



Contribution ID: 63

Type: **not specified**

HIFIS backbone transfer service: FTS for everyone

Wednesday, October 14, 2020 10:40 AM (20 minutes)

The German Helmholtz Association (HGF) encompasses 19 research institutes distributed all 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, the HGF established so-called incubator platforms. Two of those platforms, relevant for this presentation, are the Helmholtz Artificial Intelligence Cooperation Unit (Helmholtz AI) and Helmholtz Federated IT Services (HIFIS). While Helmholtz AI was established to connect domain scientists and AI experts for a stronger adoption of AI solutions for increasingly complex research tasks, HIFIS targets the exploitation of synergy effects in federated IT services offered by the different HGF centres.

During the ongoing ramp-up phase of both platforms, specific use cases of interdisciplinary research are arising and showing that there is a definitive need to transfer a significant amount of large data sets between centres. This primarily results from the fact that the currently used AI solutions are trained on specific data sets and that the processing of that data is sensitive to network latencies. Consequently, remote data access is less efficient in those cases and consequently data needs to be transferred from the domain scientists' home institutions to the AI experts' location, where the model training is taking place.

In order to cater to those needs, a file transfer service is being established by HIFIS for convenient and automated data transfer between the sites of those interdisciplinary research groups. After evaluating competing solutions like Globus Online and Onedata we agreed to go for 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 each other as well as with FTS, using a third-party copy (TPC) extension of the HTTP protocol. In order to facilitate an easy installation of endpoints, which are not the known WLCG storage systems (like dCache, DPM and EOS), we provide an Apache web server extension that complies with the needs of FTS3 and can thus act as an storage endpoint for data transfers via HTTP-TPC.

We will present the necessary prerequisites for such an endpoint, its configuration details as well as the modifications applied to the Apache modules. Adding to that, we will present insights into the performance and reliability of the TPC data transfers measured with test data sets.

Primary authors: WETZEL, Tim (Deutsches Elektronen-Synchrotron DESY); MILLAR, Paul; Dr JANDT, Uwe (Deutsches Elektronensynchrotron (DESY)); FUHRMANN, Patrick

Co-authors: Mr HELMERT, Stefan (Helmholtz-Zentrum Dresden-Rossendorf); Mr HUSTE, Tobias (Helmholtz-Zentrum Dresden-Rossendorf); Dr JUCKELAND, Guido (Helmholtz-Zentrum Dresden-Rossendorf)

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

Session Classification: Wednesday morning

Track Classification: Storage & Filesystems