

HIFIS transfer service: FTS for Helmholtz

Wednesday, 24 March 2021 13:40 (20 minutes)

The German Helmholtz Association consists of 19 research institutes distributed over Germany, covering a wide variety of research topics ranging from particle and material physics over cancer research to marine biology. In order to stimulate collaborations between different centres, Helmholtz established so-called incubator platforms. One of those platforms, HIFIS, aims to provide the infrastructure for federating IT services offered by the different Helmholtz centres. The other platforms provide thematically bound resources for domain scientists in the forms of project funding for cross-centre and cross-domain collaborations, computing resources, and consulting services.

Use cases of interdisciplinary research are arising in the platforms and show that there is a definitive need to transfer a significant amount of large data sets between centres. This results from the fact that primary data acquisition and the subsequent processing steps are increasingly distributed over multiple centers. As data processing is generally sensitive to network latencies, remote data access is not efficient in those cases and consequently data needs to be transferred from the primary institution to another, where the data processing is taking place.

In order to cater to those needs, a file transfer service is being established by HIFIS for convenient and automated data transfers between the sites of the aforementioned cross-centre research groups. After evaluating alternative solutions like Globus Online and Onedata, we agreed to use FTS3 for reasons we will elaborate on during the presentation. FTS3 is a file transfer service that can commission data transfers between storage endpoints and has been developed at CERN for the transfer of WLCG research data between CERN and several hundred LHC Tier centres. Those endpoints need to be able to communicate with FTS and each other using a third-party copy (TPC) extension of the HTTP protocol to transfer data directly. In order to facilitate an easy installation of endpoints, we provide an Apache web server extension that complies with the needs of FTS3 and can thus act as an endpoint for data transfers via HTTP-TPC.

We will present the necessary prerequisites and setup variants for such an endpoint, its configuration details, as well as a brief overview of the modifications applied to the Apache modules. Adding to that, we will present insights into the access possibilities, performance and reliability of the data transfers.

Summary

We present details of and use cases for a storage endpoint solution based on the Apache web server capable of making use of FTS3 for transferring large data sets.

Primary authors: Dr FUHRMANN, Patrick (DESY/dCache.org); Dr MILLAR, Paul (DESY); Mr WETZEL, Tim (Deutsches Elektronen-Synchrotron DESY); Dr JANDT, Uwe (DESY)

Presenter: Mr WETZEL, Tim (Deutsches Elektronen-Synchrotron DESY)

Session Classification: Data Management & Big Data Session

Track Classification: Data Management & Big Data