*HPC Data MANAGEMENT*

storage provisioning process

Version *1.0*

*04/05/2018*

**Version History**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version Number** | **Implemented**  **By** | **Revision**  **Date** | **Approved**  **By** | **Approval**  **Date** | **Description of Change** |
| 1.0 | Zhengwu Lu | 04/05/2018 |  |  | Initial Draft with Eric’s comments incorporated |
|  |  |  |  |  |  |

Table of Contents

1. PURPOSE 4
2. SUPPORT RATIONALE**…………………………………………………………………………..…**4
3. STORAGE REQUEST/APPROVAL………………………………………………………………….4
4. STORAGE PROVISIONING…………………………………………………………........................5
5. FOLLOW-UP & INTEGRATION………………… …………………………………...5
6. RECLAMATION……………………………………… ………………………………6
7. APPENDIX A: Template email to FNL Help Desk 6
8. **Purpose**

The HPC (High performance computing) Data Management Storage Provisioning Process navigates the administrative steps to be taken to request and have a production storage vault provisioned for centralized data archiving by all NCI DOCs (Division, organization and centers), and any applicable processes related to the archive storage after provisioning.

This document is a living document and will undergo revisions as NCI/FNL develops and refines its scientific data management life cycle policies and procedures.

1. **Support Rationale**

Many research laboratories or groups at NCI or FNLCR generate large volume of data or collaborates with extramural communities and have a need to archive their data in a central safely guarded location for later analysis. The magnitude of the data and speed of accumulation is beyond what can be handled through a local thumb drive or hard drive. The HPC Data Management Services group ([HPC\_DME\_Admin@nih.gov](mailto:HPC_DME_Admin@nih.gov)) under the NCI Strategic and Data Science Initiatives will be the point of contact to coordinate and facilitate the provisioning of centralized storage for all NCI DOCs with such emerging need and integration with HPC DME (Data Management Environment), where capabilities for storing, managing, transferring and sharing data across different systems securely and efficiently via RESTful API have been custom developed and in operational use since January 2017.

Communication for actual storage archive provisioning must be sent by email to the FNL Help Desk ([NCI-FHelpdesk@mail.nih.gov](mailto:NCI-FHelpdesk@mail.nih.gov)**)** currently. FNL internally creates an Alloy ticket in that system for recordkeeping. This process may be changed when the FNL IT Asset Management (i.e., ServiceNow) is in operation.

1. **Storage Request/Approval**

### Each DOC lab head or designee reaches out to their IT management lead and get approval of using NCI CleverSafe (or other applicable storage system) as archive store of their data/files

### DOC IT management lead reaches out to NCI CBIIT CIO, government sponsor or Strategic and Data Science Initiatives Director and get go-ahead for either provisioning a separate vault or using an existing vault.

### Data Science Initiatives Director assigns a HPC designee to assist with the provisioning

### HPC Designee consults with and advises of the DOC designee on the logistics of vault provisioning or creating a base path on an existing vault

### HPC designee reaches out to FNL Storage & Resources Group or Information Technologies Operations Group manager/director and confirm availability of storage and storage administration support.

### If an existing vault is desired and agreed upon, adding additional storage capacity and creating DOC specific base paths will be expected and facilitated through HPC designee. Otherwise, a new storage vault will be requested as below.

### 

1. **Storage Provisioning**

### The following are pre-requisites to provision a separate vault for a new DOC:

### The DOC designee will be the business owner and group administrator

### DOC designee will be the central point of contact. Exceptions to normal use of the data archive resource will be handled on a case-by-case basis with the request coming from the DOC designee

### HPC Data Management Services designee will work with the DOC designate helping with the actual process

### 

### Actual process in procuring the requested storage vault:

### HPC Data Services designee will work with DOC designate creating a meaningful name for the production vault, following the format of “HPCDME\_XXX\_PROD”, where "XXX" may be replaced with the short abbreviation of the DOC or laboratory unit.

### DOC designee will work with HPC Data Services designate to best estimate the initial vault storage capacity need for 2~4 years

### DOC designee submits a FNL Help Desk ticket via email (NCI-FHelpdesk@mail.nih.gov) for provisioning the storage vault. Please note that the email must explicitly indicate that the HPCDME production service account (“ncifhpcdmsvcp”) be granted read/write access rights on the new vault

### See appendix A for a template email request (NCI-FHelpdesk@mail.nih.gov) to provision a new storage vault (bucket) for a DOC/business unit from a DOC designee.

1. **Follow-up & Integration**

### HPC Data Services designee follows up with DOC designee and/or FNL Storage administrator on provisioning status for the newly requested vault using the Help Deask ticket# assigned

### HPC Data Services designee verifies that the provisioned production vault for the DOC is now accessible, and works with the DOC designee and/or others in gathering and populating data management configuration table records for the newly created DOC (BASE\_PATH, S3\_VAULT, S3\_URL, and DOC/Basepath related JSON rule files).

### HPC Data Services designee verifies that test files can be uploaded, and all test data/metadata are completely deleted from both front and back ends for the DOC in production ready use.

1. **Reclamation**

### Provisioned storage archive and the capacity, regardless of what storage technology/system, is assured for 4 years after which availability of unused space may no longer be assured without an updated request. The implementation of such process may depend on how the emerging data life cycle policies develop at NCI and NIH. It is possible that unused space in the Archive vault may be reclaimed for reallocation at the end of a four year period following the last allocation which either establishes or increases the size of the vault.

### 

1. **APPENDIX A:** **Template email to FNL Help Desk**

### Hi FNL Help Desk:

### This request is to submit to the FNL Storage&Resources group for provisioning a Production CleverSafe Vault for HPCDME XXX Production tier, where "XXX" is the abbreviation of the DOC:

### Name the new vault “HPCDME\_XXX\_PROD”, where "XXX" is the abbreviation of the DOC.

### Production CleverSafe to be provided with XXX TB capacity. This vault will be used to host initial data storage need and be accommodated for increment per future need.

### On privilege/permission setup, please make service account “ncifhpcdmsvcp” to have read/write access rights on the new vault.

### I will be the point of contact from XXX (where "XXX" indicates the abbreviation of the DOC) on requesting direct access to the storage vault for the future.

### Please let me know if there are any questions for this request.

### Thanks,

### XXXX