



HPCC System Administrator's Guide

Boca Raton Documentation Team

HPCC System Administrator's Guide

Boca Raton Documentation Team

Copyright © 2014 HPCC Systems. All rights reserved

We welcome your comments and feedback about this document via email to <docfeedback@hpccsystems.com> Please include **Documentation Feedback** in the subject line and reference the document name, page numbers, and current Version Number in the text of the message.

LexisNexis and the Knowledge Burst logo are registered trademarks of Reed Elsevier Properties Inc., used under license. Other products, logos, and services may be trademarks or registered trademarks of their respective companies. All names and example data used in this manual are fictitious. Any similarity to actual persons, living or dead, is purely coincidental.

2014 Version 5.0.0-1

Introducing HPCC Systems Administraton	4
Introduction	4
Architectural Overview	5
Routine Maintenance	11
Back Up Data	11
Log Files	14
Preflight	16
Preflight System Servers	17
Preflight Thor	20
Preflight the Roxie Cluster	23
System Configuration and Management	25
Running the Configuration Manager	28
Envrionment.conf	34
User Security Maintenance	36
Workunits and Active Directory	62
Data Handling	63
Best Practices	64
Cluster Redundancy	64
High Availability and Disaster Recovery	66
Best Practice Considerations	68
System Sizings	69
System Resources	71
HPCC Resources	71

Introducing HPCC Systems Administration

Introduction

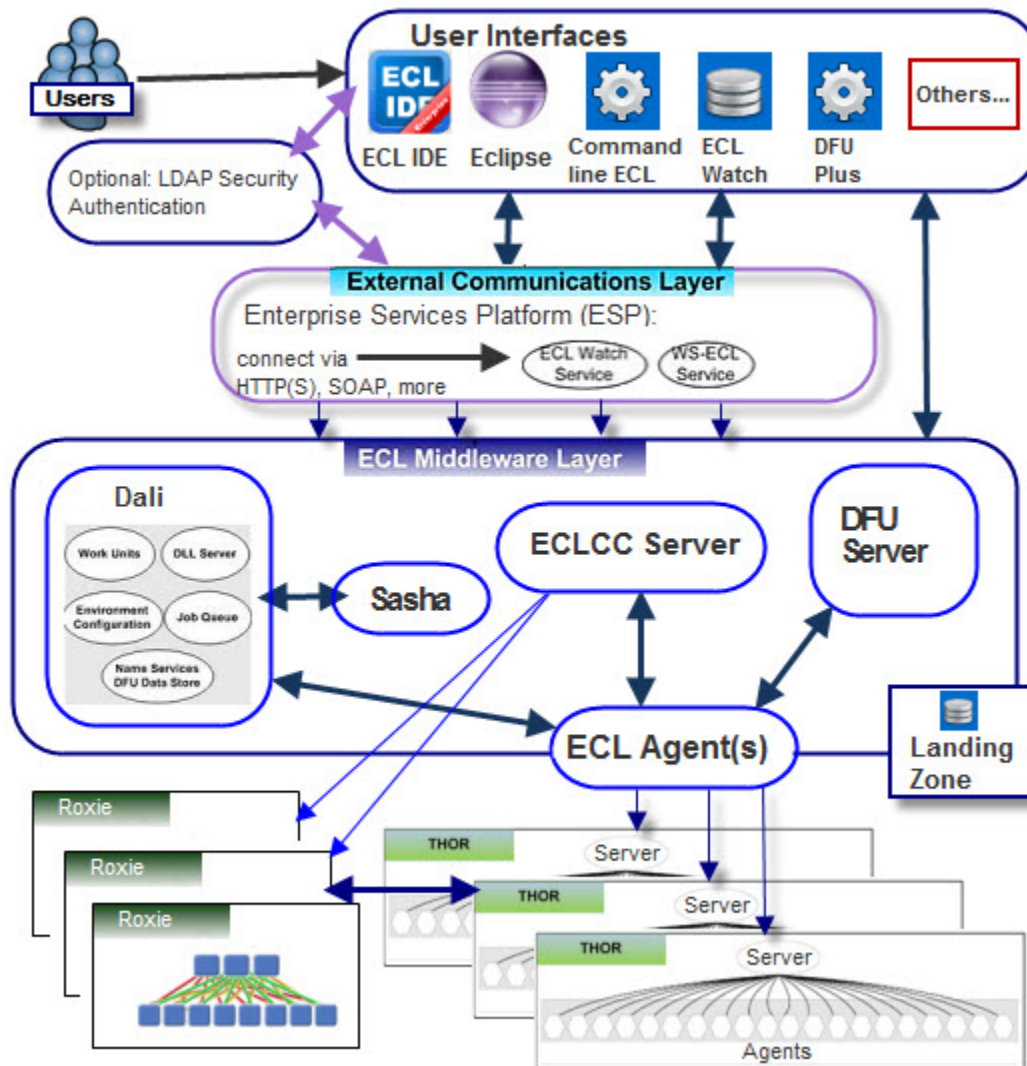
HPCC (High Performance Computing Cluster) is a massive parallel-processing computing platform that solves Big Data problems.

HPCC stores and processes large quantities of data, processing billions of records per second using massive parallel processing technology. Large amounts of data across disparate data sources can be accessed, analyzed, and manipulated in fractions of seconds. HPCC functions as both a processing and a distributed data storage environment, capable of analyzing terabytes of information.

Architectural Overview

An HPCC Systems Platform consists of the following components: Thor, Roxie, ESP Server, Dali, Sasha, DFU Server, and ECLCC Server. LDAP security is optionally available.

Figure 1. HPCC Architectural Diagram

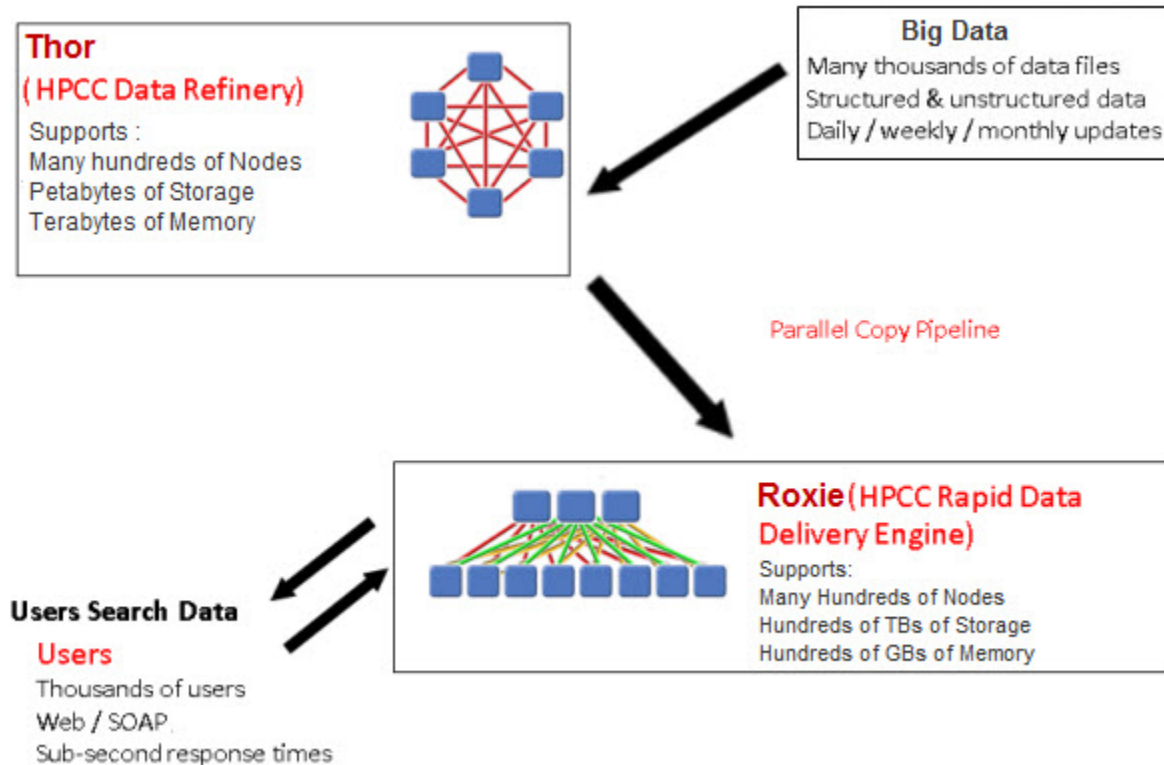


Data loading is controlled through the Distributed File Utility (DFU) server.

Data typically arrives on the landing zone (for example, by FTP). File movement (across components) is initiated by DFU. Data is copied from the landing zone and is distributed (sprayed) to the Data Refinery (Thor) by the ECL code. Data can be further processed via ETL (Extract, Transform, and Load process) in the refinery.

A single physical file is distributed into multiple physical files across the nodes of a cluster. The aggregate of the physical files creates one logical file that is addressed by the ECL code.

Figure 2. Data Processing



The data retrieval process (despraying) places the file back on the landing zone.

Clusters

HPCC environment contains clusters which you define and use according to your needs. The types of clusters used in HPCC:

Thor

Data Refinery (Thor) – Used to process every one of billions of records in order to create billions of "improved" records. ECL Agent (hThor) is also used to process simple jobs that would be an inefficient use of the Thor cluster.

Roxie

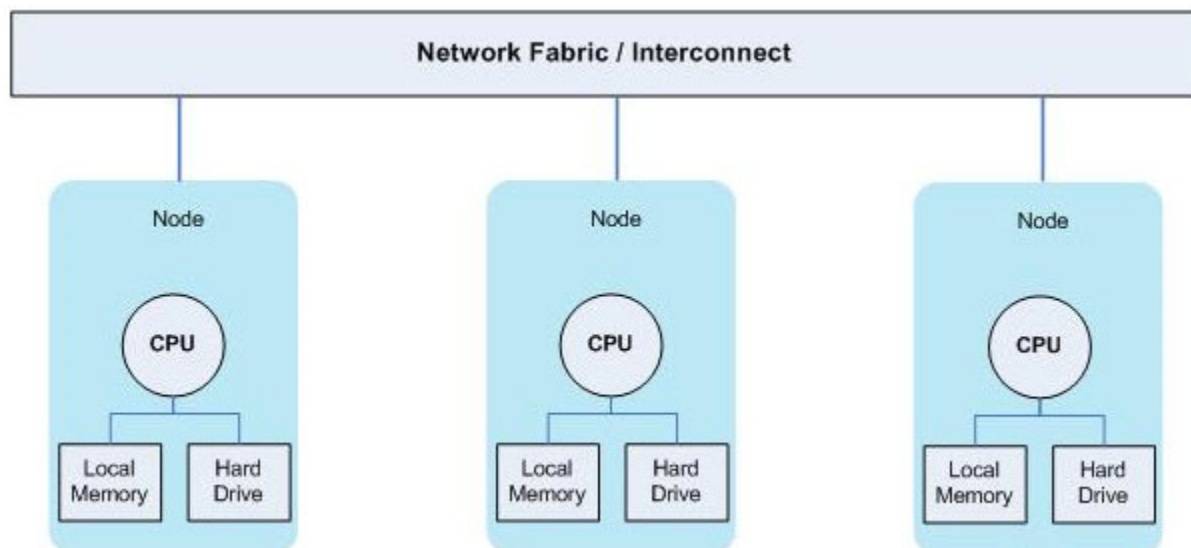
Rapid Data Delivery Engine (Roxie) – Used to search quickly for a particular record or set of records.

Queries are compiled and published, usually in ECL Watch. Data moves in parallel from Thor nodes to the receiving Roxie nodes. Parallel bandwidth utilization improves the speed of putting new data into play.

ECL Agent

The ECL Agent can act as a single-node cluster. That is called spawning an hThor cluster. hThor is used to process simple jobs that would otherwise be an inefficient use of Thor. For simple tasks, the ECL Agent will make a determination and perform the execution itself by acting as an hThor cluster.

Figure 3. Clusters



System Servers

The System Servers are integral middleware components of an HPCC system. They are used to control workflow and intercomponent communication.

Dali

Dali is also known as the system data store. It manages workunit records, logical file directory, and shared object services.

It maintains the message queues that drive job execution and scheduling. It also enforces the all LDAP security restrictions.

Sasha

The Sasha server is a companion “housekeeping” server to the Dali server. It works independently of all other components. Sasha’s main function is to reduce the stress on the Dali server. Whenever possible, Sasha reduces the resource utilization on Dali.

Sasha archives workunits (including DFU Workunits) which are stored in a series of folders.

Sasha also performs routine housekeeping such as removing cached workunits and DFU recovery files.

DFU Server

DFU server controls the spraying and despraying operations used to move data in and out of Thor.

DFU services are available from:

- Standard libraries in ECL code.
- Client interfaces: Eclipse, ECL Playground, ECL IDE, and the ECL command line interface.
- DFU Plus command line interface.

ECLCC Server

ECLCC Server is the compiler that translates ECL code. When you submit ECL code, the ECLCC Server generates optimized C++ which is then compiled and executed. ECLCC Server controls the whole compilation process.

When you submit workunits for execution on Thor, they are first converted to executable code by the ECLCC Server.

When you submit a Workunit to Roxie, code is compiled and later published to the Roxie cluster, where it is available to execute multiple times.

ECLCC Server is also used when the ECL IDE requests a syntax check.

ECLCC Server uses a queue to convert workunits one at a time, however you can have ECLCC Servers deployed in the system to increase throughput and they will automatically load balance as required.

ECL Agent

ECL Agent (hThor) is a single node process for executing simple ECL Queries.

ECL Agent is an execution engine that processes workunits by sending them to the appropriate cluster. ECL Agent processes are spawned on-demand when you submit a workunit.

ESP Server

ESP (Enterprise Service Platform) Server is the inter-component communication server. ESP Server is a framework that allows multiple services to be “plugged in” to provide various types of functionality to client applications via multiple protocols.

Examples of services that are plugged into ESP include:

- **WsECL:** Interface to published queries on a Roxie, Thor, or hThor cluster.
- **ECL Watch:** A web-based query execution, monitoring, and file management interface. It can be accessed via the ECL IDE or a web browser. See *Using ECL Watch*.

Examples of protocols supported by the ESP Server framework include: HTTP, HTTPS, SOAP, and JSON.

LDAP

You can incorporate a Lightweight Directory Access Protocol (LDAP) server to work with Dali to enforce the security restrictions for data, file, workunit scopes, and feature access.

When LDAP is configured, you need to authenticate when accessing ECL Watch, WsECL, ECL IDE, or any other client tools. Those credentials are then used to authenticate any requests from those tools.

Client Interfaces

The following Client Interfaces are available to interact with the HPCC Platform.

Eclipse

With the ECL plug-in for Eclipse, you can use the Eclipse IDE to create and execute queries into your data on an HPCC platform using Enterprise Control Language (ECL). Eclipse is open-source, and multi-platform and it can be used to interface with your data and workunits on HPCC. The ECL plug-in for Eclipse is also open source.

ECL IDE

ECL IDE is a full-featured GUI for ECL development providing access to the ECL repository and many of the ECL Watch capabilities. ECL IDE uses various ESP services via SOAP.

The ECL IDE provides access to ECL Definitions to build your queries. These definitions are created by coding an expression that defines how some calculation or record set derivation is to be done. Once defined, they can be used in succeeding ECL definitions.

ECL Watch

ECL Watch is a web-based query execution, monitoring, and file management interface. It can be accessed via ECL IDE, Eclipse, or a web browser. ECL Watch allows you to see information about and manipulate workunits. It also allows you monitor cluster activity and perform other administrative tasks.

Using ECL Watch you can:

- Browse through previously submitted workunits (WU). You can see a visual representation (graphs) of the data flow within the WU, complete with statistics which are updated as the job progresses.
- Search through files and see information including record counts and layouts or sample records.

- See status of all system servers.
- View log files.
- Add users or groups and modify permissions.

See the *Using ECL Watch* Manual for more details.

Command Line Tools

Command line tools: **ECL**, **DFU Plus**, and **ECL Plus** provide command line access to functionality provided by the ECL Watch web pages. They work by communicating with the corresponding ESP service via SOAP.

Routine Maintenance

There is some care required to ensure that your HPCC system keeps operating optimally. The following sections address the routine maintenance tasks for your HPCC system.

Back Up Data

An integral part of routine maintenance is the back up of essential data. Devise a back up strategy to meet the needs of your organization. This section is not meant to replace your current back up strategy, instead this section supplements it by outlining special considerations for HPCC Systems.

Back Up Considerations

You probably already have some sort of a back up strategy in place, by adding HPCC Systems into your operating environment there are some additional considerations to be aware of. The following sections discuss back up considerations for the individual HPCC system components.

Dali

Dali can be configured to create its own back up, ideally you would want that back up kept on a different server or node. You can specify the Dali back up folder location using the Configuration Manager. You may want to keep multiple copies that back up, to be able to restore to a certain point in time. For example, you may want to do daily snapshots, or weekly.

You may want to keep back up copies at a system level using traditional back up methods.

Sasha

Sasha itself generates no original data but archives workunits to disks. Be aware that Sasha can create quite a bit of archive data. Once the workunits are archived they are no longer available in the Dali data store. The archives can still be retrieved, but that archive now becomes the only copy of these workunits.

If you need high availability for these archived workunits, you should back them up at a system level using traditional back up methods.

DFU Server

DFU Server has no data. DFU workunits are stored in Dali until they are archived by Sasha.

ECLCC Server

ECLCC Server stores no data. ECL workunits are stored in Dali and archived by Sasha.

ECL Agent

ECL Agent stores no data.

ECL Scheduler

ECL Scheduler stores no data. ECL Workunits are stored in Dali.

ESP Server

ESP Server stores no data. If you are using SSL certificates, public and private keys they should be backed up using traditional methods.

Thor

Thor, the data refinery, as one of the critical components of HPCC Systems needs to be backed up. Back up Thor by configuring replication and setting up a nightly back up cron task. Back up Thor on demand before and/or after any node swap or drive swap if you do not have a RAID configured.

A very important part of administering Thor is to check the logs to ensure the previous back ups completed successfully.

Backupnode

Backupnode is a tool that is packaged with HPCC. Backupnode allows you to back up Thor nodes on demand or in a script. You can also use backupnode regularly in a crontab. You would always want to run it on the Thor master of that cluster.

The following example is one suggested way for invoking backupnode manually.

```
/bin/su - hpcc -c "/opt/HPCCSystems/bin/start_backupnode thor" &
```

The command line parameter must match the name of your Thor cluster. In your production environment, it is likely that you would provide descriptive names for your Thor clusters.

For example, if your Thor cluster is named thor400_7s, you would call start_backupnode thor400_7s.

```
/bin/su - hpcc -c "/opt/HPCCSystems/bin/start_backupnode thor400_7s" &
```

To run backupnode regularly you could use cron. For example, you may want a crontab entry (to back up thor400_7s) set to run at 1am daily:

```
0 1 * * * /bin/su - hpcc -c "/opt/HPCCSystems/bin/start_backupnode thor400_7s" &
```

Backupnode writes out its activity to a log file. That log can be found at /var/log/HPCCSystems/backupnode/MM_DD_YYYY_HH_MM_SS.log with the (MM) Month, (DD) Day, (YYYY) 4-digit Year, (HH) Hour, (MM) Minutes, and (SS) Seconds of the back up in the log file name. The main log file exists on the Thor master node. It shows what nodes it is run on and if it finished. You can find other backupnode logs on each of the thorslaves showing what files, if any, it needed to “restore”.

It is important to check the logs to ensure the previous back ups completed successfully. The following entry is from the backupnode log showing that back up completed successfully:

```
00000028 2014-02-19 12:01:08 26457 26457 "Completed in 0m 0s with 0 errors"
00000029 2014-02-19 12:01:08 26457 26457 "backupnode finished"
```

Roxie

Roxie data is protected by three forms of redundancy:

- **Original Source Data File Retention:** When a query is published, the data is typically copied from a remote site, either a Thor or a Roxie. The Thor data can serve as back up, provided it is not removed or altered on Thor. Thor data is typically retained for a period of time sufficient to serve as a back up copy.
- **Peer-Node Redundancy:** Each Slave node typically has one or more peer nodes within its cluster. Each peer stores a copy of data files it will read.

- **Sibling Cluster Redundancy:** Although not required, Roxie may run multiple identically-configured Roxie clusters. When two clusters are deployed for Production each node has an identical twin in terms of queries and/or data stored on the node in the other cluster. This configuration provides multiple redundant copies of data files. With three sibling Roxie clusters that have peer node redundancy, there are always six copies of each file part at any given time; eliminating the need to use traditional back up procedures for Roxie data files.

Landing Zone

The Landing Zone is used to host incoming and outgoing files. This should be treated similarly to an FTP server. Use traditional system level back ups.

Misc

Back up of any additional component add-ons, your environment files (environment.xml), or other custom configurations should be done according to traditional back up methods.

Log Files

You can review system messages and see any error messages as they are reported and captured in log files. Log files can help you in understanding what is occurring on the system and useful in troubleshooting.

Component Logs

There are log files for each component in directories below `/var/log/HPCCSystems` (default location). You can optionally configure the system to write the logs in a different directory. You should know where the log files are, and refer to the logs first when troubleshooting any issues.

There are log files which record activity among the various components. You can find the log files in subdirectories named corresponding to the components that they track. For example, the Thor logs would be found in a directory named `mythor`, the sasha log would be in the `mysasha` directory, the esp log in the `myesp` directory.

In each of the component subdirectories, there are several log files. Most of the log files use a logical naming convention that includes the component name, the date, and time in the name of the log file. There is also usually a link for the component with a simple name, such as `esp.log` which is a short cut to the latest current log file for that component.

Understanding the log files, and what is normally reported in the log files, helps in troubleshooting the HPCC system.

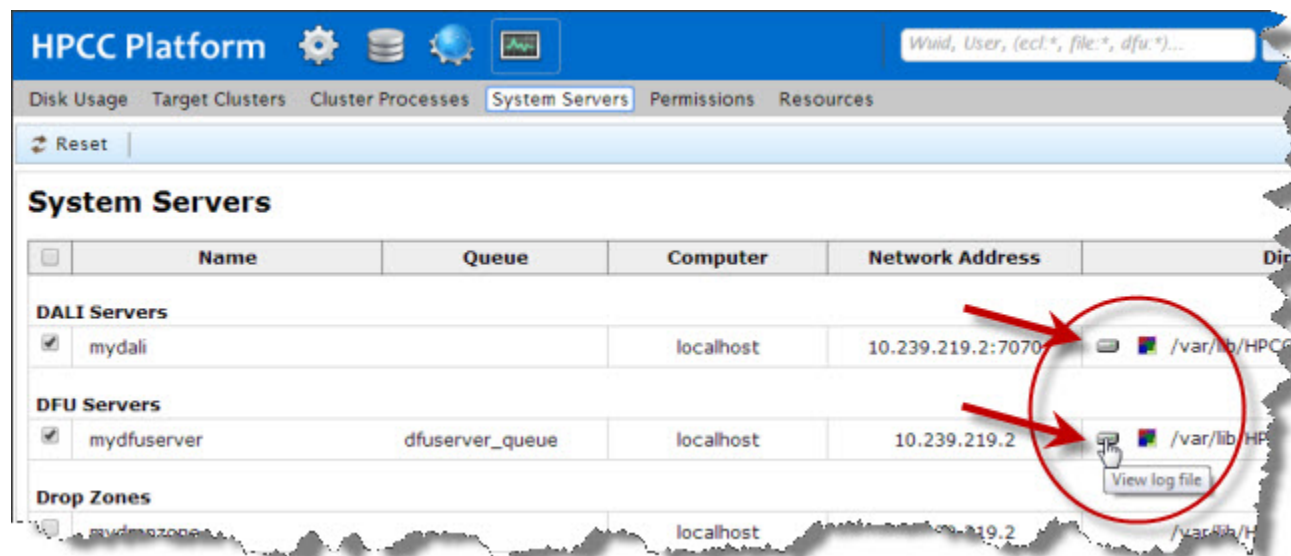
As part of routine maintenance you may want to back up, archive, and remove the older log files.

Accessing Log Files

You can access and view the log files directly by going to the component log directory from a command prompt or a terminal application. You can also view the component log files through ECL Watch.

To view logs on ECL Watch, click on the **Operations** icon, then click on the **System Servers** link. That opens the System Servers page in ECL Watch. There are several HPCC system components listed on that page. In the **Directory** column for each component there is a computer drive icon. Click the icon in the row for the component log you wish to view.

Figure 4. Logs in ECL Watch



You can also view log files from the other links under the Operations icon in ECL Watch.

1. Click on the **Target Clusters** link to open the tab with links to your system's clusters.
2. Click on the computer drive icon (circled in red in the above figure), in the row of the cluster and node of the component log you wish to view.

To view cluster process logs:

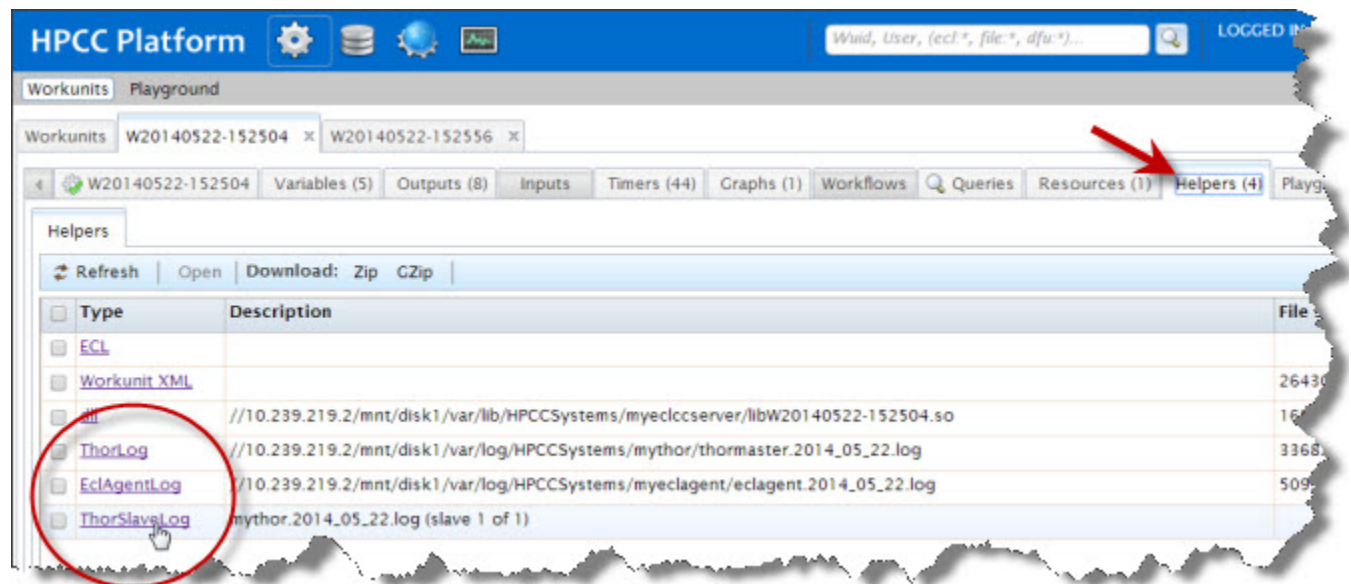
1. Click on the **Cluster Processes** link to open the tab with links to your system's clusters processes.
2. Click on the cluster process you wish to view more information about.

For example, click on the **myroxie** link. You will then see a page of all that components nodes. You will see computer drive icon, in the row of each node. Click that icon to see the logs for the cluster process for that node.

Log files in ECL Workunits

You can also access the Thor or ECL Agent log files from the ECL Workunits. (not available for Roxie workunits) In ECL Watch when examining the Workunit details, you will see a **Helpers** tab. Click on the Helpers tab to display the relevant log files for that particular workunit.

Figure 5. Logs in ECL Watch Workunits



Preflight

The first step in certifying that the platform is installed and configured properly is to run a preflight check on the components. This ensures that all machines are operating and have the proper executables running. This also confirms there is adequate disk space, available memory, and acceptable available CPU % values.

- Open ECL Watch in your browser using the following URL:

http://nnn.nnn.nnn.nnn:pppp (where nnn.nnn.nnn.nnn is your ESP Server's IP Address and pppp is the port. The default port is 8010)

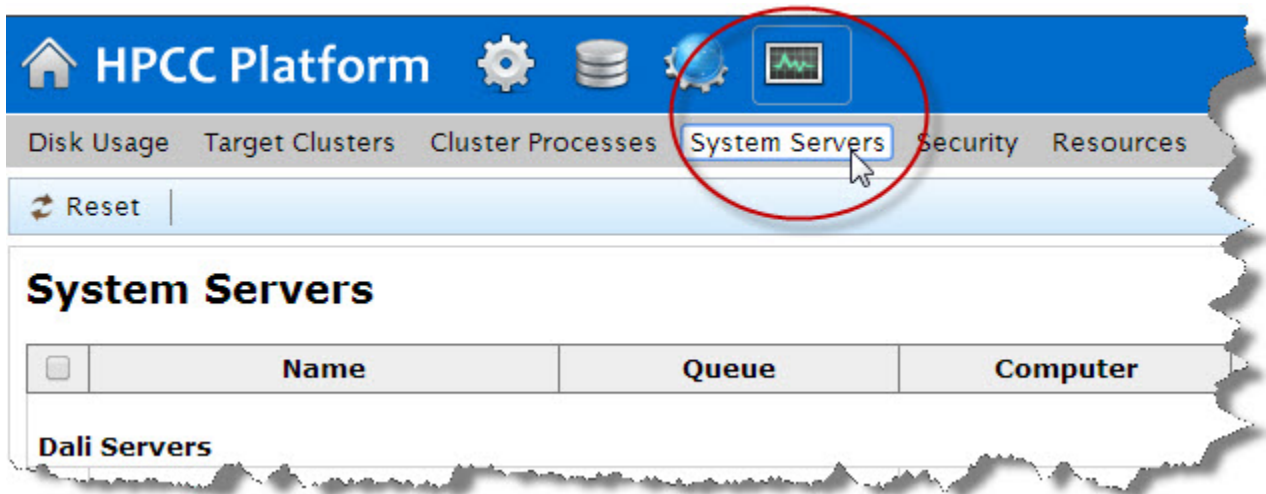


Note: That your IP address could be different from the ones provided in these figures. Please use the IP address provided by your installation.

Preflight System Servers

1. Click on the **Operations** icon then click on the **System Servers** link.

Figure 6. System Servers link



A screen similar to the following displays.

Figure 7. System Servers page

System Servers				
<input type="checkbox"/>	Name	Queue	Computer	Network
Dali Servers				
<input checked="" type="checkbox"/>	mydali		localhost	192.168.1.1
DFU Servers				
<input checked="" type="checkbox"/>	mydfuserver	dfuserver_queue	localhost	192.168.1.1
Drop Zones				
<input type="checkbox"/>	mydropzone		localhost	192.168.1.1
ECL Agents				
<input checked="" type="checkbox"/>	myeclagent		localhost	192.168.1.1

2. Press the **Submit** button at the bottom of this page to start preflight.

Figure 8. Submit

☒ Get storage information
☒ Local File Systems Only
☒ Get software information
☒ Show processes using filter
 Additional processes to filter:
☒ Auto Refresh every 5 mins

EXPECTED RESULTS:

After pressing Submit, a screen similar to the following displays.

Figure 9. System Component Information

Machine Information

<input checked="" type="checkbox"/>	Location	Component	Condition	State	Up Time	Processes Down	
<input checked="" type="checkbox"/>	10.239.219.3 /var/lib/HPCCSystems/myesp	Esp [myesp]	Normal	Ready	09:38	-	60%
<input checked="" type="checkbox"/>	10.239.219.3 /var/lib/HPCCSystems/myeclscheduler	Ecl Scheduler [myeclscheduler]	Normal	Ready	11:32	-	60%
<input checked="" type="checkbox"/>	10.239.219.3 /var/lib/HPCCSystems/myeclagent	Agent Exec [myeclagent]	Normal	Ready	11:35	-	60%
<input checked="" type="checkbox"/>	10.239.219.3 /var/lib/HPCCSystems/myeclccserver	Ecl CC Server [myeclccserver]	Normal	Ready	11:33	-	60%
<input checked="" type="checkbox"/>	10.239.219.4 /var/lib/HPCCSystems/mysasha	Sasha Server [mysasha]	Normal	Ready	11:51	-	60%
<input checked="" type="checkbox"/>	10.239.219.4 /var/lib/HPCCSystems/mydali	Dali Server [mydali]	Normal	Ready	11:54	-	60%
<input checked="" type="checkbox"/>	10.239.219.5 /var/lib/HPCCSystems/mydfuserver	Dfu Server [mydfuserver]	Normal	Ready	11:29	-	60%

☒ Select All / None
 Fetched: 11/11/11 14:13:09
 Action: Machine Information
☒ Get processor information Warn if CPU usage is over 95%
☒ Get storage information Available memory

This screen displays information on several system components. This information indicates whether several components are actually up and running appropriately. The resulting page shows useful information about each component. The component name, the condition, the component state, how long the component has been up and running, the amount of disk usage, memory usage and other information is available at a glance.

If there are any failed components, they are highlighted in orange, indicating they are not ready.

Figure 10. Failed Component

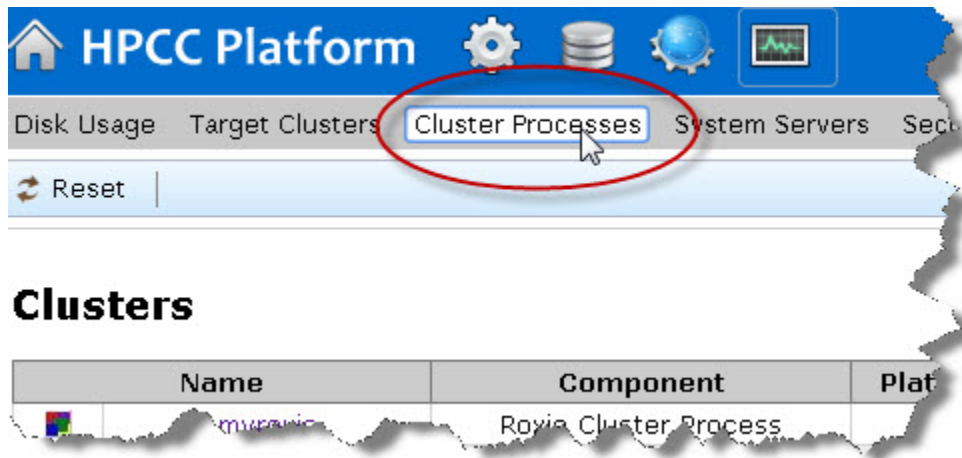
Machine Information

<input checked="" type="checkbox"/>	Location	Component	Condition	State	Up Time	Processes Down	/	/mnt/disk1	Physical Memory
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/mydali	Dali Server [mydali]	Normal	Ready	76 day(s) 00:50:02	-	43%	97%	96%
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/mydfuserver	Dfu Server [mydfuserver]	Warning	Unknown		mydfuserver	43%	97%	96%
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/myeclagent	Ecl Agent [myeclagent]	Normal	Ready	-	-	43%	97%	96%
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/myeclagent	Agent Exec [myeclagent]	Normal	Ready	76 day(s) 00:50:00	-	43%	97%	96%
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/myeclccserver	Ecl CC Server [myeclccserver]	Normal	Ready	76 day(s) 00:49:59	-	43%	97%	96%
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/myeclscheduler	Ecl Scheduler [myeclscheduler]	Normal	Ready	76 day(s) 00:49:57	-	43%	97%	96%
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/myesp	Esp [myesp]	Normal	Ready	76 day(s) 00:48:10	-	43%	97%	96%
<input checked="" type="checkbox"/>	10.239.219.2 /var/lib/HPCCSystems/mysasha	Sasha Server [mysasha]	Normal	Ready	76 day(s) 00:49:54	-	43%	97%	96%

Preflight Thor

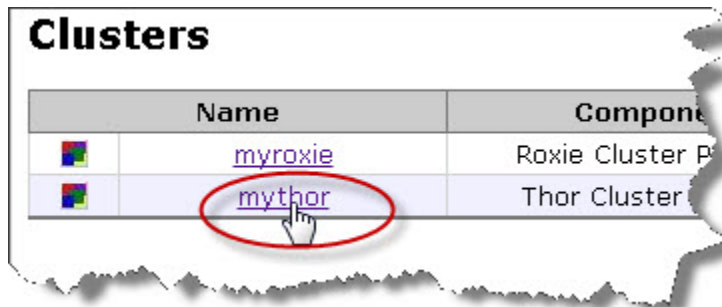
1. Click on the **Operations** icon then click on the **Cluster Processes** link.

Figure 11. Cluster Processes Link



2. Click on the **mythor** link.

Figure 12. mythor link



3. Check the **Select All** checkbox (if necessary).
4. Press the **Submit** button to start preflight.

Figure 13. Submit

☒ Get storage information
☒ Local File Systems Only
☒ Get software information
☒ Show processes using filter
 Additional processes to filter:
☒ Auto Refresh every 5 mins

EXPECTED RESULTS:

After pressing Submit, a screen similar to the following should display.

Figure 14. ESP mythor system component information

Thor Cluster 'mythor'

<input checked="" type="checkbox"/>	Location	Component	Slave Number	Condition	State	Up Time	Processes Down	/	/mnt/disk
<input checked="" type="checkbox"/>	10.239.219.4 /var/lib/HPCCSystems/mythor	Thor Slave [mythor]	2	Normal	Ready	03:17:11	-	51%	99%
<input checked="" type="checkbox"/>	10.239.219.5 /var/lib/HPCCSystems/mythor	Thor Slave [mythor]	1	Normal	Ready	03:17:11	-	51%	99%
<input checked="" type="checkbox"/>	10.239.219.3 /var/lib/HPCCSystems/mythor	Thor Master		Normal	Ready	03:17:11	-	51%	99%

☒ Select All / None
 Fetched: 06/13/14 11:56:33
 Action:

This screen displays information on Thor components. This information indicates whether the components are actually up and running appropriately. The resulting page shows useful information about each component. The component name, the condition, the component state, how long the component has been up and running, the amount of disk usage, memory usage and other information is available at a glance.

If your system has more than 1 Thor cluster, repeat these steps for each cluster.

If there are any failed components, they are highlighted in orange, indicating they are not ready.

Figure 15. Failed Component

Thor Cluster 'mythor'

<input checked="" type="checkbox"/>	Location	Component	Slave Number	Condition	State	Up Time	Processes Down	/	/mnt/disk1	Physic Memory
<input checked="" type="checkbox"/>	10.239.219.6 /var/lib/HPCCSystems/mythor	Thor Slave [mythor]	3	Warning	Unknown		mythor...	16%	95%	90%
<input checked="" type="checkbox"/>	10.239.219.5 /var/lib/HPCCSystems/mythor	Thor Slave [mythor]	2	Normal	Ready	04:32	-	52%	99%	97%
<input checked="" type="checkbox"/>	10.239.219.4 /var/lib/HPCCSystems/mythor	Thor Slave [mythor]	1	Normal	Ready	04:32	-	52%	99%	96%
<input checked="" type="checkbox"/>	10.239.219.3 /var/lib/HPCCSystems/mythor	Thor Master		Normal	Ready	04:32	-	51%	99%	97%

☒ Select All / None
Fetched: 11/08/12 11:28:41

Action: Machine Information ▾

☒ Get processor information Warn if CPU usage is over %

☒ Get storage information Warn if available memory is under % ▾

☒ Local File Systems Only

☒ Get software information Warn if available disk space is under % ▾

☒ Show processes using filter

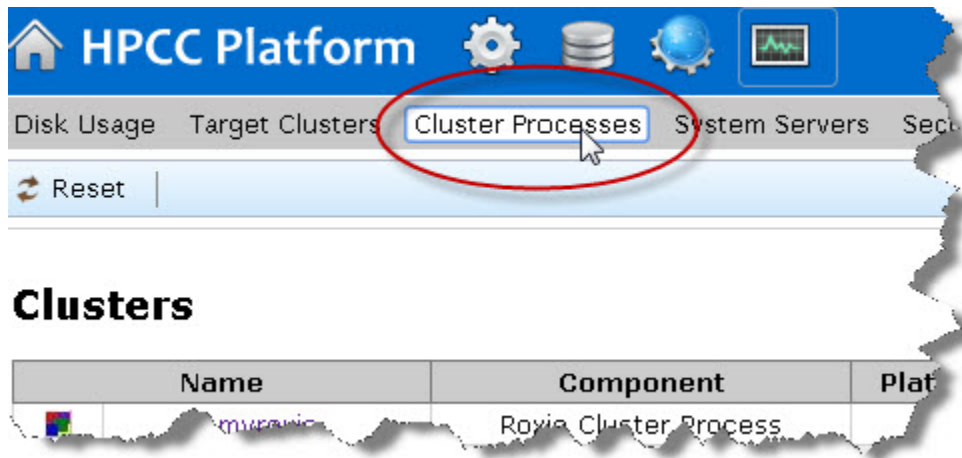
Additional processes to filter:

☐ Auto Refresh every mins.

Preflight the Roxie Cluster

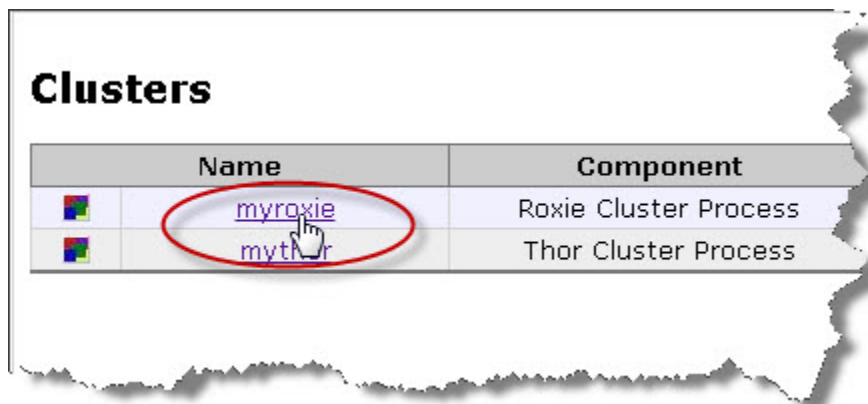
1. Click on the **Operations** icon then click on the **Cluster Processes** link.

Figure 16. Cluster Processes Link



2. Click on the **myroxie** link.

Figure 17. myroxie link



3. Press the **Submit** button to start preflight.

EXPECTED RESULTS

After pressing Submit, a screen similar to the following should display.

Figure 18. Roxie system information

Roxie Cluster 'myroxie'

<input checked="" type="checkbox"/>	Location	Component	Condition	State	Up Time	Processes Down	/
<input checked="" type="checkbox"/>	10.239.219.5 /var/lib/HPCCSystems/myroxie	Roxie Server	Normal	Ready	6 day(s) 23:27:08	-	51%
<input checked="" type="checkbox"/>	10.239.219.4 /var/lib/HPCCSystems/myroxie	Roxie Server	Normal	Ready	6 day(s) 23:27:10	-	51%

☒ **Select All / None**
Fetches: 06/13/14 12:09:27

Action: Machine Information ▾

☒ Get processor information Warn if CPU usage is over %

☒ Get storage information Warn if available memory is under % ▾

☒ Local File Systems Only

☒ Get software information Warn if available disk space is under % ▾

☒ Show processes using filter

Additional processes to filter:

☐ Auto Refresh every mins

This indicates whether the Roxie nodes are up and running, and some information about them.

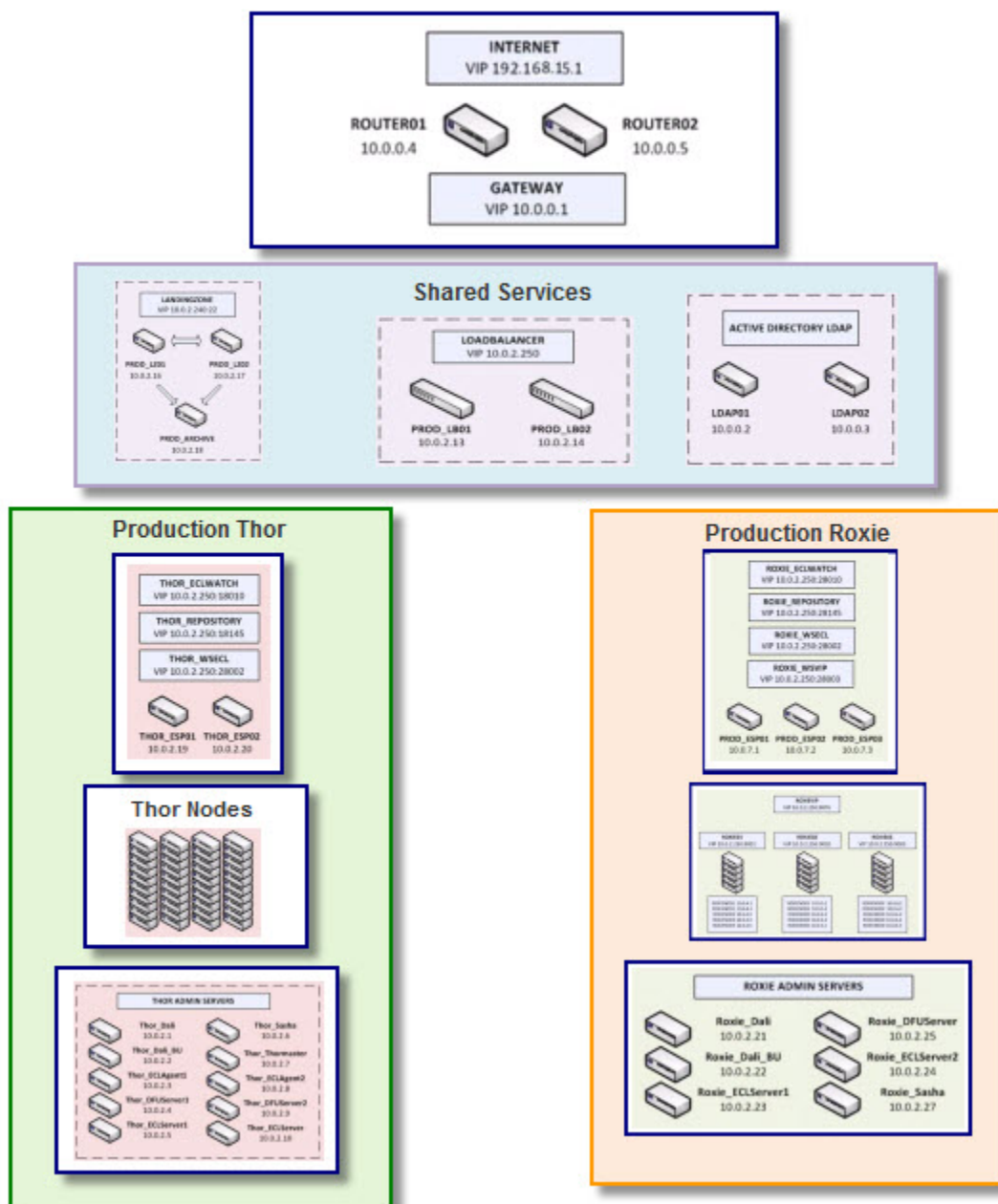


If your system has more than 1 Roxie cluster, repeat these steps for each cluster.

System Configuration and Management

The HPCC system requires configuration. The Configuration Manager tool (configmgr) included with the system software is a valuable piece of setting up your HPCC system. The Configuration Manager is a graphical tool provided that can be used to configure your system. Configuration Manager has a wizard that you can run which will easily generate an environment file to get you configured, up and running quickly. There is an advanced option available through Configuration Manager which allows for a more specific configuration, while still using the graphical interface. If desired you can edit the environment files using any xml or text editor however the file structure must remain valid.

Figure 19. Sample Production Configuration



Configuration Manager is the utility with which we configure the HPCC platform. The HPCC platform's configuration is stored in an XML file named **environment.xml**. Once you generate an environment (xml) file, it gets saved into a source directory (default is **/etc/HPCCSystems/source**). You then need to stop the system to copy it into the active HPCC directory, then distribute it into place on to each node and restart the HPCC system. At no time during configuration do you work on the live environment file.

When you install the HPCC system package, a default single-node environment.xml file is generated. After that, you can use the Configuration Manager to modify it and/or create a different environment file to configure components, or add nodes. There is a Configuration Manager wizard to help create an environment file. Give any environment file you create a descriptive name that would indicate what it is for in the source. For example, you might create an environment without a Roxie, you could call that file *environmentNoRoxie.xml*.

You would then copy the new configuration file you generate from the source directory to the **/etc/HPCCSystems** directory. Rename the file to `environment.xml`, and restart the system in order to reconfigure your system.

Configuration Manager also offers an **Advanced View** which allows more granularity for you to add instances of components or change the default settings of components for more advanced users. Even if you plan to use the Advanced View, it is a good idea to start with a wizard generated configuration file and use Advanced View to edit it.

More information and specific details for each Configuration Manager component and attributes of those components is detailed in *Using Configuration Manager*.

Running the Configuration Manager

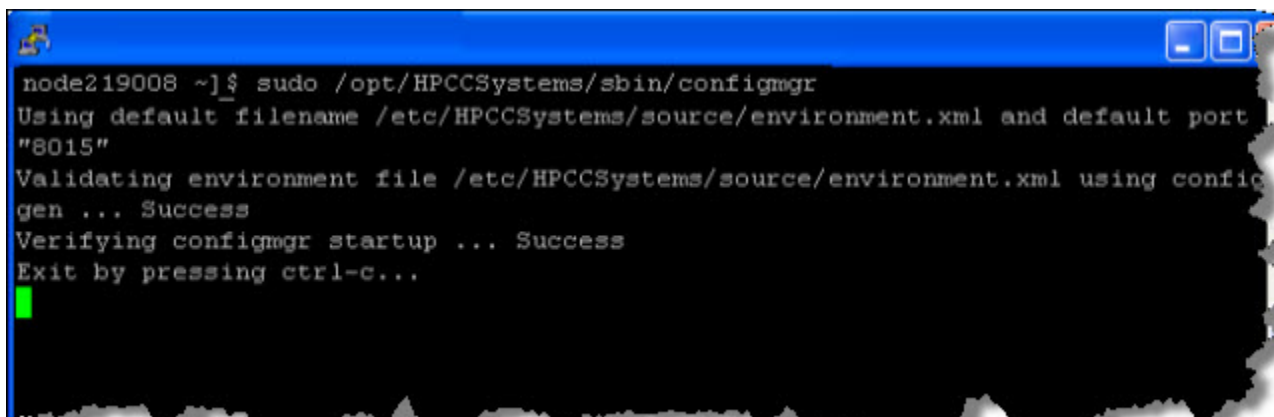
This section will guide you through configuring an HPCC environment using the Configuration Manager.

The HPCC package should already be installed on ALL nodes.

You can use any tool or shell script you choose.

1. SSH to a node in your environment and login as a user with sudo privileges. We would suggest that it would be the first node, and that it is a support node, however that is up to your discretion.
2. Start the Configuration Manager service on the node (again we would suggest that it should be on a support node, and further that you use the same node to start the Configuration Manager every time, but this is also entirely up to you).

```
sudo /opt/HPCCSystems/sbin/configmgr
```



3. Using a Web browser, go to the Configuration Manager's interface:

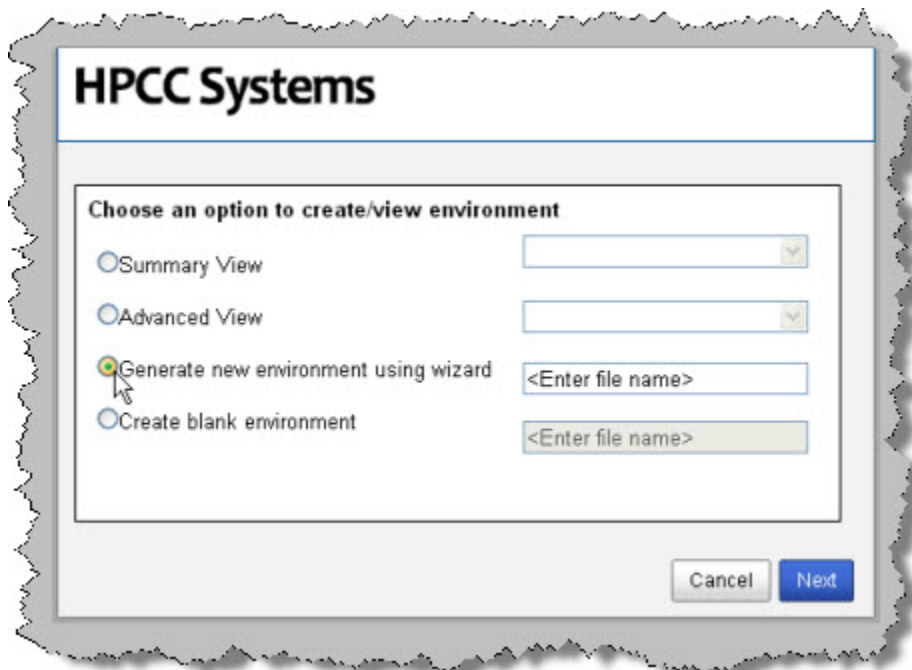
```
http://<ip of installed system>:8015
```

The Configuration Manager startup wizard displays.

There are different ways to configure your HPCC system. You can use the **Generate environment wizard** and use that environment or experienced users can then use the **Advanced View** for more specific customization. There is also the option of using **Create blank environment** to generate an empty environment that you could then go in and add only the components you would want.

Environment Wizard

1. To use the wizard select the **Generate new environment using wizard** button.



2. Provide a name for the environment file.

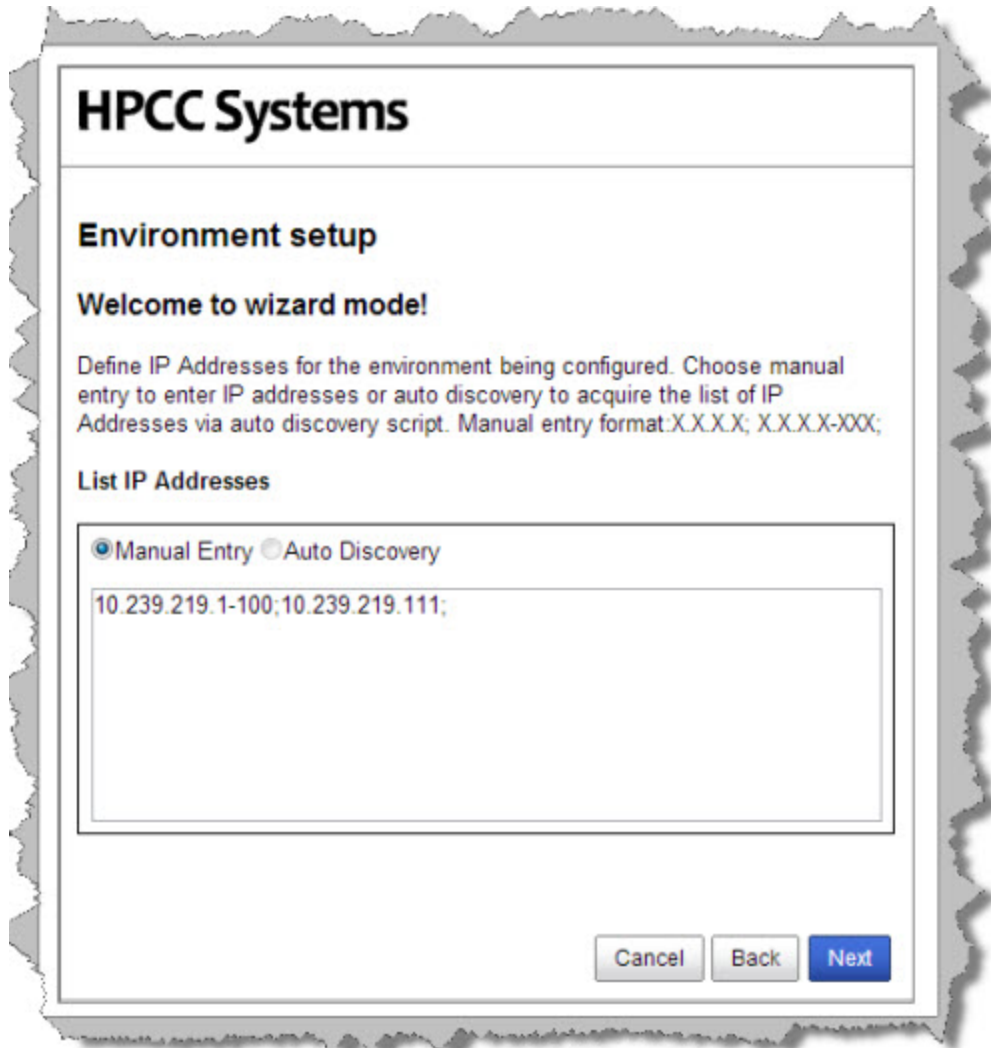
This will then be the name of the configuration XML file. For example, we will name our environment *NewEnvironment* and this will produce a configuration XML file named *NewEnvironment.xml* that we will use.

3. Press the Next button.

Next you will need to define the IP addresses that your HPCC system will be using.

4. Enter the IP addresses.

IP Addresses can be specified individually using semi-colon delimiters. You can also specify a range of IPs using a hyphen (for example, nnn.nnn.nnn.x-y). In the image below, we specified the IP addresses 10.239.219.1 through 10.239.219.100 using the range syntax, and also a single IP 10.239.219.111.



5. Press the Next button.

Now you will define how many nodes to use for the Roxie and Thor clusters.

6. Enter the appropriate values as indicated.

HPCC Systems

Environment setup

Enter number of nodes for Roxie and Thor clusters. No Roxie/Thor cluster will be generated for zero (0) number of nodes.

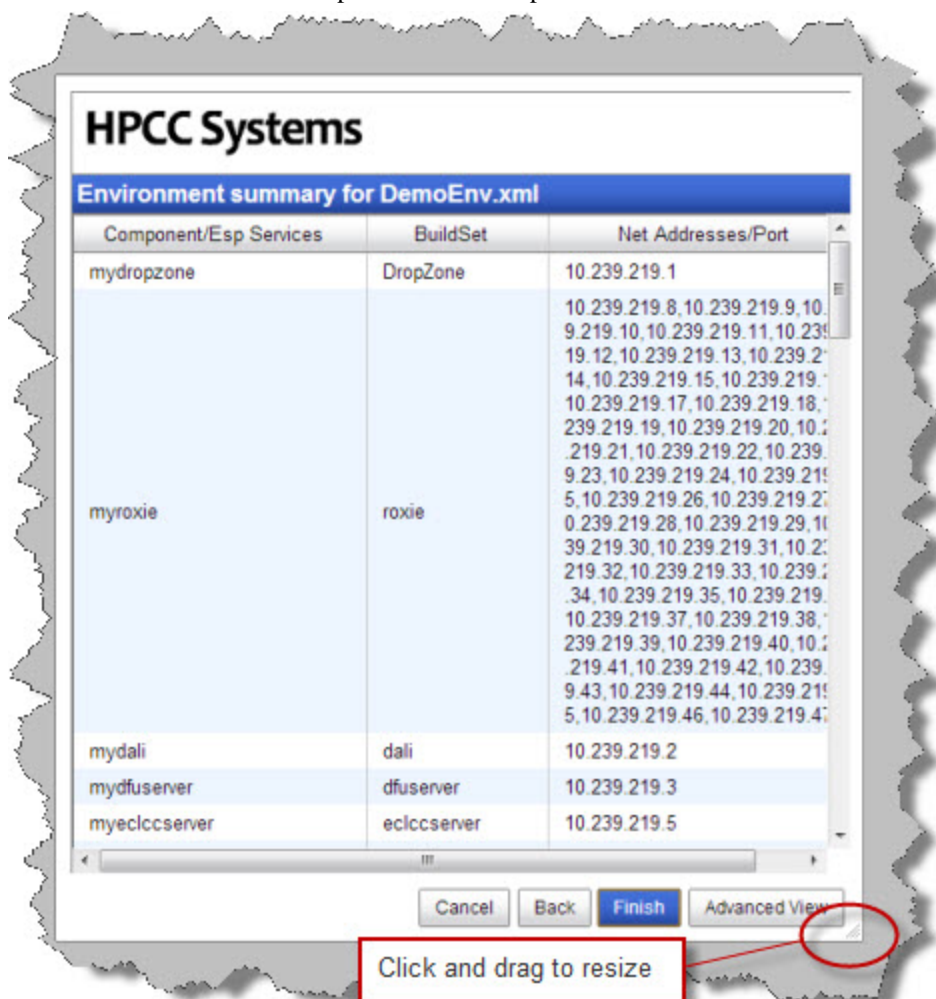
Number of support nodes	<input type="text" value="0"/>
Number of nodes for Roxie cluster	<input type="text" value="0"/>
Number of slave nodes for Thor cluster (A Thor Master will be added to the cluster and assigned to a support node)	<input type="text" value="1"/>
Number of Thor slaves per node (default 1)	<input type="text" value="1"/>
Enable Roxie on demand	<input checked="" type="checkbox"/>

Number of support nodes:	Specify the number of nodes to use for support components. The default is 1.
Number of nodes for Roxie cluster:	Specify the number of nodes to use for your Roxie cluster. Enter zero (0) if you do not want a Roxie cluster.
Number of slave nodes for Thor cluster	Specify the number of slave nodes to use in your Thor cluster. A Thor master node will be added automatically. Enter zero (0) if you do not want any Thor slaves.
Number of Thor slaves per node (default 1)	Specify the number of Thor slave processes to instantiate on each slave node. Enter zero (0) if you do not want a Thor cluster.
Enable Roxie on demand	Specify whether or not to allow queries to be run immediately on Roxie. This must be enabled to run the debugger. (Default is true)

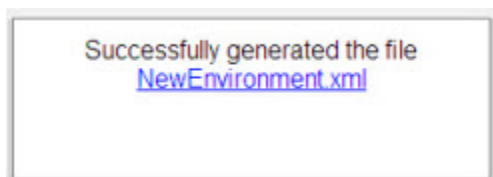
7. Press the **Next** button

The wizard displays the configuration parameters.


8. Press the **Finish** button to accept these values or press the **Advanced View** button to edit in advanced mode.



You will now be notified that you have completed the wizard.



At this point, you have created a file named NewEnvironment.xml in the **/etc/HPCCSystems/source** directory



Keep in mind, that your HPCC configuration may be different depending on your needs. For example, you may not need a Roxie or you may need several smaller Roxie clusters. In addition, in a production [Thor] system, you would ensure that Thor and Roxie nodes are dedicated and have no other processes running on them. This document is intended to show you how to use the configuration tools. Capacity planning and system design is covered in a training module.

Distribute the Configuration

1. Stop the HPCC system.

If it is running stop the HPCC system (on every node), using a command such as this:

```
sudo /sbin/service hpcc-init stop
```

Note: You may have a multi-node system and a custom script such as the one illustrated in Appendix of the [Installing and Running the HPCC Platform](#) document to start and stop your system. If that is the case please use the appropriate command for stopping your system on every node.



Be sure HPCC is stopped before attempting to copy the environment.xml file.

2. Back up the original environment.xml file.

```
# For example
sudo -u hpcc cp /etc/HPCCSystems/environment.xml /etc/HPCCSystems/source/environment-date.xml
```

Note: The live environment.xml file is located in your **/etc/HPCCSystems/** directory. Configuration Manager works on files in **/etc/HPCCSystems/source** directory. You must copy from this location to make an environment.xml file active.

You can also choose to give the environment file a more descriptive name, to help differentiate any differences.

Having environment files under source control is a good way to archive your environment settings.

3. Copy the new .xml file from the source directory to the /etc/HPCCSystems and rename the file to *environment.xml*

```
# for example
sudo -u hpcc cp /etc/HPCCSystems/source/NewEnvironment.xml /etc/HPCCSystems/environment.xml
```

4. Copy the **/etc/HPCCSystems/environment.xml** to the **/etc/HPCCSystems/** on to *every* node.

You may want to use a script to push out the XML file to all nodes. See the *Example Scripts* section in the Appendix of the [Installing and Running the HPCC Platform](#) document. You can use the scripts as a model to create your own script to copy the environment.xml file out to all your nodes.

5. Restart the HPCC platform on all nodes.

Environment.conf

Another component of HPCC system configuration is the environment.conf file. Environment.conf contains some global definitions that the configuration manager uses to configure the HPCC system. In most cases, the defaults are sufficient.



WARNING: These settings are essential to proper system operation. Only expert level HPCC administrators should attempt to change any aspects of this file.

By default the environment.conf file is located:

```
/etc/HPCCSystems
```

Environment.conf is required upon startup of HPCC. The environment.conf is where the HPCC environment file is defined.

```
/opt/HPCCSystems/environment.xml
```

This is also where the working path is defined.

```
path=/opt/HPCCSystems
```

The working path is used by several aspects of the application, changing this could cause needless complications. By default the application installs there, and sets many resources to that as well.

The default environment.conf:

```
## HPCC Systems default environment configuration file

[DEFAULT SETTINGS]
configs=/etc/HPCCSystems
path=/opt/HPCCSystems
classpath=/opt/HPCCSystems/classes
runtime=/var/lib/HPCCSystems
lock=/var/lock/HPCCSystems
# Supported logging fields: AUD,CLS,DET,MID,TIM,DAT,PID,TID,NOD,JOB,USE,SES,
#                           COD,MLT,MCT,NNT,COM,QUO,PFX,ALL,STD
logfields=TIM+DAT+MLT+MID+PID+TID+COD+QUO+PFX
pid=/var/run/HPCCSystems
log=/var/log/HPCCSystems
user=hpcc
group=hpcc
home=/Users
environment=environment.xml
sourcedir=/etc/HPCCSystems/source
blockname=HPCCSystems
interface=*
# enable epoll method for notification events (true/false)
use_epoll=true
```

Path considerations

Most of the directories are defined as absolute paths:

```
configs=/etc/HPCCSystems
path=/opt/HPCCSystems
classpath=/opt/HPCCSystems/classes
runtime=/var/lib/HPCCSystems
```

```
lock=/var/lock/HPCCSystems
```

HPCC will not run properly without the proper paths, and in some cases needs the absolute path. If a process or component can't find a path you will get an error message such as the following:

```
"There are no components configured to run on the node..."
```

If the path changes from HPCCSystems, it does NOT change in the environment.xml file. Any changes would require manually modifying the environment.xml file.

The log file, *hpcc-init.log* is written to the HPCCSystems path.

Other Environment.conf items

Some other items used by or referred to in environment.conf.

Use_epoll It is an event mechanism to achieve better performance in more demanding applications where number of watched file descriptors is large.

Logfields Categories available to be logged. These consist of Time(TIM), Date(DAT), Process ID (PID), Thread ID (TID), etc.

Interface In the default environment.conf there is a value for interface. The default value for that is:

```
interface=*
```

The default value of * assigns the interface to an open ip address, in any order. Specifying an interface, such as Eth0, will assign the specified node as the primary.

User Security Maintenance

Configuring an HPCC System to use LDAP security will give you greater control over users and the security of your HPCC system.

Introduction

HPCC systems maintain security in a number of ways. HPCC Systems can be configured to manage users' security rights by pointing either at Microsoft's Active Directory on a Windows system, or OpenLDAP on Linux systems.

Using the Permissions interface in ECL Watch, administrators can control access to features in ECL IDE, ECL Watch, ECL Plus, DFU Plus, and the ECL modules within the Attribute Repository. Additional "file access control" can be implemented over data files by configuring the Dali server to point to the Active Directory/LDAP server. This is what is known as enabling file security.

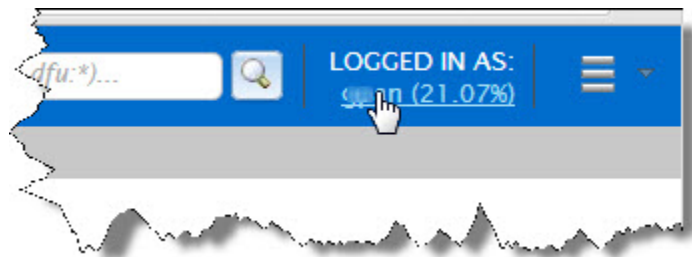
Permissions are established by group or by user and are defined as they are associated with a particular feature of the HPCC System. Only one set of permissions can be defined for each unique combination of group and feature. Permissions are separated into the following categories:

Esp Features for SMC	Controls access to features in ECLWatch and similar features accessed from ECL IDE.
File Scopes	Controls access to data files by applying permissions to File scopes
Workunit Scopes	Controls access to Workunits by applying permissions to Workunit scopes
Esp Features for WsEclAccess	Controls access to the WS-ECL web service
Repository Modules	Controls access to the Attribute Repository and Modules in the repository
Esp Features for EclDirectAccess	Controls access to the ECLDirect web service

Access to the permission settings for each of these areas is available using the **Security** area in ECL Watch.

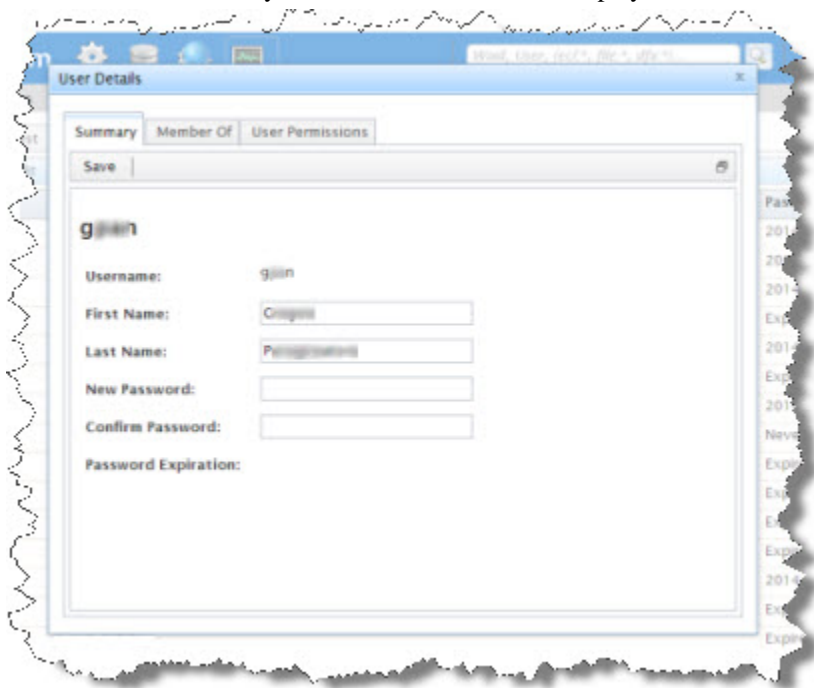
Information about your account

To find out more information about your account, in ECL Watch click on the **Logged In As:** link at the top of the ECL Watch page.



1. Click on the **Logged In As:** link.

A User Details tab with your account information is displayed.



2. Confirm the User Name that you are logged in as.

Note that Administrator rights are needed to manage users and permissions.

Ensure you are using an account with Administrator rights if you intend to manage users or permissions.

3. Verify the password expiration date, or if password is set to expire.

Security Administration using ECL Watch

Administrator rights are needed to manage permissions. Once you have administrator access rights, open ECL Watch in your browser using the following URL:

- **<http://nnn.nnn.nnn.nnn:pppp>**(where **nnn.nnn.nnn.nnn** is your ESP Server's IP Address and **pppp** is the port. The default port is 8010). For example: <http://10.150.51.27:8010/>.

Security administration is controlled using the **Security** area of ECL Watch. There are three areas where permissions may be set:

- **Users.** Shows all the users currently setup. Use this area to add or delete a user, edit a user's details, set/reset a user's password and view the permissions currently assigned to a user.
- **Groups.** Shows all the groups currently setup. Use this area to add or delete a group, view and edit the member of a group, view and edit the permissions that have been set for a group.
- **Permissions.** Shows the features of the HPCC System where permissions may be set. Use this area to view the permissions currently set for any area of the HPCC System, or to add groups and users and set/modify their permission for a specific feature



NOTE: Use caution when setting any explicit **deny** permission setting. In general, the most restrictive permission is applied.

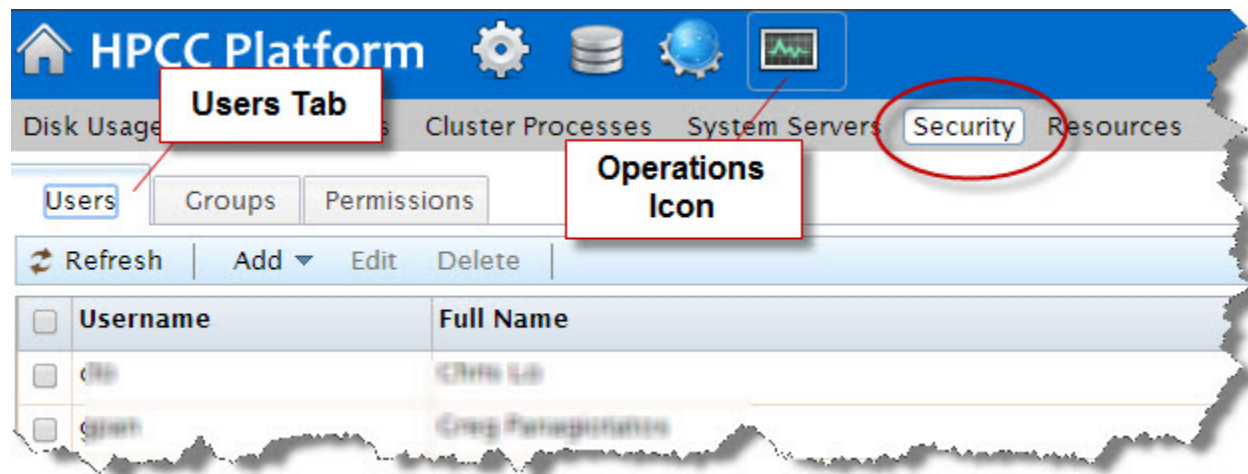
Setting and modifying user permissions

Access to ECL Watch and its features is controlled using a login and password. The **Users** area enables you to control who has access to ECL Watch and the features of your HPCC System to which they have access. Permissions can be set for users based on their individual needs and users can also be added to groups which have already been set up. Use the **Users** menu item to:

- Add a new user
- Delete a user
- Add a user to a group
- Change a user's password
- Modify the details/permissions of an individual user

Adding and editing users

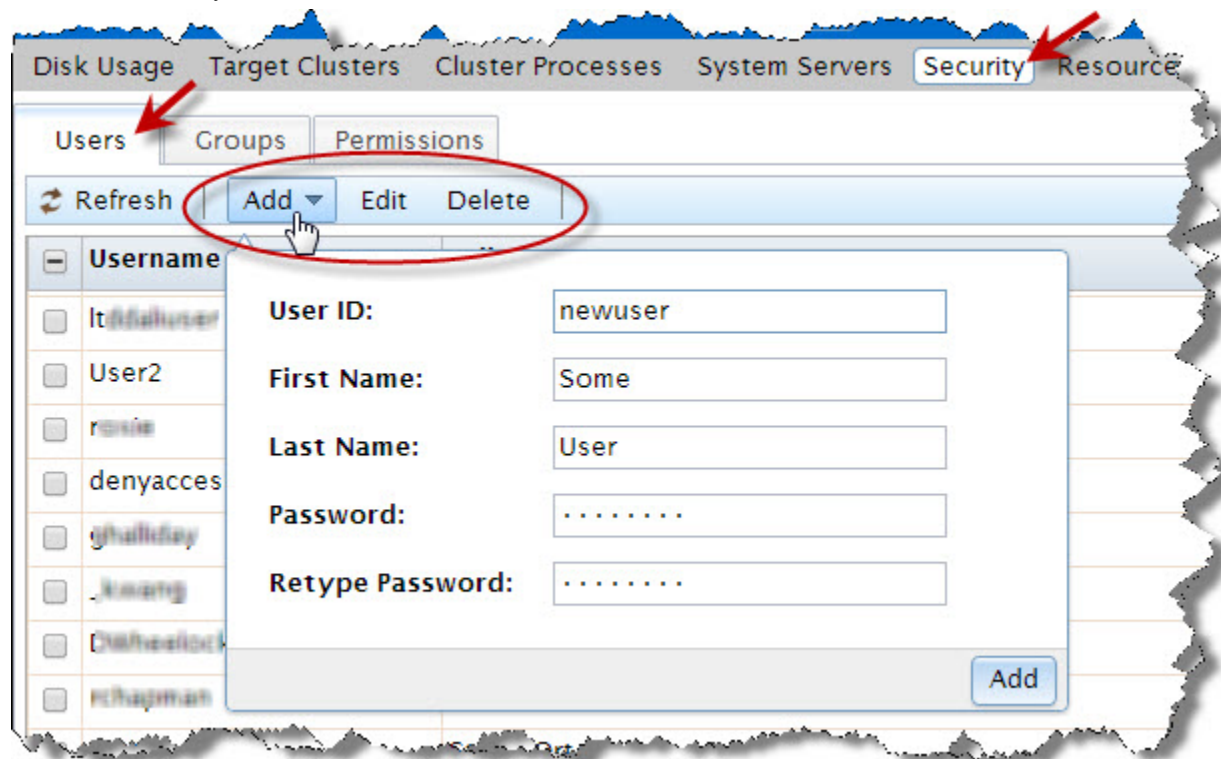
To access the permissions page click on the **Operations** icon, then click the **Security** link from the navigation sub-menu. Click on the **Users** tab to add or edit users.



All current users are identified in the list by their Username and full name.

To add a new user to the list of authenticated users:

To add a new user you must have Administrator level access.



1. Press the **Add** button.

The add user dialog is displays.

2. Enter a **Username**.

This is the login name for using ECL Watch, ECL IDE, WSECL, etc.

3. Enter the **First Name** and **Last Name** of the user.

This information helps to easily identify the user and is displayed in the **Full Name** field on the main **Users** window.

4. Enter a **Password** for the user and then confirm it in the **Retype Password** field.

5. Press **Add**.

Confirmation of the user request opens a new tab where you can verify the user's information.

6. Press **Save**.

Once added, the new user displays in the list and you can modify details and set permissions as required.

To modify a user's details:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

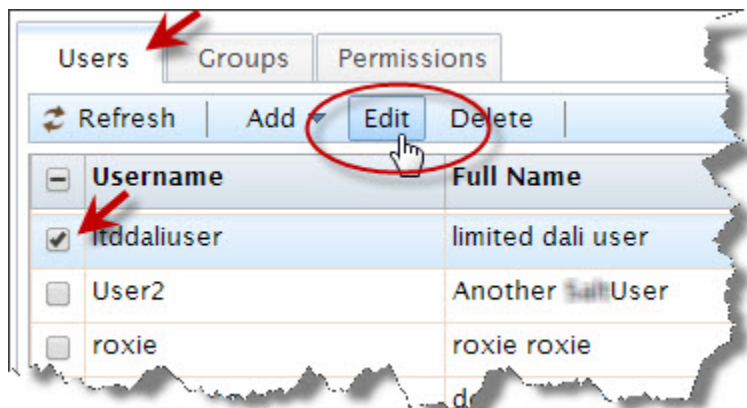
1. Click on the **Users** tab.

The users display in a list.

2. Select the user (or users) to modify. Check the box next to the Username to select.

This will enable the Users action buttons.

3. Press the **Edit** action button.



You will see a tab open for each user selected. On that tab there are three sub-tabs.

The user details are on the **Summary** tab.

4. Modify the user's details as required (if more than one user selected, repeat for each user).

Note: The **Username** cannot be changed.

5. Press **Save**.

Confirmation message displays.

To add a user to a group:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

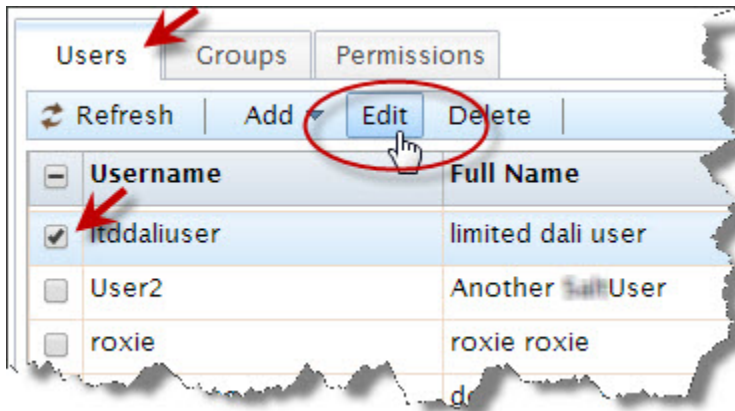
1. Click on the **Users** tab.

The users display in a list.

2. Select the user (or users) to modify. Check the box next to the Username.

This will enable the user action buttons.

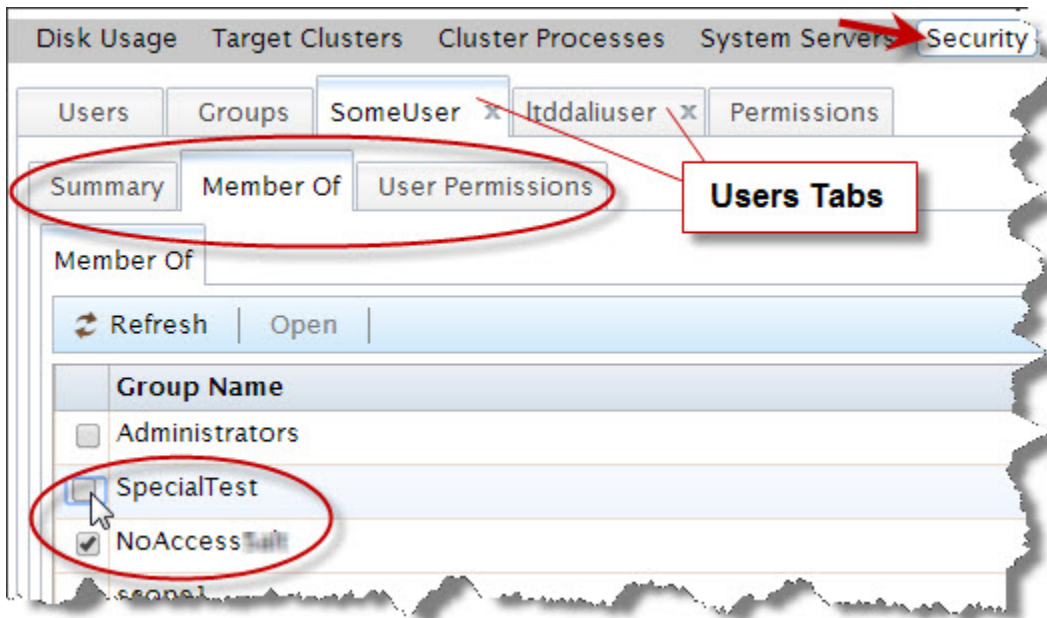
3. Press the **Edit** action button.



A new tab opens for each user selected. On that tab there are three sub-tabs.

4. Click on the tab for the user to modify (if more than one user selected, repeat for each user).

On the user's tab there are three sub-tabs.



Click on the **Member Of** sub-tab to modify that user's groups.

5. On the **Member Of** tab for that user, you will see a list of the available groups.

There is a check in the box next to each group that user belongs to.

To add that user to a group, check the box next to the desired group.

6. The changes are automatically saved. Close the tab.

To delete a user from a group:

To delete a user you must have Administrator level access.

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Users** tab.

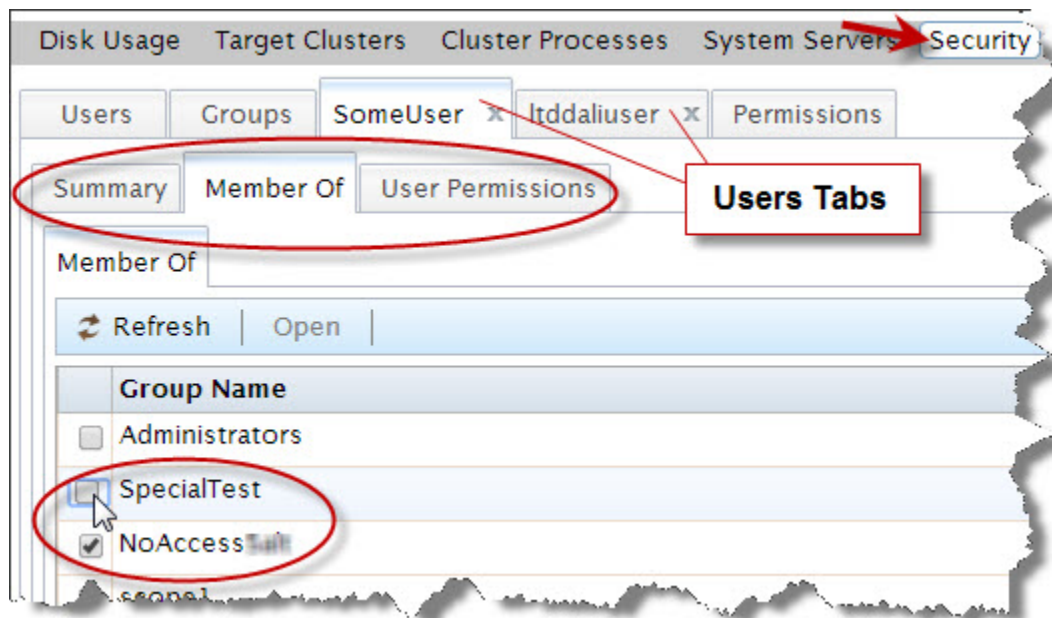
The users display in a list.

2. Select the user (or users) to modify. Check the box next to the Username.

This will enable the user action buttons. Press the **Edit** action button to modify settings for that user.

3. Click on the tab for the user to modify (if multiple users selected, repeat for each user).

On the user's tab there are three sub-tabs.



Click on the **Member Of** sub-tab to modify that user's groups.

4. On the **Member Of** tab for that user, there is a list of the available groups.

There is a check in the box next to each group that user belongs to.

To remove that user from a group, uncheck the box next to the desired group.

5. The changes are automatically saved. Close the tab.

To change a user's password:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

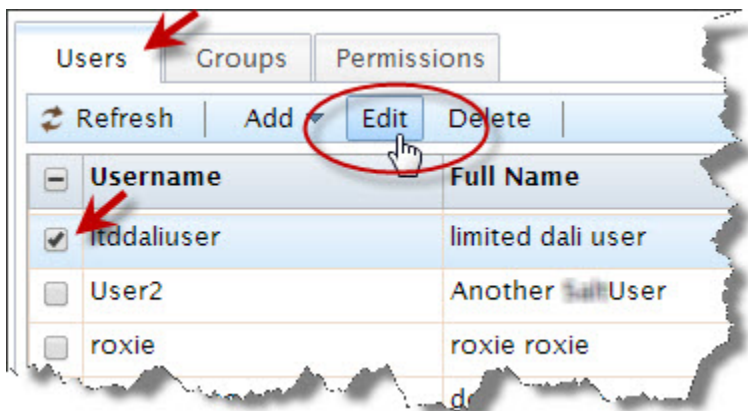
1. Click on the **Users** tab.

The users display in a list.

2. Select the user (or users) to modify. Check the box next to the Username to select.

This will enable the Users action buttons.

3. Press the **Edit** action button.



You will see a tab open for each user selected. On that tab there are three sub-tabs.

The user details are on the **Summary** tab.

4. Change the password in the **Password** and **Retype New Password** fields as required on the User details summary tab (if multiple users selected, repeat for each user).

Note: The **Username** cannot be changed.

5. Press **Save**.

A confirmation message displays.

To delete a user from the list of authenticated users:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Users** tab.

The users display in a list.

2. Check the box to the left of the user(s) you want to remove.

Note: These users will no longer have access to ECL Watch.

3. Click **Delete**.

Confirmation of the request is shown.

Setting permissions for an individual user

There may be occasions when you need to modify the permissions for individual users. For example, users may have individual security needs that are not completely covered in any group or, there may be occasions when a user requires

temporary access to an HPCC feature. Permissions set in this area of ECL Watch only affect the user you choose and any permissions you set here overwrite those set in any group to which the user belongs.

To set new permissions for an individual user:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Users** tab.

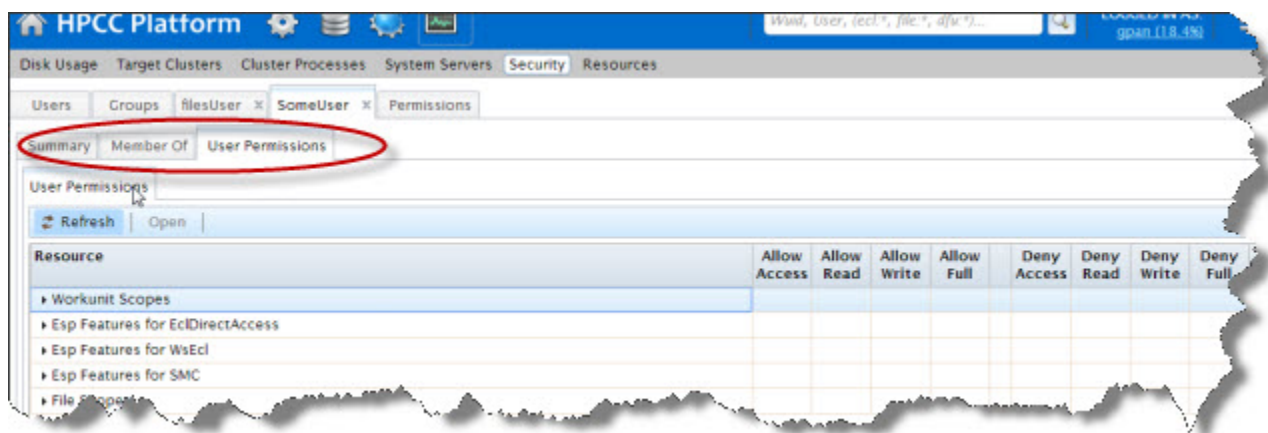
The users display in a list.

2. Select the user (or users) to modify. Check the box next to the Username to select.

This will enable the Users action buttons.

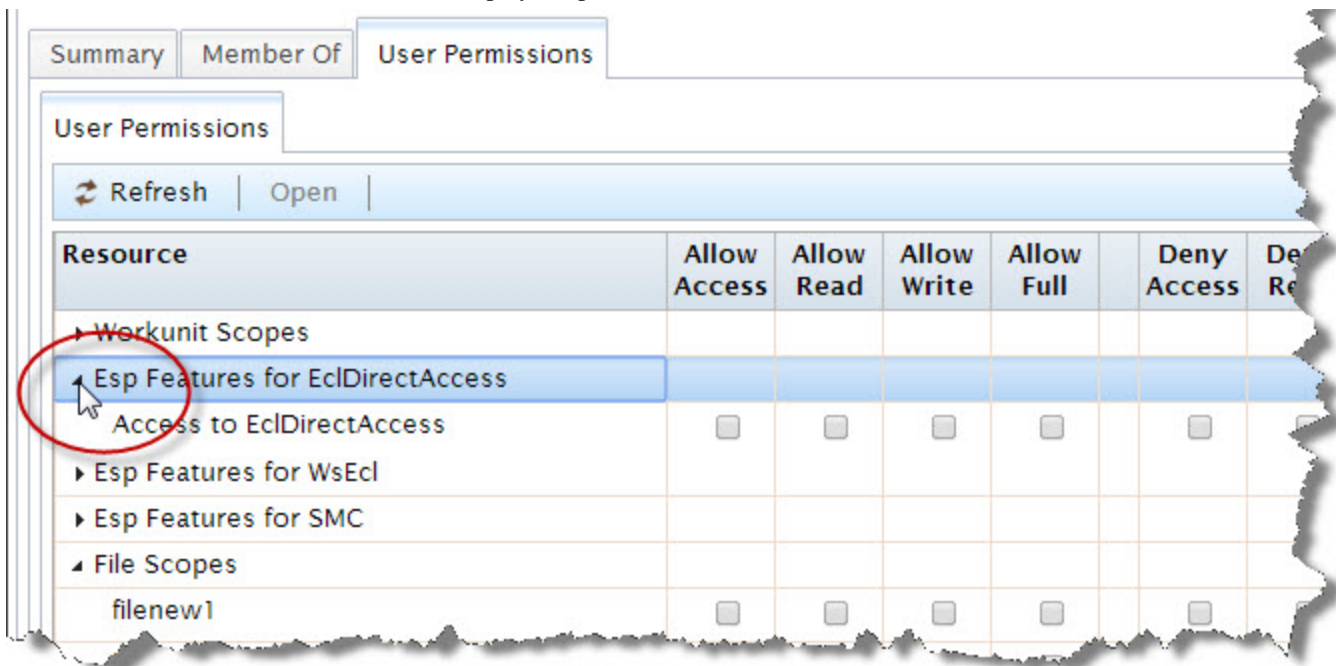
3. Click on the tab for the username to modify (if multiple users selected, repeat for each user).

On the user's tab there are three sub-tabs.



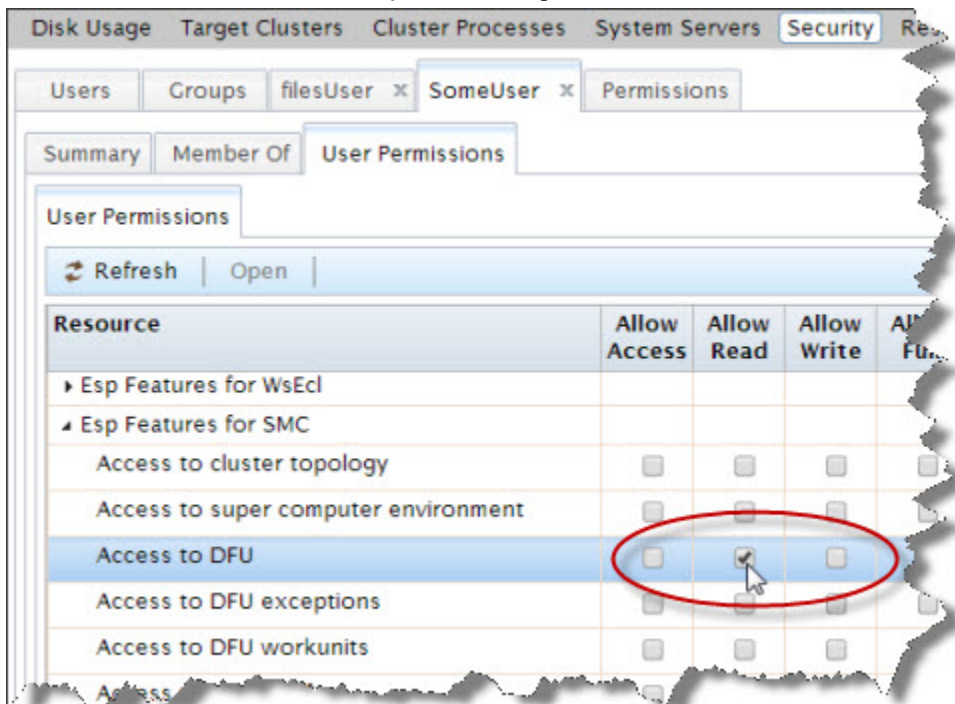
Click on the **User Permissions** sub-tab to modify that user's permissions.

- Click on the arrow next to the resource to display the permissions for that resource.



The list of permission groups currently set for this user and the ones the user has inherited are also listed. Click the arrow to allow setting the individual resource settings.

- There may be more than one resource setting available in each group, be sure to set the permissions for each setting as required.
- Check the boxes that **allow** and **deny** access as required for the user.





NOTE: Use caution when setting any explicit **deny** permission setting. In general, the most restrictive permission is applied.

7. The changes are automatically saved. Close the tab.

To modify permissions for an individual user:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Users** tab.

The users display in a list.

2. Select the user (or users) to modify. Check the box next to the Username to select.

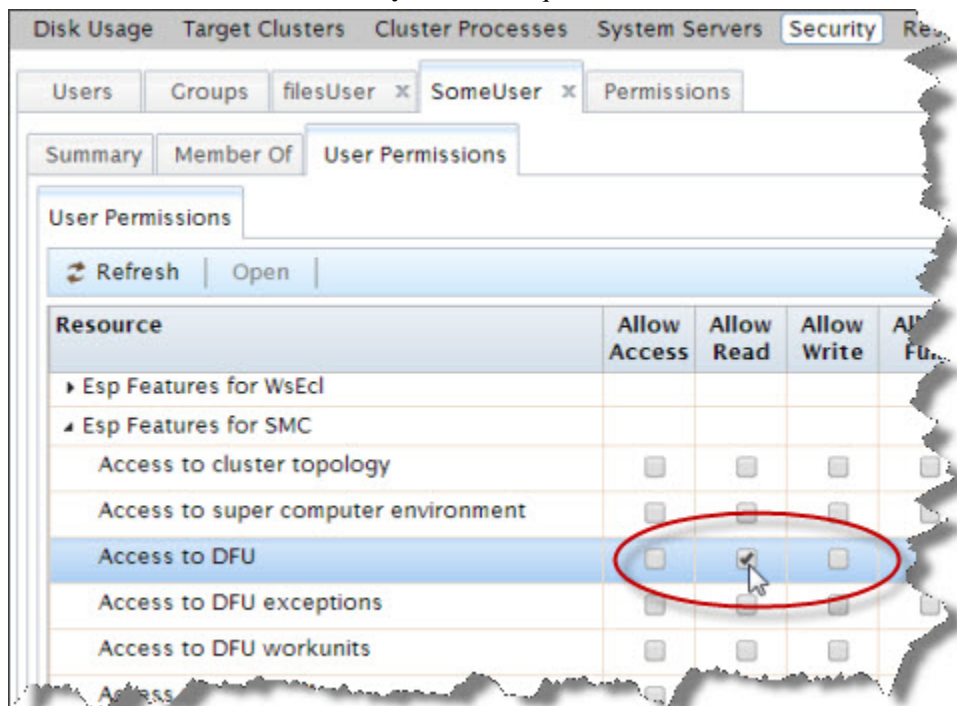
This will enable the Users action buttons.

3. Click on the tab for the user to modify (if multiple users selected, repeat for each user).

On the user's tab there are three sub-tabs. Click on the **User Permissions** sub-tab to modify that user's permissions.

4. Locate the feature you want to modify. Click on the arrow next to the resource to display the permission set for that resource.

5. Check the boxes that **allow** and **deny** access as required for the user.



NOTE: Use caution when setting any explicit **deny** permission setting. In general, the most restrictive permission is applied.

6. The changes are automatically saved. Close the tab.

Setting and modifying group permissions

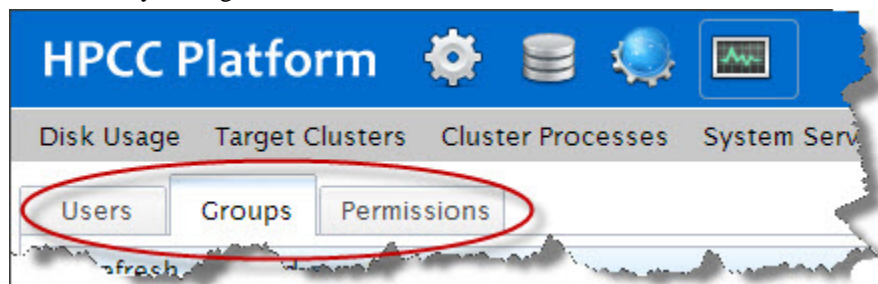
Setting up groups ensures that all users with the same permission needs have the same permission settings. You can give users the access they require to the feature areas of HPCC they need. There is no limit to the number of groups you can create. You can create as many groups as you need to control access for all your users regardless of their tasks.

Use the **Groups** menu item to:

- Add a new group
- Delete a group
- Add members to a groups
- Modify the permissions for a group

Adding and editing groups

When adding or changing the permissions for a group, all members of that group are given those permission settings. So it is important to be sure that you are giving or denying access to features appropriate for the members of that group. If you need to make a change for a single user (or small number of users), it is probably better to make that change for each individual user as illustrated in the previous sections. Since individual permission settings take precedence over the group settings, you can safely change the individual settings for a user without affecting the rest of the group(s) to which they belong.

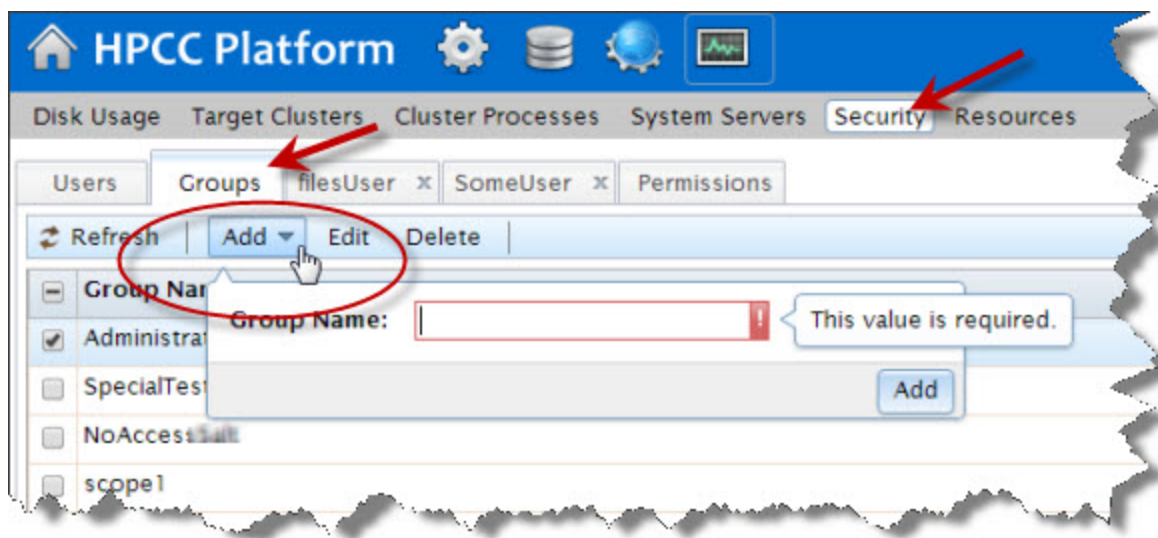


To modify groups, click on the **Operations** icon, then click the **Security** link from the navigation sub-menu. Click on the **Groups** tab.

To add a new group:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Groups** tab.
2. Press the **Add** action button.



This opens a dialog where you can enter the name for the group.

3. Enter a **Group Name**.
4. Press **Add**.

This will now open a **Summary** tab for this new group.

You can set the permissions and add members to this group from the respective sub-tabs on that group tab.

To delete a group:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Groups** tab.
2. Locate the group in the list and check the checkbox next to it.
3. Press the **Delete** action button.
4. Press **OK** to confirm.

The group no longer displays in the list.

To add new members to a group:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Groups** tab.
2. Locate the group in the list and check the box next to it.
3. Press **Edit** action button.

This will open a new tab for the group.

The group tab will have three sub-tabs: **Summary**, **Members**, and **Group Permissions**.

4. Select the **Members** tab.

The members tab displays a list of all users on the system. The users that belong to the selected group have a check in the box next to them.

5. Check the box(es) to the left for all the users you want to add to the group.
6. The changes are automatically saved. Close the tab.

To delete members from a group:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click on the **Groups** tab.
2. Locate the group in the list and check the box next to it.
3. Press **Edit** action button.

This will open a new tab for the group.

The group tab will have three sub-tabs: **Summary**, **Members**, and **Group Permissions**.

4. Select the **Members** tab.

The members tab displays a list of all users on the system. The users that belong to the selected group have a check in the box next to them.

5. Uncheck the box(es) to the left for all users you want to delete from the group.
6. The changes are automatically saved. Close the tab.

Setting permissions for a group

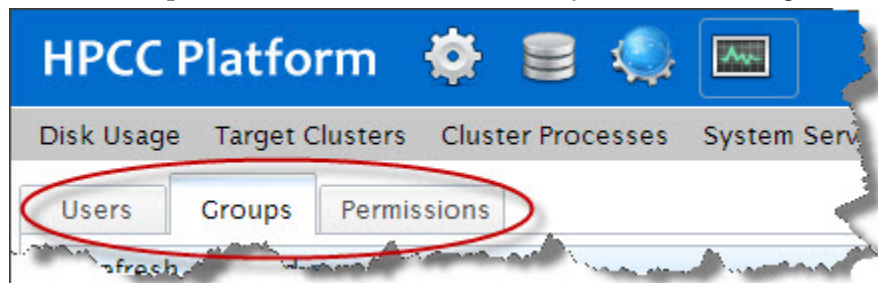
By default, all users are members of the **Authenticated Users** group. The **Authenticated users** group has access rights to almost all controls.

If you intend to restrict permissions for some users, you must remove **Authenticated Users** from the sections you wish to limit. You can then create groups with only those access rights you wish to grant. This approach allows the most flexibility since a single User ID can have multiple group memberships.

As a best practice, you should use **Allow** instead of **Deny** to control access. Denies should be used only as an exception.

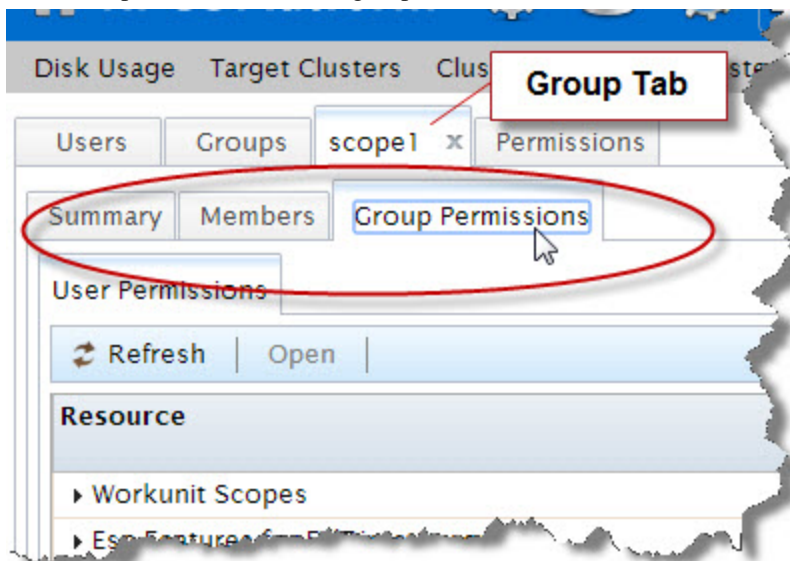
To set new permissions for a group:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.



1. Click the **Groups** tab.
2. Locate the group in the list and check the box next to it.
3. Press **Edit** action button.

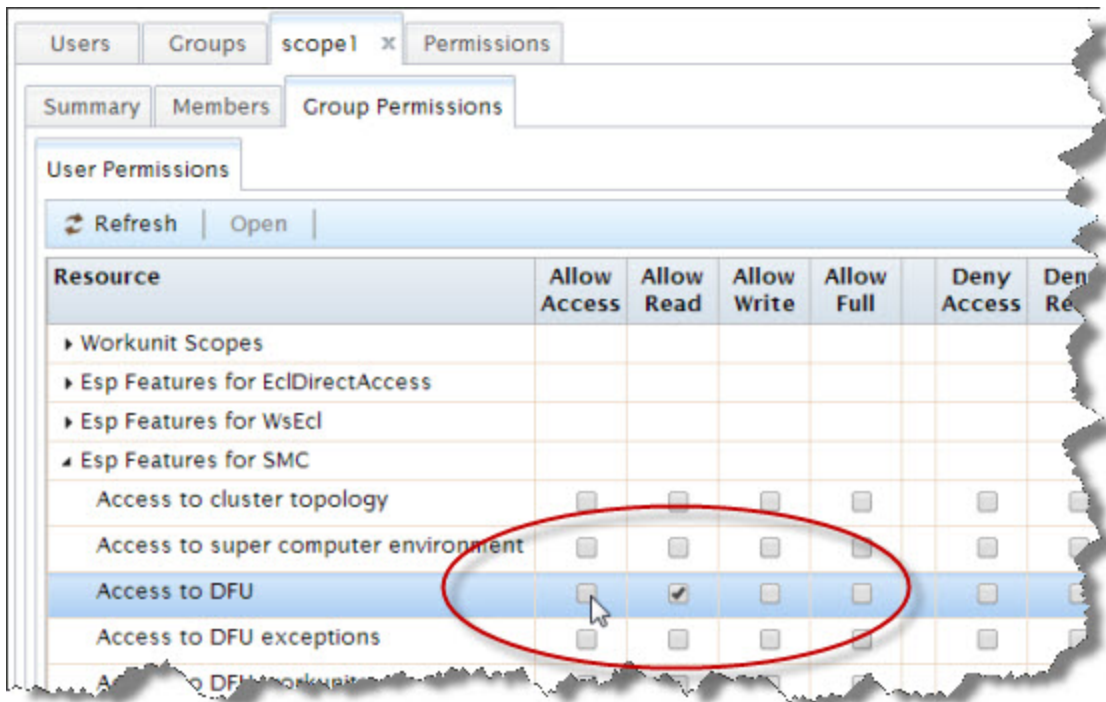
This will open a new tab for the group.



The group tab will have three sub-tabs: **Summary**, **Members**, and **Group Permissions**.

4. Select the **Group Permissions** tab.
5. Click on the arrow to the left of the resource to display the permissions for that resource. The permission groups currently set for this group and the inherited ones display.
6. There may be more than one resource setting available in each group, be sure to set the permissions for each setting as required.

7. Check the boxes for **allow** and **deny** as required for the group.



NOTE: Use caution when setting any explicit **deny** permission setting. In general, the most restrictive permission is applied.

8. There may be more than one resource setting available, select the resource(s) you require from the drop list.

Repeat for each applicable resource.

9. The changes are automatically saved. Close the tab.

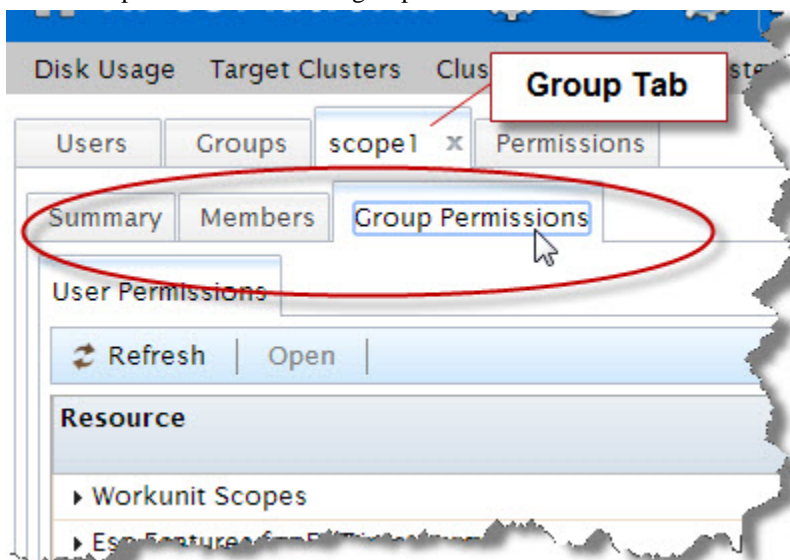
To modify permissions for a group:

Click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Click the **Groups** tab.
2. Locate the group in the list and check the box next to it.

3. Press **Edit** action button.

This will open a new tab for the group.



The group tab will have three sub-tabs: **Summary**, **Members**, and **Group Permissions**.

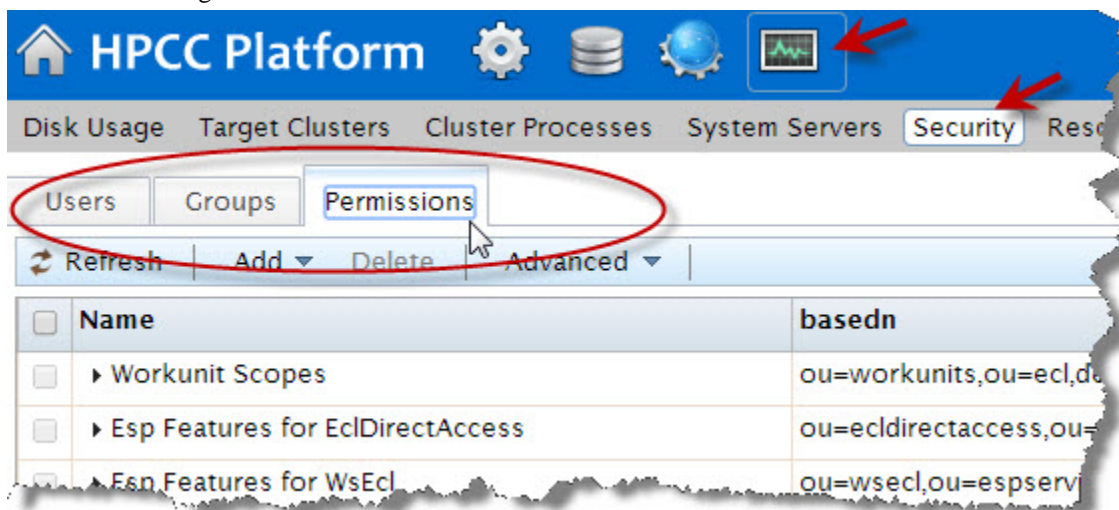
4. Select the **Group Permissions** tab.
5. Locate the feature(s) you want to modify.

Click on the arrow to the left of the resource to display the permissions for that resource.

6. Check the box(es) that **allow** and **deny** access as required for the feature.
7. The changes are automatically saved. Close the tab.

Setting and modifying feature permissions

Access to the feature permissions is available through ECL Watch. To modify feature permissions you must have Administrator level access. To access the feature permissions click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.



Use the **Permissions** menu item to:

- Edit the permissions for any feature
- Add new resources to a feature
- Set the permissions for users and groups for a specific resource
- Update the permissions for users and groups for a specific resource
- Delete a resource

Adding and editing feature permissions

Each feature contains a list of resources which are used to control access to an HPCC feature or folders containing files or workunits. The main HPCC feature permission settings are controlled using the **ESP Features for SMC** setting. When new features are added to the HPCC System, the release notes inform you that new permissions may be set. This is also true for **ESP Features for ECLDirectAccess** and **Esp Features for WsEcl**. Generally, all the permissions you require to control access to these features are already included.

However, to control access to file or workunit scopes, you must add the location as a resource before you can set permissions.

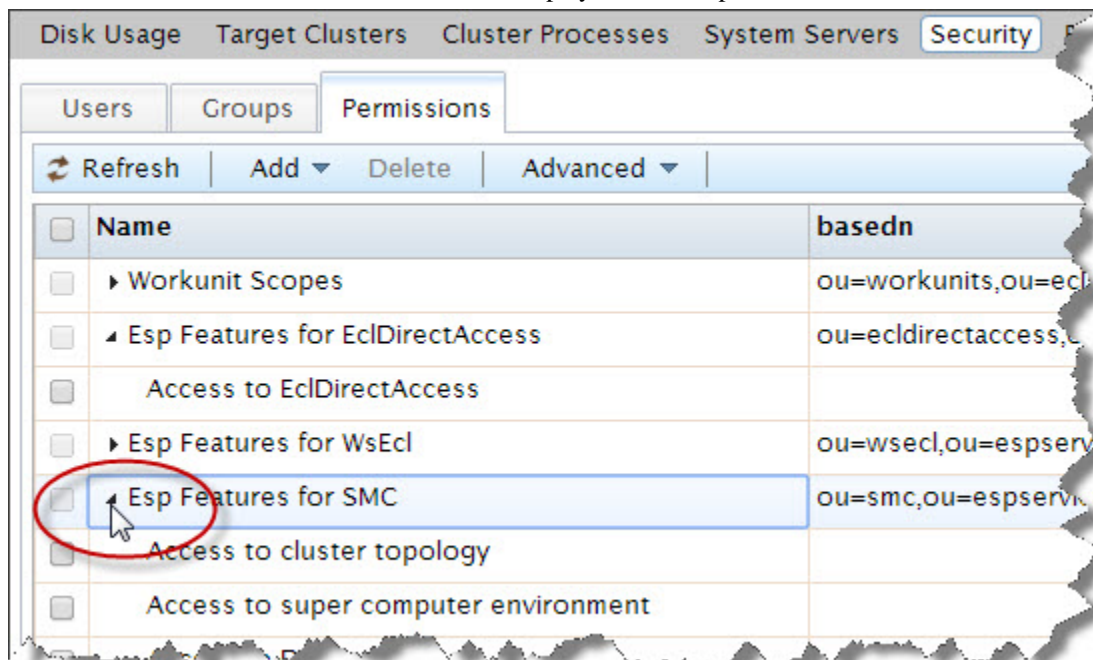
To add a scope:

To access the feature permissions click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

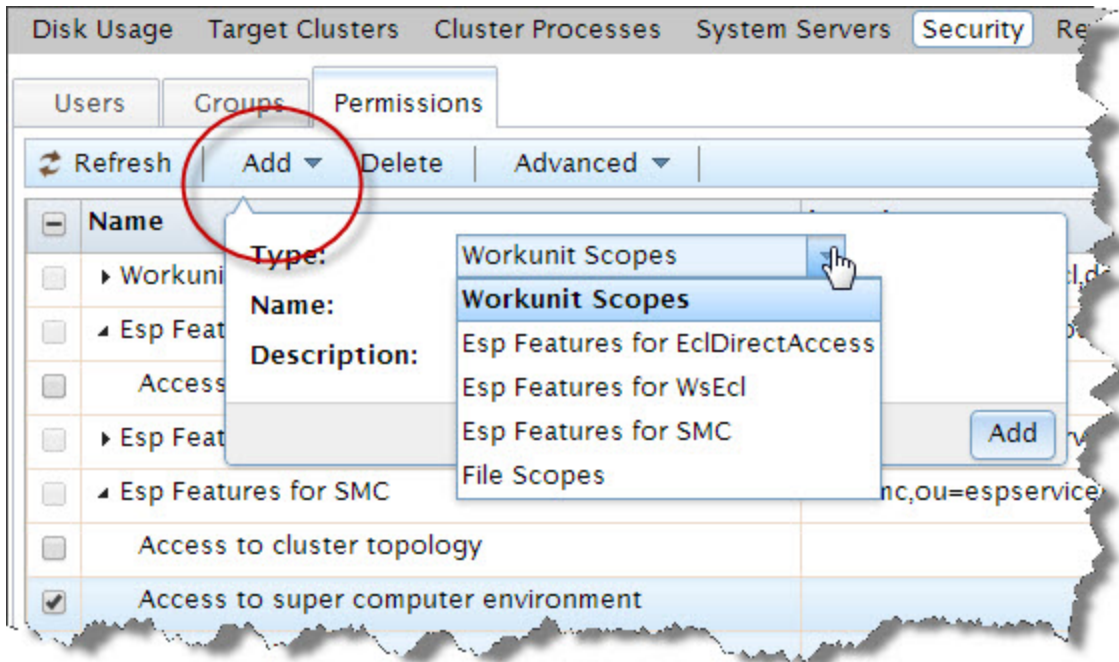
1. Click the **Permissions** tab.

The resources for that feature are listed.

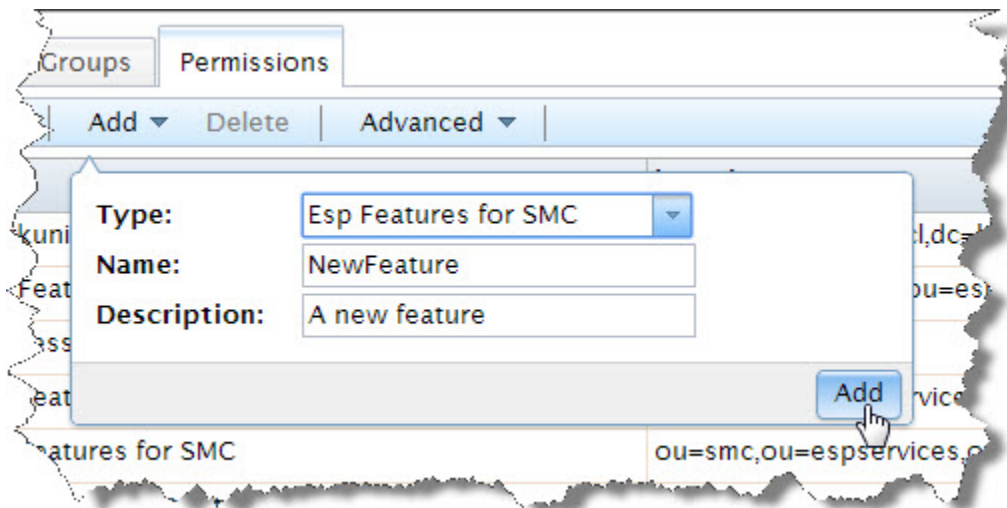
2. Click on the arrow to the left of the resource to display the feature permissions for that resource.



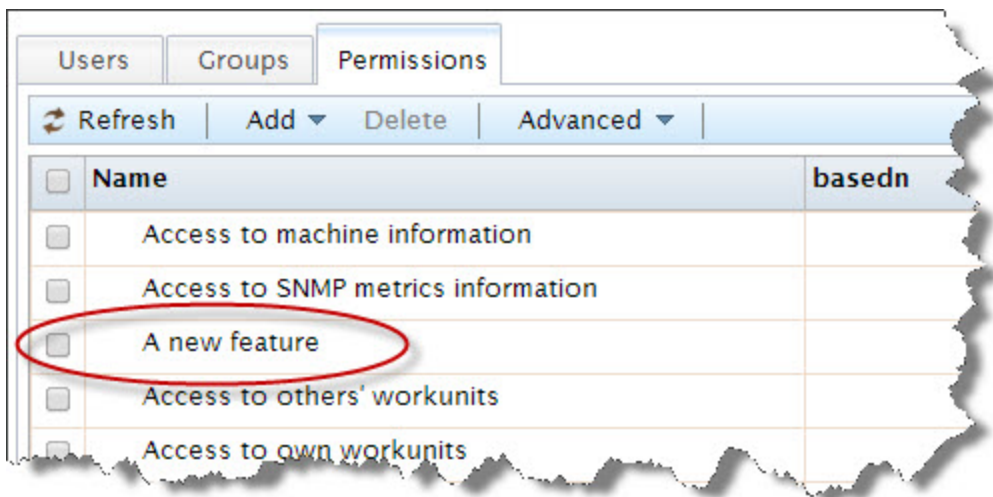
3. Press the **Add** button.
4. Choose the Type of permission from the drop list.



5. Enter the exact name of the scope you want to add (for example a file or workunit scope) in the **Name** field.
Enter a short description in the **Description** field.
6. Click **Add**.

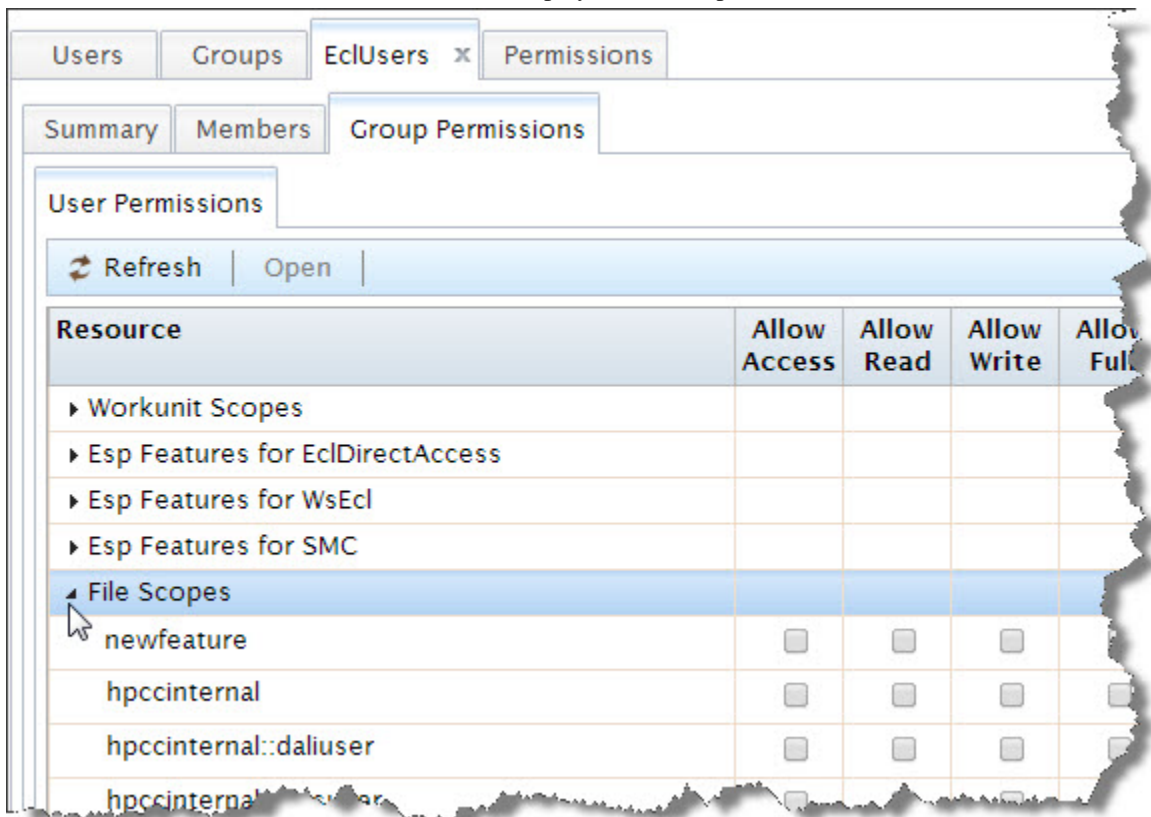


The new feature displays in the list.



7. Go to the **Groups** tab.
8. To locate your new scope, select a group by checking the box next to the group name. Press the **Edit** action button.
You can select multiple groups.

Select the Group Permissions tab in that group. (if multiple groups selected, you must repeat for each group)
9. Click on the arrow to the left of the resource to display the feature permissions for that resource.



The new feature is in the list. Check the boxes as appropriate to set the permissions for this group.

10. To set permissions for this scope for another group, go to that groups tab.

11.To set permissions for this scope for a user, go to the Users tab.

Select the user and press the Edit action button.

A new tab for that user opens. On that tab, click on the User Permissions sub-tab. You will then see your new permission listed under the appropriate Resource.

Set the permissions as appropriate for that user.

12.The changes are automatically saved. Close the tab(s).

To edit the permissions for a feature resource:

To access the feature permissions click on the **Operations** icon, then click the **Security** link from the navigation sub-menu.

1. Locate the user or group you want, select and press the **Edit** action button.
2. Click the **User Permissions** tab in ECL Watch.
3. Locate the resource you want to update.
4. Click the checkboxes in the **allow** and **deny** columns as appropriate.
5. The changes are automatically saved. Close the tab(s).

Note: You must follow this process for each user or group separately.

ECL Watch Feature Permissions

Access to features of the HPCC system is controlled by via the **ESP Features for SMC** category. These features are listed as **Resources** when setting permissions using ECL Watch.

ECL Watch feature permission settings that are not listed are not relevant and should not be used.

The following sections show the level of access required to be able to use ECL Watch features:

Login

SMCAccess is required by all users to be able to successfully login to ECL Watch.

LDAP Path	Description	Access
SmcAccess	Root Access to SMC Service	Read

Clusters

Users may be given access to the thor queue which can be manipulated by promoting/demoting queued workunits according to priority. The thor queue can also be paused or cleared and users can view thor usage statistics.

From this page, users can also click on workunit IDs to view details about the workunit. Depending on the level of access given, they can view, modify and delete their own, or others workunits.

LDAP Path	Description	Access
ThorQueueAccess	Access to Thor Job Queue Control	Full
RoxieControlAccess	Access to Roxie Process Cluster Control	Full
OwnWorkunitsAccess	Access to View Own Workunit	Read
	Access to Create or Modify Own Workunit	Write
	Access to Delete Own Workunits	Full
OtherWorkunitsAccess	Access to View Other User's Workunits	Read
	Access to Modify or Resubmit User's Workunits	Write
	Access to Delete Other User's Workunits	Full

ECL Workunits

Workunits can also be viewed using this feature of ECL Watch. The contents of the workunits list reflects whether a user has the permission to view their own and others workunits.

LDAP Path	Description	Access
OwnWorkunitsAccess	Access to View Own Workunit	Read
	Access to Create or Modify Own Workunit	Write
	Access to Delete Own Workunits	Full
OtherWorkunitsAccess	Access to View Other User's Workunits	Read
	Access to Modify or Resubmit User's Workunits	Write
	Access to Delete Other User's Workunits	Full

Topology

This section shows details about the clusters and other HPCC System components. Preflight provides diagnostic information including disc space, CPU usage and access to logs as well as the ability to swap faulty nodes out of the cluster.

LDAP Path	Description	Access
ClusterTopologyAccess	Access to Cluster Topology	Read
	Set Machine Status	Write
	Swap Node	Full
MachineInfoAccess	Access to machine/Preflight Information	Read
MetricsAccess	Access to SNMP Metrics Information (Roxie Metrics)	Read
ExecuteAccess	Access to Remote Execution in ECL Watch	Full

DFU Workunits

A user must have permission to view DFU Workunits and requires other permissions to be able to manipulate them.

LDAP Path	Description	Access
DfuWorkunitsAccess	Access to View DFU Workunits	Read
	Access to Create, Delete, Update, Submit, and Abort DFU Workunits	Write


DFU Files

Users need permission to see files on the dropzone and also to put files there. They need further permissions to be able to spray and copy files from the dropzone to their cluster and also to despray files from the cluster back to the dropzone.

XREF is used for monitoring files on the cluster(s). Reports generated show where housekeeping is required on the cluster(s) and users require additional permission to use this feature.

LDAP Path	Description	Access
DfuAccess	Access to DFU Logical Files	Read
	Delete Files, add to superfiles	Write
DfuExceptions	Access to DFU Exceptions	Read
DfuWorkunitsAccess	Access to View DFU Workunits	Read
	Access to Create, Delete, Update, Submit, and Abort DFU Workunits	Write
DfuXrefAccess	Access to DFU XREF	Read
	Clean directory	Write
	Make changes and generate XREF Reports	Full
FileDesprayAccess	Access to De-Spraying Files	Write
FileSprayAccess	Access to Spraying and Copying	Read
	Rename files	Write
	Delete from Drop zone	Full
FileIO	Access to read files in Drop zone	Read

LDAP Path	Description	Access
	Access to write to files in Drop zone	Write

	On a large system, we suggest limiting the number of users who can Generate XREF reports by setting DfuXrefAccess access to FULL for only those users.
---	--

Roxie Queries

Additional permission is required to view roxie queries in ECL Watch.

LDAP Path	Description	Access
RoxieQueryAccess	Access to Roxie Queries	Read

Users/Permissions

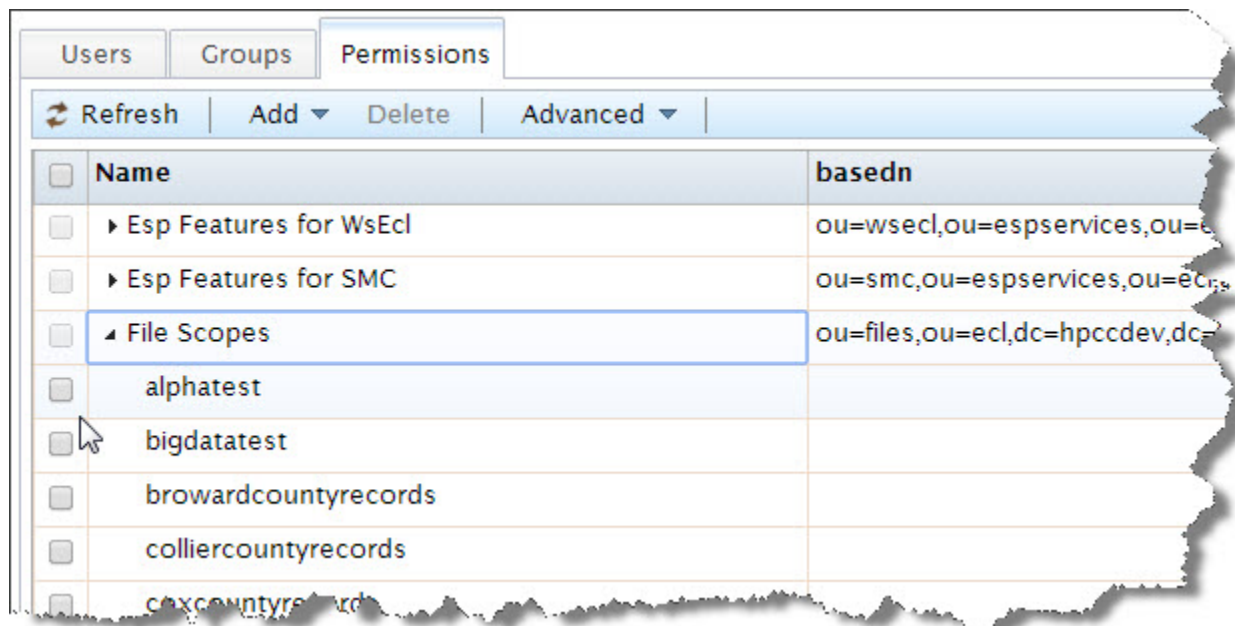
To be able to view the **Users/Permissions** area in ECL Watch, a user must be setup in LDAP as an Administrator.

File Access Control

The HPCC's LDAP **Dali Server** technology provides the ability to set secure access permissions to data file folders (or file scopes). This is controlled by organizational units (OUs) in LDAP.

An OU called **Files** is automatically created when the Dali server starts. To secure data folders, create OUs for each folder level and apply rights to each folder's OU.

Figure 20. File Scopes Page



For example, below **Files** there is an OU representing the cluster, such as **thor** (or the name that you set up for your cluster). Furthering the example, below that could be an OU named **specialdata** which contains two OUs, **public** and

secure. The **public** folder has rights granted to a large group of users and the **secure** folder has limited access granted. This allows you to prevent unauthorized users from any access to files in the **secure** folder.

The structure in LDAP corresponds to this logical structure in DFU:

specialdata::secure

Which corresponds to this physical structure:

var/lib/HPCCSystems/hpcc-data/thor/specialdata/secure

All HPCC components and tools respect LDAP file access security. The following exceptions are assumed to be system level or for administrative users:

- Network file access using UNC's, Terminal Services or SSH.
- Administrative utilities

Attempting to access a file in a folder for which access is not granted will result in one of the following errors:

```
DFS Exception: 4 Create access denied for scope <filepath>
```

or

```
DFS Exception: 3 Lookup access denied for scope <filepath>
```

(where <filepath> is the full logical file scope path)

File scope feature permissions

There are some extra features that are available for the **File Scopes** feature in the **Permissions** area of ECL Watch.

- Any file scope in the list may be reset to have the default permission settings for your system.
- Access to the physical file may also be granted separately.
- The **Check File Permissions** button allows you to quickly view the permissions that have been set for any scope listed.

Permissions settings for file scopes may be set as follows:

Description	Access
Read files in that scope	Read
Create/modify a file in that scope	Write
Delete a file in that scope	Full

Permission Caching

When you change a permission in ECL Watch, the settings are cached in the ESP server and stored in the Dali server. The information in the cache is updated at a configurable interval. This value can be set in the Configuration Manager under the LDAP Server settings Attributes tab. The default cacheTimeout is 5 minutes.

When you want a permission change to take effect immediately, you can clear the cache and force Dali to update the permission settings by pressing the **Clear Permissions Cache** button. This action transfers the settings when you press the button. Use this feature judiciously as overall system performance is affected temporarily while the LDAP settings in the Dali System Data Store repopulate.

Workunit Access Control

There are 2 aspects of workunit (WU) security:

- Feature Authentication for workunits allows you to set permissions to control whether users can view their own WUs and/or other users' WUs.
- Workunit Scope security provides the ability to set permissions for individual WU scopes. All new workunits now have a scope value.

Both methods are valid to use (either separately or together), and the strictest restriction always wins.

In other words, if someone is granted permission to see WUs in the scope *johndoe* but is denied permission to see other users' WUs in the Feature Authentication permissions, this user would still be denied access to see one of John Doe's WUs.

Conversely, if the user is allowed access to see other people's WUs but is denied access to the *johndoe* WU scope, this user will be able to see other WUs in that scope.

Note: If you do not have access to a WU, you will never be able to view it or even know of its existence.

By default, a submitted WU has a scope of the user's ID. For example, a WU JohnDoe submits has *scope=johndoe* in the WU. This value in a WU allows ESP and its services to use LDAP to check for permissions and enforce those permissions.

You can override the default scope using ECL Code:

```
#workunit('scope', 'MyScopeValue');
```

In addition, the scope of a workunit can be changed in ECL Watch by opening the WU, editing the scope field and pressing the **Save** button.

Securing workunit scopes

ESP (on startup) automatically creates an LDAP OU called **Workunits** (unless it already exists). If this OU is automatically created, the OU is made with full permissions granted to all authenticated users. All WU scopes are below the *workunits* OU either implicitly or explicitly.

If a specific scope OU does not exist in LDAP (e.g., the scope *johndoe* used in earlier example), then the parent OU's permissions are used. In other words, the scope of *johndoe* is implicitly under the *workunits* OU even though it might not be explicitly listed in the LDAP structure and therefore it would use the permissions granted for the parent, *workunits*.

Workunits feature permissions

Using the **Workunit Scopes** feature in the **Permissions** area of ECL Watch the permissions for any scope can be reset to the default permissions settings for your system. Permission settings for Workunit Scopes may be set as follows:

Description	Access
View WUs in that scope	Read
Create/modify a WU in that scope	Write
Delete a WU in that scope	Full

Workunits and Active Directory

The performance of your system can vary depending on how some components interact. One area which could impact performance is the relationship with users, groups, and Active Directory. If possible, having a separate Active Directory specific to HPCC could be a good policy. There have been a few instances where just one Active Directory servicing many, diverse applications has been less than optimal.

HPCC makes setting up your Active Directory OU's relatively easy. ESP creates all the OU's for you when it comes up, based on the settings you defined in Configuration Manager. You can then start Dali/ESP and use ECLWatch to add or modify users or groups.

You can assign permissions to each user individually, however it is more manageable to assign these permissions to groups, and then add users to these groups as appropriate. Create a group for developers and power users (people with full read/write/delete access), and another group for users that only have only read access and perhaps another group that has both read and write access. Add any other groups as appropriate for your environment. Now you can assign users to their appropriate group(s).

Active Directory, and LDAP Commonality

There are a few relevant notable terms, that may need some further explanation.

filesBasedn	Deals with restricting access to files. Also referred to as “file scoping”.
groupsBasedn	Controls the groups associated with the environment. For example, administrators, developers, ws_ecl only, etc.
modulesBasedn	Specific to systems using a legacy central repository and controls access to specific modules. Any module you create in the application will create an entry in Eclwatch>>User/Permissions>>Repository Modules
sudoersBasedn	Deprecated.
workunitsBasedn	Controls access to workunits.

Data Handling

When you start working with your HPCC system, you will want to have some data on the system to process. Data gets transferred to and the HPCC system by a process called a spray. Likewise to get data out from an HPCC system it must be desprayed.

As HPCC is a computer cluster the data gets deployed out over the nodes that make up the cluster. A *spray* or import is the relocation of a data file from one location (such as a Landing Zone) to a cluster. The term spray was adopted due to the nature of the file movement – the file is partitioned across all nodes within a cluster.

A *despray* or export is the relocation of a data file from a Data Refinery cluster to a single machine location (such as a Landing Zone). The term despray was adopted due to the nature of the file movement – the file is reassembled from its parts on all nodes in the cluster and placed in a single file on the destination.

A *Landing Zone* (or drop zone) is a physical storage location defined in your system's environment. There can be one or more of these locations defined. A daemon (dfilesrv) must be running on that server to enable file sprays and desprays. You can spray or despray some files to your landing zone through ECL Watch. To upload large files, you will need a tool that supports the secure copy protocol, something like a WinSCP.


For more information about HPCC data handling see the *HPCC Data Handling* and the *HPCC Data Tutorial* documents.

Best Practices

This chapter outlines various forms of best practices established by long time HPCC users and administrators running HPCC in a high availability, demanding production environment. While it is not required that you run your environment in this manner, as your specific requirements may vary. This section provides some best practice recommendations established after several years of running HPCC in a demanding, intense, production environment.

Cluster Redundancy

There are several aspects of cluster redundancy that should be considered when setting up your HPCC system.

	Make sure you allocate ample resources to your key components. Dali is RAM intensive. ECL Agent and ECL Server are processor dependent. Thor should have a minimum of 4GB RAM per node.
---	---

Dali

Dali should be run in an active/passive configuration. Active/passive meaning you would have two Dalis running, one primary, or active, and the other passive. In this scenario all actions are run on the active Dali, but duplicated on the passive one. If the active Dali fails, then you can fail over to the passive Dali.

Another suggested best practice is to use standard clustering with a quorum and a takeover VIP (a kind of load balancer). If the primary Dali fails, you move the VIP and data directory over to the passive node and restart the Dali service.

DFU Server

You can run multiple instances of the DFU Server. You can run all instances as active, as opposed to an active/passive configuration. There is no need for a load balancer or VIP. Each instance routinely queries the Dali for workunits. Should one fail, the other(s) will continue to pull new workunits.

ECLCC Server

You can run multiple active instances of the ECLCC Server for redundancy. There is no need for a load balancer or VIP for this either. Will routinely check for workunits. Should one fail, the other(s) will continue to compile.

ESP/ECL Watch/WsECL

To establish redundancy, place the ESP Servers in a VIP. For an active/active design, you must use a load balancer. For active/passive you can use pacemaker/heartbeat. If you run active/active, you should maintain a single client's connection to a single server for the life of a session for ECL Watch (port 8010). Other services, such as WsECL (port 8002) do not require a persistent connection to a single server.

ECL Agent

You can run multiple active instances of the ECL Agent. No need for a load balancer or VIP. Each instance routinely queries for workunits. Should one fail, the other(s) will continue to pull new workunits.

Sasha

Sasha should be run in an active/passive configuration. Active/passive meaning you would have two Sashas configured, one primary (active), and the other standing by.

ECL Scheduler

No need for a load balancer, runs active/active. Each instance routinely queries for workunits. Should one fail, the other(s) will continue to schedule workunits.

Thormaster

Set up Thor in an active/passive configuration. Active/passive meaning you would have two instances running, one primary (active), and the other passive. No load balancer needed. If the active instance fails, then you can fail over to the passive. Failover then uses the VIP (a kind of load balancer) to distribute any incoming requests.

Dropzone

This is just a fileserver that runs the dafilesrv process. Configure in the same fashion as you would any active/passive file server. One primary, or active, and the other passive. No load balancer needed. If the active instance fails, then you can fail over to the passive.

Make sure you give significant resources to your key components. Dali is RAM intensive. Eclagent and Eclserver are processor dependent. Thor should have a minimum of 4GB RAM per node.

High Availability and Disaster Recovery

If you require high availability for your HPCC system, there are some additional considerations that you should be aware of. This is not comprehensive list, and it is not meant to be step-by-step instructions for setting up disaster recovery. Instead this section just provides some more information to consider when incorporating HPCC into your disaster recovery plan.

Thor

When designing a Thor cluster for high availability, consider how it actually works -- a Thor cluster accepts jobs from a job queue. If there are two Thor clusters handling the queue, one will continue accepting jobs if the other one fails.

If a single component (thorslave or thormaster) fails, the other will continue to process requests. With replication enabled, it will be able to read data from the back up location of the broken Thor. Other components (such as ECL Server, or ESP) can also have multiple instances. The remaining components, such as Dali, or DFU Server, work in a traditional shared storage high availability fail over model.

The Downside

Costs twice as much initially because you essentially have to have two of everything.

The Upside

Almost 100% of the time you can utilize the additional processing capacity. You can run more jobs, have more space, etc.

Disaster Recovery concerns

The important factor to consider for disaster recovery (DR) is the bandwidth required to replicate your data. Your network administrator should evaluate this aspect carefully.

If you have tens of gigabytes of delta each day then an rsync type replication or some sort of hybrid model should suffice. If you have hundreds of gigabytes to petabytes of deltas, the real limit is your budget.

A best practice is to find where the data is the smallest (at ingestion, after normalization, at Roxie) and replicate from that point and rerun the processing in both locations.

The key to getting disaster recovery right is to know your data flow. For instance, if you are ingesting 20TB of raw data daily, then taking that raw data and rolling it up, scoring it, indexing it, etc. You would be better off replicating an intermediate dataset (that we call base files), rather than replicating the large ingest. If the opposite is occurring (small daily ingest and then blow the data up in size) – you would be better off to ingest the input and then re-run it.

Thor has the ability to do a “Thor copy” which copies data from one cluster to another. You can also do this through ECL code. Additionally, you may decide you don’t want, or need to have a “hot” DR Thor. In that case, the most common disasters [minor] (major switch outage, total power down, multiple fiber cuts) cause only a relatively brief, less than 1 day disaster. Since Thor is responsible for creating data updates it can take a day or a few to recover. The data just is not quite as fresh but as long as the Roxies are replicated the data is still flowing. In the case of a major disaster (a major earthquake, or a tidal wave), the likelihood of that occurring does not justify the cost of preventing against it. It could also take between 7 to 14 days to recover by building out a whole new Thor cluster.

Conclusion

Disaster recovery is a calculation. The cost of failure, times the likelihood per year of an event occurring, less than or greater than the cost to prevent against it. Taking all that into consideration can help you to put a sensible DR plan in place.

Roxie

In the case of Roxie, a best practice is to have multiple Roxie clusters and use a proxy to balance. In case of how to keep the data in sync, a pull approach is best. The Roxie automatically pulls the data it needs from the “source” listed in the package file. The data can also be pulled from another Roxie or a Thor. In most cases you would pull to your DR Roxie from the primary Roxie out of the load balancer, but it can also pull from a Thor in the primary location as well.

Middleware

Replication of some components (ECL Agent, ESP/Eclwatch, DFU Server, etc.) are pretty straight forward as they really don't have anything to replicate. Dali is the biggest consideration when it comes to replication. In the case of Dali, you have Sasha as the back up locally. The Dali files can be replicated using rsync. A better approach could be to use a synchronizing device (cluster WAN sync, SAN block replication, etc.), and just put the Dali stores on that and just allow it replicate as designed.

There isn't just a one size fits all approach. Special care, design, and planning are required to make an effective DR strategy that doesn't “over synchronize” across slow WAN links, but still provides you with an acceptable level of redundancy for your business needs.

Best Practice Considerations

There are several other aspects to best practice considerations, and these will change with your system requirements. The following sections are some best practice considerations for some aspects of the HPCC system. Keep in mind that suggested best practices are merely suggested and may not be appropriate for your needs. A thorough review of the considerations highlighted here can be very helpful if your needs align with the stated considerations.

Multiple Thors

You can run multiple Thors on the same physical hardware. Multiple Thors on the same hardware are independent and unaware of each other. The Thors run jobs as they receive them, regardless of what the other(s) is/are doing. The speed of a single job will never be faster with multiple Thors, but the throughput can be. You can run two Thors picking up jobs from two different queues or the same queue.

The downside of running multiple Thors on the same hardware is that the physical memory on the nodes needs to be shared among each of the Thors. This needs to be configured per Thor cluster definition.

You must not place multiple Thors on hardware which does not have enough CPU cores to support it. You should not have more Thors than number of cores. One good rule is to use a formula where the number of cores divided by two is the maximum number of Thor clusters to use.

System Sizings

This section provides some guidance in determining the sizing requirements for an initial installation. The following are some suggested sample configuration guides that can be helpful when planning your system.

Sample Sizing for High Data volume (Typical)

The most typical scenario for HPCC is utilizing it with a high volume of data. This suggested sample sizing would be appropriate for a site with large volumes of data. A good policy is to set the Thor size to 4 times the source data on your HPCC. Typically, Roxie would be about ¼ the size of Thor. This is because the data is compressed and the system does not hold any transient data in Roxie.

High Data Thor sizing considerations

Each Thor node can hold about 2.5 TB of data (MAX), so plan for the number of Thor nodes accordingly for your data.

If possible, SAS drives for both Thor and Roxie as they almost equal to SATA drives now. If not for both, get SAS drives at least for your Roxie cluster.

Thor replicates data, typically configured for 2 copies.

High Data Roxie sizing considerations

Roxie keeps most of its data in memory, so you should allocate plenty of memory for Roxie. Calculate the approximate size of your data, and allocate appropriately. You should either increase the number of nodes, or increase the amount of memory.

A good practice is to allocate a Dali for every Roxie cluster.

Roxie+Dali needs to have a mirror. This is because, when you need to update indexes, you update the mirror and make that primary and bring the other one down. This is not really a necessity except for high availability and performance requirements.

Sample Sizing for Heavy Processing on Low Data Volume

The following section provides some sample sizing for heavy processing with approximately the amount of data indicated.

750 GB of Raw Data

Thor = 3 (slaves) + 2 (management) = 5 Nodes

Roxie = 3 (agents) + 1 (Dali) = 4 Nodes (This will mean that the environment will be down during query deployment)

Spares = 2

Total = 13 nodes

1250 GB of Raw Data

Thor = 6 (slaves) + 2 (management) = 8 Nodes

Roxie = 4 (agents) + 1 (Dali) = 5 Nodes (This will mean that the environment will be down during query deployment)

Spares = 2

Total = 17 nodes

2000 GB of Raw Data

Thor = 8 (slaves) + 3 (management) = 11 Nodes

Roxie = 4 (agents) + 1 (Dali) = 5 Nodes (This will mean that the environment will be down during query deployment)

Spares = 2

Total = 20 nodes

3500 GB of Raw Data

Thor = 12 (slaves) + 5 (management) = 17 Nodes

Roxie = 6 (agents) + 1 (Dali) = 7 Nodes (This will mean that the environment will be down during query deployment)

Spares = 2

Total = 28 nodes

System Resources

There are additional resources available for the HPCC System.

HPCC Resources

The resources link can be found under the Operations Icon link. The resources link in ECL Watch provides a link to the HPCC Systems web portal. Visit the HPCC Systems Web Portal at <http://hpccsystems.com/> for software updates, plug-ins, support, documentation, and more. This is where you can find resources useful for running and maintaining HPCC on the web portal.

ECL Watch provides a link to the HPCC portal's download page: <http://hpccsystems.com/download>. This is the page where you can download Installation packages, virtual images, source code, documentation, and tutorials.