

Orientation to NCAR, CISL and the Outreach Services Group

Dr. Richard Loft

Director, SIParCS Program

Director, Technology Development Div.
Computational and Information Systems

Laboratory

loft@ucar.edu

13 March 2013



TDD: Technical challenges and Imperatives

- **Challenges**

- Declining Budget
- Rapidly evolving technologies

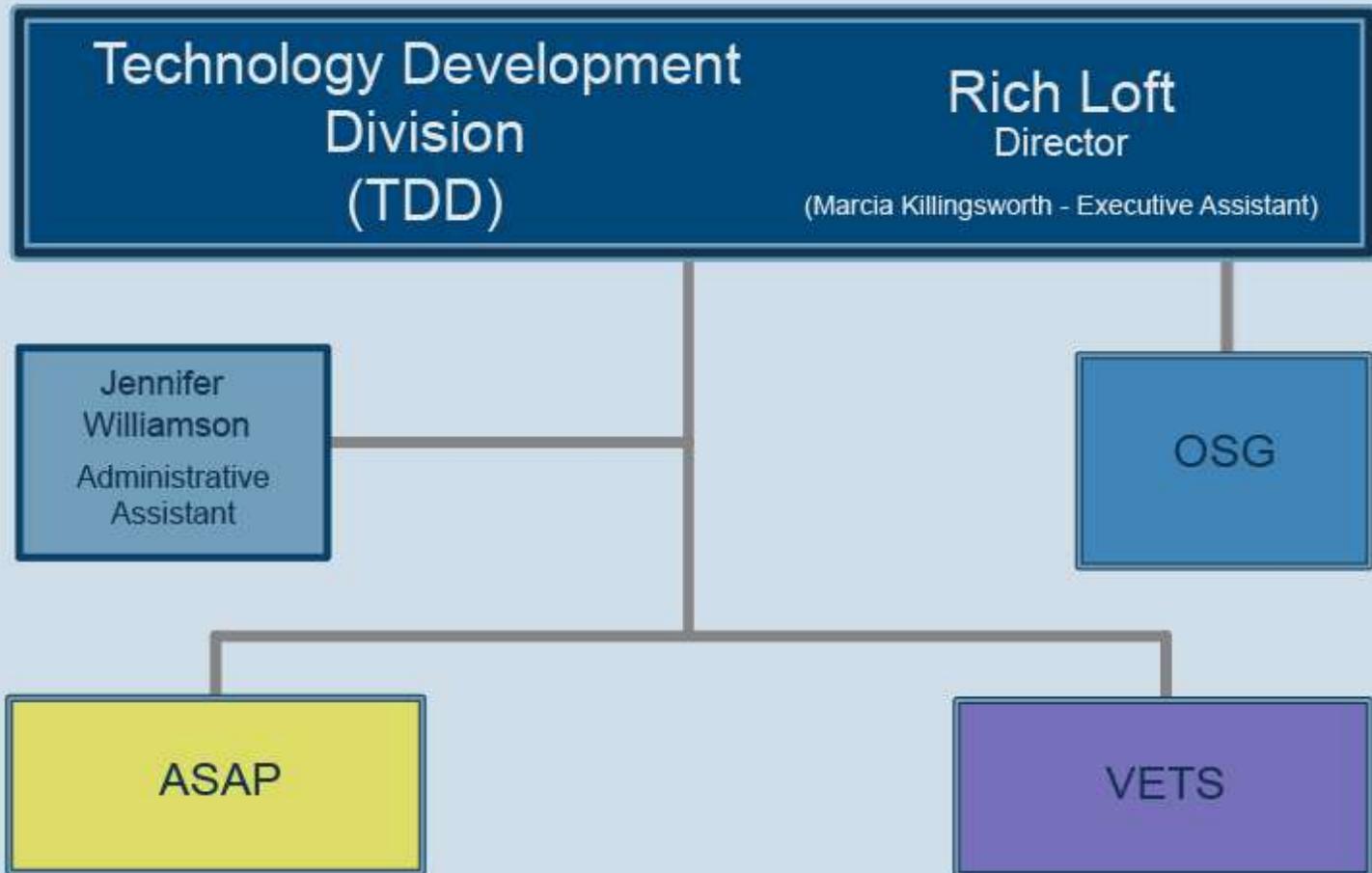
- **Imperatives**

- Must continue to innovate in tools and applications
- Must evolve services, not abandon them

- **Focus areas**

- Application performance
- Big data challenges: volume, variety, velocity
- Value add to organization core missions

TDD Organization



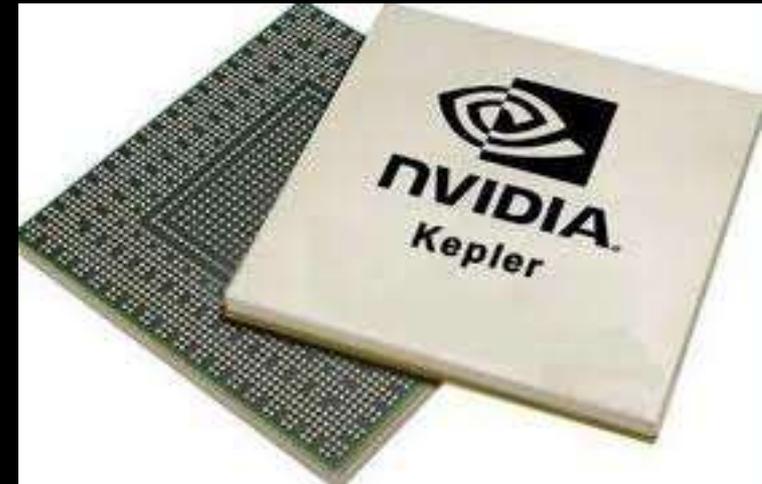
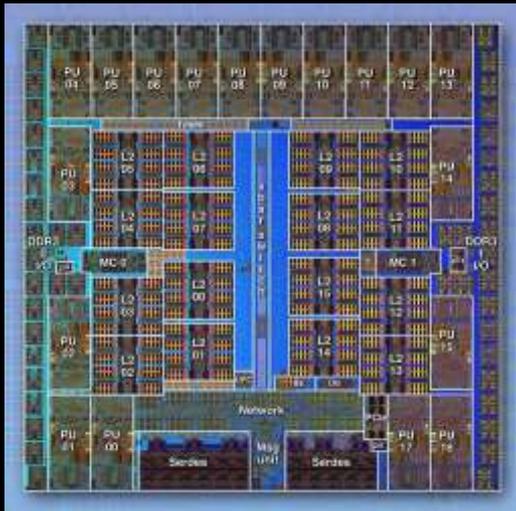
Mission of the Application Scalability and Performance Group

- Application performance
 - Characterization
 - Optimization
 - Portability
- System performance
 - Yellowstone
 - Many-core architectures
 - Exascale technology
- Data-centric computing
 - Parallel I/O
 - Analytics
 - Data compression



**Ben, Srinath, John (head),
Youngsung, Allison, Jennifer (AA)**

The current many-core architectures under evaluation in ASAP



IBM BG/Q

Cores: **16 + 2**

Multithread: **4-way**

Coprocessor: no

Boot Linux: **yes**

Intel Knights Corner

Cores: **61**

Multithread: **4-way**

Coprocessor: yes

Boot Linux: **yes**

NVIDIA Fermi->Kepler

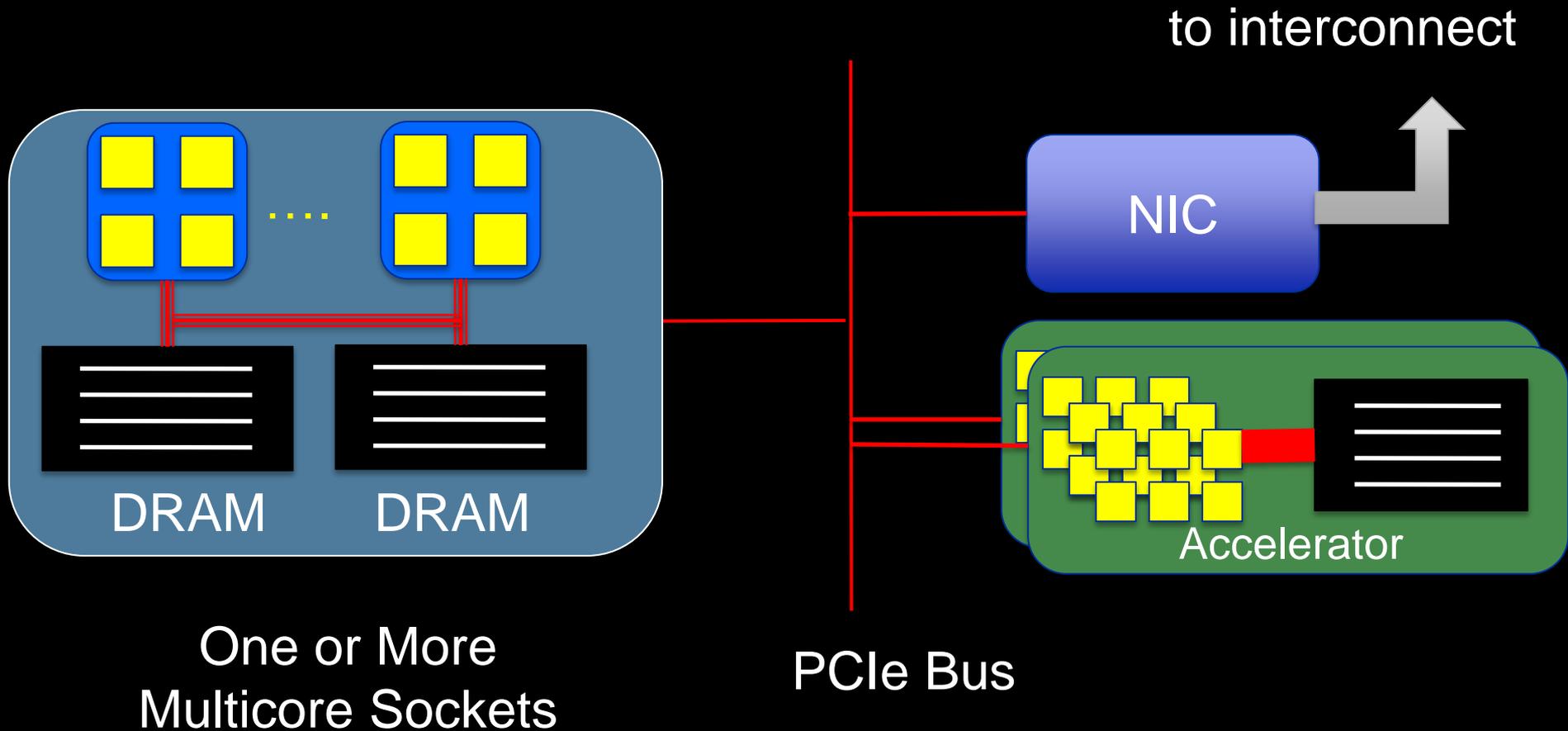
DP Cores: **512->832**

Multithread: **32-way**

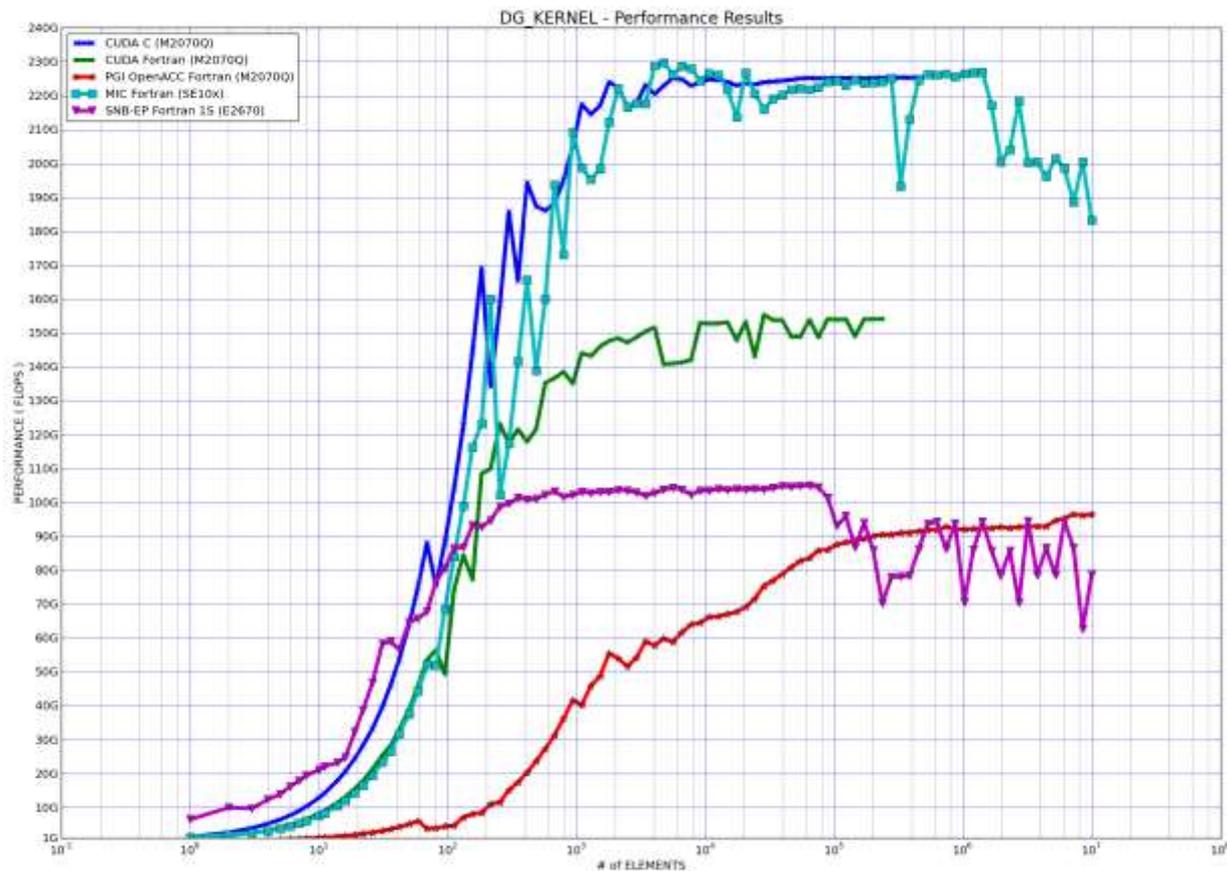
Coprocessor: yes

Boot Linux: **no** ⁵

Heterogeneous, many-core architecture



Kernel paradigm: Many-core performance



Big-Data Frontiers

- Parallel Workflows Environments
 - Workflow systems (Swift)
 - Parallel analysis tools (RMPI, MatLab, NCL)
 - On supercomputer analysis
- Optimized Analysis Architectures
 - New storage technologies: NVRAM, SSD
 - Map reduce cluster vs central file system
- Algorithms
 - Statistics
 - Data mining heuristics
 - Data Compression

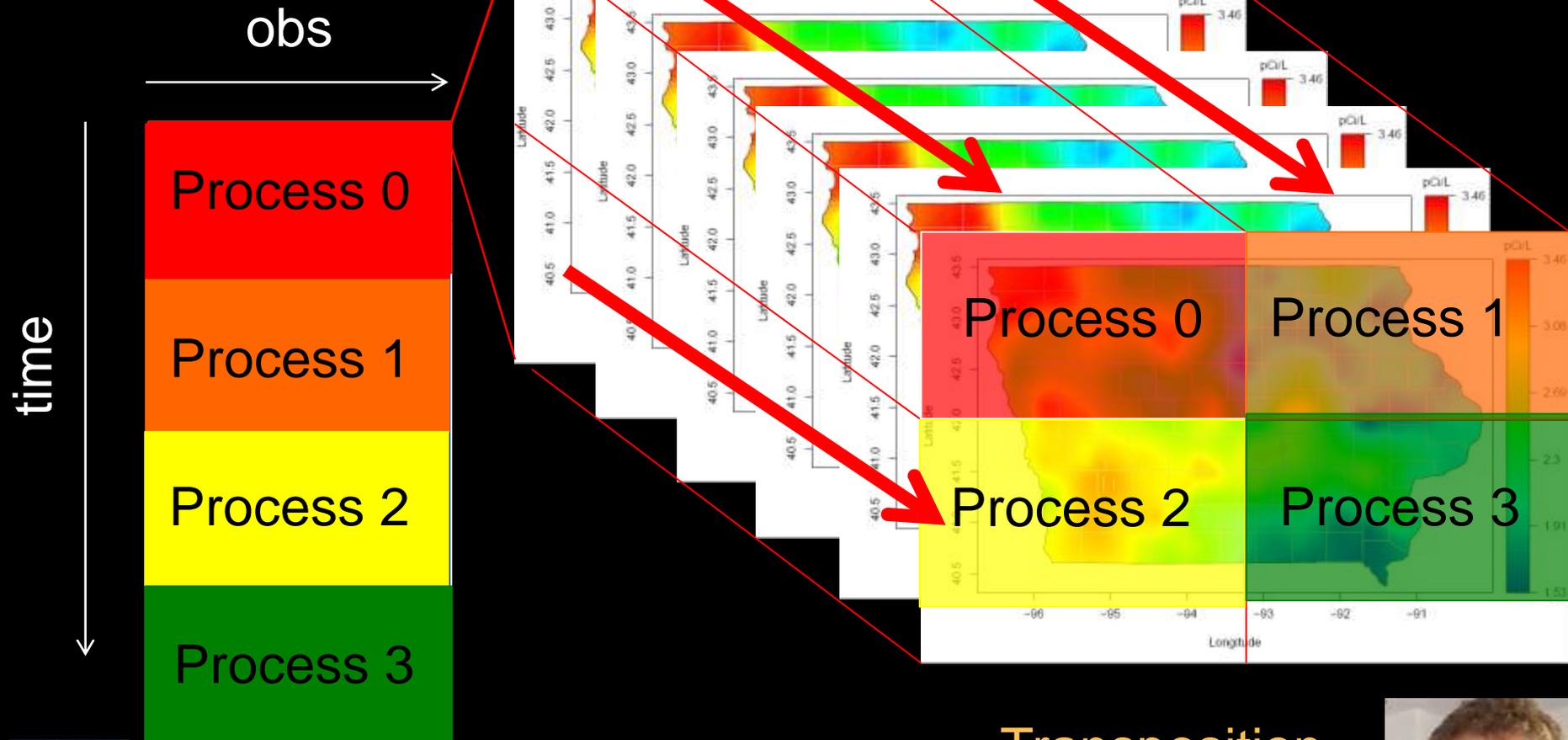
**What if atmospheric data
analysis were as fast as
“googling”?**



Heatwave workflow in RMPI

Scan Time Series for $T > T_{max}$

Kriging Interpolation

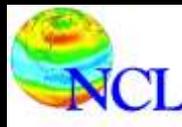


Seth McGinnis, IMAGE

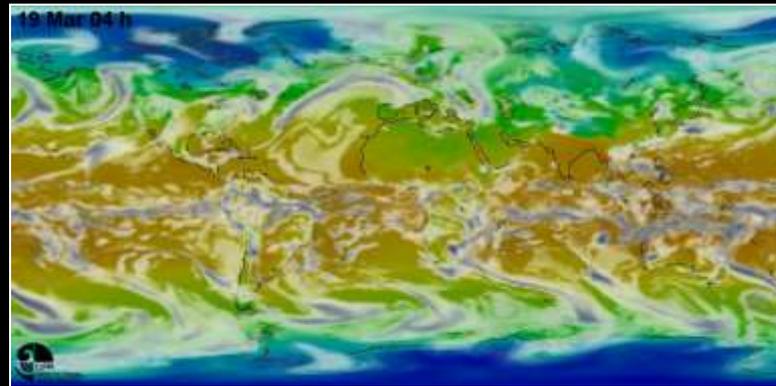
Jim Gattiker, visiting scientist

**Visualization and Enabling
Technologies Section
(VETS)**

NCAR Command Language



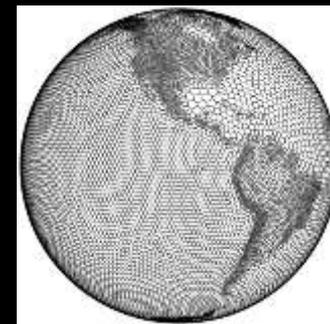
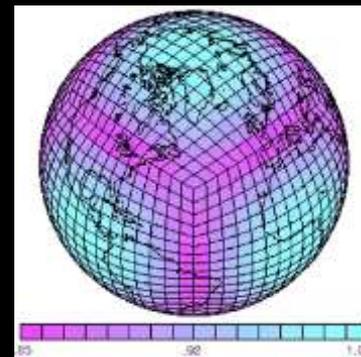
- Adopted by research centers and universities world-wide including MétéoSuisse, Max-Planck-Institut für Meteorologie, Instituto de Pesquisas Meteorológicas, WeatherUnderground, International Arctic Research Center
- Four releases in FY2013: new color display model, better support for NetCDF-4 and HDF-5 data features, ESMF regridding capabilities, oceanography functions
- Upcoming releases to include faster contouring algorithm, “quick look” graphical tool, command-line regridded, new CESM/WRF computational functions, integration of R statistical analyses
- Metrics: 17627 registered users in 160+ countries, 1700 average downloads a month, 2004 users on email list, 300+ emails a month



ParVis



- Big-data, parallel, open source analysis library and DAV applications for ultra-large and complex climate datasets
- 3-year project funded by the U.S. Dept. of Energy Office of Science, Biological and Environmental Research – led by ANL (Rob Jacob), incl. NCAR, PNNL, Sandia, and UC Davis
- ParGal: Parallel Gridded Analysis Library
- ParNCL: Compute and I/O parallel NCL
- Second beta release in April 2013



MOAB

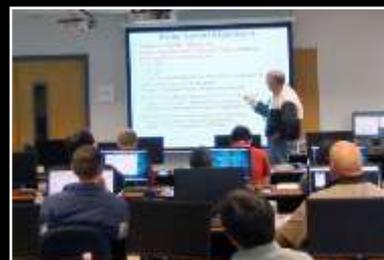
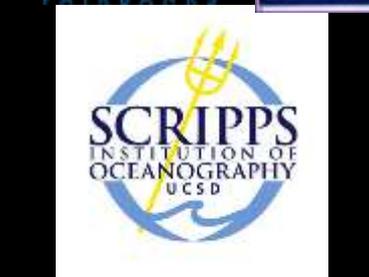


PNetCDF

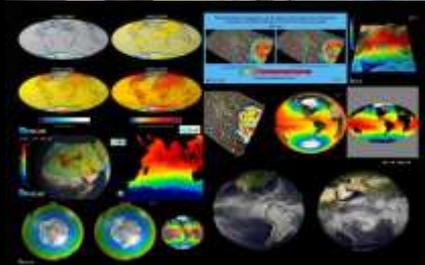


NCL Data Analysis and Visualization Workshops

- Primarily for climate and weather researchers
- 3.5 days with lectures and intensive hands-on lab sessions
- Co-taught by NESL/CISL scientific and engineering staff
- Students encouraged to BYOD (Bring Your Own Data)
- Free for all students; full funding provided for EPSCoR and MSI students
- Workshops given locally and at universities and research institutions world-wide
- **65 workshops taught to 975 students** from February 2000 to March 2013
- Workshops in FY2013:
 - Completed: CERFACS in Toulouse (2 workshops), ARSC in Fairbanks, NCAR
 - Upcoming: University of Albany (32 students), NCAR (16 students)
 - Tentative: Harvard, Naval Postgraduate School, UCLA



Vislab Highlights



- Continued productive partnerships with UCAR E&O, Government Affairs, and others to support outreach and exposure of NCAR science to government, corporate, university, scientific, and K-12 visitors.
- Supported over **2700 local and remote users** and visitors in FY2012, including US Senator Michael Bennet.
- Produced over **100 animations and graphics** for presentations by scientific staff at seminars and geoscientific conferences such as AGU and briefings to NSF. Examples of this material are currently in use online by the World Climate Research Programme, NASA, SCEC, and NESL, to name a few.
- Provided technical and physical support for the 2012 Supercomputing Conference in Salt Lake City, UT. Produced an exhibit video loop of CISL highlights and ASD visualizations.
- Provided AV support and venue for NBC and BBC interviews of NCAR scientific staff.
- Provided collaborative technology support for CISL training courses offered to local and remote users.
- Provided digital postproduction support for Hurricane Sandy, the High Park Wildfire, and NWS time lapse construction visualizations.

VETS: Science Gateways Infrastructure

- **Virtualization**

- Almost all data system moved to virtual machines
- Only databases left to virtualization, expected shortly

- **NWSC Transition**

- All data moved to NWSC in May
- VM Migration almost complete
 - 3 VMs left to transfer
 - ESG and CDP production systems running at NWSC

Data-centric science gateways

- Find the data
- Get the data
- Browse the data
- Publish data
 - Describe the data
 - Label the data
- Preserve the data



Think “digital library”

ESG-NCAR Science Gateway

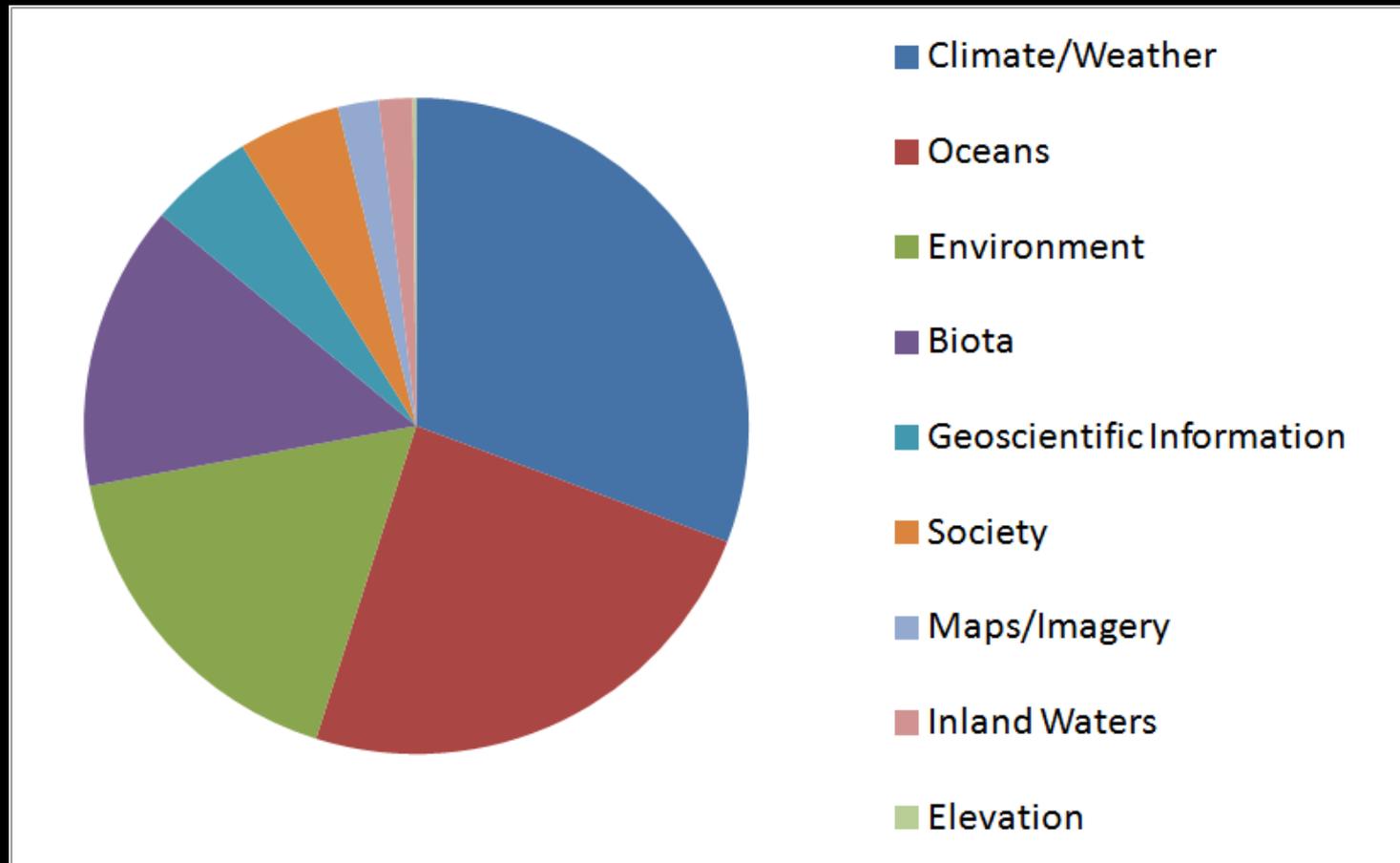
- 33,212 registered users, ~600 new user registrations each month.
- Over 1200 U.S. and international visitors monthly.
- Primary access to CESM/CCSM4, CMIP5, NARCCAP datasets.
- CESM data administered as a collaboration with CGD/NESL .
- Total download volume exceeds **110 TB monthly**.
- Total managed holdings: **2.1PB, 3.2M files**.
- Over **750 TB** of new data products published in past 12 months.
- New download metrics system deployed December, 2012.

Long Tail Problem Example: Advanced Cooperative Arctic Data Information Service (ACADIS)

- <http://www.aoncadis.org>
- Over **100** principal investigators publishing data.
- Current has about **1000** diverse data collections.



ACADIS: 1000 Diverse Data Sets

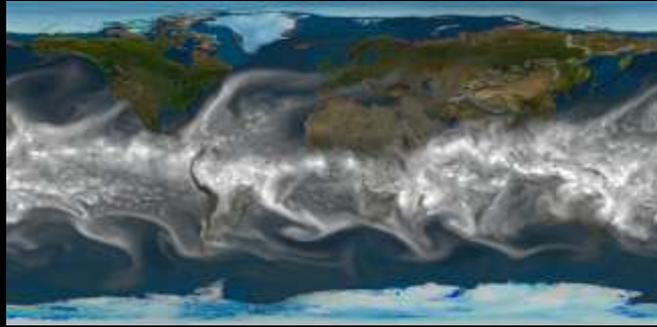


“Only” 2 Terabytes of data...

CISL Outreach mission: building communities



Public Engagement



Communicating
Research



Engaging
Community Colleges



Developing Young
Scientists



Enhancing Diversity



K-12 STEM
Education

CISL Outreach Organization



Stephanie



Brian



Steve



Marijke



Kristin

Summer Internships in Parallel Computational Science:

Students work in NCAR's Supercomputing Lab
with mentors on challenging R&D projects



SIParCS Class of 2011

- **11-week Summer internship program**
 - May – August
- **Open to:**
 - Upper division undergrads
 - Graduate students
- **In Disciplines such as:**
 - Computer Science and Software Engineering
 - Mechanical Engineering
 - Applied Math and Statistics
 - Earth System Science
- **Support:**
 - Travel, Housing, Per diem
 - 11 weeks salary
 - Conference travel and Publication costs
- **Number of interns selected:**
 - Typically ~15

**For more information go to:
<http://www.cisl.ucar.edu/siparcs>**

SIParCS Class 2013:

11 interns

Last Name	First Name	Project Title	Mentor	Institution	MSI	Academic Level
Abouali	Mohammad	Visualization of Weather Simulation Data in Google Earth (TM)	Norton	San Diego State		Grad
Cullen	Allison	Validation of Urban Heat Island Simulations	Sain	Wayne State		Grad
Fisk	Sean	Comparison of PyPy with other Scientific Computer Languages	DeIVento	Grand Valley State Univ (MI)		Grad
Grossman	Matthew	GPU Accelerated Web map Services	Brownrigg	Middlebury College		Undergrad
Guo	Ping	GPU Processing in NCL for Data Analysis and Visualization	Huang	UWyo		Grad
Kruse	Chris	WRF Performance and Scaling Assessment	DeIVento	UWyo		Grad
Liu	Zhengyang	Evaluating the Interaction between Network Routing and Application Message Passing Traffic	Dennis	U of VA		Grad
Manna	Soumi	Evaluating the Performance of the Community Atmosphere Model at High Resolutions	Jamroz	UWyo		Grad
Mizero	Fabrice	Evaluating the Interaction between Network Routing and Application Message Passing Traffic	Dennis	Philander Smith College	x	Undergrad
Olarinde	Ademola	Optimizing the Efficiency of the NCAR-Wyoming Supercomputing Facility - Grad	Andersen	Texas A&M Kingsville	x	Grad
Usset	Joseph	Determining Change Points in Balloon Based Measurements of the Atmosphere	Nychka	NC State		Grad

NWSC Visitor Center

...use many tools to help them, including
special instruments to collect information,
thermometers and satellites. They also use
powerful calculating machines called
supercomputers to figure out what those
observations mean.

The more observations we take, and the
more powerful our supercomputers, the more
discoveries we can make. And sometimes, new
clues show us that we were in the wrong track
and that we have a lot more research to do before
we understand the real story!



NWSC

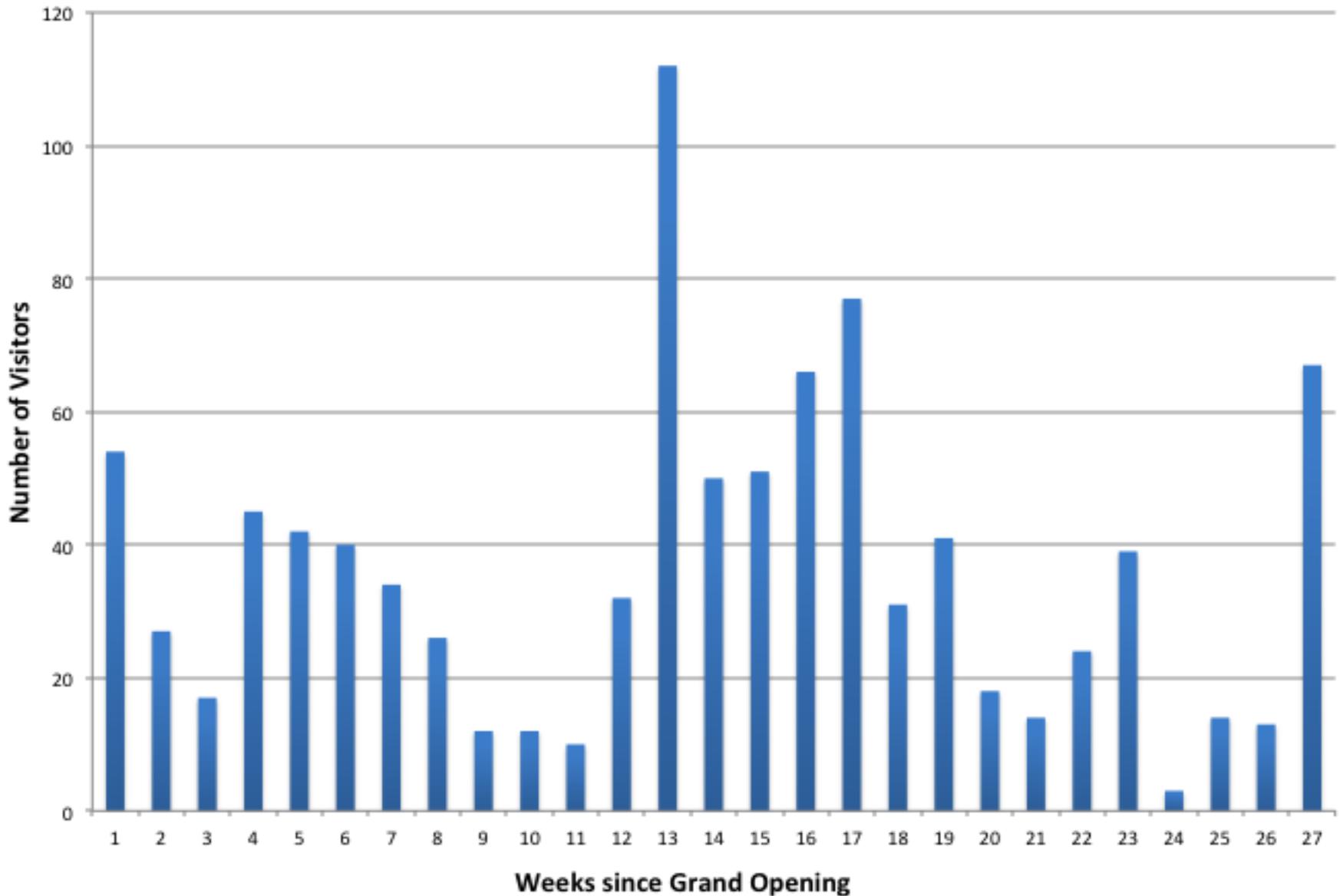
Weather supercomputers are used to help us
understand the weather. They are used to
simulate the atmosphere and predict the
weather. They are used to help us understand
the weather and to help us predict the
weather.

NWSC

The National Weather Service is a
part of the National Oceanic and Atmospheric
Administration. It is responsible for providing
weather, climate, and hydrologic information
to the public.



Visitor Statistics at NWSC





Thank You

Rich Loft
loft@ucar.edu