Contribution ID: **42**   Type: **presentation**

Architecture and prototype of a WLCG data lake for HL-LHC

Thursday 12 July 2018 11:30 (15 minutes)

The computing strategy document for HL-LHC identifies storage as one of the main WLCG challenges in one decade from now. In the naive assumption of applying today's computing model, the ATLAS and CMS experiments will need one order of magnitude more storage resources than what could be realistically provided by the funding agencies at the same cost of today. The evolution of the computing facilities and the way storage will be organized and consolidated will play a key role in how this possible shortage of resources will be addressed. In this contribution we will describe the architecture of a WLCG data lake, intended as a storage service geographically distributed across large data centers connected by fast network with low latency. Will present the experience with our first prototype, showing how the concept, implemented at different scales, can serve different needs, from regional and national consolidation of storage to an international data provisioning service. We will highlight how the system leverages its distributed nature, the economy of scale and different classes of storage to optimise the hardware and operational cost, through a set of policy driven decisions concerning data placement and data retention. We will discuss how the system leverages or interoperates with existing federated storage solutions. We will finally describe the possible data processing models in this environment and present our first benchmarks.

Primary authors: CAMPANA, Simone (CERN); BIRD, Ian (CERN); GIRONE, Maria (CERN); ESPINAL CURULL, Xavier (CERN); SCHOVANCOVA, Jaroslava (CERN); MCCANCE, Gavin (CERN)

Presenter: CAMPANA, Simone (CERN)

Session Classification: T4 - Data handling

Track Classification: Track 4 - Data Handling