

dCache: challenges and opportunities when growing into new communities

Tuesday, 18 September 2012 11:40 (20 minutes)

Description of the work

The presentation will give a summary of experiences gained from working to support non-HEP communities, such as XFEL and LOFAR and how adopting standards has been key in facilitating how dCache may be used by other communities. The reasons how (and why) the HEP community are currently using non-standard protocols are mentioned along with the outlook for the future. The talk will describe challenges when moving from the (relatively) well-organised and structured HEP communities to the more chaotic collaborations. There are challenges both within dCache (e.g., how to authenticate end-users and authorise their activity) and challenges at the boundary of dCache and other services (e.g., facilitating data movement and complex data management). The talk will describe how recent changes in dCache target these problems and how future changes are aimed to solve them. The presentation will also describe how dCache is opening up to support a more community-driven approach; a more inclusive approach will allow more flexible adoption. A new software license, along with spinning off dCache components as separate projects, allows for easier re-use within other projects. Some examples of successes are provided. In addition, reformulating the existing flexibility in dCache as support for plugins provides a low-overhead entry-point for community involvement. The presentation will also include details from the “dCache labs” project. These are experimental new features that the dCache team will be providing as technological previews, without promising that they will be adopted by dCache core; examples include our ongoing investigation with integrating shared file-systems. This is done to allow dCache administrators to provide feedback on features.

Link for further information

<http://www.dcache.org/>

Wider impact of this work

Different communities are now starting to generate HEP-like quantities of data that they now need to store and process, communities that hitherto could fit their experimental data on single server solutions are now faced with a data explosion. Many of the problems from this data explosion are problems that the HEP community has already experienced and found solutions, some of which are encoded within dCache; for example, the ability to split responsibility for digital curation and data access from scientific analysis allows a common approach across different disciplines, with a corresponding economy-of-scale. By adopting existing solutions and, importantly, by adjusting existing solutions so they are adoptable, the effort required for these scientific communities to support their Big Data problems is reduced, allowing them to focus on their scientific programme.

Printable Summary

The dCache project started in 2000 in order to provide a common commodity storage front-end to different tape systems for use within the High-Energy Particle Physics (HEP) community. Since then, the project has morphed into a generic storage system, providing both disk-only and disk-and-tape systems. Although dCache is mostly heavily used by the HEP communities, it is being adopted by ever increasing number of communities beyond. This talk will describe some of the challenges in adapting dCache to support a wider community of users, the recent changes in dCache to accommodate these challenges and our plans to support an increasingly wide user-base. This presentation is aimed at people from non-HEP communities that are looking for solutions to their Big Data problems, administrators who are interested in supporting a wider

range of users with their existing dCache infrastructure, and people generally interested in challenges when dealing with data in large communities.

Primary author: MILLAR, Paul (DESY)

Presenter: MILLAR, Paul (DESY)

Session Classification: New middleware products

Track Classification: Resource Infrastructure services (Peter Solagna: track leader)