Contribution ID: **78**   Type: **Short Talk**

Analysis of data integrity and storage quality of a distributed storage system

Thursday, 20 May 2021 15:52 (13 minutes)

CERN uses the world's largest scientific computing grid, WLCG, for distributed data storage and processing. Monitoring of the CPU and storage resources is an important and essential element to detect operational issues in its systems, for example in the storage elements, and to ensure their proper and efficient function. The processing of experiment data depends strongly on the data access quality, as well as its integrity and both of these key parameters must be assured for the data lifetime. Given the substantial amount of data, O(200PB), already collected by ALICE and kept at various storage elements around the globe, scanning every single data chunk would be a very expensive process, both in terms of computing resources usage and in terms of execution time. In this paper, we describe a distributed file crawler that addresses these natural limits by periodically extracting and analyzing statistically significant samples of files from storage elements, evaluates the results and is integrated with the existing monitoring solution, MonALISA.

Primary authors: NEGRU, Adrian-Eduard (University Politehnica of Bucharest (RO)); LATCHEZAR, Betev (CERN); GRIGORAS, Costin (CERN); CARABAS, Mihai (University Politehnica of Bucharest (RO)); WEISZ, Sergiu (University Politehnica of Bucharest (RO)); TAPUS, Nicolae (University Politehnica of Bucharest (RO))

Presenter: NEGRU, Adrian-Eduard (University Politehnica of Bucharest (RO))

Session Classification: Monitoring

Track Classification: Distributed Computing, Data Management and Facilities