BLUE WATERS

SUSTAINED PETASCALE COMPUTING

November 14, 2018

Best Practices and Lessons from Deploying and Operating a Sustained-Petascale System: The Blue Waters Experience

Gregory Bauer, Brett Bode, Jeremy Enos, William Kramer,
Scott Lathrop, Celso Mendes, Rob Sisneros

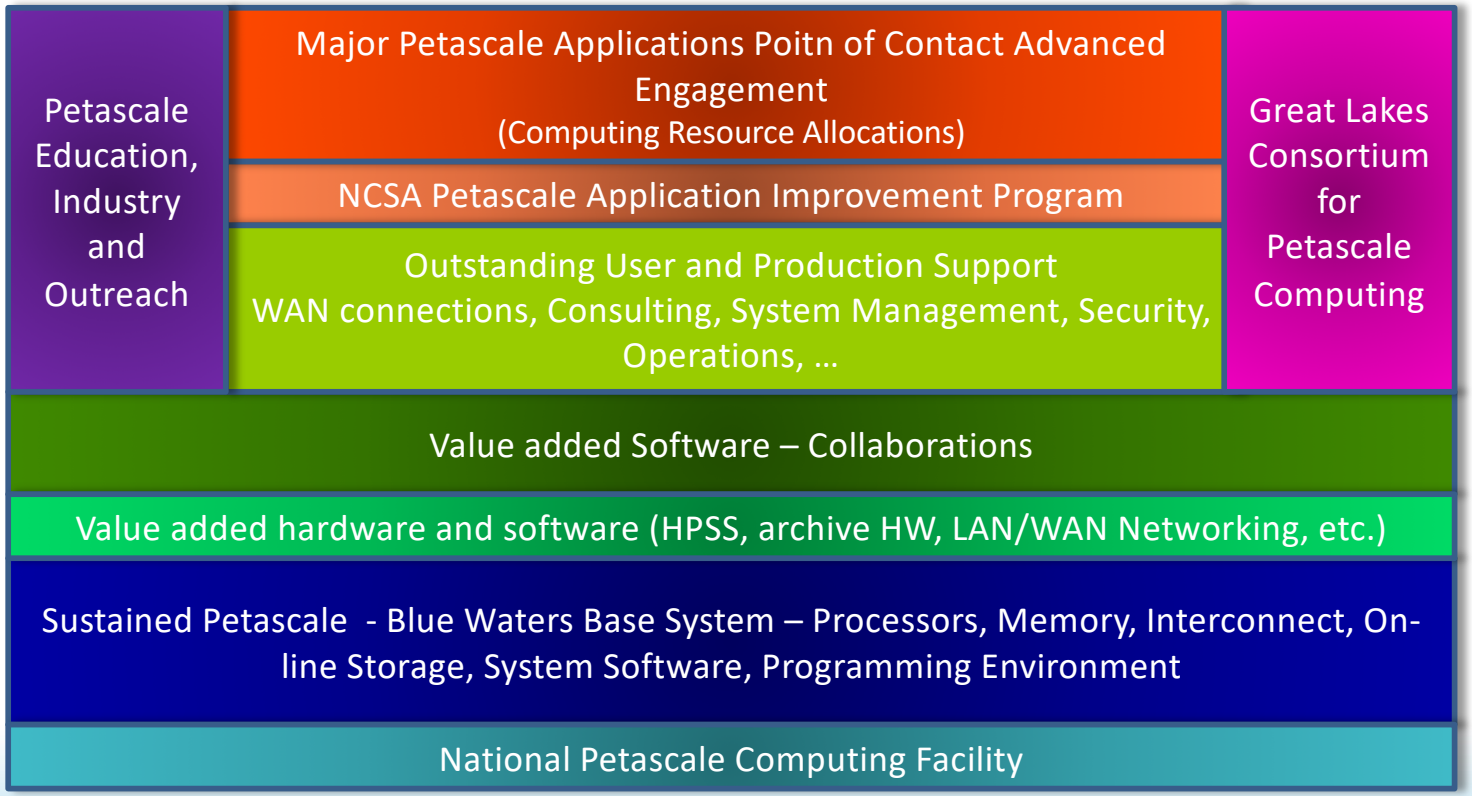
NCSA, University of Illinois

GREAT LAKES CONSORTIUM
FOR PETASCALE COMPUTATION

Introduction

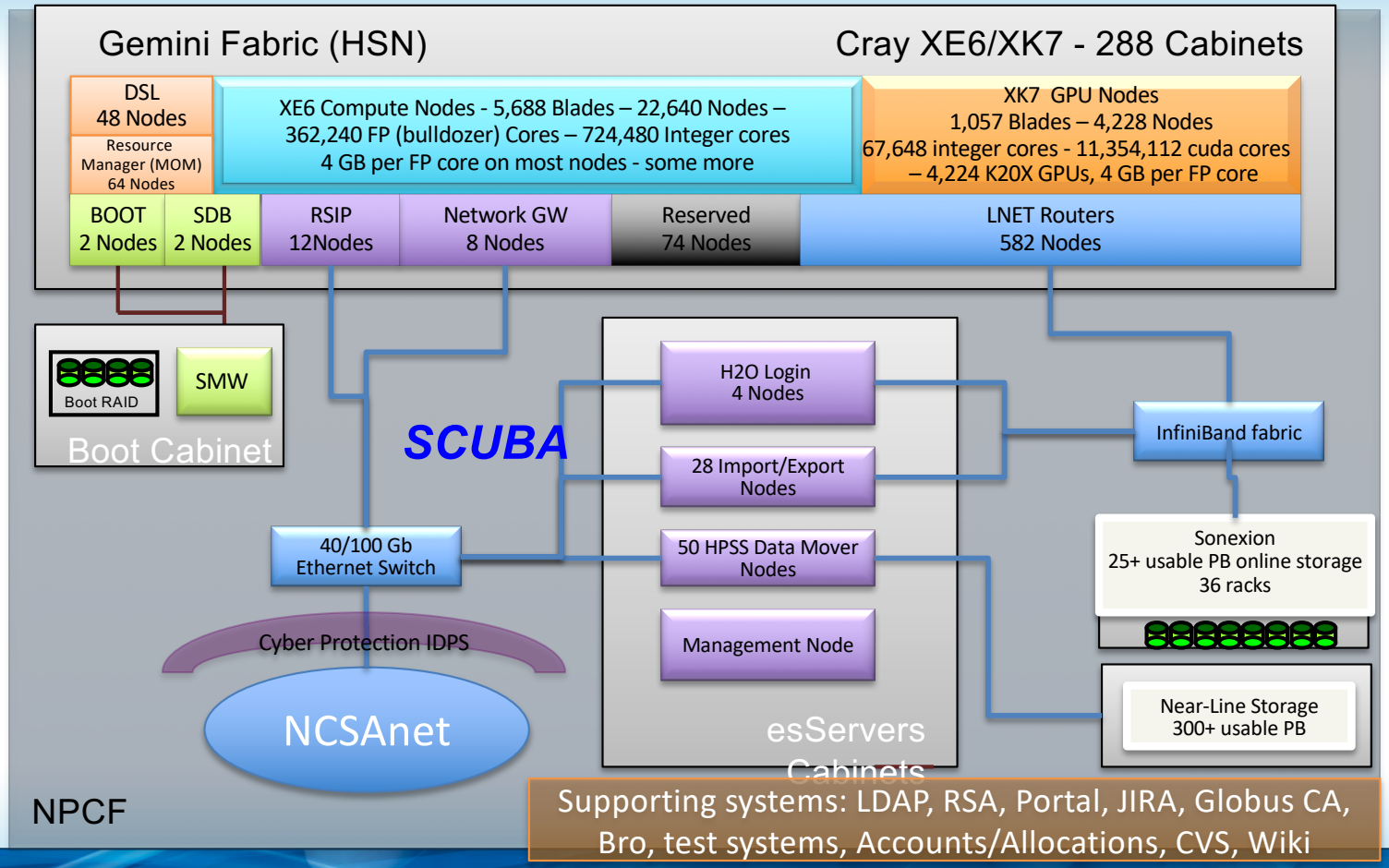
- Blue Waters Project
  - System deployment, operations, adjustments
  - User support, documentation, training, outreach, etc
  - Best practices adopted in all project fronts
- Goals of this paper:
  - Document Blue Waters' best practices.
  - Share Blue Waters' lessons learned.
  - Provide inspiration for future deployments.
  - Contribute to community advancement.
  - Improve the (scarce) literature in the area.

Blue Waters is a Multi-Focused Program



The Blue Waters Project

- Comprehensive development, deployment and service phases with co-design and other aspects
- Blue Waters is a **Universal Research Instrument** for computational and data science that has delivered over 23 Billion core hour equivalents
- Diverse Science teams are able to make excellent use of those capabilities due to the system's flexibility and emphasis on sustained performance.
- The Blue Waters system is a top ranked system in all aspects of its capabilities. – **Intentionally not listed on Top500 list**
  - 44% larger than any system Cray has ever built. The next closest Cray system was #1 on the Top500.
  - Performance and delivered cycles are approximately the same as the aggregate of all the NSF XSEDE resources.
  - Ranks in the top 15 systems in the world in peak performance – despite being over six years old
  - Largest memory capacity (1.66 PetaBytes) of any HPC system in the world until this year.
  - One of the fastest file systems (>1 TB/s) in the world!
  - Largest nearline tape system (>250 PB) in the world
  - Fastest external network capability (420 Gb/s) of any open science site.

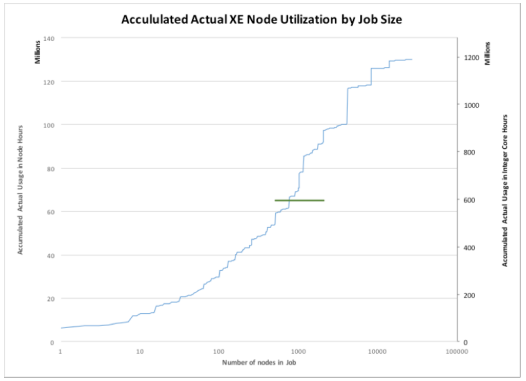

Gemini Fabric (HSN)
Cray XE6/XK7 - 288 Cabinets
DSL
48 Nodes
Resource Manager (MOM)
64 Nodes
XE6 Compute Nodes - 5,688 Blades – 22,640 Nodes –
362,240 FP (bulldozer) Cores – 724,480 Integer cores
4 GB per FP core on most nodes - some more
XK7 GPU Nodes
1,057 Blades – 4,228 Nodes
67,648 integer cores - 11,354,112 cuda cores
– 4,224 K20X GPUs, 4 GB per FP core
BOOT
2 Nodes
SDB
2 Nodes
RSIP
12Nodes
Network GW
8 Nodes
Reserved
74 Nodes
LNET Routers
582 Nodes
Boot RAID
SMW
Boot Cabinet
SCUBA
H2O Login
4 Nodes
28 Import/Export
Nodes
50 HPSS Data Mover
Nodes
Management Node
esServers
Cabinets
InfiniBand fabric
Sonexion
25+ usable PB online storage
36 racks
Near-Line Storage
300+ usable PB
40/100 Gb
Ethernet Switch
Cyber Protection IDPS
NCSAnet
NPCF
Supporting systems: LDAP, RSA, Portal, JIRA, Globus CA,
Bro, test systems, Accounts/Allocations, CVS, Wiki

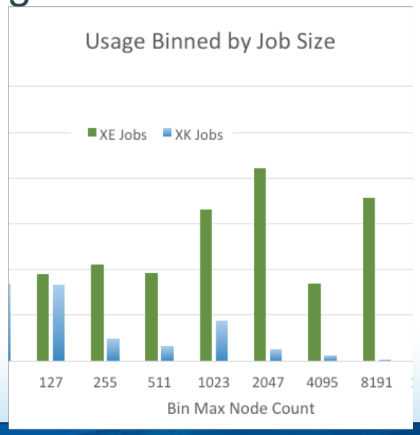
Project Mission

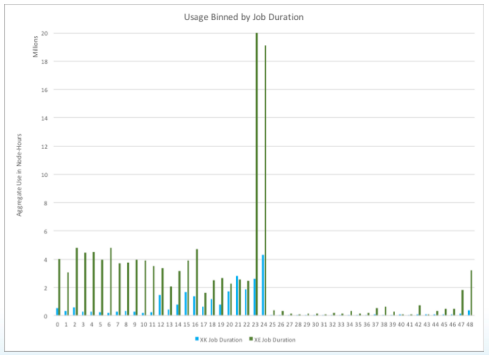
- A clear project mission is critical
  - Who are you serving?
  - What are the characteristics of the science you are supporting?
  - What behaviors do you want to encourage?
- Blue Waters -
  - Mission is to support NSF open-science projects that can not be done anywhere else.
  - Emphasis on large-scale workloads – Both large job and ensemble based.
  - Flexibility throughout the system.

Focus on Large Problems and Jobs

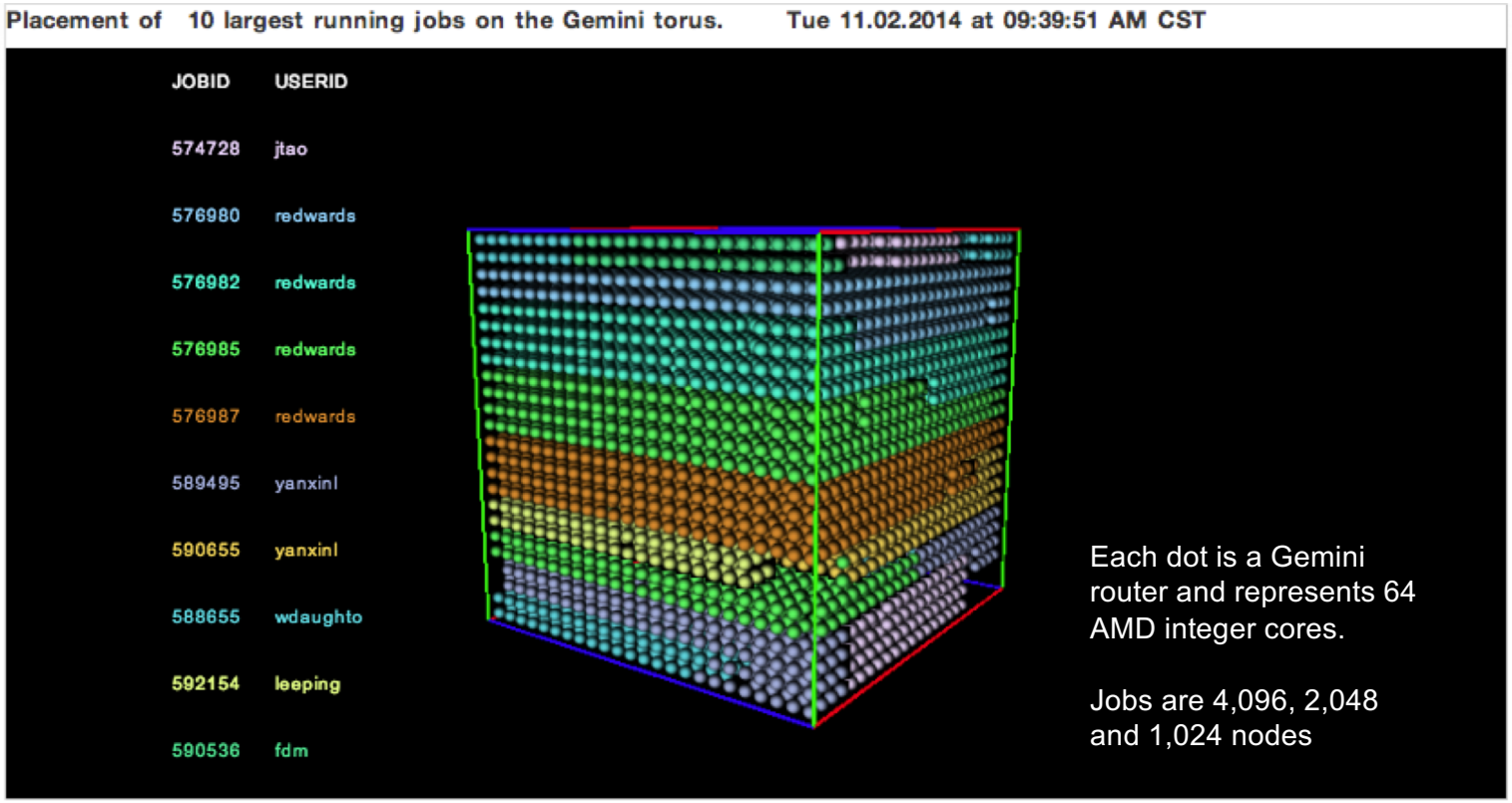
- Median submitted job size is about 1,000 nodes (32,000 core equivalents) May 2015- June 2016
- Median in first year was about double that.
- What is the difference? A lot of convergence of Extreme Computing and Big Data efforts.







Blue Waters Usage 2/11/14 – A Good Day Automatically Largest 10 Jobs-Torus View - 99.82% of nodes in use



Project Management

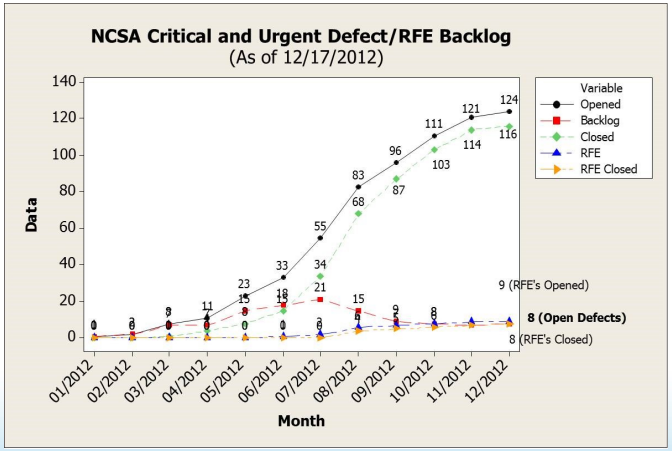
- Allocation of project personnel in *teams*
  - Project management, sys-admin, application support, storage, networking, security, facilities, education/outreach, industry, public affairs
  - Every team meets weekly, remote participation enabled.
  - Internal Wiki for all project documents
    - Includes section with vendor access to enable information sharing.

Project Management

- Help Request System - aka Ticket System
  - All requests documented in the ticket system regardless of how the request is made – email, phone, web, chat or in person.
  - Monitored by members from all project areas – all staff can interact with any ticket.
  - Metrics defined for ticket response and reviewed regularly as well as reported to funder.
    - Time to initial human response and time to close.
    - Automated tool to evaluate metrics

Project Management

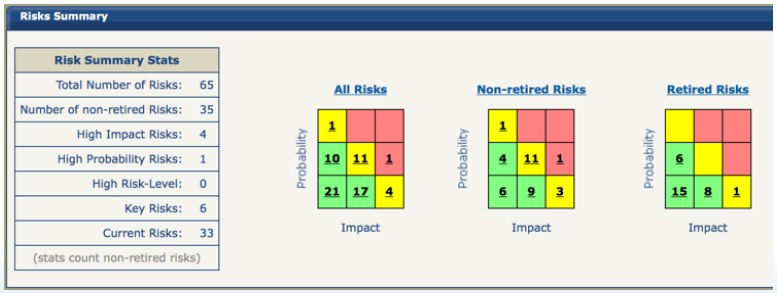
- Strong interaction with vendors
  - Weekly meetings, access to vendor's management
  - Joint effort to handle problems as quickly as possible
  - Tickets that result in a vendor bug include the vendor bug information and when possible are linked to the vendor's bug system.



Number of severe bugs filed with Cray in 2012

Risk and Change Management

- Management of project risks
  - Tracked with an internally-developed risk register tool.
    - Freely available from: https://wiki.ncsa.illinois.edu/display/ITS/NCSA+Risk+Register
  - A risk is assigned a probability and impact, mitigation plans are defined.
  - Risks are reviewed monthly and retired if no longer needed.
- Project-level change control
  - Each requested change is formally reviewed.
  - Change-control Board to analyze and approve/reject.



Point-of-Contact (PoC) User Support Mode

- Each major or strategic science team, including Blue Waters Graduate Fellows, is assigned a Point of Contact (PoC) for the duration of their Blue Waters project.
- With a dedicated PoC the science team can write in with questions to someone who is already familiar with that team's software and science approach. An informed response can be given much more quickly than the typical approach where the support staff has little knowledge of science teams.
- When a science team is planning a campaign the PoC can assist in the planning and pull in other staff experts as needed (ex: storage and system admin).

Project Management

- Approach for new component acceptance
  - Detailed planning and extensive testing.
  - Multiple levels of coordination and approval.
  - Deploy and test on small-scale test system first.
  - Goal is always find problems before they affect users.
- Early-Science System (ESS)
  - 48 cabinets in 2012, in user-friendly mode.
  - Access by “important” application teams.
  - Science use while acceptance-testing was ongoing.
    - Many bugs found during this period!

Configuration Evolution

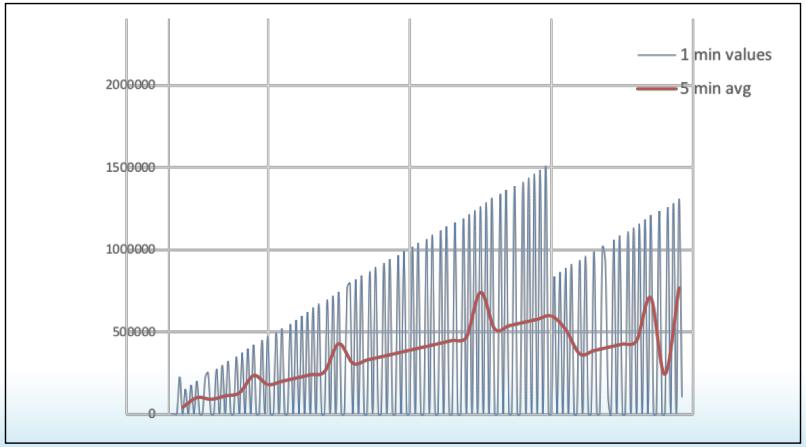
- Blue Waters' initial configuration:
  - 276 cabinets, 32 with GPUs
- Summer'2013 expansion
  - Addition of 12 more cabinets with GPUs
  - Relocation of I/O nodes was needed
- Other changes
  - Deployment of Topology-Aware Scheduler (TAS)
  - Use of declustered arrays in storage sub-system
  - Doubled the memory of some nodes.
  - Many software upgrades…

System Monitoring and Analysis

- Blue Waters: most instrumented HPC system ever!
  - Use of both off-the-shelf and custom monitoring software
  - All sub-systems are monitored (i.e. compute, storage, etc.)
  - Big-data challenge: 20+ billion datums captured per day!

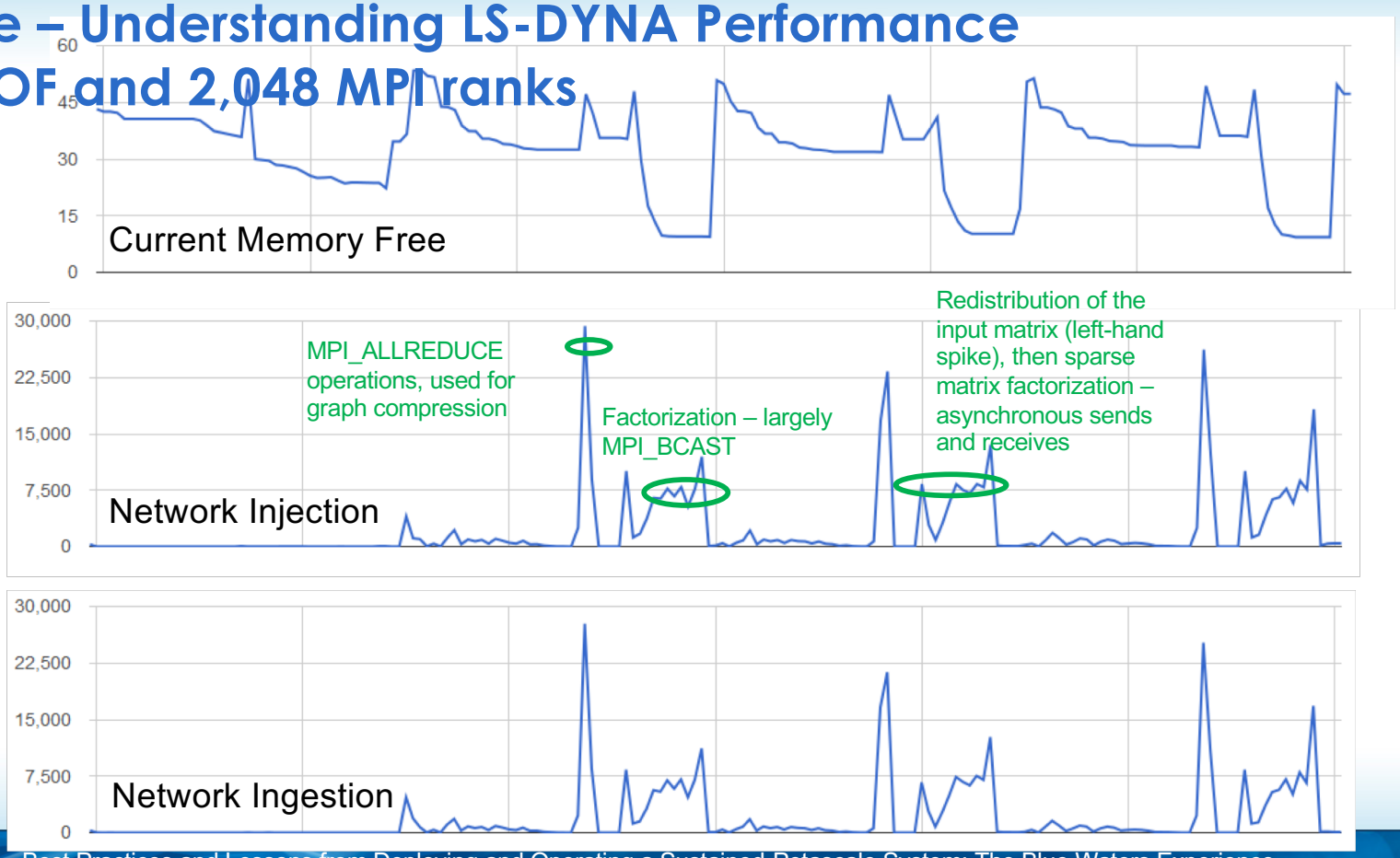
Example of the benefits from capturing data with high fidelity

(Data-collection based on OVIS, developed at LANL)

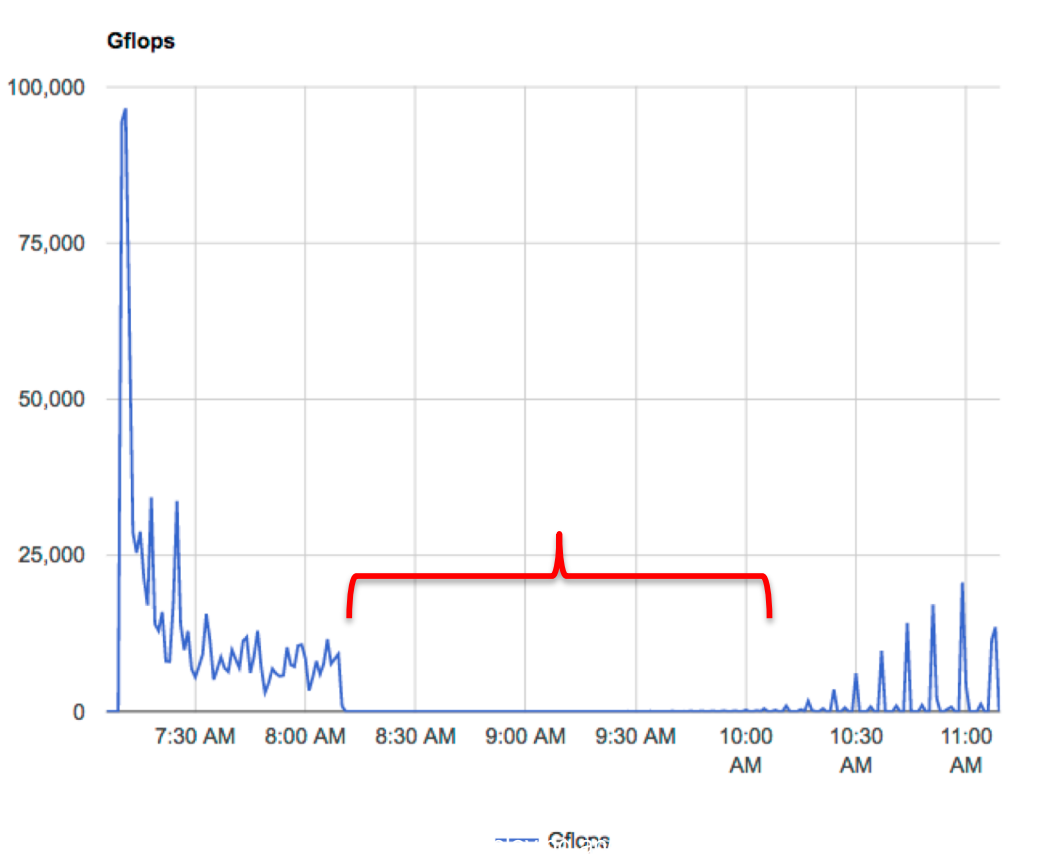


Example – Understanding LS-DYNA Performance 200M DOF and 2,048 MPI ranks

The spike is MPI_ALLREDUCE
The redistribution spike is asynchronous sends/receives (we also have an MPI_ALLTOALLV variant).
The factorization "ramp up" is largely MPI_BCAST.



System Monitoring Examples Continued

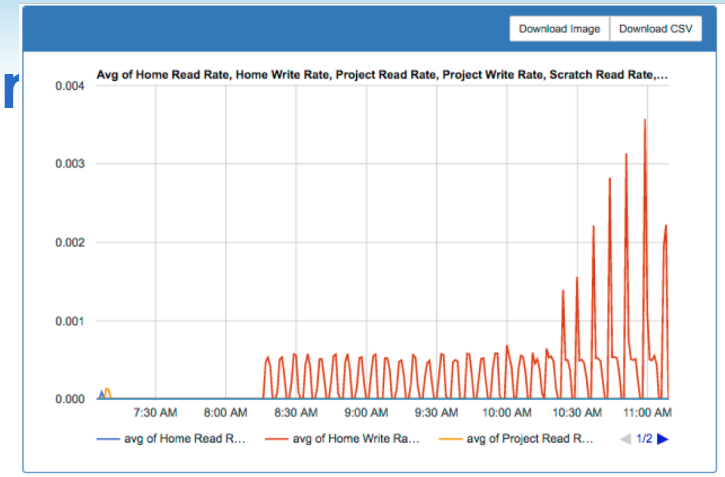


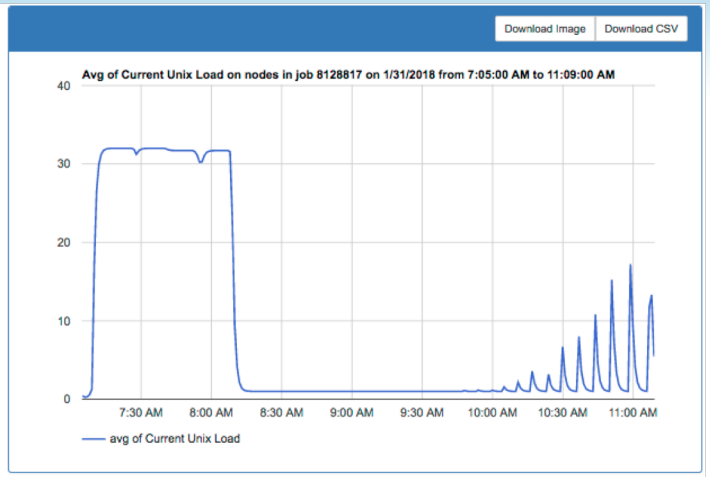
Full system job with long idle phase (no flops)!

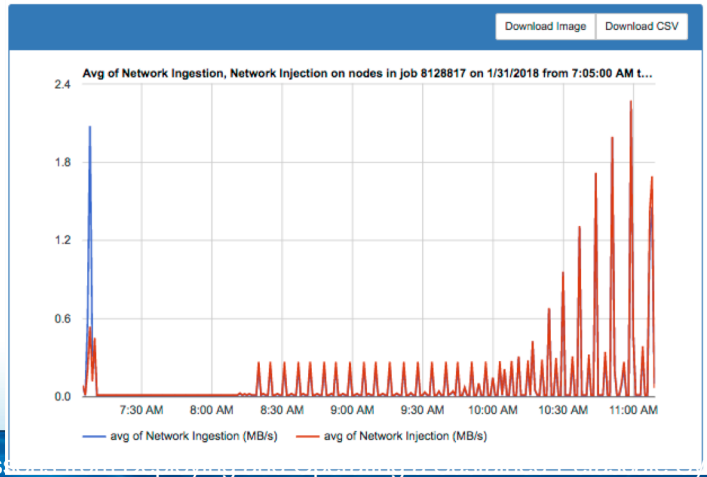
System Monitorir

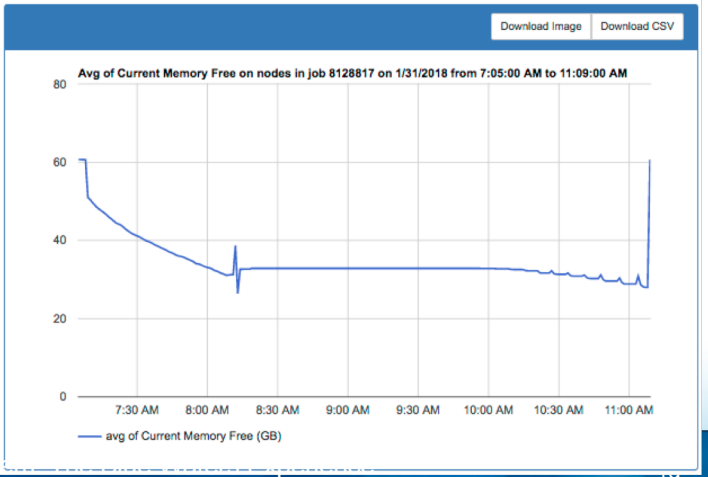
No flops phase corresponds with slow IO phase, low node load and constant memory usage.

User contacted. Issue with data on slower filesystem and not large enough system (scaled too far).





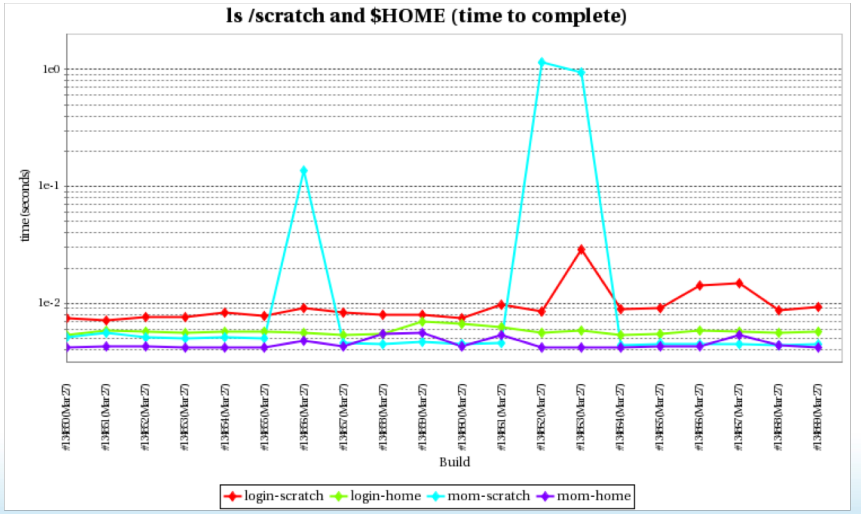




System Monitoring and Analysis (cont.)

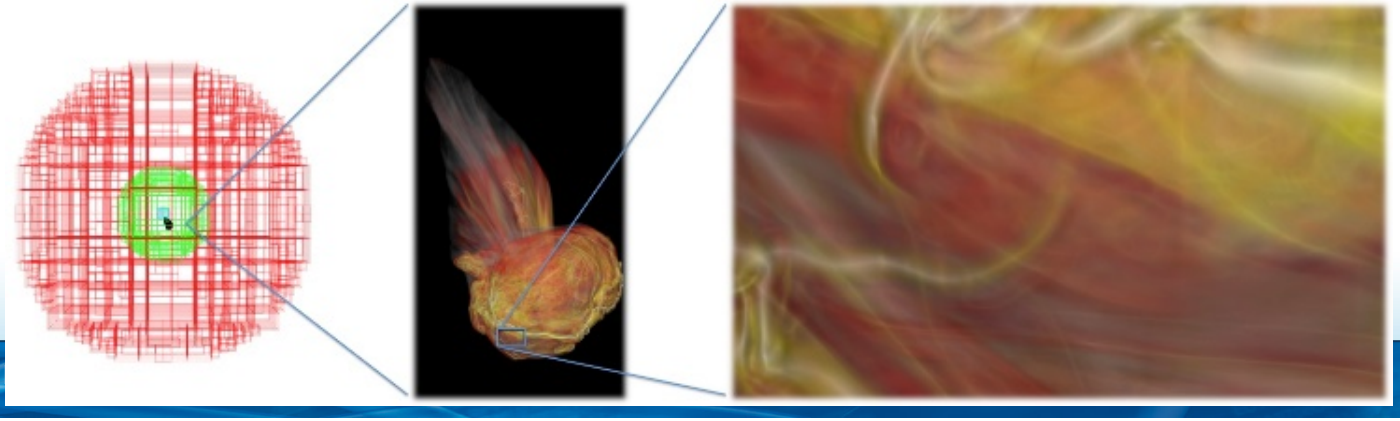
- Proactive System Monitoring
  - Jenkins server to automate periodic testing of apps.
  - Attempt to probe the user experience.
  - Multiple tools for other automated tests

Example of outcomes from periodic file system tests on *login* and *mom* nodes



Visualization Support

- Primary goal:
  - Integrate visualization into domain-science workflows.
    - Collaborative effort with science teams.
  - Requires scalable tools and a focus on data layout.
- Example: AMR dataset
  - Multiple levels of data representation may be needed



Visualization Support (cont.)

- Many opportunities for "support" activities:
  - Visual representation of system status
    - e.g. job/utilization data, I/O rates, and others
  - Main goal is to aid users and sys-admins
- Concrete potential for community-advance work
  - Example: middleware for *in-situ* processing
    - Visualizations were generated "for free"
    - Promotes cost-effective and efficient resource usage in extreme-scale systems

PAID: Petascle Application Improvement Discovery

- Third Phase of a broader program in the Blue Waters project.
- Goal: improve application performance by pairing PRAC teams and **Improvement Method Enablers** (IMEs); providing both financial and infrastructure support.
- Targeted 5 areas of need: (i) task mapping and load balancing, (ii) scalable I/O and HDF, (iii) Fourier transforms, (iv) GPU exploitation, (v) programming model best practices.
- Implemented project management:
  - Each PRAC team + IME developed a SOW with milestones/deliverables with a **measured baseline of performance**.
  - Each IME team assigned a **PoC** to facilitate interactions, resolve issues, review SOWs, review reports, etc.
  - Monthly progress meetings with all teams to highlight successes, discuss problems.

- Program metrics
  - Measured level of performance improvement over baseline
  - Implementation of new functionality
- Activities and results achieved
  - 15 PRAC teams participated with the 8 IMEs teams
  - GPU exploitation
    - CUDA speed-up: 1.3x and 1.6x over existing CUDA codes (AWP-ODC, ChaNGa)
    - OpenACC speed-up: 2.8x and 3.9x over all-CPU node (MS-FLUKKS, 3D-FDTD)
  - Topology-aware task placement and load balancing
    - task placement speed-up: 1.2x to 2.2x over default placement (PSDNS, MILC).
    - Load balancing speed-up: 2x with improved SMP (NAMD)
  - Scalable I/O and HDF
    - HDF5 speed-up: 6x to 9x of IO time (NEURON, PSDNS)
    - (New) meshio library speed-up: 20x of IO time (MILC)
  - Programming Models Best Practices
    - OpenMP refactoring speed-up: 2.9x (MS-FLUKKS)
  - FFT
    - improved ACCFFT performance 2x to 4x.

Education and Outreach

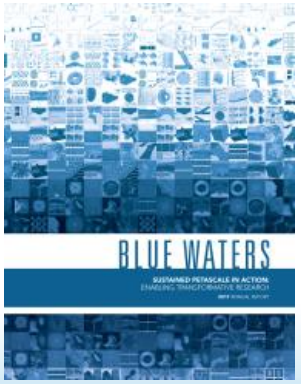
- Multifaceted program to broadly engage the community
- Education allocations and training
  - Allocations for grad and undergrad courses
  - Internship and fellowship programs for undergrad. and grad. students.
  - Webinars, workshops, tutorials, hackathons, etc.
    - Enabled by a limited use specialized authentication scheme.
- Allocations for broadening participation
  - Pro-active recruiting of underrepresented individuals
  - Reaching several minority organizations nation-wide

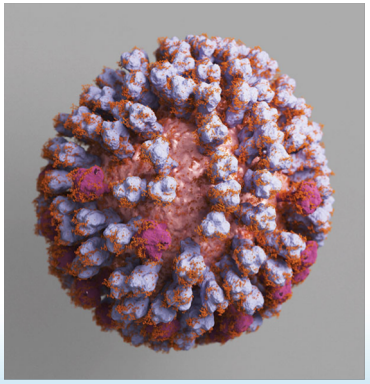
Student Engagement

- Undergraduate internships and graduate fellowships
  - Fellowship focused on computational research at scale
  - BW staff as points of contact for sustained support of fellows
  - Two week petascale institute to prepare interns
  - Mentors on research projects provide application of new skills
  - The internships have been conducted for 7 years
- Interns publish papers on their experiences in JOCSE journal

Communicating Success

- Annual Blue Waters Symposium
  - Unique cross discipline meeting: science users, NCSA staff, students
  - Engage people from different domains and fosters collaborations in a collegial setting.
  - Reports from all groups, audio/video available on Blue Waters portal.
- Blue Waters Annual Report
  - Articles convey the science enabled.
    - Understandable to the general public.
    - Societal impact is addressed.
    - Highlights why Blue Waters was essential.





Return on Investment

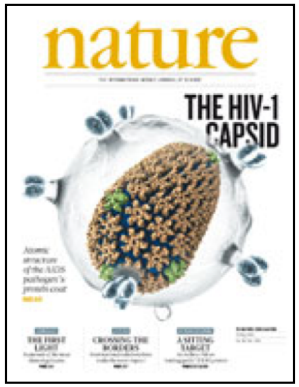
- No complete overall financial analysis yet.
- Data for specific parts of the project are available.
  a) PAID program
     - Cost: approximately $5.5M
     - Gains: smaller app runtime ~ $9.7M
  b) Topology-Aware Scheduler
     - Gains: applications run faster – estimate of 42% higher science throughput, based on network traffic measured
  c) Impact of the project on Illinois' economy
     - Direct projected impact: $1.08 billion

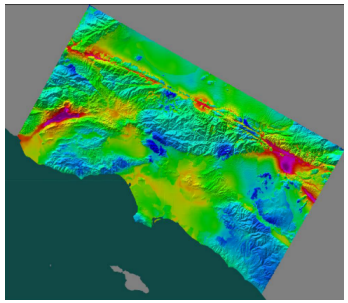
Conclusion

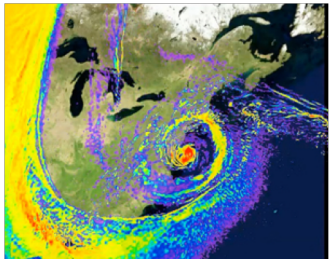
- Blue Waters Best Practices
  - Blue Waters broke new ground on many fronts resulting in many challenges and required many innovation solutions.
  - We have found these practices to be essential to our success.
    - Many have been inspired by practices at other centers.
    - We hope this work inspires you to copy and improve upon our practices!

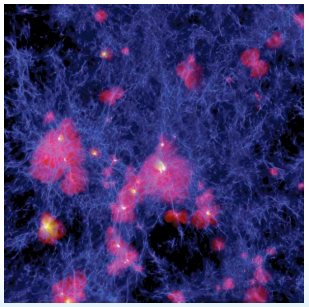
Conclusion

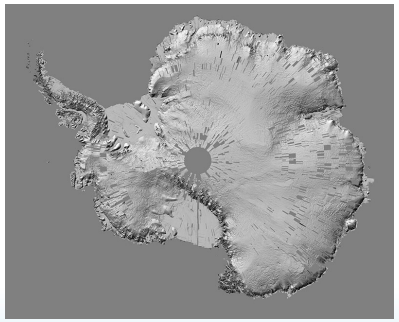
- Blue Waters: 5+ Years Enabling Discoveries











Acknowledgments

- Funding: NSF OCI-0725070/ACI-1238993 and the state of Illinois
- Personnel: The **NCSA** Blue Waters team and the Cray site team