



Contribution ID: 228

Type: oral presentation

## Data preservation for the HERA Experiments @ DESY using dCache technology

*Tuesday, 14 April 2015 17:15 (15 minutes)*

We report on the status of the data preservation project at DESY for the HERA experiments and present the latest design of the storage which is a central element for bit-preservation. The HEP experiments based at the HERA accelerator at DESY collected large and unique datasets during the period 1992 to 2007. In addition, corresponding Monte Carlo simulation datasets were produced, which are significantly larger by volume and still being added to as the final analyses are completed. As part of the ongoing DPHEP data preservation efforts at DESY, these data sets must be transferred into storage systems that guarantee a reliable long term access. At the same time, given that the experiments are still active, easy access to the data must be guaranteed for the coming years. The long term storage system is two-fold: an archive part where the data exists on two tape copies and an online part where the full dataset can be kept available and allows easy access to all HERA data. The archive and online parts are physically separate. The demanding aspect of this data is not only the size of about 1PB but also the large number (about 4 million) of files and the broad range of file sizes from a few KB to a few hundred of GB. To achieve a high level of reliability, we use the dCache distributed storage solution and make use of its replication capabilities and tape interfaces. We describe the dCache installation with tape backend that is used as mass storage together with an newly introduced small files service that allows for the automatic creation of tape friendly container files, containing many single (small) files. From the user's point of view, this is done in a fully transparent way in terms of creation and access to the data.

**Primary authors:** KRUECKER, Dirk (DESY); SCHWANK, Karsten (DESY)

**Co-authors:** LEWENDEL, Birgit (DESY); SOUTH, David (DESY); FUHRMANN, Patrick (DESY)

**Presenter:** SCHWANK, Karsten (DESY)

**Session Classification:** Track 3 Session

**Track Classification:** Track3: Data store and access