Contribution ID: **481**   Type: **Oral**

Federated data storage evolution in HENP: data lakes and beyond

Monday 11 March 2019 19:00 (20 minutes)

Storage have been identified as the main challenge for the future distributed computing infrastructures: Particle Physics (HL-LHC, DUNE, Belle-II), Astrophysics and Cosmology (SKA, LSST). In particular, the High Luminosity LHC (HL-LHC) will begin operations in the year of 2026 with expected data volumes to increase by at least an order of magnitude as compared with the present systems. Extrapolating from existing trends in disk and tape pricing, and assuming flat infrastructure budgets, the implications for data handling for end-user analysis are significant. HENP experiments need to manage data across a variety of mediums based on the types of data and its uses: from tapes (cold storage) to disks and solid state drives (hot storage) to caches (including world wide access data in clouds and "data lakes"). DataLake R&D project aims at exploring an evolution of distributed storage while bearing in mind very high demands of HL-LHC era. Its primary objective is to optimize hardware usage and operational costs of a storage system deployed across distributed centers connected by fat networks and operated as a single service. Such storage would host a large fraction of the data and optimize the cost, eliminating inefficiencies due to fragmentation. In this talk we will highlight current status of the project, its achievements, interconnection with other research activities in this field like WLCG-DOMA and ATLAS-Google DataOcean, and future plans.

Primary authors: ESPINAL, Xavier (CERN);   Mr KIRIANOV, Andrey (Petersburg Nuclear Physics Institut (RU));   SCHOVANCOVA, Jaroslava (CERN);   ZAROCHENTSEV, Andrey (St Petersburg State University (RU))

Presenter: ZAROCHENTSEV, Andrey (St Petersburg State University (RU))

Session Classification: Track 1: Computing Technology for Physics Research

Track Classification: Track 1: Computing Technology for Physics Research