The LLNL Atlas log

System: Linux Cluster (Atlas) at LLNL
Duration: Nov 2006 to Jun 2007
Jobs: 60332

This log contains several months worth of accounting records from a large Linux cluster called Atlas installed at Lawrence Livermore National Lab. For more information about Linux clusters at LLNL, see URL https://computing.llnl.gov/tutorials/linux_clusters/. This specific cluster has 1152 nodes, each with 8 processors, for a total of 9216 processors.

Atlas is considered a "capability" computing resource, meaning that it is intended for running large parallel jobs that cannot execute on lesser machines. This is in contrast with Thunder, which is a "capacity" machine, used for running large numbers of smaller jobs.

Note that the log does not include arrival information, only start times.

The LLNL Atlas workload log was graciously provided by Moe Jette, who also helped with background information and interpretation. If you use this log in your work, please use a similar acknowledgment.

Downloads:

LLNL-Atlas-2006-0	1.2 MB gz	original log
LLNL-Atlas-2006-2.swf	640 KB gz	converted log
LLNL-Atlas-2006-2.1-cln.swf	490 KB gz	cleaned log -- RECOMMENDED, see usage notes
LLNL-Atlas-2006-1.swf	640 KB gz	OLD VERSION of converted log (replaced 11 Dec 2011)
LLNL-Atlas-2006-1.1-cln.swf	490 KB gz	OLD VERSION of cleaned log (replaced 11 Dec 2011)

(May need to click with right mouse button to save to disk)

System Environment

Atlas is an 1152 node Linux cluster. Each node boasts 8 AMD Opteron processors clocked at 2.4 GHz and 16 GB of memory. The nodes are connected by an Infiniband switch.

The nodes are divided into three partitions:

login	8 nodes
debug	32 nodes
batch	1072 nodes

Scheduling is performed with the Slurm resource management system. For more information about Slurm, see URL https://computing.llnl.gov/LCdocs/slurm/.

Log Format

The original log is available as LLNL-Atlas-2006-0.

This file contains one line per completed job in the Slurm format. The fields are

1. JobId=<number>
2. UserId=xxxxx(<number>) - the x's hide the actual user name to conserve privacy
3. Name=<string> - name of executable (script), could be empty
4. JobState=<status>
5. Partition=<string>
6. TimeLimit=<number> - in minutes
7. StartTime=<date and time>
8. EndTime=<date and time>
9. NodeList=<string> - comma separated list of single nodes and ranges
10. NodeCnt=<number>

Conversion Notes

The converted log is available as LLNL-Atlas-2006-2.swf. The conversion from the original format to SWF was done subject to the following.

- The log does not indicate each job's submittal time. Therefore the jobs' submit times were set to their start times, and the wait times were set to -1.
- The number of processors used is taken from the NodeList field, by parsing the list. The number requested is taken from the NodeCnt field.
- 42 entries were truncated, typically in the middle of the NodeList field. In these cases we do not know the number of requested or allocated nodes, and these were therefore set to -1.
- The log should reflect jobs that ran on the debug and batch partitions, which total 1104 nodes. However, 9 jobs were recorded as using more than that. The conversion therefore assumes that all 1152 nodes are available.
- Requested time is the wallclock limit, not a precise estimate.
- The status field mapping used was

COMPLETED	1
FAILED	0
TIMEOUT	0
NODE_FAIL	0
CANCELLED	5

- The conversion loses the following data, that cannot be represented in the SWF:
    - Distinction between failure modes (timeout and node failure).
    - The precise list of nodes used by each job, as given in the NodeList field.
    - The actual command that was executed, as given in the Name field.
    - When the requested runtime is explicitly specified as UNLIMITED
- The following anomalies were identified in the conversion:
    - 19 jobs got more processors than they requested
    - 3771 jobs got more runtime than they requested, but in all cases the difference was less than a minute.

The difference between conversion 1 and conversion 2 is only in the wait time field, which was set to 0 in conversion 1 and to -1 in conversion 2.

The conversion was done by a log-specific parser in conjunction with a more general converter module.

Usage Notes

The original log contains several flurries of very high activity by individual users, which may not be representative of normal usage. In addition, the initial part of the log exhibits an uncharacteristically low level of activity, as the system was still quite new at the time (according to the utilization graph, it might have been operating at half the final capacity). These were removed in the cleaned version. It is recommended that the clean version be used.
The cleaned log is available as LLNL-Atlas-2006-2.1-cln.swf.

A flurry is a burst of very high activity by a single user. The filters used to remove the initial section and the five flurries that were identified are

submitted before 18 Dec 2006 (1434 jobs)
user=4 and job>3873 and job<5926 (2038 jobs)
user=28 and job>20616 and job<21532 (887 jobs)
user=7 and job>21547 and job<22295 (709 jobs)
user=19 and job>40338 and job<51898 (6783 jobs)
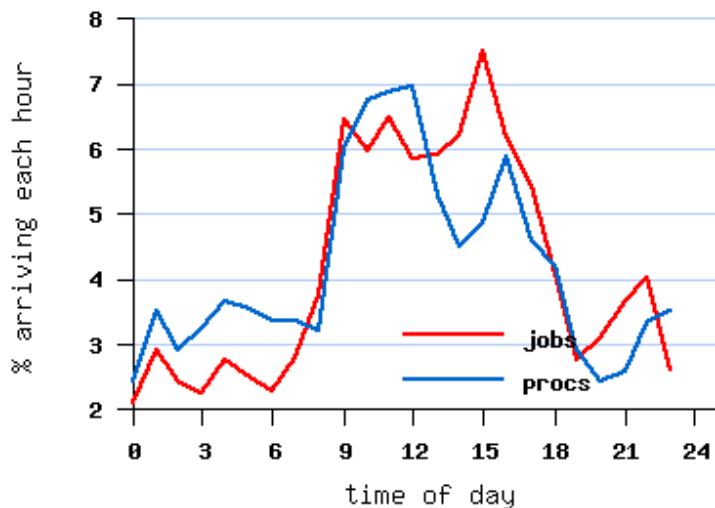user=66 and job>22438 and job<56102 (4703 jobs)

Note that the filters were applied to the original log, and unfiltered jobs remain untouched. As a result, in the filtered logs job numbering is not consecutive. Moreover, due to the fact that the whole initial part of the log is discarded, the start time indication in the header comments is also wrong.
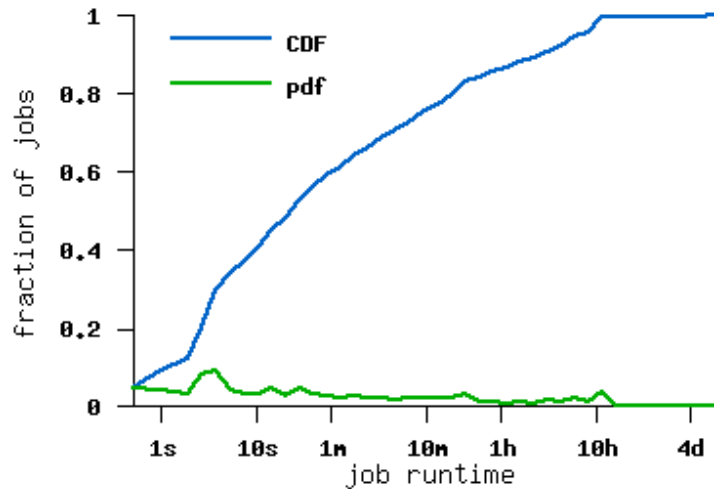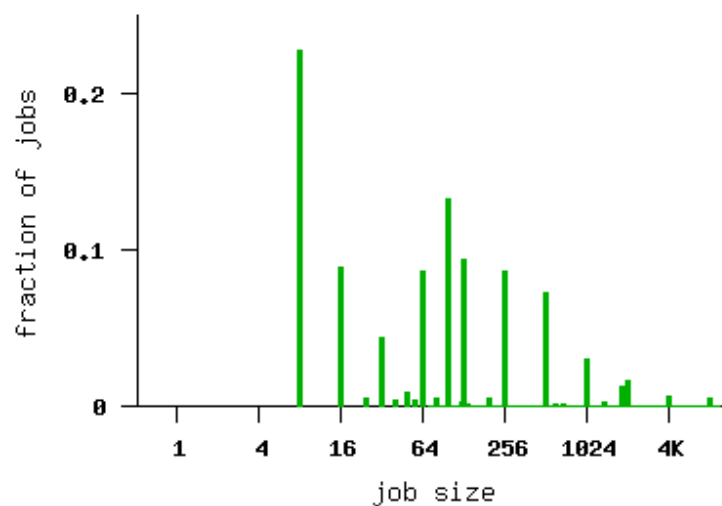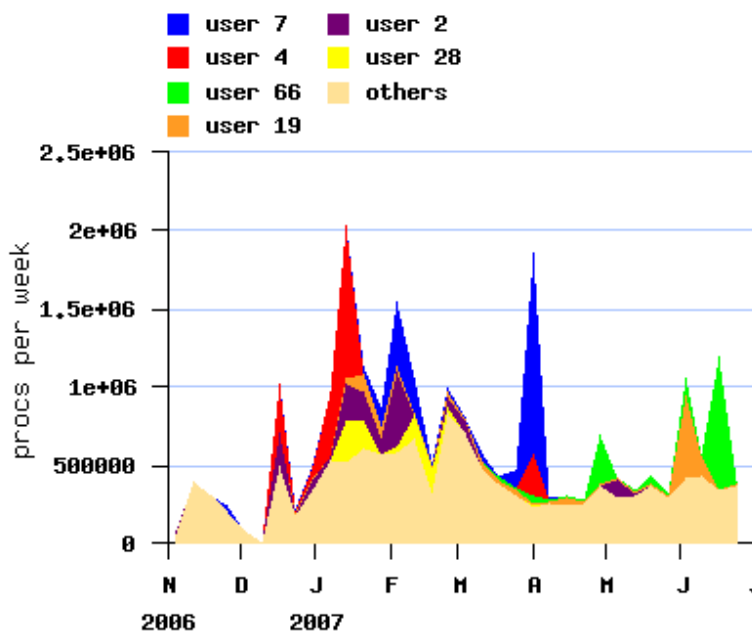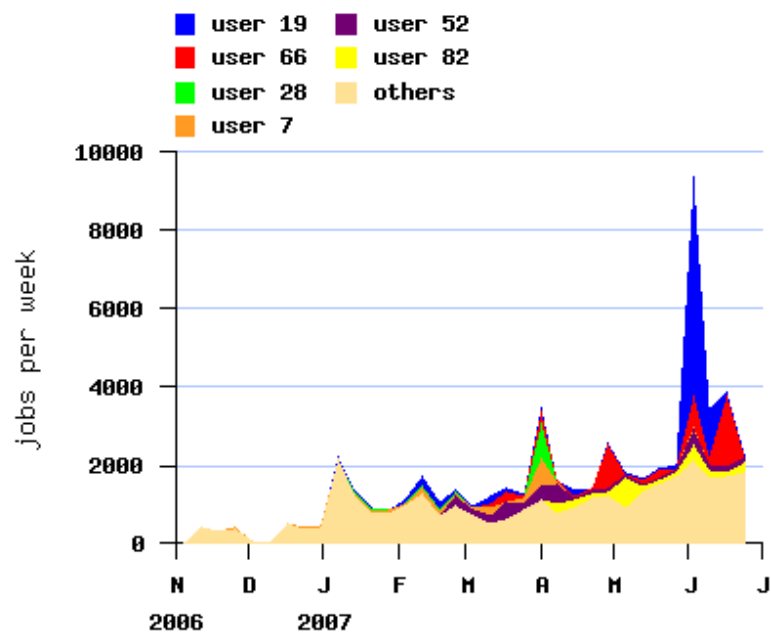
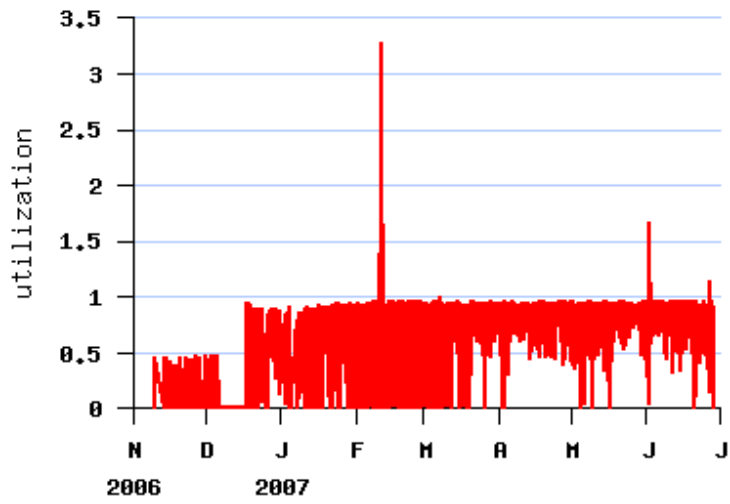Further information on flurries and the justification for removing them can be found in:

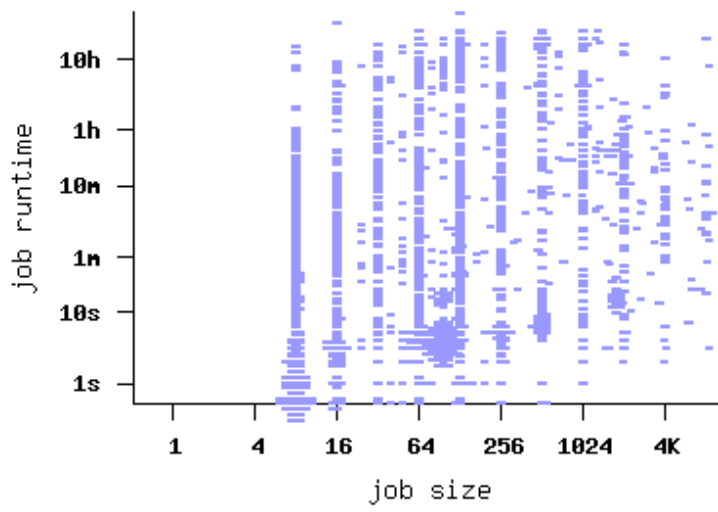- D. G. Feitelson and D. Tsafrir, ``Workload sanitation for performance evaluation''. In *IEEE Intl. Symp. Performance Analysis of Systems and Software*, pp. 221-230, Mar 2006.
- D. Tsafrir and D. G. Feitelson, ``Instability in parallel job scheduling simulation: the role of workload flurries''. In 20th *Intl. Parallel and Distributed Processing Symp.*, Apr 2006.

The Log in Graphics

File LLNL-Atlas-2006-2.swf

Top left chart:

Legend: user 19, user 52, user 66, user 82, user 28, others, user 7

y-axis: jobs per week (0, 2000, 4000, 6000, 8000, 10000)

x-axis: N D J F M A M J J, 2006, 2007

Top right chart:

Legend: user 7, user 2, user 4, user 28, user 66, others, user 19

y-axis: procs per week (0, 500000, 1e+06, 1.5e+06, 2e+06, 2.5e+06)

x-axis: N D J F M A M J J, 2006, 2007

Bottom left chart:

y-axis: fraction of jobs (0, 0.1, 0.2)

x-axis: job size (1, 4, 16, 64, 256, 1024, 4K)

Bottom right chart:

Legend: CDF, pdf

y-axis: fraction of jobs (0, 0.2, 0.4, 0.6, 0.8, 1)

x-axis: job runtime (1s, 10s, 1m, 10m, 1h, 10h, 4d)

File LLNL-Atlas-2006-2.1-cln.swf