Facilitating NASA Earth Science Data Processing Using Nebula Cloud Computing

Long Pham1, Aijun Chen1,2, Steven Kemppler3, Christopher Lynnes1, Michael Theobald1, Esfandiari Asghar3, Jane Campino4, Bruce Vollmer1

1NASAGoddardEarthSciencesData&InformationServicesCenter(GESDISC);2CenterforSpatialInformationScienceandSystems,GeorgeMasonUniversity
3ADNETSystemInc.;4ADNETSystemInc.andNASAGoddardSpaceFlightCenterSESDAIIProgramOffice

Embry-RiddleAeronauticalUniversity,1413ArkansasRoad,Room106,Andrews,MD20762

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:

a) Make NASA realize significant cost savings through efficient resource utilization, reduced energy consumption, and reduced labor costs.

b) Provide an easier way for NASA scientists and researchers to efficiently explore and share large and complex data sets.

c) Allow customers to provision, manage, and decommission computing capabilities on an as-needed basis. NASA Nebula: http://nebula.nasa.gov

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 S4PM is an open source data processing infrastructure. Based on S4PM, scientific data processing algorithms can be run to efficiently process large volumes of satellite data.

b) Using Nebula Cloud to run scientific data processing workflow The Atmospheric Infrared Sounder (AIRS) focuses on supporting climate research and improving weather forecasting. Based on S4PM, the AIRS L1 & L2 algorithms workflow can run on the local box, then the Nebula box.

c) Porting a Web-based scientific data processing application to Nebula Cloud Giovanni is a Web-based application which offers online visualization and analysis of vast amounts of Earth science data. The 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.

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) Migrate more of GES DISC’s applications/portals, e.g. Giovanni portals, to the Nebula Cloud platform;

b) Make mature migrated applications operational on the Nebula Cloud platform.

c) Testing some commercial Cloud applications designed for government, e.g. FBI.

Advantages of NASA Nebula Cloud Platform:

Advantages of NASA Nebula Cloud Platform:

- User friendly interface, access to and management of Nebula resources: dashboard & EucaTools.
- Data processing comparisons to local box 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 process 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 upgrade often!
- Take detailed notes:
  - Record each step taken to launch and install missing 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. portals are not stable, network (FTP) is slow and not stable
- Underdeveloped (e.g. Object Store) managing and monitoring tools.
- Bare-bones images, wrong location of attached packages.
- Gaps in Knowledge Base.
- Size Limitation, e.g. limited size of volume, at most 16 GB.
- 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) Migrating mature applications operational on the Nebula Cloud platform.

c) Testing some commercial Cloud applications designed for government, e.g. Amazon GovCloud.