How to access and decode ProtoDUNE-II data

Julio Ureña

IFIC - Valencia

Waffles tutorial - 5 December 2024

Outline

The steps we will go through are

1) Create a DAQ-Build-Tools (DBT) environment

   It is needed to have access to the DAQ libraries which are used, in our case, to read the PDS data from the DAQ output files (HDF5) into Waffles objects (Waveform, WaveformSet etc.)

2) Install Waffles within your DBT environment

   Waffles *[1]* readers need access to the DAQ libraries

3) Setup rucio

   Rucio is *[2]* a Data Replica service to get files from Rucio Storage Elements (RSE) around the world. RSEs are provided by collaborating institutions *[3]*. Most DUNE users are now enabled to use it. New users may not automatically be added.

4) Get an XRootD ticket

   RSEs are grid-accessible *[3]*, i.e. specific tool suites to handle data stored there is required, in our case, XRootD which *[4]* '*gives high performance, scalable fault tolerant access to data repositories of many kinds. The typical usage is to give access to file-based ones*'.

5) Create a WaveformSet object

Starting point

To be able to go through the steps shown in this slides you need to

- have access to lxplus,
- be able to kerberos-authenticate to FNAL.GOV and
- have access to rucio

1. Create a DAQ-Build-Tools (DBT) environment

Navigate to the folder where you will create your DBT environment, then:

```
@lxplus $ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
@lxplus $ setup_dbt latest
@lxplus $ dbt-create -l
@lxplus $ dbt-create fddaq-v4.4.7-a9 <dbt_dir>
@lxplus $ cd <dbt_dir>
@lxplus $ source <dbt_dir>/env.sh
```

```
[jurenago@lxplus9122 tutorial]$ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
[jurenago@lxplus9122 tutorial]$
```

1. Create a DAQ-Build-Tools (DBT) environment

Navigate to the folder where you will create your DBT environment, then:

```
@lxplus $ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
@lxplus $ setup_dbt latest
@lxplus $ dbt-create -l
@lxplus $ dbt-create fddaq-v4.4.7-a9 <dbt_dir>
@lxplus $ cd <dbt_dir>
@lxplus $ source <dbt_dir>/env.sh
```

```
[jurenago@lxplus9122 tutorial]$ setup_dbt latest
Updated /cvmfs/dunedaq.opensciencegrid.org/tools/dbt/v8.3.0/bin -> PATH
Updated /cvmfs/dunedaq.opensciencegrid.org/tools/dbt/v8.3.0/scripts -> PATH
DBT setuptools loaded
[jurenago@lxplus9122 tutorial]$
```

1. Create a DAQ-Build-Tools (DBT) environment

Navigate to the folder where you will create your DBT environment, then:

```
@lxplus $ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
@lxplus $ setup_dbt latest
@lxplus $ dbt-create -l
@lxplus $ dbt-create fddaq-v4.4.7-a9 <dbt_dir>
@lxplus $ cd <dbt_dir>
@lxplus $ source <dbt_dir>/env.sh
```

```
[jurenago@lxplus9122 tutorial]$ dbt-create -l
 - current
 - fddaq-v3.1.4-a9
 - fddaq-v3.1.4-a9-1
 - fddaq-v3.1.4-a9-2
 - fddaq-v4.1.0
 - fddaq-v4.1.0-a9
 - fddaq-v4.1.1
 - fddaq-v4.1.1-a9
 - fddaq-v4.2.0
 - fddaq-v4.2.0-a9
 - fddaq-v4.2.0-c8
 - fddaq-v4.2.1-a9
 - fddaq-v4.2.1-c8
```

.
.
.

DUNE DEEP UNDERGROUND NEUTRINO EXPERIMENT

IFIC INSTITUT DE FÍSICA CORPUSCULAR

1. Create a DAQ-Build-Tools (DBT) environment

Navigate to the folder where you will create your DBT environment, then:

```
@lxplus $ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
@lxplus $ setup_dbt latest
@lxplus $ dbt-create -l
@lxplus $ dbt-create fddaq-v4.4.7-a9 <dbt_dir>
@lxplus $ cd <dbt_dir>
@lxplus $ source <dbt_dir>/env.sh
```

```
[jurenago@lxplus9122 tutorial]$ dbt-create fddaq-v4.4.7-a9 dbt
Release "fddaq-v4.4.7-a9" requested; interpreting this as release "fddaq-v4.4.7-a9-1"
Setting up the Python subsystem.

This script is calling "spack load systems@coredaq-v4.5.7-a9-1"; it will print "Finished loading"
on successful completion.

If this is the first time the "spack load ..." command has been run in
a while on this node it may take ~15 minutes; this is because cvmfs is
populating its local cache. Please be patient; subsequent runs should
take less than a minute.

Finished loading
```

.
.
.

1. Create a DAQ-Build-Tools (DBT) environment

Navigate to the folder where you will create your DBT environment, then:

```
@lxplus $ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
@lxplus $ setup_dbt latest
@lxplus $ dbt-create -l
@lxplus $ dbt-create fddaq-v4.4.7-a9 <dbt_dir>
@lxplus $ cd <dbt_dir>
@lxplus $ source <dbt_dir>/env.sh
```

.
.
.

```
Finished loading

INFO [12/05/24 11:44:11]: creating virtual_env .venv by cloning
/cvmfs/dunedaq.opensciencegrid.org/spack/releases/fddaq-v4.4.7-a9-1/.venv.
Depending on a variety of factors this can take from several seconds to several minutes...

Total time to run /cvmfs/dunedaq.opensciencegrid.org/tools/dbt/v8.3.0/bin/dbt-create: 77 seconds
Start time: Thu Dec 5 11:44:03 CET 2024
End time:   Thu Dec 5 11:45:20 CET 2024

See https://dune-daq-sw.readthedocs.io/en/latest/packages/daq-buildtools for build instructions

Script completed successfully
```

1. Create a DAQ-Build-Tools (DBT) environment

Navigate to the folder where you will create your DBT environment, then:

```
@lxplus $ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
@lxplus $ setup_dbt latest
@lxplus $ dbt-create -l
@lxplus $ dbt-create fddaq-v4.4.7-a9 <dbt_dir>
@lxplus $ cd <dbt_dir>
@lxplus $ source <dbt_dir>/env.sh
```

```
[jurenago@lxplus9122 tutorial]$ cd dbt/
[jurenago@lxplus9122 dbt]$ source env.sh
Updated /cvmfs/dunedaq.opensciencegrid.org/tools/dbt/v8.3.0/bin -> PATH
Updated /cvmfs/dunedaq.opensciencegrid.org/tools/dbt/v8.3.0/scripts -> PATH
DBT setuptools loaded
Work area: '/afs/cern.ch/work/j/jurenago/private/tutorial/dbt'

This script hasn't yet been sourced (successfully) in this shell; setting up the build environment


This script is calling "spack load fddaq@fddaq-v4.4.7-a9-1"; it will print "Finished loading"
on successful completion.
```

.
.
.

DEEP UNDERGROUND NEUTRINO EXPERIMENT

IFIC INSTITUT DE FÍSICA CORPUSCULAR

1. Create a DAQ-Build-Tools (DBT) environment

Navigate to the folder where you will create your DBT environment, then:

```
@lxplus $ source /cvmfs/dunedaq.opensciencegrid.org/setup_dunedaq.sh
@lxplus $ setup_dbt latest
@lxplus $ dbt-create -l
@lxplus $ dbt-create fddaq-v4.4.7-a9 <dbt_dir>
@lxplus $ cd <dbt_dir>
@lxplus $ source <dbt_dir>/env.sh
```

.
.

```
If this is the first time the "spack load ..." command has been run in
a while on this node it may take ~15 minutes; this is because cvmfs is
populating its local cache. Please be patient; subsequent runs should
take less than a minute.

Finished loading

Now loading devtools for latest CMake, gcc, etc.

Found venv in the current workarea, activating it now...

Updating paths...
...done

This script has been sourced successfully
```

2. Install Waffles within your DBT environment

Activate the DBT environment (`source env.sh`), navigate to the folder where you will download waffles, then:

```
@lxplus (dbt) $ cd <repos_dir>
@lxplus (dbt) $ git clone https://github.com/DUNE/waffles.git
@lxplus (dbt) $ cd waffles
@lxplus (dbt) $ pip install -r requirements.txt
@lxplus (dbt) $ pip install .
```

```
(dbt) [jurenago@lxplus9122 repositories]$ git clone https://github.com/DUNE/waffles.git
Cloning into 'waffles'...
remote: Enumerating objects: 3020, done.
remote: Counting objects: 100% (1259/1259), done.
remote: Compressing objects: 100% (404/404), done.
remote: Total 3020 (delta 897), reused 966 (delta 769), pack-reused 1761 (from 1)
Receiving objects: 100% (3020/3020), 4.40 MiB | 6.27 MiB/s, done.
Resolving deltas: 100% (1750/1750), done.
(dbt) [jurenago@lxplus9122 repositories]$
```

2. Install Waffles within your DBT environment

Activate the DBT environment (`source env.sh`), navigate to the folder where you will download waffles, then:

```
@lxplus (dbt) $ cd <repos_dir>
@lxplus (dbt) $ git clone https://github.com/DUNE/waffles.git
@lxplus (dbt) $ cd waffles
@lxplus (dbt) $ pip install -r requirements.txt
@lxplus (dbt) $ pip install .
```

```
(dbt) [jurenago@lxplus9122 repositories]$ cd waffles/
(dbt) [jurenago@lxplus9122 waffles]$ pip install -r requirements.txt
Collecting awkward (from -r requirements.txt (line 1))
  Obtaining dependency information for awkward from https://files.pythonhosted.org/packages/c2/91/3bc90f6a3f109f41edaedba0
a23ad9b1d1a2ae6739ebef4678b97d4f0901/awkward-2.7.1-py3-none-any.whl.metadata
  Downloading awkward-2.7.1-py3-none-any.whl.metadata (7.0 kB)
Collecting inquirer (from -r requirements.txt (line 2))
  Obtaining dependency information for inquirer from https://files.pythonhosted.org/packages/a4/b2/be907c8c0f8303bc4b10089
f5470014c3bf3521e9b8d3decf3037fd94725/inquirer-3.4.0-py3-none-any.whl.metadata
  Using cached inquirer-3.4.0-py3-none-any.whl.metadata (6.8 kB)
Collecting ipywidgets (from -r requirements.txt (line 3))
  Obtaining dependency information for ipywidgets from https://files.pythonhosted.org/packages/22/2d/9c0b76f2f9cc0ebede1b9
371b6f317243028ed60b90705863d493bae622e/ipywidgets-8.1.5-py3-none-any.whl.metadata
  Using cached ipywidgets-8.1.5-py3-none-any.whl.metadata (2.3 kB)
Collecting matplotlib (from -r requirements.txt (line 4))
  Obtaining dependency information for matplotlib from https://files.pythonhosted.org/packages/4a/86/bb508f20bdda70b5e7afd
```

.
.

2. Install Waffles within your DBT environment

Activate the DBT environment (`source env.sh`), navigate to the folder where you will download waffles, then:

```
@lxplus (dbt) $ cd <repos_dir>
@lxplus (dbt) $ git clone https://github.com/DUNE/waffles.git
@lxplus (dbt) $ cd waffles
@lxplus (dbt) $ pip install -r requirements.txt
@lxplus (dbt) $ pip install .
```

```
(dbt) [jurenago@lxplus9122 waffles]$ pip install .
Processing /afs/cern.ch/work/j/jurenago/private/tutorial/repositories/waffles
  Preparing metadata (setup.py) ... done
Building wheels for collected packages: waffles
  Building wheel for waffles (setup.py) ... done
  Created wheel for waffles: filename=waffles-0.1.0-py3-none-any.whl size=136160 sha256=8c80a8558af76f5b93e16739c77594479b
91b77f331c53b46b70efc6a39c50e5
  Stored in directory: /tmp/jurenago/pip-ephem-wheel-cache-ce845vjd/wheels/60/75/25/1a4382574d13de42c0caee181daba5d15c9adf
6d1299750391
Successfully built waffles
Installing collected packages: waffles
Successfully installed waffles-0.1.0
```

3. Setup rucio

```
@lxplus (dbt) $ deactivate
@lxplus $ cd <repos_dir>/waffles/scripts
@lxplus $ source setup_rucio_a9.sh
```

```
(dbt) [jurenago@lxplus9122 waffles]$ deactivate
[jurenago@lxplus9122 waffles]$ cd scripts/
```

3. Setup rucio

```
@lxplus (dbt) $ deactivate
@lxplus $ cd <repos_dir>/waffles/scripts
@lxplus $ source setup_rucio_a9.sh
```

This script sets up rucio for Linux Alma 9. Alternatively, for Scientific Linux 7 run `setup_rucio_a9.sh`

```
[jurenago@lxplus9122 scripts]$ source setup_rucio_2.sh
Checking if /tmp/                  can be reused ... no
Authorizing ...... authorized
Fetching certificate ...../cvmfs/larsoft.opensciencegrid.org/spack-packages/opt/spack/linux-almalinux9-x86_64_v2/gcc-11.4.
1/cigetcert-1.20-h5k3aphdddyhcyomqirlgkgjc7salbvv/bin/cigetcert:1221: DeprecationWarning: PKCS#12 support in pyOpenSSL is
deprecated. You should use the APIs in cryptography.
  p12 = crypto.load_pkcs12(pkcs12cert, p12password)
/cvmfs/larsoft.opensciencegrid.org/spack-packages/opt/spack/linux-almalinux9-x86_64_v2/gcc-11.4.1/cigetcert-1.20-h5k3aphdd
dyhcyomqirlgkgjc7salbvv/bin/cigetcert:1221: DeprecationWarning: str for passphrase is no longer accepted, use bytes
  p12 = crypto.load_pkcs12(pkcs12cert, p12password)
 fetched
Storing certificate in /tmp/
Your certificate is valid until: Thu Dec 12 13:05:40 2024
```

.
.
.

3. Setup rucio

```
@lxplus (dbt) $ deactivate
@lxplus $ cd <repos_dir>/waffles/scripts
@lxplus $ source setup_rucio_a9.sh
```

This script sets up rucio for Linux Alma 9. Alternatively, for Scientific Linux 7 run `setup_rucio_a9.sh`

.
.
.

```
Your certificate is valid until: Thu Dec 12 13:05:40 2024
status     : ACTIVE
account    :
email      :
deleted_at : None
updated_at : 2024-04-12T03:34:11
suspended_at : None
account_type : USER
created_at : 2024-04-12T03:34:11
Contacting voms1.fnal.gov:15042 [/DC=org/DC=incommon/C=US/ST=Illinois/O=Fermi Research Alliance/CN=voms1.fnal.gov] "dune".
..
Remote VOMS server contacted succesfully.


Created proxy in /tmp/.

Your proxy is valid until Tue Dec 10 13:06:00 CET 2024
```

4. Get an XRootD ticket

Still in `<repos_dir>/waffles/scripts` and with the kerberos authentication ticket to FNAL.GOV, run

```
@lxplus $ source setup_ifdhc.sh
```

```
[jurenago@lxplus9122 scripts]$ source setup_ifdhc.sh
Setting up larsoft UPS area... /cvmfs/larsoft.opensciencegrid.org
Setting up DUNE UPS area... /cvmfs/dune.opensciencegrid.org/products/dune/
perl: symbol lookup error: /cvmfs/larsoft.opensciencegrid.org/products/mrb/v6_09_07/slf7.x86_64/CPAN/lib/perl5/x86_64-linu
x-thread-multi/auto/Cwd/Cwd.so: undefined symbol: Perl_xs_apiversion_bootcheck
/cvmfs/fermilab.opensciencegrid.org/products/common/db//../prd/cigetcert/v1_20_4/Linux64bit-3-10-2-17/bin/cigetcert: line
22: /usr/bin/python: No such file or directory
Contacting voms1.fnal.gov:15042 [/DC=org/DC=incommon/C=US/ST=Illinois/O=Fermi Research Alliance/CN=voms1.fnal.gov] "dune".
..
Remote VOMS server contacted succesfully.

WARNING: proxy lifetime limited to issuing credential lifetime.

Created proxy in /tmp/[redacted]

Your proxy is valid until Tue Dec 10 13:06:00 CET 2024
Attempting OIDC authentication with https://htvaultprod.fnal.gov:8200

Complete the authentication at:
    https://cilogon.org/device/?user_code=QQC-93X-2LN
No web open command defined, please copy/paste the above to any web browser
```

4. Get an XRootD ticket

Still in `<repos_dir>/waffles/scripts` and with the kerberos authentication ticket to FNAL.GOV, run

```
@lxplus $ source setup_ifdhc.sh
```

.
.
.

```
Waiting for response in web browser
Storing vault token in /run/user/      /vt
Saving credkey to /afs/cern.ch/user/j/
Saving refresh token ... done
Attempting to get token from https://htvaultprod.fnal.gov:8200 ... succeeded
Storing bearer token in /tmp/bt_token_dune_Analysis_
Copying 864189264 bytes https://fndcadoor.fnal.gov:2880/dune/tape_backed/dunepro/hd-protodune/raw/2024/detector/cosmics/No
ne/00/02/63/65/np04hd_raw_run026365_0001_dataflow0_datawriter_0_20240522T135540.hdf5 => file:///afs/cern.ch/work/j/jurenag
o/private/repositories/waffles/scripts/np04hd_raw_run026365_0001_dataflow0_datawriter_0_20240522T135540.hdf5
gfal-copy error: 17 (File exists) - The file exists and overwrite is not set
Perhaps you forgot a -D to indicate destination is a directory?
Thu Dec  5 13:24:44 2024
 program: www_cp.sh  https://fndcadoor.fnal.gov:2880/dune/tape_backed/dunepro/hd-protodune/raw/2024/detector/cosmics/None/
00/02/63/65/np04hd_raw_run026365_0001_dataflow0_datawriter_0_20240522T135540.hdf5 /afs/cern.ch/work/j/jurenago/private/rep
ositories/waffles/scripts/.exited status 17
delaying 1 ...
Thu Dec  5 13:24:47 2024
 retrying...
Copying 864189264 bytes https://fndcadoor.fnal.gov:2880/dune/tape_backed/dunepro/hd-protodune/raw/2024/detector/cosmics/No
ne/00/02/63/65/np04hd_raw_run026365_0001_dataflow0_datawriter_0_20240522T135540.hdf5 => file:///afs/cern.ch/work/j/jurenag
o/private/repositories/waffles/scripts/np04hd_raw_run026365_0001_dataflow0_datawriter_0_20240522T135540.hdf5
```

5. Create a WaveformSet object

Consider run 29297 - check if the rucio paths for such run are already available in `/eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths`

```
(dbt) [jurenago@lxplus9122 test]$ cd /eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths
(dbt) [jurenago@lxplus9122 1_rucio_paths]$ ls
025106.txt  026140.txt  026263.txt  026487.txt  026631.txt  027211.txt  027384.txt  027752.txt  027951.txt  028181.txt  028563.txt  029657.txt
025107.txt  026141.txt  026265.txt  026490.txt  026634.txt  027212.txt  027385.txt  027754.txt  027952.txt  028183.txt  028586.txt  029681.txt
025108.txt  026145.txt  026270.txt  026491.txt  026636.txt  027213.txt  027386.txt  027758.txt  027965.txt  028186.txt  028595.txt  029697.txt
025109.txt  026147.txt  026271.txt  026512.txt  026639.txt  027214.txt  027387.txt  027762.txt  027971.txt  028187.txt  028600.txt  029698.txt
025123.txt  026149.txt  026272.txt  026513.txt  026640.txt  027215.txt  027388.txt  027801.txt  027973.txt  028188.txt  028602.txt  029699.txt
025125.txt  026152.txt  026273.txt  026514.txt  026649.txt  027216.txt  027389.txt  027803.txt  027975.txt  028189.txt  028605.txt  029701.txt
025130.txt  026154.txt  026274.txt  026515.txt  026650.txt  027217.txt  027390.txt  027805.txt  027980.txt  028191.txt  028610.txt  029704.txt
025141.txt  026155.txt  026275.txt  026516.txt  026683.txt  027218.txt  027391.txt  027808.txt  027997.txt  028193.txt  028622.txt  029753.txt
025145.txt  026156.txt  026281.txt  026517.txt  026685.txt  027219.txt  027392.txt  027810.txt  028005.txt  028195.txt  028623.txt  029754.txt
```

•
•
•

They are not

```
[jurenago@lxplus9122 ~]$ cd /eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths
[jurenago@lxplus9122 1_rucio_paths]$ ls 029297.txt
ls: cannot access '029297.txt': No such file or directory
[jurenago@lxplus9122 1_rucio_paths]$
```

5. Create a WaveformSet object

They are not. Get them by running `<repos_dir>/waffles/scripts/get_rucio.py` giving the targeted runs via the `--runs` parameter

```
(dbt) [jurenago@lxplus9122 scripts]$ python get_rucio.py --runs 29297

Getting the path for run 29297:

You are the first one looking for this file. Let's get the rucio paths!.
You are the first one to look at the paths of run 029297!!
get_protodunehd_files.sh: line 54: rucio: command not found
Configuring rucio in Red Hat Enterprise Linux: 9.5
--2024-12-05 13:58:58--  https://authentication.fnal.gov/krb5conf/SL7/krb5.conf
Resolving authentication.fnal.gov (authentication.fnal.gov)... 2620:6a:0:105:f0:0:105:43, 131.225.105.111
Connecting to authentication.fnal.gov (authentication.fnal.gov)|2620:6a:0:105:f0:0:105:43|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 12552 (12K) [text/plain]
Saving to: [redacted]

[redacted]                  100%[=================================================>]  12.26K  --.-KB/s    in 0s

2024-12-05 13:58:59 (220 MB/s) - [redacted] saved [12552/12552]
```

.
.
.

5. Create a WaveformSet object

They are not. Get them by running `<repos_dir>/waffles/scripts/get_rucio.py` giving the targeted runs via the `--runs` parameter

.
.
.

```
/eos/experiment/neutplatform/protodune/dune/hd-protodune/6c/a0/np04hd_raw_run029297_0000_dataflow0_datawriter_0_20240925T104111.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/a8/74/np04hd_raw_run029297_0000_dataflow1_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/31/70/np04hd_raw_run029297_0000_dataflow2_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/e8/25/np04hd_raw_run029297_0000_dataflow3_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/39/62/np04hd_raw_run029297_0000_dataflow4_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/6b/6f/np04hd_raw_run029297_0000_dataflow5_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/d4/df/np04hd_raw_run029297_0000_dataflow6_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/3c/7a/np04hd_raw_run029297_0000_dataflow7_datawriter_0_20240925T104111.hdf5
```

5. Create a WaveformSet object

The rucio paths should have been stored to
`/eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths/029297.txt`

```
[jurenago@lxplus9122 1_rucio_paths]$ cat /eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths/029297.txt
/eos/experiment/neutplatform/protodune/dune/hd-protodune/6c/a0/np04hd_raw_run029297_0000_dataflow0_datawriter_0_20240925T104111.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/a8/74/np04hd_raw_run029297_0000_dataflow1_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/31/70/np04hd_raw_run029297_0000_dataflow2_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/e8/25/np04hd_raw_run029297_0000_dataflow3_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/39/62/np04hd_raw_run029297_0000_dataflow4_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/6b/6f/np04hd_raw_run029297_0000_dataflow5_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/d4/df/np04hd_raw_run029297_0000_dataflow6_datawriter_0_20240925T104110.hdf5
/eos/experiment/neutplatform/protodune/dune/hd-protodune/3c/7a/np04hd_raw_run029297_0000_dataflow7_datawriter_0_20240925T104111.hdf5
```

5. Create a WaveformSet object

The path to the newly created file (containing the rucio paths) can be given to the `rucio_filepath` variable defined in `<repos_dir>/waffles/test/wtest_hdf5_reader.py`. This python script is a simple demonstration on how the waffles reading tools can be used to create a waffles WaveformSet object and save it to a pickle file.

This script takes the rucio paths stored in the given file, and creates a WaveformSet encapsulating the data found in the **first** rucio path.

```
import pickle
import waffles.input.raw_hdf5_reader as reader

rucio_filepath = "/eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths/
029297.txt"
filepaths = reader.get_filepaths_from_rucio(rucio_filepath)
print(filepaths)
wfset = reader.WaveformSet_from_hdf5_file( filepaths[0],                    # path to the root file
                                           read_full_streaming_data = False, # self-triggered (False) data
                                         )                                   # subsample the data reading (read each 2 ent
ries)
with open("wfset.pkl", "wb") as f:
    pickle.dump(wfset, f)
```

5. Create a WaveformSet object

This WaveformSet object is saved to the `wfset.pkl` file alongside `<repos_dir>/waffles/test/wtest_hdf5_reader.py`

```
(dbt) [jurenago@lxplus9122 test]$ python3 wtest_hdf5_reader.py
Your files are stored in /eos/
['/eos/experiment/neutplatform/protodune/dune/hd-protodune/6c/a0/np04hd_raw_run029297_0000_dataflow0_datawriter_0_20240925
T104111.hdf5', '/eos/experiment/neutplatform/protodune/dune/hd-protodune/a8/74/np04hd_raw_run029297_0000_dataflow1_datawri
ter_0_20240925T104110.hdf5', '/eos/experiment/neutplatform/protodune/dune/hd-protodune/31/70/np04hd_raw_run029297_0000_dat
aflow2_datawriter_0_20240925T104110.hdf5', '/eos/experiment/neutplatform/protodune/dune/hd-protodune/e8/25/np04hd_raw_run0
29297_0000_dataflow3_datawriter_0_20240925T104110.hdf5', '/eos/experiment/neutplatform/protodune/dune/hd-protodune/39/62/n
p04hd_raw_run029297_0000_dataflow4_datawriter_0_20240925T104110.hdf5', '/eos/experiment/neutplatform/protodune/dune/hd-pro
todune/6b/6f/np04hd_raw_run029297_0000_dataflow5_datawriter_0_20240925T104110.hdf5', '/eos/experiment/neutplatform/protodu
ne/dune/hd-protodune/d4/df/np04hd_raw_run029297_0000_dataflow6_datawriter_0_20240925T104110.hdf5', '/eos/experiment/neutpl
atform/protodune/dune/hd-protodune/3c/7a/np04hd_raw_run029297_0000_dataflow7_datawriter_0_20240925T104111.hdf5']
run_numb= 29297
452it [00:06, 71.83it/s]
(dbt) [jurenago@lxplus9122 test]$ ls
wfset.pkl  wtest_hdf5_reader.py
```

If the given HDF5 is stored in `/eos` then `WaveformSet_from_hdf5_file()` can directly instantiate a `hdf5libs.HDF5RawDataFile` object given the HDF5 filepath. P.e. the data from run 29297 is hosted in `/eos`

5. Create a WaveformSet object

That is not always the case though. HDF5 files might be stored *somewhere around the world*, i.e. in RSEs of collaborating institutions. If the HDF5 file is not stored in `/eos`, then `WaveformSet_from_hdf5_file()`uses XRootD to temporarily copy the targeted HDF5 file from the XRootD server to `/tmp`

Take for example run 26188:

```
(dbt) [jurenago@lxplus9122 ~]$ cat /eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths/026118.txt
root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/f6/54/np04hd_raw_run026118_0000_dataflow0_datawriter_0_20240509T193347.hdf5

root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/ca/14/np04hd_raw_run026118_0000_dataflow1_datawriter_0_20240509T193346.hdf5

root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/6c/7e/np04hd_raw_run026118_0000_dataflow2_datawriter_0_20240509T193346.hdf5

root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/37/b6/np04hd_raw_run026118_0000_dataflow3_datawriter_0_20240509T193346.hdf5

root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/4d/1f/np04hd_raw_run026118_0000_dataflow4_datawriter_0_20240509T193346.hdf5

root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/9c/8c/np04hd_raw_run026118_0000_dataflow5_datawriter_0_20240509T193347.hdf5
```

5. Create a WaveformSet object

Running `<repos_dir>/waffles/test/wtest_hdf5_reader.py` for such a run gives

```python
import pickle
import waffles.input.raw_hdf5_reader as reader

rucio_filepath = "/eos/experiment/neutplatform/protodune/experiments/ProtoDUNE-II/PDS_Commissioning/waffles/1_rucio_paths/026118_.txt"
filepaths = reader.get_filepaths_from_rucio(rucio_filepath)
print(filepaths)
wfset = reader.WaveformSet_from_hdf5_file( filepaths[0],                    # path to the root file
                                           read_full_streaming_data = False, # self-triggered (False) data
                                         )                                   # subsample the data reading (read each 2 entries)
with open("wfset.pkl", "wb") as f:
    pickle.dump(wfset, f)
```

```
(dbt) [jurenago@lxplus9122 test]$ python3 wtest_hdf5_reader.py

Your files are stored around the world.
[WARNING] Check you have a correct configuration to use XRootD
['root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/f6/54/np04hd_raw_run026118_0000_dataflow0_datawriter_0_20240509T193347.hdf5', '
root://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/ca/14/np04hd_raw_run026118_0000_dataflow1_datawriter_0_20240509T193346.hdf5', 'ro
ot://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/6c/7e/np04hd_raw_run026118_0000_dataflow2_datawriter_0_20240509T193346.hdf5', 'root
://xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/37/b6/np04hd_raw_run026118_0000_dataflow3_datawriter_0_20240509T193346.hdf5', 'root:/
/xrootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/4d/1f/np04hd_raw_run026118_0000_dataflow4_datawriter_0_20240509T193346.hdf5', 'root://x
rootd.pic.es:1094/pnfs/pic.es/data/dune/RSE/hd-protodune/9c/8c/np04hd_raw_run026118_0000_dataflow5_datawriter_0_20240509T193347.hdf5']
Using XROOTD
[2.682GB/2.682GB][100%][==================================================][50.86MB/s]
run_numb= 26118
10019it [01:56, 85.84it/s]
(dbt) [jurenago@lxplus9122 test]$ ls
wfset.pkl  wtest_hdf5_reader.py
```

Reference list

[1] Waffles documentation webpage
https://waffles.readthedocs.io/en/latest/index.html

[2] Computing basics for DUNE: Data management
https://hschellman.github.io/computing-basics/03-data-management/index.html

[3] Computing basics for DUNE tutorial: Storage spaces
https://dune.github.io/computing-basics/02-storage-spaces/index.html

[4] XRootD webpage
https://xrootd.github.io/