

CERNBox: the data hub for data analysis

Tuesday, 11 October 2016 16:30 (15 minutes)

A new approach to providing scientific computing services is currently investigated at CERN. It combines solid existing components and services (EOS Storage, CERNBox Cloud Sync&Share layer, ROOT Analysis Framework) with rising new technologies (Jupyter Notebooks) to create a unique environment for Interactive Data Science, Scientific Computing and Education Applications.

EOS is the main disk storage system handling LHC data in the 100PB range. CERNBox offers a convenient sync&share layer and it is available everywhere: web, desktop and mobile. The Jupyter Notebook is a web application that allows users to create and share documents that contain live code, equations, visualizations and explanatory text. ROOT is a modular scientific software framework which provides the functionality to deal with big data processing, statistical analysis, visualisation and storage.

The system will be integrated in all major work-flows for scientific computing and with existing scientific data repositories at CERN. File access will be provided using a range of access protocols and tools: physics data analysis applications access CERNBox via xrootd protocol; Jupyter Notebooks interact with the storage via file-system interfaces provided by EOS fuse mounts; Grid jobs use webdav access authenticated with Grid certificates whereas batch jobs may use local Krb5 credentials for authentication. We report on early experience with this technology and applicable use-cases, also in a broader scientific and research context.

Tertiary Keyword (Optional)

Outreach

Primary Keyword (Mandatory)

Cloud technologies

Secondary Keyword (Optional)

Storage systems

Primary author: MOSCICKI, Jakub (CERN)

Presenter: MOSCICKI, Jakub (CERN)

Session Classification: Posters A / Break

Track Classification: Track 8: Security, Policy and Outreach