ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

ESCAPE prototypes a Data Infrastructure for Open Science

R. Bolton (SKAO), S. Campana (CERN), A. Ceccanti (INFN-CNAF), X. Espinal (CERN), P. Fuhrmann (DESY), A. Fkiaras (CERN), Y. Grange (ASTRON)

Computing in High Energy Physics, Adelaide, November 2019

ESCAPE

European Science Cluster of Astronomy & Particle physics ESFRI research infrastructures

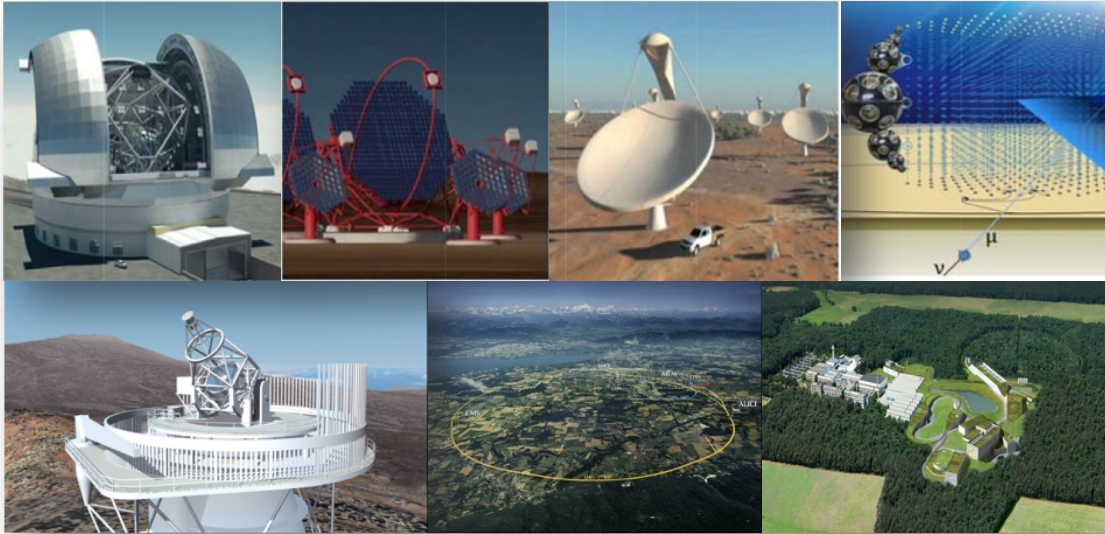Horizon 2020 funded project

EUROPEAN OPEN SCIENCE CLOUD

Goals:

Prototype an infrastructure adapted to the Exabyte-scale needs of the large science projects.
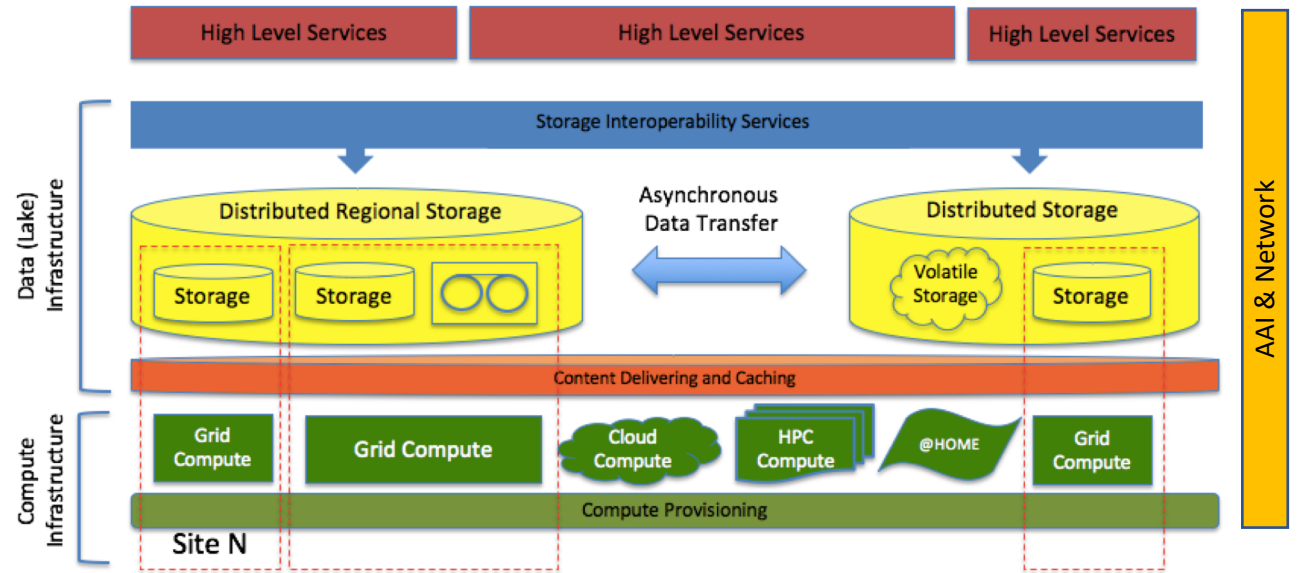
Ensure the sciences drive the development of the EOSC

Address *FAIR* data management

Science Projects

HL-LHC	SKA
FAIR	CTA
KM3Net	JIVE-ERIC
ELT	EST
EURO-VO	EGO-VIRGO
(LSST)	(CERN,ESO)

Data centres: CERN, INFN, DESY, GSI, Nikhef, SURFSara, RUG, CCIN2P3, PIC, LAPP, INAF

WP2 – Data Infrastructure for Open Science

In synergy and complementing the work of WLCG DOMA: to define, integrate and commission an ecosystem of tools and services to build a data lake

Leaves to the science projects the flexibility to choose the services and layout most suitable to their needs. Provides a reference implementation

Contributes to deliver Open Access and FAIR data services: relies on trustable data repositories; enables data management policies; hides the complexities of the underlying infrastructure providing a transparent data access layer

WP2 important milestones

Today

Jul 2019	Jan 2020	Jul 2020	Jan 2021	Jul 2021	Jan 2022	Jul 2022
6	12	18	24	30	36	42

Implementation plan and the design of the of the pilot

Initial DataLake with at least 3 centers

Analysis and assessment of the first version of the pilot DataLake
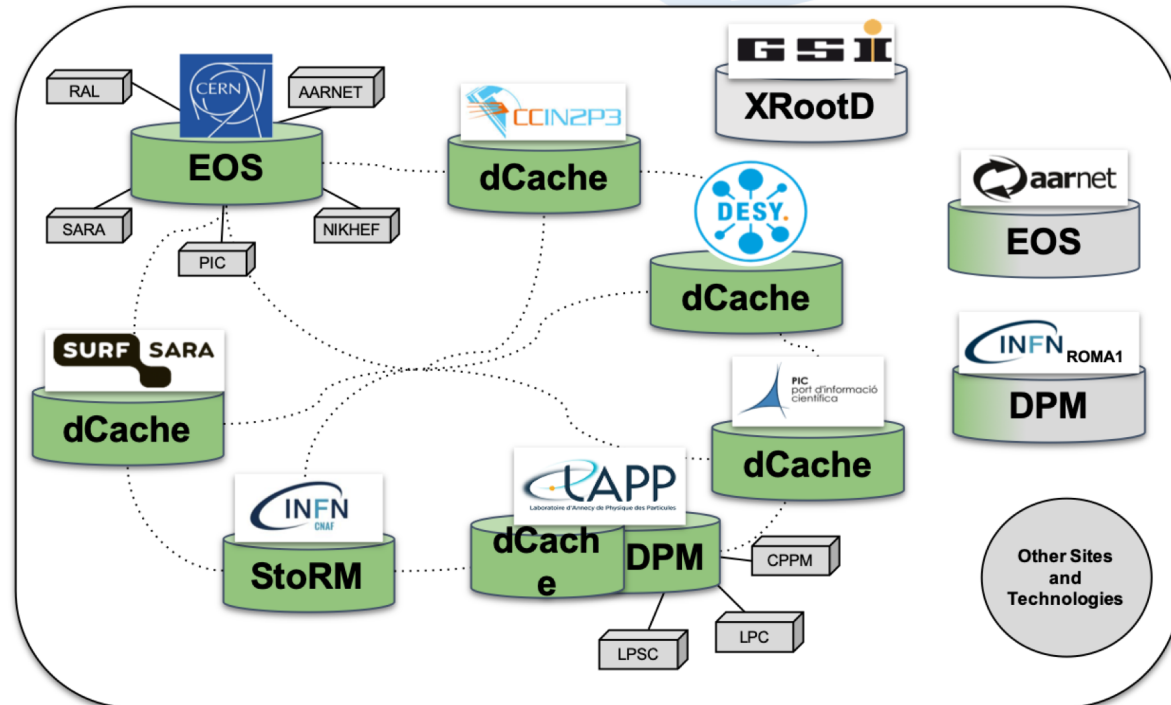
Extended Prototype with more centres and tools

Final analysis and assessment of full prototype
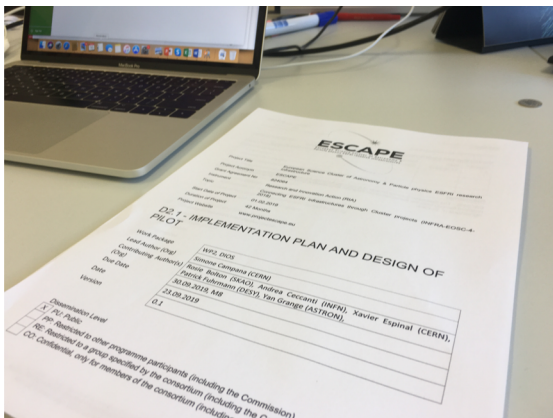
Storage Technologies

EOS
DPM
dCache
StoRM
XRootD ☐

http
xroot
gsiftp

Currently deployed (**green**) and foreseen (**grey**) storage services in ESCAPE DataLake.
➤ Some centers joining even if not funded by ESCAPE!

An heterogeneous set of technologies, deployment models and Quality-Of-Service

Asynchronous Data Transfer

https://fts3-pilot.cern.ch:8449/fts3/ftsmon/#/?vo=**escape**
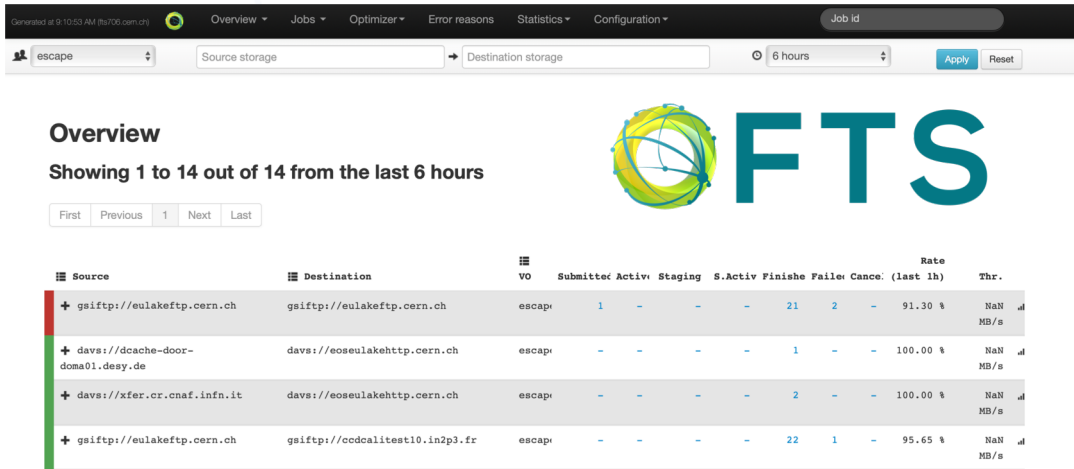
FTS is the workhorse for asynchronous point-to-point data transfer in the reference implementation

Planning for a gridFTP-free data lake: HTTP and xrootd

perfSONAR to provide low level network monitoring
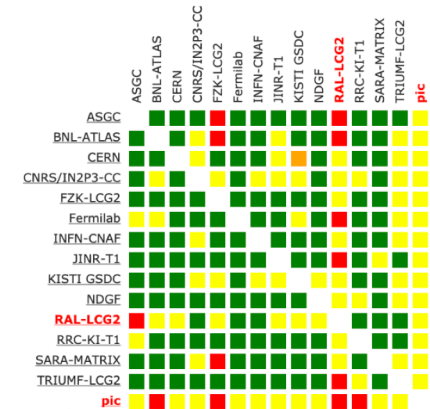
This is the WLCG OPN mesh. We should build an ESCAPE one

Orchestration Service

Rucio as the Orchestration service in the reference implementation

➤ File/dataset catalog, rule based engine

Rucio WebUI as interface for ESCAPE users
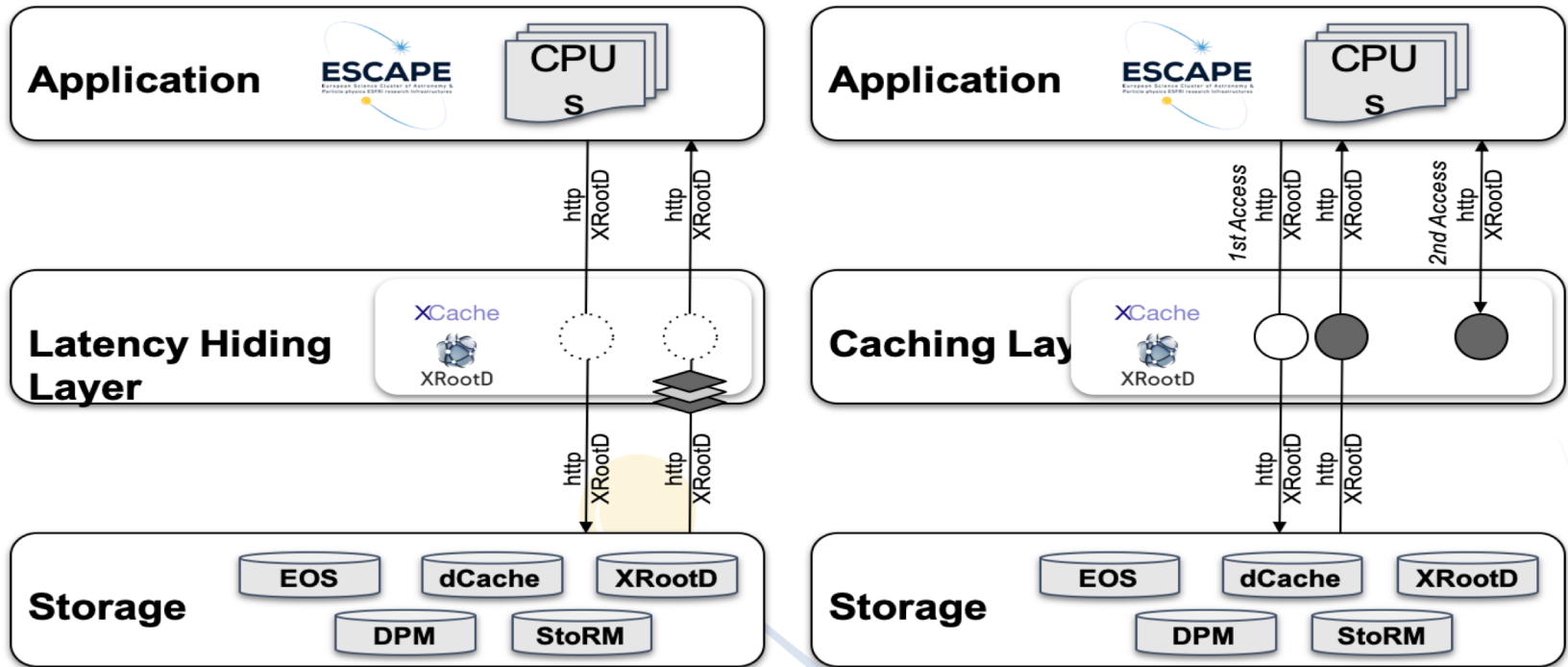

Functional Tests & Visualization
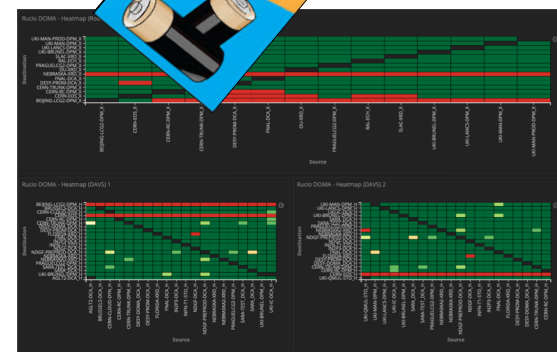
Caching and Latency hiding
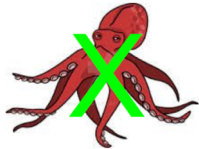
xCache technology as reference implementation.

Support HTTP and xrootd protocols

Authentication, Authorization, Identity

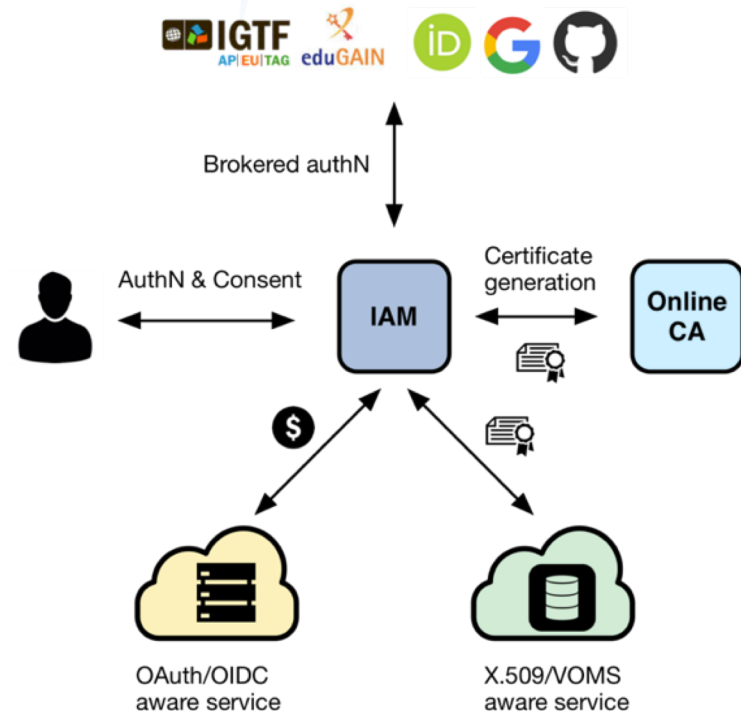AAI has a much broader scope than WP2

IAM was chosen as reference implementation service for AAI

Token-based authentication rather than X509 as baseline. IAM enables translation X509-to-Token for coexistence period

X509-free DataLake is not realistic in the timescale of ESCAPE (IMHO)

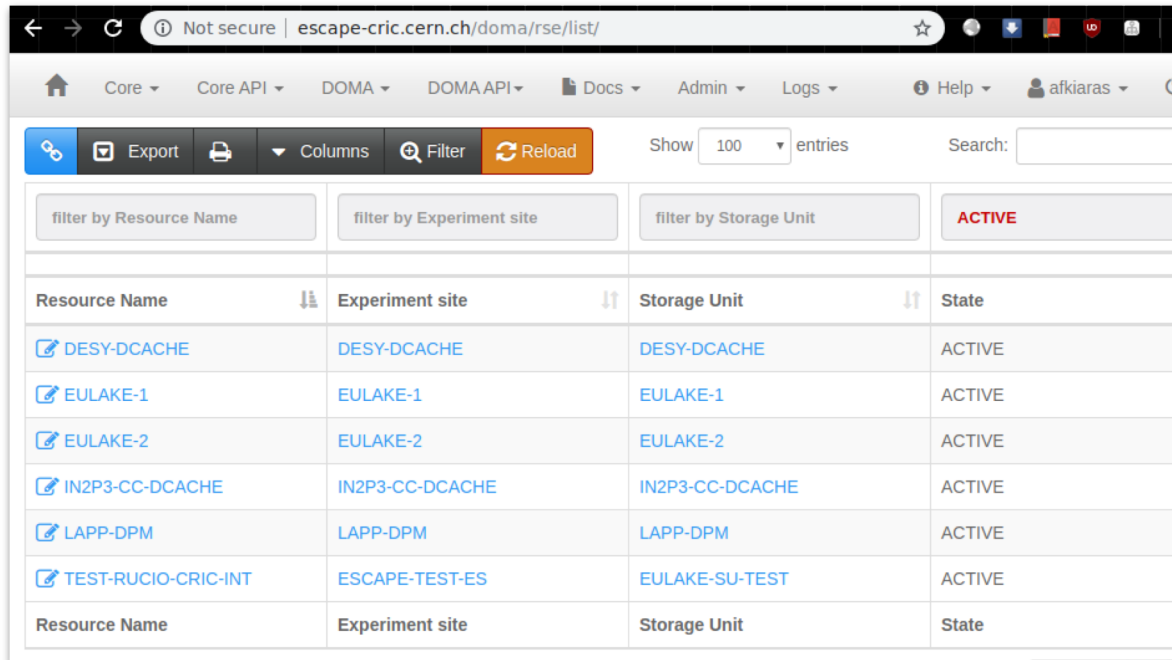Some fully token based use cases however are in range

Leveraging the work in the Federated Identity Management For Research (FIM4R) initiative

Configuration

CRIC as the DataLake configuration service

➢ In use in WLCG, integrated with Rucio

Commissioning

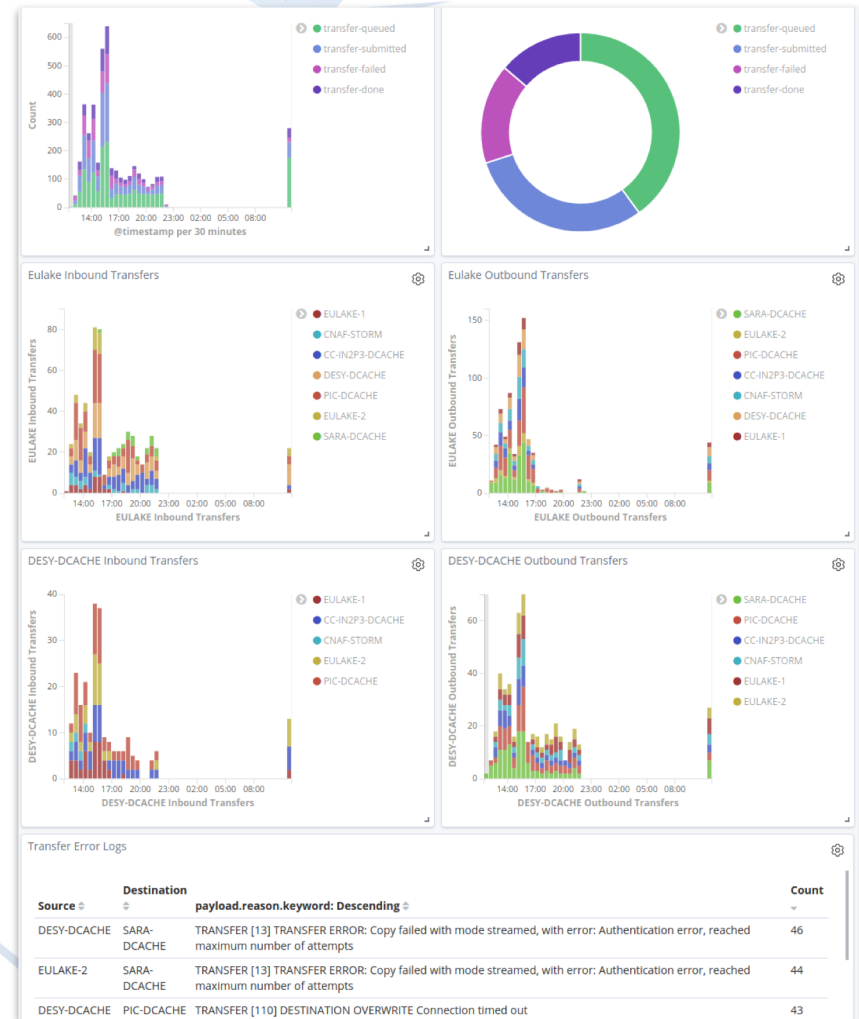Hammercloud for testing and commissioning

➢ Can run applications in distributed environments, through different plugins
➢ Comes with monitoring and analytics
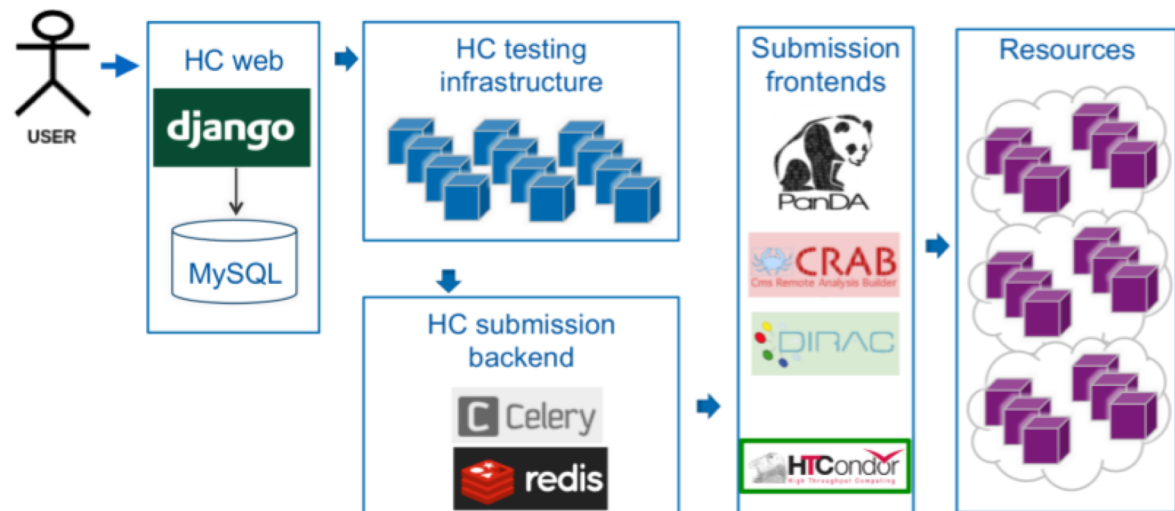
Monitoring

Plan to build an ESCAPE DataLake dashboard, based on open source tools

Most of the information (Rucio, FTS) are already collected in ElasticSearch and can be exposed e.g. via Kibana

Conclusions

As part of the ESCAPE project we are contributing to prototype a Data Infrastructure for the large volume of Open Science data in the 2020s

Building on the collaboration between High Energy and Nuclear Physics and Astronomy, driven by the sciences.

In synergy and leveraging existing projects and solutions