ASSOCIATING NODES WITH ONE OR MORE QUEUES

Hind Fathallah – Technical Support Engineer, Altair / November 25, 2020

Introduction to PBS Professional

Altair® PBS Professional® is a fast, powerful workload manager designed to improve productivity, optimize utilization and efficiency, and simplify administration for HPC clusters, clouds, and supercomputers. It automates job scheduling, management, monitoring, and reporting, and it's the trusted solution for complex Top500 systems and smaller clusters.

Challenges

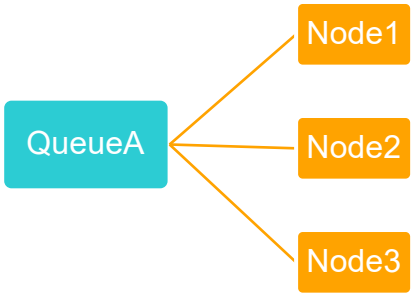
PBS Professional provides the user with an interface to submit jobs to a queue, which will be considered by the scheduler and dispatched to computational resources in a cluster, supercomputer, or cloud. This default configuration has no restrictions on the user, and it allows for maximum utilization of computational resources, which PBS calls vnodes.

As an administrator, you may be asked to allocate or restrict computational resources to a specific user, group, or project. Another common ask is from a user who needs dedicated computational resources for a special project.

PBS Professional provides many ways for an administrator to allocate or restrict computational resources to specific users, groups, or projects. In this document we will describe two common methods of satisfying the requirement by using queues and custom resources.

Method 1: Associate Vnode(s) With One Execution Queue

As an administrator, you can associate one or more vnodes with an execution queue, using the vnode's queue attribute. Using this method, each vnode can be associated with at most one execution queue. Each execution queue can be associated with more than one vnode. If you associate an execution queue and one or more vnodes using this method, any jobs in the queue can run only on the associated vnodes, and the only jobs that can run on the vnodes are the ones in the queue.



One drawback to consider with associating vnode(s) to one execution queue is that when a vnode is idle it cannot execute jobs from other queues, introducing underutilization. This configuration might be suitable for smaller organizations or departments that require users to execute jobs on specific vnodes.

Method 1: Demonstration

As root, create a new queue called queueA and associate a vnode to the newly created queueA. Assuming your vnodes have already been created via qmgr, you will set the node attribute called `queue`. Then the user will submit a job requesting the queue name.

As root, create your new queue, called queueA.

```
qmgr << EOF
create queue queueA
set queue queueA queue_type = Execution
set queue queueA enabled = True
set queue queueA started = True
EOF
```


Set the node attribute called queue to equal queueA.

```
qmgr -c "set node node1 queue=queueA"
```

As a user, submit a job to queueA and see that the job executes on node1.

```
qsub -q queueA jobscript.sh
```

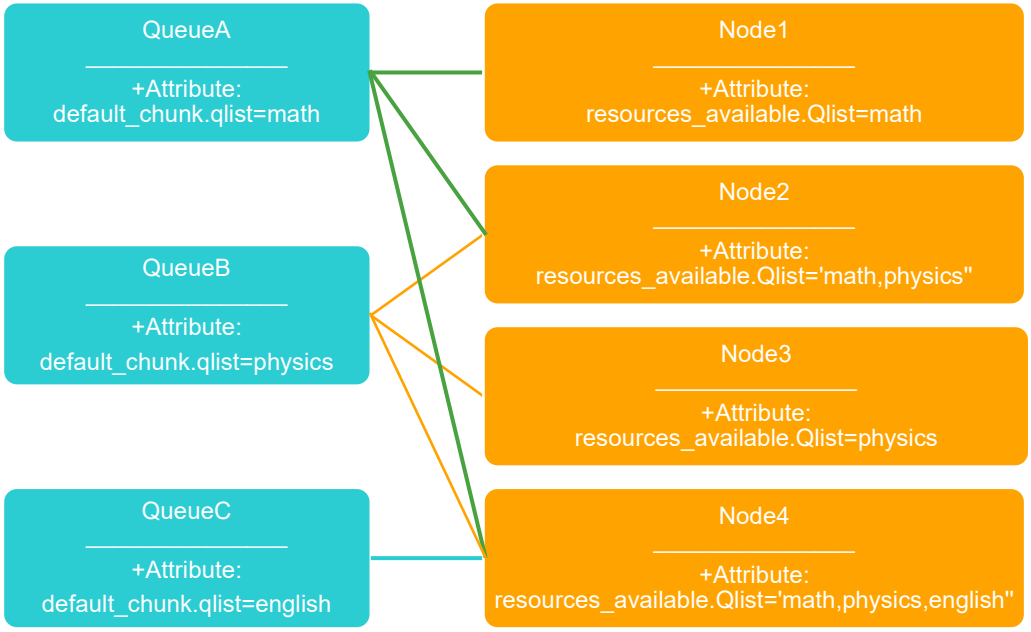
The table below summarizes the expected behavior when submitting 4 jobs, for which jobs 1 and 3 are submitted to the default queue (workq) and jobs 2 and 4 request queue queueA.

Job#	Job Submission	Effective queue qstat -f	node1	Other nodes
1	qsub jobscript.sh	queue=workq		YES
2	qsub -q queueA jobscript.sh	queue=queueA	YES	
3	qsub jobscript.sh	queue=workq		YES
4	qsub -q queueA jobscript.sh	queue=queueA	YES	

Method 2: Associate One or More Vnodes With Multiple Queues

As an administrator, you can use custom host-level resources to associate one or more vnodes with more than one execution queue. The PBS Professional scheduler will use the resources for scheduling just as it does with any resource.

To map a vnode to more than one execution queue, you must define a new host-level string array custom resource. This string array holds a string that has the same value for the queue and vnode you wish to associate. The mechanism of association is that a job that lands in the queue inherits that value for the resource, and then the job can run only on vnodes having a matching value for the resource. You can associate more than one queue with a vnode by setting the resource to the same value at each queue.



Unlike the first method, associating vnode(s) with multiple execution queues will improve node utilization. However, it is possible to idle a vnode if there are not enough jobs in the named queues.

Method 2: Demonstration

As root, you will create a custom string-array resource, update the scheduler's resources parameter, and create three new queues. Assuming your vnodes have already been created via qmgr, you will set the nodes' `resources_available.Qlist` attribute. Then the user will submit a job requesting the queue name.

As root, create your custom string-array host-level resource, called Qlist.

```
qmgr -c "create resource Qlist type=string_array, flag=h"
```

Update the scheduler's resources parameter in the `PBS_HOME/sched_priv/sched_config` to include the new string-array resource Qlist.

```
resources: "ncpus, mem, arch, host, vnode, aoe, eoe, Qlist"
```

Send the HUP signal to the pbs_sched daemon.

```
pkill -HUP pbs_sched
```

Create your new queues with the resources_min, resources_max, and default_chunk attributes.

```
qmgr << EOF
create queue queueA
set queue queueA queue_type = Execution
set queue queueA default_chunk.Qlist = math
set queue queueA enabled = True
set queue queueA started = True
create queue queueB
set queue queueB queue_type = Execution
set queue queueB resources_max.Qlist = physics
set queue queueB enabled = True
set queue queueB started = True
create queue queueC
set queue queueC queue_type = Execution
set queue queueC resources_max.Qlist = english
set queue queueC enabled = True
set queue queueC started = True
EOF
```

Set the resources_available.Qlist attribute for each node.

```
qmgr -c "set node node1 resources_available.Qlist=math"
qmgr -c "set node node2 resources_available.Qlist='math,physics'"
qmgr -c "set node node3 resources_available.Qlist=physics"
qmgr -c "set node node4 resources_available.Qlist='math,physics,english'"
```

The table below summarizes the expected behavior when submitting 3 jobs. There is an effective queue, effective Qlist, and associate nodes.

Job#	Job Submission	Effective queue qstat -f	Effective Qlist qstat -f	node1	node2	node3	node4
1	qsub -q queueA jobscript.sh	queue=queueA	Qlist=math	YES	YES		YES
2	qsub -q queueB jobscript.sh	queue=queueB	Qlist=physics		YES	YES	YES
3	qsub -q queueC jobscript.sh	queue=queueC	Qlist=english				YES

Conclusion

To control this capability and give your users access to the system and to the queue, you should own your own host. Following the basic setup steps, you should be able to quickly associate nodes with one or more queues working.

This is a very small example of what associating nodes with one or more queues can do. You can learn more about PBS Professional queues and queue attributes in the PBS Professional Administrator's Guide.