

Advancements in data management services for distributed e-infrastructures: the eXtreme-DataCloud project

Thursday, 12 July 2018 14:45 (15 minutes)

The development of data management services capable to cope with very large data resources is a key challenge to allow the future e-infrastructures to address the needs of the next generation extreme scale scientific experiments.

To face this challenge, in November 2017 the H2020 “eXtreme DataCloud - XDC” project has been launched. Lasting for 27 months and combining the expertise of 8 large European research organisations, the project aims at developing scalable technologies for federating storage resources and managing data in highly distributed computing environments. The targeted platforms are the current and next generation e-Infrastructures deployed in Europe, such as the European Open Science Cloud (EOSC), the European Grid Infrastructure (EGI), and the Worldwide LHC Computing Grid (WLCG).

The project is use-case driven with a multidisciplinary approach, addressing requirements from research communities belonging to a wide range of scientific domains: High Energy Physics, Astronomy, Photon and Life Science, Medical research.

XDC will implement data management scalable services to address the following high level topics: policy driven data management based on Quality-of-Service, Data Life-cycle management, smart placement of data with caching mechanisms to reduce access latency, meta-data with no predefined schema handling, execution of pre-processing applications during ingestion, data management and protection of sensitive data in distributed e-infrastructures, intelligent data placement based on access patterns.

Experts from the project consortium will work on combining already established data management and orchestration tools to provide a highly scalable solution supporting the computing models of the current and next generation experiments. The XDC products will be based on tools such as ONEDATA, EOS, FTS, Indigo-Orchestrator, Indigo-CDMI server, Dynafed.

This contribution will introduce the project, present the foreseen overall architecture and the developments that are being carried on to implement the requested functionalities.

Primary authors: Dr COSTANTINI, Alessandro (INFN); CESINI, Daniele (Universita e INFN, Bologna (IT)); DON-VITO, Giacinto (INFN-Bari); DUMA, Doina Cristina (INFN - CNAF); MATTHEW, Viljoen; BATTAGLIA, Serena (ECRIN); POIREAU, Vincent (Laboratoire d'Annecy-le-Vieux de Physique des Particules (LAPP)); DELL'AGNELLO, Luca (INFN-CNAF); KEEBLE, Oliver (CERN); LEMRANI, Rachid (CNRS/IN2P3); OHMANN, Christian; Mr MARCO DE LUCAS, Jesus (Instituto de Fisica de Cantabria); DUTKA, Lukasz; FUHRMANN, Patrick (DESY); AGUILAR GOMEZ, Fernando (Universidad de Cantabria (ES))

Presenter: CESINI, Daniele (Universita e INFN, Bologna (IT))

Session Classification: T4 - Data handling

Track Classification: Track 4 - Data Handling