Trace analysis report GWA-DAS2

General information

This is the trace analysis report (generated by reportgen.py) for the DAS-2 system. The trace data was taken from the filename anon_jobs.gwf, which contains job data obtained from Sun Grid Engine logs (local resource manager). Below is a summary of the contents of the trace data:

- Date first entry: Tue Feb 22 16:52:25 2005
- CPU time consumed by jobs: 68y 230d 21h 30m 0s
- Number of sites in the system: 5
- Number of CPUs in the trace: 400
- Number of jobs in the trace: 1124772
- Number of users in the trace: 333
- Number of groups in the trace: 12

System-wide characteristics

System utilization

We define the overall system utilization as the ratio between the total CPU time consumed by users, and the total CPU time available to the users. We compute the total CPU time consumed by users as the sum of CPU time consumed by each job in the system; for failed jobs, only those that have effectively spent resource time are considered. We compute the total CPU time available as the number of CPUs multiplied by the duration of a fixed time interval, c.q. 10 minutes.

Below we show the statistical properties of both the overall system utilization and the overall system for non-zero values, that is, excluding all intervals that have system utilization equal to zero. This excludes values that may account for downtime of the system.

Figure 1 shows System utilization over time.

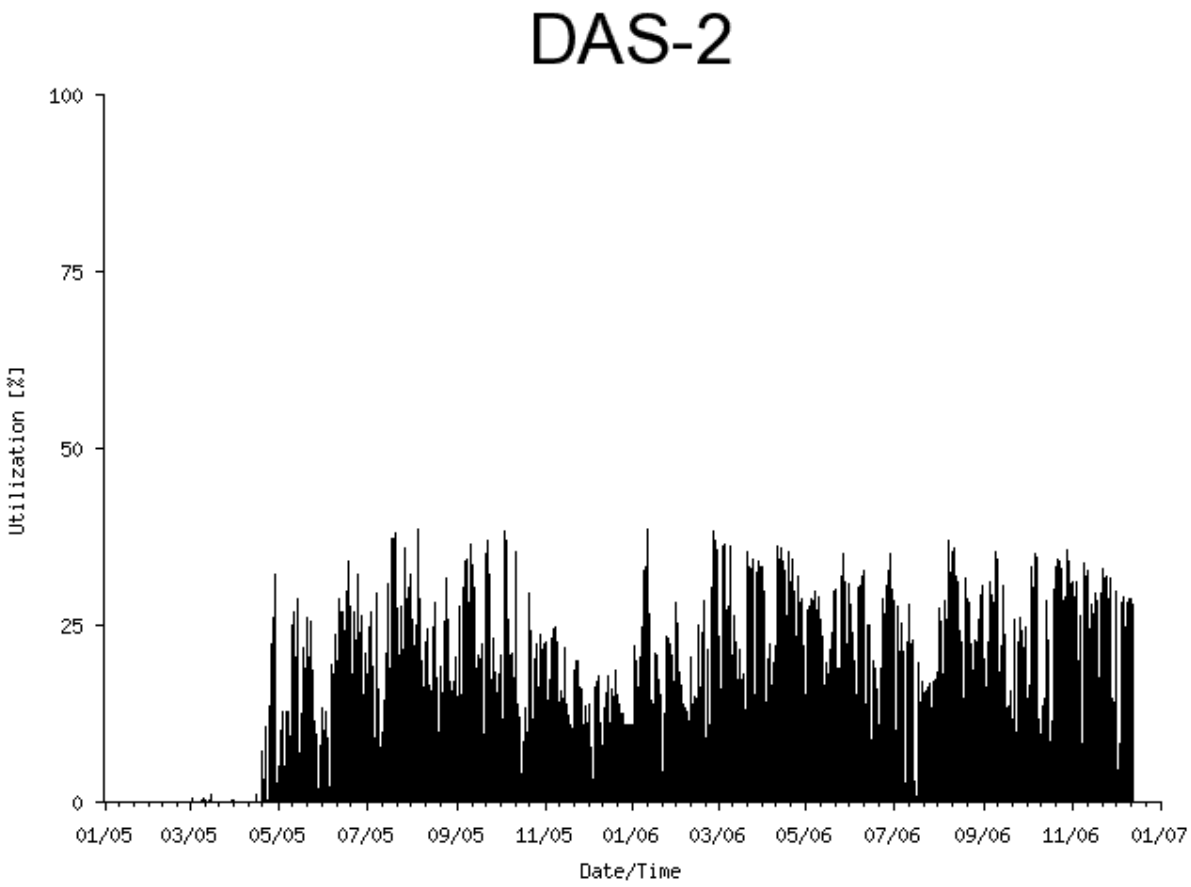

Figure 1: System utilization over time

Overall system utilization

- Minimum: 0.0 percent
- Maximum: 38.616 percent
- Average: 9.932 percent

Overall system utilization for non-zero values

- Minimum: 0.101 percent
- Maximum: 38.616 percent
- Average: 11.697 percent

Job arrival rate

We define the job arrival rate as the number of jobs that are submitted to the system in a fixed time interval. We compute the arrival rate for every hour by counting the all jobs that are recorded in the trace during that hour. This includes failed jobs and jobs that are cancelled before execution. Below we list the time periods in which the highest number of jobs were submitted to the system. We also summarize statistical properties for all job arrival rate values, and the statistical properties for arrival rate higher than zero. This excludes time periods that may account to downtime of the system.

Figure 2 shows Overall job arrival rate during hourly intervals.

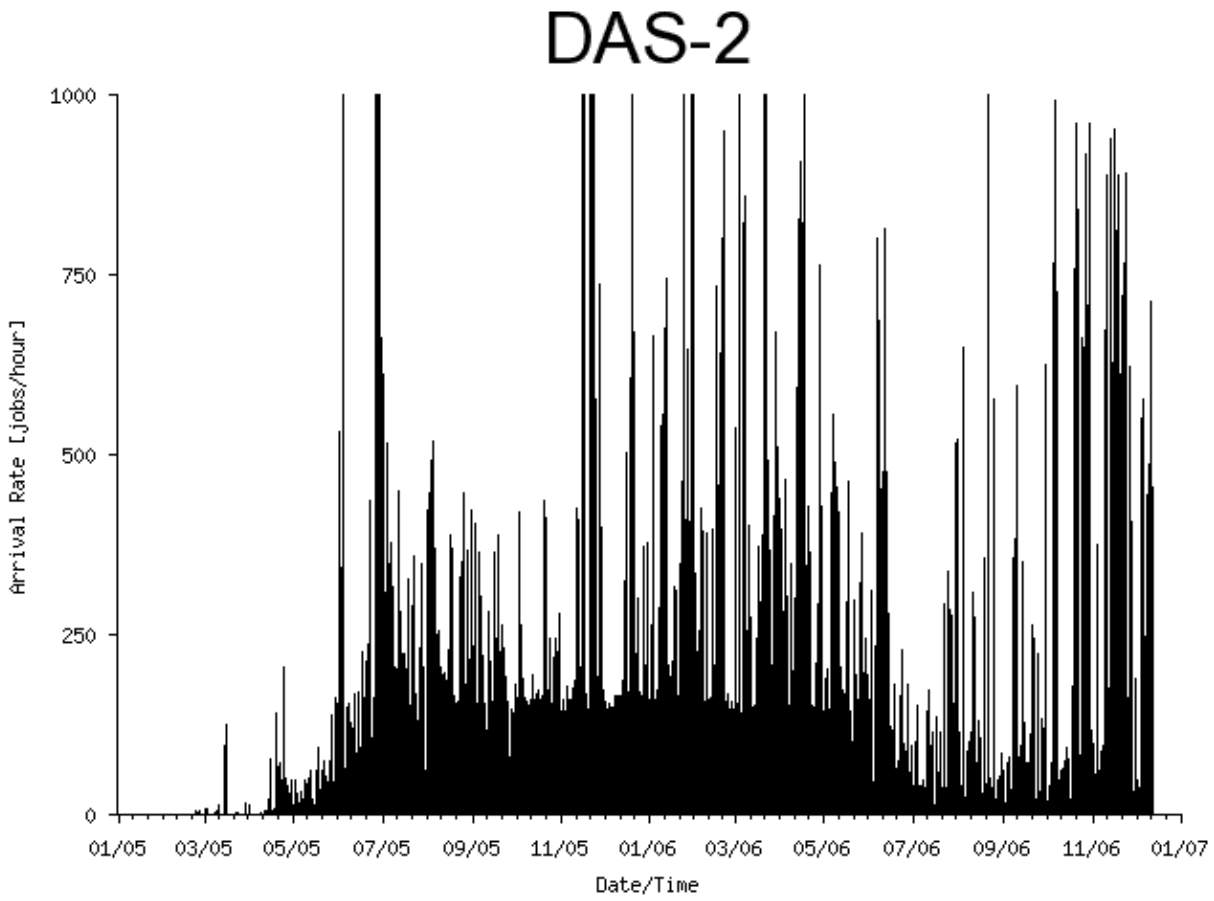

Figure 2: Overall job arrival rate during hourly intervals

Busiest time periods in terms of number of job submissions

- Busiest day: 2005-11-21
- Busiest week: 2006-15
- Busiest month: 2006-01

Overall job arrival metrics

- Minimum: 0.00 jobs/hour
- Maximum: 7009.00 jobs/hour
- Average: 71.13 jobs/hour

Overall job arrival metrics for non-zero values

- Minimum: 2.00 jobs/hour
- Maximum: 7009.00 jobs/hour
- Average: 85.56 jobs/hour

Job characteristics

We compute three important characteristics of jobs in the trace: number of CPUs used, the runtime of the job and the amount of memory used. Below we summarize the statistical properties for single jobs in the trace. We do not include jobs that were cancelled before execution, because those jobs did not consume resources from the system.

Figure 3 shows CDFs of the most important job characteristics.

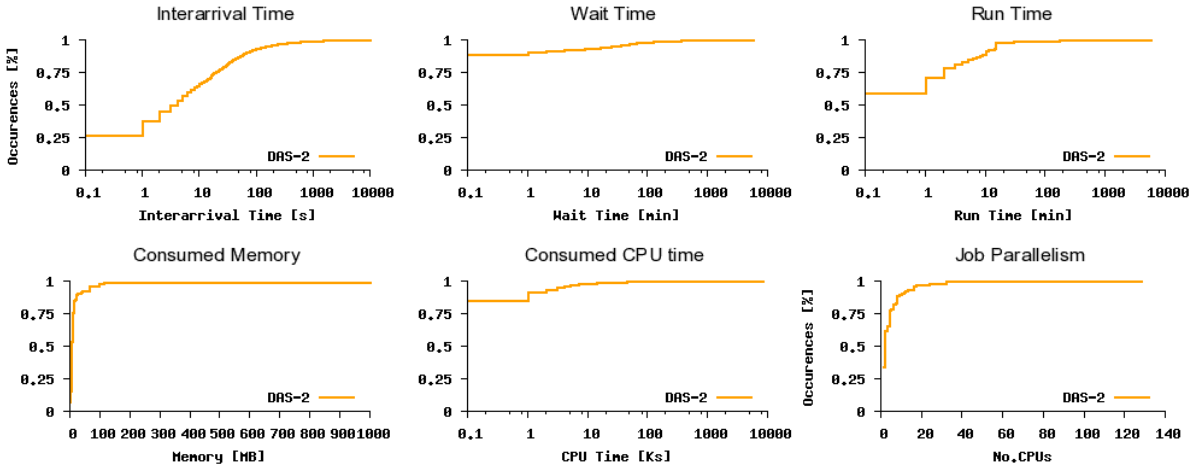

Figure 3: CDFs of the most important job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 128 processors
- Average: 4.306 processors
- Standard deviation: 6.325
- Coefficient of variation: 1.469

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 548363.00 seconds
- Average: 369.72 seconds
- Standard deviation: 3938.101
- Coefficient of variation: 10.652

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 4294.95 MB
- Average: 45.90 MB
- Standard deviation: 346.425
- Coefficient of variation: 7.548

Sequential vs. Parallel jobs

Below we summarize the resource usage of all sequential and all parallel jobs, that is all jobs that use more than one processor. First we calculate the number of sequential jobs and the number of parallel jobs that are submitted to the system. Furthermore, we compute the consumed CPU time by multiplying the runtime of a job by the number of processors allocated to the job. Again, this is divided into parallel and sequential jobs. For the number of jobs and the consumed CPU time, the percentage of all jobs is displayed.

Number of jobs

- Sequential: 384787 jobs (34.21 percent)
- Parallel: 739985 jobs (65.79 percent)

Consumed CPU Time

- Sequential: 215694324 seconds (9.97 percent)
- Parallel: 1948703076 seconds (90.03 percent)

User and group characteristics

User characteristics

Figure 4 shows The number of submitted jobs and the consumed CPU time by user.

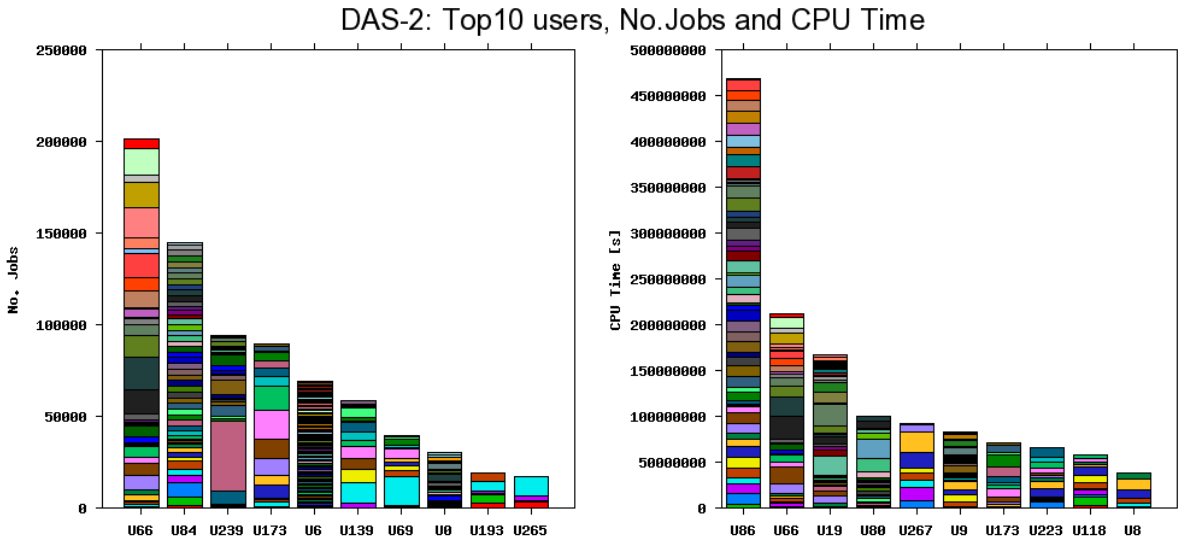

Figure 4: The number of submitted jobs (left) and consumed CPU time (right) by user. Only the top 10 users are displayed. The horizontal axis depicts the user's rank. The vertical axis shows the cumulated values, and the breakdown per week. Users have the same labels in the left and right sub-graphs

Top 10 users by number of job submitted to the system

Table 1 shows Top 10 users by number of jobs submitted to the system.

Table 1: Top 10 users by number of jobs submitted to the system

Rank	UserID	Number of jobs	Percentage
1	U66	201088	17.88%
2	U84	144876	12.88%
3	U239	94323	8.39%
4	U173	89417	7.95%
5	U6	69363	6.17%
6	U139	58361	5.19%
7	U69	39321	3.50%
8	U0	30216	2.69%
9	U193	19069	1.70%
10	U265	17295	1.54%

Table 1: Top 10 users by number of jobs submitted to the system			
Rank	**UserID**	**Number of jobs**	**Percentage**
11	Other	361443	32.13%
12	Total	1124772	100.00%

System utilization

- Minimum: 0.0 percent
- Maximum: 35.12 percent
- Average: 2.232 percent

Job arrival

- Minimum: 0.00 jobs/hour
- Maximum: 7008.00 jobs/hour
- Average: 48.27 jobs/hour

Job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 128 processors
- Average: 3.774 processors
- Standard deviation: 4.123
- Coefficient of variation: 1.092

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 163602.00 seconds
- Average: 167.81 seconds
- Standard deviation: 856.996
- Coefficient of variation: 5.107

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 4294.95 MB
- Average: 50.10 MB
- Standard deviation: 359.133
- Coefficient of variation: 7.168

Top 10 users by consumed CPU time

Table 2 shows Top 10 users by consumed CPU time (in seconds).

Table 2: Top 10 users by consumed CPU time (in seconds)			
Rank	**UserID**	**CPU seconds**	**Percentage**
1	U86	468554924	21.65%
2	U66	212065102	9.80%
3	U19	167202129	7.73%
4	U80	100551669	4.65%
5	U267	92623432	4.28%
6	U9	83391701	3.85%

Table 2: Top 10 users by consumed CPU time (in seconds)			
Rank	**UserID**	**CPU seconds**	**Percentage**
7	U173	70982290	3.28%
8	U223	65848559	3.04%
9	U118	57673112	2.66%
10	U8	38322655	1.77%
11	Other	807181827	37.29%
12	Total	2164397400	100.00%

System utilization

- Minimum: 0.0 percent
- Maximum: 36.36 percent
- Average: 6.708 percent

Job arrival

- Minimum: 0.00 jobs/hour
- Maximum: 971.00 jobs/hour
- Average: 23.46 jobs/hour

Job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 68 processors
- Average: 6.022 processors
- Standard deviation: 7.000
- Coefficient of variation: 1.162

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 302461.00 seconds
- Average: 594.45 seconds
- Standard deviation: 5038.493
- Coefficient of variation: 8.476

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 1505.22 MB
- Average: 7.39 MB
- Standard deviation: 6.181
- Coefficient of variation: 0.836

Group characteristics

Figure 5 shows The number of submitted jobs and consumed CPU time by group.

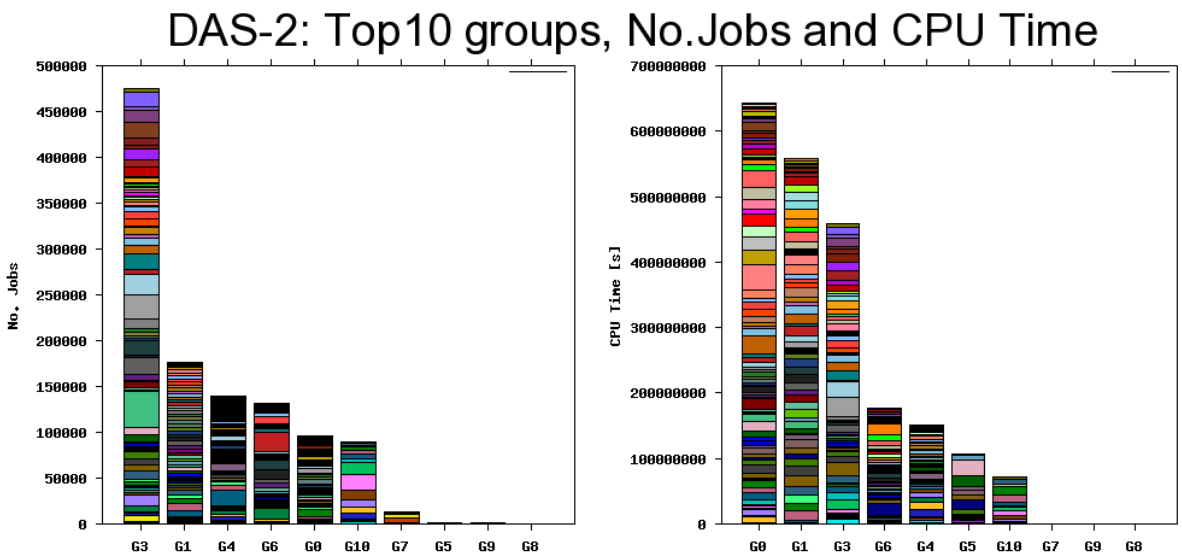

Figure 5: The number of submitted jobs (left) and consumed CPU time (right) by group. Only the top 10 groups are displayed. The horizontal axis depicts the groups rank. The vertical axis shows the cumulated values, and the breakdown per week. Groups have the same labels in the left and right sub-graphs

Table 3 shows Top 10 groups by number of jobs submitted to the system.

Table 3: Top 10 groups by number of jobs submitted to the system

Rank	GroupID	Number of jobs	Percentage
1	G3	475361	42.26%
2	G1	176842	15.72%
3	G4	139552	12.41%
4	G6	132019	11.74%
5	G0	96071	8.54%
6	G10	89417	7.95%
7	G7	13010	1.16%
8	G5	1526	0.14 %
9	G9	662	0.06%
10	G8	309	0.03%
11	Other	3	0.00%
12	Total	1124772	100.00%

Table 4 shows Top 10 Groups by consumed CPU time (in seconds).

Table 4: Top 10 Groups by consumed CPU time (in seconds)

Rank	GroupID	CPU seconds	Percentage
1	G0	642187067	29.67%
2	G1	558828859	25.82%

Table 4: Top 10 Groups by consumed CPU time (in seconds)			
Rank	**GroupID**	**CPU seconds**	**Percentage**
3	G3	457906596	21.16%
4	G6	176464378	8.15%
5	G4	150287662	6.94%
6	G5	106568051	4.92%
7	G10	70982290	3.28%
8	G7	647723	0.03%
9	G9	441177	0.02%
10	G8	83582	0.00%
11	Other	15	0.00%
12	Total	2164397400	100.00%

Performance analysis

Waiting and running jobs

Figure 6 shows The number of running and of waiting jobs during hourly intervals. The vertical axis is limited to 7500 for better visibility.

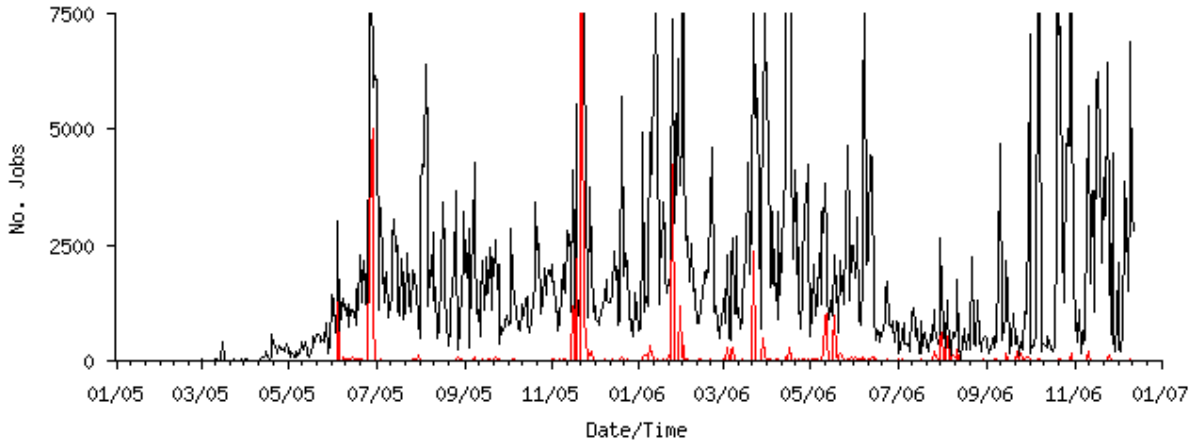

Figure 6: The number of running and of waiting jobs during hourly intervals. The vertical axis is limited to 7500 for better visibility

We compute the number of running and waiting jobs by considering a fixed time interval. In each time interval, we count in the trace the amount of jobs that have been submitted but not yet started, that is, waiting. We also count the number of jobs that have been submitted, and have started executing in the time interval, but did not finish executing, and thus are running. Below we show the values for an interval value of 3600 seconds, summarized in amounts per day. Also the summary for values higher than zero are displayed, which excludes the possible effect of downtime of the system.

Number of waiting jobs per day

- Minimum: 0 jobs
- Maximum: 12347 jobs
- Average: 97.79 jobs

Number of waiting jobs per day (non-zero values)

- Minimum: 1 jobs
- Maximum: 12347 jobs
- Average: 122.74 jobs

Number of running jobs per day

- Minimum: 0 jobs
- Maximum: 19562 jobs
- Average: 1780.92 jobs

Number of running jobs per day (non-zero values)

- Minimum: 2 jobs
- Maximum: 19562 jobs
- Average: 1830.93 jobs

Throughput

We compute the job throughput by considering a fixed time interval. In each time interval, we count in the trace the amount of jobs that have been submitted, started and finished executing. Below we show the values for an interval value of 3600 seconds, summarized in amounts per day. Also the summary for values higher than zero are displayed, which excludes the possible effect of downtime of the system.

Figure 7 shows Throughput during hourly intervals. The vertical axis of each individual site graph is limited to 7500 for better visibility.

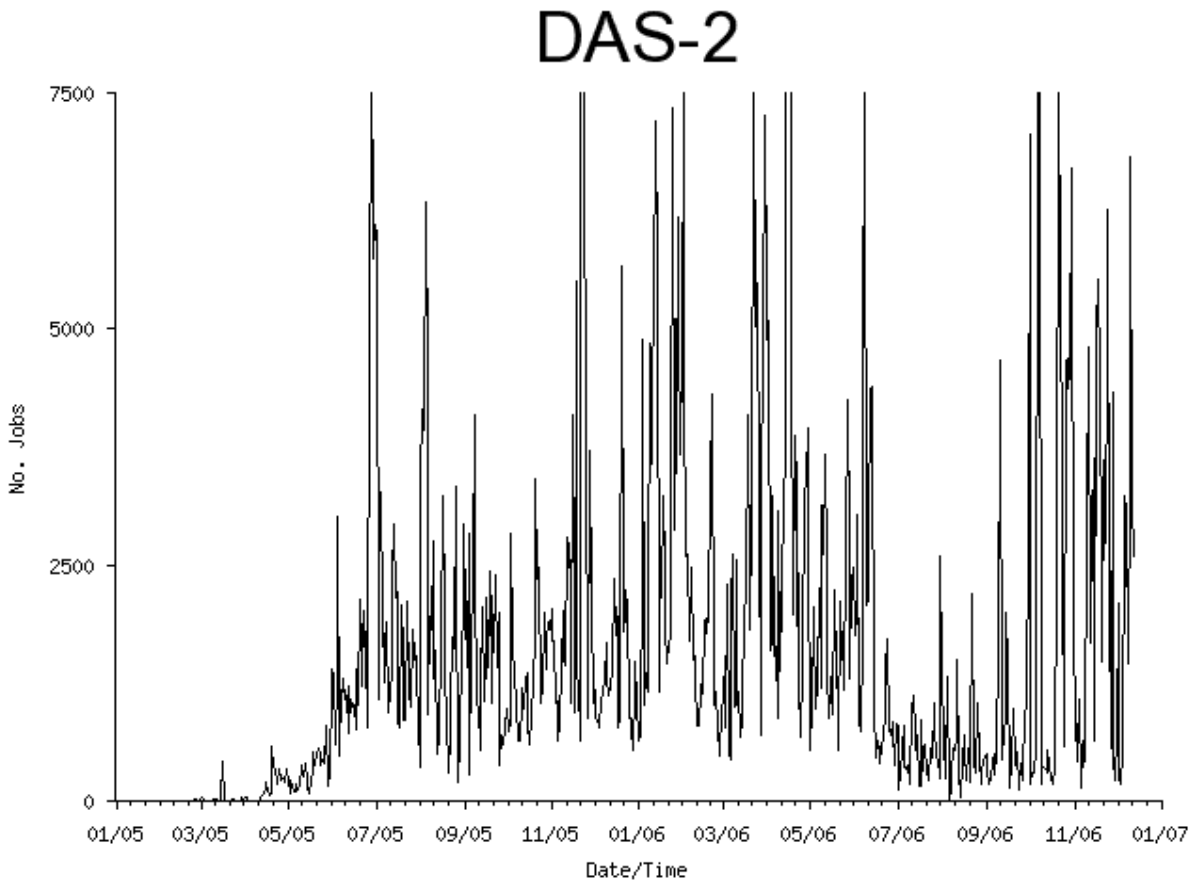

Figure 7: Throughput during hourly intervals. The vertical axis of each individual site graph is limited to 7500 for better visibility

Throughput per day

- Minimum: 0 jobs
- Maximum: 19399 jobs
- Average: 1705.16 jobs

Throughput per day (non-zero values)

- Minimum: 2 jobs
- Maximum: 19399 jobs
- Average: 1753.04 jobs

Completed jobs

Figure 8 shows The number of completed jobs during hourly intervals.

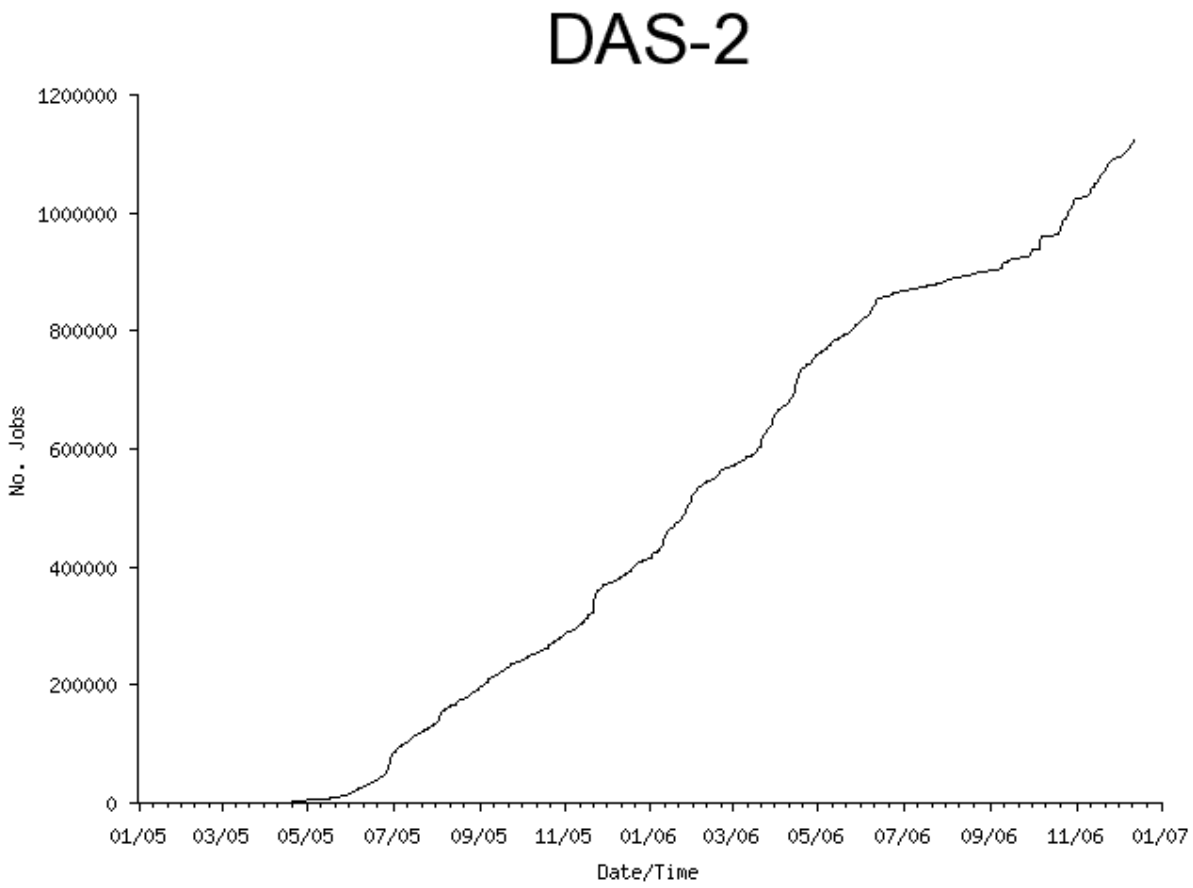

Figure 8: The number of completed jobs during hourly intervals