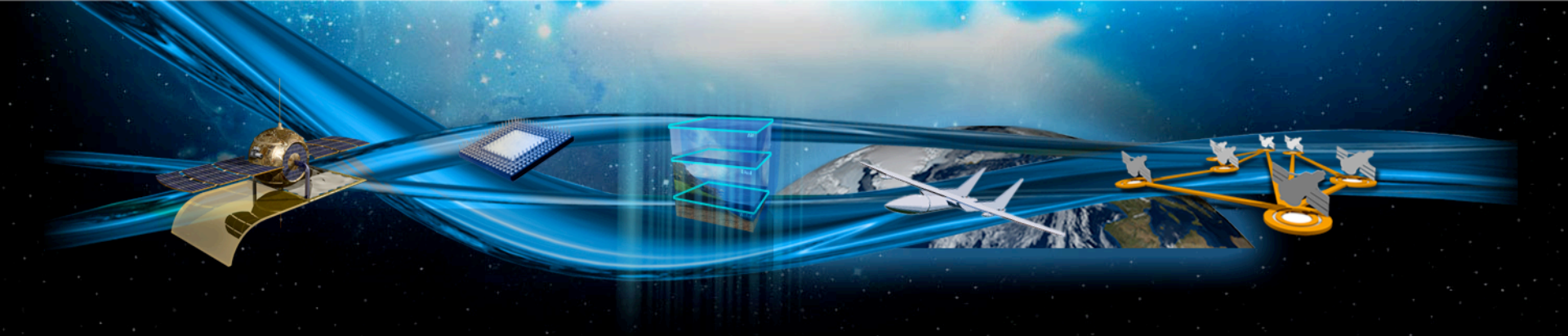


Big Data Infrastructures for Earth Science Data-Intensive Analysis

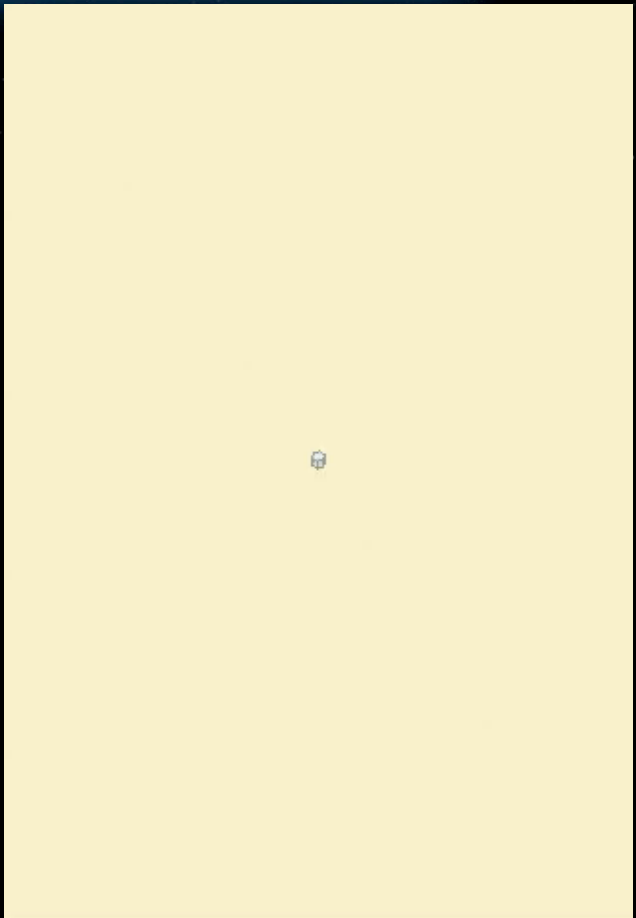
K-S Kuo, TL Clune, MS Seablom, &
R Ramachandran

My Colleagues

- Dr. Thomas Clune
  - Lead, Advanced Software Technology Group, NASA Goddard Space Flight Center
- Mr. Michael Seablom, MBA
  - Chief Technologist of Science, NASA HQ
  - Formerly, Program Manager of the Advanced Information System Technology (AIST) program at NASA Earth Science Technology Office.
- Dr. Rahul Ramachandran
  - DAAC Manager, NASA Marshall Space Flight Center

K-S Kuo – Self-introduction

- Affiliations, all in Maryland, USA:
  - Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA.
  - Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD, USA.
  - Bayesics, LLC, Bowie, MD, USA.
- Trained in atmospheric radiative transfer.
  - Took a number of computer science classes!
- Practicing in remote remote sensing.
  - Currently involved in precipitation remote sensing, associated with NASA/JAXA Global Precipitation Measurement mission.
- Advised, as a scientist user, NASA Earth Science Data Information System project for 3 years.
- Certified by IEEE Computer Society as a Software Development Professional.
- Passionate in crafting *Power Tools* for research.



Outline

- What is *Big Data*?
  - How is it different from *small data*?
- Established practices and architecture.
  - File-based archive and distribution.
  - Consequences of such practice.
- Addressing the Problem.
  - NASA's investments in improving the process.
  - Compute-intensive (HPC) architecture,
  - Data-intensive (shared-nothing) architecture.
- A closer look at a data-intensive experiment.
- Concluding Remarks

What is *Big Data*?

- *Big Data* is an abbreviation of this situation:
  - **The accumulation of data outpaces our ability to fully analyze them and maximize knowledge extraction in a timely manner.**
- *Big Data* is characterized by the following challenges:
  - ***Volume***: accumulated volume growing exponentially!
  - ***Variety***: *in situ*, remote sensing, model output…
  - ***Velocity***: need for vastly improved speed in analysis.
  - ***Veracity***: beyond data quality into reproducibility.
- The Goal of *Big Data* technology/architecture/practice:
  - Maximizing return on investment, i.e. ***Value***!

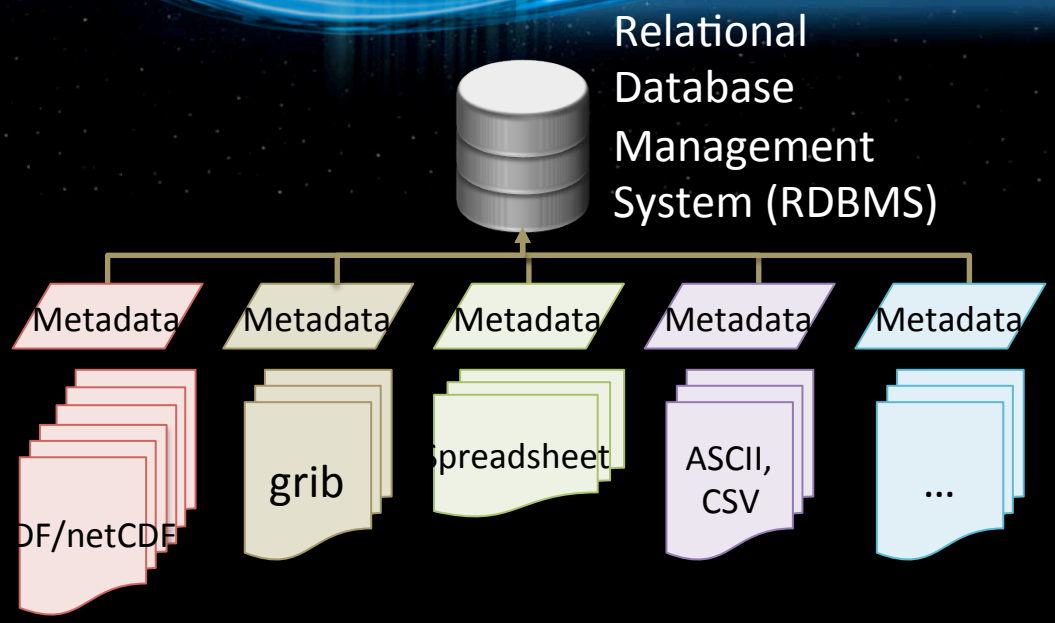
How is *Big Data* different?

- Era of "Private Data"
  - Prior to 1990s
  - Example: Kepler using Tycho Brahe's observation of 2' accuracy as opposed to Copernicus's 10'.
- Era of "Open Data"
  - NASA is a frontrunner of the "open data" policy with Earth Observing Systems Data Information System (EOSDIS).
  - Data volume accessible by individual researchers explodes!

Established Practices and Architecture

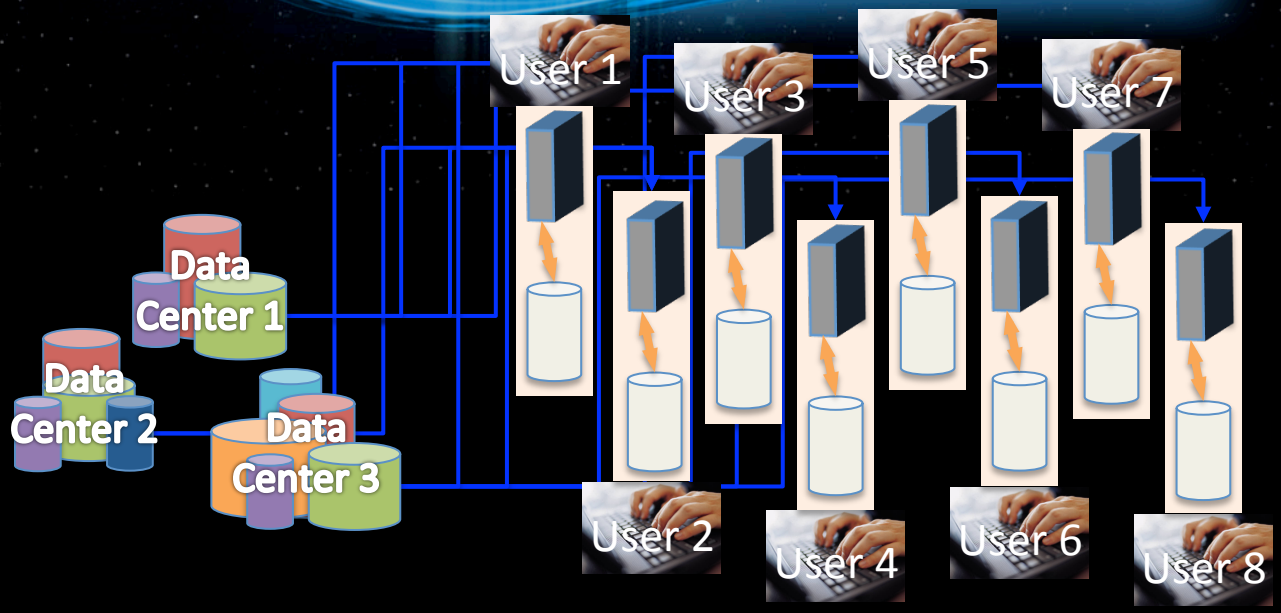
- File-based archive and distribution.
  - Root of all ills?
- Established, prevailing architecture.
- Stove-piped vs. collaborative.
- An *optimistic* data-analysis lifecycle.
- The full cost of data-analysis research?
- IT cost at NASA Goddard Space Flight Center (GSFC) – a sample

File-based Archive and Distribution



- The great majority of Earth science data are archived and distributed as files, ⋯
- Standardized through APIs, e.g. HDF and netCDF, for access, ⋯
- Only the metadata of which are cataloged into RDBMS and are thus searchable.
- Searching for data not contained in the metadata becomes slow, even if it can be done (e.g. through OPeNDAP).

Established, Prevailing Architecture

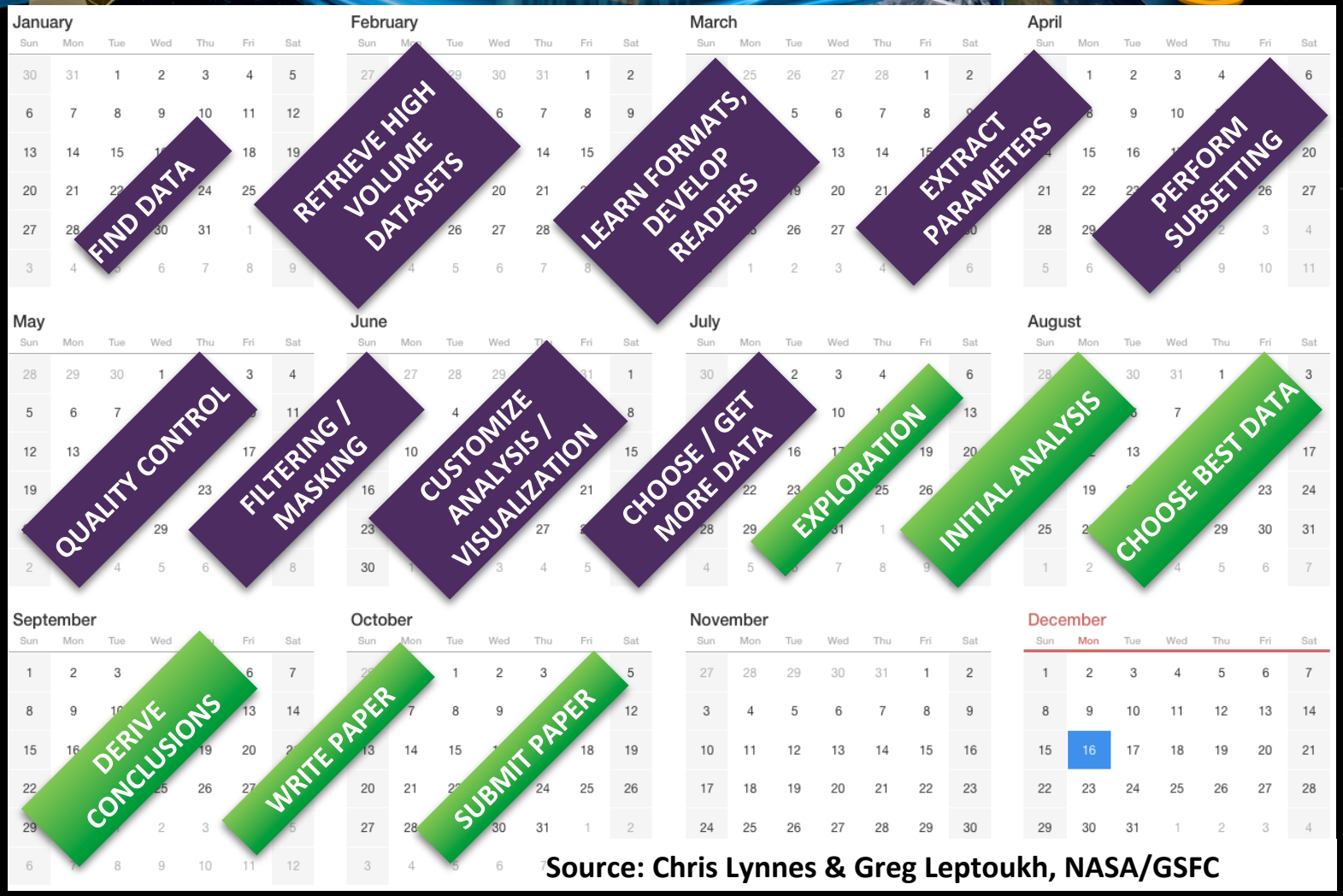


- Everyone has to transfer data to local storage.
- Processing is done locally, mostly in serial.
- Scientists need to procure the resources for local compute/storage.
- Scientists must engage in data management for local data.
- Analysis becomes highly personal, subject to individual's preferences.

Stove-piped vs. Collaborative

- We are largely "stuck" in this *stovepipe mode* for data analysis.
- On individual level, scientists today still labor
  - to find the data they need,
  - to learn formats to read unfamiliar data,
  - to perform subsetting on the quantities desired,
  - to implement quality control,
  - to visualize and interpret the products they have generated,
  - and, in many instances, repeat all of these steps until a satisfactory resultant product has been generated.
- The process above frequently represents more than 6 months of preparation, just so the actual research may commence.

Optimistic Data Analysis Lifecycle



Source: Chris Lynnes & Greg Leptoukh, NASA/GSFC

The Full Cost?

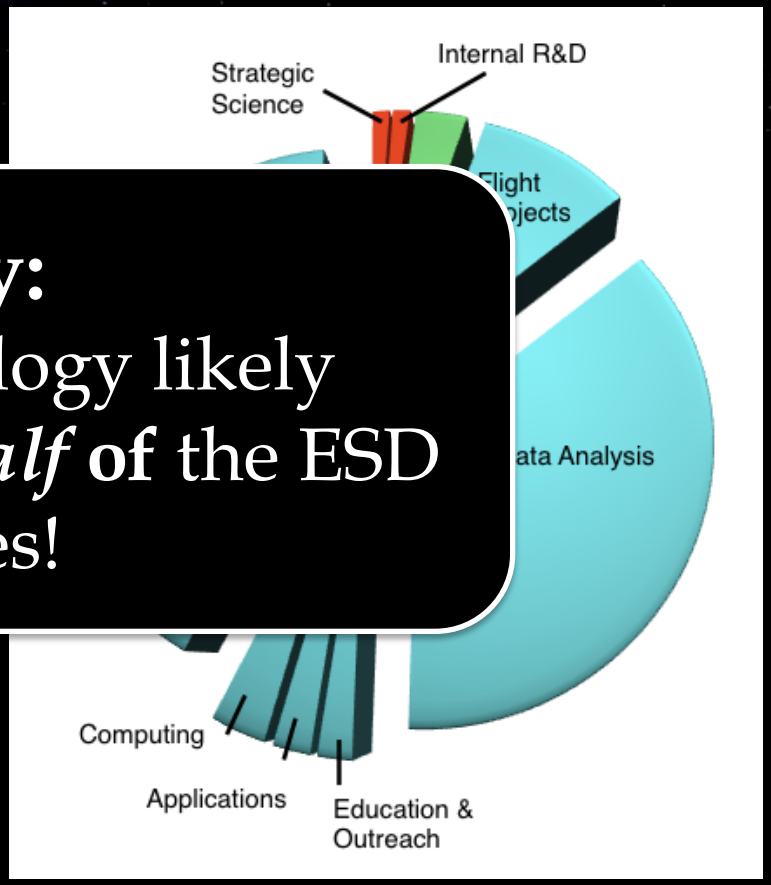
- Furthermore, on the institutional level, the process requires IT support, such as
  - Procuring the hardware and software resources,
  - Housing and maintaining the resources, and
  - Downloading and backing up the data.
- Established practices cause enormous repetition, duplication, and hence inefficiency:
  - Every scientist repeats at the individual level.
  - Every institute, or even every research group within the same institute, repeats at the institutional level.

NASA Goddard Earth Science Division

2009 Profile

- 195
- 215
- 639

Estim



Takeaway:
Information Technology likely consumes ***more than half*** **of** the ESD expenditures!

Addressing the Problem

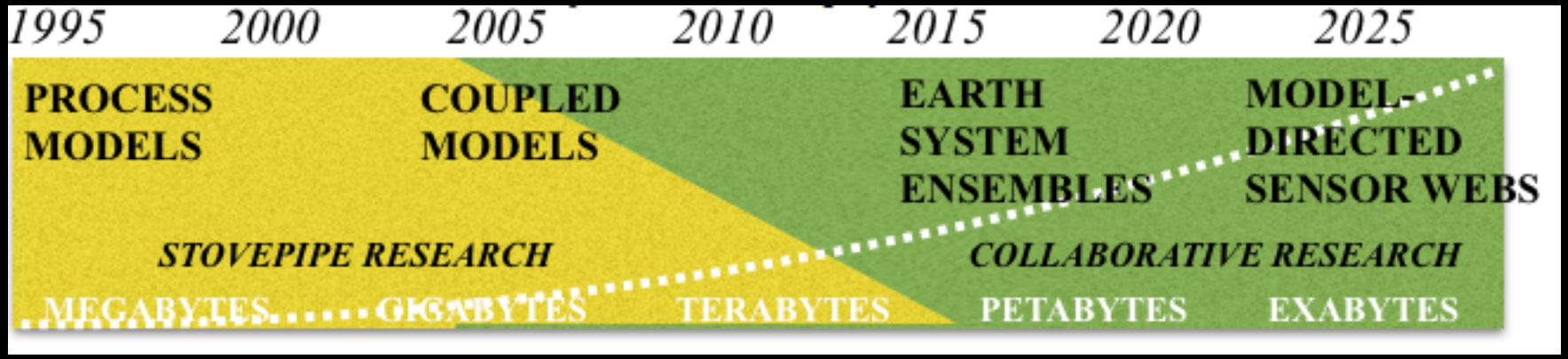
- A brief history of NASA IT investments.
- NASA IT investment strategy.
- Examples
  - Compute-intensive architecture
    - AIST investment: NASA Earth eXchange, NEX
  - Data-intensive architecture
    - AIST investment: Automated Event Service, AES
- A near-term vision.

A Brief History

- 1980's and 1990's Information Technology (IT) investments tilted heavily towards hardware.
  - Return on investment (RoI) for hardware is more obvious than that for software.
- By the turn of the century, the awareness of unwieldy and non-interoperable software shifted the hardware-centric focus toward software quality and reuse.
- NASA's past Computing Information and Communications Technologies (CICT) and on-going support for Earth System Modeling Framework (ESMF) represent examples of such a shift.
- Other Federal (non-military or -intelligence) agencies have similar, and some larger, programs.
- While these issues remain largely unsolved, another challenge, i.e. Big Data, has (re)surfaced.

Model Development Trend

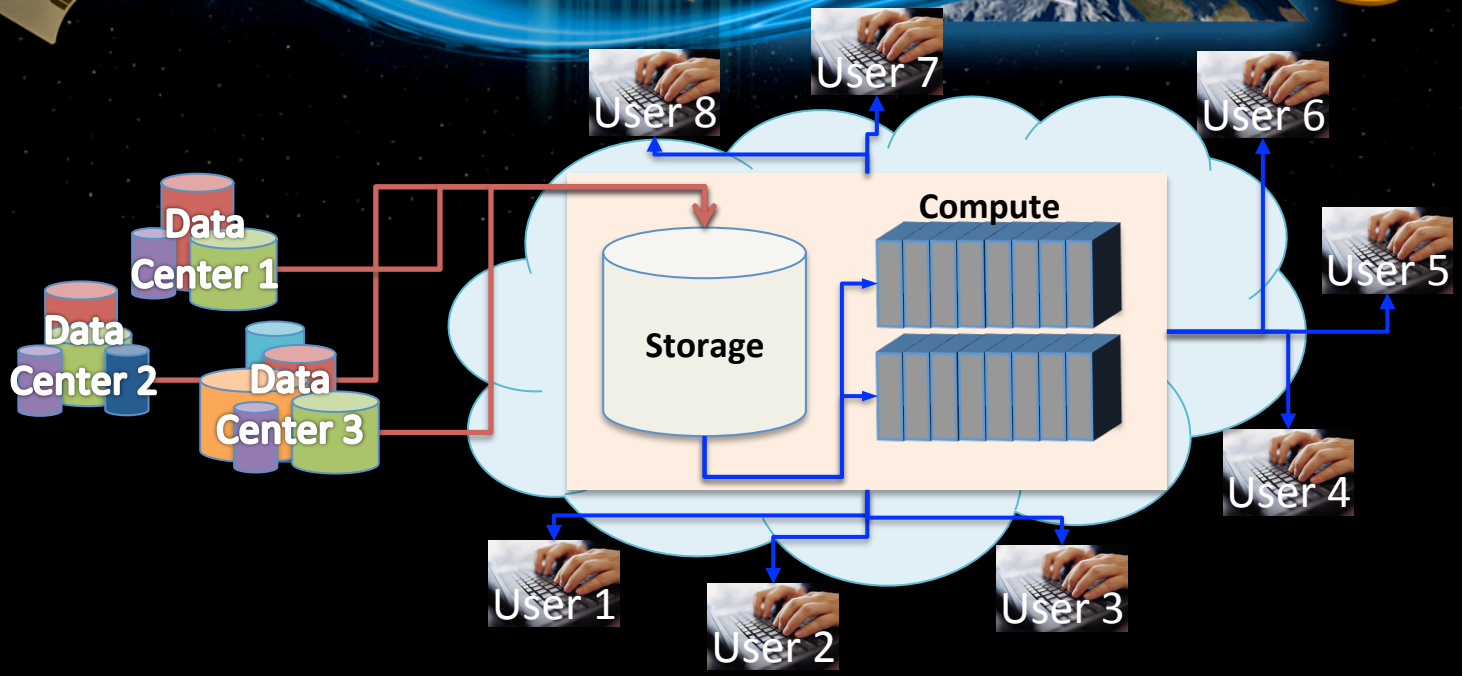
- Model simulation output is a significant driver in data volume increase:
  - MERRA products total ~100 TB
  - A planned NASA GEOS-5 2-year Nature Run to support OSSE will produce ~3 PB *compressed*!
- Model development is in an evolution to become more synergistic.



NASA IT Investment Strategy

- Consolidate resources and services to reach a scale that is more economical.
- Leverage modern software architectures and technologies to capitalize hardware parallel capabilities.
- Provide collaboration-friendly infrastructure to foster easier interdisciplinary collaborations.
- A couple examples from AIST investments:
  - NASA Earth eXchange, NEX
  - Automated Event Service, AES

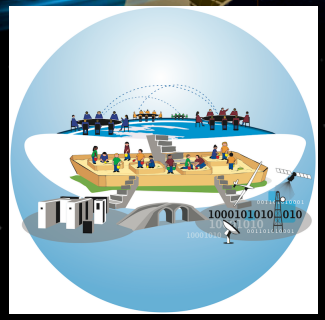
Compute-intensive Architecture



- Data transferred to fast, central storage.
- High-bandwidth connection from storage to compute.
- Infrequent access to storage once compute starts.
- Parallelism often involves tight coupling through MPI
  - Inter-process communication is essential.

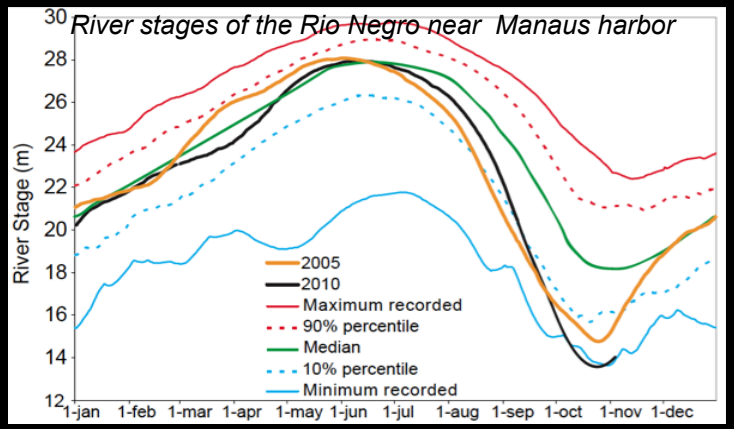
Sample AIST Investment:
NASA Earth Exchange (NEX)

Rama Nemani, NASA / ARC



NASA's first Collaboratory brings computing and large data stores together to engage and enable the Earth science community address global environmental challenges

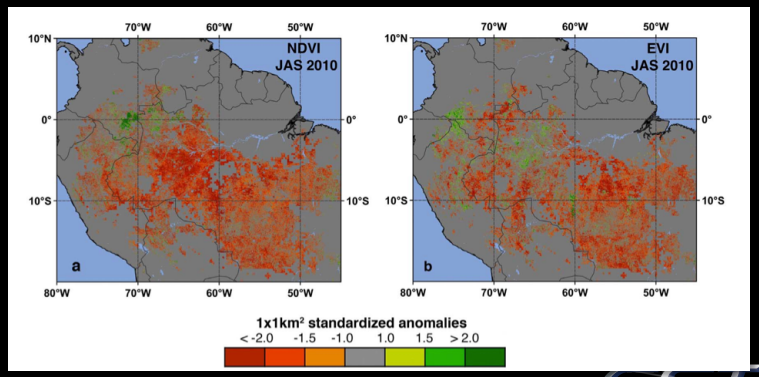
- ***Current capability: 10K+ cores, 1PB online data***
- *New paradigm in "big data" analysis and scientific discovery*



Samanta *et al.*, 2010 and Xu *et.al*, 2011 used NEX to process large amounts of data to examine the 2005 and 2010 severe Amazon droughts

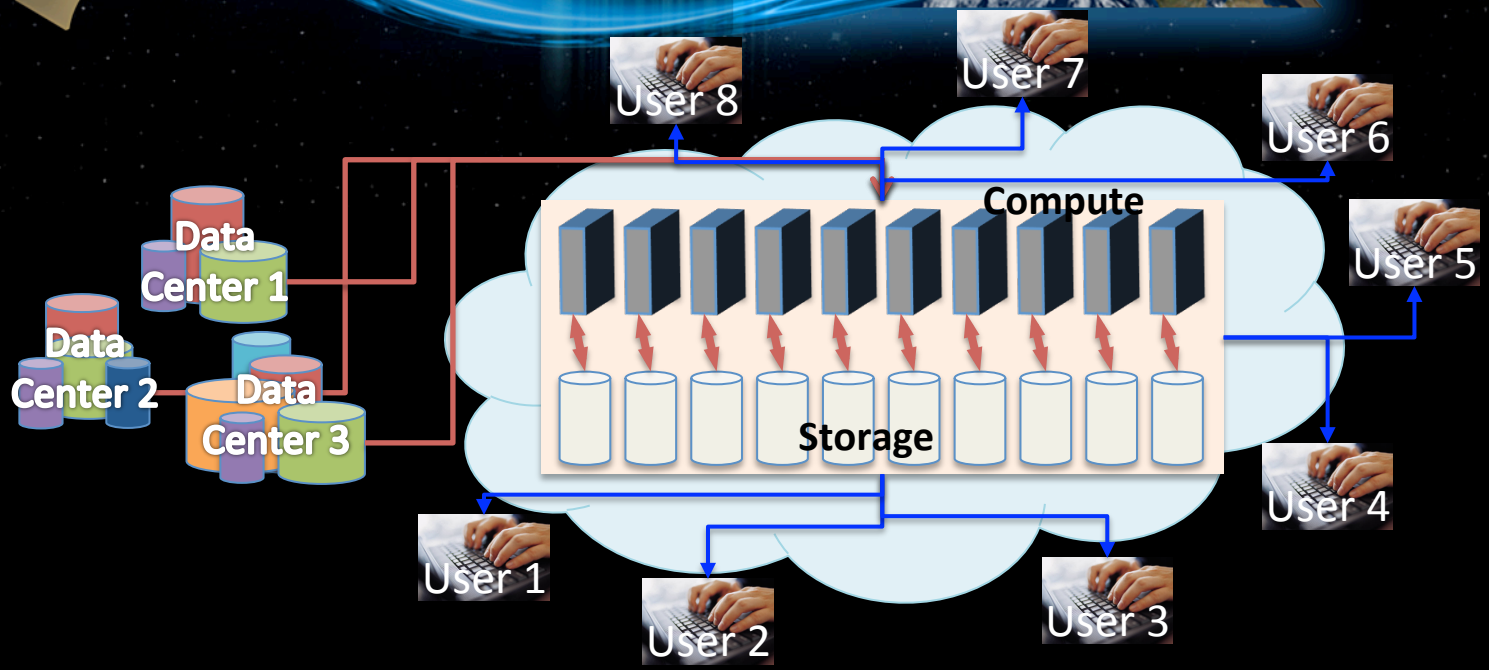
- *Prior to NEX (2005) case: 18 months to analyze data and submit paper*
- *After NEX (2010 case): 4 months to analyze data and submit paper*

Work begun in 2012 will build semi-automated workflows for science analysis



Primary funding for NEX is provided by NASA's High End Computing Program with support from ESTO

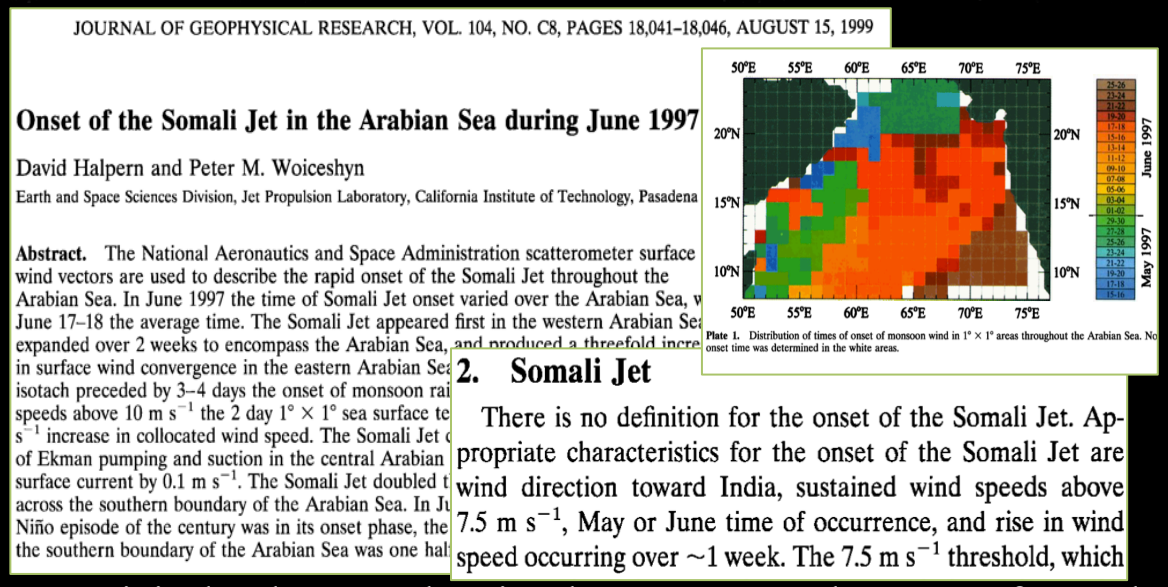
Data-Intensive Architecture



- Data transferred through coordinator nodes of a cluster.
- Data are distributed and indexed to local node storage.
  - High-bandwidth local storage access.
- Frequent data transfer from and to the storage is the norm.
- Analysis compute is primarily loosely coupled.
  - Little or no inter-process communication.

Sample AIST Investment:
Automated Event Service

- Use *SciDB* on a shared-nothing architecture.
- Support custom-defined event discovery and identification.
- Incorporate a collaborative infrastructure.



JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. C8, PAGES 18,041–18,046, AUGUST 15, 1999

Onset of the Somali Jet in the Arabian Sea during June 1997

David Halpern and Peter M. Woiceshyn

Earth and Space Sciences Division, Jet Propulsion Laboratory, California Institute of Technology, Pasadena

Abstract. The National Aeronautics and Space Administration scatterometer surface wind vectors are used to describe the rapid onset of the Somali Jet throughout the Arabian Sea. In June 1997 the time of Somali Jet onset varied over the Arabian Sea, w June 17–18 the average time. The Somali Jet appeared first in the western Arabian Sea expanded over 2 weeks to encompass the Arabian Sea, and produced a threefold incre in surface wind convergence in the eastern Arabian Sea isotach preceded by 3–4 days the onset of monsoon rai speeds above 10 m s^{-1} the 2 day 1° × 1° sea surface te s^{-1} increase in collocated wind speed. The Somali Jet of Ekman pumping and suction in the central Arabian surface current by 0.1 m s^{-1}. The Somali Jet doubled t across the southern boundary of the Arabian Sea. In Ju Niño episode of the century was in its onset phase, the the southern boundary of the Arabian Sea was one hal

Plate 1. Distribution of times of onset of monsoon wind in 1° × 1° areas throughout the Arabian Sea. No onset time was determined in the white areas.

2. Somali Jet

There is no definition for the onset of the Somali Jet. Appropriate characteristics for the onset of the Somali Jet are wind direction toward India, sustained wind speeds above 7.5 m s^{-1}, May or June time of occurrence, and rise in wind speed occurring over ~1 week. The 7.5 m s^{-1} threshold, which

Installed on the MERRA Analytic Service (MAS) cluster

- Current capacity:
  - ~1K cores,
  - ~1 PB online data
- Data parallelism

- Original Halpern and Woiceshyn (1999) study was performed using 1 year, i.e. 1997, of NASA Scatterometer, NSCAT, data
- Two years later Halpern and Woiceshyn (2001) published a follow-on paper using 12 years of NSCAT data
- AES extended it to ~24 years of DMSP SSMI data and got it done in minutes!

Comparison of the Architectures

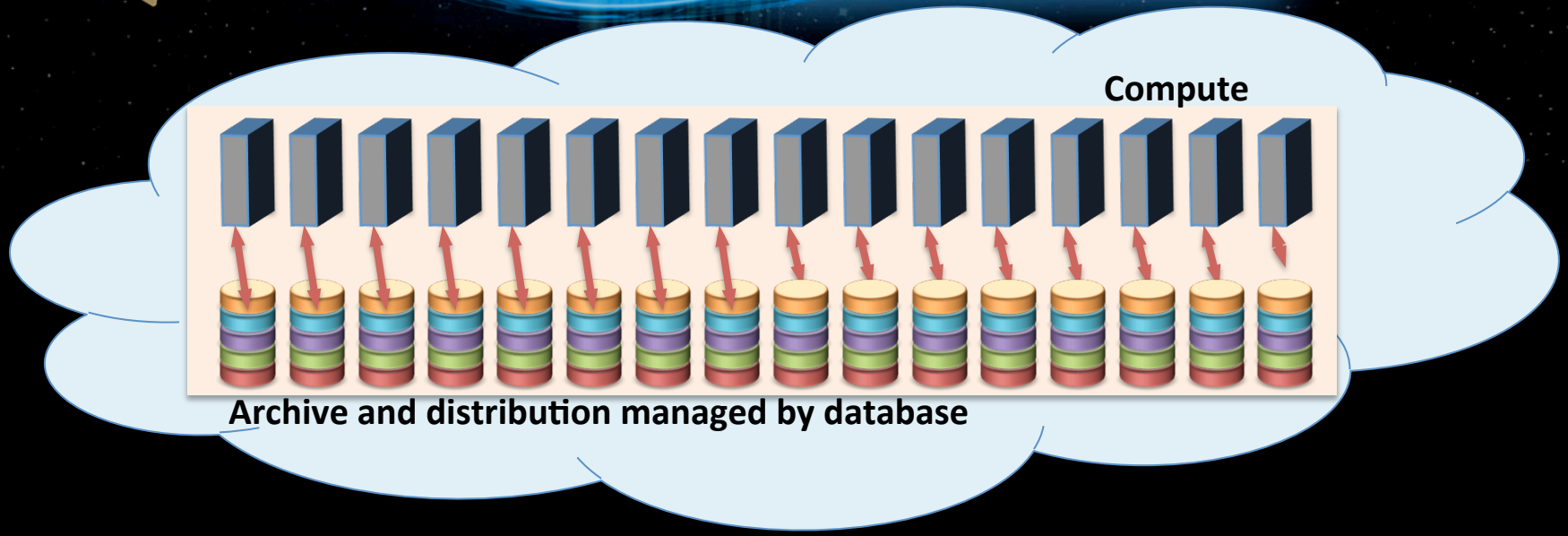
Compute-intensive

- Suits tightly-coupled parallel problems.
- Reuses existing code.
  - However, parallelization often needed.
- Operates largely in batch mode.

Data-intensive

- Suits mainly pleasingly parallel problems.
- Needs new code to take advantage of the new architecture.
- Emphasizes real-time, dynamic responses.

A Near-term Vision



- Data are archived directly into distributed parallel database.
- The same shared-nothing architecture supporting the database also supports in-place data analysis.
- Advantages:
  - Intrinsically data parallel.
  - Reducing need for data download.
  - Data management largely handled by the database.

Closer Look at AES

- Introduction
- SciDB - *Big Data* technology of choice
- SciDB architecture
- Hardware environment
- Blizzard science scenario
  - Blizzard definition and considerations
  - Visibility vs. snowfall rate
  - Visibility vs. wind speed (blowing snow)
  - Steps to blizzard identification
  - 2010 Winter CONUS blizzard animation

Automated Event[1] Service

A NASA Advanced Information System Technology (AIST) project funded through the Earth Science Technology Office to develop the technology to

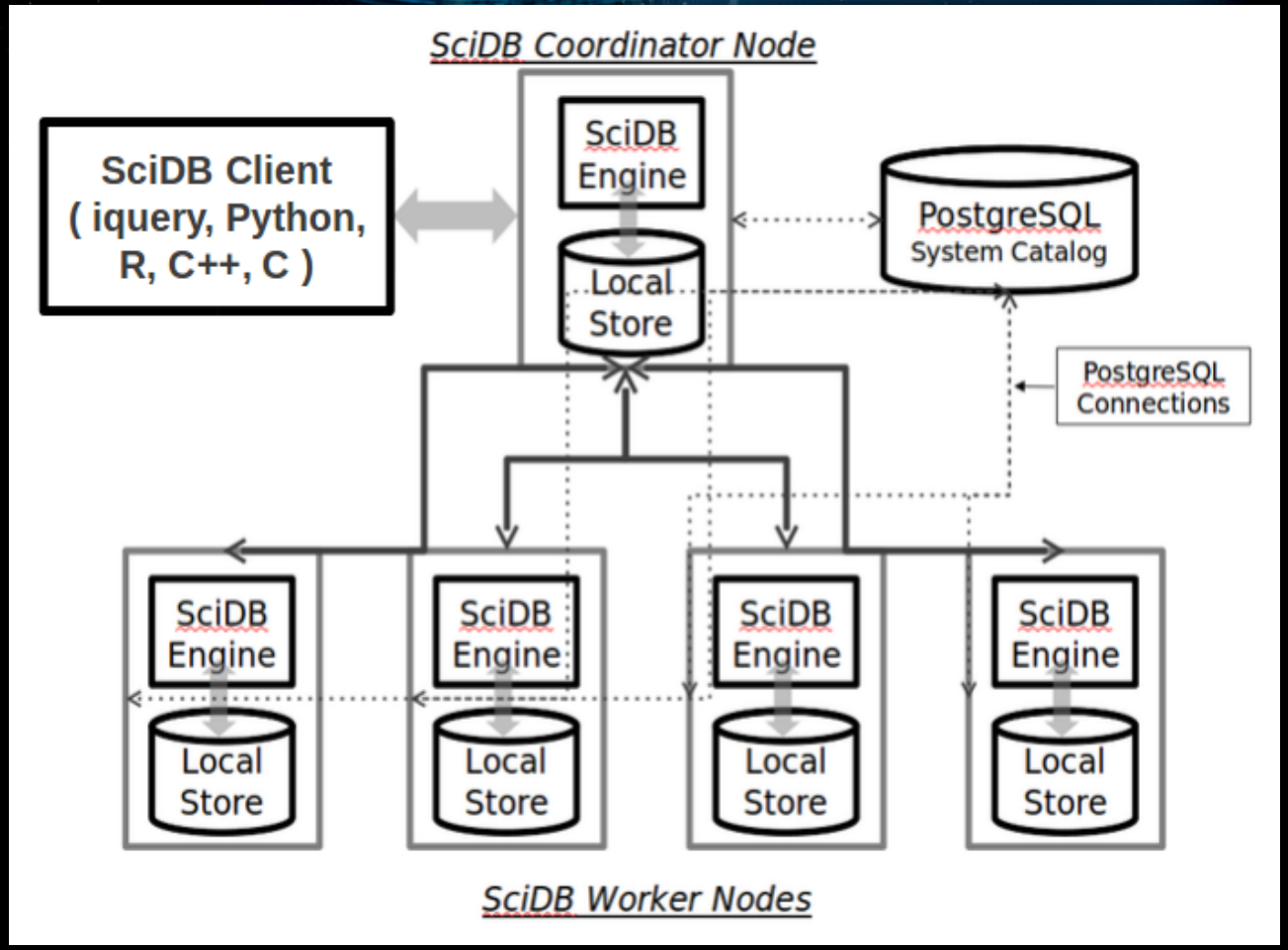
- Enable **systematic** identification of **investigator-defined** Earth science events from reanalysis and satellite data.
  - Addressing a significant portion of ES research;
  - Reducing duplication of effort among research teams;
  - Improving return on investment (ROI) for NASA data and compute resources.
- Provide driver to improve affinity of computing and data resources
  - Move computing to the data rather than data to computing.
- Greatly improve interactive data exploration and analysis.

[1] Events are occurrences of phenomena, usually 4D (space and time) in nature.

A *Big-Data* Technology: SciDB

- An all-in-one ***data management*** and advanced ***analytics platform*** that features:
  - Complex analytics inside a next-generation parallel ***array*** database,
    - *i.e.* not *row-based* or *column-based* like RDBMS's based on *table* data model
  - Based on the "shared nothing architecture" for data parallelism,
  - data versioning and provenance to support science applications, and
  - Open source (currently in beta).
- A better performer than Hadoop (MapReduce):
  - 2-10 times faster, in almost all benchmarks that we have performed so far.

Basic SciDB Architecture



NCCS[1] Experimental MAS Cluster

- 36 nodes
  - 30 in a main cluster
  - 6 in a test/development cluster
- 2x8 SandyBridge Intel Cores, i.e. 16 cores per node,
- 32 GB memory per node,
- 36 TB local storage per node,
- FDR Infiniband
- Gigabit ethernet

[1] NCCS – NASA Center for Climate Simulation

Blizzard Science Scenario

- **Definition**: According to NWS, a blizzard means that the following conditions are expected to prevail for a period of 3 hours or longer:
  - **Sustained** wind or **frequent** gusts to 15.6 m/s (35 mph) or greater; and
  - **Considerable** falling and/or blowing snow, *i.e.*, reducing visibility **frequently** to less than 400 m (¼ mile)
- The definition contains **imprecise** adjectives/qualifiers.
- Point-based (local and/or instantaneous) definitions do not translate directly to space/time-averaged parameters as in MERRA[1] reanalysis data.
- It is obvious that **visibility** is the crucial criterion in defining blizzard, but MERRA does not yet include visibility observation.
- Visibility is directly related to in-air snow mass concentration and dependent upon snow particle mass-dimension (morphology) property.

[1] MERRA – Modern Era Retrospective analysis for Research and Applications for the satellite era using NASA's GEOS-5 model, focusing on historical analysis of the hydrological cycle. It is composed of multiple regularly gridded data sets, **~100 TB** total.

Additional Considerations

- Most of the ES phenomenon definitions are, like blizzard, not based on space/time-averages.
- It is not possible to define ES phenomena unequivocally (e.g. nothing but grid cells containing blizzard condition) using space/time-averaged data sets such as MERRA re-analysis.
- The goal thus becomes finding **the smallest possible superset** that, for example,
  - captures all blizzard grid cells and
  - minimizes the number of false-positive, non-blizzard grid cells.
- Not all fields may be available. In-air snow mass concentration is contributed primarily by
  - Falling snow – using snow rates in MERRA
  - Blowing snow – using snow accumulation on surface and wind speed at 10 m above surface as proxy
- Events found using MERRA serve as a basis to locate other useful data sets for validation or refinements.

Visibility vs. Snowfall Rate

- Rasmussen et al (1999) plot visibilities and snowfall rates
  - based on theoretical calculations of various snow crystal types, and
  - supported by observations.
- Variations among crystal types are considerable.
- The blue line is described by:

$$\log v = -\log s + 3$$

v: visibility in meter

s: snowfall rate in mm hr-1

- The above relation is in turn used to find extinction as a function of snowfall rate.

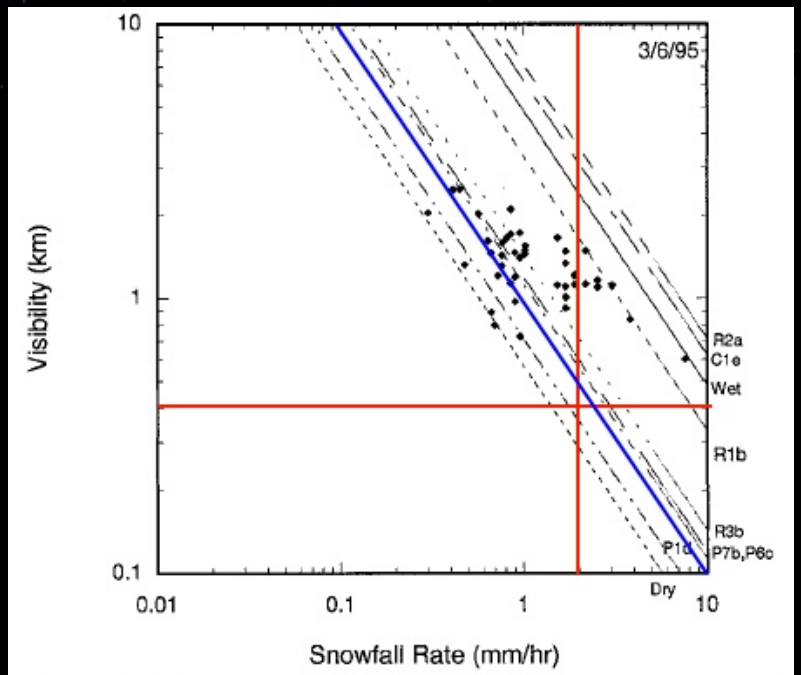


FIG. 16. Theoretical visibility–snowfall relationships from Eq. (19) compared to the observed visibility–snowfall data from 6 Mar 1995. The theoretical curves correspond to the crystal types observed for this event.

Visibility vs. Wind Speed in Blowing Snow

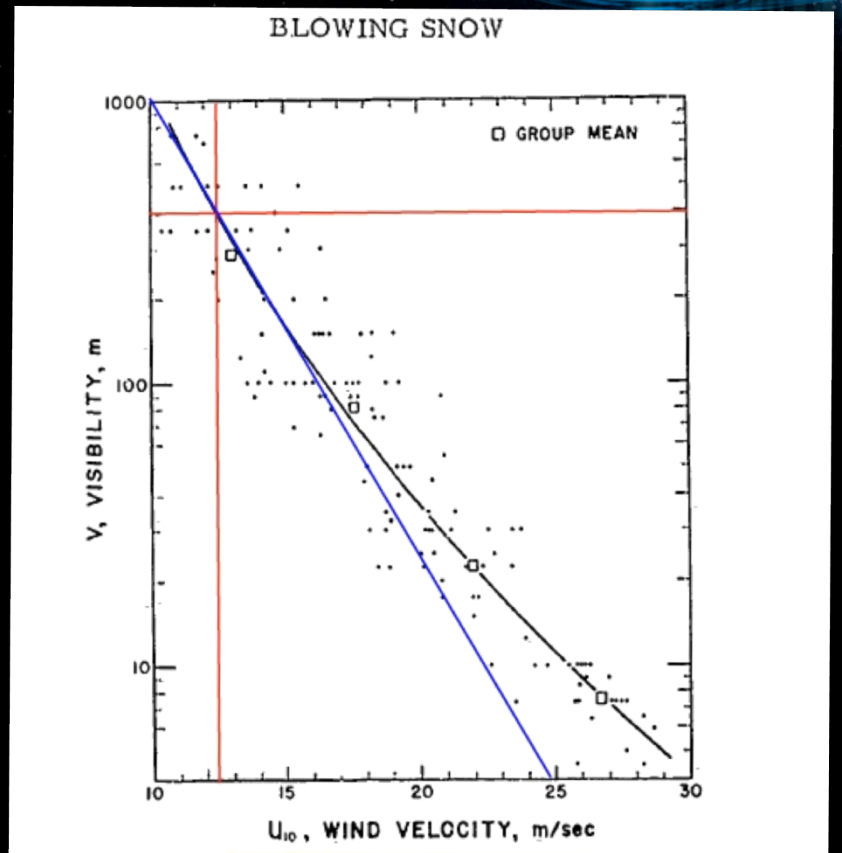


Figure 65. Logarithm of visibility as a function of wind speed at 10-m height for blowing snow conditions. (From Liljequist[53]).

- Most studies relates visibility to in-air snow mass concentration (g m^{-3}).
  - A combination of falling and blowing snow.
- Liljequist (1957) relates visibility directly to wind speed based on typical blowing snow situations in Antarctica.
- The blue line is described by

$$\log v = -0.1592w + 4.5918$$

v: visibility in meter

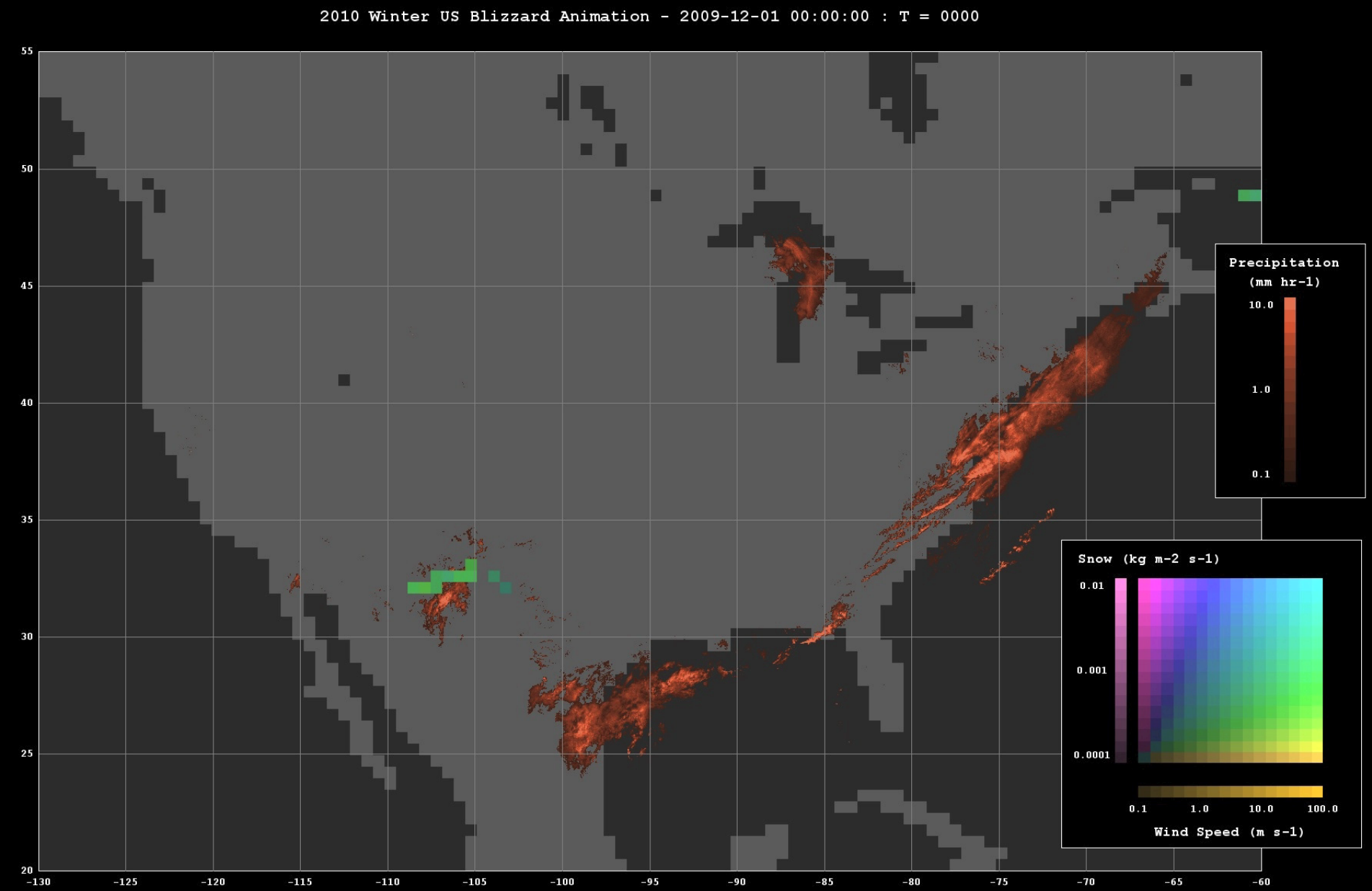
w: wind speed 10-m above surface in m s^{-1}

Steps to Blizzard Definition using MERRA

Using MERRA high-resolution (.5°/.67° in latitude/longitude) **hourly** data sets

- Extract 2010 Winter US subset.
- Calculate:
  - wind speed 10-m above surface, w, using the east-west (U10M) and south-north (V10M) components,
  - grid-area-weighted mean (μ_w) and standard deviation (σ_w) of w, and
  - grid-area-weighted mean (μ_s) and standard deviation (σ_s) of **log 10** snow rates in **snow-only** grids.
- Conduct trial experiments to define blizzard:
  - Experiment with snow rate threshold defined as rational multiples of σ_s above μ_s, i.e. $\theta_s = \mu_s + q\sigma_s$,
  - Find corresponding wind speed yielding same visibility as snow rate threshold in blowing snow conditions, and
  - Apply wind speed criterion to grids with snow accumulation over 3-cm.
- It is found that $q \approx 1.6$ yields satisfactory results
- For the **global** data sets
  - Find μ_w, σ_w, μ, and σ_s,
  - Use the same q to determine snowfall threshold, and
  - Determine corresponding wind speed criterion.

2010 Winter US Blizzard



Sister Project 1 of AES: PROBE

Enable routine use of process-based diagnostics (PBDs) as a means for targeted improvement of weather/climate models.

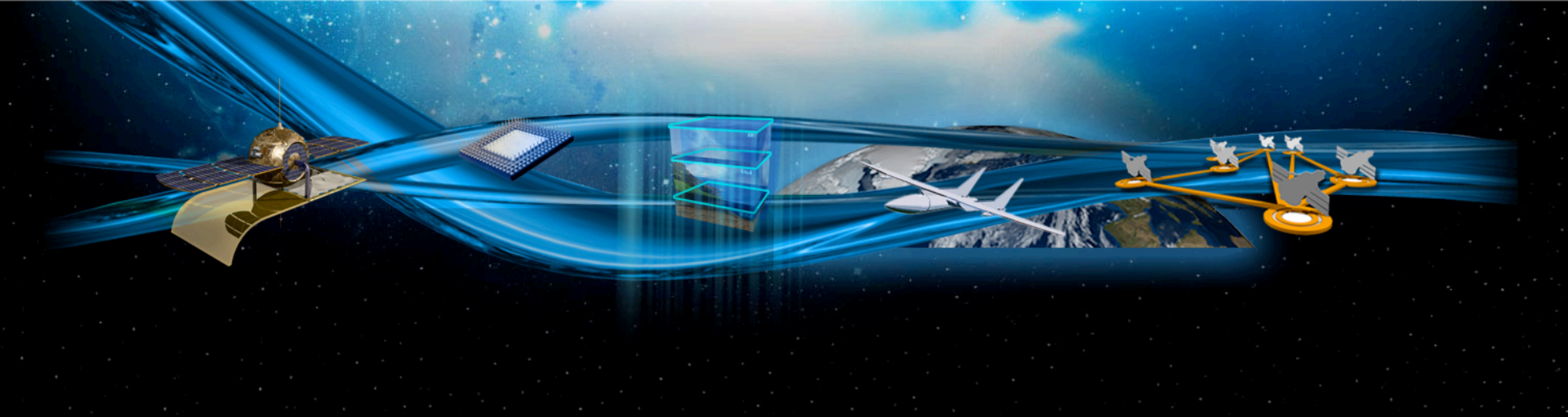
- Conventional diagnostics are inadequate:
  - Strong coupling - difficult to point finger
  - Signal is diluted by portions of domain where process is inactive
- Builds upon Automated Event Service (AES)
  - Step 1 is to identify regions in which a selected process is active. I.e. "events" by AES definitions.
  - PROBE requires more analysis and customized operators.
- PROBE must also extend AES to support:
  - Ensembles and versioning
  - Routine automations for iterative evaluation

Sister Project 2 of AES: Cloud

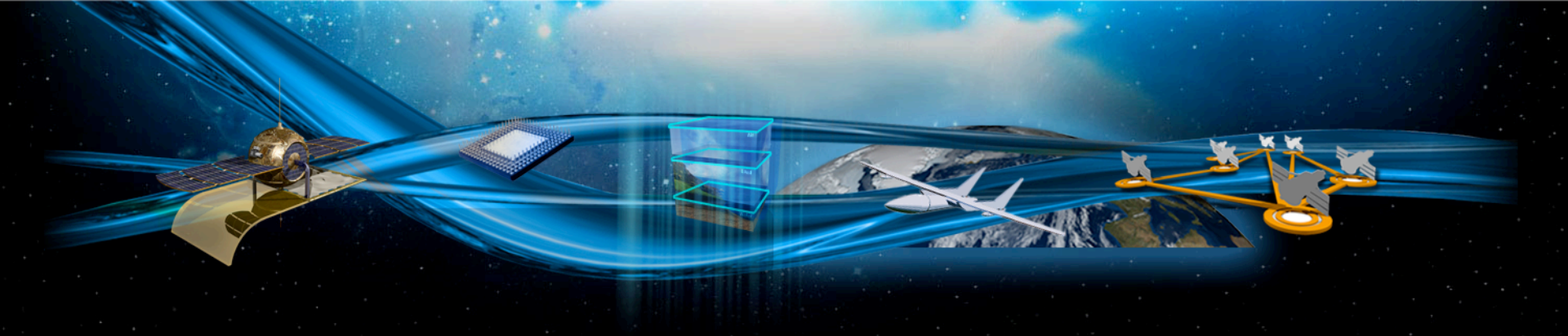
- Goals:
  - Explore feasibility of providing AES as a cloud-service
  - Demonstrate scalability of AES/SciDB on large systems
    - 4-node ➛ 30-node ➛ 300-node ➛ (3000-node?)
- Limitations of existing resources:
  - Traditional HPC is not well-suited for interactive exploration and distributed I/O
  - Traditional data centers are not well-suited for custom (user-defined) analysis. Also data is segregated across DAACs.
- <u>Cloud-based computing appears well-suited.</u>

Concluding Remarks

- Constrained resources necessitate a holistic solution, requiring the research communities to adapt as well.
  - Science communities can no longer rely on the IT programs alone to offer solutions.
- Science Programs need to acclimatize themselves to these changes.
- Science communities and Science Programs need to craft a culture of utilizing advanced information technologies.
  - The full cost of inefficient research ecosystems is under-appreciated.
- Together we should embark on an evolution toward more productive architectures, collaboration-minded infrastructures and practices, and finally a self-sustaining and -rejuvenating research IT ecosystem.
- Some hopeful trends:
  - EarthCube
  - NASA Earth Exchange (NEX)
  - Big Data initiatives from OSTP, DoE, NIST

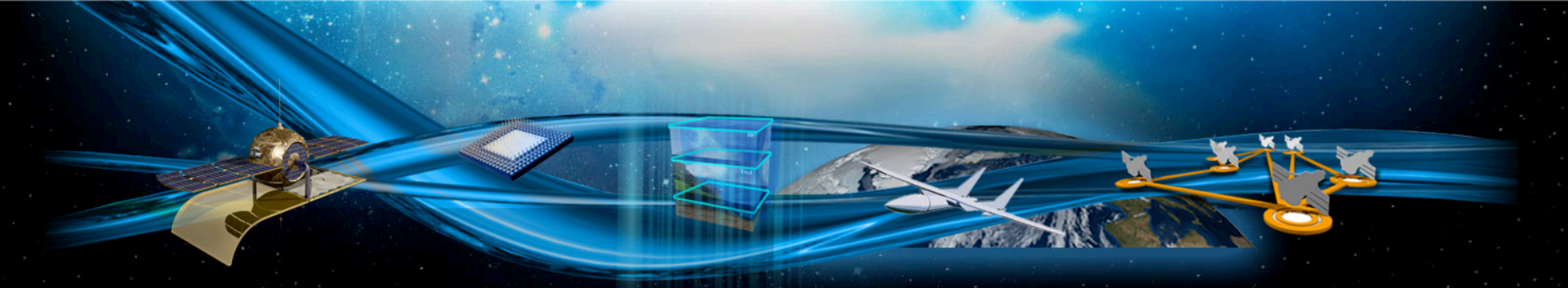


THANK YOU!



	Code 612	MAS
# Processors	184	480
RAM	616 GB	960 GB
Hard Disk Storage	355 TB	1.1 PB

Do we want to talk about Code 612 and AES comparisons?



- “Did Amazon forests green up during the 2005 drought?” – An exercise on reproducible science using SciDB.
- Should we talk about this?

Big Data Challenge

Earth Science as a science for "system of systems" demands interdisciplinary research.

- Need to analyze more ***varieties*** of data:
  - Scientists are required to collaborate outside of their traditional focus areas.
- Need to analyze great ***volumes*** of data:
  - To achieve better understanding we are resorting to ever finer resolutions in both observations and model simulations.
- Need to analyze data with greater ***velocity***:
  - Current processes are taking too long to digest the data and turn them into knowledge.
- Need to analyze data with ***veracity***:
  - Irreproducibility of science research is undermining validity.