

Data Avenue: a flexible data transfer service among various DCIs

Wednesday, 21 May 2014 14:10 (20 minutes)

Globus On-line is a popular service to transfer data among Globus Grid storages. Data Avenue is a similar service but with much more flexibility. It is a file commander tool for data transfer, enabling easy data moving between various storages services (such as grid, cloud, cluster, supercomputers) by various protocols: http, https, sftp, gsiftp, srm.

With Data Avenue users can up- and download their data to storage services for scientific computation. Additionally, they can copy, move and delete files as well as they can create and copy folders.

The Data Avenue family consists of three members:

The Data Avenue Blacktop is the core service that provides unified access to different types of storages. It offers a web service interface for performing different operations on the supported storages (SFTP, GSIFTP, HTTP, HTTPS, SRM, Amazon), and an HTTP server for convenient data up- and download. The Data Avenue core service can be accessed with an easy-to-use Java API making integration with existing software components an easy task.

The Data Avenue is the user interface that makes data transfer very easy and user friendly.

The Data Avenue@SZTAKI is the SZTAKI installation of the Data Avenue service. It can be used by those who choose not to install their own service or just would like to see how it works!

Data Avenue is also realized as a Liferay portlet and hence every science gateway based on Liferay can use this portlet to provide the service for their users.

Wider impact and conclusions

The Data Avenue service enables the fast and easy transfer of data among different grid and cloud storages. This enables easy collaboration between applications and communities that use, for example, gLite, globus and cloud infrastructures. Beyond that the service can be integrated to any Liferay gateways as an independent Liferay portlet.

Furthermore it can be integrated with workflow systems enabling workflow nodes to process data from various DCI storages. This flexibility of accessing and processing data from various DCI storages tremendously simplifies collaboration of user communities and applications/services using different DCIs.

URL(s) for further info

<https://data-avenue.eu/>

Description of work

The Data Avenue service has been developed by MTA SZTAKI in the frame of the EU FP7 SCI-BUS project based on the data usage requirement analysis performed by the project partners.

The service has been implemented in a very flexible way. The core part realizes the common file transfer related activities while the storage protocol specific actions are executed by various data storage plugins. The most versatile data storage plugin is based on the JSAGA data management API.

The core service called as Data Avenue Blacktop is running on the SZTAKI Cloud and can be accessed in several ways:

1. Via the Data Avenue user interface that also runs in the SZTAKI Cloud
2. Via a Liferay portlet. In this way users of Liferay based portals can easily access the Data Avenue service if the Data Avenue portlet is added to the portlet set of the Liferay based portal. For example, every WS-PGRADE portal contains this portlet and hence users of the more than 80 production WS-PGRADE gateways can use the service.

3. Since a Java API is available every user community can set up their own Data Avenue user interface.

Also based on the JAVA API every workflow system can be integrated with the Data Avenue service. It means that nodes of a workflow running in different DCIs can use data files that are stored in various grid and cloud storages.

As a reference implementation the Data Avenue service is integrated with WS-PGRADE workflows so WS-PGRADE gateway users can easily develop workflows that use data from different DCI storages.

Primary authors: Dr HAJNAL, Akos (MTA SZTAKI); KACSUK, Peter (MTA SZTAKI); FARKAS, Zoltan (MTA SZTAKI)

Presenter: FARKAS, Zoltan (MTA SZTAKI)

Session Classification: New data management solutions for EGI

Track Classification: Requirements and solutions for data management and computing (Track Leaders: B. Konya, H. Heller, S. Tarkoma)