Airborne Cloud Computing Environment (ACCE)

ESIP Federation Summer Meeting
July 14, 2011

Sean Hardman, Dana Freeborn, Dan Crichton, Emily Law, Liz Kay-Im
Jet Propulsion Laboratory, California Institute of Technology
What is ACCE?

- ACCE is JPL’s internal investment to improve the return on airborne missions.
  - Improve development performance of the data system
  - Improve return on the captured science data

- The investment is to develop a common science data system capability for airborne instruments that encompasses the end-to-end lifecycle covering planning, provisioning of data system capabilities, and support for scientific analysis in order to improve the quality, cost effectiveness, and capabilities to enable new scientific discovery and research in earth observation.
Why ACCE?

- Currently, no coordinated data system effort to collect, process, distribute and archive data from airborne missions
- Data systems are not architected for use across several instruments; limited data sharing
- Limited infrastructure for developing, testing and integrating algorithms
- Each PI team develops their own “one off” leading to cost and performance challenges
Approach

• Develop an Airborne Science Data System
  – Capture, process, archive and distribute airborne instrument data
  – Leverage existing product line software at JPL being used for OCO, SMAP, NPP Sounder PEATE, EDRN, VODC, …
    • Utilizing the full suite of CAS-related Apache OODT components
  – Provide options to either host PI data system instances or allow them to deploy their own instance

• Deliver shared services as part of a “cloud” strategy enabling:
  – PI integration and access to airborne data system services from their own facilities and labs (processing, data management, distribution, etc.)
  – Use of both internal and external cloud services
A full service stack is deployed for each mission, utilizing any mission-provided resources as well as a cloud computing infrastructure.

Mission proprietary data is presented in a mission-specific secure portal while publicly available data is aggregated in the public portal.
Cloud Trade Study

- Initiated a trade study of cloud solutions to explore the benefits of performing science data processing for airborne missions in the cloud.

*A typical science data processing job (PGE) is an executable that performs a task on one or more inputs to produce one or more outputs.*
Cloud Challenges

• Host Environment
  – Each PGE is compiled for a specific OS platform, but not all PGEs are compiled on the same platform
  – Dynamically linked third party libraries must be available
    • May require locally available licenses
  – Sufficient CPU and memory available on-demand
  – Sufficient storage available on-demand

• Security
  – ITAR-compliance is required for handling ITAR-sensitive data

• Network
  – File transfer rates for large input and output files
  – Ability to communicate through the JPL Firewall
Technologies Explored

- Apache’s Object Oriented Data Technology’s (OODT) Catalog and Archive Service (CAS) components
  - Provides basic services for File, Workflow and Resource management.

- Amazon EC2/S3 (public/commercial cloud)
  - Elastic compute resources and on-demand storage

- Eucalyptus (private cloud infrastructure)
  - Infrastructure software for establishing a private cloud

- Hadoop – Distributed File System (DFS) and Map/Reduce
  - Increased processing performance on large data sets
## Trade Study Analysis

<table>
<thead>
<tr>
<th>Trade Dimension</th>
<th>Mission Maintained</th>
<th>Private Cloud</th>
<th>Commercial Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processor Speed</td>
<td>High$/Med$/Low</td>
<td>Low</td>
<td>High*/Med*/Low</td>
</tr>
<tr>
<td>Interconnect Speed</td>
<td>High$/Med$/Low</td>
<td>Low</td>
<td>High*/Low</td>
</tr>
<tr>
<td>Transfer Speed</td>
<td>Med-High</td>
<td>Med-High</td>
<td>Low-Med</td>
</tr>
<tr>
<td>ITAR Concerns</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Offsite Data Backup</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Growth/Bursting Potential</td>
<td>Step-wise capitol investment</td>
<td>Managed growth, minimal bursting</td>
<td>Linear scaling, high burst capability</td>
</tr>
<tr>
<td>Upfront Costs (Processing)</td>
<td>Capitol investment</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ongoing Costs (Processing)</td>
<td>SA Support, Maintenance, HW Replenishment</td>
<td>Pay-by-drink</td>
<td>Pay-by-drink</td>
</tr>
<tr>
<td>Upfront Costs (Disk)</td>
<td>Capitol investment</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ongoing Costs (Disk)</td>
<td>SA Support, Maintenance, HW Replenishment</td>
<td>Pay-by-drink</td>
<td>Pay-by-drink</td>
</tr>
</tbody>
</table>

$ ^\text{Increased Upfront Cost} \quad * ^\text{Increased Ongoing Costs}$
Trade Study Summary

- Cloud storage benefits vary depending on the provider

- Cloud computing (public or private) decreases processing cost
  - *No investment in capital required (upfront or refresh costs)*
  - *Pay only for what you use*

- Possible limiting factors for public cloud viability
  - *Support for ITAR-sensitive data*
  - *Data transfer rates between JPL and the cloud*
  - *Access through JPL firewall*
The Carbon in Arctic Reservoirs Vulnerability Experiment (CARVE) is the first operational customer/partner for ACCE.

Their Fourier Transform Spectrometer (FTS) instrument will produce over 700K soundings over the life of the mission (3 years).

Their hardware cost was driven by the requirement for Level 2 Full Physics processing.

Their analysis concluded the following processing costs:
- 481K - *Purchase and operation of a cluster-based system*
- 196K - *Utilize Amazon EC2*

Therefore the FTS Level 2 Full Physics processing will be performed on the cloud.
- *All other processing performed locally on commodity hardware with the ACCE software stack deployed in the CARVE environment.*
Current Efforts with Amazon

- JPL’s preferred public cloud provider is Amazon.
  - Don’t have too many details on the CIO office’s decision but contracts are in place to facilitate charging JPL project accounts.
- Amazon is making an effort to address the “limiting factors”.
  - An “ITAR Region” is planned for Fall 2011.
  - Continued benchmarking and optimization has demonstrated increased data transfer rates, 250 - 400 mbps.
  - Their Virtual Private Cloud (VPC) solution alleviates JPL firewall issues.
- Working with the developers of Polyphony to provide the ACCE interface to Amazon.
  - Polyphony is a workflow orchestration framework for cloud computing.
  - The framework is not limited to Amazon and will allow ACCE to interface with other cloud solutions.
  - More information on Polyphony:
    - [http://usrp.usra.edu/technicalPapers/jpl/Rossi.pdf](http://usrp.usra.edu/technicalPapers/jpl/Rossi.pdf)
• Although Amazon is JPL and CARVE’s choice, ACCE is looking into other cloud and computing solutions.
  – *Pleiades Supercomputer*
  – *NEBULA Cloud Computing Platform*
• Investment to date has allowed ACCE to standardize deployment of the service stack and a generalized portal.
• The last year of ACCE’s internal funding will focus on new portal features and cloud integration.
Questions/Comments

This effort was supported by the Jet Propulsion Laboratory, managed by the California Institute of Technology under a contract with the National Aeronautics and Space Administration.