PSAAP University Alliance Presentation

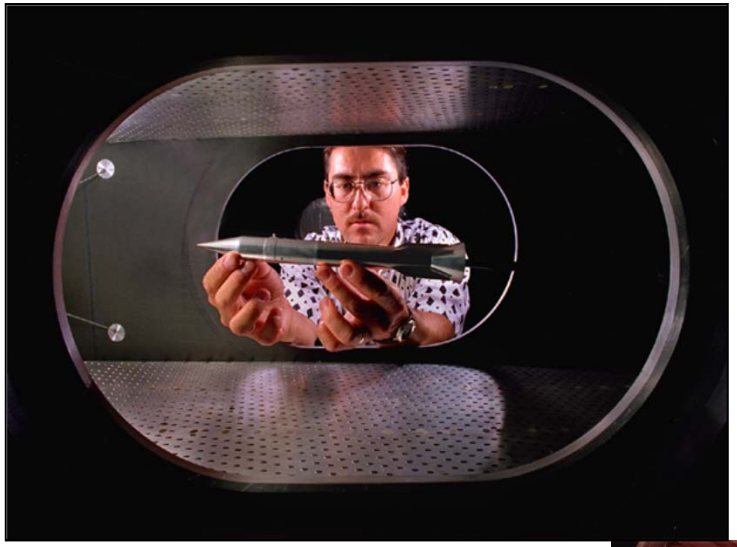
Sandia National Laboratories

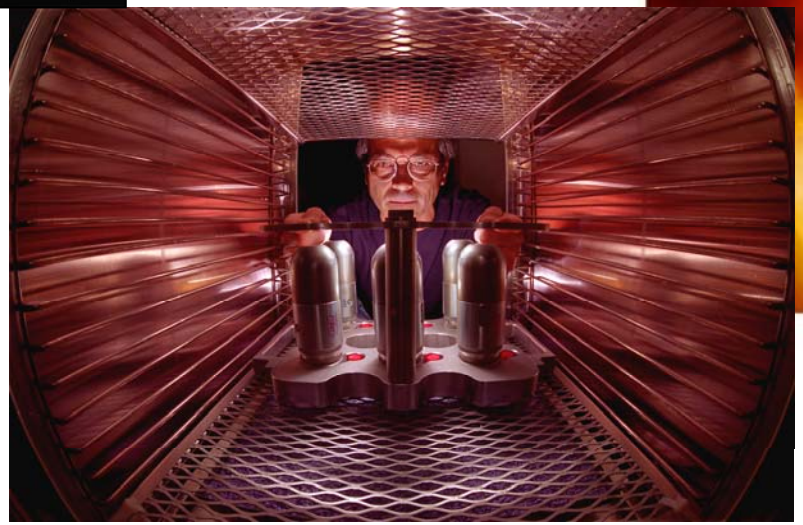
September 4, 2008

Barbara Jennings, Joel Stevenson &
Lilia Martinez

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Who are we?





- **National Security Laboratory**
- **Primary Mission in Nuclear Weapons**
- **Broader Mission in Science and Engineering to meet National Security Needs**

Advanced Simulation & Computing™

Sandia National Laboratories

Where are we?

- New Mexico
- California

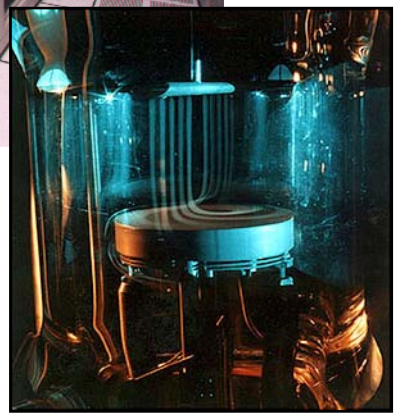
- Nevada
- Hawaii
- Texas
- Washington, D.C.

Advanced Simulation & Computing™

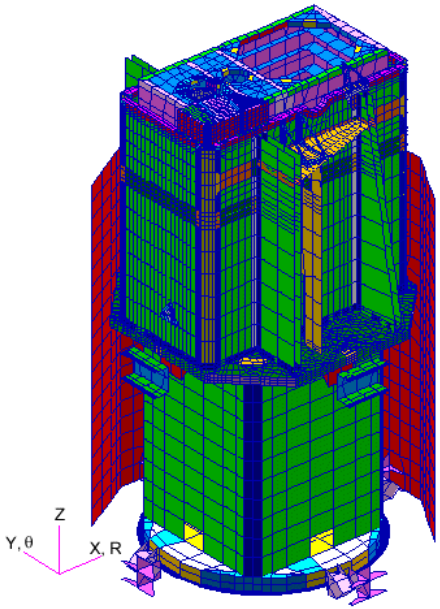
Sandia National Laboratories

Example of our Science Base

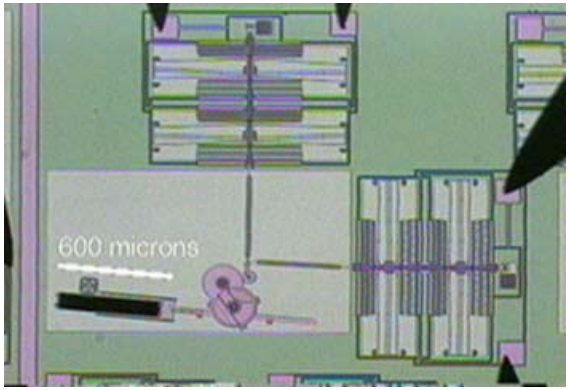
Computational and information sciences



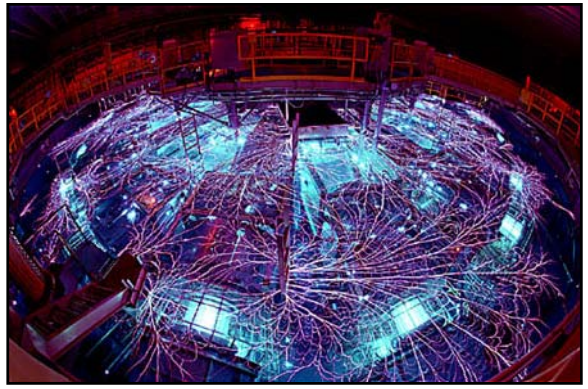
Materials and processes



Engineering sciences

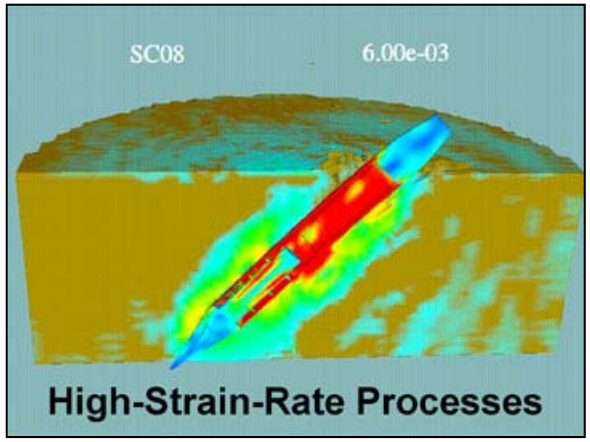


Microelectronics and photonics



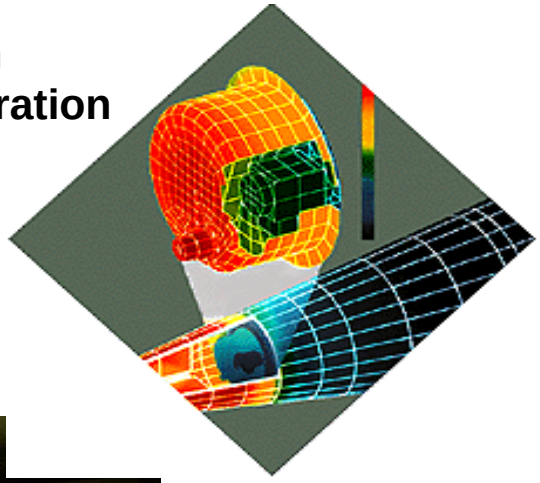
Pulsed power

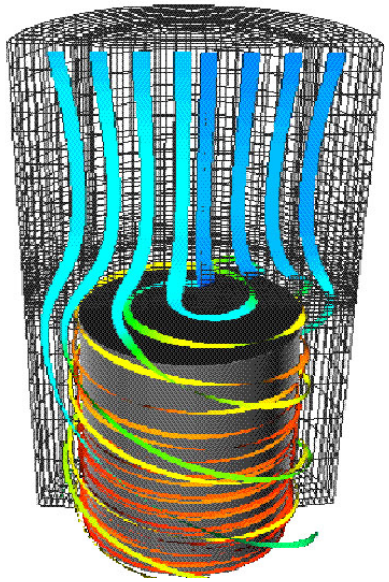
Modeling and Simulation



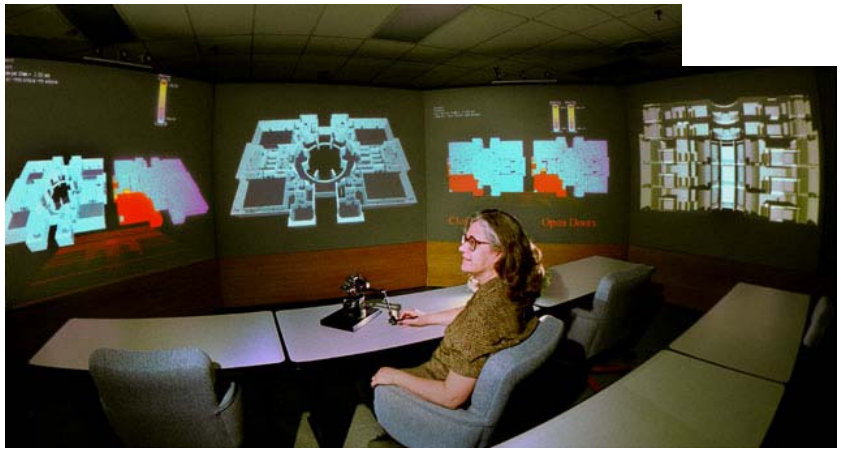
System performance modeling

Mesh generation





Manufacturing process model



Engineering Design Center

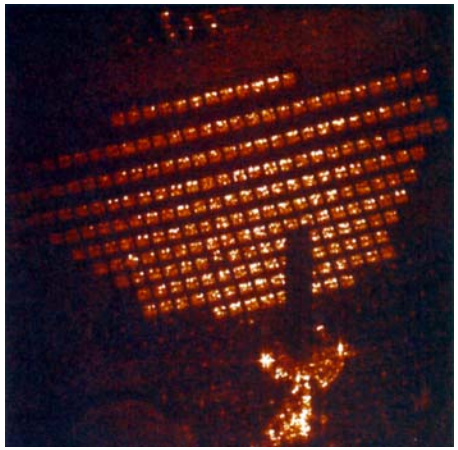
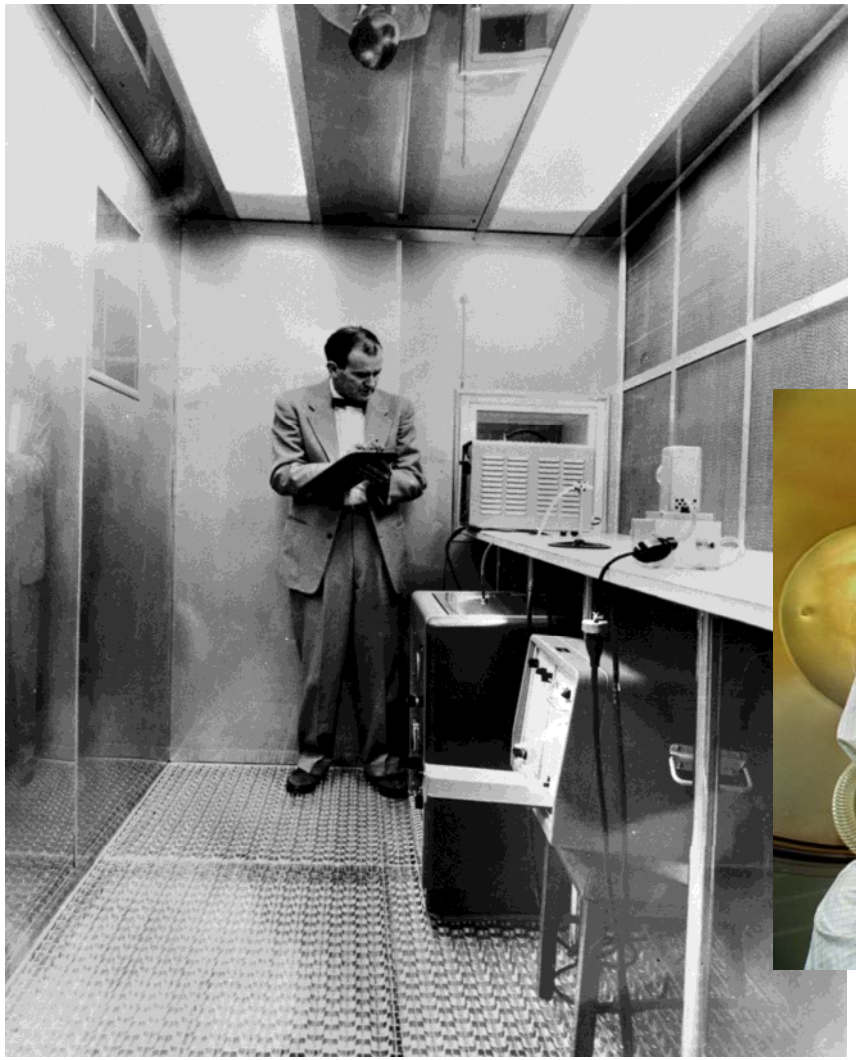


Image processing

Advanced Simulation & Computing™

Sandia National Laboratories

50 Years of Exceptional Service



Cleanroom invented 1963

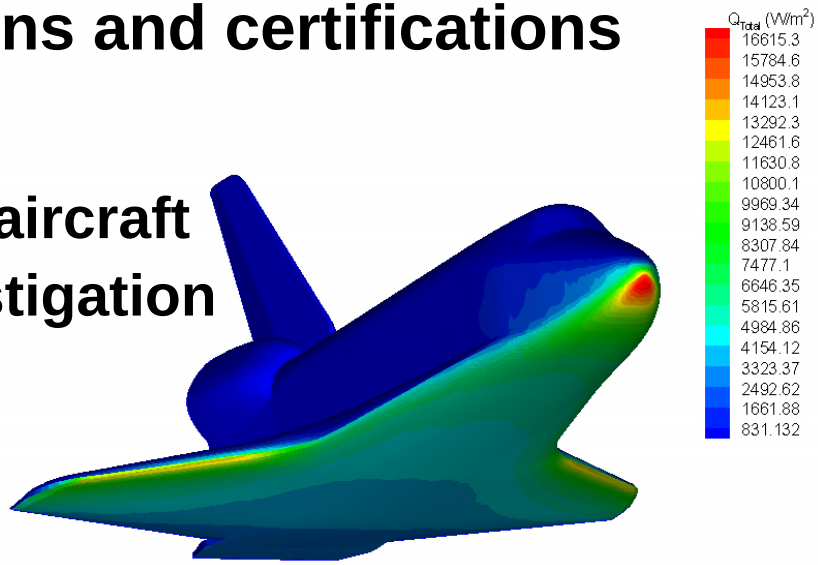
Advanced Simulation & Computing™

Sandia National Laboratories

50 Years of Exceptional Service

Investigations and certifications

- TWA Flight 800
- Forest Service aircraft
- Columbia investigation

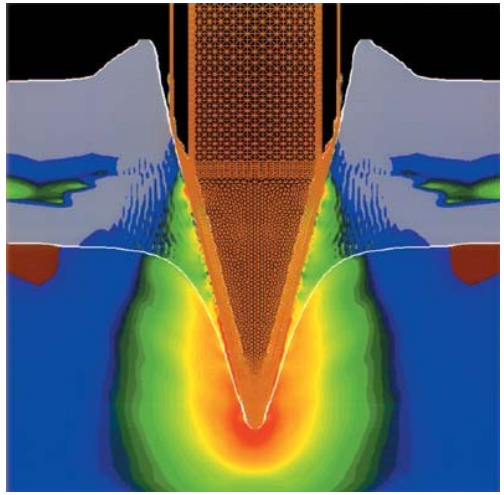


Columbia wing section

Our Primary Mission is the Surety of the Nuclear Stockpile

- *Primary Mission:* Stewardship of the nuclear weapons stockpile
- As a Systems Integrator, Sandia is Responsible for more than 95% of U.S. Nuclear Weapons Components
- Sandia is also Responsible for Maintaining Weapons in the Stockpile

Computer modeling and simulation



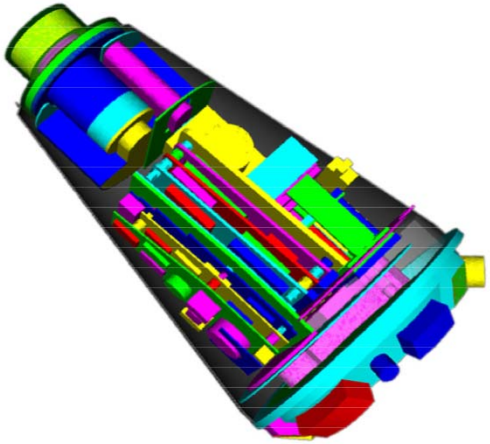
Flight test

Advanced Simulation & Computing™

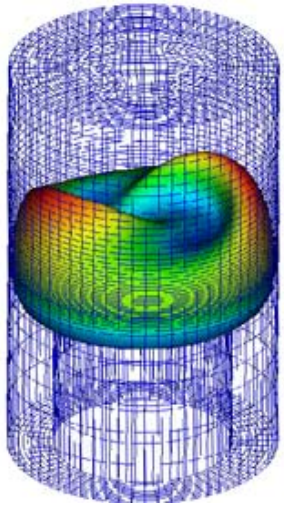
Sandia National Laboratories

Revolutionizing Engineering & Manufacturing Through Supercomputing

Model fidelity



- Design and simulation
- Life-cycles engineering
- World's fastest computers in a labs-wide collaborative environment



Computational simulation

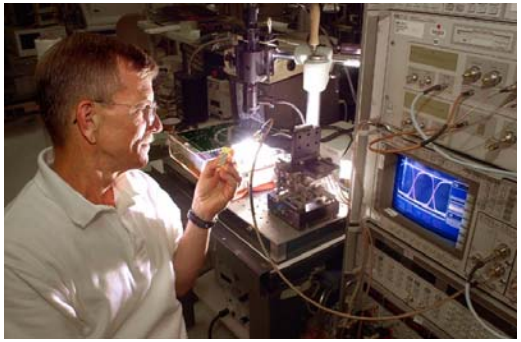
Red Storm
124 teraflops

Advanced Simulation & Computing™

Sandia National Laboratories

We Actively Seek New Partnerships

- Research facilities available to industry
- 250-acre Science and Technology Park
- Teamwork benefits industry and government
- Computer Science Research Institute
  http://www.sandia.gov/CSRI



EMCORE partnership

Goodyear CRADA

Extreme Ultraviolet Lithography Consortium

Advanced Simulation & Computing™

Sandia National Laboratories

Research to Production

- Advanced Simulation and Computing (ASC) applications in support of DOE mission
- Modeling and Simulation lab-wide
- Software Development Research
- Parallel Algorithm Research
- Software/Hardware Systems Research
- Production Computing in Support of National Needs and Requirements

Sandia HPC History

- **1987 – 1024 node nCUBE 10**
  - Sandia fielded first true MPP
  - Won Karp Challenge
  - Won inaugural Gordon-Bell Prize
  - Awarded patents for nearly every aspect of MPP software
- **1990 – 2 1024-node nCUBE-2's**
  - Outperformed Cray vector computers @ 1/7 the cost!
  - Sandia began research on Light Weight Kernel (LWK) operating systems (SUNMOS)

- **1993– Sandia fielded first ~1850-node (3900 processors) Intel Paragon**
  - Sandia (Intel) Paragon is #1 on Top 500 list (first for Sandia)
  - Wins Sandia's second Gordon-Bell Prize
  - First use of Sandia developed Light Weight Kernel (LWK) Operating System software (PUMA) for production computing (16 MB memory/node.)

1997– Tflops, World's First Terascale Computer

4600+ nodes (9200+ processors – initial 200 MHz, upgraded to 333 MHz; memory doubled)

- –Ran Sandia system developed LWK System Software (Cougar)
- –Number 1 on the "Top 500" list for 7 consecutive lists from June, 1997 through June, 2000, a record still unmatched! **(1.068, 1.338, 2.123, 2.3796 TF)**
- –Red-Black switching
- –Approx 16 Cray YMP's
- –Portals networking protocol

- **ASCI Option Red – The Intel Tflops system was *operational until October of 2005*. Subsequently retired and reduced to rubble for security reasons.**

Advanced Simulation & Computing™

Sandia National Laboratories

Sandia HPC History

- **1997– Sandia began development of world's first Linux "Super-cluster", Cplant**
  - **Sandia integrated DEC/Compaq HW with Myrinet high performance interconnect network**
  - **Sandia developed all run-time, file system and messaging software**
  - **Open source Cluster Management Toolkit**
  - **World's first terascale cluster– eventually achieves nearly 1 TF on Linpack.** (996.9 GF in November 2003)
- **2001– Sandia invented Red Storm architecture – issues RFQ to industry**
  - **Responses were being reviewed on Sep 11**
- **2002 – Sandia contracted with Cray, Inc to jointly develop Red Storm/XT3**
  - **Sandia architecture; Cray built hardware; Sandia and Cray developed system software**

Computational Plant

Red Storm

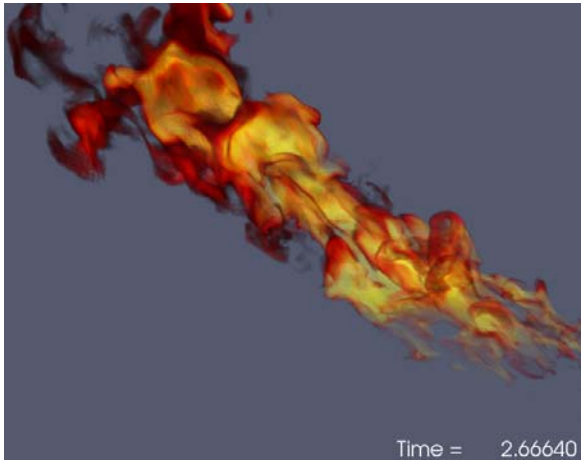
True MPP, designed to be a single system
- Full 3-D mesh interconnect
- 12,960 compute nodes (Upgrade 48% Qaud Core)
- 39 Terabytes of memory
- 340 Terabytes of disk storage

Sandia contributions include
- MPP system architecture
- Helped design interconnect
- Lightweight kernel strategy
- Red/Black switching

Excellent performance
- 102.2 TeraOps on HPL
- Second system ever to exceed 100TF
- First general-purpose system to exceed 100TF

Successful technology transition
- Cray now has now > over 30 systems based the Red Storm model

Serves LANL, LLNL, SNL, Alliance Partners
- Remote user support a given
- Remote networking support active since ASCI Red



Fire simulations are being used to certify new test facility and will be used to analyze weapons safety issues

Advanced Simulation & Computing™

Sandia National Laboratories

10 Years HPC Computing at Sandia
2.379 TF -> 4.92 PF

- **ASCI Red - Intel**
  - 1997 - 2006
  - World's First TOPs (Tera Operations per Second) System; Upgraded to 3TF 2.379 TP
  - 4510 Nodes/9298 Processors
  - Fastest Computer on TOP500 list for 3 years remained on list until decomissioned in 2006
- **Thunderbird - Dell**
  - 2006 - Current
  - Linux Cluster 65.4 TF
  - 4,480 Nodes/8960 Processors
  - Linux OS
  - June 2006 #6 In TOP500 list
- **Red Storm - Cray**
  - 2006 - Current
  - TFLOP (Tera Flops Operations Per Second) 4.92 PFLOP's 284 TF Theoretical Peak
  - 12960 Nodes/38,400 Processors:6240 quad-6720 dual
  - Catamount OS
  - June 2008 Listed #12 in TOP500 list

Current CapViz/ESHPC Resources

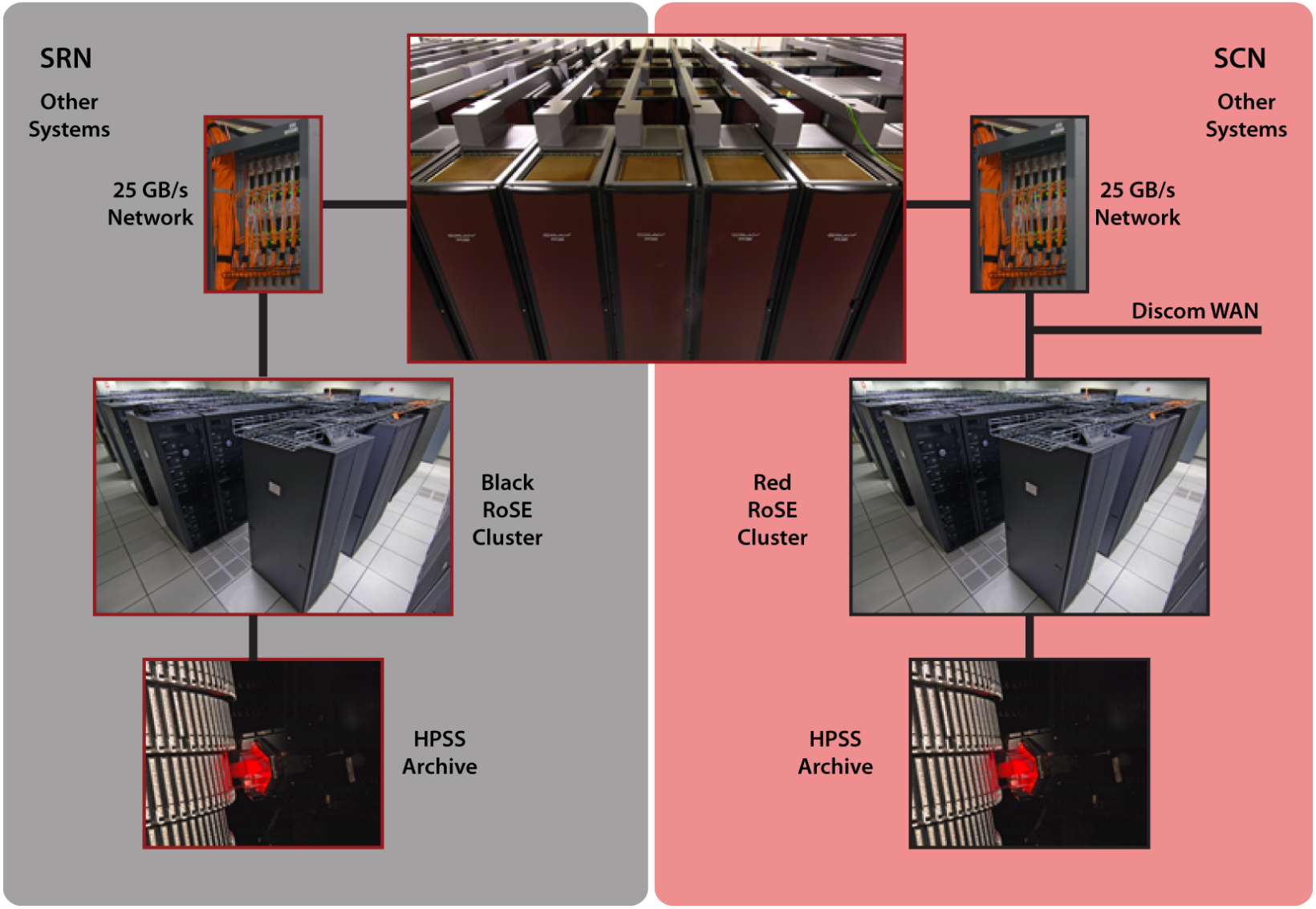
System	Nodes/Processor cores	Processor	Memory	TFLOPS	Proc .Hrs per Year
Tbird	4480 nodes/8960 cores	3.6 GHz EM64T	6 GB/node	65	78,489,600
Spirit	512 nodes/1024 cores	3.4 GHz EM64T	2 GB/node	6.9	8,970,240
Shasta	198 nodes/396 cores	3.0 GHz IA32	2 GB/node	2.4	3,468,960
Black RoSE	128 nodes/512 cores	2.8 GHz AMD	8 GB/node	2.8	4,485,120
Rogue	407 nodes/814 cores	2.8 GHz EM64T	4 GB/node	4.6	7,130,640
Razor	280 nodes/1120 cores	3.0 GHz Woodcrest	8 GB/node	13.4	9,811,200
Justice	512 nodes/1024 cores	3.4 GHz EM64T	2 GB/node	6.9	8,970,240
Lassen	192 nodes/384 cores	3.4 GHz EM64T	2 GB/node	2.6	3,363,840
Red RoSE	264 nodes/528 cores	3.6 GHz Xeon	4 GB/node	3.8	4,625,280
Unity	272 nodes/4352cores	2.2 GHz AMD Quad Core	32 GB/node	38	38.123.520
Glory	272 nodes/4352 cores	2.2 GHz AMD Quad Core	32 GB/node	38	38.123.520
Whitney	272 nodes/4352 cores	2.2 GHz AMD Quad Core	32 GB/node	38	38.123.520

Red Storm at Albuquerque Site

Advanced Simulation & Computing™

Sandia National Laboratories

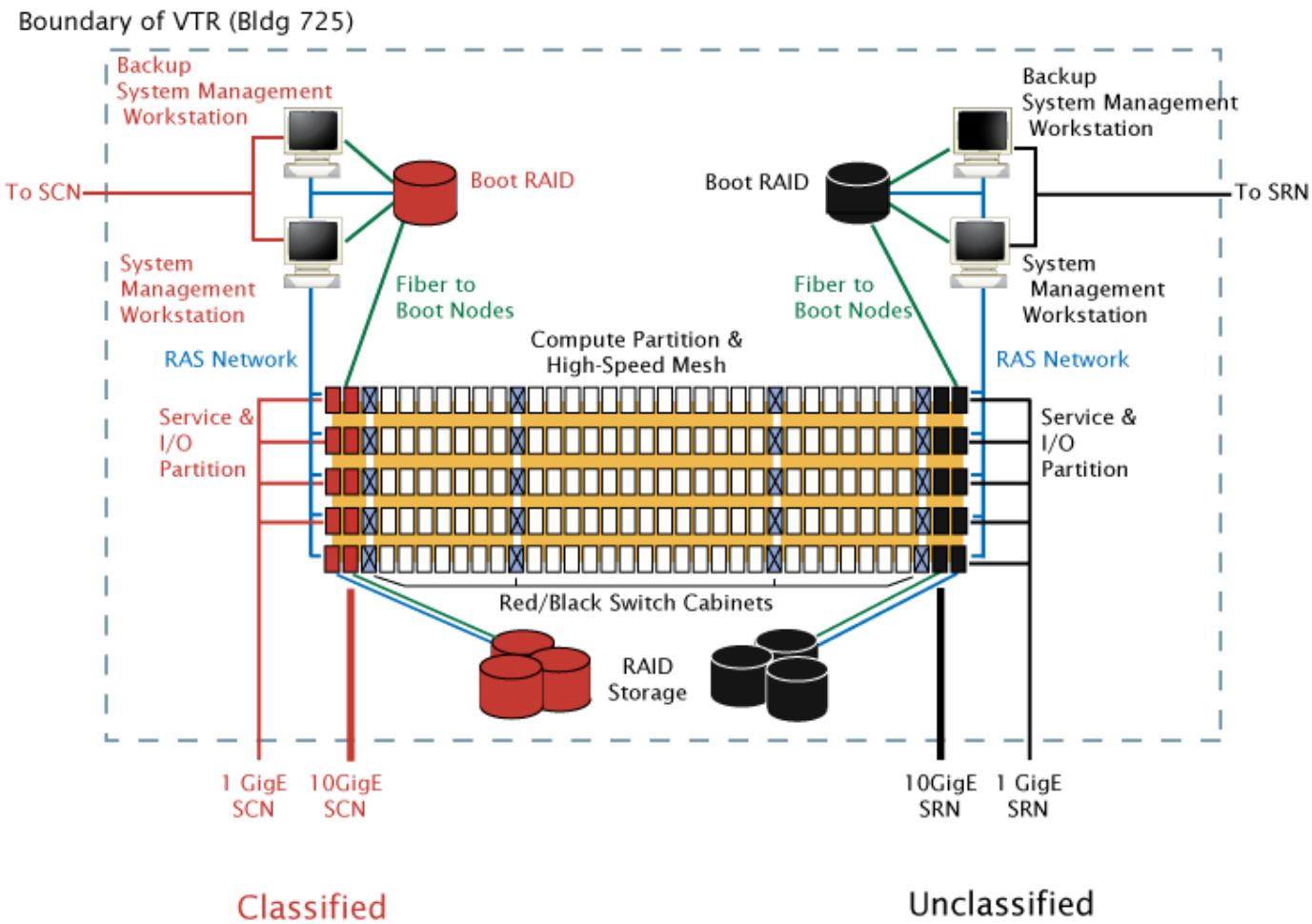
Architected Red Storm Environment



Advanced Simulation & Computing™

Sandia National Laboratories

Red Storm Architecture



Red Storm System Software

Operating Systems

- LINUX on service and I/O nodes (SuSE Enterprise Server)
- Catamount LWK on compute nodes
- Linux on RAS monitors

Run-Time System

- Logarithmic job launch (yod)
- Node allocator (CPA)
- Batch system – MOAB

File Systems

- High performance file system (Lustre)

User Environment

- PGI Compilers - Fortran, C, C++
- Libraries - MPI, I/O, Math, MPI-2
- Showmesh
- Debugger - TotalView
- Performance Monitor

Network

- 10x1GigE to login nodes
- 1 GigE to Mgmt Workstations

System Mgmt and Admin

- Accounting
- Red Storm Management System

Reliability Features of Red Storm

- Catamount Light Weight Kernel (LWK) OS on compute partition
  - Less code fails less often
- Monitoring of correctible errors
  - Fix soft errors before they become hard
- Hot swapping of components
  - Overall system keeps running during maintenance
- Redundant power supplies
- RAS System

Red Storm – Lightweight Kernel

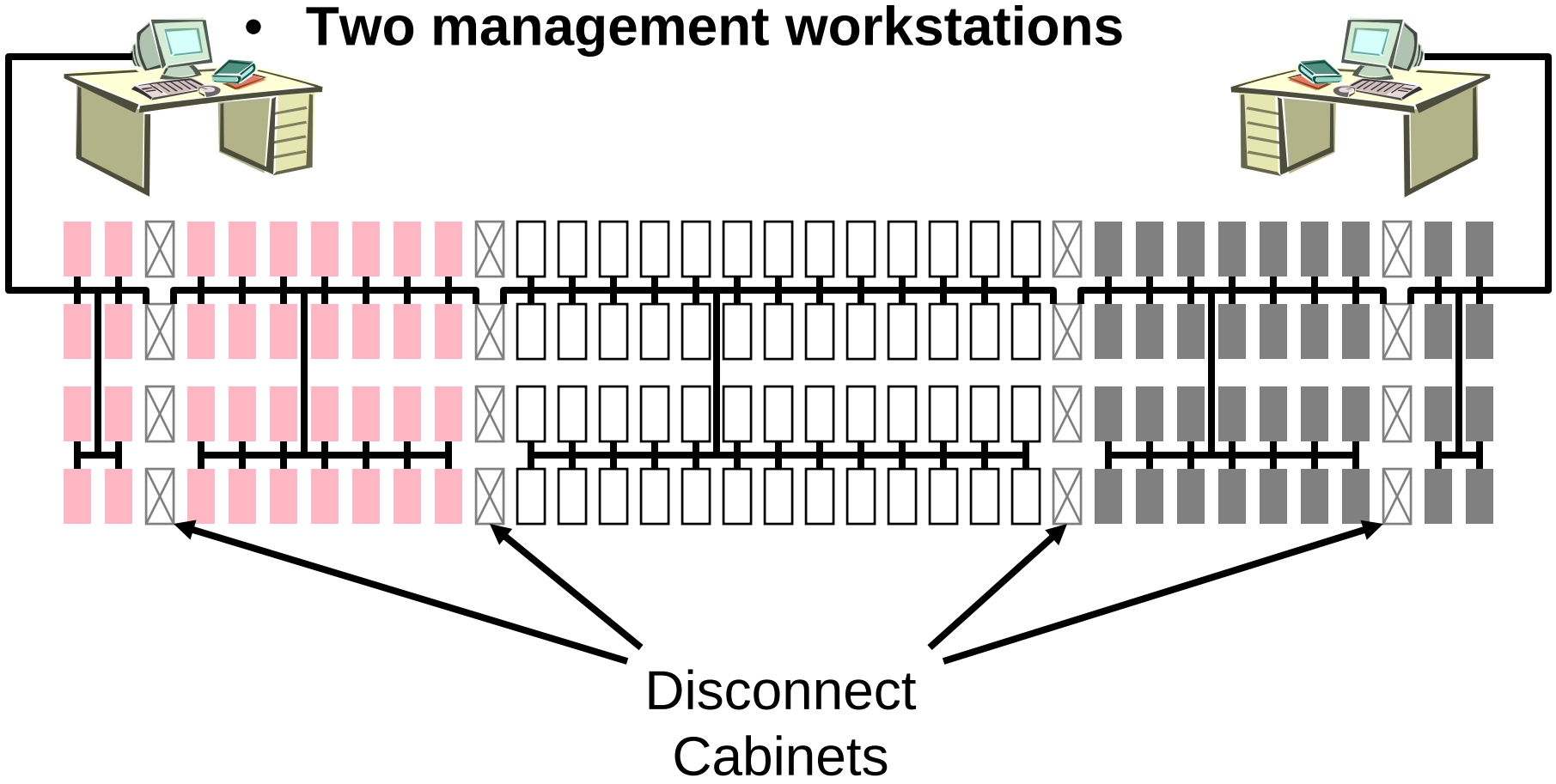
- **Lightweight Kernels are a key contribution from Sandia to the HPC community**
- **Based on several key principles:**
  - **The purpose of compute nodes is to compute**
    - **Whenever any processor takes an interrupt, other processors in the computation may have to wait**
    - **Your computation is only as fast as its slowest component**
  - **Maximize the resources devoted to the computation**
  - **Simplify the code to enhance reliability**
- **This may not be an issue for dozens of nodes**
  - **Red Storm has 25,920 processors to coordinate!**

The Light Weight Kernel (LWK)

- A lightweight compute node OS is a fundamental part of the Sandia architecture
- It is essential for
  - Maximizing CPU resources
    - Reduce OS and runtime system overhead
  - Maximizing memory resources
    - Small memory footprint, large page support
  - Maximizing network resource
    - No virtual memory, physically contiguous address mapping
  - Increasing reliability
    - Small code base, reduced complexity
  - Deterministic performance
    - Repeatability
  - Scalability
    - OS resources must be independent of job size
- Others have realized these benefits
  - nCUBE (Vertex), Cray T3 (UNICOS/mk), IBM BG/L (HPK)
- Similar trend in Linux-based cluster SW – reducing overhead and unused capabilities

Reliability, Availability, and Serviceability

- **Separate LAN to all nodes for RAS**
- **Two management workstations**



File Systems

- Storage
  - SMSS - (pftp, hsi, htar, (c)hopper)
  - http://dmt.sandia.gov tools for moving files
- Luster Volumes for Jobs to run - don't run from home
  - /scratch1
  - /scratch2
  - /scratch_grande
  - /scratch3 (Biggest - 300TB Uncl/1PB Class)
- Compile on login node
  - /home
  - /projects (nfs mounted get backed up)
- Reddish: Compile and build and move to RS
- Blackrose: VIZ nodes, and apps (paraview, ensight)
- Red Storm System Maintenance 7:00 am - 1:00 pm MST
  - Tuesdays - Classified
  - Thursdays - UnClassified

Compilers/Libraries

Compilers

- pgi
- gnu
- gcc
- pathscale

Libraires

- C Language Runtime Library
- AMD Core Math Library
- Cray XT3 LibSci Scientific Libraries
- Cray MPICH2 Message Passing Library
- Cray Shared Memory Access (SHMEM) Library

Estimating Jobs - Required

HERT

- HPC (High Performance Computing) Estimations & Requirements Tool
- Captures estimates of future HPC system usage from users
- Allows users to specify urgency and importance
- Organizes and reports this data
- Tied directly to HPC system queues and job schedulers

- https://computing.sandia.gov/hert
- Hert-dev@sandia.gov

Job Submission

- The MOAB workload manager adopted by Tri-Lab
- Working toward uniform interface for batch scheduler across the ASC Tri-Labs
  - Common submit, status, and control
- Based on native MOAB command set
  - Currently support PBS interface and will provide MOAB wrapped commands for backward compatibility
- Provides fair-share mechanism
- Quick Start guide for MOAB and PBS:
  - http://redstormweb.sandia.gov/RedStorm/usage.html

Interactive

- **Method 1 – allocated from the large pool of compute nodes**
  - **Step 1 → qsub –I –lsize=<size> –lwalltime=<time> –A project/task**

    This places you on a service node and gives you an interactive session for <time>, where <size> is the maximum number of nodes that you need during the session
  - **Step 2 → yod [–sz <cores>] [–SN, –VN, –VN4] <executable>**

    You can run yod as often as you like during the session, specifying the number of cores for each run.
    - **Example 1 (one core on one node) % yod -sz 1 -SN <exe>**
    - **Example 2 (two cores on two nodes) % yod -sz 2 -SN <exe>**
    - **Example 3 (two cores on one node) % yod -sz 2 -VN <exe>**
    - **Example 4 (<size> cores on <size> nodes) % yod <exe>**
    - **Example 5 (2*<size> cores on <size> nodes) % yod -VN <exe>**
- **Method 2 – allocated from the much smaller pool of interactive nodes**
  - **yod –sz <size> –A project/task <executable>**

Batch

- qsub <script>
  where script looks like:
  - "single-core" mode

    #!/bin/sh
    #PBS –lsize=64 (number of nodes that you need)
    #PBS –lwalltime=10:00:00
    cd $PBS_O_WORKDIR
    yod [–sz 64] [–SN] <executable> (64 cores on 64 nodes)
  - "dual-core" mode

    #PBS –lsize=32 (number of nodes that you need)
    yod [–sz 64] –VN <executable> (64 cores on 32 nodes)
  - "quad-core" mode

    #PBS –lsize=16 (number of nodes that you need)
    yod [–sz 64] –VN4 <executable> (64 cores on 16 nodes)

PBS Wrapper Scripts

- qsub
- qstat
- qel, qsig
- qhold
- qrls
- tracejob

Scheduling via File System Requests

- Users can voluntarily add a file system "feature" at job submission as follows:

```
#PBS -lsize=1:ppn=2:file_system_name
```

or

```
qsub -lsize=1:ppn=2:file_system_name
```

where file_system_name is one of the following:

SRN: gscratch1 or gscratch3 (on Blackrose)

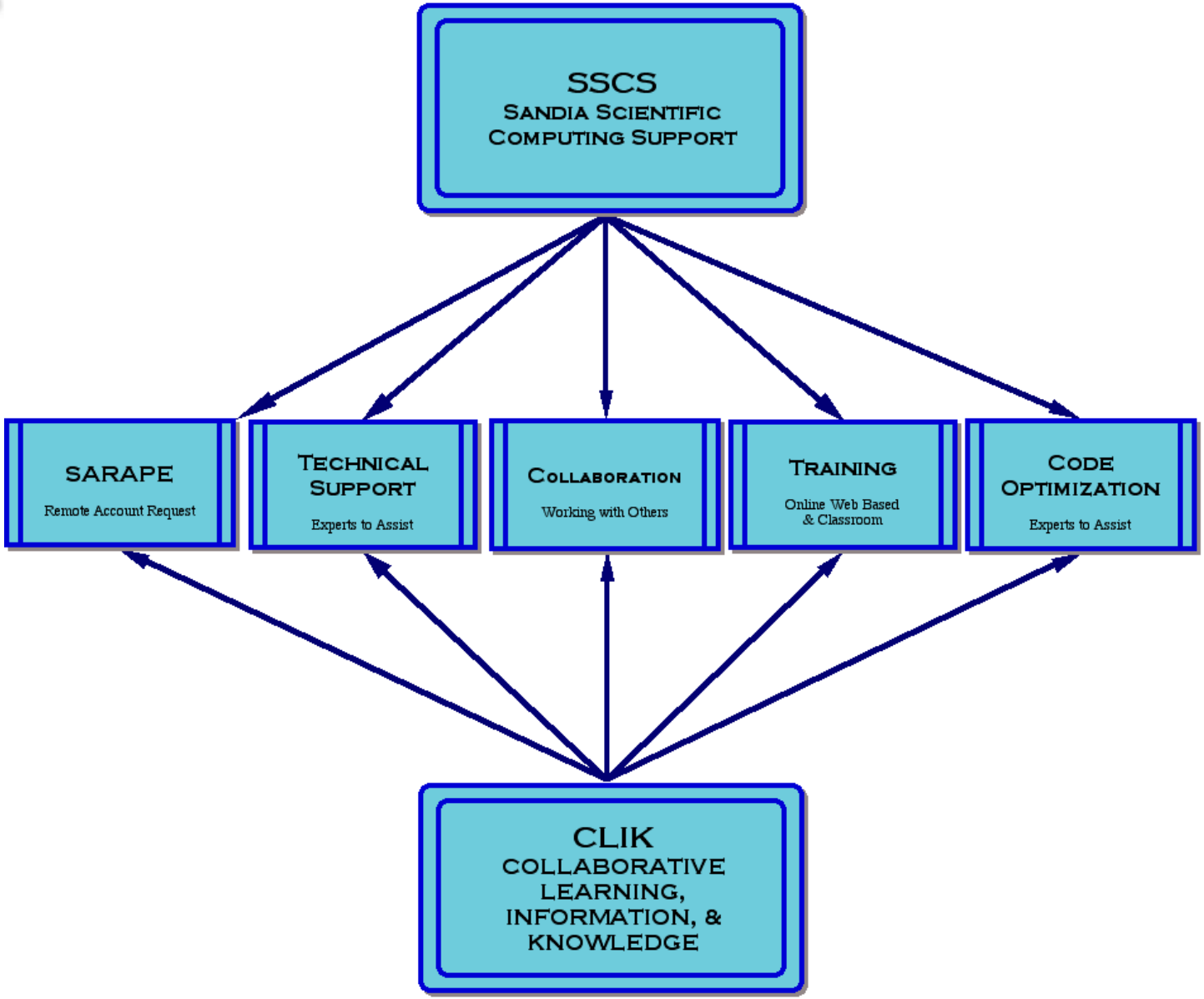
- To check the file system for a specific job, you can run the following command:

```
checkjob -v <job_id> | grep -i attr
```

HPC User's Needs Require Unique Technical Support

We have created a wisdom community to provide support for HPC computing at SNL

- **Focus**
  - **High Performance Computing Environment**
- **Getting Help**
  - **Experts as consultants**
  - **Access via phone, web, and email**
- **Knowledge Base**
  - **Web based queries**
  - **User/Expert Contributions**
- **Training**
  - **Web based on-line training**



Advanced Simulation & Computing™

Sandia National Laboratories

Knowledge Culture Support for HPC Community

Unique support is provided to all SNL users by means of a custom tool http://clik.sandia.gov

CLIK

COLLABORATIVE LEARNING, INFORMATION, & KNOWLEDGE

Advanced Simulation & Computing™

Sandia National Laboratories

User Support/Knowledge Base

- **clik.sandia.gov**
  - **Technical Assistance Requests (TARs) Knowledge Base (KBs) Entries/Submissions**
  - **MyClik - History of your TARs and KBs**
- **redstorm-help@sandia.gov**
- **redstorm-outage@sandia.gov**
  - **All users on having accounts on RedStorm (Unclassified) are invited to subscribe to redstorm-outage by sending email to majordomo@sandia.gov majordomo@sandia.gov with the text subscribe redstorm-outage in the message body.**
- **sscs.sandia.gov**
  - Documentation, meeting notes, code optimization and other links
- **Phone**
  - 505-845-2243

Advanced Simulation & Computing™

Sandia National Laboratories

User Training

- **Driver's License**
  - http://hpctraining.sandia.gov - Driver's License
- **Face-to-Face**
  - Optimization on Cray T3 - Instruction from Cray
  - Red Storm - SNL Provided
  - As required
- **OSC**
  - On-line: www.osc.edu/supercomputing/training
    - Parallel Programming
    - Performance Tuning
    - MPI
    - MatLab

Remote Account Requests via SARAPE

- **Automated, Web Based**
- **May request accounts to be created or deleted**
- **Need to know the name and the location of the resource**
- **Questions or assistance with SARAPE call 1(800)417-2634 ext 284-7727 or email sarape-help@sandia.gov**
- **http://sarape.sandia.gov**

Requesting Accounts

http://sarape.sandia.gov

Advanced Simulation & Computing™ **SARAPE** Sandia National Laboratories

Synchronized Account Request Automated ProcEss

Home | User Help | Admin Information | FAQs | CCHD Area

Welcome to SARAPE

Initiate Request

Getting Started

- Available Resources
- Example Requests
- Who is the contact at my site?

SARAPE is a web-based application that allows users within restricted domains (Tri-Labs, NWC, and ASC University Alliance Partners) to request remote accounts for selected resources at each of the Tri-Labs. An enhancement to the current manual forms, SARAPE gathers the same information electronically from a user that is contained on the manual forms. The SecureNet Guest Account Request, Intersite Unclassified Guest Account Request Form, and SecureNet Classified Guest Account Request Form are currently maintained at Livermore National Laboratory from a location on the ExtraNet Network that is accessible to any user within the restricted domains.

Lab staff, contractors, and visitors with need-to-know authorization may request computer accounts, access to applications, and authorized access such as Kerberos and DCE cross-cell. Renewal of existing accounts may also be requested, as well as the deletion of accounts that are no longer needed.

SARAPE does not replace the internal process established at each site to process access requests. It provides additional verification and approvals for guest requests.

Example Requests

Advanced Simulation & Computing™ **SARAPE** Sandia National Laboratories

Synchronized Account Request Automated ProcEss

Home | User Help | Admin Information | FAQs | CCHD Area

Welcome to SARAPE

Initiate Request

SARAPE is a web-based application that allows users within
NWC, and ASC University Alliance
ccounts for selected resources at
ncement to the current manual
same information electronically from
e manual forms. The SecureNet
rsite Unclassified Guest Account
et Classified Guest Account
naintained at Livermore National
the ExtraNet Network that is
the restricted domains.

itors with need-to-know authorization may request computer
ions, and authorized access such as Kerberos and DCE
g accounts may also be requested, as well as the deletion of
eeded.

e internal process established at each site to process access
al verification and approvals for guest requests.

Getting Started

- Available Resources
- Example Requests
- Who is the contact at my site?

or an application. Authorized access is required. If you are not requesting guest access or changes to an existing guest account, please use the internal process at your site for requesting an account. If you need assistance with SARAPE, please contact sarape-help@sandia.gov.

If you are at an ASC Alliance Partner requesting an unclassified machine at LLNL, you must also complete the Livermore Computing Policies form at **https://www.llnl.gov/lcforms/policies/Policies.htm** and the Livermore Computing Computer Security Briefing at **http://www.llnl.gov/computing/forms/policies/briefing_i.pdf**. Give these forms to your GPA. Example requests can be found under *Examples,* below.

Back to top of page

Example Requests

- New Access to an Application
- New Computer Account
- New Other Access
- Renew Access and/or Account(s)
- Delete Access and/or Existing Computer Account(s)

Back to top of page

Initiate Request

© Sandia Corporation | Lilia Martinez | Privacy and Security

Sandia National Laboratories

Filling out the form

*Kerberos User Name(s) at Requestor's Site: Uncl.: Secure:
*Required for certain resources.
(Note: User Names are case sensitive.)

*U.S. Citizen? Yes No

*Date of Birth: (Use MM/DD/YYYY format.)

*Place of Birth: (city and state/country)
*Required for LANL secure resources and for SNL resources if non-U.S. citizen.

*Social Security Number: (Use 888-88-8888 format.)

*Reenter to Verify SSN:
*SSN not required for deletions. Required for U.S. Citizens for all other requests. If non-U.S. citizen, provide if available.

*Requestor's site:

CIP Request? (NISAC) Yes No

If CIP Request, Company Affiliation:
(Note: For CIP requests only, if Requestor's site is not listed.)

*Request is to:
- Access an Application (e.g., U/IMS, U/WFS)
- Open a New Computer Account
- Obtain Other Access (e.g., cross cell, kerberos, VPN)
- Renew Access and/or Account(s) at
- Delete Access and/or Existing Computer Account(s)

Next...

Purdue University

guest processing agent (GPA)

Ananth Grama, ayg@cs.purdue.edu, 765-494-6964

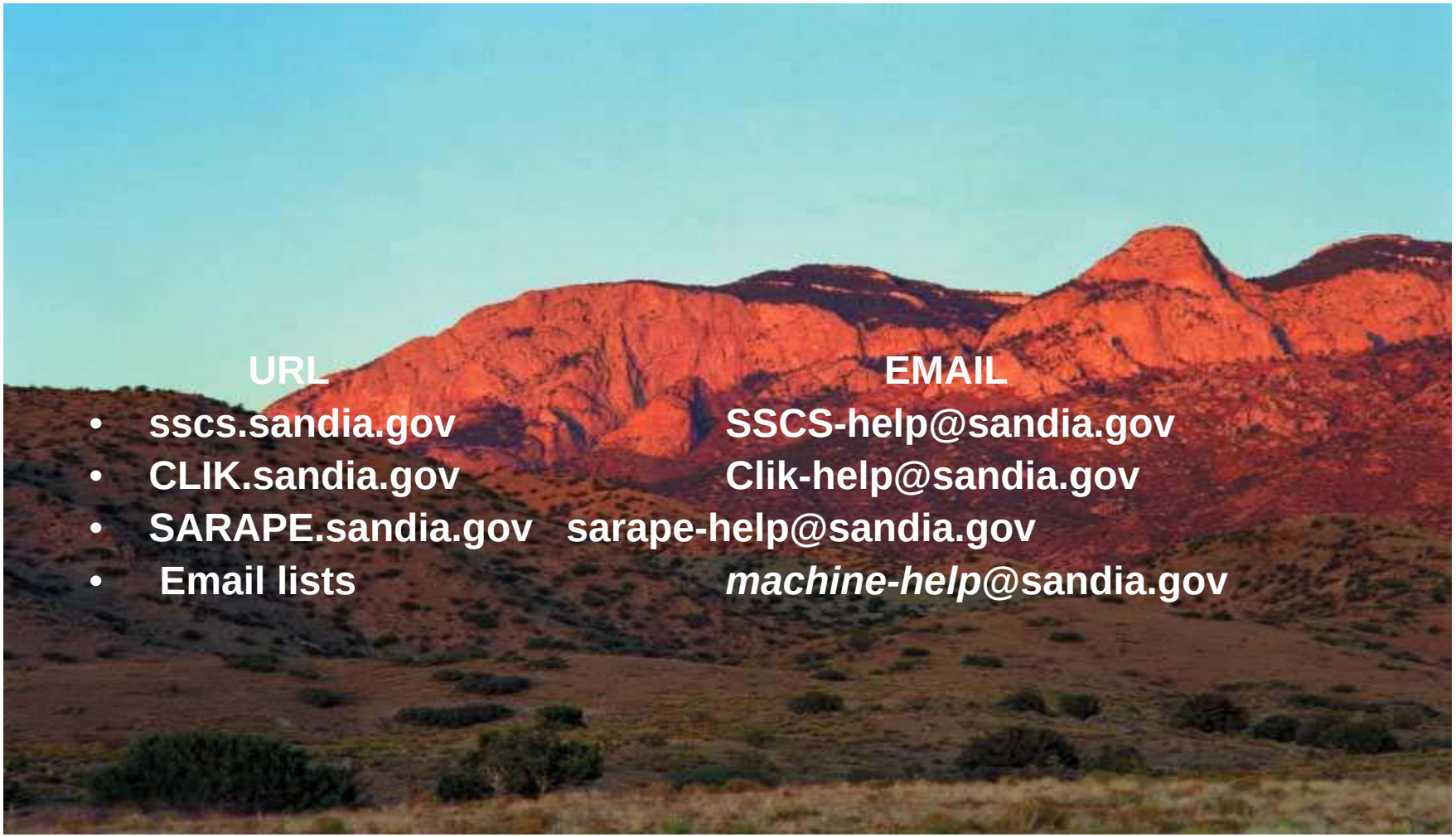
SARAPE Help

email: sarape-help@sandia.gov

Advanced Simulation & Computing™

Sandia National Laboratories

Sandia Scientific Computing Support - SSCS



URL	EMAIL
• sscs.sandia.gov	SSCS-help@sandia.gov
• CLIK.sandia.gov	Clik-help@sandia.gov
• SARAPE.sandia.gov	sarape-help@sandia.gov
• Email lists	*machine-help*@sandia.gov

In Conclusion ….

Sandia provides an exciting atmosphere to work and an exceptional opportunity to learn

in a unique part of the country.

Please let us know if you are interested in obtaining an internship with us.