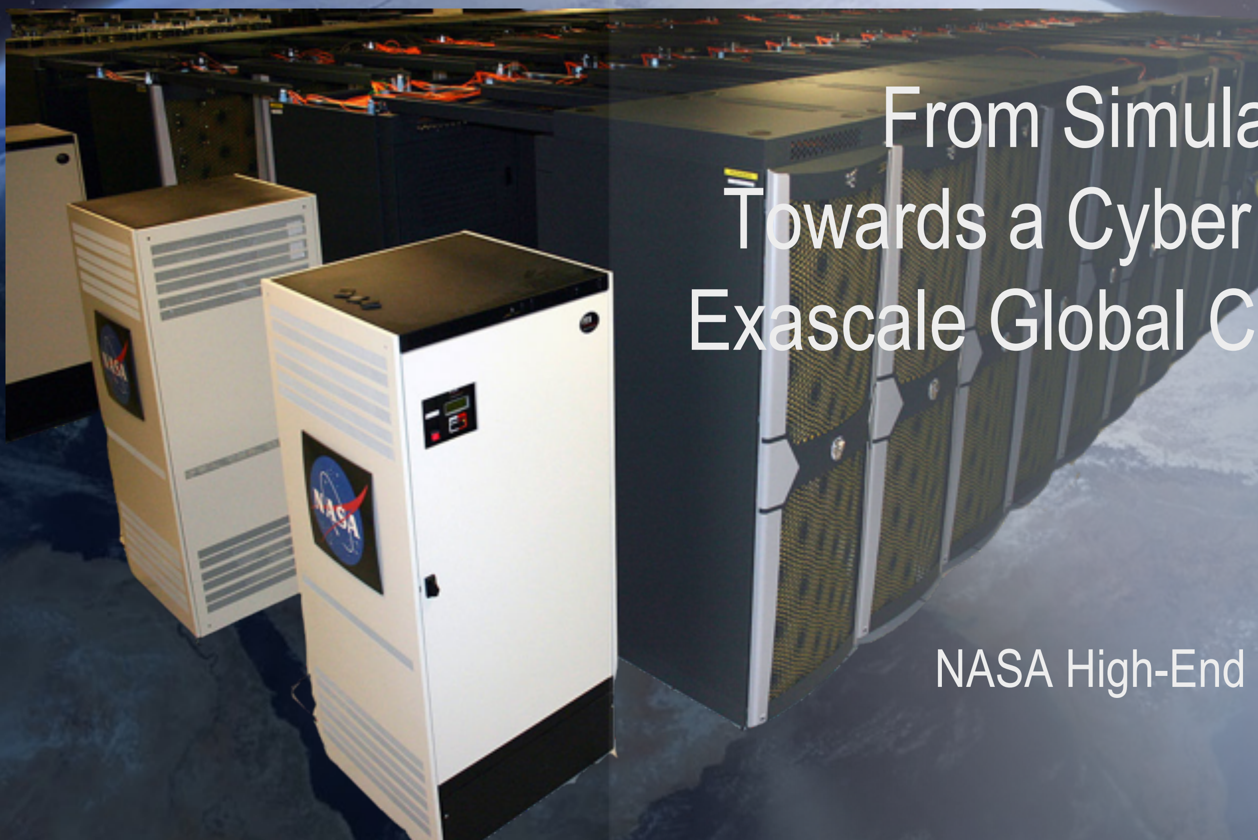
National Aeronautics and
Space Administration



From Simulation to Analytics: Towards a Cyber Infrastructure for Exascale Global Circulation Models

Dr. Tsengdar Lee
NASA High-End Computing (HEC) Program
NASA Headquarters

September 14, 2017

Authors

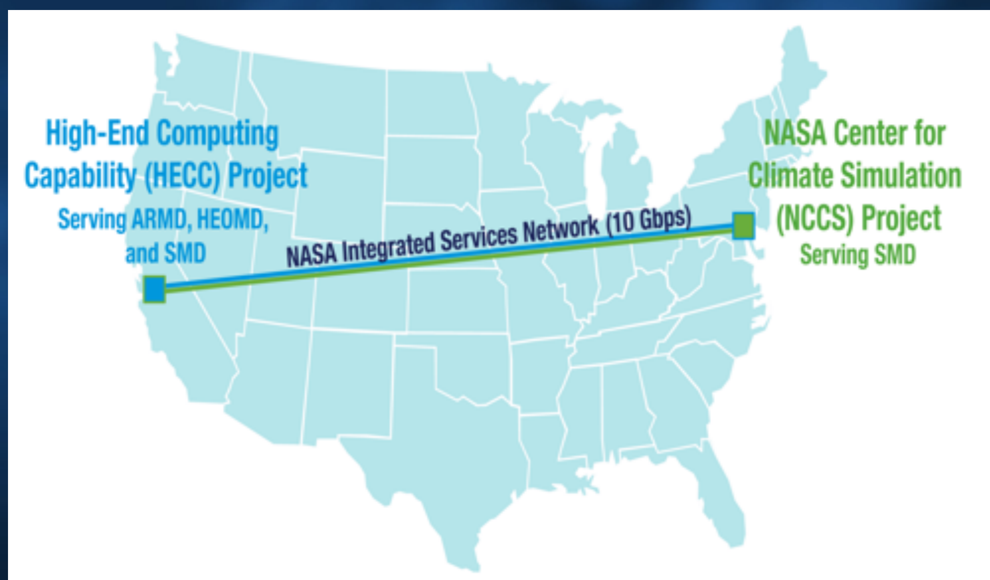
Tsengdar Lee (1), Dan Duffy (2), Carrie Spear (2), Michael Bowen (2), John Thompson (2), Fei Hu (3), Chaowei Yang (3), Tom Maxwell (2), Daniel Pierce (2), Laura Carriere (2), William Putman (4)

1. NASA High-End Computing (HEC) Program, NASA Headquarters, Washington, DC, USA.
2. NASA Center for Climate Simulation (NCCS), NASA Goddard Space Flight Center, Greenbelt, MD, USA.
3. George Mason University, Fairfax, VA, USA.
4. Global Modeling and Assimilation Office (GMAO), NASA Goddard Space Flight Center, Greenbelt, MD, USA.

HEC Program Mission and Content

- The mission of NASA's High-End Computing (HEC) Program is to:
  - Plan and provision HEC systems and services to support NASA's mission needs.
  - Operate and manage these HEC resources for the benefit of agency users, customers, and stakeholders.
- The HEC Program manages two projects, which serve all four NASA Mission Directorates: Aeronautics (ARMD), Human Exploration and Operations (HEOMD), Science (SMD), and Space Technology (STMD).

HECC provides world-class high-end computing, storage, and associated services to enable scientists and engineers supporting NASA missions to broadly and productively employ large-scale modeling, simulation, and analysis to achieve successful mission outcomes. It supports 1,200 users from NASA and around the U.S. HECC is run by the NASA Advanced Supercomputing (NAS) Division at Ames Research Center.



NCCS offers an integrated set of supercomputing, visualization, and data interaction technologies to enhance NASA capabilities in weather and climate prediction and enable future scientific discoveries that will benefit humankind. It serves hundreds of users at NASA centers, laboratories, and universities across the U.S. NCCS is run by the Computational and Information Sciences and Technology Office (CISTO) at Goddard Space Flight Center.

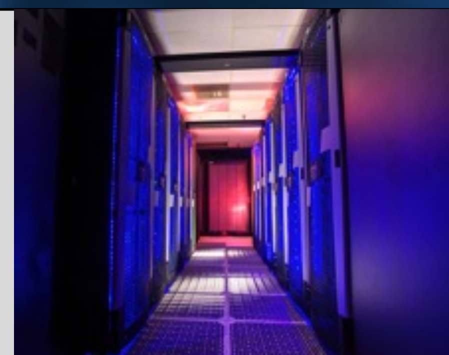
NASA Science Mission Diractorate/Earth Science Division
High-End Computing Portfolio (1)

SMD/ESD HEC Portfolio (a 7120.8 terminology) is comprised of two supercomputing center projects and a small research and analysis (R&A) element in Earth science.

Supercomputing Centers:
Supercomputing centers provide shared computing, networking, data analysis, visualization and user services.

ARC/High End Computing Capability (HECC)

GSFC/NASA Center for Climate Simulation (NCCS)



ARC Visualization Capability

GSFC Visualization Capability & SMD Hyperwall

SMD/ESD High-End Computing Portfolio (2)

- ESD ROSES element “Computational Modeling Algorithms and Cyber Infrastructure”:
  - Computational algorithms in GPU and new computing architectures
  - Tools and technologies to support management of data
  - Tools and technologies to support distributed model development
  - Tools and technologies to support model longevity
  - Summer Education and Internship Program
- SBIR
  - Technologies for Large-Scale Numerical Simulation
  - Algorithms and Tools for Science Data Processing, Discovery and Analysis, in State-of-the-Art Data Environments
  - Integrated Mission Modeling for Opto-mechanical Systems
  - Fault Management Technologies

HECC Advanced Computing Environment

HECC Compute Portfolio

4 Compute Clusters

•	Pleiades	161 Racks / 11,376 nodes / 7.58 PF / 32,308 SBU/hr
•	Electra	16 Racks / 1,152 nodes / 1.24 PF / 4,654 SBU/hr
•	Merope	56 ½ Racks / 1,792 nodes / 252 TF / 1,792 SBU/hr
•	Endeavour	3 Racks / 2 nodes / 32 TF / 140 SBU/hr

1 Visualization Cluster — 245 million pixel display / 128 node / 703 TF

Shared Storage Environment

• 7 Lustre File Systems	39.6 PB
• 6 NFS File Systems	1.5 PB
• Archive System	490 PB

Experimental Quantum D-Wave System with 2,031 qubits

Modular Computing Facility Deployment

Modular Computing Facility (MSF) Deployment

- Modular Supercomputer Facility Prototype fully operational. Provided Proof-of-Concept for major future expansion.
- Power Usage Effectiveness (PUE) of under 1.03, surpassing the 1.06 original target, saving half a million kilowatt-hours of energy and more than a million gallons of water annually.
- NIST National Renewable Energy Laboratory (NREL) visited the site and studied the design as it became a shining successful energy saving design.

Electra Installation

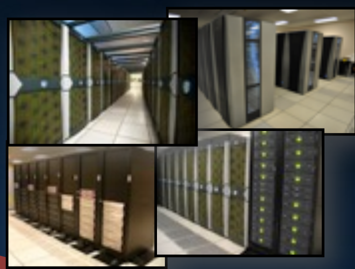
- 16 Broadwell racks providing 1.2 PF peak compute capacity. Achieved 1.096 PF on LINPACK, placing system at #96 on TOP500 list. Achieved 0.0252 PF on HPCG, placing system at #46 in the world

Storage Expansion

- Expanded the file systems from 42.8 PB to 61.0 PF

HECC Conducts Work in Four Major Technical Areas

Supercomputing Systems



Provide computational power, mass storage, and user-friendly runtime environment through continuous development of management tools, IT security, systems engineering

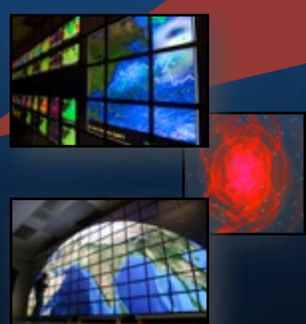
Application Performance and User Productivity

Facilitate advances in science and engineering for NASA programs by enhancing user productivity and code performance of high-end computing applications of interest

Supporting Tasks

- **Facility, Plant Engineering, and Operations:** Necessary engineering and facility support to ensure the safety of HECC assets and staff
- **Information Technology Security:** Provide management, operation, monitoring, and safeguards to protect information and IT assets
- **User Services:** Account management and reporting, system monitoring and operations, first-tier 24x7 support
- **Internal Operations:** NASA Division activities hat support and enhance the HECC Project areas

Data Analysis and Visualization



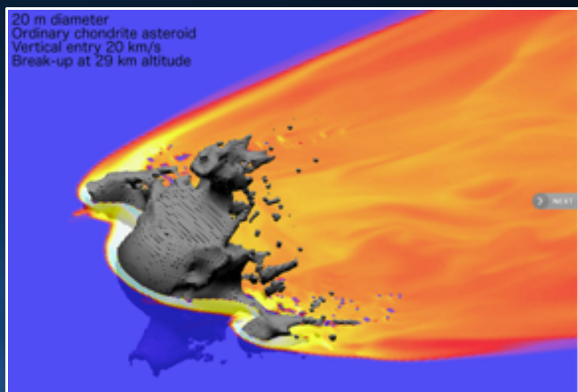
Create functional data analysis and visualization software to enhance engineering decision support and scientific discovery by incorporating advanced visualization technologies

Networking

Provide end-to-end high-performance networking analysis and support to meet massive modeling and simulation distribution and access requirements of geographically dispersed users

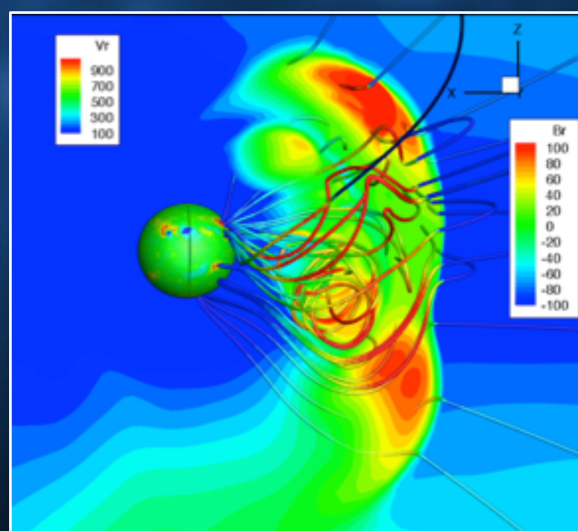
Science Support (94M SBUs*)

*October 1, 2015 to September 30,2016



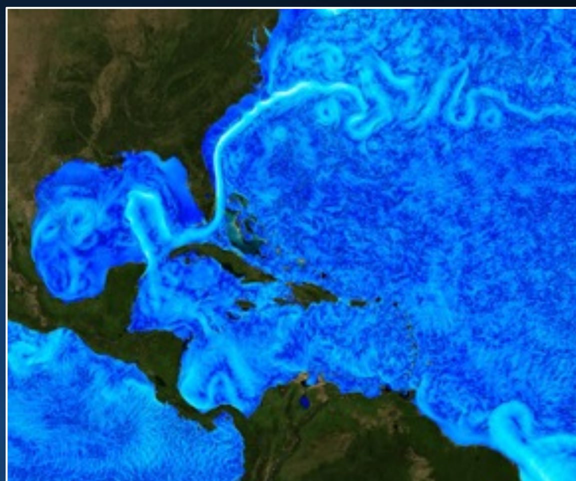
Researchers are running computational fluid dynamics simulations on Pleiades to determine the fragmentation and blast propagation that occurs when asteroids impact the atmosphere or the ground.

Researchers ran simulations on Pleiades to study coronal mass ejections (CMEs) – powerful expulsions of plasma from the sun that interact with Earth's magnetosphere, causing "space weather."

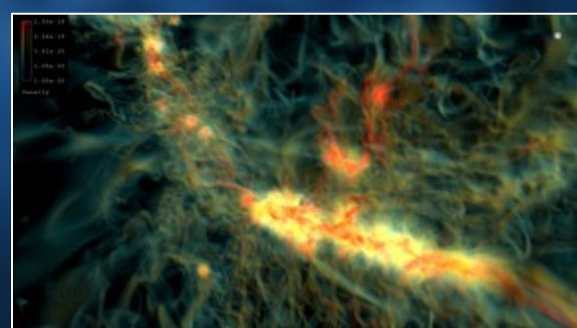


To support a growing number of projects that are developing new, large-scale science data products, the NEX team developed a hybrid HPC-cloud data production system during the last year.

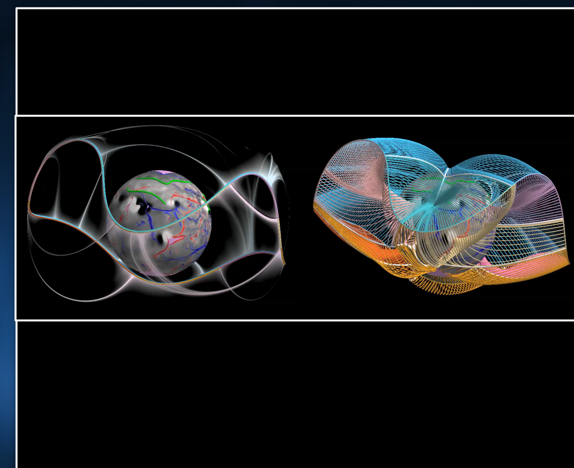
Science Support (Continued)



Researchers from the Estimating the Circulation and Climate of the Ocean (ECCO) consortium are running simulations on Pleiades to produce global, time-evolving maps of Earth's ocean and sea-ice system at an unprecedented resolution—about 1 kilometer (km) horizontal grid spacing.



Researchers are running large-scale radiation-magnetohydrodynamics simulations on Pleiades to explore the origin of massive stars and star clusters—one of the most fundamental unsolved problems in astrophysics.



Heliophysicists are running innovative simulations on Pleiades to investigate the structure and evolution of solar storms, with the ultimate goal of contributing to the development of predictive capabilities.

NASA Earth Exchange (NEX)

Provides a Complete Work Environment "Science as a Service"

COLLABORATION

Over 650 Members

CENTRALIZED DATA REPOSITORY

Over 2.3 PB of Data

COMPUTING

Scalable
Diverse
Secure/Reliable

KNOWLEDGE

Workflows
Machine Images
Model codes
Re-useable software

Public-Private-Partnerships
Helping NEX Community

Engaging a broader community of Earth scientists worldwide

OpenNEX

Lectures | Labs | Tutorials | Workshops

Commercial Cloud Partner (AWS)

Leveraging developments in the industry to help NASA's Earth science mission
Commercial Partner (NVIDIA)

NASA Center for Climate Simulation (NCCS)

Provides an integrated high-end computing environment designed to support the specialized requirements of Climate and Weather modeling.

- High-performance computing, cloud computing, data storage, and networking technologies
- High-speed access to petabytes of Earth science data
- Collaborative data sharing, publication, and analysis services

Primary Customers (NASA Science)

- NASA-funded science projects can get access to these resources
- Global Modeling and Assimilation Office (GMAO)
- Goddard Institute for Space Studies (GISS)
- Land Information Systems (LIS)
- Variety of other Research and Development (R&D) and Engineering customers
    - ABoVE, HiMAT, CALET, WFIRST

Funded by the High-End Computing (HEC) Program under SMD

- Dr. Tsengdar Lee, Program Manager

Based at NASA Goddard Space Flight Center in Greenbelt, MD

NCCS Climate Computing Facility

Discover Cluster
- Since 2006, NCCS has deployed a scalable cluster approach to meet the application requirements of large-scale climate simulations, with the number of cores increasing more than 35x and computational capability nearly 55x.
- Current Configuration
  - 3,394 nodes; 90,336 processor cores
  - 78.856 terabytes of memory
  - 45 petabytes of global file system storage
  - 3.478 petaflops peak
- Commodity Compute Cluster
  - Intel processors: Haswell and SandyBridge
  - IBM and SGI/SuperMicro nodes
  - 20, 40, and 56 gigabit-per-second (Gbps) InfiniBand networks
- Scalable Compute Unit Architecture
  - Add units over time with minimal down time
  - Applications can run on any node

GPU Cluster
– IBM iDataPlex

– 384 Intel Westmere cores; 64 NVIDIA Tesla M2070 GPUs with 28,672 GPU streaming cores

– 1.92 terabytes of memory

– 40 Gbps InfiniBand network

– 36.4 teraflops peak (double precision)

Dali GPU Analysis Nodes
- IBM iDataPlex m360
- 12 nodes; 144 Intel Westmere cores
- 24 NVIDIA Tesla M20708 GPUs with 10,752 streaming GPU cores
- 40 Gbps InfiniBand network
- 2.448 terabytes of memory

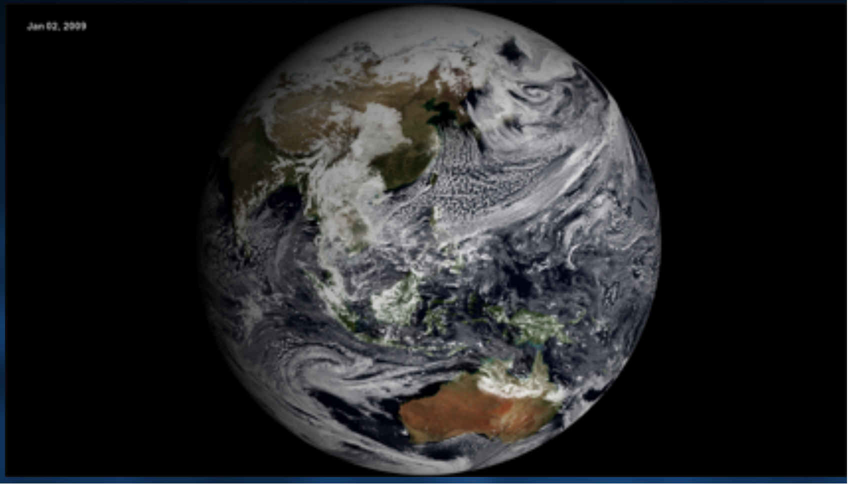
ADAPT—Advanced Data Analytics Platform
- Managed Virtual Machine Environment
- 100+ Hypervisors – Intel Xeon Westmere (2.8 GHz) and Intel Xeon Ivy Bridge (2.5 GHz) processor cores
- High-speed InfiniBand and 10 Gigabit Ethernet networks
- Linux and Windows Virtual Machines
- 10 petabytes of raw storage under Gluster file system management

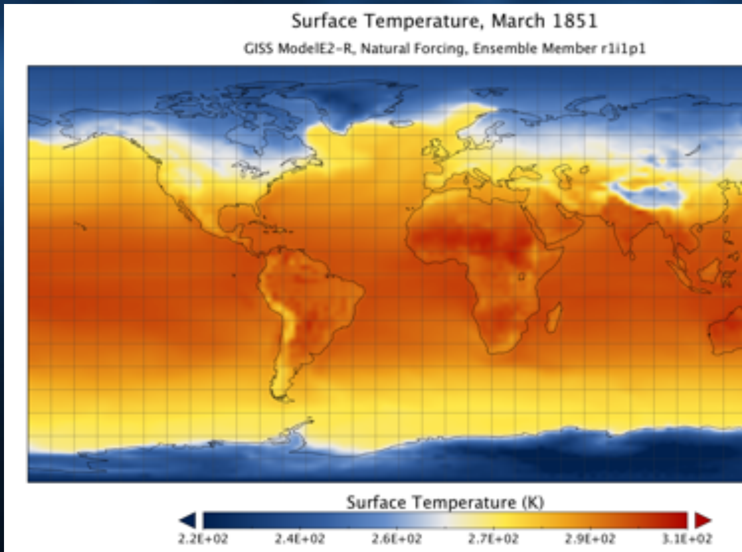
Mass Storage Archive
- StorageTek Silos
- T10KC and T10KD tape drives
- SL8500 tape libraries
- 94 petabytes of capacity

Data Portal
- HP ProLiant DL380p Gen8 Hypervisors
- 8 nodes; 80 Intel Haswell cores
- 128 gigabytes of memory

NASA's Climate and Weather Modeling

- Spans timescales from weather to short-term climate prediction to long-term climate change.
- Spans weather, climate, atmospheric composition, water & energy cycles, carbon cycle.
- "Unique" in bringing models and observations together through assimilation and simulation.
- Products to support NASA instrument teams and atmospheric chemistry community.
- Contributes to international assessments: WMO/UNEP, IPCC 5th Assessment Report – Earth System Grid, a new paradigm of "data delivery" for NASA modeling community (partnership of NCCS and LLNL/ PCMDI), over 400 terabytes of NASA contributions.
- Contributes to WWRP and WCRP.





Climate Science Computing

Climate Science is "Data Intensive". . . . .

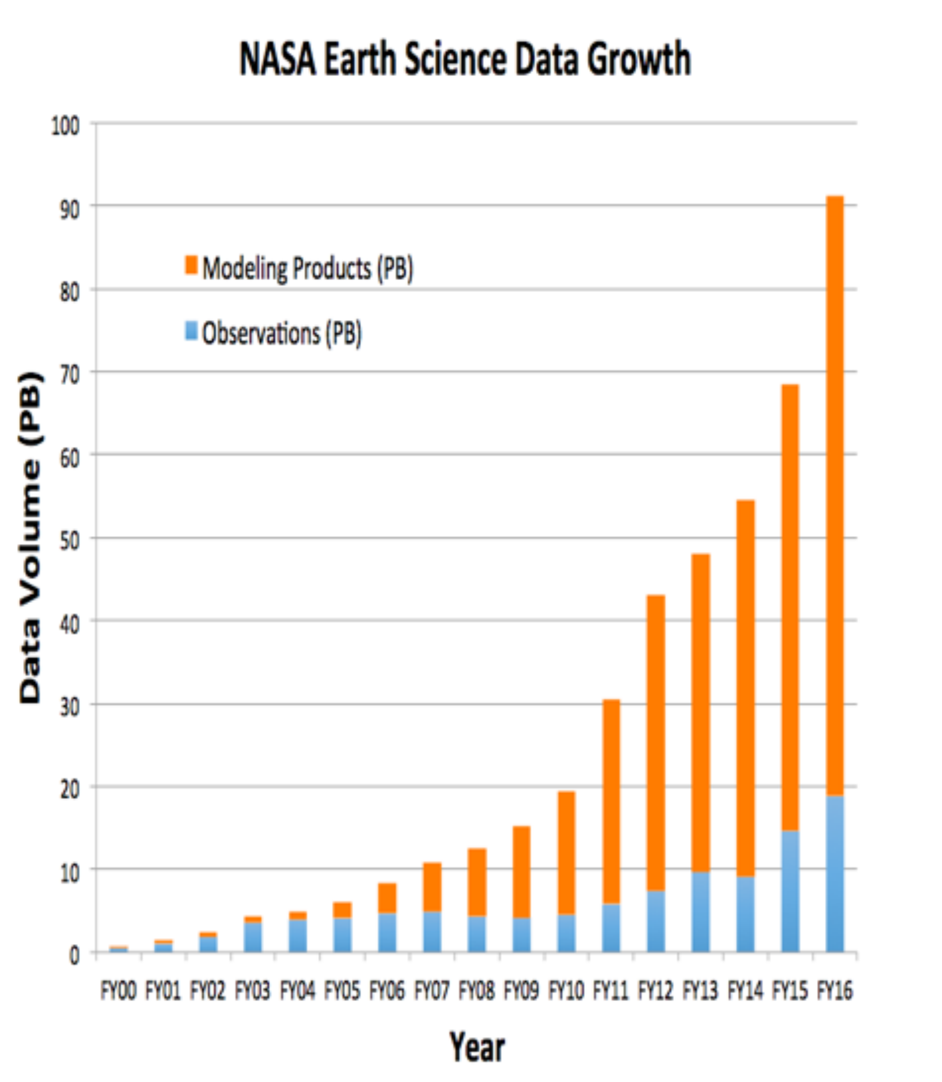
- Our understanding of Earth processes is based in the observational data record and is expressed as mathematical models.
- Data assimilation combines observational data with model prediction.
- Climate models produce large data sets (100s of terabytes) for the scientific community as well as decision makers.
- Reanalysis with improved models results in vast data sets (100s of terabytes) for the scientific community.

. . . . . . .And Requires "Data Analytics-Centric" Computing

- Designed for effective manipulation of large data sets.
- Global file system makes data available to all services. Effective data management tools are required.
- Access to data beyond the computing center must be transparent to the user.
- Efficient data analysis needs to have "supercomputing" capability with data sets online.
- Data sets must be made easily accessible to "external users" with analysis and visualization capability.

Current NASA Earth Science Data Holding

- Observations include all products archived at EOSDIS (http://earthdata.nasa.gov/).
- Modeling products include all the high resolution climate modeling and data assimilation products at NASA Center for Climate Simulations (http://www.nccs.nasa.gov/) and NASA High-End Computing Capabilities (http://www.nas.nasa.gov/hecc/).
- Significant growth in modeling data is triggered by the availability of high resolution Earth observations and the computational resources.



Climate Simulation Data Communities

NASA Scientific Community
- Simulation data consumers
- Advance scientific knowledge
- Direct access to systems
- *Supercomputer* capability required for effective analysis

NASA Modeling Community
- Model development, testing, validation, and execution
- Data creation
- Largest HPC usage
- Requires observational data as input

Simulation Data

External Applications Community
- Huge opportunity for impact from climate change data
- Simulation data consumers
- Limited Earth science data expertise
- Web-based access to systems

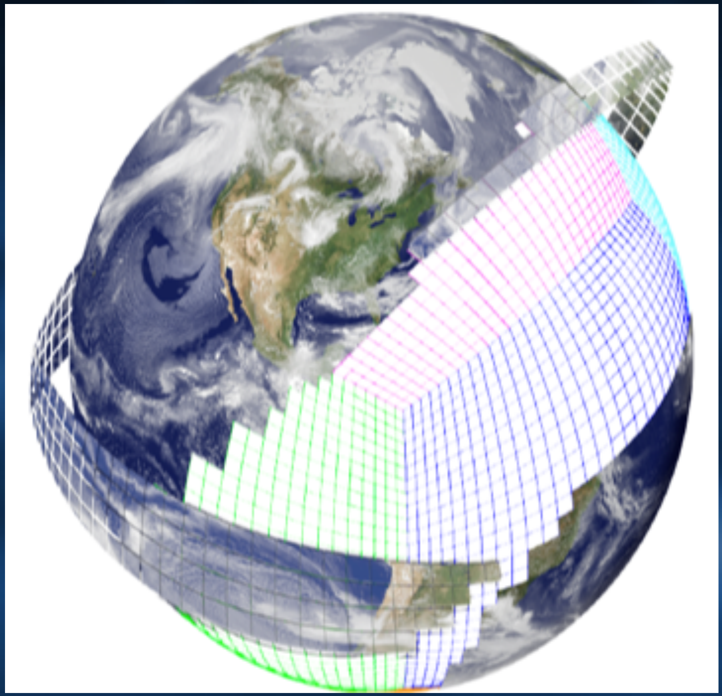
External Scientific Community
- Simulation data consumers
- Advance scientific knowledge
- Web-based access to data

Each community has different capabilities and data usage requirements.

Goddard Earth Observing System (GEOS) Model
NASA Global Modeling and Assimilation Office (GMAO)

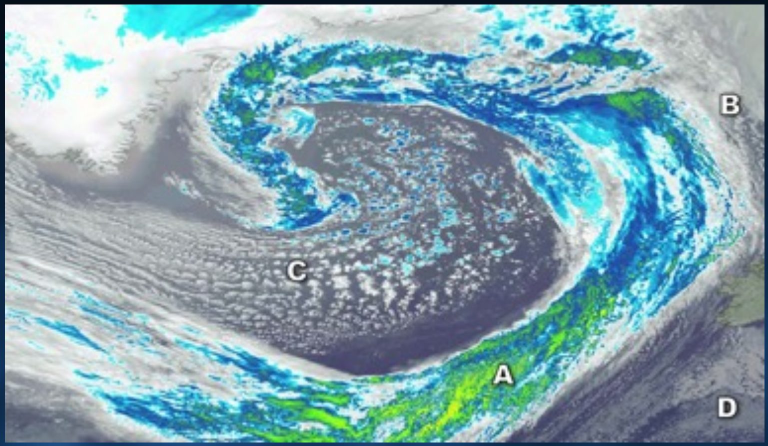
•FV3 Dynamical Core uses a Cubed-Sphere which maps the Earth onto faces of a cube
- There are 6 faces of the cube and multiple vertical layers
- Total number of grid points = X * Y * Z * 6 Faces of the Cube

•Current GMAO Research
- Operational research forecasts are running at 12.5 KM resolution
- MERRA-2 Reanalysis (including chemistry)
- Nature Runs at 7 KM and 3.5 KM
- Dynamic downscaling of reanalysis and forecasts down to 6 KM
- Highest-resolution research runs are at 1.5 KM global resolution



GMAO Simulates Icelandic Low at 50-, 12-, and 1.5-km Resolutions (1)

- Low-pressure systems departing the east coast of the United States intensify in the North Atlantic near Iceland, forming what are known as Icelandic Lows. The Global Modeling and Assimilation Office (GMAO) simulated 5 days of an April 2017 Icelandic Low at three resolutions on the NCCS Discover supercomputer. The simulations show that grid cell size has a significant influence on the cloud types represented.
- At 50 km (216 cores, 200 days/day)—MERRA-2 resolution—deep clouds associated with the frontal system and large-scale structure of the low are well resolved, while GEOS model parameterizations adequately represent low clouds ahead of and behind the front.
- At 12 km (5,400 cores, 80 days/day)—modern global weather prediction resolution—smaller clusters of convection appear within the deep frontal clouds, and low clouds begin to show their broken nature rather than appear simply as broad regions of thick low clouds.
- At 1.5 km (30,000 cores, 3.5 days/day)—leading-edge, cloud-resolving global model resolution—GEOS explicitly resolves detailed cloud structures, including convective clouds within the front and individual stratocumulus cloud clusters behind the front.
- More information: https://gmao.gsfc.nasa.gov/research/science_snapshots/2017/IcelandicLow.php



A snapshot from a 1.5-km simulation of an April 2017 Icelandic Low showing the different clouds types represented. Deep cumulus clouds develop at point A along the frontal boundary, feeding into the low-pressure system. Thin high cirrus clouds are blown off ahead of the low at point B. Shallow stratocumulus clouds form around point C in streets and gyres. Thick marine stratus clouds appear at point D in the warm air near the surface ahead of the system. Research and visualization by William Putman, GMAO.

Increasing the GEOS-5 Model Resolution for Research

The following table shows the requirements to simulate 1 year of the Earth's atmosphere.

Year	Resolution (m)	X	Y	Z	Grid Points	Cores	PB
2017	1,736	5,760	5,760	72	14,332,723,200	30,000	60.00
2019	868	11,520	11,520	144	114,661,785,600	240,000	480.00
2021	434	23,040	23,040	144	458,647,142,400	960,000	1,920.00
2023	217	46,080	46,080	144	1,834,588,569,600	3,840,000	7,680.00
2025	109	92,160	92,160	288	14,676,708,556,800	30,720,000	61,440.00

What we know…

- The data sets needed for science are getting too big to move outside of the computing center or to simply serve through traditional data services.

What we need to be considering…

- May not be able to store everything within each simulation; not all fields are stored
- Need the ability to run simulations to generate the data needed for specific science questions
- This is just the atmosphere! Fully coupled models will require even more capabilities!

Analytics as a Service

Moving Toward a Data and Analytics Centric Environment

Model Data Sets

- Forecasts
- Seasonal Forecasts
- Reanalyses
- Nature Runs
- Weather and Climate Data

Observation Data Sets

- Science Driven

Exposure

Convenient and Extensible

Capabilities need to be easy to use and facilitate community engagement and adaptive construction …

Interfaces

- System prompt
- APIs
- Web Services
- Python Notebooks
- Zeppelin Notebooks

Data Analytics Service

- 1,000's of cores
- TF's of compute
- PB's of storage
- High Speed Networks
- ***Operational Spring 2017***

Data

Relevance and Collocation

Data have to be significant, sufficiently complex, and physically or logically co-located to be interesting and useful …

High-Performance

Compute/Storage Fabric

Storage-proximal analytics with simple canonical operations

Data do not move, analyses need horsepower, and leverage requires something akin to an analytical assembly language …

MERRA Analytic Services: Meeting the Big Data challenges of climate science through cloud-enabled Climate Analytics-as-a-Service, Schnase, et al., *Computers, Environment, and Urban Systems* (doi:10.1016/j.compenvurbsys.2013.12.003)

Evolution of Major NCCS Systems

FY15

Data Portal

Mass Storage | **HPC - Discover**

FY16

ADAPT

Mass Storage | **HPC - Discover**

FY17

ADAPT

Mass Storage | **DASS** | **HPC - Discover**

In addition to increasing compute and storage, the NCCS is evolving its services toward ***Analytics-as-a-Service (in-situ, large-scale data analytics).***

FY16: Creation of a High-Performance Science cloud (virtual environment) designed for traditional data services, data analytics, and web services: move the data to the analysis.

FY17: Creation of the Data Analytics Storage Service (DASS), a combined High-Performance Computing and Data environment to enable emerging analytics: move the analytics to the data.

NCCS Data Analytics Centric Environment
Major NCCS Services

Data Portal/Data Sharing
- Capability to share data & results
- Supports community-based development
- Data distribution and publishing

ADAPT
High-Performance Science Cloud
- Large-scale data analytics
- Loosely coupled applications
- Low barrier to entry
- Agile virtual environment
- Compute and storage

Analysis and Visualization
- Interactive analysis environment
- Software tools for image display
- Easy access to data archive
- Specialized visualization support

User Services
- Help desk
- Account/allocation support
- Computational science support
- User teleconferences
- Training & tutorials

DASS
Storage with Embedded Analytics
- ~15 PB of stored data
- Tightly coupled compute and storage
- Traditional and emerging data analytics stacks

Code Development
- Code repository for collaboration
- Environment for code development and test
- Code porting and optimization support
- Web-based tools

Dirac
Mass Storage Environment
- ~65 PB of stored data
- Front end disk (~5 PB)
- Backend tape
- SGI DMF System

Zoned security architecture, common configuration management, High-Speed Local Area Networks (10 and 40 Gbps)

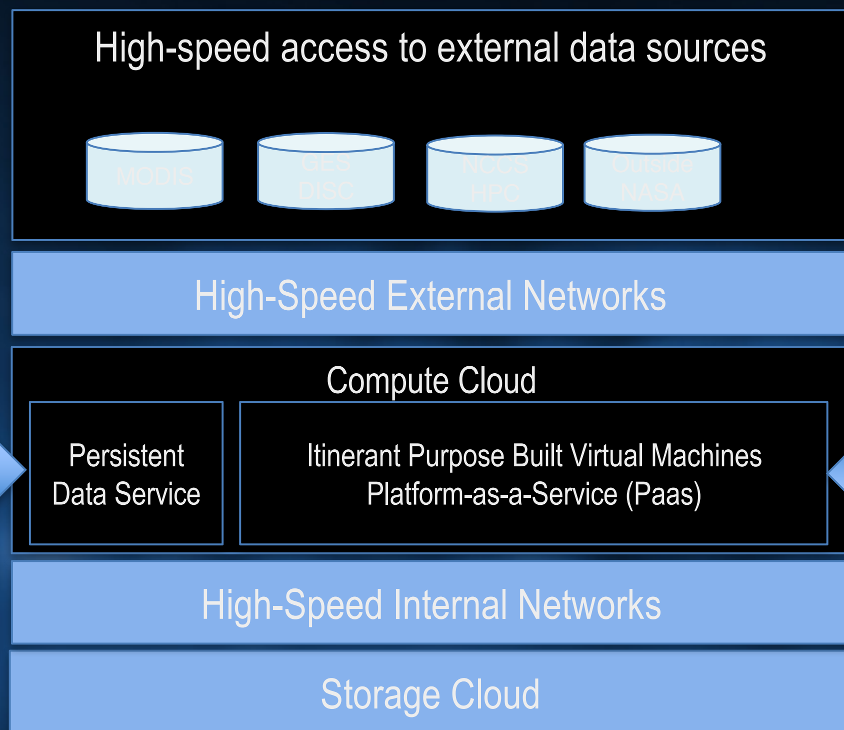
Discover
HPC Cluster
- ~3.5 PB of peak computing
- >90K cores
- >50 PB of storage
- Shared parallel file system
- High-speed networks
- Tightly coupled applications

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

Data Services

Long lived virtual machines specifically designed for data or web services. Examples include:

- NCCS Data Portal
- ESRI ArcGIS
- Web Portals
- Web Map Service
- OpenDAP
- GDS
- THREDDS
- HTTP
- FTP
- Earth System Grid Federation (ESGF)
- Database
- Visualization Servers



- ~10 PB of storage
- “Data Lake” concept – storage is available as needed to all virtual environments
- Low-cost, commodity-based storage
- Multiple petabytes in size and easily expandable
- High-performance file system using IBM GPFS

Compute Services

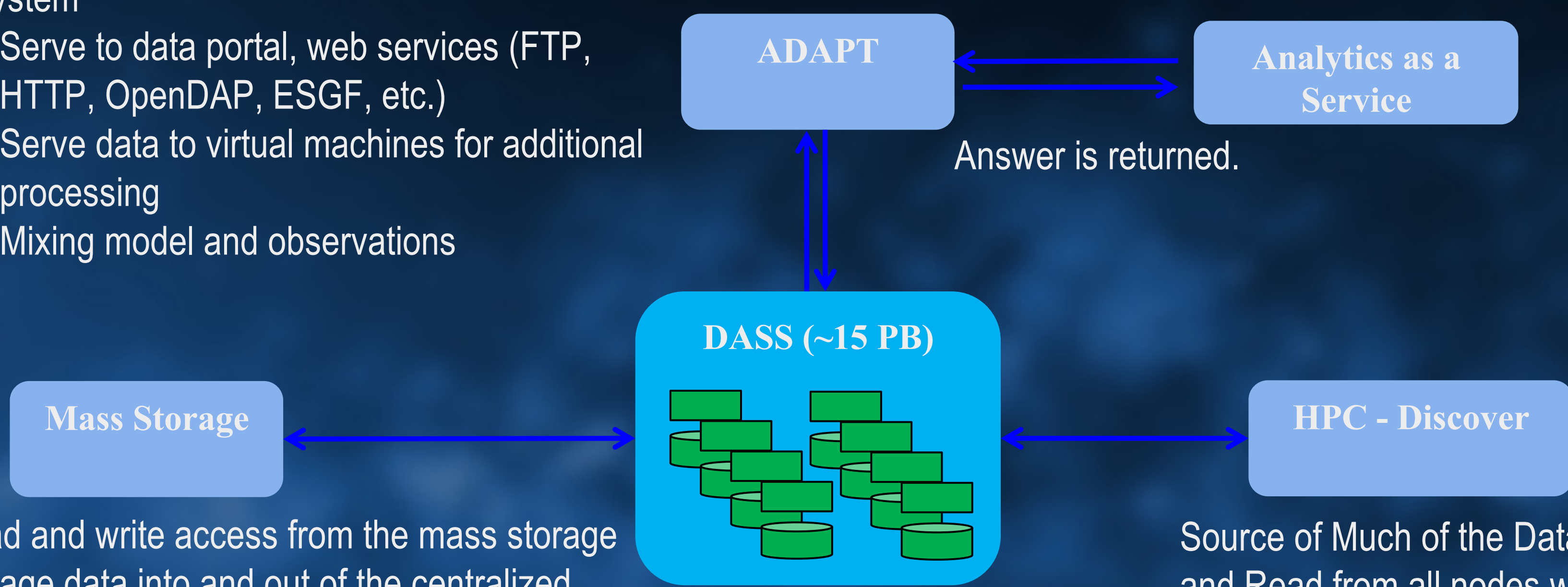
Itinerant purpose built virtual machines are customized for each user/project.

- Compute systems are older, ***repurposed high performance compute nodes***
- ~340 nodes, ~5,000 cores
- Vary in memory sizes from 32 to 64 to 256 GB

Data Analytics Storage System (DASS) Concept

Read access from all nodes within the ADAPT system
- Serve to data portal, web services (FTP, HTTP, OpenDAP, ESGF, etc.)
- Serve data to virtual machines for additional processing
- Mixing model and observations

Analysis request is sent to a service.



Read and write access from the mass storage
- Stage data into and out of the centralized storage environment as needed

Source of Much of the Data: Write and Read from all nodes within Discover – models write data into GPFS which is then staged into the centralized storage (burst buffer like).

Note that all the services will still have local file systems to enable local optimized writes and reads as needed within their respective security domains.

Future of HPC and Big Data at Exascale

Analytics Intensive

ADAPT
Virtual Environment
HPC and Cloud
Existing Size
~1,000 cores
~10 PB of storage

Designed for Big Data Analytics

Future Exascale Environment
Merging of HPC and Big Data Analytics Capabilities

Ability for in-situ analytics throughout the environment … known analytics and machine learning

DASS
Tiered Storage
Memory, SSD, Disk
Existing Size
~15 PB of storage

Designed for both compute and long term storage.

Mass Storage
Tiered Storage
Disk and Tape
Existing Size
~10 to 100 PB of storage

Designed for long-term storage and recall; not compute

HPC/Discover
HPC Cluster
Existing Size
~100,000 cores
~ 50 PB of storage

Designed for Large-Scale Climate Simulations

Computational Intensive

More Information

NASA High-End Computing Program
https//www.hec.nasa.gov

High-End Computing Capability Project
https://www.nas.nasa.gov/hecc

NASA Center for Climate Simulation
https://www.nccs.nasa.gov

Global Modeling and Assimilation Office
https://gmao.gsfc.nasa.gov

Tsengdar Lee
Tsengdar.j.lee@nasa.gov