Information Technology Infrastructure Committee (ITIC)

Report to the NAC

August 2011

Larry Smarr

Chair ITIC

Committee Members

Membership

- *Dr. Larry Smarr, Director- California Institute of Telecommunications and Information Technology*
- *Mr. Alan Paller, Research Director- SANS Institute*
- *Dr. Robert Grossman, Professor- University of Chicago*
- *Dr. Charles Holmes (Vice-Chair), Retired- NASA*
- *Ms. Debra Chrapaty, Chief Information Officer-Zynga*
- *Dr. Alexander Szalay, Professor- Johns Hopkins University*
- *Dr. Alexander H. Levis, Professor- George Mason University*
- *Ms. Tereda J. Frazier (Exec Sec), Special Assist. to CIO, NASA*

New Tools for Distributed Collaboration

NASA Ames Virtual Institute Leading the Way

- ***Workshop Without Walls: Broadening Access to Science Around the World***
  - Accepted by PLoS Biology
  - Authors: Betul Arslan, Eric Boyd, Wendy Dolci, Marco Boldt, Estelle Dodson, Carl Pilcher
  - Provides an overview of the methodology for hosting virtual science workshops, the challenges, benefits and lessons learned.

PLoS PUBLIC LIBRARY of SCIENCE

Letter to Science
"Leading the Charge to Virtual Meetings"

NAI *Leading the Charge to Virtual Meetings* was published in Science 11 February 2011

NAI letter was written in response to December 2010 letter *Travel Trade-Offs for Scientists* suggesting that scientists demonstrate new ways of doing business that don't contribute so substantially to climate change

NAI letter key points

- -NAI successfully demonstrated a new paradigm with two NAI *Workshops Without Walls* in March and November 2010
- -Surveys indicated that 98% of participants would attend another such workshop
- -Virtual meetings are just the beginning: emerging technologies are enabling virtual workspaces of the future

LETTERS

Leading the Charge to Virtual Meetings

IN HER LETTER "TRAVEL TRADE-OFFS FOR SCIentists" (10 December 2010, p. 1476), I. C. Burke wrote that science organizations should lead the charge to reduce greenhouse gas emissions from travel and that scientists should demonstrate a different way of doing business that doesn't contribute so substantially to climate change. These suggestions are right on the mark.

The NASA Astrobiology Institute (NAI) is pioneering the use of collaborative tools for virtual communication—i.e., meetings and scientific data analysis across distance. The NAI hosted two Workshops Without Walls during 2010 (*1*, *2*), which drew together hundreds of researchers from around the globe for scientific exchange, with no travel required. The integration of high-definition videoconferencing and online meeting software allowed attendees to participate in the workshops with just a computer and an Internet connection.

Surveys by the NAI after both Workshops Without Walls showed that 98% of 175 respondents would attend another one, although some noted that they missed the social interaction of in-person meetings (*1*, *2*). One NAI survey respondent wrote, "The energy, money, and time saving of this method of scientific communication is spectacular. I believe it should be encouraged and expanded."

A majority of NAI survey respondents participated in at least part of the workshops from their office or home. Others gathered in videoconferencing facilities at their institutions. Some moved freely between home, office, or conference room, catching talks most relevant to their research from wherever they happened to be. Experience has shown that holding meetings 4 hours per day over several days helps accommodate people in different time zones, as does recording and posting the talks in almost real time.

Virtual meetings are just the beginning. Emerging technologies such as 3-dimensional immersive environments to share, visualize, and analyze data sets across distance are enabling virtual workspaces for science and the potential for further savings of time, money, and natural resources.

WENDY W. DOLCI,[1]* MARCO S. BOLDT,[1,2] K. ESTELLE DODSON,[1,2] CARL B. PILCHER[1]

[1]NASA Astrobiology Institute, NASA Ames Research Center, Moffett Field, CA 94035, USA. [2]Lockheed Martin, NASA Ames Research Center, Moffett Field, CA 94035, USA.

*To whom correspondence should be addressed. E-mail: wendy.w.dolci@nasa.gov

References

1. D. Scalice, "NASA Astrobiology Institute hosts science 'Workshop Without Walls,'" NASA Astrobiology Institute, 19 March 2010 (http://astrobiology.nasa.gov/articles/nasa-astrobiology-institute-hosts-science-workshop-without-walls-/).
2. W. Dolci, "NAI hosts second Workshop Without Walls," NASA Astrobiology Institute, 29 November 2010 (http://astrobiology.nasa.gov/articles/nai-hosts-second-workshop-without-walls/).

No Mystery About the Polio Outbreak

THE POLIO OUTBREAK IN THE REPUBLIC OF THE Congo ("Polio outbreak breaks the rules," L. Roberts, News of the Week, 24 December 2010, p. 1730) is a tragic example of the so-called "virgin soil" epidemics, described by Macfarlane Burnet (*1*), that occurred in isolated regions such as St. Helena, Guam, New Guinea, Samoa, and the Hudson Bay Arctic, where the population had not been exposed to the poliovirus early in life and had not acquired any immunity. These epidemics were characterized by severe paralysis and a high mortality. For example, the Eskimos suffered a paralysis rate of 40% (although there were no paralyzed cases in infants under 3 years of age) and the mortality was 14%. The disease's effect on males between 15 and 25 in the Congo was not an "unusual twist." It was in fact predicted

CORRECTIONS AND

Letters: "Hopes for merit-based grants in China" by Q. Wang (2 reference 1 was listed as X. Heping. The correct name is H. Xie.

News of the Week: "From *Science*'s online daily news site" (14 Now story "First Earth-sized exoplanet discovered" states that th in fact 1.6 times that of Earth.

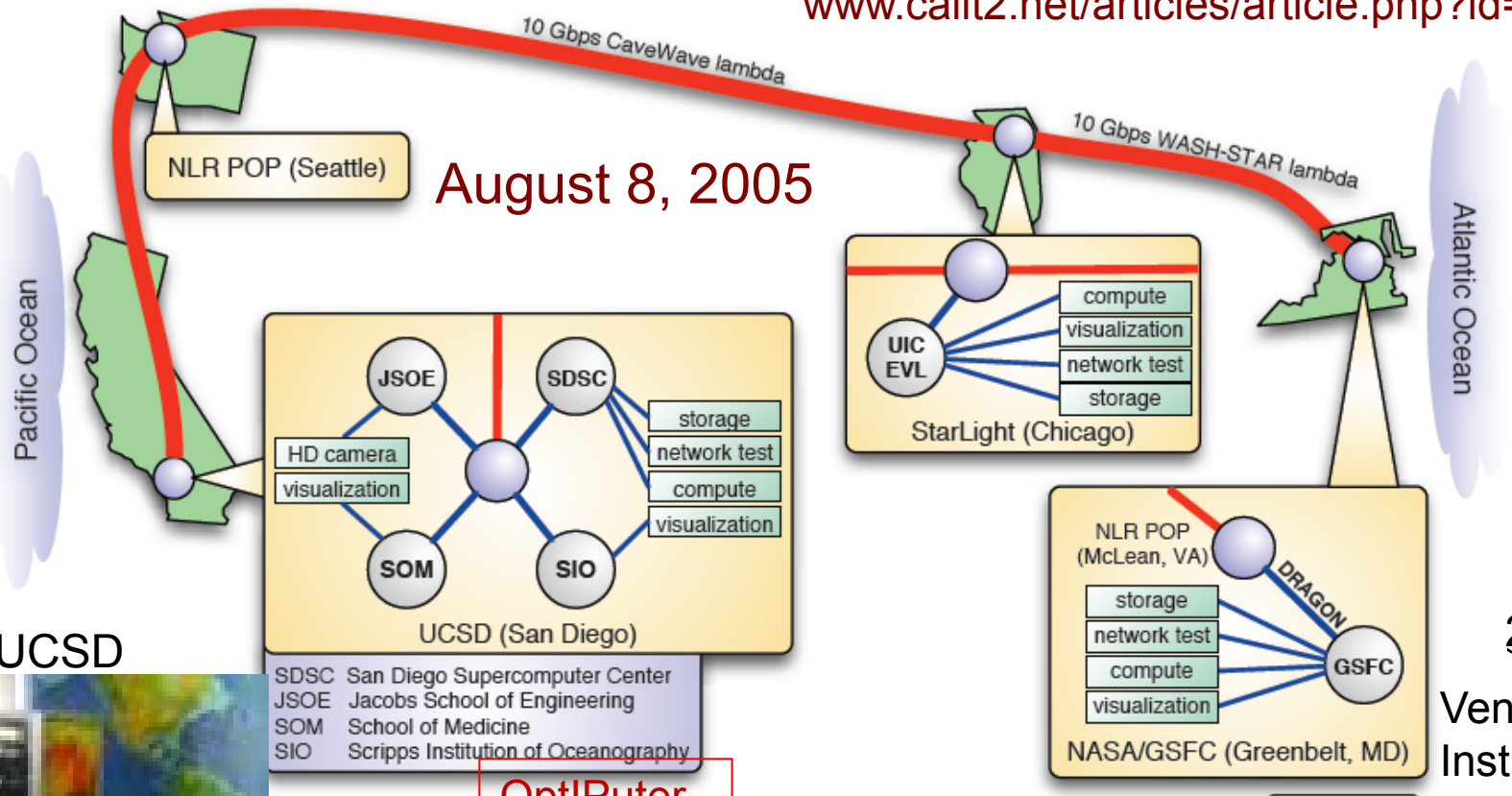
Reports: "A basal dinosaur from the dawn of the dinosaur era in ary, p. 206). The phylogenetic definitions for three taxa (Dinosa as cited (www.taxonsearch.org) and as follows: Dinosauria, the l 1889) and *Passer domesticus* (Linnaeus 1758); Ornithischia, the (Marsh 1889) but not *Passer domesticus* (Linnaeus 1758) or *Sal* rischia, the most inclusive clade containing *Passer domesticus* (L and including *Saltasaurus loricatus* (Bonaparte and Powell 198

Editors' Choice: "Sensing worms" by B. Hanson (17 December should have been "von Neumann."

Google+ Video Hangouts Are Now Part of Your Web Browser!

Extending Telepresence with Remote Interactive Analysis of Data Over NLR

www.calit2.net/articles/article.php?id=660



SIO/UCSD

25 Miles

Venter Institute

OptIPuter Visualized Data

HDTV Over Lambda

NASA Goddard

Remote Control of Scientific Instruments: Live Session with JPL and Mars Rover from Calit2

Source: Falko Kuester, Calit2; Michael Sims, NASA

High Definition Video Connected OptIPortals: Virtual Working Spaces for Data Intensive Research

NASA Supports Two Virtual Institutes

LifeSize HD

Calit2@UCSD 10Gbps Link to
NASA Ames Lunar Science Institute, Mountain View, CA

Source: Falko Kuester, Kai Doerr Calit2;
Michael Sims, Larry Edwards, Estelle Dodson NASA

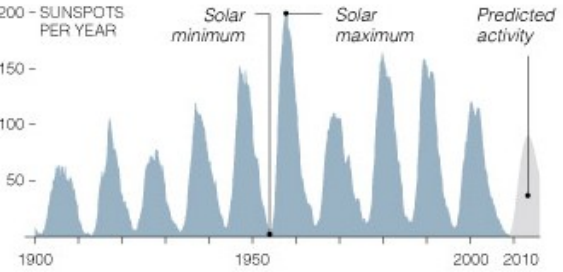
The Rise of Data-Intensive Computing

SCIENCE ILLUSTRATED | # A Mission to Study the Sun

Scheduled for launching this week, the Solar Dynamics Observatory is the first mission in NASA's Living With a Star program. The craft will observe the Sun in detail, studying how its magnetic field is generated and measuring its extreme ultraviolet radiation. This information will help scientists predict solar activity, which can affect Earth's climate, knock out power grids and disrupt communications.

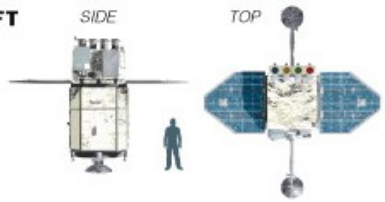
THE SOLAR CYCLE

Solar activity occurs in waves of roughly 11 years. The current lull has been unexpectedly long.



THE SPACECRAFT

S.D.O. will be placed in a geosynchronous orbit. The craft weighs 3,440 pounds and is 14.8 feet tall.



SOLAR PANELS
Produce 1,450 watts of power. Tapered shape avoids blocking antennas.

STAR TRACKER

PROPULSION SYSTEM
S.D.O. carries 3,117 pounds of fuel, enough to keep it in orbit up to 10 years.

THRUSTERS

HIGH-GAIN ANTENNAS
Two antennas transmit 18 megabytes of data per second

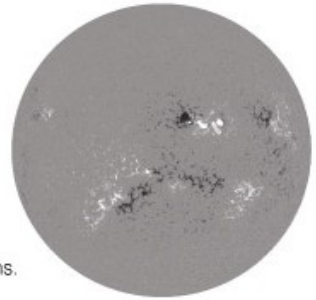
S.D.O.'s SCIENCE INSTRUMENTS

Extreme Ultraviolet Variability Experiment

Will measure variations in the Sun's ultraviolet output, which can fluctuate by factors of 2 to 100 in a matter of minutes.

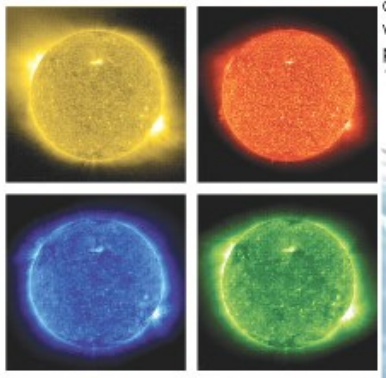
Helioseismic and Magnetic Imager

Will be able to study the flow of material in the Sun's interior by observing waves rippling across the Sun's surface. It will also measure erupting magnetic fields. In the magnetogram at right, black and white areas indicate magnetically intense active regions.



Atmospheric Imaging Assembly

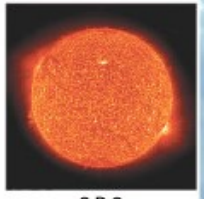
Made up of four multiwavelength telescopes, it will produce high-definition images of the Sun's corona in 8 of 10 selected wavelengths every 10 seconds. Previous spacecraft have been able to study only four wavelengths at different and much slower cadences, as in the images below.



S.D.O. will send back images in higher resolution and at a much higher frame rate than earlier spacecraft, like the Solar and Heliospheric Observatory (SOHO).

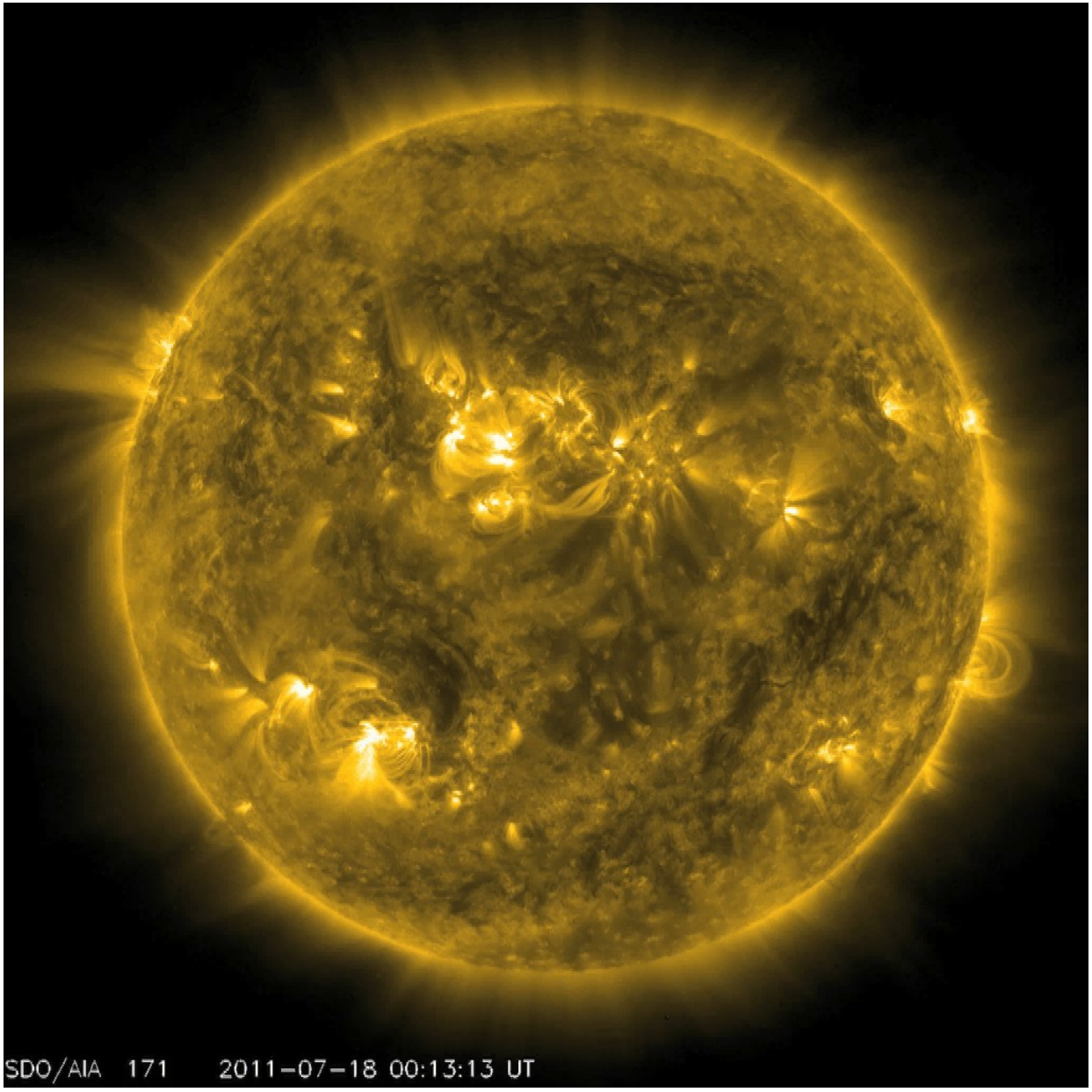
NUMBER OF PIXELS

SOHO
1024 x 1024



S.D.O.
4096 x 4096

SDO/AIA & HMI Last Week 171A + Ic



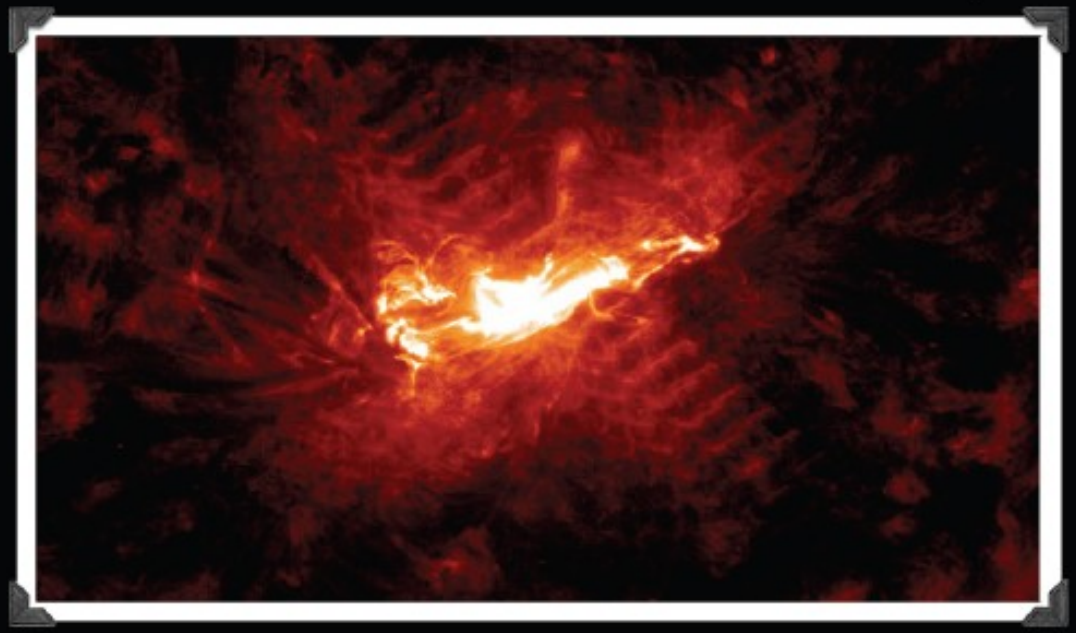
SDO's JSOC Produces Enormous Data Volumes

- **HMI: 2 cameras with 4096x4096**
  - each 3.85s = 46,080 images per day
- **AIA: 4 cameras, 4096x4096**
  - avg 6s each = 57,600 images per day
- **Combined raw data c. 1.4TB/day**
- **JSOC is archiving about 5TB/day including higher level products**
- **Leads to over 1 Petabyte per year**

JSOC-SDP Pipeline Hardware

Data-processing cluster, file- and tape-server machine, and T950 tape library
[COMPLETE as of July 2008]

Compute Intensive Data Processing
NASA's Solar Dynamics Observatory



- Example of a compute-intensive operation on AIA images:
  - The mesh in the entrance filters of the AIA telescopes results in a diffraction pattern consisting of diagonal spikes. This is especially prominent during a solar flare.

Mark Cheung on behalf of the LMSAL science team

GPUs Provide 100x Speedup of Key Task

- GPU implementation
  - CUDA (C + extensions for NVIDIA GPU)
  - NVIDIA's implementation of parallel FFTW
- 25 Richardson-Lucy iterations
  - Using 1 GPU (448 cores): 1 sec
- Speed up of >100x compared to an implementation in IDL (a commercially available data analysis package popular with the astronomy community).

Data-Driven Simulations Are Becoming Major HPC Driver

Radiative MHD Simulation of Active Region (Cheung & Rempel)	Simulation of Convection Zone to Corona (Bart De Pontieu, Juan Martinez-Sykora)	Data-Driven Simulations using HMI Data (Cheung & DeRosa)
C, MPI	**Fortran, MPI**	**Fortran, MPI CUDA/OpenCL**
6-8 Million CPU-hours (NASA + NICS)	2-4 Million CPU-hours (NASA)* *20 Million CPU-hours if we include time allocated for science studies for the IRIS mission	5-8 Million CPU-hours (NASA) Access to GPU-enabled nodes

NASA Cloud Activities

NASA Has East and West Coast Nebula Cloud Services

IT Committee Visited Both and Had Extensive Discussions With NASA Staff

NASA Ames

NASA Goddard

Moving NASA Apps & Workflows to the Cloud

- **Is the cloud deployment model appropriate?**
  - NASA private?
  - NASA community?
  - Commercial?

- **Nebula provides opportunity for a new standard for security**
  - Scientists build prototypes in Nebula
  - Instrumented for security from the beginning
  - Cost savings by baking security in from start

- **What are appropriate migration strategies to move to the cloud?**

- **What is an appropriate research investment to meet the science needs of NASA?**

Science Clouds-a NASA Leadership Opportunity?

- **Commercial clouds are not designed to for big science data**

- **Science needs support for:**
  - Long term permanent archive for data
  - Large data and data-intensive work flows
  - Large data flows into and out of science clouds
  - Integration of data-driven simulations

- **Examples**
  - DOE Magellan Cloud
  - NSF Future Grid
  - Open Cloud Consortium's Open Science Data Cloud
  - NASA SMD actively evaluating Nebula this year

- **Need for design and implementation of a NASA Science Cloud distributed across the Centers**

NASA Marshall's Short-term Prediction Research and Transition Center is Powered by Nebula

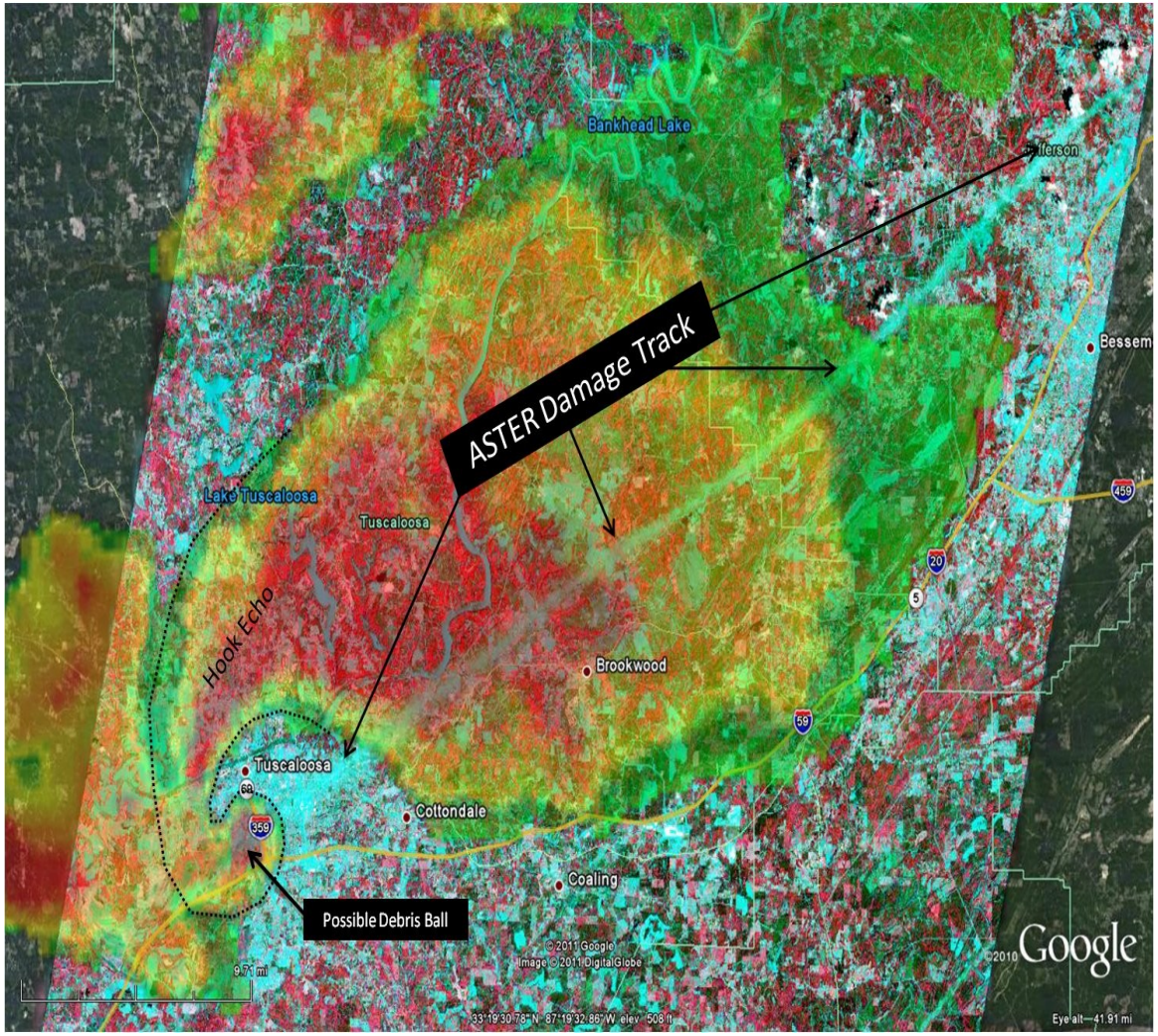
SPoRT used Nebula to process datasets provided to National Weather Service through Google Earth to verify path length and width of tornado combining spectral channels to obtain false color imagery of damages impacting vegetation and ecology

SPoRT, helps to incorporate NASA Earth science research into the National Weather Service forecast operations. The SpoRT Center is managed by Marshall scientists

SPoRT Identified Damage Track from Tuscaloosa Tornado

Advanced Space Borne Thermal Emission and Reflection Radiometer (ASTER)

- Nebula hosted tiling application with large hi-res images
- Rapidly configured
- Tiles created pushed back to local web server
- Made available via Google Earth



That's Fantastic!

NASA SPoRT says they're pleased with Nebula's scalability capabilities….

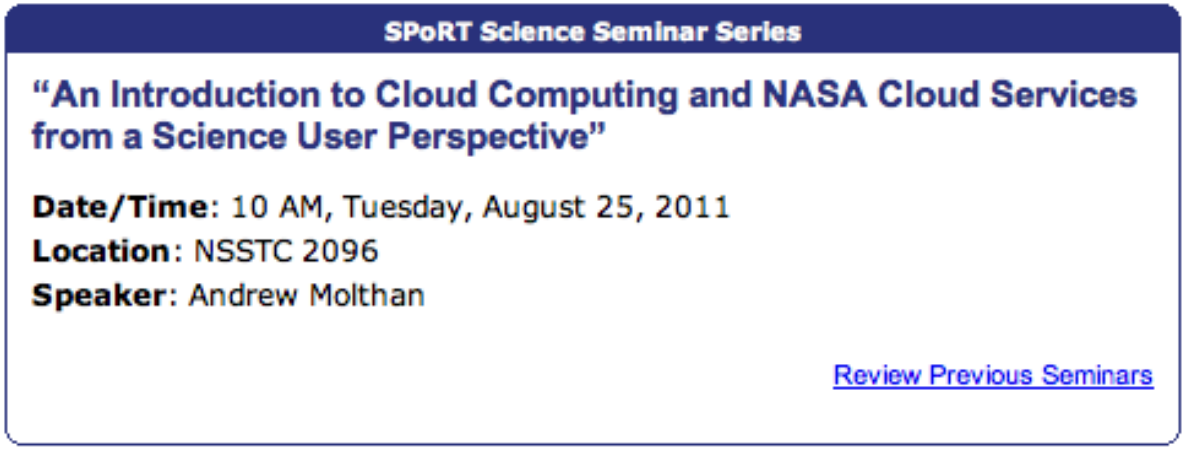
"An earlier test run of my forecast model ran for 54 consecutive days without issue before I brought it down. That's fantastic!"

Andrew Molthan

Senior Meteorologist

Servir and SPoRT

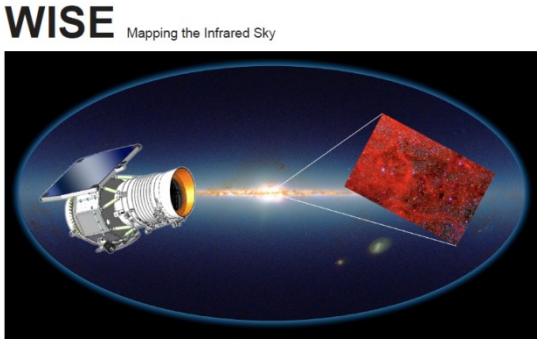
NASA MSFC



WISE Mission Powered by Nebula

- **Wide Field Infrared Survey Explorer (WISE): Images the sky with greater than 8X redundancy**
  - Helping NASA find the most luminous galaxies in the universe and the closest stars to the sun. 2.7 million images captured.
  - Issue: Encountered a short-term need for a large number of small servers and also needed a server with a large memory footprint, did not have access and could not justify cost for his needs alone

- **Nebula Project #1: 2000 distant galaxies**
  - Increase resolution with processing
  - 100 CPU hrs per galaxy
  - We upped instance quota to get started



- **Nebula Project #2: Some sky areas require huge RAM-based processing.**
  - We set up an 80 GB RAM instance
  - Finished first phase on Nebula in two days.

Nebula Value to WISE Project

- Hundreds of galaxies can be hi-res processed to provide higher angular resolution images for better studies of star formation, galactic structure, etc.
- Thousands of galaxies could be hi-res processed to resolve & measure source sizes, etc.
- Hundreds of thousands of galaxies can be hi-res processed to differentiate point-like (active galactic nucleus or nuclear star formation) from distributed (merger or spiral galaxy) emission.
- Several large regions nearby can be surveyed for distributed star formation in our galaxy.
- A few large regions can be processed to tremendous depth.
- The whole sky can be hi-res processed.

Quotes from NASA WISE Project Scientist

"After having access to Nebula for only a few days, I find that I am already able to accomplish more data-intensive calculations than I can do on any of the local servers we have here, and with no difficulties at all!

"The system is easy to access and use, and offers a capability that I absolutely need occasionally but for which I could never justify the expense if it were for my needs alone.

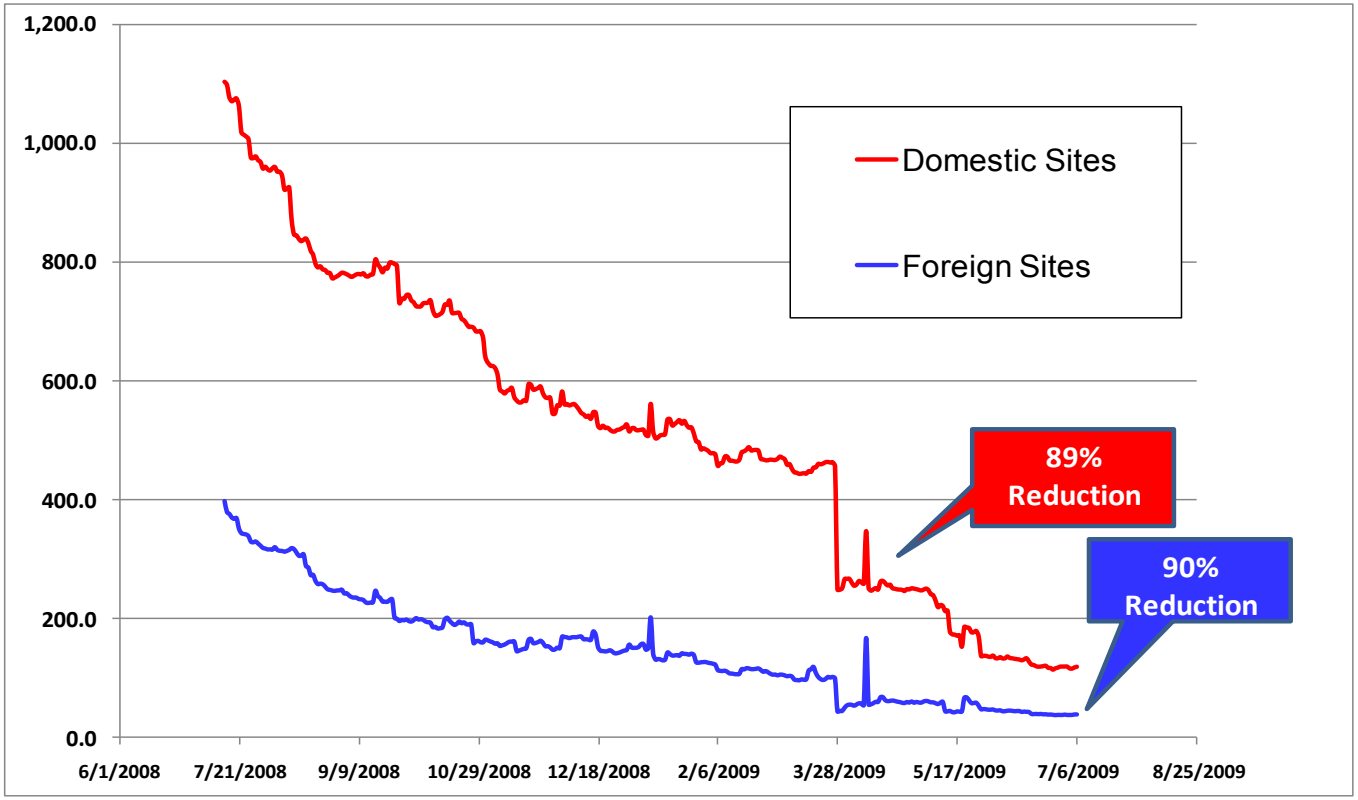
"With the recent addition of a large-RAM instance, I am now able to conduct calculations that could not be done on our project's large server farm. Nebula has provided me with a tool for science data analysis that far surpasses anything that I could envision in a single-user context. NASA Cloud computing may be the way forward for our data-intensive projects in the future, since only a NASA system could provide the necessary reliability and proprietary controls on our data"

Cloud powered by Nebula can help offload from Pleiades; which would allow scientist to use Pleiades at it's maximum potential.

Cyber Security

Cybersecurity at NASA

- Huge opportunity for improvement
- Goddard and Ames interested in leading by example
- Need to separate mission, science, & administrative computing
- Learning from other agencies to lower risk



ITIC Projected Work Plan for FY 2012

ITIC Work Plan for FY 2012

1. Examine the ongoing and planned efforts for the IT Infrastructure and mission areas.

    a. *Examine best practices across the government and the private sector and suggest approaches that are appropriate for NASA's unique mission.*

    b. *Investigate the state of NASA's software and infrastructure support for collaborative teams. What collaborative systems and services should NASA make available to enhance distributed teams?*

    c. *Recommend areas of disruptive technology where NASA needs to be prototyping innovative systems which may require major changes in established NASA IT procedures, processes and acquisitions.*

.

ITIC Projected Work Plan for 2011 (cont'd)

2. Investigate the state of NASA's high performance networks, computing and visualization systems, as well as data intensive computing and storage systems. How do these compare to other federal agencies and other large scale efforts, including commercial facilities? Recommend opportunities for NASA to use leading edge capabilities in these areas.

3. Examine NASA's data and communications environment for its aerospace operations and point out areas in need of attention in order to improve efficiency and reliability for all users end to end.

ITIC Projected Work Plan for 2011 (cont'd)

4. Examine the role of the OCIO, its strategic plans and projected resources, and IT governance across NASA. Make recommendations on how IT infrastructure should be managed at NASA. Identify deficiencies and opportunities for improvement and innovation.