

The eXtreme-DataCloud project

Solutions for data management services in distributed e-infrastructures



D.Cesini¹, A. Costantini¹, G. Donvito², F. Aguilar Gomez³, D.C. Duma¹, P. Fuhrmann⁴, L. Dutka⁵, O. Keeble⁶, L. Dell'Agnello¹, C. Ohmann⁷, S. Battaglia⁷, V. Poireau⁸, R. Lemrani⁹

¹INFN-CNAF, ²INFN-BARI, ³Universidad de Cantabria/IFCA, ⁴DESY, ⁵AGH University of Science and Technology, ⁶CERN, ⁷ECRIN, ⁸CNRS/LAPP, ⁹CNRS/IN2P3

About

The Horizon2020 eXtreme DataCloud project aims at developing scalable technologies for federating storage resources and managing data in highly distributed computing environments, as required by the most demanding, data-intensive research experiments in Europe and worldwide.

XDC is based on existing, production quality tools, that the project enriched with new functionalities and plugins.

Research Communities



Particle Physics

Photon Science

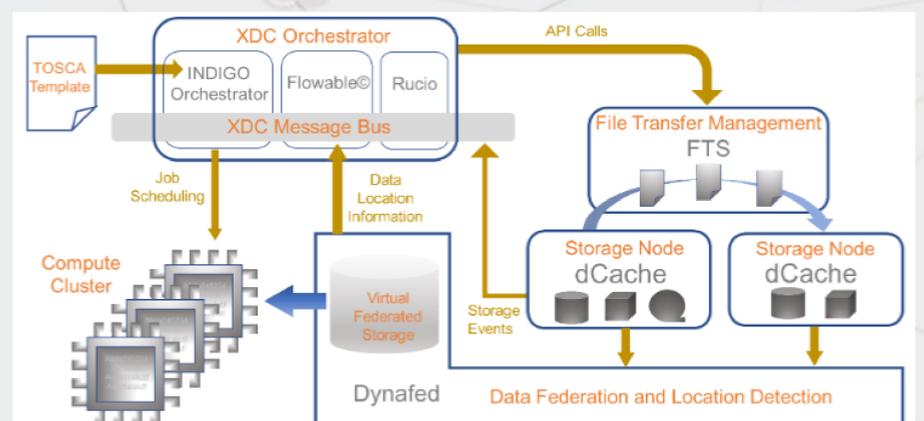
Example use case

The X-ray Free-Electron Laser (XFEL) data management challenges

- Large number of comparatively small files
- Bursty ingestion rates with very high peaks
- Very large amount of data (hundreds of Hz)
- Scalability is crucial, polling across file systems no longer feasible
- Data processing needed both locally and at remote partner institutes
- Automated preprocessing and data transshipment over WAN

XDC Solution

The XFEL use case has been addressed developing the storage events notification system in dCache that implements two event systems, based on Server-Sent Events (SSE) and Kafka. Storage events are used for user-requested, policy-driven or adaptive data movements and trigger actions without polling. They are injected directly into microservices/FaaS resources or into the XDC Orchestration system via message bus. Policies are described by TOSCA templates prepared by users and stored into the XDC Orchestrator that exploits RUCIO and FTS to apply them.



XDC tools used to generate and process storage events notifications for the XFEL use case



XDC High Level Architecture and Tools

XDC release Highlights

The project produced two main releases, Pulsar and Quasar, that can be downloaded from the eXtreme-DataCloud repositories and the INDIGO-DataCloud DockerHub organization.

XDC 1 – Pulsar

- OpenIDConnect support for token-based authentication
- New QoS types integration and support in dCache, FTS, GFAL
- Orchestrator integration
- Support for groups and roles and performance improvements in Onedata
- EOS-dCache integration
- Caching systems instantiation
- Storage events notification in dCache
- EOS caching with Xcache for geographic deployment and EOS external storage adoption



XDC 2 - Quasar

- XDC Message bus
- Full orchestration
- Finalize integration of RUCIO
- ECRIN data ingestion in Onedata
- Secure storage and encryption in Onedata
- Complete caching reference workflows with HTTP based systems



www.extreme-datacloud.eu
@XtremeDataCloud
info@extreme-datacloud.eu

