

Machine Learning analysis of CMS data transfers

Tuesday, 7 March 2017 15:00 (20 minutes)

Tens of Petabytes of collision and simulated data have been collected and distributed across WLCG sites in Run-1 and Run-2 at LHC. A low latency in transfers among dozens of computing centres is crucial to make an efficient use of the computing resources. Despite on average the desired level of throughput has been successfully achieved to serve the LHC physics programs, it is not uncommon to observe transfer latencies caused by a large variety of causes, from file corruptions to site issues, most of which require operator intervention. To improve on this front, in particular, the CMS experiment equipped the PhEDEx dataset replication system with a system to collect the latency data, and a mechanism to categorise and analyse them promptly, matching them to quick and focussed operators intervention. The transfer latencies data has also been the target of Machine Learning techniques - already used in CMS to study and predict the dataset popularity - and preliminary results on the predictability potential of this approach will be presented and discussed.

Primary authors: Prof. BONACORSI, Daniele (University of Bologna); Dr MAGINI, Nicolò (Fermilab (US)); Dr KUZNETSOV, Valentin (Cornell University)

Co-authors: TOMMASO, Diotalevi (University of Bologna); KIPRAS, Kančys (Vilnius University); ZYGIMANTAS, Matonis (University of Vilnius); AURIMAS, Repečka (Vilnius University)

Presenter: Prof. BONACORSI, Daniele (University of Bologna)

Session Classification: Data Management & Big Data

Track Classification: Data Management & Big Data