The XDC project

Alessandro Costantini
INFN
alessandro.costantini<at>extreme-datacloud.eu

Daniele Cesini
INFN - Project Coordinator
daniele.cesini<at>extreme-datacloud.eu
Outline

- XDC: an introduction
- The XDC-1/Pulsar Release: the first year achievements
- Next Steps & Conclusion
The eXtreme DataCloud is a software development and integration project

Develops scalable technologies for federating storage resources and managing data in highly distributed computing environments
- Focus efficient, policy driven and Quality of Service based DM

The targeted platforms are the current and next generation e-Infrastructures deployed in Europe
- European Open Science Cloud (EOSC)
- The e-infrastructures used by the represented communities

Addresses the EINFRA-21-2017 (b)-2: “Computing e-infrastructure with extreme large datasets”
- Deal with heterogeneous datasets
- Bring to TRL8 and include in a unified service catalogue services and prototype at least at TRL6
# XDC Consortium

<table>
<thead>
<tr>
<th>ID</th>
<th>Partner</th>
<th>Country</th>
<th>Represented Community</th>
<th>Tools and system</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>INFN (Lead)</td>
<td>IT</td>
<td>HEP/WLCG</td>
<td>INDIGO-Orchestrator</td>
</tr>
<tr>
<td>2</td>
<td>DESY</td>
<td>DE</td>
<td>Research with Photons (XFEL)</td>
<td>dCache</td>
</tr>
<tr>
<td>3</td>
<td>CERN</td>
<td>CH</td>
<td>HEP/WLCG</td>
<td>EOS, DYNAFED, FTS, RUCIO</td>
</tr>
<tr>
<td>4</td>
<td>AGH</td>
<td>PL</td>
<td></td>
<td>ONEDATA</td>
</tr>
<tr>
<td>5</td>
<td>ECRIN [ERIC]</td>
<td>[ERIC]</td>
<td>Medical data</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>UC</td>
<td>ES</td>
<td>Lifewatch</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>CNRS</td>
<td>FR</td>
<td>Astro [CTA and LSST]</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>EGI.eu</td>
<td>NL</td>
<td>EGI communities</td>
<td></td>
</tr>
</tbody>
</table>

- 8 partners, 7 countries
- 6 research communities represented + EGI
- XDC Total Budget: 3.07Meuros
The Approach

✗ Improve already existing, production quality Data Management services

- By adding missing functionalities requested by research communities
- Based mainly on technologies provided by the partners and by the INDIGO-Datacloud project
- Must be coherently harmonized in the European e-Infrastructures
XDC Approach

- The partners owning/involved in each of the tools are the main developers for that solution in XDC
- We always aim to push back the code in the main development tree on the original projects
  - This widely increase the sustainability of the services
XDC standards and protocols

We always rely on standard and widely adopted protocols for services-to-services and for user-to-services interaction.

<table>
<thead>
<tr>
<th>Functionalities</th>
<th>Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access/transport</td>
<td>http/webdav, gridftp, etc</td>
</tr>
<tr>
<td>Metadata</td>
<td>RDF, JSON</td>
</tr>
<tr>
<td>Orchestrating services</td>
<td>TOSCA, Ansible</td>
</tr>
<tr>
<td>Application sw distribution</td>
<td>Docker</td>
</tr>
<tr>
<td>AAI</td>
<td>OpenIDConnect, X509</td>
</tr>
<tr>
<td>Storage QoS</td>
<td>SNIA CDMI</td>
</tr>
<tr>
<td>Inter process communication</td>
<td>REST APIs, SSE, Message bus</td>
</tr>
</tbody>
</table>

This concretely opens the possibility to interact with external services.

At the infrastructure level we provide few reference implementations.

Others services/initiative are implementing the same functionalities using the same standard/protocols.
Participation to standardization bodies

- **RDA - Research Data Alliance**
  - Participating (DESY) to the Storage Service Definition WG

- **SNIA - Storage Networking Industry Association**
  - XDC presented at the Storage Development Conference 2018
  - DESY is leading the standardization work on QoS for the SNIA CDMI protocol

- **DOMA – Data Organization Management Access in WLCG**
  - Not a proper standardization body, but affects the computing models of a huge community
  - XDC leads the QoS working groups (Paul Millar, DESY)
  - XDC participates to all other working groups
    - Thirdparty copy
    - Caching (Access)
A User Driven Project

- XDC Project Overview
- EOSC
- LifeWatch ERIC
- ECRIN

The long tail of science
The New Functionalities

- Intelligent & Automated Dataset Distribution
  - Orchestration to realize a policy-driven data management
  - Data distribution policies based on Quality of Service (i.e. disks vs tape vs SSD) supporting geographical distributed resources (cross-sites)
  - Data lifecycle management

- Data pre-processing during ingestion

- Metadata management

- Data management based on storage events

- Smart caching
  - Transparent access to remote data without the need of a-priori copy
    - To support dynamic inclusion of diskless sites
    - To improve efficiency in multi-site storage systems and storage federations (i.e. Datalakes)

- Sensitive data handling
  - Secure storage and encryption
General Architecture Definition

- XDC acts at all the e-infrastructure levels
  - Storage systems at sites
  - Federations of storage systems
    - regional and global
  - High level orchestration
  - User experience

The “toolbox” was mapped in those levels to define the general architecture
- Taking into account the user requirements
First XDC Release

Involved tools
- CachingOnDemand
- dCache
- Dynafed
- EOS
- FTS, GFAL
- Onedata
- Onedata
- PaaS Orchestrator plugin
- TOSCA types & templates plugin

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

Key technical highlights

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

26/03/2019
A. Costanini, D.Cesini - XDC Project Overview – EOSC-hub week 2019
First XDC Release

Key technical highlights

- OpenIDConnect support for token based authentication
- New QoS types integration and support in dCache, FTS, GFAL
- Orchestrator integration with other components
- Performance improvements in Onedata
- Support for groups and roles in Onedata
- EOS-dCache integration
- **Caching systems instantiation**
- Storage events notification in dCache
- EOS caching with XCache for geographic deployment
- EOS external storage adoption
- Deployment of Geo-distributed caches
- Network of unmanaged storage for hot data
- On-demand cache resources

Slide from Diego Ciangottini
First XDC Release

Key technical highlights

- OpenIDConnect support for token based authentication
- new QoS types integration and support in dCache, FTS, GFAL
- Orchestrator integration with other components
- Performance improvements in Onedata
- Support for groups and roles in Onedata
- EOS-dCache integration
- Caching systems instantiation
- **Storage events notification in dCache**
- EOS caching with XCache for geographic deployment
- EOS external storage adoption
First XDC Release

Key technical highlights

- OpenIDConnect support for token based authentication
- New QoS types integration and support in dCache, FTS, GFAL
- Orchestrator integration with other components
- Performance improvements in Onedata
- Support for groups and roles in Onedata
- EOS-dCache integration
- Caching systems instantiation
- **Storage events notification in dCache**
- EOS caching with XCache for geographic deployment
- EOS external storage adoption
First XDC Release

Key technical highlights

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

Onedata Transparent POSIX File System
Processing transparently cached data - 37GBytes/sec

A. Costatnini, D.Cesini - XDC Project Overview – EOSC-hub week 2019
Exploitation and sustainability path

- The Service Providers Board (SPB) goal is to link XDC with Service Provider within and outside of the project consortium in order to have a regular dialog among XDC and all the Service Providers
  - First meeting @EOSC-hub week 2019
- Making XDC products available through EGI distribution channels
  - UMD release already delivers dCache, FTS, GFAL, XRootd
- Identification and interaction with the EOSC-HUB Service Providers
  - To be included in the XDC SPB (if not already present)
- Pushing developments in the upstream repositories of all the services to ensure sustainability beyond the project
  - Double path
    - XDC repositories
    - Upstream repository
- XDC solutions included in other projects or actions
  - Dealing with BigData infrastructures
  - Dealing with the DataManagement of extreme scale experiments

26/03/2019
A. Costatnini, D.Cesini - XDC Project Overview – EOSC-hub week 2019
Next Steps

❌ Start working on the second release (codenamed Quasar)
  ➡️ Will complete the implementation of the XDC architecture
  ➡️ Message bus, full orchestration, finalize integration of RUCIO, secure storage in Onedata, finalize the ECRIN use case developments, complete caching reference workflows

❌ Continue the dissemination activities in events and standardization bodies
  ➡️ Finalize the XDC Service Catalogue dissemination material
  ➡️ Organize a training event towards the EOSC

❌ Closer interaction with Service Providers and EOSC-HUB

❌ Scalability testing and verification

26/03/2019
A. Costatnini, D.Cesini - XDC Project Overview – EOSC-hub week 2019
XDC Contacts

Website: www.extreme-datacloud.eu

@XtremeDataCloud on Twitter

Mailing list: info<at>extreme-datacloud.eu
Internal Events

23-25 January 2018
Bologna, Italy

26/03/2019
A. Costatnini, D.Cesini - XDC Project Overview – EOSC-hub week 2019