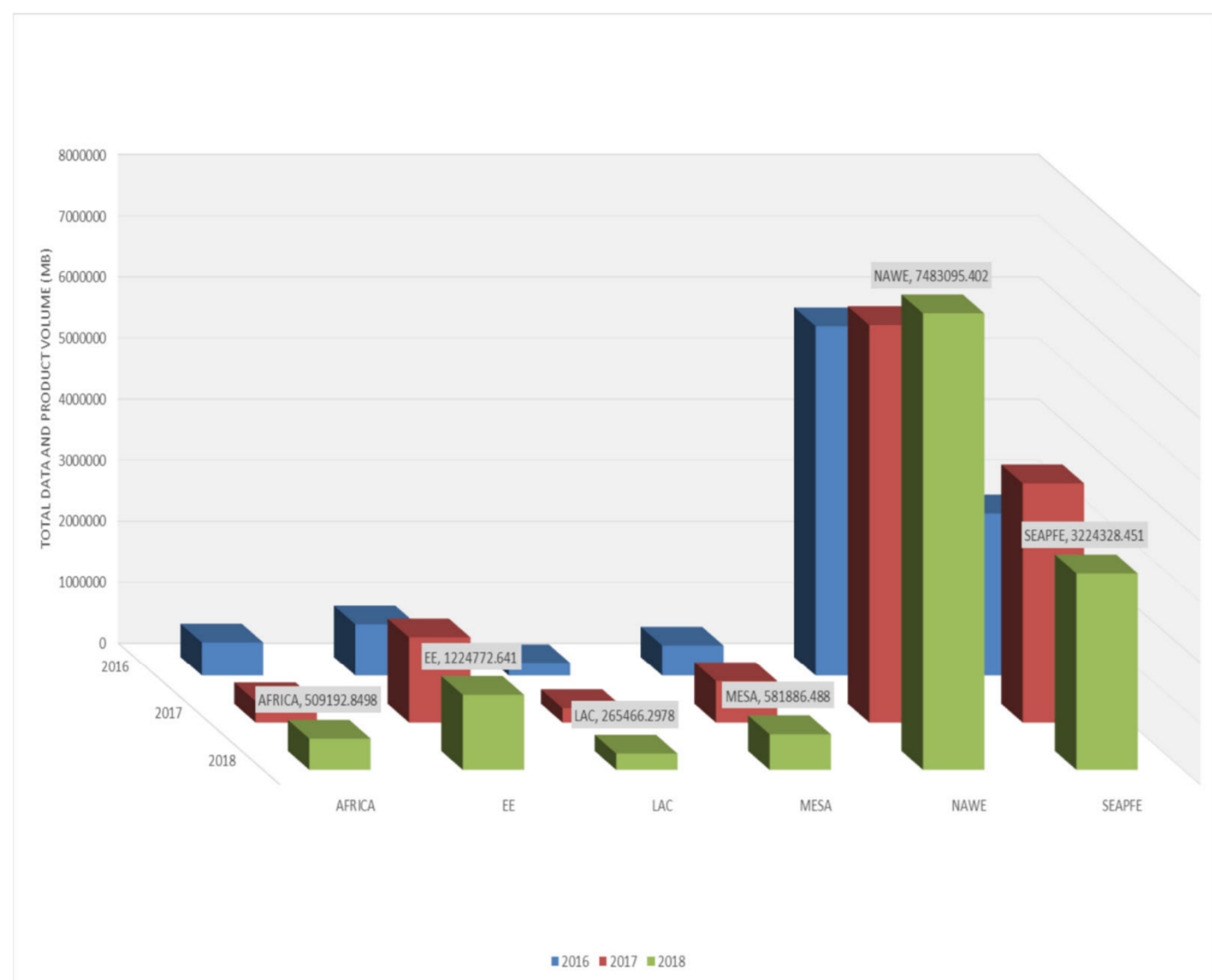




Introduction

- The Comprehensive Nuclear Test-Ban Treaty is a near-universal treaty with 184 Signatory States. Each of these States has the Treaty right to the International Monitoring System data and the International Data Centre products. The ability to request and then analyze the data is not so universal.



- There are many reasons for the unevenness in data usage.
 - Infrastructure
 - Band-Width
 - Expertise
- Capacity Building is addressing many of these issues and great strides have been made but there is the possibility to now utilize the cloud to equalize the infrastructure challenges.

Objective

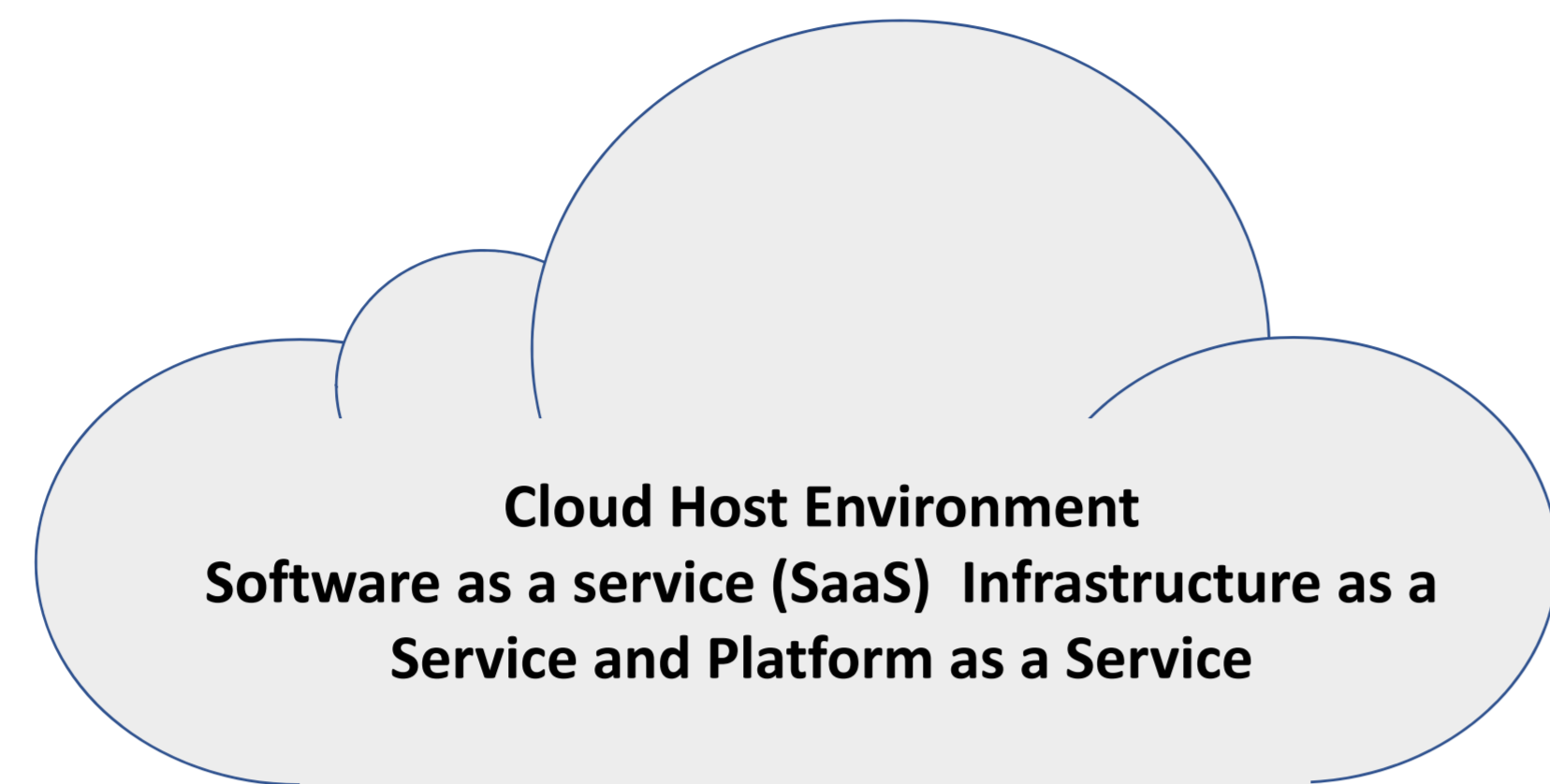
- To demonstrate how the use of cloud computing can work around infrastructure inequities to enable all National Data Centers to be fully engaged with the Treaty.
- NDC infrastructure can be augmented by performing the data intensive calculations in the cloud.
- My colleagues at LANL for another study utilized the Amazon Web Services (AWS) cloud to download a year's worth of US transportable array seismic data (1690 arrays) from IRIS and processed the data in the cloud. The results of the data analysis and the sorted data were downloaded to a local PC.

LA-UR-19-24413

The views expressed here do not necessarily reflect the views of the United States Government, the United States Department of Energy, the National Nuclear Security Administration, or the Los Alamos National Laboratory.

Disclaimer: The views expressed on this poster are those of the author and do not necessarily reflect the view of the CTBTO

Cloud Services



In the Software as a Service model, **software is licensed on a subscription basis and is centrally hosted by a Cloud Service Provider (CSP)**. "out of the box, really is ready to go on day one and allows you use innovative products to enhance your mission very, very quickly."

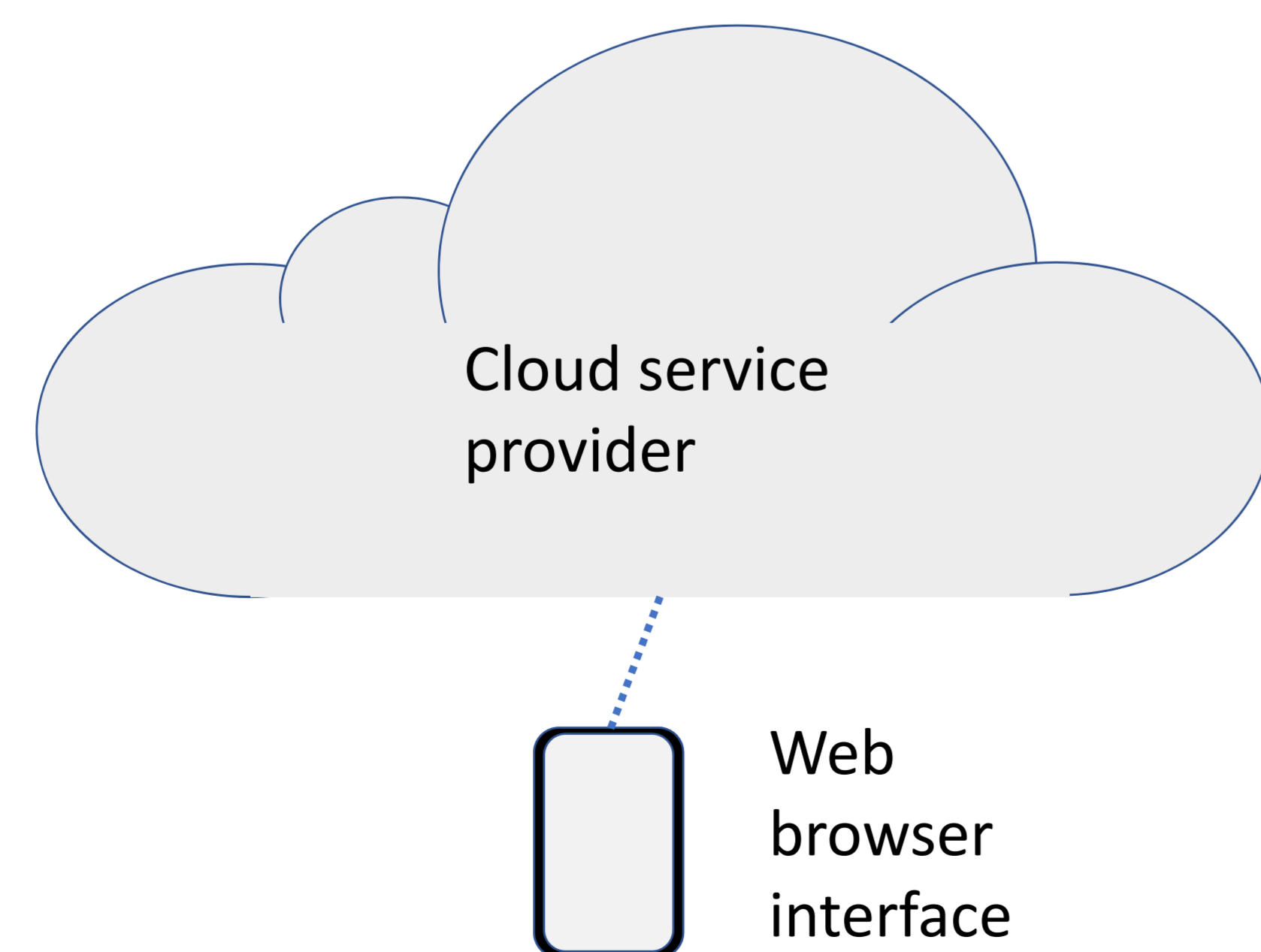
Infrastructure as a Service allows agencies to **shift major IT functions, such as computing, storage, networking and databases, to the cloud**. Agencies can effectively "rent" all of the equipment needed for such functions in a CSP's data center and pay only for the capacity and space they use. "build their own custom environments,"

Platform as a Service enables agencies to **develop, run and manage apps without having to worry about the infrastructure** they would usually need to develop and launch an app. gives agencies the agility to use higher-end services like AI and machine learning, and it **boosts the speed of application development and lowers the cost of development**.

<https://fedtechmagazine.com/article/2019/02/iaas-vs-paas-vs-saas-what-cloud-strategy-right-your-agency-perfcon>

"The traditional cloud deployment models reflected a progression of increasing vendor-ownership through system layers, from Infrastructure as a Service (IaaS) where vendors provide only the infrastructure and hardware, to Platform as a Service (PaaS) where vendors provide hosting and infrastructure management, to Software as a Service (SaaS) where "users" only need to provide their data and most other capabilities and functionality are provided by a vendor," <https://cloud.cio.gov/strategy/>

Ultimate End Point

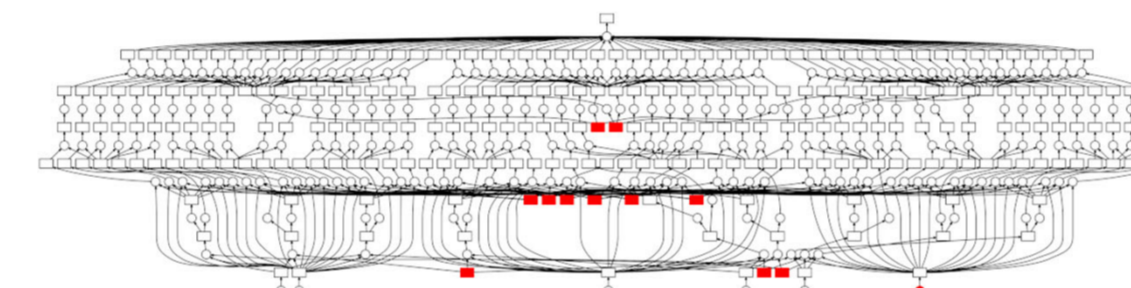


What we did



A DIY Custom Research Cluster in AWS

- Describe your complete computation as a directed acyclic graph (DAG).
- Use familiar Python code.
- Submit it to the Dask scheduler in AWS
- Distributes the download tasks across the computation, good for IRIS DMC

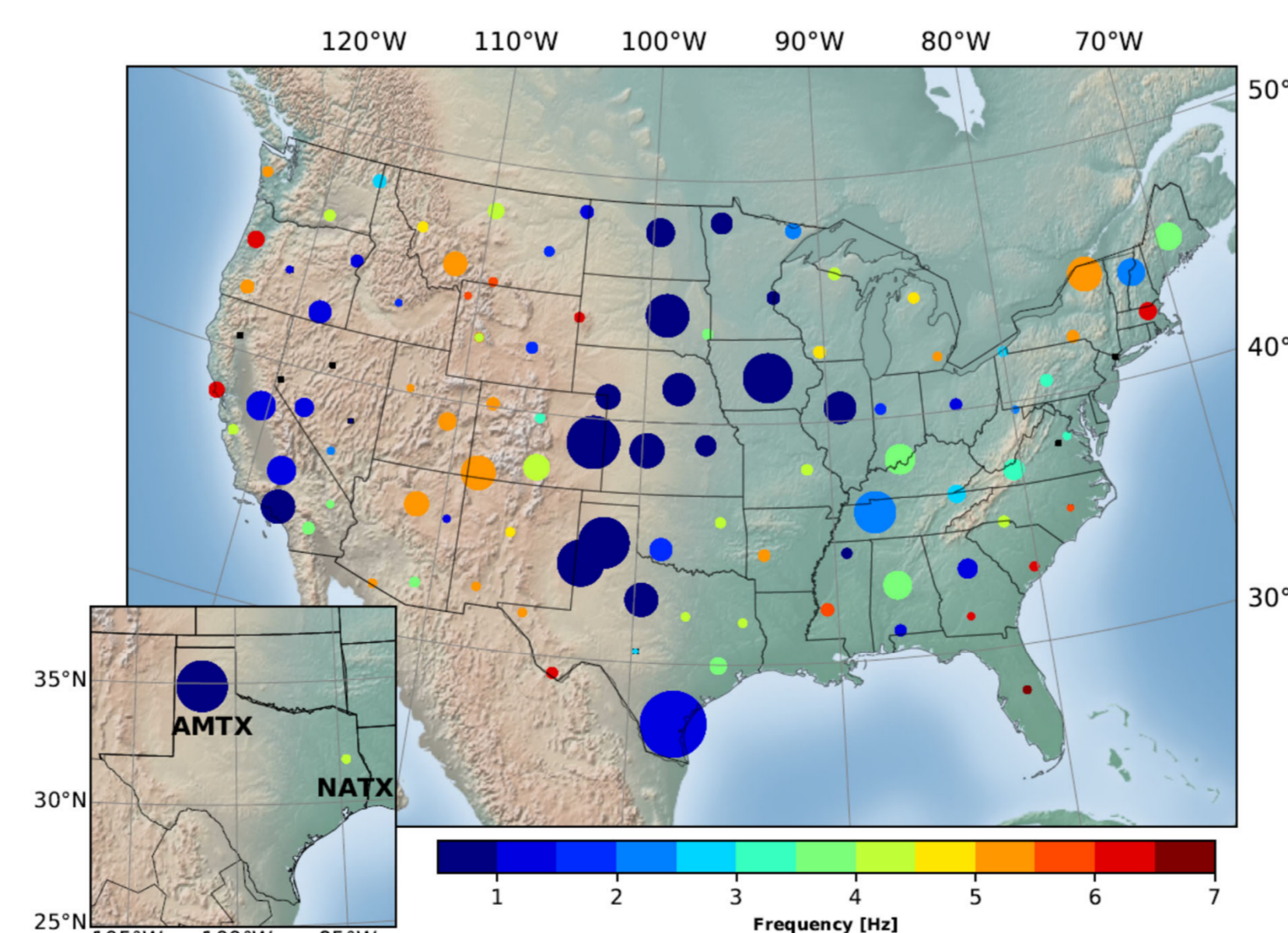
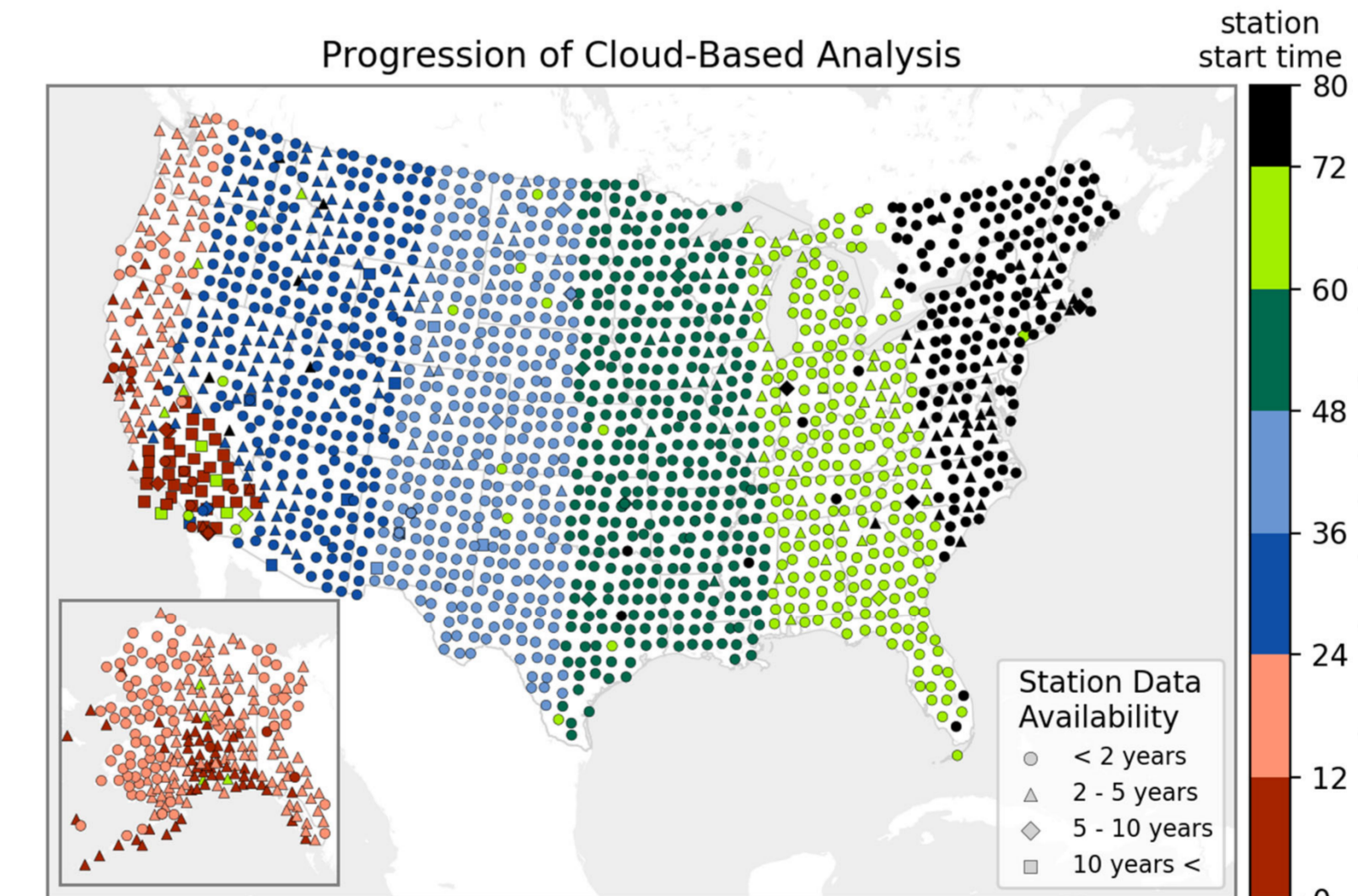
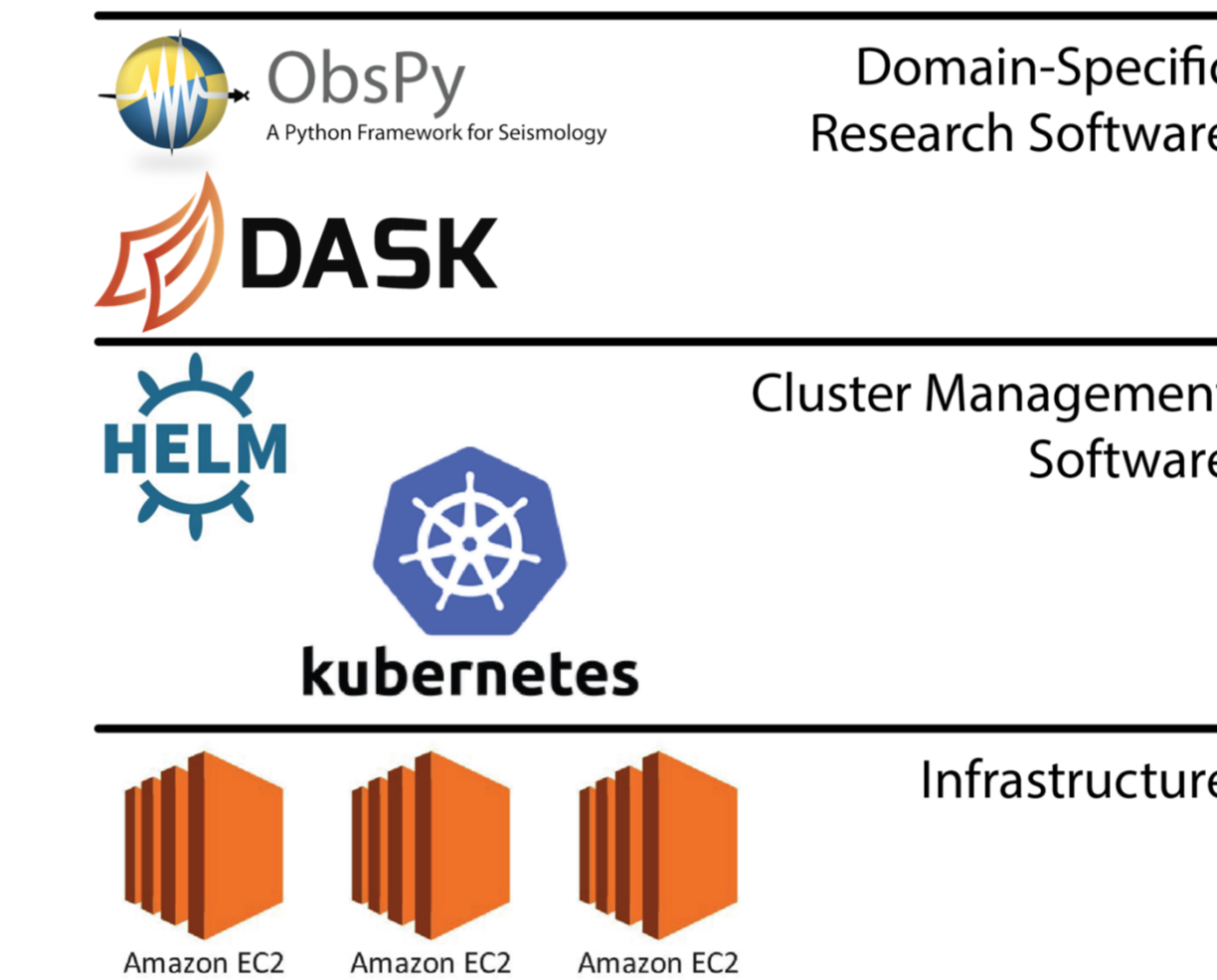
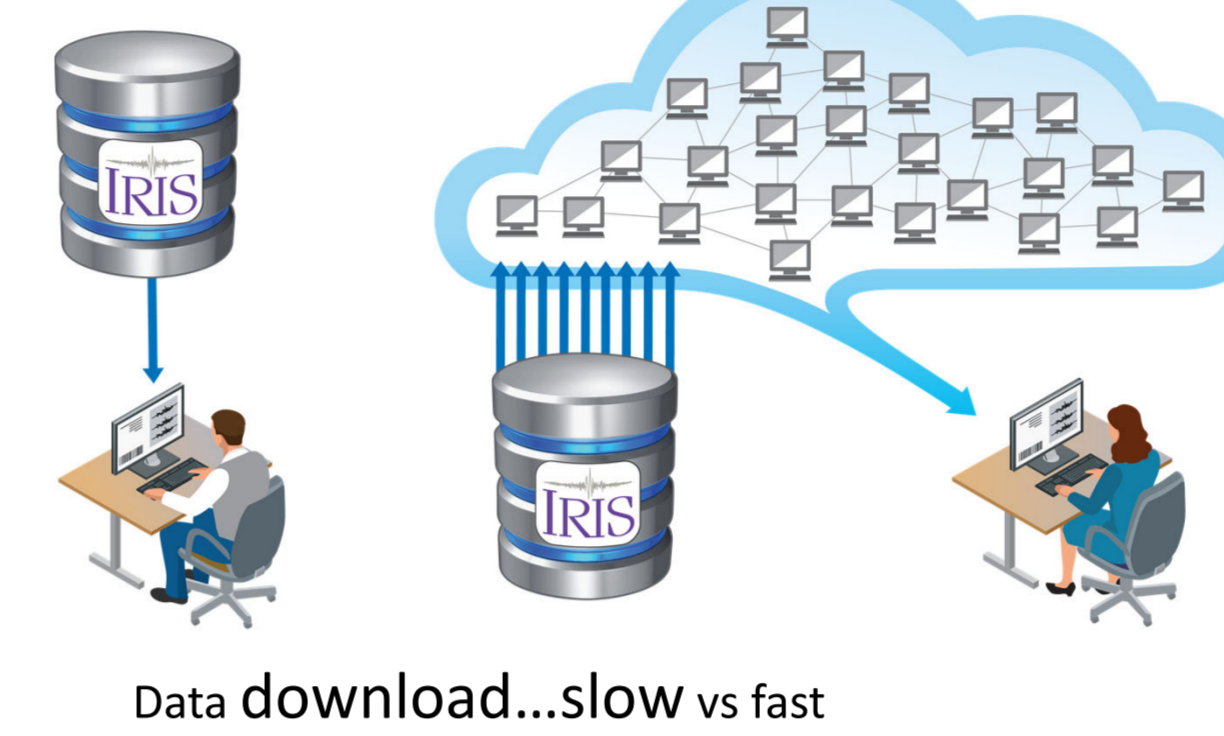


A Example of a typical DAG Grid Search

- 2099 stations, BHZ, all available data
- 80 hours to complete
- \$120 / day
- 50 nodes on EC2: -8 GB RAM

As impressive as the time was to do the analysis if the data was resident in the cloud. The analysis would have been a magnitude faster

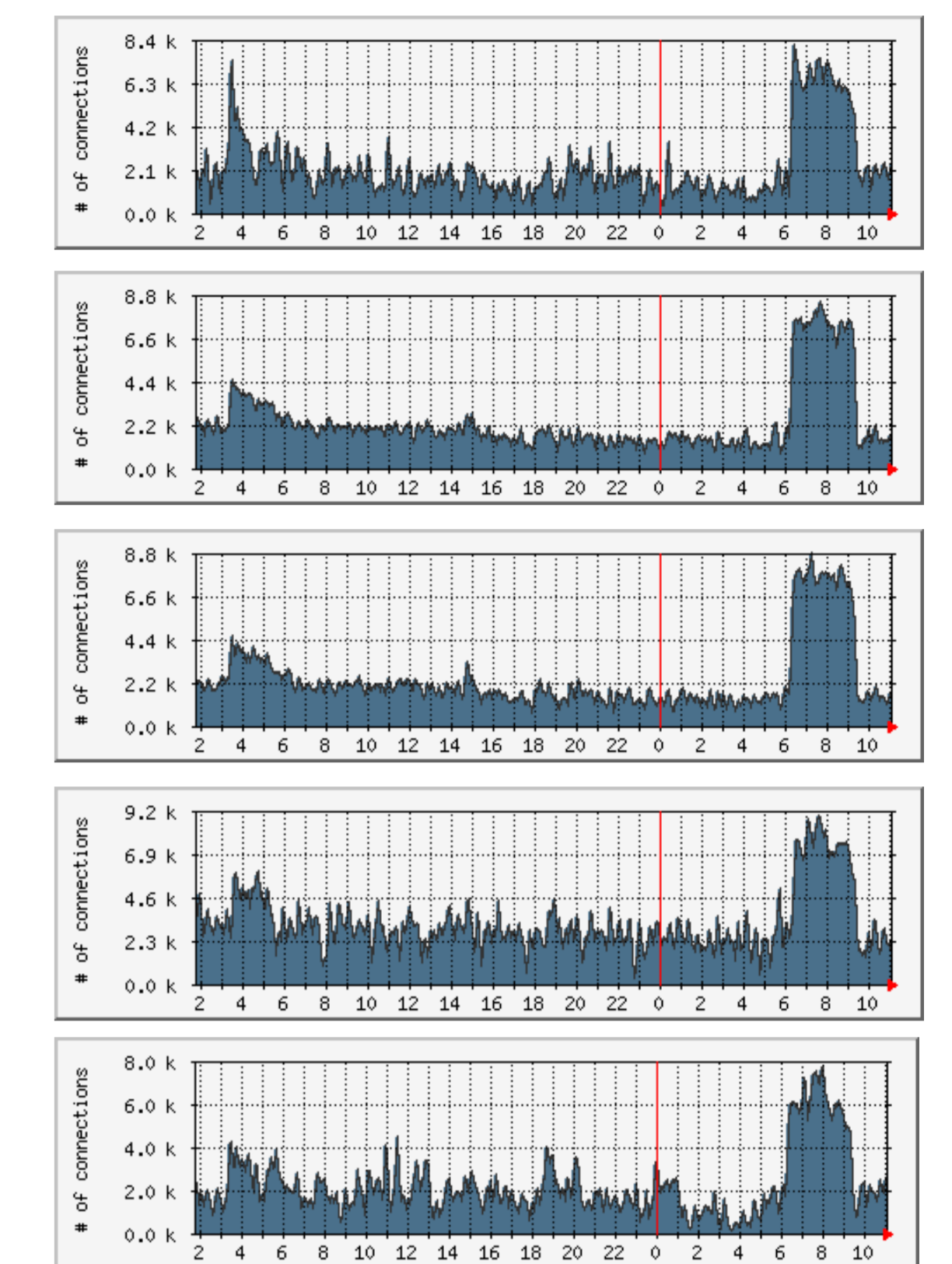
A US Transportable Array 2099 3C stations >12 TB data



- A streaming harmonic tonal noise (HTN) survey was run on 100 stations of the USNSN over 10 years. Circles are located at seismic stations, colored according to the frequency content of the HTN detections, and scaled in proportion to the number of detections.

Impact on Data Centers

- Our 50-node cluster tripled the request volume at IRIS Data Management Center (DMC).
- The streaming model represents a significant change in data center usage
- Scaling doesn't scale:
 - Unused service capacity is a limited resource.
 - Coordination with the data center is required.
- This was one researchers impact on the IRIS DMC



How could this work with an NDC

Data Storage

- Data from the IDC is based upon a subscription.
- The data is then pulled from the IDC into a database.
 - The software to pull the data and the database could be resident in the cloud.
 - The data then stored in the cloud for use.
 - Avoids the DMC impact.

Analysis Software

- Analysis software "NDC in a Box" could be placed and utilized in the cloud environment.

Ukrainian NDC presented a PAAS system at the NDC workshop in Algeria and at WGB52. ECS-PRES-WGB-52-18

OR

- Utilize cloud based analysis software

Similar to Ukrainian NDC proposal

Python scripts for example

Analysis and Results

- The data pull for the analysis would be from the cloud and the analysis done in the cloud so there would be no NDC infrastructure or bandwidth issues.
- The number of nodes required for the analysis could right-sized depending upon the data set and the analysis software requirements.
- Analysis inputs and results are communicated through a web browser interface.

Using the cloud can address infrastructure and band-width issues and reduce the required personnel footprint

This work was completed for the U.S. Department of Energy, National Nuclear Security Administration, Office of Nonproliferation and Arms Control by Los Alamos National Laboratory under award number 89233218CNA000001.

