



CI pipeline triggering analysis execution on Analysis Facility

Matteo Bartolini¹, **Mattia Lizzo**¹, **Lorenzo Viliani**¹, **Tommaso Tedeschi**²

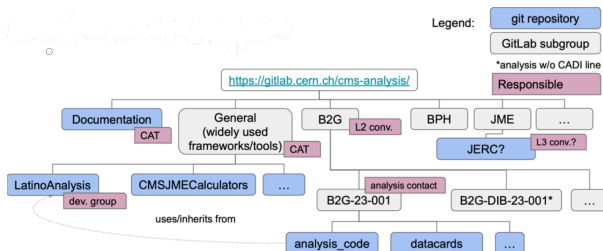
¹Università e INFN, Firenze

²Università e INFN, Perugia

Spoke2 annual meeting

18-20 December 2023

Continuous integration with CMS dataset



- CMS has established a group for the management of analysis tools (CAT)
- CAT provides an area on gitlab.cern.ch for analysis code with instructions to disseminate the use of gitlab CI:
 - Issue: difficult, because authentication is often needed in order to access the dataset → CERN gitlab CI runners typically don't have it
 - Solution: Use IAM token to authenticate on AF and voms proxy to access the datasets

How do we do it?

These are the steps:

- Create a standalone docker image of the needed software stack containing all the libraries needed to run it
 - In our case we eliminated the dependency on the CMS software stack**
- have a separate repository with the analysis specific software
- The image is used by the analysis code running on the CI runners everytime a new commit is made.
- The CI runners offload the condor jobs to workers running on the Analysis Facility
- The workers on the [Analysis Facility](#) will also run the docker image of the framework and perform all the operations

CI for software stack container

- To build a docker image of your framework the **CI tool** is used
- This project has the objective to supply an easy to use gitlab CI template to build images

```
.gitlab-ci.yml 432 B
Edit Lock Replace Delete
1 stages:
2 - build
3
4 include:
5 - project: 'ci-tools/container-image-ci-templates'
6   file:
7     - 'kaniko-image.gitlab-ci.yml'
8
9 variables:
10 CONTEXT_DIR: ""
11 DOCKER_FILE_NAME: "Dockerfile"
12 GIT_SUBMODULE_STRATEGY: recursive
13 PUSH_IMAGE: "true"
14 ACCELERATED_IMAGE: "false"
15 BUILD_ARGS: ""
16 SCAN_IMAGE: "false"
17 REGISTRY_IMAGE_PATH: ${CI_REGISTRY_IMAGE}
18
19 add_to_image:
20 extends: .build_kaniko
21 stage: build
22 tags:
23 - cvmfs
```

CI for analysis specific code

- The docker image is loaded every time a commit is made in the analysis code

```
.gitlab-ci.yml 2.17 KIB [Edit] [Lock] [Replace] [Delete] [Icons] [Download]
1 default:
2   image:
3     name: gitlab-registry.cern.ch/Lenzip/mkshapesrdf
4     entrypoint: ["/bin/sh", "-c"]
5
6 test:
7   tags:
8     - cvmfs
9
10  before_script:
11    - source /code/start.sh
12    - source /code/fix_xrdafs.sh
13
14  script:
15    - . .gitlab/init_infra_AF_token.sh
16    - ls /ca.crt
17    - condor_q
18    - condor_q -debug
19    - printf $proxy | base64 -d > myproxy
20    - export X509_USER_PROXY=$(pwd)/myproxy
21    - export X509_CERT_DIR=/cvmfs/cms.cern.ch/grid/etc/grid-security/certificates/
22    - source /code/fix_xrdafs.sh
23    - echo $X509_USER_PROXY
24    - echo $X509_CERT_DIR
25    - xrdafs root://eoscms.cern.ch ls /eos/cms/store/group/phys_higgs/cmshww/amassiro/HWWNano/Summer20UL18_106x_nA0C
26    - voms-proxy-info
27    - root -l -q root://eoscms.cern.ch//store/group/phys_higgs/cmshww/amassiro/HWWNano/Run2018_UL2018_nA0Dv9_Full20
28    - which checkCondor
29    - ls -a
30    - cd Full2017_v9
31    - ls -a
32    - mkShapesRDF -c 1
33    - ls -a
34    - condor_q
```

The Analysis Facility token

- Access to the AF resources is done via token (using IAM instance)

```
1 IAM_TOKEN_ENDPOINT=https://cms-auth.web.cern.ch/token
2
3 #IAM_USER=dciangot
4
5 result=$(curl -s -L \
6   -d client_id=${IAM_CLIENT_ID} \
7   -d client_secret=${IAM_CLIENT_SECRET} \
8   -d grant_type=client_credentials \
9   -d username=${IAM_CLIENT_ID} \
10  -d password=${IAM_CLIENT_SECRET} \
11  -d scope="openid profile offline_access wlcg" \
12  ${IAM_TOKEN_ENDPOINT})
13
14 if [[ $? != 0 ]]; then
15   echo "Error!"
16   echo $result
17   exit 1
18 fi
19
20
21
22 access_token=$(echo $result | jq -r .access_token)
23 refresh_token=$(echo $result | jq -r .refresh_token)
24
25 echo $access_token > my_access_token
26
27 export _condor_SCHEDD_NAME=131.154.96.124.myip.cloud.infn.it
28 export _condor_SCHEDD_HOST=131.154.96.124.myip.cloud.infn.it
29 export _condor_COLLECTOR_HOST=131.154.96.124.myip.cloud.infn.it:30618
30 export _condor_SCITOKENS_FILE=$(pwd)/my_access_token
31 export _condor_AUTH_SSL_CLIENT_CAFILE=/ca.crt
32 export _condor_SEC_DEFAULT_AUTHENTICATION_METHODS=SCITOKENS
33 export _condor_TOOL_DEBUG=D_FULLDEBUG,D_SECURITY
```

Final result

cms-analysis > ... > WpWmJJ_polarizations > analysis_code > Jobs > #33957811

Search job log

```
4682 hadd Target path: merged_output.root:/top_2j_em/mjj
4683 hadd Target path: merged_output.root:/DY_2j_em
4684 hadd Target path: merged_output.root:/DY_2j_em/events
4685 hadd Target path: merged_output.root:/DY_2j_em/dnn_isVBS
4686 hadd Target path: merged_output.root:/DY_2j_em/dnn_isLL
4687 hadd Target path: merged_output.root:/DY_2j_em/dnn_isTT
4688 hadd Target path: merged_output.root:/DY_2j_em/dnn_TTvsLL_49
4689 hadd Target path: merged_output.root:/DY_2j_em/dnn_TTvsLL_36
4690 hadd Target path: merged_output.root:/DY_2j_em/dnn_TTvsLL_25
4691 hadd Target path: merged_output.root:/DY_2j_em/dnn_TTvsLL_16
4692 hadd Target path: merged_output.root:/DY_2j_em/dnn_TTvsLL_16v2
4693 hadd Target path: merged_output.root:/DY_2j_em/mjj
4694 $ rm -rf mkShapes__RDF*.root
4695 $ ls -a
4696 .
4697 ..
4698 merged_output.root
✓ 4700 Uploading artifacts for successful job 00:01
4701 Uploading artifacts...
4702 Full2017_v9/root_output/*.root: found 1 matching artifact files and directories
4703 Uploading artifacts as "archive" to coordinator... 201 Created id=33957811 resp
onseStatus=201 Created token=64_z3YTr
✓ 4705 Cleaning up project directory and file based variables 00:00
4707 Job succeeded
```

Finished: 6 days ago
Queued: 2 seconds
Timeout: 1h (from project) ⓘ
Job ID: #33957811
Runner: #33539 (z-wxdVgFU)
runners-k8s-cvmfs-
runners-58d5567dbf-
qsg6m

Tags: [cvmfs](#)

Job artifacts ⓘ

These artifacts are the latest. They will not be deleted (even if expired) until newer artifacts are available.

[Keep](#) [Download](#) [Browse](#)

Commit [c5235038](#) ⓘ

Update [.gitlab-ci.yml](#) file

Pipeline [#6481799](#) passed for
master ⓘ

test

Related jobs

→ test

Conclusions

- We have been setting up a CI pipeline running a full analysis on an INFN analysis facility
- We overcame the initial struggles with authentication and tokens
- Detailed instructions will be thoroughly documented and made available
- Job submission is entirely based on condor at the moment, but we plan to start experimenting soon the use of dask to improve handling and merging of the full dataset
- It would be nice to understand if we can test this using the cineca infrastructure
- The authentication methods used so far will change in the future → using solutions for credential protections