



Enabling Grids for E-science

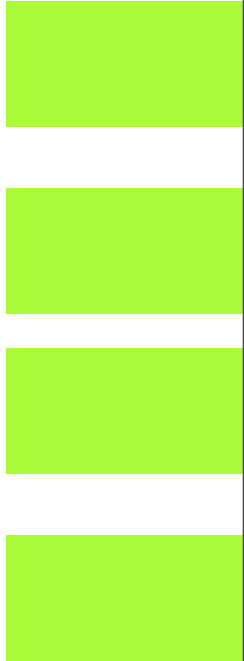
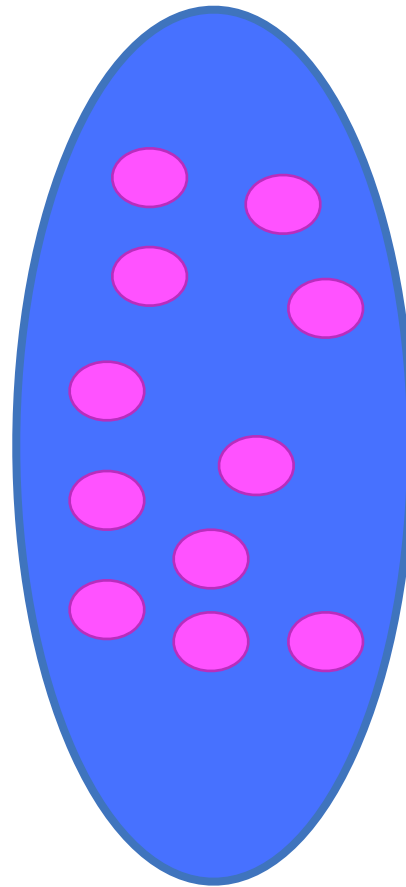
TORQUE and MAUI Tutorial GD White Areas April 2008

Steve Traylen, CERN, steve.traylen@cern.ch

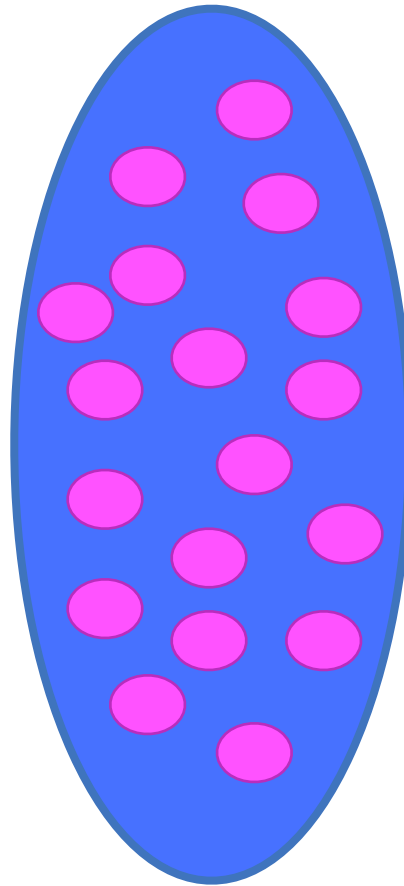


- **Torque and MAUI easily the most prominent in EGEE.**
 - MAUI can be used with many batch systems SGE, LSF,
 - § As well as CASTOR and DIRAC....
 - § Not Condor AFAIK.
- **Overview,**
 - Maui Priorities,
 - Hard and Soft Limits
 - Maui Reservations
 - Extra configurations needed for LCG.

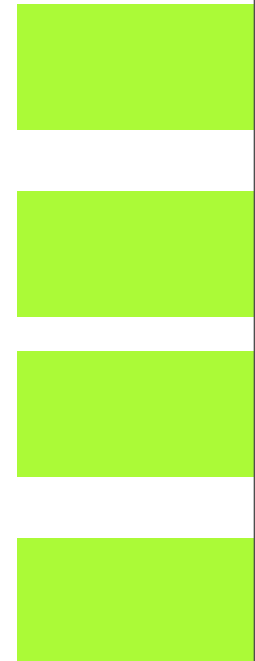
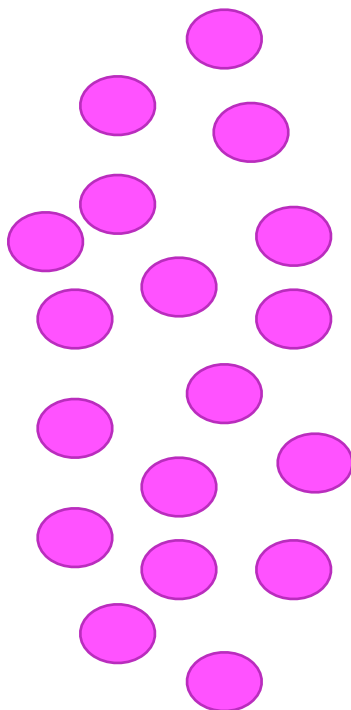
- **What is TORQUE's job as the resource manager.**
 - Accepting and starting jobs across a batch farm.
 - Stopping jobs.
 - Monitoring the state of jobs.
 - Collecting return codes.
- **What is MAUI's Job?**
 - MAUI makes all the decisions.
 - Should a job be started asking questions like:
 - § Should I kill any jobs for over running?
 - § Is there enough resource to start the job?
 - § Given all the jobs I could start which one should I start?
- **MAUI runs a scheduling iteration:**
 - When a job is submitted or ends
 - § Torque pokes Maui.
 - At regular configurable intervals.



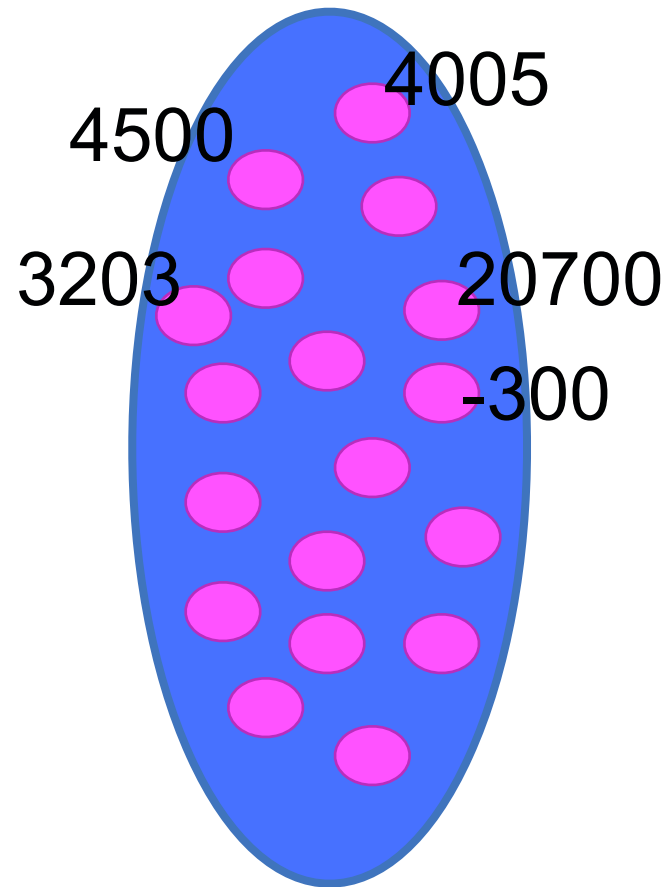
- ◆ Jobs are submitted into a pool of jobs.
- ◆ Forget about queues, MAUI considers all jobs.



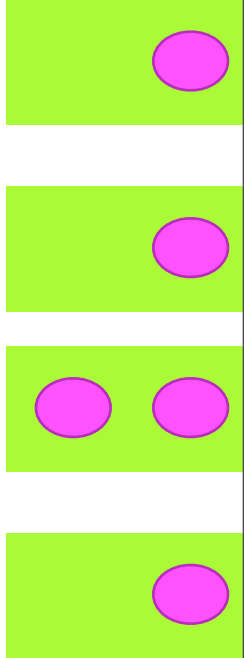
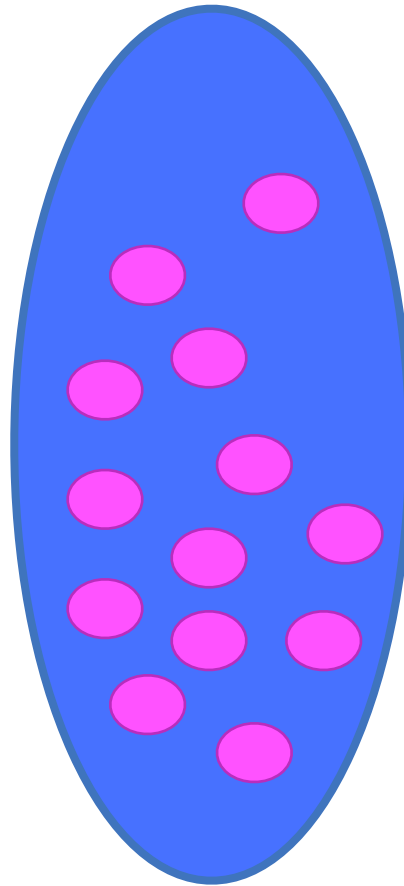
- ◆ Maui scans through all the jobs and nodes:
 - ◆ When a job is submitted.
 - ◆ When a job completes.
 - ◆ And at periodic intervals.



- ◆ Each job has a priority number calculated.



- ◆ Each job has a priority number calculated.
- ◆ The highest priority is executed first.



- **A job's priority is made up from components:**
 - CRED* = Credentials, e.g user or group name, submission queue, ...
 - FS* = Fairshare, e.g considers historical usage of user, group,
 - RES = Resources, e.g. Number of nodes requested, length of job, ..
 - SERV* = Service, e.g Time job has been queued,
 - TARGET = Target, e.g User must run 500 jobs in two days.
 - USAGE = Usage e.g Time consumed by jobs running now.

- **A job's priority is made up from components:**
 - CRED* = Credentials, e.g user or group name, submission queue, ...
 - FS* = Fairshare, e.g considers historical usage of user, group,
 - RES = Resources, e.g. Number of nodes requested, length of job, ..
 - SERV* = Service, e.g Time job has been queued,
 - TARGET = Target, e.g User must run 500 jobs in two days.
 - USAGE = Usage e.g Time consumed by jobs running now.
- **Each component is weighted and summed to form the priority,**

$$\text{PRIORITY} = \text{CREDWEIGHT} * (\text{CREDComp}) + \text{FSWEIGHT} * (\text{FSComp}) + \dots$$

- **A common mistake is to leave say FSWEIGHT at 0 having configured FS.**
- **Components, e.g. CREDComp are made up of SubComponents Will only look at *s today.**

- **CRED components are static contributions to the overall priority number. e.g username, groupname, submission queue (class).**

| Config Attribute | Value | Summary |
|--------------------------|-----------------------|--|
| CREDWEIGHT | 10 | Component Weight |
| USERWEIGHT | 20 | SubComp' Weight |
| USERCFG[straylen] | PRIORITY=1000 | Static Priority for Me. |
| CLASSWEIGHT | 5 | SubComp' Weight |
| CLASSCFG[short] | PRIORITY=10000 | Static Priority for short Queue |

- CRED components are static contributions to the overall priority number. e.g username, groupname, submission queue (class).

| Config Attribute | Value | Summary |
|-------------------|----------------|---------------------------------|
| CREDWEIGHT | 10 | Component Weight |
| USERWEIGHT | 20 | SubComp' Weight |
| USERCFG[straylen] | PRIORITY=1000 | Static Priority for Me. |
| CLASSWEIGHT | 5 | SubComp' Weight |
| CLASSCFG[short] | PRIORITY=10000 | Static Priority for short Queue |

$$\text{PRIORITY} = \text{CREDWEIGHT} * (\text{CREDComp}) + \text{FSWEIGHT} * (\text{FSComp}) + \dots$$

- CRED components are static contributions to the overall priority number. e.g username, groupname, submission queue (class).

| Config Attribute | Value | Summary |
|-------------------|----------------|---------------------------------|
| CREDWEIGHT | 10 | Component Weight |
| USERWEIGHT | 20 | SubComp' Weight |
| USERCFG[straylen] | PRIORITY=1000 | Static Priority for Me. |
| CLASSWEIGHT | 5 | SubComp' Weight |
| CLASSCFG[short] | PRIORITY=10000 | Static Priority for short Queue |

$$\text{PRIORITY} = \text{CREDWEIGHT} * (\text{CREDComp}) + \text{FSWEIGHT} * (\text{FSComp}) + \dots$$

$$\text{CREDComp} = \text{USERWEIGHT} * (\text{USERCFG[straylen] priority}) + \text{CLASSWEIGHT} * (\text{CLASSCFG[short] priority}) + \dots$$

- The the “diagnose -p” command is used for this.

```
[root@lxb1407 root]# diagnose -p
diagnosing job priority information (partition: ALL)

Job                PRIORITY*      Cred( User:Class)
                Weights      -----
34                 700000         100.0(1000.:10000)
35                 700000         100.0(1000.:10000)
36                 700000         100.0(1000.:10000)
37                 700000         100.0(1000.:10000)
38                 700000         100.0(1000.:10000)
39                 700000         100.0(1000.:10000)
40                 700000         100.0(1000.:10000)

Percent Contribution ----- 100.0( 28.6: 71.4)

* indicates system prio set on job

[root@lxb1407 root]#
```

- Provides a break down of priority number components.

- **FS considers historical usage of the batch service.**
 - FSUSER => FS by username
 - FSGROUP => FS by groupname
 -
- **The USAGE**
 - MAUI calculates usage for each USER, GROUP, CLASS, QOS and ACCOUNT.
- **The TARGET**
 - SysAdmin can specify a TARGET for every USER, GROUP, CLASS, QOS or ACCOUNT.
- **Comparison of USAGE and TARGET.**
 - For each FSSubComponent e.g. FSUSER the used and target values are compared to give a contribution to a queued jobs priority value.

- **FSPOLICY=DEDICATEDPS**, uses **walltime** as the metric.

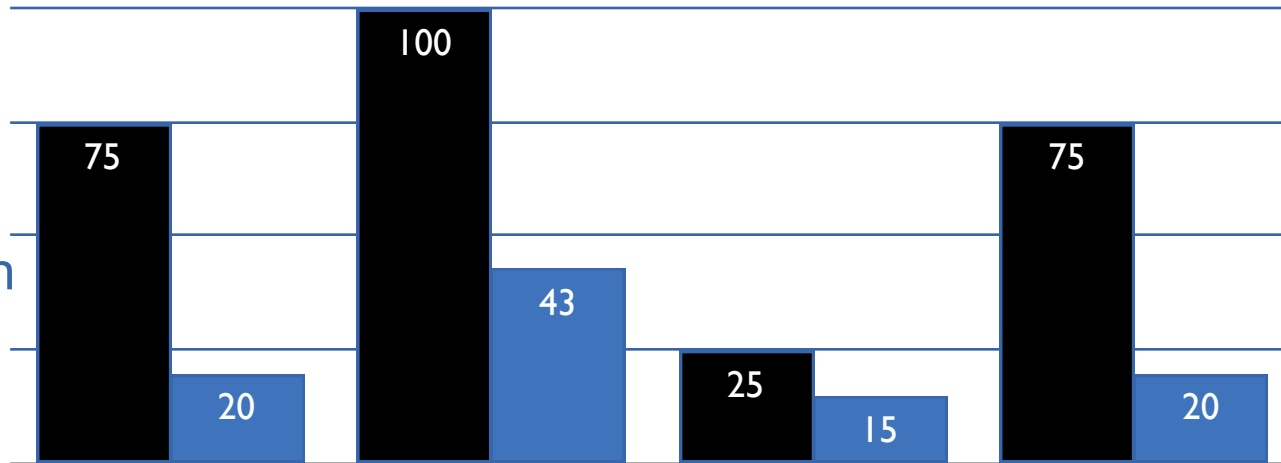
$$Usage = \frac{\sum_{i=0}^{DEPTH-1} (U_i * DECAY^i)}{\sum_{i=0}^{DEPTH-1} (T_i * DECAY^i)}$$

- **FSPOLICY=DEDICATEDPS**, uses **walltime** as the metric.

$$Usage = \frac{\sum_{i=0}^{DEPTH-1} (U_i * DECAY^i)}{\sum_{i=0}^{DEPTH-1} (T_i * DECAY^i)}$$



FSDEPTH=4
FSDECAY=0.5
FSINTERVAL=24h



0-24 hours 24-48 hours 48-72 hours 72-96 hours

$$USAGE = \frac{20 * 0.5^0 + 43 * 0.5^1 + 15 * 0.5^2 + 20 * 0.5^3}{75 * 0.5^0 + 100 * 0.5^1 + 25 * 0.5^2 + 20 * 0.5^3}$$

- For each user, group, class a target can be specified in the configuration.

| Config Attribute | Value | Summary |
|-------------------|---------------|------------------------------|
| FSWEIGHT | 10 | Component Weight |
| FSUSERWEIGHT | 20 | SubComp' Weight |
| USERCFG[straylen] | FSTARGET=1000 | FS target for me. |
| USERCFG[fred] | FSTARGET=500 | FS target for Fred. |
| USERCFG[DEFAULT] | FSTARGET=20 | FS target for everyone else. |

- Note: The share will be 1000:500:20:20:20:.....
 - Number of users can make a large difference.
 - Solution: Avoid [DEFAULT] ,easy for groups, ...
- Have your FSTARGETS add to 100 if possible.
- USAGE is reported as a % so diagnosis easier.

- A comparison of the target and usage for the user, group or class then gives the contribution to the jobs overall priority.
- There are two configurations for this calculation:
 - Difference - `FSPOLICY=DEDICATEDPS` is rubbish.
 - Ratio - `FSPOLICY=DEDICATEDPS%` is much better.

- A comparison of the target and usage for the user, group or class then gives the contribution to the jobs overall priority.
- There are two configurations for this calculation:
 - Difference - FSPOLICY=DEDICATEDPS is rubbish.
 - Ratio - FSPOLICY=DEDICATEDPS% is much better.

$$\begin{aligned}
 \text{PRIORITY} &= \text{CREDWEIGHT} * (\text{CREDComp}) + \text{FSWEIGHT} * (\text{FSComp}) + \dots \\
 \text{FSComp} &= \text{FSUSERWEIGHT} * (1 - \text{straylen's fsusage/straylens' fstarget}) \\
 &+ \text{FSGROUPWEIGHT} * (1 - \text{dteam's fsuage/dteam's fstarget}) + \dots
 \end{aligned}$$

- To interrogate fairshare status use “diagnose -f”.

```

fred@lxb1407:~
[fred@lxb1407 fred]$ diagnose -f
FairShare Information

Depth: 4 intervals   Interval Length: 00:01:00   Decay Rate: 0.50

FS Policy: DEDICATEDPS
System FS Settings: Target Usage: 0.00   Flags: 0

FSInterval      %      Target      0      1      2      3
FSWeight        -----
TotalUsage      100.00 -----
                0.1      0.1      0.1      0.1

USER
-----
straylen*       93.67 1000.00 100.00  95.65  75.00  75.00
fred*           6.33  500.00 -----  4.35  25.00  25.00

GROUP
-----
straylen        93.67 ----- 100.00  95.65  75.00  75.00
fred            6.33 ----- -----  4.35  25.00  25.00

CLASS
-----
batch           100.00 ----- 100.00 100.00 100.00 100.00

[fred@lxb1407 fred]$
    
```

- **Now we are using two components, CRED and FS.**
 - The components are in direct competition with another, they must be tuned. Use “diagnose -p” again.

```

root@lxb1407:/var/spool/maui
[root@lxb1407 maui]# diagnose -p
diagnosing job priority information (partition: ALL)

Job              Weights  PRIORITY*  Cred( User;Class)  FS( User)
-----
77              25300    99.2(123.0; 10.0)  0.8(200.0)
78              25300    99.2(123.0; 10.0)  0.8(200.0)
79              25300    99.2(123.0; 10.0)  0.8(200.0)
80              25300    99.2(123.0; 10.0)  0.8(200.0)
81              25300    99.2(123.0; 10.0)  0.8(200.0)
82              25300    99.2(123.0; 10.0)  0.8(200.0)
55              20680    99.1(100.0; 10.0)  0.9(180.0)
56              20680    99.1(100.0; 10.0)  0.9(180.0)
57              20680    99.1(100.0; 10.0)  0.9(180.0)
58              20680    99.1(100.0; 10.0)  0.9(180.0)
59              20680    99.1(100.0; 10.0)  0.9(180.0)

Percent Contribution  -----  99.2( 97.0;  2.2)  0.8(  0.8)

* indicates system prio set on job

[root@lxb1407 maui]#
    
```

- Allows us to group types of jobs together based on a credential. Can be queues, users, groups,....
- Required for recommendations of job priority working group.
 - Starting point is jobs are submitted in groups lhcba, lhccb, lhcbc, cmsa, cmsb, cmsc representing different roles with LHCb and CMS.

```

GROUPCFG[lhcba]  FSTARGET=20  QDEF=qlhcb
GROUPCFG[lhccb]  FSTARGET=20  QDEF=qlhcb
GROUPCFG[cmsa]   FSTARGET=80  QDEF=qcms
GROUPCFG[cmsb]   FSTARGET=20  QDEF=qcms
QOSCFG[qcms]     FSTARGET=40
QOSCFG[qlhcb]    FSTARGET=60
FSGROUPWEIGHT 100
FSQOSWEIGHT 1000
    
```

$$\text{FSComp(cmsa)} = \text{FSGROUPWEIGHT} * (1 - \text{cmsa's fsusage}/\text{cmsa' fstarget}) + \text{FSQOSWEIGHT} * (1 - \text{qcms's fsuage}/\text{qcms's fstarget}) + \dots$$

- **Hard Limits**

- Allow an absolute cap to be introduced for a credential.

| Credential | Value | Details |
|-------------------|------------------------|--|
| USERCFG[straylen] | MAXJOB=20 | Limits me to 20 running jobs. |
| GROUPCFG[dteam] | MAXWC=1h | Limits dteam to only have 1hour of walltime remaining. |
| CLASSCFG[short] | MAXJOB[GROUP]=5 | Any group can run 5 jobs in the short queue. |
| CLASSCFG[short] | MAXJOB[GROUP:dteam]=10 | Group dteam can run 10 jobs in the short queue. |

- Jobs can be in three states:
 - RUNNING (on cpu), IDLE (eligible to run), BLOCKED (Non-Eligible)
- Can easily result in idle CPUs , not good.....

- **showq is your friend. USERCFG[straylen] MAXJOB=2**

```

root@lxb1407:~
ACTIVE JOBS-----
JOBNAME          USERNAME      STATE  PROC  REMAINING          STARTTIME
91               straylen     Running  1    00:58:54  Tue Jan 23 10:12:34
92               straylen     Running  1    00:58:54  Tue Jan 23 10:12:34
102              fred        Running  1    00:59:16  Tue Jan 23 10:12:56
103              fred        Running  1    00:59:16  Tue Jan 23 10:12:56

    4 Active Jobs      4 of      4 Processors Active (100.00%)
                       2 of      2 Nodes Active      (100.00%)

IDLE JOBS-----
JOBNAME          USERNAME      STATE  PROC  WCLIMIT          QUEUETIME
106              fred         Idle   1    1:00:00  Tue Jan 23 10:12:54
107              fred         Idle   1    1:00:00  Tue Jan 23 10:12:54
108              fred         Idle   1    1:00:00  Tue Jan 23 10:12:55
109              fred         Idle   1    1:00:00  Tue Jan 23 10:12:56

4 Idle Jobs



BLOCKED JOBS-----
JOBNAME          USERNAME      STATE  PROC  WCLIMIT          QUEUETIME
94               straylen     Idle   1    1:00:00  Tue Jan 23 10:12:29
95               straylen     Idle   1    1:00:00  Tue Jan 23 10:12:30
97               straylen     Idle   1    1:00:00  Tue Jan 23 10:12:31

Total Jobs: 11  Active Jobs: 4  Idle Jobs: 4  Blocked Jobs: 3
[root@lxb1407 root]#
    
```




































| Credential | Value | Details |
|-----------------|-------------|--|
| GROUPCFG[atlas] | MAXJOB=2,3 | Run a max 2 jobs unless all soft limits are reached. |
| GROUPCFG[alice] | MAXJOB=5,10 | Run a max 3 jobs unless all soft limits are reached. |

- **Jobs over soft limit wait for other jobs below soft limit.**
 - **Soft limits are satisfied first.**
- **Can be used for non historical fairshare.**
 - **e.g 100 slot farm, MAXJOB=25,1000 will give 25%**
- **Can be used for offering a basic level of service.**
 - **e.g 100 slot farm,**
 - **GROUPCFG[DEFAULT] MAXJOB=10,1000**
 - **For any group starting they will rush up to 10 jobs**


Example of Soft and Hard Limits

 ATLAS JOB MAXJOB=1,4
 CMS JOBS MAXJOB=3,5

 Job Slots (4)

| TRANSITION | BLOCKED | IDLE | | | | | | | RUNNING | | | | |
|-------------------|---------|---|---|---|---|---|---|---|---|---|---|---|--|
| Submit Jobs(1-7) | |  |  |  |  |  |  |  |  |  |  |  | |
| Jobs Start | | | | | | | | |  |  |  |  | |
| Submit Jobs(8-10) | | | | | | | | |  |  |  |  | |
| Jobs End | | | | | | | | |  |  |  |  | |
| Jobs Start | | | | | | | | |  |  |  |  | |
| Jobs End | | | | | | | | |  |  |  |  | |
| Jobs Start | | | | | | | | |  |  |  |  | |


Example of Soft and Hard Limits

 ATLAS JOB MAXJOB=1,4
 CMS JOBS MAXJOB=3,5

































 Job Slots (4)

| TRANSITION | BLOCKED | IDLE | | | | | | | RUNNING | | | |
|-------------------|---------|------|--|--|--|--|--|--|---------|--|--|--|
| Submit Jobs(1-7) | | | | | | | | | | | | |
| Jobs Start | | | | | | | | | | | | |
| Submit Jobs(8-10) | | | | | | | | | | | | |
| Jobs End | | | | | | | | | | | | |
| Jobs Start | | | | | | | | | | | | |
| Jobs End | | | | | | | | | | | | |
| Jobs Start | | | | | | | | | | | | |

Example of Soft and Hard Limits

 ATLAS JOB MAXJOB=1,4
 CMS JOBS MAXJOB=3,5

 Job Slots (4)

| TRANSITION | BLOCKED | IDLE | | | | | | RUNNING | | | | | | | | | | | | | |
|-------------------|---|---|---|--|--|--|--|---------|---|---|---|---|---|----|---|---|---|---|---|---|---|
| Submit Jobs(1-7) | | | | | | | | 7 | 6 | 5 | 4 | 3 | 2 | 1 |  |  |  |  | | | |
| Jobs Start | | | | | | | | | | | | | | 7 | 6 | 3 |  |  |  |  | |
| Submit Jobs(8-10) |  |  |  |  | | | | | | | | | | 10 | 8 |  |  |  |  | | |
| Jobs End | | | | | | | | | | | | | | | | | |  |  |  |  |
| Jobs Start | | | | | | | | | | | | | | | | | |  |  |  |  |
| Jobs End | | | | | | | | | | | | | | | | | |  |  |  |  |
| Jobs Start | | | | | | | | | | | | | | | | | |  |  |  |  |

Example of Soft and Hard Limits



ATLAS JOB

CMS JOBS

MAXJOB=1,4

MAXJOB=3,5



Job Slots (4)

| TRANSITION | BLOCKED | IDLE | | | | | | RUNNING | | | | | | | | | | | | | | |
|-------------------|---------|------|---|---|--|--|--|---------|---|---|---|---|---|---|---|---|----|---|---|---|---|---|
| Submit Jobs(1-7) | | | | | | | | 7 | 6 | 5 | 4 | 3 | 2 | 1 | | | | | | | | |
| Jobs Start | | | | | | | | | | | | | | 7 | 6 | 3 | 2 | 5 | 4 | 1 | | |
| Submit Jobs(8-10) | 9 | 7 | 6 | 3 | | | | | | | | | | | | | 10 | 8 | 2 | 5 | 4 | 1 |
| Jobs End | 9 | 7 | 6 | 3 | | | | | | | | | | | | | 10 | 8 | 2 | 5 | 4 | |
| Jobs Start | | | | | | | | | | | | | | | | | | | | | | |
| Jobs End | | | | | | | | | | | | | | | | | | | | | | |
| Jobs Start | | | | | | | | | | | | | | | | | | | | | | |

Example of Soft and Hard Limits



ATLAS JOB

MAXJOB=1,4

CMS JOBS

MAXJOB=3,5



Job Slots (4)

| TRANSITION | BLOCKED | IDLE | RUNNING |
|-------------------|---------|---------------|---------|
| Submit Jobs(1-7) | | 7 6 5 4 3 2 1 | ■ ■ ■ ■ |
| Jobs Start | | | 2 5 4 1 |
| Submit Jobs(8-10) | 9 7 6 3 | | 2 5 4 1 |
| Jobs End | 9 7 6 3 | | 2 5 4 ■ |
| Jobs Start | | 10 9 8 7 6 3 | 2 5 4 8 |
| Jobs End | | | ■ ■ ■ ■ |
| Jobs Start | | | ■ ■ ■ ■ |

Example of Soft and Hard Limits



ATLAS JOB

CMS JOBS

MAXJOB=1,4



MAXJOB=3,5

















































Job Slots (4)

| TRANSITION | BLOCKED | IDLE | | | | | | RUNNING | | | | | | | | | | | | | | | | | | | | | | | |
|-------------------|---------|------|---|---|--|--|--|---------|---|---|---|---|---|---|---|---|----|---|---|----|---|---|---|---|---|---|---|---|---|---|---|
| Submit Jobs(1-7) | | | | | | | | 7 | 6 | 5 | 4 | 3 | 2 | 1 | | | | | | | | | | | | | | | | | |
| Jobs Start | | | | | | | | | | | | | | 7 | 6 | 3 | 2 | 5 | 4 | 1 | | | | | | | | | | | |
| Submit Jobs(8-10) | 9 | 7 | 6 | 3 | | | | | | | | | | | | | 10 | 8 | 2 | 5 | 4 | 1 | | | | | | | | | |
| Jobs End | 9 | 7 | 6 | 3 | | | | | | | | | | | | | 10 | 8 | 2 | 5 | 4 | | | | | | | | | | |
| Jobs Start | | | | | | | | | | | | | | | | | | | | 10 | 9 | 8 | 7 | 6 | 3 | 2 | 5 | 4 | 8 | | |
| Jobs End | | | | | | | | | | | | | | | | | | | | | | | | 9 | 7 | 6 | 3 | | 5 | 4 | 8 |
| Jobs Start | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Example of Soft and Hard Limits

 ATLAS JOB MAXJOB=1,4
 CMS JOBS MAXJOB=3,5

 Job Slots (4)

| TRANSITION | BLOCKED | IDLE | | | | | | RUNNING | | | | | | | | | | | | | | | |
|-------------------|---|---|---|---|--|--|--|---------|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|
| Submit Jobs(1-7) | | | | | | | | 7 | 6 | 5 | 4 | 3 | 2 | 1 |  |  |  |  | | | | | |
| Jobs Start | | | | | | | | | | | | | | |  |  |  |  | | | | | |
| Submit Jobs(8-10) |  9 |  7 |  6 |  3 | | | | | | | | 10 | 8 |  |  |  |  | | | | | | |
| Jobs End |  9 |  7 |  6 |  3 | | | | | | | | 10 | 8 |  |  |  |  | | | | | | |
| Jobs Start | | | | | | | | | | | | 10 | 9 |  |  |  |  |  |  |  |  |  |  |
| Jobs End | | | | | | | | | | | | | | | | | |  |  |  |  | | |
| Jobs Start | | | | | | | | | | | | 10 | 8 |  |  |  |  |  |  |  |  | | |

- Reserve particular resources to a certain type of job.
- Reserve a CPU for a queue, say the short one.

```
SRCFG[sdj] HOSTLIST=grid21.lal.in2p3.fr  
SRCFG[sdj] PERIOD=INFINITY  
SRCFG[sdj] ACCESS=DEDICATED  
SRCFG[sdj] TASKCOUNT=1  
SRCFG[sdj] RESOURCES=PROCS:1  
SRCFG[sdj] CLASSLIST=short
```

- **1 task (slot) is reserved of a size 1 processor.**
- **The reservation can only be accessed using the short queue(class).**
- **ACCESS=DEDICATED** blocks the slot being used by any jobs not in the short list.
- **ACCESS=SHARED** allows res' to be used by others.....?

- **Overlaying Jobs. Running say 4 jobs on a 2 CPU node under certain conditions.**
 - e.g. You may want to run monitoring jobs everywhere on top of existing jobs.
 - e.g. System administrators may want their whole farm to stress test their latest dcache.
- **You must lie in TORQUE first. i.e. np=4 for each node.**
 - Any published information needs fixing afterwards.
- **Set up two reservations on each node for two queues.**

```

SRCFG[ad] HOSTLIST=grid21.lal.in2p3.fr
SRCFG[ad] PERIOD=INFINITY
SRCFG[ad] ACCESS=DEDICATED
SRCFG[ad] TASKCOUNT=2
SRCFG[ad] RESOURCES=PROCS:1
SRCFG[ad] CLASSLIST=ops
    
```

```

SRCFG[lhc] HOSTLIST=grid21.lal.in2p3.fr
SRCFG[lhc] PERIOD=INFINITY
SRCFG[lhc] ACCESS=DEDICATED
SRCFG[lhc] TASKCOUNT=2
SRCFG[lhc] RESOURCES=PROCS:1
SRCFG[lhc] CLASSLIST=atlas,cms
    
```

- **MAUI and TORQUE both have default values.**
- **Many of these may need changing.**
 - Especially for larger farms.

- **RMPOLLINTERVAL** default 60 seconds.
- **JOBAGGREGATIONTIME** default 0 seconds.
 - Define the Maximum and Minimum times between schedule runs.
 - By default since a MAUI run is triggered at every job submission or completion by TORQUE it will run sequentially for large sites.
- **MAXIJOBS** and similar.
 - Tune number of jobs to be considered for scheduling.
 - Can stop to many jobs belonging to lhcb starting at once and melting your storage element.

- **poll_jobs**
 - default was FALSE in previous gLite version but now TRUE.
 - Previously a qstat would contact every node to get it's status every time.
 - When TRUE the pbs_server will poll each node periodically to check there status. qstat will not block as a result.
- **job_stat_rate**
 - default is 30 seconds.
 - This is the TTL for the polled information from batch workers.
 - This value should definitely be increased on large farms.
 - SuperCluster vaguely recommends as much as 5 minutes.
- **Make sure you have nscd or similar.**

- -'ve priorities are by default handled in an “odd” way.
- FairShare components include 1 - (used/target)
 - It is very easy to have a -'ve priority for a job.
- **ENABLENEGJOBPRIORITY** default is **FALSE**
 - With this setting -'ve priorities will be reset to 1.
 - This is not what you want, set it to true.
- **REJECTNEGPRIOJOBS** default is **TRUE**
 - Defines that -'ve priority jobs will never start.
 - This is not what you want, set it to false.

| MAUI config | Default Value | Details |
|------------------------|---------------|----------------------------------|
| SERVICWEIGHT | 1 | Priority Component Weight |
| QUEUETIMEWEIGHT | 1 | Sub-Component Weight |

- **The only Priority component and sub-component that are not disabled by default.**
- **By default queued jobs increase their priority by**
 - $SERVICWEIGHT * QUEUETIMEWEIGHT * \text{minutes queued} *$
- **This is the fifo component.**
 - If you have fairshare configured then you may wish to switch this off. i.e. $SERVICWEIGHT=0$.

- **For LHC computing Fairshare, Throttling and Reservations are possibly enough.**
 - Multi CPU jobs not covered, e.g. Backfill policies are critical for this.
- **Extra Help**
 - MAUI and Torque mailing lists, documentation and bugzilla.
<http://www.clusterresources.com>
 - Purchase MOAB, the commercial version of MAUI.
 - § Contains a lot of extra code and stuff - annoying.
- **Also Interesting.**
 - Early Castor 2 used Maui rather than LSF.
 - LHCb's DIRAC 3 uses Maui/Moab.
 - Globus Workspaces are integrated with torque.
 - § Other XEN or similar Torque integrations floating around.
- **How Big?**
 - RAL is largest in EGEE but mailing list has mentioned 4000 node clusters. (No Reference)