Contribution ID: **24** Type: **Short Talk**

Methods of Data Popularity Evaluation in the ATLAS Experiment at the LHC

Thursday, 20 May 2021 15:39 (13 minutes)

The ATLAS Experiment at the LHC generates petabytes of data that is distributed among 160 computing sites all over the world and is processed continuously by various central production and user analysis tasks. The popularity of data is typically measured as the number of accesses and plays an important role in resolving data management issues: deleting, replicating, moving between tapes, disks and caches. These data management procedures were still carried out in a semi-manual mode and now we have focused our efforts on automating it, making use of the historical knowledge about existing data management strategies. In this study we describe sources of information about data popularity and demonstrate their consistency. Based on the calculated popularity measurements, various distributions were obtained. Auxiliary information about replication and task processing allowed us to evaluate the correspondence between the number of tasks with popular data executed per site and the number of replicas per site. We also examine the popularity of user analysis data that is much less predictable than in the central production and requires more indicators than just the number of accesses.

Primary authors: GRIGORYEVA, Maria (M.V. Lomonosov Moscow State University (RU)); Dr SCIABÀ, Andrea (CERN); CHUCHUK, Olga (Université Côte d'Azur (FR)); KLIMENTOV, Alexei (Brookhaven National Laboratory (US)); TRETYAKOV, Evgeny (National Research Nuclear University MEPhI (RU)); DI GIROLAMO, Alessandro (CERN); LASSNIG, Mario (CERN); SCHULZ, Markus (CERN); BEERMANN, Thomas (Bergische Universitaet Wuppertal (DE))

Presenter: GRIGORYEVA, Maria (M.V. Lomonosov Moscow State University (RU))

Session Classification: Monitoring

Track Classification: Distributed Computing, Data Management and Facilities