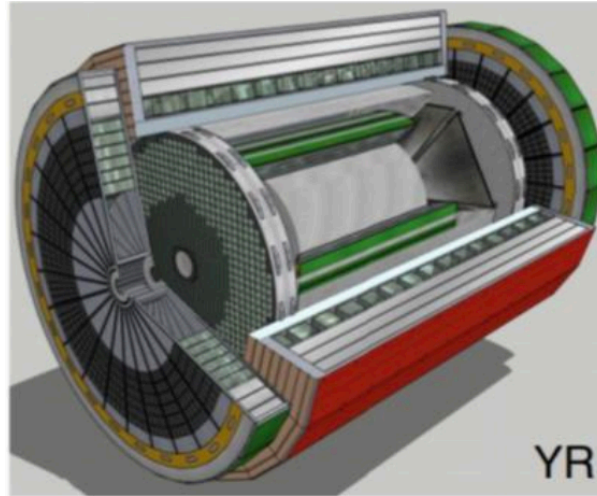


Running fun4all via PanDA



Everything on a Post-it

```
pip3 install panda-client  
# start from  
~eickolja/scratch/spacktest/example_workflows/fun4all-  
ecce/panda_fun4all  
# adapt, maybe using locate panda_setup.sh and  
source env.sh #or env.csh  
# On mac, install ssl certs inside the python3 bundle  
# adapt paths and  
python3 ./submit_to_panda.py  
# Once every 24 hours, follow link to panda-iam-doma.cern.ch  
# Monitor on  
https://panda-doma.cern.ch/tasks/?display\_limit=300
```

Preparation

- Get python3
- Get panda-client

```
$ pip3 install panda-client
```

- Don't *need* pip

```
$ spack load py-setuptools  
$ git clone git://github.com/PanDAWMS/panda-client.git  
$ python3 setup.py install --prefix=~/.install
```

- Adapt and source env.[c]sh

```
setenv PANDA_URL_SSL https://ai-ids-01.cern.ch:25443/server/panda  
setenv PANDA_URL http://ai-ids-01.cern.ch:25080/server/panda  
setenv PANDA_AUTH oidc  
setenv PANDA_VERIFY_HOST off  
source /direct/eic+u/eickolja/install/etc/panda/panda_setup.csh  
setenv PANDA_AUTH_VO EIC
```



```
$ source env.csh
```

Prepare Submission Script

- Hard-coded paths → Can be submitted from anywhere

```
import sys
from pandatools import Client
import os
import stat

OUTPUT_DIR = '/gpfs02/eic/eickolja/scratch/submittest/fun4all-ecce/panda/work'
SOURCEDIR = '/gpfs02/eic/kkauder/scratch/spacktest/example_workflows/fun4all-ecce'
PAYLOAD_DIR = SOURCEDIR+'/payload'
EXECUTABLE = SOURCEDIR+'/panda_fun4all/RunFun4allPanda.sh'
```

- They do need to exist, with the right permissions
→ prepare at some point with (some version of)

```
os.makedirs(OUTPUT_DIR, exist_ok=True)
os.chmod(OUTPUT_DIR, 0o775)
os.chmod(EXECUTABLE, 0o775)
```

See below for why this matters

Prepare Task

```
task_command_line = "{EXECUTABLE} {OUTPUT_DIR} {PAYLOAD_DIR}".format(  
    EXECUTABLE=EXECUTABLE, OUTPUT_DIR=OUTPUT_DIR, PAYLOAD_DIR=PAYLOAD_DIR)
```

```
taskParamMap = {  
    taskParamMap['nFilesPerJob'] = 1  
    inFileList = [str(i) for i in range(0, LIMIT)]  
    taskParamMap['nFiles'] = len(inFileList)  
    taskParamMap['noInput'] = True  
    taskParamMap['pfnList'] = inFileList  
    taskParamMap['coreCount'] = 1  
    taskParamMap['ramCount'] = 3000  
    taskParamMap['jobParameters'] = [  
        {'type': 'constant',  
         'value': ""  
        },  
        {'type': 'command',  
         'value': echo ""+task_command_line+"";  
                ""+task_command_line+"";""  
        },  
    ]  
}
```

```
print(Client.insertTaskParams(taskParamMap, verbose=True))
```

Use as a dummy for
multiple jobs

Resource constraints

Could do everything here (in bash);
cleaner with a wrapper script

More Task Parameters

```
taskParamMap['vo'] = 'wlcg'  
taskParamMap['site'] = 'BNL_OSG_1'  
taskParamMap['workingGroup'] = 'EIC'  
taskParamMap['taskName'] = 'test EIC fun4all submission'  
taskParamMap['userName'] = 'A user Name'  
taskParamMap['processingType'] = 'step1'  
taskParamMap['prodSourceLabel'] = 'test'  
taskParamMap['taskType'] = 'test'
```

We can see gpfs (and cvmfs)

```
taskParamMap['transPath'] = 'https://atlpan.web.cern.ch/atlpan/bash-c'  
taskParamMap['taskPriority'] = 900  
taskParamMap['architecture'] = ''  
taskParamMap['transUses'] = ''  
taskParamMap['transHome'] = None  
taskParamMap['skipScout'] = True  
taskParamMap['cloud'] = 'US'
```

- Most of these, I'm not certain of their meaning 😊

Wrapper Script

```
#!/bin/env bash

# Allow group members to change or delete created files
umask 0002

# set up ECCE environment
source /cvmfs/eic.opensciencegrid.org/ecce/gcc-8.3/opt/fun4all/core/bin/ecce_setup.sh -n

# define various convenience variables
[...]

# change to Working Directory
cd $WRKDIR

# And do the work
root -l -b -q ${PAYLOAD_DIR}/Fun4All_G4_EICDetector.C \
    1> ${WRKDIR}/${NAMEBASE}.out 2> ${WRKDIR}/${NAMEBASE}.err

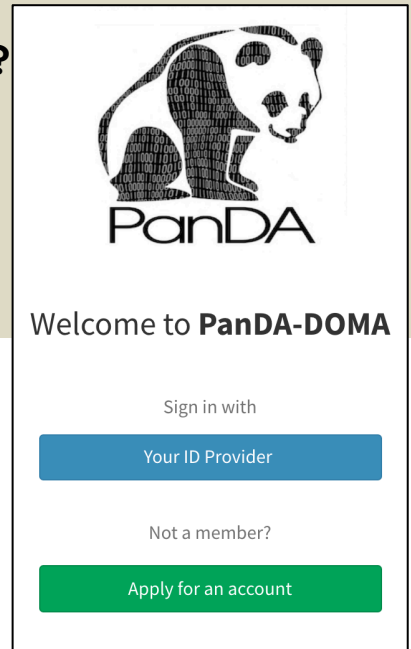
# Could move or copy files to destinations here, via
# (mc) mv, cp
```

See below for why this matters

- Mostly just busy-work for output redirection

Submit

```
$ python3 ./submit_to_panda.py
DEBUG : cached token unavailable
DEBUG : getting device code
DEBUG : b'{"user_code":"XXXXX","device_code":"XXXXXXXXXX",
"verification_uri_complete":"https://panda-iam-
doma.cern.ch/device?user_code\u003dXXXXX","verification_uri":"https://pand
a-iam-doma.cern.ch/device","expires_in":1800}'
INFO : Please go to https://panda-iam-doma.cern.ch/device?
and sign in. Waiting until authentication is completed
INFO : Ready to get ID token?
[y/n]
y
DEBUG : getting ID token
```



Need to get a token every 24 hours

- This is the step where you need an account
- Lab ID works, but you need to be added

Monitor

https://panda-doma.cern.ch/tasks/?display_limit=300

300 tasks, sorted by jedidtaskid

ID	Task name	Task status	Input files	Total/Remaining events	Modified	State changed	Priority	Nucleus	Cloud
Parent	TaskType/ProcessingType Campaign Group User Errors Logged status	Nfiles	finish% fail% Nfinish Nfail						
1892	test EIC fun4all ECCE submission anal /step1 EIC Kolja Kauder Errors	done 1	100% 1		2021-05-11 18:21:27	2021-05-11 18:21:27			
1891	test GOOGLE GKE submission anal /step1 Shuwei Ye Errors	failed 1	100% 1		2021-05-11 19:21:34	2021-05-11 19:21:34			

1 job files

Filename (Type)	Scope	Size (MB)
000000:0 (pseudo_input)		0.00

Other key job parameters

Last state change	to failed at 2021-04-30 18:03:15
Transformation exit code	1 (non-zero exit code from job payload)
Working group	EIC
Attempt number	3 of a maximum 3
CPU consumption time (s)	2
Job parameters	source /cvmfs/sphenix.sdcc.bnl.gov/gcc-8.3/opt/sphenix/core/bin/sphenix_setup.sh -n new; echo /gpfs/n/gpfs02/eic/eickolja/scratch/submittest/fun4all/panda/; /gpfs/mnt/gpfs02/eic/kkauder/scratch/spacktest/e

Logs

- No logs available
- Pilot job stdout
- Pilot job stderr
- job stdout
- Pilot job jdl
- Pilot records
- Action logger (Kibana)
- Open all logs

Sergey can say much much more!

Output

```
$ ls -lt ~/scratch/submittest/fun4all-ecce/panda/work/
total 7169
-rw-rw-r-- 1 osgeic 69997 20216 May 11 14:14 fun4all-ecce.out
-rw-r--r-- 1 osgeic 69997 17934 May 11 14:14 G4EICDetector_g4tracking_eval.root
-rw-r--r-- 1 osgeic 69997 10351 May 11 14:14 g4fwdjets_ClusterFwd_10_eval.root
[...]
-rw-r--r-- 1 osgeic 69997 10756 May 11 14:14 g4fwdjets_full_05_eval.root
-rw-r--r-- 1 osgeic 69997 10591 May 11 14:14 g4fwdjets_track_05_eval.root
-rw-r--r-- 1 osgeic 69997 12899 May 11 14:14 G4EICDetector_g4jet_eval.root
[...]
-rw-r--r-- 1 osgeic 69997 55631 May 11 14:14 G4EICDetector_g4cemc_eval.root
-rw-rw-r-- 1 osgeic 69997      0 May 11 14:13 fun4all-ecce.err
```

- Task writes directly to gpfs
- Everything belongs to the group account
- **Only one (maybe a few) production tsars can do this**
- **May need some afterburner to turn over ownership**
- One option: Production tsar also gives osgeic S3 account?

Discussion Topics

- Better option (on a good timescale) for ownership?
- Is there need for different access to pilot job stdout?
- Can you send code and extract output without the job seeing gpfs?
- Priority – Requested 900, monitor shows 1000
 - Also, how does it currently coexist with preferred EIC slots?
 - Are there automatic ways to keep things fair if IP6, others ramp up production?
- Note, can be steered via Jupyter Notebook (not tried)

All code here is at

`/gpfs02/eic/kkauder/scratch/spacktest/example_workflows/fun4all-ecce/panda_fun4all`

and soon on https://github.com/eic/example_workflows