



Contribution ID: 59

Type: **Talk**

A Workflow for Identifying Jobs with Similar I/O Behavior Utilizing Time Series Analysis

Wednesday, 2 December 2020 14:00 (30 minutes)

One goal of support staff at a data center is to identify inefficient jobs and to improve their efficiency. Therefore, a data center deploys monitoring systems that capture the behavior of the executed jobs. While it is easy to utilize statistics to rank jobs based on the utilization of computing, storage, and network, it is tricky to find patterns in 100.000 jobs, i.e., is there a class of jobs that aren't performing well.

In this talk, a methodology to rank the similarity of all jobs to a reference job based on their temporal IO behavior is described.

A study is conducted to explore the effectivity of the approach which starts from three reference jobs and investigates related jobs.

The data stems from DKRZ's supercomputer Mistral and includes more than 500.000 jobs that have been executed for more than 6 months of operation.

Student?

No

Supervisor name

Supervisor email

Primary author: Dr KUNKEL, Julian (University of Reading)

Presenter: Dr KUNKEL, Julian (University of Reading)

Session Classification: HPC

Track Classification: Storage and IO