

# DAC21 take-aways and forward look

ESCAPE WP2 Workshop 21st-23rd of March 2022

---

Paul Kramp

March 22, 2022

GSI

## CBM Use Cases

- mCBM ingestion
- CBM simulation
- mCBM reconstruction

## PANDA Use Cases

- Ingestion
- Reconstruction

- 2.3TB of fallback data uploaded to Data Lake from 125 parallel batch jobs
- 2.3 TB of data was uploaded live from simulation jobs
- Large number of files stuck in REPLICATING state

- Goal: have batch jobs read from Data Lake and upload processed data
- Small set of test data was successfully processed and the results were uploaded
- Large scale job unsuccessful due to data in REPLICATING state
- Solution could not be found during the remaining time

# DAC 21 mCBM ingestion

- mCBM data taking delayed until December, therefore data acquisition replayed
- Functions the same way as data coming from detector
- Achieved asynchronous zero-copy injection of mCBM data (100 replicas) into the data lake
- Registering a 4GB file and a replication rule took 30-60 seconds
- Bottleneck: calculation of Adler32 checksums on client side
- Proposed solution: Add support for more modern checksum algorithms

## DAC 21 CBM simulation

- Setup in Docker containers containing all required software
- Initial Monte Carlo data uploaded to Data Lake, verified and followed by extraction
- Simulate transport of particles and upload data files
- Third script to extract some of the transport data and generated digitalisation data
- Scripts would run via cron job every 5 minutes
- Simulations successfully run in prepared docker images
- Protocol related issues led to partial success
- Timeout of voms proxy during automatic tests

- Task completed inside a docker container on Gentoo Linux
- Developed demo environment required update during DAC21
- Tracking Reconstruction Code merged to Development Branch during DAC21
- Reconstruction of 2021 data ran successfully, after CbmRoot updates during DAC21
- Protocol related issues on client side

- Positive experiences with WP2 and DIOS
- Preparatory work required to access data lake from GSI cluster due to network constraints
- In general successful operations during DAC21
- CBM errors due to minor client issues
- PANDA errors due to replication state issues



# Conclusions and Forward look

- Observe progress of Rucio development
- Research and compare alternatives
- Follow decisions on infrastructure from other institutes in collaboration
- More involvement of the experiments in testing and defining the requirements
- Integrate the data lake concept in the user analysis of the FAIR experiments