Moab Cluster Manager®

User Guide

version 6.1

Legal Notices

Copyright

Trademarks

Moab Cluster Manager™ User's Guide

version 6.0

Administrative Topics

List of Tables

Chapter 1.0 Getting Started

1.1. Getting Started Overview

Moab Cluster Manager (MCM) is a Java-based graphical interface for managing the Moab Workload Manager. It allows users to submit jobs, schedule reservations, view job statistics, etc., in an easy, user-friendly way.

This chapter explains how to get started using the Moab Cluster Manager by installing it, connecting it to a Moab Workload Manager, and describing its main window.

1.2. Installation

1.2.1. Unix-Based installation

Moab Cluster Manager is written in Java 1.6. You will need to have Java 1.6 or higher installed on your system to run MCM. This may be an existing copy on your system, or the JRE bundled with the MCM distribution. These install instructions assume a basic familiarity with Unix/Linux file systems, and commands such as ls, tar, mv, etc.

The installation steps are as follows:

1. Download the latest version of the tar file from the Adaptive Computing web site:
    a. The mcm-[version]-[build number].linux.tar.gz comes with a bundled JRE.
    b. The mcm-[version]-[build number].tar.gz does NOT include a JRE.
2. Move the tar file to your home directory, or another directory you have access to (i.e. /home/username/). Unpack the tar file (tar -xzvf xxxx.tar.gz).
3. Change to the newly unpacked MCM directory.
4. You may now start MCM at any time by running the mcm script (i.e. ./mcm). This script will check for the existence of Java and then run MCM.

1.2.2. Windows installation

1. Download the installation executable from the Adaptive Computing web site. The mcm-[version]-[build number].exe is the Windows installer that will setup MCM on your system.
2. Double-click the installation file. The MCM installer will guide you through the installation process.

    Note: The default target folder is "C:\Program Files\Moab Cluster Manager".

3. The installer will create Start Menu and Desktop icons that can be used to run MCM.
4. Double-click the Moab Cluster Manager icon on the Desktop to run MCM.

1.3. Connection Wizard

The Connection Wizard provides four connection options for the Moab Cluster Manager: Remote Connection, Local Connection, Offline Demonstration and Online Demonstration.

1. Remote Connection - Connects to a remote Moab Workload Manager server over SSH.
2. Local Connection - Connects to a locally hosted Moab Workload Manager server.
3. Offline Demonstration - Allows a user to view a demonstration snapshot.
4. Online Demonstration - Automatically logs in to an online demonstration cluster for a preview of the product.

1.3.1. Remote Connection

The Remote Connection feature allows you to securely connect to a remote Moab Workload Manager server. Here is a description of each of the connection options:

- Host Name & Port
  - Host Name (or IP Address) - The Host Name (or IP address) of the server that is running Moab Workload Manager. If you do not know the host name or IP address of the server, please consult your system administrator.
  - Port - The port on which SSH is running on the remote server (the default is 22). If you do not know which port to use, please consult your system administrator.
- Authentication Options
  - Password Authentication - This option tells Moab Cluster Manager to authenticate by prompting the user for a password. Consult with your system administrator for information regarding your user name, password and the type of authentication used.
  - Keyboard Interactive Authentication - This option tells Moab Cluster Manager to interactively prompt for authentication information.
  - SSH Key Authentication - This option tells Moab Cluster Manager to connect to the remote computer using only the user name and a private key file. Consult your system administrator for information regarding your user name, private key and the type of authentication used.
  - Ask for SSH Key Passphrase - Some private keys require a passphrase to be entered before it will allow a user to authenticate. In this is the case this box should be checked, otherwise an empty passphrase will be used for authentication.
- Connection Settings
  - User Name - This is the name used to login to the remote computer. Consult your system administrator for information regarding your user name or password.
  - Path to Moab Workload Manager Client Commands (i.e. showq) - The directory containing the Moab Client Commands, such as showq, mschedctl, mdiag, etc. This is not the location of the Workload Manager, but instead the location of the commands that control the Workload Manager. This location is usually "/usr/local/bin".
  - Private Key Path - If SSH key Authentication is being used, this field is for the path of the private key file.
- Load, save or delete stored sessions

    - Auto Connect On Next Session - This option sets the Moab Cluster Manager to automatically connect to the specified saved session the next time it is run.
    - Saved Sessions - This field is where a user is able to save his/her remote connection settings so that they don't need to be entered each time.
    - Load Button - This button will load the selected saved session in the list to the left.
    - Save Button - This button will save a session according to the name typed in the Saved Sessions field.
    - Delete Button - This button will delete the saved session that is selected in the list to the left.
- Open Button - This button will open a connection to a remote Moab Workload Manager. It will attempt to authenticate the user based on the given settings, and then open the Moab Cluster Manager.
- Cancel Button - Clicking this button will close the Moab Cluster Manager Connection Wizard

1.3.2. Local Connection

This option connects to a Moab Workload Manager running on the local machine. The only option for this mode is the Path for the Moab Workload Manager Client Commands.

- Path to Moab Workload Manager Client Commands (i.e. showq) - The directory containing the Moab Client Commands, such as showq, mschedctl, mdiag, etc. This is not the location of the Moab Workload Manager, but instead the location of the commands that interact with the Moab Workload Manager. This location is usually "/usr/local/bin".
- Open Button - This button will open a connection to a local Moab Workload Manager and then open the Moab Cluster Manager.
- Cancel Button - Clicking this button will close the Moab Cluster Manager Connection Wizard

1.3.3. Offline Demonstration

Moab Cluster Manager is capable of recording all the data gathered from a cluster and saving it to a demonstration snapshot. This connection option will allow a user to view a previously recorded demonstration snapshot. Here is a description of what each of the buttons on this screen do:

- Import Button - The import button allows the user to select a file to copy to the appropriate Moab Cluster Manager directory. This allows users to import demonstration snapshots from other Moab Cluster Managers.
- Delete Button - This button will delete a saved demonstration snapshot.
- Open Button - This button will open the selected Moab Cluster Manager demonstration snapshot.
- Cancel Button - Clicking this button will close the Moab Cluster Manager Connection Wizard

1.3.3. Online Demonstration

The Online Demonstration is a free, online demonstration cluster for users to preview. This connection option will automatically log in to the demonstration cluster.

- Open Button - This button will connect Moab Cluster Manager to the Adaptive Computing demonstration Moab Workload Manager.
- Cancel Button - Clicking this button will close the Moab Cluster Manager Connection Wizard

1.4. View Summary

The Moab Cluster Manager main window provides an overview of the current state of the cluster. There are four parts of this main window: the Main Menu Bar, the Dashboard, the Main Info Screen, and the System Utilization Bar.

1.4.1. Main Menu Bar

The Main Menu Bar is located across the top of the Moab Cluster Manager window. Through this menu (File, Configure, Manage, etc.), all Moab Cluster Manager features can be accessed. The services unique to this menu are:

- Console window
- Save System Snapshot window
- Moab Cluster Manager Preferences window
- Plugin Manager window
- About window

1.4.2. Dashboard

The Dashboard is a directory of all the services that the Moab Cluster Manager can provide to users and administrators. The availability of some services depends on the user's privileges, as determined by the ADMINCFG level defined in the moab.cfg file. The chapters of this User Guide mimic the layout found in the Dashboard.

1.4.3. Main Info Screen

The largest area in the Moab Cluster Manager main window is the Main Info Screen. This screen is intended to give general information about the system that Cluster Manager is currently connected to.

1.4.3.1. Scheduler Information

This panel displays the following information about the scheduler:

- Name - The name of the scheduler. Has no impact on operation.
- Host - This refers to the host computer where the Resource Manager is running.
- Port - The specific port that the scheduler is operating on.
- Mode - The operating mode of the scheduler. Mode options are shown in this table:

NORMAL (default)	Normal operation, controls the Resource Manager as configured
MONITOR	MONITOR mode behaves identical to NORMAL mode except the ability to start, cancel, or modify jobs is disabled. This allows safe diagnosis of the scheduling state and behavior using the various diagnostic client commands.
INTERACTIVE	Like NORMAL mode, except Moab sends the desired change request to the screen and asks for permission to complete it.

- Status - Indicates whether the scheduler is running, down, or paused.

1.4.3.2. Node Summary

This panel displays a high-level view of the state of the nodes found within the cluster. Click on any label to obtain a detailed list of nodes in the given category. Category Descriptions:

- Busy Nodes - *Busy* nodes include all nodes which are actively executing batch jobs. A node will be listed as busy even if it is only partially loaded with jobs.
- Idle Nodes - *Idle* nodes include all nodes which are available but are currently not running any jobs.
- Down Nodes - *Down* nodes include all nodes which have reported major software, hardware, or batch failures or have been marked 'down' or 'offline' by an administrator.
- Total Nodes -The *total* nodes category includes all nodes in the cluster and is a sum of the *busy*, *idle*, and *down* nodes listed above.

1.4.3.3. Job Summary

This panel displays a high-level view of the state of the jobs found within the cluster. Click on any label to obtain a detailed list of jobs in the given category. Category Descriptions:

- Running Jobs - *Running* jobs include all jobs which are actively executing or performing post-execution clean-up. This includes jobs in the states *starting*, *running*, or *exiting*.
- Eligible Jobs - *Eligible* jobs include all jobs which are in state *idle* and are not blocked by holds, usage limits, or other policies. Eligible jobs typically will run as soon as resources become available and may already have a reservation in place.
- Blocked Jobs - *Blocked* jobs include all jobs which cannot run due to reasons other than resource availability. Jobs may be blocked by job holds, resource manager level policies, scheduler job usage policies, job deadline constraints, or other factors. Clicking on the label will bring up the detailed blocked job list which will include additional information in the *blocked reason* column.
- Total Jobs - The *total* jobs category includes all jobs in the cluster and is a sum of the *running*, *eligible*, and *blocked* jobs listed above.

1.4.3.4. User Information

This panel displays the user information of whoever started the Moab Cluster Manager.

- User - The name of the user running Moab Cluster Manager.
- Group - The name of the user's group.
- Account - Any accounts the user may belong to.
- Class - Classes the user has access to.
- QoS - Any QoS accounts the user may belong to.

1.4.3.5. User Job Summary

Displays information concerning jobs run by the current user.

Running Jobs - Jobs run by the current user.
- Eligible Jobs - Users jobs that are waiting in the queue.
- Blocked Jobs - Users jobs that have been blocked, either by policy or user.
- Total Jobs - Total jobs from the user.

1.4.4. System Utilization Bar

The System Utilization bar displays *historical* system utilization as it pertains to utilized processors as captured by Moab profiling intervals. Also note that processors utilized are only measured once a job is finished.

Chapter 2.0 Workload

2.1. Workload Overview

The workload category of features deals with the submitting and viewing of jobs, reservations, and triggers. These functions are used to get work done by the system.

2.2. Jobs

2.2.1. Create Job

A cluster runs programs. A job tells a cluster when, where, and how to run the programs. The create job window, often referred to as a job submission window, is how a user creates a job.

2.2.1.1. Job Creation

Basic

Field	Field Information
Template	If there are job templates that are selectable, the user can do so here. Any attributes associated with the job templates will be mapped onto the submitted job.
Script File	This field allows an administrator to search for a script to run.
Nodes	This allows an administrator to select where the job will run from the following options: • Number of Nodes - A node is a computer consisting of 1 or more processors. A job requires at least 1 processor to execute and therefore must use at least 1 node. If this field is set to 0, Workload Manager assigns the job to 1 node unless the *Node List* field is populated. • Node List - A node is a computer consisting of 1 or more processors. This field allows the user to define which nodes a job requires to execute. If a node list is not specified the nodes needed for the job are gathered from the nodes field. • VM List - The specific virtual machines the job will use.
Duration	The amount of time the job will run.

Job Information

Field	Field Information
Job Name	A user can attach a custom name to a job to assist him/her in identifying the job. The name is provided only for the user's convenience and does not affect any policies or settings.
User Job Priority	The higher a job's priority, the sooner it will start. By changing this field, a user can reduce the job's priority and change the order in which the job starts. This field is usually utilized to execute a user's jobs in a specific order. This field only supports negative numbers with the exception of 0. A 0 User Job Priority will not delay the job from starting. However, the higher a negative number, the lower a job's priority. For example, a job with a User Job Priority of -100 will allow more jobs to start before it starts, thus postponing the job from starting.

Resources

Field	Field Information
Processors per Node	All jobs require at least 1 processor. If this field is not used, the processors are calculated using the available processors on a node. If a processor and a node are not requested, Workload Manager assigns 1 node to the job.
Memory per Node	Some jobs require a specific amount of memory. This field allows a job to request the memory it needs for each node. It should be noted that this field is not the total memory across the entire cluster but only the memory on each node needed by the job. Workload Manager will start this job only on the nodes that have sufficient memory. If this field is not used, Workload Manager will start the job on any available node.
Swap per Node	Some jobs require a specific amount of swap. This field allows a job to request the swap it needs for each node. It should be noted that this field is not the total swap across the entire cluster, but the swap on each node needed by the job. Workload Manager will start this job only on the nodes that have sufficient swap. If this field is not used, Workload Manager will start the job on any available node.
Operating System	If an operating system is selected, Moab will try to run the job on any nodes with the specified operating system.
Architecture	If an architecture is selected, Moab will try to run the job on any nodes with the specified architecture.
Required Node Features	Some jobs require a specific feature on a node. A feature is a custom tag attached to a specific list of nodes. Consult your system administrator for specific information regarding each tag.
Partition	Clusters are often divided into different sections. These sections are commonly called partitions. Users can only request one specific partition for their job. Consult your system administrator to learn which partition is the best suited for your job.
Reservation	If a reservation is specified, the job will execute only on the nodes of the selected reservation.

Data Management

Field	Field Information
Execution Directory	Some jobs need to be executed in a specific location on each node. This field allows a user the ability to define that location. By default, the job is executed in the user's home directory. Consult your system administrator for information regarding your home directory.
Output Directory	All scripts, executables, programs, and applications use an output directory. This field gives the user the ability to customize the location of the output directory.
Error Directory	All scripts, executables, programs, applications use an error directory. This field gives the user the ability to customize the location of the error directory.

Credential Information

Field	Field Information
User	This field defines the name of the user under whom the job will execute. Only users with Admin1 rights can change this field. See User Access for more information.
Group	This field defines the name of the group under which the job will execute. Only the groups available to the user are displayed.
Account	This field defines the name of the account under which the job will execute. Only the accounts available to the user are displayed.
Class	This field defines the name of the class under which this job will execute. Only the classes available to the user are displayed.
Quality of Service (QoS)	This field defines the name of the quality of service (QoS) under which this job will execute. Only the QoS's available to the user are displayed.

Job Dependencies

Field	Field Information
Job Dependencies	This field allows an administrator to specify which other jobs this job is dependent on. The dependency types are as follows: • After - The job starts any time after specified jobs start executing. • Afterok - The job may start any time after all specified jobs have completed successfully. • Afternotok - The job may start any time after all specified jobs have completed unsuccessfully. • Afterany - The job may start at any time after all specified jobs have completed regardless of status.

Job Flags

Field	Field Information
Preemptible	A job that is preemptible can be suspended or re-queued by higher priority jobs.
Restartable	If a job experiences a failure during execution, the user must resubmit the job to Workload Manager. However, a job that is restartable is automatically restarted by Workload Manager in the event of a failure.
Dynamic	The job grows and shrinks dynamically based on job requests.
Preemptor	The job may preempt other jobs which have the **PREEMPTEE** flag.
Hold	A hold can only be placed upon jobs that have not yet executed. A hold stops or halts a job from running until the user or an administrator releases the hold.

Time Frame

Field	Field Information
Start Time	Some jobs require a specific amount of time before they are allowed to start. This field allows the user the ability to define the earliest time that the job can begin. By default, a job may start as soon as resources become available.
Estimated Start Time Calculator	This button displays Moab's text-based output that determines when a job can start.
Estimated Start Time Table	This button displays a table of the estimated start times for jobs of different processor sizes.

Email Notification

Field	Field Information
On Job Completion	When a job finishes execution an email notification will be sent to the user stating this.
On Job Execution	When a job begins execution an email notification will be sent to the user stating this.
On Job Failure	When a job cannot start or crashes during execution an email notification will be sent to the user stating this.

User email addresses may be specified in the 'Create/Manage Users' page.

2.2.2. Modify a Job

Summary:

A cluster runs programs. A job tells a cluster when, where, and how to run the programs. The Modify Job window allows a user to modify an existing job. The fields that can be modified are user job priority, system priority, and duration. If multiple jobs are selected, then QoS can also be modified.

Overview

Field	Field Information
ID	This field displays the unique ID given to the job by Workload Manager.
Name	Users can attach a custom name to the job to allow them to easily identify their jobs. The name does not change any Workload Manager settings or prioritizations. If a name has been attached, it will appear in this field.

State	This field displays the execution status of the job. For example, running, stopped, executing, idle, blocked, etc.
Hold	This field displays whether the job has a hold. A hold can only be placed upon jobs that have not started. A hold prevents a job from running until the user or an administrator releases the hold.
Messages	This field displays informational messages relating to the job.
Blocked Reason / Error	This field displays diagnostic messages related to the job.

Credentials

Field	Field Information
User	This field displays the user ID under which the job executes.
Group	This field displays the group ID under which the job executes.
Account	This field displays the account ID used by the job.
Class	This field displays the class/queue ID used by the job.
Quality of Service (QoS)	This field displays the quality of service (QoS) ID used by the job.

Resources

Field	Field Information
Node List	A node is a computer consisting of 1 or more processors. A job requires at least 1 processor to execute and therefore must use at least 1 node. The node list is a list of the nodes that the job is using.
Required Features	Some jobs require a specific feature on a node. A feature is a custom tag attached to a specific list of nodes. This field allows a user to view the required feature for the job. Consult your system administrator for specific information regarding each tag.
Processor per Task	A *task* is a group of resources that must all be on the same node. One resource in that group is a processor. This field displays the number of processors in each task that the user's job requires.
Operating System	Some jobs require a specific operating system. This field allows a user to view the operating system required by this job.
Required Minimum Task Count	A task is a group of resources that must all be on the same node. This field displays the minimum number of tasks, or groups of resources, the user's job requires on each node.

Job Attributes

Field	Field Information

Flags	Cluster Manager schedules jobs differently according to their flags. Possible flags are hold, interactive, restartable, and preemptible.
Bypassed in Queue	This field displays the number of times another job of a lower priority started before this job.
Partition Access List	This field displays the partitions available for this user. Clusters can be divided into different sections commonly called partitions. Consult your system administrator to learn which partition is the best suited for your job.
Start Priority	This field displays the start priority for the job. With some exceptions, the higher a job's priority over other jobs, the sooner it will begin to execute.
User Job Priority	With a few exceptions, a job with a high priority will begin sooner depending on how much greater its priority is to that of other jobs. A user has the ability to reduce the job's priority and, in effect, delay the job's start time by changing this field. This option is usually utilized by users who desire their jobs to execute in a specific order. This field only supports negative numbers with the exception of 0. A 0 user job priority will not delay the job from starting. However, a job's priority will decrease as the priority number decreases. For example, a user job delay priority of -100 will allow more jobs to start before it than a job with a user job priority of -10. It should be noted that the user job priority literally lowers the start priority of a job.
System Priority	With some exceptions, the higher a job's priority over other jobs, the sooner it will begin to execute. This field allows an administrator the ability to start jobs immediately. This field adds 1,000,000,000 and the administrator priority to the start priority of the job.

Time Frame

Field	Field Information
Duration	This field displays the job's estimated execution time. If a job requires more time than the specified duration, duration violation policies come into effect. Consult your system administrator for more information regarding these policies. If no duration is specified, a default walltime is applied. Consult your system administrator for more information regarding your cluster's default walltime.
Queue Time	This field displays the amount of time the job has been waiting to start.

2.2.3. List Jobs/Job Templates

Default Categories

Summary:Job ID, Job Name, State, User, Used Wallclock, Allotted Wallclock, Nodes, Procs

Credentials: Job ID, State, User, Group, Class, Account, QoS

Time: Job ID, State, Start Time, Used Wallclock, Completion Time, Submission Time, System Minimal Start Time, Wallclock, Earliest Start Time, Latest Completion Time

Node Information: Job ID, State, Allocated Node List, Master Node, Node List, Executed Node List

IDs: Job ID, State, Global Job ID, System ID, Step ID, RM Job ID

Required Resources: Job ID, State, Allocated Node List, Partitions, Allocated Nodes, Required Procs, Req. Node Feature, Req. Node Memory

Utilized Resources: Job ID, State, Memory-Seconds Utilized, Processors-Seconds Dedicated, Processor-Seconds Utilized, Utilized Memory, Utilized Processors

Reservation: Job ID, State, Reservation

Resource Manager: Job ID, State, RM Job ID

Executable: Job ID, State, Input File, Executable, Arguments, Initial Working Directory, Executable Type

Priority: Job ID, State, Run Priority, System Priority, User Priority, Start Priority

Diagnostic: Job ID, State, Suspend Duration, Hold, Blocked Reason, Expected State, Bypass

Comments: Job ID, State, Messages

List Job Fields

Field	Field Information
Account	This field displays the account id used by the job.
Allocated Node List	A node is a computer consisting of 1 or more processors. A job requires at least 1 processor to execute and therefore must use at least 1 node. The allocated node list is a list of the nodes that the job is using.
Allocated Nodes	This field displays the number of node allocated to the job.
Arguments	Some programs provide users with options. This field allows the user the ability to view those options. A user should consult the program documentation to learn about the available options.
Block Reason	This field displays diagnostic messages related to the job.
Bypass	This field displays the number of times another job of a lower priority started before this job.
Class	This field displays the class id used by the job.
Completion Time	This field displays the date and time the job finished execution.
Earliest Start Time	This field displays the user-specified date and time the job is available to start. Workload Manager will not start the job until after this specified date and time.
Execution Eligibility Time	This field displays the time the job was eligible for execution. The format is hours:minutes:seconds.
Expected	This field displays the estimated execution status of the job. For example,

State	running, stopped, executing, idle, blocked, etc.
Flags	Cluster Manager schedules jobs differently according to their flags. Possible flags are hold, interactive, restartable, and preemptible. Refer to the create job documentation for definitions of the flags.
Generic Attributes	This field displays a custom attribute attached to the job.
Global Job ID	The *global job id* is used when multiple resource managers are being used.
Group	This field displays the group id under which the job executes.
Hold	This field displays whether the job has a hold. A hold can only be placed on jobs that have not started. A hold stops or halts a job from running until the user or an administrator releases the hold.
Initial Working Directory	Some jobs must execute in a specific location on each node. This field allows a user the ability to define that location. By default, the job is executed in the user's home directory. Consult your system administrator for information regarding your home directory.
Is Template Selectable?	This field displays whether the job was created from a selectable template.
Is Template?	This field displays whether the job was created from a template.
Job ID	This field displays the job's unique ID given by Workload Manager.
Job Name	A user can attach a custom name to the job to allow him/her to easily identify their job. The name does not change any Workload Manager settings or prioritizations. If a name has been attached it will appear in this field.
Latest Completion Time	This field displays the date and time the job must finish execution.
Master Node	In a cluster, one specific node is in charge of communication with all the other nodes on the cluster. This node is often referred to as the master node or the head node. This field will display the name of the master node.
Memory-Seconds Utilized	Memory seconds utilized is defined as the total amount of memory used by the job times the number of seconds the memory was used. The value is calculated as a sum total of all the memory on the cluster and not on a per node basis.
Messages	This field allows the user to add a comment to the job.
Node Features	This field displays whether the job prefers a node with a specified feature.
Node List	This field displays the list of nodes that the job requires to execute.
Partition	Clusters are often divided into different sections. These sections are commonly called partitions. Users can only request one specific partition for the job. Consult your system administrator to learn which partition is the best suited for your job.
Processor-	Processor seconds dedicated is the total number of processors reserved by

Seconds Dedicated	Workload Manager for the job times the amount of time, in seconds, that the processors were reserved. The value is calculated as a sum total of all the processors on the cluster and not on a per node basis.
Processor-Seconds Utilized	Processor seconds utilized is the total number of processors used by the job times the number of seconds the processors were reserved. The value is calculated as a sum total of all the processors on the cluster and not on a per node basis.
Quality of Service (QoS)	This field will display the quality of service (QoS) id used by the job.
Quality of Service (QoS) Required	This field displays the required QoS for this job.
Required Access	This field displays the required access for the job.
Required Allocated Node List	A node is a computer consisting of 1 or more processors. A job requires at least 1 processor to execute and therefore must use at least 1 node. The allocated node list is a list of the nodes that the job is using.
Required Allocated Partition	Clusters are often divided into different sections. These sections are called partitions. Users can only request one specific partition for each job. Consult your system administrator to learn which partition is the best suited for your job.
Required Architecture	Some jobs require a specific node architecture. This field allows a user to view the architecture required by this job.
Required Class	This field displays the required class for this job.
Required Disk per Task	A task is a group of resources that must all be on the same node. One resource in that group is disk space. This field displays the amount of disk in each task that the job requires.
Required Memory Per Task	A task is a group of resources that must all be on the same node. One resource in that group is memory. This field displays the amount of memory in each task that the job requires.
Required Minimum Nodes	A node is a computer consisting of 1 or more processors. This field displays the minimum number of nodes required for the job to execute.
Required Network	This field displays the network required by this job.
Required Node Disk	This field displays the required amount of disk space the job needs on each node. This field is not the total disk across the entire cluster but only the disk space on each node.
Required Node Memory	This field displays the requested amount of memory it needs for each node. This field is not the total memory across the entire cluster but only the memory on each node.
Required Node	This field displays the processors required by this job.

Processors	
Required Node Swap	This field displays the required swap space the job needs for each node. It should be noted that this field is not the total swap across the entire cluster but only the swap on each node.
Required Operating System	This field allows a user to view the operating system required by this job.
Required Partition	Clusters are often divided into different sections. These sections are called partitions. Users can only request one specific partition for each job. Consult your system administrator to learn which partition is the best suited for your job.
Required Processors Per Task	A task is a group of resources that must all be on the same node. One resource in that group is a processor. This field displays the number of processors in each task that the job requires.
Required Set	This field displays the groupings of nodes this job requires.
Required Swap Per Task	A task is a group of resources that must all be on the same node. One resource in that group is swap space. This field displays the amount of swap in each task that the job requires.
Required Tasks Per Nodes	A task is a group of resources that must all be on the same node. This field displays the number of tasks that the job requires on each node.
Required Memory	This field allows a job to request the memory it needs for each node. This field is not the total memory across the entire cluster but only the memory on each node needed by the job. Workload Manager will start this job only on the nodes that have sufficient memory.
Required Nodes	A node is a computer consisting of 1 or more processors. A job requires at least 1 processor to execute and therefore must use at least 1 node.
Required Processors	This field displays the number of processors used by this job.
Reservation	A user can specify a reservation for this job. If a reservation is specified the job will execute only on that reservation's nodes.
Run Priority	This field is used by jobs that are preemptible to decide which job should be preempted. With a few exceptions, the higher a job's priority, the more likely it will be preempted.
Script File	This field displays which script file the job executes.
Start Count	This field displays the number of times the job has attempted to start executing.
Start Priority	This field displays the start priority of the job. With a few exceptions, a job with a high priority will begin sooner depending on how much greater its priority is to other jobs.
Start Time	This field displays the date and time the job started.
State	This field displays the execution status of the job. For example, running, stopped, executing, idle, blocked, etc.

Submission Time	This field displays the time the job was first created. The format is hours:minutes:seconds. If the exact date is desired moving the mouse over the value will display the exact date of the submission time.
Suspend Duration	This field displays the time the job was in a suspended state. The format is hours:minutes:seconds.
System ID	The *system job id* is used when multiple resource managers are being used.
System Priority	With a few exceptions, a job with a high priority will begin sooner depending on how much greater its priority is to other jobs. This field allows an administrator the ability to start jobs immediately. This field adds 1,000,000,000 and the administrator priority to the job's priority.
System Start Time	This field displays the time when the job started.
Used Wallclock	This field displays the actual execution time of the job. The format is hours:minutes:seconds. The white space indicates the remaining time left before the job reaches its requested Wallclock time. The colored section indicates the amount of Wallclock that has been used. A red bar indicates that the job has violated it Wallclock limit. Refer to the *Wallclock* field for the job Wallclock.
User	This field will display the user id under which the job is executing.
User Job Priority	With a few exceptions, a job with a high priority will begin sooner depending on how much greater its priority is to other jobs. A user has the ability to reduce the job's priority and, in effect, delay its start time by changing this field. This option is usually utilized by users who desire their jobs to execute in a specific order. This field only supports negative numbers with the exception of 0. A 0 user job priority will not delay the job from starting. However, a job's priority will decrease as the priority number decreases. For example, a user job delay priority of -100 will allow more jobs to start before it starts than a user job priority of -10.
Wallclock	This field displays the estimated amount of time that the job will execute. If a job requires more time than the specified duration, duration violation policies come into effect. Consult your system administrator for more information regarding these policies. If no duration is specified, a default walltime is applied. Consult your system administrator for more information regarding your cluster's default wall time.

2.2.4. Job Timeline

The Job Timeline window displays each executing job. On the left side of the Now line is the amount of a job that has completed. The right side is the remaining execution time.

2.2.5. Job Outlines

Job outlines are settings saved from the create/submit job window. This window allows job outlines to be saved, deleted, or opened either locally on the machine that Moab Cluster Manager is running on, or remotely on the machine that Moab Workload Manager is running on.

Outline Information

This section displays information of the currently loaded job outline.

Field	Field Information
Outline Name	This field allows a user to change the saved file name of the outline.
Name	This field displays the custom name for the job outline.
Script	This field displays the script that is used in the outline.

2.2.5.1. Outline Windows

Local Job Outlines - Outlines found on the machine running Moab Cluster Manager.

Personal Remote Outlines - Outlines found in a personal directory on the machine running Moab Workload Manager.

Shared Remote Outlines - Outlines found in a shared remote directory on the machine running Moab Workload Manager.

Outline Directories - Directories where outlines can be found.

Outline Operations

Field	Field Information
Load	This button retrieves the selected outline information and place it in the outline information fields
Delete	This button deletes the selected outline.

2.2.6. Job Workflows

A job workflow is a visual representation of the dependency relationships between jobs. The workflow shows which jobs depend on others. Since a job can depend on any number of jobs and vice versa, the workflow can show the whole dependency relationship for any given job.

To view a job workflow, select a job from the input options at the bottom of the page. The system then determines all jobs that depend or are dependent on this target job. The results are displayed graphically in a timeline window of the main page. Jobs are placed on the graph according to their start and end times.

On the graph, time spans from left (earliest time) to right (latest time). The time window is dependent on the workflow, as the earliest time shown is the earliest job's start time and the end time is it latest job's end time. If the current time is included in the time window, a vertical bar on the graph displays the current time. Keep in mind however that some workflows are entirely in the future and others are only in the past, in which case no vertical bar will be shown.

A dependency is represented as an arrow on the graph. Jobs that are dependent on others will have an arrow pointing to it representing the expected flow of the jobs, or in other words the order in which the jobs will run. There are various dependency types and therefore various job states that must be represented. Double clicking on any job will bring up more details about the job in the modify job window.

2.2.7. Dynamic Job Allocation

Dynamic job allocation allows a user with mjobctl privileges to manually allocate or deallocate nodes for a dynamic job. This allows complete control over how many nodes a dynamic job has. Keep in mind that if performance metric ranges are specified (such as TARGETLOAD, TARGETBACKLOG, etc.) that the dynamic job may reallocate or deallocate nodes that were just modified to meet their metrics.

If this window is accessed from a node-based window (such as "Visual Cluster" or "List/Modify Nodes"), the node list will be locked down. The jobs will be filtered s0 that they make sense in the add or release action. Similarly, if this window is accessed from a job based window (such as "List/Modify Jobs"), the job list will be locked down. The nodes will be filtered so that they make sense in the add or release action.

2.3. Reservations

2.3.1. Create Reservation

Summary:

A *reservation* sets apart resources during a particular time frame for a particular owner. Reservations usually reserve resources such as nodes or processors on the cluster. The Create Reservation window allows you to define what resources a reservation requires, as well as the time frame for the reservation.

In addition to being able to create a basic reservation, this window also allows you to create a recurring reservation. A *recurring reservation*, also referred to as a standing reservation or a reservation generator/creator, provides the user with the option of having reservations automatically created according to a desired time frame. For example, if a user wants to create a reservation every Tuesday and Thursday starting at 11 am and ending at 4 pm, a recurring reservation would fulfill this need. It should be noted that a recurring reservation could potentially not be able to create a reservation if the resources are already dedicated to another reservation or job. To reduce the possibility of this occurring, the day/week depth field can be increased so Workload Manager will create the reservations further in advance.

Panels Containing Required Parameters:

Reservation Information

This panel allows the user to specify the reservation name and owner.

Field	Additional Information
Reservation Name	This field allows a user to create a name for the reservation. Workload Manager appends a numerical value to the end of the reservation name which allows users the ability to enter duplicate reservation names without affecting a previous reservation.
Owner	An *owner* is a user, group, account, class, or quality of service. A reservation can reserve only the resources that the owner can access. This field allows a user to select the owner of the reservation.

Access Control List

This panel allows the user to specify what credentials have permission to access the reservation. At least one credential is required to be in the Access Control List for a reservation, otherwise it would not be very useful. The user may select from 5 different types of credentials - users, groups, accounts, classes, and quality-of-services. Any of the credentials in the Access Control List have permission to use the reservation.

Button	Additional Information
Add	Pops up a window that allows the user to select credentials of a specific type to add to the Access Control List

Clear	Removes all credentials from the Access Control List

Resources

Allows the user to specify what resources will be.

Field	Additional Information
Host List / Host Expression	Users often require specific nodes for their applications. A list of nodes required by the user is called a *host list*. If a host list is not specified, the task count must be specified.
Task Count	A *task* is a group of resources that must all be on the same node. This field defines how many groups of resources will be required to create this reservation. If the task count is not specified the host list must be specified.
Tasks	A *task* is a group of resources that must all be on the same node. This button displays a window allowing a user to define what those resources will be in a task. • Memory Per Task - This field provides the user with the option of requesting the amount of memory, in megabytes, a job needs on a node. • Swap Per Task - This field provides the user with the option of requesting the amount of swap, in megabytes, a job needs on a node. • Disk Per Task - This field provides the user with the option of requesting the amount of disk space, in megabytes, a job needs on a node. • Processors Per Task - This field provides the user with the option of requesting the number of processors a job needs on a node.

Reservation Time Frame Tabs

Once

Basic reservation will be created for the desired start and end time.

Field	Additional Information
Start Time	This field allows the user to select the day and time when the reservation begins.
End Time	This field allows the user to select the day and time when the reservation ends.

Daily Recurring Reservation

A recurring reservation is initialized to automatically create a reservation on the desired days at the desired start and end time. The recurring reservation continually generates new reservations. The *day depth* is used to decide when a reservation is created. For example, if a reservation starts 4 days from now and the *day depth* is set to 2 days the reservation will not be created for 2 more days.

Field	Additional Information
Days	This field allows the user to select the days of the week a reservation is created.
Start Time	This field allows the user to select the time of day the reservation begins.
End Time	This field allows the user to select the time of day the reservation ends.
Day Depth	This field allows the user to specify how many days in advance Workload Manager should create recurring reservations.

Weekly Recurring Reservation

A recurring reservation will be initialized to automatically create a reservation for the week starting from the desired start day and ending on the desired end day of the week. The recurring reservation continually generates new reservations. The *week depth* is used to decide when a reservation is created. For example, if a reservation starts 4 weeks from now and the *week depth* is set to 2 weeks the reservation will not be created for 2 more weeks.

Field	Additional Information
Start Day	This field allows the user to select the day and time in the week the reservation begins.
End Day	This field allows the user to select the day and time in the week the reservation ends.
Week Depth	This field allows the user to specify how many weeks in advance Workload Manager should create recurring reservations.

Infinite Reservation - A reservation will be created that will continue indefinitely. No start or end time is required.

Panels Containing Advanced Options:

Option Tabs

Single Reservation Options

Field	Additional Information
Exclusive	The exclusive option allows only this reservation, and no other reservation, access to the requested resources.

Recurring Reservation Options

Field	Additional Information
Single	The single use option allows only one job to run in this reservation. Once that

Use	job has finished execution, the basic reservation, not the recurring reservation, will be canceled.
By Name	Only jobs that request this reservation are allowed to execute within it.
Owner Preempt	This option allows jobs that are running inside of this reservation to be preempted by jobs owned by the same owner as this reservation.
Space Flex	The *space flex* option gives Workload Manager permission to alter the number of requested resources for this reservation.
Time Flex	The *time flex* option gives Workload Manager permission to alter the time frame for this reservation. The space flex option must be enabled if time flex is desired.

Event Triggers

Button	Additional Information
Create Trigger 1-6	This field allows the user the ability to attach triggers to a reservation.

Grid Sandboxing

Field	Additional Information
Allow grid sandboxing	By default, this allows only the resources in the recurring reservation to be visible to grid peers.
Cluster List	List of clusters that have access to the grid sandbox.

Misc. Options

Field	Additional Information
Partition	Clusters can be divided into different sections. These sections are called *partitions*. Users can only request one specific partition for each reservation. Consult with your system administrator to learn which partition is best suited for your reservation.
Node Features	Some jobs require a specific feature on a node. A *node feature* is a custom tag attached to a specific list of nodes. Consult your system administrator for specific information regarding each tag.

2.3.2. Modify Reservation

Summary

A *reservation* is a time frame on the cluster reserved for a particular need. Reservations usually reserve resources such as nodes or processors on the cluster. The Modify Reservation window allows you to view and modify existing reservations.

Basic Information

Field	Field Information
Name	This field allows a user to create a name for the reservation. Workload Manager appends a numerical value to the end of the reservation name allowing users the ability to enter duplicate reservation names without affecting any other reservation.
Owner	An *owner* is a user, group, account, class, or quality of service. A reservation can reserve only the resources that the owner has access to. This field displays the owner of the reservation. If the reservation is a job reservation, this field will be blank.
Messages	This field allows a user the option of adding a message or comment to a reservation.
Type	This field displays whether a reservation is a user or job reservation.
Sub Type	This is the type of reservation. Some examples of the available types are grid, standing reservation, user, maintenance, etc.
ACL	This field displays who is allowed to access resources on the reservation.
Statistics	This field displays the percentage of processors seconds reserved by the reservation that were used by a job or multiple jobs.

Credentials

Field	Field Information
User	This field displays the user ID used by the reservation. If this field is empty no user can directly access this reservation.
Group	This field displays the group ID used by the reservation. If this field is empty no group can directly access this reservation.
Account	This field displays the account ID used by the reservation. If this field is empty no account can directly access this reservation.
Class	This field displays the class ID used by the reservation. If this field is empty no class can directly access this reservation.
Quality of Service (QoS)	This field displays the quality of service (QoS) ID used by the reservation. If this field is empty no quality of service can directly access this reservation.

Cluster Information

Field	Field Information
Flags	Cluster Manager schedules reservations differently according to their flags. This field displays the reservation flags.
Node Expression	This field displays a list of nodes required by the job to execute. The list of nodes is a regular expression. A *node* is a computer consisting of 1 or more processors.

Partition	This field displays the partition on which the reservation will be configured.
Required Node Count	This field displays the number of nodes required by the reservation. A *node* is a computer consisting of 1 or more processors.
Required Node List	This field displays a list of nodes required by the reservation. A *node* is a computer consisting of 1 or more processors.
Required Task Count	This field displays the number of processors required by the reservation.
Resources	This field displays what type of resource is reserved by the reservation.

Time Frame

Field	Field Information
Start Time	Some jobs require a specific amount of time before they are allowed to start. This field allows the user to define the earliest time the job can start. By default, a job may start as soon as resources become available.
End Time	This field displays the time at which the reservation will finish.
Duration	The *duration* is the job's estimated execution time. If a job requires more time than the specified duration, duration violation policies come into effect. Consult your system administrator for more information regarding these policies. If no duration is specified, a default walltime is applied. Consult your system administrator for more information regarding your cluster's default walltime.

Node List

Field	Field Information
Nodes	Each button displayed represents a node that the reservation has reserved. When the button is selected the view/modify node window will appear containing information about the node.

2.3.3. List Reservations

Summary

A *reservation* is a time frame on the cluster reserved for a particular needed. Reservations usually reserve resources such as nodes or processors on the cluster. Reservations are created either by a user or by a job. A reservation created by a user is called a *user reservation* while a reservation created by a job is called a *job reservation*. All executing jobs have reservations.

List Reservation Fields

Field	Field Information
ACL	This field displays who is allowed to access resources on the reservation.
Account	This field displays which accounts can access this reservation. If this field is blank, no accounts have been given access to the reservation.
Accounts in ACL	This field displays which accounts can access resources on the reservation.
Allocated Nodes	A *node* is a computer consisting of 1 or more processors. The allocated node list is a list of the nodes that the reservation is using.
Architecture	This field displays the hardware architecture of the reservation. The exact hardware information displayed will depend upon the information the resource manager supplies to Workload Manager.
Class	This field displays which classes can access this reservation. If this field is blank, no classes have been given access to the reservation.
Classes in ACL	This field displays which classes are allowed to access resources on the reservation.
Duration	The duration is the job's estimated execution time. The format used is days:hours:minutes:seconds.
End Time	If the reservation will end in less than 12 hours, the value is displayed in the format of hours:minutes:seconds, where a negative value indicates that the reservation ended that many hours:minutes:seconds ago. A positive value indicates that the reservation will end in that many hours:minutes:seconds. Resting the mouse over the value displays the exact date that the reservation ended or will end. An end time that is years in the future often indicates that the reservation was created without any end time specified and Workload Manager inserted a default end time.
Flags	Cluster Manager schedules reservations differently according to their flags. This field displays the reservation flags.
Global ID	This field only displays information when multiple resource managers are present.
Group	This field displays which groups can access the reservation. If this field is blank, no groups have been given access to the reservation.
Groups in ACL	This field displays which groups can access resources on the reservation.
Max Tasks	This field displays the maximum number of processors a reservation can use.
Memory	This field displays how much memory the reservation requires per task.
Messages	This field allows users the option of adding a message or comment to a reservation.
Name	This field allows users to create a name for the reservation. Workload Manager appends a numerical value to the end of the Reservation ID allowing users the ability to enter duplicate Reservation IDs without affecting any other reservation.
Node Expression	A *node* is a computer consisting of 1 or more processors. This field displays a list of nodes required by the job to execute. The list of nodes is a regular

	expression.
Node Set Policy	This field displays the policy that the reservation will use to select the nodes.
OS	This field displays the operating systems the job requires.
Owner	This field displays the owner of the reservation. A reservation can reserve only the resources that the owner has access to. An *owner* is a user, group, account, class, or quality of service.
Partition	Clusters can be divided into different sections. These sections are called *partitions*. Users can only request one specific partition for each reservation. Consult your system administrator to learn which partition is the best suited for your reservation.
QoS	This field displays which quality of service (QoS) will be able to access this reservation. If this field is blank, no qualities of service (QoS) have been given access to the reservation.
QoS in ACL	This field displays which qualities of service can access resources on the reservation.
Required Feature List	A feature is a custom attribute attached to a node. This field displays the node features required for a node to be used for the reservation.
Required Feature Policy	This field displays the policy that the reservation will use to select the features.
Required Node Count	A *node* is a computer consisting of 1 or more processors. This field displays the number of nodes required by the reservation.
Required Node List	A *node* is a computer consisting of 1 or more processors. This field displays a list of nodes required by the reservation.
Required Task Count	This field displays the number of processors required by the reservation.
Resources	This field displays what type of resource is reserved by the reservation.
Spec Name	This field displays information for multiple resource managers.
Start Time	If the reservation will start in less than 12 hours, the value is displayed in the format of hours:minutes:seconds, where a negative value indicates that the reservation will start in that many hours:minutes:seconds. A positive value indicates that the reservation started that many hours:minutes:seconds ago. Resting the mouse over the value will display the exact date that the reservation started or will start. The colored bar shows the percentage of the reservation that has completed. The white space indicates the remaining reservation time.
Statistics	This field displays statistical information relating to the reservation.
Sub Type	This displays the type of reservation. Some examples of the available types are grid, standing reservation, user, maintenance, etc.
Task Count	A *task* is a group of resources that must all be on the same node. This field displays how many groups of resources will be required to create this reservation.
Timeline	This displays the reservation time lines. The green bar indicates the used

	amount of the reservation while the blue bar indicates the remaining amount of the reservation. The display options on the left side allow a user change how much of the time line is displayed.
Trigger	This field displays information about any trigger that is attached to the reservation.
Type	This field displays whether a reservation is a user or job reservation.
User	This field displays which users can access this reservation. If this field is blank, no users have been given access to the reservation.

Default Category Settings

Summary: Name, Type, User, Group, Account, Class, Quality of Service (QoS), Start Time, End Time, Duration

Credentials: Name, Type, User, Group, Account, Class, Quality of Service (QoS) *Time:* Name, Type, Start Time, End Time, Duration *Resources:* Name, Type, Partition, Resources *Required Resources:* Name, Type, Required Feature List, Required Node count, Required Node List *Flags:* Name, Type, Flags *Nodes:* Name, Type, Allocated Node List, Node Expression, Node Count, Node List, Node Set Policy *Statistics:* Name, Type, Statistics *Comments:* Name, Type, Messages *Tasks:* Name, Type, Maximum Tasks, Required Task Count, Task Count *Identification:* Name, Type, Global ID, Owner *Trigger:* Name, Type, Trigger

2.3.4. List Recurring Reservations

Summary

A *recurring reservation*, also referred to as a *standing reservation* or a *reservation generator*, creates reservations according to user-defined settings. To choose which fields you would like to view in the chart, click on the customize table columns icon, which is the second icon to the left. To add fields, create a check mark by clicking in the box to the left of the field you would like to view. To remove fields, click on the checked box.

List Reoccurring Reservations

Field	Field Information
Account	This field displays which accounts can access the created reservation.
Class	This field displays which classes can access the created reservation.
Days	This field displays which days of the week the reservations will start. This field will only display information if the period is set to daily.
Disable	This field displays whether a particular reservation generator setting has been disabled.
Group	This field displays which groups can access the created reservation.
Host List	The *host list* is a list of the nodes that the reservation is using. A *node* is a computer consisting of 1 or more processors.

Name	This field displays the reservation's name.
Owner	A reservation generator can reserve only the resources that the owner can access. This field displays the owner of the reservation generator. An *owner* is a user, group, account, class, or quality of service.
Period	This field displays the interval in which the reservations will be created. The display options are daily, weekly or infinitely.
QoS	This field displays which quality of service can access the created reservation.
User	This field displays which users can access the created reservation.

2.3.5. Reservation Timeline

Summary

The Reservation Timeline window displays each reservation. On the left side of the Now line is the amount of the reservation that has been used. The right side is the remaining amount of the reservation.

2.3.6. Reservation Calendar

Summary:

The reservation calendar displays reservations color coded by AAccount. If a reservation does not have an AAccount, it is shown in gray. The height of a reservation indicates the number of processors it needs. Thus, tall reservations require more processors. Detailed information about the reservation can be seen upon mouseover. To modify a reservation, simply click on it and a reservation modification window will appear.

2.4. Triggers

2.4.1. List Triggers

Summary:

Workload Manager can launch events or triggers based on certain events. For example, an administrator may want an email sent when the reservation usage falls below a certain percentage, or a user may want to launch an evaluation script 5 minutes before his or her job is scheduled for completion.

List Triggers

Field	Field Summary
Error File	This field displays the location of the file containing all the trigger error messages.
Estimated Start Time	This field displays the date and time the trigger will execute if it is possible to calculate it. For example, reservation end times and job completion times can be calculated. Node or scheduler failures cannot be calculated ahead of time.
Flags	This field displays which flags have been set for this trigger.
Messages	This field displays the status information indicating possible failures or unexpected conditions.
Offset	This field displays the number of seconds after a resource event occurs that the trigger will execute. If this value is negative, the trigger will execute that many seconds before the resource event occurs.
Output File	This field displays the location of the file containing all the trigger output messages.
Resource Event	This field displays the event that must occur for the trigger to execute. The possible events are when the resource is created, when the resource starts, when the resource ends, or when a failure occurs in the resource.
Resource ID	This field displays the ID of the job, reservation, or node to which the trigger is attached.
Resource Type	This field displays whether the trigger is attached to a job, reservation, node, or the scheduler.
Script	This field displays the script, application, program, or executable that will be executed when the trigger is starts.
Start Time	This field displays the date and time the trigger started. This field is only populated after the trigger has been executed.
Threshold	This field displays the reservation usage threshold for this trigger. If the reservation falls below the displayed usage, the trigger will execute.
Trigger Action	This field displays the type of trigger action that will occur when the trigger is executed. The possible trigger types are cancel the resource the trigger is

	attached to; email the administrator; or execute a script, application, program, or executable.
Trigger ID	This field displays the unique ID assigned to the trigger by Workload Manager.
Trigger State	This field displays the execution status of the trigger. If the state is Idle, the trigger is waiting to execute. If the state is Active, the trigger is executing. Once the trigger has executed, the state displayed will be Successful or Failure, depending on the outcome of the trigger action.

Chapter 3.0 Resources

3.1. Resources Overview

The resources category gives adminstrators the ability to view, modify, and set policies and attributes while effectively diagnosing various system resources.

3.2. Moab Workload Manager

3.2.1. Scheduler Settings

This window provides a control center for the basic operations of Workload Manager.

Control Panel

- Pause/Resume- The Pause button stops Workload Manager from scheduling any new jobs but will not turn Workload Manager off. The resume button, which replaces the pause button when Workload Manager is paused, will allow Workload Manager to begin scheduling jobs again.
- Restart - The Restart button shuts down Moab and restarts it using the original execution environment and command line arguments.
- Shutdown- The Shutdown button turns Workload Manager off. Please note that Workload Manager cannot be restarted from Cluster Manager.
- Stop Iteration- The Stop Iteration button causes Workload Manager to stop scheduling once it reaches the iteration defined in the field. The iteration is defined as the cycle that Workload Manager is currently on. When Workload Manager starts its cycle is 0. Workload Manager increases the cycle by one approximately every 30 seconds.

3.2.2. Log Settings

Summary

Logging is defined as recording error, diagnostic, and informational messages to a file. This window allows an administrator to configure the logging that occurs in Workload Manager.

Options

Field	Additional Information
Log Level	This field allows an administrator the option of specifying the amount of data recorded in the log files. A value of 1 means almost no data is recorded while a value of 9 means all the data is recorded. Each value increment means that approximately double the amount of data is logged to the log files. The default log level is 3.
Log Facilities	This field determines what is recorded in the log file. • fCore - This option records Workload Manager core messages. • fSched - This option records messages that involve the scheduler. • fSock - This option records messages that involve the socket communication. • fUI - This option records messages that involve the user interface. • fLL - This option records messages that come from LoadLeveler Resource Manager. • fRM - This option records resource manager messages. • fSDR - This option records messages that involve system data repository. • fCONFIG - This option records messages that involve the configuration file. • fSTAT - This option records messages that involve statistics.

	• fSIM - This option records messages that occur during the simulation operating mode. • fSTRUCT - This option records messages that involve Workload Manager's structure. • fFS - This option records messages that involve fairshare. • fCKPT - This option records messages that involve the checkpoint file. • fBANK - This option records messages that involve QBank. • fPBS - This option records messages that come from the Torque/OpenPBS Resource Manager. • fWIKI - This option records messages that involve WIKI. • fALL - This option records all the events that occur.
Log Directory	This field allows an administrator to specify the directory in which log files will be maintained.
Log File	This field allows an administrator to specify the name of the Workload Manager log file.
Log File Maximum Size	This field allows an administrator the option of specifying the maximum allowed size (in bytes) of the log file before it will be *rolled*.
Log File Roll Depth	When a log file reaches its maximum size, it is *rolled* or renamed to another filename and a new log file is created using the original file name. This field allows an administrator the option of defining the number of renamed files Workload Manager should maintain.

3.2.3. System Settings

Summary

This window provides an administrator the option of changing numerous Workload Manager settings.

Options

Field	Additional Information
Name	This field allows an administrator to name the cluster. The name is only available for administrator convenience and is not used by Workload Manager.
Host	This field allows an administrator to define the host name that the Workload Manager subcomponents or clients use to connect to Workload Manager.
Port	This field allows an administrator to define the port that the Workload Manager subcomponents or clients use to connect to Workload Manager.
Home Directory	This field allows an administrator to define the directory where Workload Manager's configuration, statistics, and log files are located.
Feedback Program	This field allows an administrator to define a program that will be run at the completion of each job. Usually the program is used to contact the user through email informing him/her that the job completed execution.
Notify Program	This field allows an administrator to define a program that will be run when messages or alerts occur in Workload Manager.
RM Poll	This field is the time in between which Workload Manager will communicate

Interval	with the resource manager.
Node Purge Time	This field allows an administrator to define the amount of time Workload Manager will keep track of a node which is no longer reported by the resource manager. This value should be increased when using a resource manager that often loses information about a node due to internal failures.
Job Purge Time	This field allows an administrator to define the amount of time Workload Manager will keep track of a job which is no longer reported by the resource manager. This value should be increased when using a resource manager that often loses information about a job due to internal failures.
Charge Metric	This field allows an administrator to specify how quality of service charging should occur. 1. DEBITALLCPU - This policy will charge according to the number of processors used. 2. DEBITALLPE - This policy will charge according to the number of processors used times the number of processor equivalent* nodes used. 3. DEBITSUCCESSFULWC - This policy will charge jobs that successfully completed according to the number of hours they were on the cluster. 4. DEBITSUCCESSFULCPU - This policy will charge jobs that successfully completed according to the number of processors used. 5. DEBITSUCCESSFULPE - This policy will charge jobs that successfully completed according to the number of processor equivalent* nodes they used.
Charge Rate Policy	This field allows an administrator to specify how the quality of service charging should occur. 1. QOSREQ - This policy will charge based upon the quality of service requested. 2. QOSDEL - This policy will charge based upon the quality of service dedicated or given.

*Processor equivalence is a relative measure of how much of a node is taken by a job, even if only one type of node resource is requested. For example, if a job requires 1 processor and 1 GB of memory, and it is running on a 4 processor node with 1GB of memory, the PE of the job is 4. All of the processors are considered to be taken because the first job is using all of the memory, which prevents any other job from running on that node.

3.2.4. Statistics Settings

Summary

Credential statistics are disabled by default although cluster-wide statistics are always enabled.

Enable Credentials Statistics

This section allows an administrator to enable or disable user, group, account, class, or

quality of service statistics. Because statistics increase Workload Manager's memory usage, an administrator can decrease the memory footprint of Workload Manager by disabling credential statistics.

Number of intervals in each day

Workload Manager combines statistics into intervals. The Daily Statistical Count allows an administrator the option of increasing or decreasing the number of intervals in each day. A higher amount of intervals creates more precise statistics, but Workload Manager uses more memory when intervals are higher

3.2.5. High Availability

Summary

High availability provides a backup Workload Manager in the unlikely situation of a failure.

High availability allows an administrator to state the connection information for the primary Workload Manager and a backup secondary Workload Manager.

The fields for both the primary and secondary Workload Manager are as follows:

- Status - This field displays information regarding the state reported by a Workload Manager. There are a few states reported by a Workload Manager:
    1. Running - Workload Manager is executing as expected.
    2. Hibernating - Workload Manager is operating as a backup scheduler
    3. Unknown - Contact with Workload Manager has failed or has not been correctly set up.
- Host - The host name where Workload Manager is located.
- Port - The port which the Workload Manager communicates.

3.3. Nodes

3.3.1. Modify Nodes

Node Information

Field Name	Description
Node ID	This field assigns the node profile to a desired node.
Node State	This field displays the operating status of the node. For example, unknown, draining, busy, running, down, idle, etc.

Node Usage Limits

Field Name	Description
Maximum Jobs	This field allows the user to specify the maximum number of simultaneous jobs allowed to run on this node.
Maximum Jobs Per User	This field allows the user to specify the maximum number of simultaneous jobs per end user allowed to run on the node.
Maximum Load	This field allows the user to specify the maximum percentage of load allowed to run on this node. *Load* is the number of jobs divided by the number of processors.

General Attributes

Field Name	Description
Priority	This field allows a user to specify the fixed node priority relative to other nodes. The default node priority is 0.
Speed	This field allows a user to specify the speed as a relative value. A base node generally has a speed of 1.0, but may be changed based on memory, networking interface, etc. It does not have to be proportional to processor speeds.
Processor Speed	This field allows a user to specify the processor(s) speed on this node. This provides Workload Manager the information needed to schedule nodes with similar processor speeds.
Load	**This field allows the user to specify the maximum percentage of load allowed to run on this node. Load is the number of jobs divided by the number of processors.**
Partition	Clusters can be divided into different sections. These sections are called

	partitions. This field displays the partition to which the node is assigned.
Rack	This field allows a user to specify the rack number where the node is located.
Slot	This field allows a user to specify the slot number where the node is located.
Features	This field allows a user to specify features assigned to this node.

Statistics

Field Name	Description
Active Time	This field displays the total time the node has actively been executing jobs.
Up Time	This field displays the total time the node has been available to execute jobs.
Total Time	This field displays the total time the node has been on the cluster.

Node Software

Field Name	Description
OS	The operating system the node uses.
OS List	The list of possible operating systems the node may use.

Node Attributes

Field Name	Description
Architecture	This field displays the hardware architecture of the node. The exact hardware information displayed depends on the information the resource manager supplies to Workload Manager.

Node Priority Function

If Moab is set up to use the *NODEALLOCATIONPOLICY* of *PRIORITY*, you will also see the Node Priority Function Editor panel. This editor allows you to modify the function that is used to assign the node its priority. This function corresponds to the *PRIORITYF* attribute in Moab's configuration.

A valid priority function consists of any number of function expressions. A function expression is a coefficient multiplied by an attribute. For example, one function expression could be *5 * Available Memory*. All the function expressions are summed up to get an effective priority value. This summation is then used to calculate the node's effective priority.

For more information about node allocation policies or the PRIORITY policy, see the Moab documentation.

Over Commit

Field Name	Displayed

Disk	Allows overcommitting of the disk space of the node.
Memory	Allows overcommitting of the memory of the node.
Processors	Allows overcommitting of processors of the node.
Swap	Allows overcommitting of the swap space of the node.

Node Hardware

Field Name	Description
Free Disk Space	This field displays the available disk space, measured in megabytes (MB), on the node.
Free Memory	This field displays the available memory, measured in megabytes (MB), on the node.
Free Processors	This field displays the number of available processors on the node.

Job List

Job List displays all the jobs scheduled on the node.

Reservation List

Reservation List displays all the reservations on the node.

3.3.2. List Nodes

List Node Fields

Field	Additional Information
Active Time	This field displays the total time the node has spent doing work.
Architecture	This field displays the hardware architecture of the node. The exact hardware information displayed will depend upon the information the resource manager supplies to Workload Manager.
Available Class	This field displays the classes that can access the node.
Available Disk	This field displays the available disk space, measured in megabytes (MB), on the node.
Available Memory	This field displays the available memory, measured in megabytes (MB), on the node.
Available Processors	This field displays the number of available processors on the node.

Available Swap	This field displays the available swap, measured in megabytes (MB), on the node.
Block Reason	This field displays any error messages related to the node.
Charge Rate	This field displays the charging rate to the usage of particular resources.
Class	This field displays the classes that can access the node.
Comments	This field allows a user to attach a comment to the node.
Consumable Resources	This field displays the restricted resources and the current number available for the node. Some clusters have restrictions placed upon certain nodes. Usually these restrictions are in the form of software licenses. Sometimes a software license can restrict the number of jobs that can simultaneously be using the software on a node. *Consumable resources* allow a system administrator to define the number of licenses or other restricted resources available on a particular node.
Dynamic Job List	A list of the jobs that may require different resources at different times.
Features	A feature is a custom attribute often describing a unique hardware or software configuration associated with the node. This field displays the features associated with the node.
Job List	A node can execute one or more jobs simultaneously. This field displays a list of jobs currently executing on the node.
Load	The *load* is the number of processors on the node divided by the number of jobs on the node. This field displays what the current node load is.
Maximum I.O. Input	This field displays the maximum disk input in bytes that can occur before the node state is changed to busy.
Maximum I.O. Load	This field displays the maximum disk input and output in bytes that can occur before the node state is changed to busy.
Maximum I.O. Output	This field displays the maximum disk output in bytes that can occur before the node state is changed to busy.
Maximum Jobs	This field displays the maximum number of jobs allowed on the node at one time.
Maximum Jobs Per User	This field displays the maximum number of jobs for a single user allowed on the node at one time.
Maximum Load	The *load* is defined as the number of processors on the node divided by the number of jobs on the node. This field displays the maximum load for the node.
Maximum Processor Equivalent Per Job	This field displays the maximum number of processor equivalents* per job allowed on this node at one time.
Maximum Processors	This field displays the maximum number of utilized processors allowed on this node at one time.

Maximum Processors Per Class	This field displays the maximum number of utilized processors per class allowed on this node at one time.
Messages	This field displays information messages provided by Workload Manager relating to the node.
Network	This field displays the network hardware on the node.
Node ID	All nodes require a unique ID. This field displays that ID.
Node Type	A *node type* is a custom tag attached to a node. It is usually used in conjunction with an allocation manager such as QBank to assign different charge rates according to the specific node type. This field displays the node type attached to the node.
Operating System	A node is configured with a specific operating system. This field displays the node's configured operating system.
Operating System List	A node is configured with a specific operating system. This field displays the node's configured operating system, as well as other operating systems that are compatible with the configured operating system.
Partition	Clusters can be divided into different sections. These sections are called partitions. This field displays the partition to which the node is assigned.
Priority	This field displays the priority of the node. The default priority is 0.
Priority Function	This field displays which priority function will be used to calculate a node's priority.
Processor Speed	This field displays the processor speed as gathered from the resource manager.
Rack	This field displays the rack number where the node is logically located.
Reservation Count	This field displays the number of reservations on the node.
Size	This field displays a description of the size of the node such as 1u or 2u.
Slot	This field displays the slot number where the node is logically located.
State	This field displays the operating status of the node. For example, unknown, draining, busy, running, down, idle, etc.
Total Disk	This field displays the total disk space, measured in megabytes (MB), on the node.
Total Memory	This field displays the total memory, measured in megabytes (MB), on the node.
Total Processors	This field displays the total number of processors on the node.
Total Swap	This field displays the total swap space, measured in megabytes (MB), on the node.
Total Time	This field displays the total time the node has been on the cluster.
Up Time	This field displays the total time the node has been available to execute jobs

3.3.3. Node Calendar

Summary

The node calendar displays the jobs and reservations on a calendar. Current and future reservations and jobs are shown as well as historical jobs and reservation. The top bar or x-axis is the displayed time frame. The left bar or y-axis is the nodes on the cluster. The colored boxes/cells in the table are identified in the display key.

The node calendar supports 4 time frames. Days in Month, Days in Week, Hours in Day, Minutes In Hour. The top left tabs allow the user to choose the desired time frame. When the **Display Selected Time Frame** button is selected, the currently selected time frame will be displayed.

3.3.3.1. Display Key

The Display Key panel allows the user to show or hide resources on the node calendar.

The Current Time displayed in green allows a user to see what the current time on the cluster is according to Moab's time reporting.

Jobs displayed in light blue shows all jobs according to the job reservations created for that job.

Reservations displayed in dark blue shows all reservations that fit within the specified time frame.

3.3.3.2. Resize Calendar

The resize calendar panel, when enabled, reduces the size of each box/cell in the table to allow the user to see more nodes at once. The node names are not visible when the table is compacted.

3.4. Partitions

3.4.1. Modify a Partition Profile

Clusters can be divided into different sections. These sections are commonly called *partitions*. A *partition* is a semi-permanent division of the cluster and is most often used when certain nodes contain unique hardware. It should be recognized that jobs are not allowed to run in more than one partition. If jobs need the ability to span multiple partitions an infinite reservation should be used instead of a partition.

The ability to modify partitions is not currently available in Cluster Manager.

3.4.2. List Partitions

List Partition

Field	Description
ID	This field displays the partition ID, or the name of the partition.
Reserved Disk	This field displays the amount of disk space reserved by this partition. The disk space is measured in megabytes (MB).
Reserved Memory	This field displays the amount of memory reserved in this partition. The memory is measured in megabytes (MB).
Reserved Nodes	This field displays the number of nodes reserved in this partition.
Reserved Swap	This field displays the amount of swap space reserved by this partition. The swap space is measured in megabytes (MB).
Resource Manager	This field displays the resource manager ID of which this partition is a member.
Total Disk	This field displays the total disk space available in the partition. The disk space is measured in megabytes (MB).
Total Memory	This field displays the total memory available in this partition. The memory is measured in megabytes (MB).
Total Nodes	This field displays the total number of nodes available in this partition.
Total Processors	This field displays the total number of processors available in this partition.
Total Swap	This field displays the total swap space available in the partition. The swap space is measured in megabytes (MB).
Utilized Disk	This field displays the amount of disk space currently being used by this partition. The disk space is measured in megabytes (MB).
Utilized	This field displays the amount of memory currently being used by this

Memory	partition. The memory is measured in megabytes (MB).
Utilized Nodes	This field displays the number of nodes currently being used in this partition.
Utilized Processor	This field displays the number of processors currently being used in this partition.
Utilized Swap	This field displays the amount of swap space currently being used by this partition. The swap space is measured in megabytes (MB).

3.5. Cluster

3.5.1. Visual Cluster

Summary

The visual cluster is an easy and concise way of viewing your entire cluster and the status of each node. The table and explanation below explain how to interpret the visual cluster:

Table 3-1. Visual Cluster Example

	Slot 1	Slot 2	Slot 3
Rack 1	Node A		Node D
Rack 2		Node B	
Rack 3			Node C

A rack is a physical frame that holds a node. The slot is the location of the node inside the rack. The racks make up the first column of the table. The slot locations increase from left to right. For example, Node A is located on Rack 1 in Slot 1. Node D is also located on Rack 1 but instead of Slot 1 it's located in Slot 3. In the visual cluster Node A through Node D are displayed as icons. The different icons can represent node state, node attributes, reservations, jobs, and/or nodes. The subpanel sections below describe these states in more detail. Further information can be gathered about nodes by hovering the mouse over any nodes.

It should be noted that the visual cluster is for display purposes only and the location of the node does not play any part in how Workload Manager schedules.

3.5.1.1. Node Attribute Selector

The node attribute selector gives the user the power to see various attributes of the nodes displayed in the Visual Cluster. This allows the user to compare and contrast attributes of interest. Node attributes include standard categories such as architecture, OS, hardware metrics (memory, disk, swap, etc.), as well as any metric read in through Moab as a generic metric (a node's GMETRIC). The default display for node attributes is the corresponding color of the outer rim of each node cell. This can be changed in the section titled "Node Display Preferences".

The Clear Attribute button clears any current selection and node attribute.

Once a node attribute is selected, Moab Cluster Manager determines the number of nodes and processors that describe each attribute and displays a corresponding key value that matches the Visual Cluster by color. Each attribute's display can be individually controlled via the check box next to each node attribute name and color.

If the node attribute is a numerical value - for example, a generic metric - then Moab Cluster Manager will attempt to place the values into a reasonable range as to effectively categorize the values.

3.5.1.2. Highlight Jobs and/or Nodes

This section provides a user with the option of highlighting resources in the visual cluster table. The three sections are divided into jobs and nodes. Each section can be simultaneously displayed by giving each border a different color. The white box displays the names of the selected resources with the headers "Job: " and "Node: " respectively. The Select buttons open the list jobs or list nodes window depending on which resource the user has selected. The visual cluster window appears with the desired resources highlighted. The Clear buttons remove the highlight from the visual cluster table and erase the names from the colored box. The Color button changes the highlight color for the specified resource. The new highlight color is displayed in the colored box.

3.5.2. Processor Usage

Summary

This graph displays how the cluster's processors are being used over time. The left bar, or y-axis, displays the number of processors. The bottom bar, or x-axis, displays time. The light yellow color displays the total available processors on the cluster. The dark yellow color displays the processors used by jobs and job reservations. The blue color displays the processors used by reservations other than job reservations.

The switch statistics option allows for the Available Processors and Jobs Reservations colors to be switched.

3.5.3. Event Calendar

Summary

The events calendar helps you view the events you are interested in while filtering out those you are not interested in. This can be much more convenient than searching through logs and event files. Once you have finished filtering, you can view the details of the remaining events and correlate this with your nodes' state history.

Filtering

The first filter you should probably apply is the time filter. After setting a start and end time, you can then select what type of events you wish to see. Use the tree control at the top left of this window to click the event types and/or subtypes you desire. For example, you might choose to see all events, all job events, or job end events only. After making your selection, click the Apply Filter button.

If you are looking for events for a specific object, and you know the object's name, you can filter out all other events by specifying the object type and ID in the fields at the lower left. For example, you might wish to see only events relating to the job "moab.2". To do this, you would specify an object type of "Job" and an object ID of "moab.2".

Viewing Event Details

To view the details of desired events, simply click their colored icon representations in the timeline. Note that on busy schedulers, an icon may represent multiple events. You will be

able to see the event type, ID, time, object type, object ID, and any messages from the scheduler pertaining to this event.

The Node State/Node Category Chart

Below the timeline is a colored line chart. Vertically below any point on the timeline one can see the percentage of nodes that were active, idle, down, etc. This can be useful in determining the events you might want to filter. For example, if you see that 30% of your nodes suddenly went idle, yet you know there was a large backlog during that time, you might want to view all reservation start events. This might help you find an unused reservation responsible for the idle nodes.

3.6. Virtual Private Cluster

3.6.1. VPC Profiles

VPC Profiles allow sites to set default parameters and resources for the construction of Virtual Private Clusters (VPC).

- Name - The name of the VPC profile
- Description - A description of the VPC profile
- Reservation Profile - The default reservation profile
- Node Hour Charge Rate - The per node charge rate multiplier associated with using resources and services in a VPC
- Node Set Up Cost - The per node cost in credits associated with provisioning resources and when setting up a VPC
- Required Start Pad - The amount of additional time required before each explicitly specified requirement
- Required End Pad - The amount of additional time required after each explicitly specified requirement
- Access Control List - Specifies which credentials can view this profile and can create Virtual Private Clusters using this profile
- Node Sets - A list of node set constraints
- Required Attributes - A list of resource requirements

3.6.2. Virtual Private Clusters

A Virtual Private Cluster (VPC) is a resource collection including such items as network, data, and security objects. These resources are then available to customers as a complete package or customized to fit individual needs. This greatly simplifies the process of requesting needed resources.

- ID - The unique name of the VPC
- Host List - The hosts this VPC will run on
- Owner - The owner of the VPC
- Owner Type - The credential type of the owner. For example, it could be user, group, account, etc.
- VPC Profile - The VPC profile associated with this VPC
- Start Time - The time this VPC will start
- Duration - How long this VPC will last
- State - The state of this VPC
- Cost - The cost (in credits) of this VPC

The control panel allows one to destroy or attach messages to this VPC

- Destroy - Will destroy this VPC
- Send Message - Will attach a message to this VPC

Chapter 4.0 Credentials

4.1. Credentials Overview

The credentials section allows an adminstrator to view all credentials in the system and their various roles. New credential profiles can be added as well as modified.

4.2. Visual Credential Access

Summary

This window allows a user to view which credentials can access which credentials. The arrows symbolize that the credential can access the other credential. There are three distinct sections displayed in the window. The first section displays all the credentials that can access the second section. The second section contains only the selected credential. The arrows from the second section to the third section show what the selected credential can access. The first and third sections may not be displayed if they do not contain credentials that are accessible to the credential in the second section.

- Display All Credentials - This field allows a user to display all of the credentials of a specific credential type.
- Display Listed Credentials - This field allows a user to display only the listed credentials.
- Display who can access the selected credential - This field allows a user to specify whether the credentials that access the selected credentials should be displayed.
- Display who the selected credential can access - This field allows a user to specify whether the credentials that the selected credential can access should be displayed.

4.3. User Access

Summary

The User Access Settings page allows an administrator to assign roles to each user in the system. These settings are saved in Workload Manager and affect command-line operations, as well as permissions within Cluster Manager. The set of default roles available is:

1. *Admin1* - Users with this role are called *administrators* . They have complete control of Workload Manager.
2. *Admin2* - Users with this role are called *operators* . They have complete control of jobs, nodes, reservations, etc., but cannot modify Workload Manager configuration parameters.
3. *Admin3* - Users with this role are called *help desk personnel* . They can control various aspects of Workload Manager, but cannot modify workload.
4. *Admin4* - Users with this role are most likely trusted or experienced users. They have a subset of Admin2 permissions that is different from Admin3's.
5. *Admin5* - Users with this role can only view workload and resource information.

4.4. List Credentials

Summary

This window displays all of the information regarding users, groups, accounts, classes, and qualities of service (QoS) commonly called credentials.

List Credential Fields

Field	Additional Information
Access Resources QT Threshold	This field displays the number of minutes that an idle job must wait before it can access any reservations owned by the Quality of Service.
Access Resources XF Threshold	This field displays the expansion factor value that an idle job must be greater than or equal to in order to access any reservations owned by the Quality of Service.
Account	This field displays the accounts that this particular credential ID can access.
Account Default	This field displays the account that will be used by this credential ID's job if no account is specified.
Class	This field displays the classes that this particular credential ID can access.
Class Default	This field displays the class that will be used by this credential ID's job if no class is specified.
Comments	This field allows a user to enter any comments relating to the Credential.
Create Reservation QT Threshold	This field displays the number of minutes that an idle job must wait before a job reservation will be created for it. A job reservation will guarantee it specific resources, as well as a specific start time.
Create Reservation XF Threshold	This field displays the expansion factor value that an idle job must be greater than or equal to before a job reservation will be created for it. A job reservation guarantees it specific resources, as well as a specific start time.
Credential	This field displays whether the credential is a user, group, account, class, or quality of service (QoS).
Credits	This field displays the total credits available to the credential ID.
Dedicated Resource Cost	What Workload Manager charges for each resource unit dedicated - whether used or not - to a job.
Default QoS	The default quality of service (QoS) is the quality of service (QoS) that will be used by this credential ID's job if no quality of service (QoS) is specified
E-Mail	This field allows a user to enter the e -mail address for the Credential.

Address	
Enable Statistics	Statistics are tracked for each credential ID. This field allows the user the option of enabling/disabling statistics for each credential ID.
Fairshare Target	Refer to the fairshare section for information regarding fairshare target.
Fairshare Type	Refer to the fairshare section for information regarding fairshare type.
Group	The operating system is usually responsible for the creation of groups. This field displays the groups that this particular credential ID can access.
Group Default	This field displays the group that will be used by this credential ID's job if no group is specified.
ID	All credentials must have an identification unique to its credential type. This field displays the credential identification.
Maximum Job	The default maximum job is the maximum job value that will be used by this credential ID's job if no maximum job is specified.
Maximum Nodes	This field displays the maximum nodes value that will be used by this credential ID's job if no maximum nodes value is specified.
Maximum Proc	This field displays the maximum processors value that will be used by this credential ID's job if no maximum processor is specified.
Maximum Proc Sec	This field displays the maximum processor seconds value that will be used by this credential ID's job if no maximum processor seconds value is specified.
Partition	Clusters can be divided into different sections. These sections are commonly called *partitions*. This field displays the partitions this credential ID can access.
Partition Default	This field displays the partition that will be used by this credential ID's job if no partition is specified.
Preemption QT Threshold	This field displays the number of minutes that an idle job must wait before it will be given preemptor access.
Preemption XF Threshold	This field displays the expansion factor value that an idle job must be greater than or equal to before it will be given preemptor access.
Priority	This field displays the priority of the credential.
QoS	This field displays the quality of services (QoS) that this particular credential ID can access.
QoS Flags	This field displays the quality of service (QoS) settings for this credential ID.
Queue Time Weight	This field displays the quality of service weight factor. If an idle job is submitted to this quality of service, the number of minutes that it has been in the queue is multiplied by this value. This increases the job's start priority.
Reservation	The required reservation that any job submitted has to use.
Soft	The soft maximum job limit restricts the number of jobs allowed to execute

Maximum Job	for this credential ID. If, however, additional resources are available after all the soft maximum job limits are met, then the hard maximum job limits are used.
Soft Maximum Nodes	The soft maximum node limit restricts the number of nodes used by any job for this credential ID. If, however, additional resources are available after all the soft maximum node limits are met, then the hard maximum node limits are used.
Soft Maximum Proc	The soft maximum processor limit restricts the number of processors used by any job for this credential ID. If, however, additional resources are available after all the soft maximum processor limits are met, then the hard maximum processor limits are used.
Soft Maximum Proc Sec	The soft maximum processor seconds limit restricts the number of processor seconds used by any job for this credential ID. If, however, additional resources are available after all the soft maximum processor seconds limits are met, then the hard maximum processor seconds limits are used.
Used Credits	This field displays the credits used by this credential ID.
User	This field displays the users that this particular credential ID can access.
User Default	This field displays the user that will be used by this credential ID's job if no user is specified.
Utilized Jobs	This field displays the number of jobs currently executing for this credential ID.
Utilized Nodes	This field displays the number of nodes currently being used by this credential ID's jobs.
Utilized Proc	This field displays the number of processors currently being used by this credential ID's jobs.
Utilized Proc Sec	This field displays the total number of processors used by executing jobs for this particular credential ID times the number of seconds each processor has been used.
Utilized Resource Cost	This field displays what Workload Manager charges for each resource unit consumed/utilized by a job.
XFactor Weight	This field displays the quality of service weight factor. If an idle job is submitted to this quality of service, its expansion factor is multiplied by this value. This increases the job's start priority.

Categories

Membership: Credential, ID, User, Group, Class, Account, Quality of Service (QoS)

Utilized Resources: Credential, ID, Utilized Job, Utilized Processors, Utilized Nodes, Utilized Processor Seconds

Soft Maximum Limits: Credential, ID, Soft Maximum Jobs, Soft Maximum Processors, Soft Maximum Nodes, Soft Maximum Processor Seconds

Priority: Credential, ID, Credential Priority

Fairshare: Credential, ID, Fairshare Type, Fairshare Target

Partition & Reservation: Credential, ID, Partition, Reservation

Statistics: Credential, ID, Enable Statistics

Credits: Credential, ID, Total Credits, Used Credits,

Default Credentials: Credential, ID, User Default, Group Default, Class Default, Account Default, Quality of Service (QoS) Default

Default Resource: Credential, ID, Partition Default, Maximum Job Default, Maximum Processor Default, Maximum Nodes Default, Maximum Processor Seconds Default

Comments & Email: Credential, ID, Comments, E-Mail Address

4.5. Create / Modify a User Profile

Summary

Users are created by the operating system while user profiles are created by Workload Manager. When a user submits a job, that user becomes visible to Workload Manager and at that moment a credential profile is automatically created for the user.

Credential Access

Field	Description
User Name	This field allows an administrator to define the name of the user. Usually this is the user's login name.
Group Access List	The group access is defined by the operating system and cannot be defined by Workload Manager.
Account Access List	This field allows an administrator to define which accounts this user can access.
Default Account	This field allows an administrator to define which accounts will be automatically used if the user doesn't specify an account.
Quality of Service (QoS) Access List	This field allows an administrator to define which qualities of service (QoS) this user can access.
Default Quality of Service (QoS)	This field allows an administrator to define which quality of service (QoS) will automatically be used if the user doesn't specify a quality of service (QoS).

Resource Access

Field	Description
Partition	This field allows an administrator to define which partitions this user can access
Default Partition	This field allows an administrator to define which partitions this user can access by default.
Reservation	This field allows an administrator to define which reservation this user can access.

Job Usage Limits

Field	Description
Idle Maximum Jobs	This field allows an administrator the option of setting the user's maximum number of simultaneously idle jobs.

Maximum Jobs	This field allows an administrator to set the user's maximum number of simultaneously executing jobs.
Maximum Procs	This field allows an administrator to set the user's maximum number of simultaneously utilized processors.
Maximum Proc Seconds	This field allows an administrator to set the user's maximum number of simultaneously utilized processor seconds. *Processor seconds* is the number of processors utilized times the number of seconds they are utilized.
Maximum Nodes	This field allows an administrator to set the user's maximum number of simultaneously utilized nodes. A *node* is a computer consisting of 1 or more processors
Maximum Wallclock	This field allows an administrator to set the user's maximum wallclock time available when submitting a job.
Minimum Wallclock	This field allows an administrator to set the user's minimum wallclock time when submitting a job.
Default Wallclock	This field allows an administrator to set the user's default wallclock time set when submitting a job with no specified wallclock.

Fairness

Field	Description
Fairshare Policy	*Fairshare* is a method of enforcing cluster sharing between credentials. A *credential* is a user, group, account, class/queue, or quality of service (QoS). Fairshare tracks each credential's usage for a desired amount of time and decreases a job's start priority if the fairshare policy is violated. By decreasing a job's start priority, a user specifies that the job will wait longer in the queue before it starts allowing other jobs to execute first. • Fairshare Floor Policy - If the user's cluster usage is below the fairshare target, then the user's start priority for the job will increase. The user's cluster usage is measured as the total percentage amount of the cluster used by the user. • Fairshare Target Policy- If the user's cluster usage is above or below the fairshare target, then the user's start priority for the job will increase or decrease accordingly. The user's cluster usage is measured as the total percentage amount of the cluster used by the user. • Fairshare Cap Policy - If the user's cluster usage is above the fairshare target, then the user's start priority for the job decreases. The user's cluster usage is the total percentage of the cluster used by the user. • Absolute Fairshare Policy - If a user's cluster usage exceeds the fairshare target, then the start priority for the job will decrease. The user's cluster usage is the total number of processor hours used by a user. • Relative Fairshare Policy - If a user's cluster usage exceeds the fairshare target, then the user's start priority for the job will decrease. The user's cluster usage is the total percentage of the cluster used by the user.
Fairshare Target	This field allows an administrator to define the fairshare target for this user. Refer to the fairshare policy for an understanding of how fairshare target will be used.
Priority	This field allows an administrator to define a user's job priority. A user's job

	priority will increase or decrease the start priority of this user's jobs. Workload Manager, with some exceptions, will start the jobs with the highest start priority first.

Comments

Field	Description
Comments	This field allows an administrator to enter comments regarding the user.

Email Address

Field	Description
Email Address	This field allows an administrator to add a user's email address to Workload Manager. The email address is only for contact information and is not used by Workload Manager or the resource manager.

Usage Statistics

This is only visible if a profile is being modified.

Field	Description
System Utilization	Shows what percent of the entire system has recently been dedicated to this credential. Each value in this line graph represents the number of dedicated processor seconds used by this credential during the profiling duration divided by the number of dedicated processor seconds in the entire system for the profiling duration. The number of dedicated processor seconds for the entire system for a profiling duration is equal to the size of the profiling duration in seconds multiplied by the number of processors.

4.6. Create / Modify a Group Profile

Summary

Groups are created by the operating system while group profiles are created by Workload Manager. When a user submits a job, that user's group becomes visible to Workload Manager and at that moment, a credential profile is automatically created for the group.

Credential Access

Field	Description
Group Name	This field allows an administrator to define the identification name of the group. Usually this is the login name for the group.
User Access List	This field allows an administrator to define which users can access this group.
Group Access List	This field allows an administrator to define which accounts this group can access.
Quality of Service (QoS) Access List	This field allows an administrator to define which qualities of service (QoS) this group can access.
Default Quality of Service (QoS)	This field allows an administrator to define which quality of service (QoS) will automatically be used if the group doesn't specify a quality of service (QoS).

Resource Access

Field	Description
Partition List	This field allows an administrator to define which partitions this group can access.
Default Partition	This field allows an administrator to define which partitions this group can access by default.
Reservations	This field allows an administrator to define which reservation this group can access.

Job Usage Limits

Field	Description
Idle Maximum Jobs	This field allows an administrator the option of setting the group's maximum number of simultaneously idle jobs.
Maximum Jobs	This field allows an administrator to set the group's maximum number of simultaneously executing jobs.

Maximum Procs	This field allows an administrator to set the group's maximum number of simultaneously utilized processors.
Maximum Proc Seconds	This field allows an administrator to set the group's maximum number of simultaneously utilized processor seconds. *Processor seconds* is the number of processors utilized times the number of seconds they are utilized.
Maximum Nodes	This field allows an administrator to set the group's maximum number of simultaneously utilized nodes. A *node* is a computer consisting of 1 or more processors.
Maximum Wallclock	This field allows an administrator to set the group's maximum wallclock time available when submitting a job.
Minimum Wallclock	This field allows an administrator to set the group's minimum wallclock time when submitting a job.
Default Wallclock	This field allows an administrator to set the group's default wallclock time set when submitting a job with no specified wallclock.

Fairness

Field	Description
Fairshare Policy	*Fairshare* is a method of enforcing cluster sharing between credentials. A *credential* is a user, group, account, class, or quality of service (QoS). Fairshare tracks each credential's usage for a desired amount of time and decreases a job's start priority if the fairshare policy is violated. By decreasing a job's start priority, the job will wait longer in the queue before it starts, allowing other jobs to execute first. • Fairshare Floor Policy - If the group's cluster usage is below the fairshare target , then the group's start priority for the job increases. The group's cluster usage is measured as the total percentage amount of the cluster used by the group. • Fairshare Target Policy- If the group's cluster usage is above or below the fairshare target , then the group's start priority for the job increases or decreases accordingly. The group's cluster usage is measured as the total percentage of the cluster used by the group. • Fairshare Cap Policy - If the group's cluster usage is above the fairshare target , then the group's start priority for the job decreases. The group's cluster usage is the total percentage of the cluster used by the group. • Absolute Fairshare Policy - If a group's cluster usage exceeds the fairshare target , then the group's start priority for the job decreases. The group's cluster usage is measured as the total number of processor hours used by a group. • Relative Fairshare Policy - If a group's cluster usage exceeds the fairshare target , then the group's start priority for the job decreases. The group's cluster usage is the total percentage of the cluster used by the group.
Fairshare Target	This field allows an administrator to define the fairshare target for this group. Refer to the Fairshare Policy for an understanding of how fairshare target will be used.
Priority	This field allows an administrator to define a group's job priority. A group's job priority will either increase or decrease the start priority of this group's jobs.

	Workload Manager, with some exceptions, will start the jobs with the highest start priority first.

Comments

Field	Description
Comments	This field allows an administrator to enter any comments regarding the group.

Usage Statistics

This is only visible if a profile is being modified.

Field	Description
System Utilization	Shows what percent of the entire system has recently been dedicated to this credential. Each value in this line graph represents the number of dedicated processor seconds used by this credential during the profiling duration divided by the number of dedicated processor seconds in the entire system for the profiling duration. The number of dedicated processor seconds for the entire system for a profiling duration is equal to the size of the profiling duration in seconds multiplied by the number of processors.

4.7. Create / Modify an Account Profile

Summary

Account creation occurs in Workload Manager.

Credential Access

Field	Description
Account Name	This field allows an administrator to define the name of the account. This is usually the login name for the account.
User Access List	This field allows an administrator to define which users can access this account.
Group Access List	This field allows an administrator to define which groups can access this account.
Quality of Service (QoS) Access List	This field allows an administrator to define which qualities of service (QoS) this account can access.
Default Quality of Service (QoS)	This field allows an administrator to define which quality of service (QoS) will automatically be used if the account doesn't specify a quality of service (QoS).

Resource Access

Field	Description
Partition	This field allows an administrator to define which partitions this account can access
Default Partition	This field allows an administrator to define which partitions this account can access by default.
Reservation	This field allows an administrator to define which reservation this account can access.

Job Usage Limits

Field	Description
Idle Maximum Jobs	This field allows an administrator the option of setting the account's maximum number of simultaneously idle jobs.
Maximum Jobs	This field allows an administrator to set the account's maximum number of simultaneously executing jobs.
Maximum	This field allows an administrator to set the account's maximum number of

Procs	simultaneously utilized processors.
Maximum Proc Seconds	This field allows an administrator to set the account's maximum number of simultaneously utilized processor seconds. *Processor seconds* is defined as the number of processors utilized times the number of seconds they are utilized.
Maximum Nodes	This field allows an administrator to set the account's maximum number of simultaneously utilized nodes. A *node* is a computer consisting of 1 or more processors
Maximum Wallclock	This field allows an administrator the option of setting the account's maximum wallclock time available when submitting a job.
Minimum Wallclock	This field allows an administrator the option of setting the account's minimum wallclock time when submitting a job.
Default Wallclock	This field allows an administrator the option of setting the account's default wallclock time set when submitting a job with no specified wallclock.

Fairness

Field	Description
Fairshare Policy	*Fairshare* is a method of enforcing cluster sharing between credentials. A *credential* is a user, group, account, class, or quality of service (QoS). Fairshare tracks each credential's usage for a desired amount of time and decreases a job's start priority if the fairshare policy is violated. By decreasing a job's start priority, the job will wait longer in the queue before it starts, allowing other jobs to execute first. • Fairshare Floor Policy - If the account's cluster usage is below the fairshare target, then the account's start priority for the job will increase. The account's cluster usage is the total percentage of the cluster used by the account. • Fairshare Target Policy- If the account's cluster usage is above or below the fairshare target, then the account's start priority for the job increases or decreases accordingly. The account's cluster usage is measured as the total percentage amount of the cluster used by the account. • Fairshare Cap Policy - If the account's cluster usage is above the fairshare target, then the account's start priority for the job increases or decreases. The account's cluster usage is the total percentage of the cluster used by the account. • Absolute Fairshare Policy - If an account's cluster usage exceeds the fairshare target, then the account's start priority for the job decreases. The account's cluster usage is measured as the total number of processor hours used by an account. • Relative Fairshare Policy - If an account's cluster usage exceeds the fairshare target, then the account's start priority for the job decreases. The account's cluster usage is the total percentage of the cluster used by the account.
Fairshare Target	This field allows an administrator to define the fairshare target for this account. Refer to the Fairshare Policy for an understanding of how fairshare target will be used.

Priority	This field allows an administrator to define an account's job priority. An account's job priority increases or decreases the start priority of this account's jobs. Workload Manager, with some exceptions, starts the jobs with the highest start priority first.

Comments

Field	Description
Comments	This field allows an administrator to enter comments regarding the account.

Usage Statistics

This is only visible if a profile is being modified.

Field	Description
System Utilization	Shows what percent of the entire system has recently been dedicated to this credential. Each value in this line graph represents the number of dedicated processor seconds used by this credential during the profiling duration divided by the number of dedicated processor seconds in the entire system during the profiling duration. The number of dedicated processor seconds for the entire system for a profiling duration is equal to the size of the profiling duration in seconds multiplied by the number of processors.

4.8. Create / Modify a Class Profile

Summary

Classes are created by the resource manager while class profiles are created by Workload Manager.

Credential Access

Field	Description
Class Name	This field allows an administrator to define the name of the class. Usually this is the login name for the class.
User Access List	This field allows an administrator to define which users can access this class.
Group Access List	This field allows an administrator to define which groups can access this class.
Account Access List	This field allows an administrator to define which accounts this class can access.
Default Account	This field allows an administrator to define which accounts will automatically be used if the class doesn't specify an account.
Quality of Service (QoS) Access List	This field allows an administrator to define which qualities of service (QoS) this class can access.
Default Quality of Service (QoS)	This field allows an administrator to define which quality of service (QoS) will automatically be used if the class doesn't specify a quality of service (QoS).

Resource Access

Field	Description
Partition	This field allows an administrator to define which partitions this class can access.
Default Partition	This field allows an administrator to define which partitions this class can access by default.
Reservation	This field allows an administrator to define which reservation this class can access.

Job Usage Limits

Field	Description
Idle Maximum	This field allows an administrator to set the class's maximum number of

Jobs	simultaneously idle jobs.
Maximum Jobs	This field allows an administrator to set the class's maximum number of simultaneously executing jobs.
Maximum Procs	This field allows an administrator to set the class's maximum number of simultaneously utilized processors.
Maximum Proc Seconds	This field allows an administrator to set the class's maximum number of simultaneously utilized processor seconds. *Processor seconds* is the number of processors utilized times the number of seconds they are utilized.
Maximum Nodes	This field allows an administrator to set the class's maximum number of simultaneously utilized nodes. A *node* is a computer consisting of 1 or more processors
Maximum Wallclock	This field allows an administrator to set the class's maximum wallclock time available when submitting a job.
Minimum Wallclock	This field allows an administrator to set the class's minimum wallclock time when submitting a job.
Default Wallclock	This field allows an administrator to set the class's default wallclock time set when submitting a job with no specified wallclock.

Fairness

Field	Description
Fairshare Policy	*Fairshare* is a method of enforcing cluster sharing between credentials. A *credential* is a user, group, account, class, or quality of service (QoS). Fairshare tracks each credential's usage for a desired amount of time and decreases a job's start priority if the fairshare policy is violated. By decreasing a job's start priority, the job will wait longer in the queue before it starts, allowing other jobs to execute first. • Fairshare Floor Policy - If the class's cluster usage is below the fairshare target, then the class's start priority for the job increases. The class's cluster usage is the total percentage of the cluster used by the class. • Fairshare Target Policy- If the class's cluster usage is above or below the fairshare target, then the class's start priority for the job increases or decreases accordingly. The class's cluster usage is the total percentage of the cluster used by the class. • Fairshare Cap Policy - If the class's cluster usage is above the fairshare target, then the class's start priority for the job decreases. The class's cluster usage is the total percentage of the cluster used by the class. • Absolute Fairshare Policy - If a class's cluster usage exceeds the fairshare target, then the class's start priority for the job decreases. The class's cluster usage is the total number of processor hours used by a class. • Relative Fairshare Policy - If a class's cluster usage exceeds the fairshare target, then the class's start priority for the job decreases. The class's cluster usage is the total percentage of the cluster used by the class.
Fairshare Target	This field allows an administrator to define the fairshare target for this class. Refer to the Fairshare Policy for an understanding of how fairshare target will be used.

Priority	This field allows an administrator to define a class's job priority. A class's job priority will increase or decrease the start priority of this class's jobs. Workload Manager, with some exceptions, will start the jobs with the highest start priority first.

Comments

Field	Description
Comments	This field allows an administrator to enter comments regarding the class.

Usage Statistics

This is only visible if a profile is being modified.

Field	Description
System Utilization	Shows what percent of the entire system has recently been dedicated to this credential. Each value in this line graph represents the number of dedicated processor seconds used by this credential for the profiling duration divided by the number of dedicated processor seconds in the entire system for the profiling duration. The number of dedicated processor seconds for the entire system for a profiling duration is equal to the size of the profiling duration in seconds multiplied by the number of processors.

4.9. Create / Modify a Quality of Service Profile

Summary

Quality of services are created by the operating system while quality of service profiles are created by Workload Manager. When a quality of service submits a job then that quality of service becomes visible to Workload Manager and at that moment a credential profile is automatically created for the quality of service.

Credential Access

Field	Description
Quality of service Name	This field allows an administrator to define the name of the quality of service. Usually this is the login name for the quality of service.
User Access List	This field allows an administrator to define which users can access this quality of service.
Group Access List	The group access is defined by the operating system and cannot be defined by Workload Manager.
Account Access List	This field allows an administrator to define which accounts this quality of service can access.
Default Account	This field allows an administrator to define which accounts will automatically be used if the quality of service doesn't specify an account.

Partitions & Reservations

Field	Description
Partition	This field allows an administrator to define which partitions this quality of service can access
Default Partition	This field allows an administrator to define which partitions this user can access by default.
Required Reservation	This field allows an administrator to define which reservations jobs that access this quality of service must use.

Fairness

Field	Description
Fairshare Policy	Fairshare is a method of enforcing cluster sharing between credentials. A *credential* is a user, groups, account, class/queue, or quality of service (QoS). It consists of tracking each credentials usage for a desired amount of time and decreasing a jobs start priority if the *fairshare policy* is violated. By decreasing

	a job's start priority the job will wait longer in the queue before it starts allowing other jobs to execute first. • Fairshare Floor Policy - If the quality of services cluster usage is below the *fairshare target* the quality of services start priority for the job increases. The quality of services cluster usage is the total percentage of the cluster used by the quality of service. • Fairshare Target Policy- If the quality of services cluster usage is above or below the *fairshare target,* the quality of services start priority for the job increases or decreases accordingly. The quality of services cluster usage is the total percentage of the cluster used by the quality of service. • Fairshare Cap Policy - If the quality of services cluster usage is above the *fairshare target,* the quality of services start priority for the job decreases. The quality of services cluster usage is the total percentage of the cluster used by the quality of service. • Absolute Fairshare Policy - If a quality of services cluster usage exceeds the *fairshare target,* the quality of services start priority for the job decreases. The quality of services cluster usage is the total number of processor hours used by a quality of service. • Relative Fairshare Policy - If a quality of services cluster usage exceeds the *fairshare target*, then the quality of services start priority for the job decreases. The quality of services cluster usage is the total percentage of the cluster used by the quality of service.
Fairshare Target	This field allows an administrator to define the *fairshare target* for this quality of service. Refer to the *Fairshare Policy* for an understanding of how *fairshare target* will be used.

General Attributes

Field	Description
Comments	This field allows an administrator to enter comments regarding the quality of service.

Job Usage Limits

Field	Description
Idle Maximum Jobs	This field allows an administrator to set the quality of services maximum number of simultaneously idle jobs.
Maximum Jobs	This field allows an administrator to set the quality of services maximum number of simultaneously executing jobs.
Maximum Procs	This field allows an administrator to set the quality of services maximum number of simultaneously utilized processors.
Maximum Proc Seconds	This field allows an administrator to set the quality of services maximum number of simultaneously utilized processor seconds. *Processor seconds* is the number of processors utilized times the number of seconds they are utilized.
Maximum Nodes	This field allows an administrator to set the quality of services maximum number of simultaneously utilized nodes. A *node* is a computer consisting of 1

	or more processors
Maximum Wallclock	This field allows an administrator to set the quality of services maximum wallclock time available when submitting a job.
Minimum Wallclock	This field allows an administrator to set the quality of services minimum wallclock time when submitting a job.
Default Wallclock	This field allows an administrator to set the quality of services default wallclock time set when submitting a job with no specified wallclock.

Credits & Charging

Field	Description
Dedicated Cost	The cost to have dedicated access to this resource, regardless of whether it is being utilized.
Utilized Cost	The cost to utilize this resource.

Prioritization

Field	Description
Queue Time Weight	This field displays the quality of service weight factor. If an idle job is submitted to this quality of service the number of minutes that it has been in the queue will be multiplied by this value. This increases the jobs start priority.
Expansion Factor Weight	This field displays the quality of service weight factor. If an idle job is submitted to this quality of service it's expansion factor is multiplied by this value. This increases the job's start priority.

Quality of Service Flags

Flag Name	Description
Deadline	This option makes any job with this quality of service with a completion deadline to finish before that deadline.
Dedicated	This option makes any job submitted to this quality of service to require a dedicated node. A dedicated node is a node that is completely reserved for only one job.
Enable User Reservation	This option makes any user that is a member of this quality of service able to create user/personal reservations.
Ignore All Policies	This option makes any job submitted to this quality of service exempt from all resource usage policies.
No Backfill	This option makes any job submitted to this quality of service exempt from the backfill algorithm.
No Reservation	This option makes any job submitted to this quality of service unable to create a job reservation and, therefore, only able to share resources.

Next To Run	This option makes any job submitted to this quality of service run next. This is accomplished by increasing the start priority of a job, than all of the other queued jobs.
Preemptee	This option makes any job submitted to this quality of service preemptable. A preemptable job can be stopped and requeued if a high priority preemptor job needs to execute.
Preemptor	This option makes any job submitted to this quality of service able to preempt any preemptable job.
Provision	This option makes any job submitted to this quality of service that requests unavailable resources such as an operating system or software to have Workload Manager setup a number of nodes with the correct resources.
Reserve Always	This option makes any job submitted to this quality of service create a reservation. Usually job reservations are created when the job starts but with this option enabled the job will create a reservation immediately.

Usage Statistics

This is only visible if a profile is being modified.

Field	Description
System Utilization	Shows what percent of the entire system has recently been dedicated to this credential. Each value in this line graph represents the number of dedicated processor seconds used by this credential during the profiling duration divided by the number of dedicated processor seconds in the entire system during the profiling duration. The number of dedicated processor seconds for the entire system for a profiling duration is equal to the size of the profiling duration in seconds multiplied by the number of processors.

Chapter 5.0 Policies

5.1. Policies Overview

Moab Workload Manager has many powerful policies that can be managed to effectively handle as large a workload as possible while satisfying other desires. The policies section is intended to give control over Workload Manager's various policies.

5.2. Fairshare

Summary

Fairshare allows the cluster to be shared between different individuals and/or organizations without allowing any individual or organization the ability to monopolize the cluster. This is achieved by tracking how the cluster is used over time by each credential or user, group, class, account, and quality of service (QoS) and by increasing or decreasing the start priorities of jobs waiting to execute. It should be noted that the start priority is used by Workload Manager to decide which jobs get executed first. The higher the start priority, the sooner a job will execute. The information collected about each credential is inserted into what is called a usage window. The amount of time tracked in a window is defined by the system administrator.

System administrators often cannot achieve the cluster sharing they desire without using multiple usage windows. Multiple usage windows allow Workload Manager to balance cluster usage differently by making the most recent window more important than more outdated windows. This is achieved by using the Usage Window Decay Factor. Essentially, the lower the decay factor, the less important outdated usage windows are.

5.2.1. Fairshare Options

- Interval Length - This field allows an administrator to define how long each window lasts.
- Depth - This field allows an administrator to define how many windows should exist.
- Usage Metric - This field allows an administrator to define how credential usage is tracked.
    1. Dedicated PES - This field tracks credential usage according to the number of processor equivalent seconds reserved for a job.
    2. Dedicated PS - This field tracks credential usage according to the number of processor seconds reserved for a job.
    3. PDEDICATEDPS - This field tracks dedicated processor seconds scaled by the processor speed of the node.
    4. SDEDICATEDPES - This field tracks dedicated processor-equivalent seconds scaled by the speed of the node.
    5. Utilized PS - This field tracks credential usage according to the number of processor seconds used by a job.
- Decay Factor - This field allows an administrator to define how much of an influence outdated usage windows have in calculating a job's start priority. The lower the Usage Window Decay Factor, the less important the outdated usage windows.

5.2.2. Fairshare Table

- Credential Type - This field displays the credential type (user, group, account, class, or quality of service (QoS)).
- Credential ID - This field displays the credential's unique name.
- Fairshare Credentials Policy - This field displays the fairshare credentials policy. Consult the create user, group, class, account, or quality of service (QoS) documentation for more information regarding the policies.
- Fairshare Credentials Target - This field displays the fairshare credentials policy.

Consult the create user, group, class, account, or quality of service (QoS) documentation for more information regarding the targets.

- Percentage Cluster Usage - This field displays the percentage of the cluster that has been used by this credential ID in comparison to the other credential IDs for this Credential Type.
- Current Interval - This field displays the first usage window. The decay factor does not affect this window at all.
- Interval 1 - 31 - This field displays the usage interval windows 1 through 31. The decay factor affects these windows with the most outdated window being window 31 and the most recent window being window 1.

5.3. Priority

Summary:

A job has one start priority that determines when a job will start. The higher the start priority, the sooner a job will start. The job start priority can be anywhere between 1,000,000,000 and -1,000,000,000.

How to read priorities.

A job has one start priority that determines when a job will start. The higher the start priority, the sooner a job starts. Workload Manager uses the priority policies to calculate a job's start priority.

A subcomponent priority of 0 means the subcomponent will be ignored. A positive subcomponent priority means the start priority will be increased. A negative subcomponent priority means the start priority will be decreased. Refer to the documentation below for information about Main Components.

How to understand the priority window layout.

Workload Manager uses 39 components to calculate the start priority. These components are grouped into tabs according to their functionality. The Main Component is different from the subcomponents. Refer to the documentation below for further information about Main Components.

The table shows only idle/queued jobs and their start priority. Only idle/queued jobs are displayed because priority policies do not affect running jobs.

What are components and subcomponents?

The 7 component groupings are crucial to understanding priorities. The Main Component tab is used only to increase or decrease the subcomponents priorities. The subcomponents increase or decrease the job start priority.

How the job start priority is calculated.

A job has one start priority which is used to decide when a job will start. The higher the start priority the sooner a job will start. Workload Manager uses the priority policies to calculate a job's start priority.

The start priority is calculated by adding all the subcomponents in a group together and multiplying the total of these subcomponents by the Main Component priority. This process is repeated 7 times, or once for each main component. The start priority is a summation of these 7 totals.

It should be noted that if the Main Component priority is set to 0, all of the subcomponent priorities for that Main Component will be ignored.

How to enable a priority

To enable a priority, two priorities must change. The first priority is the sub component

priority and the second priority is the Main Component priority.

For example, to apply a priority of 1 for a user's priority, the user priority in the subcomponent credential's priorities and the Credential Priorities in the Main Component must both be set to 1.

What does a 0 mean?

If the Main Component priority is set to 0, all of the subcomponent priorities for that Main Component will be ignored. A sub component priority of 0 means the sub component will be ignored.

5.3.1. Main Priority Components

- Wait Time Job Services - This field allows an administrator to increase or decrease all of the Wait Time Job Services priorities. If this is set to 0, all of the subcomponent's priorities for Wait Time Job Services are ignored.
- Wait Time Job Targets - This field allows an administrator to increase or decrease all of the Wait Time Job Target's priorities. If this is set to 0, all of the subcomponent's priorities for Wait Time Job Targets will be ignored.
- Fairshare Usage - This field allows an administrator to increase or decrease all of the Fairshare Usage priorities. If this is set to 0, all of the subcomponent's priorities for Fairshare Usage will be ignored.
- Resource Requests - This field allows an administrator to increase or decrease all of the Resource Request's priorities. If this is set to 0, all of the subcomponent's priorities for Resource Request's will be ignored.
- Credential Priorities - This field allows an administrator to increase or decrease all of the Credential Priority's priorities. If this is set to 0, all of the subcomponent's priorities for Credential Priorities will be ignored.
- Job Attributes - This field allows an administrator to increase or decrease all of the Job Attributes priorities. If this is set to 0, all of the subcomponent's priorities for Job Attributes will be ignored.
- Executing Job Usage - This field allows an administrator to increase or decrease all of the Executing Job Usage priorities. If this is set to 0, all of the subcomponent's priorities for Executing Job Usage will be ignored. Unlike the other components, this component only effects executing jobs and is only applicable when preemption is used.

5.3.2. Wait Time Job Services

- Queue Time - This field allows an administrator to set the priority of a job according to the minutes the job has waited in the queue.
- Expansion Factor (X-Factor) - This field allows an administrator to set the priority of a job according to the expansion factor* of the job.
- Policy Violation - This field allows an administrator to set the priority of a job according to whether the job has violated a usage limit. If the job has violated a usage limit, the job is assigned a policy violation value of 1; otherwise, the job is assigned a policy violation value of 0.
- By Pass - This field allows an administrator to set the priority of a job according to the number of other jobs that started execution before this job. The other jobs are only counted if Workload Manager started the other jobs because of a backfill policy.

5.3.3. QoS Targets

- Queue Time - This field allows an administrator to set the priority of a job according to Quality of Service queue time target. The closer the job is to this target, the higher the Queue Time value.
- Expansion Factor (X-Factor) - This field allows an administrator to set the priority of a job according to Quality of Service expansion factor target. The closer the job is to this target, the higher the Expansion Factor value.

5.3.4. Credential Priority

- User - This field allows an administrator to set the priority of a job according the user's priority.
- Group - This field allows an administrator to set the priority of a job according the group's priority.
- Account - This field allows an administrator to set the priority of a job according the account's priority.
- Class - This field allows an administrator to set the priority of a job according the class's priority.
- QoS - This field allows an administrator to set the priority of a job according the Quality of Service (QoS)'s priority.

5.3.5. Job Attribute/State Priority

- Job Attribute - This field allows an administrator to set the priority of a job according a job's attributes. Refer to the Workload Manager Priority Factors documentation for information on how to set the Job Attributes.
- Job State - This field allows an administrator to set the priority of a job according a job's state.

5.3.6. Fairshare Usage Priority

- User - This field allows an administrator to set the user's priority of a job according to the fairshare usage.
- Group - This field allows an administrator to set the group's priority of a job according to the fairshare usage.
- Account - This field allows an administrator to set the account's priority of a job according to the fairshare usage.
- Class - This field allows an administrator to set the class's priority of a job according to the fairshare usage.
- Quality of Service (QoS) - This field allows an administrator to set the qualities of service (QoS)'s priority of a job according to the fairshare usage.
- Jobs Per User - This field allows an administrator to set the priority of a job according to the number of jobs currently executing for this user.
- Processor Seconds - This field allows an administrator to set the priority of a job according to the number of processor seconds currently being used by this user.
- Processors Per User - This field allows an administrator to set the priority of a job according to the number of processors currently being used by this user.

5.3.7. Resource Requests Priority

- Node - This field allows an administrator to set the priority of a job according to the total number of nodes requested by the job. The more nodes requested, the higher the Node value.
- Disk - This field allows an administrator to set the priority of a job according to the total amount of disk space requested by the job. The more disk space requested, the higher the disk value.
- Memory - This field allows an administrator to set the priority of a job according to the total amount of memory in megabytes requested by the job. The more memory requested, the higher the Memory value.
- Swap - This field allows an administrator to set the priority of a job according to the total amount of swap in megabytes requested by the job. The more swap requested, the higher the swap value.
- Processor-Equivalent - This field allows an administrator to set the priority of a job according to the total number of processor equivalents requested by the job. The more processor equivalents requested, the higher the processor equivalent's value.
- Walltime - This field allows an administrator to set the priority of a job according to the total amount of walltime seconds requested by the job. The more walltime requested, the higher the Walltime value.

5.3.8. Executing Job Usage Priority

- Consumed - This field allows an administrator to set the priority of a job according to the total number of processor seconds it has consumed. Unlike other components, this component only effects executing jobs and is only applicable when preemption is used.
- Hunger - This field allows an administrator to set the priority of a job according to the total number of processors needed to balance a dynamic job's performance. This field is only applied to dynamic jobs.
- Remaining - This field allows an administrator to set the priority of a job according to the total number of processor seconds it has remaining. Unlike other components this component only effects executing jobs and is only applicable when preemption is used.
- Percentage Consumed - This field allows an administrator to set the priority of a job according to the percentage of the Wallclock that has been consumed. Unlike other components, this component only affects executing jobs and is only applicable when preemption is used.

5.3.9. Priority Display Options

- Display start priority pie chart - This option displays a pie chart of the priority components. Negative components are not displayed.
- Display start priority bar graph - This option displays a bar chart of the priority components. Positive and negative components are displayed.

5.4. Job Policies and Settings

Contained in this section:

5.4.1. Job Limit Settings

Summary

This window is used to place system wide restrictions on jobs.

- Maximum Jobs Started per Iteration - This field allows an administrator to specify the maximum number of times Workload Manager will attempt to start the job.
- Maximum Jobs Preempted per Iteration- This field allows an administrator to define the maximum number of times a job can be preempted by Workload Manager for higher priority jobs.
- Max Processors per Job - This field allows an administrator to define the maximum number of processors that can be requested by any jobs.
- Maximum Wallclock - This field allows an administrator to specify the maximum amount of Wallclock time that can be requested by any single job.
- Wallclock Violation Soft Limit - This field allows an administrator to define the amount of time Workload Manager allows a job to exceed its wallclock limit before Moab sends mail to the primary administrator.
- Wallclock Violation Limit - This field allows an administrator to define the amount of time Workload Manager allows a job to exceed its wallclock limit before it is terminated.
- Wallclock Violation Action - This field allows an administrator to define whether Moab should cancel or preempt a job that has exceeded its wallclock limit.

5.4.2. Job Defer Settings

Summary

- Defer Wait Time - This field allows an administrator to define the amount of time a job will be held in the deferred state before being released back to the idle job queue.
- Starts Before Defer - This field allows an administrator to define the amount of time a job will be allowed to fail in its start attempts before being deferred.
- Defers Before Hold - This field allows an administrator to define the number of times a job can be deferred before it will be placed in batch hold.
- Job Sync Time - This field allows an administrator to define the length of time after which Workload Manager will change a job's expected state to an unexpected reported state. It should be noted that Workload Manager will not allow a job to run as long as its expected state does not match the state reported by the resource manager.

5.4.3. Global Job Policy Settings

Summary

This window contains job-specific global settings.

- Preempt Policy - This field allows an administrator to specify how Moab should respond to a preempted job.
    1. CANCEL - The job is terminated and removed from the queue.
    2. CHECKPOINT - The job is terminated, checkpointed, and left in the queue.
    3. REQUEUE - The job is terminated and left in the queue.
    4. SUSPEND - The active job is suspended.
- Node Match Policy - This field allows an administrator to specify how compute nodes are selected.
    1. EXACTNODE - Moab selects as many nodes as requested, regardless of how many tasks it assigns to the same node.
    2. EXACTPROC - Moab selects nodes with exactly the same number of processors configured as are requested per node, even if other nodes have more processors available.
- Job Priority Policy - This field allows an administrator to specify how Moab should track the dynamic aspects of a job's priority.
    1. ACCRUE - The job will accrue queuetime based priority from the time it is submitted unless it violates any of the policies not specified in JOBPRIOEXCEPTIONS.
    2. RESET - The job will accrue priority from the time it is submitted unless it violates any of the JOBPRIOEXCEPTIONS. If the job does violate JOBPRIOEXCEPTIONS, its queuetime based priority will be reset to 0.
- Use Machine Speed - By checking this box, an administrator specifies that a job's wallclock should be increased if the job is executing on a slower node and that a job's wallclock should be decreased if the job is executing on a faster node. The speed of the node is assessed by examining the node speed option located in the list nodes window.
- Enable Multiple Node Jobs - By checking this box, an administrator specifies that resources may be allocated from multiple nodes to a job.
- Enable Multiple Requirement Jobs - By checking this box, an administrator specifies that jobs may have more than one requirement.

5.5. Reservation Policies

Summary

Workload Manager uses reservations to guarantee that a specific amount of resources will be available for a given job or set of users at a particular time. For example, Workload Manager can reserve 20 processors and 10 GB of memory for users Bob and John from Friday 6:00 AM to Saturday 10:00 PM. Workload Manager uses reservations internally to manage backfill, protect job resources, allow service guarantees, support deadlines and QoS, and enable grid scheduling. Workload Manager supports infinite, reoccurring, and one-time reservations.

When backfill is enabled, Workload Manager attempts to schedule lower-priority jobs ahead of a higher-priority job that can't start immediately. In order to ensure that those low-priority jobs don't delay the high-priority job's start time, Workload Manager can reserve the resources needed by the high-priority job. These are called *priority reservations*. The reservation policy determines how Workload Manager handles priority reservations.

Reservation Settings

Field	Description
Policy	This is the policy Workload Manager uses when creating priority reservations. These reservations protect the resources a job is using until the job completes. • *CurrentHighest* - Existent priority reservations will be relinquished to new jobs with higher priority • *Highest* - All idle jobs that receive a reservation will keep it until they run even if new jobs are higher priority • *Never* - No idle jobs receive reservations
Depth	This is how many priority reservations Workload Manager will create. A higher value protects the start time of high-priority jobs but may decrease backfill efficiency.
Reservations Per Node	This is the maximum number of priority reservations that can be created on any single node. On large SMP systems, this value should be set to approximately twice the number of reservations that exist on the system.
Retry Time	This is the period of time Workload Manager attempts to restart a job with a priority reservation that originally failed to start.
Creation Policy	This determines which users can create one-time reservations (also called *Administrative* reservations). This setting is unrelated to priority reservations.

5.6. Node Policies

Summary

Node Task Allocation

A *task* is a request for resources that must exist on a single compute node. Each job may have one or more tasks. Workload Manager allocates resources to jobs based on the tasks in the job. This is useful because nodes with multiple processors are usually able to support more than one task at a time. For example, if a job has 2 tasks where each task requires 1 processor and 256 MB of memory, Workload Manager may choose to allocate the job to a dual processor node with 512 MB of memory or to 2 single processor nodes with 256 MB of memory each. The node-task allocation policy determines which tasks may run on the same node.

Field	Field Information
Policy	This is the algorithm Workload Manager uses to determine which tasks may run on the same node. • Shared optional - Tasks from any job and any user may use available resources on any node. • Shared only - Tasks from any job and any user may use available resources on any node. • Single Job - Only tasks from the same job may run for any given node. • Single Task - Only one task may run on each node. • Single User - For any given node, only tasks from jobs submitted by the same user may run. • Unique User - Any number of tasks from a job may allocate resources from a node if the user has no other jobs running on that node.

Node Allocation Policy

Node allocation is the process of selecting the best resources, from a list of available resources, to assign to a job. Making this decision intelligently is important in environments with heterogeneous resources or nodes that can support multiple jobs at the same time.

Field	Field Information
Policy	This is the algorithm Workload Manager uses to allocate nodes. • FirstAvailable - Nodes are allocated in the order they are presented by the resource manager. • LastAvailable - Resources are selected so as to minimize the amount of time the resources remain unused after the job completes. This minimizes node-time fragmentation and is useful in systems that have a large number of reservations. • MinResource - Nodes that have the smallest amount of resources that meet the job's requirements are selected. • CPULoad - Nodes that have the maximum amount of available, unused CPU power are selected. This is good for timesharing systems, but is only

	applicable to jobs starting immediately. For future jobs, the MinResource policy is used. • Contiguous - Nodes are allocated in contiguous (linear) blocks. This is required by the Compaq RMS system. • MaxBalance - Nodes that are as similar as possible to each other are allocated to each job. The most important consideration in determining node similarity is node speed. • Fastest - The fastest available nodes are allocated to each job. Workload Manager determines which nodes are fastest based upon first the node speed and then the processor speed of each node. If neither of these values is available, the nodes are selected randomly.

Node Availability Policy

Workload Manager will start jobs on nodes that are not full and are not considered busy. Workload Manager considers a node busy according to which Node Availability Policy is set.

Field	Field Information
Policy	This is the algorithm Workload Manager uses to determine if a node is busy. • Combined - Either of the above two conditions is met. • Dedicated - The dedicated (assigned or reserved) resources on the node equal or exceed the configured resources. • Utilized - The utilized (in use) resources on the node equal the configured (total available) resources. Resources may be dedicated to a user, group, or account for a specific period of time. But some of those resources may not be used during the entire period. This setting allows Workload Manager to differentiate between the two possibilities.

5.7. Partition Policies

Summary:

This section deals with policies relating partitions and their behavior. Below is a list of partition policies.

5.7.1. Partition Allocation Policy

This is a direct way to assign a peer allocation algorithm when multiple partitions are available for a job. Because clusters are considered partitions, this defines how jobs can be migrated to remote resources if multiple remote clusters can be found. Values and their descriptions are listed in the table below.

- BESTFIT - Allocate resources from the eligible peer with the fewest available resources - measured in tasks (minimizes fragmentation of large resource blocks).
- BESTPFIT - Allocate resources from the eligible peer with the fewest available resources - measured in percent of configured resources (minimizes fragmentation of large resource blocks).
- FIRSTCOMPLETION - Allocate resources from the eligible peer which can complete the job the soonest (takes into account data staging time and job-specific machine speed.
- FIRSTSTART - Allocate resources from the eligible peer which can start the job the soonest.
- ROUNDROBIN - Allocate resources from the eligible peer which has been least recently allocated.
- LOADBALANCE - Allocate resources from the eligible peer with the most available resources - measured in tasks (balances workload distribution across potential peers).
- LOADBALANCEP - Allocate resources from the eligible peer with the most available resources - measured in percent of configured resources (balances workload distribution across potential peers).

5.8. Backfill

Summary

Backfill is an optimization policy that allows a scheduler to make better use of available resources by running jobs out of order. When using Backfill, Workload Manager prioritizes the jobs in the queue into a sorted list with the highest priority job first. Beginning at the top of the list, it starts the jobs one by one until it reaches a job that it cannot start because the necessary resources are not available. Using the start times and wallclock limits of the currently running jobs, Workload Manager then calculates when it will be able to start the job. It reserves that spot (in the future) for the job, and attempts to schedule some of the remaining lower-priority jobs in the gaps left over from the higher-priority jobs. This process continues until Workload Manager has attempted to start all the jobs in the list, until all resources are consumed, or until Workload Manager has considered a specific number of jobs. Backfill allows Workload Manager to achieve a higher utilization than would be otherwise possible, while remaining mostly true to the original job priorities.

Backfill Settings

Field	Field Information
Policy	In this field you specify the kind of backfill algorithm Workload Manager uses to schedule jobs. • *FirstFit* - Considers jobs in the queue sequentially, beginning with the highest priority and moving down the list. • *BestFit* - Considers all jobs in the queue and selects the job that best fits the available resources (see the *Attribute* parameter below). • *Greedy* - Considers all possible combinations of jobs that can run on the available resources and selects the best combination (see the *Attribute* parameter below). • *Optimistic* - If any resource manager reports a state of up, that state will be used. • *Preempt* - Allows the scheduler to start backfill jobs even if the required walltime is not available. • *None* - Backfill is not enabled.
Metric	This is the criteria used by the backfill algorithm to determine the best jobs to backfill. For example, if *Procs* is selected, a job that requires the exact amount of available processors will be considered the best. This parameter only applies to the *BestFit* and *Greedy* backfill policies. • *Procs* - This is the number of processors. • *ProcSeconds* - This is the number of processors multiplied by the duration of the job, in seconds. • *Nodes* - This is the number of nodes. • *Seconds* - This is the duration (or wallclock time) of the job, in seconds.
Depth	This is the number of jobs in the queue Workload Manager should consider for backfill. By default, all jobs are considered. If *Depth* is set, Workload Manager

	will only consider that number of jobs for backfill scheduling. For example, if there are idle jobs in the queue and *Depth* is set to 10, only 10 jobs would be considered for backfill. If there are fewer than 10 jobs in the queue, all will be considered. Setting this number higher will result in a higher utilization and better turn-around times, especially for smaller jobs, but may result in low-priority jobs being started before medium-priority jobs. This parameter should be tuned for your specific situation.
Priority Policy	• DURATION - The length of the jobs determines which backfill job to preempt. • HWDURATION • NONE - No priority policy is used. • RANDOM
Chunking Duration	This field allows an administrator to specify the duration during which freed resources will be aggregated for use by larger jobs.
Chunk Size	This field allows an administrator to specify the minimum job size which can utilize chunked resources.
Virtual Walltime Scaling Factor	This field allows an administrator to specify the factor by which eligible jobs' wallclock time is virtually scaled.
Minimum Virtual Walltime	This field allows an administrator to specify the minimum job wallclock time for virtual scaling.

5.9. Green Computing

5.9.1. Live Dashboard

Shows power usage for the current day and month and estimates the cost of that power. Estimates how much the power would cost assuming the current consumption stays constant throughout the month.

Summary

- Current Power Usage - The number of kilowatts currently being used by all nodes in the cluster.
- Month To Date Power Usage - The amount of kilowatt hours that have been used from the beginning of the month until the present.
- Projected Power Cost This Month - The projected cost of the kilowatt hours used for the month if the power usage remains consistent with what it has been from the beginning of the month until know.

Energy Usage

- Low kW Usage since Midnight - The least amount of power (in kilowatts) being used by the entire cluster at any one sampling since midnight this morning.
- Peak kW Usage since Midnight - The highest amount of power (in kilowatts) being used by the entire cluster at any one sampling since midnight this morning.
- Peak kW Usage This Month - The highest amount of power (in kilowatts) being used by the entire cluster at any one sampling since midnight on the first day of the current month.

Energy Usage

- kWh Used Since Midnight - The amount of power in kilowatt hours used by the entire cluster since midnight this morning.
- Daily Average kWh Usage - The average amount of power used per day since the beginning of the month in kilowatt hours.

Power Usage over Time

A line graph showing power usage (in kilowatts) for the entire cluster over time.

5.9.2. Green Power Settings

On Demand Settings

- Node Idle Power Threshold - The maximum number of nodes that can be idle and stay powered on when green power management is on.
- Green Standby Pool Size - The number of nodes to keep powered on in the standby pool.
- Node Power On Duration - The estimated amount of time it will take to power on a node.

Charts and Graphs

- Metric Type - The type of metric to generate statistics about.
- Chart Data Category Type - The type of usage that will be represented in the statistics. The estimated amount of time it will take to power on a node.
- Chart Data Organized By - How the statistics are organized.
- Start Time - The beginning of the time period to be represented in the statistics.
- End Time - The end of the time period to be represented in the statistics.

Cost per kWh

The cost per kilowatt hour. This is used in savings and cost estimations in the green charts form.

Chapter 6.0 Statistics

6.1. Statistics Overview

Cluster Manager offers a wide assortment of customizable statistics whether they be quick charts, customized charts, graphs, and reports, or estimation matrix statistics.

6.2. Quick Charts/Graphs

Quick Charts provides a simple interface for viewing the most common statistics. Statistics are gathered from the first day of the current month to the last day of the current day, week, or month.

Available Charts

1. System Overview
2. Total Processor Hours Per Account
3. Queue Time Per Account
4. Total Processor Hours Per User
5. Resource Requests Per User
6. Queue Time Per Quality of Service (QoS)

6.3. Matrix Statistics

Summary

Matrix statistics are used both to analyze historic workload and to predict future workload. The left column of the table displays different job processor sizes. The top column displays relative time frames in the format Hours:Minutes:Seconds.

For example, the table below would be understood as follows. For jobs using 1 processor, 12 jobs are completed within the first 15 minutes of execution and 8 jobs are completed after the first 15 minutes and before the first hour of execution. For jobs using 4 processors, 5 jobs are completed within the first 15 minutes of execution and 0 jobs are completed after the first 15 minutes and before the first hour of execution.

Total Completed Jobs	00:15:00	01:00:00
1 Processor	12	8
4 Processors	5	-

Matrix Statistics Types

- Estimated Start Time - This field displays the predicted start time of a submitted job according to the number of processors the job would use. This information can help users determine how many processors they should submit a job to for optimal start time. For example, it may take less time to start a four-hour job submitted to four processors, than to one processor for a sixteen-hour job.
- Average Expansion Factor - This field displays the historic average expansion factor of a job according to the number of processors it used. The expansion factor is calculated using the following equation: ((queue time of a job + job's duration) / job duration).
- Maximum Expansion Factor - This field displays the historic maximum job expansion factor of a job according to the number of processors it used. The expansion factor is calculated using the following equation: ((queue time of a job + job's duration) / job duration).
- Average Queue Time - This field displays the historic average wait time before a job starts executing, according to the number of processors it used. Queue time is the number of hours a job waited before it began execution.
- Average Bypass - This field displays the historic average by pass of a job according to the number of processors it used. By pass is the number of jobs that started execution before this job because of backfill policies. This is useful in recognizing which types of jobs are being by passed by backfill.
- Maximum Bypass - This field displays the historic maximum by pass of a job according to the number of processors it used. Bypass is the number of jobs that started execution before this job because of backfill policies. This is useful in recognizing which type of jobs are being by passed by backfill.
- Total Completed Jobs - This field displays the total number of jobs that completed in the time interval according to the number of processors they used.
- Cluster Processor Hours Requested (%) - This field displays a breakdown of the requested time on the cluster according to the number of processor hours. Processor hours are the number of processors times the number of hours that they were requested. The sum total of the table is 100%. Each cell inside the table gives the

percentage of the total cluster processor hours requested by jobs of that size and duration.

- Cluster Processor Hours Utilized (%) - This field displays a breakdown of the utilized time on the cluster according to the number of processor hours. Processor hours are the number of processors times the number of hours that they were utilized. The sum total of the entire table is 100%. Each cell inside the table gives the percentage of the total cluster processor hours utilized by jobs of that size and duration.
- Wallclock Accuracy (%) - This field displays the average Wallclock accuracy, or user estimate accuracy of how long a job would execute, according to the number of processors it used. A value greater than 100 indicates the average user overestimates the job Wallclock time. A value less than 100 indicates the average user underestimates the job wallclock time. A value of 100 indicates the average user estimates the job wallclock accurately.
- Backfill Count (%) - This field displays the percentage of jobs that were delayed in executing because the backfill policy made them execute later.
- Backfill Processor Hours Utilized (%) - This field displays the percentage of processor hours for jobs that were delayed in executing because of the backfill policy that later executed.
- Job Efficiency (%) - This field displays the average percentage of the CPU that jobs used according to the number of processors of each job.
- Quality of Service (QoS) Delivered (%) - This field displays the average percentage of jobs that received their desired quality of service (QoS) according to the number of processors they used.

6.4. Custom Charts/Graphs

Summary

This window allows you to create charts and graphs showing statistics over a custom time interval. You can create pie, bar, line, and stacked line charts and, if desired, export those charts as PNG files. To export a chart, right click it and select "save as" from the menu. Zoom in on sections of the line charts by using the mouse to drag a box over the area of interest. The domain and range axes will automatically be adjusted. To return to the original zoomed out view, hold down the left mouse button, drag towards the left, and release the mouse button.

Note that, since charts with large time ranges are sampled for efficiency, the zoomed-in chart may not be as detailed as directly creating a chart for the zoomed-in time range. For example, you might create a line chart for all of 2008 and zoom in on September 1st. The zoomed-in portion of this chart will be less detailed than if you specified the time parameters as 12:00 am Sept 1 to 11:59 pm Sept. 1st when initially creating the chart.

6.4.1. Credential Based Charts

Select Credentials

The drop down box allows you to select which credential for which Moab Cluster Manager will calculate statistics or if the cluster-wide statistics should be used. The Display All Credentials option displays all the credentials that have been tracked by the statistics regardless of whether they have any activity recorded. The Display Credentials with Statistics option displays only the credentials that have recorded some type of activity. The Display Listed Credentials option displays only the requested credentials.

Select Metric

One can select from the following metrics when creating a chart.

- Backlog Hours**** - The number of processor hours of the jobs belonging to this credential that are waiting in the queue.
- Dedicated Processor Hours - The number of processor hours dedicated to jobs run by this credential, regardless of whether those jobs are actually using those processors.
- Executed Jobs - The number of jobs belonging to this credential that have finished execution. If a line graph is selected, each y value is the number of jobs that completed during the profiling interval around the value on the x axis.
- Jobs That Met QOS Target** - The percentage of completed jobs that met their QOS target to total jobs.
- Queue Hours** - The number of hours a credential's jobs waited in the queue before starting. Note that queue hours are floored to the closest lower integer value. For example, 23.8 would be floored to 23. This also means that if there is only .9 queue hours then 0 will be reported.
- Resource Requests - Causes three fields to be charted - processor-seconds, memory, and wallclock per credentials that either exceeded the requested resource or under-utilized the requested resource.
- System Utilization* - This field displays the system utilization percentage and should range between 0% and 100%. Each value is calculated as dedicated processor hours

for that profiling interval divided by the total number of processor hours available on that system in that profiling interval (times 100). For example, if the profiling interval was 1 hour, there were 10 total processors, and 5 of those processors were in use for the entire profiling interval, the calculation would be (5 proc hours/10 procs*1 hour)*100=50%. This makes sense intuitively as half the processors were in use during the profiling interval.

- Utilized Processor Hours - The number of processor hours utilized by this credential. Processors that are idle yet are dedicated to the credential are not counted. It is recommended that one use dedicated instead of utilized processor hours where possible since utilized processor hours are dependent on the accuracy of the resource manager, and some resource managers are known to report inaccurate utilized processor hours. Dedicated processor hours are calculated based on data from within Moab Workload Manager and are a more dependable statistic.
- XFactor** (Expansion Factor) - The expansion factor of a job is defined as ((Queue Time + Execution Time) / Requested Wallclock Time).

In addition to the above metrics, one can select any generic metrics that are configured in the system. Generic metric charts are line graphs that show the cumulative generic metric for all nodes dedicated to this credential. For example, if at a certain time user Bob used 3 nodes which each used 10 WATTS, where WATTS was defined as a generic metric, the corresponding point on the line graph for that time would be 30 WATTS.

Because Workload Manager operates by averaging usage across its statistical intervals the values displayed can be misleading. For example, if the statistical interval for Workload Manager was set to 10 minutes for a cluster of 256 processors and one job, which used all 256 processors, was submitted that started and ended in 5 minutes, then an administrator would assume the System Utilization would display 100% of the processors used. However, because the interval was 10 minutes long and the job only ran for 5 minutes, the average System Utilization for the 10 minute interval was 50%.

This field can only be calculated when a job finishes execution.

Requires that Moab is running under a dedicated node model where a node can be running only 1 job at a time.

The accuracy of backlog hour charts is somewhat less precise when a NODEACCESSPOLICY is being used.

Select Calculation Parameters

The pie charts and bar graphs can be customized to display averages, maximums, sums, and/or line graphs, depending on what calculation parameters are appropriate to the usage metric. Some calculation parameters make little sense for certain usage metrics and thus are disabled.

1. If the Average option is selected, the charts will display either the average per job or average per profiling interval, depending on the usage metric. For example, average backlog per profiling interval will show the average amount of backlog in any given profiling interval.
2. If the Max option is selected, the charts will display the maximum value of this usage

metric that was achieved in a profiling interval. For example, the maximum System Utilization in a Profiling Interval will show you the highest system utilization level achieved in a single profiling interval.

3. If the Sum/Total option is selected, the charts will display the sum of the usage metric for all profiling intervals in the selected time window. For example, a chart that shows Total Executed Jobs for April 2008 will display the total number of jobs that executed for each of the selected credentials in April of 2008.
4. If the Line Graph option is selected, one can see a line graph showing the values of the usage metric at each profile interval in the requested time frame.

Chart Type

For formatting, you can choose from Pie Chart, 3D Pie Chart, Bar Graph, 3D Bar Graph, and Line Graph.

Time Frame

Choose a time frame for the graph. Time frames can be chosen from the basis of Month, Week, Day, Hour, or Custom. The Month time frame gathers data from the first of the month to the end of the month. For example, if today was the 13th of October selecting Month would cause a chart to display data from the 1st through the 31st of October. The Week time frame gathers data from midnight Sunday morning of the current week to the current time. The Day time frame gathers data from the midnight this morning to just before midnight tonight. The Hour time frame gathers data starting at the 0th second of the current hour. For example, if the current time was 11:15am selecting Hour would cause a chart to be created with data from 11:00am to the 11:15am. The Custom time frame gathers data from the start time and ends at the end time.

6.4.2. Node Based Charts

Node Categorization Charts

Creates charts that show node categorization over time. For example, one might create a line chart that will show when nodes were in a hardware failure state or create a bar graph to show how much time nodes spent in user reservations. Node categories are a superset of the standard idle, busy, and down states commonly found in batch systems. Some examples include Hardware Failure, Software Failure, Maintenance, etc. For more detailed information, see the Moab Workload Manager Administrator's Guide.

Creating a stacked line graph brings up a chart window that allows one to dynamically make node categories visible or invisible. It shows these node categories in a tree structure where branches corresponding to node categories can be expanded or contracted. As these node categories are expanded or contracted the chart is immediately updated. For example, one might collapse the hardware failure and software failure categories into the parent down time node category.

Why do only 4 of the node categories appear on the chart?

By default, Moab will only categorize nodes into the states idle, active, hardwareFailure, and [NONE]/Other. Usually [NONE]/Other means that Moab was not running. In order to see the other node categories, one must set a reservation on the node indicating the desired category. For example, for node001 to appear as being in the hardware maintenance state one would set a reservation on node001 whose duration matched the

length of the hardware maintenance. Administrators wanting accurate node categorization charts should consider using triggers to set reservations when nodes change state.

Generic Metric Charts

Creates charts that show the generic metric values of nodes over time. For example, if there was a generic metric called WATTS one could create a chart that showed the WATTS value for each in a set of selected nodes. For information on how to configure generic metrics, see the Moab Workload Manager Administrator's Guide.

To create a chart, simply select the desired nodes and generic metric and click the create button. A chart should appear. If there are too many nodes in your system to view in a single chart, the paging buttons will be enabled. Click the Next Page button to see the next 10 nodes.

You can also filter out data sequences which are not of interest. To do this, click the Advanced checkbox at the lower left of the chart. A lower panel should appear allowing you to select nodes that either spike or average above or below a certain threshold. Clicking the Apply Filter button causes this change to be reflected in the chart. For example, one might only choose to see nodes that spike above 80 WATTS. The Aggregate Nodes into One Sequence option allows one to see a single line where each point represents the sum of the values for all selected nodes for the corresponding time. This would allow one to see how the total number of WATTS used by all nodes in the cluster changes over time.

What to do if you see a warning that generic metrics aren't configured

If a chart cannot be created due to a warning informing you that no generic metrics are configured in Moab Workload Manager, you should check to see that your resource manager is returning generic metric information to Moab Workload Manager. Generic Metrics are usually returned to Moab Workload Manager through the CLUSTERQUERYURL configured in your moab.cfg.

The accuracy of generic metric charts are dependent on the resource manager, since this is where Moab gets generic metric information.

6.4.3. Job Template Charts

This window allows one to create charts that show statistics relating to job templates. To create a chart, simply select the desired job templates, usage metric, calculation parameters, chart type, and time frame and click the create button.

Why am I warned that no job templates are configured?

If you cannot create a chart due to a warning that statistics for job templates is not configured in Moab Workload Manager, modify your moab.cfg. Most likely you either have no job templates configured or you did not add the JSTAT attribute. Moab Workload Manager will not collect statistical data for job templates not configured with a JSTAT attribute. For more information on configuring job templates with JSTAT, see the Moab Workload Manager Administrators Guide.

6.5. Command Line Charts/Graphs/Reports

Command Line Charts & Graphs

You can create many of the pie, bar, and line charts available from the credential tab in the Charts & Graphs window using a command line interface, which can be faster than navigating through windows and menus. The command line interface also facilitates scripting. For example, you could create a cron job that generates charts at regular intervals and puts them in a specific directory.

To create charts from the command line, navigate to the scripts directory. There should be a script entitled createchart.sh and a sample properties file called cmdlinechart.cfg. If you are running on Windows, you will want to use createchart.bat instead of createchart.sh. The cmdlinechart.cfg file is a properties file that lets you specify information about what timeframe, credentials, and metric you would like to see in your chart. It includes comments explaining what each property does and has example values. The properties file lets you specify how to connect to Moab Workload Manager. You can use this script to connect to a Moab Workload Manager running either locally or remotely over SSH.

If running on a Unix based operating system you would run the following:

./createchart.sh cmdlinechart.cfg

If running Windows you would use:

createchart.bat cmdlinechart.cfg

Command Line Reports

You can also use the command line to create many of the phone bill style reports available from the Reports window. Similar to command line charts, there is a creatreport.sh script for Unix based operating systems and a createreport.bat script for Windows users. The cmdlinereport.cfg file is a sample properties file that contains parameters on how to connect to Moab Workload Manager.

To create reports on a Unix system:

./createreport.sh cmdlinereport.cfg

If running Windows you would run:

createreport.bat cmdlinereport.cfg

Chapter 7.0 Diagnostics

7.1. Diagnostics Overview

Diagnostics are intended to give an adminstrator quick and easy system information for diagnosing potential problems.

7.2. Diagnostics Support

Summary: The diagnostics support features allows the user to run a set of commands that will check the status of various parts of the system. These commands are controlled by a script named, "support.diag.pl". This should be in your $WORKLOADHOMEDIR/tools directory or wherever your Moab tools directory has been installed.

This window will not work without the support.diag.pl script.

The diagnostics support screen allows the user to select from the tree what commands to run using a built in script. The script will then package the output of each command into a file. This file is saved in an output directory specified by the user or by default in "/tmp/".

If anything should go wrong in the support process, a pop-up box will open saying what happened. If this is not sufficient, please consult the mcm.log file for more information.

Chapter 8.0 Miscellaneous

8.1. Miscellaneous Overview

Various Cluster Manager sections that don't fit in other categories are contained here.

8.2. Console

Cluster Manager communicates directly with Workload Manager. This console displays the commands submitted to Workload Manager from Cluster Manager as well as any information returned by Workload Manager. Workload Manager output messages are highlighted in green while error messages are highlighted in red.

- Prepend path to Moab Workload Manager commands - This checkbox indicates whether the path the Moab client commands should be prepended to the commands being run. For example, if checked, the command 'showq' would become `/opt/moab/bin/showq` (assuming the Moab path was `/opt/moab/bin`). If false, it would stay `showq` and might result in an error if the Moab client commands were not in the default $PATH environment variable. This should be checked when running Moab commands and unchecked when running anything else. If the path was prepended for a non-Moab command like `pwd`, it would become `/opt/moab/bin/pwd`, which would result in an error
- Process Commands - This field will submit to Workload Manager any text in the Commands text window.

8.3. Debugging and Log Levels

Allows users to select the log level in Moab Cluster Manager, which can be used to help prepare logs to be accompanied with bug reports. Logs are written to the <MCM_HOME>/logs/mcm.log file. Higher logging levels create more detailed logging information, which facilitates debugging but may slow performance.

Below are the logging levels available, listed in order of increasing verbosity.

- 0 - Off : Turns off logging.
- 1 - Fatal : Logs only server events that cause the application to abort.
- 2 - Error : Logs all events that Fatal logs plus error events that might allow the application to continue running.
- 3 - Warn : Logs everything that Error logs plus other minor problems.
- 4 - Info : Logs everything that Warn logs does, plus informational messages that highlight the progress of the application at coarse-grained level.
- 5 - Info-With-Moab-Cmd-Debug : Logs everything that Info logs plus all the interaction with Moab Workload Manager except the frequent (defaults to every 2 seconds) commands to determine if Moab Cluster Manager should refresh data from Moab Workload Manager.
- 6 - Debug : Logs everything that Info-With-Moab-Cmd-Debug level logs plus fine-grained informational events that are useful in debugging.
- 7 - Trace : The most verbose logging level. The only level that logs all interaction with Moab Workload Manager, including the frequent refresh checks ignored by other logging levels.

Users can configure extremely fine grained logging information by editing the <MCM_HOME>/conf/log4j.properties file. Using this file, one can set the log level on individual classes or packages within Moab Cluster Manager. It should be noted that configuring individual package or class loggers in the log4j.properties file other than the root logger or the logger for com.moab.api.XMLDebuggingTools may override the settings applied from this window.

Verbose log levels cause a small performance penalty. Because levels 5 and above log all the interaction with Moab Workload Manager, they can use substantially more memory when connected to larger systems.

8.4. Cluster Manager Preferences

These preferences control Cluster Manager-specific settings.

Refresh Rate

The more often Cluster Manager communicates with Workload Manager the more up-to-date the information; however, when Cluster Manager communicates with Workload Manager more often, Workload Manager operates at a slower pace.

- Fast - At this refresh rate Cluster Manager updates its information every minute.
- Medium - At this refresh rate, which is the default refresh rate, Cluster Manager updates its information every 10 minutes.
- Slow - At this refresh rate Cluster Manager updates its information every hour.

Advanced Settings

Check this box to enable more advanced options throughout Cluster Manager. Currently, the plugin Manager is the only option available.

Font Size

This option specifies the font size used throughout MCM. Valid sizes are from 9 - 14. Default is 11. MCM must be restarted in order for font size changes to take effect.

Moab Commands

- Automatically Process Moab Commands - If checked, all commands submitted to Moab are automatically processed. If unchecked, all commands are displayed in the commands console for review. Unprocessed commands will have to be manually submitted via the Console.
- Show Command Output in Pop up Window - If checked, the output of all commands submitted to will be shown in a pop up window after the commands have run. If not checked, no pop up window will display.

Chapter 9.0 License

SOFTWARE from any organization that is not an 'Authorized Distribution and Services Partner' of Adaptive Computing Enterprises Inc. Any unauthorized partner that desires to become an 'Authorized Distribution and Services Partner' of Adaptive Computing Enterprises Inc. may contact us at support@adaptivecomputing.com. 'End User' organizations that desire services from Adaptive Computing Enterprises Inc., or an 'Authorized Distribution and Services Partner' may contact us using the same email listed above.

- Distribution

  'End User' organizations that are academic and government agencies may redistribute this SOFTWARE subject to the condition that the distribution contains conspicuous publication of the acknowledgement statement found within the LICENSE agreement distributed with this SOFTWARE.

  Organizations that are commercial and other for-profit organizations may not redistribute this code or derivations of this code in any form whatsoever, including parts of SOFTWARE incorporated into other software programs without express written permission from Adaptive Computing Enterprises Inc.

  Redistribution of the SOFTWARE in any form whatsoever, including parts of the code that are incorporated into other software programs, must include a conspicuous and appropriate publication of the following acknowledgement:

  'This product was developed by Adaptive Computing Enterprises Inc. Moab Scheduling System is a trademark of Adaptive Computing Enterprises Inc.'

  Any such modification of the SOFTWARE must, when installed, display the above language, the copyright notice, and the warranty disclaimer.

  Each time the SOFTWARE is redistributed (or any work based on the SOFTWARE), the recipient must automatically receive this LICENSE, copyright notice, and the warranty disclaimer as described in this license agreement, which govern the ability to copy, distribute or modify the SOFTWARE subject to these terms and conditions, and has the choice of accepting or declining the LICENSE. As the LICENSEE, you shall automatically provide the recipient with a copy of this LICENSE. Further restrictions are not to be imposed on recipients of the SOFTWARE by the LICENSEE beyond those expressly described herein.

- Use of Modifications

  LICENSEES with a redistribution agreement that wish to distribute their modifications (including government and academic institutions) must first send a copy of the modifications along with a brief explanation of why the modification was made and the resulting performance or functionality of the modifications to Adaptive Computing Enterprises Inc. at support@adaptivecomputing.com. Failure to send a copy of distributed modifications renders the LICENSE invalid, as well as any LICENSES granted to third parties subsequent to the incorporation of the modifications into SOFTWARE. Any such modification of the SOFTWARE must, when installed, display the LICENSE, the copyright notice, and the warranty disclaimer as described in the LICENSE agreement/s distributed with this SOFTWARE. Those without a LICENSE to redistribute may send modifications to Adaptive Computing Enterprises Inc. for evaluation and possible incorporation into SOFTWARE.

Copyright owners of modifications to SOFTWARE hereby grant Adaptive Computing Enterprises Inc., Inc. a non-exclusive, royalty-free, worldwide, irrevocable right and LICENSE to install, use, distribute, sublicense, and prepare derivative works of said modifications. Only organizations receiving an express prior written exclusion to this condition are exempted from providing these non-exclusive rights to Adaptive Computing Enterprises Inc.

- Communications about and Endorsement of SOFTWARE and Products/Software Derived from the SOFTWARE

  The name 'Moab Scheduling System' or 'Moab Scheduler' or any of its variants must not otherwise be used to endorse or to promote products derived from the SOFTWARE without prior written permission from ACE.

  Products derived from or incorporating the SOFTWARE in whole or in part shall not contain as part of the product's name any form of the terms 'Adaptive Computing Enterprises Inc.', 'ACE', 'Moab', 'Moab Scheduling System', 'Moab Scheduler', or 'Supercluster Development Group' unless prior written permission has been received from Adaptive Computing Enterprises Inc.

  All advertising materials for products that use or incorporate features of the SOFTWARE must display the following acknowledgement: 'This product includes software developed by Adaptive Computing Enterprises Inc. for use in the Moab Scheduling System.'

- Acceptance of this LICENSE

  It is not required that you accept this LICENSE; however, if you do not accept the terms of this LICENSE, you are prohibited by law from installing, using, modifying or distributing the SOFTWARE or any of its derivative works. Therefore, by installing, using, modifying or distributing the SOFTWARE (or any of its derivative works), you have agreed to this LICENSE and have accepted all its terms and conditions.

  If any portion of this LICENSE is held invalid or unenforceable under any particular circumstance, the balance of the LICENSE will continue to apply.