High Performance Computing (HPC) and Climate Model Data

Ad Hoc Big Data Task Force of the
NASA Advisory Council Science Committee
June 28-30, 2016

Daniel Duffy (daniel.q.duffy@nasa.gov and on Twitter @dqduffy)
High Performance Computing Lead
NASA Center for Climate Simulation
http://www.nccs.nasa.gov

NASA Center for Climate Simulation (NCCS)
NASA Goddard Space Flight Center

Topics

- Introduction to the NCCS
- Climate Models
- Growth of computing
- Growth of storage
- Toward exascale
- Analysis is different than HPC
- Need a different infrastructure
- ADAPT
- DASS
- Final Thoughts

NASA High-End Computing Program

HEC Program Office
NASA Headquarters
Dr. Tsengdar Lee
Scientific Computing Portfolio Manager

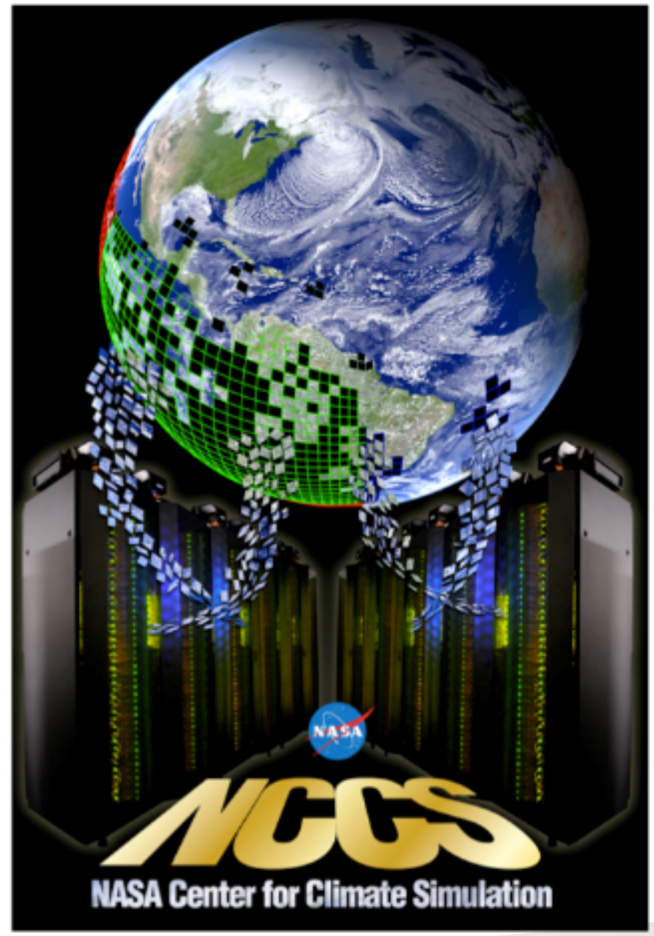
NASA Center for Climate Simulation (NCCS)
Goddard Space Flight Center (GSFC)
Dr. Daniel Duffy

High-End Computing Capability (HECC) Project
NASA Advanced Supercomputing (NAS)
NASA Ames
Dr. Piyush Mehrotra

NASA Center for Climate Simulation (NCCS) High Performance Science

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

- High-performance computing, data storage, and networking technologies
- High-speed access to petabytes of Earth Science data
- Collaborative data sharing and publication services
- Advanced Data Analytics Platform (ADAPT)
- Data Analytics Storage System (DASS)



Primary Customers (NASA Climate Science)

- Global Modeling and Assimilation Office (GMAO)
- Goddard Institute for Space Studies (GISS)

High-Performance Science

- http://www.nccs.nasa.gov
- Code 606.2
- Located in Building 28 at Goddard

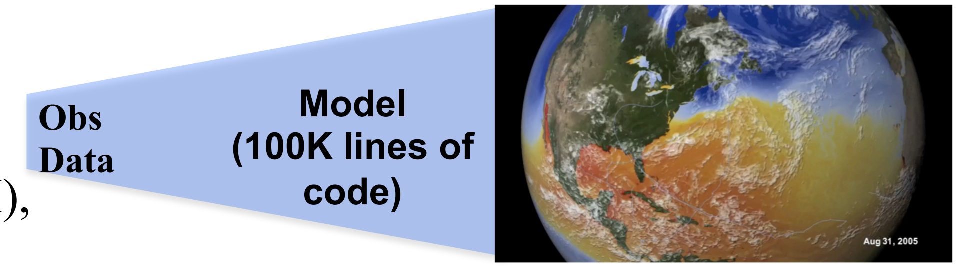
Example HPC Applications

Takes in small input and creates large output

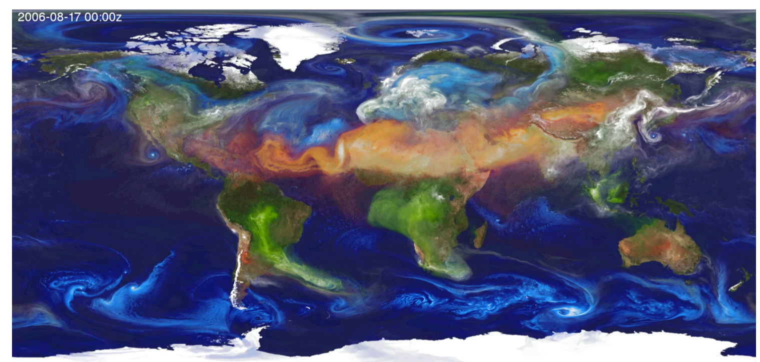
- Using relatively small amount of observation data, models are run to generate forecasts
- Fortran, Message Passing Interface (MPI), large shared parallel file systems
- Rigid environment – users adhere to the HPC systems

Examples

- GMAO GEOS-5 High-Resolution Nature Runs for Observing System Simulation Experiments (OSSE)
- Evaluation of dynamical downscaling and comparison of regional versus global models
- Extremely high-resolution global circulation models



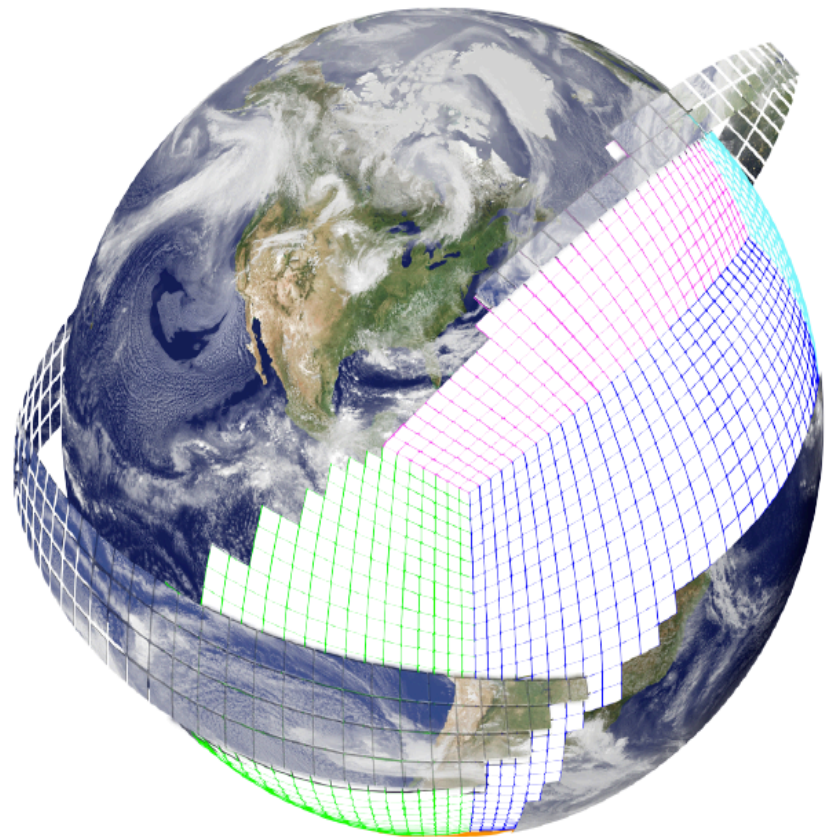
10-km GEOS-5 meso-scale simulation for Observing System Simulation Experiments(OSSEs)



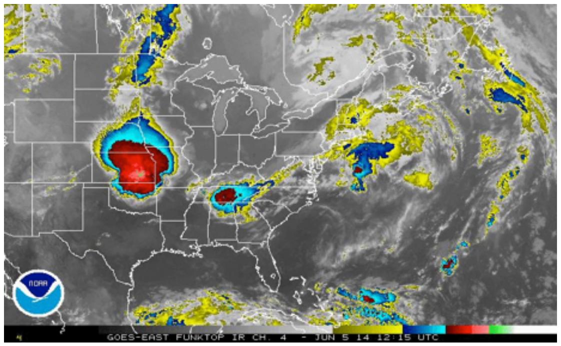
The Goddard Chemistry Aerosol Radiation and Transport (GOCART) model, Courtesy of Dr. Bill Putman, Global Modeling and Assimilation Office (GMAO), NASA Goddard Space Flight Center.

Goddard Earth Observing System (GEOS) Model

- 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 operational forecast is running at 27 KM resolution using about 27 million grid points
- Target operational resolution of 13 KM by the end of 2016
- Highest resolution research runs are at 1.5 KM global resolution

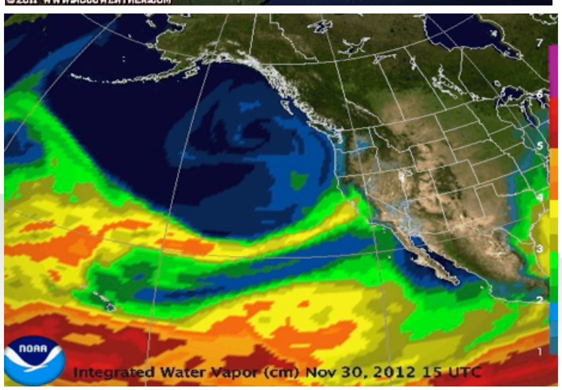


Dynamic Downscaling Assessment
Narrow Scope – Focus only on 3 Impactful Phenomena



Northeast Wintertime Storms (NESs)
- Extreme precipitation/snowfall events
- Extreme wind events

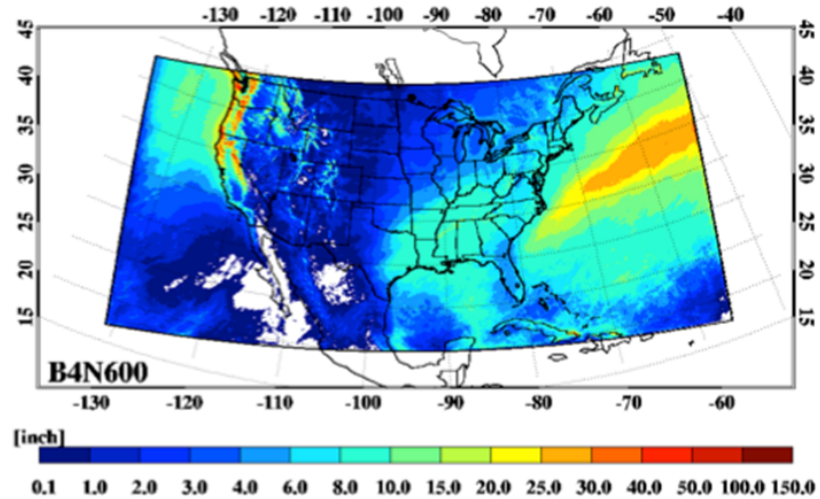
Midcontinent Summertime MCSs
- Warm / Dry Climate Model Biases
- Extreme weather events



West Coast Wintertime Atmospheric Rivers (ARs)
- Crucial for water resources/availability
- Associated with most flooding events

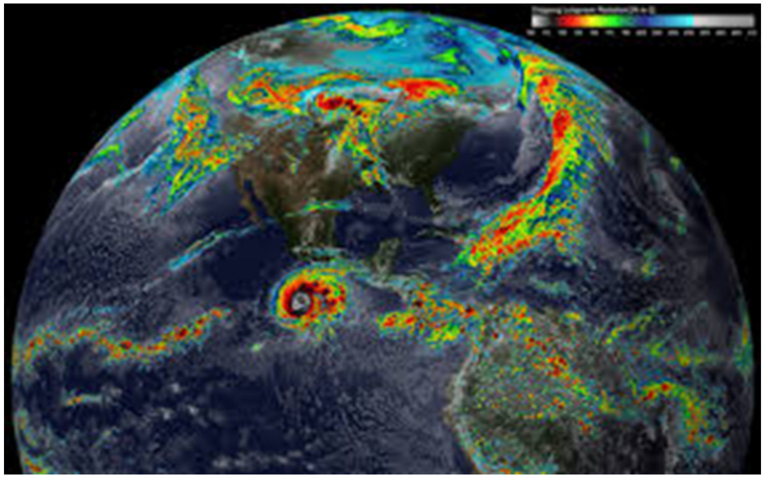
NASA Downscaling Models

- Regional Climate Model
  - NASA Unified-WRF (NU-WRF) Based on WRF-ARW v3.5.1
- Initial/Boundary Conditions
  - MERRA-2 six-hourly re-analyses over CONUS
- Land Initial Conditions
  - Land Information System (LIS) 10-yr spin-up of Noah LSM
- Nudging (large scale forcing of certain variables to the synoptic scale)
  - Simulations with and without spectral nudging of p, t, and horizontal winds above the PBL



NASA Downscaling Models (Continued)

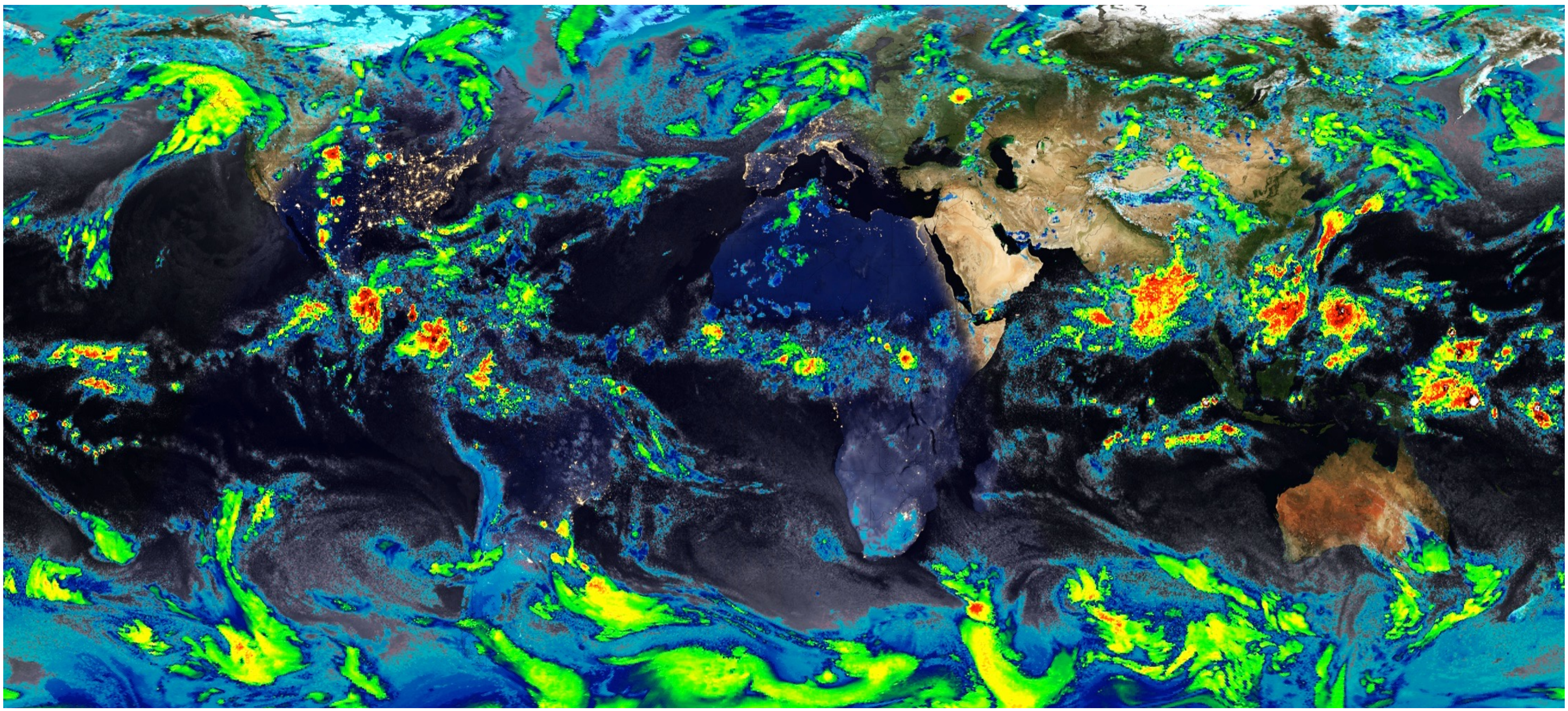
- MERRA-2 Replay at 12km – M2R12K
  - Goddard Earth Observing System Model, Version 5 (GEOS-5) in replay mode at 12km resolution (global simulation)
- Boundary Conditions
  - MERRA-2 six-hourly re-analyses
- Land Initial Conditions
  - MERRA-2 CLSM
- Nudging
  - Replay capability adds a forcing term to constrain the 12km run to MERRA-2 p, t, winds, and humidity (q)



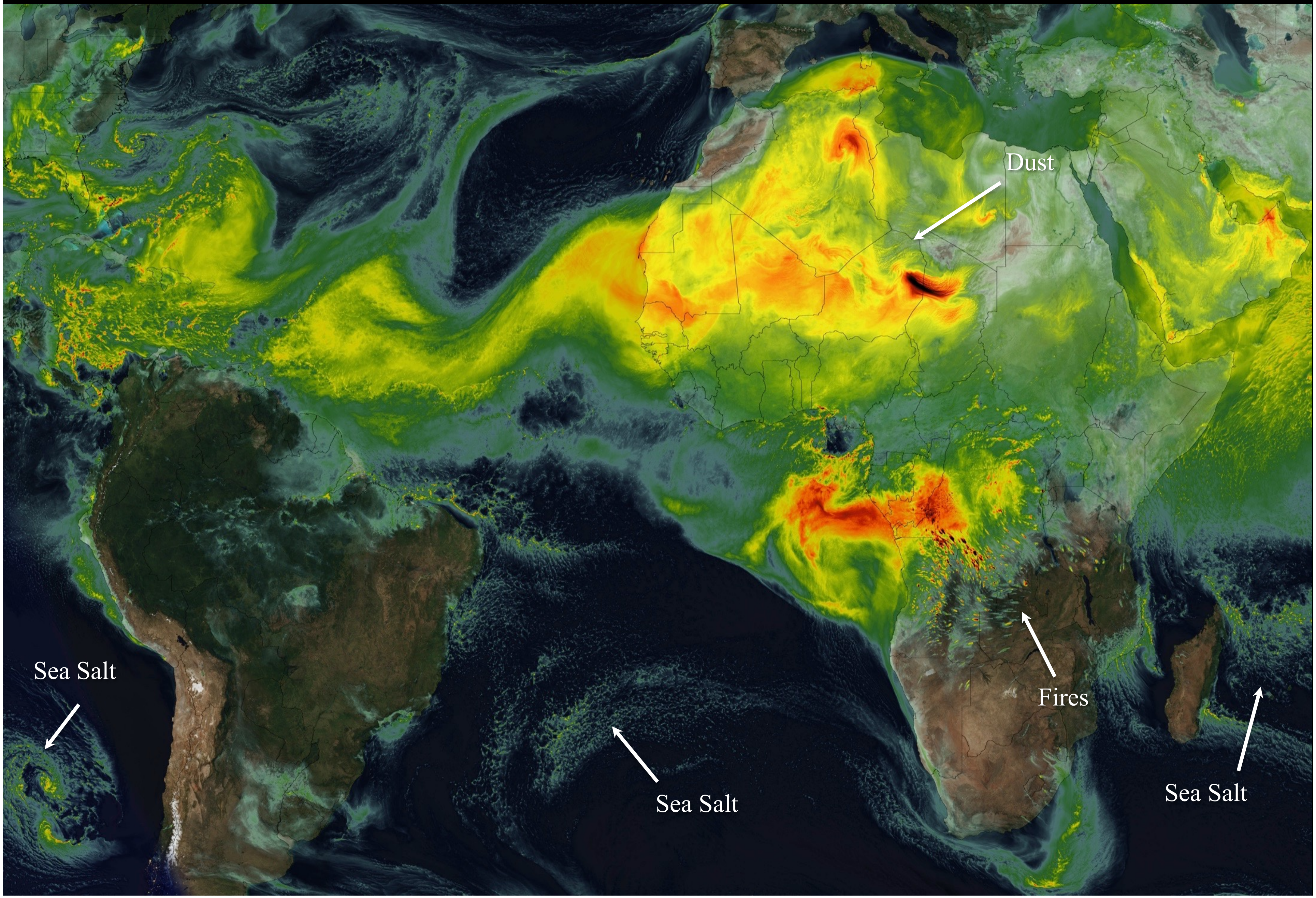
- Replay allows the model to develop its own internally developed mesoscale while following the large scale trajectory of the underlying MERRA-2 reanalysis (~50km)

1.5km Global Simulation with GEOS-5 at NCCS

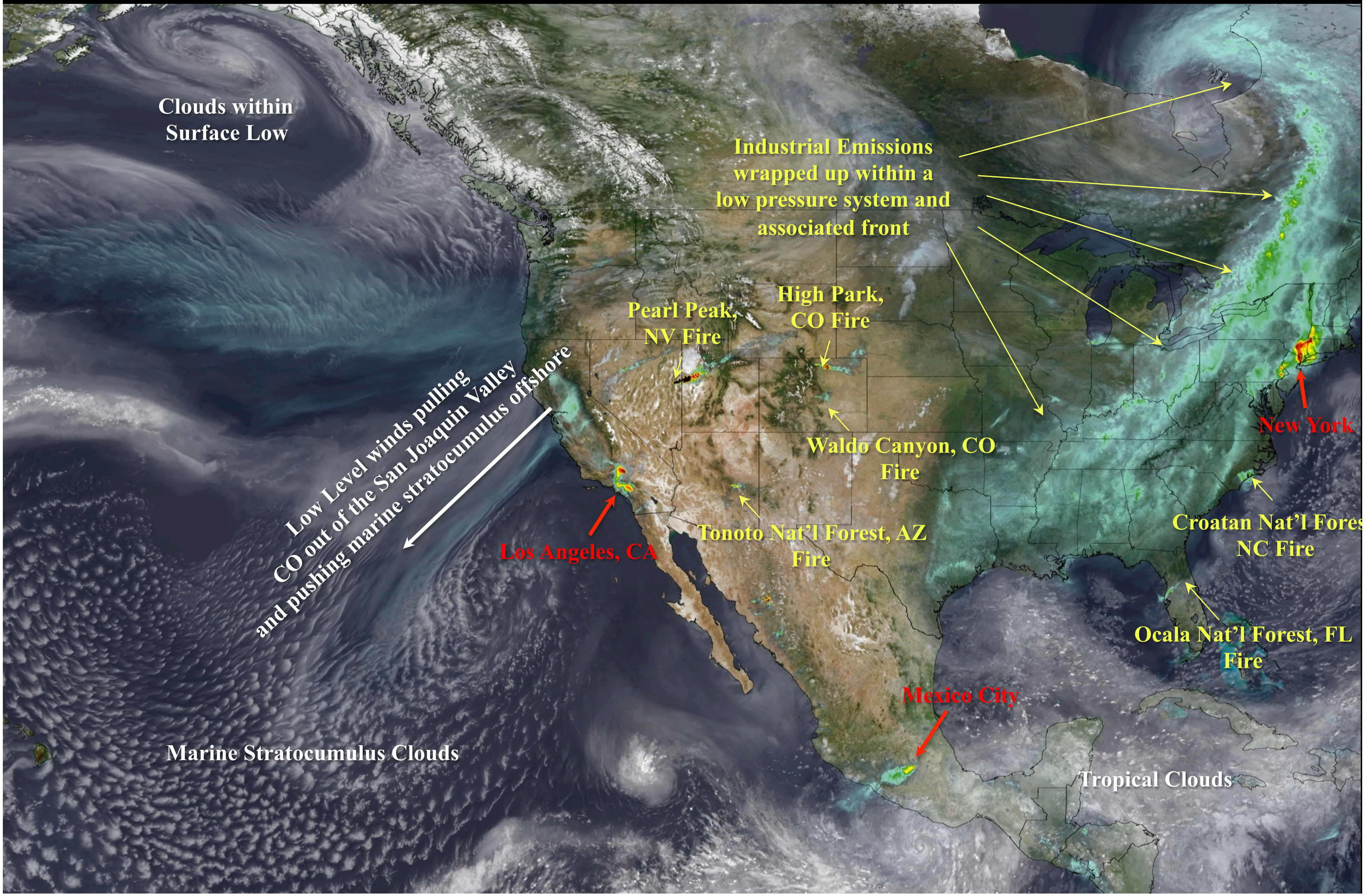
- The highest resolution simulation performed with any US global model
  - Includes updated 2-moment micro-physics and interactive aerosols/chemistry with GOCART
  - Executed on 1,055 nodes of the NCCS 'Discover' SCU10 cluster (1,024 nodes for compute and 31nodes for IO)
  - 1.5km GEOS-5 completed ~1 simulated-day/wallclock-day
  - 0.87KM NICAM global simulation was run on the K-Computer (10.5 Petaflop system) with a throughput of ~0.5 simulated-days/wallclock-day



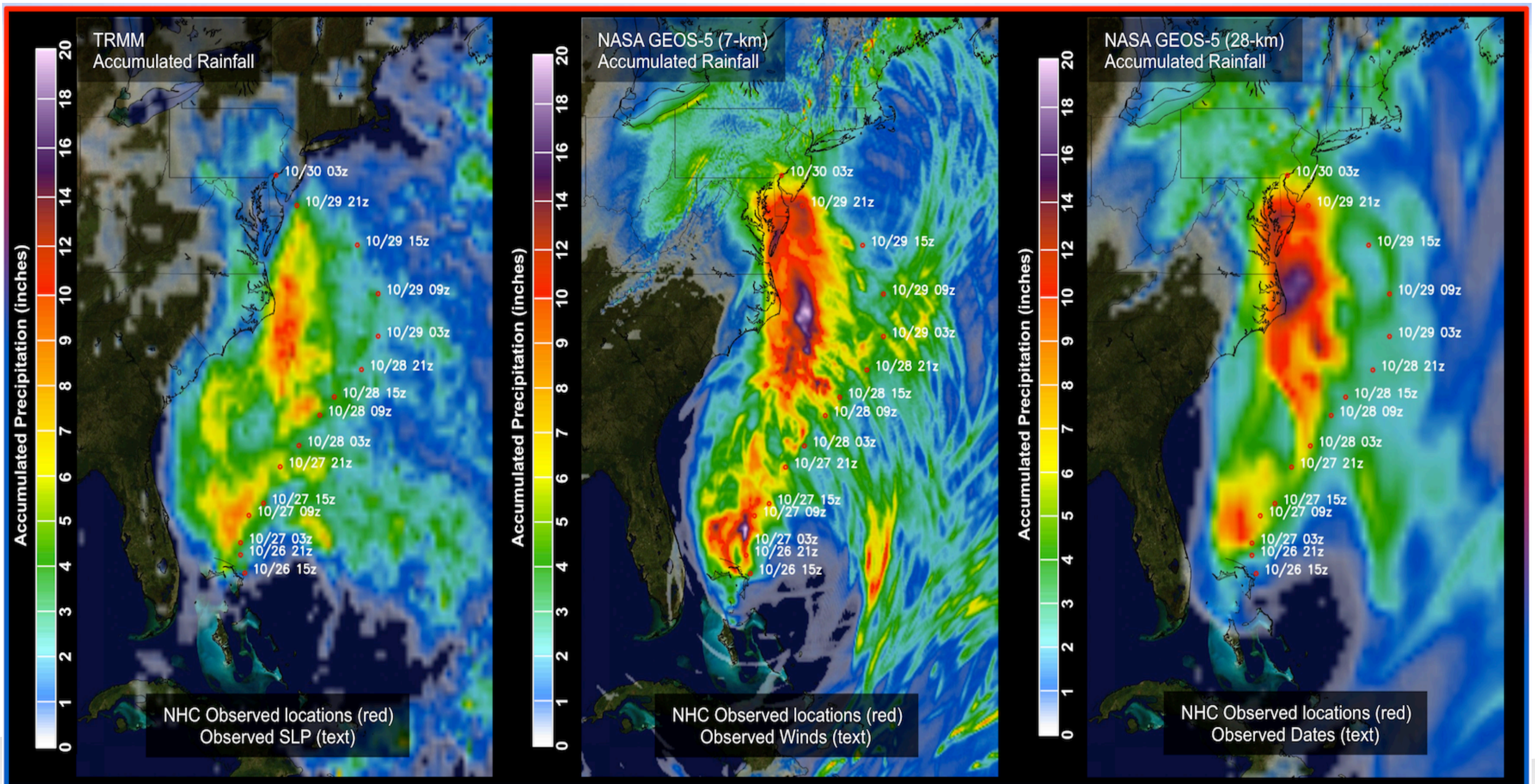
Total Aerosol Optical Thickness from GEOS-5 at 1.5-km, June 15, 2012 12:00z
Global Modeling and Assimilation Office



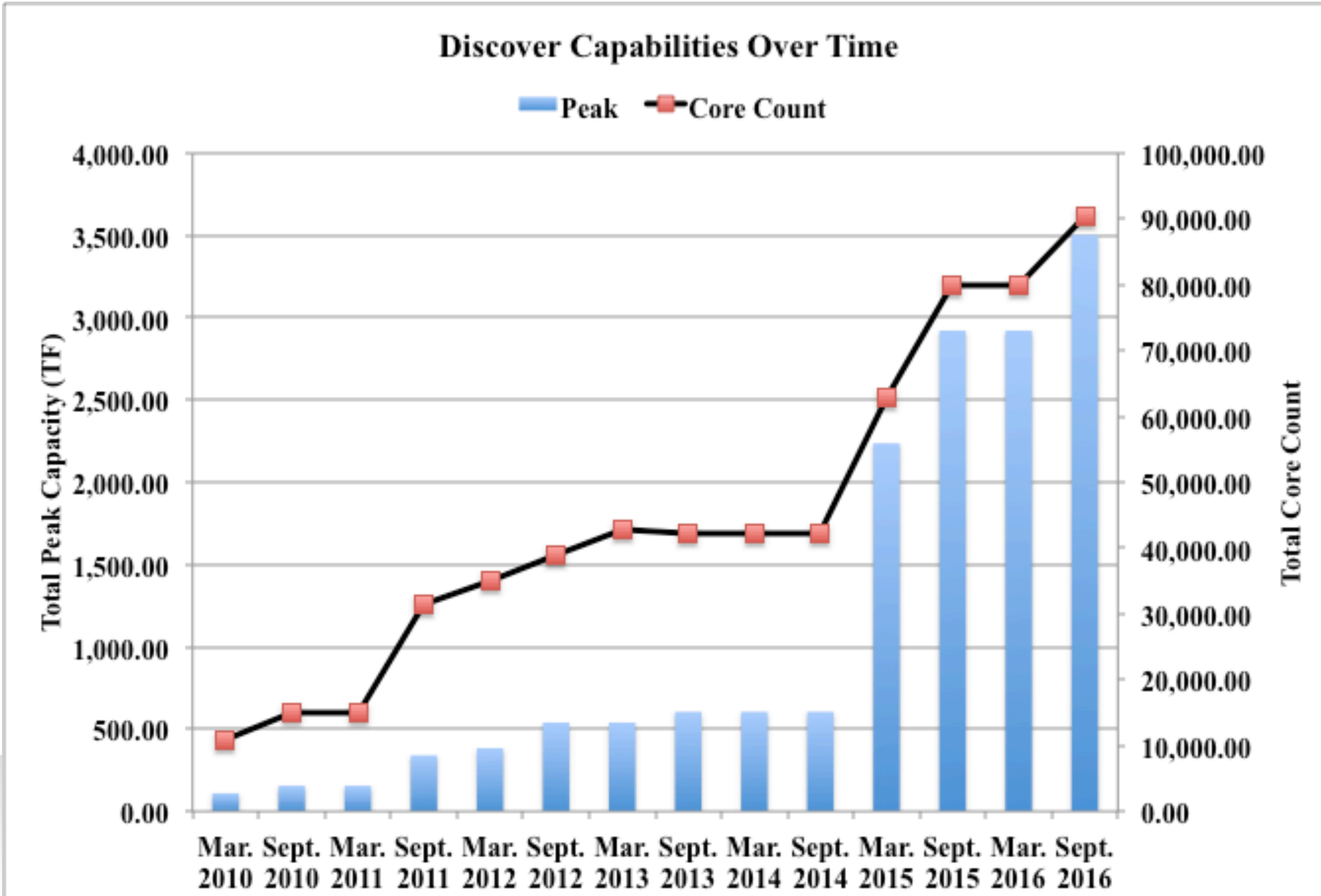
Low clouds and column carbon monoxide from GEOS-5 at 1.5-km, June 17, 2012 23:00z, from wildfires throughout the western United States and from industrial emissions in the East (Global Modeling and Assimilation Office)



Hurricane Sandy



NCCS Computing Growth Over Time



HPC Storage Capacity Growth

Calendar	Description	Decommission	Total Usable Capacity (TB)
2012	Combination of DDN disks	None	3,960
Fall 2012	NetApp1: 1,800 by 3 TB Disk Drives; 5,400 TB RAW	None	9,360
Fall 2013	NetApp2: 1,800 by 4 TB Disk Drives; 7,200 TB RAW	None	16,560
Early 2015	DDN10: 1,680 by 6 TB Disk Drives, 10,080 TB RAW	DDNs 3, 4, 5	~26,000
Mid 2015	DDN11: 1,680 by 6 TB Disk Drives, 10,080 TB RAW	DDNs 7, 8, 9	~33,000
Mid 2016	DDN11: 1,680 by 6 TB Disk Drives, 10,080 TB RAW	None	~40,000 (40 PB)

Requirements Continue to Grow: GMAO Science Portfolio

Forward Processing System	Satellite-Era Reanalysis 1979 - Present	EOS-Era Reanalysis 2000 – Present	Nature Runs (OSSEs)	Seasonal Forecast System	Coupled Simulations (Decadal, CMIP6)
3D-Hybrid Ensemble-Var (25km) 32 ensemble members Hydrostatic 1-Moment Cloud Microphysics *Current GEOS-5 FP system*	**MERRA (50km)** **Ending Feb. 2016** 3D-Var ~200 TB **MERRA-2 (50km)** 3D-Var Aerosols and CO, SO_2, O_3 1-Moment Cloud Microphysics ~400 TB	**M2R12K (12km)** MERRA2 downscaled to 12 km Aerosols CO_2, CO, SO_2, O_3 Non-Hydrostatic 1-Moment Cloud Microphysics	**G5NR (7km)** Simulated 2005-2007 Aerosols, CO_2, CO, SO_2, O_3 Non-Hydrostatic 1-Moment Cloud Microphysics 4 PB	**GEOS SFS (50km)** MERRA-2 replay 50km, 40L ocean analysis 31 members per month Include aerosols, CO, CO_2 *M2-driven EnOI ocean analysis*	**GEOS CMIP (25km)** 25km Atmosphere 25km 50L ocean Include aerosols greenhouse gases Hydrostatic 2-Moment Cloud Microphysics *Planning/discussion and system evaluation in progress* *Will align with "MERRA-3" SFS and strategic direction of ESD*
3D-Hybrid Ensemble-Var (12km) 32 ensemble members Atmosphere, ocean surface Hydrostatic 2-Moment Cloud Microphysics *Parallel FP stream in 1Q-2016*	**MERRA-2 GMI replay (50km)** Replay GMI Chemistry 1 streams, 1,000 cores each 12 to 18 months ~ 1 PB	**IESA (12km)** 3D-Hybrid Ensemble-Var 32 ensemble members atmosphere, land, ocean surface Aerosols, CO_2, CO, SO_2, O_3 Non-Hydrostatic 2-Moment Cloud Microphysics 5,000 cores ; 40 simulation days/ day 150 days total wallclock ~3 to 4 PB of data	**G5NR-CHEM (12km)** Simulated 2013-2014 Replay to M2R12K Full Reactive Chemistry Non-Hydrostatic 1-Moment Cloud Microphysics 1 PB of data *4Q-FY2016*	**GEOS SFS (25km)** Alignment with "MERRA-3" 25km, 50L ocean analysis System design under review *FY2019 target*	
4D Ensemble-Var (9km) ~100 ensemble members Atmosphere, ocean surface Non-Hydrostatic 2-Moment Cloud Microphysics (The first GEOS-6 system) *Parallel FP stream in 4Q-2016*	**Coupled Reanalysis ("MERRA-3")** Atmosphere-land-ocean-cryosphere (alignment with SFS and CMIP6) *FY2019 target*	**IESAR4K (4km)** IESA Downscaled to 4km downscallng evaluation for NCA Aerosols, CO_2, CO, SO_2, O_3 Non-Hydrostatic 2-Moment Cloud Microphysics 5,000 cores ; 40 simulation days/ day 150 days total wallclock ~3 to 4 PB of data	**G6NR (3km)** Simulated 2015 Aerosols CO_2, CO, SO_2, O_3, CH_4 Non-Hydrostatic 2-Moment Cloud Microphysics ~4 PB *Planning/evaluation*		

Core GMAO projects completed, in-progress or planned.

Pathfinding projects toward GMAO core efforts.

Proposed project

SMD Reserve projects.

Increasing the GEOS-5 Model Resolution

- Target – Run approximately 100 meter global resolution research runs in 10 to 15 years
- Each doubling of resolution requires 8x the grid points; hence, 8x the memory.

X Values	Grid Points	Resolution (meters)	RAM (PB)
5,760	14×10^9	1,736	0.1
11,520	112×10^9	868	0.8
23,040	896×10^9	434	6.4
46,080	$7,168 \times 10^9$	217	51.2
92,160	$57,344 \times 10^9$	109	4,096

Bad News – This is only one component of the application (the atmosphere). To truly simulate the climate, we need a coupled model including Atmosphere, Ocean, Waves, Ice, and More; We expect the model to required much more memory pushing us toward a higher memory to flop ratio.

Compare HPC to Data Analysis

High Performance Computing

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

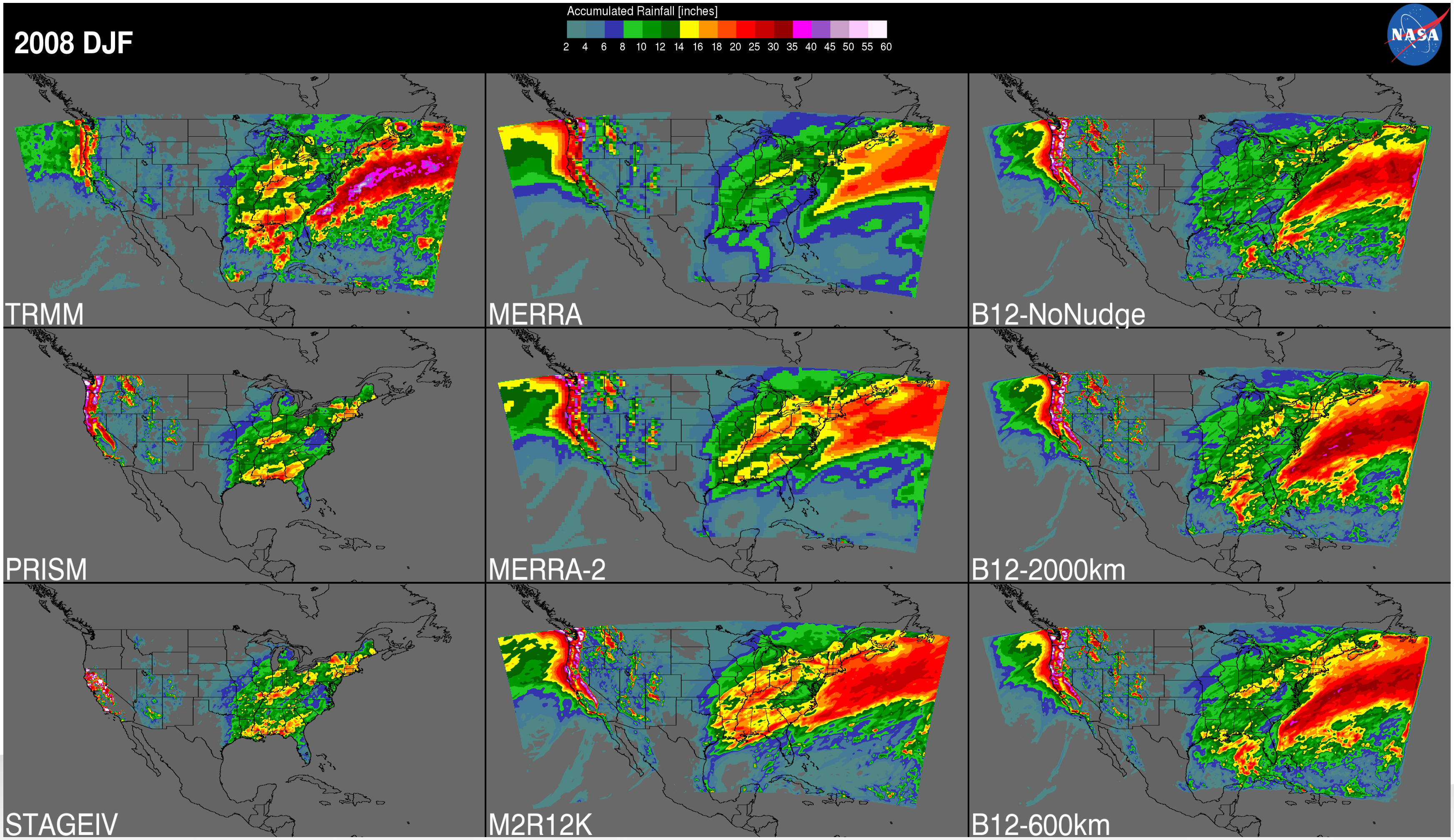
Data Analysis

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

Data Analysis is inherently different than High Performance Computing applications.

Example of Downscaling Data Analysis

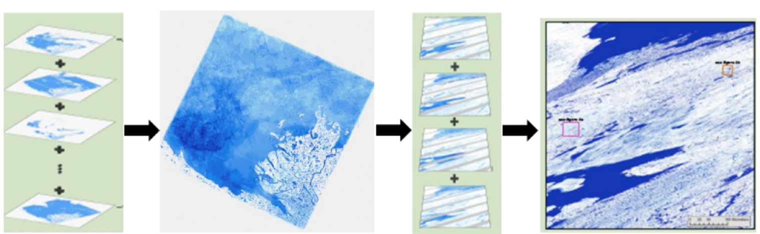


PacificNW Spatial Correlation	MERRA	MERRA-2	M2R12K	B12_600km	B12_2000km	B12_NoNudge	TRMM
	0.17	0.36	0.63	0.84	0.83	0.82	0.61

Example Data Analysis Using LandSat

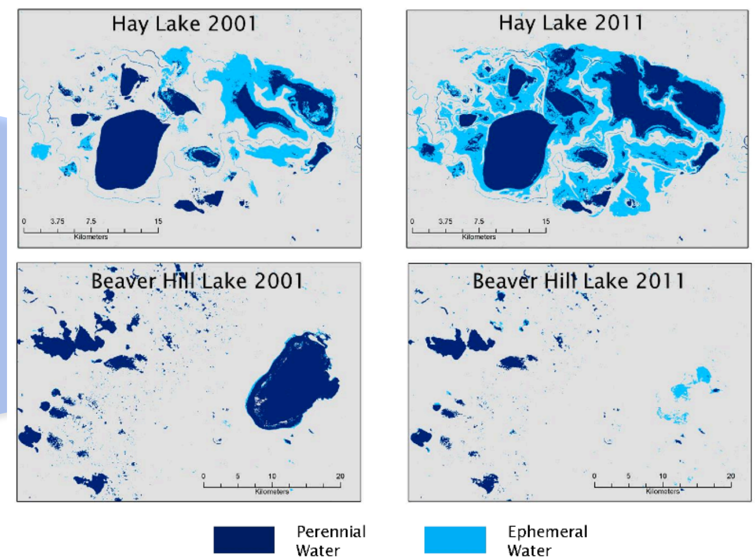
Takes in large amounts of input and creates small output

- Using large amounts of observation or model data
- Python code of 100's of lines
- Easily run in parallel across multiple virtual machines



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.

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

Reducing Data Friction Using MODIS Data

Goal: Build a cloud climatology record over the entire MODIS data record (TERRA & Aqua) for circumpolar Arctic 60 degrees north and above

Data Input requirements: Daily MODIS data from two sensors, 2003 to present. (36 tiles and approximate 56 TB)

Data Download and Processing WITHOUT Science Cloud

- Order and download MODIS data by tile to local workstation
  - Managing download against available storage
- Process data, QA, store outputs, delete inputs, and repeat

Total Time Estimate
~ 9 months

Data Download and Processing with Science Cloud

- Order and download MODIS data to science cloud
- Process data, QA, store outputs
  - Processing over multiple virtual machines (benchmark 16 VMs / more available)

Total Time Estimate
~ 3 to 4 weeks

Staged Data and Processing with Science Cloud

- Access staged MODIS data in Science Cloud
- Process data, QA, store outputs
  - Processing over multiple virtual machines (benchmark 16 VMs) more available

Total Time Estimate
~ 1-4 days

Different Infrastructure Needed for Analytics

- Big data analytics needs a different infrastructure than HPC
  - Analytics has traditionally used HPC resources, but it is beginning to limit its effectiveness and is causing issues within HPC
- An infrastructure that enables
  - Both the traditional use of data while enabling future analytics
  - Storage proximal analytics (dramatic reduction in data movement)
  - Builds towards emerging data analysis paradigms, such as deep learning
- NCCS is evolving its services using commercial Big Data technologies
  - Such as Virtualization, Hadoop and MapReduce, Object Store
- Major Challenge
  - These technologies were built for unstructured text data and do not easily integrate with structure scientific, binary data

Advanced Data Analytics Platform (ADAPT) Platform-as-a-Service (PaaS) Architecture

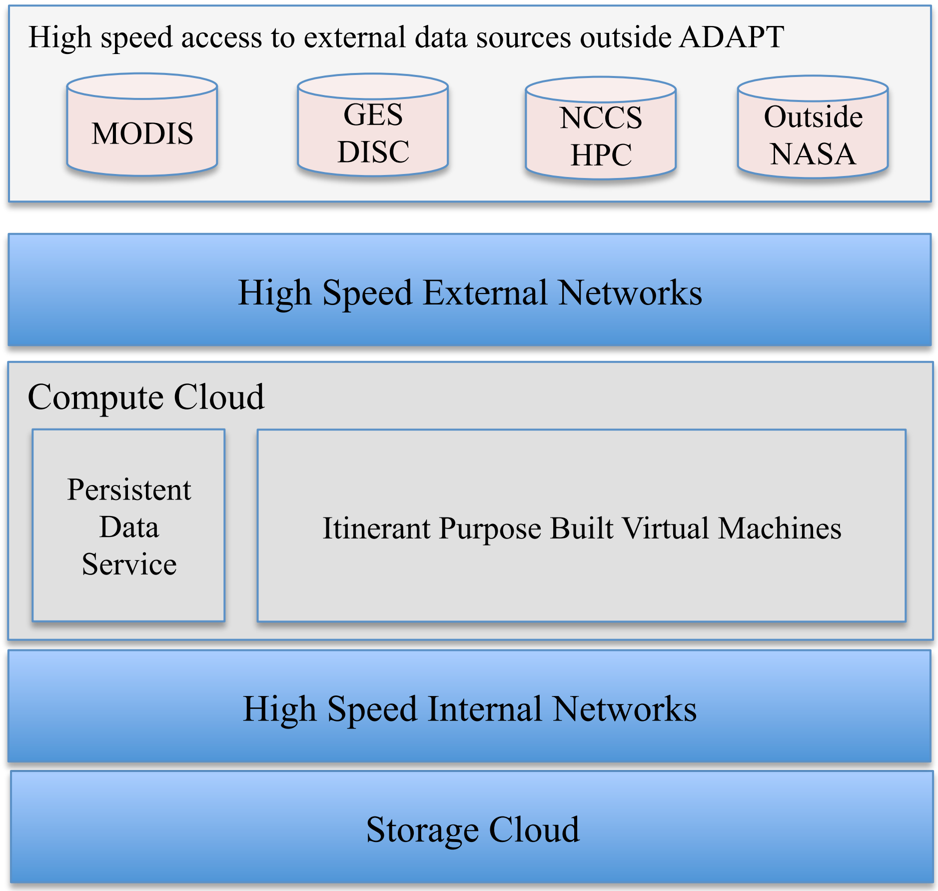
Compute systems are older, ***repurposed high performance compute nodes***

- 100's of nodes currently with plans to expand over the next 6 months
- Capable of 1,000s of virtual machines

Persistent Data Services are long lived virtual machines specifically designed for data or web services. Examples include:

- Web Portals
- Web Map Service
- FTP
- OpenDAP
- Earth System Grid Federation (ESGF)
- ESRI ArcGIS

Itinerant purpose built virtual machines are customized for each user/project. These virtual machines are not persistent and can be spun up and down as needed.



- "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

High speed external networks capable of 10 GbE and 40 GbE are available to transfer data into and out of ADAPT. In addition, remote mounts to external data sources, such as MODIS, are being served over these networks.

High speed internal networks use repurposed high performance Infiniband switches along with more traditional Ethernet switching.

What applications and data work best in ADAPT?

- Not designed for Message Passing Interface (MPI)
  - These are highly coupled processes performing large amounts of data movement over high speed networks and synchronization
  - Recommend to use HPC systems for this
- ADAPT is designed more for inherently parallel processing of big data
  - Independent processes written to analyze large data sets
  - ADAPT has tools to assist users in submitting independent parallel scripts across multiple virtual machines
- ADAPT is not an archive
  - Large data sets relevant to multiple research projects can be made available, such as LandSat and NGA data
  - Remote access to commonly used data sets is also available, such as for MODIS
  - Projects must work with the NASA DAACs to provide long term data management for their products
- Publishing of data
  - Persistent data services created to provide a capability to share large data
  - NCCS can support this through a variety of technologies

Data Locally Stored in ADAPT

- User directories: Home and Scratch (/nobackup) directories
- ADAPT has large volumes of relevant data that are locally stored for ABoVE researchers
- Data stored is driven by the science requirements
- NGA data data covers more than just the Arctic Boreal Vulnerability Experiment (ABoVE) study domain with more data coming in every day
- Will store the Digital Elevation Maps (DEMs) created both by the Polar Geospatial Center (PGC) and in the NCCS once those are created

Data Locally Stored in ADAPT	Volume
LandSat	101 TB
MODIS	252 TB
MERRA (Atmospheric Reanalysis)	200 TB
NGA	> 1 PB

Making Remote Data Look Local

- ADAPT also has the ability to remotely mount data to a virtual machine to make it seem like it is local
- Uses the Lightweight Virtual File System (LVFS) developed by the MODIS team
- ***No need to download data; access what you need when you need it directly from OpenDAP and FTP sites as long as the application can tolerate the latency.***

MODIS Data Example of ALL MODIS collection 5 and 6 data

- Local Directory: /att/pubrepo/MODIS/remote

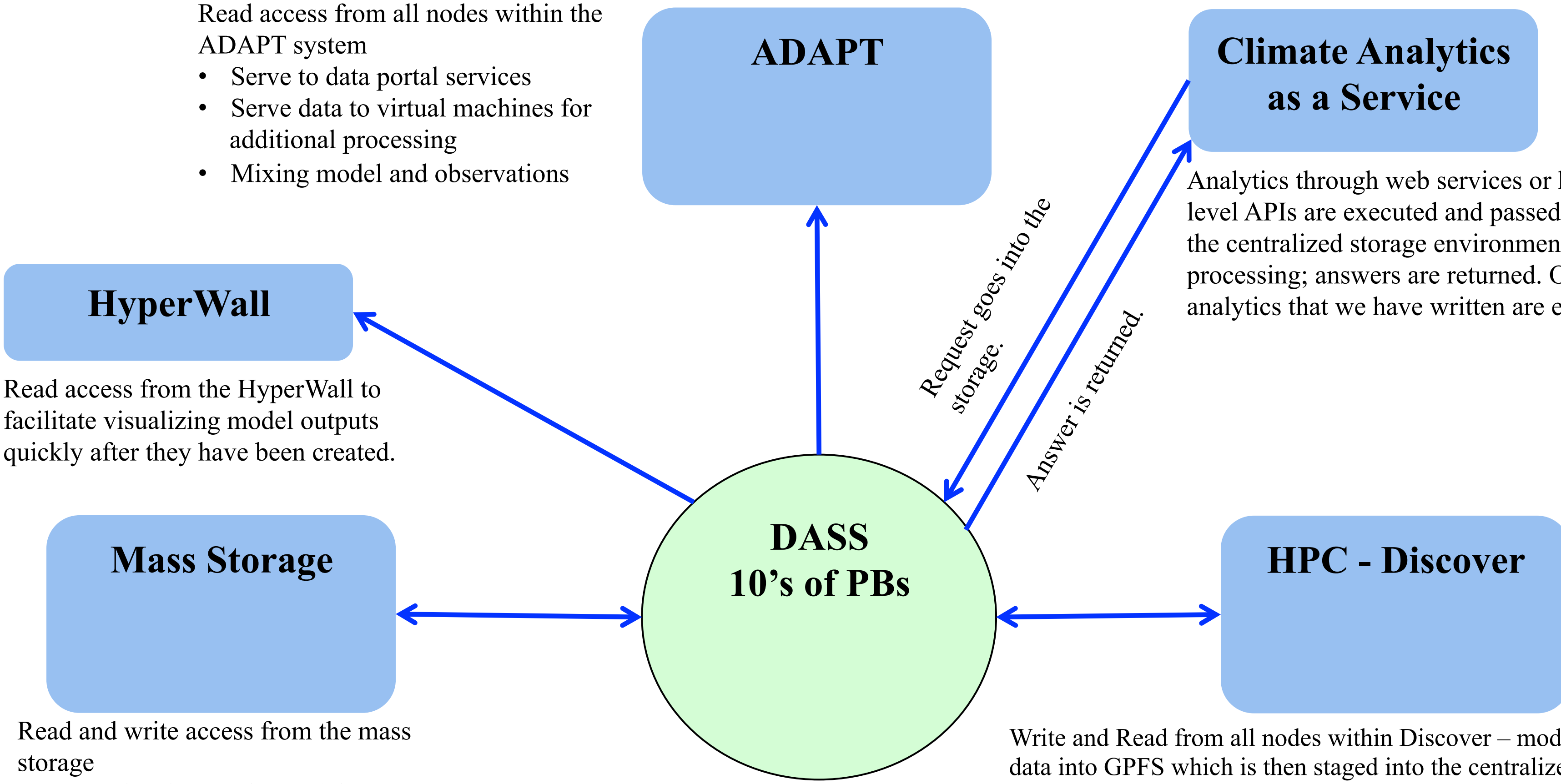
NOAA Data Example of Global Brightness Temperature

- Remote Link: ftp://ftp.cpc.ncep.noaa.gov/precip/global_full_res_IR/
- Local Directory: /att/lvfs/ncep/data/precip/global_full_res_IR

Science Projects Using ADAPT

- Artic Boreal Vulnerability Experiment (ABoVE)
  - Partnership with the Carbon Cycle Ecosystems Office to provide cloud services for ABoVE field campaign
  - Examples include comparing multiple, time displaced images of the same geographical area looking for changes in surface water extent using LANDSAT data; long-term multi-sensor record of fire disturbances in high northern latitudes using LANDSAT and MODIS
- NCCS Web Services
  - The NCCS is using ADAPT to host its own web services
- Asteroid Hunter
  - Evaluating design of next generation space-based telescope using a combination of asteroid projection modeling and simulated telescope
- National Geospatial Agency (NGA) High Resolution Image Processing
  - Examples include converting native NGA NTF to GTIFF (GeoTIFF); computing the vegetation index by counting trees, shrubs; computation of mosaics; digital elevation maps
- CALET Processing
  - Simulation of detectors for high energy particles

Data Analysis Storage System (DASS)



Read access from all nodes within the ADAPT system

- Serve to data portal services
- Serve data to virtual machines for additional processing
- Mixing model and observations

Analytics through web services or higher level APIs are executed and passed down into the centralized storage environment for processing; answers are returned. Only those analytics that we have written are exposed.

Read access from the HyperWall to facilitate visualizing model outputs quickly after they have been created.

Read and write access from the mass storage

- Stage data into and out of the centralized storage environment as needed

Write and Read from all nodes within Discover – models write data into GPFS which is then staged into the centralized storage (burst buffer like). Initial data sets could include:

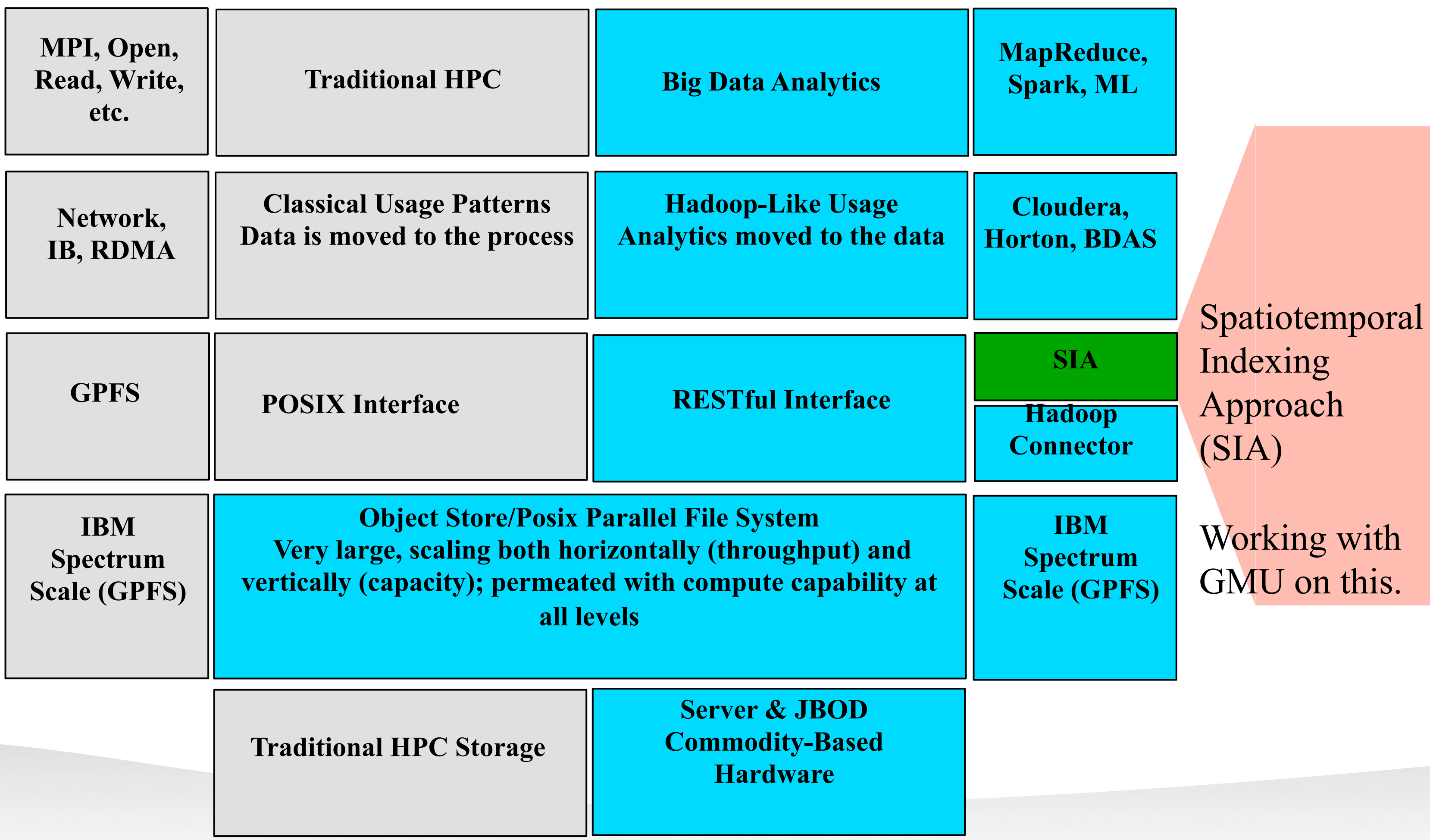
•Nature Run
•Downscaling Results
•Reanalysis (MERRA, MERRA2)
•High Resolution Reanalysis

Note that more than likely all the services will still have local file systems to enable local writes within their respective security domain.

DASS Software Stack

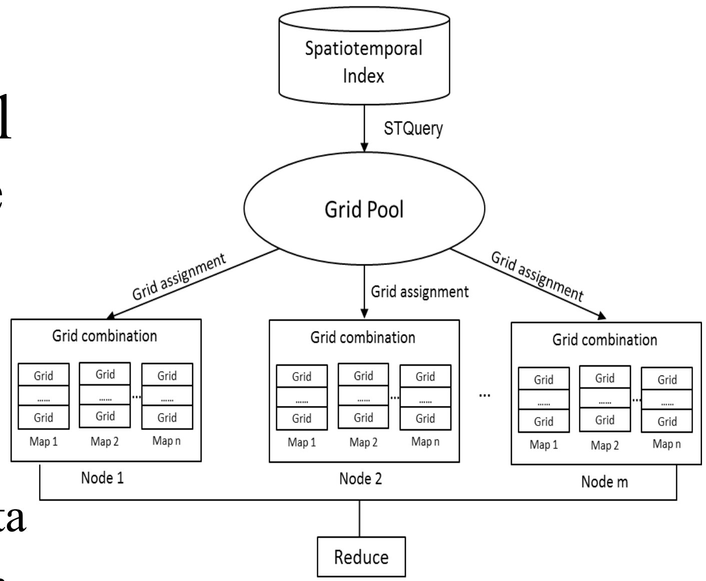
MPI, Open, Read, Write, etc.	Traditional HPC	Big Data Analytics	MapReduce, Spark, ML
Network, IB, RDMA	Classical Usage Patterns Data is moved to the process	Hadoop-Like Usage Analytics moved to the data	Cloudera, Horton, BDAS
GPFS	POSIX Interface	RESTful Interface	Hadoop Connector
IBM Spectrum Scale (GPFS)	Object Store/Posix Parallel File System Very large, scaling both horizontally (throughput) and vertically (capacity); permeated with compute capability at all levels		IBM Spectrum Scale (GPFS)
	Traditional HPC Storage	Server & JBOD Commodity-Based Hardware	

How do you make this work with NASA science data?



Spatiotemporal Index and Hadoop File System

- Use what we know about the structured scientific data
- Create a spatiotemporal query model to connect the array-based data model with the key-value based MapReduce programming model using grid concept
- Built a spatiotemporal index to
  - Link the logical to physical location of the data
  - Make use of an array-based data model within HDFS
  - Developed a grid partition strategy to
  - Keep high data locality for each map task
  - Balance the workload across cluster nodes



Analytics Infrastructure Testbed

Test Clusters using decommissioned HPC servers

Test Cluster 1
SIA
Cloudera
HDFS

- 20 nodes
- Cloudera
- HDFS
- Sequenced data
- Native NetCDF data
  - Put only

Test Cluster 2
SIA
Cloudera
Hadoop Connector
GPFS

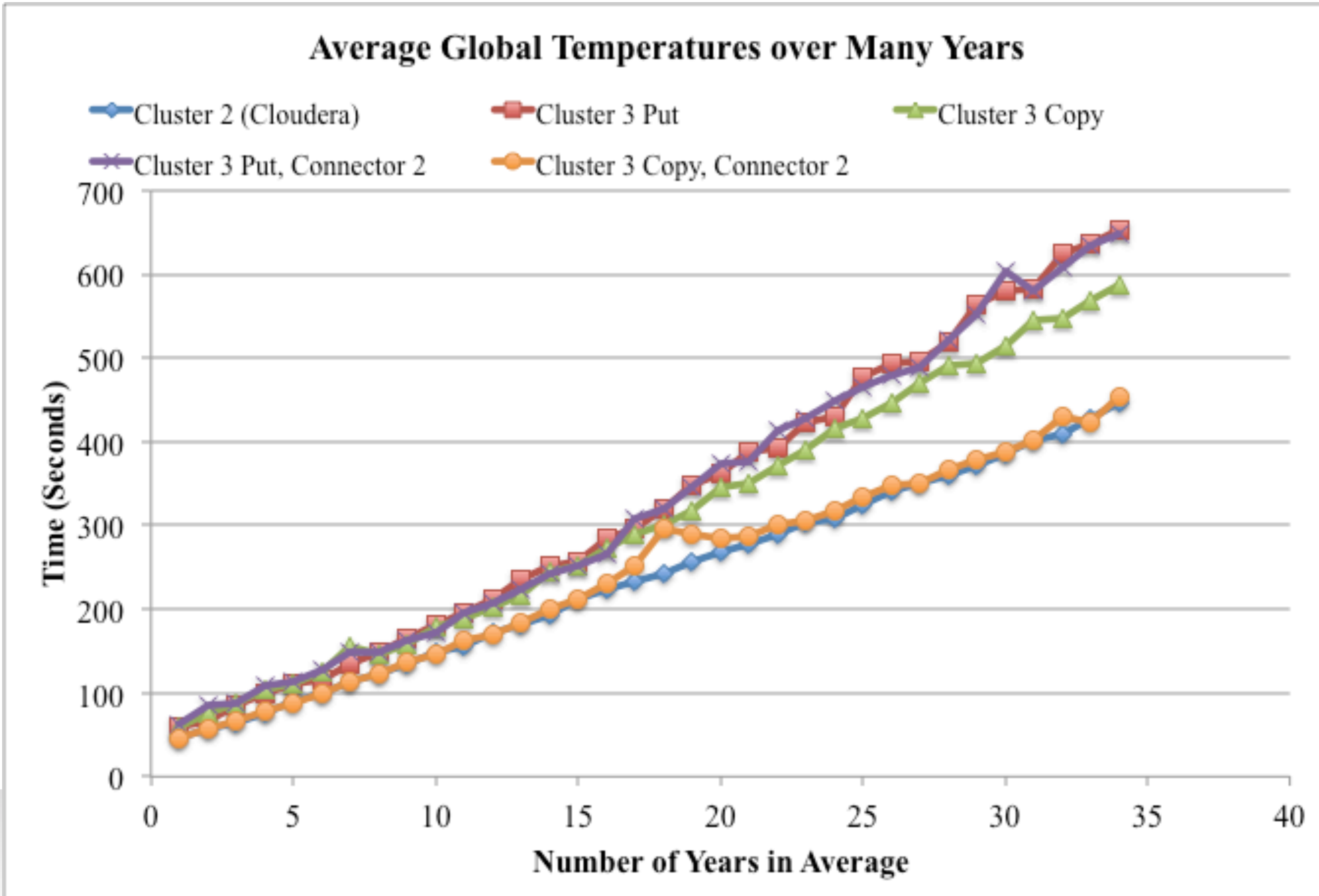
- 20 nodes
- Cloudera
- GPFS
- GPFS Hadoop Connectors
- Sequenced data
  - Put and Copy
- Native NetCDF Data
  - Put and Copy

Test Cluster 3
SIA
Cloudera
Hadoop Connector
Lustre

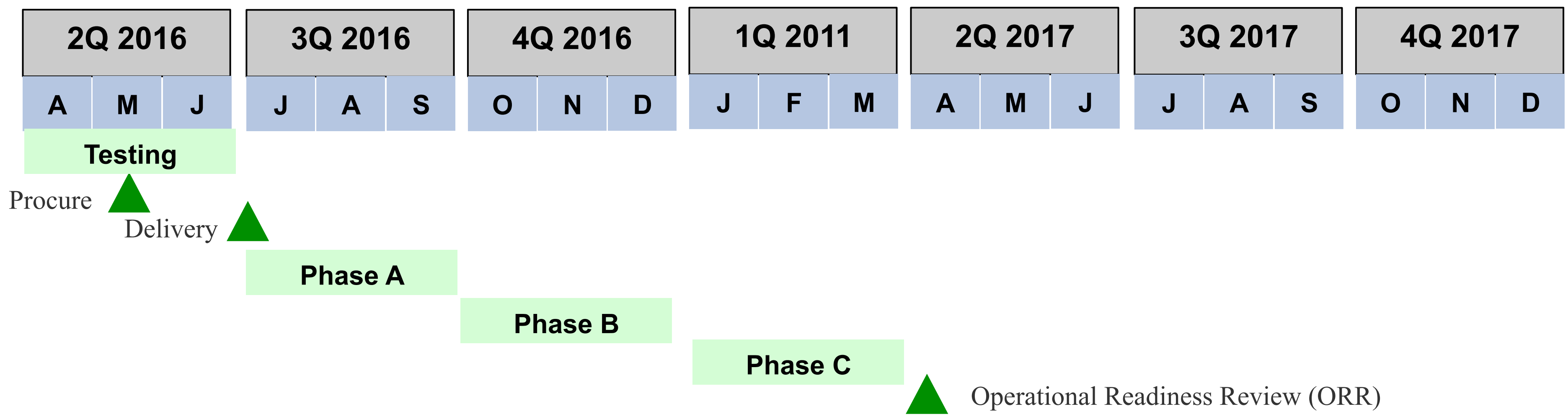
Almost completed

- 20 nodes
- Cloudera
- Lustre
- Lustre HAM and HAL
- Sequenced data
  - Put and Copy
- Native NetCDF Data
  - Put and Copy

Initial Testing Results
(No SIA at this point – sequenced data only)



DASS Timeline



Testing
- Cloudera
- GPFS
- Lustre

Phase A
- Installation and burn-in
- Acceptance
- Testing file system configurations
- Security planning and reviews
- Planning for coupling to other NCCS systems

Phase B
- Copy larger data sets into the DASS (such as MERRA and MERRA2)
- Install higher software layers, such as HDFS and SIA
- Performance and load testing
- Test coupling/mounts of file systems to other NCCS systems (ADAPT, TDS)

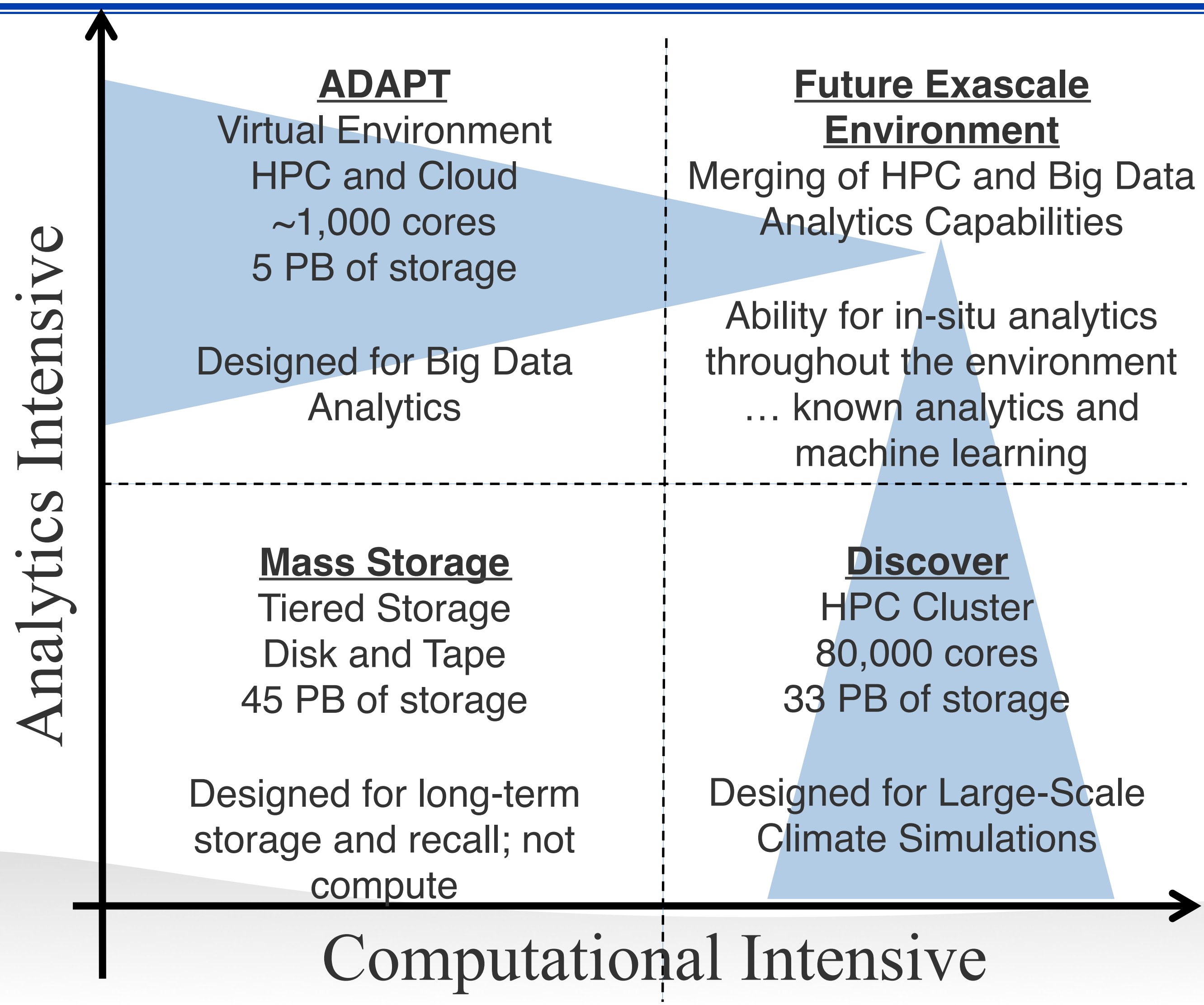
Phase C
- Pioneer users
- Complete coupling to ADAPT and DISCOVER

Compute Where the Data Resides

CPU	On chip high bandwidth memory – think NVIDIA GPUs and Intel Phi architectures.
In Package Memory	High Bandwidth Memory on the chip.
Node Memory	NVME, emerging technologies, etc. Large quantities of persistent storage close to the CPU.
File System in Memory	Technologies to enable shared memory across many nodes as well as collective operations at the network level.
Network	Very fast reads and writes into and out of the network and the HPC environment.
High Performance File System	Large aggregate space and throughput designed for longer term persistent storage.
Tiered Storage Subsystems	Hierarchy of storage systems from SSD's to spinning disks to …
Cloud	Ability to store data in the cloud and burst when appropriate for data analytics.

Challenge: Must be able to perform data analysis at EVERY layer. Be able to efficiently move the data to the appropriate layer for computation or move the thread to the data.

Future of HPC and Big Data at Exascale



Just for fun
The 5 V's of Big Data … and More

- Let's start with the 5 V's of data that everyone knows…
- Volume, Velocity, Veracity, Variety, Value

Others are adding more V's …

- Visualization, Variability, Viability

Here are a few more that we are keeping in mind as we move forward …

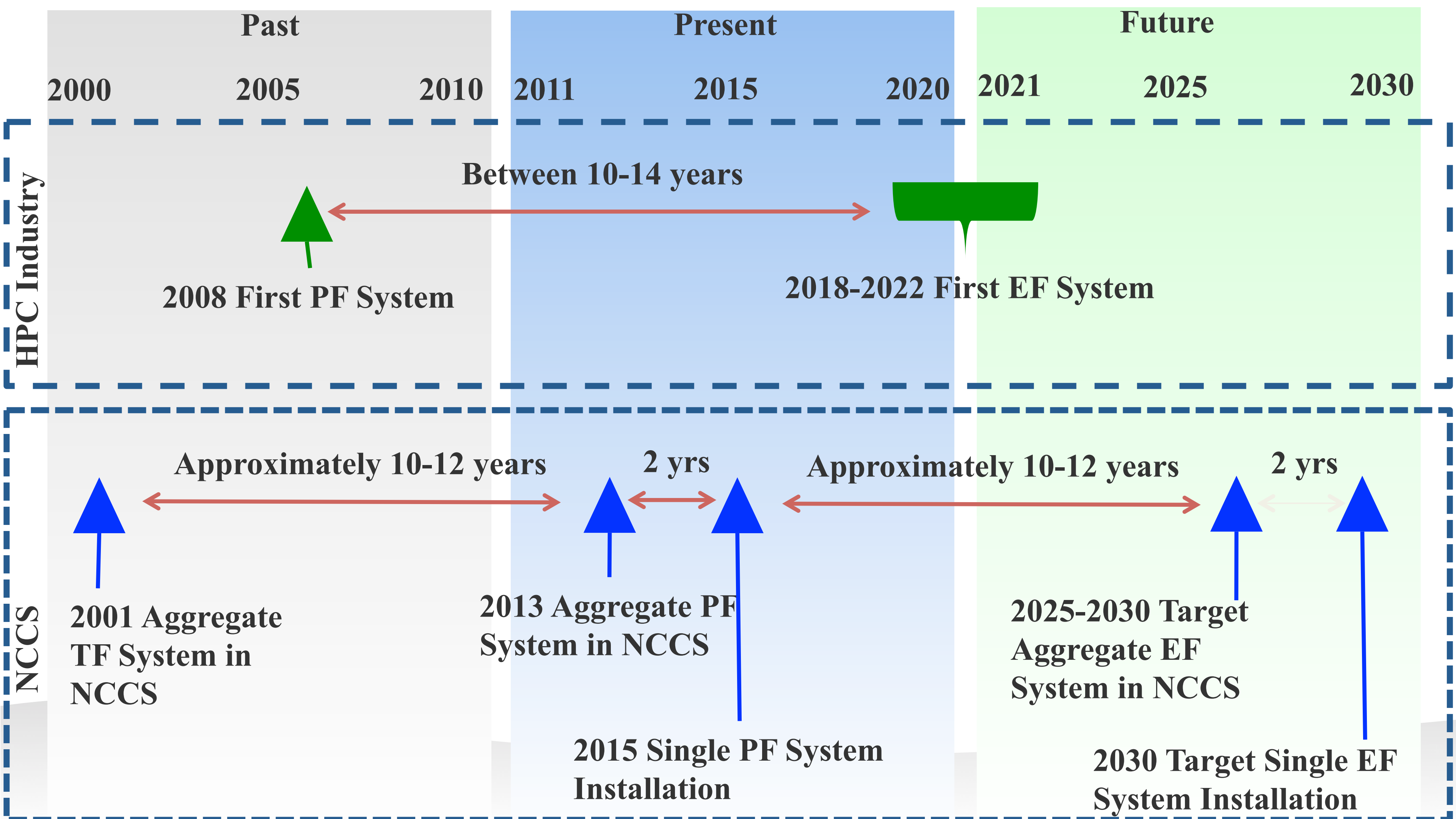
Lifecycle of Data
- Viva La Data
- Vintage
- Vindictive
- Vicious

Data Security
- Vandalized
- Victimized
- Velociraptor
- Voldemort

Just for fun
- Vortex
- Vice
- Venomous
- Vivacious

Backup Slides

Moving toward Exascale



What would an Exascale system look like for the NCCS?

NCCS Projections by Mid 2016 (traditional processors only)

- Peak Computing ~3,500 TF
- Total RAM ~403,968 GB and Total Storage ~54,667 TB

Ratio of Memory to Peak Computing

- Peak Computing ~3,500 TF
- Total RAM ~403,968 GB
- Ratio = ~116 GB/TF (0.1 Byte/Flop)

Ratio of Storage to RAM

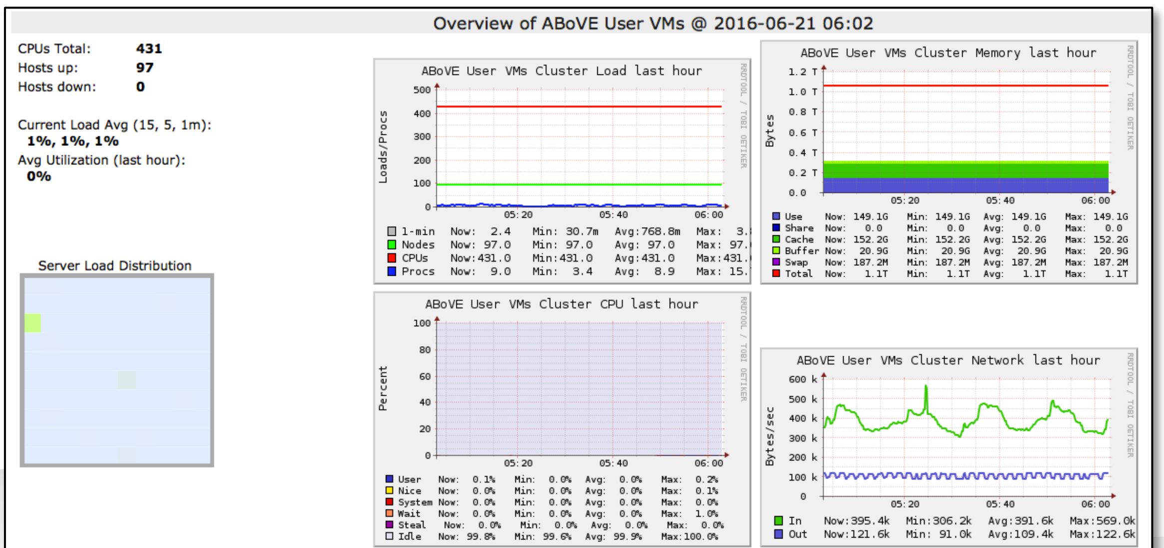
- Total Storage ~54,667 TB
- Ratio = ~138

Extrapolating to Exascale

- 1 EF Peak
- 100 PB of RAM
- 14 EB of Storage

Start Small and Scale Up

- Start users with 1 to 4 virtual machines each with 4 cores and 9 GB of RAM
- 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 a summary of ABoVE virtual machine environment.

- 431 CPUs or cores
- 97 virtual machines

Software Stack (mostly open source)

External License Servers

- 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
Python, NetCDF, R, etc.

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

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

Operating Systems
Linux (Debian, CentOS) and Windows

Platform-as-a-Service

- Compute virtual machines are open souce Linux
- Windows used for remote desktop and ArcGIS

Data Publication and Distribution Services

Data Publication Services	Protocol	Download	Subsetting	2D Visualization
Web Access For downloading small files	HTTP	✓		
File Transfer Protocol (FTP) Anonymous FTP supporting wget	FTP	✓		
GRads Data Server (GDS) Data subsetting and analysis services	OPENDAP	✓	✓	
Live Access Server (LAS) Data subsetting and analysis services	OPENDAP	✓	✓	
THREDDS Data Server (TDS) subsetting , analysis, & visualization	OPENDAP	✓	✓	✓
Earth System Grid Federation (ESGF) Data access to IPPC CMIP data	OPENDAP	✓		✓
Web Map Service (WMS) Data publication to IPPC CMIP Format	OPENDAP	✓	✓	✓

Future: Working on an architecture to support ArcGIS server, portal, and desktop.