

The ATLAS Data Carousel Project

Thursday 17 October 2019 14:00 (25 minutes)

The ATLAS Experiment is storing detector and simulation data in raw and derived data formats across more than 150 Grid sites world-wide: currently, in total about 200 PB of disk storage and 250 PB of tape storage is used.

Data have different access characteristics due to various computational workflows. Raw data is only processed about once per year, whereas derived data are accessed continuously by physics researchers. Data can be accessed from a variety of mediums, such as data streamed from remote locations, data cached on local storage using hard disk drives or SSDs, while larger data centers provide the majority of offline storage capability via tape systems. Disk is comparatively more expensive than tape, and even for disks there are different types of drive technologies that vary considerably in price and performance. Slow data access can dramatically increase costs for computation.

The HL-LHC era data storage estimate requirements are several factors bigger than the present forecast of available resources, based on a flat budget assumption. On the computing side, ATLAS Distributed Computing (ADC) was very successful in the last years with HPC and HTC integration and using opportunistic computing resources for the Monte-Carlo production. On the other hand, equivalent opportunistic storage does not exist for HEP experiments. ADC started the “Data Carousel” and “Hot/Cold Storage” projects to increase the usage of less expensive storage, i.e., tape or even commercial cloud storage, so it is not limited to tape technologies exclusively. Data Carousel orchestrates data processing between workload management, data management, and storage services with the bulk data resident on offline storage. The processing is executed by staging and promptly processing a sliding window of inputs onto faster buffer storage, such that only a small percentage of input data are available at any one time. With this project we aim to demonstrate that this is the natural way to dramatically reduce our storage costs. The first phase of the project was started in the fall of 2018 and was related to I/O tests of the sites archiving systems. Now we are at Phase II, which requires a tight integration of the workload and data management systems and more intensive data migration between hot (disk) and cold (tape) storage systems. Additionally, the Data Carousel will study the feasibility to run multiple competing workflows from tape. The project is progressing very well and the results will be used before LHC Run 3. In addition we will present the first results related to our R&D project with Google Cloud Platform for a similar studies.

Speaker release

Yes

Desired slot length

20

Authors: KLIMENTOV, Alexei (Brookhaven National Laboratory (US)); ZHAO, Xin (Brookhaven National Laboratory (US)); BORODIN, Misha (University of Iowa (US)); BARISITS, Martin (CERN); LASSNIG, Mario (CERN); GOLUBKOV, Dmitry (Institute for High Energy Physics of NRC Kurchatov Institute (R)); WALKER, Rodney (Ludwig Maximilians Universitat (DE)); DI GIROLAMO, Alessandro (CERN); ELMSHEUSER, Johannes (Brookhaven National Laboratory (US))

Presenter: KLIMENTOV, Alexei (Brookhaven National Laboratory (US))

Session Classification: Storage and Filesystems

Track Classification: Storage & Filesystems