**High Performance Computing (HPC) Data Management Environment (DME) Getting Started: Command Line Utilities**

**Purpose**

This document is intended to serve as a guide for getting started on using the High Performance Computing (HPC) Data Management Environment (DME) for the National Cancer Institute (NCI). Specifically, this guide is intended to help the reader perform bulk registrations of files into the DME using Bash shell utilities. In this context, bulk means multiple files per registration transaction.

**Assumptions**

Prerequisites are met (described in next section).

Reader is comfortable: working with Bash shell command-line environment (or Cygwin Bash emulator on Windows), making minor edits to properties files, and decompressing archive files (using any tool of reader’s choice).

(Optional) If reader intends to use Globus with DME, then it is assumed they are familiar with Globus or can figure out how to perform various tasks with Globus without guidance from this document. For background information about Globus and how it can be used with DME, please read the DME User Guide available at

<https://github.com/CBIIT/HPC_DME_APIs/blob/master/doc/guides/HPC_User_Guide.docx>.

The Pre-requisites section of the User Guide contains the relevant content.

**Prerequisites**

1. Java runtime environment (JRE)
   1. To test if you have JRE installed and ready to use, in command window, type “java -version”. If it is installed, command returns version information about JRE.

**Instructions on Setting Up HPC DME Client Utilities**

1. From GitHub, download applicable archive depending on whether you are using Production DME or User Acceptance Testing (UAT) DME.
   1. For Production DME, visit <https://github.com/CBIIT/HPC_DME_APIs/tree/master/archives/production> and download desired version and format (zip file or gzipped tar ball)
   2. For UAT DME, visit <https://github.com/CBIIT/HPC_DME_APIs/tree/master/archives/uat> and download desired version and format
2. Extract archive to directory of your choice. These instructions shall refer to that directory as <CLIENT\_UTILS\_HOME>.
3. In <CLIENT\_UTILS\_HOME>, edit *hpcdme.properties* file to make configuration changes. Sub-items a-d below offer elaboration about this. It is recommended to make backup copy of this file before editing.
   1. Change *hpc.ssl.keystore.password* property to correct value. Request this password from appropriate HPC DME technical point of contact.
   2. Change *hpc.user* property to your username (a.k.a. your NIH username).
   3. Optionally for Globus integration, uncomment and change *hpc.globus.user* property to your Globus user ID.
      1. Please use Globus ID having fully qualified form resembling [someuser@globusid.org](mailto:someuser@globusid.org)
      2. For setting property value, use brief form of Globus ID without *@globusid.org* suffix
   4. Optionally for Globus integration, uncomment and change *hpc.default.globus.endpoint* property to UUID of Globus endpoint to utilize.
   5. For other properties not mentioned above, only change settings if you are confident you know what you are doing. Otherwise, leave them as-is.
4. For Cygwin users only, run following command to alter *functions* file to use Unix standard end of line delimiters. $ sed -I "s/\r$//" functions
5. Edit your *~/.bashrc* or *~/.profile* file to append following lines in Courier New font but replacing <CLIENT\_UTLS\_HOME> with absolute path to that directory.

# export environment variable HPC\_DM\_UTILS pointing to directory where

# HPC DME client utilities are, then source functions script in there

export HPC\_DM\_UTILS=<CLIENT\_UTILS\_HOME>

source $HPC\_DM\_UTILS/functions

1. Source whichever file you modified in Step 5 to make the changes take effect in current Bash session.

$ source ~/.bashrc

or

$ source ~/.profile

**How to Register Collections with HPC DME Client Utilities**

Preliminary steps – Do these prior to running utilities that perform DME transactions

1. Open or switch to a command window.
2. Run DME client utility *dm\_generate\_token* to generate DME API authentication token. With a token, you avoid repeatedly typing your password for every command requiring authentication to DME API server.

$ dm\_generate\_token

The subsequent table describes DME client utilities you can use to register/put multiple files in DME using one transaction; in other words, the utilities support bulk registration. Please refer to the User Guide at <https://github.com/CBIIT/HPC_DME_APIs/raw/master/doc/guides/HPC_User_Guide.docx>, In particular, refer to the section entitled "Executing HPC DME API with Command Line Utilities".

|  |  |
| --- | --- |
| **Summary: DME Client Utilities for Bulk Registration** | |
| **Utility** | **Description & Usage Example** |
| dm\_register\_directory | Register a Collection to DME using local directory.  Usage example:  $ dm\_register\_directory /home/joeuser/projX \  $ /someDOC/joeuser/projX  Notes:   1. First argument is path to local directory. 2. Second argument is destination path in DME. 3. Command as shown above selects specified directory and all its contents. 4. There are additional options not shown in above usage example which support finer-grained control. Refer to User Guide mentioned above this table for details. 5. To register metadata for the resulting Collection, there must be a JSON metadata file that is located in the same parent directory as the local directory. That file must have name that is the name of the local directory appended with file extension of ".metadata.json". 6. To register metadata for any resulting Sub Collection or Data Object in the resulting Collection, there must be a JSON metadata file that is in the same parent directory as the originating sub directory or file in the local file system. The metadata file must have name that is the corresponding item's name including file extension, if applicable, appended with the file extension of ".metadata.json". |
| dm\_register\_globus\_directory | Register a Collection to DME using a Globus directory.  Usage example:  $ dm\_register\_globus\_directory \  $ /myWork/studyXyz/ \  $ /theDOC/studies/studyXyz  Notes:   1. First argument is path to Globus directory at default Globus endpoint.    1. Default Globus endpoint is set in the properties file as described in this guide in the section entitled **Instructions on Setting Up HPC DME Client Utilities**   specifically Step 3, part d.   1. Second argument is destination path in DME. 2. If there is no default Globus endpoint configured or you wish to use a Globus endpoint different from the default one, you may specify a third argument to indicate which Globus endpoint. This optional third argument must be UUID of intended Globus endpoint. 3. To register metadata for the resulting Collection, there must be a JSON metadata file that is at the Globus endpoint in the same parent directory as the specified Globus directory. That file must have name that is the name of the Globus directory appended with file extension of ".metadata.json". 4. To register metadata for any resulting Sub Collection or Data Object in the resulting Collection, there must be a JSON metadata file that is at the Globus endpoint in the same parent directory as the originating Globus sub directory or Globus file. The metadata file must have name that is the corresponding item's name, including file extension if applicable, appended with the file extension of ".metadata.json". |

You can use HPC DME web GUI client to examine Collections (directories/folders) and Data Objects (files) that are in HPC DME storage. For production web client, go to <https://hpcdmeweb.nci.nih.gov>. For UAT web client, go to <https://fr-s-hpcdm-uat-p.ncifcrf.gov/>. Information on how to use HPC DME web client is available in the User Guide at <https://github.com/CBIIT/HPC_DME_APIs/raw/master/doc/guides/HPC_User_Guide.docx>, specifically the section entitled "Executing HPC DME API with Web Client".