# HPC-DME Data Transfer w/ Globus S3 connector

# Architecture Overview

# Overview

HPC-DME provides an API and UI to transfer data between its IBM Cleversafe archive and user’s Globus endpoint. The team is working on utilizing the Globus S3 connector to perform the data transfer instead of the existing home-grown solution. The purpose of this document is to provide an overview of the as-is solution and the new architecture where the Globus S3 connector is used.

# As-is Solution

The as-is solution to transfer data between Globus endpoint and Cleversafe consists of a 2-step process. The file is first transferred completely to an HPC-DME server owned Globus managed endpoint. From the HPC-DME server’s Globus endpoint - the file is then transferred to Cleversafe (in an upload use-case) or to the user’s Globus endpoint (in a download use-case).

The HPC-DME API server has direct filesystem access to its Globus endpoint (NFS mount), which allows using the AWS S3 SDK to transfer files from/to Cleversafe to the HPC-DME Globus endpoint. Because the file is fully transferred twice, this 2-step process often referred to as the ‘2-hop solution’. The 2-hop solution performs checksum verification end-to-end.

A screenshot of a cell phone

Description automatically generated

# Solution w/ Globus S3 Connector

In this solution, a Globus S3 connector was configured to expose the Cleversafe archive as a Globus endpoint, and thus having the HPC-DME API server use the Globus transfer API to transfer files ‘directly’ between the user’s Globus endpoint and Cleversafe. The requirement is to have Globus perform checksum verification.

A close up of a piece of paper

Description automatically generated

# Performance Benchmark

The team performed a benchmark test to understand the performance gain/loss of moving to the new architecture utilizing the Globus S3 connector. The test included transferring of a small file (2.5GB), medium-size file (10GB) and a large file (35GB) – each repeated 5 times.

The transfer w/ Globus S3 connector was tested with and without checksum verification to understand the performance impact of the checksum itself. The 2-hop solution was tested with checksum verification only.

The results show the 2-hop solution is favorable across all file sizes but with different margins. The results suggest that performing checksum verification w/ Globus S3 connector doubles the transfer time. Also – in transferring a large file (35GB), the 2-hop solution was three times faster.

