ProtoDUNE storage need estimates. H-Schellman , Nov 24, 2019

The spreadsheet attached to this document is estimates of data sizes for protoDUNE DP and SP.

**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 LEM’s for the 2019 run. Observed data size without compression is 110MB. For the 2020-2022 we project a compression factor of 10, with 4 LEMS for the 2021-2022 beam runs.

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

**Reconstructed data estimates:**

Reconstruction is assumed to drop raw hits, leading to output sizes around 3 times smaller for SP.

Simulation is assumed to be 1-2 PB/year. Simulation time estimates are XXXX.

**Data lifetimes and # of copies:**

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

We assume 2 reconstruction passes/year which will only be kept for 3 years.

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

.

**Table 1 – projections for storage and CPU time.**

The top 2 sets show the raw single copy data sizes and CPU based on the year by year estimates. At the bottom, we assume 2 copies of raw data on tape and 2 processing passes/year with 2 copies on disk at different DUNE institutions.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **2018** | **2019** | **2020** | **2021** | **2022** | **1st FD module** |
|  |  | **Asbuilt** |  |  |  |  | **uncompressed** |
| SP | Events, M | 10.9 | 19.4 | 6.5 | 23.5 | 30.0 | 4.3 |
|  | Raw data, TB | 751 | 1197 | 448 | 1624 | 2072 | 10790 |
|  | Reco data, TB | 1501 | 395 | 134 | 487 | 621 | 108 |
|  | CPU, MH | 1.8 | 3.2 | 1.1 | 3.9 | 5.0 | 22.5 |
|  |  |  |  |  |  |  |  |
| DP | Events, M | 0.0 | 2.6 | 21.6 | 55.2 | 55.2 | 0.0 |
|  | Raw data, TB | 0 | 285 | 238 | 1214 | 1214 | 0 |
|  | Reco data, TB | 0 | 342 | 238 | 1214 | 1214 | 0 |
|  | CPU, MH | 0.0 | 0.4 | 3.6 | 9.2 | 9.2 | 0.0 |
|  |  |  |  |  |  |  |  |
| **yearly** |  |  |  |  |  |  |  |
| total | Events, M | 10.9 | 22.0 | 28.1 | 78.7 | 85.1 | 4.3 |
| 2x | Raw data, TB | 1501 | 2963 | 1371 | 5674 | 6570 | 21580 |
| 2 passes | Reco data, TB | 3003 | 1474 | 744 | 3401 | 3670 | 216 |
| copies | reco data, TB | 6005 | 2948 | 1488 | 6803 | 7340 | 432 |
|  | CPU, MH x2 | 3.6 | 7.3 | 9.4 | 26.2 | 28.4 | 22.5 |
|  | total storage | 7507 | 5911 | 2859 | 12477 | 13910 | 22011 |
|  |  |  |  |  |  |  |  |
|  | sim | 0.5 | 1 | 2 | 2 | 2 | 2 |
|  | CPU for sim |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
| **Cumulative** | |  |  |  |  |  |  |
| Raw \*2 |  | 1501 | 4465 | 5836 | 11510 | 18080 | 39660 |
| Reco 2 versions for 3 years | | 3003 | 4477 | 5221 | 5619 | 7815 | 8031 |
| Reco \*2 copies | | 6005 | 8953 | 10441 | 11238 | 15630 | 16062 |