

GridKa Migration to Grid Engine

WLCG GDB 2013-05-08

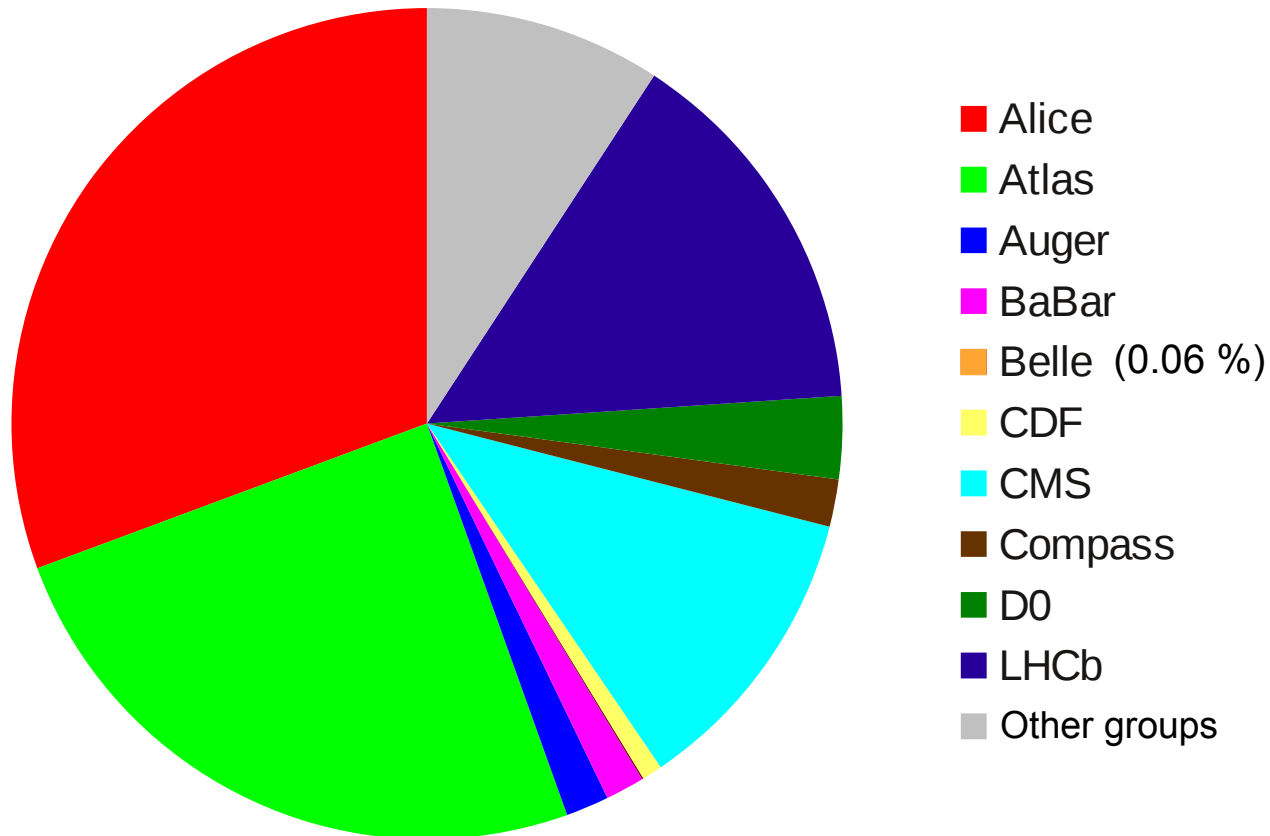
Manfred Alef

STEINBUCH CENTRE FOR COMPUTING — SCC

job-ID	prior	ntckts	name	user	department	state	cpu	me
6216459	0.50001	0.00115	cream_6466	auger	Auger	r	4:06:35:47	0.
6217490	0.50001	0.00115	cream_5296	auger	Auger	r	4:05:57:09	0.
6226049	0.50001	0.00115	cream_5674	auger	Auger	r	4:02:12:07	0.
6226051	0.50001	0.00115	cream_8115	auger	Auger	r	4:02:12:01	0.
6229774	0.50001	0.00115	cream_4640	auger	Auger	r	3:23:23:47	0.
6234079	0.50000	0.00032	cream_1757	auger	Atlas-Pilot	r	3:20:56:31	0.
6234225	0.50000	0.00032	cream_2947	auger	Atlas-Pilot	r	3:20:51:05	0.
6234226	0.50000	0.00032	cream_2107	auger	Atlas-Pilot	r	3:20:51:22	0.
6234255	0.50000	0.00032	cream_6667	auger	Atlas-Pilot	r	3:20:49:50	0.
6234276	0.50000	0.00032	cream_6717	auger	Atlas-Pilot	r	3:20:47:46	0.
6234277	0.50000	0.00032	cream_3350	auger	Atlas-Pilot	r	3:20:47:32	0.
6234284	0.50000	0.00032	cream_5255	auger	Atlas-Pilot	r	3:20:47:03	0.
6234286	0.50000	0.00032	cream_4315	auger	Atlas-Pilot	r	3:20:47:36	0.
6234287	0.50000	0.00032	cream_4665	auger	Atlas-Pilot	r	3:20:47:04	0.
6234289	0.50000	0.00032	cream_6523	auger	Atlas-Pilot	r	3:20:47:04	0.
6234292	0.50000	0.00032	cream_1092	auger	Atlas-Pilot	r	3:20:47:17	0.

GridKa Compute Farm

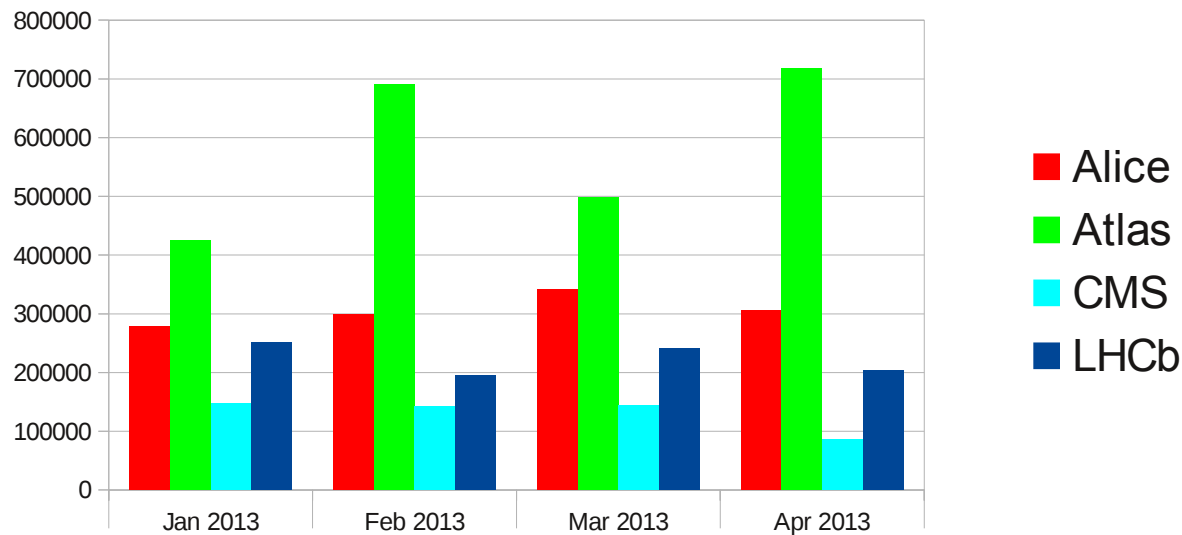
- Single cluster
 - Fair-share policy configured according to pledged VO shares



GridKa Compute Farm

■ Current cluster size

- 146 kHS06
- 1,000 WNs
- 11,800 (physical) cores
- 17,400 logical (hyperthreaded) cores
- **14,600 job slots**
- **~ 1,500,000 jobs per month**



LRMS Used in the Past

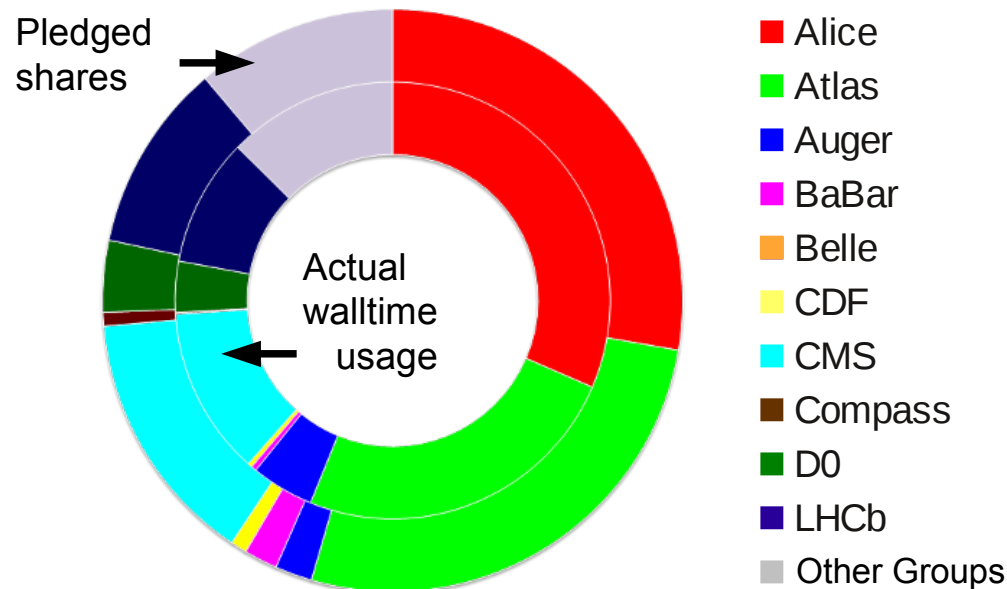
- Batch system used at GridKa until 2012:
PBS Professional
 - Many troubles, e.g.
 - Black hole node issues
 - Corrupted job descriptions in database, couldn't restart batch scheduler to recover
 - More and more issues with every cluster expansion
 - Bug in new release of licence management tool has caused unscheduled downtime of 3+ days

LRMS Used in the Past

- Batch system used at GridKa until 2012:
PBS Professional (cont.)
 - 2010: Split into 2 sub-clusters to improve stability
 - More stable than before, but still recurring troubles
 - 2012: New WNs shipped, added to batch farm:
 - (Only the number of WNs had been increased, but constant number of running or queued jobs!)
 - Stability of the batch farm decreased significantly

LRMS Used in the Past

- Batch system used at GridKa until 2012: PBS Professional (cont.)
 - *In average* good correlation of pledged shares and accounting records [HS06 hours]



- Nevertheless many complains by users about fair-share scheduling / oscillating number of running jobs / high waiting time of jobs

Comparative Studies and Tests

- Looking for alternative batch systems (2010 ... 2011):
 - Several LRMS' tested
 - SGE+UGE test installations on complete farm (in parallel with PBS)
 - Up to 1000 job slots per WN configured
 - Hammered with Millions of jobs (sleepers)
 - Tests successfully completed
 - Stability + performance

- Small test cluster (~ 2,000 slots) managed by UGE since summer 2012

Final LRMS Migration

- Dezember 2012:
Migration of the whole cluster to UGE
 - Testing phase till March 2013
 - No serious problems so far

- March 2013:
Licence agreement with UNIVA extended for 2 years

Final LRMS Migration

- Some configuration details:
 - Single server, no failover
 - Flat files (no DB)
 - Certificate Security Protocol (CSP) enabled
 - Fair-share configurations:
 - Based on reserved usage (aka walltime)
 - Share-tree (using historical data) and functional ticket policy, weighting: 50% each

Final LRMS Migration

- Experiences (cont.):
 - Steep learning curve at the beginning



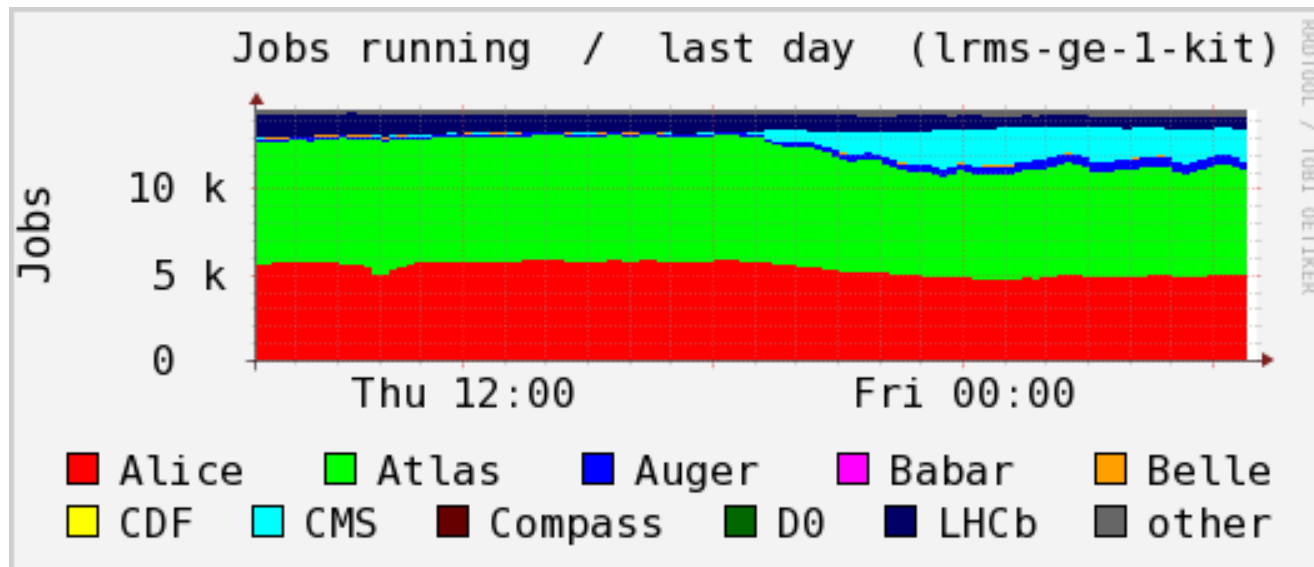
- Documentations partially not clear and intelligible, or incomplete
 - A lot of developers' slang
- Some bugs found (uncritical)



- Very quick and efficient responses by Univa support desk (located in Germany; no time shift – same time zone as Karlsruhe)

Final LRMS Migration

- Experiences:
 - Stable operation, no crashes, no black hole node issues
 - Smooth fair-share scheduling, no complains by VOs / users so far



Final LRMS Migration

- Missing important feature:
 - Monitoring of job efficiencies (cpu-per-walltime ratio of running jobs)
 - Essential for site and users VOs
 - CPU usage reported by (original) qstat is reserved CPU usage (aka walltime), according to fair-share settings
 - Has been discussed with developer – Univa has already provided a patch to provide consumed CPU as well as walltime



```
# qstat -j \* | grep ^usage | head -5 | cut -d, -f-2
usage 1: wallclock=4:11:08:30, cpu=4:08:59:00 ✓
usage 1: wallclock=4:10:54:56, cpu=4:10:25:41 ✓
usage 1: wallclock=4:10:19:25, cpu=4:10:02:35 ✓
usage 1: wallclock=4:07:06:39, cpu=00:00:09 ✗
usage 1: wallclock=4:07:05:33, cpu=4:04:57:06 ✓
#
```

Questions, Comments?