Contribution ID: **352** Type: **presentation**

Advancing throughput of HEP analysis work-flows using caching concepts

Monday, 9 July 2018 14:30 (15 minutes)

High throughput and short turnaround cycles are core requirements for the efficient processing of I/O-intense end-user analyses. Together with the tremendously increasing amount of data to be processed, this leads to enormous challenges for HEP storage systems, networks and the data distribution to end-users. This situation is even compounded by taking into account opportunistic resources without dedicated storage systems as possible extension of traditional HEP computing facilities for end-user analyses.

Enabling data locality via local caches on the processing units is a very promising approach to solve throughput limitations and to ensure short turnaround cycles of end-user analyses. Therefore, two different caching concepts have been studied at the Karlsruhe Insitute of Technology. Both are transparently integrated into the HTCondor batch system in order to avoid job specific adaptations for end-users.

The first concept relies on coordinated caches on SSDs in the worker nodes. Data locality is taken into account by custom developed components around the HTCondor batch system ensuring that jobs are assigned to nodes holding its input data.

The second concept utilizes CEPH as a distributed file system acting as a system-wide cache. In this case no data locality specific adjustments need to be applied to the HTCondor batch system. In combination with developed XRootD caching and data locality plug-ins, this approach is also very well suited to tackle bandwidth limitations on opportunistic resources like HPC centers offering parallel file systems.

In this talk an overview about the utilized technologies, the data locality concepts and the current status of the project will be presented.

Primary author: HEIDECKER, Christoph (KIT - Karlsruhe Institute of Technology (DE))

Co-authors: QUAST, Gunter (KIT - Karlsruhe Institute of Technology (DE)); GIFFELS, Manuel (KIT - Karlsruhe Institute of Technology (DE)); FISCHER, Max (GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)); SCHNEPF, Matthias Jochen (KIT - Karlsruhe Institute of Technology (DE)); KUHN, Eileen (KIT - Karlsruhe Institute of Technology (DE))

Presenter: HEIDECKER, Christoph (KIT - Karlsruhe Institute of Technology (DE))

Session Classification: T4 - Data handling

Track Classification: Track 4 - Data Handling