Analysis Facilities & DOMA

A. Forti

HSF Analysis Ecosystem Workshop

23 May 2022

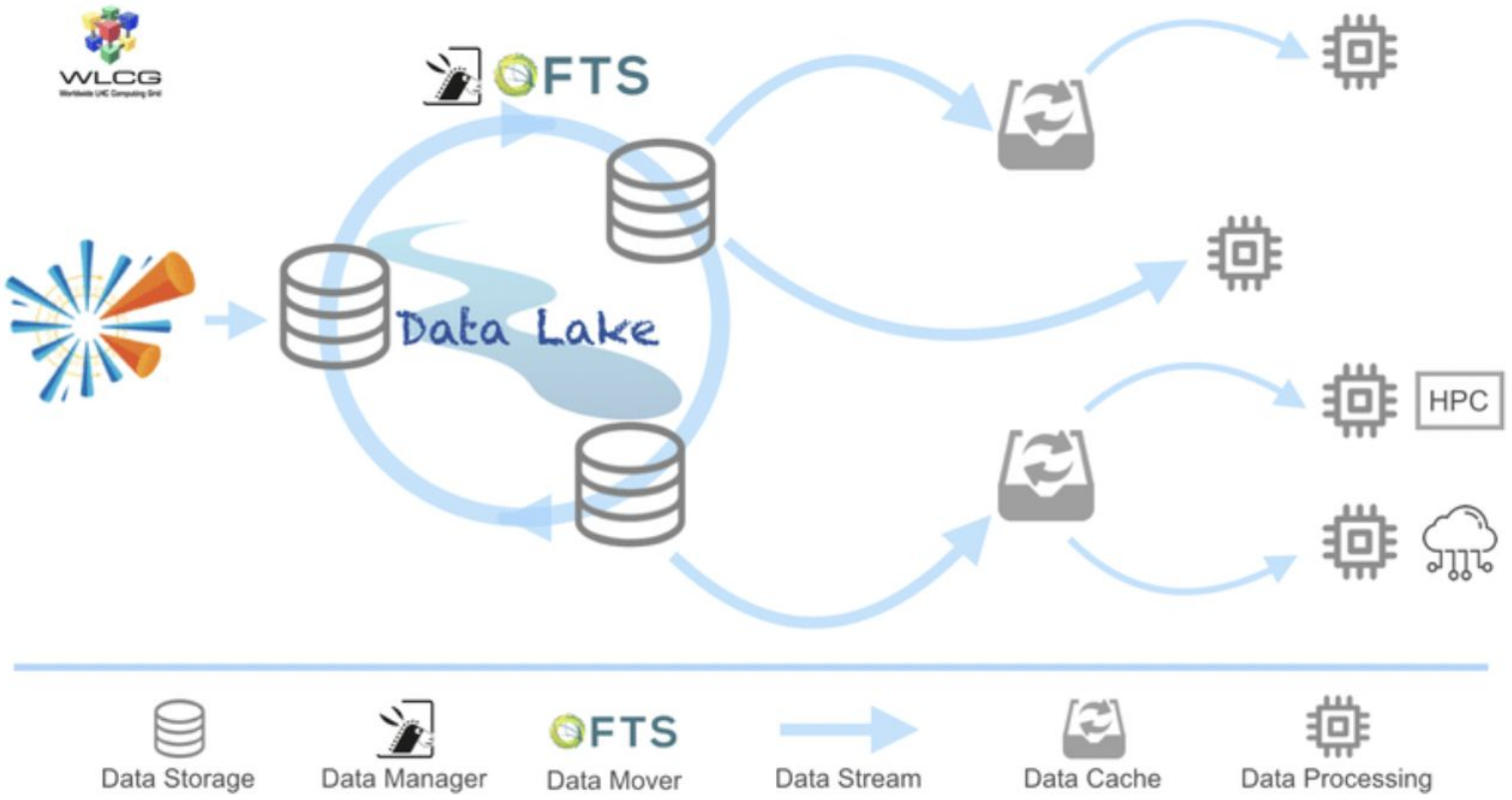
In the beginning….

- The grid was built on distributed independent fat and not so fast commodity storages with moderate QoS strategies and with POSIX access.
  - This is not enough anymore with the progressive expansion of resources and diversification of workflows.
- The latency kept low by sending the jobs where the data is.
  - Sometimes the data is not there…. can the job go somewhere else or the data arrive to the job?
- Grid also built around a unique data format for all the experiments that can be analysed only within one of the available HEP specific frameworks (either experiment or root)
  - Formats and tools differentiating need to cater for new workflows

DOMA Evolution

- HL-LHC scale of data has initiated a revision of this model introducing DOMA activities to look at
  - Network (latency/bandwidth/network orchestration),
  - Caches (hide latency, protect source storages, use diskless sites)
  - Storage QoS (reliability, vs cost vs performance)
  - Storage Deployment and Operation models (full storage, vs caches, vs remote access)
  - Access protocols (gridftp vs https, root, s3)
  - AAI (x509 vs tokens)
  - Storage type (object stores vs posix)
  - Data formats (data transformation and delivery services)
  - Introduction of DataLake and CDN concepts
- So what has all this to do with the Analysis Facilities?
  - AFs use a lot of these concepts to optimise data access for new workflows

DOMA Data Lake



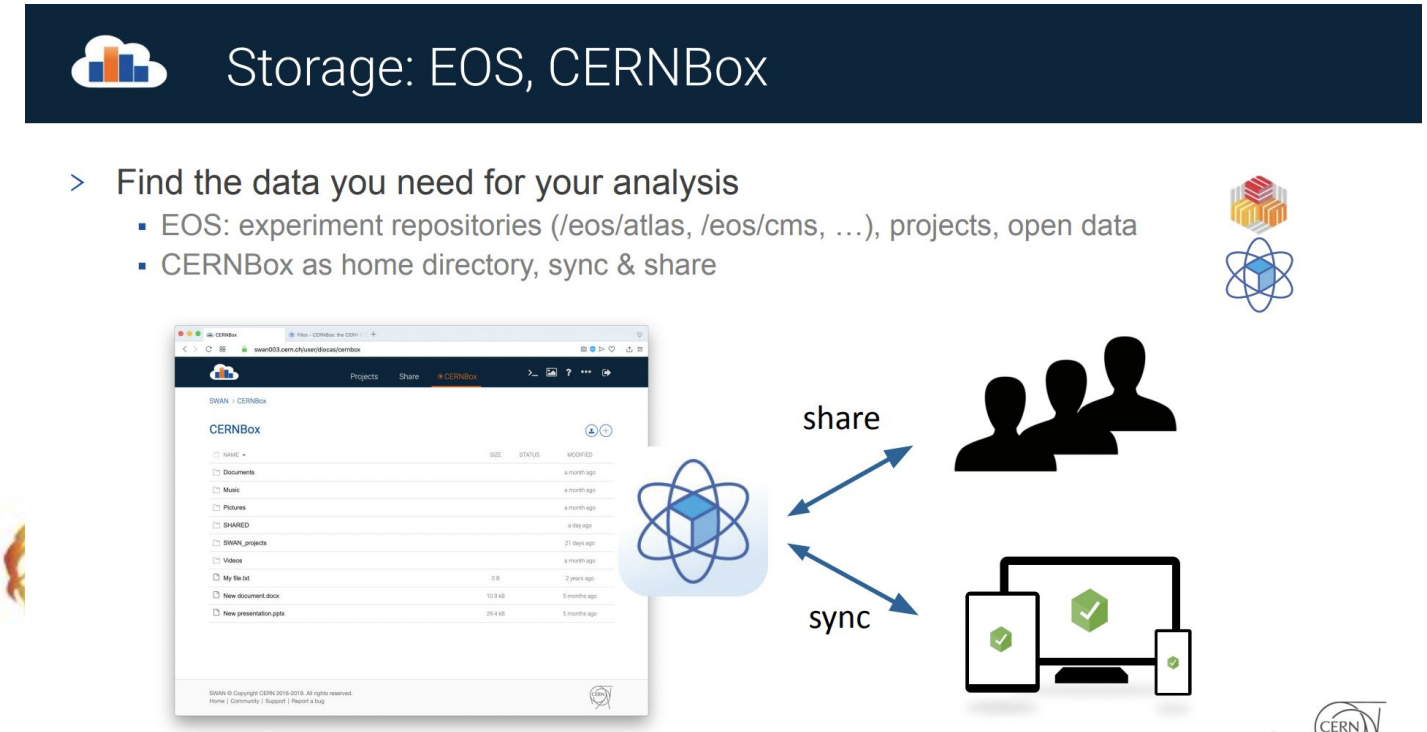
DOMA Access, 2020

Analysis Facilities

- HSF AF definition of analysis facility is quite loose
  - *Infrastructure and services that provide integrated data, software and computational resources to execute one or more elements of an analysis workflow. These resources are shared among members of a virtual organization and supported by that organization.*
- Many types of deployment which are part of the conversation
- 3 types which are emerging
  - Large sites with all data local CERN
  - Distributed among set of sites with caches USATLAS
  - Remote access of storage providers ESCAPE
- Excluding because of requirement of shared resources
  - T3 private
  - HPC sort of private
  - Cloud accessible via WFMS
  - some solutions like Xcache work for these cases too

CERN Lake

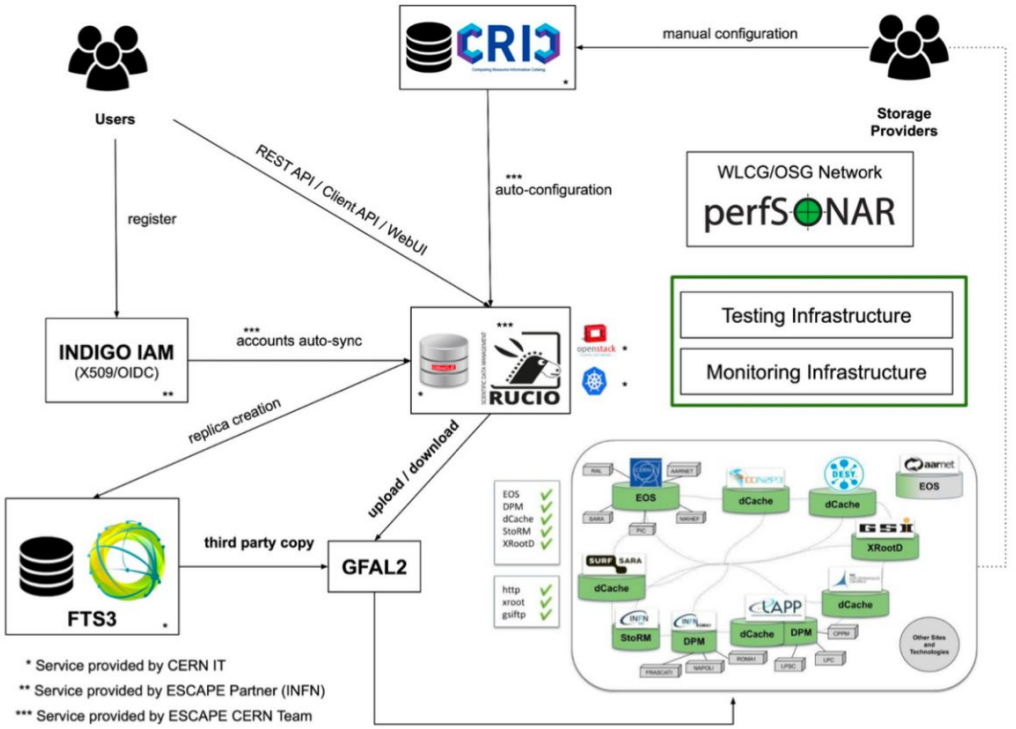
- Interactive, jupyter notebooks, batch system….
- EOS storage accessible by all the systems
  - Experiment data, user output, software sharing any type of data or file they can use CERN box
  - EOS closer to the commercial vendors definition of data lake as a “centralised repository of (un)structured data”
  - Recurring question can this be replicated?



- Part of the ScienceBox (EOS, CERNBox, SWAN) which is installable either as single box or on kubernetes on larger scale

ESCAPE Data Lake

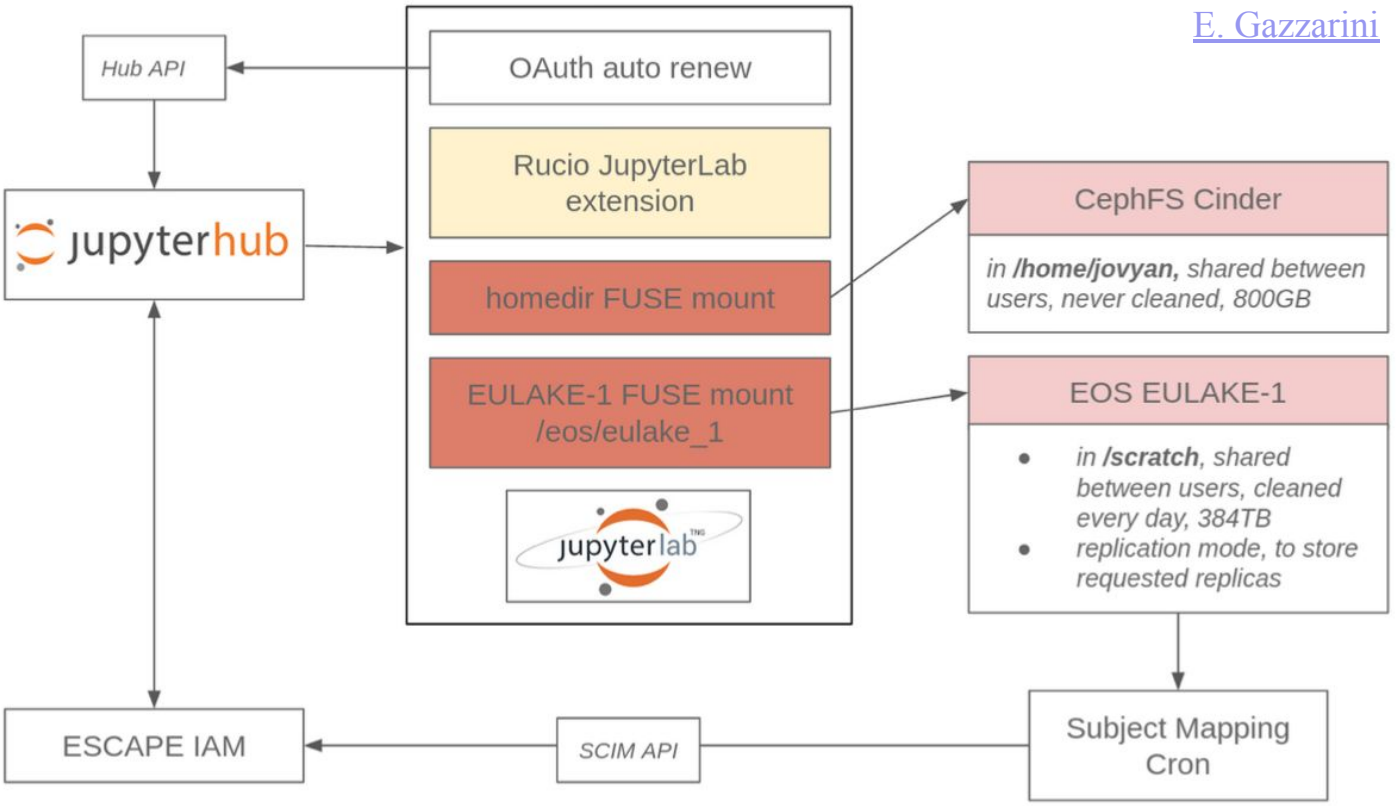
- Concepts evolved from DOMA.
  - \+ FTS traffic monitoring infrastructure
    - Reworked later in DOMA DC
  - \+ Network monitoring (perfsonar)
  - \+ IAM and token integration



E. Gazzarini

ESCAPE DLaaS

E. Gazzarini



Escape DLaaS: jupyterhub users have a rucio plugin and can pull their data in a scratch space that is cleaned every day

US shared T3s

- USATLAS is setting up a shared T3 at 3 sites
  - BNL, SLAC, UChicago
  - All ATLAS users are allowed access to all 3 facilities
- An attempt to offer AF resources to the whole collaboration
  - Users need to register at each site
- And the sites are independent in terms of resources and storage
  - Tools specific for users analysis shared ia CVMFS
  - Xcaches at each site allows to soften the fact that not all data might be at the site of choice by simply caching it
  - No possibility to share things across but access is granted so people can for now move from one to the other
- CMS is also opening their US facilities to CERN SSO
  - INFN-T1 also

Xcache

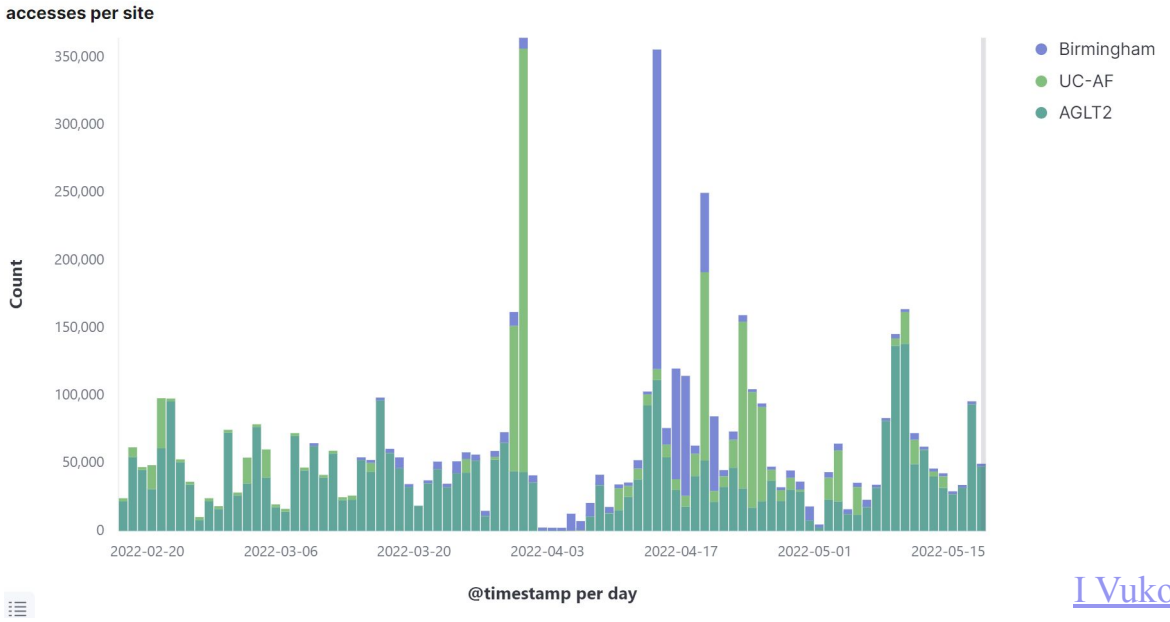
- Analysis highly repetitive data access pattern
  - Adding an Xcache between the user and the data has positive aspects
  - Cost less/GB
    - Only small caches are needed and non redundant hardware
  - Reduces latency / increase CPU eff.
  - Costs less manpower to run

I Vukotic transfer rates after first access

Other uses of Xcache

Virtual Placement is a service used by ATLAS to enhance the use of Xcaches on the grid by combining the information the caches can give to the service with what rucio knows about the data and the topology.

Xcache so configured help handle the AF spiky nature of data access



I Vukotic

ServiceX with Xcache

- Xcache can also be used by ServiceX
- SerivceX is a data transformation service.
  - It transforms more standard HEP data in a format more suitable to columnar analysis and can also perform data reduction
  - It is also integrated with rucio via a DID discovery service which finds the data on the lake using rucio.
  - It also benefits from having an Xcache behind
- It is one of the services that will be tested at scale during the IRIS-HEP AGC but recent numbers
  - About 15 min to transform 3TB of input data into 3GB of columnar output (R. Gardner)
    - Still leaves a lot of questions open when the number of users goes up.
- Can also be combined with VP to have the DID service produce the final file URL which includes the cache.

Cross contamination

- Type of interactions we should have
  - Jupyter rucio plugin from ESCAPE DLaaS.
  - Xcache at US AF
  - This is could be also an interesting combination to explore in the future getting the rucio plugin to work with VP and maybe expand ESCAPE DLaaS with Xcaches
- CMS interested in using ScienceBox
- ……

Last slide

- There is a lot of scope for cooperation between different projects
- Many DOMA concepts have been evolving for few years adding bits and pieces to the diagrams circulating.
- We are starting to see the changes appear at sites on the grid and at AF.
- Some more mainstream users starting to look at new workflows
  - They'll need support
    - also to test AFs tools and services
- AGC an important target to test things particularly for what concerns data access and delivery at existing facilities