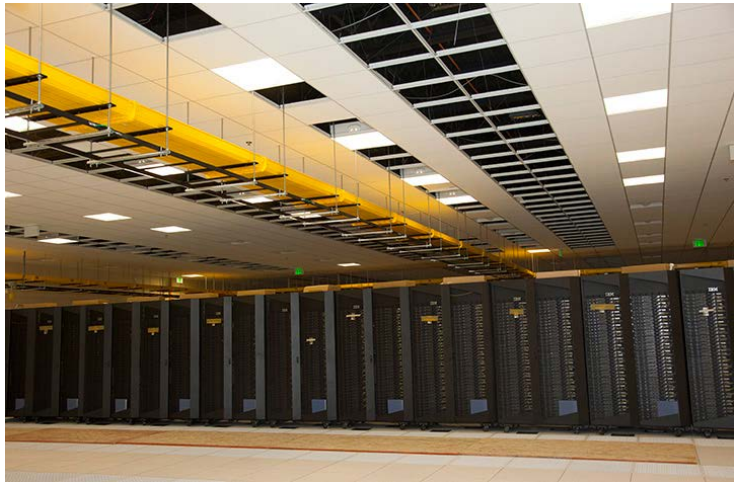
Modeling Research in the Cloud:
Workshop Goals, Opportunities and Challenges

31 May 2017
Boulder, CO

Mohan Ramamurthy
Unidata Program Center
University Corporation for Atmospheric Research
Boulder, CO 80301
USA

High Performance Computing

- Historically, the atmospheric modeling community has relied mainly on high performance computing facilities and local systems for modeling studies.



Cloud Computing

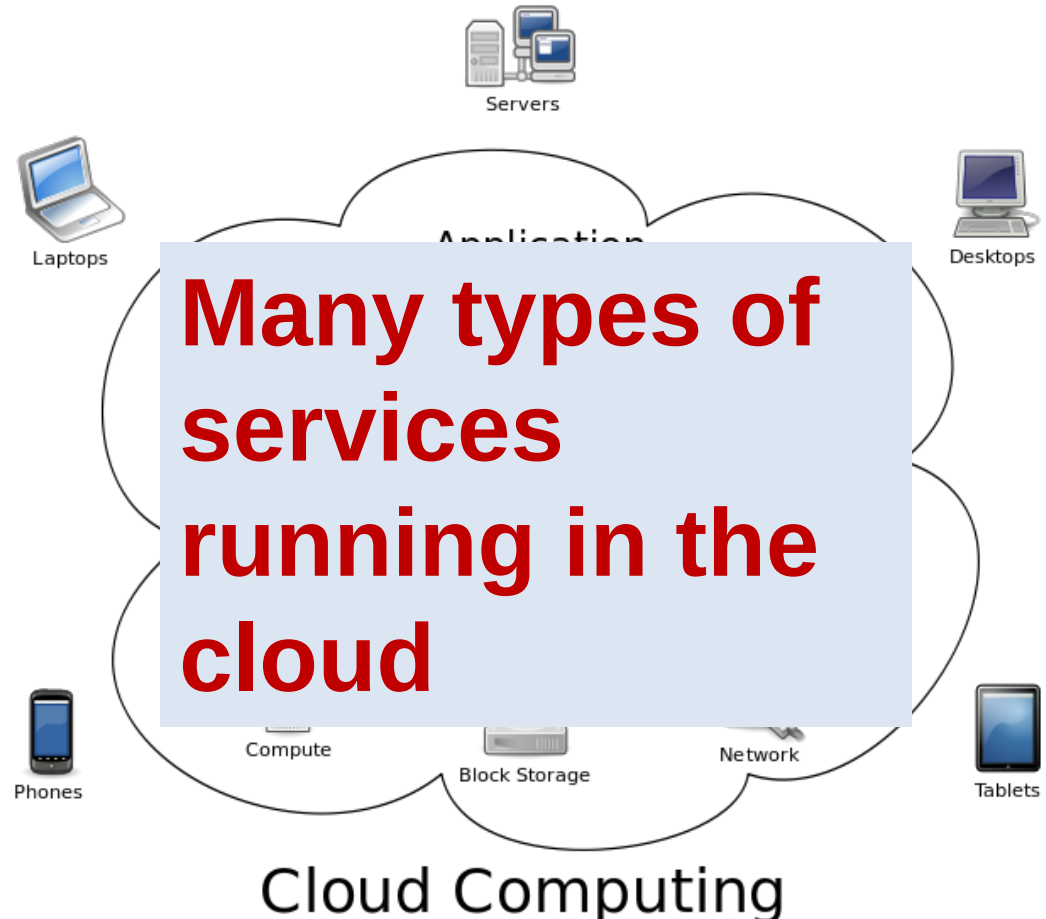
"A pool of abstracted, highly scalable, and managed computer infrastructure capable of hosting end-customer applications and billed by consumption[1]"

"Cloud computing is an emerging approach to shared infrastructure in which large pools of systems are linked together to provide IT services."

Cloud computing came into existence circa 2000, with Amazon introducing *Elastic Cloud Computing* in 2006 and Microsoft launching *Azure* in 2008.

Cloud Computing

- The “cloud” provides elasticity, virtualization, and on-demand (rapid provisioning of) computational, storage, networking, and software services via the Internet.



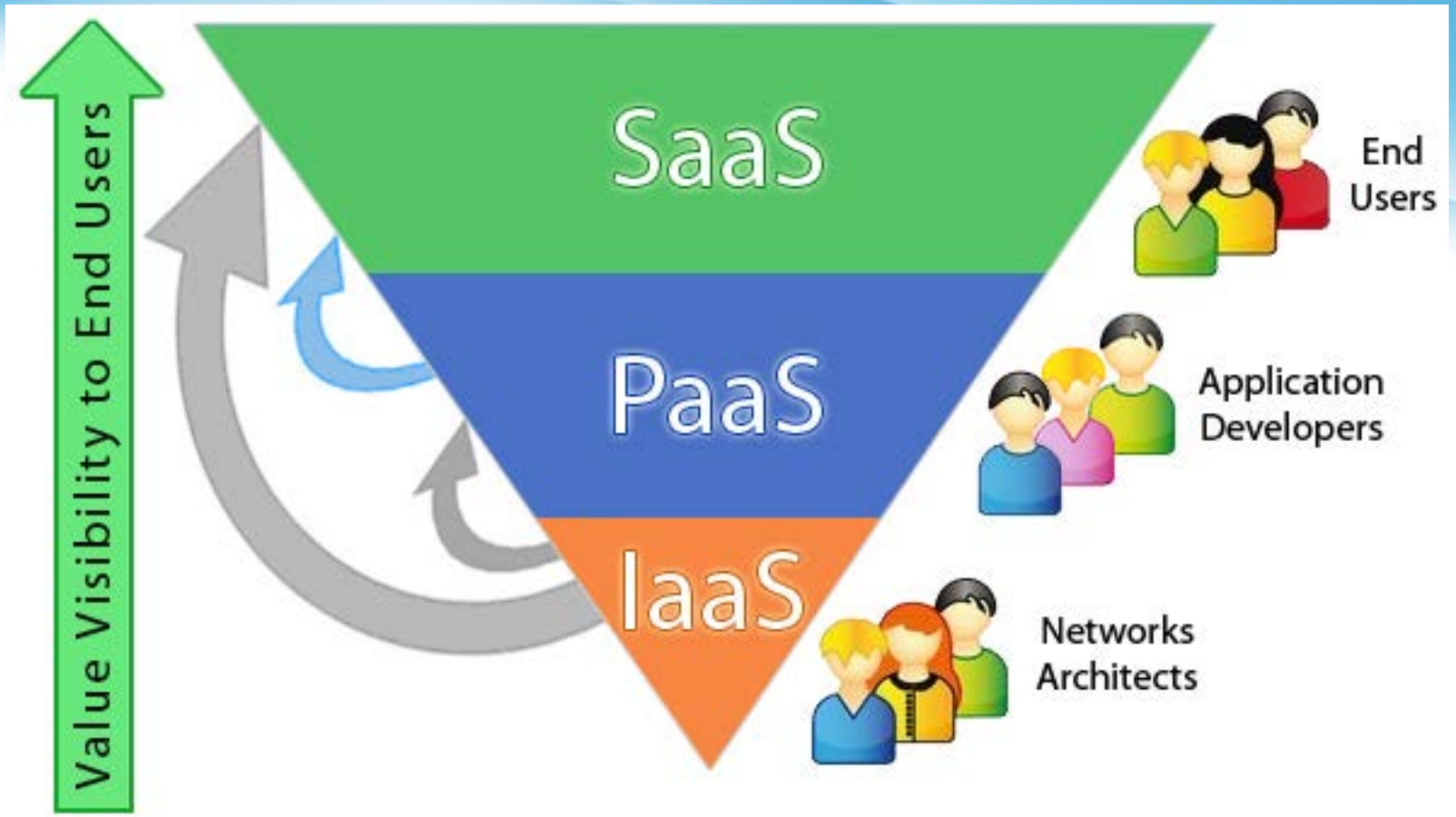
Many types of services running in the cloud

- Any Resource
- Any Time
- Any Device
- Any Place

Cloud Computing

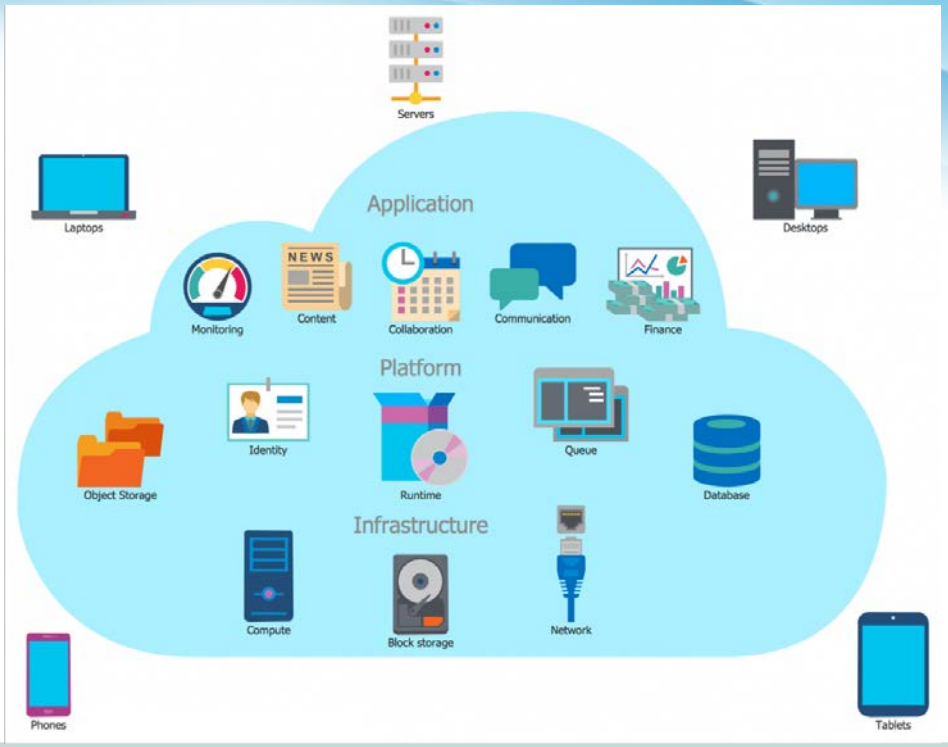
- In the past decade, cloud computing has matured significantly and has become a robust and reliable computing environment.
- It represents a fundamental shift in the way IT services are developed, deployed, operated, and paid for, creating a major paradigm shift.
- The cloud has the potential for researchers to gain access to computing resources beyond the traditional supercomputing centers for modeling studies.
- **As a result, more organizations are either exploring or have already migrated their workflows and operations to the cloud.**

Cloud Computing: Types of Services



In the last few years, there has been a remarkable surge in the adoption of cloud services in all kinds of enterprises.

A Typical Cloud Environment

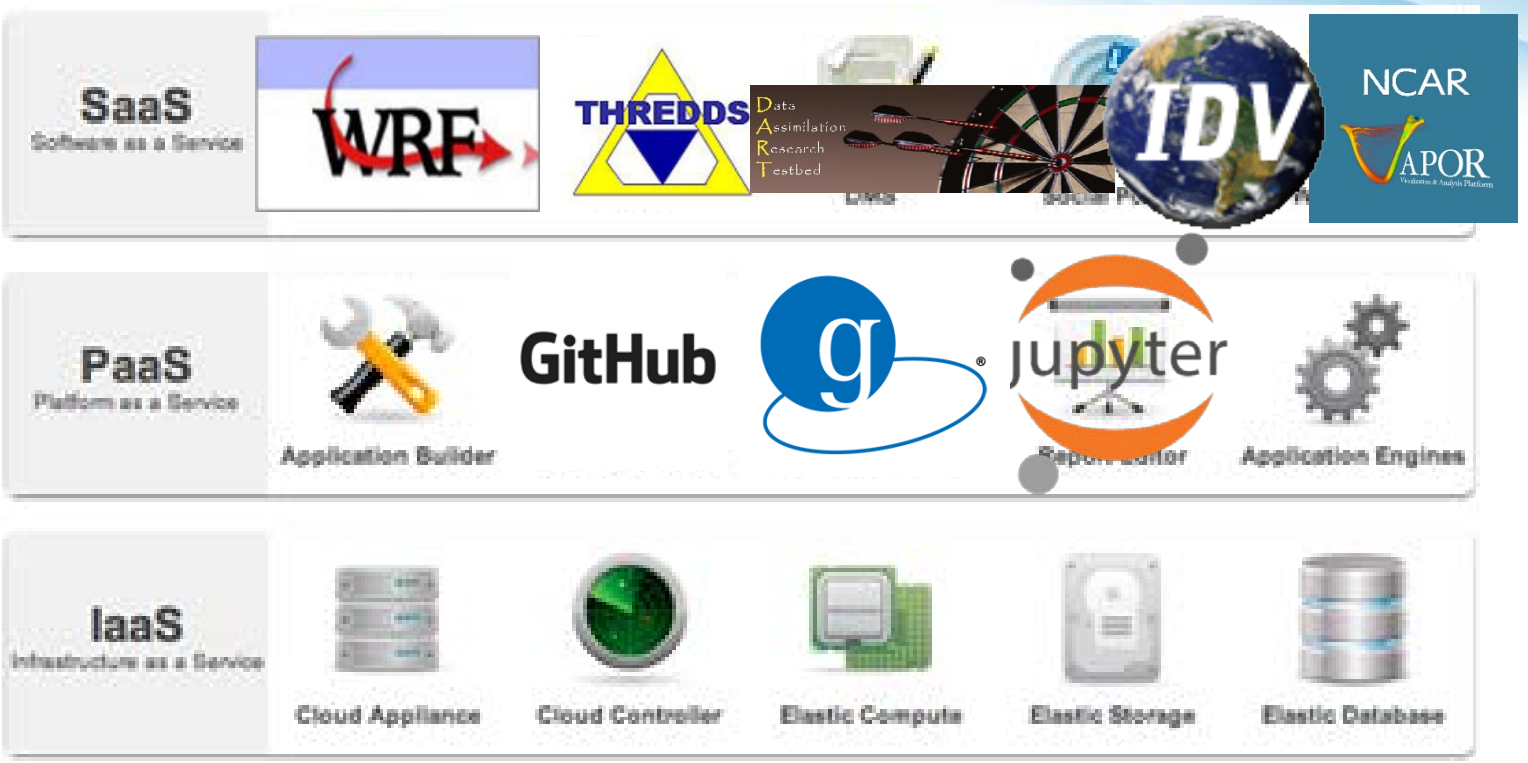


The rapid adoption and fierce competition among cloud vendors has resulted in a dramatic decrease in the cost of cloud computing and a sharp increase in the number of applications available. Cloud computing is now commodity that is traded in the spot market!

Rationale

- Data volumes are getting to be too large to keep moving data around.
- Need to keep data close to the point of origin or dissemination and provide an ecosystem of data, tools and services, **creating a workbench in the cloud with end-to-end capabilities that researchers can use**.
- End to end capabilities include: first guess fields and observations for data assimilation, initial and boundary conditions, the model (e.g., WRF) and model output, verification data, and tools for post-processing, analysis, visualization, verification and product generation.
- **We need to move from "bringing the data to the processing" to "bringing processing to the data".**

A Typical Cloud Stack



Imagine an environment that enables interlinked applications, services, and workflows that facilitate various components of modeling research.

Data-proximate Analysis and Visualization

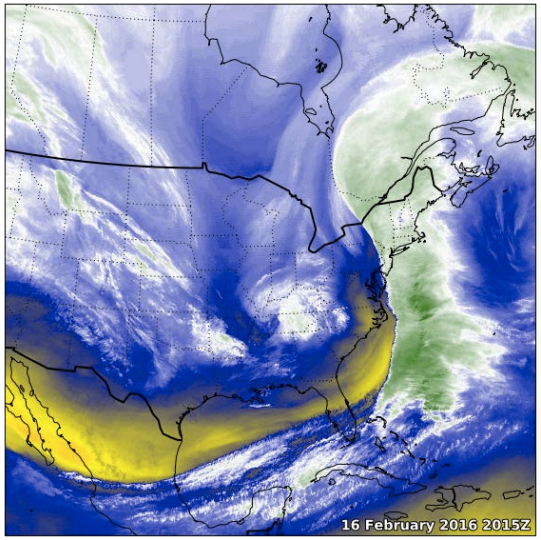
THREDDS TDS

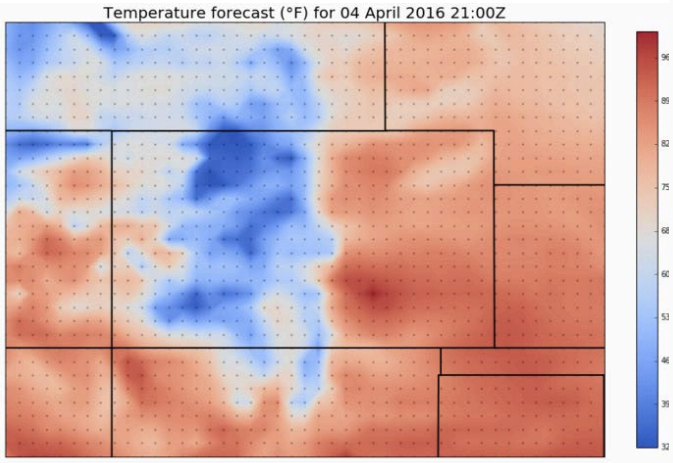
jupyter

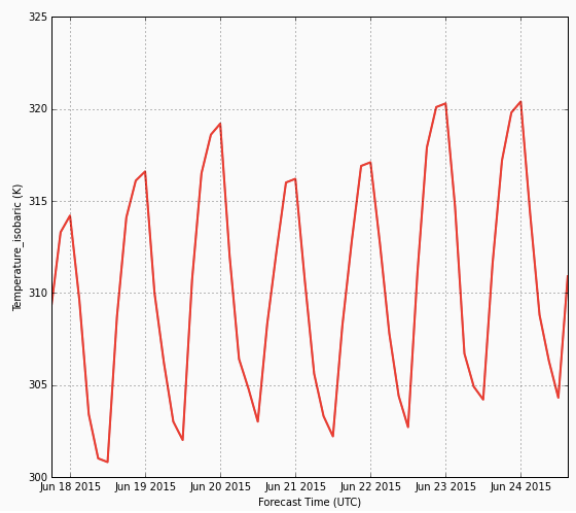
Siphon

MetPy

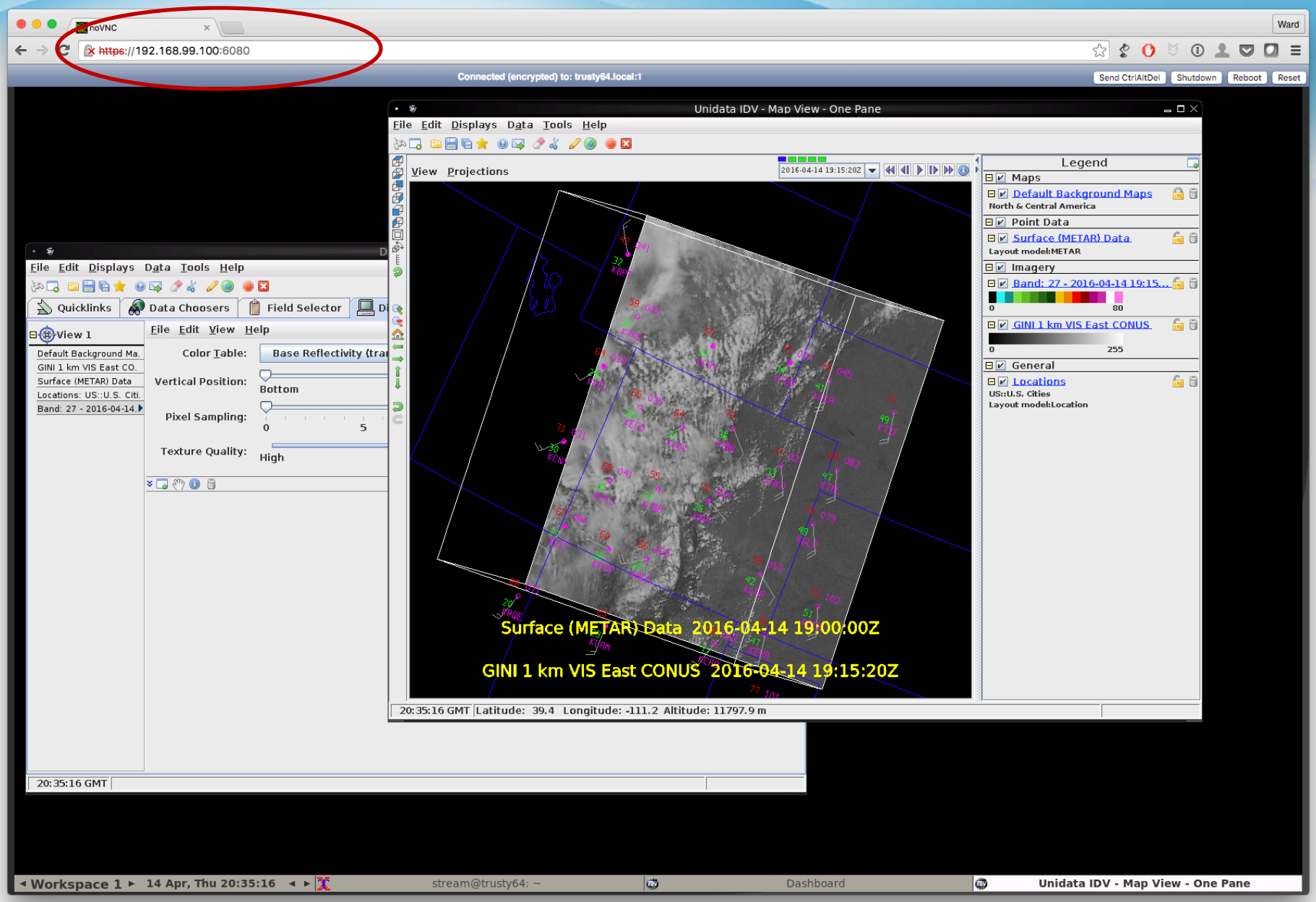
Using Siphon to query the NetCDF Subset Service and plotting it to a map







CloudIDV


noVNC
https://192.168.99.100:6080
Ward
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
2016-04-14 19:15:20Z
Legend
Maps
Default Background Maps
North & Central America
Point Data
Surface (METAR) Data
Layout model:METAR
Imagery
Band: 27 - 2016-04-14 19:15...
0
80
GINI 1 km VIS East CONUS
0
255
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:
Base Reflectivity (tra
Vertical Position:
Bottom
Pixel Sampling:
0
5
Texture Quality:
High
20:35:16 GMT
Workspace 1
14 Apr, Thu 20:35:16
stream@trusty64: ~
Dashboard
Unidata IDV - Map View - One Pane

Opportunities & Challenges

- The ease of acquiring and provisioning resources and services, the elasticity to scale, and virtualization provide a significant opportunity for the prediction community, much like it has done for the business community.
- While cloud computing has tremendous potential to transform the conduct of science in general and modeling research in particular, there remain key challenges as well.
  - Cost of procuring and paying for resources on an ongoing basis – it is a very different business model;
  - Understanding the sweet spot – what types of problems and situations are best suited for migration to the cloud;
  - Building trust and confidence in the paradigm;
  - Researchers' workflows will need to be altered;
  - Social and cultural challenges

Workshop Goals

- Raise awareness of the emerging technologies in the age of cloud computing and data-driven science in the atmospheric prediction community;
- Understand the different cloud options that are available, discuss opportunities, challenges and benefits of conducting modeling research in the cloud;
- Investigate what types of modeling research, scales of problems, and operational situations will benefit from using the cloud vs. using local clusters and HPC facilities;
- discuss a way forward and future steps to explore an end-to-end data, computational, software and modeling ecosystem in the cloud to advance modeling research.

Thank you!
Questions and Discussion

The planned workshop is funded by the National Science Foundation Award NSF-1651316.