LLNL ASC Resources

Blaise Barney
LLNL
Services & Development Division

Los Alamos NATIONAL LABORATORY

Sandia National Laboratories

NNSA National Nuclear Security Administration

Lawrence Livermore National Laboratory

SDSC

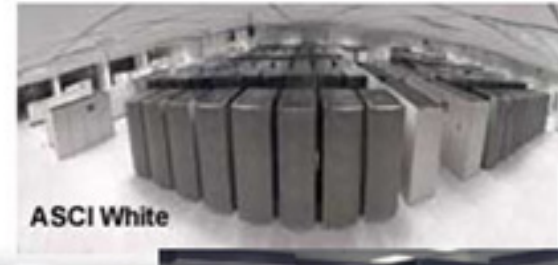
ILLINOIS

THE UNIVERSITY OF CHICAGO

CALIFORNIA INSTITUTE OF TECHNOLOGY

ASC™

THE UNIVERSITY OF UTAH



UCRL-PRES-208318

Overview

- **Hardware Environment**
  - ASC Blue
  - ASC White
  - ASC Purple
  - Parallel File Systems
  - HPSS
  - ALC & Frost Batch Queues
  - Alliance YTD Usage
- **Software Environment**
  - AIX, PSSP, CHAOS, SLURM
  - Compilers
  - Tools
- **Training**
- **Futures**
  - Future Plans
  - Blue Gene/L
  - Terascale Simulation Facility
  - SC2004

The Matrix

System	Program	Manufacturer & Model	Operating System	Interconnect	Nodes	CPUs	Memory (GB)	Peak GFLOP/s
Unclassified Network (OCF)								**420,292**
BlueGene/L	ASCI	IBM	Linux		65536	131,072	32,768	367,002
Thunder	M&IC	California Digital	CHAOS 2.0	Elan4	1024	4,096	8,192	22,938
MCR	M&IC	Linux NetworX	CHAOS 2.0	Elan3	1152	2,304	4,608	11,059
ALC	ASCI	IBM xSeries	CHAOS 2.0	Elan3	960	1,920	3,840	9,216
UV (pEDTV)	ASCI	IBM p655	AIX 5.2	Federation	128	1,024	2,048	6,144
Frost	ASCI	IBM SP	AIX 5.1	Colony DS	68	1,088	1,088	1,632
TC2K	M&IC	Compaq SC ES40	Tru64 5.1b	Elan3	128	512	280	683
iLX	M&IC	RAND Federal	CHAOS 2.0	N/A	67	134	268	678
PVC	VIEWS	Acme Micro	CHAOS 2.0	Elan3	64	128	128	614
GPS	M&IC	Compaq GS320/ES45	Tru64 5.1b	N/A	49	160	344	277
Qbert	M&IC	Digital 8400	Tru64 5.1b	MC 1.5	2	20	24	25
Riptide	VIEWS	SGI Onyx2	Irix 6.5.13f	8 IR2 Pipes	1	48	37	24
Classified Network (SCF)								**134,482**
Purple	ASCI	IBM SP	AIX 5.3	Federation	1528	12,224	48,896	99,503
White	ASCI	IBM SP	AIX 5.1	Colony DS	512	8,192	8,192	12,288
Lilac (xEDTV)	ASCI	IBM xSeries	CHAOS 1.2	Elan3	768	1,536	3,072	9,186
UM (pEDTV)	ASCI	IBM p655	AIX 5.2	Federation	128	1,024	2,048	6,144
Adelie	ASCI	Linux NetworX	CHAOS 1.2	Elan3	128	256	512	1,434
Emperor	ASCI	Linux NetworX	CHAOS 1.2	Elan3	128	256	512	1,434
Ace	ASCI	Rackable Systems	CHAOS 1.2	N/A	128	256	512	1,434
S (Blue-Pacific)	ASCI	IBM SP	AIX 5.1	TB3	488	1,952	1,164	1,296
GViz	VIEWS	Rackable Systems	CHAOS 1.2	Elan3	64	128	256	717
Ice	ASCI	IBM SP	AIX 5.1	Colony DS	28	448	448	672
SC Cluster	ASCI	Compaq ES40/ES45	Tru64 5.1b	N/A	40	160	384	235
Whitecap	VIEWS	SGI Onyx3800	Irix 6.5.13f	4 IR3 Pipes	1	96	96	77
Tidalwave	VIEWS	SGI Onyx2	Irix 6.5.13f	16 IR2 Pipes	1	64	24	38
Edgewater	VIEWS	SGI Onyx2	Irix 6.5.13f	10 IR2 Pipes	1	40	18	24

ASC Blue

- **Blue (OCF)**
  - decommissioned 8/5/04
- **SKY (SCF)**
  - decommission scheduled 10/04

R.I.P.
"Gone but not forgotten"

ASC White

- **Frost (OCF)**
  - 68 total nodes
  - 64 compute nodes
  - 16 GB memory/node
  - 1.6 TFlop
  - 20+ TB parallel file system

- **White / Ice (SCF)**
  - 512 / 28 total nodes
  - 489 / 26 compute nodes
  - 16 GB memory/node
  - 109 / 5.7 TB parallel file systems

- **IBM POWER3 technology**
  - 16 CPUs/node
  - 375 MHz clock
  - 1500 Mflops/CPU
  - 8MB L2 cache/CPU
  - 64-bit architecture

- **Frost to be retired in Jan 2005 timeframe!**
  - Necessitated by move to TSF
  - Still being heavily utilized by Illinois, Stanford and Chicago
  - Need to begin moving to ALC

ASC Purple

- **ALC**
  - Unclassified component available for Alliance use
- **UM and UV**
  - "Identical" 128 node, POWER4 systems; 8 cpus/node
  - 6.1 TFlop each
  - UM is classified, UV will be moved to classified side in Jan 2005 timeframe
- **PURPLE**
  - 1528 node, POWER5 system; 8 cpus/node
  - 100 TFlop
  - Classified
  - Delivery in mid 2005 timeframe
  - Will reside in our new TSF building

Servers > UNIX servers > News >

Press release

IBM reaches POWERful milestone

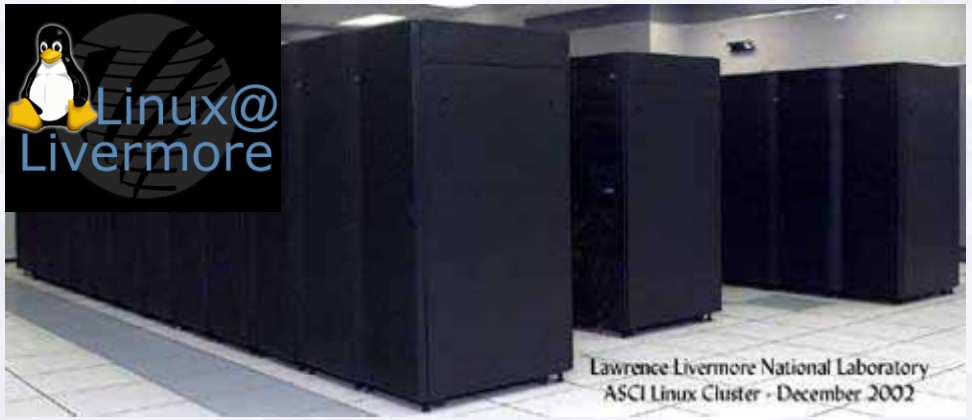
Next generation POWER5-based servers up and running

ARMONK, NY, June 27, 2003—IBM today announced that the first servers based on its next generation POWER5™ microprocessors are up and running in IBM's Poughkeepsie Labs. Initial internal performance tests indicate that POWER5 based IBM @server™ systems are expected offer four times the system performance over the first POWER4™ based servers[1].

POWER5 based IBM @server systems are expected to start shipping to customers next year in IBM @server pSeries™ and iSeries™ systems as well

ALC

- **ASC Linux Cluster**
  - Unclassified component of ASC Purple
  - 9.2 TFlop system
  - 960 nodes
  - Each node has 2 Pentium4 Xeon processors
  - 4 GB memory per node
  - Quadrics switch
  - 72 TB Lustre parallel I/O file system

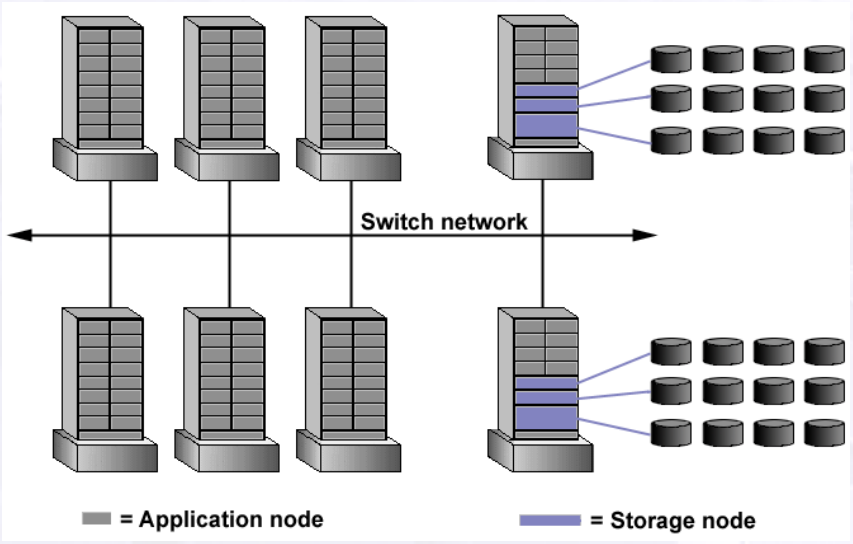


- **Alliance Resource**
  - 1/2 of ALC is devoted to Alliance use
  - 1/2 remains in testing and development mode for Lustre file system
  - Whole system runs can be requested

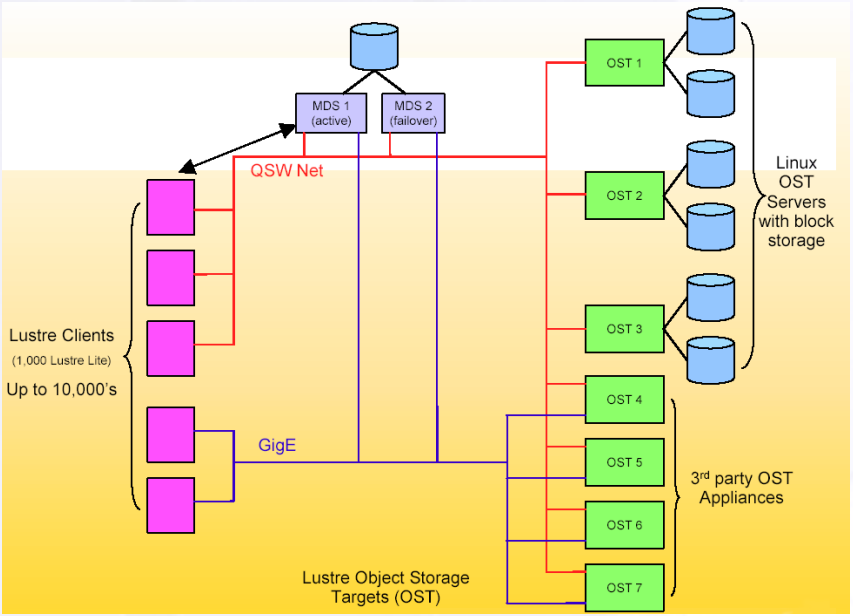
- **Configuration (usual)**
  - 454 node batch pool
  - 8 node debug pool
  - 462 node lustre testing pool

Parallel File Systems

- **GPFS**
  - IBM's General Parallel File System
  - All LLNL ASCI systems have their own, multi-terabyte GPFS file system(s)
  - Frost performance with 60 client nodes and 2 server nodes:
    Write = 550 MB/sec
    Read = 600 MB/sec

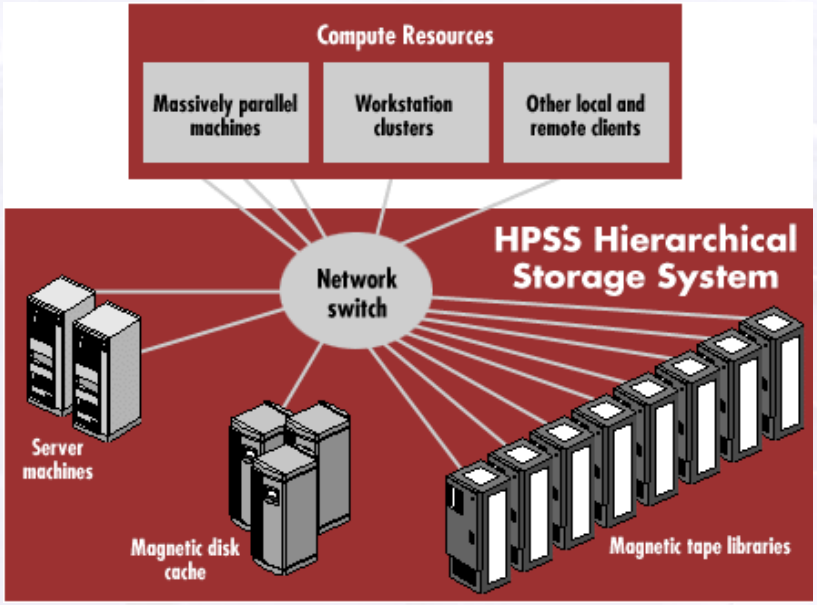


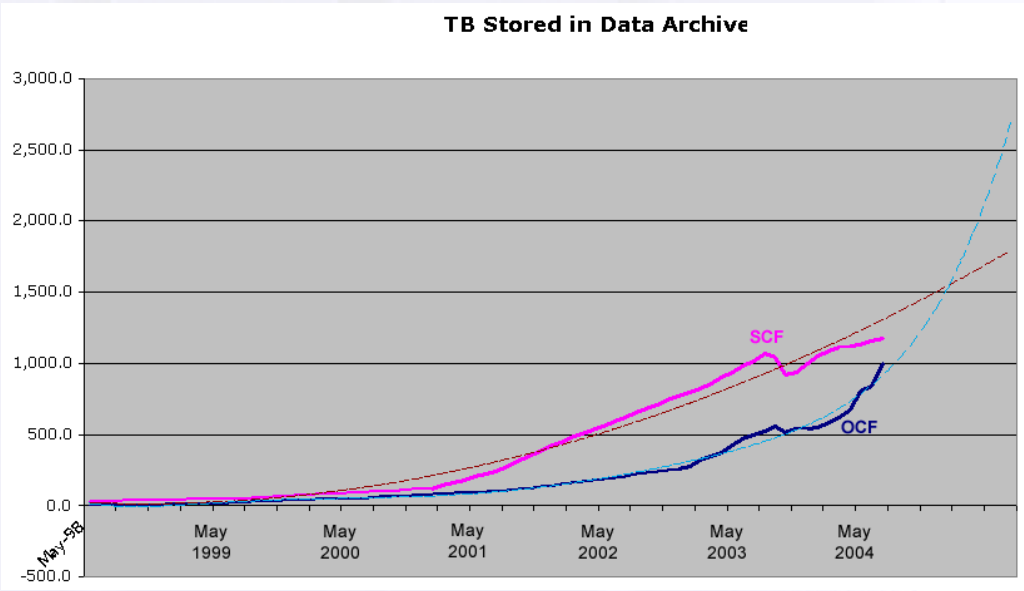
- **Lustre**
  - Linux cluster based parallel file system from Cluster File Systems, Inc.
  - Goals: clusters with 10,000's of nodes, petabytes of storage, move 100's of GB/sec with state of the art security and management infrastructure.
  - Currently running on Livermore's ALC, PVC, Thunder and MCR machines.
  - 6.1–8.0 GB/s parallel I/O
  - Still being developed and tested
  - See **www.lustre.org** for more info



HPSS Archival Storage

- **Integrated into the OCF and SCF gigabit ethernet networks**
- **Some metrics**
  - OCF: 2.2 PB capacity; @821 TB used
  - SCF: 3.3 PB capacity @1.11 PB used
  - 250 MB/s aggregate writes
  - 150 MB/s on a per file basis

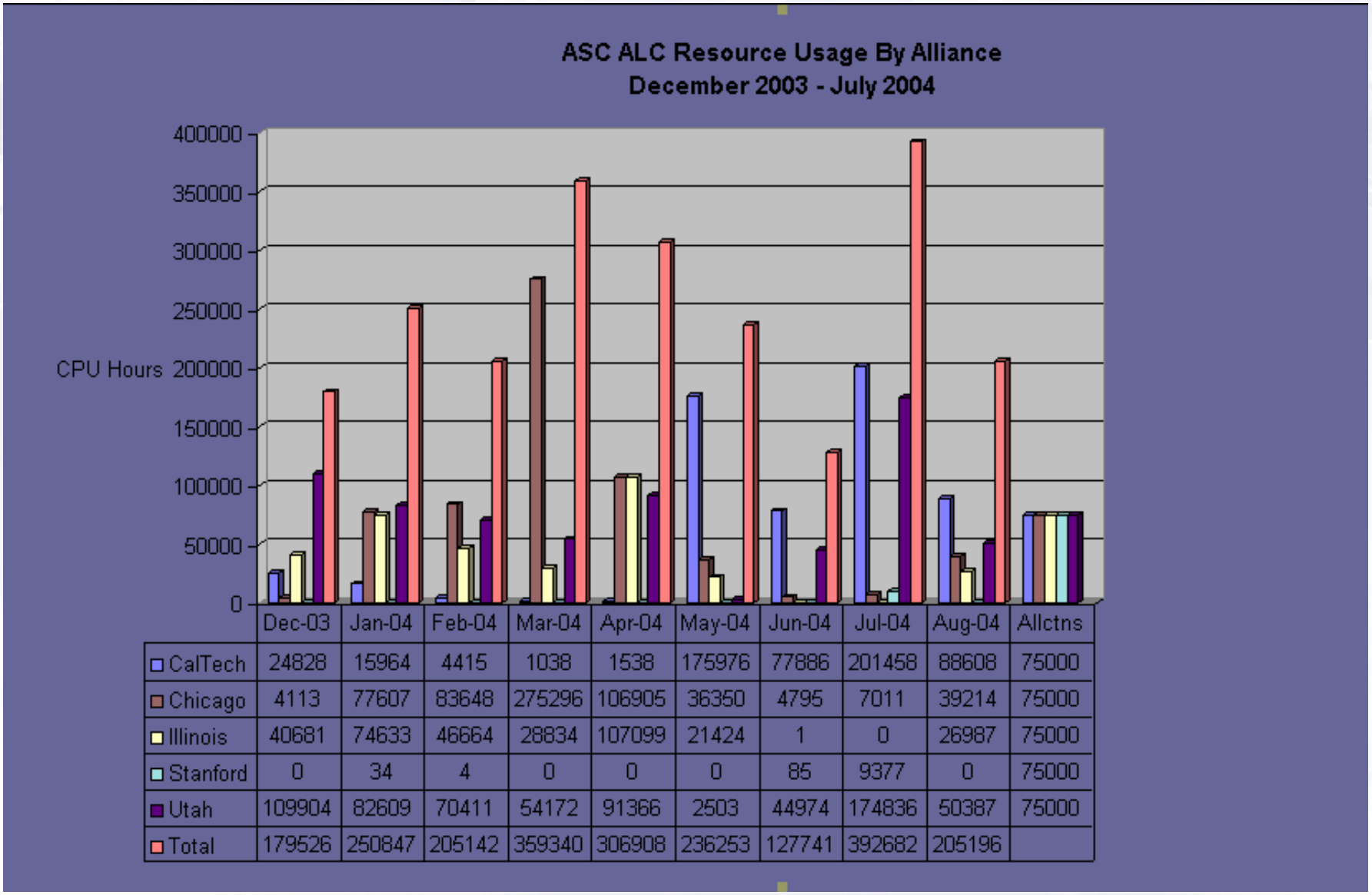




Frost & ALC Batch Queues

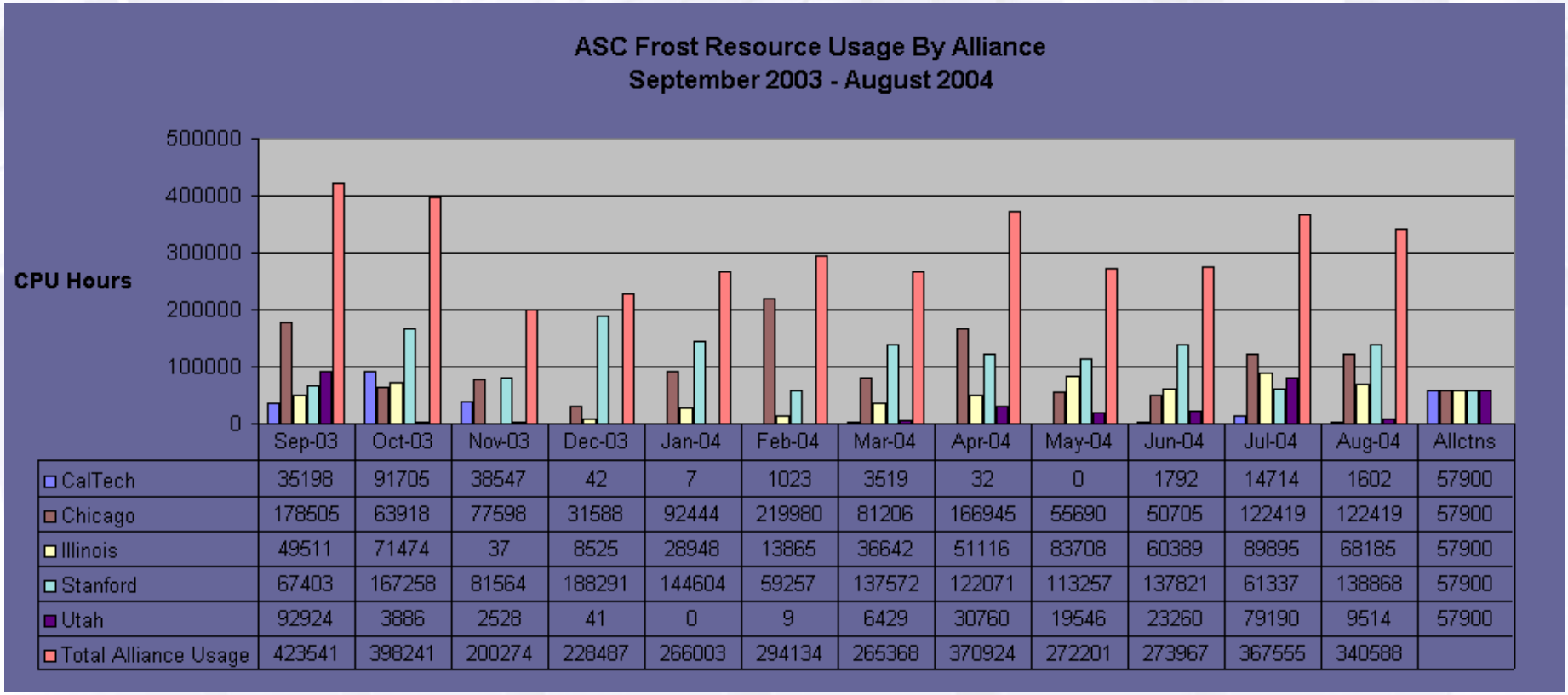
System	Batch Pool	Shift	Max Time	Max Nodes	Max Jobs
FROST	pbatch	Day (7am-7pm)	12 hr or 96 node-hr	24	4
		Night / Weekend (7pm-7am)	12 hr or 384 node-hr	32	4
	pdebug	All shifts	1 hr	1	1
ALC	pbatch	All shifts	24 hr	TBD	TBD
	pdebug	All shifts	TBD	TBD	TBD

ALC Alliance Usage



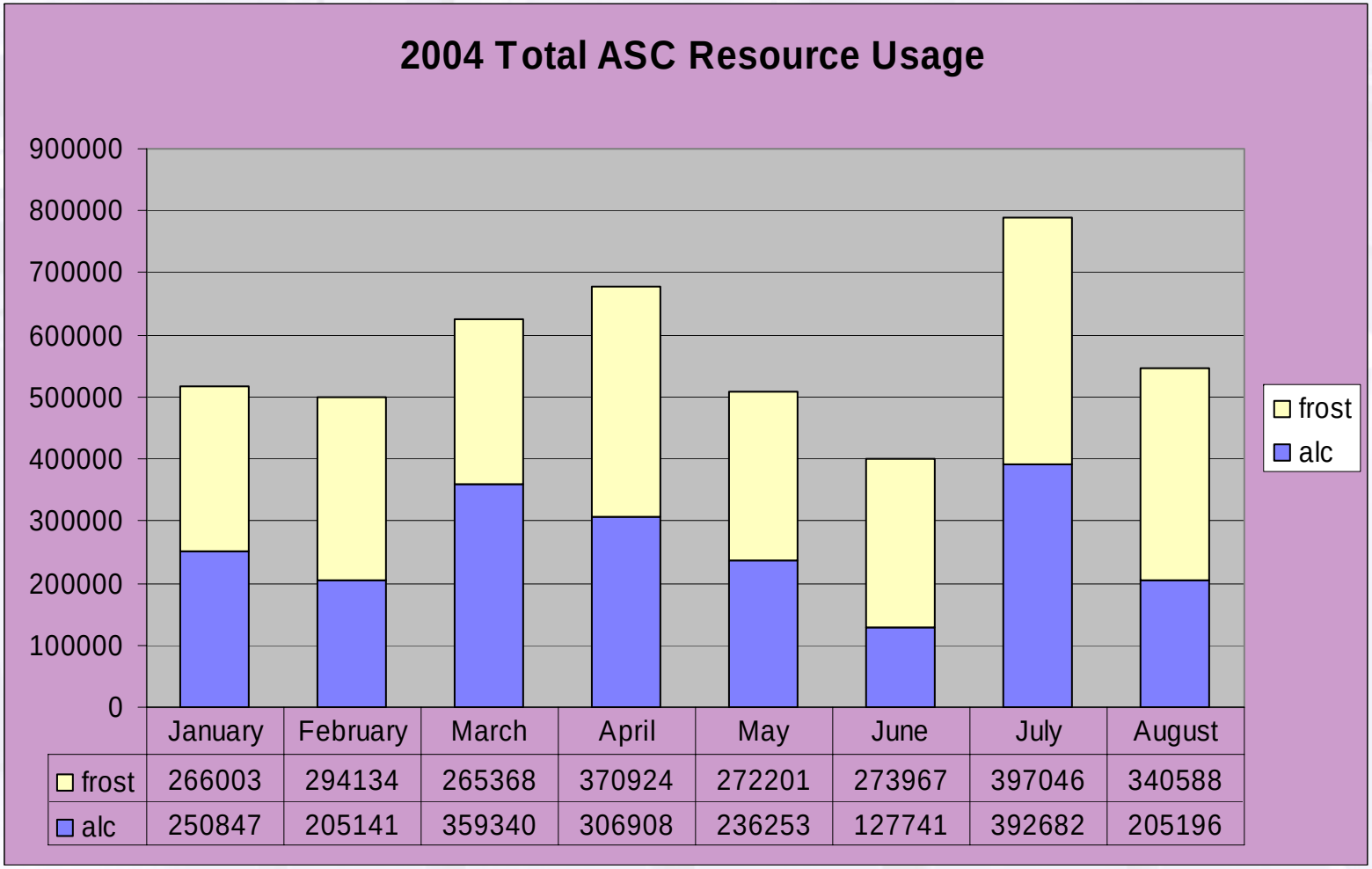
	Dec-03	Jan-04	Feb-04	Mar-04	Apr-04	May-04	Jun-04	Jul-04	Aug-04	Allctns
CalTech	24828	15964	4415	1038	1538	175976	77886	201458	88608	75000
Chicago	4113	77607	83648	275296	106905	36350	4795	7011	39214	75000
Illinois	40681	74633	46664	28834	107099	21424	1	0	26987	75000
Stanford	0	34	4	0	0	0	85	9377	0	75000
Utah	109904	82609	70411	54172	91366	2503	44974	174836	50387	75000
Total	179526	250847	205142	359340	306908	236253	127741	392682	205196	

Frost Alliance Usage



	Sep-03	Oct-03	Nov-03	Dec-03	Jan-04	Feb-04	Mar-04	Apr-04	May-04	Jun-04	Jul-04	Aug-04	Allctns
CalTech	35198	91705	38547	42	7	1023	3519	32	0	1792	14714	1602	57900
Chicago	178505	63918	77598	31588	92444	219980	81206	166945	55690	50705	122419	122419	57900
Illinois	49511	71474	37	8525	28948	13865	36642	51116	83708	60389	89895	68185	57900
Stanford	67403	167258	81564	188291	144604	59257	137572	122071	113257	137821	61337	138868	57900
Utah	92924	3886	2528	41	0	9	6429	30760	19546	23260	79190	9514	57900
Total Alliance Usage	423541	398241	200274	228487	266003	294134	265368	370924	272201	273967	367555	340588	

Total Alliance Usage



	January	February	March	April	May	June	July	August
frost	266003	294134	265368	370924	272201	273967	397046	340588
alc	250847	205141	359340	306908	236253	127741	392682	205196

Frost cpu-hr = 5.4 Tflop/hr
ALC cpu-hr = 17.3 Tflop/hr

Software Environment

- **Frost: AIX 5.1**
  PSSP 3.4
  GPFS 1.5

  - Planning to upgrade to AIX 5.2 and Parallel Environment 4.1 in the late Sep timeframe.

- **ALC: CHAOS 2.0**
  - Clustered High Availability Operating System
  - LC's developmental Linux cluster OS
  - Based upon RedHat Linux
  - Used on ALC and all other LC Linux clusters
  - How did the upgrade from 1.0 affect you?

- **SLURM**
  - Simple Linux Utility for Resource Management
  - Collaboration between Livermore and Linux NetworX
  - Under development
  - Used on ALC and all other LC Linux clusters

Software Environment

- **Compilers**
  - Issues? Intel version 8.0?
- **Debuggers, correctness tools**
  - Assure
  - décor
  - Great Circle
  - Insure++
  - TotalView
  - Umpire
  - pdbx
  - ZeroFault
- **Profiling**
  - gprof
  - HPM
  - MPX
  - mpiP
  - papi
  - prof
  - Xprofiler
- **Performance analysis**
  - Dimemas
  - Paradyn
  - Paraver
  - PE Benchmarker
  - Tau
  - Vampir/Guideview
  - Vtune (coming)
  - PeekPerf (coming)
- **What tools do you use?**
- **What are we missing?**

Training

- **Regular introductory workshops at LLNL**
  - Parallel programming
  - Linux & Compaq clusters
  - POE
  - Pthreads
  - TotalView
  - LC resources and environment
  - IBM hardware/software
  - MPI
  - OpenMP
  - …

- **Other workshops**
  - Performance analysis tools and topics for the IBM SP
  - MPI performance topics
  - Vampir/GuideView, Paraver, Dimemas
  - Advanced MPI
  - Advanced TotalView
  - Python, Linux topics

- **Tri-lab and Alliance workshops**
  - Combined training for multiple ASC platforms held at any Tri-lab or Alliance location
  - Customized workshops delivered at Alliance's location

Training

- **Collaborative HPC training via the Access Grid and and informal HPC Training Consortium**

Berkeley Nat'l Lab	NCSA
Sandia Nat'l Lab	SDSC
Los Alamos Nat'l Lab	U. Hawaii
Ohio Supercomputer Ctr.	U. Michigan
Texas Adv. Comp. Ctr.	Purdue U.
Argonne Nat'l Lab	General Atomics
Livermore Nat'l Lab	Maui HPC
Princeton Plasma Physics Lab	

HPC-training-consortium@purdue.edu

Future Plans

- **Frost**
  - Decommission in Jan 2005 time frame
- **Purple**
  - Not much to say here for the Alliances
- **BlueGene/L**
- **Move to the TSF**
- **SC2004**

Blue Gene/L

- **Blue Gene/L**
  - BlueGene/L is a computational sciences research and evaluation platform designed by IBM research for the DOE/NNSA ASC Program
  - New architecture optimized for cost, performance and scalability
  - 180-360 Tflops
  - 65,536 dual processor nodes with 512 MB memory/node; Torus network
  - IBM PowerPC ASIC processor @700MHz; dual FPU
  - More info: **www.llnl.gov/asci/platforms/bluegenel**

- **Proposals Update**
  - Alliances were asked early this year for proposals:
    - (1) identify the specific application(s) to run on BG/L
    - (2) identify personnel to work porting/scaling issues
    - (3) provide details on approach to using BG/L
  - Four responses were received and evaluated: Caltech, Chicago, Illinois, Utah
  - Only one site will be offered “early” (First Wave) time on BG/L

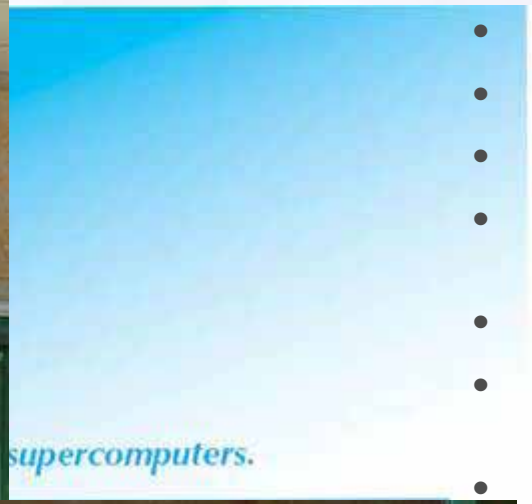
Blue Gene/L

- **Probable Timeline**
  - Oct ‘04 - Selected Alliance site determined/announced
  - Jan ‘05 - Early access to LLNL hardware for selected site
  - Jul ‘05 - Likely start date for early runs on full system
  - Jan ‘06 - General access to BG/L for other Alliance sites

- **Allocation strategies for LLNL, LANL, SNL, Alliances and other external users have not yet been determined…**

- **"Second Wave" apps will be targeted similarly to "First Wave" apps**

Terascale Simulation Facility



- Home for Purple
- Home for BG/L
- 253,000 ft^2 total
- 2 machine rooms totaling 48,000 ft^2
- 22 MW power total
- 9 MW for machines
- $92 million cost

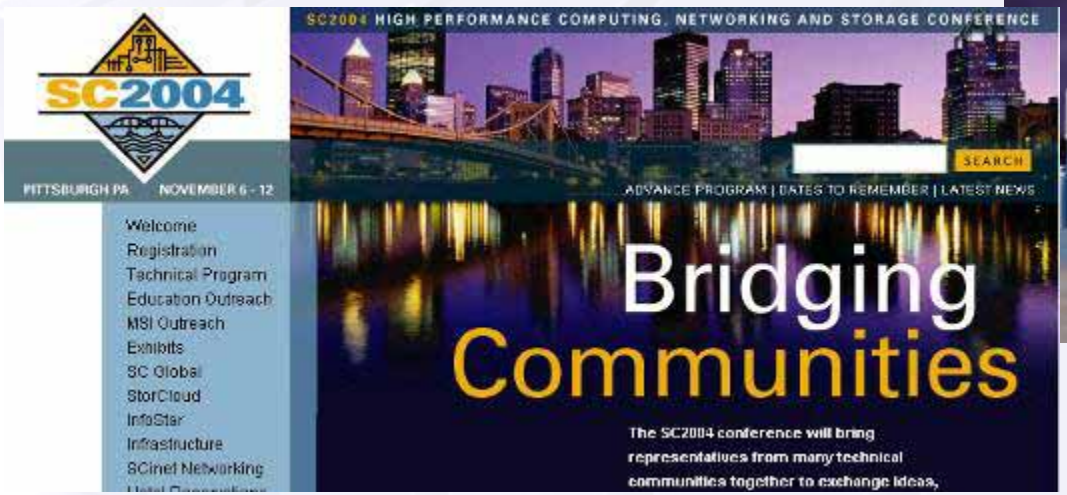
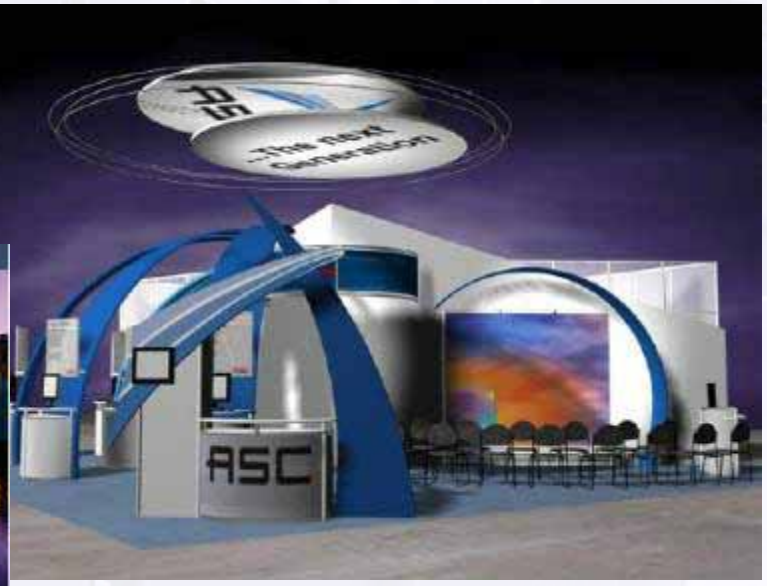
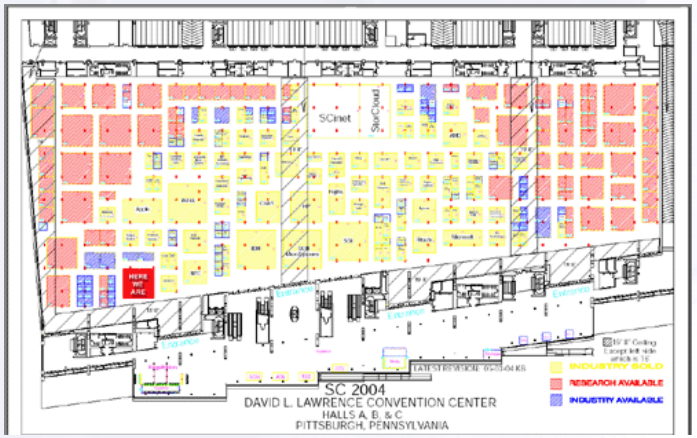
Office space for 288 scientists, engineers and support staff

Staff will begin moving in late this year

Ribbon cutting ceremony held July 8th

Ribbon cutting. From left to right are Camille Yuan-Soo Hoo (NNSA Livermore Site Manager), Robert Dynes (University of California President), Bruce Goodwin (Associate Director for Defense & Nuclear Technologies), Spencer Abraham (Secretary of Energy), Michael Anastasio (LLNL Director), Dona Crawford (Associate Director for Computation), and Robert Foley (UC Vice President for Laboratory Management).

SC2004 - ASC Exhibit Booth

- **Pittsburgh, Nov 6-12**
- **Call for participation!!**

  http://public.lanl.gov/sc04/