Dark Storm: Further Adventures In XT Architecture Flexibility

John P. Noe

Robert A. Ballance

Geoff McGirt

Jeff Ogden

Sandia National Laboratories

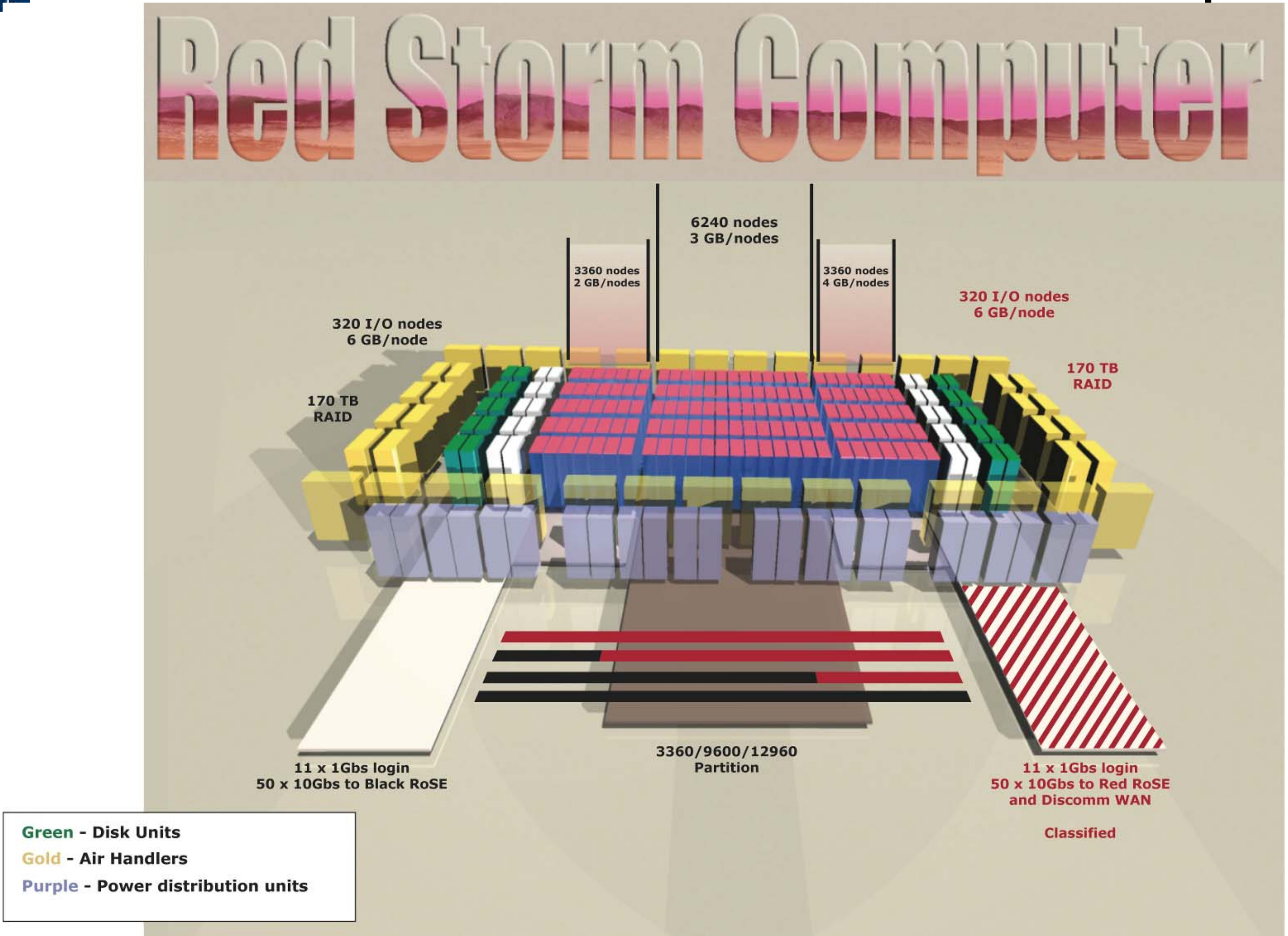
May 27, 2010

Sandia National Laboratories is a multi-program laboratory operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin company, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Outline

- Cray/Sandia Red Storm system
- Flexibility aspects of Red Storm
- Follow-on mission for Sandia and Red Storm
- Experiences in building Dark Storm
- Things remaining to accomplish
- Questions

Diagram of Red Storm shows Red /Black switch configurations and dual head concept.



Red Storm, Architected for ”ility”: Reliability, Adaptability, Flexibility

Cray/Sandia Red Storm functional segregation explicitly enabled upgrades and flexibility

- Service and I/O nodes
- Compute nodes (single to dual to quad core!)
- Interconnect network (mezzanine card)
- System Management
  - SMW – hold many configurations
  - Boot disk – easily replicated

Red Storm: Architected for Flexibility

Sandia Red/Black switch adds further flexibility for system configuration and resource allocation

- Upgrades done on small system first
  - S/W upgrades
  - Firmware updates
  - Quad-core upgrades done on small side
- After successful implementation system reconfigured to LARGE and subsequently to JUMBO
- All upgrades done at lowest security level initially

Red Storm: Architected for Flexibility

This flexibility extended to disk subsystem/file system.

- Initial disks deployed were DDN - < 400 TB
- Augmented by LSI - > 1.2 PB

- Lustre supported on Catamount as well as CLE

Dark Storm: New National Security Mission for Red Storm

NNSA User Facility for Capability Computing decision removed Sandia from ASC capability system rotation (2007). NNSA conceived new expanded mission for Red Storm in National Security area inspired by Operation Burnt Frost (2008).

Mission Requirement: Support high classification customers (plural) whose data must not intermingle. Provide access to full capabilities of Red Storm system as required to address urgent National Security issues.

Dark Storm: New National Security Mission for Red Storm

NNSA agrees to transition period for Dark Storm operational model, but requires NS customers to commit to support.

Operational Challenge: Demonstrate high capacity parallel I/O system support with potential for multiple clients with serial access to Dark Storm system. Goal: 50 GB/sec.

Technical constraints: Sanitizing disk not viable (or sufficient) for Petabyte sized file systems. Need more flexible solution.

Dark Storm Challenge

Demonstrate network attached file storage with sufficient bandwidth to support existing applications at scale of 30,000+ processors.

Provide access to data independent of state of Dark Storm system

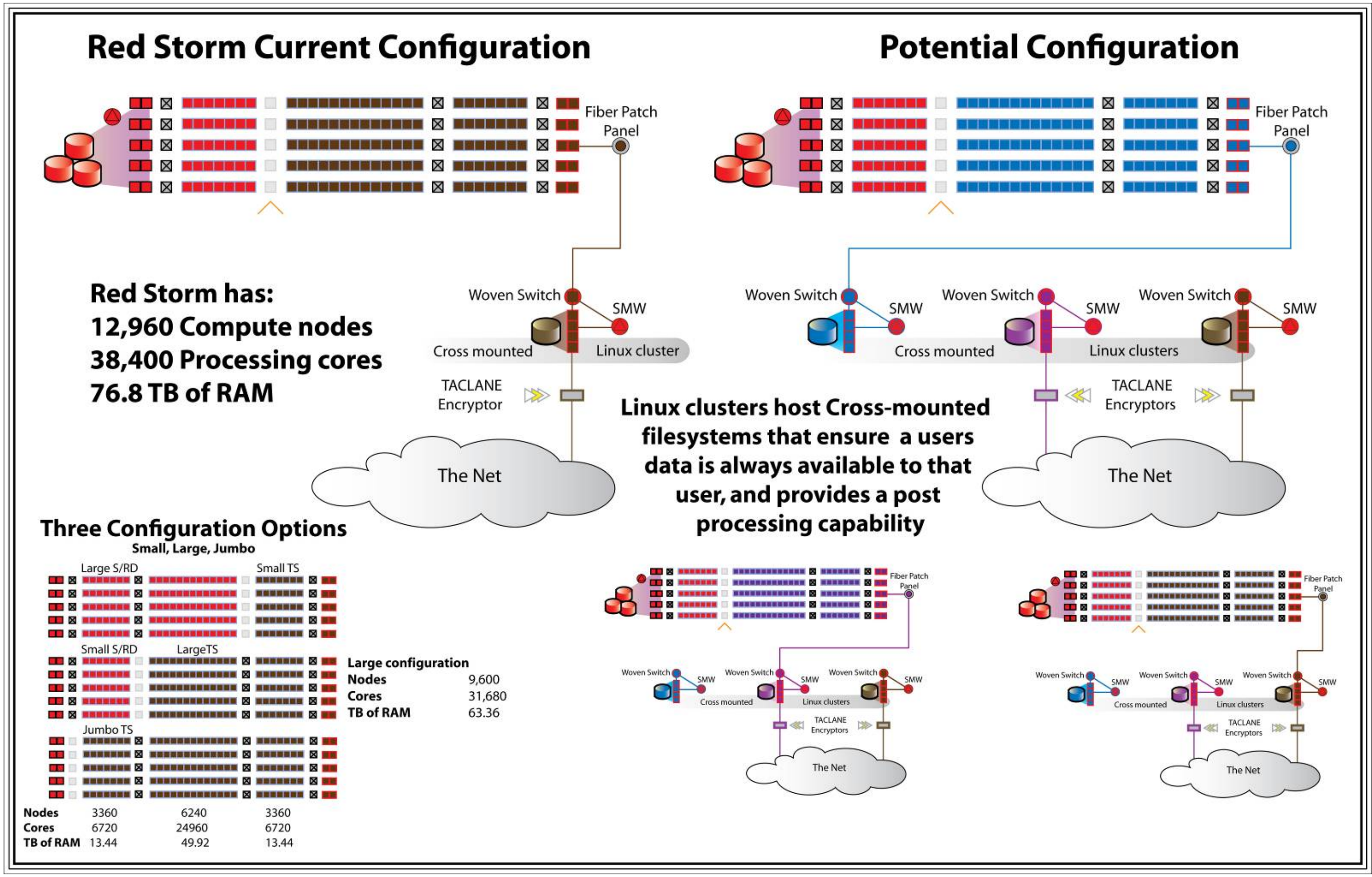
Provide viable solution for multiple clients

Dark Storm Solution

Integrate Linux cluster serving Lustre through High Performance Woven 10GE switch.

- Catamount limitation: Lustre 1.4
- Initial implementation should support up to 20 GB/sec bandwidth with higher potential
  - Employ LNET router option
  - Throttle application code I/O if necessary
- Boot raid and SMW also switch with Lustre server. Login nodes is through the Woven switch. Analogous concept to the Red/Black switch.

Dark Storm Solution



Red Storm Current Configuration

Red Storm has:
12,960 Compute nodes
38,400 Processing cores
76.8 TB of RAM

Potential Configuration

Linux clusters host Cross-mounted filesystems that ensure a users data is always available to that user, and provides a post processing capability

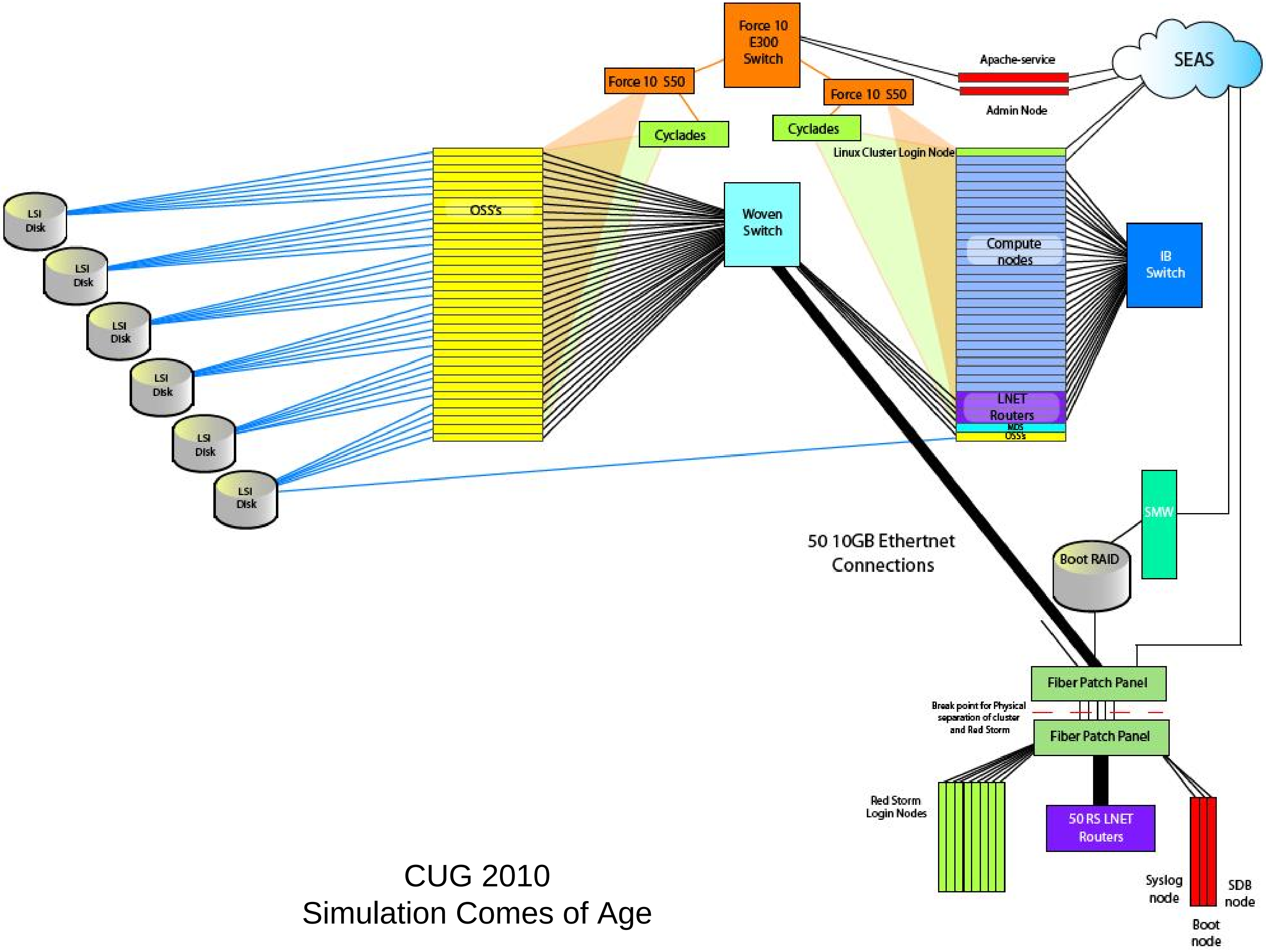
Three Configuration Options

Small, Large, Jumbo

Nodes	3360	6240	3360
Cores	6720	24960	6720
TB of RAM	13.44	49.92	13.44

Large configuration	
Nodes	9,600
Cores	31,680
TB of RAM	63.36

Dark Storm Solution: Apache Lustre Cluster



Issues identified and addressed

Application I/O overdrives Lustre 1.4 on cluster

Solution: Employ baton passing within application to limit number of nodes writing output simultaneously. Demonstrated ability to support runs up to 12,000 cpus.

Running Catamount Lustre 1.4 clients and Lustre 1.4 servers. Investigated upgrade to 1.6 server but did not implement (yet!).

Issues identified and addressed

Application stalls on I/O, aka Lustre “thrashing” sometimes for hours, sometimes cures itself, resumes and continues on.

Solution: Deep dive debugging includes watching network traffic, checking Lustre kernel debug logs and LNET router logs. Discovered timeout issue with LNET side sending unsolicited RPCs to clients. Lustre Bug # 18938 patched on 1/22/2010 corrected the problem.

More detail on condition

- In Catamount, LibLustre only processes RPC traffic when application does I/O
- LibLustre assumes it is free to process unsolicited RPCs whenever (even hours later)
- LNET routers are connection oriented, timeout any RPCs delivered to LibLustre, drop connection to clients
- Catamount client tries I/O, LNET tries to reconnect clients and server, thrashing created by large number of clients!

- What caused the problem in the first place? We shot ourselves in the foot!
- chmod o-rwx /scratch/USERDIR …. run on Linux cluster to enforce policy!

- directory mods force async lock cancellations, RPCs to clients.
- Lustre mod corrects by permitting LNET small RPC messages to Catamount
- Unique Catamount and LNET router configuration…..

Issues Remaining

Lustre support declining for older releases and for Catamount in particular.

Potential Solution: Move to CLE and newer Lustre (1.6 exhibits better behavior)

Potential Solution: Move to CLE and Panassas (ala Cielo configuration)

Contra-indicators: All Applications would need to be rebuilt/validated. New I/O environment. Need to revamp all development platforms. 1.6 nearing end of life support also!

Issues Remaining

Demonstrate multiple customers usability. Requires additional hardware and security tests.

Demonstrate full scale I/O rates without application throttling. (pretty confident about this…need more disk and Lustre 1.6 or more)

Settle on upgrade path to current Lustre version

Improve debugging ability in high security environments!

Questions?

Thanks to Cray, and Sun/CFS support, Ruth Klundt, Lee Ward at Sandia.