CMS plans for DAC21

D. Ciangottini (INFN) on behalf of the CMS-ESCAPE team

ESCAPE - The European Science Cluster of Astronomy & Particle Physics ESFRI Research Infrastructures has received funding from the European Union's Horizon 2020 research and innovation programme under the Grant Agreement n° 824064.

Overall objectives

Demonstrate the **interactions with data-lake of CMS workflows**

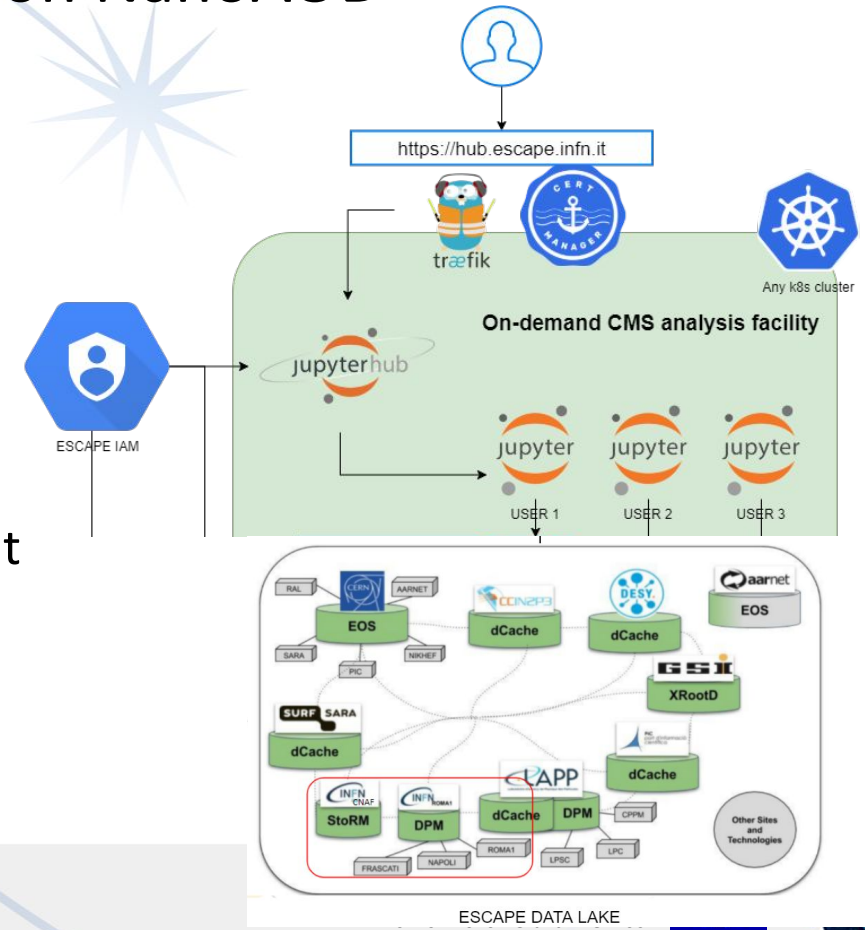
- **Input:** OpenData (possibly embargoed)
- **Access to data** via:
  - task on a batch system
  - Interactive workflow (jupyterlab)
    - Dynamic scale over batch system with Dask framework
- **Upload outputs** back to the lake
  - Simple plots
  - Reduced datasets

In other words: Reproducing and managing base plots and reduced "Ntuples" starting from data on the lake via different workflow and infrastructures

Funded by the European Union's Horizon 2020 - Grant N° 824064

Interactive CMS analysis on NanoAOD

- **NanoAOD opendata O(100GB)**→ **plain ROOT files**→ simple python **ROOTDataFrame** as framework
- We will perform the analysis on a JupyterLab instance on a dedicated JupyterHUB hosted at CNAF
- We target opendata, but, of course, in terms of volume embargoed will be a bit more… “realistic”
- **QoS:** we can imagine a simulation of a scenario where inputs are on FAST and output on CHEAP-ANALYSIS

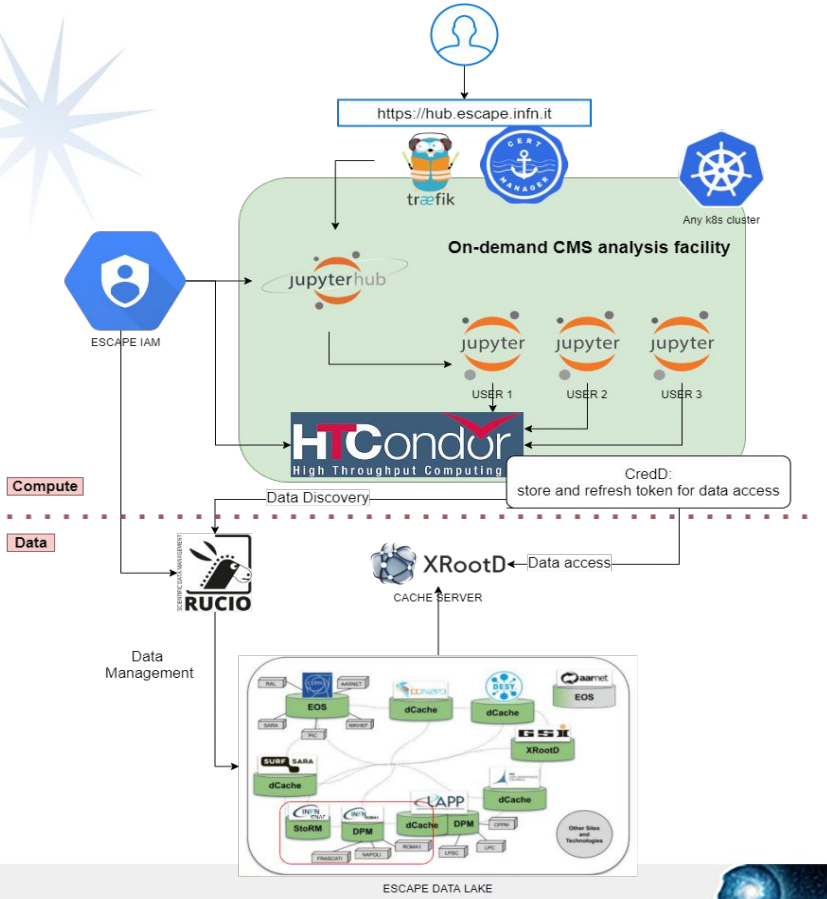


Access data through jobs running on a HTCondor cluster

Demonstrate distributed access
performing NanoAOD analysis via
HTCondor cluster (hosted at CNAF as well)

Focus on:

- **testing the propagation of the user credentials** (both x509 and token, but clearly more interest would be on latter)
- **Collaboration XCache integration** (see later)

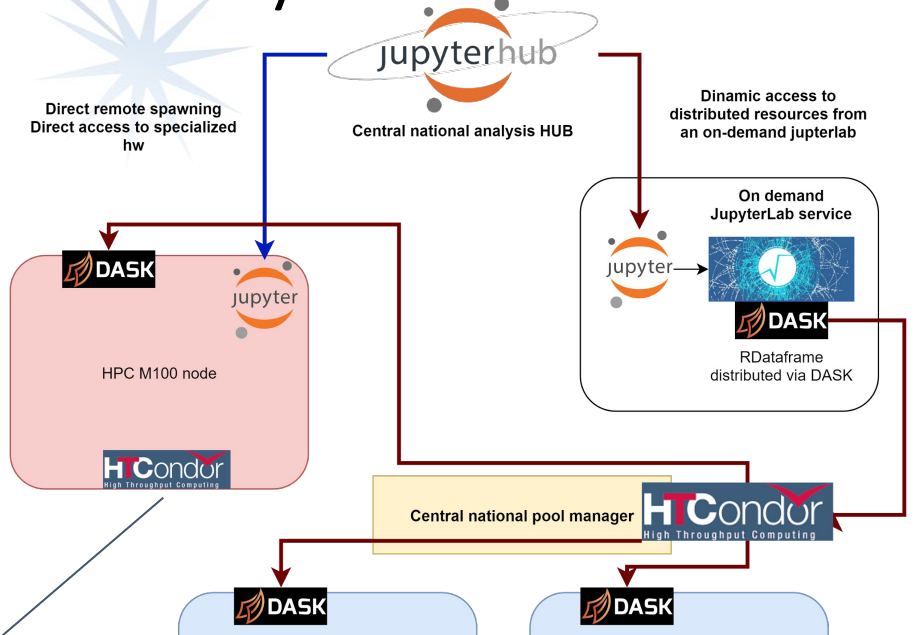


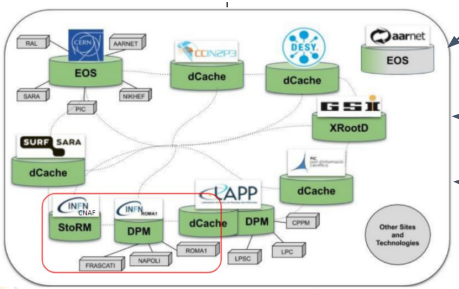
Funded by the European Union's
Horizon 2020 - Grant N° 824064

All together:
Interactive extension over batch systems with Dask

Adopting Python and ROOTDataFrame is possible to allow **scaling out over a batch system (in our case condor) thanks to Dask**

This will mainly enable a realistic simulation of a possible future CMS analysis infrastructure scenario





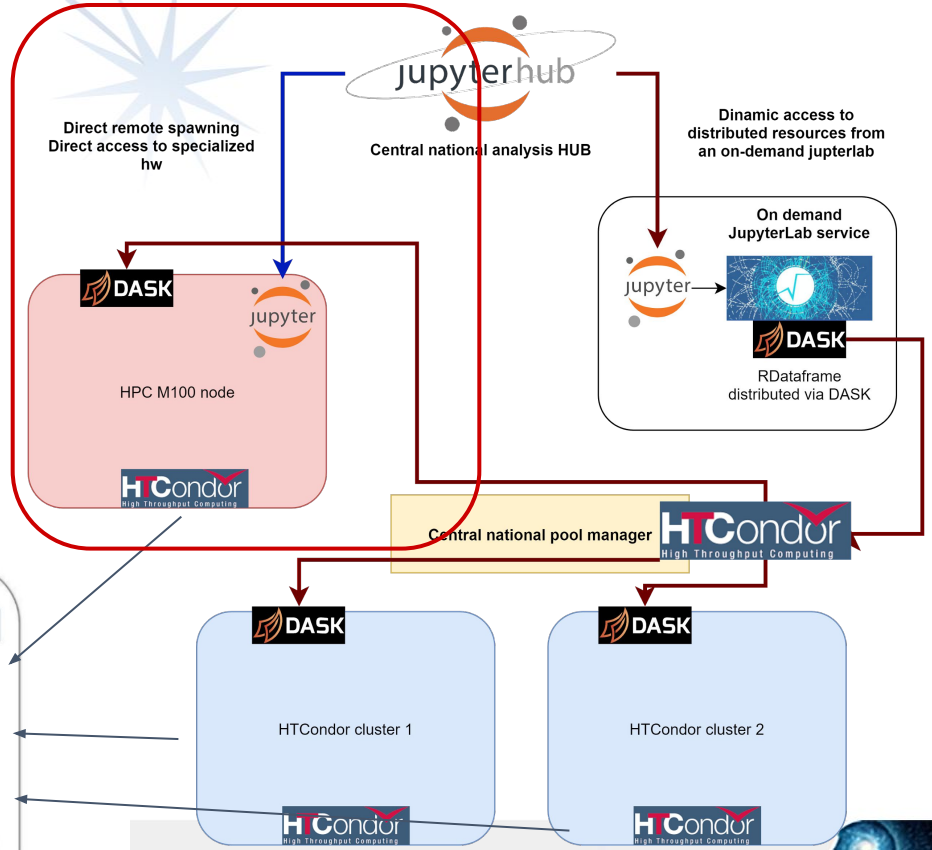
ESCAPE DATA LAKE

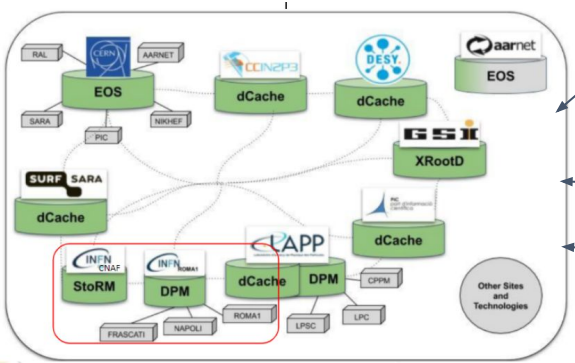
Funded by the European Union's Horizon 2020 - Grant N° 824064

HPC integration

Moreover, in the same system we have the possibility (depending of available slots) to **run on HPC Marconi100 a few tests:**

- With spawn of a JupyterLab directly on slurm
- With extension via Dask as shown before





Funded by the European Union's Horizon 2020 - Grant N° 824064

XCache integrations

We'd like to explore yet few other scenarios:

- **Group managed XCache**
  - The cache server authenticate with the lake and than it's open (no auth, or custom auth) to the clients on a certain network
- **Protocol translation cache** (e.g. reading xrootd and serving http or vice versa)
- (optional) XRootD(s) protocol authn via token

Funded by the European Union's Horizon 2020 - Grant N° 824064

Conclusions

Just to remark our interest in testing/helping with this scenario where through tokens we could be able to access on CMS embargoed data

Every task presented can potentially be repeated on embargoed data as well!

Task tables: https://wiki.escape2020.de/index.php/CMS_DAC21

Actions in summary:

- Infrastructure and code for task 1 and 2 is ready
- To do:
  - Upload and registration of input data (testing QoS as well)
  - Verifying the possibility to have slots available for HPC tests
  - Integrate XCache servers into the infrastructure

Funded by the European Union's Horizon 2020 - Grant N° 824064

BACKUP

Funded by the European Union's
Horizon 2020 - Grant N° 824064

Input data list

/eos/root-eos/cms_opendata_2012_nanoaod/Run2012B_DoubleElectron.root

/eos/root-eos/cms_opendata_2012_nanoaod/Run2012B_DoubleMuParked.root

/eos/root-eos/cms_opendata_2012_nanoaod/Run2012C_DoubleElectron.root

/eos/root-eos/cms_opendata_2012_nanoaod/Run2012C_DoubleMuParked.root

/eos/root-eos/cms_opendata_2012_nanoaod/SMHiggsToZZTo4L.root

/eos/root-eos/cms_opendata_2012_nanoaod/ZZTo2e2mu.root

/eos/root-eos/cms_opendata_2012_nanoaod/ZZTo4e.root

/eos/root-eos/cms_opendata_2012_nanoaod/ZZTo4mu.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/DYJetsToLL.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/GluGluToHToTauTau.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/Run2012BC_DoubleMuParked_Muons.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/Run2012B_TauPlusX.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/Run2012C_TauPlusX.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/TTbar.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/VBF_HToTauTau.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/W1JetsToLNu.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/W2JetsToLNu.root

/eos/opendata/cms/derived-data/AOD2NanoAODOutreachTool/W3JetsToLNu.root

Funded by the European Union's Horizon 2020 - Grant N° 824064