The XDC project

eXtreme DataCloud

Data Management for extreme scale computing

Giacinto DONVITO
XDC – Technical Coordinator
INFN
donvito@infn.it

eXtreme DataCloud is co-funded by the Horizon2020 Framework Program – Grant Agreement 777367

Outline

- XDC introduction and objectives
- The XDC-1/Pulsar Release: the first year achievements
- Next Steps && Conclusion

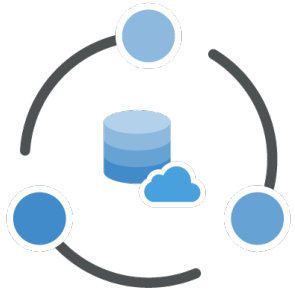
XDC

- 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 on 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

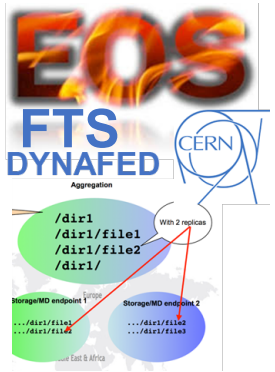
The Approach

- **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

INDIGO PaaS Orchestrator



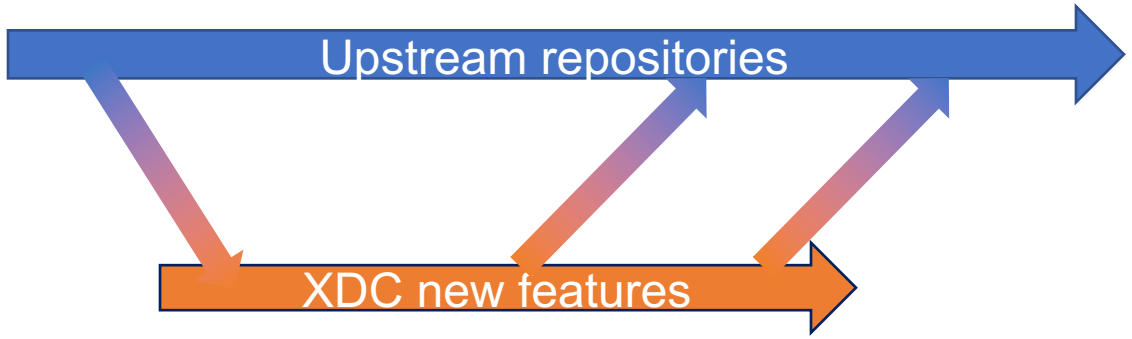
INDIGO CDMI Server



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

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

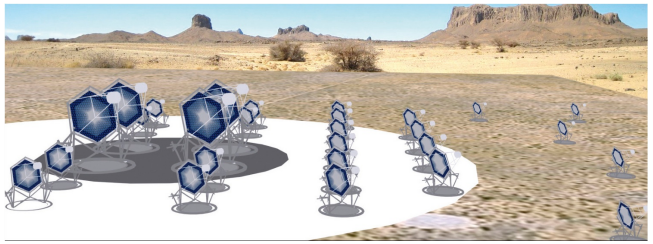
- This concretely opens the possibility to **interact with external services**
  - At the infrastructure level we provide few **reference implementations**
  - **others** services/initiative are **implementing** the same functionalities **using same** standard/**protocols**

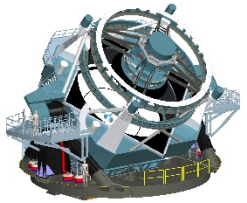
A User Driven Project

eXtreme DataCloud

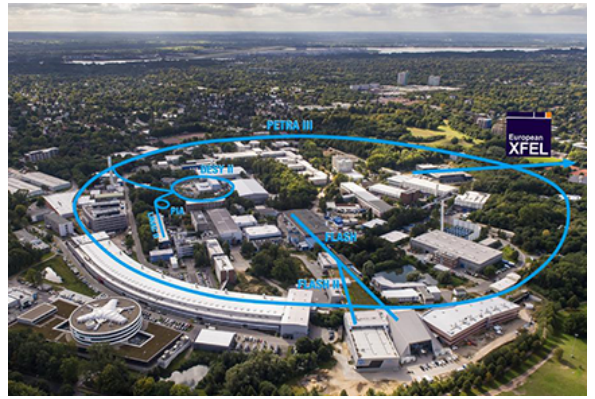
cta cherenkov telescope array

LSST

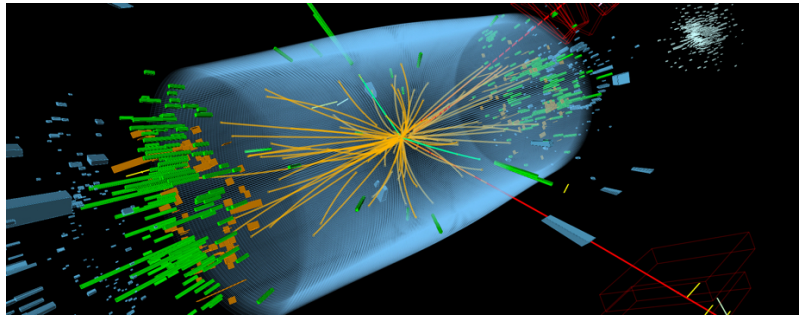




European XFEL

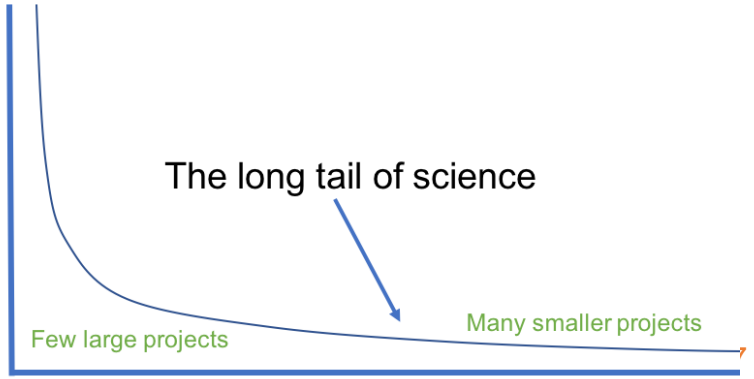


WLCG
Worldwide LHC Computing Grid



LifeWatch ERIC

ECRIN
EUROPEAN CLINICAL RESEARCH INFRASTRUCTURE NETWORK



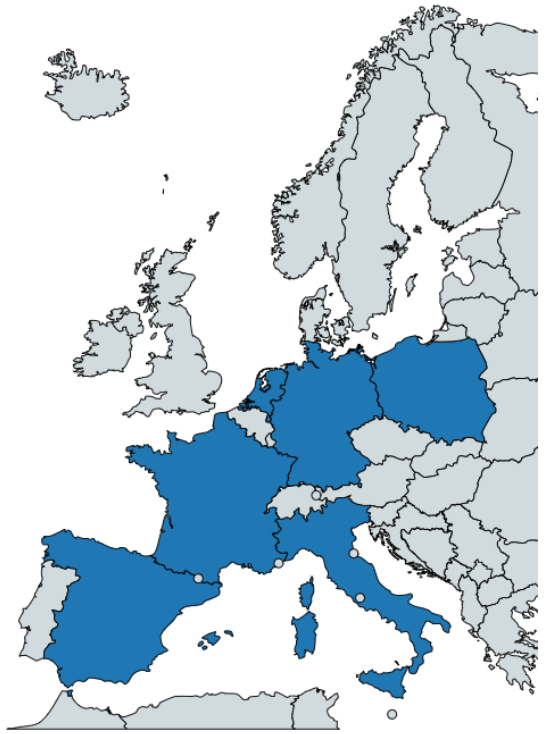
XDC Topics

- 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

XDC Consortium

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

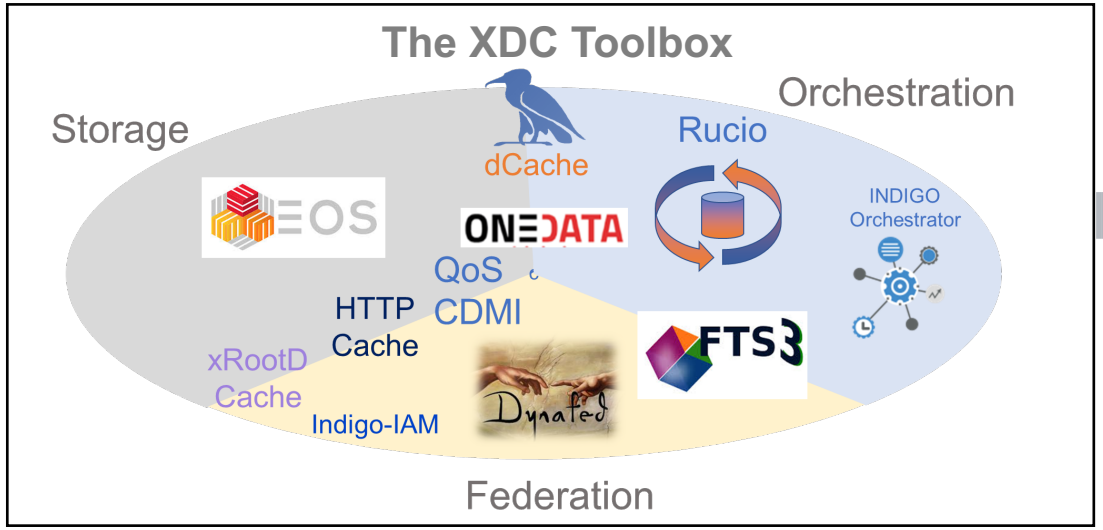
- 8 partners, 7 countries
- 6 research communities represented + EGI
- XDC Total Budget: 3.07Meuros

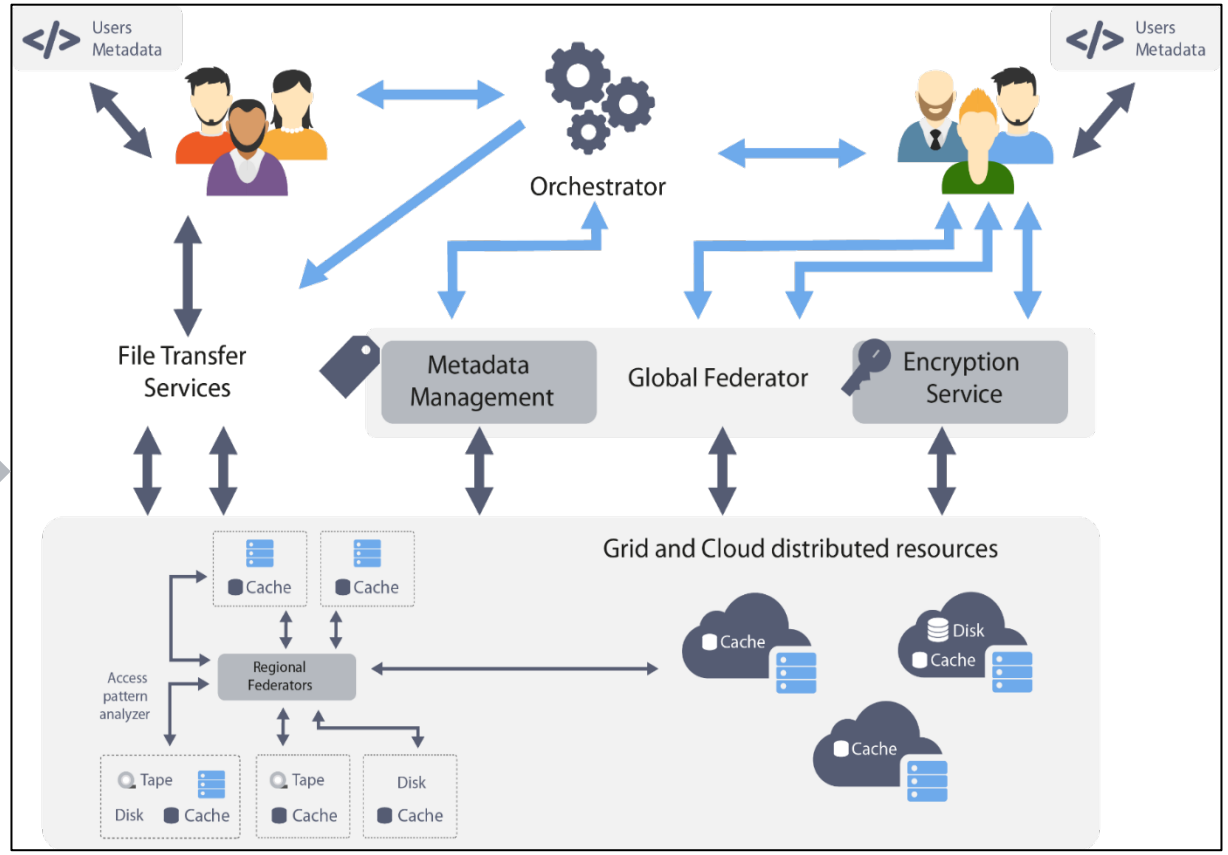


General Architecture Definition

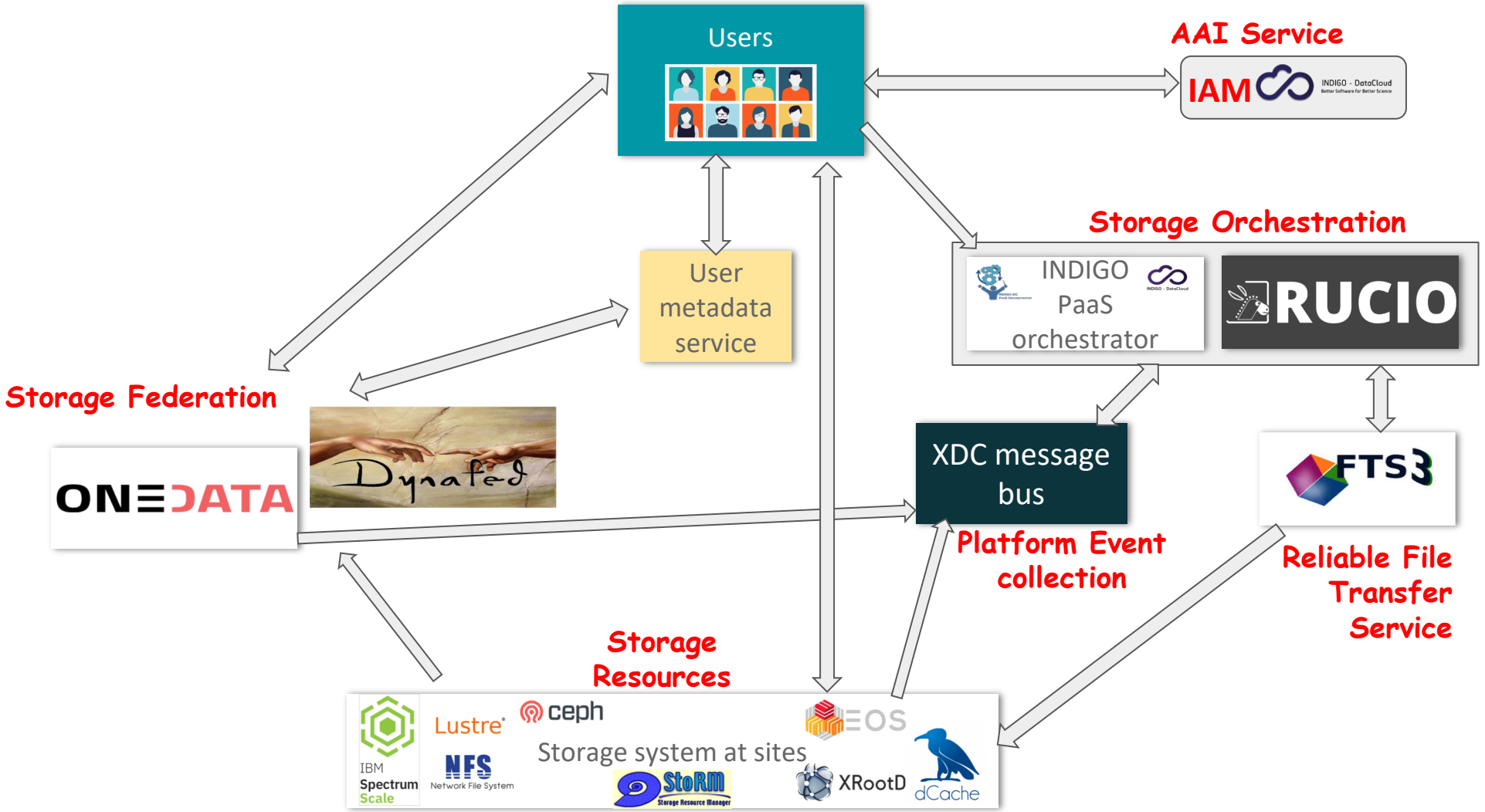
- XDC acts at all the e-infrastructure levels
  - Storage systems at sites
  - Federations of storage systems
    - regional and global
  - High level orchestration
  - User experience

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





XDC General Architecture



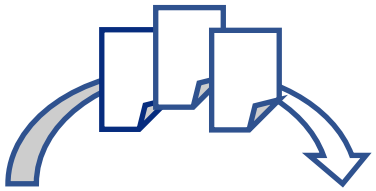
Connection to external Entities

- DEEP-DataCloud
  - Connection to compute workloads
- DOMA (Data Organization, Management, Access)
  - DOMA "Access"
  - DOMA "QoS" (XDC Workgroup Leader)
  - DOMA "3rd Party Copy"
- ESCAPE – WP2 (XDC Task Leader)
  - Deployment of a European Data Lake
- RDA
  - Defining QoS vocabulary
- SNIA
  - Rendering CDMI to reflect QoS
  - XDC introduced at 2018 SNIA conference

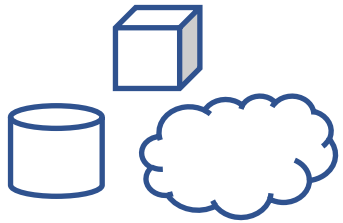
XDC Components

The Components

- Orchestration and Federation Components
  - XDC Orchestrator
    - INDIGO PaaS Orchestrator
    - Flowable © (BPM)
    - Rucio Data Management System



- Data Transfer and Data Federation technologies
  - FTS, File Transfer Service,
  - Dynafed, Data Federator, Onedata



- Storage Systems
  - dCache
  - EOS
  - StoRM
  - Onedata

XDC Orchestration Components

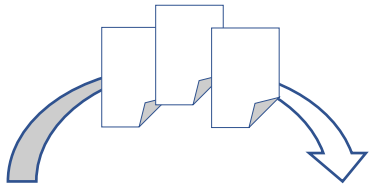
- INDIGO PaaS Orchestrator
  - Based on INDIGO-DataCloud developments.
  - Allows to coordinate complex deployments on hybrid clouds featuring advanced scheduling and federation capabilities
  - Orchestrates compute resources and provides data-aware scheduling of jobs through data placement plugins (XDC extensions)
  - Integrates with Rucio for data location and transfer orchestration (XDC developments)
  - Operates with an professional BPM system. (Flowable)

- Flowable © (BPM)
  - Provides a workflow and Business Process Management (BPM) platform for developers, system admins and business users

- Rucio
  - Originally LHC ALTAS data management tools.
  - Recently adopted by a growing number of other communities.
  - Already provides interfaces to most XDC components

XDC Transport Components

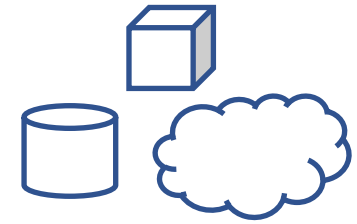
The Components



- FTS, File Transfer Service
  - WLCG data transfer workhorse.
  - Transfers around 1 Exabytes of WLCG data per year between hundreds of storage sites around the word.
  - Performs request queueing and network shaping.
  - Can be used as “micro service” or with GUI (WebFTS).
  - Support X509 and token based authentication for endpoints.
- Dynafed, Data Federator
  - Federates storage endpoints to a single root namespace.
  - Supported Protocols: http/WebDAV, S3.
  - Performs metadata prefetching.
  - Provides location meta data to high level services.

XDC Storage Components

The Components



- dCache
  - Open Source Storage system provided by DESY, Fermilab and NDGF.
  - Handling 150 PBytes at more than 60 big data centers, including 7 WLCG Tier 1 centers.
  - Supports industry standard data access and security protocols on top of a geo-aware multi tier storage stack.
- EOS
  - Scalable storage running at CERN and elsewhere.
  - Geo-aware management of hundred of PBs.
  - HTTP interface.
- StoRM
  - Provided by INFN/CNAF
  - Engine providing multiple data transport and control protocols on top of GPFS and Lustre.

XDC Storage Components

- Onedata
  - Unified Data management platform
- INDIGO CDMI Reference Implementation
  - INDIGO re-implementation of the SNIA CDMI reference implementation, now hosted by SNIA.
  - Provided the CDMI protocol engine and forwards the requests to a plug-in system.
  - Provides plug-ins for a REST protocol dialect as well as for CEPH and GPFS.
- XCache
  - Read-only, block-level data cache
  - Deployed close to CPU to hide latency and reduce WAN traffic
  - HTTP interface

First XDC Release

Involved tools

- CachingOnDemand
- dCache
- Dynafed
- EOS
- FTS,GFAL
- 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

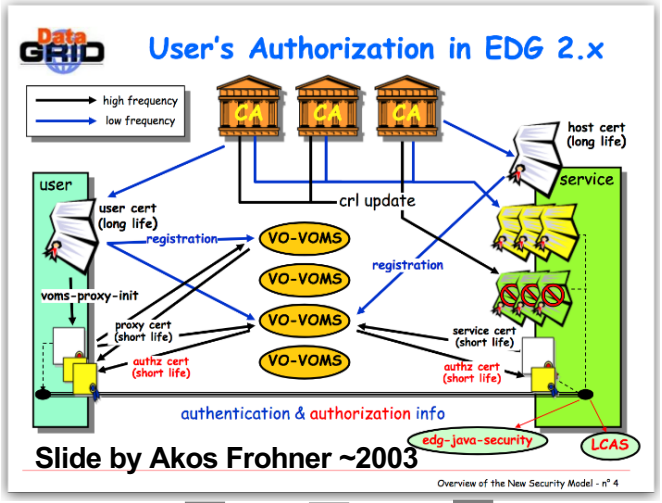
XDC-1/Pulsar

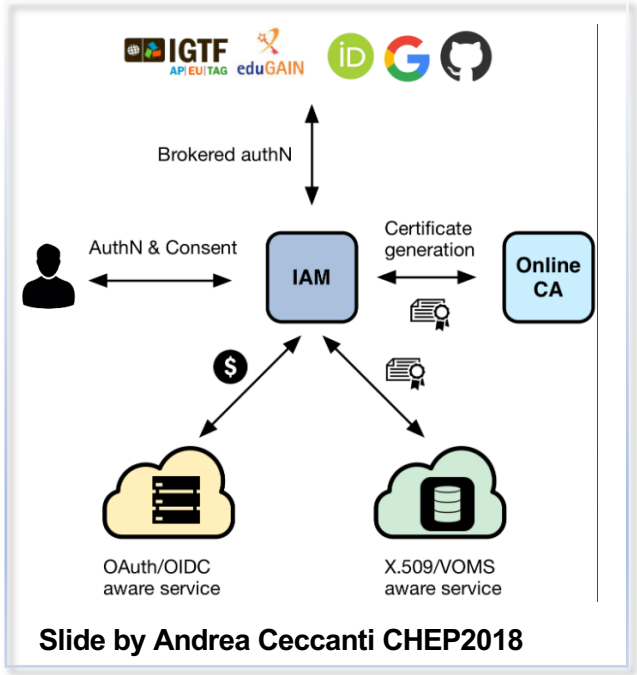
https://releases.extreme-datacloud.eu/en/latest/releases/pulsar/index.html

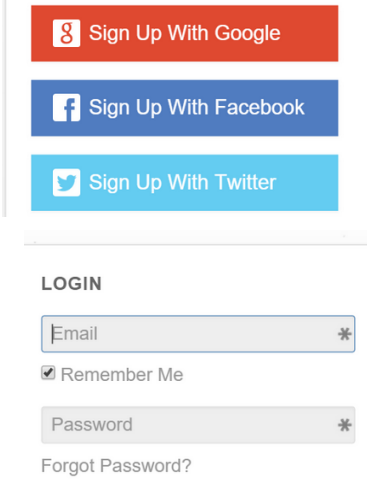
First XDC Release

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



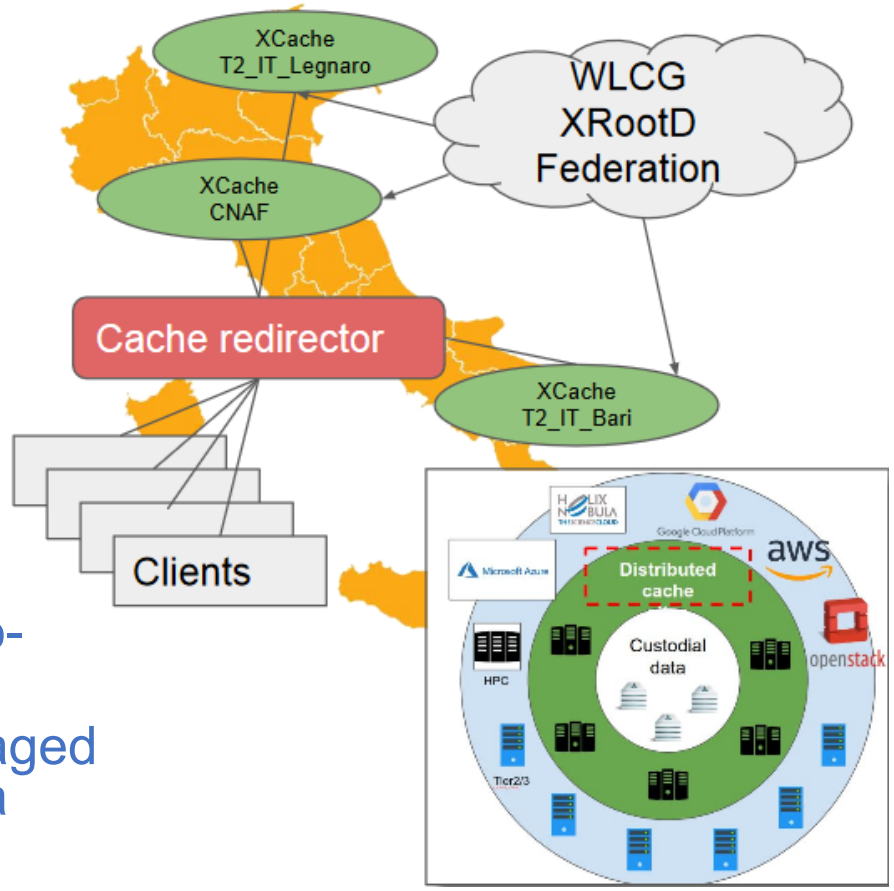




First XDC Release

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



Slide © Diego Ciangottini

- Deployment of Geo-distributed caches
- Network of unmanaged storage for hot data
- On-demand cache resources

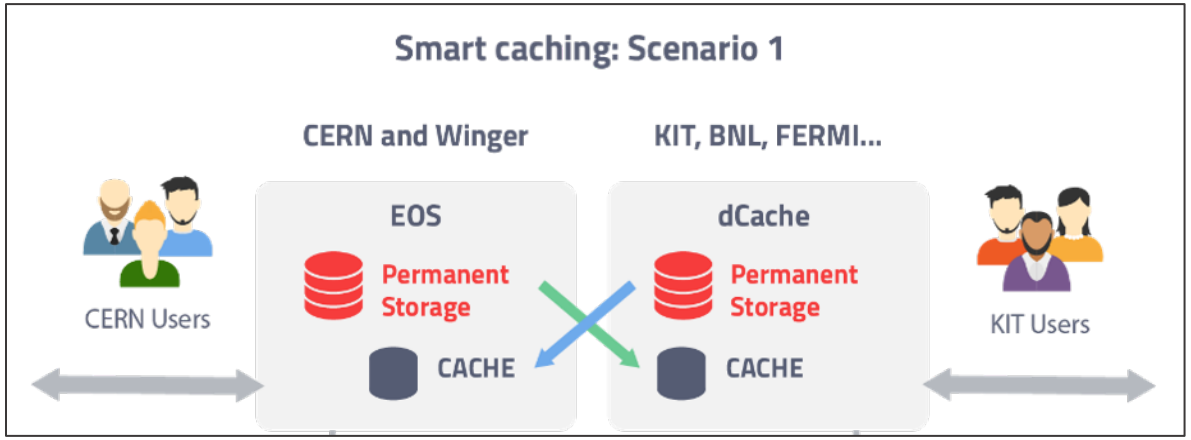
Based on xRootD/xCache

See D.Ciangottini talk on "Integration of the Italian cache federation within CMS computing model": https://indico4.twgrid.org/indico/event/8/session/23/contribution/45

First XDC Release

Key technical highlights

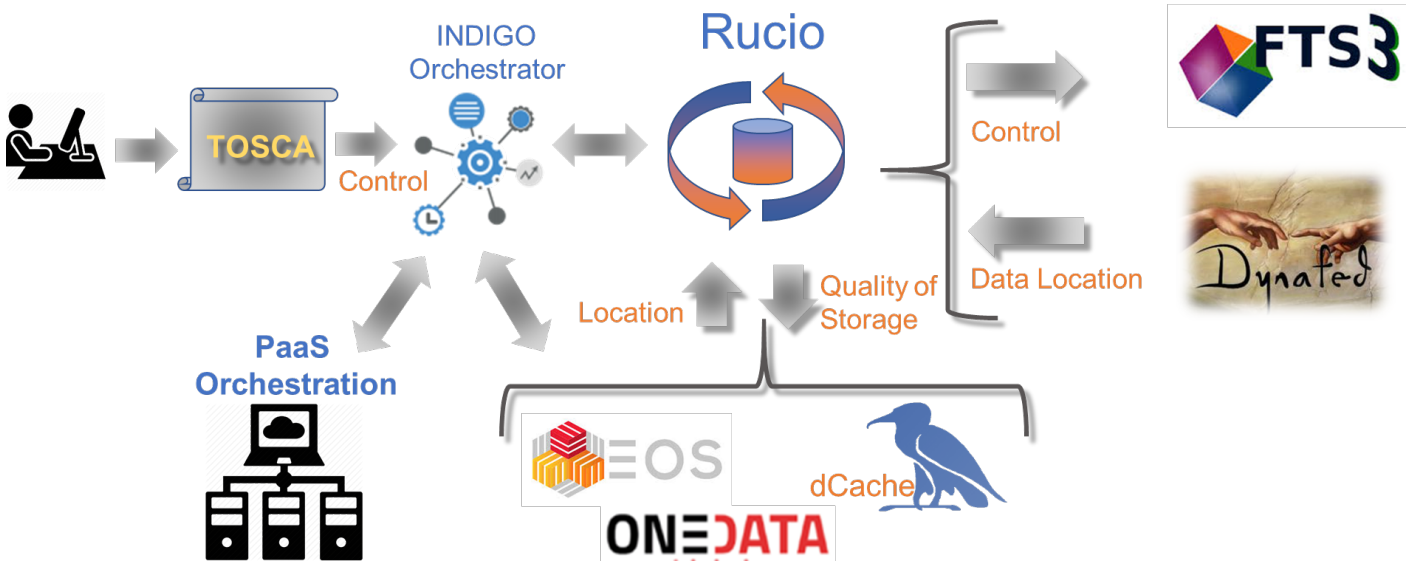
- OpenIDConnect support for token based authentication
- new QoS types integration and support in dCache, FTS, GFAL
- Orchestrator integration with other components
- Performance improvements in Onedata
- Support for groups and roles in Onedata
- EOS-dCache integration
- Caching systems instantiation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



First XDC Release

Key technical highlights

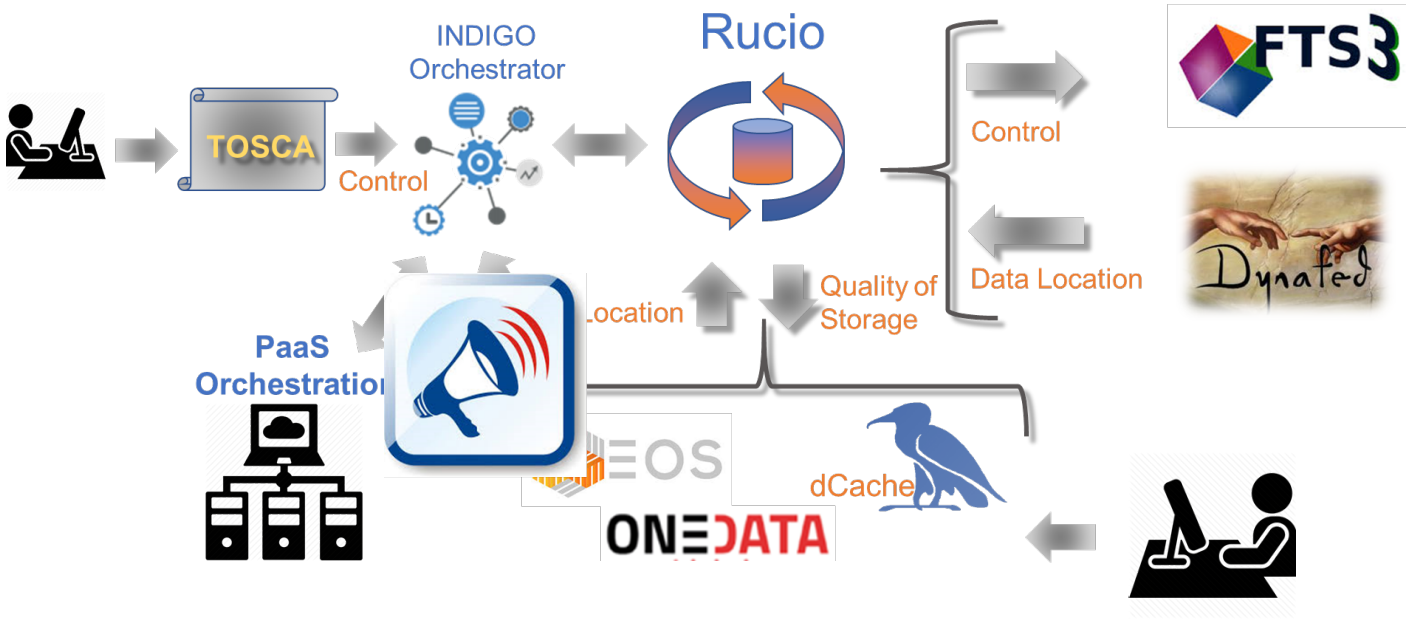
- OpenIDConnect support for token based authentication
- new QoS types integration and support in dCache, FTS, GFAL
- Orchestrator integration with other components
- Performance improvements in Onedata
- Support for groups and roles in Onedata
- EOS-dCache integration
- Caching systems instantiation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



First XDC Release

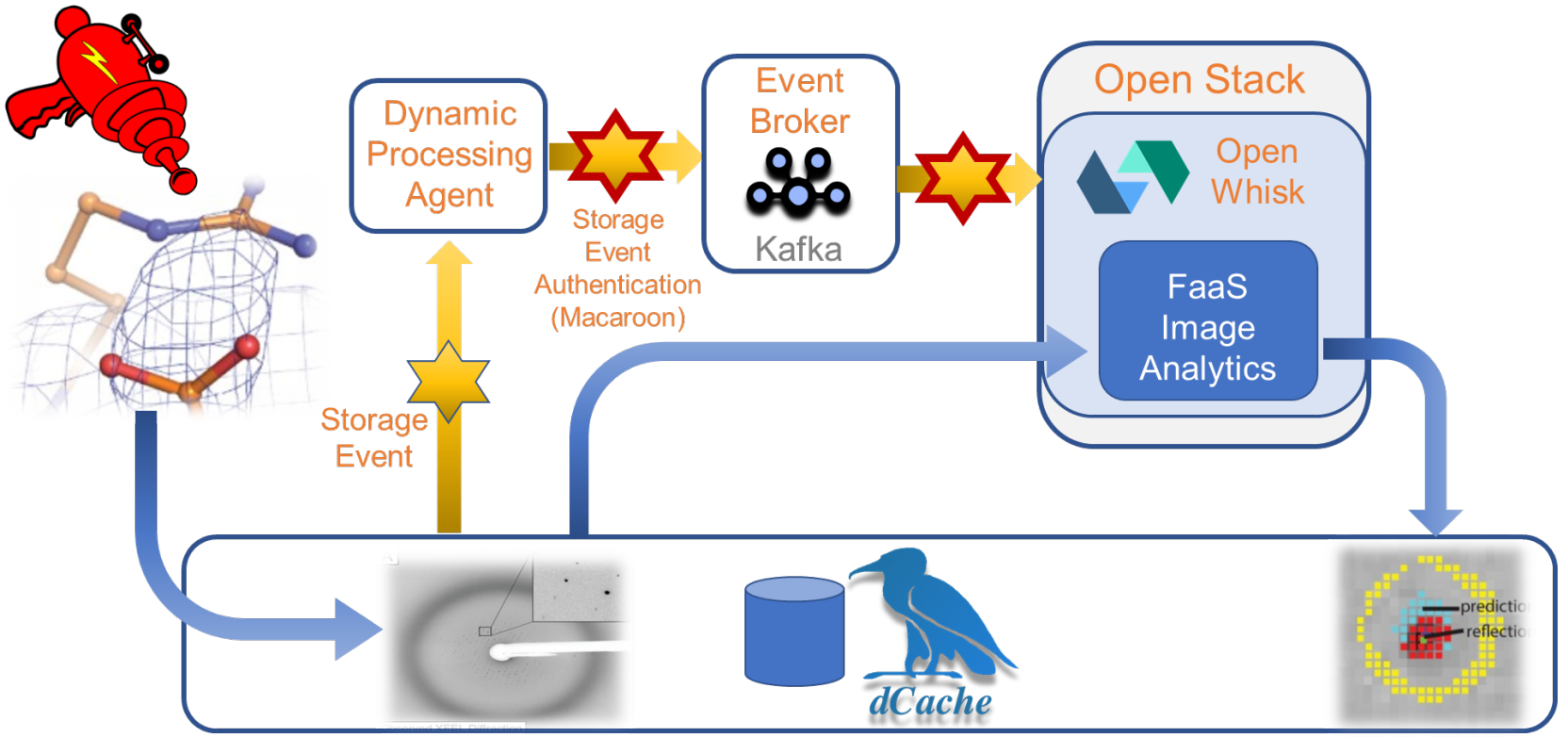
Key technical highlights

- 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



XFEL Use Case in XDC

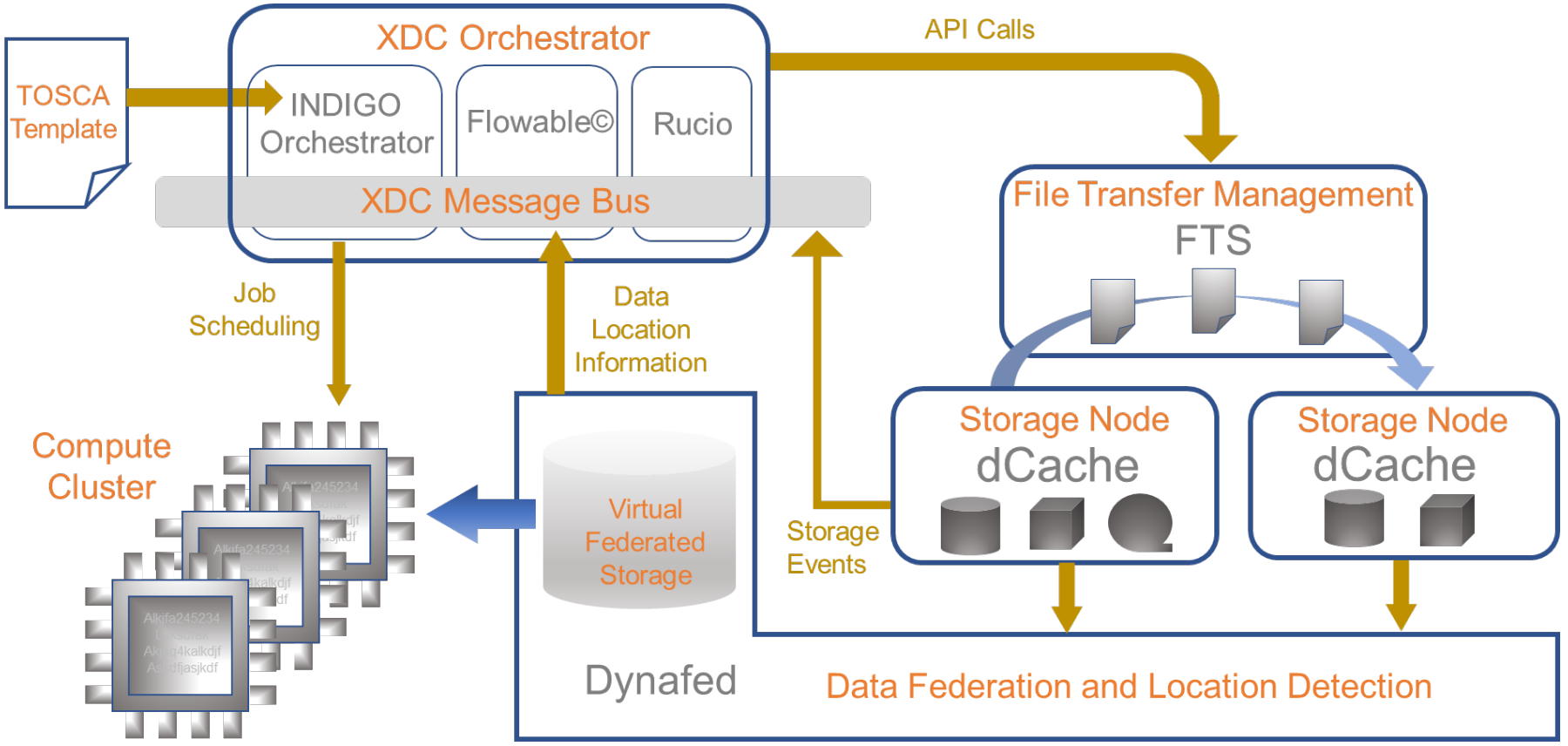
- The XFEL UseCase is driving the developments on storage events notifications support
  - A reference implementation is done using dCache as backend
- Refer to the Patrick's presentation:
  - https://indico4.twgrid.org/indico/event/8/session/15/contribution/9



© Patrick's Fuhrmann

XFEL Use Case in XDC

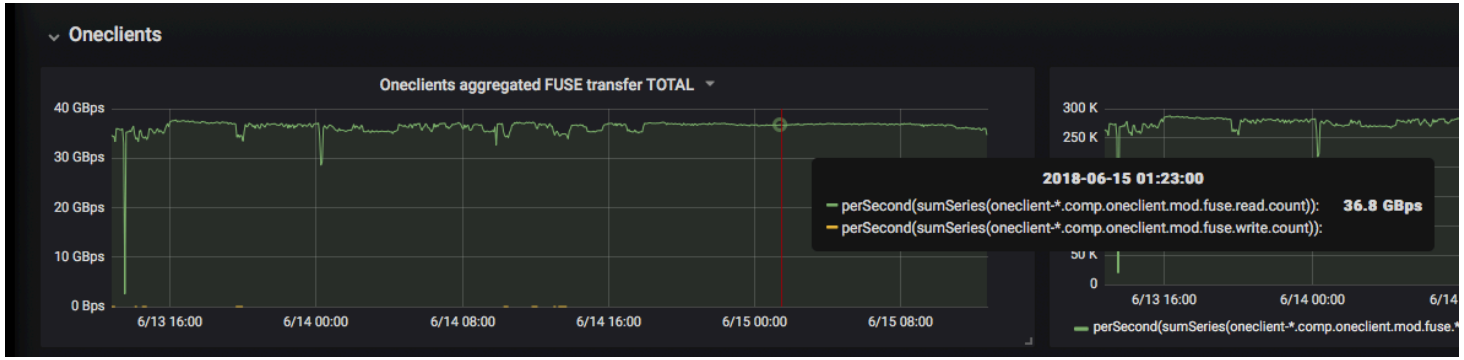
- The XFEL UseCase is driving the developments on storage events notifications support
  - A reference implementation is done using dCache as backend
- Refer to the Patrick's presentation:
  - https://indico4.twgrid.org/indico/event/8/session/15/contribution/9



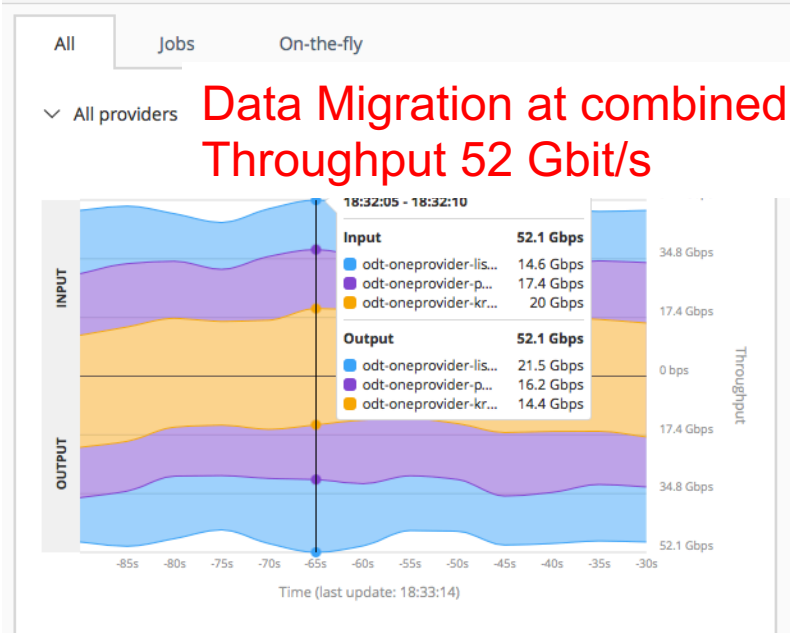
First XDC Release

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 and metadata handling 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

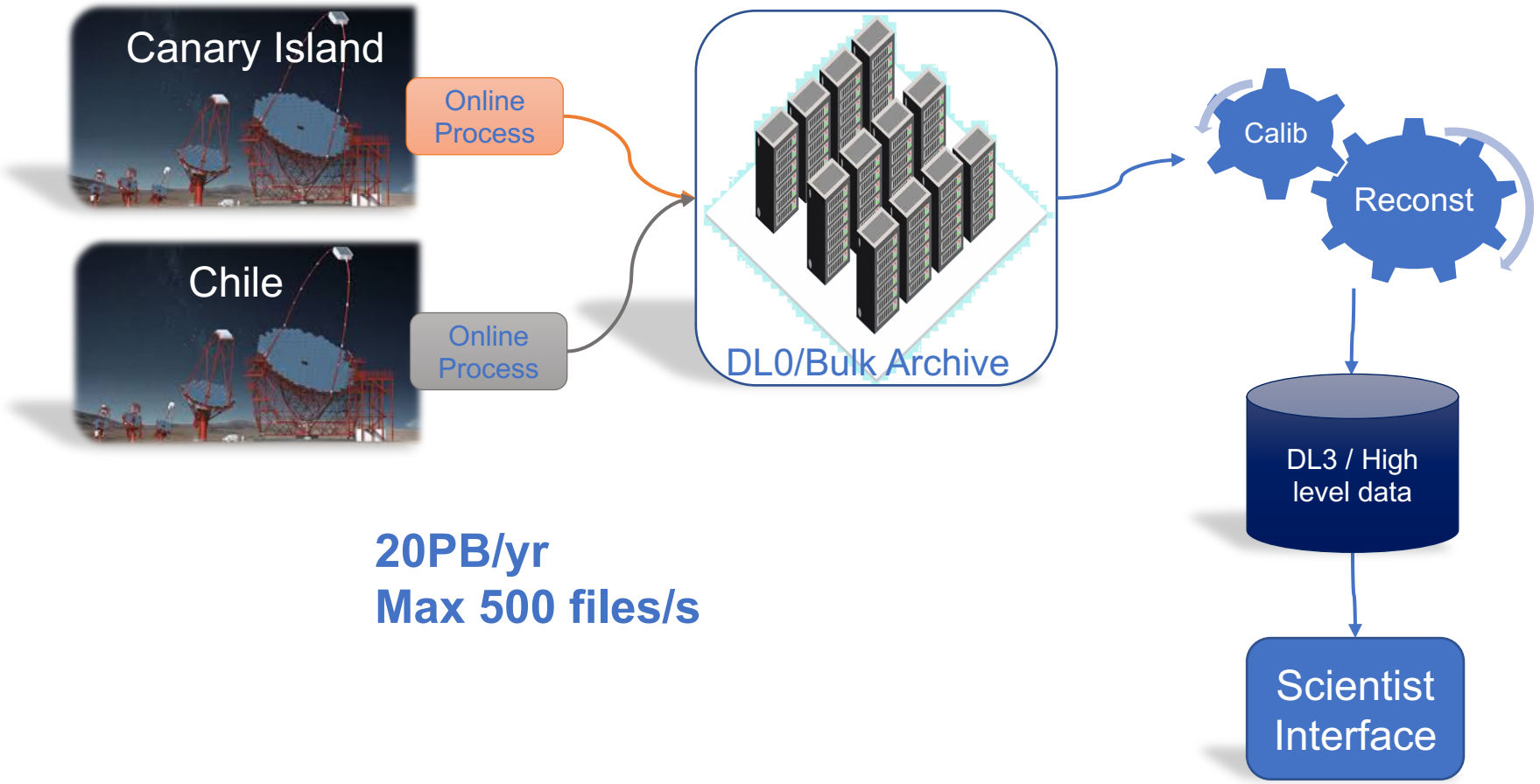


Onedata Transparent POSIX File System
Processing transparently cached data - 37GBytes/sec



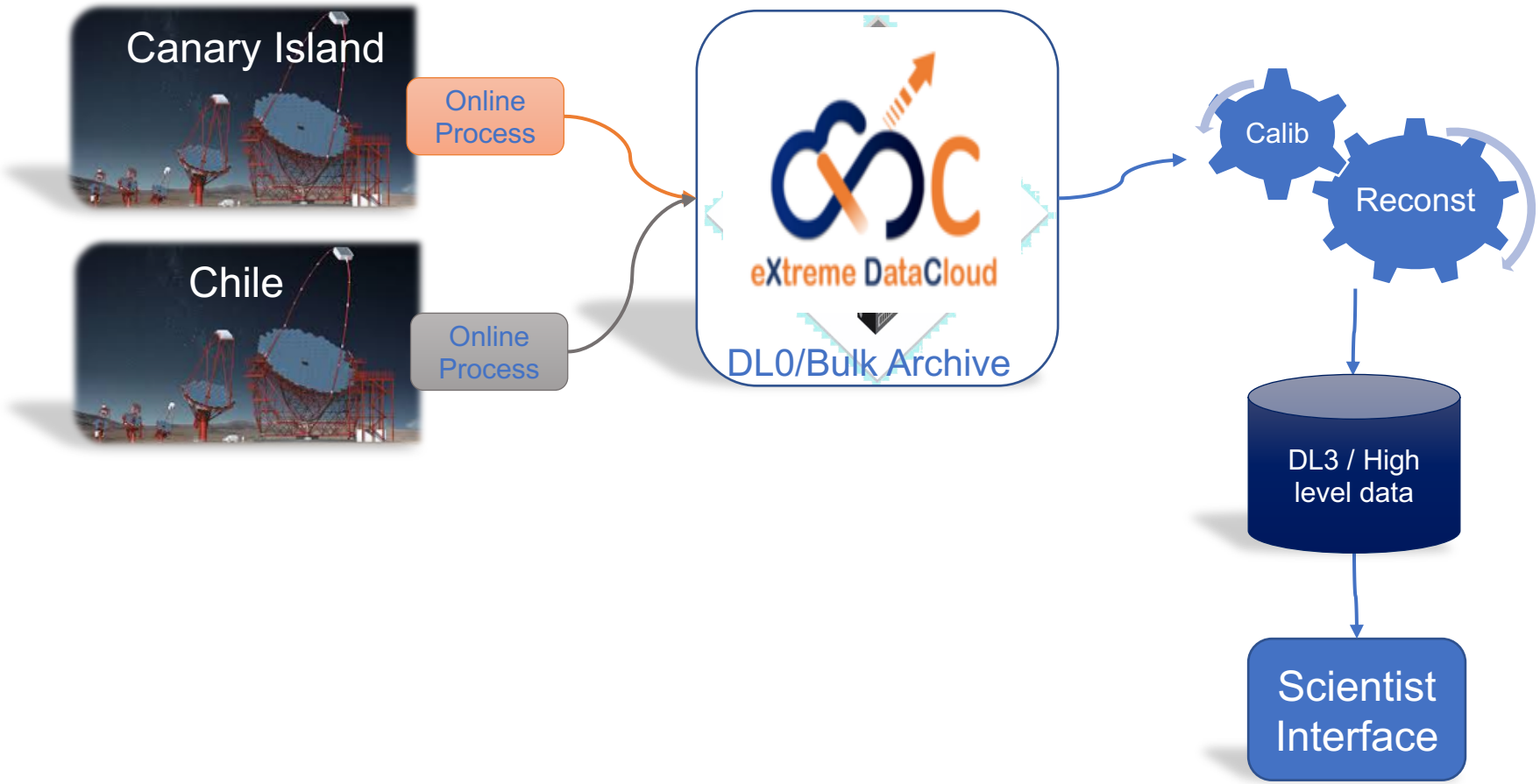
- **Data Transfer Mesh**
- **3 Oneproviders connected by 20+Gbit/s links**
- **Transfer data between all them**
- **Single VM Node per Provider**
- **Linear scalability**

CTA Use Case Workflow in XDC

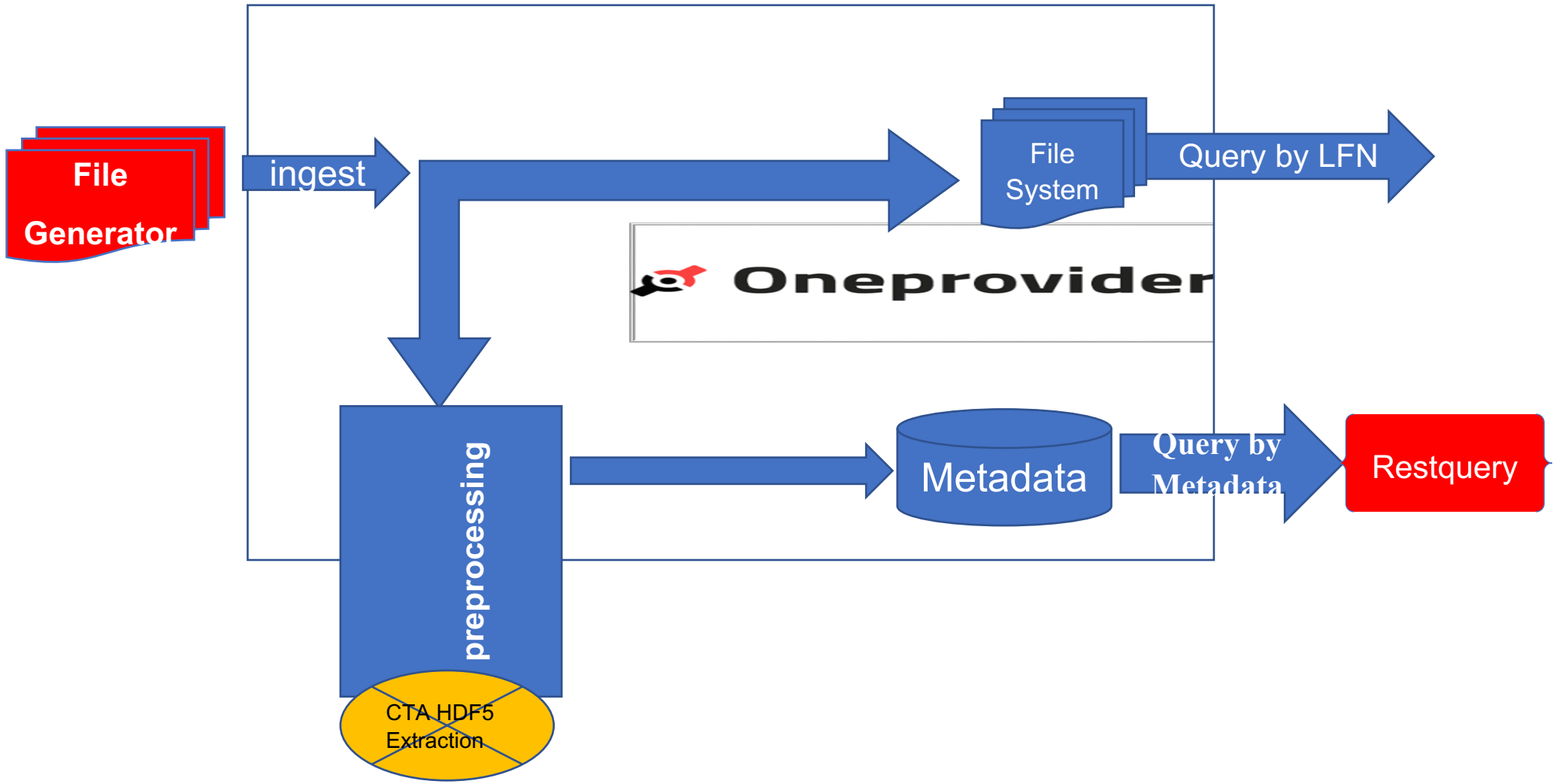


20PB/yr
Max 500 files/s

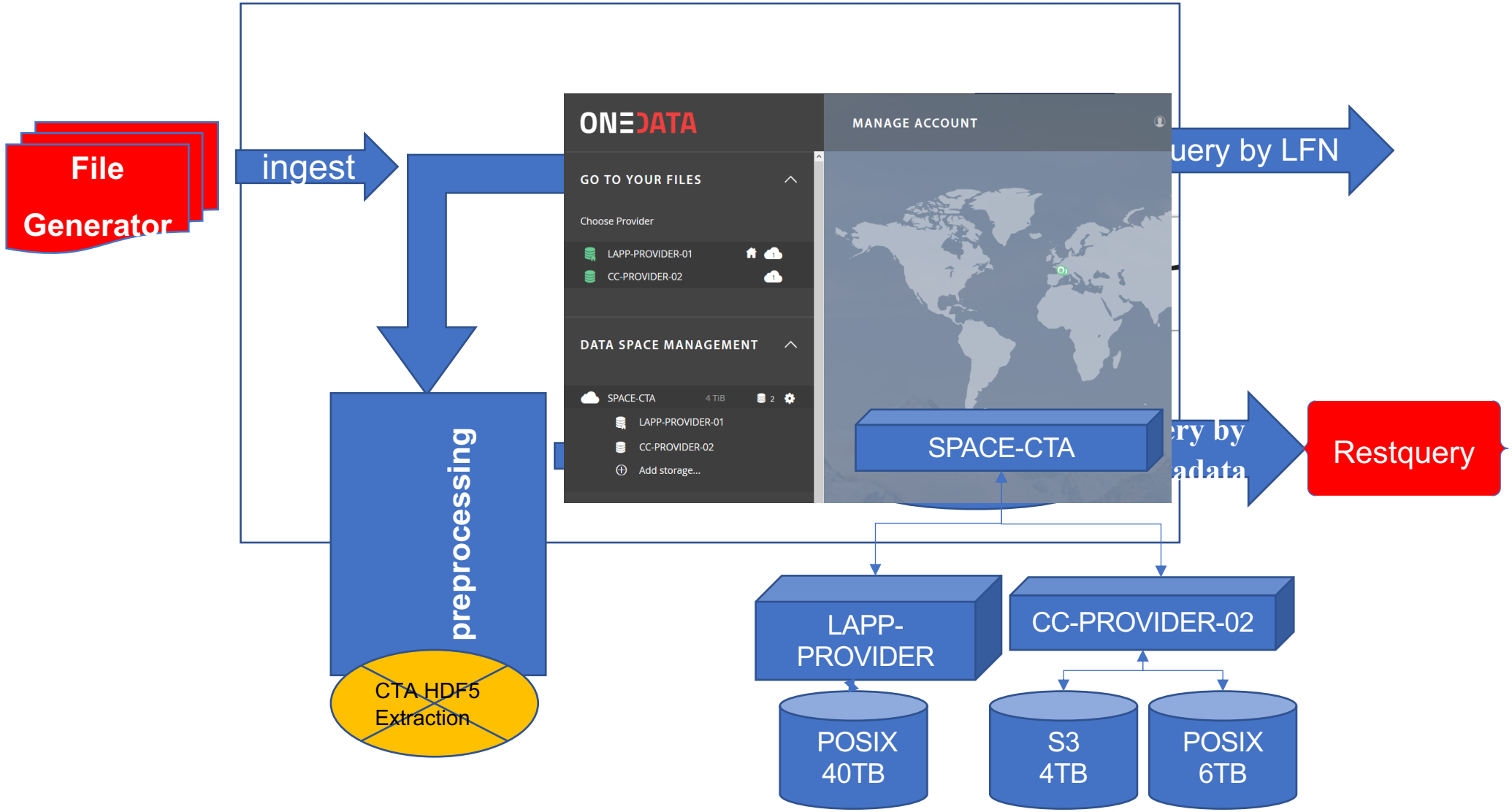
CTA Use Case Workflow in XDC



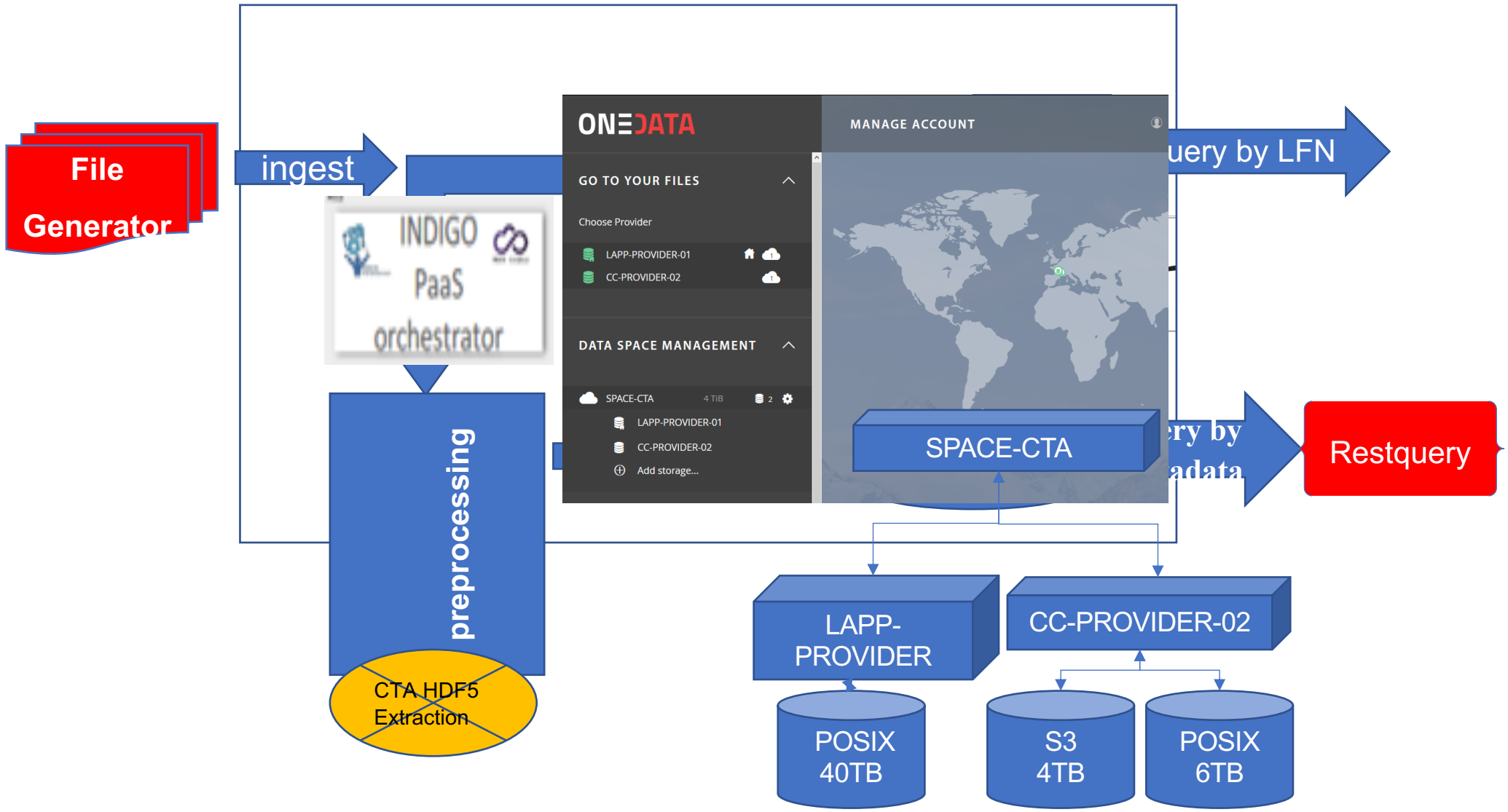
CTA Use Case Workflow in XDC



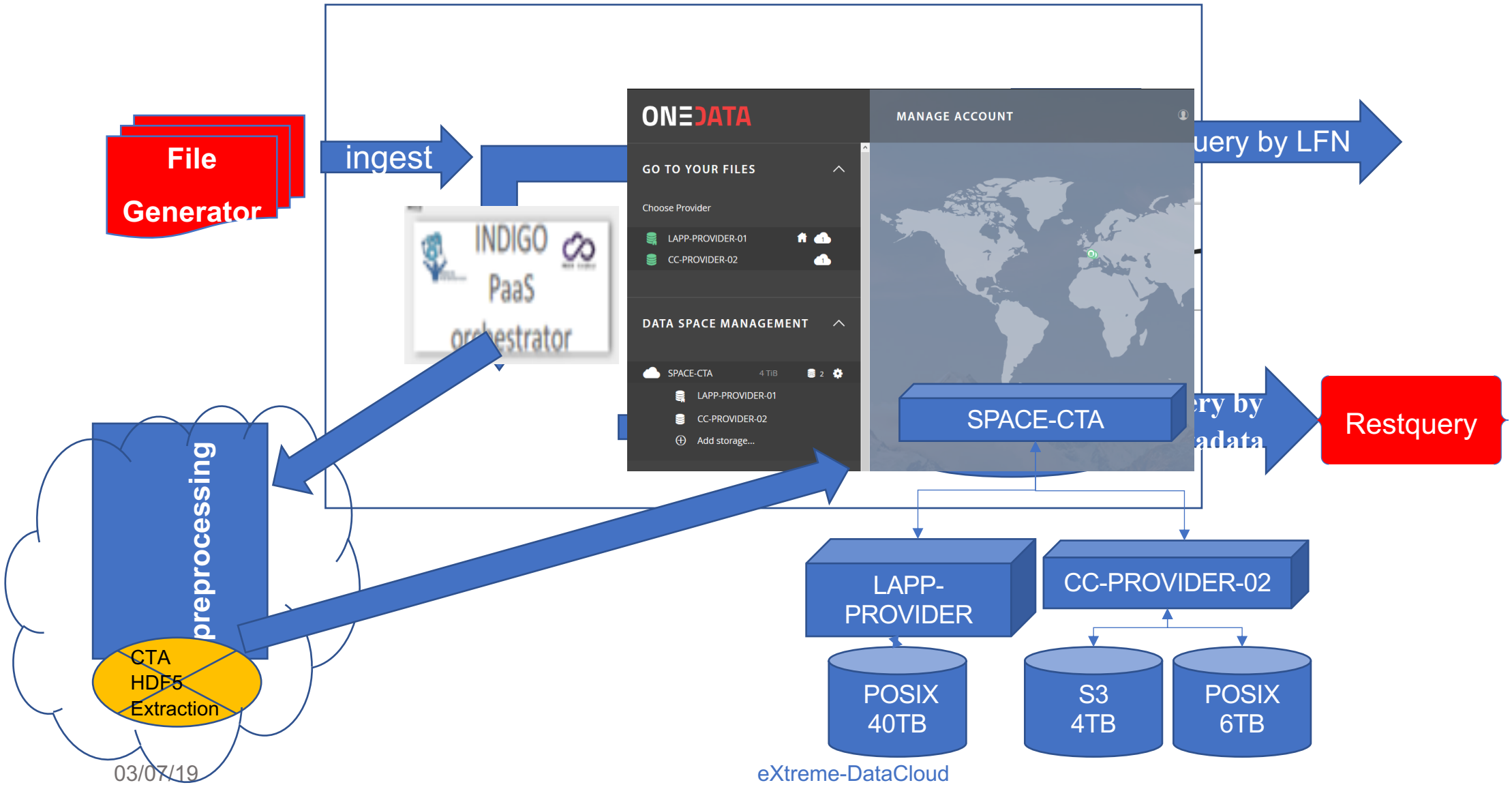
CTA Use Case Workflow in XDC



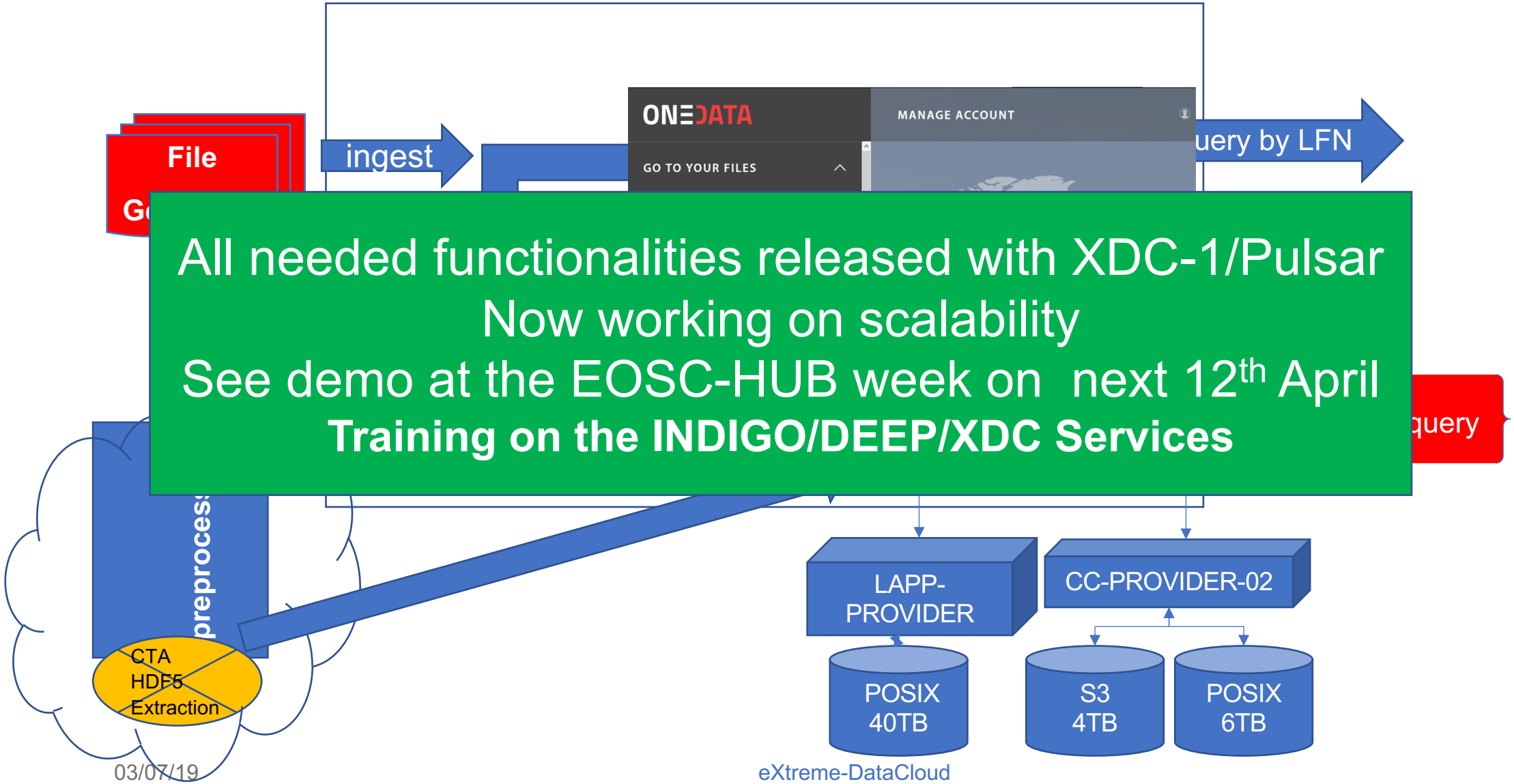
CTA Use Case Workflow in XDC



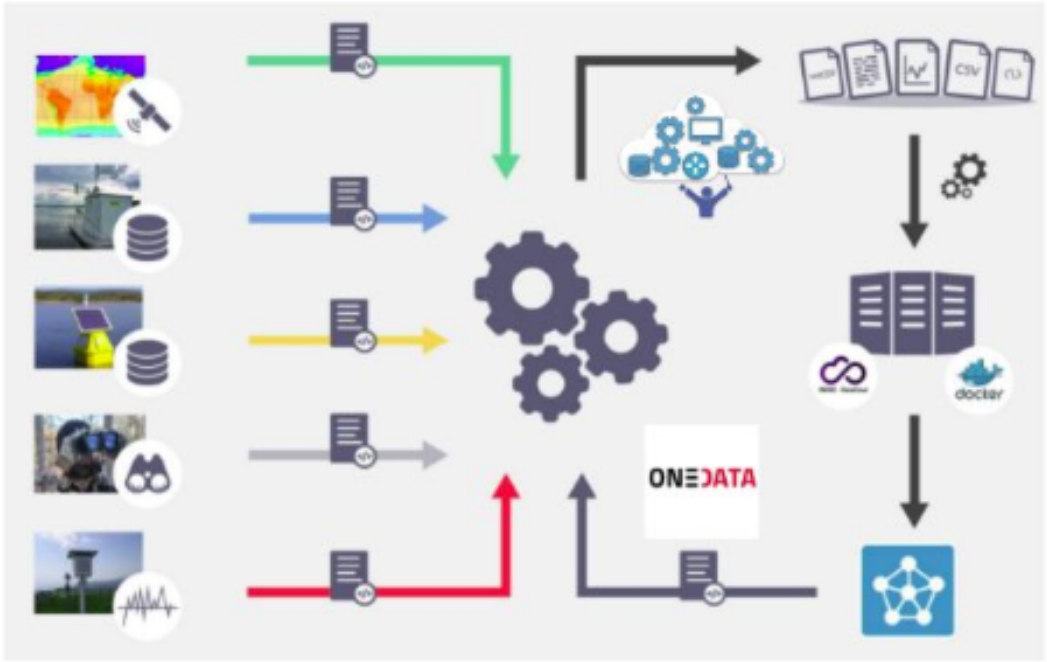
CTA Use Case Workflow in XDC



CTA Use Case Workflow in XDC



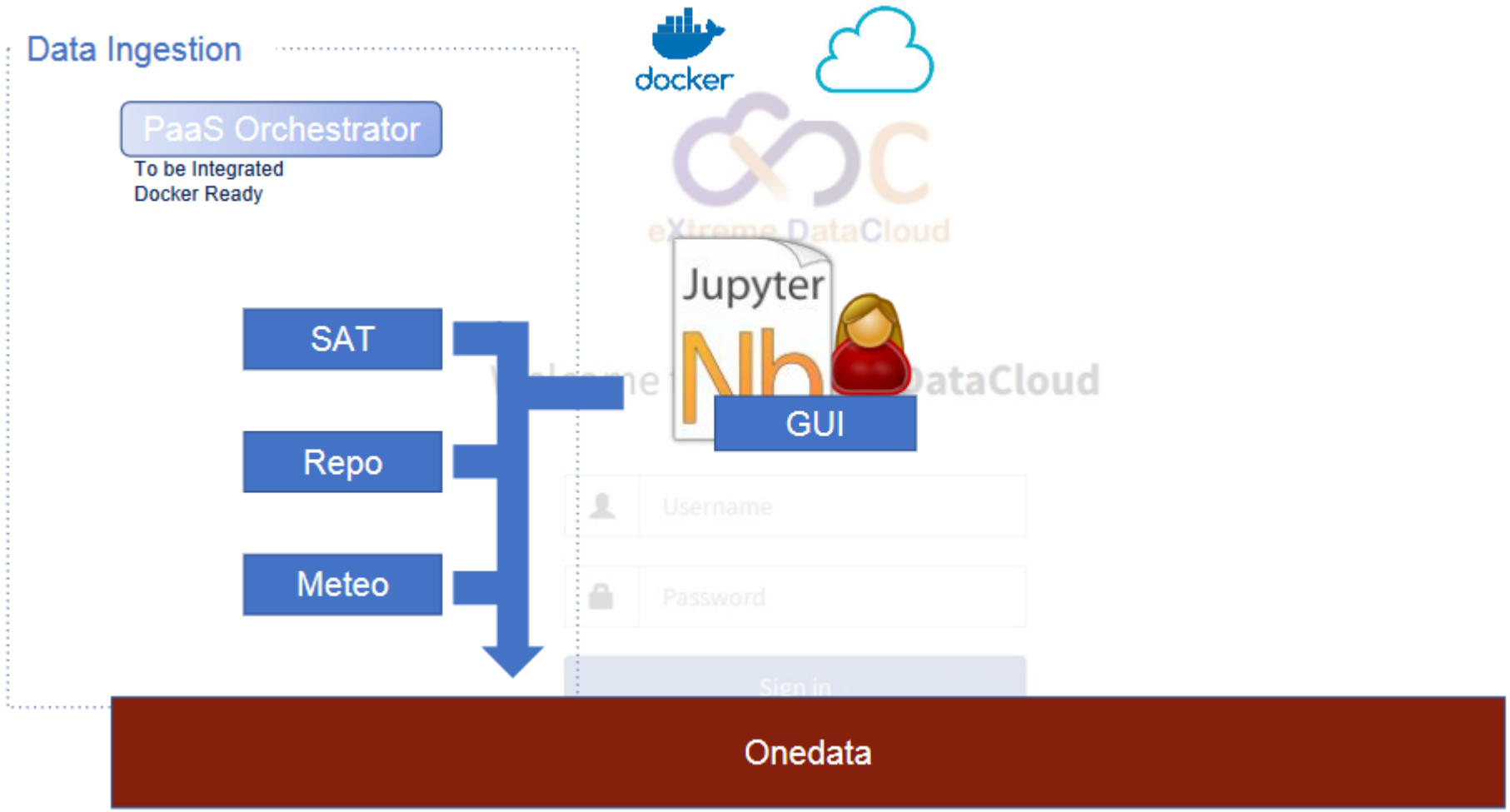
LifeWatch Use Case in XDC



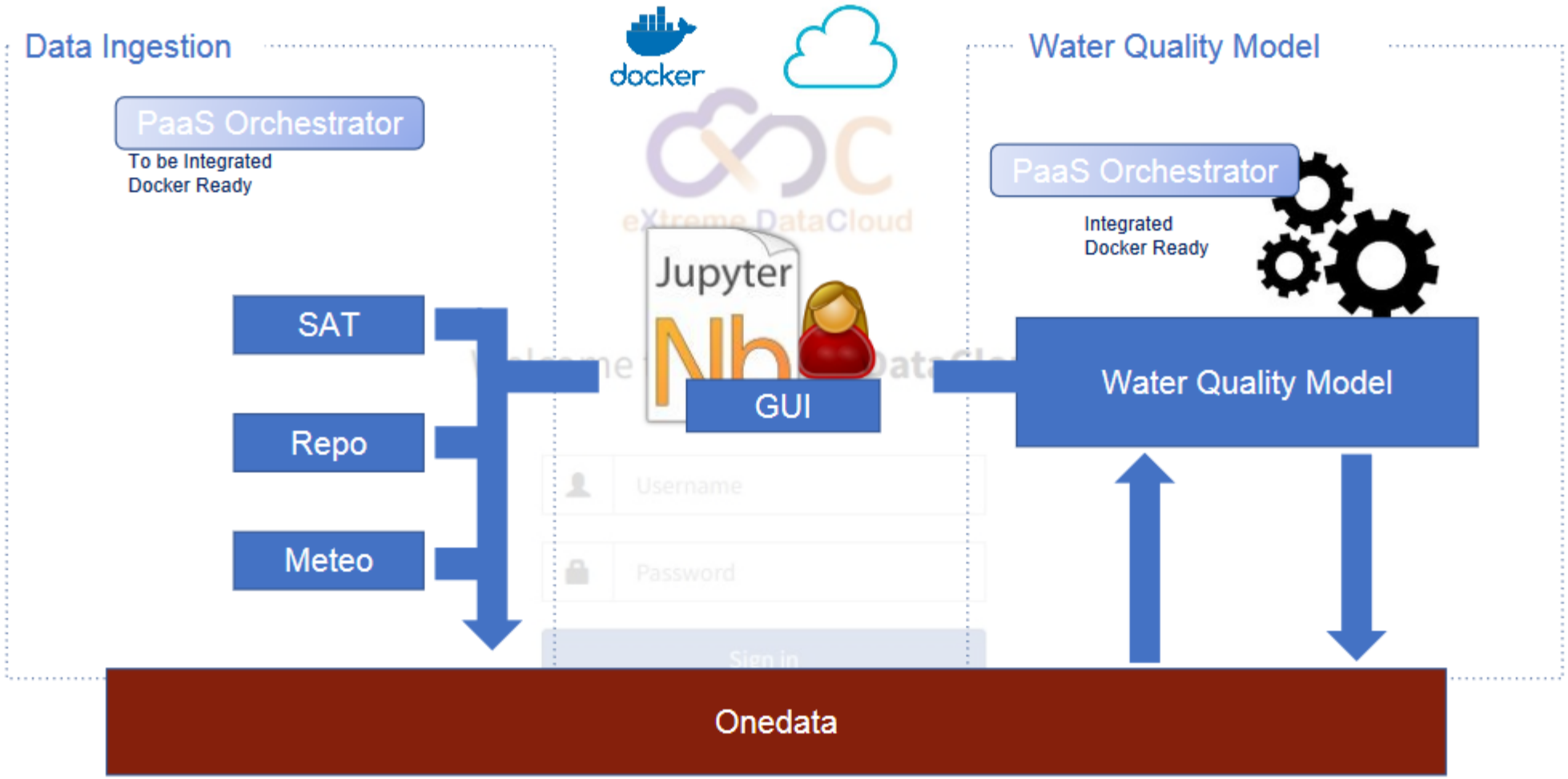
- **Objetives:** Integrate different and heterogeneous data sources: **satellite data, real-time monitoring system based on sensors, observations, and meteorological data** to feed the **hydrological and water quality models**, thus automating modeling and prediction of water quality.

- **XDC Services Requirements:**
  - **XDC IAM**
  - **Onedata:**
    - Onedata Attachment
    - Onedata Discovery
  - **PaaS Orchestrator**

LifeWatch Use Case in XDC

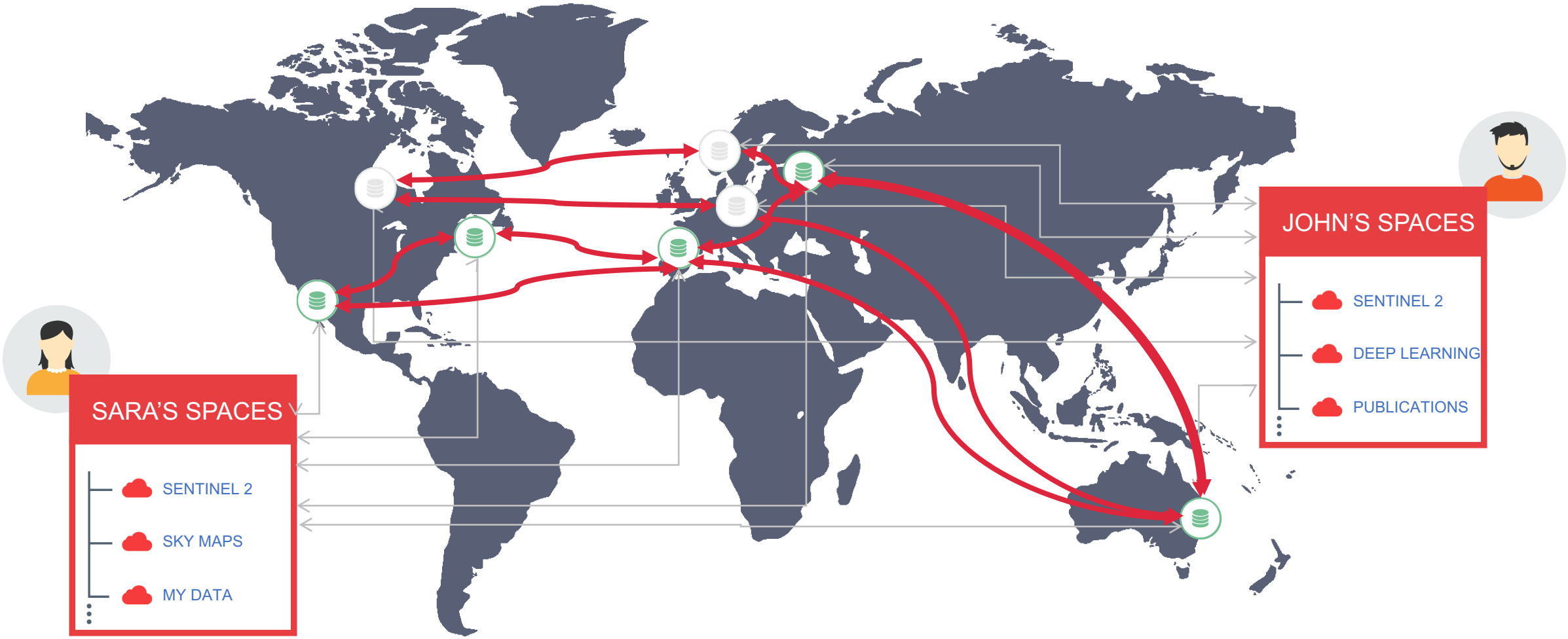


LifeWatch Use Case in XDC

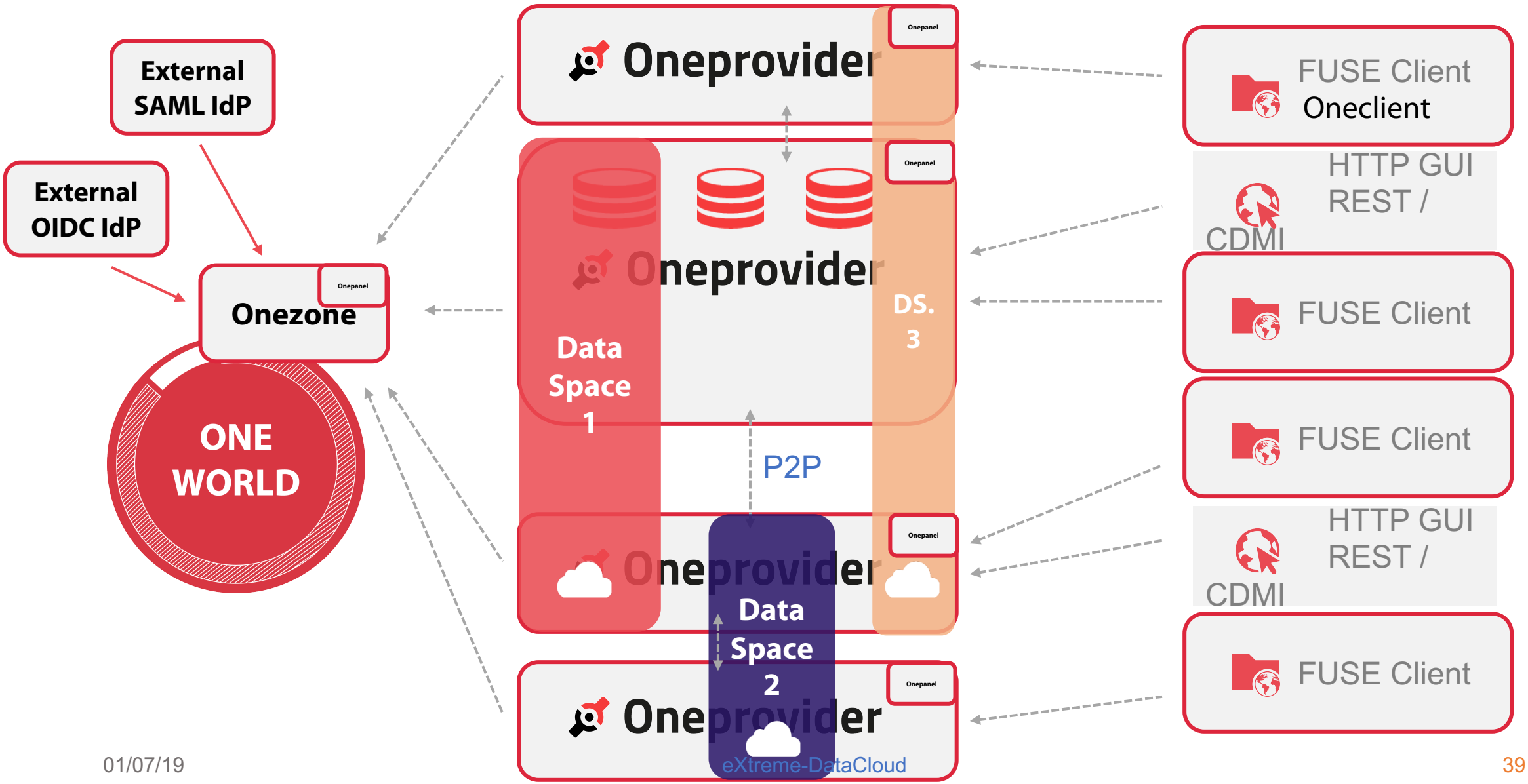


Additional slides

ONEDATA DISTRIBUTED DATA IN HYBRID CLOUDS



ONEDATA SYSTEM ARCHITECTURE



Authentication and Delegation

Token based authentication (OpenID Connect, Oauth2, Macaroons)

- Token based authentication using decentralized IdP infrastructures.
  - Avoids the X509 management and delegation overhead.
  - Will be used by WLCG in the future.
- XDC Components supporting Open ID Connect
  - FTS has demonstrated a transfer authorised with an OIDC access token
    - Still needs token translation to X509 to support legacy storage systems
  - Dynafed allows access via OIDC (redirects to the storage endpoint with same token)
    - Can even translate to S3 (Signed redirection)
  - Storage Endpoints: dCache, Onedata
  - INDIGO PaaS orchestrator delivers OIDC tokens to FTS and Dynafed.

Authentication and Delegation

Token based authentication (OpenID Connect, Oauth2, Macaroons)

eXtreme DataCloud

- Macaroons
  - Fast anonymous delegation token allowing caveats, like max lifetime, restricted IP range, number of accesses.
  - Available already in dCache.
  - Candidate for 3rd party copy with FTS and WebDAV.
- Identity and Group Management
  - Using the INDIGO AIM for XDC.
  - Expecting WLCG to use AIM as well (decision imminent)
  - We assume, our solution is compatible with the EGI Checkin.
  - Demos provided by INFN
    - IAM -> EGI Checkin -> home IdP or EGI SSO, for talking to the Orchestrator. Will be presented at the EOSC Hub week (Prague)
    - Deployment of VMs on the EGI federated Cloud by the Orchestrator using the IAM through EGI SSO.

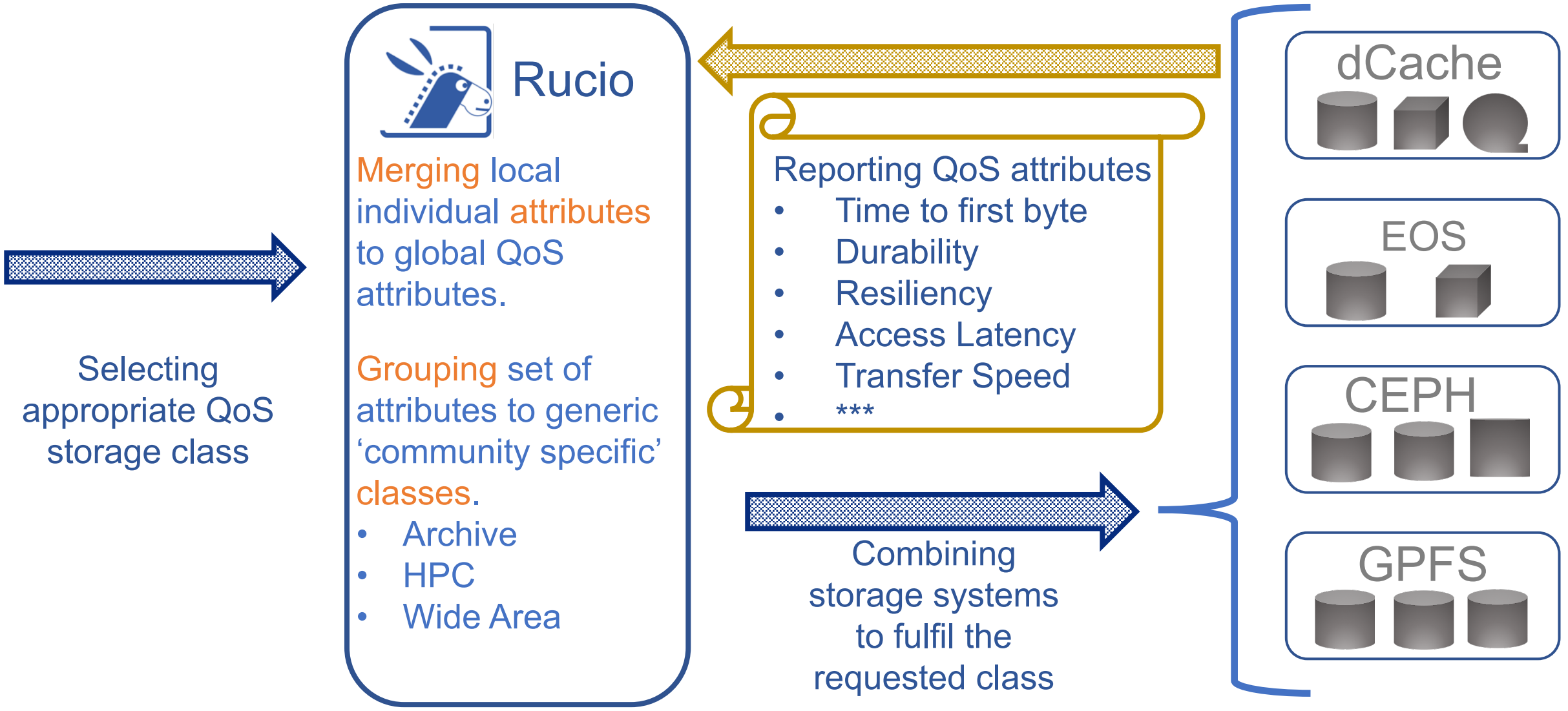
QoS in storage

- Why QoS in storage
  - Different points in the data life cycle path have different requirements on the quality of the underlying storage
  - Different storage providers come with different storage capabilities and consequently with different prices models
  - Matching the above may safe significant resources (money)
- Prerequisites
  - Storage Endpoints need to understand the quality of their attached storage devices (Tape, Spinning Disk, SSD, Cloud Storage)
  - Storage Endpoint must report those qualities in a standard way and must allow the quality to be chosen and should allow transitions between qualities.
  - Orchestration System must be able to query the endpoints for the quality capabilities and must be able to match requirements and availabilities on the global level.

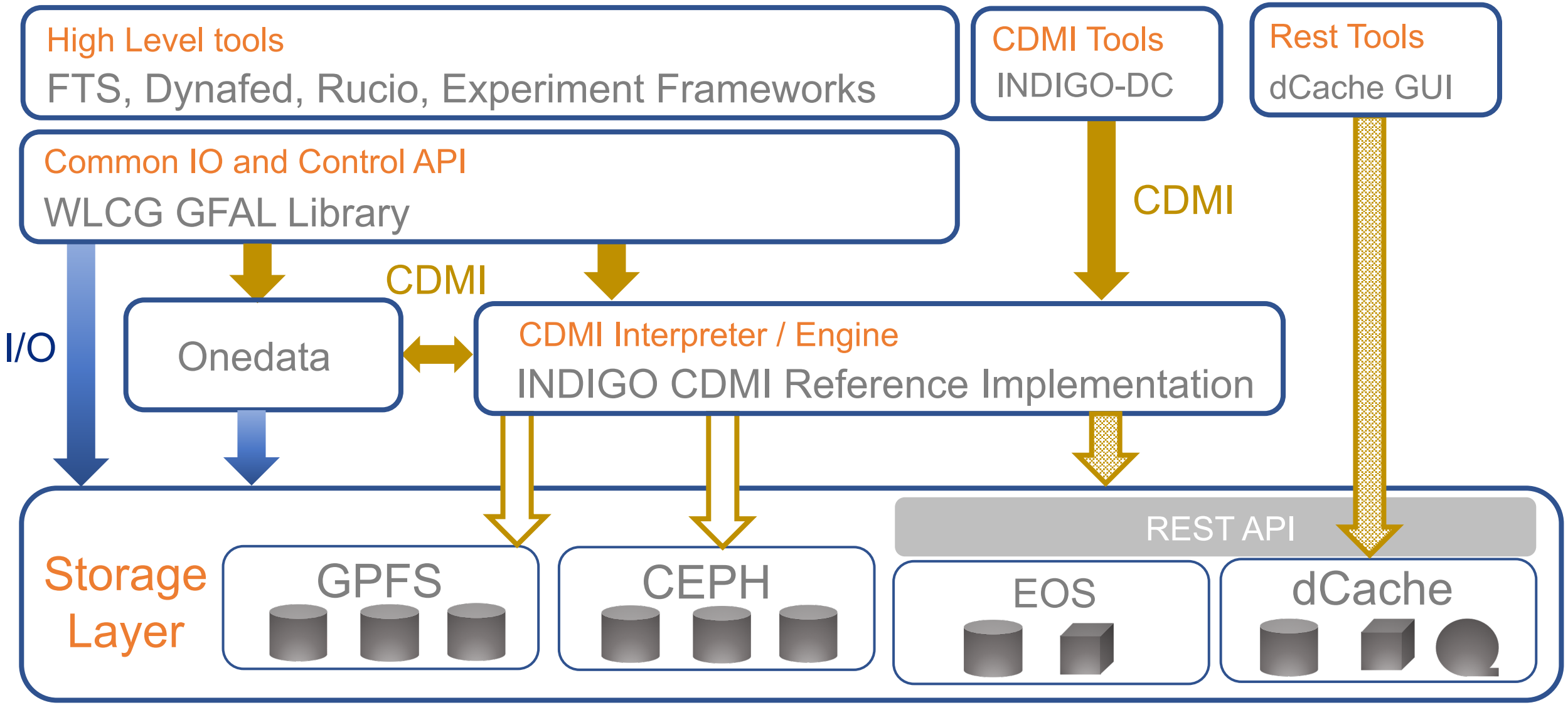
XDC Activities to support QoS

- Building upon the work of INIDIGO DataCloud
  - dCache already supports the idea of QoS, steered with REST
  - INDIGO reference implementation provides a plug-in system to convert CDMI requests into REST, CEPH and GPFS calls.
  - Collaboration with RDA and SNIA
- XDC work with sustainability in mind.
  - Agreement with CERN software providers (EOS, FTS, Rucio) and ONEDATA on a REST and CDMI core set of functionalities.
  - XDC leading the WLCG DOMA working group.
    - Commitment of the DOMA group to work with XDC.
  - XDC involvement in ESCAPE on the same topic.

Global QoS Architecture



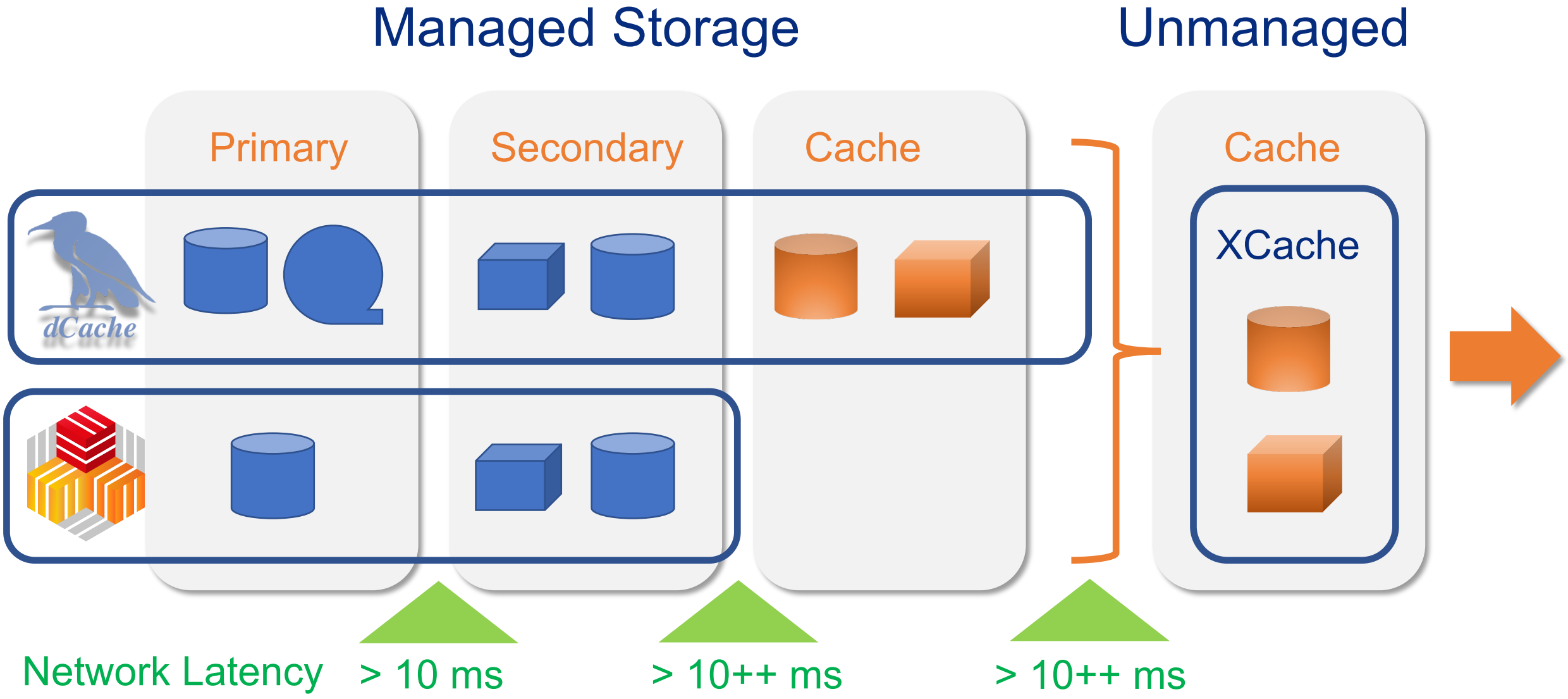
Design of QoS with CDMI



Caching Scenarios

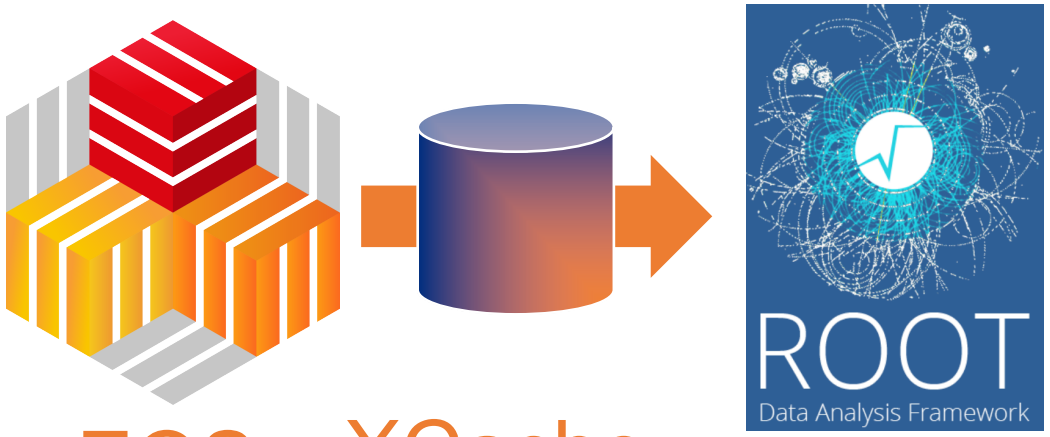
- Caching, depending on the use case, can significantly reduce performance degradation due to network latency.
- Why do we need different caching systems?
  - Caching is infrastructure and application dependent.
  - Available cache systems:
    - Managed secondary storage and cache
    - Unmanaged cache.
- Currently we provide three scenarios
  - EOS with XCache
  - INFN National Cache, using XCache
  - dCache distributed cache. (Part of the WLCG DOMA group)

XDC Cache Models



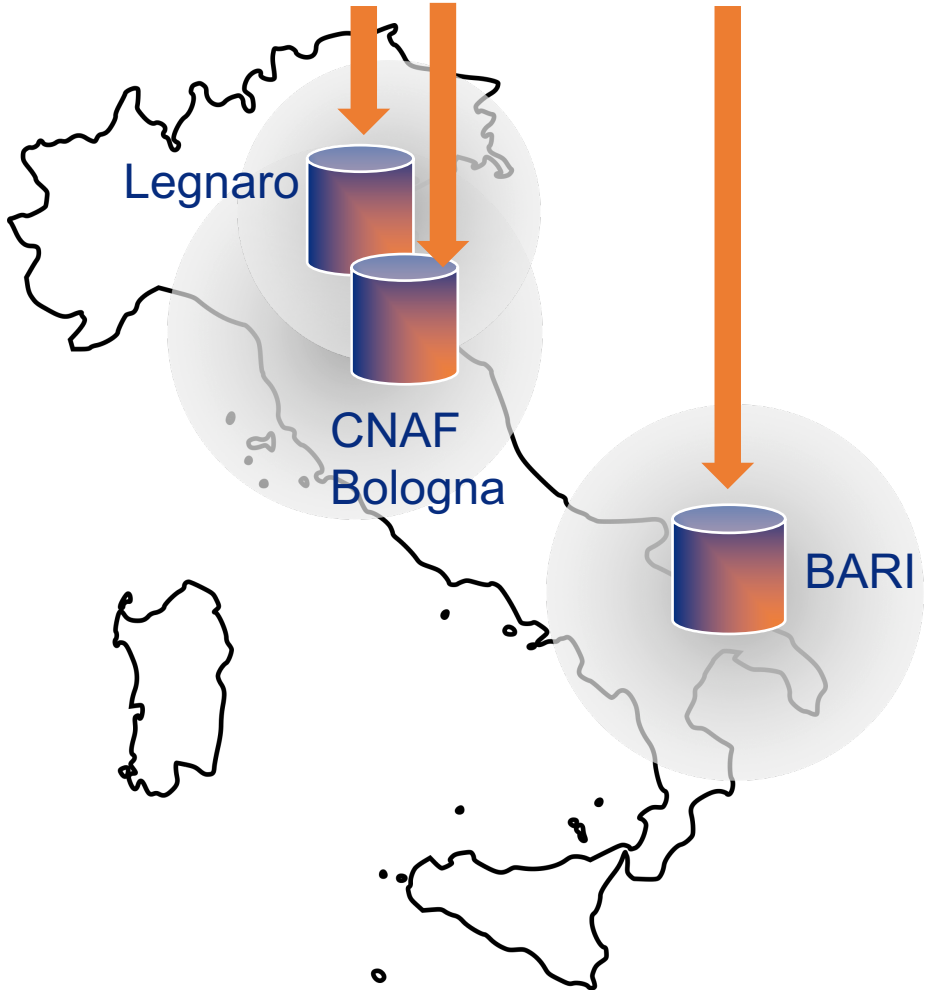
Cache Demos (XCache Deployment)

EOS and XCache

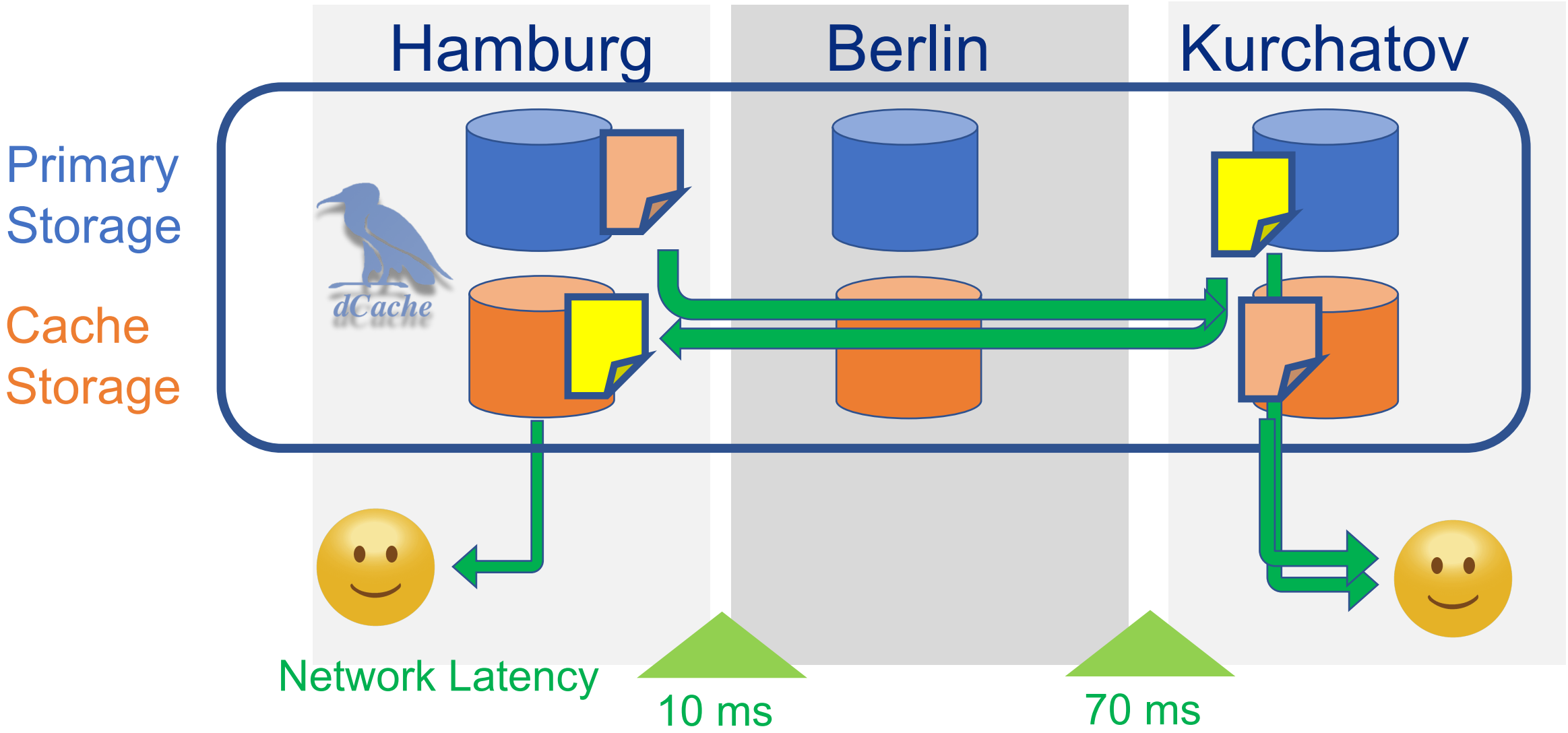


EOS XCache

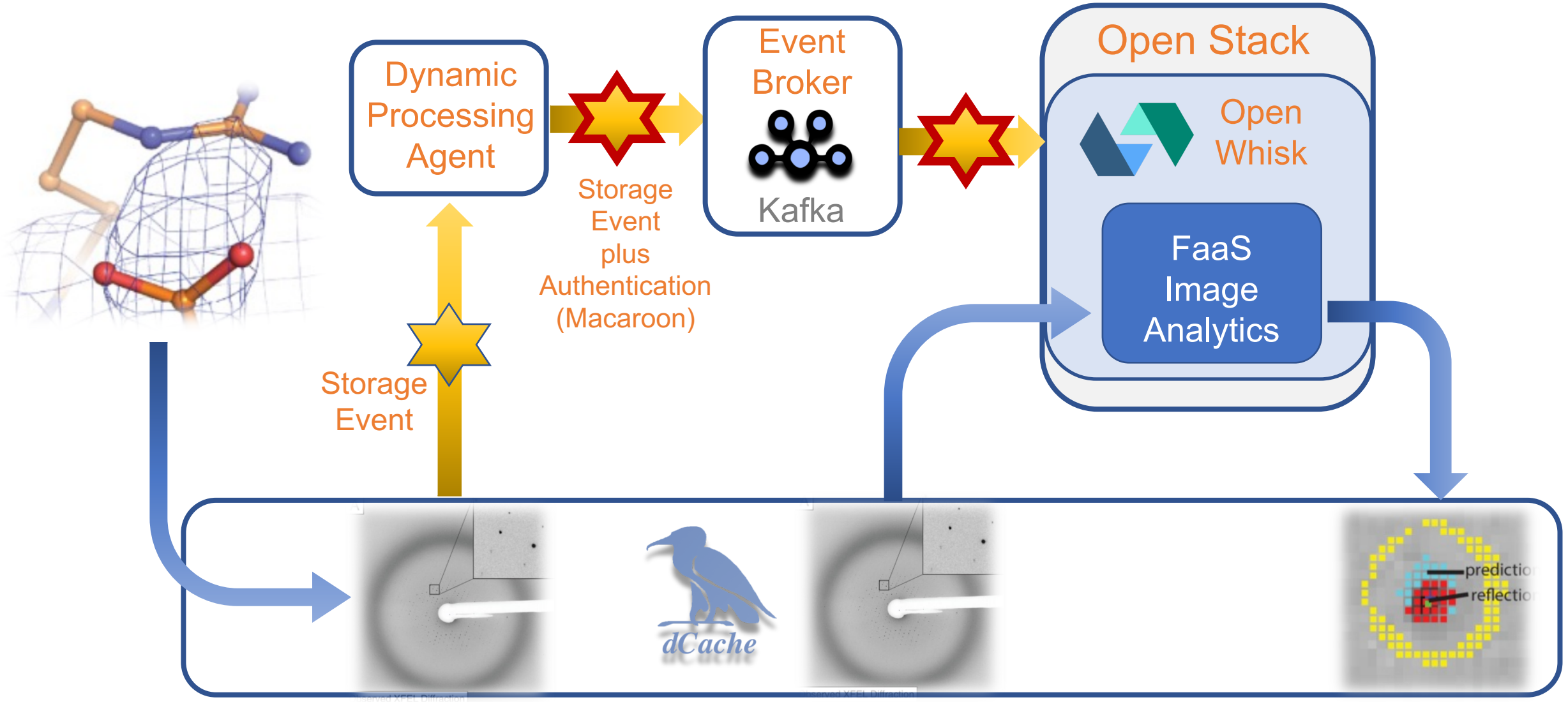
Italian (INFN) National Cache



Caching for WLCG DOMA



Hardwired Compute Orchestration



Onedata Achievements (1)

- Unified data access platform for PaaS at large
  - 100M+ files per collection
  - 100GB/s range local processing throughput
  - 100Gbps range distributed meshed transfers
- Simplified deployment
  - Onedatify
  - Integrated DNS and subdomain delegation
  - Integrated Let's encrypt
- Onedata can be used for high throughput data acceleration at exascale level
- Advanced file popularity and caching mechanism implemented
- Next generation integrated GUI interface
- Simplified rights management
- Implemented driver for WebDAV for integration with dCache and EUDAT- B2X

Onedata Achievements (2)

- Redesigned and reimplemented internal map reduce system for large scale indices based on single collections
- Performance improvements in metadata management subsystem
- Interface for external logic depending on metadata streams
- Event based streams interfaces optimized and extended filtering options
- Integration attributes with POSIX API (xattr command works from command line)
- Public Shares integrated with DOI/PID minting

Onedata Achievements (3)

- Prepared architecture for distributed metadata gathering and indexing
- Introduction new concept at Onezone level – Harvesters
  - Collecting metadata from multiple: spaces, providers
  - Using external indexing services (pluggable) - at the moment Elasticsearch, but other to come like MySQL, SemanticDBs etc.
  - Pluggable data discovery portal
  - Detailed access control

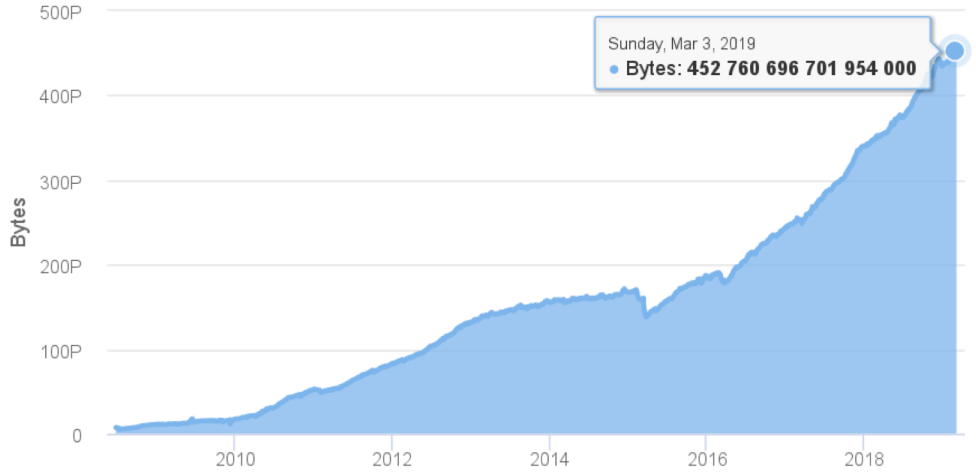
RUCIO Slide

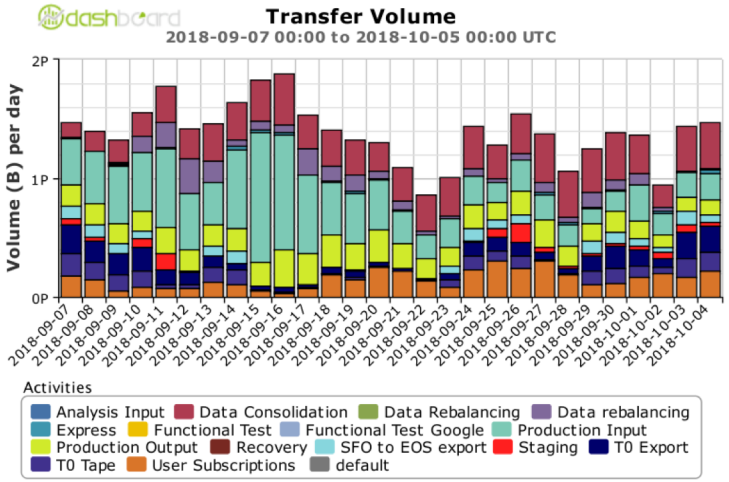
Rucio in a nutshell

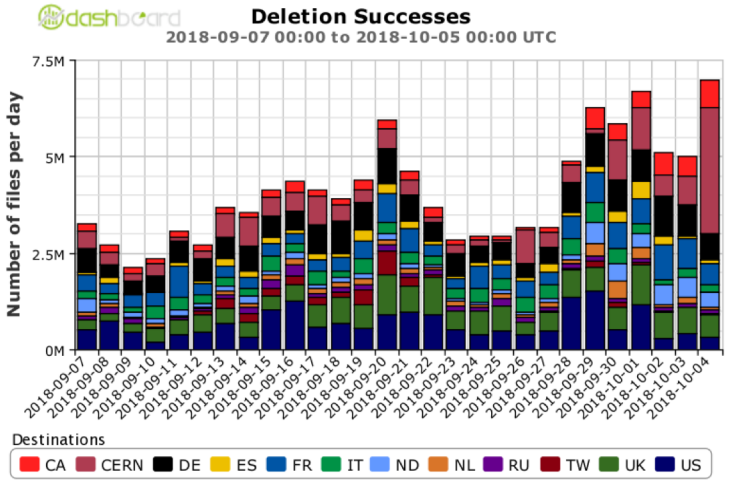
- Developed by the High-Energy Physics experiment ATLAS
- Rucio provides a complete and generic scientific data management service
  - Data can be scientific observations, measurements, objects, events, images saved in files
  - Facilities can be distributed at multiple locations belonging to different administrative domains
  - Designed with more than 10 years of operational experience in large-scale data management!
- Rucio manages multi-location data in a distributed environment
  - Creation, location, transfer, and deletion of replicas of data
  - Orchestration according to both low-level and high-level driven data management policies (usage policies, access control, and data lifetime)
- Rucio (arXiv) is open source and available under Apache 2.0 license
- Make use of established open source tools

Data management at ATLAS

- ATLAS instance in a few numbers
  - 1B+ files, 450 PB of data, 400+ Hz frontend interaction rate
  - Up to 4M files/2.5 PB transferred per day
  - 10PB access via Rucio mover; >1000 active users
- Expect to increase one order of magnitude for Run4







Community


Advanced European Network of E-infrastructures
for Astronomy with the SKA


LCLS


Belle II


Science & Technology
Facilities Council

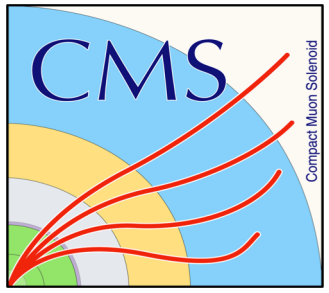

COMPASS


3D
EISCAT


ATLAS
EXPERIMENT


Alpha Magnetic Spectrometer
AMS-02
Europe • Asia
North America


Xe
XENON
Dark Matter Project


CMS
Compact Muon Solenoid


DUNE
DEEP UNDERGROUND
NEUTRINO EXPERIMENT


egi


ARC


cta
cherenkov telescope array


LSST


IceCube
SOUTH POLE NEUTRINO OBSERVATORY


FTS


eXtreme DataCloud


Fermilab


BROOKHAVEN
NATIONAL LABORATORY


SKA
SQUARE KILOMETRE ARRAY


LIGO


Open Science Grid

Rucio main functionalities

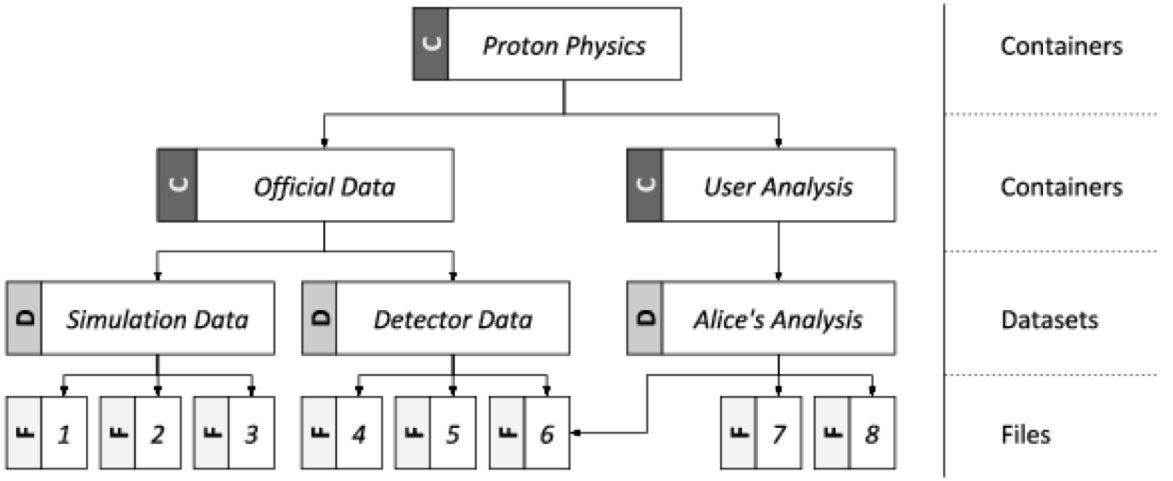
- Provides many features (Can be enabled selectively)
  - File and dataset catalog (logical definition and replicas)
  - Transfers between sites and staging capabilities
  - User Interface and Command Line Interface to discover/download/upload/transfer data
  - Extensive monitoring
  - Powerful policy engines (rules and subscriptions)
  - Bad file identification and recovery
  - Dataset popularity based replication
  - …
- Rucio can be integrated with Workload and Workflow Management System
  - Already supporting PanDA (ATLAS WFMS)
  - Planned integration with DIRAC

More advanced features

Rucio concepts - Namespace with DIDs

- All data stored in Rucio is identified by a Data IDentifier (DID)
- There are different types of DIDs
  - Files
  - Datasets: Collection of files
  - Container: Collection of dataset and/or container
- Each DID is uniquely identified and composed of
  - Scope
  - Name
  - Example:

    `user.martin:test.file.001`



Rucio concepts - Metadata

- Rucio supports different kinds of metadata
  - System-defined, e.g., size, checksum, creation time, status
  - Physics, e.g., number of events, lumiblock
  - Production, e.g., which task or job produced the file
  - Data management internal: necessary for the organisation of data, e.g., replication factor
- Metadata are custom attributes on data identifiers
  - Enforcement possible by type, e.g., enum
  - Naming convention enforcement and automatic metadata extraction
- Provides additional namespace to organise the data
  - Searchable via name and metadata
  - Aggregation based on metadata searches
  - Can also be used for long-term reporting (e.g., evolution of particular metadata selection over time)

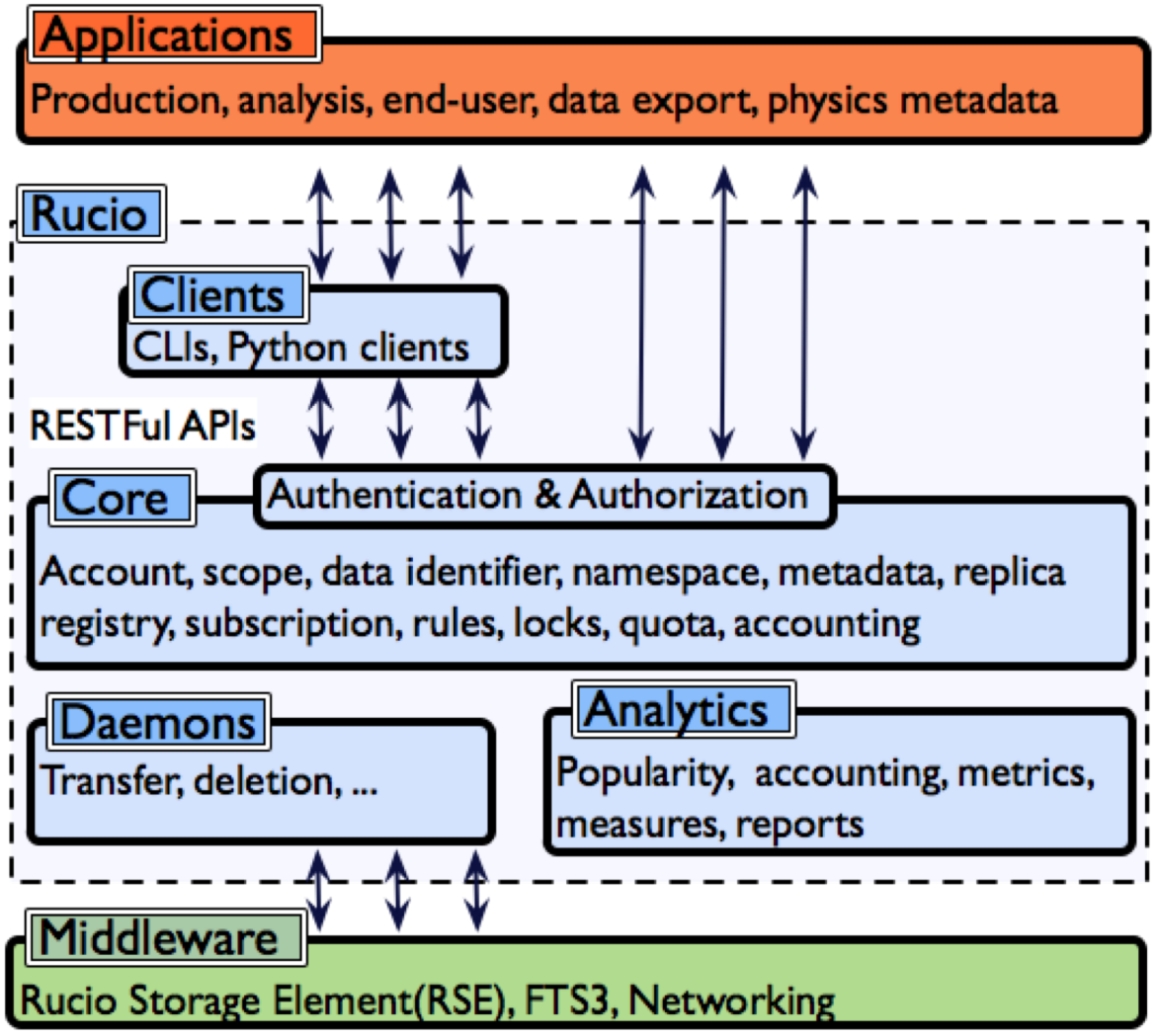
Rucio concepts - RSEs

- Rucio Storage Elements (RSEs) are logical entities of space
  - No software needed to run at the site
  - RSE names are arbitrary (e.g., "CERN-PROD_DATADISK", "AWS_REGION_USEAST", …)
  - Usually one RSE per site and storage data class
- RSEs collect all necessary metadata for a storage system
  - protocols, hostnames, ports, prefixes, paths, implementations, …
  - data access priorities can be set (e.g. to prefer a protocol for LAN access)
- RSEs can be assigned meta data
  - Key/Value pairs (e.g., *country=UK*, *type=TAPE*, *support=brian@unl.edu*)
  - You can use RSE expressions to describe a list of RSEs (e.g. `country=UK&type=TAPE`

Rucio concepts - Declarative data management

- Express what you want, not how you want it
  - *e.g., "Three copies of this dataset, distributed evenly across two continents, with one copy on TAPE"*
- Replication rules
  - Rules can be dynamically added and removed by all users
  - Evaluation engine resolves all rules and tries to satisfy them by requesting transfers and deletions
  - Lock data against deletion in particular places for a given lifetime
  - Primary replicas have indefinite lifetime rules
  - Secondary replicas are dynamically created replicas based on traced usage and popularity
- Subscriptions
  - Automatically generate rules for newly registered data matching a set of filters or metadata
  - e.g., *project=data17_13TeV* and *data_type=AOD* evenly across *T1s*

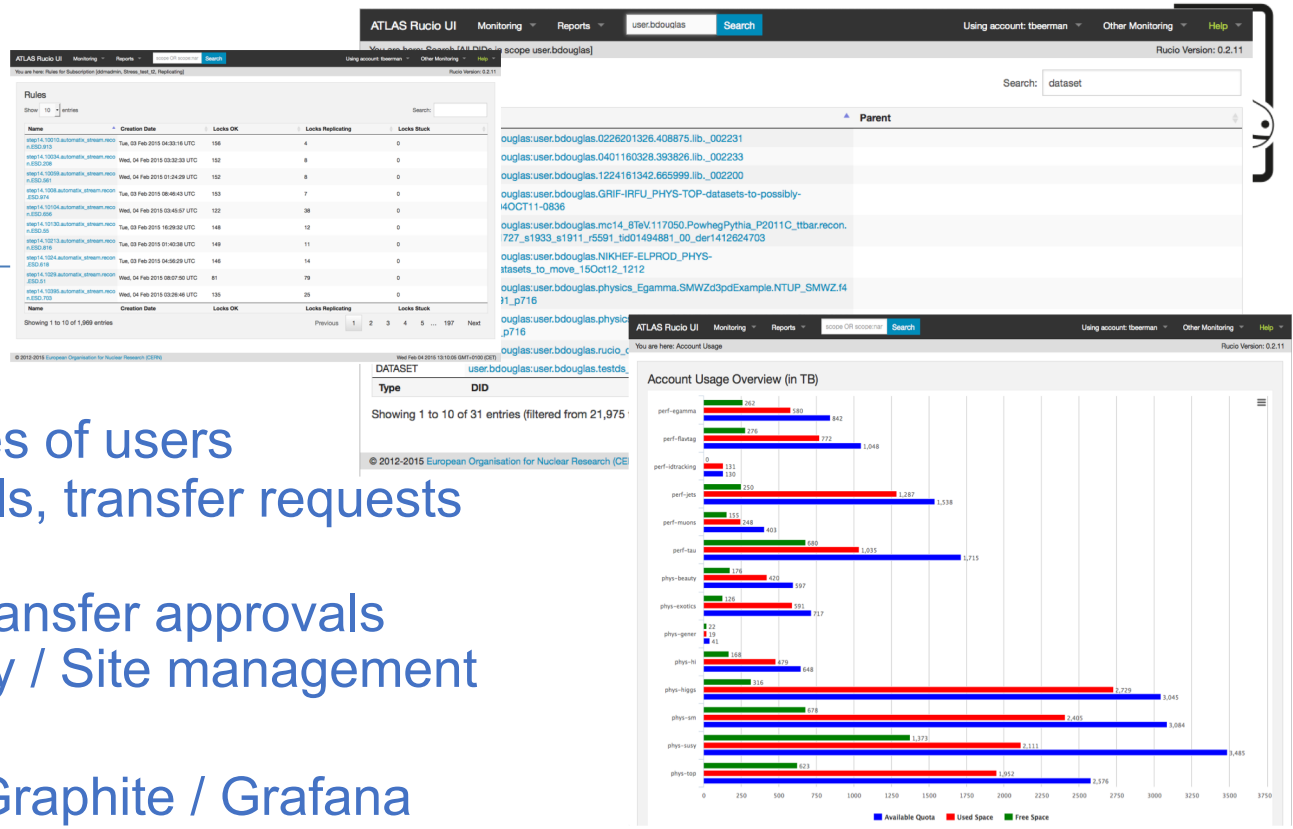
Architecture



Fully built on open standards and frameworks!

- **Servers**
  - HTTP REST/JSON APIs
  - Token-based authentication (x509, ssh, kerberos, …)
  - Horizontally scalable
- **Daemons**
  - Orchestrates the collaborative work
    e.g., transfers, deletion, recovery, policy
  - Horizontally scalable
- **Messaging**
  - STOMP / ActiveMQ-compatible
- **Persistence**
  - Object relational mapping
  - Oracle, PostgreSQL, MySQL/MariaDB, SQLite
- **Middleware**
  - Connects to well-established products,
    e.g., FTS3, DynaFed, dCache, EOS, S3, …
- **Python**
  - Clients: 2.6, 2.7, 3
  - Server: 2.7, 3

Monitoring & analytics



- RucioUI
  - Provides several views for different types of users
  - Normal users: Data discovery and details, transfer requests and monitoring
  - Site admins: Quota management and transfer approvals
  - Central administration: Account / Identity / Site management
- Monitoring
  - Internal system health monitoring with Graphite / Grafana
  - Transfer / Deletion / … monitoring built on HDFS, ElasticSearch, and Spark
  - Messaging with STOMP
- Analytics and accounting
  - e.g., Show which the data is used, where and how space is used, …
  - Data reports for long-term views
  - Built on Hadoop and Spark

Future developments

- Generic & arbitrary metadata support
- Workload aware system components
  - Auto-scaling depending on load
- Multi-experiment data management features on shared infrastructures
- Quality of Service - Following the evolution of storage
  - Declarative Data Management based on QoS
- Expand support for commercial cloud providers
  - Transparent Google Cloud integration showed good results
- Capability-based authentication and authorisation
  - Bearer tokens, Sci-Tokens, Macaroons, OpenID, EduGain
- Event level data management
  - Include events and event metadata into Rucio - `rucio download <event>`

XDC Main Releases

- A second major release is foreseen before the end of the project
  - XDC Message bus implementation
  - full orchestration
  - finalize integration of RUCIO
  - secure storage in Onedata
  - finalize the ECRIN Use Case
  - complete caching reference workflows with HTTP based systems

	Release Date	End of Full Updates	End of Standard Updates	End of Security Updates & EOL
XDC 1	Jan 2018	May 2019	Sep 2019	Nov 2019
XDC 2	Oct 2019	Apr 2020	Jul 2020	Sep 2020

XDC products can be downloaded from XDC repositories or from each components upstream repositories after they have been pushed back

Conclusion

- XDC is adding new functionalities to already existing, production quality, data management services
- XDC-1/Pulsar was released in January 2019
  - A step towards the complete implementation of the defined architecture
  - Research communities can already start implementing their use cases using Pulsar
- A second release is foreseen by next October
- Scalability verification is in progress and will be one of the core activities in 2019
- XDC consortium members will act as service providers to facilitate the uptake of the XDC services by the EOSC communities
  - We are involving external service providers to increase the uptake of new user communities

XDC Contacts

Giacinto DONVITO
XDC – Technical Coordinator
INFN
donvito@infn.it

- Website: www.extreme-datacloud.eu
- **@XtremeDataCloud on Twitter**
- **Mailing list: info<at>extreme-datacloud.eu**