Scheduling on the Top 50 machines

Carsten Ernemann, Martin Krogmann, Joachim Lepping, Ramin Yahyapour

Computer Engineering Institute, University Dortmund, 44221 Dortmund, Germany
(email: {carsten.ernemann,martin.krogmann,joachim.lepping,ramin.yahyapour}@udo.edu)

Abstract. The well-known TOP500 list ranks the 500 most powerful high-performance computers. However, the list lacks details about the job management and scheduling on these machines. As this statistic is interesting for researchers and system designers, this paper gives an overview and survey on scheduling relevant information for the first 50 entries in the TOP500 list.

1 Introduction

The task of scheduling computational jobs on parallel computers is subject to research for quite a long time. Despite many different approaches from theory, only a few scheduling strategies are practically in use. The actual statistics of the actual implementations are of interest to researchers, system administrators and manufacturers. The most known statistic about high-performance computers is the TOP500 list which is published every half year [2]. The list contains the 500 most powerful computers according to the LINPACK benchmark [5].

Unfortunately, the TOP500 list focuses on the benchmark result, peak performance, machine size, manufacturer and installation site. That is, there are no information about the scheduling systems that are deployed on these machines. To this end, this paper gives an survey about additional information of the top 50 machines on the TOP500 list from November 2003. The information has been collected from available web sites, publications and by querying the corresponding system administrators. The following section gives a description about the data in the list.

2 List Description

TOP500: Position in the TOP500 ranking for the November 2003 edition of the TOP500 list.
Name: Installation name from the TOP500 list.
Country and City: Location of the installation.
Year: Year of installation or last significant update.
Computer Family Model/Manufacturer: Information about the system model and the manufacturer.
Type: Type of the computer, e.g. parallel computer (MPP), vector computer, cluster.
Inst. Type: Classification of the application field of the installation (research, academic, industry).
Processors: Number of processors.
Op. System: Operating System of the machine.
Max. Mem./Total Mem.: Maximum available main memory on a single processing node/cummulative total memory.
R_{max}/R_{peak}**:** Maximal LINPACK performance achieved and the theoretical peak performance respectively (both in GFlops).
N_{max}/N_{half}**:** LINPACK problem size for achieving R_{max} and for achieving half of R_{max}.
Queues: Information about the existing queues in the job management system.
Scheduling: Information about the used job scheduling system and strategies.
Prioritization: shows whether priorities are assigned to users and/or jobs.
Backfilling: whether backfilling is used as a job scheduling strategy [4, 3]
Reservations: whether processor allocations are reservable in advance.

Checkpointing: The local management supports the checkpointing of a job. A file of a checkpointed job is generated that allows a later continuation from that point. The checkpoint file may also be migratable to other resources, but this feature is not required.

Preemption: A job is preempted on a given processor allocation and later continued [1]. In this case the corresponding application is stopped but remains resident on the allocated processors and can be resumed later. This preemption is not synonymous with the preemption in a multitasking system that typically happens in the time range of milliseconds.

Gang Scheduling: A parallel job can be preempted and continued on a given processor allocation. The scheduling system assures that all tasks of a parallel jobs are active at the same time, so that no process of a job has to wait for communication with another process of the job which is not currently active. That is preemption is synchronized for all processes of a job; within a "gang" all processes are active at the same time. This strategy can be used to allow time-shared execution of several parallel applications within different gangs.

Partitions: Many systems use partitioning to split the existing number of processors into groups for special applications. For instance, dedicated partitions for interactive jobs or data-intensive applications.

Average Utilization: Information about the average utilization of the complete machine.

3 List

TOP500[1]**:**	1	**Name:**	Earth Simulator Center	
Country:	Japan	**City:**	Yokohama	**Year:** 2002
Computer **Family Model:**	Earth-Simulator	**Manufacturer:**	NEC	
Type:	Parallel vector	**Inst. Type:**	Research	
Processors:	5120	**Op. System:**	ESOS (SUPER-UX)	
Max. Mem.:	16 GB	**Total Mem.:**	10 TB	
$\mathbf{R_{max}}$[2] :	35860	$\mathbf{R_{peak}}$[3] :	40960	
$\mathbf{N_{max}}$[4] :	$1{,}0752\times10^6$	$\mathbf{N_{half}}$[5] :	266240	
Queues: • S-queue : small scale batch requests (Max 8 AP and 16 GB within 1 node) • L-queue : large scale batch requests (Max 512 nodes)				
Scheduling: • NQS-II (ERS-II : S-queue, customized scheduler : L-queue), NEC				
Prioritization:	No	**Backfill:**	Yes	
Reservations:	No	**Checkpointing:**	Yes	
Preemption:	No	**Gang Scheduling:**	No	
Partitions: • 2048 Banks				
Average Utilization: not given				

TOP500:	2	**Name:**	Los Alamos National Lab		
Country:	USA	**City:**	Los Alamos, NM	**Year:**	2002
Computer Family Model:	ASCI Q-AlphaServer SC 45, 1.25 GHz	**Manufacturer:**	HP		
Type:	Cluster	**Inst. Type:**	Research		
Processors:	8192	**Op. System:**	Tru64 Unix		
Max. Mem.:	not given	**Total Mem.:**	22 TB		
$\mathbf{R_{max}}$:	13880	$\mathbf{R_{peak}}$:	20480		
$\mathbf{N_{max}}$:	633000	$\mathbf{N_{half}}$:	225000		
Queues: • 8-9 active queues per cluster • 4-5 queues per cluster that are activated for special purposes • Queue configuration is changed according to customer input on current needs averaging once per month. • Queues maybe set up for a project with a deadline to give it on-demand access (without preemption), special debugging queues, queues that allow very long running jobs, etc.					
Scheduling: • LSF (Fair Share Scheduling)					
Prioritization:	Yes	**Backfill:**	Yes		
Reservations:	Yes	**Checkpointing:**	Yes		
Preemption:	Yes	**Gang Scheduling:**	Yes		
Partitions: • No login nodes in the Unix/RMS sense. • All access is through LSF scheduled/controlled jobs. • 128 nodes on each cluster are file serving nodes and permit the interactive login to one or two whole nodes via a LSF interactive job. • This provides immediate access for ”login jobs” since there are adequate resources for our typical interactive development workload. These nodes are not normally used for large parallel jobs. • All queues support LSF interactive access up to the maximum size allowed by the queue. • User can schedule up to 384 whole nodes (1356 processors) interactively via an LSF job using the large queue.					
Average Utilization: For 2003 the utilization was approximately 55% on 8192 processors or 2048 nodes.					
Information from: Manuel Vigil, Los Alamos, NM email: mbv@lanl.gov					

TOP500:	3	**Name:**	Virginia Tech		
Country:	USA	**City:**	Falls Church, VA	**Year:**	2003
Computer Family Model:	1100 Dual 2.0 GHz Apple G5, Mellanox Infiniband 4X	**Manufacturer:**	Self-made		
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	2200	**Op. System:**	Mac OS X		
Max. Mem.:	4 GB	**Total Mem.:**	4,4 TB		
$\mathbf{R_{max}}$:	10280	$\mathbf{R_{peak}}$:	17600		
$\mathbf{N_{max}}$:	520000	$\mathbf{N_{half}}$:	152000		
Queues: not given					
Scheduling: • Deja vu					
Prioritization:	No	**Backfill:**	No		
Reservations:	No	**Checkpointing:**	Yes		
Preemption:	No	**Gang Scheduling:**	No		
Partitions: not given					
Average Utilization: not given					

TOP500:	4	**Name:**	NCSA		
Country:	USA	**City:**	Champaign, IL	**Year:**	2003
Computer Family Model:	PowerEdge 1750, P4 Xeon 3.06 GHz, Myrinet	**Manufacturer:**	Dell		
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	2500	**Op. System:**	Linux (Red Hat 9.0)		
Max. Mem.:	3 GB	**Total Mem.:**	3,75 TB		
$\mathbf{R_{max}}$:	9819	$\mathbf{R_{peak}}$:	15300		
$\mathbf{N_{max}}$:	630000	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: • Maui Scheduler					
Prioritization:	Yes	**Backfill:**	Yes		
Reservations:	No	**Checkpointing:**	No		
Preemption:	Yes	**Gang Scheduling:**	No		
Partitions: not given					
Average Utilization: not given					

TOP500:	5	**Name:**	Pacific Northwest National Lab		
Country:	USA	**City:**	Richland, WA	**Year:**	2003
Computer Family Model:	Integrity rx2600 Itanium2 1.5 GHz, Quadics	**Manufacturer:**	HP		
Type:	Cluster	**Inst. Type:**	Research		
Processors:	1956	**Op. System:**	Linux (Red Hat 7.2)		
Max. Mem.:	not given	**Total Mem.:**	6,8 TB		
$\mathbf{R_{max}}$:	8633	$\mathbf{R_{peak}}$:	11616		
$\mathbf{N_{max}}$:	835000	$\mathbf{N_{half}}$:	140000		

Queues:

- three main queues for normal user jobs
- A large job queue that has a slightly higher priority and only runs jobs requiring 256 CPU's.
- A short queue for jobs of 8 CPU's or less and less than 30 minutes of run time and a normal queue of other user jobs.
- All of these jobs will backfill if possible.
- In addition to these we have some other queues for testing system issues and for running special jobs that we need to tend.
- Also we have the SLURM queue for other extremely low priority jobs that we can kill when we need the node for a "real" job.

Scheduling:

- LSF as a scheduler on top of the Quadrics RMS resource management system.
- SLURM resource manager for some of the lowest priority, preemptable backfill, jobs.
- SLURM jobs to backfill also but preempt them when LSF jobs are scheduled to run.

Prioritization:	Yes	**Backfill:**	Yes
Reservations:	Yes	**Checkpointing:**	No
Preemption:	Yes	**Gang Scheduling:**	Yes

Partitions:

- Partition for the user login nodes and the management nodes (4 nodes).
- Partition for the Lustre filesystem nodes (34 nodes).
- The remaining nodes are in a single partition (940 nodes).
- These nodes consist of "Fat" nodes (8 GB memory and 400 GB local scratch disk at 200MB/s).
- "Thin" nodes (6 GB memory, 12 GB local scratch disk)

Average Utilization:

We average over 95% node utilization for the last 30 days.

Information from:

Gary B. Skouson
email: Gary.Skouson@pnl.gov

TOP500:	6	**Name:**	Los Alamos National Lab		
Country:	USA	**City:**	Los Alamos, NM	**Year:**	2003
Computer Family Model:	Opteron 2 GHz, Myrinet	**Manufacturer:**	Linux Networx		
Type:	Cluster	**Inst. Type:**	Research		
Processors:	2816	**Op. System:**	Linux (Red Hat)		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	8051	$\mathbf{R_{peak}}$:	11264		
$\mathbf{N_{max}}$:	761160	$\mathbf{N_{half}}$:	109208		
Queues: not given					
Scheduling: • LCRM • SLURM • Fair Share with Half-Life					
Prioritization:	Yes	**Backfill:**	No		
Reservations:	No	**Checkpointing:**	No		
Preemption:	Yes	**Gang Scheduling:**	Yes		
Partitions: not given					
Average Utilization: not given					

TOP500:	7	**Name:**	Lawrence Livermore National Lab		
Country:	USA	**City:**	Livermore, CA	**Year:**	2002
Computer Family Model:	MCR Linux Cluster Xeon 2.4 GHz, Quadrics	**Manufacturer:**	Linux Networx		
Type:	Cluster	**Inst. Type:**	Research		
Processors:	2304	**Op. System:**	Chaos 1.2 (modified Red Hat 7.3)		
Max. Mem.:	4 GB	**Total Mem.:**	4,5 TB		
$\mathbf{R_{max}}$:	7634	$\mathbf{R_{peak}}$:	11060		
$\mathbf{N_{max}}$:	350000	$\mathbf{N_{half}}$:	75000		
Queues: not given					
Scheduling: • LCRM • SLURM • Fair Share with Half-Life					
Prioritization:	Yes	**Backfill:**	No		
Reservations:	No	**Checkpointing:**	No		
Preemption:	Yes	**Gang Scheduling:**	Yes		
Partitions: not given					
Average Utilization: not given					

TOP500:	8	**Name:**	Lawrence Livermore National Lab		
Country:	USA	**City:**	Livermore, CA	**Year:**	2000
Computer Family Model:	ASCI White, SP Power3 375 Mhz	**Manufacturer:**	IBM		
Type:	Parallel	**Inst. Type:**	Research		
Processors:	8192	**Op. System:**	AIX		
Max. Mem.:	16 GB	**Total Mem.:**	8 TB		
$\mathbf{R_{max}}$:	7304	$\mathbf{R_{peak}}$:	12288		
$\mathbf{N_{max}}$:	640000	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: • DPCS • LoadLeveler • GangLL					
Prioritization:	Yes	**Backfill:**	No		
Reservations:	No	**Checkpointing:**	Yes		
Preemption:	Yes	**Gang Scheduling:**	Yes		
Partitions: • Debug Partition • Batch Partition					
Average Utilization: not given					

TOP500:	9	**Name:**	NERSC/LBNL		
Country:	USA	**City:**	Berkeley, CA	**Year:**	2002
Computer Family Model:	SP Power3 375 Mhz 16way	**Manufacturer:**	IBM		
Type:	Parallel	**Inst. Type:**	Research		
Processors:	6656	**Op. System:**	AIX		
Max. Mem.:	16 GB - 64 GB	**Total Mem.:**	7 TB		
$\mathbf{R_{max}}$:	7304	$\mathbf{R_{peak}}$:	9984		
$\mathbf{N_{max}}$:	640000	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: • LoadLeveler					
Prioritization:	No	**Backfill:**	Yes		
Reservations:	No	**Checkpointing:**	No		
Preemption:	No	**Gang Scheduling:**	Yes		
Partitions: not given					
Average Utilization: not given					

TOP500:	10	**Name:**	Lawrence Livermore National Lab		
Country:	USA	**City:**	Livermore, CA	**Year:**	2003
Computer Family Model:	xSeries Cluster Xeon 2.4 GHz, Quadrics	**Manufacturer:**	IBM/ Quadrics		
Type:	Cluster	**Inst. Type:**	Research		
Processors:	1920	**Op. System:**	not given		
Max. Mem.:	4 GB	**Total Mem.:**	3,75 TB		
$\mathbf{R}_{\mathbf{max}}$:	6586	$\mathbf{R}_{\mathbf{peak}}$:	9216		
$\mathbf{N}_{\mathbf{max}}$:	425000	$\mathbf{N}_{\mathbf{half}}$:	90000		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	11	**Name:**	National Aerospace Lab of Japan		
Country:	Japan	**City:**	Tokyo	**Year:**	2002
Computer Family Model:	PRIMEPOWER HPC2500 1.3 GHz	**Manufacturer:**	Fujitsu		
Type:	Parallel	**Inst. Type:**	Research		
Processors:	2304	**Op. System:**	not given		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R}_{\mathbf{max}}$:	5406	$\mathbf{R}_{\mathbf{peak}}$:	11980		
$\mathbf{N}_{\mathbf{max}}$:	658800	$\mathbf{N}_{\mathbf{half}}$:	100080		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	12	**Name:**	Pittsburgh Supercomputing Center	
Country:	USA	**City:**	Pittsburgh, PA	**Year:** 2001
Computer Family Model:	AlphaServer SC45, 1GHz	**Manufacturer:**	HP	
Type:	Cluster	**Inst. Type:**	Academic	
Processors:	3016	**Op. System:**	Tru64 UNIX	
Max. Mem.:	32 GB	**Total Mem.:**	3 TB	
$\mathbf{R_{max}}$:	4463	$\mathbf{R_{peak}}$:	6032	
$\mathbf{N_{max}}$:	280000	$\mathbf{N_{half}}$:	85000	

Queues:

- one large job queue (>= 256 nodes (>= 1024 cpus))
- one smaller job queue (< 256 nodes (< 1024 cpus))

Scheduling:

- OpenPBS with the custom scheduler Simon (written in TCL).
- Simon features advance reservations, backfilling, and co-scheduling special purpose visualization nodes.
- Supports various job prioritizations based on job size and queue priority to accommodate the user base and desired workload mix.

Prioritization:	Yes	**Backfill:**	Yes
Reservations:	Yes	**Checkpointing:**	Yes
Preemption:	No	**Gang Scheduling:**	No

Partitions:

- One partition to which jobs are scheduled.
- 1 node (an SMP) is comprised of 4 cpus and 4 GB of memory.
- Scheduling at the node level so that no nodes are shared.

Average Utilization:

- Typical utilization runs about 90%.
- Allocating nodes is done by using a reserved resource model. That is, once a node has been allocated to a job, it's up to the user to decide how to use the resources of the node or nodes assigned as they are assigned exclusively to the user.
- Billing and measuring utilization is based on the number of nodes allocated to jobs.

Information from:

Chad Vizino
email: vizino@psc.edu

TOP500:	13	**Name:**	NCAR		
Country:	USA	**City:**	Boulder, CO	**Year:**	2003
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM		
Type:	Cluster	**Inst. Type:**	Research		
Processors:	1600	**Op. System:**	AIX		
Max. Mem.:	2 GB	**Total Mem.:**	3 TB		
$\mathbf{R_{max}}$:	4184	$\mathbf{R_{peak}}$:	8320		
$\mathbf{N_{max}}$:	550000	$\mathbf{N_{half}}$:	93000		
Queues: 27					
Scheduling: • LoadLeveler					
Prioritization:	No	**Backfill:**	Yes		
Reservations:	No	**Checkpointing:**	No		
Preemption:	No	**Gang Scheduling:**	No		
Partitions: not given					
Average Utilization: not given					

TOP500:	14	**Name:**	Cinese Academy of Science		
Country:	China	**City:**	Beijing	**Year:**	2003
Computer Family Model:	DeepComp 6800, Itanium2 1.3 GHz, QsNet	**Manufacturer:**	Legend		
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	1024	**Op. System:**	not given		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	4183	$\mathbf{R_{peak}}$:	5324,8		
$\mathbf{N_{max}}$:	491488	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: 2 • Climate Simulation Laboratory jobs • Community Computing Jobs					
Average Utilization: not given					

TOP500:	15	**Name:**	Comm. a l'Energie Atomique		
Country:	France	**City:**	St.-Paul-lez-Durance	**Year:**	2001
Computer Family Model:	AlphaServer SC45, 1GHz	**Manufacturer:**	HP		
Type:	Cluster	**Inst. Type:**	Research		
Processors:	2560	**Op. System:**	Tru64 UNIX 5.1a		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	3980	$\mathbf{R_{peak}}$:	5120		
$\mathbf{N_{max}}$:	360000	$\mathbf{N_{half}}$:	85000		
Queues: • LSF batch management system					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	16	**Name:**	HPCx	
Country:	UK	**City:**	Edinburgh	**Year:** 2002
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Academic	
Processors:	1280	**Op. System:**	AIX	
Max. Mem.:	1 GB	**Total Mem.:**	1,2 TB	
$\mathbf{R_{max}}$:	3406	$\mathbf{R_{peak}}$:	6656	
$\mathbf{N_{max}}$:	317000	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: • LoadLeveler				
Prioritization:	no	**Backfill:**	yes	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	no	
Partitions: not given				
Average Utilization: not given				

TOP500:	17	**Name:**	Forecast Systems Laboratory	
Country:	USA	**City:**	Washington, DC	**Year:** 2002
Computer Family Model:	Aspen Systems, Dual Xeon 2.2 GHz,Myrinet2000	**Manufacturer:**	HPTi	
Type:	Cluster	**Inst. Type:**	Research	
Processors:	1536	**Op. System:**	Linux (Red Hat 6)	
Max. Mem.:	1 GB	**Total Mem.:**	0,75 TB	
$\mathbf{R_{max}}$:	3337	$\mathbf{R_{peak}}$:	6758	
$\mathbf{N_{max}}$:	285000	$\mathbf{N_{half}}$:	75000	
Queues: not given				
Scheduling: • PBS Pro				
Prioritization:	no	**Backfill:**	yes	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	no	
Partitions: not given				
Average Utilization: not given				

TOP500:	18	**Name:**	Naval Oceanographic Office	
Country:	USA	**City:**	Stennis SC, MS	**Year:** 2002
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Research	
Processors:	1184	**Op. System:**	AIX 5.1	
Max. Mem.:	8 GB-64 GB	**Total Mem.:**	1,4 TB	
$\mathbf{R_{max}}$:	3160	$\mathbf{R_{peak}}$:	6156,8	
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given	
Queues: 7 • batch • priority • bigmem • share • transfer • debug • background				
Scheduling: • LoadLeveler				
Prioritization:	no	**Backfill:**	yes	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	no	
Partitions: not given				
Average Utilization: not given				

TOP500:	19	**Name:**	Government	
Country:	USA	**City:**	not given	**Year:** 2003
Computer Family Model:	Cray X1	**Manufacturer:**	Cray Inc.	
Type:	Parallel vector	**Inst. Type:**	not given	
Processors:	252	**Op. System:**	UNICOS/mp	
Max. Mem.:	not given	**Total Mem.:**	5 TB	
$\mathbf{R_{max}}$:	2932,9	$\mathbf{R_{peak}}$:	3225,6	
$\mathbf{N_{max}}$:	338688	$\mathbf{N_{half}}$:	44288	
Queues: not given				
Scheduling: • PBS Pro • Load Balancer • Gang Scheduler				
Prioritization:	no	**Backfill:**	no	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	yes	
Partitions: not given				
Average Utilization: not given				

TOP500:	20	**Name:**	Oak Ridge National Laboratory	
Country:	USA	**City:**	Oak Ridge, TN	**Year:** 2003
Computer Family Model:	Cray X1	**Manufacturer:**	Cray Inc.	
Type:	Parallel vector	**Inst. Type:**	Research	
Processors:	252	**Op. System:**	UNICOS/mp	
Max. Mem.:	not given	**Total Mem.:**	5 TB	
$\mathbf{R_{max}}$:	2932,9	$\mathbf{R_{peak}}$:	3225,6	
$\mathbf{N_{max}}$:	338688	$\mathbf{N_{half}}$:	44288	
Queues: not given				
Scheduling: • PBS Pro • Load Balancer • Gang Scheduler				
Prioritization:	no	**Backfill:**	no	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	yes	
Partitions: not given				
Average Utilization: not given				

TOP500:	21	**Name:**	Cray Inc.	
Country:	USA	**City:**	Seattle, WA	**Year:** 2003
Computer Family Model:	Cray X1	**Manufacturer:**	Cray Inc.	
Type:	Parallel vector	**Inst. Type:**	Vendor	
Processors:	252	**Op. System:**	UNICOS/mp	
Max. Mem.:	not given	**Total Mem.:**	5 TB	
$\mathbf{R_{max}}$:	2932,9	$\mathbf{R_{peak}}$:	3225,6	
$\mathbf{N_{max}}$:	338688	$\mathbf{N_{half}}$:	44288	
Queues: not given				
Scheduling: • PBS Pro • Load Balancer • Gang Scheduler				
Prioritization:	no	**Backfill:**	no	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	yes	
Partitions: not given				
Average Utilization: not given				

TOP500:	22	**Name:**	Korea Institute of Science	
Country:	Korea	**City:**	Seoul	**Year:** 2003
Computer **Family Model:**	eServer Cluster 1350 xSeries Xeon 2.4 GHz, Myrinet	**Manufacturer:**	IBM	
Type:	Cluster	**Inst. Type:**	Research	
Processors:	1024	**Op. System:**	Linux (Red Hat 7.3)	
Max. Mem.:	not given	**Total Mem.:**	1024 GB	
$\mathbf{R_{max}}$:	3067	$\mathbf{R_{peak}}$:	4915,2	
$\mathbf{N_{max}}$:	300000	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: • PBS Pro • Maui Scheduler				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: no partitions				
Average Utilization: not given				

TOP500:	23	**Name:**	ECMWF		
Country:	UK	**City:**	Reading	**Year:**	2002
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM		
Type:	Parallel	**Inst. Type:**	Research		
Processors:	960	**Op. System:**	AIX		
Max. Mem.:	8 GB	**Total Mem.:**	2,7 TB		
$\mathbf{R_{max}}$:	2560	$\mathbf{R_{peak}}$:	4992		
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given		

Queues: 5 classes

- classes os and ns in the 3 LPAR for serial jobs
- classes op, debug and np in the 116 LPAR for parallel jobs.

Scheduling:

- The standard IBM LL backfill scheduling scheme aided by own combined job-filter
- runtime history files that ensures most job are given an accurate wall_clock_limit plus a base-time of 24 hours.

Prioritization:	yes	**Backfill:**	yes
Reservations:	yes	**Checkpointing:**	no
Preemption:	no	**Gang Scheduling:**	no

Partitions:

- Each system has 30 × p690 compute frames and 2 × Nighthawk I/O frames.
- The 30 × p690 frames are subdivided.
- 4 LPAR/frame, so 120 compute LPAR in total, each with 8 CPU so in total 960 CPUs.
- 2 memory types in the 30 × p690 frames.
- 27 frames have 32 GB memory and 3 frames 128 GB memory.

Average Utilization: between 94% and 97.5%

Information from:

Graham Holt
Technical Group Leader
HPCF Scheduling Specialist
ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK
email: graham.holt@ecmwf.int

TOP500:	26	**Name:**	Texas Advanced Computing Center		
Country:	USA	**City:**	Austin, Texas	**Year:**	2003
Computer Family Model:	PowerEdge 1750, Pentium4 Xeon 3.06 GHz, Myrinet	**Manufacturer:**	Dell-Cray		
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	600	**Op. System:**	Linux		
Max. Mem.:	not given	**Total Mem.:**	0,6 TB		
$\mathbf{R_{max}}$:	2455	$\mathbf{R_{peak}}$:	3672		
$\mathbf{N_{max}}$:	252000	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: • Job Mix Scheduler					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	27	**Name:**	Sandia National Laboratory		
Country:	USA	**City:**	Livermore, CA	**Year:**	1999
Computer Family Model:	ASCI Red, Pentium II Xeon	**Manufacturer:**	Intel		
Type:	Parallel	**Inst. Type:**	Research		
Processors:	9632	**Op. System:**	Paragon OS		
Max. Mem.:	256 MB/ 512 MB	**Total Mem.:**	1,2 TB		
$\mathbf{R_{max}}$:	2379	$\mathbf{R_{peak}}$:	3207		
$\mathbf{N_{max}}$:	362880	$\mathbf{N_{half}}$:	75400		
Queues: not given					
Scheduling: • Gang Scheduler					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	yes		
Partitions: not given					
Average Utilization: not given					

TOP500:	28	**Name:**	Oak Ridge National Laboratory		
Country:	USA	**City:**	Oak Ridge, TN	**Year:**	2002
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM		
Type:	Parallel	**Inst. Type:**	Research		
Processors:	864	**Op. System:**	AIX		
Max. Mem.:	8 GB	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	2310	$\mathbf{R_{peak}}$:	4492,8		
$\mathbf{N_{max}}$:	275000	$\mathbf{N_{half}}$:	62000		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	29	**Name:**	IBM		
Country:	Canada	**City:**	Markham, Ontario	**Year:**	2003
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM		
Type:	Parallel	**Inst. Type:**	Vendor		
Processors:	864	**Op. System:**	AIX		
Max. Mem.:	8 GB	**Total Mem.:**	not given		
$\mathbf{R}_{\mathbf{max}}$:	2310	$\mathbf{R}_{\mathbf{peak}}$:	4492,8		
$\mathbf{N}_{\mathbf{max}}$:	275000	$\mathbf{N}_{\mathbf{half}}$:	62000		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	30	**Name:**	Louisiana State University		
Country:	USA	**City:**	Baton Rouge, LA	**Year:**	2002
Computer Family Model:	P4 Xeon 1.8 GHz Myrinet	**Manufacturer:**	Atipa		
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	1024	**Op. System:**	Linux (Red Hat 7.2)		
Max. Mem.:	2 GB	**Total Mem.:**	1 TB		
$\mathbf{R}_{\mathbf{max}}$:	2207	$\mathbf{R}_{\mathbf{peak}}$:	3686,4		
$\mathbf{N}_{\mathbf{max}}$:	280000	$\mathbf{N}_{\mathbf{half}}$:	56000		
Queues: not given					
Scheduling: • PBS Pro					
Prioritization:	no	**Backfill:**	yes		
Reservations:	no	**Checkpointing:**	no		
Preemption:	no	**Gang Scheduling:**	no		
Partitions: not given					
Average Utilization: not given					

TOP500:	31	**Name:**	Max-Planck-Gesellschaft MPI/IPP	
Country:	Germany	**City:**	Garching	**Year:** 2003
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Research	
Processors:	832	**Op. System:**	AIX	
Max. Mem.:	21 × 64 GB + 2 × 96 GB + 2 × 256 GB	**Total Mem.:**	2 TB	
$\mathbf{R}_{\mathbf{max}}$:	2198,4	$\mathbf{R}_{\mathbf{peak}}$:	4326,4	
$\mathbf{N}_{\mathbf{max}}$:	not given	$\mathbf{N}_{\mathbf{half}}$:	not given	
Queues: • 12 queues with different number of nodes (processors) and different runtimes. • One special queue for the two ”fat” nodes with 256 GB main memory each.				
Scheduling: • IBM Loadleveler				
Prioritization:	yes	**Backfill:**	yes	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	not in use	
Partitions: • 25 compute (batch) nodes and 2 I/O nodes				
Average Utilization: 93% on 25 compute nodes				
Information from: Dr. Ingeborg Weidl, Max-Planck-Gesellschaft, D-85748 Garching email: weidl@rzg.mpg.de				

TOP500:	32	**Name:**	NASA	
Country:	USA	**City:**	Greenbelt, MD	**Year:** 2002
Computer Family Model:	AlphaServer SC45, 1GHz	**Manufacturer:**	HP	
Type:	Cluster	**Inst. Type:**	Research	
Processors:	1392	**Op. System:**	Tru64 UNIX 5.1a	
Max. Mem.:	not given	**Total Mem.:**	0,6 TB	
$\mathbf{R}_{\mathbf{max}}$:	2164	$\mathbf{R}_{\mathbf{peak}}$:	2784	
$\mathbf{N}_{\mathbf{max}}$:	320000	$\mathbf{N}_{\mathbf{half}}$:	40000	
Queues: • LSF batch management system				
Scheduling: not given				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: not given				
Average Utilization: not given				

TOP500:	33	**Name:**	Lawrence Livermore National Lab	
Country:	USA	**City:**	Livermore, CA	**Year:** 1999
Computer Family Model:	ASCI Blue-Pacific SST, IBM SP 604e	**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Research	
Processors:	5808	**Op. System:**	AIX 5	
Max. Mem.:	1,5-2,5 GB (432 nodes with 2,5 GB)	**Total Mem.:**	1,9 TB	
$\mathbf{R_{max}}$:	2144	$\mathbf{R_{peak}}$:	3856,5	
$\mathbf{N_{max}}$:	431344	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: • Parallel Op. System (POE)				
Prioritization:	no	**Backfill:**	no	
Reservations:	no	**Checkpointing:**	no	
Preemption:	no	**Gang Scheduling:**	yes	
Partitions: • 976 4-CPU SMP nodes consisting of 2 × 488-node sectors, S and K • 4 Login Nodes				
Average Utilization: not given				

TOP500:	34	**Name:**	US Army Research Laboratory	
Country:	USA	**City:**	Adelphi, MD	**Year:** 2002
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Research	
Processors:	800	**Op. System:**	AIX 5	
Max. Mem.:	8 GB	**Total Mem.:**	not given	
$\mathbf{R_{max}}$:	2140	$\mathbf{R_{peak}}$:	4160	
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: not given				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: not given				
Average Utilization: not given				

TOP500:	35	**Name:**	NCSA		
Country:	USA	**City:**	Champaign, IL	**Year:**	2003
Computer Family Model:	TeraGrid, Itanium2 1.3 GHz, Myrinet		**Manufacturer:**	IBM	
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	512	**Op. System:**	Suse SLES 8		
Max. Mem.:	4 GB/ 12 GB		**Total Mem.:**	2 TB	
$\mathbf{R_{max}}$:	2110		$\mathbf{R_{peak}}$:	2662,4	
$\mathbf{N_{max}}$:	308350		$\mathbf{N_{half}}$:	not given	
Queues: not given					
Scheduling: • PBS Pro • Maui Scheduler					
Prioritization:	yes		**Backfill:**	yes	
Reservations:	no		**Checkpointing:**	no	
Preemption:	yes		**Gang Scheduling:**	no	
Partitions: not given					
Average Utilization: not given					

TOP500:	36	**Name:**	Atomic Weapons Establishment		
Country:	UK	**City:**	Reading	**Year:**	2002
Computer Family Model:	SP Power3 375 Mhz 16way		**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Research		
Processors:	1920	**Op. System:**	AIX		
Max. Mem.:	16 GB (2 Nodes of 64 GB)		**Total Mem.:**	not given	
$\mathbf{R_{max}}$:	2106		$\mathbf{R_{peak}}$:	2880	
$\mathbf{N_{max}}$:	not given		$\mathbf{N_{half}}$:	not given	
Queues: not given					
Scheduling: not given					
Prioritization:	not given		**Backfill:**	not given	
Reservations:	not given		**Checkpointing:**	not given	
Preemption:	not given		**Gang Scheduling:**	not given	
Partitions: 120 nodes with 16 proccessors					
Average Utilization: not given					

TOP500:	37	**Name:**	Deutscher Wetterdienst		
Country:	Germany	**City:**	Offenbach	**Year:**	2003
Computer Family Model:	SP Power3 375 Mhz 16way		**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Research		
Processors:	1920	**Op. System:**	AIX 5.1		
Max. Mem.:	not given		**Total Mem.:**	1,24 TB	
$\mathbf{R_{max}}$:	2106		$\mathbf{R_{peak}}$:	2880	
$\mathbf{N_{max}}$:	not given		$\mathbf{N_{half}}$:	not given	
Queues: not given					
Scheduling: not given					
Prioritization:	not given		**Backfill:**	not given	
Reservations:	not given		**Checkpointing:**	not given	
Preemption:	not given		**Gang Scheduling:**	not given	
Partitions: not given					
Average Utilization: not given					

TOP500:	38	**Name:**	University at Buffalo		
Country:	USA	**City:**	Buffalo, NY	**Year:**	2002
Computer Family Model:	PowerEdge 2650 Cluster	**Manufacturer:**	Dell		
	P4 Xeon 2.4 GHz - Myrinet				
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	600	**Op. System:**	Linux (RedHat 7.3, 2.4 Kernel)		
Max. Mem.:	2 GB	**Total Mem.:**	not given		
$\mathbf{R}_{\mathbf{max}}$:	2004	$\mathbf{R}_{\mathbf{peak}}$:	2880		
$\mathbf{N}_{\mathbf{max}}$:	253400	$\mathbf{N}_{\mathbf{half}}$:	42200		
Queues: not given					
Scheduling: • PBS Pro • Maui Scheduler					
Prioritization:	yes	**Backfill:**	yes		
Reservations:	no	**Checkpointing:**	no		
Preemption:	yes	**Gang Scheduling:**	no		
Partitions: 258 Nodes					
Average Utilization: not given					

TOP500:	39	**Name:**	NC for Environmental Prediction		
Country:	USA	**City:**	Camp Springs, MD	**Year:**	2002
Computer Family Model:	pSeries 690 Turbo 1.3 GHz	**Manufacturer:**	IBM		
Type:	Parallel	**Inst. Type:**	Research		
Processors:	704	**Op. System:**	AIX		
Max. Mem.:	8 GB	**Total Mem.:**	not given		
$\mathbf{R}_{\mathbf{max}}$:	1849	$\mathbf{R}_{\mathbf{peak}}$:	3660,8		
$\mathbf{N}_{\mathbf{max}}$:	240000	$\mathbf{N}_{\mathbf{half}}$:	32500		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	40	**Name:**	SARA	
Country:	Netherlands	**City:**	Almere	**Year:** 2003
Computer Family Model:	SGI Altix 1.3 GHz	**Manufacturer:**	SGI	
Type:	Parallel	**Inst. Type:**	Academic	
Processors:	416	**Op. System:**	Linux (Red Hat)	
Max. Mem.:	not given	**Total Mem.:**	0,83 TB	
$\mathbf{R_{max}}$:	1793	$\mathbf{R_{peak}}$:	2163	
$\mathbf{N_{max}}$:	298799	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: not given				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: 6 batch nodes / 1 interactive node				
Average Utilization: not given				

TOP500:	41	**Name:**	KISTI Supercomputing Center	
Country:	South Korea	**City:**	Daejeon City	**Year:** 2003
Computer Family Model:	pSeries 690 Turbo 1.7 GHz	**Manufacturer:**	IBM	
Type:	Parallel	**Inst. Type:**	Research	
Processors:	544	**Op. System:**	AIX	
Max. Mem.:	8 GB	**Total Mem.:**	not given	
$\mathbf{R_{max}}$:	1760	$\mathbf{R_{peak}}$:	3699,2	
$\mathbf{N_{max}}$:	400000	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: not given				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: not given				
Average Utilization: not given				

TOP500:	42	**Name:**	Semiconductor Company	
Country:	USA	**City:**	not given	**Year:** 2003
Computer Family Model:	xSeries Cluster Xeon 2.4 GHz, Gig-E	**Manufacturer:**	IBM	
Type:	Cluster	**Inst. Type:**	Industry	
Processors:	1834	**Op. System:**	Linux	
Max. Mem.:	not given	**Total Mem.:**	not given	
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	8803,2	
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: not given				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: not given				
Average Utilization: not given				

TOP500:	43	**Name:**	WETA Digital		
Country:	New Zealand	**City:**	Wellington	**Year:**	2003
Computer Family Model:	BladeCenter Cluster Xeon 2.8 GHz, Gig-E	**Manufacturer:**	IBM		
Type:	Cluster	**Inst. Type:**	Industry		
Processors:	1176	**Op. System:**	Linux (Red Hat)		
Max. Mem.:	6 GB	**Total Mem.:**	3,4 TB		
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	6585,6		
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	44	**Name:**	Semiconductor Company		
Country:	USA	**City:**	not given	**Year:**	2003
Computer Family Model:	xSeries Cluster Xeon 2.8 GHz, Gig-E	**Manufacturer:**	IBM		
Type:	Cluster	**Inst. Type:**	Industry		
Processors:	1140	**Op. System:**	Linux		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	6384		
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	47	**Name:**	PGS		
Country:	USA	**City:**	Houston, TX	**Year:**	2003
Computer Family Model:	xSeries Cluster Xeon 3.06 GHz, Gig-E	**Manufacturer:**	IBM		
Type:	Cluster	**Inst. Type:**	Industry		
Processors:	1024	**Op. System:**	Linux		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	6266,88		
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	48	**Name:**	WETA Digital		
Country:	New Zealand	**City:**	Wellington	**Year:**	2003
Computer Family Model:	BladeCenter Cluster Xeon 2.8 GHz, Gig-E	**Manufacturer:**	IBM		
Type:	Cluster	**Inst. Type:**	Industry		
Processors:	1080	**Op. System:**	not given		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	6048		
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	52	**Name:**	CGG		
Country:	USA	**City:**	Houston, TX	**Year:**	2003
Computer Family Model:	xSeries Cluster Xeon 2.4 GHz, Gig-E	**Manufacturer:**	IBM		
Type:	Cluster	**Inst. Type:**	Industry		
Processors:	1100	**Op. System:**	Linux		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	5280		
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	53	**Name:**	Arizona State University/TGEN		
Country:	USA	**City:**	Phoenix, AZ	**Year:**	2003
Computer Family Model:	xSeries Cluster Xeon 2.4 GHz, Gig-E	**Manufacturer:**	IBM		
Type:	Cluster	**Inst. Type:**	Academic		
Processors:	1100	**Op. System:**	Linux		
Max. Mem.:	not given	**Total Mem.:**	not given		
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	5030,4		
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given		
Queues: not given					
Scheduling: not given					
Prioritization:	not given	**Backfill:**	not given		
Reservations:	not given	**Checkpointing:**	not given		
Preemption:	not given	**Gang Scheduling:**	not given		
Partitions: not given					
Average Utilization: not given					

TOP500:	54	**Name:**	Paradigm Geophysical	
Country:	USA	**City:**	Houston, TX	**Year:** 2003
Computer Family Model:	BladeCenter Cluster Xeon 2.4 GHz, Gig-E	**Manufacturer:**	IBM	
Type:	Cluster	**Inst. Type:**	Research	
Processors:	1024	**Op. System:**	not given	
Max. Mem.:	not given	**Total Mem.:**	not given	
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	4915,2	
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: not given				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: not given				
Average Utilization: not given				

TOP500:	55	**Name:**	TotalFinaElf	
Country:	France	**City:**	not given	**Year:** 2003
Computer Family Model:	xSeries Cluster Xeon 2.4 GHz, Gig-E	**Manufacturer:**	IBM	
Type:	Cluster	**Inst. Type:**	Industry	
Processors:	1024	**Op. System:**	not given	
Max. Mem.:	not given	**Total Mem.:**	not given	
$\mathbf{R_{max}}$:	1755	$\mathbf{R_{peak}}$:	4915,2	
$\mathbf{N_{max}}$:	not given	$\mathbf{N_{half}}$:	not given	
Queues: not given				
Scheduling: not given				
Prioritization:	not given	**Backfill:**	not given	
Reservations:	not given	**Checkpointing:**	not given	
Preemption:	not given	**Gang Scheduling:**	not given	
Partitions: not given				
Average Utilization: not given				

4 Acknowledgements

While some of the data have been gathered from the available web pages, the authors are grateful to the different contributions from system administrations. The names are given in the tables for the corresponding entries.

References

1. Dror G. Feitelson, Larry Rudolph, Uwe Schwiegelshohn, Kenneth C. Sevcik, and Parkson Wong. Theory and practice in parallel job scheduling. In *IPPS'97 Workshop: Job Scheduling Strategies for Parallel Processing*, volume 1291 of *Lecture Notes in Computer Science (LNCS)*, pages 1–34. Springer, Berlin, April 1997.
2. 22nd TOP500 List introduced during the Supercomputer Conference (SC2003) in Phoenix, AZ. http://www.top500.org November 2003.
3. D.G. Feitelson and A.M. Weil. Utilization and Predictabillity in Scheduling the IBM SP2 with Backfilling. In *Proceedings of IPPS/SPDP 1998*, IEEE Computer Society, pages 542–546, 1998.
4. D. A. Lifka. The ANL/IBM SP Scheduling System. In D. G. Feitelson and L. Rudolph, editors, *IPPS'95 Workshop: Job Scheduling Strategies for Parallel Processing*, pages 295–303. Springer, Berlin, Lecture Notes in Computer Science LNCS 949, 1995.
5. A. Petitet and R.C. Whaley and J. Dongarra and A. Cleary HPL - A Portable Implementation of the High-Performance Linpack Benchmark for Distributed-Memory Computers http://www.netlib.org/benchmark