



**National Aeronautics and
Space Administration**

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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.



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

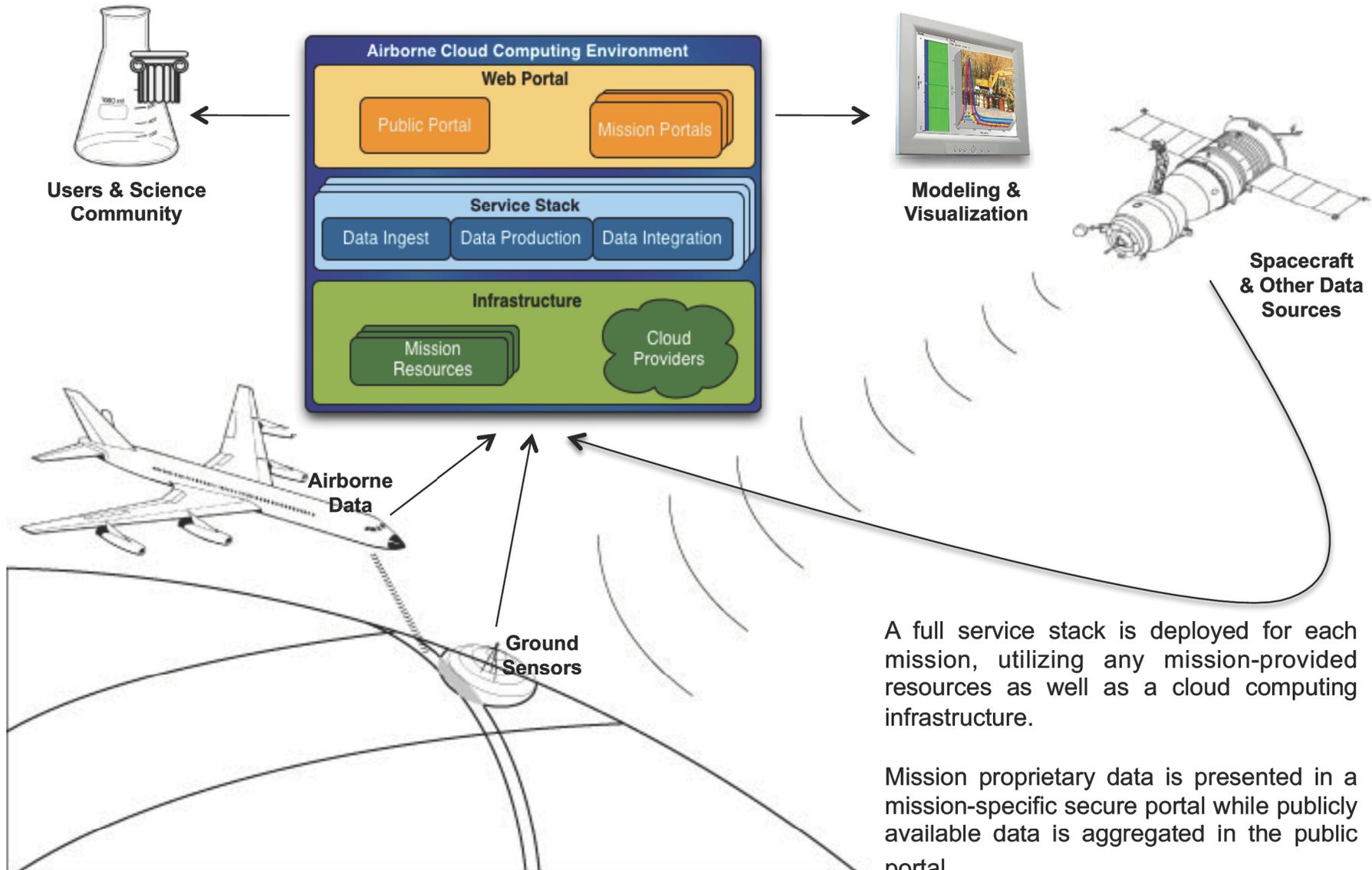
- Develop an Airborne Science Data System
 - *Capture, process, archive and distribute airborne instrument data*
 - *Leverage existing product line software at JPL being used for OCO, SMAP, NPP Sounder PEATE, EDRN, VODC, ...*
 - Utilizing the full suite of CAS-related **Apache OODT** components
 - *Provide 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*



National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Data System Concept



A full service stack is deployed for each mission, utilizing any mission-provided resources as well as a cloud computing infrastructure.

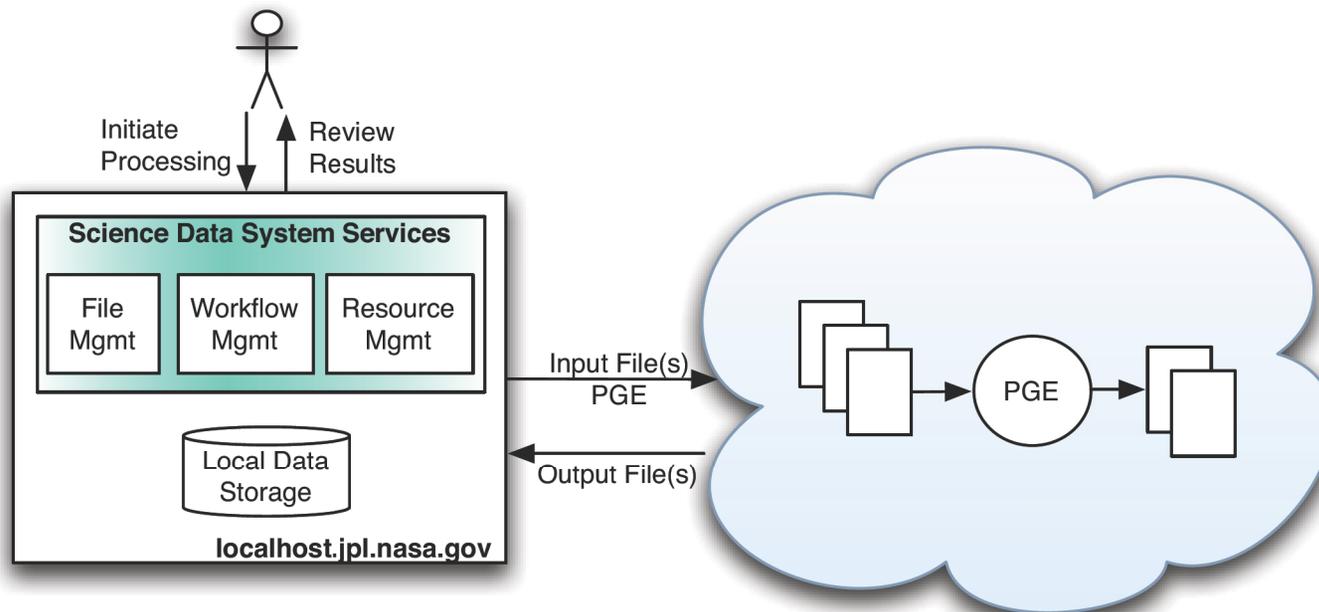
Mission proprietary data is presented in a mission-specific secure portal while publicly available data is aggregated in the public portal.

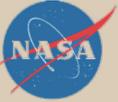


Cloud Trade Study

- 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*



National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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

	Trade Dimension	Mission Maintained	Private Cloud	Commercial Cloud
Performance	Processor Speed	High [§] /Med [§] /Low	Low	High*/Med*/Low
	Interconnect Speed	High [§] /Med [§] /Low	Low	High*/Low
	Transfer Speed	Med-High	Med-High	Low-Med
Key Considerations	ITAR Concerns	No	No	Yes
	Offsite Data Backup	No	No	Yes
	Growth/Bursting Potential	Step-wise capitol investment	Managed growth, minimal bursting	Linear scaling, high burst capability
Costs	Upfront Costs (Processing)	Capitol investment	-	-
	Ongoing Costs (Processing)	SA Support, Maintenance, HW Replenishment	Pay-by-drink	Pay-by-drink
	Upfront Costs (Disk)	Capitol investment	-	-
	Ongoing Costs (Disk)	SA Support, Maintenance, HW Replenishment	Pay-by-drink	Pay-by-drink

§ Increased Upfront Cost * 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*



Partnering with CARVE

- 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.*



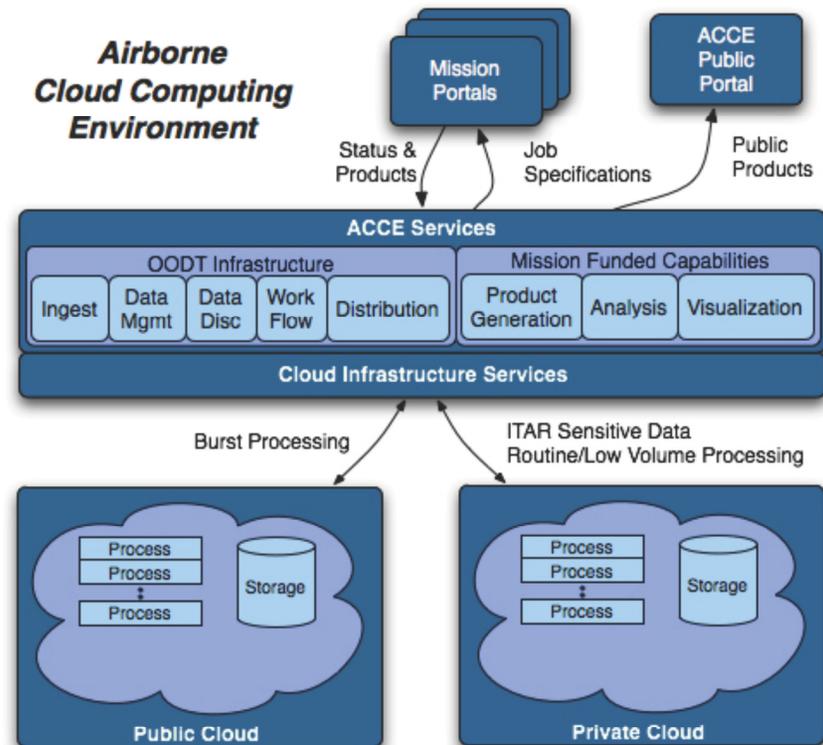
Current Efforts with Amazon

- 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://aws.amazon.com/solutions/case-studies/nasa-jpl/>



Wrap Up

- 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.





**National Aeronautics and
Space Administration**

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

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.