Science in the Cloud: Accelerating Science through Cloud Computing

Science in the Cloud
GSFC
31 January 2018

Daniel Duffy (606) and Ed Masuoka (619)
daniel.q.duffy@nasa.gov
edward.j.masuoka@nasa.gov

Alternative Titles

Cloudy with a chance of science

GSFC Cloud Strategy (sounded a bit pretentious)

What the heck is this cloud thing and how will it affect me?

Why are there so many commercials about cloud computing?

When will I stop hearing about cloud computing?

It's in the cloud – what the heck is my kid saying?

Partnerships

NASA Center for Climate Simulation (NCCS – Code 606.2)

Terrestrial Information Systems Laboratory (Code 619)

Computing and Communications Division (C&C – Code 720)

NASA CIO Computing Services Office (CSSO – NASA HQ)

Lots of people doing great work

- Too many to name, and I am going to forget people.

Planning Science/Proposal Writing

What question am I trying to answer?

- Example: Suppose we want to generate maps of surface water from 1990 to 2012 in the arctic boreal region (problem courtesy of Mark Carroll, Code 618)

What data are available and where are they?

- Landsat time series available at the LP DAAC

How much data is needed?

- Full time series requires >100,000 scenes and ~20TB of data storage

Can I store all that data? If not, how can I process it?

- No. So download chunks of 5TB to local machine. Process. Delete. Download.
- Projected time – 9 months – without any mistakes!

That's too long, so how can I modify my science question accordingly?

- Average across three epochs (1990, 2000, 2010)
- 25,000 scenes and ~7TB of data
- Projected time – 2 to 3 months

Scientists are limiting their questions (and science) based on the IT resources of their desktops!

Conversations "We Don't Want" Between Scientists

Scientist 1

Hey, what are you working on these days?

Scientist 2

Oh, you know, just processing data from the new satellite for my ROSES project.

You know, I need that same data for my project. Where did you get that?

I downloaded it from the web.

How long did it take you?

Quite awhile; several weeks.

Oh, man, I don't want to have to download all That data and take several weeks. Do you think I could get a copy from you?

Sure, I am just not sure how to get it to you. I could NFS serve it from my machine to yours or just give you access to my system.

That would be great. You don't think the security guys would mind do you?

No, I'm sure they wouldn't. It is in the name of science after all.

Conversations “We DO Want” Between Scientists

Scientist 1
Hey, what are you working on these days?

Scientist 2

Oh, you know, just processing data from the new satellite for my ROSES project.

You know, I need that same data for my project. Where did you get that?

It’s in the cloud.

Awesome, I will work on it there.

From HTPC to the Cloud

	High Performance Computing	Advanced Data Analytics Platform (ADAPT)	Goddard Private Cloud	AIST Managed Cloud Environment (AMCE)	Public Clouds
Public/Private	Private	Private/Public	Private	Public	Public
Location	GSFC B28	GSFC B28	GSFC B28/B32	AWS	AWS/Other
Users	NASA	NASA	NASA	NASA/Non-NASA	NASA
Managed	Yes	Yes	Yes	Yes/No	Yes/No
Networks	100 Gbps	10 to 40 Gbps	10 Gbps	1 Gbps	1 Gbps
Data	PBs	PBs	PBs	GBs to TBs	GBs to TBs and beyond
Shared File System	Yes	Yes	NFS/S3	NFS/S3	NFS/S3
ITAR	Small	No	Yes	No	Maybe
Managed By	NCCS	NCCS	600 & 700	NCCS & ESTO	700
Customization	Very Little	Yes	Yes	Yes	Yes

Advanced Data Analytics Platform (ADAPT)
Private High Performance Science Cloud

Why did the NCCS create ADAPT?

High Performance Computing	Data Analysis

Takes in small amounts of input and creates large amounts of output…

- Using relatively small amount of observation data, models are run to generate forecasts
- Tightly coupled processing requiring synchronization within the simulation
- Simulation applications are typically 100,000’s of lines of code
- Production runs of applications push the utilization of HPC systems to be very high
- Fortran, Message Passing Interface (MPI), large shared parallel file systems
- Rigid environment – users adhere to the HPC systems

Takes in large amounts of input and creates a small amount of output…

- Use large amounts of distributed observation and model data to generate science
- Loosely coupled processes requiring little to no synchronization
- Analysis applications are typically 100’s of lines of code
- Require more agile development with many small runs; utilization can be low on average
- Python, IDL, Matlab, custom
- Agile environment – users run in their own environments
- Steep learning curve for these users to take advantage of HPC resources or even public clouds

Where is this cloud located?

Physically located at GSFC, Building 28

Combination of new and old hardware

- New equipment for storage and management
- New servers for compute
- Retired High Performance Computing servers for virtual machines

Welcome to drop by and take a tour!

- Doesn’t really look this clean.

System Components/Configuration

	Capability and Description	Configuration
	**Persistent Data Services** Virtual machines or containers deployed for web services, examples include ESGF, GDS, THREDDS, FTP, etc.	128 GB of RAM, 10 GbE, and FDR IB
	**DataBase** High available database nodes with solid state disk.	128 GB of RAM, 3.2 TB of SSD, 10 GbE, and FDR IB
	**Remote Visualization** Enable server side graphical processing and rendering of data.	128 GB of RAM, 10 GbE, FDR IB, and GPUs
	**High Performance Compute** More than 3,800 cores coupled via high speed Infiniband networks for elastic or itinerant computing requirements.	~300 nodes with between 24 and 256 GB of RAM; Small set of nodes with 6 TB of SSD
	**High-Speed/High-Capacity Storage** Petabytes of storage accessible to all the above capabilities over the high speed Infiniband network.	Storage nodes configured with ~10 PB’s of usable capacity
	**High Performance Networks** Internal networks enable high speed access to storage, while external networks provide high performance data movement.	External: 10 and 40 GbE Internal: 10 GbE and Infiniband

Software Stack

External License Servers

Open Source Tools
Python, NetCDF, GDAL, R, etc.

Commercial Tools
Intel Compiler (C, C++, Fortran), IDL (4 seats)

Operating Systems
Linux (Debian, CentOS) and Windows

- Virtual machines can be set up to reach out to external license servers
- Time is needed to make requests to poke holes through various NASA firewalls

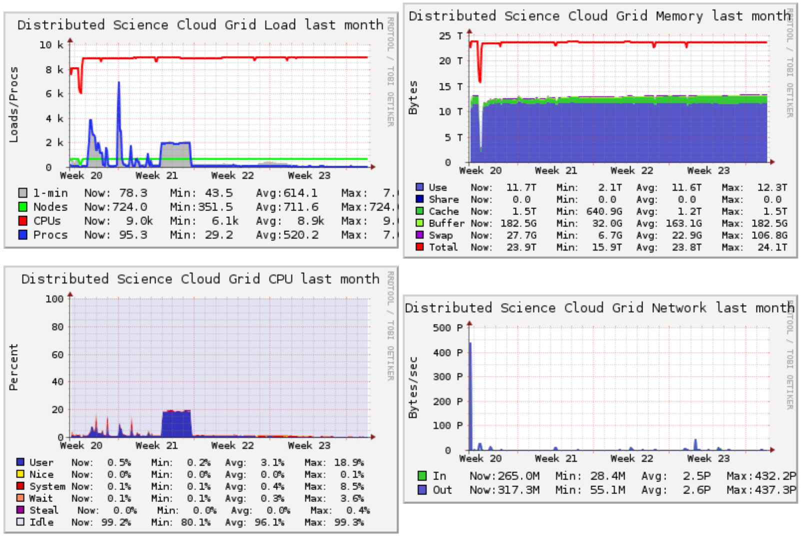
Open source tools:
- Very flexible
- If the open source tool does not need elevated privileges to install, the user can install the software in their home or scratch directories
- Commonly used tools may be installed in a shared directory for multiple users; the NCCS can assist with this as needed
- If the tool requires elevated privileges, users should submit a ticket to the NCCS for assistance. That tool will have to go through a security vetting process before it can be installed

Platform-as-a-Service
- Compute virtual machines are open souce Linux
- Windows used for remote desktop and ArcGIS

Virtual machines can be customized based on the end user application needs. The NCCS will work with you to create customized VMs specific to meet your needs.

Start Small and Scale Up

- Typically start users with 1 to 4 virtual machines each with 4 cores and 9 GB of RAM (about 2 GB/core)
- Once the user has developed or ported their application, the system administrators will scale up the number of VMs
- Need to make sure the resources are available prior to scaling up
- Scheduling resources so far has not been a problem
- As is typical with virtual environments/clouds, utilization varies greatly over time



Snapshot of the Ganglia monitoring tool showing the utilization of resources over a period of time.

Staged/Common Data Sets in ABoVE Science Cloud

Common datasets "Staged" for ABoVE investigators in ABoVE Science Cloud

- Staged and available for direct use
- Individual investigators don't have to invest time to locate and transfer data into system
- Avoids duplications of large datasets on system
- Additional datasets can be added, including generated data from ABoVE PI
- Data Services Manager to locate data

Example Download Times For 80TB

Speed	Time HH:MM:SS
9.6 Kbps	18518518:31:06
14.4 Kbps	12345679:00:44
28.8 Kbps	6172839:30:22
33.6 Kbps	5291005:17:27
56 Kbps	3174603:10:28
64 Kbps (ISDN)	2777777:46:40
128 Kbps (ISDN-2)	1388888:53:20
256 Kbps	694444:26:40
512 Kbps	347222:13:20
1.024 Mbps	173611:06:40
1.544 Mbps (DS1, T1)	115141:02:52
2.048 Mbps (E1, ISDN-32)	86805:33:20
10 Mbps (10Base-T)	17777:46:40
25.6 Mbps (ATM25)	6944:26:40
34 Mbps (E3)	5228:45:29
45 Mbps (DS3, T3)	3950:37:02
51 Mbps (OC1)	3485:50:19
100 Mbps (100Base-T)	1777:46:40
155 Mbps (OC3)	1146:57:12
622 Mbps (OC12)	285:48:58
1 Gbps (1000Base-T)	177:46:40
2.4 Gbps (OC48)	74:04:26
10 Gbps (OC192)	17:46:40

7.4 Mbps average US internet speed

1 Gbps NASA / Other Gov

@10 Mbps
Days: 741
Weeks: 106
Months: 24

@1 Gbps
Days: 7
Weeks: 1.1
Months: 0.25

Current ABoVE Science Cloud Data Holdings

Large Collections	Amount
Landsat	186 TB
MODIS	MODAPS collection remotely mounted
MERRA & MERRA2	406 TB
DigitalGlobe Imagery	2.8 PB
Total	> 3 PB

Other Data Sets

- Elevation datasets
  - ArcticDEM
  - CDEM
  - ASTER GDEM
  - Etc.
- Vegetation products
- Land cover products
- Products generated by the science team
- *Others as the team requests…*

Note that the ABoVE Science Cloud is not a permanent repository or the definitive source for this data. Official ABoVE products will be curated by the DAACs.

NGA/DigitalGlobe High Resolution Commercial Satellite Imagery

National Geospatial Intelligence Agency (NGA) has licensed all DigitalGlobe ≥ 31 cm satellite imagery for US Federal use, i.e., NSF, NASA and NASA funded projects.

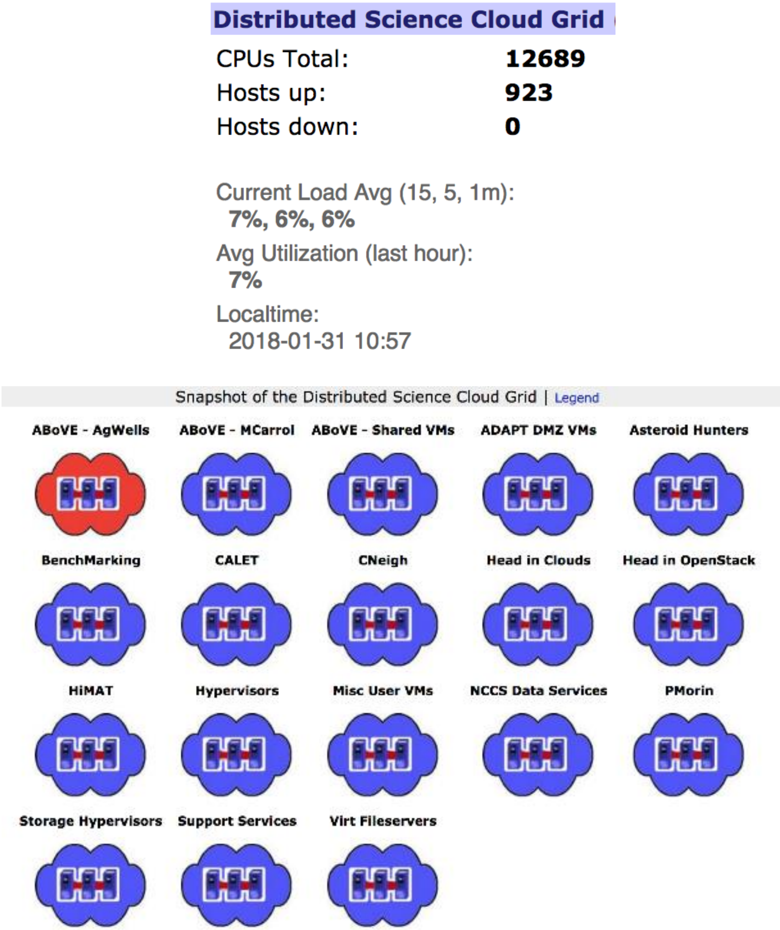
- Archive of >4.2 billion km^2 of data from 2000 to present
- Data from six different satellites: Worldview-1, 2 and 3; Ikonos; Quickbird; and Geoeye-1

Satellite	Bands	Nadir Panchromatic Resolution (m)	Nadir Multispectral Resolution (m)
Ikonos	Pan, R, G, B, Near IR	0.82	3.2
GeoEye	Pan, R, G, B, Near IR	0.41	1.65
Quickbird	Pan, R, G, B, Near IR	0.55	2.16
WorldView-1	Panchromatic only	0.5	N/A
WorldView-2	Pan, R, G, B, Near IR 1, Near IR 2, Coastal, Red Edge, Yellow	0.46	1.85
WorldView-3	Same as WV-2 plus 8 SWIR bands and 12 CAVIS bands	0.31	1.24

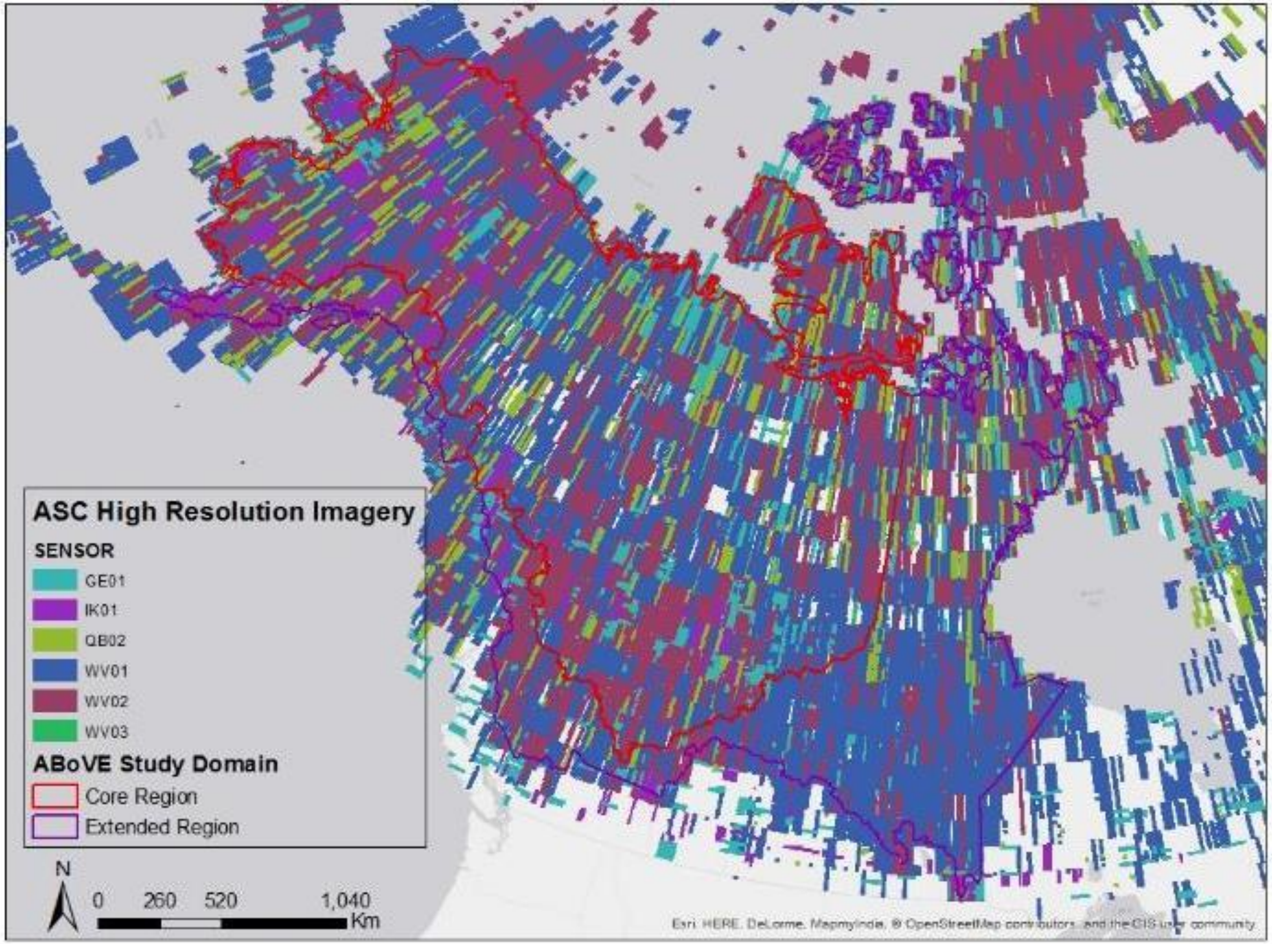
DigitalGlobe Satellite Fleet

ADAPT Use Cases

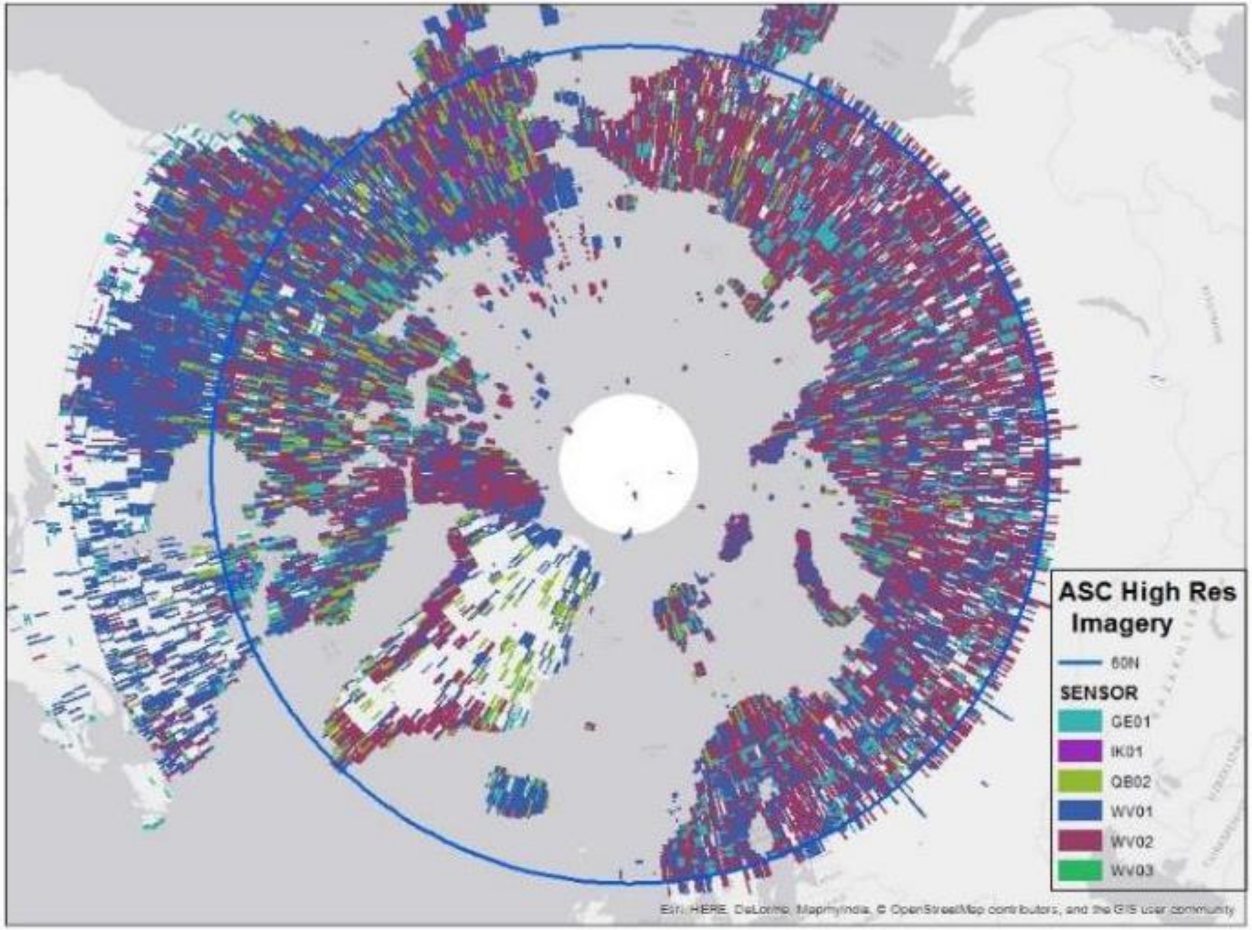
- Arctic Boreal Vulnerability Experiment (ABoVE)
- CALET (CALorimetric Electron Telescope)
- High Mountain Asia Terrain (HiMAT)
- Asteroid Hunters – Near Earth Objects
- Biomass in South Sahara
- NCCS Data Services
- Laser Communications Relay Demonstration (LCRD) Project -



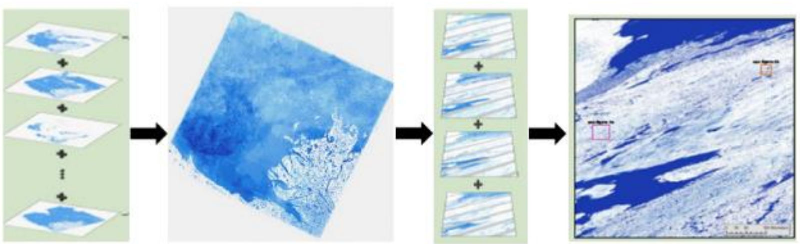
ABoVE Science Cloud DigitalGlobe Imagery: Study Domain



ABoVE Science Cloud DigitalGlobe Imagery: Circumpolar

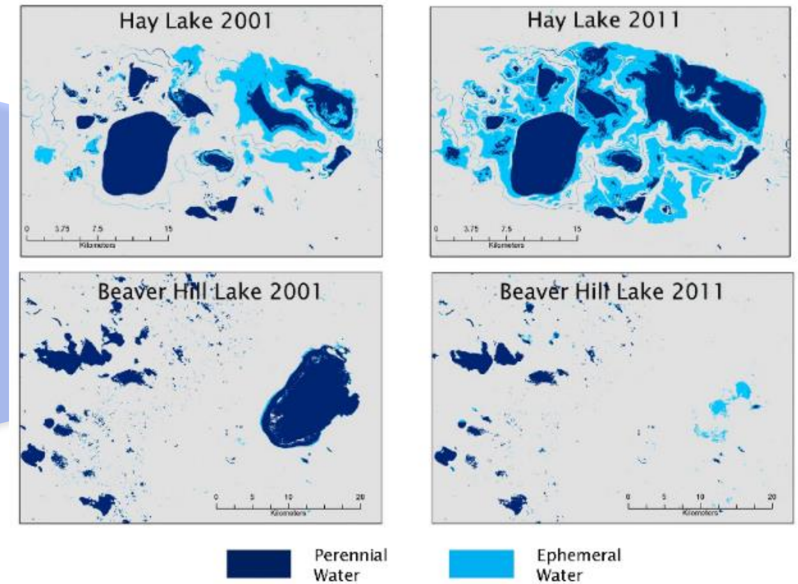


AboVE Water Maps: 30 meter spatial resolution surface water 1991-2011," M.L. Carroll, et. al



Processing work flow for the generation of the ABoVE water maps from Landsat scenes to ABoVE tiles.

100,000 LandSat Scenes
20 TB of Data



AWM for 2001 and 2011 for Hay Lake and Beaver Hill Lake in Canada. Hay Lake has clearly expanded over this time frame while Beaver Hill Lake has diminished.

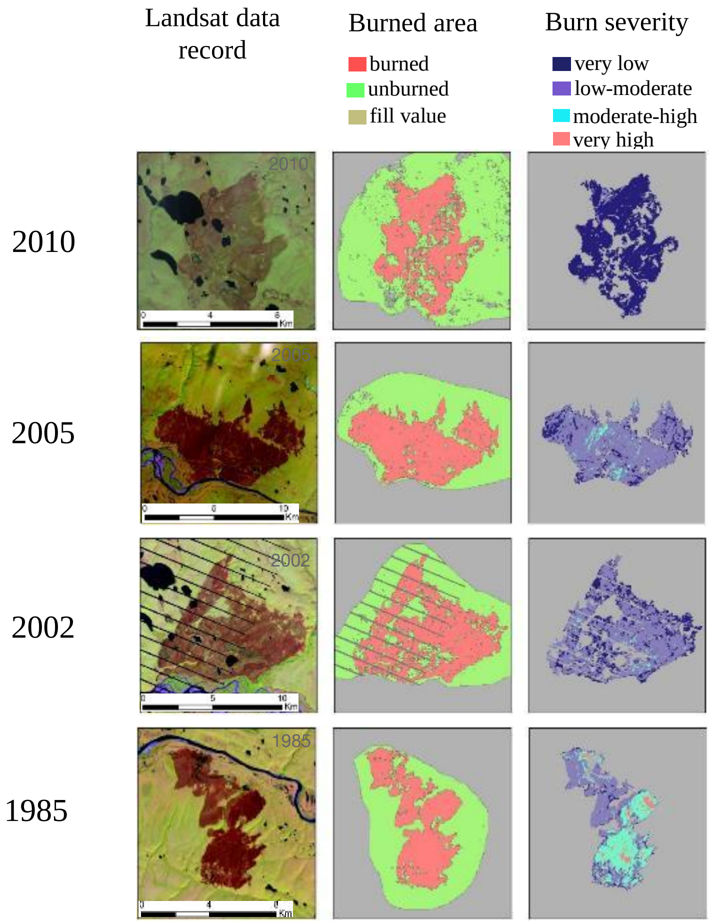
- Takes in large amounts of input and creates small output
- Using large amounts of observation or model data
- Python code of 100's of lines
- Easily run in parallel across multiple virtual machines

Taken from "AboVE Water Maps: 30 meter spatial resolution surface water 1991-2011," M.L. Carroll, et. al, http://above.nasa.gov/pdfs/ABoVE_water_maps_user_guide_05102016.pdf

Fire History for ABoVE – T. Loboda & M. Miller

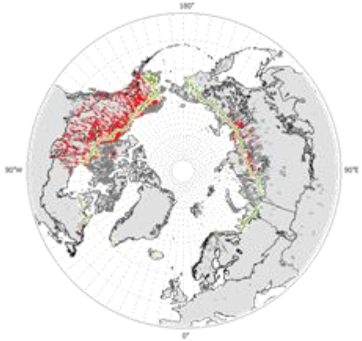
- Fire history across the ABoVE study region is compiled from available and new (Miller et al. in prep) data products and enhanced

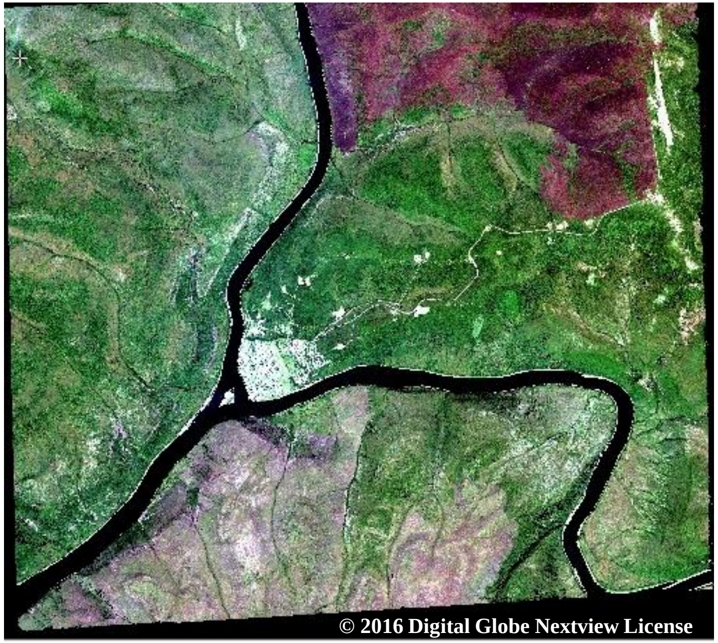
- Multiple VMs on the ASC are used to process Landsat and MODIS data to develop the burn severity characterization



Forest Canopy Surface Elevations – C. Neigh & P. Montesano

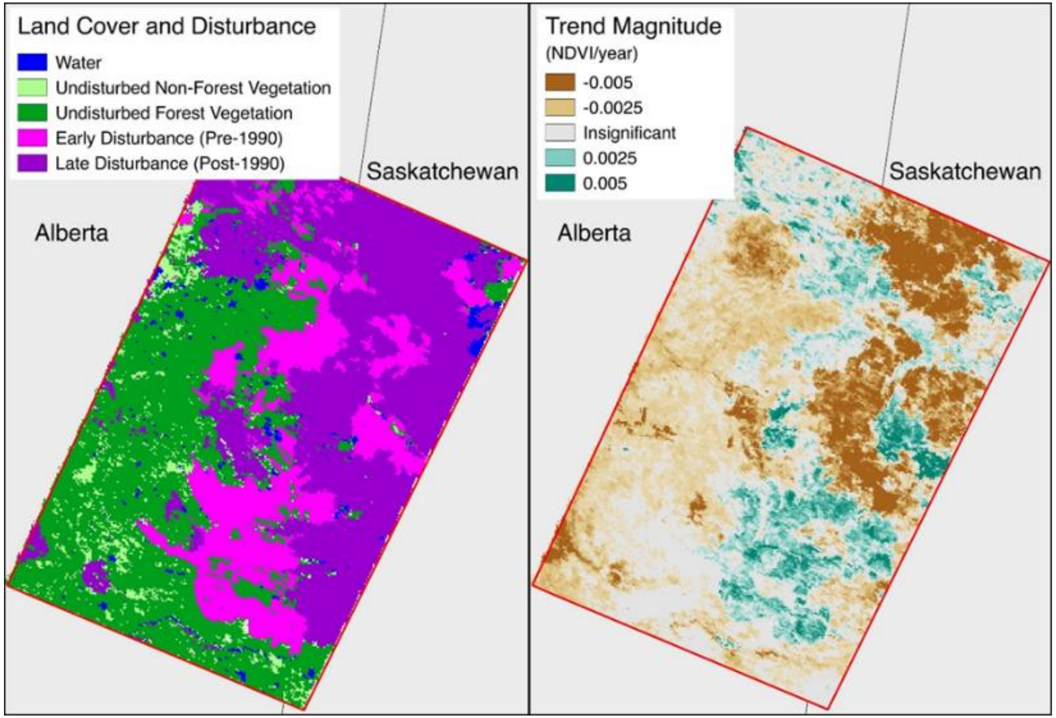
- Understanding forest patterns using DigitalGlobe high-resolution satellite imagery
- Using multiple VMs and Ames Stereo Pipeline (ASP) on the ASC to process Digital Elevation Models





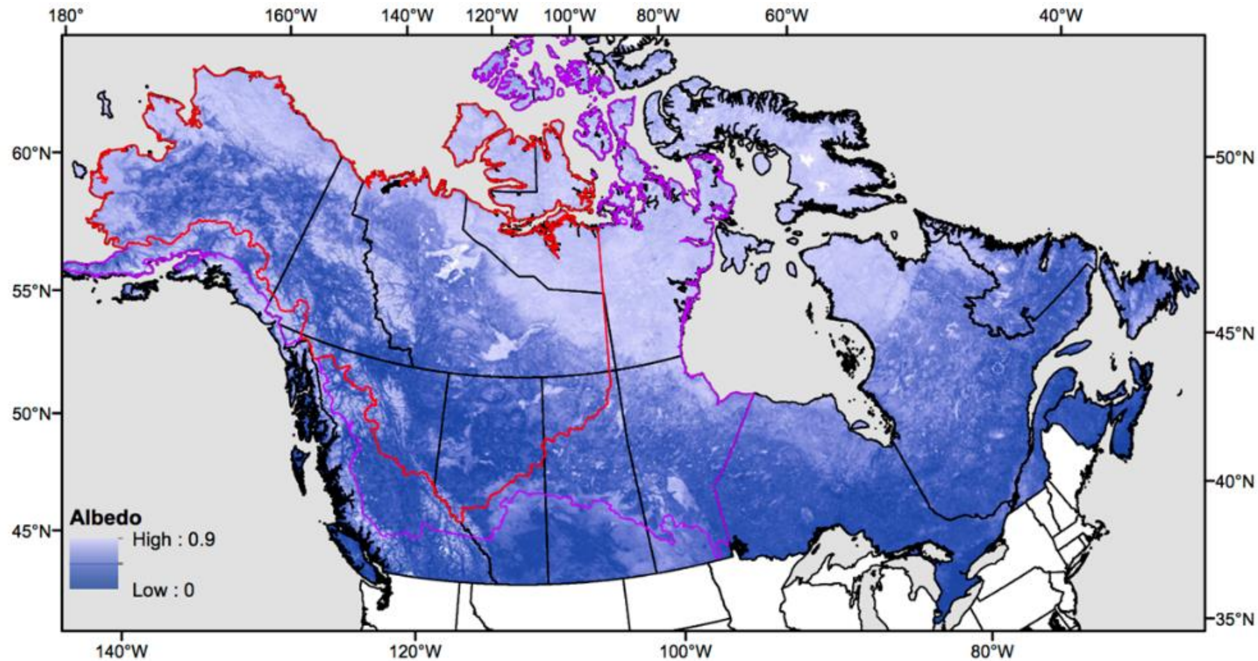
Landscape-Scale Histories of Disturbance, Seasonality and Greenness Trends - C. Woodcock & D. Sulla-Menashe

- 30+ year historical record and ongoing characterization of disturbance events and phenology across the ABoVE study domain
- Using multiple VMs to move Landsat data into the ABoVE grid and then develop the landscape histories



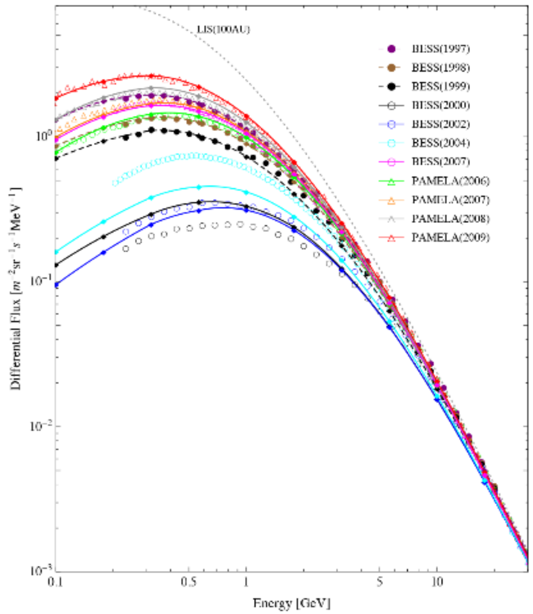
Understanding the Causes and Implications of Enhanced Seasonal CO_2 Exchange in Boreal and Arctic Ecosystems – B. Rogers

- Modeling driving factors of post-fire albedo trajectories
- Creation of mean albedo maps
- Fire combustion mapping



ADAPT Enables Cosmic Ray Science Advances (1 of 2)

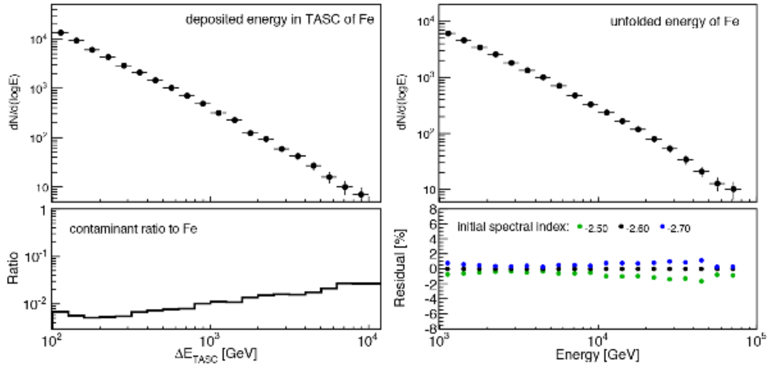
- The NCCS Advanced Data Analytics Platform (ADAPT) enabled science advances presented at the 35th International Cosmic Ray Conference, July 2017, Busan, South Korea.
- Shoko Miyake, National Institute of Technology, Ibaraki College, Japan, led an investigation into how the solar wind modulated the charges of galactic cosmic ray (GCR) protons and antiprotons throughout the heliosphere over solar cycle 23 (1997–2009).
- Using ADAPT, they calculated the energy spectra of GCR protons and antiprotons and compared the numerical results with observations by BESS (Balloon-borne Experiment with a Superconducting Spectrometer) and PAMELA (Payload for Antimatter Matter Exploration and Light-nuclei Astrophysics).
- Calculated antiproton/proton ratios show good agreement with observations, except for a 2004 discrepancy due to the calculated energy spectrum of GCR protons being lower than observed.
- More information:
  - S. Miyake, 2017: Charge-Sign Dependence in the Solar Modulation During the Solar Cycle 23. PoS(ICRC2017), 018.



Calculated and observed energy spectra of galactic cosmic ray GCR protons. Solid or dashed lines indicate the energy spectra of GCR protons obtained by the calculation. Figure by Shoko Miyake, National Institute of Technology, Japan.

ADAPT Enables Cosmic Ray Science Advances (2 of 2)

- Yosui Akaike, GSFC/University of Maryland, Baltimore County, led research on measuring cosmic-ray fluxes with the CALorimetric Electron Telescope (CALET) installed on the International Space Station.
- ADAPT assisted Akaike and the CALET Collaboration in obtaining the first preliminary results of heavy cosmic-ray nuclei from CALET data over the period October 13, 2015 to March 31, 2017.
- Monte Carlo simulations run on ADAPT estimated the efficiency of each analysis procedure, e.g., estimating the contaminant ratio from neighboring charged nuclei onto iron nuclei, the heaviest element in the study (see figure).
- The researchers demonstrated CALET’s ability to measure heavy cosmic-ray nuclei and also derived preliminary energy spectra for the primary cosmic ray elements.
- More information:
    - Y. Akaike and on Behalf of the CALET Collaboration, 2017: Measurements of Heavy Nuclei With the CALET Experiment. PoS(ICRC2017), 181.

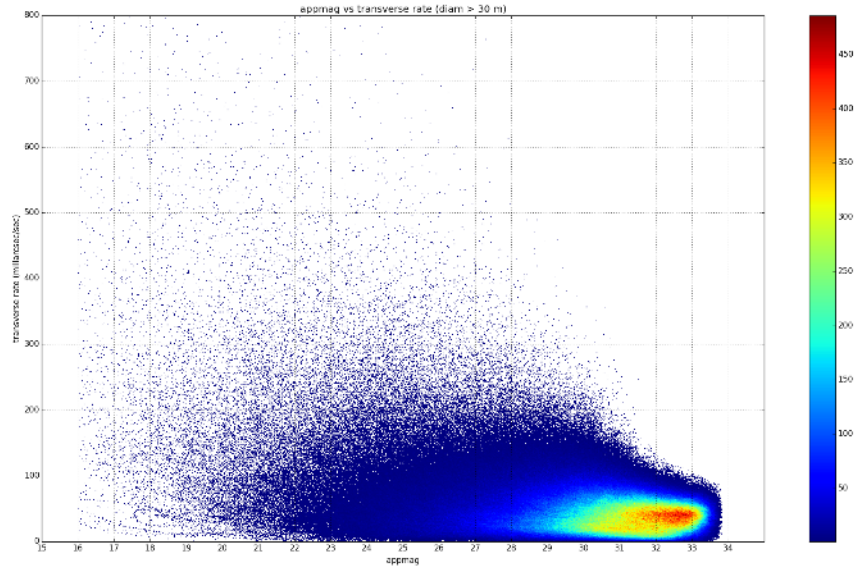


(Left top) The energy distribution in Total AbSorption Calorimeter (TASC) observations of iron. (Left bottom) Contaminant ratio from neighboring charged nuclei to iron estimated from Monte Carlo simulations. (Right top) The unfolded energy distribution of iron. (Right bottom) Systematic uncertainty depending on the initial spectral index in the unfolding procedure compared to an index of -2.60. Figure by Yosui Akaike, GSFC/UMBC.

ADAPT Enables NEO Survey Simulations

New NEO survey simulations and studies facilitated by the ADAPT system help meet a number of GSFC and NASA NEO research needs

- NEODAC simulation models the performance of both GSFC and NASA proposed survey missions
- Supports modelling of a complex sky survey and exploration of the duty-cycle/pointing-scheme trade space
- Supports rapid testing of various detection models
- With ADAPT, a sim with 60~ million objects propagating at time-steps of a 5-15 seconds over a few months can be completed in 2-4 days. Outputs can be processed with new detection models and scan patterns in minutes.



Heatmap of synthetic NEO population's brightness and speed over a few days.

Moving Toward a Hybrid Cloud

Ability to burst into commercial clouds as needed depending on science requirements.

Cloud services paid for by the science project.

Leveraging the NASA CIO Enterprise Cloud Computing services, science projects would provide funding to a WBS in NASA to get to commercial cloud offerings.

Potentially burst into multiple commercial clouds.

Leverage the best value solution for the science application.

Would like to test with Azure.

Capability currently available from ADAPT to AWS.
Would like to find a project willing to test bursting into Azure.

Estimate Biomass in South Sahara Bursting from ADAPT to AWS

Project Goal

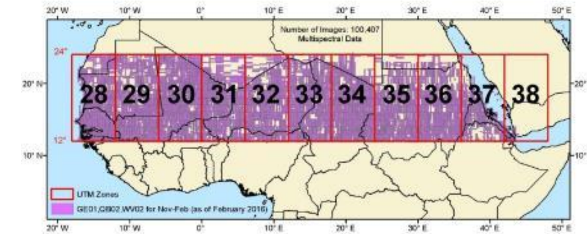
- Using National Geospatial Agency (NGA) data to estimate tree and bush biomass over the entire arid and semi-arid zone on the south side of the Sahara

Project Summary

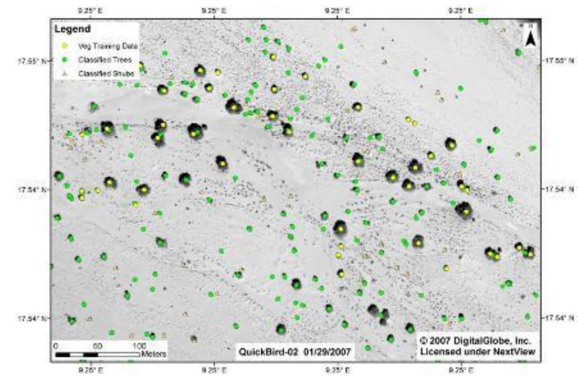
- Estimate carbon stored in trees and bushes in arid and semi-arid south Sahara
- Establish carbon baseline for later research on expected CO_2 uptake on the south side of the Sahara

Principal Investigators

- Dr. Compton J. Tucker, NASA Goddard
- Dr. Paul Morin, University of Minnesota
- Dr. Tsengdar Lee, NASA HQ



The project focuses on the Sub-Sahara region of Africa and encompasses Universal Transverse Mercator (UTM) zones 28 through 38.



Vegetation classifier algorithms identify trees and bushes in UTM Zone 32. Image by Jamon Van Den Hoek, Gerasimos Michalitsianos, and John L. David, GSFC.

Reference: Tucker and Morin are extending earlier tree and bush mapping work published by Gonzalez, Tucker, and Sy entitled "Tree density and species decline in the African Sahel attributable to climate" in the Journal of Arid Environments in 2012.

Goddard Private Cloud (GPC)

Goddard Private Cloud

Joint Project between Code 600 and 700

- Component of the Goddard Strategic Information Technology strategy
- Leverage code 600's experience in ADAPT
- Leverage code 600's and 700's experience in cloud computing
- Security and security plan covered by Code 600/700

Bridge to Public Clouds

- Can handle both non-ITAR and ITAR (working on containers)
- Many applications can be virtualized but may not be truly be ready for a cloud
- Many uses don't know exactly how to leverage cloud computing
- This provides a cost effective solution for consolidation of applications and services while learning more about cloud computing
- Users can work on cloud enabling their application in a low-cost, low-risk environment
- Works best if we can partner with a system administrator familiar with the user application

Current Prototypes

Prototypes

- Purchasing new equipment for the cloud management and storage
- Reuse HPC compute servers and some storage to keep the overall costs low
- Equipment in B28 and B32
- Using the SEN network
- Built using OpenStack
- Windows and Linux
- Non-ITAR and ITAR
- Containers using OpenShift

Prototype Hardware Inventory
Server and Storage Summary (as of January 2018)

Function	Openstack	Server Count	CPU-Cores	Sum of RAM (GB)	Storage (TB)
hypervisor-galaxy (non-ITAR)	Mataka	8	128	2,048	
hypervisor-nebula (ITAR)	Mataka	12	192	2,304	
hypervisor - B28 not deployed yet	Pike	28	448	5,376	
hypervisor - B32 not deployed yet	Pike	76	1216	14,592	
NFS Fileserver-galaxy		1	16	32	
NFS Fileserver-nebula		1	16	32	
Network Storage - galaxy		2	32	64	960
Network Storage-nebula		8	48	192	192
Network Storage -B32 test string	Pike	8	96	1,168	180
openstack-control		7	28	160	
openstack-control (nebula Neutron)		2	32	64	
openstack-control (virtual control plane)		4	64	256	
xcat server (cluster management)		1	4	24	
Grand Total		**158**	**2,320**	**26,312**	**1,332**

** CPU Cores denotes physical Cores per CPU. SCU8 (IBM) servers are dual CPU servers*

NAME	MANUFACTURER	MODEL	SPEED
switch-621f60	Mellanox		40gbps (running @10gpbs)
gpc-r1-esw01	Dell	S3048-ON	1gbps
gpc-r1-eswx01	Dell		10gbps
gpc-r2-esw01	Dell	S3048-ON	1gbps

GPC Projects

- **HelioViewer**
- **KeyShot**
- **Wind Chill/Creo**
- **SEEC (Sellers Exoplanet Environments Collaboration)**
  - EMAC (Exoplanet Modeling and Analysis Center) server transition to cloud
- **Mini-LHR**
  - Last analysis showed run time cut in half versus local machine
- **NIMO (Code 592)**
  - Support for model based system engineering
- **MESA (Code 590)**
- **HFSS (Code 500)**
  - High frequency Electromagnetic Field Simulation
- **Haisam Ido (Code 444)**
- **G-LiHT (Code 618)**

Plans

Prototype and Future Plans

- Currently running in Building 28
- Establish a more cloud storage environment (blocks and NFS)
- Work toward merging ITAR and non-ITAR using encryption (single cloud with multiple tenants)
- Expand use cases and work with users to help move their applications into the cloud

Production capability

- Initial production capability in late 2018 in Building 32
- Expand the cloud to be in two buildings (28 and 32) in 2019

Billing

- Currently still working this out
- Target to be similar or better costs to AWS

AIST Managed Cloud Environment (AMCE)

AIST Managed Cloud Environment (AMCE)

System Owner

- Mike Little
- Earth Science Technology Office (ESTO)
- Advanced Information Systems Technology Program (AIST)

Operations

- Jointly operated by ESTO and the NASA Center for Climate Simulation (NCCS – Code 606.2)

Concept of Operations

AMCE provides and manages

- Amazon Web Services infrastructure, platform, and services
- Management without interference
- Security - shifting the burden of security to AMCE
- Cloud overlay tool

Projects compute and customize

- In an environment protected by logical isolation
- Exercising freedom and control
- Outside of the NASA network (reduced risk)
- NASA credentialed and non-NASA credentialed users
  - Projects decide who has access and progress is not held up waiting for NASA credentials
- Projects supply the funding and an system administrator

By using the AMCE, projects will have NASA-compliant work more easily infused into NASA applications.

Benefits

Flexibility

- "Fail smartly, fail cheaply"
- Very little start-up time
- Quickly access new tools
- Scalable

Technology Infusion

- Increase the number of AIST projects adopted into operations
- Reduce total cost and time of research to operations
- Quick, safe, and inexpensive test and development environment
- Incorporates NASA standards from the start
- End users can test and provide feedback early in the process
- Increase use of NASA science data

Access

- Allows for NASA and non-NASA partners
- Try and share new tools
- Store, share, and analyze large datasets
- Reduces duplicative efforts

Security

- Ensure NASA IT security compliance from the beginning
- Provide security compliant images to build upon
- No need for projects to spend significant time modifying images to be security compliant
- R&D outside the NASA firewall

AMCE Projects So Far

21 Projects on-boarded to AMCE overall

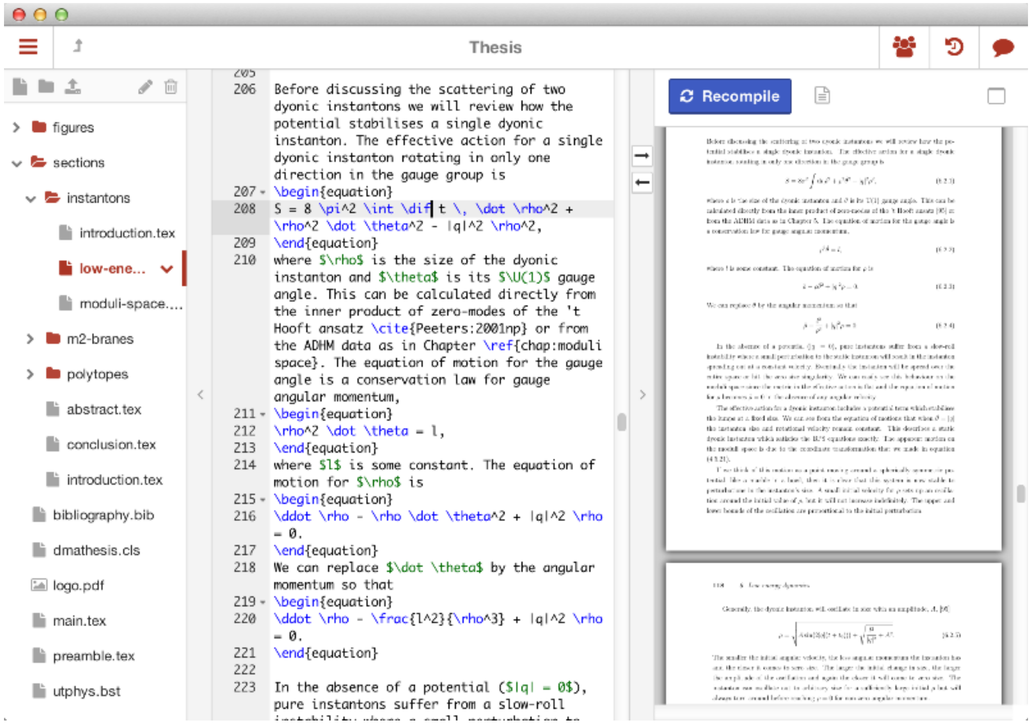
- 18 active projects on-boarded to AMCE
- 3 in-active projects that completed their work and either archived or moved their project data
- 3 projects preparing for on-boarding

Variety of Projects using AMCE

- Science and technology development research projects
- Document development and sharing projects
- Data analytics services projects
- Mission data access services projects

AMCE Example Project ShareLatex

- Tom Clune (GMAO)
- Virtual machine with ShareLatex software
- Facilitates shared document creation for collaborators working on papers or grant proposals
- Both NASA and non-NASA credentialed users
- Tom wants to get more users in order to share the cost of this service
- Estimated monthly total budget of $200



Wrap Up

Future Remains Cloudy

ADAPT

- Integration of systems with GPUs configured specifically for machine learning and deep learning
- Integration of analytics-as-a-service capabilities to enable remote computing capability from desktops

GPC

- Move to production support
- Expand capabilities

AMCE

- Expand capabilities to leverage more of AWS services

Lots more to Discuss

- Containers – Project to containerize LIS for the AMCE
- Best Practices and How to
- Software as a service
- AWS Services such as Lambda (serverless or transactional processing)

Sounds Great … When can we get access?

All are ready now!

- Though there is always paperwork involved!

AMCE

- Mike Little – m.m.little@nasa.gov

ADAPT

- Daniel Duffy – daniel.q.duffy@nasa.gov
- Ellen Salmon – ellen.m.salmon@nasa.gov

GPC

- Hoot Thompson – john.h.thompson@nasa.gov

In all cases

Backup Slides

AMCE Overview and Status

ATO granted July 27, 2016 and renewed each year.

Data Categorization

- Low - No ITAR or SBU

Operations

- AMCE Helpdesk
- Resource Adjustments as needed and on demand

Users

- Role-based model - user permissions are based on the role's required functions

Machine Maintenance

- Scanning (Nessus) – project machines and management SW weekly, image templates monthly
- Projects review scan reports and patch - Projects patch their own machines, AMCE patches management software and image templates
- Backups - Projects backup their own machines, AMCE backs up image templates weekly

Logging

- Splunk logs user activity; continuous real-time alerts; weekly log reports to projects; monthly log

AMCE Benefits

For the PI:	For the Program:
Enables quick and easy access to a lot of data and computing	Lowers program costs
Secure template machine images	Security and compliance
Supports elasticity of resources	Facilitates user and expense monitoring
Accelerates project research	Reduces risk
Encourages collaboration	Pre-defines roles/permissions
Supports end-user implementation	Supports end-user implementation
Facilitates expense tracking	
Facilitates transitions	
Limits interference	

Flexibility

- “Fail smartly, fail cheaply”
- Perform research using new computing capabilities without extensive start-up time.
- Ability to try new tools and methods from AWS’s product catalogue easily and affordably.
- Easily scalable computing for high-demand, high-bandwidth needs.

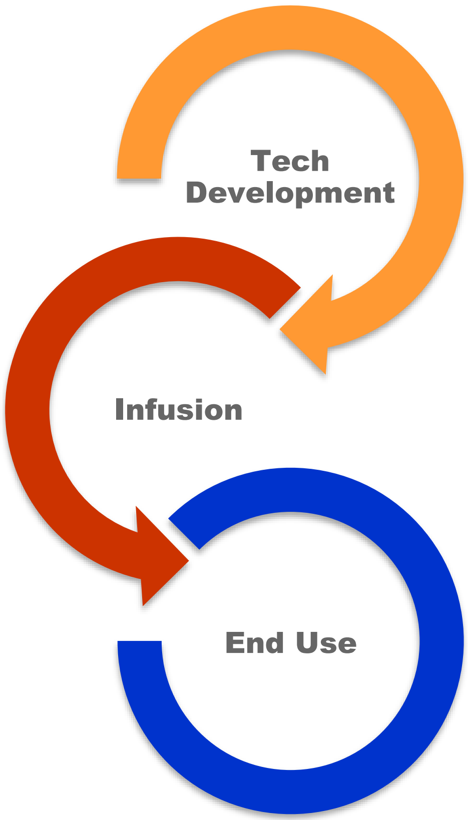
Access

- Provides a secure and isolated workspace that also allows for partnering opportunities.
- The collaborative environment is conducive to flexible teams with diverse sets of users.
- Easily use multiple datasets and try new tools without common transfer problems.

- Allows for storing, sharing and analyzing terabyte and petabyte-scale science data without extensive hardware investment or securing computing time.
- End users have a secure area to explore and test the newly-developed tools.

Technology Infusion

- Increase the number of AIST projects that are ready to be adopted and used by end users.
- Ends users can test and provide feedback to research teams early to accelerate the infusion process.
- Make it easier for end users to infuse new work into their systems and processes by incorporating NASA standards from the beginning.
- Reduce the total cost and amount of time it takes to typically infuse a new tool.
- Increase PI awareness of, and collaboration with, similar projects and initiatives.



Security

- Reduce risk of susceptibility and threats by ensuring NASA-compatible security requirements are followed.
- AMCE teams inside and outside of NASA have access to security-compliant platform that's outside the NASA firewall.
- The system manages established NASA security procedures in background.
- There's no need to set new security parameters with each instance. All secure functionality is built