Cloud Computing has been implemented in several commercial areas. 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 bases. NASA Nebula: http://nebula.nasa.gov

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
4SPM is an open source data processing infrastructure. Based on 4SPM, 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 4SPM, the AIRS L1 & L2 algorithms workflow can be 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 MAPSS (Multi-sensor Aerosol Products Sampling System) portal focuses on visualizing aerosol relationships among ground-based data and satellite data.

Migrating AIRS L1/L2 Algorithms Workflow
Running 4SPM requires installation of auxiliary packages. The AIRS L1/L2 algorithm workflow runs based on 4SPM infrastructure and involves quite a few libraries, e.g., HDF, sqlite, 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 4SPM and AIRS algorithms first on the local box, then the Nebula box.

Performance Comparison between Nebula & Local

Cost Comparison between Nebula & Local
Using Nebula Cloud to run scientific data processing is faster, with much lower cost, compared with the current data processing system.

Advantages of NASA Nebula Cloud Platform:
- User friendly interface, access to and management of Nebula resources; dashboard & EucaTools.
- Better performance compared to local box
- Scalability, on-demand provisioning of resources in near real-time, and no user involvement for peak loads
- Lower cost (only pay for used time and resources)
- Cloning, simple bundling process to save a modified/ improved image.
- An excellent feature to maintain, back up, and monitor 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 missing software packages.
- Use same directory structure
- Use Euro2ools
- Expect the process to be time-consuming

Advantages:
- Speed - e.g. portals are not stable, network (FTP/WeGet) 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 bundled images.
- 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) Making mature migrated applications operational on the Nebula Cloud platform.
c) Testing some commercial Cloud applications designed for government, e.g. Amazon GovCloud.

Hardware Performance Analysis

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Local Linux box</th>
<th>Nebula virtual Linux Box</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware</td>
<td>DELL PowerEdge E8600 with Dual-Core Xeon Processor 7160 series / 4 CPU</td>
<td>DELL PowerEdge c2110 with Quad-Core Xeon Processor 5500 series / 2 CPU (2x1000 offers 4 CPU virtually)</td>
</tr>
<tr>
<td>CPU (GHz)</td>
<td>4 * 3.16 (8 cores available)</td>
<td>4 * 2.8 (16 cores available)</td>
</tr>
<tr>
<td>RAM (GB)</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>Storage</td>
<td>11TB</td>
<td>300GB (200GB in default)</td>
</tr>
<tr>
<td>CPU Microarchitecture</td>
<td>65nm NetBurst</td>
<td>45nm Nehalem/Westmere</td>
</tr>
</tbody>
</table>

Netburst Microarchitecture (Local Box) Nehalem/Westmere Microarchitecture (Nebula)
Cache L3                    | N/A             | 2MB/core                 |
PFS                         | Dual Independent 800MHz | QPI=4x60T (QuickPath Interconnects) |
Memory                     | DDR2 266 ECC DRAM (double channel) | DDR3 3 (triple channel) |

Netburst (65nm) -> Core (65nm) | Penryn (45nm) -> Nehalem (45nm) | Westmere (32nm) |
Core = 2.5 Netburst          | Penryn = 1.8 Core | Nehalem/Westmere         |
Nebula/Westmere=[5.4-5.0] x Netburst

<table>
<thead>
<tr>
<th>Two days (2012.133-134)</th>
<th>Local Linux box</th>
<th>Nebula Test 1</th>
<th>Nebula Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Volume (GB)</td>
<td>29.11</td>
<td>29.11</td>
<td>29.11</td>
</tr>
<tr>
<td>Output Volume (GB)</td>
<td>12.14</td>
<td>12.14</td>
<td>12.14</td>
</tr>
<tr>
<td>Elapsed Time (h)</td>
<td>63.05</td>
<td>63.05</td>
<td>63.05</td>
</tr>
<tr>
<td>CPU Time (h)</td>
<td>10.11</td>
<td>10.11</td>
<td>10.11</td>
</tr>
<tr>
<td>System Time (m)</td>
<td>29.02</td>
<td>34.43</td>
<td>29.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Two days (2012.133-134)</th>
<th>Input Data (GB)</th>
<th>Output Volume L2 data (GB)</th>
<th>Elapsed Time (h)</th>
<th>CPU Time (h)</th>
<th>System Time (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIRS L1 &amp; L2 Processing</td>
<td>120.98</td>
<td>12.14</td>
<td>42.80</td>
<td>35.67</td>
<td>35.80</td>
</tr>
</tbody>
</table>

Two days (2012.133-134) | Input Data (GB) | Output Volume L2 data (GB) | Elapsed Time (h) | CPU Time (h) | System Time (m) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The Atmospheric Infrared Sounder (AIRS)</td>
<td>2010.123-124</td>
<td>120.98</td>
<td>12.14</td>
<td>42.80</td>
<td>35.67</td>
</tr>
</tbody>
</table>

Using Nebula Cloud to run scientific data processing is faster, with much lower cost, compared with the current data processing system.

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.: NNX08AD39A), Center Director: Dr. Lijing Diao.

IN41A-1398