



Contribution ID: 477

Type: oral presentation

## High performance data analysis via coordinated caches

*Tuesday, April 14, 2015 5:45 PM (15 minutes)*

With the second run period of the LHC, high energy physics collaborations will have to face increasing computing infrastructural needs.

Opportunistic resources are expected to absorb many computationally expensive tasks, such as Monte Carlo event simulation.

This leaves dedicated HEP infrastructure with an increased load of analysis tasks that in turn will need to process an increased volume of data.

In addition to storage capacities, a key factor for future computing infrastructure is therefore input bandwidth available per core.

Modern data analysis infrastructure relies on one of two paradigms:

data is kept on dedicated storage and accessed via network or distributed over all compute nodes and accessed locally.

Dedicated storage allows data volume to grow independently of processing capacities, whereas local access allows processing capacities to scale linearly.

However, with the growing data volume and processing requirements, HEP will require both of these features.

For enabling adequate user analyses in the future, the KIT CMS group is merging both paradigms:

High-throughput data is spread over a local disk layer on compute nodes,

while any data is available from an arbitrarily sized background storage.

This concept is implemented as a pool of distributed caches, which are loosely coordinated by a central service.

A Tier 3 prototype cluster is currently being set up for performant user analyses of both local and remote data.

The contribution will discuss the current topology of computing resources available for HEP user analyses.

Based on this, an overview on the KIT CMS analysis cluster design and implementation is presented.

Finally, operational experience in terms of performance and reliability is presented.

**Primary authors:** METZLAFF, Christian (KIT - Karlsruhe Institute of Technology (DE)); KUHN, Eileen (KIT - Karlsruhe Institute of Technology (DE)); FISCHER, Max (KIT - Karlsruhe Institute of Technology (DE))

**Co-authors:** JUNG, Christopher (KIT - Karlsruhe Institute of Technology (DE)); QUAST, Gunter (KIT - Karlsruhe Institute of Technology (DE)); GIFFELS, Manuel (KIT - Karlsruhe Institute of Technology (DE)); HAUTH, Thomas (CERN)

**Presenter:** FISCHER, Max (KIT - Karlsruhe Institute of Technology (DE))

**Session Classification:** Track 8 Session

**Track Classification:** Track8: Performance increase and optimization exploiting hardware features