



Contribution ID: 39

Type: **Presentation**

## Sync & share solution to replace HPC home?

*Wednesday, January 31, 2018 12:20 PM (20 minutes)*

Handling 100s of Terabytes of data at the speed of 10s of GB/s is nothing new in HPC. However, high performance and large capacity of the storage systems rarely go together with their ease of use. HPC storage systems are specifically difficult to access from outside the HPC cluster. While researchers and engineers tolerate the fact that they need to use rigid tools and applications such as textual SFTP client or globus-url-copy and SSH/text consoles in order to store, retrieve and manipulate the data within HPC system, such inconvenience does not support productivity. Oposite, the burden related to exchanging data with HPC systems makes learning curve for HPC systems adopters steep and is often a source of errors and delays in implementing and running computing workflows.

Moreover, in the classical HPC storage architecture, the data flow to and from the HPC system has to be explicitly managed by users and their applications or workflow systems. For instance, downloading the computations results from the cluster to the user workstation or mobile device has to be triggered manually by user or steered by a mechanism that is synchronised with computing jobs management system. Compared to ease fo use of cloud storage systems, such a data flow control reminds 80's or 90's of the previous century.

In our presentation we will demonstrate how a high-performance and robust sync & share storage system based on well-optimised synchronisation mechanism and equipped with functional and well-optimised user tools such as GUI and virtual drive can provide efficient, convenient and reliable interface among to HPC storage systems. We will also show that such an interface is also capable of handling really large volumes of data at proper speed as well as offer relevant responsiveness to user I/O operations.

The discussed solution is based on Seafile, a scalable and reliable sync&share software deployed in PSNC's HPC department servers and storage infrastructure. Seafile has been for years a basis of sync & share services that HPC Department at PSNC offers internally and to academic community in Poland since 2015 through the PIONIER network. While, this service was initially targeted mostly regular users who store their documents, graphics and other typical data sets, it attracted also researchers. These power users started to challenge our systems with 10s of Terabytes of data. However, the real challenge was yet to come.

In autumn 2017, we started a pilot deployment of Seafile for research groups within CoeGSS EU project. They use our synchronisation and sharing solution as an equivalent of the home directory in HPC systems, in order to store, access and exchange results of the simulations rrun in the PSNC's flagship HPC cluster 'Eagle'(1,4 PFlops, #172 at Top500 on Nov 2017).

While the computations are performed using high-performance Lustre filesystem attached through Infiniband to the HPC cluster, the input and output data are transferred, accessed and synchronised over regular network using Seafile. This solution provides ease of access to data sets, as users can interact with their files stored in the HPC system using Web interface, desktop GUI applications and virtual drive solution available accross Windows, Linux and MacOS. Automation of the data flow is also ensured as the data can be selectively synchronised with the HPC storage as they are produced by computing jobs in the cluster. In the same time high performance of storage and retrieval is possible as Seafile scales to many MB/s of transmission througput and seamlessly serves lots of small files.

**Authors:** BRZEZNIAK, Maciej (PSNC Poznan Poland); Mr WADÓWKA, Krzysztof (PSNC Poznan Poland); Mr

POŚPIESZNY, Marcin (PSNC Poznan Poland); Mr BRONA, Piotr (PSNC Poznan Poland); Mr JANUSZEWSKI, Radosław (PSNC Poznan Poland)

**Presenter:** BRZEZNIAK, Maciej (PSNC Poznan Poland)

**Session Classification:** Scalable Storage Backends for Cloud and HPC: Integration