Data intensive ATLAS workflows in the Cloud

Tuesday, 11 October 2016 16:30 (15 minutes)

This contribution reports on the feasibility of executing data intensive workflows on Cloud infrastructures. In order to assess this, the metric ETC = Events/Time/Cost is formed, which quantifies the different workflow and infrastructure configurations that are tested against each other.

In these tests ATLAS reconstruction Jobs are run, examining the effects of overcommitting (more parallel processes running than CPU cores available), scheduling (staggered execution) and scaling (number of cores). The desirability of commissioning storage in the cloud is evaluated, in conjunction with a simple analytical model of the system, and correlated with questions about the network bandwidth, caches and what kind of storage to utilise.

In the end a cost/benefit evaluation of different infrastructure configurations and workflows is undertaken, with the goal to find the maximum of the ETC value.

Tertiary Keyword (Optional)

Data processing workflows and frameworks/pipelines

Secondary Keyword (Optional)

Distributed data handling

Primary Keyword (Mandatory)

Cloud technologies

Primary author: RZEHORZ, Gerhard Ferdinand (Georg-August-Universitaet Goettingen (DE))

Co-authors: QUADT, Arnulf (Georg-August-Universitaet Goettingen (DE));   KAWAMURA, Gen (Georg-August-Universitaet Goettingen (DE));   KEEBLE, Oliver (CERN)

Presenter: RZEHORZ, Gerhard Ferdinand (Georg-August-Universitaet Goettingen (DE))

Session Classification: Posters A / Break

Track Classification: Track 4: Data Handling