



# Statistics in Moab Cluster Manager

# Presentation Outline

- Moab Cluster Manager (MCM) Overview
- Enabling Statistics in Moab Workload Manager (Moab)
- Profiling Interval Based Statistics
  - Credential Based
  - Node Categorization
- Non-Profiling Interval Based Statistics
  - Matrix
  - Visual Cluster
- Conclusion



# Moab Cluster Manager (MCM) Overview

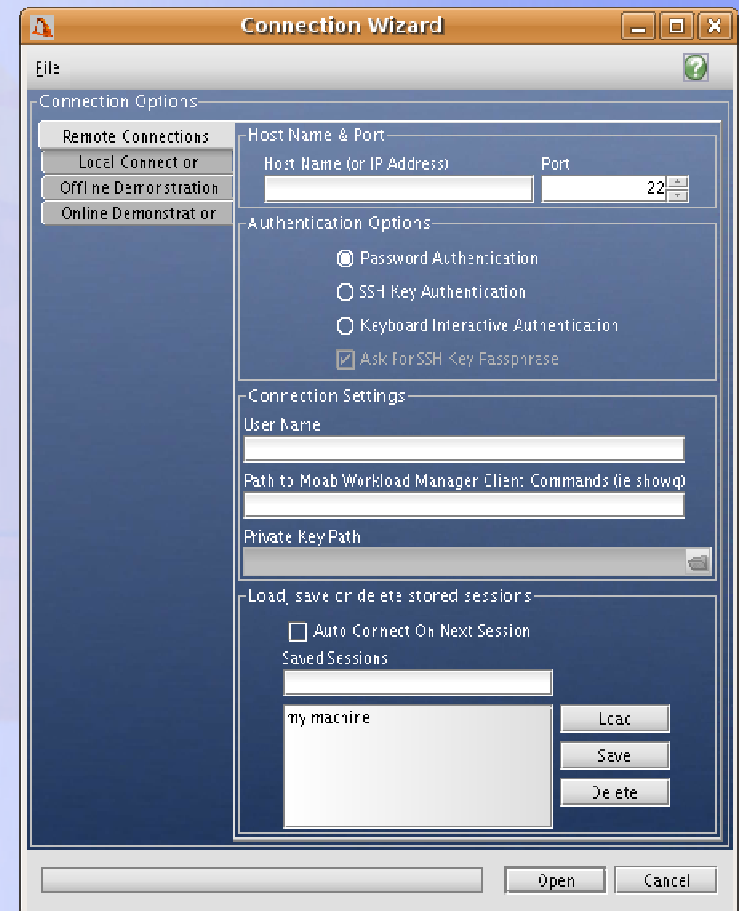
# Moab Cluster Manager

- Administer Resources and Workload Policies Through an Easy-to-Use Graphical User Interface
- Monitor, Diagnose and Report Resource Allocation, Usage, and Failures



# Connecting To Moab

- MCM can connect to Moab in a few different ways
  - SSH is used to connect remotely
  - Local connections are used to connect to Moab if running MCM on the head node



# MCM Main Screen

**Moab Cluster Manager**

File Configure Manage Monitor Reporting Help

Directory

- Workload
  - Jobs
    - Submit
    - List/Modify Jobs
    - Timeline
    - Job Templates
  - Reservations
  - Triggers
- Resources
  - Nodes
  - Partitions
  - Licenses
  - Cluster
    - Visual Cluster
    - Processor Usage
    - Configure
- Organization
  - Visual Credentials
  - User Access Settings
  - User
    - Group
    - Account
    - Class
    - QoS
- Policies
  - Fairshare
  - Fairshare Tree
  - Priority
  - Credentials
  - Other
- Reporting
  - Quick Charts/Graphs
  - Charts & Graphs
  - Matrix Statistics
  - Reports
- Diagnostics
- Support

**Cluster Information**

Name	puka
Host	compa
Port	36880
Mode	SIMULATION
Status	Running

**User Information**

User	bkimball
Group	bkimball
Account	
Class	batch, batch
QoS	

**Node Summary**

Busy Nodes	251	98 %
Idle Nodes	4	1.6 %
Down Nodes	1	0.4 %
Offline Nodes	0	0 %
Other Nodes	0	0 %
Total Nodes	256	

**bkimball Job Summary**

Running Jobs	0	0 %
Eligible Jobs	0	0 %
Blocked Jobs	0	0 %
Total Jobs	0	

**Cluster Resources, Inc.**

Professional Support
Documentation

**Job Summary**

Running Jobs	31	23.7 %
Eligible Jobs	95	72.5 %
Blocked Jobs	5	3.8 %
Total Jobs	131	



# Simplified Job Submission

**Submit Job**

Options

Job Information	
Script File*	<input type="text"/>
Job Name	<input type="text"/>
User Job Priority	<input type="text"/>

Resources	
Partition	<input type="text"/>
Reservation	<input type="text"/>
Job Dependency	<input type="text"/>
Node Features	<input type="text"/>
<input type="radio"/> Node List	<input type="text"/>
<input checked="" type="radio"/> Num. Of Nodes	<input type="text" value="0"/>
Procs. Per Node	<input type="text" value="0"/>
Memory Per Node	<input type="text" value="0"/> MB
Swap Per Node	<input type="text" value="0"/> MB
Operating System	<input type="text"/>
Architecture	<input type="text"/>

Search For Resources

Estimated Start Time Calculator

Estimated Start Time Table

Data Management	
Execution Directory*	<input type="text"/>
Output File	<input type="text"/>
Error File	<input type="text"/>

Credentials Information	
User	bkimball
Group	bkimball
Account	<input type="text"/>
Class	<input type="text"/>
Quality of Service	<input type="text"/>

Flags / Options	
<input type="checkbox"/> Preemptible	<input type="checkbox"/> Restartable
<input type="checkbox"/> Hold	

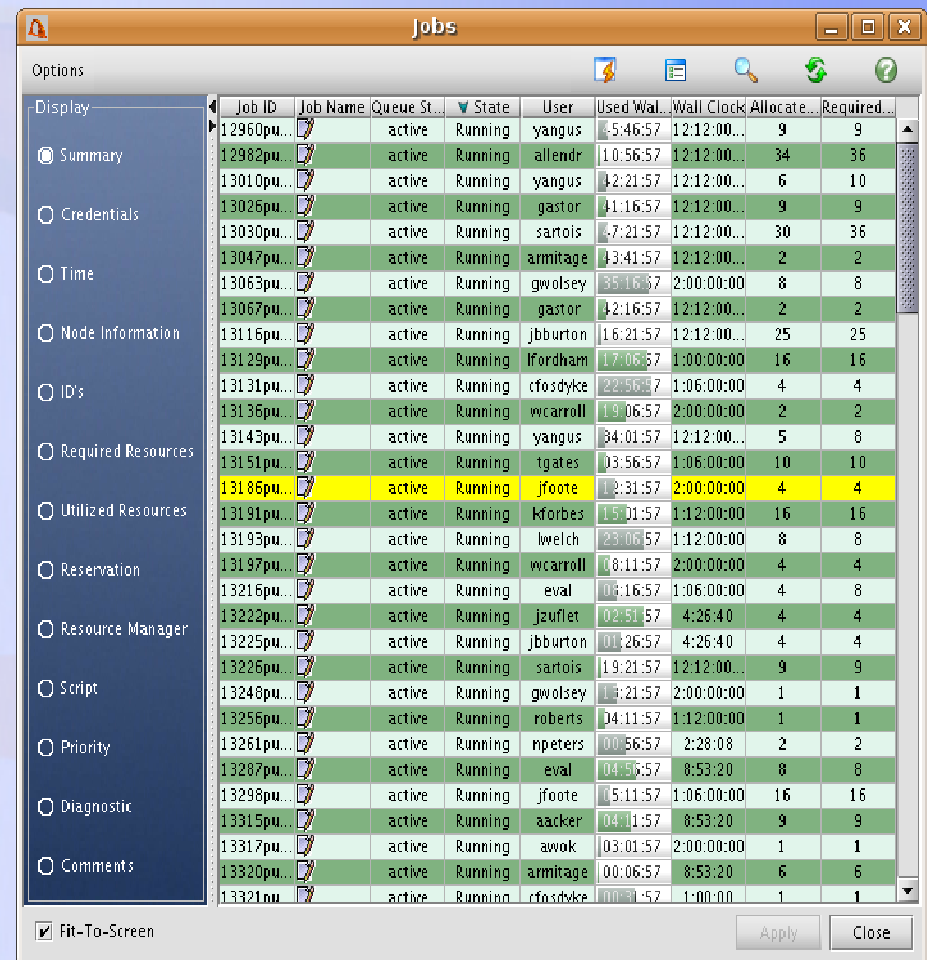
Email Notifications	
<input type="checkbox"/> On Job Completion	<input type="checkbox"/> On Job Failure
<input type="checkbox"/> On Job Execution	

Time Frame	
Start Time	<input type="text"/>
Duration	<input type="text"/>

Submit Close

# Diagnosing Your System

- Information is available for various parts of a system
  - Jobs
  - Nodes
  - Reservations
  - Credentials
  - Policies – Priority, Fairshare, etc.
  - Licenses
  - and many more
- MCM can submit commands to Moab as well



Job ID	Job Name	Queue	State	User	Used Wal...	Wall Clock	Allocate...	Required...
12960pu...			active	Running	yangus	4:54:57	12:12:00...	9
12982pu...			active	Running	allendr	10:56:57	12:12:00...	34
13010pu...			active	Running	yangus	12:21:57	12:12:00...	6
13026pu...			active	Running	gastor	11:16:57	12:12:00...	9
13030pu...			active	Running	sartois	47:21:57	12:12:00...	30
13047pu...			active	Running	armitage	13:41:57	12:12:00...	2
13063pu...			active	Running	gwolsey	35:16:57	2:00:00:00	8
13067pu...			active	Running	gastor	12:16:57	12:12:00...	2
13116pu...			active	Running	jbburton	16:21:57	12:12:00...	25
13129pu...			active	Running	lfordham	17:06:57	1:00:00:00	16
13131pu...			active	Running	cfosdyke	22:56:57	1:06:00:00	4
13136pu...			active	Running	wcarroll	19:06:57	2:00:00:00	2
13143pu...			active	Running	yangus	34:01:57	12:12:00...	5
13151pu...			active	Running	tgates	13:56:57	1:06:00:00	10
13186pu...			active	Running	jfoote	12:31:57	2:00:00:00	4
13191pu...			active	Running	kforbes	15:01:57	1:12:00:00	16
13193pu...			active	Running	lwelch	23:06:57	1:12:00:00	8
13197pu...			active	Running	wcarroll	18:11:57	2:00:00:00	4
13216pu...			active	Running	eval	05:16:57	1:06:00:00	4
13222pu...			active	Running	jzuffet	02:51:57	4:26:40	4
13225pu...			active	Running	jbburton	01:26:57	4:26:40	4
13226pu...			active	Running	sartois	19:21:57	12:12:00...	9
13248pu...			active	Running	gwolsey	13:21:57	2:00:00:00	1
13256pu...			active	Running	roberts	14:11:57	1:12:00:00	1
13261pu...			active	Running	npeters	00:56:57	2:28:08	2
13287pu...			active	Running	eval	04:05:57	8:53:20	8
13298pu...			active	Running	jfoote	15:11:57	1:06:00:00	16
13315pu...			active	Running	aacker	04:11:57	8:53:20	9
13317pu...			active	Running	avok	03:01:57	2:00:00:00	1
13320pu...			active	Running	armitage	00:06:57	8:53:20	6
13321pu...			active	Running	cfosdyke	00:00:57	1:00:00:00	1



# Enabling Statistics in Moab Workload Manger (Moab)

# Enabling Statistics in Moab

- Credential based statistics
  - Credential types
    - user, group, account, class, qos
  - In moab.cfg
    - USERCFG[DEFAULT] ENABLEPROFILING=true
    - GROUPCFG[DEFAULT] ENABLEPROFILING=true
    - etc.
- Node based statistics
  - In moab.cfg
    - NODECFG[default] ENABLEPROFILING=true

## Profiling Intervals

- Resource manager poll intervals
  - How often moab contacts the resource manager
  - Set in moab.cfg
    - RMPOLLINTERVAL 00:05:00
- Statistical profiling intervals
  - Should be a positive integer multiple of resource manager polling interval
    - never smaller
  - Should be  $\geq 5:00$
  - PROFILEDURATION 00:15:00
- Caution: Moab will not force this to happen
  - defaults to 30:00

# Setting the Correct Profiling Interval

RM interval	RM interval	RM interval
Profiling Interval		

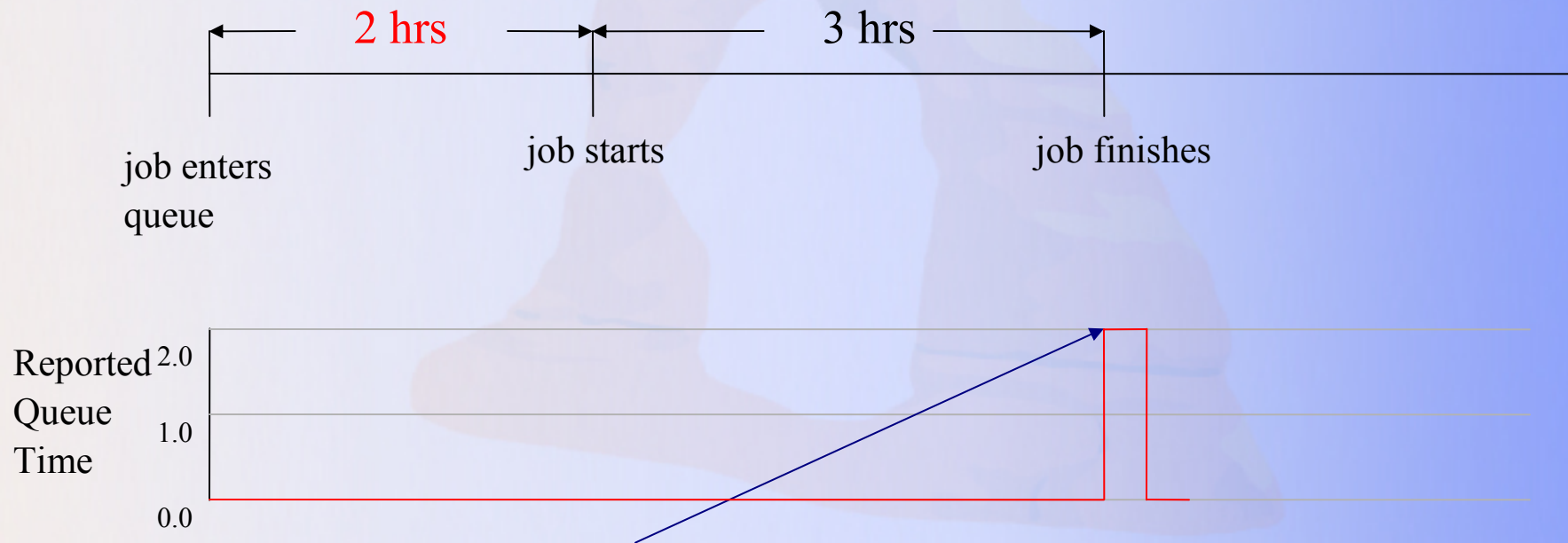
- Caution – Moab will adjust the profiling interval length to evenly divide 24 hours
  - resource manager poll interval not adjusted
  - Profiling interval of 7:00 becomes 6:40
- To verify, use the commands
  - `mschedctl -l --flags=verbose | grep RMPOLLINTERVAL`
  - `mschedctl -l --flags=verbose | grep PROFILEDURATION`

## When Statistics are Calculated

- **Continuously**
  - Reported when each profiling interval
  - Examples -
    - node state
    - number of available processors
    - backlog
- **Upon Job Completion**
  - Updated only when a job completes
  - Currently running jobs are “invisible”
  - Examples
    - number of executed jobs
    - job execution time
    - queue time



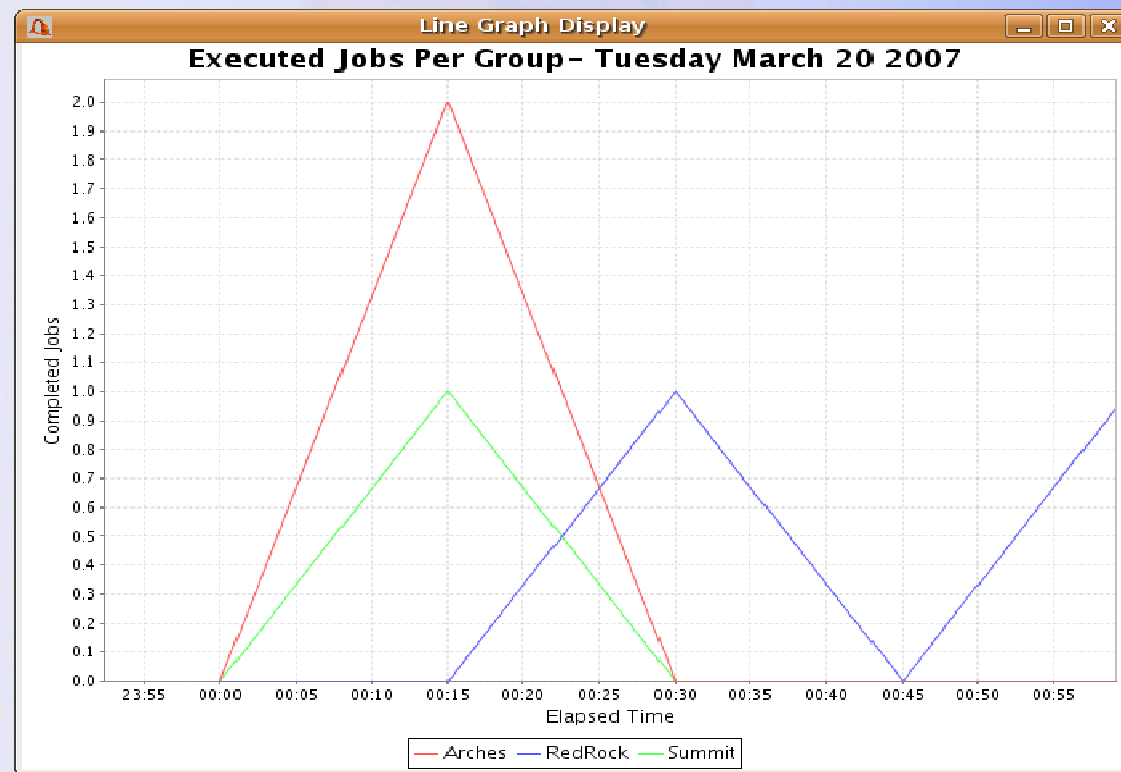
# Statistics Calculated Upon Job Completion



Notice the 2 hours of queue time is not reported until the job finishes.



# Example – Executed jobs not calculated until jobs complete



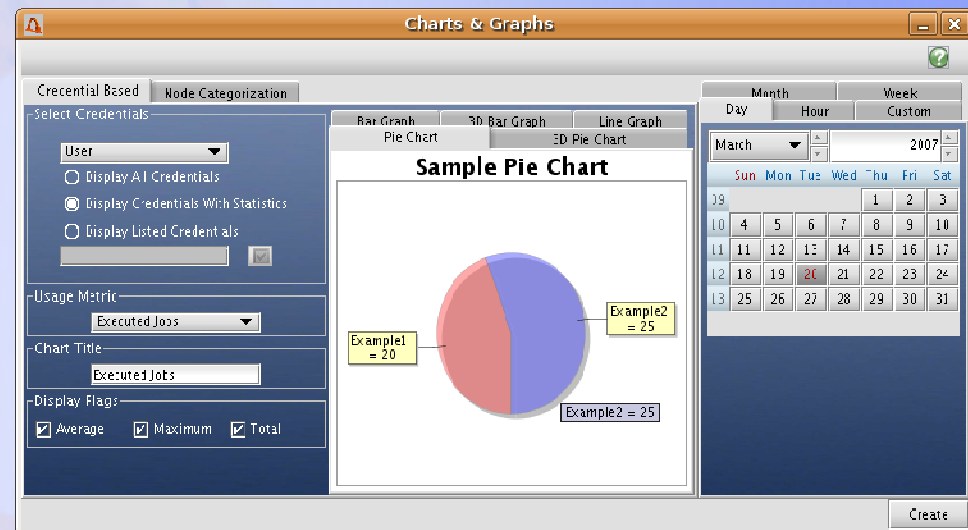
# Statistics Categorized by Credential

## Credential Based Statistics

- Allow historical statistics to be categorized by credential
  - Allows administrators to determine
    - who is using the most resources
    - if resources are being allocated fairly
  - Examples:
    - average job queue time PER user
    - total executed jobs PER account

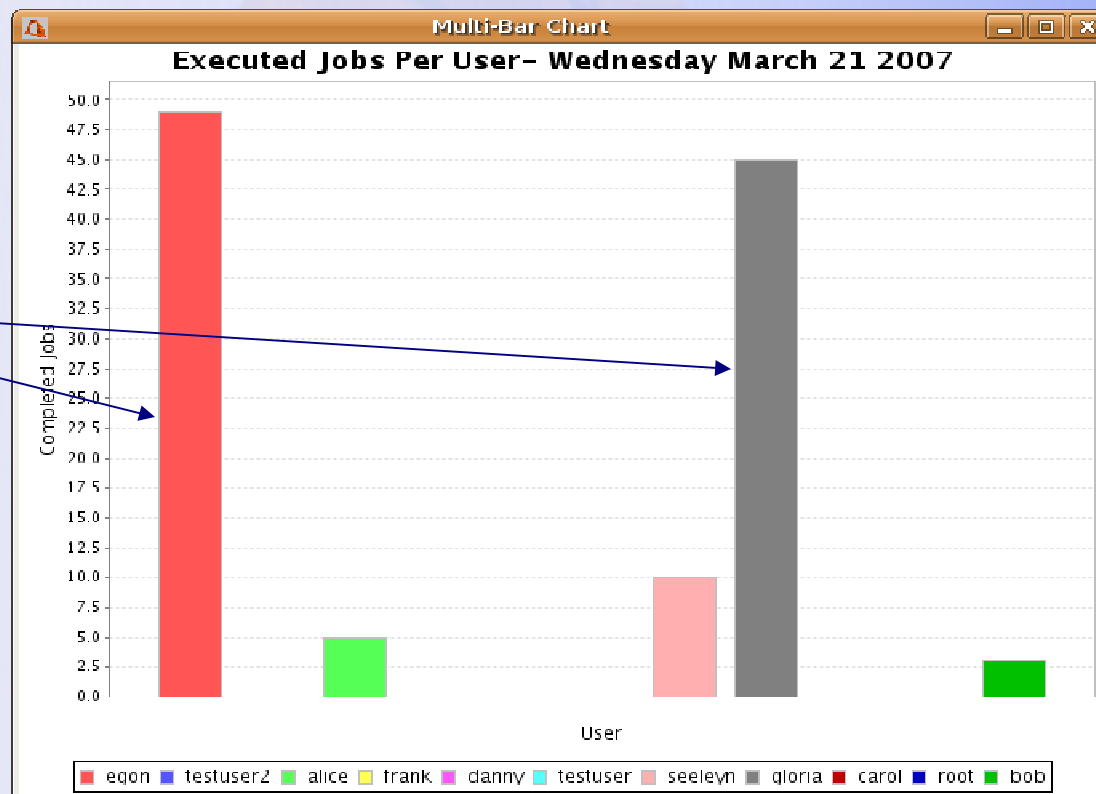
# Charts & Graphs

- Allows one to select
  - Credential
  - Usage Metric
  - Chart Type
  - Time Window
  - Display Flags
  - Title
- Exportable as .png files

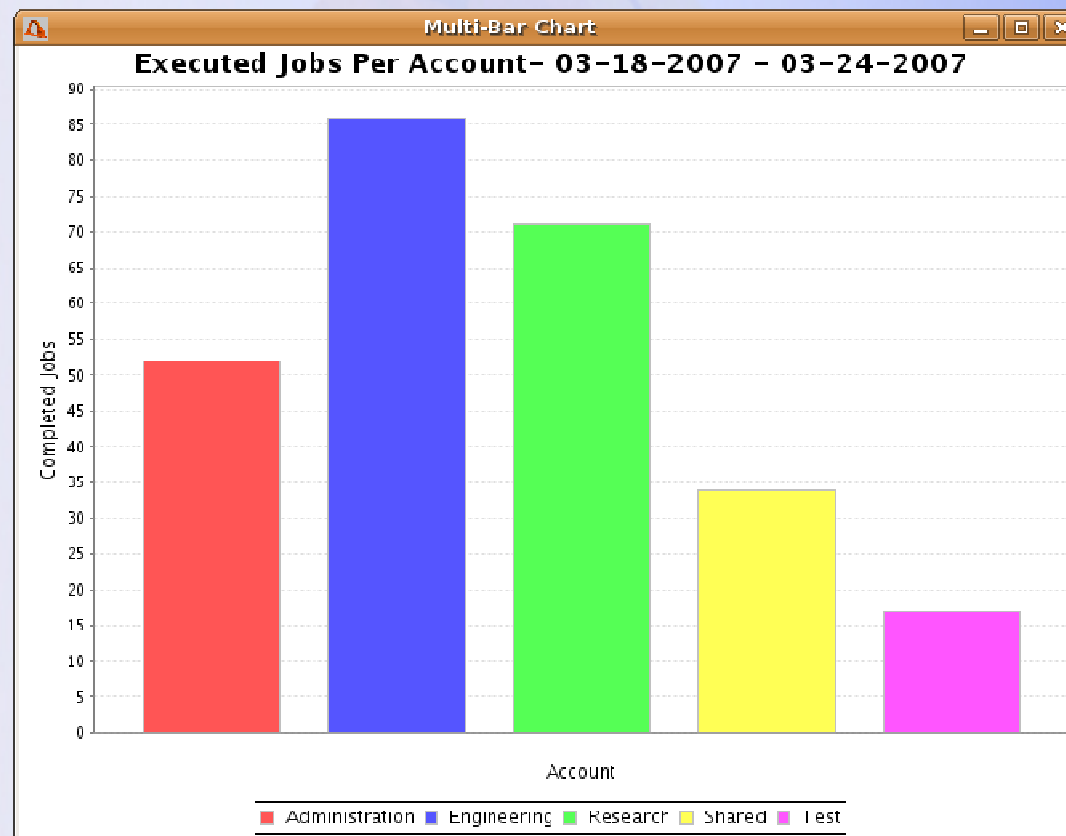


# Example – Determining resource allocation

Gloria and Egon  
are executing the  
majority of jobs

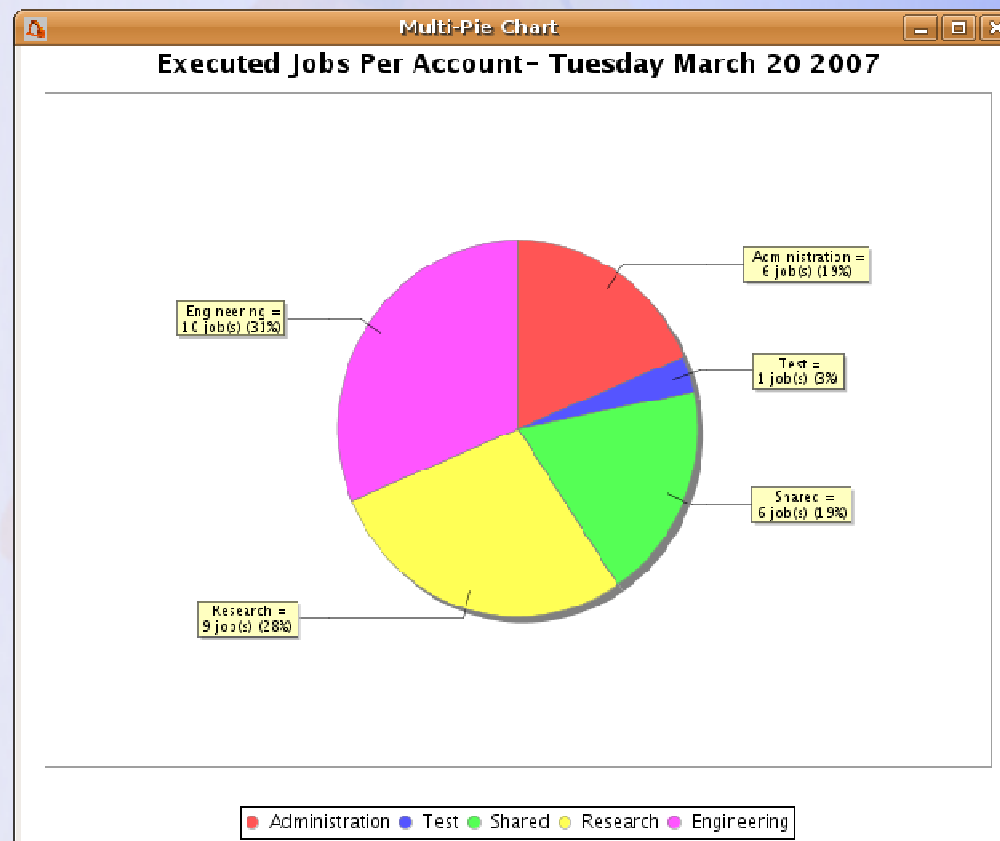


## Example – Total jobs executed per account in bar graph

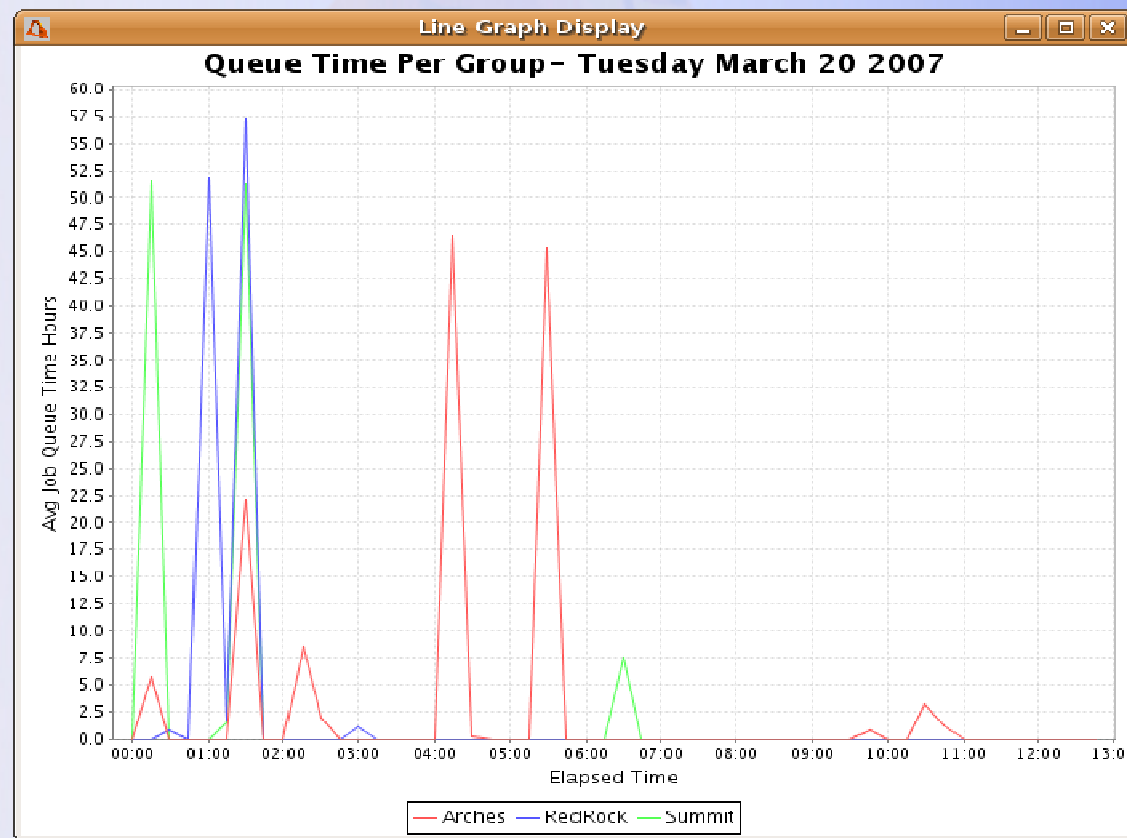




## Example – Total jobs executed per account in bar graph

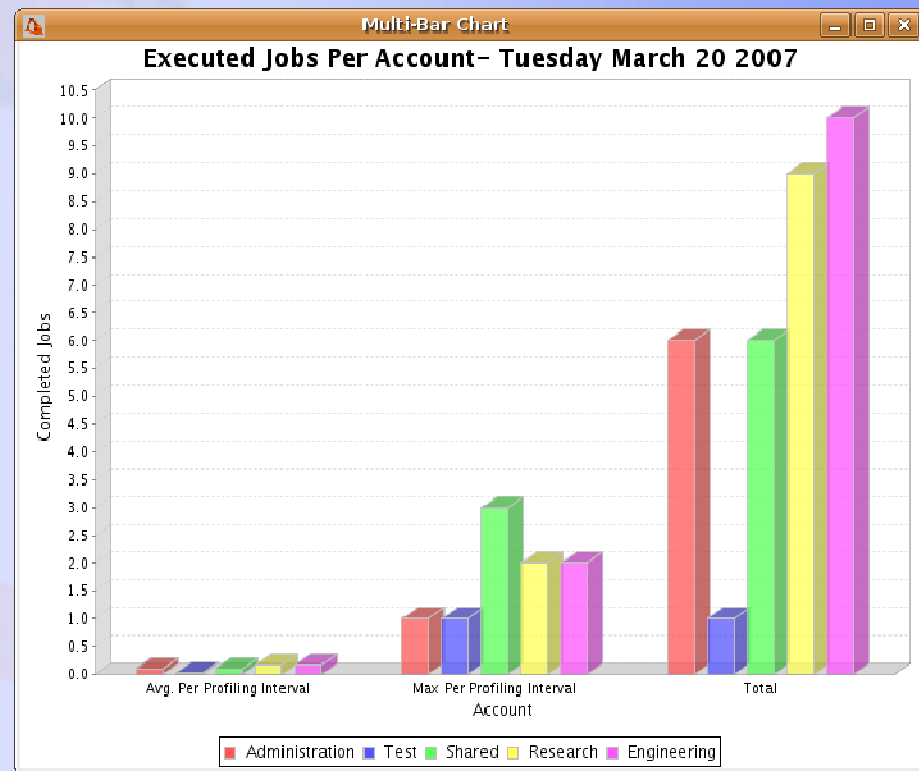


# Example – Queue time per group in line graph



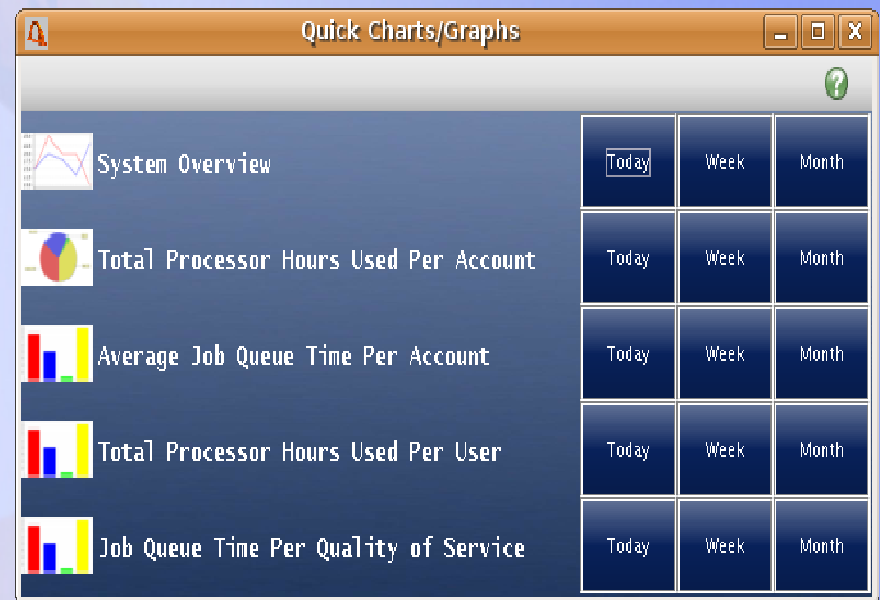
# Display Flags

- Average
  - the average per profiling interval
- Maximum
  - the maximum in one profiling interval
- Total
  - The total over all profiling intervals



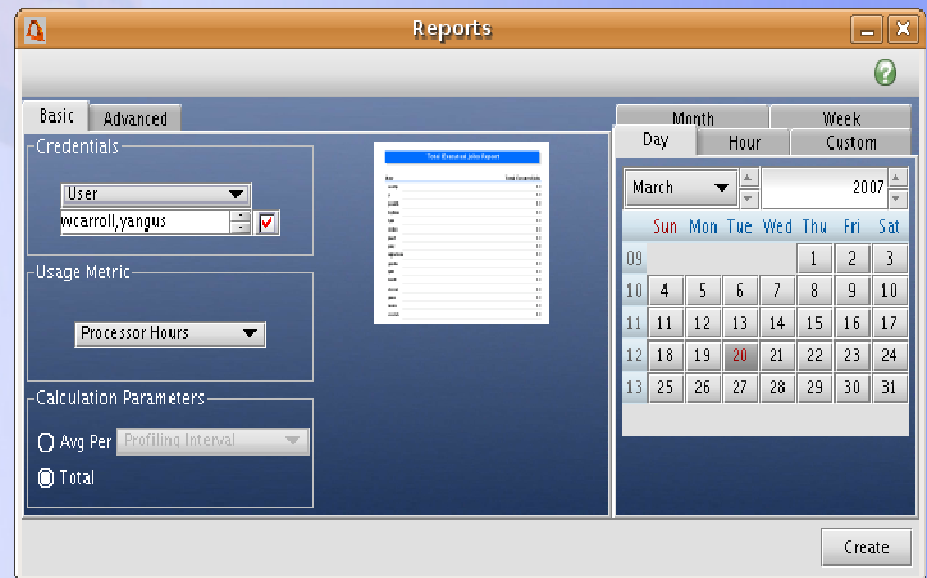
## Quick Charts

- Common credential based charts
- Time windows
  - day
    - 12:00am to 11:59 pm
  - week
    - Sun to Sat
  - month
    - 1<sup>st</sup> to end of month



# Reports

- Credential based statistics in a phone bill format
- Can be exported to PDF
- Can be categorized by group, account, class, or qos



The screenshot shows the 'Reports' application window with the 'Basic' tab selected. The interface includes a 'Credentials' section with a 'User' dropdown menu set to 'wcarroll,yangus' and a checked checkbox. Below this is a 'Usage Metric' section with a 'Processor Hours' dropdown. The 'Calculation Parameters' section has two radio buttons: 'Avg Per' (selected) and 'Total'. A 'Profiling Interval' dropdown is also present. On the right, there is a calendar for March 2007, showing the 10th as the selected date. A 'Create' button is located at the bottom right of the window.

# Example Report

Custom Reports - Print Preview

File Navigation Zoom Help

100%

Tuesday March 20 2007

User	Count : 45	Total Proc. Hours
axdeer		0
allende		420.25
amadsen		0
arnitage		116.35
awok		26.97
cfosdyle		7.5
cjackson		0
drahman		1.41
erai3		120.1
fothgill		0
gastor		148.32
gwolsey		101.81
ibentley		0
jkhurton		242.66
jbethune		0
frooxe		0.34
jsmith		61.53
jvela		0.17
jzuffet		170.19
ldorbes		7.88
lbeverly		0.85
lfordham		8.32
twelch		0
mbentley		1.96
mhirch		0
mshaw		0
mullis		45.58
npeters		0
nsmith		38.02

Page 1

Page 1 of 2



## Example Categorized Report

- Categorized by account
- Note that a user can belong to more than one account
  - If billing be careful not to double charge

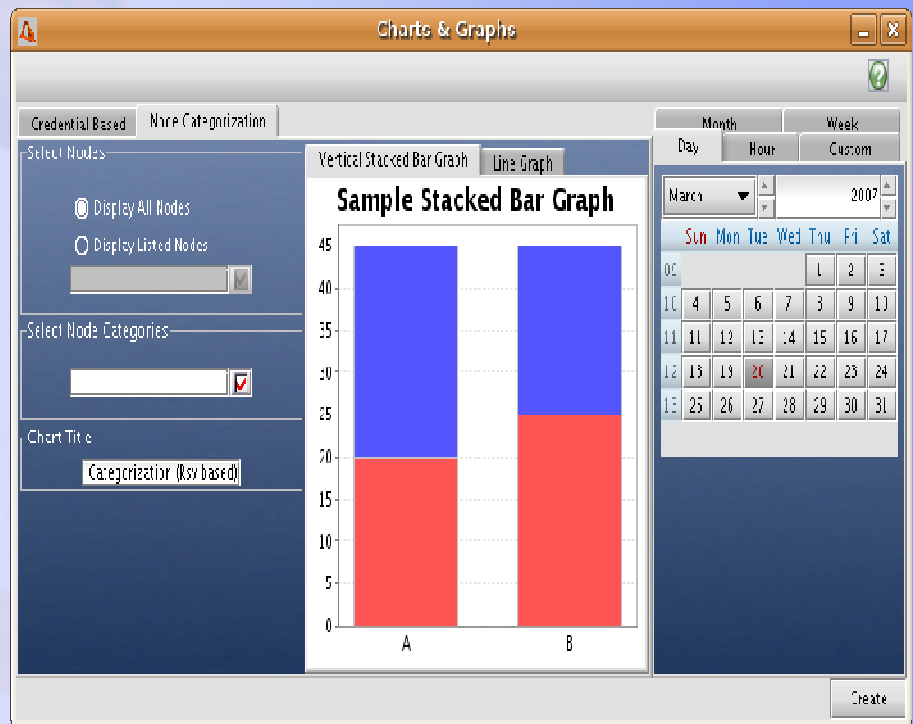
engineering items # 4	
	Total Executed Jobs
alice	5
bob	1
carol	0
see leyn	10
Total	16
Average	4

marketing items # 2	
	Total Executed Jobs
grima	40
see leyn	10
Total	50
Average	25

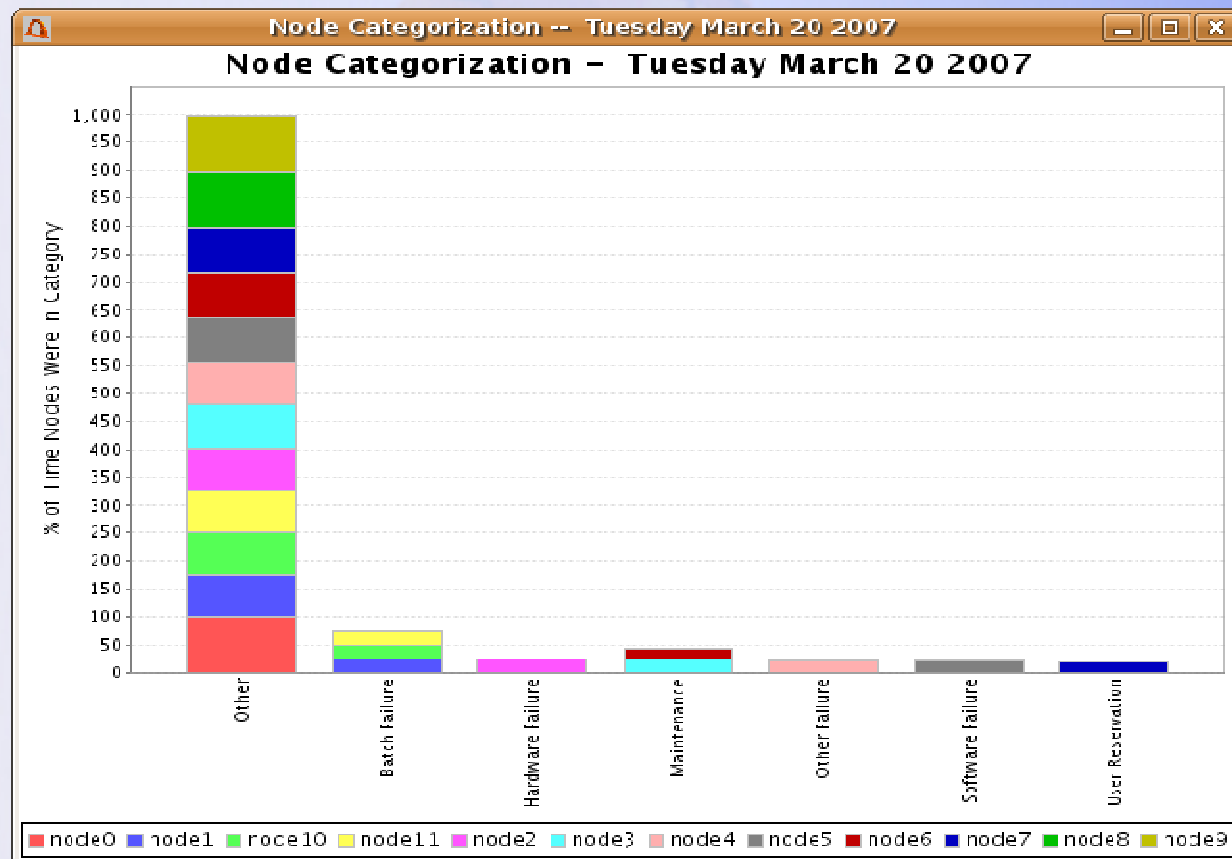
# Node Categorization Statistics

# Node Categorization

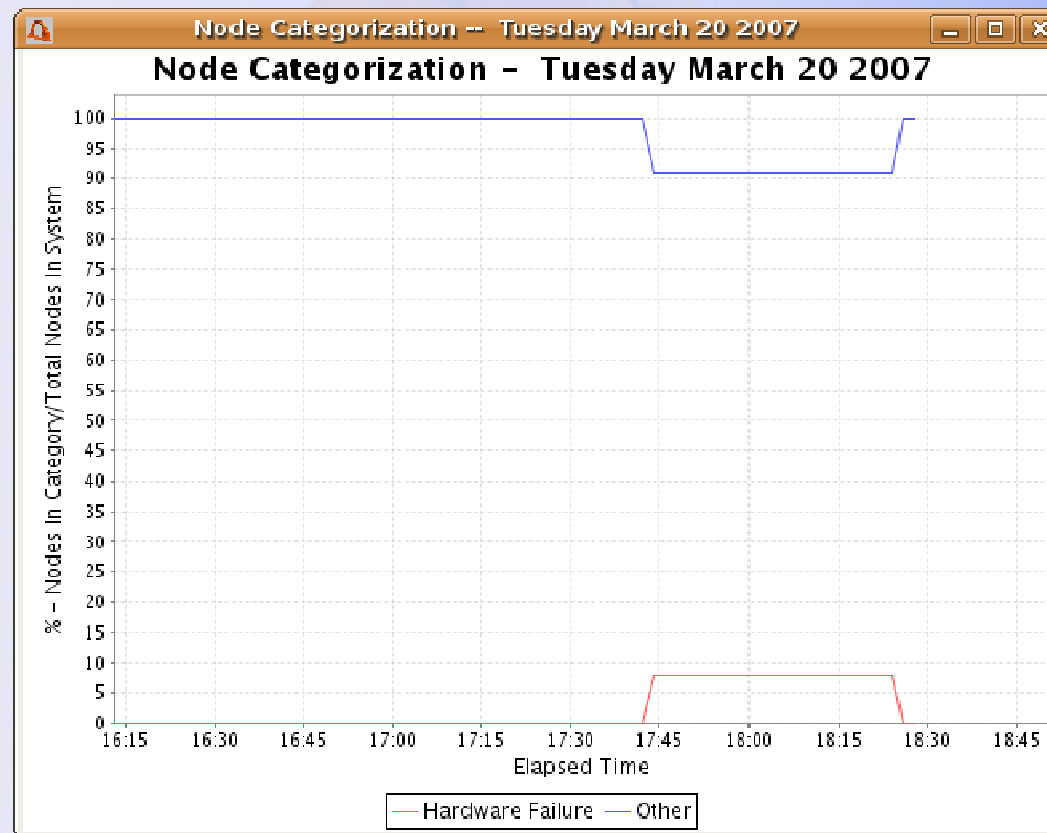
- Display node categories over time
- Category examples
  - hardware failure
  - software failure
  - active
  - personal reservation



# Example – Percent time nodes spent in different categories



## Example – Percent of nodes in different categories





# Matrix Statistics

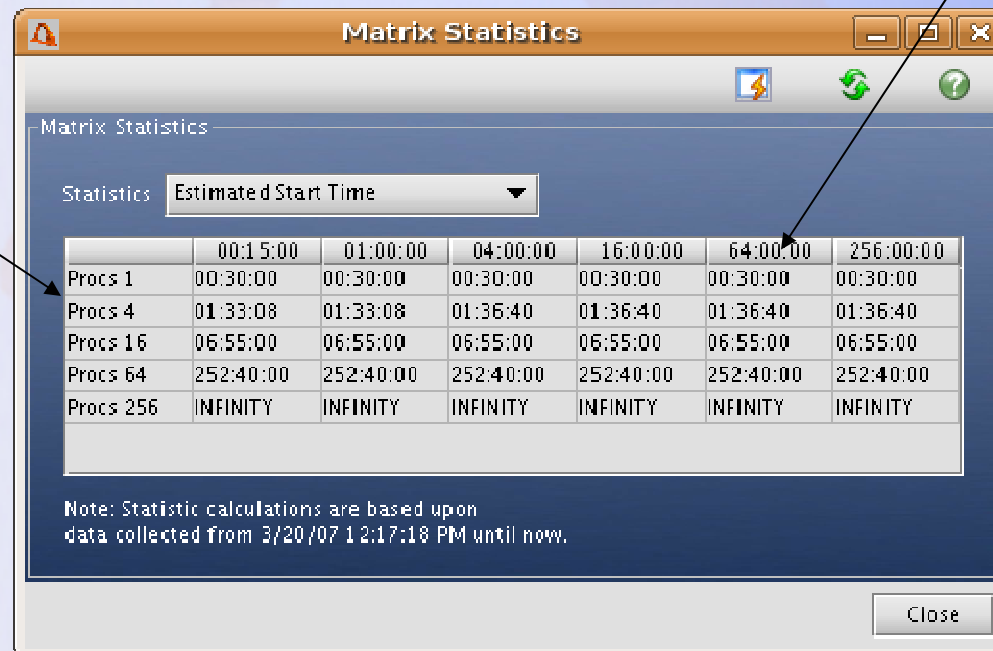


# Matrix Statistics

- Matrix statistics are used both to analyze historic workload and predict future workload.

Relative Time Frames

Job  
Processor  
Size



The screenshot shows a window titled "Matrix Statistics" with a toolbar containing a lightning bolt, a refresh icon, and a help icon. Below the toolbar is a "Statistics" section with a dropdown menu set to "Estimated Start Time". The main area contains a table with 7 columns representing time intervals and 6 rows representing different processor counts. The first row of the table is highlighted. Below the table is a note about the data source and a "Close" button.

	00:15:00	01:00:00	04:00:00	16:00:00	64:00:00	256:00:00
Procs 1	00:30:00	00:30:00	00:30:00	00:30:00	00:30:00	00:30:00
Procs 4	01:33:08	01:33:08	01:36:40	01:36:40	01:36:40	01:36:40
Procs 16	06:55:00	06:55:00	06:55:00	06:55:00	06:55:00	06:55:00
Procs 64	252:40:00	252:40:00	252:40:00	252:40:00	252:40:00	252:40:00
Procs 256	INFINITY	INFINITY	INFINITY	INFINITY	INFINITY	INFINITY

Note: Statistic calculations are based upon data collected from 3/20/07 12:17:18 PM until now.

Close

# Matrix Statistics Types

- Matrix Statistics have a variety of statistics for display
  - Proc Hours Requested
  - Job Queue Times
  - Expansion Factor
  - Backfill
  - QoS Met
  - Job Efficiency
  - and more

**Matrix Statistics**

Matrix Statistics

Statistics: Cluster Proc Hours Requested %

	00:15:00	01:00:00	04:00:00	16:00:00	64:00:00	256:00:00	total
Procs 1	0.00	0.03	0.11	0.14	2.01		2.30
Procs 4	0.00	0.25	1.77	1.65	8.20		11.87
Procs 16	0.00	2.12	10.69	7.58	54.45		74.85
Procs 64	0.00	0.31	3.92	2.92	3.83		10.98
Procs 256	-	-	-	-	-		-
Procs total	0.01	2.71	16.49	12.29	68.50		

Note: Statistic calculations are based upon data collected from 3/20/07 12:17:18 PM until now.

Close

**Matrix Statistics**

Matrix Statistics

Statistics: Average Job Queue Time

	00:15:00	01:00:00	04:00:00	16:00:00	64:00:00	256:00:00	total
Procs 1	12.67	0.74	3.26	0.48	2.48		2.32
Procs 4	2.83	5.63	6.99	8.17	7.66		6.90
Procs 16	9.83	10.47	10.28	5.76	8.56		9.51
Procs 64	18.75	9.58	4.71	10.14	1.42		7.45
Procs 256	-	-	-	-	-		-
Procs total	9.51	8.13	8.34	6.54	6.74		7.78

Note: Statistic calculations are based upon data collected from 3/20/07 12:17:18 PM until now.

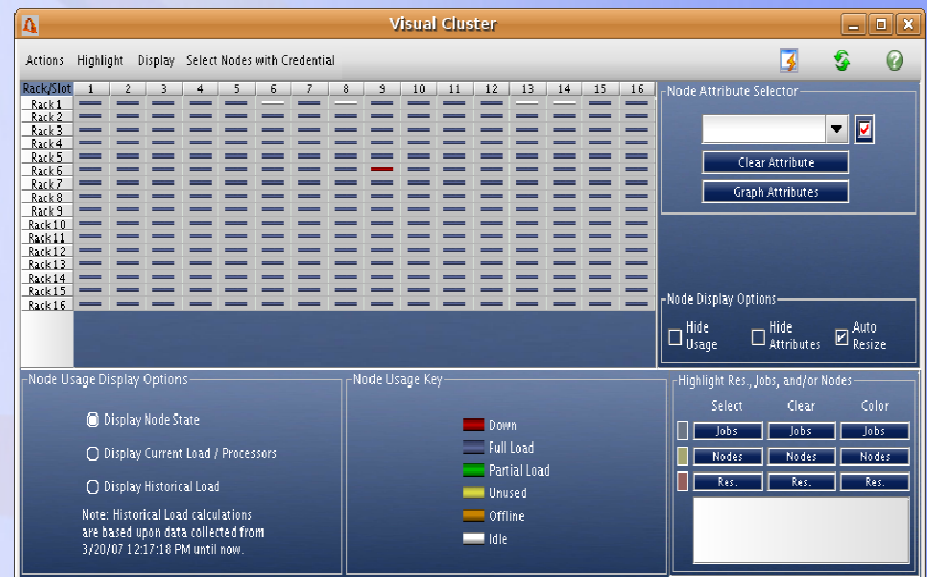
Close



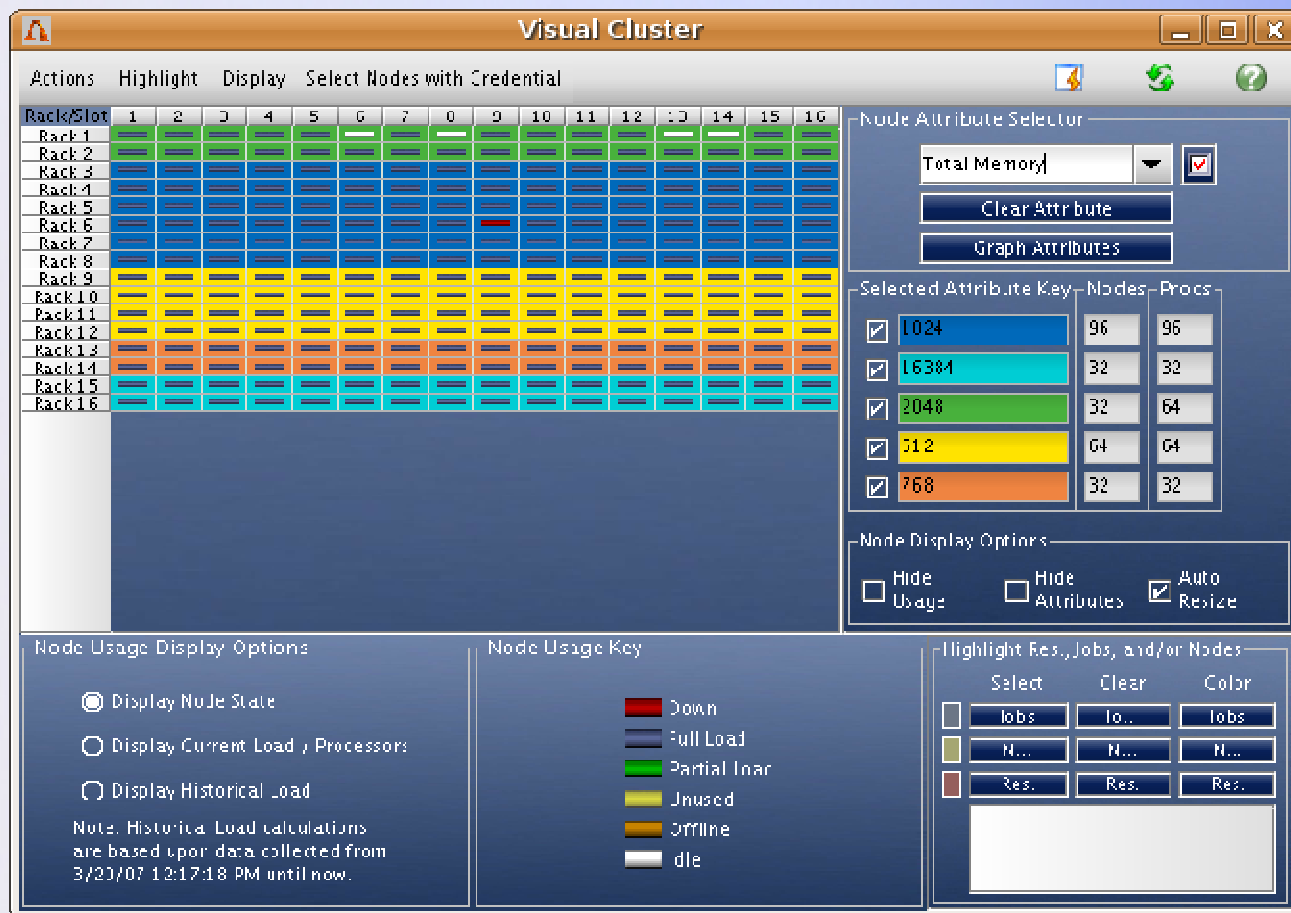
# Visual Cluster Statistics

# Visual Cluster Diagnostics and Statistics

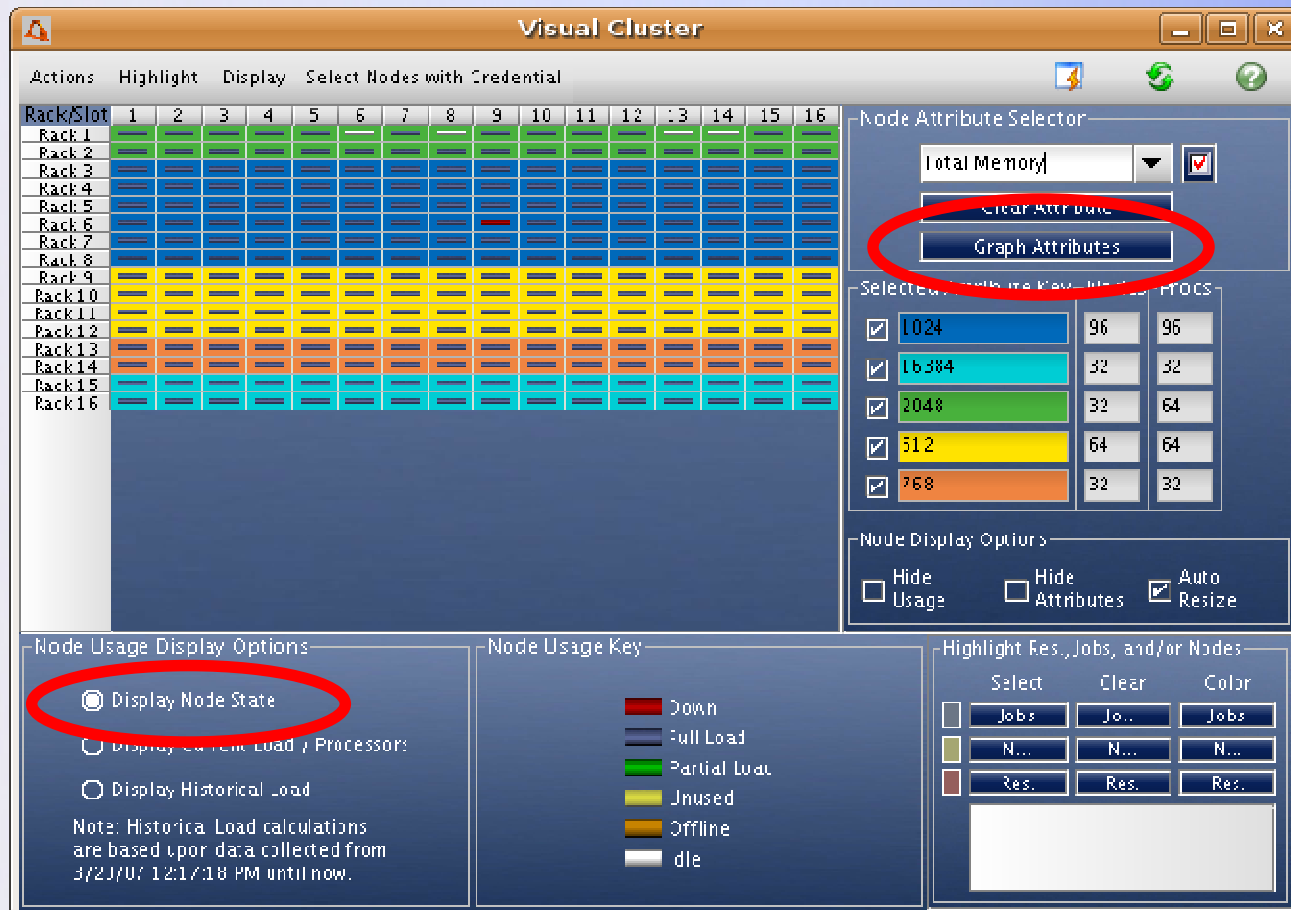
- The visual cluster gives an easy and concise way of viewing your entire cluster and the status of each node
- Visual Cluster Uses
  - Diagnose Failures
  - Control Nodes
  - View Node Attributes
  - View Jobs On Nodes
  - View Credentials
  - View Reservations
  - and more



# Visual Cluster Screen



# Graphing Node Attributes



The screenshot shows the Visual Cluster software interface. The main window displays a grid of node attributes for 16 racks, each with 16 slots. The grid is color-coded by node state: green for 'Down', blue for 'Full Load', yellow for 'Partial Load', orange for 'Unused', and red for 'Offline'. The 'Display Node State' option is selected in the 'Node Usage Display Options' panel. The 'Node Attribute Selector' panel shows 'Total Memory' selected, and the 'Graph Attributes' button is highlighted with a red circle. The 'Node Usage Key' panel shows the color coding for node states. The 'Highlight Res., Jobs, and/or Nodes' panel shows options for highlighting resources, jobs, and nodes.

**Visual Cluster**

Actions Highlight Display Select Nodes with Credential

Rack/Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Rack 1																
Rack 2																
Rack 3																
Rack 4																
Rack 5																
Rack 6																
Rack 7																
Rack 8																
Rack 9																
Rack 10																
Rack 11																
Rack 12																
Rack 13																
Rack 14																
Rack 15																
Rack 16																

**Node Attribute Selector**

Total Memory [v] [x]

Clear Attributes

Graph Attributes

**Select Nodes with Key Attributes**

Attribute	Value	Proc
1024	96	96
16384	32	32
2048	32	64
512	64	64
768	32	32

**Node Display Options**

☐ Hide Usage ☐ Hide Attributes ☒ Auto Resize

**Node Usage Display Options**

☒ Display Node State

☐ Display Current Load / Processor

☐ Display Historical Load

Note: Historical Load calculations are based upon data collected from 3/23/07 12:17:18 PM until now.

**Node Usage Key**

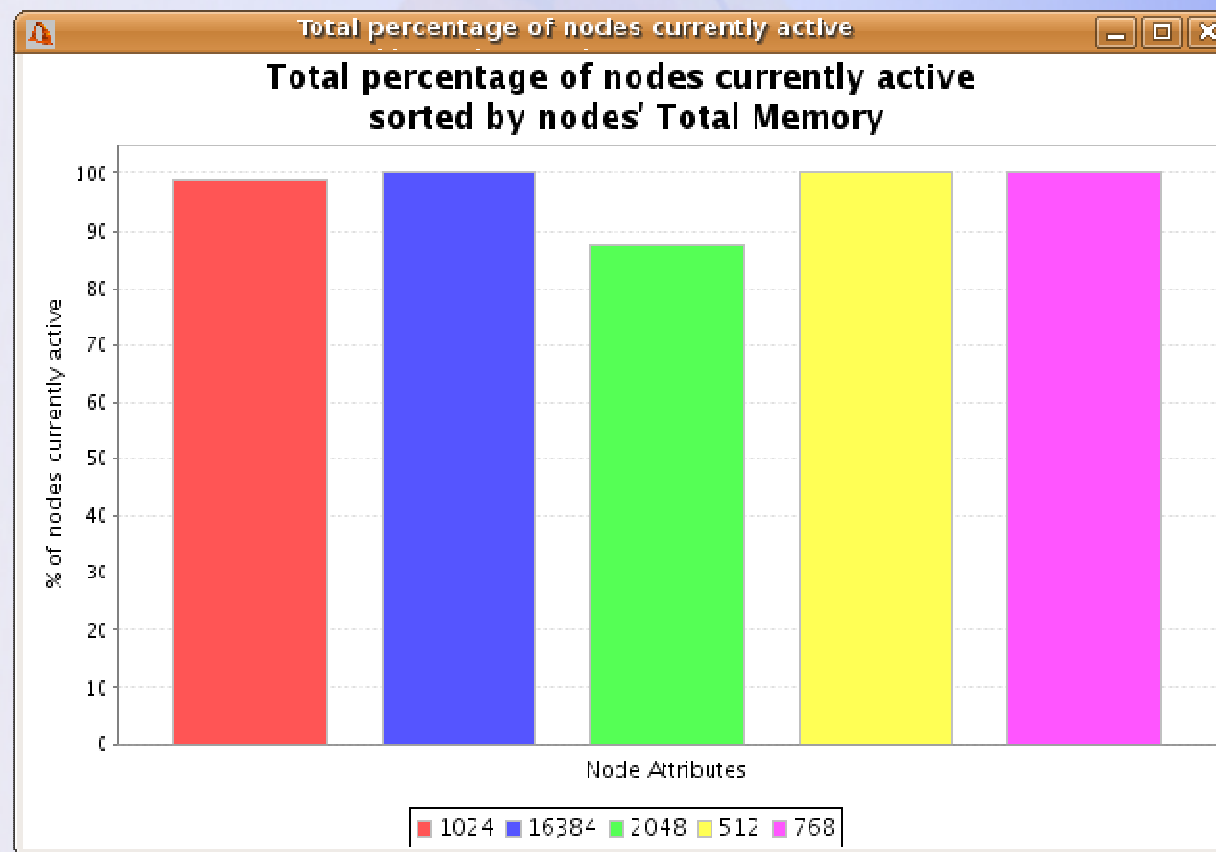
- Down
- Full Load
- Partial Load
- Unused
- Offline
- die

**Highlight Res., Jobs, and/or Nodes**

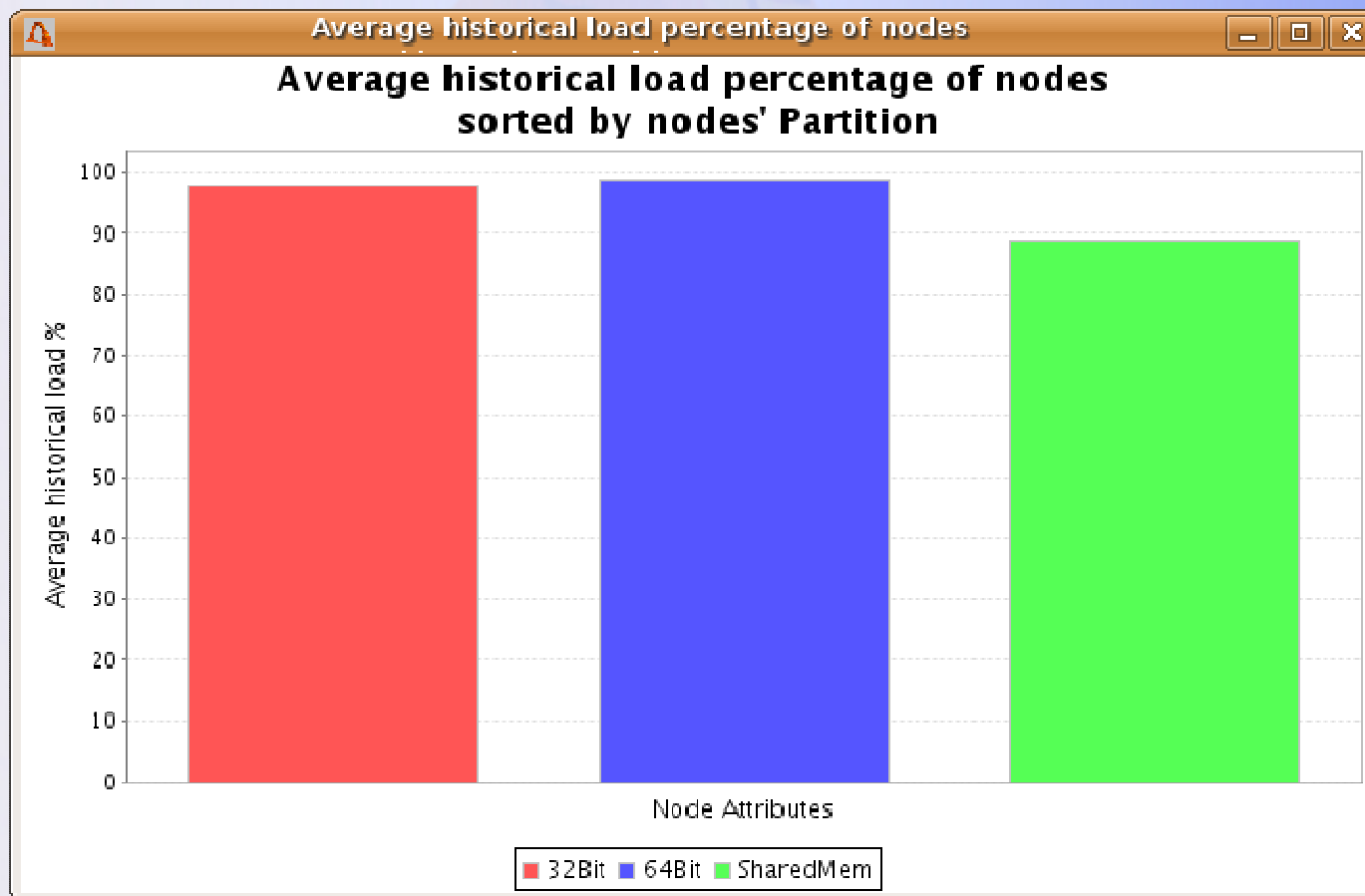
Select	Clear	Color
Jobs	Jobs	Jobs
N...	N...	N...
Res.	Res.	Res.



# Graphing Node Attributes Based On Node State



# Graphing Node Attributes Based On Historical Load



## Conclusion

- MCM gives administrators many tools to help them diagnose their clusters
- In order to use MCM's statistical tools one needs to know
  - how to properly configure Moab
  - how to interpret the statistics reported
- MCM's statistical tools allows administrators to be more productive by allowing them to
  - determine if their system is being optimally utilized
  - decide if policies need adjustment
  - identify failures