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

Daniele Cesini[1,*]*, Giacinto* Donvito[2]*, Alessandro* Costantini[1]*, Fernando* Aguilar Gomez[3], *Doina Cristina* Duma[1]*, Patrick* Fuhrmann[4]*, Lukasz* Dutka[5]*, Matthew* Viljolen[6]*, Serena* Battaglia[7]*, Vincent* Poireau[8]*, Luca* Dell'Agnello[1]*, Oliver* Keeble[9]*, Rachid* Lemrani[10]*, Christian* Ohmann[11] and *Jesus* Marco de Lucas[3]

[1]Italian Institute for Nuclear Physics (INFN), INFN-CNAF, Bologna, Italy
[2]Italian Institute for Nuclear Physics (INFN), INFN-BARI, Bari, Italy
[3]Istituto de Fisica University of Cantabria (IFCA), Santander, Spain
[4]Stiftung Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany
[5]AGH University of Science and Technology (AGH-UST), Kracow, Poland
[6]EGI Foundation, Amsterdam, the Netherlands
[7]European Clinical Research Infrastructure Network (ECRIN), Paris, France
[8]Centre National de la Recherche Scientifique (CNRS), IN2P3-LAPP, Annecy, France
[9]European Organization for Nuclear Research (CERN), Genève, Switzerland
[10]Centre National de la Recherche Scientifique (CNRS), IN2P3, Villeurbanne, France
[11]European Clinical Research Infrastructure Network (ECRIN), Dusseldorf, Germany

Abstract. 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 eight 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 is aimed at implementing data management scalable services, combining already established data management and orchestration tools, 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. This contribution introduces the project, presents the foreseen overall architecture and the developments that are being carried on to implement the requested functionalities.

Corresponding author: daniele.cesini@cnaf.infn.it

1 Introduction

The eXtreme DataCloud (XDC) project [1] develops scalable technologies for federating storage resources and managing data in highly distributed computing environments. The provided services are capable of operating at the unprecedented scale required by the most demanding, data intensive, research experiments in Europe and Worldwide. The targeted platforms for the released products are the already existing and the next generation e-Infrastructures deployed in Europe, such as the European Open Science Cloud (EOSC) [2], the European Grid Infrastructure (EGI) [3], the Worldwide LHC Computing Grid (WLCG) [4] and the computing infrastructures that will be funded by the upcoming H2020 EINFRA-12 call. XDC is funded by the H2020 EINFRA-21-2017 Research and Innovation action under the topic Platform-driven e-Infrastructure innovation [5]. It is carried on by a Consortium that brings together technology providers with a proven long-standing experience in software development and large research communities belonging to diverse disciplines: LifeScience, Biodiversity, Clinical Research, Astrophysics, High Energy Physics and Photon Science. XDC started on 1st November 2017 and will run for 27 months until January 2020. The EU contribution for the project is 3.07 million euros. XDC is a use case driven development project and the Consortium has been built as a combination of technology providers, Research Communities and Infrastructure providers. New developments will be tested against real-life applications and use cases. Among the high level requirements collected from the Research Communities, the Consortium identified those considered more general (and hence exploitable by other communities), with the greatest impact on the user base and that can be implemented in a timespan compatible with the project duration and funding.

2 Project Objectives

The XDC project develops open, production quality, interoperable and manageable software that can be easily plugged into the target European e-Infrastructures and adopts state of the art standards in order to ensure interoperability. The building blocks of the high-level architecture foreseen by the project are organized to avoid duplication of development effort. All the interfaces and links are developed exploiting the most advanced techniques for authorization and authentication. Services are scalable to cope with most demanding, extreme scale scientific experiments like those in operation at the Large Hadron Collider at CERN and the Cherenkov Telescope Array (CTA), both of them represented in the project Consortium. The project will enrich already existing data management services by adding missing functionalities as requested by the user communities. The project continues the effort invested by the now ended INDIGO-DataCloud project [6] in the direction of providing friendly, web-based user interfaces and mobile access to the infrastructure data management services. The project builds on the INDIGO-DataCloud achievements in the field of Quality of Services and data lifecycle management, developing smart orchestration tools to realize easily an effective policy driven data management. One of the main objectives of the project is to provide data management solutions for the following use cases:

- Dynamic extension of a computing center to a remote site providing transparent bidirectional access to the data stored in both locations.
- Dynamic inclusion of sites with limited storage capacity in a distributed infrastructure, proving transparent access to the data stored remotely.
- Federation of distributed storage endpoints, i.e. a so-called WLCG DataLake, enabling fast and transparent access to their data without a-priory copy.

These use cases are addressed implementing intelligent, automatic and hierarchical caching mechanisms.

3 User Communities Requirements

In the following paragraphs the user communities represented in XDC are presented together with the list of the main requirements injected into the project development process.

3.1 Lifewatch

LifeWatch [7] is the e-Science and Technology European Infrastructure for Biodiversity and Ecosystem Research that aims to advance science in these disciplines and to address the big environmental challenges as well as to support knowledge-based strategic solutions to environmental preservation. This mission is achieved by providing access to a multitude of datasets, services, tools and computing resources in general, enabling the construction and operation of Virtual Research Environments (VREs). Requirements injected into the project are:

1. Integration of different tools based on the cloud, that allow the management of the data lifecycle, the production of data based on FAIR (+R, Reproducibility) principles
2. Integration of tools that allow the automation of saving the data produced by the sensors stored in distributed resources and the management of the metadata for the integration of the different sources.

3.2 CTA

Very high-energy electromagnetic radiation reaches Earth from a large part of the Cosmos, carrying crucial and unique information about the most energetic phenomena in the Universe. CTA [8] (Cherenkov Telescope Array) will answer many of the persisting questions by enabling the detection of more than thousands sources over the whole sky. CTA builds on the proven technique of detecting gamma ray induced particle cascades in the atmosphere through their Cherenkov radiation, simultaneously imaging each cascade stereoscopically with multiple telescopes, and reconstructing the properties of the primary gamma ray from those images. Requirements injected into the project are:

1. Intelligent and automated dataset distribution: The CTA computing model should be distributed among four data centers and two geographic array sites. The event reconstruction consists of six levels of data, each with specific and complex policies for data distribution: disk vs tape, number of versions, number of replicas per version, etc. Some versions must be automatically erased/removed by the system. Data placement is important to distribute load between the four data centers, not only for storage but also for reconstruction/analysis jobs. The data access rights are also complex.
2. Data ingestion pre-processing: The CTA distributed archive is based on the Open Archival Information System (OAIS) ISO standard. Event data are in files (FITS format) containing all metadata. Metadata need to be extracted from the ingested files, with an automatic filling of the metadata database. Metadata will be used to further query of archive. The system should be able to manage replicas, tapes, disks, etc, with data from low-level to high-level.

3.3 ECRIN

ECRIN (European Clinical Research Infrastructure Network) [9] is a not-for-profit intergovernmental organisation that provides support for the development and implementation of multinational clinical research projects in Europe. These are mostly investigator initiated (rather than industry sponsored) clinical trials, run by non-commercial Clinical Trials Units (CTUs) based within universities or hospitals, though ECRIN does also support trials initiated by biotech and medical device small and medium enterprises (SMEs). As a result more and more researchers are making such material (generically, clinical trial data objects) available for sharing with others. The datasets are rarely freely available - instead a variety of access mechanisms (e.g. individual request and review, membership of pre-authorised groups, or web based self-attestation) are used in combination with different access types (e.g. download versus in-situ perusal). Furthermore, the various data objects are stored in a wide variety of different locations making the discoverability of such data an important challenge to address.

Requirements injected into the project are:
1. Availability of metadata managements tools, easy to use, web based that allow researchers to provide the necessary metadata themselves (and therefore most accurately), aggregating that metadata centrally
2. Availability of tools to harvest and map the metadata from existing data repositories, using available APIs and/or data mining techniques
3. Availability of a generic metadata schema that can not only describe the data objects themselves, but also link them to the source trials and provide information on how the data objects can be accessed (ECRIN has recently proposed such a schema, created by extending DataCite).

3.4 WLCG

The WLCG (Worldwide LHC Computing Grid) [4] is both a worldwide collaboration that counts thousands of researchers and a research distributed e-infrastructure shared by the LHC (Large Hadron Collider) experiments at CERN. WLCG is, in fact, composed by at least four main distinct communities, one per experiment: ALICE, ATLAS, CMS and LHCb. WLCG was historically built for the 4 LHC experiments but it is now shared, in several sites, with many other High Energy Physics (and astrophysics) Virtual Organizations, at national and international level. This created a very variegated and heterogeneous ecosystem of HEP user communities injecting requirements into the same infrastructure. Thousands of researchers exploit the infrastructure on a daily basis to run their applications workflows. Due to the advent of the High Luminosity (HI-LUMI) LHC, expected to enter into production in 2025, has been estimated that the infrastructure and the provided services will need to handle more than 1.5 Exabyte. This extreme scale poses several challenges, technical and economical, on the data management infrastructure. In fact, the storage resources counts roughly for the 50% of the total infrastructure cost. To support these new computing models, improved infrastructure-level data management services would greatly help in lowering the effort to maintain the experiment frameworks. Requirements injected into the project: XDC has to provide smart caches mechanisms to be added to the storage systems currently deployed in the WLCG infrastructures. These caching mechanisms will be the basis to realize a distributed storage system deployed across centers connected by fat networks (10Tb in 2025) and operated as a single service. These so-called "DataLakes" will simplify the dynamic extension of sites to remote locations and allowing for less datasets distribution. The inclusion into the infrastructure of non-standard configuration of sites (i.e. diskless, only tape, HPC farm, etc) will be favoured

by the availability of such systems. The increased usage of caches will allow less disk usage and a reduced a-priori data transfers among the distributed sites.

3.5 EXFEL

The European XFEL [10] is a X-ray Light Source, located in Hamburg, Germany. It is installed mainly in underground tunnels, which can be accessed on three different sites. The 3.4 kilometre-long facility will run from the DESY campus in Hamburg to the town of Schenefeld in Schleswig-Holstein. At the research campus in Schenefeld, teams of scientists from all over the World will carry out experiments using the X-ray flashes. Using the X-ray flashes of the European XFEL, scientists will be able to map the atomic details of viruses, decipher the molecular composition of cells, take three-dimensional images of the nano-world, film chemical reactions, and study processes such as those occurring deep inside planets. The expectation is to make derived data (images) available for further processing at the Kurchatov Institute NRC as fast as possible after the raw data has been calibrated at DESY. As the European XFEL started in Sep 2017, this process is not yet in place, so the XDC work will be an integral part of setting up the data analysis chain for the collaboration with the Moscow Kurchatov Institute NRC and the XFEL central facilities.

4 XDC Overall architecture and related components

The XDC project aims at providing advanced data management capabilities that require the interaction among several components and services. Those capabilities include but are not limited to Quality-of-Services management, pre-processing at ingestion and automated data transfers. Therefore, a global orchestration layer is needed to take care of the execution of those complex workflows. Figure 1 depicts the high-level architecture of the XDC project by describing the components, their roles and the related connections. The following sections describe the main areas of development for the project.
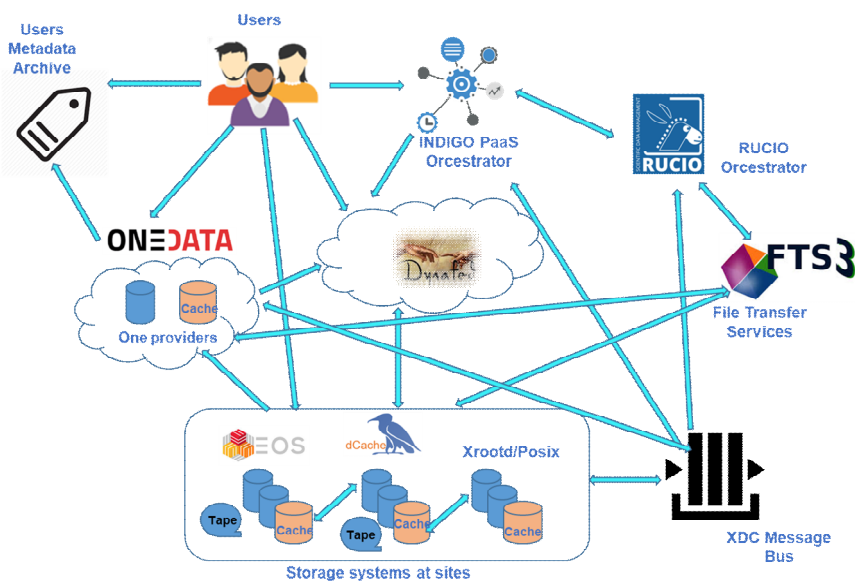

Figure 1 - The XDC high-level architecture and relationships among components

4.1 XDC Orchestration system

In XDC, the global orchestration layer covers two essential aspects: i) The overall control, steering and bookkeeping including the connection to compute resources, ii) The orchestration of the data management activities like data transfers, and data federation. Consequently, it was decided to split the responsibilities between two different components: the INDIGO Orchestrator [11] and Rucio [12]. The INDIGO PaaS Orchestrator, the system wide orchestration engine, is a component of the PaaS layer that allows to instantiate resources on Cloud Management Frameworks (like OpenStack and OpenNebula) and Mesos clusters. It takes the deployment requests, expressed through templates written in TOSCA YAML Profile 1.0, and deploys them on the best cloud site available. The Rucio project, the data management orchestration subsystem, is the new version of ATLAS Distributed Data Management (DDM) service, allowing the experiment to manage the large volumes of data, both taken by the detector as well as generated or derived, in the ATLAS distributed computing system. Rucio is used to manage accounts, files, datasets and distributed storage systems. FTS [13] is exploited by Rucio to execute the data movements. The two components, the PaaS Orchestrator and Rucio, provide different capabilities and can complement each other to offer a full set of features to meet the XDC requirements. Rucio is being integrated in the INDIGO Orchestrator as a plugin to be used to steer the data movement. The Orchestrator will be the main entry point for the user requests and, in particular, it will also interact directly with the storage backends (i.e. dCache[14], EOS[15], Onedata [16]) in order to get information about the data availability needed to trigger the right processing flow. The Dynamic Federations system (Dynafed) [17] is used to have a very fast dynamic namespace exposed via HTTP and used by the other components to gather information for the data movements and policy creation. Components used in XDC communicate with the Indigo Orchestrator by means of a message bus (which is under development), allowing the orchestrator itself to subscribe for storage events related to data transfers and data access. For what concern authentication, all the components involved in the XDC Orchestration system are being extended in order to support tokens based on the OpenID-Connect protocol.

4.2 XDC QualityOfService implementation

The idea to provide scientific communities with the ability to specify a particular Quality of Service when storing data, e.g. the maximum access latency or minimum retention policy, was introduced within the INDIGO-DataCloud project that carried on three interlinked activities: the definition of a common vocabulary within the framework of the Research Data Alliance (RDA), the rendering of those attributes by a standard protocol or API and a reference implementation of the selected standard. In INDIGO, the network protocol and the implementation was based on an extension of the SNIA [18] Cloud Data Management Protocol CDMI [19]. The XDC QoS-related developments are a continuation of the INDIGO work to consistently compliment all data related activities. In other words, whenever storage space is requested, either manually by a user or programmatically, the quality of that space can be negotiated between the requesting entity and the storage provider essential.

4.3 Caching within XDC

XDC is building a hierarchy of components whose goal is to maximise the accessibility of data to clients while minimising global infrastructure costs. The architecture considers a set of multisite storage systems, potentially accessed through caches, which are aggregated

globally through federations. While large, multi-site storage systems may hold the majority of a community custodial data, XDC does not foresee that they will necessarily host all the compute capacity. In particular, CPU-only resource centres or cloud procurements must be supported ensuring as efficient as possible wide area data access. Such resource centres may access custodial data through a standalone proxy cache. The simplest cache envisaged by XDC, a kind of "minimum viable product", would have the following characteristics: i) read-only operations, ii) fetch data on miss with service credentials iii) data can be chunked or full-file, iv) cache residency management features, evicting data when necessary, v) HTTP frontend with group-level authorisation. Beyond this scenario, certain extensions are investigated: write support, synchronisation and enforcement of ACLs, namespace synchronisation (allowing discovery of non-resident data), call-out to FTS for data transport, QoS support, integration with storage notifications. In the XDC caching architecture the EOS and dCache storage systems are integrated, each of them operating in a caching mode for the other, staging data in when a cache miss occurs. Dynafed acts as data and cache federator and all the functionalities needed for linking it with other components are under development. Standalone HTTP caches are built from existing web technology, such as NGINX [20], modified for horizontal scalability and relevant AAI support.

4.4 XDC developments for the Onedata storage federator

Onedata is a data management system providing easy, scalable and high-performance access and management of distributed storage resources. It was developed with the contribution of the INDIGO-DataCloud project to support a wide range of use cases. Functionalities for end users are available via the Onezone that provides single-sign on authentication and authorization for users that can create access tokens to perform data access management via the Web browser, REST APIs or using Onedata POSIX virtual filesystem. Onezone can federate multiple storage sites by deploying Oneprovider services on top of the actual storage resources provisioned by the sites. For the purpose of job scheduling and orchestration, Onedata is being extended to provide notification via the XDC message bus. This will allow the Orchestrator to react to changes in the overall system state (e.g. a new file in a specific directory or space, data distribution changes initiated by manual transfers, cache invalidation or on-the-fly block transfers). On the data access layer, Onedata will provide a WebDAV storage interface. This will enable integration of other HTTP transfer based components such as FTS or EOS. Furthermore, Onezone will allow for semi-automated creation of data discovery portals, based on metadata stored in the federated Oneprovider instances and on a centralized ElasticSearch engine indexing the metadata. In XDC Onedata will also improve its already existing caching capabilities with machine learning techniques used on detailed I/O trace logs generated while running actual user applications. Furthermore, Onedata will provide a federated level encryption key management service, allowing users to securely upload symmetric encryption keys (e.g. AES-256).

5 Conclusion

In the present contribution, the XDC objectives have been discussed and presented. Those objectives are the real driver of the project and derive directly from the several scientific communities represented into the Consortium. The overall architecture of the XDC developments has been presented describing its components, which are already well established, open source solutions that XDC is enriching with new functionalities. The project is also interacting with other European and international initiatives. As an example,

XDC is collaborating with the Designing and Enabling E-Infrastructures for intensive processing in Hybrid Data Clouds (DEEP-Hybrid-DataCloud) [21] project aimed at promoting the integration of specialized hardware under a Hybrid Cloud platform, and targeting the evolution of the corresponding Cloud services supporting these intensive computing techniques to production level. Both projects (XDC and DEEP) have the common objective to open new possibilities to scientific research communities in Europe by supporting the evolution of e-Infrastructure services for Exascale computing. Those services are expected to become a reliable part of the final solutions for the research communities available in the European Open Science Cloud Service Catalogue.

References

1. XDC project, "XDC" [web page], www.extreme-datacloud.eu [accessed 2019-03-13]
2. EC EOSC, "EOSC" [web page], https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud, [accessed 2019-03-13]
3. EGI Foundation, "EGI" [web page], https://www.egi.eu/, [accessed 2019-03-13]
4. WLCG infrastructure, "WLCG" [web page], wlcg.web.cern.ch/, [accessed 2019-03-13]
5. EC H2020 EINFRA21 call, "EINFRA-21"[web site], http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/einfra-21-2017.html, [accessed 2019-03-13]
6. D. Salomoni et al., "INDIGO-DataCloud: a Platform to Facilitate Seamless Access to E-Infrastructures" in Journal of Grid Computing, vol. 16, number 3, pp. 381–408, 2018
7. Lifewatch ERIC, "LifeWatch" [web site], https://www.lifewatch.eu, [accessed 2019-03-13]
8. CTA Experiment, "CTA" [web site], https://www.cta-observatory.org, [accessed 2019-03-13]
9. ECRIN ERIC, "ECRIN" [web site], https://www.ecrin.org, [accessed 2019-03-13]
10. XFEL Organization, "XFEL" [web site], https://www.xfel.eu, [accessed 2019-03-13]
11. INDIGO Paas Orchestrator, "Orchestrator" [software], version 2.1.1-FINAL, 2019. Available from https://github.com/indigo-dc/orchestrator, [accessed 2019-03-13]
12. Barisits et al., "The ATLAS Data Management System Rucio: Supporting LHC Run-2 and beyond", ACAT, Seattle, 2017
13. CERN FTS, "FTS" [software], version 3.9.0, 2019. Available from https://github.com/indigo-dc/fts-rest, [accessed 2019-03-13]
14. dCache storage, "dCache" [software], version vXDC-1,2019. Available from https://github.com/indigo-dc/dcache/tree/xdc, [accessed 2019-03-13]
15. EOS storage, "EOS" [software], version v4.4, 2019. Available from https://github.com/indigo-dc/eos, [accessed 2019-03-13]
16. onedata.org, "Onedata" [software], version v18.02.0-rc13, 2019. Available from https://github.com/indigo-dc/onedata, [accessed 2019-03-13]
17. Dynafed federator, "DYNAFED" [software], version v1.3.3, 2019. Available from lcgdm.web.cern.ch/dynafed-dynamic-federation-project, [accessed 2019-03-13]
18. SNIA organization, "SNIA" [web site], www.snia.org, [accessed 2019-03-13]
19. CDMI specification, "CDMI" [document], http://www.snia.org/sites/default/files/CDMI_Spec_v1.1.1.pdf, [accessed 2019-03-13]
20. NGINX web server, "NGINX" [web site], https://www.nginx.com/, [accessed 2019-03-13]
21. DEEP-Hybrid-DataCloud project, "DEEP-HDC" [web site], https://deep-hybrid-datacloud.eu/, [accessed 2019-03-13]