*HPC Data MANAGEMENT*

adminISTRATION guide

Version *1.4.1*

*11/30/2017*

**Version History**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version Number** | **Implemented**  **By** | **Revision**  **Date** | **Approved**  **By** | **Approval**  **Date** | **Description of Change** |
| 1.0 | Prasad Konka | 3/9/2016 |  |  | Initial Draft |
| 1.0 | Zhengwu Lu | 3/16/16 |  |  | Review |
| 1.0 | Prasad Konka | 3/17/16 |  |  | Updates |
| 1.0 | Prasad Konka | 3/21/16 |  |  | Updates from George Zaki comments |
| 1.0 | Zhengwu Lu | 11/27/16 |  |  | Minor update |
| 1.0 | Prasad Konka | 2/23/2017 |  |  | API Security |
| 1.1 | Prasad Konka | 5/22/2016 |  |  | Service account, database setup |
| 1.2 | Zhengwu Lu/Prasad Konka | 6/14/2017 |  |  | Update for 1.2 release |
| 1.3 | Prasad Konka | 7/13/17 |  |  | Updates addressing Zhengwu comments |
| 1.3 | Prasad Konka | 9/19/2017 |  |  | Updated application/system properties |
| 1.3 | Prasad Konka | 9/26/2017 |  |  | Update, delete, general info |
| 1.4 | Prasad Konka | 10/30/2017 |  |  | Addressed reviewers comments |
| 1.4.1 | William Liu | 11/30/2017 |  |  | Added notes about how SSH connecting to production servers may require first SSH connecting to DEV/UAT server. |

TABLE OF CONTENTS

[1. Purpose 5](#_Toc497123963)

[2. Introduction 5](#_Toc497123964)

[2.1. What is HPC DME? 5](#_Toc497123965)

[2.2. Intended Users 5](#_Toc497123966)

[3. Setting up HPC DME 6](#_Toc497123967)

[3.1. Pre-requisiteS for API server 6](#_Toc497123968)

[3.1.1 Java 6](#_Toc497123969)

[3.1.2 Install Maven 6](#_Toc497123970)

[3.1.3 Install Git Client 6](#_Toc497123971)

[3.1.4 Install ServiceMix 6](#_Toc497123972)

[3.2 Set up irods Database server 6](#_Toc497123973)

[3.2.1 Install PostgreSQL 6](#_Toc497123974)

[3.3 Set up irods server 7](#_Toc497123975)

[3.3.1 Install iRODS 7](#_Toc497123976)

[3.3.2 Install iRODS iCommands CLI 7](#_Toc497123977)

[3.2. Build and Deploy API Server 8](#_Toc497123978)

[3.2.1 Set HPC DME Services properties 9](#_Toc497123979)

[3.2.2 Set up CA Certificate for SSL port 16](#_Toc497123980)

[3.2.3 Build HPC DME 17](#_Toc497123981)

[3.2.4 Build HPC DME keystore 17](#_Toc497123982)

[3.2.5 Deploy HPC DME into ServiceMix 17](#_Toc497123983)

[3.2.6 Setup HPC DME Database 18](#_Toc497123984)

[3.2.7 Import SSL Certs into Java Keystore on the API Server 18](#_Toc497123985)

[4. Accessing HPC DME 18](#_Toc497123986)

[5. HPC DME Services ADMIN Operations 19](#_Toc497123987)

[5.1 Set up System Admin account 19](#_Toc497123988)

[5.2 Set up Service account 20](#_Toc497123989)

[5.2.1 Setup or Update Cleversafe service account credentials 20](#_Toc497123990)

[5.2.2 Setup or Update Globus service account credentials 21](#_Toc497123991)

[5.2.3 Setup or Update iRODS Service account credentials 22](#_Toc497123992)

[5.2.4 NIH AD Service Account 22](#_Toc497123993)

[5.2.5 Password Policy 23](#_Toc497123994)

[5.3 Set Up A Base PATH 23](#_Toc497123995)

[5.4 Managing Security 24](#_Toc497123996)

[5.4.1 Roles 25](#_Toc497123997)

[5.4.2 Service API Security 26](#_Toc497123998)

[5.5 Create a User 27](#_Toc497123999)

[5.6 Updating an Existing User 28](#_Toc497124000)

[5.7 Assigning permissions to HPC DME Group administrators 29](#_Toc497124001)

[5.8 Create System Administrator group and assign System Admins 29](#_Toc497124002)

[5.9 Accessing HPC DME Database 30](#_Toc497124003)

[5.10 Register a DOC 31](#_Toc497124004)

[5.11 Updating Existing DOC Configuration 32](#_Toc497124005)

[5.12 Connecting to iRODS through iCommand 33](#_Toc497124006)

[5.13 Delete or Edit a Collection on iRODS 33](#_Toc497124007)

[5.14 Delete or Edit a Data File on iRODS and Cleversafe 34](#_Toc497124008)

[5.15 Globus Application Account Configuration 34](#_Toc497124009)

[5.16 Accessing Application Log files 36](#_Toc497124010)

[5.17 Clean up and reset initial test data load for Development, UAT Tiers 36](#_Toc497124011)

[5.18 General Configuration 36](#_Toc497124012)

[5.19 Technology Stack 37](#_Toc497124013)

# Purpose

The HPC DME Admin guide describes pre-requisites and installation instructions to setup HPC Data Management Services. The Admin guide further details different functions admin users would be performing to support HPC DME user needs.

# Introduction

## What is HPC DME?

The HPC DME, High Performance Computing Data Management Environment (Services), is a highly adaptable and an open ended data storage environment supporting storage and management of data, produced from high performance computing or other research systems. HPC DME provides capabilities for storing, managing, transferring and sharing data across different systems securely and efficiently.

Users can store data objects for a long term on HPC DME, share and transfer their data such that they do not have to redistribute or maintain copies of the data on other systems by eliminating the data integrity issues. HPC DME stores and associates user defined metadata to any registered data at different levels of data life cycle, enabling the system not only to help identify the data but also to enhance the search capabilities and to be able to attach a value factor to each dataset.

HPC DME has two main components, HPC DME API – RESTful API and HPC DME Web – Web UI interacting with the API.

## Intended Users

The HPC DME has been built to cater to the data storage and data management needs of NCI cancer community. Any user with a valid HPC DME user account can run its client interface. This guide intends to guide HPC system administrators in managing user accounts, security and data integrity.

Note: HPC DME supports only NIH active directory account holders at this point of time.

# Setting up HPC DME

## Pre-requisiteS for API server

### Java

Download and install Java 1.8.x

<https://www.java.com/en/download/>

Please visit the following link for installation instructions

<https://www.java.com/en/download/help/index_installing.xml>

After successful installation of Java, update PATH with Java executable folder (Ex: c:/jdk1.8\_13/bin)

<https://java.com/en/download/help/path.xml>

Open cmd/shell window and type “java -version” to make sure it displays the version information you installed.

**Installation on CentOS**

Download the Java 8 RPM from:

<http://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>

Open a shell, and install the JDK with the following command (you will either need to be root, or using an account with sudo access):

(sudo) yum install <rpm name>

Open a command shell and type java -version to confirm that it displays the correct version. If not, in your shell type:

(sudo) alternatives –config java

And select the correct version of java from the list.

You can additionally enter:

(sudo) alternatives –config javac

This will set the default version of the java compiler.

Set JAVA\_HOME to the directory where you just installed Java. This is usually accomplished by editing /etc/profile (which will set it for all users) or individual .profile files for each account. Adding the following lines will usually suffice:

export JAVA\_HOME=/path/to/java

export PATH=$JAVA\_HOME/bin:$PATH

### Install Maven

Install Maven 3.3.3: [https://maven.apache.org/download.cgi#](https://maven.apache.org/download.cgi)

After successful installation of Maven, update PATH with Maven executable folder (Ex: c:/apache-maven-3.3.3/bin)

Open cmd/shell window and type “mvn -version” to make sure it displays the version information you installed.

**Installation on CentOS**

Server installation of maven is straight forward. Download the appropriate version for your OS at the link above (for the configuration intstructions below, we are assuming you downloaded the file ‘apache-maven-3.3.3-bin.tar.gz’; modify the instructions accordingly if you downloaded a different version).

In a terminal, navigate to the directory where you downloaded this file and enter the following commands:

(sudo) tar xvf apache-maven-3.3.3-bin.tar.gz

(sudo) cp apache-maven-3.3.3 /usr/share {this can be any directory you choose}

(sudo) ln -s apache-maven-3.3.3 maven {optional, but you’ll need to use the full path below if you don’t do this}

(sudo) vi /etc/profile.d/maven.sh

Enter the following lines into this file:

export M2\_HOME=/usr/share/maven {update this to whatever path you chose above}

export PATH=${M2\_HOME}/bin:${PATH}

Now open another terminal and type ‘mvn -version’ to confirm that you configured it correctly.

Generic instructions for installing Maven in any environment are available here: <https://maven.apache.org/install.html>. All that is really required is that you add the maven bin directory to your PATH environment variable. You may need to change ownership or make the mvn binary executable for anyone.

### Install Git Client

For Windows: <https://tortoisegit.org/>

Ubuntu: sudo apt-get -y install git

CentOS: (sudo) yum install git

### Install ServiceMix

Downlad Servicemix 7.0.0: <http://servicemix.apache.org/downloads.html>

Extract ServiceMix to any folder.

On CentOS, this can be accomplished by the following commands:

(sudo) unzip <download file.zip>

(sudo) cp <unzipped directory name> /usr/share {or whatever path you choose}

Additional instructions for installing servicemix can be found here:

<https://servicemix.apache.org/docs/7.x/users-guide/installation.html>

Note that you may see an error message starting Servicemix if you haven’t set JAVA\_HOME as described above. You can either do so, or edit the file ‘setenv’ in the bin directory where you installed Servicemix to set it there.

## Set up irods Database server

### Install PostgreSQL

Install PostgreSQL 9.5.x for iRODS iCAT database

<https://help.ubuntu.com/community/PostgreSQL>

Detailed nstructions for installing PostgreSQL on CentOS can be found here:

<https://linode.com/docs/databases/postgresql/how-to-install-postgresql-relational-databases-on-centos-7/>.

Note that if the command ‘(sudo) yum search postgres’ shows that the packages ‘postgresql95-server’ and ‘postgres95-contrib’ are available, you can simply enter the following command at a terminal to install the database:

(sudo) yum install postgresq l95-server postgresql95-contrib

Once the database is installed, you’ll need to initialize it and set it to start when the server boots:

(sudo) /usr/pgsql-95/bin/postgresql95-setup initdb

(sudo) systemctl start postgresql-9.5

(sudo) systemctl enable postgresql-9.5

## Set up irods server

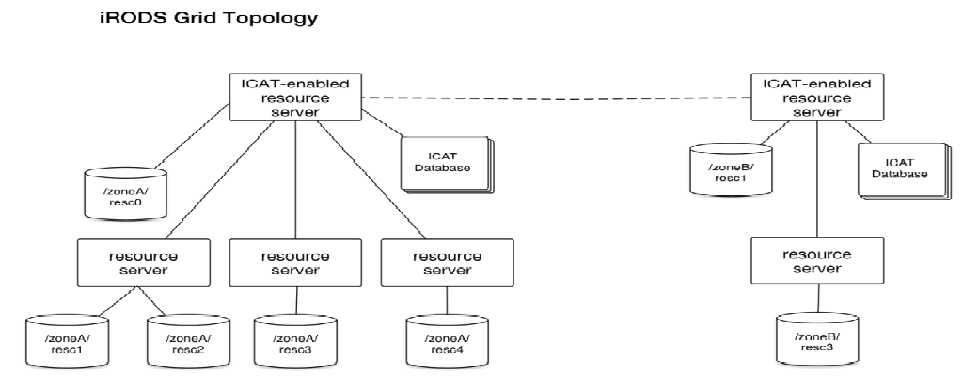
### Install iRODS

Download iRODS 4.2.x from the following link:

<http://irods.org/download/>

There are two types of iRODS server installation, iCAT and Resource Server:

1. An iCAT server manages a Zone, handles the database connection to the iCAT metadata catalog (which could be either local or remote), and can provide Storage Resources. An iRODS Zone will have exactly one iCAT server.
2. A Resource server connects to an existing Zone and can provide additional storage resource(s). An iRODS Zone can have zero or more resource servers.



An iCAT server is just a resource server that also provides the central point of coordination for the Zone and manages the metadata. A single computer cannot have both an iCAT server and a Resource server installed. The simplest iRODS installation consists of one iCAT server and zero Resource servers.

Please visit following link for detailed instructions on how to install iRODS iCAT and Resource servers. Current implementation of HPC DME supports setting up one iRODS zone with no resource servers: <https://docs.irods.org/master/manual/installation/>

### Install iRODS iCommands CLI

If you are planning to interact with iRODS remotely, download iCommands CLI package from <http://irods.org/download/>

The iCommands client is a command line interface that can be installed on Unix-based operating systems (Linux, Solaris, Macintosh, AIX) as well as Windows platforms.

This package installs just the client iCommands. Install this package if you want to have a system-wide installation of the iCommands (in /usr/bin), but do not want the entire iRODS iCAT or resource server. The iCommands are already included in both the iCAT and Resource packages – so installing this package in addition is unnecessary (and will conflict).

Once the iCommands client has been installed, you must then configure the iRODS environment. This can be done by editing $HOME/.irods/irods\_environment.json. You can set the contents of this file using the following configuration.

{

"irods\_host": "<iRODS host name>",

"irods\_port": <iRODS port>,

"irods\_user\_name": "rods",

"irods\_zone\_name": "<Zone name>"

}

The default installation of iRODS comes with a system user account 'rods' with a default password 'irods'. You should change the password before letting anyone else into the system:

iinit - Enter rods user password

irods@hostname:~/ $ iadmin moduser rods password <newpassword>

To make sure everything is successful, you will need to re-authenticate and check the new connection:

irods@hostname:~/ $ iinit

Enter your current iRODS password:

irods@hostname:~/ $ ils

/tempZone/home/rods:

“rods” user is defaulted with “rodsadmin” privileges. You should not use “rods” user account for your system admin functions. It is recommended to create a separate admin user account associated with NIH ID for your system admin functions.

## Build and Deploy API Server

* Set your $PATH environment variable

If you are using Windows, set the following environment variables

MAVEN\_HOME=<Path to Maven folder>

SERVICE\_MIX\_HOME=<Path to ServiceMix folder>

PATH=%MAVEN\_HOME%/bin;%SERVICE\_MIX\_HOME%/bin;%PATH%

If you are using Unix, add the following to your shell configuration file.

export MAVEN\_HOME=<Path to Maven folder>

export SERVICE\_MIX\_HOME=<Path to ServiceMix folder>

export PATH=$MAVEN\_HOME/bin:$SERVICE\_MIX\_HOME/bin:$PATH

* First, Clone the code from HPC DME GIT repository:

Repository URL: <https://github.com/CBIIT/HPC_DME_APIs>

* Set HPC\_HOME environment variable to the ‘src’ directory path in the source tree.

### Set HPC DME Services properties

HPC DME build scripts supports preparing the build for “dev”, “preprod” and “prod” environments.

Edit < HPC\_HOME >\ hpc-server\hpc-features\src\main\resources\<properties file> to set system configuration properties.

To set a system property, you would need to uncomment it and set a valid value to it. If a property is commented, it uses its default value.

The following system properties are related to REST services

|  |  |  |
| --- | --- | --- |
| Property Name | Description | Default Value |
| **The following system properties are related to REST services** | | |
| hpc.ws.rs.host | This is host name of the machine where HPC DME Rest services will be running | Localhost |
| hpc.ws.rs. port | This is the secured port of the machine where HPC DME Rest services will be running. This port should be secured with CA cert as detailed below in section 3.2.2 | 7738 |
| hpc.ws.rs.path | HPC DME Server URL context name | hpc-server |
| hpc.ws.rs.stackTraceEnabled | Enable stacktrace as part of the error log | True |
| hpc.ws.rs.api-version | Do not change this value. This is updated by developers for any new release | 1.3.0 |
| hpc.ws.rs.ldapAuthentication | Enable NCI LDAP Authentication. If LDAP authentication is set to false, HPC DME uses authentication based on iRODS accounts | True |
| hpc.ws.rs.attachment-tmp-directory | Temporary staging directory where servicemix writes files while receiving from a client request |  |
| hpc.ws.rs.attachment-memory-threshold | Memory threshold before servicemix starts writing into “hpc.ws.rs.attachment-tmp-directory” folder | 104857600 |
| hpc.ws.rs.ssl.keystore.path | Name of the keystore where HPC DME certificate is stored | ${karaf.base}/etc/hpc-server/keystore.jks |
| hpc.ws.rs.ssl.keystore.type | Type of the keystore where HPC DME certificate is stored | JKS |
| hpc.ws.rs.ssl.keystore.password | Password of the keystore where HPC DME certificate is stored | hpc-server-store-pwd |
| hpc.ws.rs.ssl.truststore.path | Name of the trust store where HPC DME certificate is stored | ${karaf.base}/etc/hpc-server/keystore.jks |
| hpc.ws.rs.ssl.truststore.type | Type of the trust store where HPC DME certificate is stored | JKS |
| hpc.ws.rs.ssl.truststore.password | Password of the trust store where HPC DME certificate is stored | hpc-server-store-pwd |
| hpc.ws.rs.ssl.wantClientAuth | Set this property to the value true to configure the server to request a client certificate. This property is ignored when configured for a client component such as SSLSocketAppender. | False |
| hpc.ws.rs.ssl.needClientAuth | Set this property to the value true to configure a server to require a valid client certificate. This property is ignored when configured for a client component such as SSLSocketAppender. | False |
| hpc.ws.rs.jetty-engine.minThreads | Specifies the min size of the thread pool used by the specific Jetty instance | 5 |
| hpc.ws.rs.jetty-engine.maxThreads | Specifies the max size of the thread pool used by the specific Jetty instance | 15 |
| **The following system properties are related to REST services access permissions** | | |
| hpc.ws.rs.auth.register-user | Mapping user registration function to Allowed roles to register user | SYSTEM\_ADMIN GROUP\_ADMIN |
| hpc.ws.rs.auth.security.update-user | Mapping user registration function to Allowed roles to update user | SYSTEM\_ADMIN |
| hpc.ws.rs.auth.get-user | Allowed roles to query user | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.security.get-invoker | Allowed roles to query invoker | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.security.get-active-users | Allowed roles to query active users | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.security.get-all-users | Allowed roles to query all users | SYSTEM\_ADMIN |
| hpc.ws.rs.auth.authenticate | Allowed roles to authenticate user | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.security.register-group | Allowed roles to register a group | SYSTEM\_ADMIN GROUP\_ADMIN |
| hpc.ws.rs.auth.security.update-group | Allowed roles to update group | SYSTEM\_ADMIN GROUP\_ADMIN |
| hpc.ws.rs.auth.security.get-group | Allowed roles to get group | SYSTEM\_ADMIN GROUP\_ADMIN |
| hpc.ws.rs.auth.security.get-groups | Allowed roles to get groups | SYSTEM\_ADMIN GROUP\_ADMIN |
| hpc.ws.rs.auth.security.delete-group | Allowed roles to delete group | SYSTEM\_ADMIN GROUP\_ADMIN |
| hpc.ws.rs.auth.security.register-system-account | Allowed roles to register system account | SYSTEM\_ADMIN |
| hpc.ws.rs.auth.register-collection | Allowed roles to register collections | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.get-collection | Allowed roles to query a collection | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.download-collection | Allowed roles to download collection | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.set-collection-permissions | Allowed roles to set collection permissions | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.get-collection-permissions | Allowed roles to get collection permissions | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.register-data-object | Allowed roles to register a data object | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.get-data-object | Allowed roles to query data object | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.download-data-object | Allowed roles to download data object | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.set-data-object-permissions | Allowed roles to set data object permissions | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.get-data-object-permissions | Allowed roles to get data object permissions | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.get-data-management-model | Allowed roles to get data management model | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-management.get-data-management-tree | Allowed roles to get data management tree | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.query-collections | Allowed roles to query collections | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.query-data-objects | Allowed roles to query data objects | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.add-query | Allowed roles to save a named query | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.update-query | Allowed roles to update a named query | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.delete-query | Allowed roles to delete a named query | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.get-query | Allowed roles to get a named query | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.get-queries | Allowed roles to get a named queries | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.get-metadata-attributes | Allowed roles to get all metadata attributes with levels | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.data-search.refresh-metadata-views | Allowed roles to refresh search views in the database | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.notification.subscribe-notifications | Allowed roles to subscribe to system generated notifications | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.notification.get-notification-subscriptions | Allowed roles to get system generated notifications | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| hpc.ws.rs.auth.report.generate-report | Allowed roles to get reports | SYSTEM\_ADMIN GROUP\_ADMIN USER |
| **The following system properties are related to the application** | | |
| hpc.service.security.authenticationTokenSignatureKey | Signature key for the authentication token | User defined |
| hpc.service.security.authenticationTokenExpirationPeriod | Authentication token expiration in mins | 120 |
| hpc.service.dataTransfer.downloadDirectory | Staging directory for 2 hop data transfer. 1st hop is from client to API (stage). And the 2nd hop is from API (stage) to data archive | ${karaf.base}/data/tmp/S3 |
| hpc.service.dataSearch.searchResultsPageSize | Search page results size | 100 |
| hpc.service.dataSearch.defaultCollectionLevelFilter.level | Default search filter for ANY collection level | 1 |
| hpc.service.dataSearch.defaultCollectionLevelFilter.operator | Default search filter operator for ANY collection level | NUM\_GREATER\_OR\_EQUAL |
| hpc.service.dataSearch.defaultDataObjectLevelFilter.level | Default search filter for ANY data object level | 1 |
| hpc.service.dataSearch.defaultDataObjectLevelFilter.operator | Default search filter operator for ANY data object level | EQUAL |
| hpc.service.notification.formatsPath | Notification formats file path | ${karaf.base}/etc/hpc-server/notificationFormats.json |
| hpc.service.notification.mail.host | Email host name to send out email notifications | mailfwd.nih.gov |
| hpc.service.notification.mail.port | Email host port to send out email notifications | 25 |
| hpc.service.notification.notificationDeliveryReceiptsPageSize | Page size to display notifications on UI Dashboard | 100 |
| hpc.service.dataManagement.systemAdminSubjects | System accounts not allowed to set/update permissions by users | rods rodsadmin |
| hpc.service.event.invokerCollectionUpdateNotification | Flag to generate notification to the invoker of collection update | false |
| **The following system properties are related to business services called by REST API for Asynchronous data transfer.** | | |
| hpc.integration.globus.globusAuthUrl | Globus authentication URL. Please do not change this value. | https://auth.globus.org/v2/oauth2/token |
| hpc.integration.globus.globusAuthScope | Globus authentication scope. Please do not change this value. | urn:globus:auth:scope:transfer.api.globus.org:all |
| hpc.integration.globus.archive.type | Archive type. Valid values are TEMPORARY\_ARCHIVE, ARCHIVE. Please do not change this value. | TEMPORARY\_ARCHIVE |
| hpc.integration.globus.archive.endpoint | Globus endpoint to support 2 hop data transfer upload. This is data endpoint must be shared with Globus application account with Read, Write permissions | UUID of the endpoint |
| hpc.integration.globus.archive.path | Archive path on the shared globus endpoint. Upload files are written into this location before they are transferred to Archive storage. The background process clean up the files at globus location after successful registration | /Upload |
| hpc.integration.globus.archive.directory | Archive directory on the mounted drive of the globus endpoint. | /mnt/IRODsTest/FNL\_SF\_Archive/HPCDME\_PROD/Upload |
| hpc.integration.globus.archive.type | Archive type. Valid values are TEMPORARY\_ARCHIVE, ARCHIVE. Please do not change this value. | TEMPORARY\_ARCHIVE |
| hpc.integration.globus.download.endpoint | Globus endpoint to support 2 hop data transfer download. This is data endpoint must be shared with Globus application account with Read, Write permissions | UUID of the endpoint |
| hpc.integration.globus.download.path | Archive path on the shared globus endpoint. Dowload files are written into this location before they are transferred to Archive storage. The background process clean up the files at globus location after successful registration | /Download |
| hpc.integration.globus.download.directory | Archive directory on the mounted drive of the globus endpoint. | /mnt/IRODsTest/FNL\_SF\_Archive/HPCDME\_PROD/Download |
| **The following system properties are related to application services called by integration services.** | | |
| hpc.integration.irods.host | iRODS host name | DEV - fr-s-hpcdm-gp-d.ncifcrf.gov  UAT - fr-s-hpcdm-uat-p.ncifcrf.gov  Production - fr-s-hpcdm-irods-p.ncifcrf.gov |
| hpc.integration.irods.port | iRODS host port number | 1247 |
| hpc.integration.irods.zone | iRODS zone name | tempZone |
| hpc.integration.irods.resource | iRODS resource name | demoResc |
| hpc.integration.irods.basePath | iRODS zone base path | / tempZone/home |
| hpc.dao.postgresql.dbName | iRODS iCAT database name | Please refer to prod/preprod/dev properties file |
| hpc.dao.postgresql.host | iRODS iCAT database host name | Please refer to prod/preprod/dev properties file |
| hpc.dao.postgresql.port | iRODS iCAT database port number | Please refer to prod/preprod/dev properties file |
| hpc.dao.postgresql.username | iRODS iCAT database user name | Please refer to prod/preprod/dev properties file |
| hpc.dao.postgresql.password | iRODS iCAT database user password | Please refer to prod/preprod/dev properties file |
| hpc.dao.postgresql.encryptor.key | Encryptor key used to encrypt integrating system credentials | \*\*\*128-bit-key\*\* |
| **The following system properties are related to application scheduler that runs background tasks.** | | |
| hpc.scheduler.cron.summaryreport.delay | Background task to generate daily summary report notification. This runs every day at 11PM. | 0 0 23 1/1 \* ? |
| hpc.scheduler.cron.weeklysummaryreport.delay | Background task to generate weekly summary report notification. This runs on every Sunday at 11PM. | 0 23 \* \* 7 ? |
| hpc.scheduler.cron.processevents.delay | Background task to process application events. Runs at every 10th second of a minute | 10 0/1 \* \* \* ? |
| hpc.scheduler.cron.processDataTranferUploadReceived.delay | Background task to process data transfer upload requests. Runs at every 5th second of a minute | 5 0/1 \* \* \* ? |
| hpc.scheduler.cron.processDataTranferUploadInProgress.delay | Background task to process “In progress” state upload requests . Runs at every 15th second of a minute | 15 0/1 \* \* \* ? |
| hpc.scheduler.cron.processTemporaryArchive.delay | Background task to process files from temporary archive to permanent archive. Runs at every 25th second of a minute | 25 0/1 \* \* \* ? |
| hpc.scheduler.cron.completeDataObjectDownloadTasks.delay | Background task to process download requests. Runs at every 35th second of a minute | 35 0/1 \* \* \* ? |
| hpc.scheduler.cron.processCollectionDownloadTasks.delay | Background task to process download collection requests. Runs at every 50th second of a minute | 50 0/1 \* \* \* ? |
| hpc.scheduler.cron.completeCollectionDownloadTasks.delay | Background task to complete download collection requests. Runs at every 55th second of a minute | 55 0/1 \* \* \* ? |
| hpc.scheduler.cron.refreshMaterializedViews.delay | Background task to refresh search materialized views. Runs at every 15 minutes | 0 0/15 \* \* \* ? |
| hpc.scheduler.cron.refreshReportViews.delay | Background task to refresh reports materialized views. Runs at every 60 minutes | 0 0/60 \* \* \* ? |

### Set up CA Certificate for SSL port

CA Certificates can be located at: /etc/ssl/certs

1. Locate Certificate files:
   1. Primary certificate – host.domain.crt
   2. Intermediate certification – DigiCertCA.crt
   3. Private Key – host.domain.key
2. Generate a Keystore & Import Intermediate Certification

*keytool –importcert –trustcacerts –file DigiCertCA.crt –alias root –keystore keystore.jks*  
You will be asked to create a password for the keystore (Make sure to remember it).

1. Combine Primary Certificate and Private Key

*openssl pkcs12 –export –name [domain] –in [host.domain.crt] –inkey [host.domain.key] –out keystore.p12*  
You will be asked to create a password. For simplicity, use the same as in step #2

1. Import the Primary Certificate & Private Key into the Keystore

*keytool –importkeystore –destkeystore keystore.jks –srckeystore keystore.p12 –srcstoretype pkcs12 –alias [domain]*

You will be asked to provide passwords. Use what you created in step #2.

You now have keystore.jks that contains the cert. Copy it to *$SERVICEMIX\_HOME/etc/hpc-server*

1. Configure HPC-DME

Uncomment and change these properties as needed to set the path and the password for the keystore.jks file.

$SERVICEMIX\_HOME/etc/gov.nih.nci.hpc.cfg:

#hpc.ws.rs.ssl.keystore.path=${karaf.base}/etc/hpc-server/keystore.jks

#hpc.ws.rs.ssl.keystore.password=hpc-server-store-pwd

#hpc.ws.rs.ssl.truststore.path=${karaf.base}/etc/hpc-server/keystore.jks

#hpc.ws.rs.ssl.truststore.password=hpc-server-store-pwd

### Build HPC DME

Open cmd window and navigate to $HPC\_HOME/

Run “*mvn clean install -P<env>*”

Valid values for <env> are dev, preprod, prod

### Build HPC DME keystore

To setup SSL port certificate with ServiceMix, you would create your own keystore with the certificate or could use build utility to create a keystore with self-signed certificate.

*cd $HPC\_HOME/hpc-server/hpc-ws-rs-impl*

*mvn clean install –Pkeystore*

### Deploy HPC DME into ServiceMix

Run Servicemix from the cmd prompt: servicemix

After successful start, type the following commands in the Servicemix Console

*features:addUrl mvn:gov.nih.nci.hpc/hpc-features/1.0.0-SNAPSHOT/xml/features*

*features:install hpc-server*

### Setup HPC DME Database

HPC DME users iRODS iCAT database to host its own schema. Once iCAT database is setup, please run the following scripts to create HCP DME custom schema in iCAT Postgres database.

<HPC\_HOME>\hpc\hpc-server\hpc-dao-impl\src\main\scripts

If you are doing a new setup, please run all the scripts from “schema” subfolder. If you have already had HPC DME database schema setup, run scripts from migration subfolder based on your current version.

### Import SSL Certs into Java Keystore on the API Server

HPC DME API integrates with Cleversafe, iRODS, NIH AD over SSL to enable secure communication.

Cleversafe, iRODS and NIH AD certs are located on fr-s-hpcdm-uat-p at /opt/certs

* + 1. Please import NIH AD SSL Cert into Java keystore

keytool -importcert -file NIHDeviceRoot.pem -keystore cacerts -alias NIHRoot

keytool -importcert -file NIHDeviceIntermediate.pem -keystore cacerts -alias NIHIntermediate

* + 1. Please import Cleversafe SSL Cert into Java keystore

keytool -importcert -file Cleversafe.crt -keystore cacerts -alias Cleversafe

* + 1. Please import iRODS SSL Cert into Java keystore

keytool -importcert -file iRODS.pem -keystore cacerts -alias iRODS

## Build and Deploy WEB UI Application

Open cmd window and go to to *$HPC\_HOME/hpc-web*

Edit $HPC\_HOME/hpc-web/src/main/resources/application.properties file

|  |  |  |
| --- | --- | --- |
| Property Name | Value | Comments |
| gov.nih.nci.hpc.server | Ex: https://fr-s-hpcdm-gp-d.ncifcrf.gov:7738/hpc-server | API Server URL |
| gov.nih.nci.hpc.web.server | https://fr-s-hpcdm-gp-d.ncifcrf.gov | Web application URL |
| server.port | Ex: 8080 | Web application container port number |

You don’t have to change remaining properties unless you know what you are doing.

From the cmd window, run “mvn clean build”

To run in built container with the war file, run “java -jar target\hpc-web-<version>.war

To deploy the war into Tomcat application container, copy war to <Catalina\_home>/webapps

Update conf/server.xml to include following statement in the <host> element:

<Context path="" docBase="hpc-web-<version>" privileged="true"/>

Open browser and go to Web UI URL.

# Accessing HPC DME

HPC DME can be accessed through its Service API client interface. HPC DME Service APIs are developed based on REST standard. These APIs can be accessed by any interface that is compatible with REST standard. For example, these Service APIs can be accessed through Java, Perl, Ruby, Curl, JavaScript languages programmatically or through UI clients like SOAP UI, REST Client.

The following is the HPC DME Service API URL:

https://<host name>:<port>/<hpc.ws.rs.path>/<Resource Name>

DEV: https://fr-s-hpcdm-gp-d.ncifcrf.gov:7738/hpc-server/

UAT: <https://fr-s-hpcdm-uat-p.ncifcrf.gov:7738>/hpc-server/

Production: https://hpcdmeapi.nci.nih.gov/

HPC DME can also be accessed through its interactive client interface. Please refer to [HPC DME User guide](https://github.com/CBIIT/HPC_DME_APIs/blob/master/doc/guides/HPC_User_Guide.docx) and [API Specification](https://github.com/CBIIT/HPC_DME_APIs/blob/master/doc/guides/HPC_API_Specification.docx) for details.

# HPC DME Services ADMIN Operations

HPC Admin is responsible for HPC user accounts and authorizations setup to facilitate end users in using the solution to manage scientific data/metadata workflow. Managing hardware, storage, network is out of scope for this document.

## Set up System Admin account

HPC DME users are configured to get authenticated with NIH AD pool. When HPC DME is deployed initially, it does not have any accounts created. In order to set up a system admin account, please perform following steps one time only.

1. Login into PostgreSql database.
   1. psql <userId> -h <host> -d <database>
   2. After successfully login into the database, execute following statement

INSERT INTO "HPC\_USER"("USER\_ID", "FIRST\_NAME", "LAST\_NAME", , "IRODS\_USERNAME", "CREATED", "LAST\_UPDATED", , "IRODS\_PASSWORD", "DOC") VALUES ('<NCI Login ID>', '<First Name>', '<Last Name>', '<NCI Login ID>', '<MM-DD-YYYY>', '<MM-DD-YYYY>', null, 'FNLCR/CCBR');

1. Login into iRODS through its iCommands CLI and execute following command.

iadmin mkuser <NCI UserId>#<Zone name> Type rodsadmin

Following are the zone names in different environments:

DEV: tempZone

UAT: ncifHpcZone

Production: ncifprodZone

1. Assign permission on iRODS home to System Admin

ichmod -r own <adminUserId> <iRODS home>

## Set up Service account

HPC DME uses service account to integrate with Globus, Cleversafe and iRODS. While HPC DME security enforces authentication and authorization, the service account is used to perform actions on behalf of the users. For example, the service account is used to upload data into Cleversafe data archive object store. It is used to update system metadata in iRODS. It is also used to create shared endpoints used by HPC DME application account for 2 hop data transfer process. Please section 5.15 for details on application accounts and the setup around 2 hop data transfer. It is also used to manage asynchronous data transfer using Globus. It is important to keep service accout credentials in a safe location. HPC DME stores service accounts (Globus, Cleversafe and iRODS) information in its database in an encrpted manner. HPC DME provides a REST API to register service account with the system initially. This API is accessible to System Administrator role only. Please see API Specification section below for API details.

The following service account credentials are expected to be setup and attached with the service account for HPC DME to function properly.

1. Cleversafe AWS access key and secret access key

2. iRODS user name (rods) and password with “rodsadmin” role and has “own” access to the root folder

3. Globus Application account Id and Key. Please see section 5.15 to setup Globus Application Id.

Globus user name and password with read/write access to “nihfnlcr#gridftp1”

After updating service account credentials, application should be restarted to get changes in effect.

### Setup or Update Cleversafe service account credentials

|  |
| --- |
| Set Cleversafe credentials into a file called Cleversafe.json  {  "account": {  "username": "AWS Access Key Id",  "password": "AWS Secret access key",  "integratedSystem": "CLEVERSAFE"  },  "dataTransferType" : "S3"  }  **Production:**  **AWS Access Key Id:** XmddNHnGowAPX7th9eSb  **DEV/UAT:**  **AWS Access Key Id:** 025ff462-07e1-483b-8dbb-1fc26c7eb17e DelCtesWPOLiDL1aRLh5  Execute following curl command with correct parameter values:  curl -k -H "Content-Type: application/json" -d @Cleversafe.json -X PUT <server URL>/systemAccount --user <userId>:<password> -H "Accept: application/json" |

### Setup or Update Globus service account credentials

|  |
| --- |
| Set Globus application account credentials into a file called Globus.json  {  "account": {  "username": "<Globus Application Client Id>",  "password": "<Globus Application Client Secret>",  "integratedSystem": "GLOBUS"  },  "dataTransferType" : "GLOBUS"  }  **Production:**  **Globus Application Client Id:** 07cec4f4-8d99-4b9d-a513-d9a1148ac054  **DEV/UAT:**  **Globus Application Client Id:** 025ff462-07e1-483b-8dbb-1fc26c7eb17e 025ff462-07e1-483b-8dbb-1fc26c7eb17e  Globus service account uses NIH AD service account credentials.  NIH AD service account is used to create application project and its accounts in <https://developers.globus.org>. Please see section 5.15 for details.  Execute following curl command with correct parameter values:  curl -H "Content-Type: application/json" -d @Globus.json -X PUT <server URL>/ systemAccount --user <userId>:<password> -H "Accept: application/json" |

### Setup or Update iRODS Service account credentials

HPC DME application uses NIH AD account to authenticate with iRODS PAM and to perform service account activities. In order to not to lock yourself into the system, before updating the service account password in NIH AD, first update the application with the new password and then update NIH AD. After updating the password, you will need to restart the application.

|  |
| --- |
| Set iRODS credentials into a file called iRODS.json  {  "account": {  "username": "iRODS UserId",  "password": "iRODS Password",  "integratedSystem": "IRODS"  }  Production Service Account: ncifhpcdmsvcp  DEV/UAT Service Account: ncif-hpcdm-svc  This iRODs service account is a NIH AD account with RODSADMIN privileges.  Execute following curl command with correct parameter values:  curl -H "Content-Type: application/json" -d @iRODS.json -X PUT <server URL>/ systemAccount --user <userId>:<password> -H "Accept: application/json" |

### NIH AD Service Account

HPC DME API integrates with NIH AD for user authentication. To connect to NIH AD Pool, HPC DME API users [ncihpcdmsvcad@nih.gov](mailto:ncihpcdmsvcad@nih.gov) service account to connect to it. Following are the application properties used to connect to NIH AD Pool. Please see section 3.2.1 to set application properties.

hpc.integration.ldap.URL=ldaps://ldapad.nih.gov:636

#hpc.integration.ldap.URL=ldaps://NIHDCADHUB3.nih.gov

hpc.integration.ldap.username=ncihpcdmsvcad@nih.gov

hpc.integration.ldap.password=<Password>

hpc.integration.ldap.base=OU=NIH,OU=AD,DC=nih,DC=gov

hpc.integration.ldap.userIdFilter=cn

hpc.integration.ldap.userIdDomainName=nih.gov

### Password Policy

CleverSafe credentials are generated by ITOG Group and they are in the format of S3 keys. Again, whenever there is an update to these keys, they need to be updated in HPC DME database. Please use HPC DME API specification to update Globus credentials in HPC DME database.

iRODS is setup to use PAM authentication with NIH AD. So iRODS uses NIH AD service account credentials listed in section 5.17 below. Following NIH password policy, this account credentials must be updated using step [5.2.3](#_Setup_or_Update) mentioned above.

## Set Up A Base PATH

Each DOC in HPC DME is assigned with a at least one base path. Each base path can have its own hierarchy, metadata policy and storage configuration. This base path will have to be created in the iRODS system by the System Admin.

|  |
| --- |
| Edit and save the following JSON into basepath.json file.  {  "metadataEntries": [  {  "attribute": "name",  "value": "Base path for DOC XYZ"  }  ]  }  For Example:  For FNLSF\_Archive base path following is the JSON input.  {  "metadataEntries": [  {  "attribute": "name",  "value": "Base path for FNLCR"  }  ]  }  Execute following curl command with correct parameter values:  curl -H “Content-Type: application/json” -d @basepath.json -X PUT <server>/collection/<Base path Name> -H "Accept: application/json" -D <response-header> -o <response-message.json>  Example:  curl -H “Content-Type: application/json” -d @basepath.json -X PUT <server>/collection/FNLSF\_Archive -H "Accept: application/json" -D response.json -o responsemessage.json |

This Base path should be associated with a DOC to start assign permissions and to enable data archival. Please see section 5.10 to register a DOC and set its base path. A DOC may be associated with more than one base path.

## Managing Security

HPC DME has different layers of security. First layer of security is enforced at HPC DME API level through NIH AD credentials. By default, HPC DME is setup to work with NIH AD pool server “ldaps://ldapad.nih.gov:636”. NIH AD connection requires user credentials to be set in application properties file. This service User account also follows NIH password policy. Whenever the password is changed, it has to be updated in gov.nih.nci.hpc.cfg file in <SERVICE\_MIX\_HOME>/etc folder.

HPC DME integrates with iRODS for data management and with Globus for data transfer. iRODS instance supports PAM authentication that integrates its authentication with NIH AD server. Please see the following link to setup PAM authentication for iRODS.

<https://github.com/DICE-UNC/jargon/wiki/Setting-up-PAM>

Execute following command to verify PAM authentication setup

/var/lib/irods/iRODS/server/bin/PamAuthCheck

In order to authenticate iRODS user with NIH AD, iRODS users should be created with NIH loginId only, password is not required.

HPC DME stores iRODS, Globus and Cleversafe service account credentials in its PostgreSql database in encrypted format. HPC DME API uses AES encryption in combination of encryption key set in HPC DME services properties. The following is a property in gov.nih.nci.hpc.cfg. Access to this file should be restricted.

hpc.dao.postgresql.encryptor.key=\*\*\*128-bit-key\*\*

### Roles

HPC DME supports the following roles:

1. USER – HPC DME user role with access to the following functions:
   * Update self-profile (Only through API)
   * Register a data object with metadata
   * Register a collection with metadata
   * Update collection metadata
   * Update metadata on data object
   * Download data object
   * Share data object
   * Search collections and data objects
   * Generate reports
2. GROUP\_ADMIN – HPC DME group admin role with access to following functions in addition to all USER role allowed functions:
   * List group info
   * Create or Delete a group
   * Create user
   * Add user to a group
   * Remove Data Object
   * Query All Active Users
3. SYSYEM\_ADMIN - HPC DME system admin role with access to following functions in addition to all GROUP\_ADMIN role allowed functions:
   * Update User
   * Remove User
   * Query All Users (Including inactive users)
   * Generate system usage reports

HPC DME security is integrated with iRODS security. iRODS roles “rodsadmin”, “groupadmin”, “rodsusers” are mapped to HPC DME types as follows:

SYSTEM\_ADMIN 🡪 rodsadmin

GROUP\_ADMIN 🡪 groupadmin

USER 🡪 rodsusers

### Service API Security

Access to each of HPC DME API services is configured through Spring Security configuration at \hpc-server\hpc-ws-rs-impl\src\main\resources\META-INF\spring\ hpc-ws-rs-bundle-context-osgi.xml. Making a change to this file requires rebuilding the API packages and deploying it to run-time environment to have an effect. This configuration is not expected to change without approval from the Project Management.

|  |
| --- |
| <prop key="hpc.ws.rs.auth.security.register-user">SYSTEM\_ADMIN</prop>  <prop key="hpc.ws.rs.auth.security.update-user">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.security.get-user">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.security.set-group">SYSTEM\_ADMIN GROUP\_ADMIN</prop>  <prop key="hpc.ws.rs.auth.security.authenticate-user">SYSTEM\_ADMIN</prop>  <prop key="hpc.ws.rs.auth.security.register-system-account">SYSTEM\_ADMIN</prop>    <prop key="hpc.ws.rs.auth.data-management.register-collection">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-management.get-collection">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-management.register-data-object">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-management.get-data-object">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-management.download-data-object">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-management.set-permissions">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-management.get-data-management-model">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>    <prop key="hpc.ws.rs.auth.data-search.query-collections">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.query-data-objects">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.add-query">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.update-query">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.delete-query">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.get-query">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.get-queries">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.get-metadata-attributes">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.data-search.refresh-metadata-views">SYSTEM\_ADMIN</prop>    <prop key="hpc.ws.rs.auth.notification.subscribe-notifications">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>  <prop key="hpc.ws.rs.auth.notification.get-notification-subscriptions">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop>    <prop key="hpc.ws.rs.auth.report.generate-report">SYSTEM\_ADMIN GROUP\_ADMIN USER</prop> |

## Create a User

Both System Admin and Group Admin can create HPC user accounts. System Admin can do that by accessing HPC DME REST API. Any REST client can be used to access the API. This admin function can also be performed through UI. Please see HPC\_User\_Guide for detailed instructions.

For example, a system admin would perform the following using curl.

* Create input.json with the following content

|  |
| --- |
| {  "firstName": "John",  "lastName": "Doe",  "doc" : "FNLCR",  "userRole" : "SYSTEM\_ADMIN"  } |

Execute the following command:

*curl -H "Content-Type: application/json" -d @input.json -X PUT $<SERVER>/user/<UserId> --user <Admin NCIUserId:Password> -H "Accept: application/json"*

Valid values for “dataManagementUserType” are "SYSTEM\_ADMIN", "GROUP\_ADMIN" and "USER". Known issue: If you create a user with "GROUP\_ADMIN" role, iRODS Jargon API has a known issue in creating that. A workaround is to create user with “USER” role and update User role with “groupadmin” through iCommand interface. Once you create the user, you would need to login into iRODS iCommand and execute following command to assign “groupadmin” role. This is a workaround until the issue with iRODS Jargon API is fixed.

*iadmin moduser <userId>#<ZoneName> Type groupadmin*

Please see User registration API specification for more details on the API and any errors.

## Updating an Existing User

System Admin is responsible for updating HPC user accounts. System Admin can do that by accessing HPC DME REST API. Any REST client can be used to access the API.

For example, a system admin would do the following using curl.

* Create input.json with the following content

|  |
| --- |
| {  "firstName": "Eran",  "lastName": "Rosenberg",  "doc" : "FNLCR",  "userRole" : "SYSTEM\_ADMIN"  } |

Execute the following command:

*curl -H "Content-Type: application/json" -d @input.json -X POST $<SERVER>/user/<UserId> --user <Admin NCIUserId:Password> -H "Accept: application/json"*

Please see User registration update API specification for more details on the API and any errors.

## Assigning permissions to HPC DME Group administrators

In order to complete user registration with HPC DME and start using HPC DME API, a system administrator should complete the following steps for each group administrator created for a DOC.

* Group Admin User should be NCI user with account in NIH AD
* Assign “own” permission to archive folder to the group administrator.

|  |
| --- |
| Edit ad save the following json into permission.json  {  "entityPermissionRequests": [  {  "path": "/<DOC Base Path> ",  "userPermissions": [  {  "userId": "<UserId>",  "permission": "WRITE"  }  ]  }  ]  }  curl -H “Content-Type: application/json” -d @permission.json -X POST <server>/acl -H “Accept: application/json” -D <response-header> -o <response-message.json> |

## Create System Administrator group and assign System Admins

Create System Admin users group and add system admin users to the group. Assign read permission to each DOC basepath.

|  |
| --- |
| 1. Create Group and add user   Edit ad save the following json into sysadmin.json  {  "addUserIds": [ "abc"]  }  curl -H “Content-Type: application/json” -d @ sysadmin.json -X PUT <server>/group/HPC\_DME\_SYSTEM\_ADMINS -H “Accept: application/json” -D <response-header> -o <response-message.json>   1. Assing READ permission to a DOC basepath   Edit ad save the following json into permission.json  {  "entityPermissionRequests": [  {  "path": "/<DOC Base Path> ",  "userPermissions": [  {  "userId": "<UserId>",  "permission": "READ"  }  ]  }  ]  }  curl -H “Content-Type: application/json” -d @permission.json -X POST <server>/acl -H “Accept: application/json” -D <response-header> -o <response-message.json |

## Accessing HPC DME Database

HPC DME database holds iRODS iCAT database and its own custom tables and views. To access the database, following are the steps:

* SSH to database machine
  + In order to establish SSH connection to production database machine, one must first SSH to either the development (DEV) virtual machine or the user acceptance testing (UAT) virtual machine. Then, from the connected virtual machine, regardless of whether DEV or UAT, one may SSH a second time to the production database machine. This is due to how the production database machine is configured for SSH connectivity.
* sudo su
* su postgres
* psql
* \c ICAT

## Register a DOC

System Admin can register a DOC with the application so that its policies and users can be added to access the application. Registering a DOC is currently a manual process with following steps.

1. Pick an abbreviated name for DOC in consultation with the users.
   1. Example: CCR (Center for Cancer Research)
2. Login into iRODS icommands to create DOC base path.
   1. SSH to iRODS machine
      * In order to establish SSH connection to production iRODS machine, one must first SSH to either the development (DEV) virtual machine or the user acceptance testing (UAT) virtual machine. Then, from the connected virtual machine, regardless of whether DEV or UAT, one may SSH a second time to the production iRODS machine. This is due to how the production iRODS machine is configured for SSH connectivity.
   2. sudo su
   3. su irods
   4. icd .. (You will be on iRODS home path, ex: /tempZone/home
   5. imkdir <Base path name>
   6. ichmod -r own <iRODS service account> <base path name>
3. Define DOC Base path metadata policy
   1. Metadata policy its hierarchy should be defined in JSON format. Please see Section 4.4 in API Specification document for details.
4. Get DOC Base path data store configuration. Each DOC Base path may have its dedicated S3 based vault.
5. Login into Database to insert DOC record
   1. SSH to database machine
      * In order to establish SSH connection to production database machine, one must first SSH to either the development (DEV) virtual machine or the user acceptance testing (UAT) virtual machine. Then, from the connected virtual machine, regardless of whether DEV or UAT, one may SSH a second time to the production database machine. This is due to how the production database machine is configured for SSH connectivity.
   2. sudo su
   3. su postgres
   4. psql
   5. \c ICAT
   6. Execute following SQL statement

INSERT INTO public."HPC\_DATA\_MANAGEMENT\_CONFIGURATION" ("ID",

"DOC", "BASE\_PATH", "S3\_URL", "S3\_VAULT", "S3\_OBJECT\_ID", "S3\_ARCHIVE\_TYPE",

"DATA\_HIERARCHY", "COLLECTION\_METADATA\_VALIDATION\_RULES", "DATA\_OBJECT\_METADATA\_VALIDATION\_RULES")

VALUES (?, ?, ?, ?, ?, ?,

?, ?, ?);

*Example:*

*INSERT INTO public."HPC\_DATA\_MANAGEMENT\_CONFIGURATION"("ID",*

*"DOC", "BASE\_PATH", "S3\_URL", "S3\_VAULT", "S3\_OBJECT\_ID", "S3\_ARCHIVE\_TYPE",*

*"DATA\_HIERARCHY", "COLLECTION\_METADATA\_VALIDATION\_RULES", "DATA\_OBJECT\_METADATA\_VALIDATION\_RULES")*

*VALUES ('c2b5cb88-da37-4ddc-853f-87d431f7e840', 'ESIEB', '/ESIEB\_Archive', 'fr-s-clvrsf-01', 'DSE-TestVault2', 'ESIEB\_Archive',*

*'ARCHIVE', null, null, null);*

1. Restart API Server
   1. Login into API Server machine
   2. sudo su
   3. su <service account>
   4. cd /opt/apache-servicemix-7.0.0.M3
   5. sh starthpcserver.<Version>.sh

## Updating Existing DOC Configuration

System Admin can update a DOC configuration including its metadata hierarchy, policies, data archive connection settings. Updating a DOC configuration is currently a manual process with following steps.

1. Login into Database to update DOC record
   1. SSH to database machine
      * In order to establish SSH connection to production database machine, one must first SSH to either the development (DEV) virtual machine or the user acceptance testing (UAT) virtual machine. Then, from the connected virtual machine, regardless of whether DEV or UAT, one may SSH a second time to the production database machine. This is due to how the production database machine is configured for SSH connectivity.
   2. sudo su
   3. su postgres
   4. Execute following SQL statement

psql ICAT -c “select \* from \” HPC\_DATA\_MANAGEMENT\_CONFIGURATION \” where \”ID\”=’<UUID>’” > <DOC>.txt

* 1. Edit <DOC>.txt and update values as needed and use those value with the following update statement.
  2. psql ICAT -c “

UPDATE public.\"HPC\_DOC\_CONFIGURATION\" SET (

<Column Name>=<Value> …

) WHERE \"DOC\"=’<DOC>’”

1. Restart API Server
   1. Login into API Server machine
   2. sudo su
   3. su <service account>
   4. cd /opt/apache-servicemix-7.0.0.M3
   5. sh start-deploy-servicemix.sh

## Connecting to iRODS through iCommand

Following are the steps to connect to iRODS through its iCommand. On DEV, UAT and Production environments, iRODS is setup to use PAM authentication. To access iRODS directly, you should have your account (with NIH AD UserId) registered with iRODS. iRODS port is enabled with SSL. To connect to iRODS, you need to have its pubic certificate located at /etc/apache2/ssl folder. Please refer to <https://docs.irods.org/4.2.0/icommands/user/> for commands and details.

1. SSH to iRODS server
   1. In order to establish SSH connection to production iRODS server one must first SSH to either the development (DEV) virtual machine or the user acceptance testing (UAT) virtual machine. Then, from the connected virtual machine, regardless of whether DEV or UAT, one may SSH a second time to the production iRODS server. This is due to how the iRODS server is configured for SSH connectivity.
2. From your home folder, type “mkdir .irods” and then “cd .irods”
3. If you have not already setup iRODs client environment, please follow these steps:
   1. From your home folder, type “mkdir .irods” and then “cd .irods”
   2. vi irods\_environment.json
   3. Paste following text and update highlighted text with correct values

{

"irods\_port": 1247,

"irods\_host": "fr-s-hpcdm-gp-d.ncifcrf.gov",

"irods\_user\_name": "NIH ID",

"irods\_zone\_name": "tempZone",

"irods\_authentication\_scheme": "PAM",

"irods\_ssl\_ca\_certificate\_file": "/etc/apache2/ssl/fr-s-hpcdm-gp-d.ncifcrf.gov.crt",

"irods\_ssl\_verify\_server": "none"

}

* 1. Type “iinit” and enter your NIH AD password
  2. Type any iRODS commands

## Delete or Edit a Collection on iRODS

Following are the steps to delete a collection from iRODS directly.

1. Following steps from section 5.12 to connect to iRODS
2. To recursively remove a collection and its contents and metadata “irm -rf <Collection full path>”
3. To move a collection from one location to another location: “imv <source location> <destination location>
4. To update permissions on a collection and its contents: “ichmod -r <null | read | write | own> <UserId> <Collection path>”
5. To update metadata of a collection: “imeta mod -d <Collection path> <Attribute name> <Attribute value>
6. To add metadata of a collection: “imeta add -d <Collection path> <Attribute name> <Attribute value>

## Delete or Edit a Data File on iRODS and Cleversafe

Following are the steps to delete a data from iRODS directly. Removing a data file from iRODS would delete data file logical reference and its metadata is deleted. Physical data file residing on Cleversafe would not be deleted.

1. Following steps from section 5.12 to connect to iRODS
2. To remove a data file and its metadata “irm <Data file full path>”
3. To remove data file from Cleversafe, use “S3 Browser” or “CyberDuck” to connect to Cleversafe vault with appropriate credentials. Navigate to the data file and right click on the file name to get pop-up menu. Click on delete to permanently delete the file from Archive.

## Globus Application Account Configuration

HPC DME Application uses Globus to transfer data asynchronously to and from the archive. At this time, setting up a Globus endpoint CleverSafe object store is not tried or assessed. To enable asynchronous data transfer, our application took 2 hop data transfer approach.

HPC DME uses Globus application account to perform any transfers through our System Globus endpoint. DEV and UAT environments share one Globus application account and endpoint. Production environment has its own Globus application account and service endpoint. Globus Service endpoint is the staging area to keep the data during 2 hop upload or download process.

HPC DME is using Globus endpoint “nihfnlcr#gridftp1” hosted at Frederick Data center. The file system attached to this Globus endpoint is mounted on API Server to move files for 2 hop transfer using application accounts.

**/mnt/IRODsTest** is the Globus endpoint storage path mounted on both DEV/UAT and Production API Servers.

Following are the steps to setup 2 hop transfer configuration:

Create Application Service Account:

* Login into developers.globus.org with NIH AD Service account (Production/DEV-UAT) credentials.
* Create Project with Project name and description.
* Click on “Add new app” to create application account.
* Enter “App Name” and other details as below
* 
* Click on “Generate New Client Secret” to generate secret key for the account.
* Use generated “Client Id” and “Secret Key” to setup Globus service account detailed in section 5.2.2
* Keep “Client Id” and “Secret Key” in safe place.

Creating a service Endpoint:

* Login into Globus.org with NIH AD Service account (Production/DEV-UAT) credentials.
* Browse nihfnlcr#gridftp1 and select a folder to be used a stage area for 2 hop transfer and click on “Share”
* Enter Share display name and click on Create button.
* Update “Share” permissions on this endpoint to the Globus Application account created in the section above. Make sure to give “Write” permission to the Globus application account.
* Go to Transfer Files and browse the endpoint you just created. Create a folder for upload and folder for download
* Give file system level write permission to NIH AD service account to the upload and download paths created on the endpoint
* Go to endpoint details to get UUID

Setting up Application Configuration:

* Go to HPC\_DME\_APIs\src\hpc-server\hpc-features\src\main\resources and open DEV/UAT/Production properties file based on the environment you are working with.
* Update following properties

hpc.integration.globus.archive.endpoint=< Endpoint UUID >

hpc.integration.globus.archive.path=< Upload folder name >

hpc.integration.globus.archive.directory=< Physical path on the mounted Globus file system >

#hpc.integration.globus.archive.type=TEMPORARY\_ARCHIVE

hpc.integration.globus.download.endpoint=< Endpoint UUID >

hpc.integration.globus.download.path=< Download folder name >

hpc.integration.globus.download.directory=< Physical path on the mounted Globus file system >

## Accessing Application Log files

HPC DME API runs on Apache ServiceMix. To access application logs:

1. SSH to API server
   1. In order to establish SSH connection to production API server, one must first SSH to either the development (DEV) virtual machine or the user acceptance testing (UAT) virtual machine. Then, from the connected virtual machine, regardless of whether DEV or UAT, one may SSH a second time to the production API server. This is due to how the production API server is configured for SSH connectivity.
2. Go to /opt/ apache-servicemix-<7.0.0.M3> folder
3. Go to /data/log folder to see application log files

At every restart of the application server, the log files are backed up to /backuplogs folder under /opt/ apache-servicemix-<7.0.0.M3>.

## Clean up and reset initial test data load for Development, UAT Tiers

## General Configuration

VM/Hardware:

* 1. DEV: fr-s-hpcdm-gp-d.ncifcrf.gov
  2. UAT: fr-s-hpcdm-uat-p.ncifcrf.gov
  3. Production:
     1. API: fr-s-dmeapi-t-p.ncifcrf.gov
     2. Database: fr-s-dmedb-t-p.ncifcrf.gov
     3. iRODS: fr-s-hpcdm-irods-p.ncifcrf.gov
     4. Web UI: fr-s-hpcdm-web-p.ncifcrf.gov
  4. Production mirrored VM
     1. API: fr-s-hpcdm-api-p.ncifcrf.gov
     2. Database: fr-s-hpcdm-db-p.ncifcrf.gov
  5. Database Daily Backup files storage: fr-s-hpcdm-gp-p.ncifcrf.gov

Service User Account:

DEV: ncif-hpcdm-svc

UAT: ncif-hpcdm-svc

Production: ncifhpcdmsvcp

NIH AD Account: ncihpcdmsvcad@nih.gov

Globus Endpoint for 2 hop data transfer: nihfnlcr#gridftp1

Mounted drive to access Globus Isilon storage: /mnt/IRODsTest

## Technology Stack

|  |  |
| --- | --- |
| Software/Package | Version |
| Java | 1.8 |
| Apache ServiceMix | 7.0.0 |
| Apace Tomcat | 8.0.x |
| PostgreSQL | 9.5.x |
| iRODS | 4.2 |
| Jargon API | 4.1.10.0.RELEASE |
| S3 API | 1.11.125 |
| Spring | 3.2.17.RELEASE |