



ESCAPE

European Science Cluster of Astronomy &
Particle physics ESFRI research Infrastructures

CMS report

D. Ciangottini on behalf of the CMS team

Outline

- Target scenario: recap
 - Prototyping a CMS Analysis Facility at INFN
 - Compute
 - Data access
- Tested scenario
 - Dress rehearsal:
 - CMS open data import into data lake
 - Data discovery through Rucio
 - Recap: what do we have
- Looking forward to...

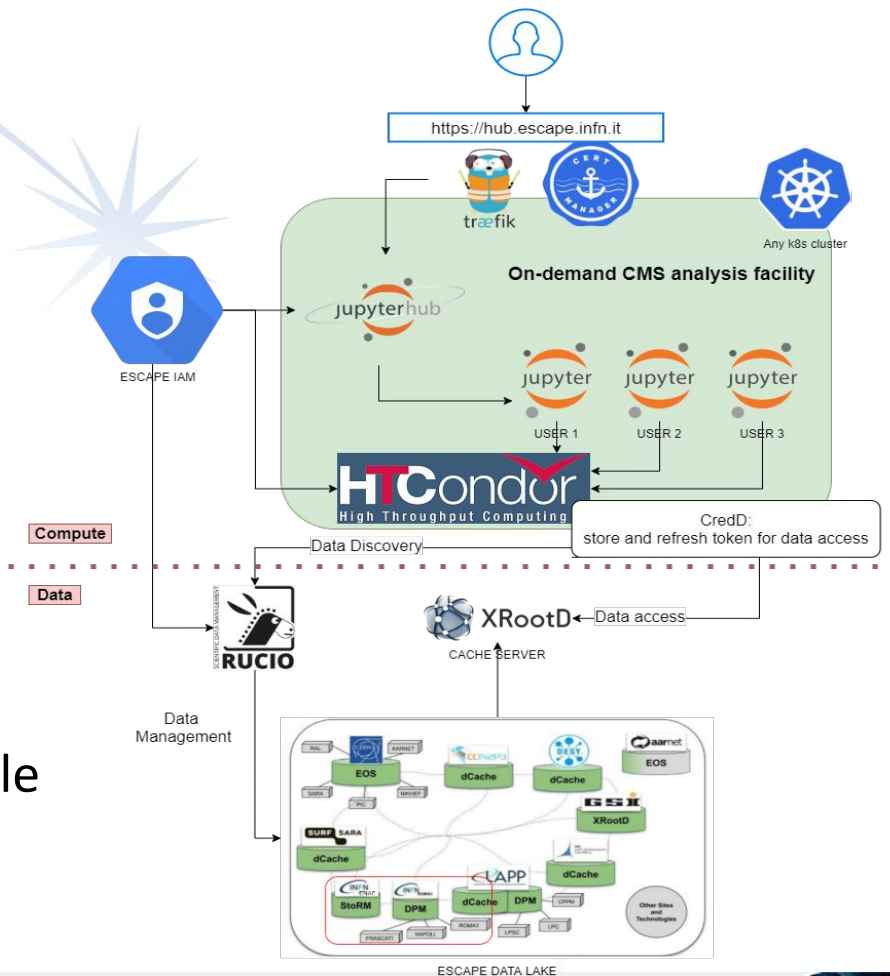


CMS evaluation: recap

- Objectives:
 - On demand analysis facility to deploy on any k8s cluster
 - Stateless and reproducible
 - IAM based AuthN/Z for both data and compute resources access
 - X509 free
 - Support both interactive and batch workflows

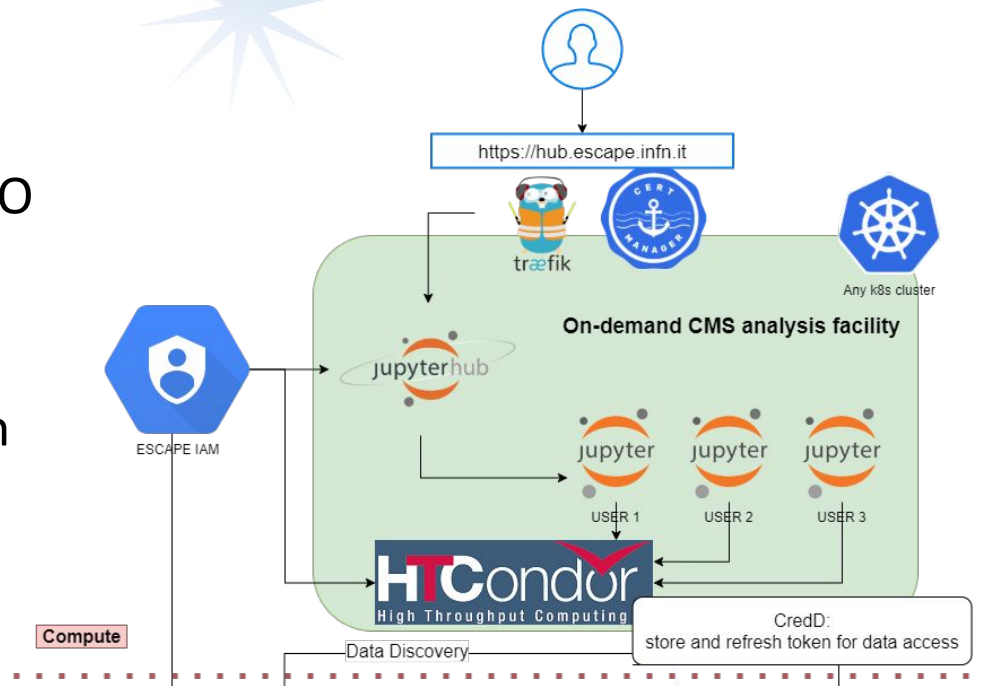
Bonus: keep it as much general as possible

Sharable approach with other communities, e.g. in the context of the INFN cloud federation



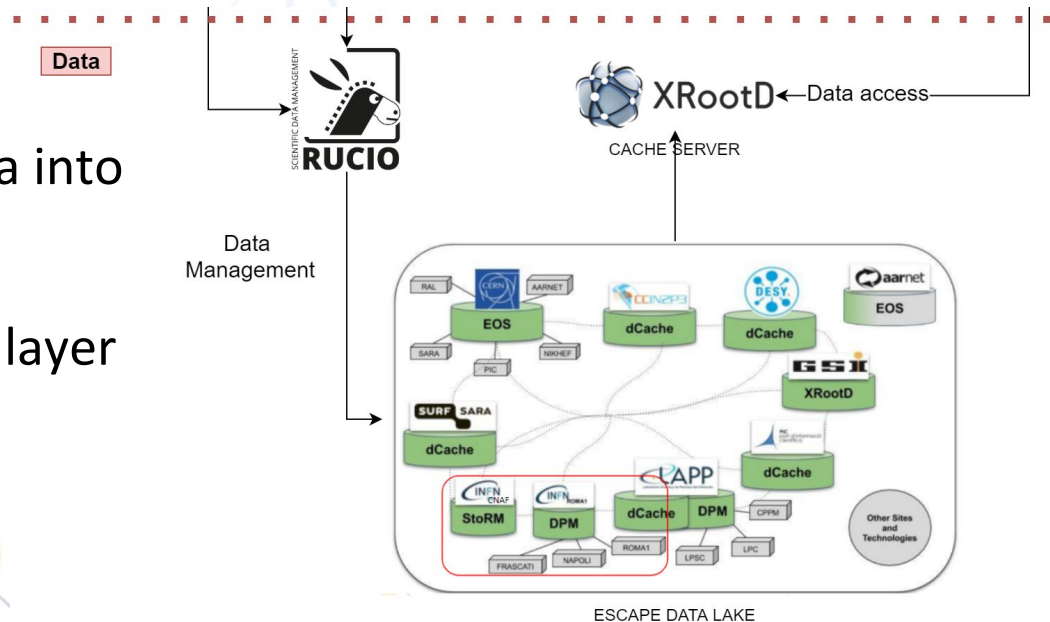
On-demand compute

- Provide software packed into Helm chart
- Multi user and possibly multi VO
 - Access IAM managed
- User containers already configured to access HTCondor batch and remote data through token



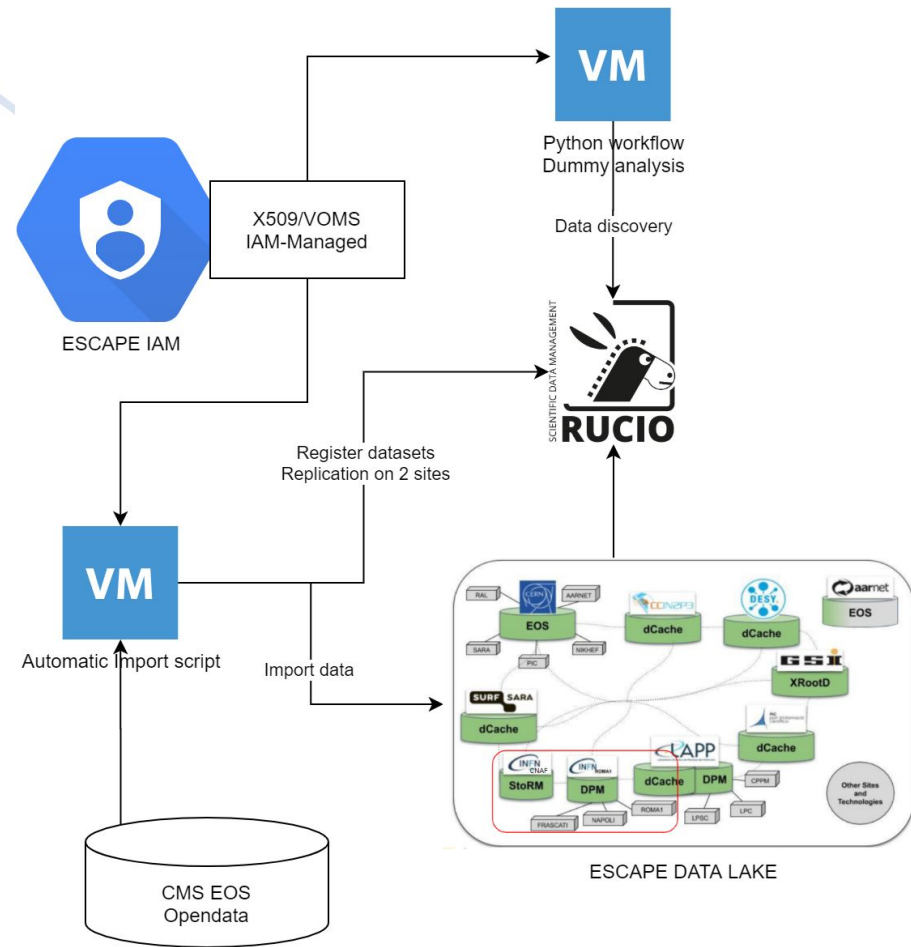
Data access and caches

- Importing and managing data into the data lake
- Rucio as data discovery tool
- Data access through a cache layer for better performances



Dress rehearsal scenario

- Open data as first easy scenario
- Useful for data import and discovery mechanics validation
- X509 based, but OK for now
 - Although in the hybrid approach of the IAM based Voms
 - Intended just for a gradual transition to no X509
- Simple demo workflows (also due to old legacy format of open data)



Results of the exercise

- Exercise code:
 - https://github.com/dciangot/rucio_escape_inject
- O(300GB) open **data correctly imported** and analyzed using rucio for **data discovery** of the location.
- 2 imports at different time done, to test the append data to dataset functionality
- Checksum bug (on Rucio side AFAIU) found when trying to upload to STORM CNAF
 - <https://github.com/rucio/rucio/issues/4127>

```
$ rucio rule-info f479b8893f4c4701a1e887f23e47f260
Id: f479b8893f4c4701a1e887f23e47f260
Account: dciangot
Scope: CMS_INFN_DCIANGOT
Name: opendata_test
RSE Expression: CNAF-STORM
Copies: 1
State: OK
Locks OK/REPLICATING/STUCK: 380/0/0
Grouping: DATASET
Expires at: 2020-11-21 20:09:16
Locked: False
Weight: None
Created at: 2020-11-17 16:09:16
Updated at: 2020-11-17 18:02:14
Error: None
Subscription Id: None
Source replica expression: None
Activity: User Subscriptions
Comment: None
Ignore Quota: False
Ignore Availability: False
Purge replicas: False
Notification: NO
End of life: None
Child Rule Id: None
```

Import in EULAKE1

```
2020-11-17 18:17:12,305 - simple_example - INFO - Successfully uploaded file 68F89B12-1C70-E011-A893-E0CB4E1A11A2.root
2020-11-17 18:17:12,318 - simple_example - DEBUG - Registering file
2020-11-17 18:17:12,338 - simple_example - DEBUG - Skipping dataset registration
2020-11-17 18:17:12,338 - simple_example - INFO - DID: CMS_INFN_DCIANGOT opendata_cms_Run2010B_Mu_AOD_Apr21ReReco-v1_0002_68F89B12-1C70-E011-A893-E0CB4E1A11A2.root
2020-11-17 18:17:12,401 - simple_example - INFO - Successfully added replica in Rucio catalogue at EULAKE-1
2020-11-17 18:17:12,483 - simple_example - INFO - Successfully added replication rule at EULAKE-1
```

Conclusions

- Importing data into the data lake succeeded without major issues
 - Also replication based on QoS label was OK
- Data discovery for an example payload worked as well
- Looking forward to:
 - **X509-free RUCIO**
 - **Embargoed data** in data lake behind **IAM token authN/Z**
 - As presented in the past, functionalities are there and functionally tested (also for caches)
 - Real workflows performance studies **reading embargoed data from different Qos endpoints** and cache layers