**HPC DATA MANAGEMENT**

**Data Transfer Process**

**Version 1.0**

03/27/2019

**Revision History**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Implemented**  **By** | **Revision**  **Date** | **Description of Change** |
| 1.0 | Eran Rosenberg | 03/27/2019 | Initial Draft |

Table of Contents

[1. Introduction 4](#_Toc5819009)

[2. Upload (Data Registration) 5](#_Toc5819010)

[2.1. Sync -> Cleversafe 5](#_Toc5819011)

[2.2. AWS S3 -> Cleversafe 6](#_Toc5819012)

[2.3. Upload URL -> Cleversafe 6](#_Toc5819013)

[2.4. Globus -> Cleversafe (2-Hop Upload) 7](#_Toc5819014)

[2.5. Sync -> POSIX 7](#_Toc5819015)

[2.6. Globus -> POSIX 7](#_Toc5819016)

# Introduction

This document describes the process in which data is transferred (upload/download) to/from HPC-DME archive. The system supports multiple archive and source types through a single API. This document provides details on the process implemented to transfer in each combination of source and archive type.

HPC-DME is using 2 technologies to transfer data – Globus and S3.

At the integration layer, HpcDataTransferProxy defines a generic data transfer interface. This interface is implemented by 2 concrete classes, one for Globus and one for S3.

The following table shows the various types of archives, and the data transfer proxy implementation that is used to communicate with it:

|  |  |
| --- | --- |
| **Archive** | **Data Transfer Proxy Impl** |
| POSIX (File System) | Globus |
| Temporary (Server’s Globus Endpoint) | Globus |
| Cleversafe | S3 |

The following table shows the various types of user’s source/destination, and the data transfer proxy implementation that is used to communicate with it:

|  |  |
| --- | --- |
| **User’s Source/Destination** | **Data Transfer Proxy Impl** |
| Globus Endpoint | Globus |
| AWS S3 | S3 |

# Upload (Data Registration)

The process of file registration starts with registering the data object and metadata in iRODs. If this part completes successfully, the data transfer process to upload is kicked off. The upload data transfer process is controlled by 2 system metadata:

* data\_transfer\_type – The data transfer proxy that is engaged in upload (S3 or GLOBUS). Please note that in a particular case (wil be discussed later) both data transfer proxies are used to upload a file.
* data\_transfer\_status – represents a ‘state’ the transfer is in (e.g IN\_PROGRESS)

The following upload scenarios are supported:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Source** | **Sync / Async** | **Archive** |
| 2.1 | API attachment | Sync | Cleversafe |
| 2.2 | AWS S3 | Async | Cleversafe |
| 2.3 | Upload URL | Async | Cleversafe |
| 2.4 | Globus | Async | Cleversafe |
| 2.5 | API Attachment | Sync | POSIX |
| 2.6 | Globus | Async | POSIX |

For each scenario, a statechart shows the process and the values of the 2 system-metadata used to track and complete the upload

## Sync -> Cleversafe

A drawing of a person

Description automatically generated

## AWS S3 -> Cleversafe

A screenshot of a cell phone

Description automatically generated

## Upload URL -> Cleversafe

A screenshot of a cell phone

Description automatically generated

## Globus -> Cleversafe (2-Hop Upload)

A screenshot of a cell phone

Description automatically generated

## Sync -> POSIX

A drawing of a person

Description automatically generated

## Globus -> POSIX

