Cloud computing is no longer an emerging technology. It has become a key player in many of our architectural considerations. This panel discussion consists of experts from science data systems, data centers, and enterprise systems to discuss their current projects, results from their studies, architectural approaches, and deployment strategies for Cloud computing.

- Jeff de La Beaujardière, PhD - NOAA Big Data Project
- Mike Little - ESTO
- Hook Hua - JPL/SWOT Science Data System deployment strategies
- Myche McAuley - JPL/PO.DAAC/SWOT Data Archive
- Dan Pilone - Moving to Cloud-Native Architecture for EOSDIS Applications

Notes:
1. The NOAA Big Data Project
NOAA has “Big Data”: 10 satellites, 150+ radars, 3 buoy networks, 200+ tide gauges, human observers, animal telemetry, 17 ships, 10 aircraft, numerical models, extramurally funded data

Traditional Data Services Approach (pre-cloud) is not efficient for large datasets;

- Conceptual Overview of NOAA Big Data project: agency-provided services, Cloud IaaS providers, application & product providers, and new customers & lines of business
- Keep previous efforts; copy some of the data to the cloud;
- NOAA big data project - 5 BDP CRADAs announced April 2015 (Amazon, Google, IBM, Microsoft, OCC)


2. AIST management cloud system: use computing resource in a more efficient way

- Automatically upscale and downscale
- Pay what you use
- Security has been taken care
- Help manage different projects
3. SDS Development Strategies for NISAR and SWOT

Conceptual data product flow:
SDS->(3TB/day)SDS->(150TB)DAAC

Data storage, processing, movement, costs are the biggest challenges; (NISAR SDS 90TB)

NISAR SDS Data Volumes: impossible to transfer data by network
High-level SDS View:

- SDS in Private Cloud
  - Forward processing at data center
  - Buick processing at public cloud
- Share data to avoid data copying by co-located
- If not co-located, cost of moving data is about $5.17 egress out to DAAC

Data Lake: Minimize data movement; maximize user services; run on public cloud provider;
Hot data lake: use object store data lake for hot data

4. PO.DAAC could scenarios for SWOT

S3 for hot data, glacier, -> reduce storage cost -> regenerate SLCs on-demand

Cost considerations: Egress costs - bring algorithms to the data

PO.DAAC solution may change over time:

- The egress costs are effectively unbounded
- Archive storage doesn’t scale like compute
- Need a malleable solution, which can migrate over time
- Looking to “buy time”

5. Assessing Cloud-Native Architectures for EOSDIS

How to deal these big data about 15 PB?

EOSDIS Vision 2020:

- Data analysis at scale
- Data and processing mobility
- Dataset upgrading
- Virtual collections
- Combining data, combining tools

Evolution review: developed the prototypes to explore the advantages, risks, and costs of using commercial cloud environments for storage and data transfer, processing, and improved data access.

Application architecture: bridge scalability and usability

- Looking forward:
  - Broader knowledge of cloud architecture and patterns(NGAP, serverless, 12 factor application)
  - Rapid pace of change
  - Potentially dramatic change in data usage patterns
  - What about cloud native data formats?
  - What about Cloud Native services?
Questions:
How to measure cost? Cloud economy -> estimate/analyze the cost for different strategies; the best one is fully on amazon; different discounts; avoid egress, and invite users to cloud (TCO);
How to configure the service on the cloud? You are free to configure anything on vms Compute the cost for different options; For example, egress would cost a lot of money for the hot datasets in the AWS.
AIST management cloud system is flexible and work for some, but have limitation on some other.

Attachments/Presentations:
01 20160721.NOAA_BDP.ESIP_JeffDLB.pptx [10]
05 ESIP Summer 2016_Big Data and Cloud Panel_Pilone.pptx [12]

Session Leads:
Name: Thomas Huang [13]
Organization(s): Jet Propulsion Lab [14]
Name: Phil Yang [15]
Organization(s): SIESIP [16]
Email: cyang3@gmu.edu [17]

Creative Common License: Creative Commons Attribution 3.0 License
Accepted:
Keywords: Cloud Computing [18]
Big Data [19]
Geosciences [20]
Architecture [21]

Source URL: http://commons.esipfed.org/node/9104

Links:
[17] mailto:cyang3@gmu.edu