

Trace analysis report NorduGrid

General information

This is the trace analysis report (generated by reportgen.py) for the NorduGrid system. The trace data was taken from the filename anon_jobs.gwf, which contains job data obtained from. Below is a summary of the contents of the trace data:

- Date first entry: Mon Mar 03 09:31:10 2003
- CPU time consumed by jobs: 2443y 220d 15h 29m 8s
- Number of sites in the system: 75
- Number of CPUs in the trace: 2000
- Number of jobs in the trace: 781370
- Number of users in the trace: 387
- Number of groups in the trace: 107

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. Unfortunately, utilization info for this trace is incomplete.

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 1 shows Overall job arrival rate during hourly intervals.

NorduGrid

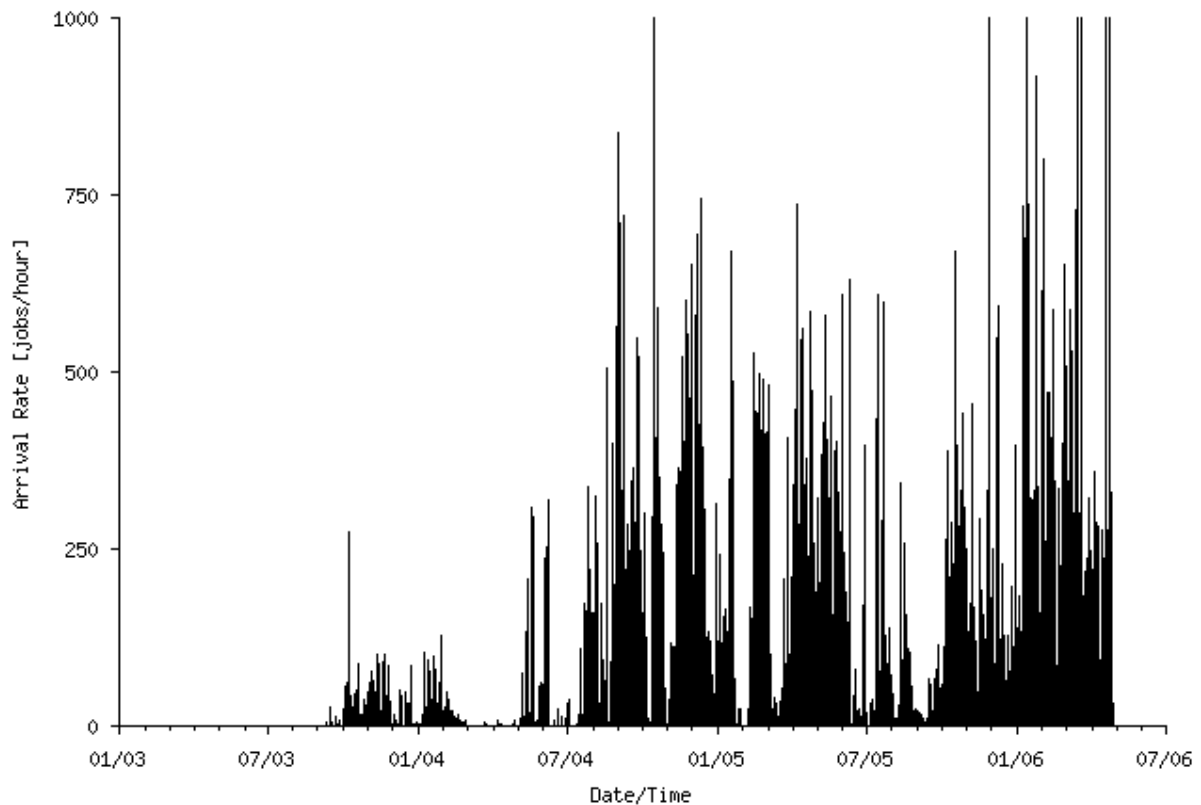


Figure 1: Overall job arrival rate during hourly intervals

Busiest time periods in terms of number of job submissions

- Busiest day: 2004-12-06
- Busiest week: 2004-50
- Busiest month: 2005-04

Overall job arrival metrics

- Minimum: 0.00 jobs/hour
- Maximum: 3614.00 jobs/hour
- Average: 28.11 jobs/hour

Overall job arrival metrics for non-zero values

- Minimum: 2.00 jobs/hour
- Maximum: 3614.00 jobs/hour
- Average: 67.35 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 2 shows CDFs of the most important job characteristics.

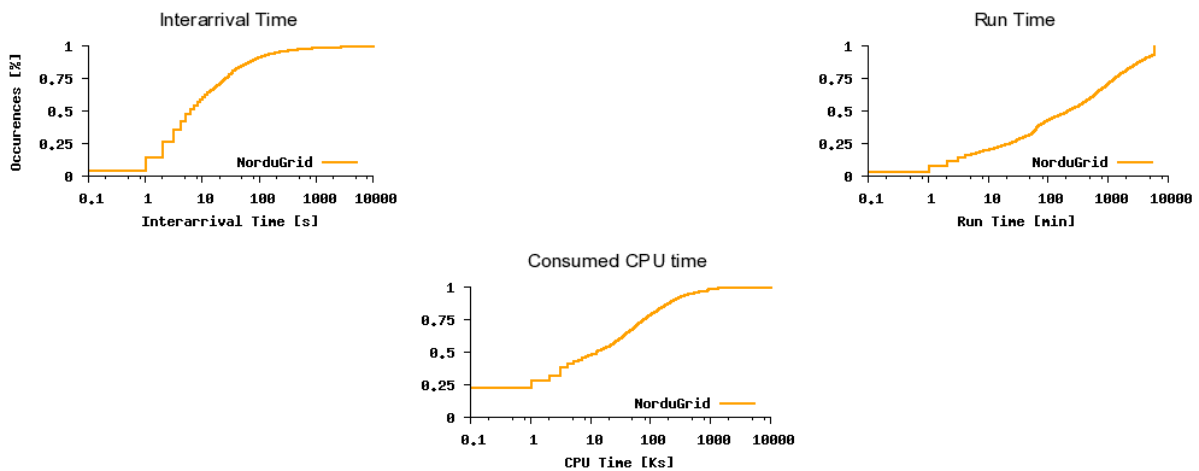


Figure 2: CDFs of the most important job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 64 processors
- Average: 1.073 processors
- Standard deviation: 1.000
- Coefficient of variation: 0.932

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 18071901.00 seconds
- Average: 89273.95 seconds
- Standard deviation: 284299.676
- Coefficient of variation: 3.185

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 2147.48 MB
- Average: 199.89 MB
- Standard deviation: 306.557
- Coefficient of variation: 1.534

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: 776089 jobs (99.32 percent)
- Parallel: 5281 jobs (0.68 percent)

Consumed CPU Time

- Sequential: 69216125790 seconds (89.82 percent)
- Parallel: 7845359571 seconds (10.18 percent)

User and group characteristics

User characteristics

Figure 3 shows The number of submitted jobs and the consumed CPU time by user.

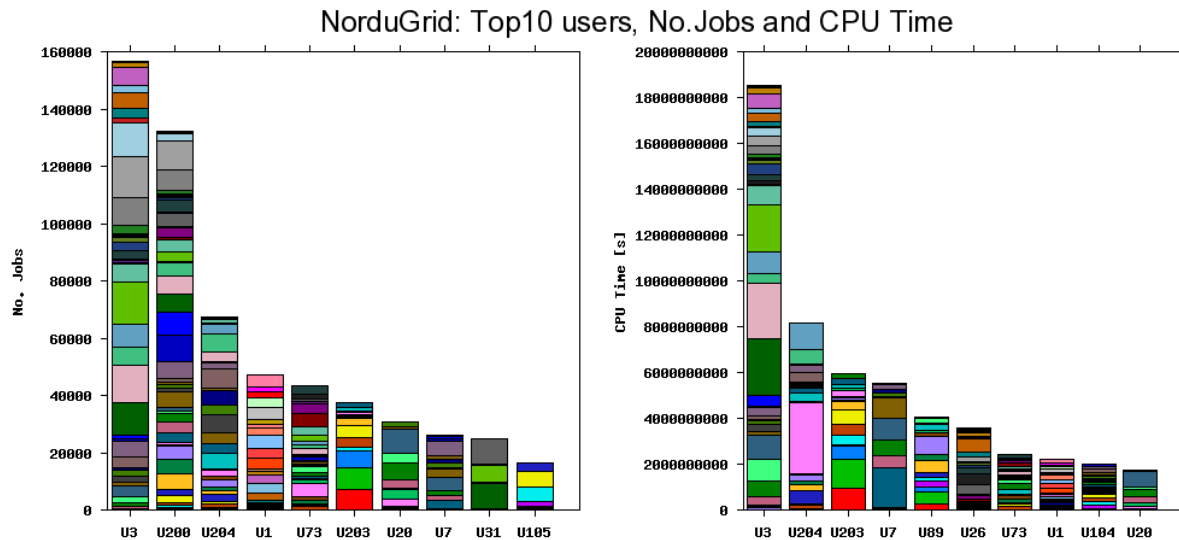


Figure 3: The number of submitted jobs (left) and consumed CPU time (right) by user. Only the top 10 users are displayed. The horizontal axis depicts the user's rank. The vertical axis shows the cumulated values, and the breakdown per week. Users have the same labels in the left and right sub-graphs

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			
Rank	UserID	Number of jobs	Percentage
1	U3	156706	20.06%
2	U200	132112	16.91%
3	U204	67164	8.60%
4	U1	47213	6.04%
5	U73	43475	5.56%
6	U203	37639	4.82
7	U20	30538	3.91%
8	U7	26105	3.34%
9	U31	24771	3.17%
10	U105	16599	2.12%

Table 1			
Rank	UserID	Number of jobs	Percentage
11	Other	199048	25.47%
12	Total	781370	100.00%

Job arrival

- Minimum: 0.00 jobs/hour
- Maximum: 3598.00 jobs/hour
- Average: 25.21 jobs/hour

Job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 1 processors
- Average: 1.000 processors
- Standard deviation: 0.000
- Coefficient of variation: 0.000

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 15138878.00 seconds
- Average: 80135.62 seconds
- Standard deviation: 201451.298
- Coefficient of variation: 2.514

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 2097.15 MB
- Average: 244.32 MB
- Standard deviation: 316.476
- Coefficient of variation: 1.295

Top 10 users by consumed CPU time

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

Table 2			
Rank	UserID	CPU seconds	Percentage
1	U3	18507070921	24.02%
2	U204	8166304553	10.60%
3	U203	5939880670	7.71%
4	U7	5507256824	7.15%
5	U89	4031940747	5.23%
6	U26	3601069398	4.67%
7	U73	2412409829	3.13%
8	U1	2200187706	2.86%
9	U104	1980319076	2.57%
10	U20	1751951254	2.27%

Table 2

Rank	UserID	CPU seconds	Percentage
11	Other	22963120770	29.80%
12	Total	77061511748	100.00%

Job arrival

- Minimum: 0.00 jobs/hour
- Maximum: 1732.00 jobs/hour
- Average: 19.14 jobs/hour

Job characteristics

Number of CPUs used by a single job

- Minimum: 1 processors
- Maximum: 1 processors
- Average: 1.000 processors
- Standard deviation: 0.000
- Coefficient of variation: 0.000

Runtime of a single job

- Minimum: 0.00 seconds
- Maximum: 12114363.00 seconds
- Average: 122380.71 seconds
- Standard deviation: 298947.075
- Coefficient of variation: 2.443

Memory usage of a single job

- Minimum: 0.00 MB
- Maximum: 2097.15 MB
- Average: 204.21 MB
- Standard deviation: 325.944
- Coefficient of variation: 1.596

Group characteristics

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

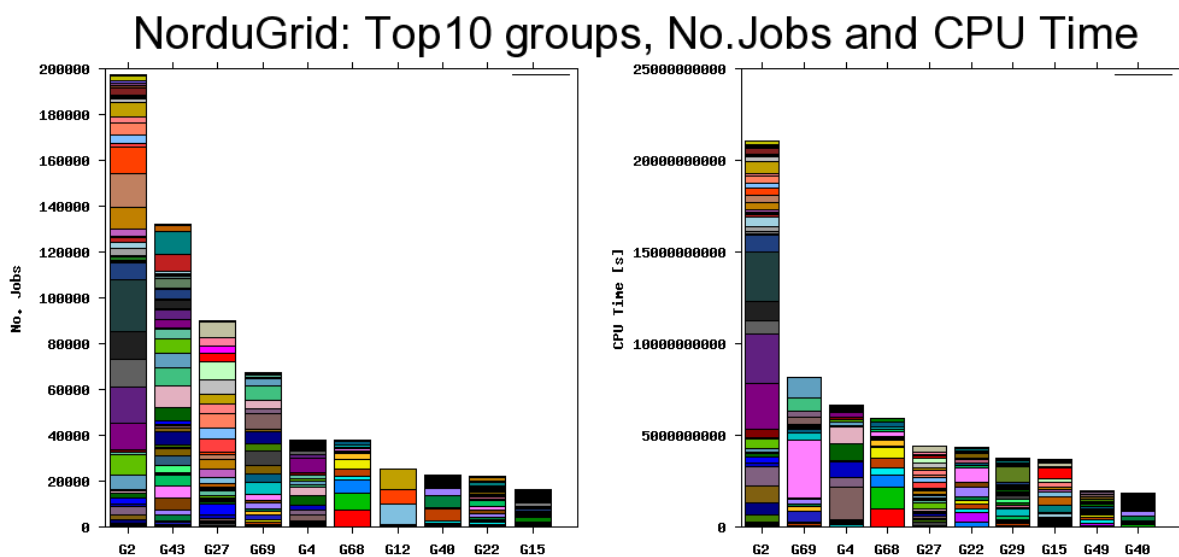


Figure 4: The number of submitted jobs (left) and consumed CPU time (right) by group. Only the top 10 groups are displayed. The horizontal axis depicts the groups

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			
Rank	GroupID	Number of jobs	Percentage
1	G2	197289	25.25%
2	G43	132331	16.94%
3	G27	89777	11.49%
4	G69	67164	8.60%
5	G4	38116	4.88%
6	G68	37639	4.82%
7	G12	25311	3.24%
8	G40	22392	2.87%
9	G22	21961	2.81%
10	G15	16304	2.09%
11	Other	133086	17.03%
12	Total	781370	100.00%

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

Table 4			
Rank	GroupID	CPU seconds	Percentage
1	G2	21067622576	27.34%
2	G69	8166304553	10.60%
3	G4	6620296351	8.59%
4	G68	5939880670	7.71%
5	G27	4400579603	5.71%
6	G22	4351686347	5.65%
7	G29	3780782574	4.91%
8	G15	3695243859	4.80%
9	G49	1980319076	2.57%
10	G40	1853578643	2.41%
11	Other	15205217496	19.73%
12	Total	77061511748	100.00%

Performance analysis

Waiting and running jobs

Figure 5 shows The number of running and of waiting jobs during hourly intervals. The vertical axis is limited to 7500 for better visibility.

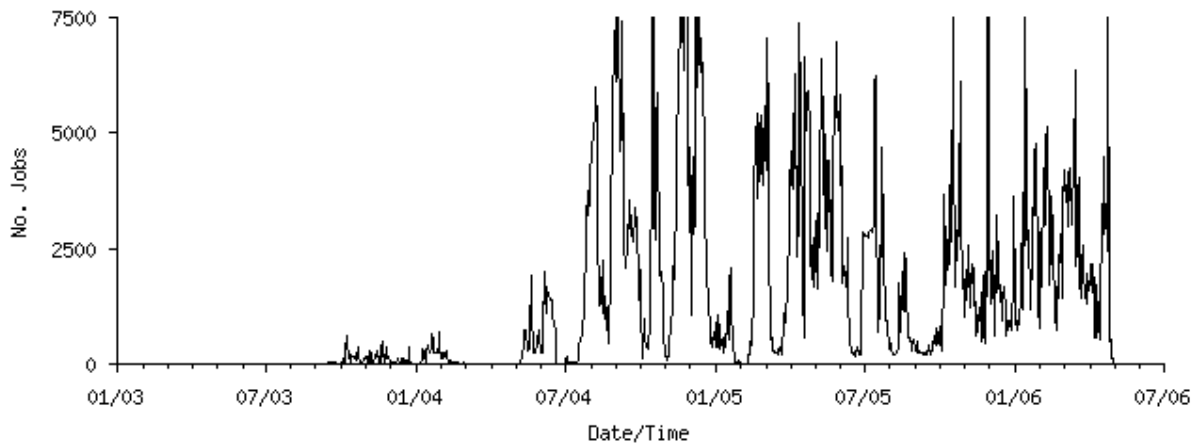


Figure 5: The number of running and of waiting jobs during hourly intervals. The vertical axis is limited to 7500 for better visibility

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: 0 jobs
- Average: 0.00 jobs

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

- Minimum: 0 jobs
- Maximum: 0 jobs
- Average: 0.00 jobs

Number of running jobs per day

- Minimum: 0 jobs
- Maximum: 10901 jobs
- Average: 1410.70 jobs

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

- Minimum: 1 jobs
- Maximum: 10901 jobs
- Average: 1737.86 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 6 shows Throughput during hourly intervals. The vertical axis of each individual site graph is limited to 7500 for better visibility.

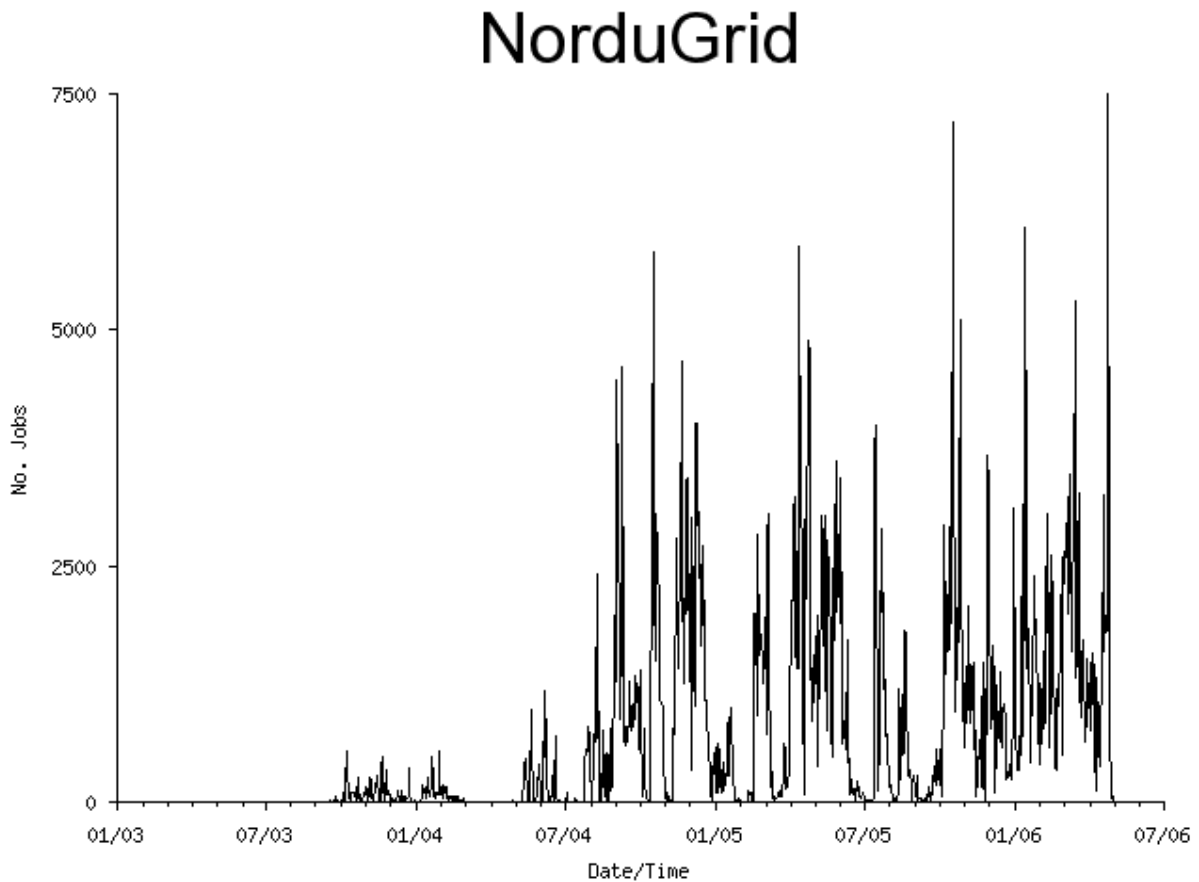


Figure 6: Throughput during hourly intervals. The vertical axis of each individual site graph is limited to 7500 for better visibility

Throughput per day

- Minimum: 0 jobs
- Maximum: 8569 jobs
- Average: 674.76 jobs

Throughput per day (non-zero values)

- Minimum: 1 jobs
- Maximum: 8569 jobs
- Average: 897.09 jobs

Completed jobs

Figure 7 shows The number of completed jobs during hourly intervals.

NorduGrid

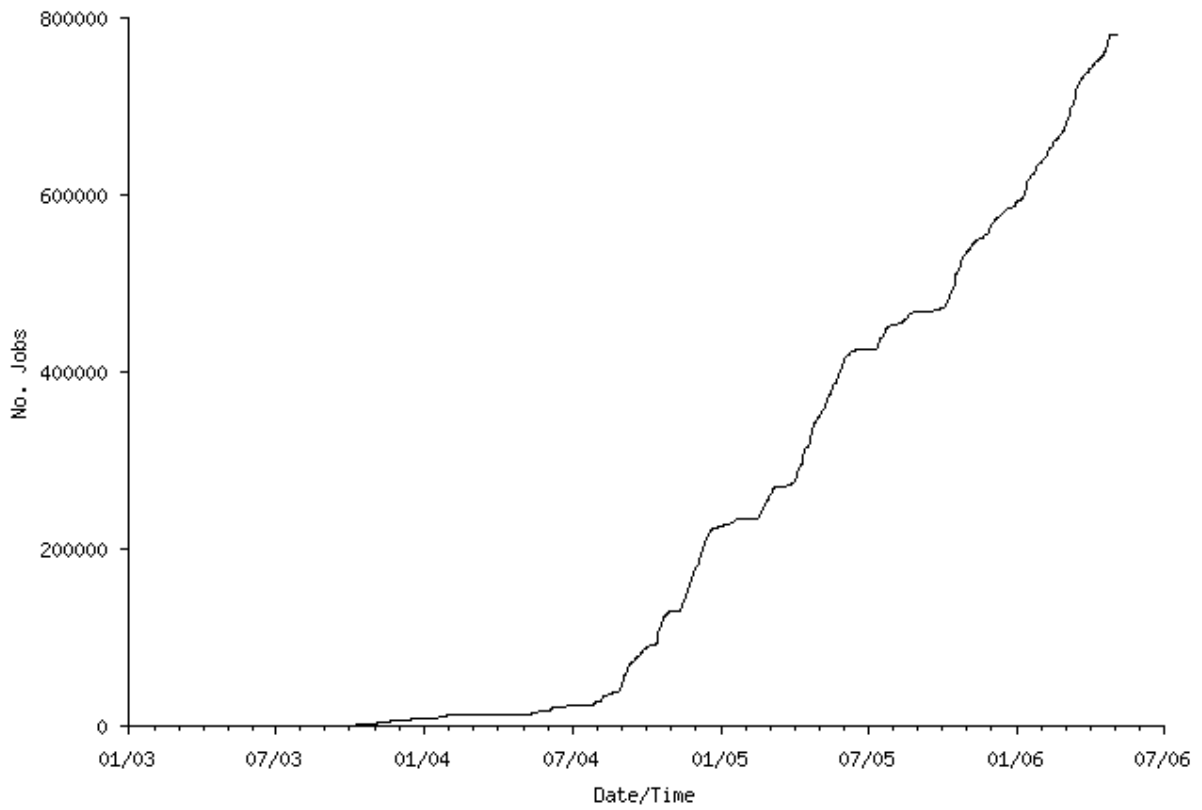


Figure 7: The number of completed jobs during hourly intervals