Red Storm Update

Jim Tomkins

Fall Creek Falls Conference

September 7 - 10, 2008

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.

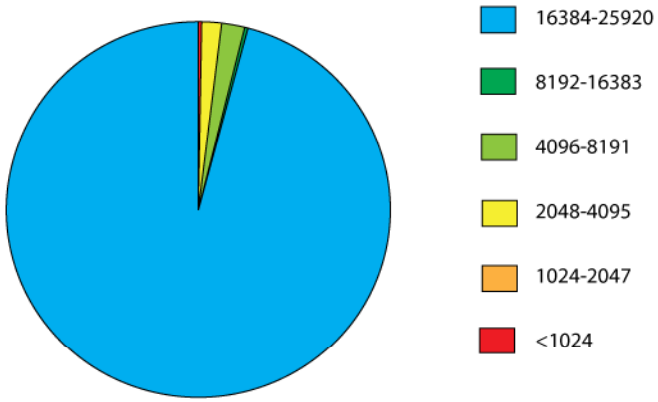
Outline

- Red Storm Changes
  - Catamount N-Way (CNW) Work
  - Quad-Core Performance Measurements
  - Additional Storage
  - Quad Core Upgrade

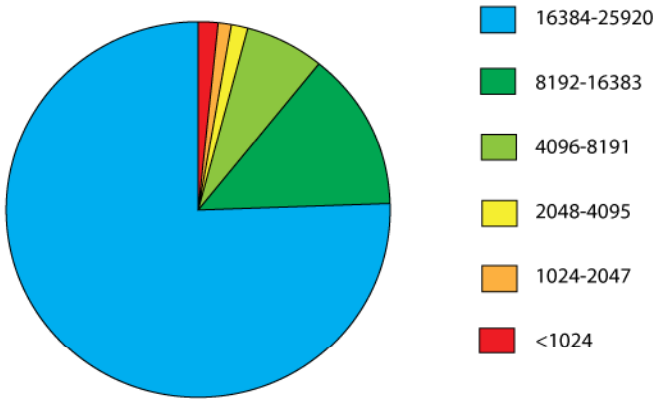
Architectural Goals

- Balanced System Performance: CPU, Memory, Interconnect and I/O
- **Scalability:** System Hardware and System Software scale, a single cabinet system to 32K node system
- Functional Partitioning: Hardware and System Software
- Reliability: Full system Reliability, Availability, Serviceability (RAS) designed into Architecture
- **Upgrade-ability:** Designed in path for system upgrade
- Red/Black Switching: Flexible support for both classified and unclassified Computing in a single system
- Custom Packaging: High density, relatively low power system
- Price/Performance: Excellent performance per dollar, use high volume commodity parts where feasible

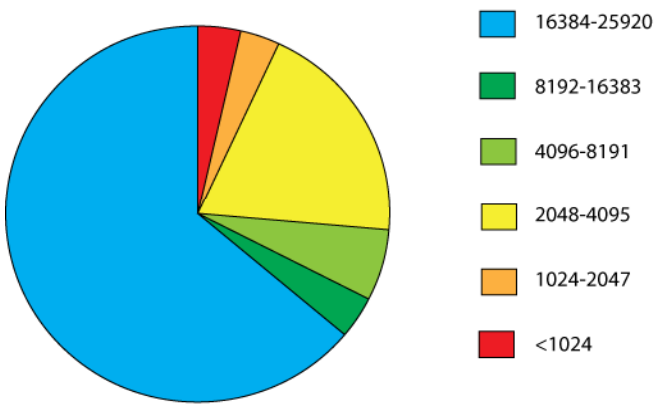
Red Storm Jumbo Mode, 12/8/07 – 3/11/08
System CPU Hours by Job Size



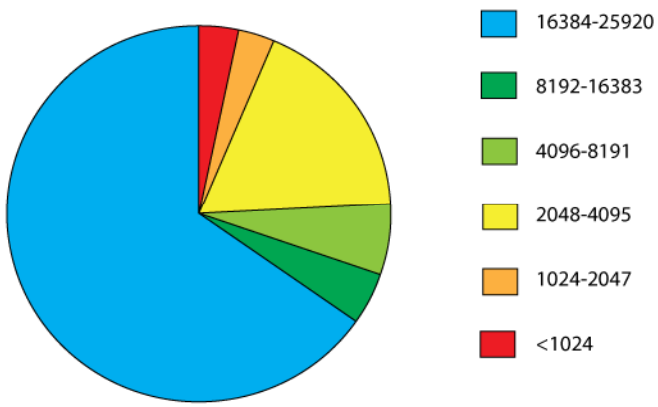
Classified
Opus Computation



Unclassified
All Computations



Classified
All Computations



All Computations

National Nuclear Security Administration

Sandia National Laboratories

Red Storm Impact on Supercomputing

- Cray and Sandia with the support of NNSA/ASC have created one of the most successful new supercomputers ever.
- There are now over 20 sites with Red Storm (XT3 or XT4) computer systems with over 35 system installed.
- The Red Storm impact is continuing with the XT5.

Catamount N-Way LWK OS

- SUNMOS, research project on nCUBE2 that became production OS on Sandia Paragon Computer
- PUMA, second generation LWK
- Cougar, production version of PUMA on Sandia ASCI Red Computer
- Catamount, production OS for initial Red Storm
- Catamount VN, production OS for Dual-Core Red Storm
- Catamount N-Way, production OS for heterogeneous (Dual-Core and Quad-Core) Red Storm - (Funded by DOE Office of Science)
  - Tested on Jaguar (Quad-Core) at full scale
  - Operational on Red Storm dual-core since June
  - Running on heterogeneous Red Storm now

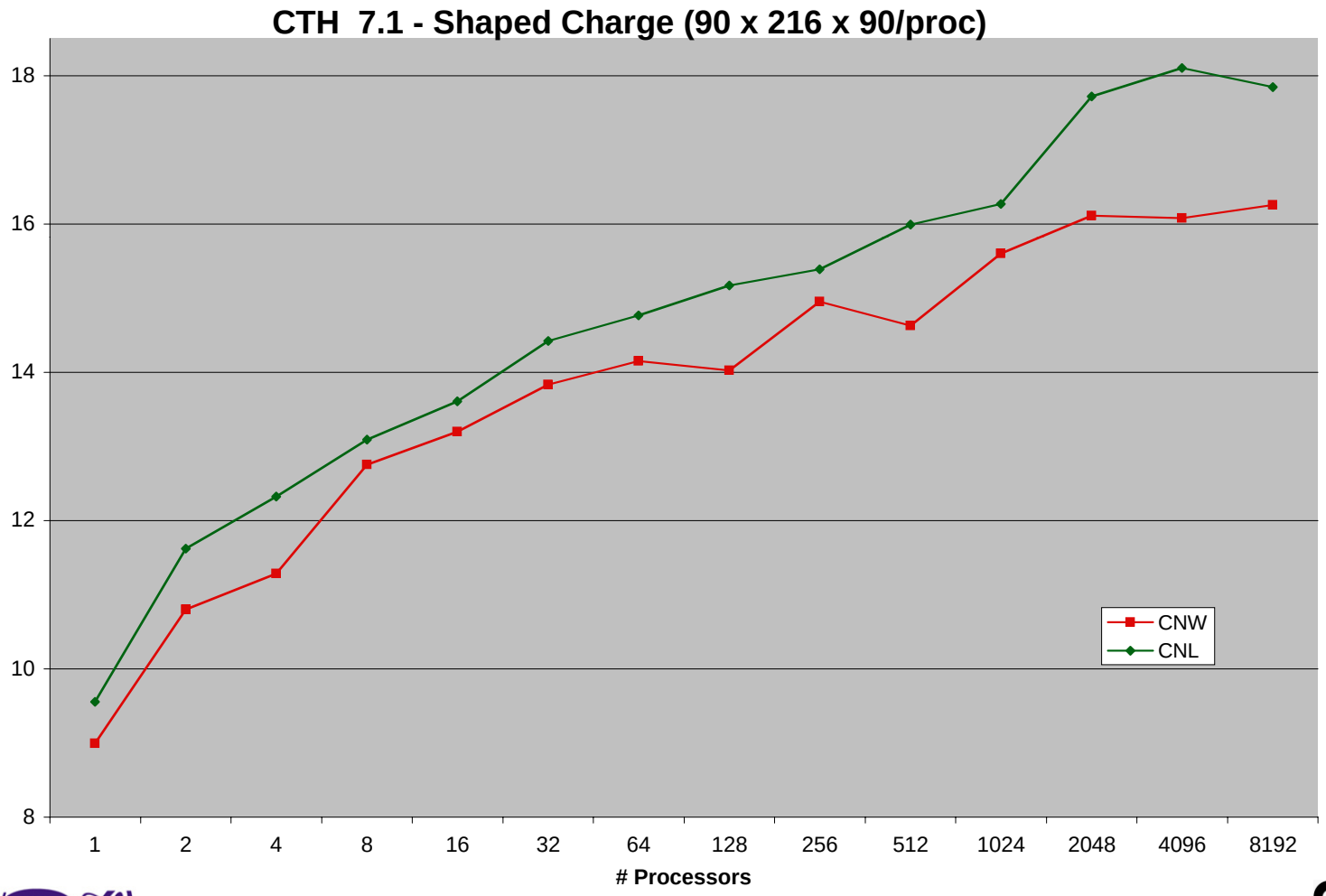
Catamount N-Way Testing
HPCC Quad-Core

HPCC Benchmark	Mumber of MPI Ranks	Cores per Node	CNL	CNW	CNW/ CNL
PTRANS	4	1	0.567	1.606	2.83
HPL	4	1	17.88	17.90	1.00
STREAMS	4	1	25.21	25.84	1.02
Random	4	1	0.0064	0.0118	1.83
FFT	4	1	1.609	1.646	1.02
PTRANS	4	2	0.488	1.551	3.18
HPL	4	2	17.78	18.03	1.01
STREAMS	4	2	16.45	18.11	1.10
Random	4	2	0.0061	0.0115	1.88
FFT	4	2	1.337	1.36	1.02

Catamount N-Way Testing
HPCC Quad-Core

HPCC Benchmark	Mumber of MPI Ranks	Cores per Node	CNL	CNW	CNW/ CNL
PTRANS	4	4	0.287	1.244	4.33
HPL	4	4	17.59	17.72	1.01
STREAMS	4	4	7.85	9.95	1.27
Random	4	4	0.0060	0.0115	1.92
FFT	4	4	0.902	0.959	1.06

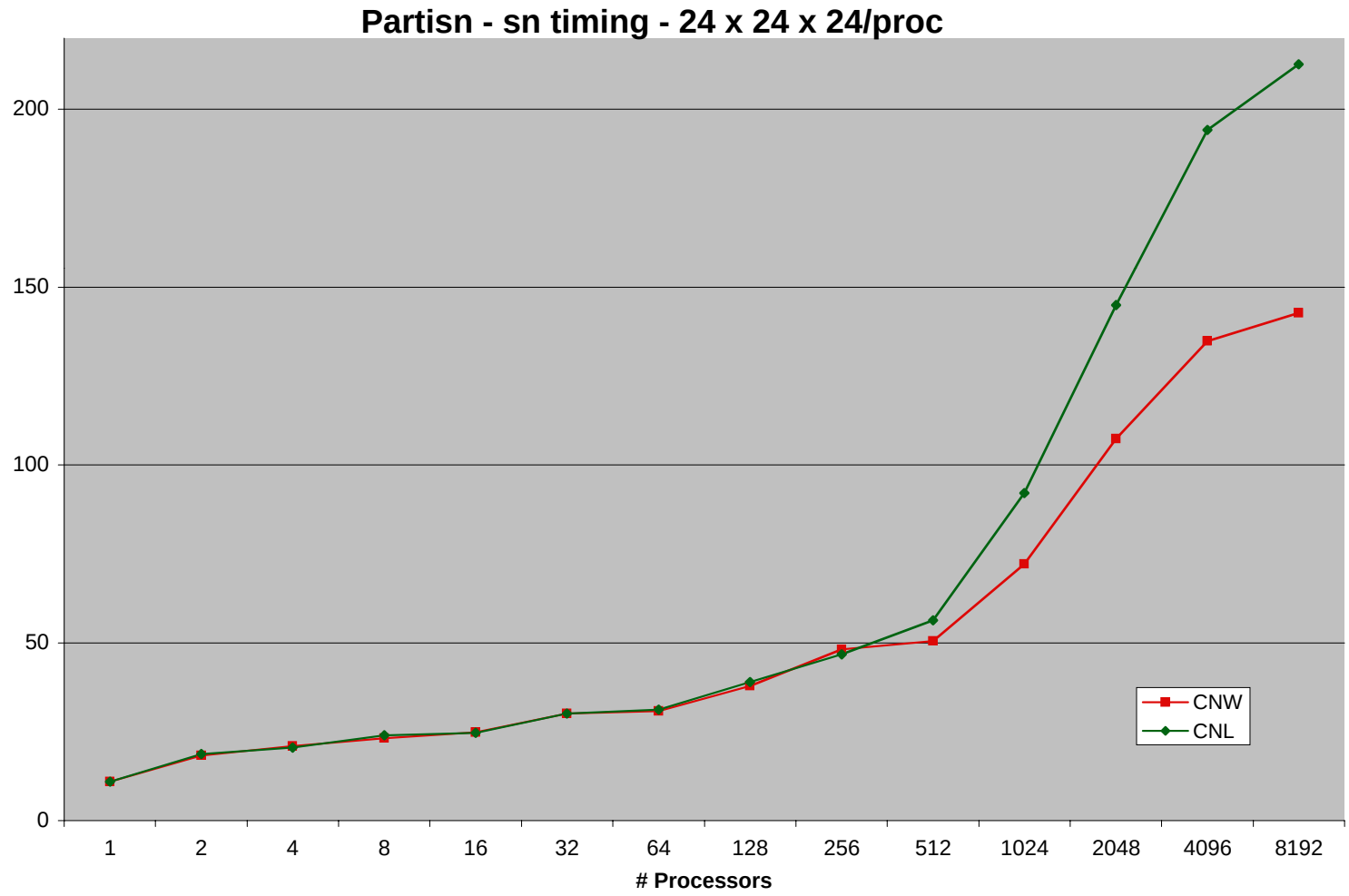
Catamount N-Way Testing
Large Scale Dual-Core - CTH



National Nuclear Security Administration

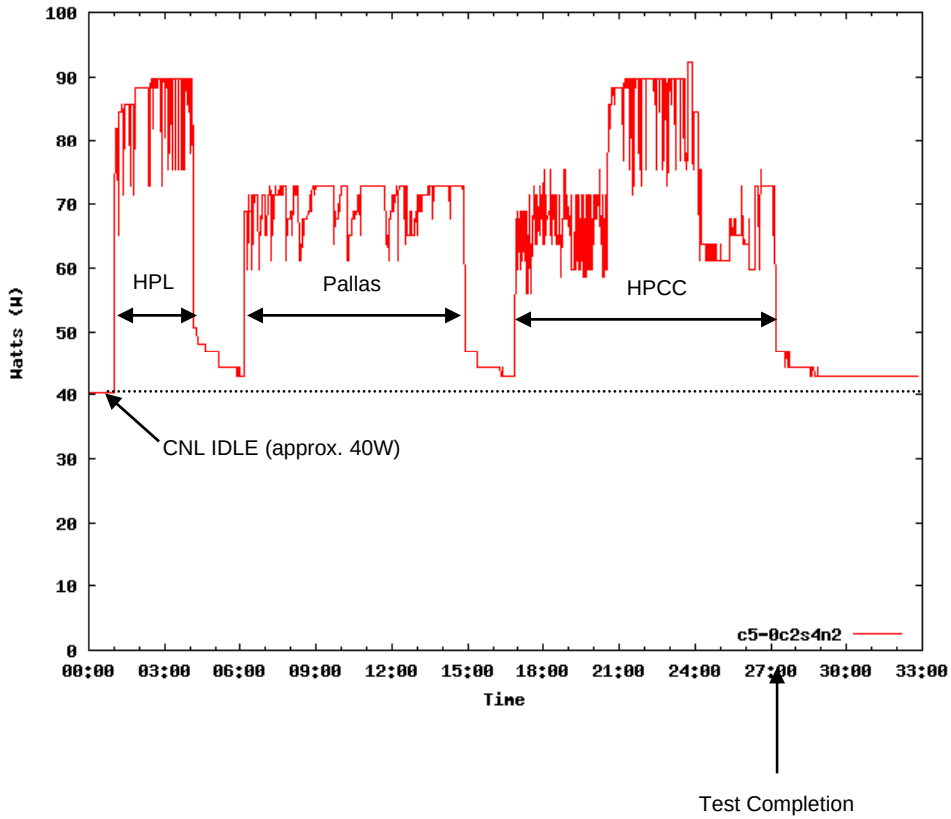
Sandia National Laboratories

Catamount N-Way Testing
Large Scale Dual-Core - Partisn

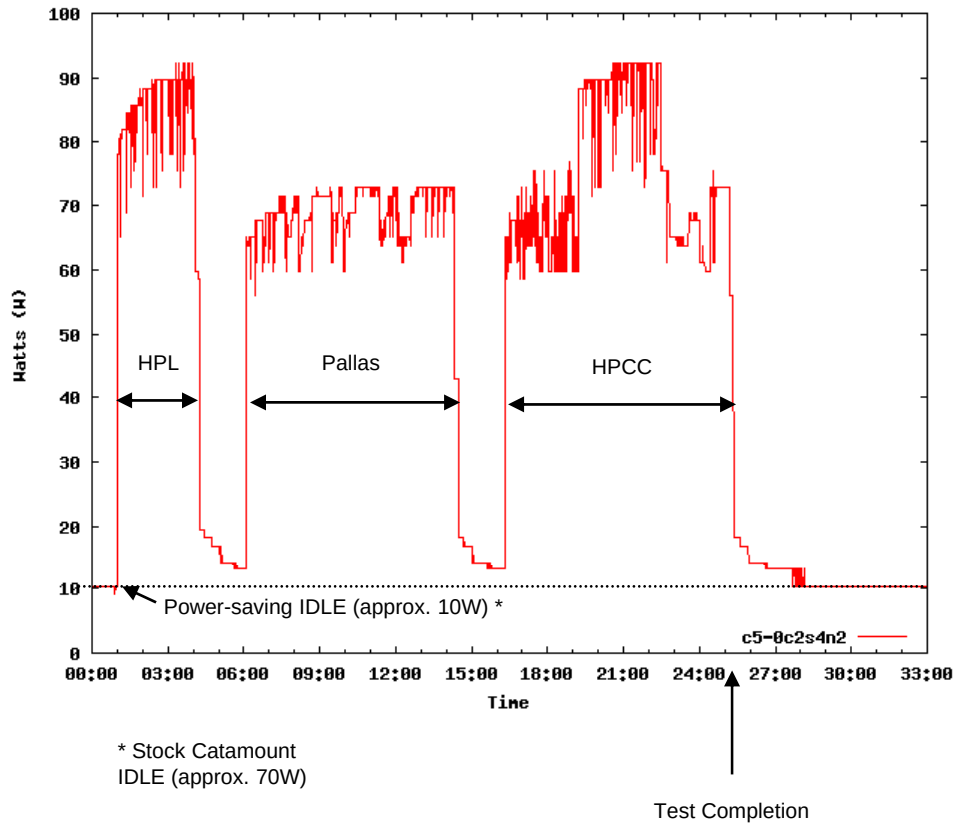


Application Power Signatures

Compute Node Linux (CNL) Power Signatures



Catamount N-Way Power Signatures



Quad-Core Testing
CTH

- Shock hydrodynamics
- Shaped charge problem with explosives and meso-scale problem with a large number of materials
- VN2 mode takes 10% to 13% more time than SN mode
  - Consistent with dual core results
- VN4 mode takes 39% to 52% more time than SN mode

time on 4 cores	SN mode	VN2 mode	VN4 mode
shaped 4	8.385	9.338	11.832
shaped 2	4.534	5.056	6.901
meso 2	12.312	13.863	18.023
meso 1	6.482	7.076	8.953

Quad-Core Testing
Sage

- hydrodynamics code
- timing_c problem with 250000 cells/PE
- Took 5.8% longer in VN2 mode than SN mode
  - Consistent with dual core results
- Took 26.3% longer with VN4 mode than SN mode

4 cores	SN mode	VN2 mode	VN4 mode
CC/sec/PE	6181.6	5823.6	4556.2

Quad-Core Testing
Partisn

- Time-Dependent, Parallel Neutral Particle Transport Code
- Reports a Transport and Diffusion Grind Time in nanoseconds
- Takes 28% to 46% longer in VN2 mode than in SN mode
- VN4 mode takes 2.2 to 2.8 times as long as SN mode

grind times 4 cores	SN mode	VN2 mode	VN4 mode
184 MB/core	32.56 / 28.23	47.55 / 39.45	90.78 / 74.07
367 MB/core	41.73 / 29.11	53.52 / 38.51	92.02 / 71.02
1444 MB/core	34.70 / 26.93	48.43 / 35.37	89.45 / 65.53

Quad-Core Testing
Small Pages vs Large Pages

- Have 128 TLB entries for large pages (2MB) with quad core instead of 8 on our original Opterons
  - Number for small pages (4KB) still 512
- Most applications run slightly (< 3%) better with large pages
  - Change from original Opterons
- Some benchmarks ran significantly better with large pages
  - Random Access - 14% to 23% better
  - PTRANS - 2.8 to 3.2 times better

Effective Use of Cores

Application	Utilization of each core	Cores effectively used
CTH	71%	2.84
PRONTO	79%	3.18
SAGE	74%	2.95
UMT2K	91%	3.62
PARTISN	40%	1.60

Performance Comparison
Dual-Core vs Quad-Core

	Time in Seconds		
Processor (Socket)	2.4 GHz Dual (940)	2.2 GHz Quad (AM2)	2.6 GHz Dual (AM2)
CTH SN	1938.5	1949.3	1715.9
CTH VN	1357.0	1184.3	1092.4
Pronto SN	3504.0	3503.0	3198.2
Pronto VN	1674.7	1594.4	1456.6
Partisn SN	619.3	425.6	513.7
Partisn VN	297.1	172.2	283.9
UMT2K SN	7108.8	5517.6	6488.4
UMT2K VN	3933.8	2939.9	3512.9

Disk Storage Upgrade

- 2.36 PB of New Disk (~1.8 PB of Formatted Storage)
  - 1640 TB on the Red End
  - 720 TB on the Black End
  - LSI XBB2 Controllers
- 70 GB/s of Aggregate Bandwidth
- Successfully Passed Acceptance Testing
- Currently Undergoing Lustre Testing

Red Storm Upgraded to ~284TF

- Upgrade 65 cabinets of Existing System to Quad Core XT4 Specifications
  - New Compute Boards
  - New 2.2 GHz AMD Opteron Quad Core Processors
  - New DDR2 6400 Memory - ~49 TB (2 GB/core)
  - Memory Upgrade to other 70 cabinets - ~26 TB (2 GB/core)
  - New VRMs
  - Reuse Seastar Mezzanines, L0s, Cabinets, Cables, PDUs, etc.

- Timeframe
  - 50 of the 65 Cabinets Completed
  - Last 15 Cabinets are Being Installed Now
  - Acceptance Testing will begin on 9/19/08

Red Storm Configuration Over Time

	Red Storm (05)	Red Storm (06)	Red Storm (08)
Theoretical Peak Performance	43.52 TF	130.56 TF	290.30 TF
MP-Linpack Performance	36.19 TF	102.2 TF	TBD
Total Memory	33.4 TB	39.2 TB	78.75 TB
System Memory B/W	57.99 TB/s	78.14 TB/s	126.29 TB/s
Disk Storage (User Formatted) (Red Black)	170 TB 170TB	170 TB 170 TB	1.5 PB 710 TB
Parallel File System B/W (Red Black)	100 GB/s 50 GB/s 50 GB/s	100 GB/s 50 GB/s 50GB/s	170 GB/s 100 GB/s 70 GB/s
External Network B/W (Red Black)	25 GB/s 25 GB/s	25 GB/s 25 GB/s	25 GB/s 25 GB/s

National Nuclear Security Administration

Sandia National Laboratories