ProtoDUNE storage need estimates.   H-Schellman , Dec 11, 2019. Docdb 17086 – v7

The spreadsheet attached to this document is estimates of data sizes for protoDUNE DP and SP. This is an updated version with projections for big DUNE added.

https://docs.dunescience.org/cgi-bin/private/ShowDocument?docid=16028

https://docs.dunescience.org/cgi-bin/private/ShowDocument?docid=14983

are inputs for the far detector.

Raw data estimate:

Estimates are a combination of bottoms up estimates based on rates and channel counts and observations from real data.

For example, uncompressed single phase data were observed to be around 178 MB in size, which is the amount expected for the # of TPC channels read + a 20% overhead for other detectors and headers. Compressed SP data averages 71 MB, consistent with compression by a factor of 2.5

Dual phase data includes 2 CRPs for the 2019 run. Observed data size without compression is 110MB. For the 2020-2022 we project a compression factor of 10, with 4 CRPs for the 2021-2022 beam runs.

Numbers for 2018 and 2019 have been adjusted to reflect actual #'s.

Add 1 TB of test data/year with retention policy of 2 years.

Run Plan:

We assume 300 days of cosmic running in 2019 and 2022 and 150 days in 2020/2021 for detector downtime

We assume 50 days of beam in both 2021 and 2022.

Assume run cosmics in 2023 and keep reprocessing the data until 2025.

Assume commission one SP-FD module in 2026 and one (DP) in 2028

FD modules are assumed to have compression factor of 2.5(SP) and 10(DP) for raw data.

Reconstructed data estimates:

After the first pass in 2018-2019, reconstruction is assumed to drop raw hits, leading to output sizes that are 10% of the original uncompressed size for both DP and SP protoDUNE. For FD events without the large cosmic ray content, we assume a factor of 100 reduction.

Reconstruction for protoDUNE currently takes ~ 600 sec/event, measured for both SP and DP on interactive machines at Fermilab.

Simulation is assumed to be 1-2 PB/year. Simulation time estimates are 0.75 hrs/event based on high occupancy ProtoDUNE events.

Data lifetimes and # of copies:

For ProtoDUNE

We assume 2 raw data copies on tape – at different centers, kept indefinitely.

Test data is written to tape and retained for 2 years.

We assume 2 reconstruction passes/year which will only be kept for 2 years. (ie 4 versions) – One copy is also to tape and becomes the backup once the disk copy is flushed.

We assume 1 simulation passes/year, kept for 2 years.

We assume that reconstructed and simulated data is available on disk at 2 institutions.

We assume reconstruction continues through 2025 at which time the samples are frozen.

For DUNE

We assume 2 raw data copies on tape – at different centers, kept indefinitely.

We assume 2 reconstruction passes/year which are run over the **full existing data sample**.

We assume 1 simulation passes/year, kept for 2 years

We assume that reconstructed and simulated data is available on disk at 2 institutions and one copy on tape.

Analysis

We assume analysis CPU is similar to production CPU based on experience to date but that analysis storage sizes are << reconstructed samples.

.
Table 1 – projections for storage and CPU time 2020-2030

The top 2 sets show the cumulative raw single copy data sizes and yearly CPU based on the year by year estimates in the attached spreadsheet. We assume 2 copies of raw data on tape and 2 processing passes/year with 2 copies on disk at different DUNE institutions.   Reconstructed samples are assumed to be dumped to tape after 2 years.

CPU estimates are doubles to reflect observed 1:1 ratio of analysis to production

Cumulative	2018	2019	2020	2021	2022	2023	2024
Raw 2 copies, PB	1.5	4.1	5.5	11.5	18.4	19.8	19.8
Reco 2 copies PB	6.0	8.1	3.6	8.9	15.5	16.3	16.3
sim 2 copies PB	1.0	5.0	8.0	8.0	8.0	8.0	8.0
total disk, PB	7.0	13.1	11.6	16.9	23.5	24.3	24.3
total tape, PB	1.5	4.1	9.0	18.0	24.2	28.2	31.6
CPU, MH x2	11	29	34	72	78	78	78

Cumulative	2025	2026	2027	2028	2029	2030
Raw 2 copies, PB	19.8	31.7	43.6	58.4	73.3	88.2
Reco 2 copies PB	16.3	8.7	1.2	1.8	3.0	4.2
sim 2 copies PB	8.0	8.0	8.0	8.0	8.0	8.0
total disk, PB	24.3	16.7	9.2	9.8	11.0	12.2
total tape, PB	32.0	43.8	55.7	66.8	77.9	93.1
CPU, MH x2	78	39	78	117	194	311