

DUNE Computing Resource Model for 2024-2025

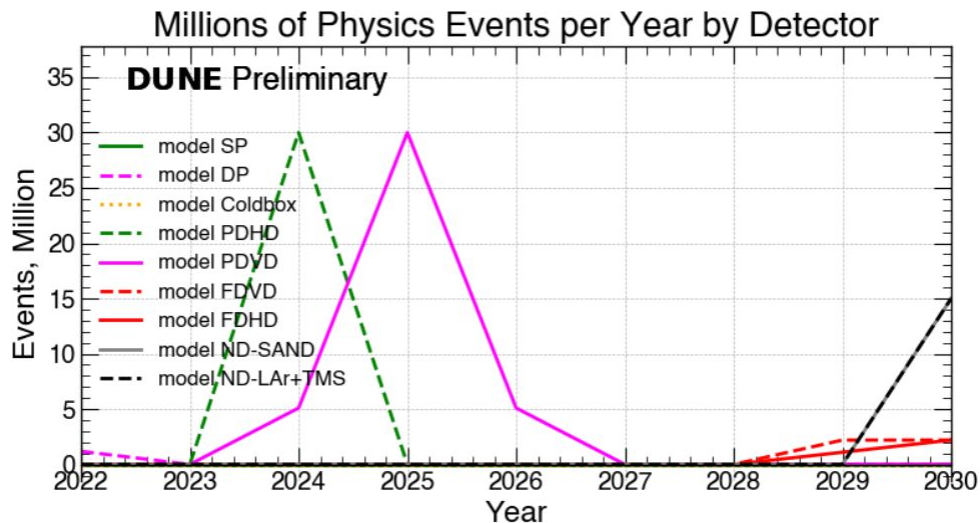
Computing Contributions Board Meeting
September 2024

Basics of the resource model

- Keep raw data on disk for 2-3 years, on tape to end of expt.
 - 2 tape copies, one disk copy. (For 2025 Increased lifetime for raw data from 1 to 3 years, reduced copies to 1 to allow both NP02 and NP04 on disk through 2026)
- Reconstruct full sample every year (protoDUNE for 3 years, DUNE to end of expt.)
- Do new simulation campaigns each year
- Keep simulation and reconstructed data on disk for 2 years (always have 2 versions)
 - One copy in Americas, one in Europe where possible (model assumes 1.5 copies for sim)
 - No need to stage from tape until it ages out
- After that, one copy of reconstruction/simulation -> tape as it can be redone if necessary.
- **CPU estimates** are based on measurements from ProtoDUNE data and existing simulations **ND estimates have large uncertainties.**

Updates to the model

- **Summer 2024: Complete rewrite of the underlying code for clarity and flexibility.**
- **Site splits now based on detector (PD, FD, ND), not year.**
- **Delayed ProtoDUNE II operations** until Spring 2024 but then more data than expected.
- With the **successful large-scale FD simulation campaigns**, we have considerably better understanding of both our processing time expectations and our simulation needs
- Not all of that additional information has been included in the model just yet (minor tweaks)
- new estimates for FD processing time based upon multithreaded processing and subsequent smaller memory footprint
- updated files sizes for reconstruction and simulation output – no longer estimate based upon retention of the raw waveforms in data or all rawdigits in the simulation
- still working on understanding the GPU requirements for 2x2 and ND-LAr and how those estimates can be translated
- Transitioned **from MWC to HEPsScore23 (HS23)** as the metric for CPU resources

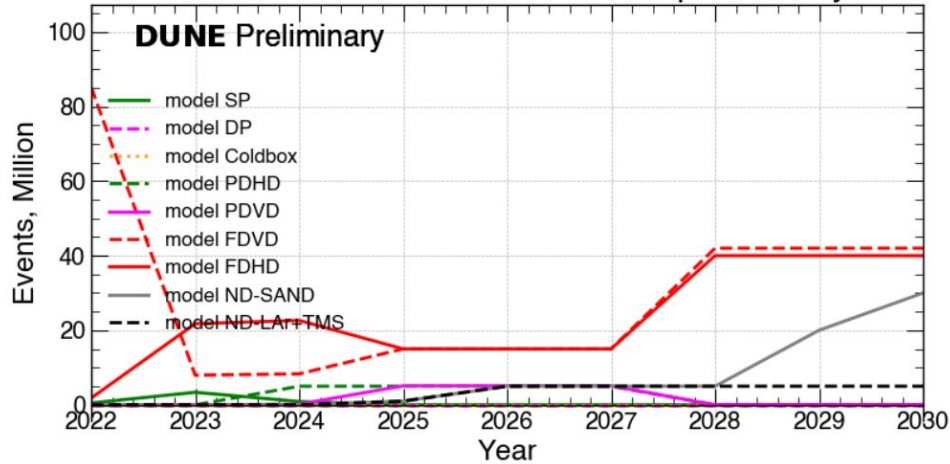


Model input, # of events

Detectors	2022	2023	2024	2025	2026	2027	2028	2029	2030
SP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DP	1.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Coldbox	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDHD	0.0	0.0	30.0	0.0	0.0	0.0	0.0	0.0	0.0
PDVD	0.0	0.0	5.1	30.0	5.1	0.0	0.0	0.0	0.0
FDVD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	2.2
FDHD	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1	2.2
ND-SAND	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0
ND-LAr+TMS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.0

Figure 2: Number of raw data events per year, units are Millions.

Millions of Reconstructed Simulated Events per Year by Detector



- Estimates for # of simulated events.

Detectors	2022	2023	2024	2025	2026	2027	2028	2029	2030
SP	0.5	3.4	0.9	0.0	0.0	0.0	0.0	0.0	0.0
DP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Coldbox	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDHD	0.0	0.0	5.0	5.0	5.0	5.0	0.0	0.0	0.0
PDVD	0.0	0.0	0.2	5.1	5.1	5.1	0.0	0.0	0.0
FDVD	85.1	8.0	8.3	15.0	15.0	15.0	42.0	42.0	42.0
FDHD	1.8	21.8	22.6	15.0	15.0	15.0	40.0	40.0	40.0
ND-SAND	0.0	0.0	0.0	1.0	5.0	5.0	5.0	20.0	30.0
ND-LAr+TMS	0.0	0.0	0.0	1.0	5.0	5.0	5.0	5.0	5.0

Figure 3: Number of simulated/reconstructed events per year, units are Millions. There was a large simulation campaign in 2022 for the FDVD technical design report.

Lifetimes and copies

Parameters	DiskCopies	DiskLifetimes	TapeCopies	TapeLifetimes
Raw-Data	1	3	2	100
Test	1	0.5	1	1.0
TP	1	0.5	1	100
Reco-Data	2	2	1	15
Reco-Sim	1.5	2	1	15
Analysis-Data	2	5	1	15
Analysis-Sim	2	5	1	15

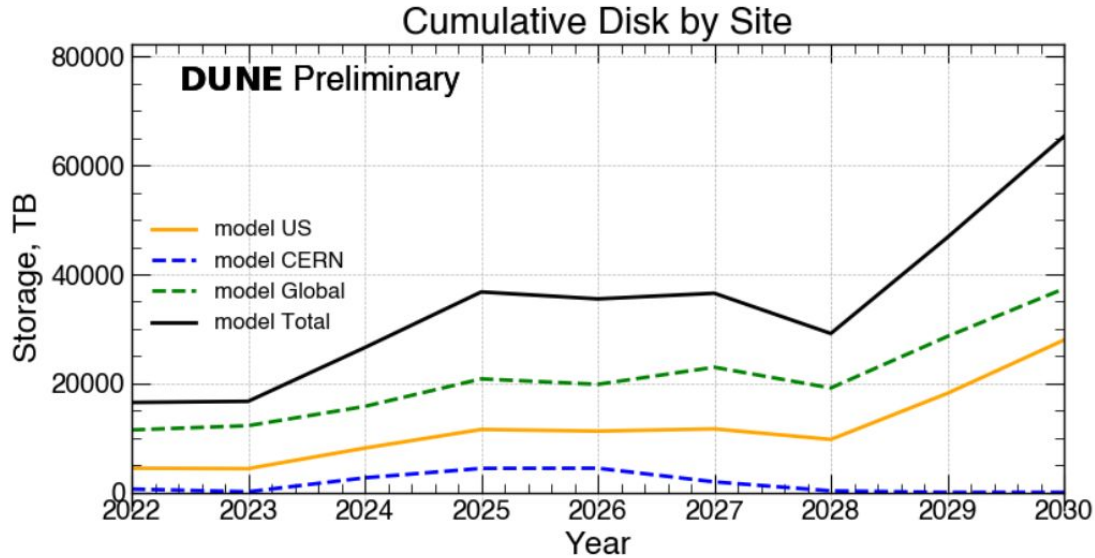
Table 3: Lifetimes and number of copies for different kinds of data. An exception, we assume protoDUNE raw radata will stay on disk for up to three years for reprocessing. Far detector data are assumed to stay for 2 years.

Proposed divisions between nations/labs

Detector Class	Data Type	US	CERN	Global	
	PD	Raw-Data	0.5	0.5	0.0
	PD	Reco-Sim	0.25	0.0	0.75
	PD	Reco-Data	0.25	0.0	0.75
	PD	Test	0.5	0.5	0.0
	PD	TP	0.5	0.5	0.0
	FD	Raw-Data	0.5	0.0	0.5
	FD	Reco-Sim	0.25	0.0	0.75
	FD	Reco-Data	0.25	0.0	0.75
	FD	Test	0.5	0.0	0.5
	FD	TP	0.5	0.0	0.5
	ND	Raw-Data	0.5	0.0	0.5
	ND	Reco-Sim	0.25	0.0	0.75
	ND	Reco-Data	0.25	0.0	0.75
	ND	Test	0.5	0.0	0.5
	ND	TP	0.5	0.0	0.5

Table 4: Assumptions about splits of Disk resources between the US, CERN and Global.

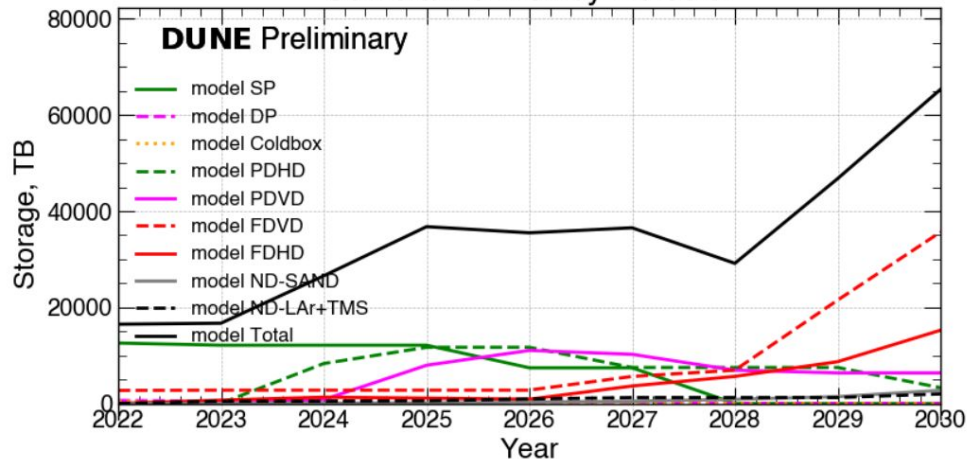
Projected Disk Needs



Locations	2022	2023	2024	2025	2026	2027	2028	2029	2030
US	4432	4359	8126	11548	11245	11652	9716	18171	27951
CERN	612	112	2670	4405	4436	1930	280	0	0
Global	11459	12240	15768	20828	19827	22964	19158	28616	37369
Total	16504	16713	26565	36782	35508	36547	29156	46787	65320

Figure 9: Cumulative Disk by site.

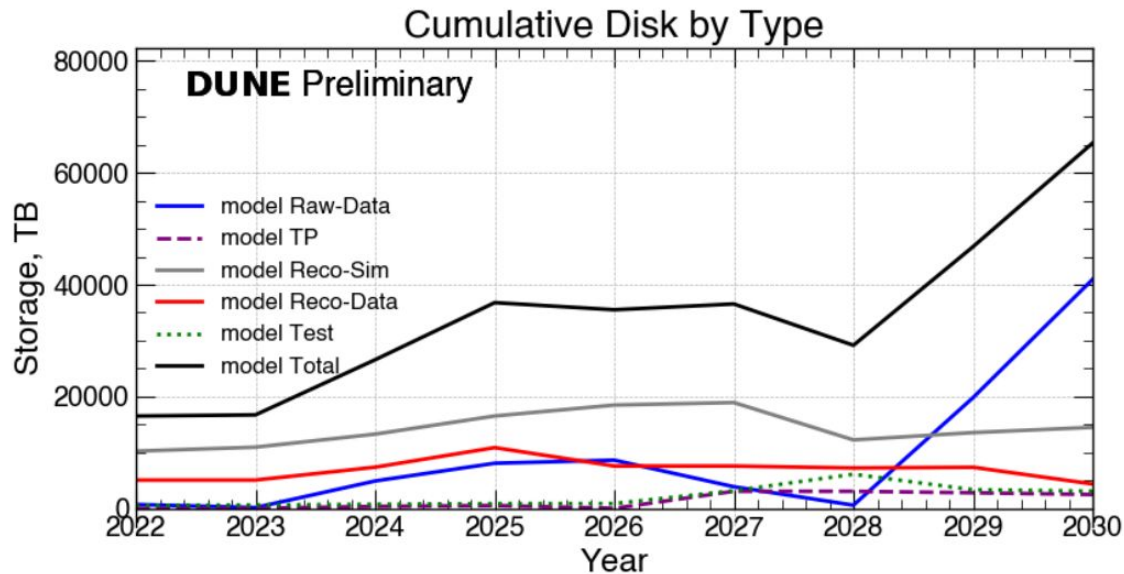
Cumulative Disk by Detector



- Shows the extensions for analysis.
- Normally 2-3 years after last reconstruction/simulation run.

Detectors	2022	2023	2024	2025	2026	2027	2028	2029	2030
SP	12587	12132	12132	12132	7435	7435	0	0	0
DP	652	485	485	355	355	355	0	0	0
Coldbox	468	20	75	0	0	0	0	0	0
PDHD	4	71	8323	11700	11700	7500	7500	7500	3300
PDVD	0	2	824	7961	11050	10239	6939	6378	6378
FDVD	2736	2792	2792	2792	2792	5642	6942	21513	35651
FDHD	54	707	1331	1128	900	3650	5650	8708	15241
ND-SAND	0	0	100	137	325	475	875	1437	2750
ND-LAr+TMS	0	500	500	575	950	1250	1250	1250	2000
Total	16504	16713	26565	36782	35508	36547	29156	46787	65320

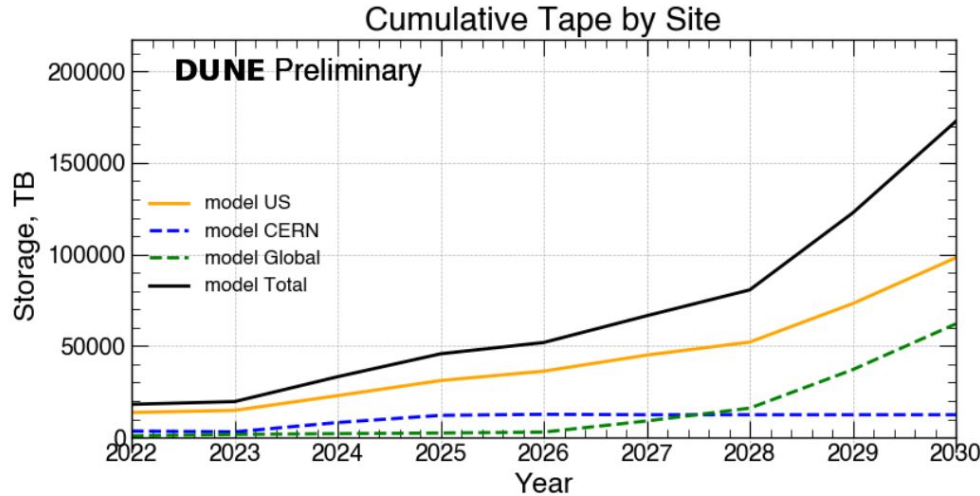
Figure 7: Cumulative Disk by detector.



DataTypes	2022	2023	2024	2025	2026	2027	2028	2029	2030
Raw-Data	687	130	4891	8061	8622	3861	561	19794	40951
TP	0	0	334	500	0	3050	3050	2779	2469
Reco-Sim	10227	10935	13269	16511	18469	18919	12233	13545	14483
Reco-Data	5052	5052	7356	10860	7567	7567	7212	7344	4352
Test	537	595	715	850	850	3150	6100	3323	3064
Total	16504	16713	26565	36782	35508	36547	29156	46787	65320

Figure 8: Cumulative Disk by data type.

Tape needs by site

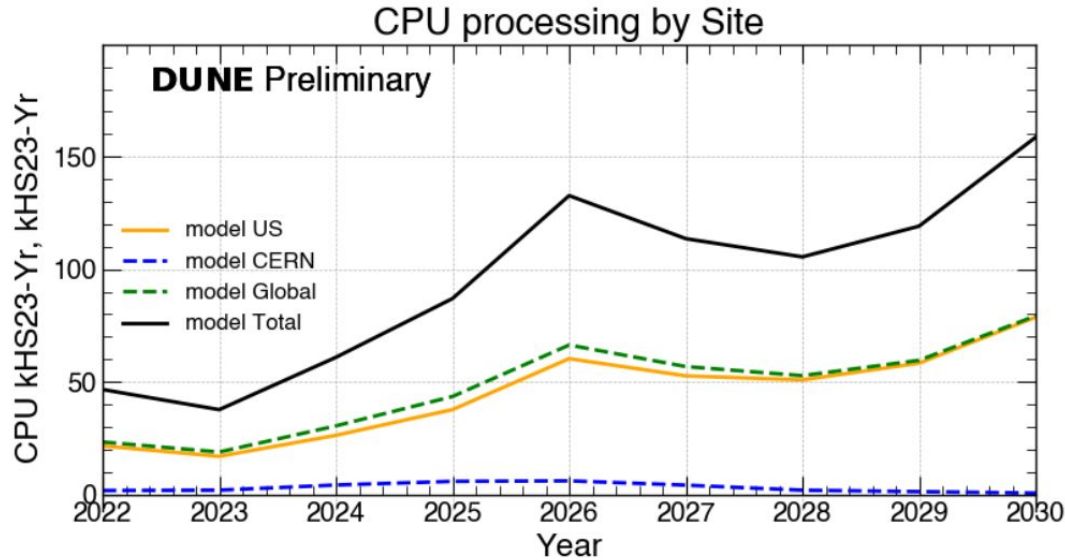


- The “Global” tape need is for reconstructed/simulated data from FD/ND.
- Doesn’t have to be there immediately but we need to prepare for start of data in 2029.

Locations	2022	2023	2024	2025	2026	2027	2028	2029	2030
US	13692	14831	22884	31130	36255	45016	52126	73151	98140
CERN	3522	3080	8195	12130	12691	12441	12441	12441	12441
Global	930	1727	2137	2474	2962	9049	16057	37082	62071
Total	18145	19639	33217	45735	51908	66507	80624	122675	172654

Figure 12: Cumulative Tape by site.

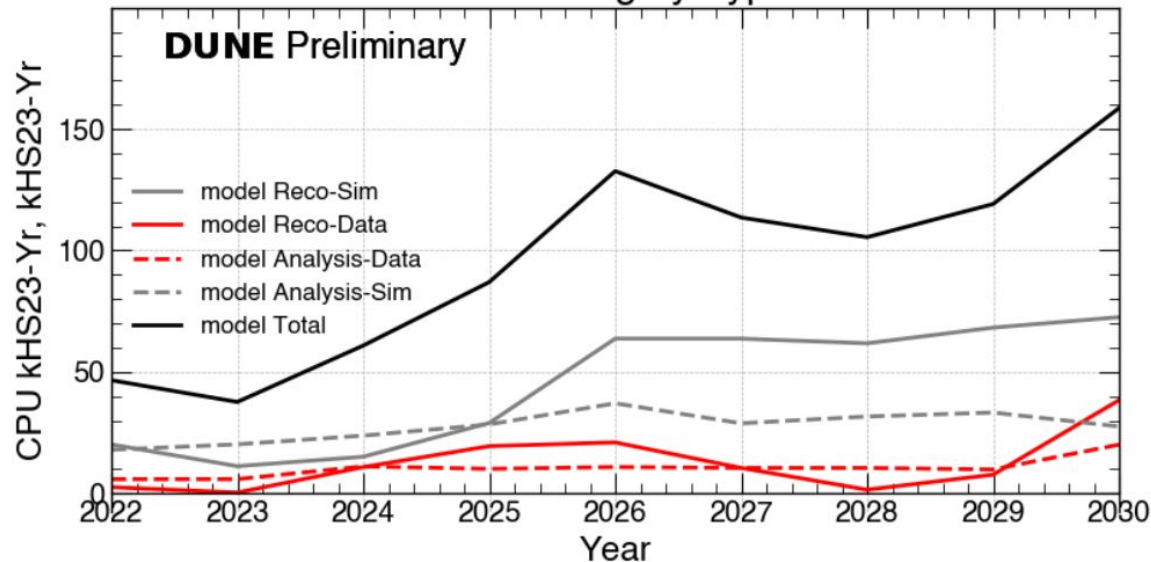
Processing needs by site



Locations	2022	2023	2024	2025	2026	2027	2028	2029	2030
US	21.6	16.9	26.3	37.7	60.3	52.7	50.9	58.3	78.7
CERN	1.7	1.9	4.2	5.8	6.0	4.1	1.9	1.3	0.6
Global	23.3	18.8	30.4	43.5	66.3	56.8	52.7	59.6	79.3
Total	46.6	37.6	60.9	87.0	132.7	113.6	105.5	119.1	158.6

Figure 15: Projected CPU use by location. Units are CPU kHS23-Yr

Processing by Type



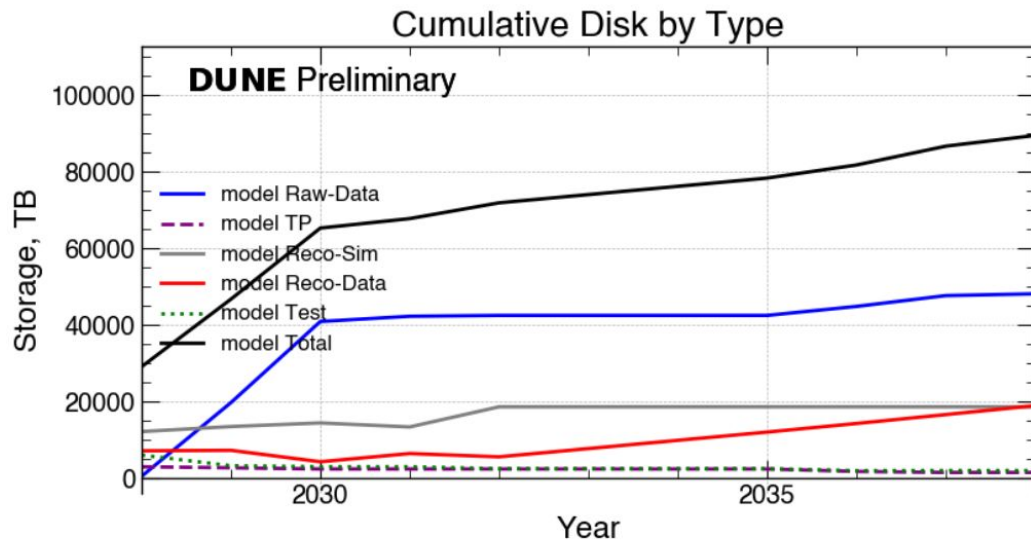
DataTypes	2022	2023	2024	2025	2026	2027	2028	2029	2030
Reco-Sim	20.2	11.2	15.1	29.0	63.7	63.7	61.8	68.3	72.6
Reco-Data	2.6	0.4	10.8	19.5	21.0	10.5	1.5	7.7	38.4
Analysis-Data	5.9	5.9	11.1	10.1	10.9	10.5	10.5	9.9	20.1
Analysis-Sim	18.0	20.2	23.8	28.4	37.1	28.9	31.7	33.3	27.6
Total	46.6	37.6	60.9	87.0	132.7	113.6	105.5	119.1	158.6

Figure 14: CPU kHS23-Yr resources by data types by year.

Long term projections

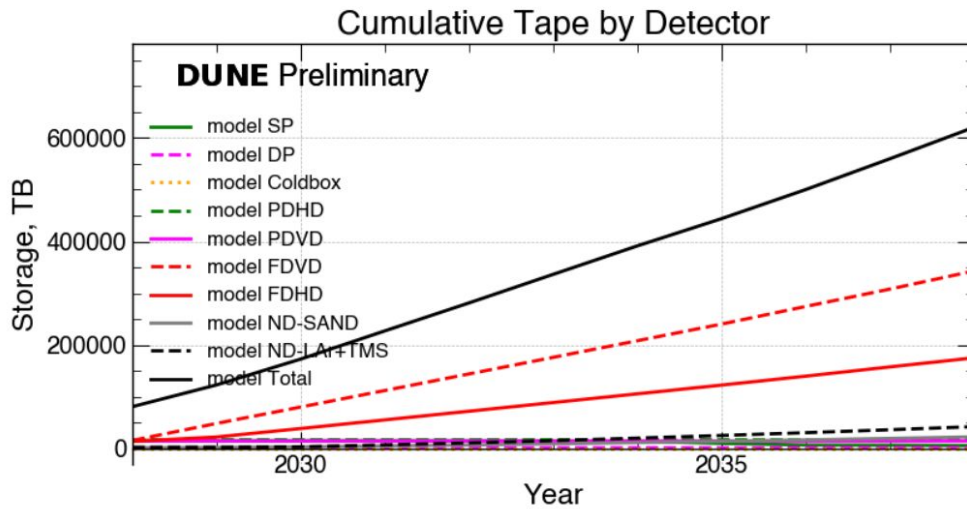
- The model can extend to 2038-40
- Storage needs will be dominated by the Far Detector
 - FD raw data capped at 30 PB/year but still dominates
- CPU/GPU dominated by the Near Detectors
 - 60 interactions/spill
 - More channels
 - Need to 2x2 data/sim to get better estimates
 - Likely to use HPC instead of grid resources
- Analysis processing is currently estimated as a % of reconstruction/simulation CPU

Long term projections



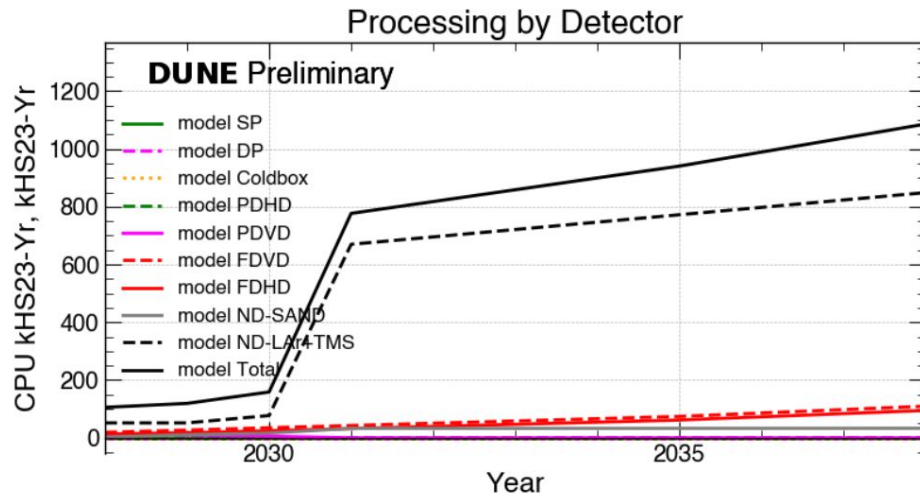
Data Types	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038
Raw-Data	561	19794	40951	42312	42537	42537	42537	42537	44882	47703	48178
TP	3050	2779	2469	2469	2469	2469	2469	2469	1830	1562	1562
Reco-Sim	12233	13545	14483	13442	18692	18692	18692	18692	18692	18692	18692
Reco-Data	7212	7344	4352	6504	5644	7796	9948	12100	14340	16668	18996
Test	6100	3323	3064	3064	2564	2564	2564	2564	2030	2062	2062
Total	29156	46787	65320	67792	71907	74059	76211	78363	81776	86687	89491

Figure 22: Cumulative Disk by data type.



Detectors	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038
SP	15021	15021	15021	15021	15021	14640	13579	9600	6428	6090	5347
DP	1208	1208	1208	1208	1208	1208	1152	1063	975	900	859
Coldbox	0	0	0	0	0	0	0	0	0	0	0
PDHD	16618	16618	16618	16618	16618	16618	16618	16618	16618	16618	16618
PDVD	15661	15661	15661	15661	15661	15661	15661	15661	15661	15661	15661
FDVD	15289	48240	79916	111900	143928	176001	208117	240277	274381	307926	343049
FDHD	13624	21974	38376	55076	71819	88607	105438	122313	139600	157505	175046
ND-SAND	1400	1900	2950	4900	6500	8750	11150	13700	16400	19250	22250
ND-LAr+TMS	1800	2050	2900	6800	10500	15000	19800	24900	30300	36000	42000
Total	80624	122675	172654	227187	281259	336487	391518	444135	500365	559953	620833

Figure 24: Cumulative Tape by detector.



Detectors	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038
SP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DP	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Coldbox	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDHD	7.8	3.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDVD	11.0	9.4	6.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
FDVD	19.0	26.8	34.7	42.6	50.5	58.3	66.2	74.1	85.9	97.7	109.5
FDHD	13.5	17.2	24.6	32.0	39.4	46.8	54.2	61.6	72.7	83.8	94.9
ND-SAND	2.7	10.8	16.2	32.4	32.5	32.6	32.7	32.7	32.8	32.9	33.0
ND-LAr+TMS	51.6	51.6	77.1	670.0	695.6	721.1	746.7	772.3	797.8	823.4	848.9
Total	105.5	119.1	158.6	777.0	817.9	858.8	899.7	940.6	989.2	1037.7	1086.3

Figure 27: CPU kHS23-Yr resources by detector by year.