



Contribution ID: 464

Type: Poster

Boosting Performance of Data-intensive Analysis Workflows with Distributed Coordinated Caching

Data-intensive end-user analyses in High Energy Physics requires high data throughput to reach short turnaround cycles.

This leads to enormous challenges for storage and network infrastructure, especially when facing the tremendously increasing amount of data to be processed during High-Luminosity LHC runs.

Including opportunistic resources with volatile storage systems into the traditional HEP computing facilities makes this situation more complex.

Bringing data close to the computing units is a very promising approach to solve throughput limitations and improve the overall performance.

We focus on coordinated distributed caching, where we coordinate the placement of critical data on distributed caches and match work-flows to the most suitable host in terms of cached files.

The coordination of data allows to efficiently use limited cache volume by reducing redundant data storage on distributed caches.

In addition, workflow coordination optimizes overall processing efficiency by improving data access for data-intensive analysis workflows.

The NaviX coordination service developed at KIT realizes this concept by connecting an XRootD cache proxy server infrastructure with an HTCondor batch system.

The usage of distributed caches on opportunistic resources was tested to enable efficient processing of data-intensive workflows there.

In addition, after successfully running a prototype system, we are building a Throughput-Optimized Analysis-System (TOPAS), where about 600 CPU cores are directly connected to a distributed 1PB cache and 11 NVME SSD 1TB caches.

Our system with coordinate distributed caches enables fast analysis of large amounts of data as required for future HEP experiments.

In this contribution, we provide an overview of the concept and the experience gained in coordinated distributed caching.

Primary authors: QUAST, Gunter (KIT - Karlsruhe Institute of Technology (DE)); GIFFELS, Manuel (KIT - Karlsruhe Institute of Technology (DE)); FISCHER, Max (GSI - Helmholtzzentrum für Schwerionenforschung GmbH (DE)); KUEHN, Eileen (KIT - Karlsruhe Institute of Technology (DE)); SCHNEPF, Matthias Jochen (KIT - Karlsruhe Institute of Technology (DE)); VON CUBE, Ralf Florian (Rheinisch-Westfälische Tech. Hoch. (DE)); SAUTER, Martin Benedikt (KIT - Karlsruhe Institute of Technology (DE)); HEIDECKER, Christoph (KIT - Karlsruhe Institute of Technology (DE))

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

Session Classification: Poster Session

Track Classification: Track 1: Computing Technology for Physics Research