

ErUM-Data Wuppertal update

T. Harenberg, M. Sandhoff, M. Vogel, C. Zeitnitz



BERGISCHE
UNIVERSITÄT
WUPPERTAL

Containerization of releases and conditions data

- The Wuppertal group is currently active in topic area A: work-package A1 (Containerization of user jobs, Containerization of services)
- Within this context, we have prepared Docker/Singularity images for running production workflows for data and MC: detector simulation, digitization and reconstruction
- The images are prepared by installing a single full release of ATLAS software and a conditions database (SQLite). The images are fully standalone and do not need a network connection
- The conditions database is copied from the central Oracle database with a C++ tool, which connects to an abstraction layer (COOL/CORAL). COOL manages the time variation and versioning, and CORAL manages the actual connections to the storage backend (SQL-free interfaces)

ErUM-Data contributions

- ATLAS distributed computing has shown a lot of interest in these images, and plans to use them to run simulation jobs at HPCs. The images are currently being used at SuperMUC and Theta (ALCF)
- A task force was formed to automate the production of images for detector simulation at HPCs using GitLab CI and PanDA (Production and Distributed Analysis System)
- We presented a poster in CHEP2019 summarizing this work titled “Standalone containers with ATLAS offline software”

<https://indico.cern.ch/event/773049/contributions/3473850/>

- We also submitted an associated paper to CHEP2019 Proceedings, which is being reviewed for publication

<https://cds.cern.ch/record/2712564>

Current development projects in non-event data handling

- ATLAS is currently re-designing its conditions database management framework (CREST).
 - A new abstraction layer based on RESTful client-server interactions
 - Payloads are stored as BLOBs (Binary Large Objects) with a simpler relational database structure in the Oracle backend
- This is the architecture currently used by CMS, and there are plans to converge into a common framework with ATLAS
- In this context there could be an opportunity for developing a common tool for copying from central databases into customized ones (e.g. SQLite), which could be containerized. Unfortunately, this is still not clear and we continue to consult with experts
- At the moment LHC experiments have diverged too much in the way they handle non-event data, and a common tool is not possible
- Another possibility that we are exploring is the dynamic instantiation of database proxy services
- We will document in detail the current and future status of the projects that handle conditions and databases, and offer recommendations for the feasibility of containerization of these services

Other projects: user job monitoring

- There is currently a lot of activity at our institute in the upgrade of the ATLAS validation framework of jobs in preparation for the multi-threaded analysis of data
- Jobs are scanned for anomalies in regular expressions and a report is produced
- Start to develop a generic framework (python) that validates logs generated by user jobs and produce reports
- The framework scans for keywords that point to known errors
- More complex types of analysis could be explored such as anomaly detection via ML techniques (an integration with Elasticsearch will be considered)

