

Using the Advanced dCache API (ADA) tool for Big Data processing

Tuesday, 3 November 2020 14:45 (15 minutes)

This demo will go through ADA (Advanced dCache API), a tool for interacting with dCache which is a powerful data storage platform tailored to data intensive applications, offered by SURF. dCache is optimised for processing huge datasets from many terabytes to petabytes, examples of datasets this large often include instrument data from sensors, DNA sequencers, telescopes, and satellites.

Several communities from various scientific domains are using the SURF dCache service to achieve high throughput data analysis. Our SURF dCache service is used, among others, by CERN and LIGO-VIRGO experiments in High energy physics domain, the Lofar radio telescope community in Astronomy and ProjectMine ALS research in Life Sciences. Lately we notice an increasing demand from Earth Observation projects using the SURF dCache service, for example Tropomi S5P and other projects dealing with Sentinel missions data in the Copernicus program.

The growing demand for our SURF dCache service has increased the need to simplify the access and data transfer methods with the dCache storage while enabling easy and secure ways to collaborate on the data. As a result, SURF developed a new tool to enable users access dCache from anywhere. Our new tool is called ADA (Advanced dCache API) and it is based on the dCache API and webdav.

Inspired by the first computer programmer 'Ada Lovelace', our ADA tool enables users to access and process their data on dCache from any platform and with various authentication methods by using industry standard tools. For several years, dCache was mainly accessible by Grid storage clients and protocols (SRM, GridFTP, Xrootd) and using x509 certificate authentication, which was limiting usage to Grid computing experts and from Grid enabled platforms. ADA was developed to unload the burden of dealing with dependencies with the Grid infrastructure and offer a portable solution to explore the storage space.

Although ADA supports various authentication methods (x509, LDAP, OpenIDconnect), this demo will cover our recommended authentication method, macaroons. Macaroons are tokens that can be used to give access to dCache data in a very granular way. This gives data managers autonomy to share their data in dCache with project members and external collaborators at local, national and international level. Finally, we will demonstrate the ADA event-driven features for triggering tasks automatically when data is uploaded or staged from tape to disk as an option for automating workflows in High Throughput Computing applications.

Primary author: DANEZI, Natalie (SURF)

Presenter: DANEZI, Natalie (SURF)

Session Classification: Demos 4