Facilitating NASA Earth Science Data Processing Using Nebula Cloud Computing

Long Pham1, Aijun Chen1,2, Steven Kempler1, Christopher Lynnes1, Michael Theobald1, Esfandiar Asghar1, Jane Campino4, Bruce Vollmer1

1NASA Goddard Earth Sciences Data & Information Services Center (GES DISC); 2Center for Spatial Information Science and Systems, George Mason University
3ADNET System Inc.; 4ADNET System Inc. and , NASA Goddard Space Flight Center SESDA II Program Office

Embry-Riddle Aeronautical University, 1413 Arkansas Road, Room 106, Andrews, MD 20762

IN4A-1398

Cloud Computing has been implemented in several commercial arenas. The NASA Nebula Cloud Computing platform is an infrastructure as a Service (IaaS) built in 2008 at NASA Ames Research Center and 2010 at GSFC. Nebula is an open source Cloud platform intended to:

- Make NASA realize significant cost savings through efficient resource utilization, reduced energy consumption, and reduced labor costs.
- Provide an easier way for NASA scientists and researchers to efficiently explore and share large and complex data sets.
- Allow customers to provision, manage, and decommission computing capabilities on an as-needed basis.

Nebula Cloud Computing Platform

Cloud Computing Projects at NASA GES DISC

NASA GES DISC has been evaluating the feasibility and suitability of migrating GES DISC’s applications to Nebula platform by porting the following projects:

- a) Using Nebula Cloud to run scientific data processing infrastructure
- b) Using Nebula Cloud to run scientific data processing workflow
- c) Porting a Web-based scientific data processing application to Nebula Cloud

GISDISC: A Web-based application which offers online visualization and analysis of vast amounts of Earth science data.

Giovanni MAPPS: Multi-sensor Aerosol Products Sampling System (portal) focuses on visualizing aerosol relationships among ground-based data and satellite data.

Running S4PM requires installation of auxiliary packages. The AIRS L1/L2 algorithm workflow runs based on S4PM infrastructure and involves quite a few libraries, e.g. HDF, sdpkt, and basic data, e.g. DEM, MODIS, AVHRR. Migrating it can be time-consuming. The diagram at right shows the procedures for pre-installation and testing of S4PM and AIRS algorithms first on the local box, then the Nebula box.

Performance Comparison between Nebula & Local

Advantages of NASA Nebula Cloud Platform:

- User friendly interface, access to and management of Nebula resources; dashboard & EucaTools.
- Better performance compared to local box.
- Lower cost (only pay for used time and resources)
- Scalability, on-demand provisioning of resources in near real-time, and no user involvement for peak loads.
- Cloning, simple bundling processes to save a modified/improved image.
- An excellent feature to maintain, back up, and mirror the systems; hence, increased reliability.
- Knowledge base, including detailed instructions, tutorial, and FAQ.

Lessons Learned:

- Bundle early, bundle and backup often!
- Take detailed notes:
  - Record each step taken to launch and install needed and required software packages.
  - Acquire SA assistance
  - Use same directory structure
  - Use EucaTools
  - Expect the process to be time-consuming

Challenges Faced:

- Stability – e.g. ports are not stable, network (FTP/Wget) is slow and not stable
- Underdeveloped (e.g. Object Store) managing and monitoring tools
- Bare-bones images, wrong location of attached volumes, some defects in the bonded images.
- Gaps in Knowledge Base.
- Size Limitation, e.g. limited size of volume, at most 16 cores
- Commercial Software installation and licenses.

Future: Making operational system at Nebula

- a) Migrate more of GES DISC’s applications/portals, e.g. Giovanni portals, to the Nebula Cloud platform.
- b) Making mature migrated applications operational on the Nebula Cloud platform.
- c) Testing some commercial Cloud applications designed for government, e.g. Amazon GovCloud.

Acknowledgements: Authors affiliated with Center for Spatial Information Science and Systems (CSISS), Georgia Institute of Technology have a cooperative agreement with GES DISC (Agreement No.: NNX08AB39A, Center Director: Dr. Ning Li).