Storage backends for scalable sync&share service based on Seafile

Monday 30 January 2017 17:10 (20 minutes)

Seafile is a scalable and reliable sync&share solution. Its synchronisation engine and data model is based on git concept adapted to dealing with large files and datasets. Seafile synchronises data based on filespace snapshots rather than per-file or per-data object versioning and involves deduplication with Content Defined Chunking algorithm. The architecture and implementation introduces small overheads as the relational database usage is reduced to minimum - only head commit ID and user-library mappings are kept there, while the actual data and meta-data are handled by the storage back-end.

Well-optimised synchronisation engine of Seafile has a potential to put a lot of stress on the storage back-end while serving a large number of I/O operations. In fact it constitutes an interesting killer application for the storage system.

Seafile deployment at PSNC targets a country-wide scale, therefore we expect to deal with large user base as well as millions of files and I/Os to be served on time. While Seafile supports various storage back-ends including filesystem and object storage as well as enables Load-Balancing and High Availability for the synchronisation engine, the decisions on choosing and configuring storage back-end for the planned scale are not trivial.

In our presentation we will overview and summarize the I/O requirements of Seafile server as well as analyse several storage systems in this context, including traditional and clustered filesystems based on Fibre Channel disk arrays as well software defined storage systems based on disk servers and 10Gbit Ethernet.

We will share the results of our analysis and benchmarks performed with Seafile sever as well as draw out general conclusions and lessons learnt on architecting storage back-ends for sync & share services.

Author: BRZEZNIAK, Maciej (PSNC Poznan Poland)

Co-author: Mr WADOWKA, Krzysztof (PSNC Poznan Poland)

Presenter: BRZEZNIAK, Maciej (PSNC Poznan Poland)

Session Classification: Technology