Fair-share

SCITAS machines use the SLURM workload manager in order to schedule users' jobs. In particular, SLURM arbitrates the jobs' queue contention by using a fair-share algorithm in order to prioritize jobs and ensure that the users' usage matches their share as much as possible. In particular, SCITAS clusters use a particular flavor of the fair-share algorithm called fair-tree.

In order to check their priority, the Sshare command is available on any SCITAS cluster. A typical output will be as follow:

```bash
$ sshare
Account  User  Raw Shares  Norm Shares  Raw Usage  Norm Usage  Effectv Usage  FairShare  Level
----------  ------  ------------  -----------  -----------  -----------  -------------  --------  -------
scitas-ge  abourt  1  0.007752  1376  0.000003  0.000005  1468.763590
scitas-ge  clemenco  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  cubuk  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  culpo  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  degiorgi  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  eroche  1  0.043478  344  0.000001  0.250000  0.253333  0.173913
scitas-ge  nvarini  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  qubit  1  0.043478  351  0.000001  0.255072  0.250000  0.170455
scitas-ge  rezzonic  1  0.043478  681  0.000001  0.494928  0.246667  0.087848
scitas-ge  richart  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  rmsilva  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  sue  1  0.043478  0  0.000000  0.000000  0.290000  inf
scitas-ge  topf  1  0.043478  0  0.000000  0.000000  0.290000  inf

The value used to decide the priority of a job is the “Level FS”. The higher the Level FS, the higher the priority. Level FS is the ratio of “Norm Shares” and “Effectv Usage” values, therefore a Level FS of less than 1 represents an overconsumption and more than 1 represents an underconsuming.

In this formula, the "Norm Shares" is the percentage of the cluster which is allocated to the account whereas “Effectv Usage” augments the normalized usage (the users’ raw usage normalized to the total number of cpu-seconds of all jobs run) to account for usage from sibling accounts. Within a group all users have equal weight and so 1 share each.

More informations about SLURM, fair-share and fair-tree can be found here:

https://slurm.schedmd.com/overview.html
https://slurm.schedmd.com/priority_multifactor.html
https://slurm.schedmd.com/fair_tree.html

Related articles

- Licensed Software on the Clusters
- Running Docker images using Shifter
- How to use Tensorflow on the GPU nodes
- Using the clusters
- FAQ