WRITING DUNE MC RAW DATA TO HDF5 FILES USING ART

Barnali Chowdhury, Tom Junk, Peter Van Gemmeran

DUNE Pre-Collaboration Computing meeting

05/09/2022

MOTIVATION

- The data size for one entry in ROOT for a single TBranch is limited to 1GB.
- Challenge is writing out one event's worth of data/uncompressed RawDigits in MC for a full FD module.
- Get around the ROOT 1-GB limit on a trigger record
  - 2 choices – ROOT, HDF5

- To be able to write output data for a smaller unit than a trigger record
  - without having to store it all in memory before writing it out.

INTRODUCTION

- Started with Tom writing and most importantly, reading the binary files back in an art job.
- The code, for now, lives in /dune/app/users/barnali/hdf5_MCRd_May2022/srcs/dunetpc/dune/HDF5Gen
- Working with larsoft_v09_38_00
- Will migrate HDF5Gen to the new GitHub repos

THE MC DATA MODEL

- The writer module writes the output with the following data structure

GROUP "APA_0" {

```
DATASET "ChannelHeader_0" {
  DATATYPE  H5T_COMPOUND {
    H5T_STD_I32LE "Chan";
    H5T_IEEE_F32LE "Pedestal";
    H5T_IEEE_F32LE "Sigma";
    H5T_STD_I32LE "nADC";
    H5T_STD_I16LE "Compression";
  }
```

→ Writes MetaData

Header Info will be replicated nChannels * nTriggerRecords times per APA.

```
DATASET "ChannelID_0" {
      DATATYPE  H5T_STD_U16LE
      DATASPACE  SIMPLE { ( 6000 ) / ( 6000 ) }
      DATA {
```

→ Writes Data

Channel Data will also be replicated nChannels * nTriggerRecords times per APA.

You Inst Logo

DAQ STYLE OF DATA MODELLING

```
GROUP "TriggerRecord00028" {
    GROUP "TPC" {
      GROUP "APA000" {
        DATASET "Link00" {
          DATATYPE  H5T_STD_I8LE
          DATASPACE  SIMPLE { ( 4640080, 1 ) / ( 4640080, 1 ) }
        }
        DATASET "Link01" {
          DATATYPE  H5T_STD_I8LE
          DATASPACE  SIMPLE { ( 4640080, 1 ) / ( 4640080, 1 ) }
        }
.....................

 DATASET "TriggerRecordHeader" {
       DATATYPE  H5T_STD_I8LE
       DATASPACE  SIMPLE { ( 368, 1 ) / ( 368, 1 ) }
    }
  }
```

WHY DIFFERENT MODELS ?

- We are not formatting the data generated in the MC to look exactly like what is produced online.
- There is no need to have the same data format for data and MC as long as there is a lossless map between them.
- MC groups the data by APA. That way searching through the trigger record for all the channels in an APA will be quicker and easier.
- All we need is a separate tool to read each kind of file format/data format in the existing tools to accommodate all different kinds of data.
- We may need to produce DAQ style MC files in future.

INPUT & OUTPUT FILES

- The input file → /dune/data/users/trj/pdsp_data/np04_raw_run005177_0004_dl1_decode_gen.root
- It reads one event out of ProtoDUNE-SP data file with RawDigits in it from 6 APAs and writes out 1 art.h5 file.
- The output file → *dump_binary.h5*
- The size of the output file ***to write 1 event*** → **191M**

DUMPING THE OUTPUT FILE

```
GROUP "APA_1" {
   DATASET "ChannelHeader_2560" {
     DATATYPE  H5T_COMPOUND {
       H5T_STD_I32LE "Chan";
       H5T_IEEE_F32LE "Pedestal";
       H5T_IEEE_F32LE "Sigma";
       H5T_STD_I32LE "nADC";
       H5T_STD_I16LE "Compression";
     }
     DATASPACE  SIMPLE { ( 1 ) / ( 1 ) }
     DATA {
     (0): {
          2560,
          2305,
          8.49856,
          6000,
          0
       }
     }
}
```

```
DATASET "ChannelID_2560" {
      DATATYPE  H5T_STD_I16LE
      DATASPACE  SIMPLE { ( 6000 ) / ( 6000 ) }
      DATA {
      (0): 2271, 2274, 2274, 2275, 2272, 2272, 2271, 2269,  2271, 2270,
      (10): 2271, 2269, 2270, 2272, 2275, 2276, 2275, 2275, 2275, 2272,
      (20): 2271, 2271, 2269, 2269, 2267, 2270, 2271, 2275, 2275, 2275,
      (30): 2275, 2273, 2271, 2269, 2271, 2272, 2275, 2277, 2276, 2274,
      (40): 2271, 2269, 2269, 2270, 2271, 2274, 2279, 2283, 2284, 2283,
      (50): 2278, 2279, 2280, 2281, 2282, 2279, 2274, 2271, 2272, 2275,
      ------------------------------------------------
      ------------------------------------------------
```

- 6000 ADC values are written here.
- They are in signed 16-bit integers.

Comments, Suggestions, Questions ??