

# Update on LANL Data and Information Availability

**James Nunez**  
**Los Alamos National Lab**

**August 6, 2007**  
**LA-UR 07-5301**

# Publicly Available Data

---

- **Previously Announced Data**

- Machine Information
- Failure/Usage/Event/Disk Failure Data for LANL Supercomputers

- **Additional Data Available**

- Machine Layout Information for LANL Supercomputers

- **Soon to be Released Data**

- File Systems Statistics Survey Information (CMU/Panasas)
- Traces

## LANL: Failure/Usage/Event Data Sets Available (almost 9 years coverage in some cases)

description	size	records	name
all systems failure/interrupt data 1996-2005	2963538	23741	LA-UR-05-7318-failure-data-1996-2005.csv
system 20 usage with domain info	51675641	489376	LA-UR-06-0803-MX20_NODES_0_TO_255_DOM.TXT
system 20 usage with node info nodes number from zero	43926669	489376	LA-UR-06-0803-MX20_NODES_0_TO_255_NODE-Z.TXT
system 20 event info nodes number from zero	33120015	433490	LA-UR-06-0803-MX20_NODES_0_TO_255_EVENTS.csv
system 15 usage with node info nodes number from zero	2416139	17823	LA-UR-06-0999-MX15-NODE-Z.TXT
system 16 usage with node info nodes number from one	321293488	1630479	LA-UR-06-1446-MX16-NODE-NOZ.TXT
system 23 usage with node info nodes number from one	60674531	654927	LA-UR-06-1447-MX23-NODE-NOZ.TXT
system 8 usage with node info nodes number from one	67291020	763293	LA-UR-06-3194-MX8-NODE-NOZ.TXT

# LANL: Machine Information

system CMU paper number	system data machine number	system type	number nodes	number cpus	cpus/node	install date	production date	decommission date	fru	mem per node	cpu type	number of interconnects	use type	
1	7	smp	1	8	8	before tracking	before tracking	Dec-99	part	16	3	0	compute	
2	24	smp	1	32	32	before tracking	before tracking	Dec-03	part	8	7	1	compute	
3	22	smp	1	4	4	before tracking	before tracking	Apr-03	part	1	6	0	compute	
4	8	cluster	164	328	2	Mar-01	Apr-01	current	part	1	4	1	compute	<a href="#">GET Machine Layout</a>
5	20	cluster	512	2048	4	Oct-01	Dec-01	current	part	16	2	2	compute	<a href="#">GET Machine Layout</a>
6	21	cluster	128	512	4	Aug-01	Sep-01	Jan-02	part	16	2	2	compute	
7	18	cluster	1024	4096	4	Mar-02	May-02	current	part	16	2	2	compute	<a href="#">GET Machine Layout</a>
8	19	cluster	1024	4096	4	Aug-02	Oct-02	current	part	16	2	2	compute	<a href="#">GET Machine Layout</a>
9	3	cluster	128	512	4	Aug-03	Sep-03	current	part	4	2	1	compute	<a href="#">GET Machine Layout</a>

EST. 1943

# LANL: Machine Layout Information – Example

---

Machine 8,Bldg 1,Room 1,,  
,RackPosition,RackPosition,Position in Rack,  
NODE NUM,East/West,North/South,Vertical Position,  
N#,1 to 26,28 to 35,"1 to 37, top to bottom",  
1,23,28,1,  
2,23,28,2,  
3,23,28,3,  
4,23,28,4,  
5,23,28,5,  
6,23,28,6,  
7,23,28,7,  
8,23,28,8,  
...  
72,23,29,35  
frontend,23,29,36  
server,23,29,37  
73,23,32,1

# File Systems Statistics Survey (fsstats)

---

- **Purpose**
  - Develop better understanding of file system files
  - Gather and Build large db of static file tree statistics
  - (LANL) Test tool and validate correctness
- **Usage**
  - In parallel with LANL's MPI-File Tree Walk
- **LANL Data**
  - Production Supercomputers & Testbeds
  - From backups (database) of workstations Lab-wide
- **Output: Histogram Information on**
  - File Size
  - Capacity Used
  - Directory Size
  - File Name Size
  - Links Information

# File Systems Statistics Survey (fsstats) Data

---

file size:

count=2992940 avg=27961.25 KB

min=0.00 KB max=2147483648.00 KB

[ 0- 2 KB): 2107742 (70.42%) ( 70.42% cumulative)

[ 2- 4 KB): 323235 (10.80%) ( 81.22% cumulative)

[ 4- 8 KB): 231610 ( 7.74%) ( 88.96% cumulative)

[ 8- 16 KB): 117887 ( 3.94%) ( 92.90% cumulative)

[ 16- 32 KB): 84308 ( 2.82%) ( 95.72% cumulative)

[ 32- 64 KB): 63225 ( 2.11%) ( 97.83% cumulative)

...

filename length:

count=3236524 avg=13.42 chars

min=1.00 chars max=142.00 chars

[ 0- 7 chars]: 1119404 (34.59%) ( 34.59% cumulative)

[ 8- 15 chars]: 989571 (30.58%) ( 65.16% cumulative)

[ 16- 23 chars]: 699386 (21.61%) ( 86.77% cumulative)

[ 24- 31 chars]: 346496 (10.71%) ( 97.48% cumulative)

# Tracing Mechanisms and Additional Information

---

- **Desired Attributes in a Trace Mechanism**
  - Minimum overhead
  - Bandwidth preserving
- **Methods being Reviewed**
  - Linux ltrace/strace
  - Tracefs from SUNY Stony Brook
  - ||trace from CMU
- **Additional Information to Collect**
  - Time Stamps
  - Listing of directory
  - What else?
- **Traces to be Released**
  - Open Source Benchmark: MPI-IO based synthetic
  - Real Application Traces



# Tracing Mechanisms and Additional Information - Input

---

## Cluster Architecture

- Number compute nodes
- Number I/O nodes
- Interconnection network topology (flat, tree, etc.)

## Compute Node Architecture

- Host OS
- CPU
- RAM
- Network Types
- Simultaneous messaging capability

## I/O Node Architecture

- Host OS
- CPU
- RAM
- Network Types
- Simultaneous messaging capabilities
- Storage Subsystem
  - Performance Characteristics
    - Throughput
    - Latency
    - interface (scsi, fibre, sata, etc.)
    - disk arch (RPMs, Cache, Settle time, Acceleration in sectors, time to max rpm, cylinders, tracks, sectors)
  - Redundancy Capability (RAID?)

# LANL: Sample MPI-Based Benchmark Trace

---

## ■ Timing Information

# Barrier before benchmark\_call

7: cadillac113.ccstar.lanl.gov (10378) Entered barrier at 1159808385.170918

7: cadillac113.ccstar.lanl.gov (10378) Exited barrier at 1159808385.173167

3: cadillac117.ccstar.lanl.gov (11335) Entered barrier at 1159808385.166396

3: cadillac117.ccstar.lanl.gov (11335) Exited barrier at 1159808385.168893

5: cadillac115.ccstar.lanl.gov (10373) Entered barrier at 1159808385.168842

5: cadillac115.ccstar.lanl.gov (10373) Exited barrier at 1159808385.171370

## ■ Traced Application Data

10:59:47.092996 MPI\_File\_open(92, 0x80675c0, 37, 0x80675a8, 0xbfdfe5e4 <unfinished ...>

10:59:47.093718 SYS\_statfs64(0x80675c0, 84, 0xbfdfe410, 0xbfdfe410, 0xbd3ff4) = 0 <0.011131>

10:59:47.108352 SYS\_open("/panfs/REALM1/scratch/johnbent/O"..., 32832, 0600) = 3 <0.000745>

10:59:47.109189 SYS\_close(3) = 0 <0.000063>

10:59:47.109310 SYS\_open("/panfs/REALM1/scratch/johnbent/O"..., -2147450814, 0600) = 3 <0.000564>

10:59:47.110912 <... MPI\_File\_open resumed> ) = 0 <0.017855>

10:59:47.110955 MPI\_Wtime(0x8063830, 0xb7fee8dc, 0xbfdfe568, 0xb7f66ad7, 0x8059020 <unfinished ...>

# LANL: Sample MPI-Based Benchmark Trace

---

## ■ Summary Information

# SUMMARY COUNT OF TRACED CALL(S)

# Function Name	Number of Calls	Total time (s)
-----------------	-----------------	----------------

#=====

MPIO_Wait	2	0.000118
MPI_Barrier	29	2.156431
MPI_File_close	2	0.108482
MPI_File_delete	1	0.102532
...		

# SUMMARY COUNT OF CALLS WITHIN 1 MPI\_File\_delete CALL(S)

# Function Name	Number of Calls	Total time (s)
-----------------	-----------------	----------------

#=====

SYS_ipc	8	0.000140
SYS_statfs64	1	0.009227
SYS_unlink	1	0.092411

## Links to Data & Codes

---

- **Machine Failure/Usage/Event/Location Data Sets**
  - <http://institutes.lanl.gov/data/>
- **MPI-IO based synthetic & MPI-File Tree Walk**
  - <http://institutes.lanl.gov/data/software/>
- **File Systems Statistics Survey (fsstats) Code**
  - <http://www.pdsi-scidac.org/fsstats/>
- **Contact E-mail:**
  - [jnunez@lanl.gov](mailto:jnunez@lanl.gov)