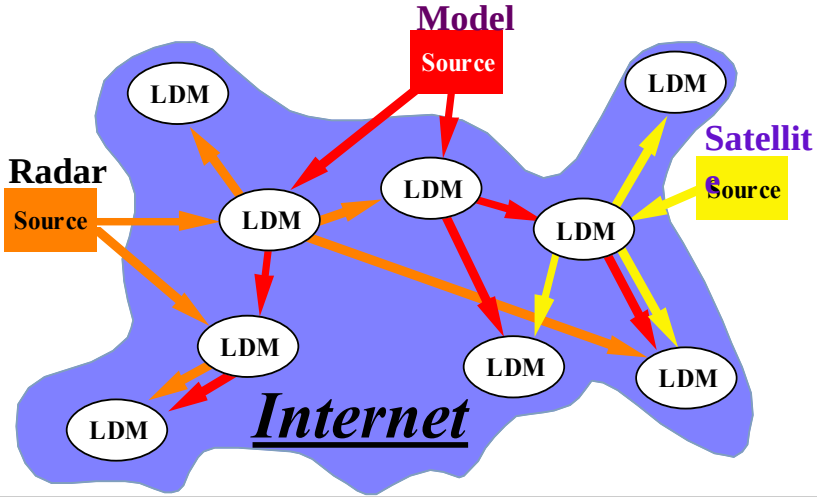


http://science-gateway.unidata.ucar.edu/

Gateway Services

- Access to data via OPeNDAP, ADDE, WCS, WMS, netCDF APIs, including subsets of data
- Data transformation and format conversion
- Extensive data analysis capabilities
- Visualization of data provided by the gateway
- Access to a collection of Jupyter Notebooks
- Access to AWIPS EDEX data server

Real-time Data Hosting



More than a terabyte of data from 30 different streams of real-time data are moved each day and hosted on the Unidata Science Gateway.

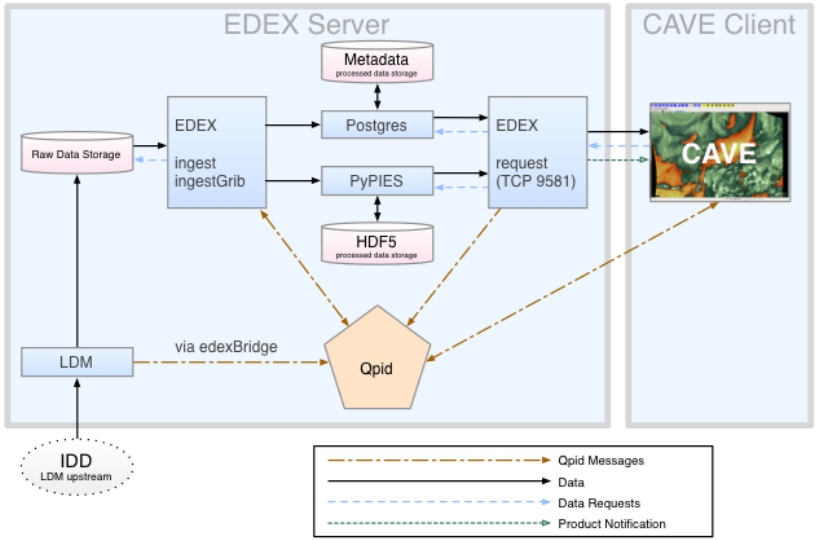
Satellite
Radar
Lightning
Surface, upper-air observations
Operational Weather Model output, …
…

AWIPS Data Servers in the Cloud

- AWIPS is the system for processing, integrating, and displaying all types of meteorological and hydrological data in National Weather Service offices.
- Unidata distributes a version of AWIPS to the university community.

AWIPS has a data-server component called **EDEX** and a client side application, **CAVE**.

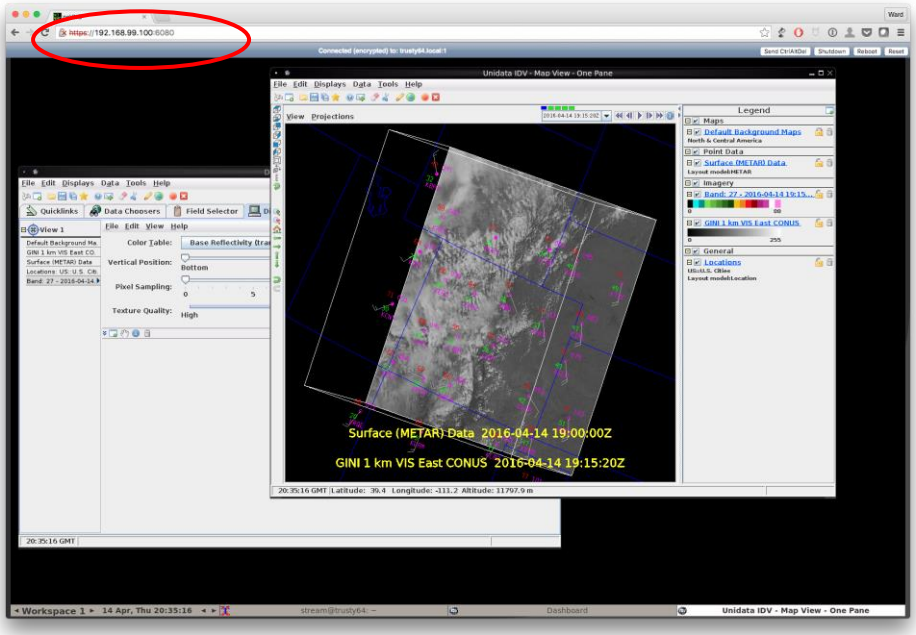
Unidata is running AWIPS-EDEX data server in the cloud.



Universities Using Cloud AWIPS

alaska.edu
calit2.net
clemson.edu
cmich.edu
cornell.edu
colorado.edu
colostate.edu
csusb.edu
erau.edu
fit.edu
gatech.edu
hawaii.edu
iowa.edu
Illinois.edu
Indiana.edu
iugaza.edu.ps
iuk.edu
Iupui.edu
kansas.edu
kctcs.edu
ku.edu
lyndonstate.edu
louisville.edu
millersville.edu
missouri.edu
msstate.edu
msudenver.edu
mtu.edu
ncsu.edu
niu.edu
nps.edu
ohio-state.edu
oregonstate.edu
ou.edu
plymouth.edu
psu.edu
purdue.edu
sdsmt.edu
sjsu.edu
slu.edu
southern.edu
stonybrook.edu
tamu.edu
uah.edu
uark.edu
ucar.edu
ucdavis.edu
ucla.edu
ucsd.edu
udel.edu
ufl.edu
uga.edu
uiuc.edu
ulm.edu
umd.edu
umich.edu
unca.edu
und.edu
unl.edu
utah.edu
utexas.edu
uvic.edu
uwm.edu
uwyo.edu
valpo.edu
vt.edu
wisc.edu
wku.edu

CloudIDV



Benefits

- No need to download, install, and keep updating software
- IDV can be co-located with data, reducing data movement.
- No longer limited by desktop/laptop memory and network bandwidth.

Unidata is currently working with a small group of early adopters to work out the details.

Remote Data Analysis & Visualization

- Unidata is also leveraging other cloud technologies to enable data proximate analysis and visualization capabilities.

- Specifically, Unidata is integrating the capabilities of THREDDS Data Server, Jupyter Notebook platform, Siphon Python data access tool, and MetPy/CartoPy/Matplotlib/xarray and GEMPAK analysis and visualization applications.

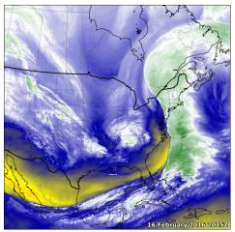
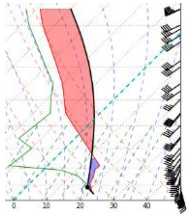
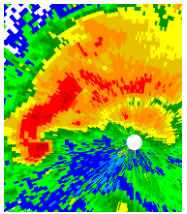
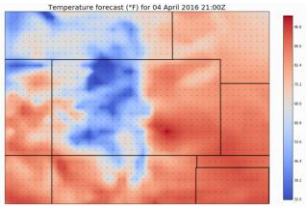
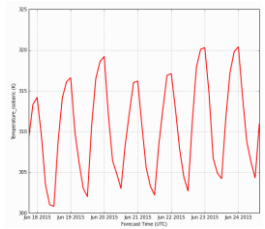
Geospatial Analysis& Visualization Jupyter Notebooks:TDS+Siphon+MetPy

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))
```

Universities are now using these Notebooks in their classes.

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 on GitHub as downloadable software to our users.

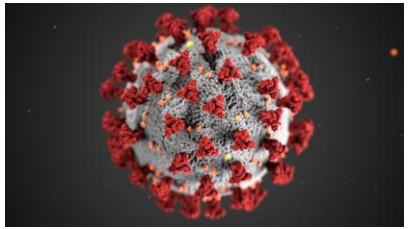
Open data access, together with the use of Container Applications, IDV bundles, and Jupyter Notebooks, is practical way to advance Reproducibility and Open Science

Scaling of JupyterHub Services

Thanks to the help from XSEDE Extended Collaborative Support Services (ECSS), Unidata is using Kubernetes to scale its JupyterHub Services so that it can support large number of users.

COVID-19: Use of Unidata Science Gateway

- Unidata offered its JupyterHub resources on its Science Gateway on Jetstream to universities for Remote-Learning Situations.
  - Because many universities are responding to local public health mandates by transitioning to the use of remote-learning techniques and online-only courses, Unidata is actively extending its offer of JupyterHub resources to universities.
  - Seven universities are using these resources to augment their teaching.



Acknowledgement

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 AGS-1901712.