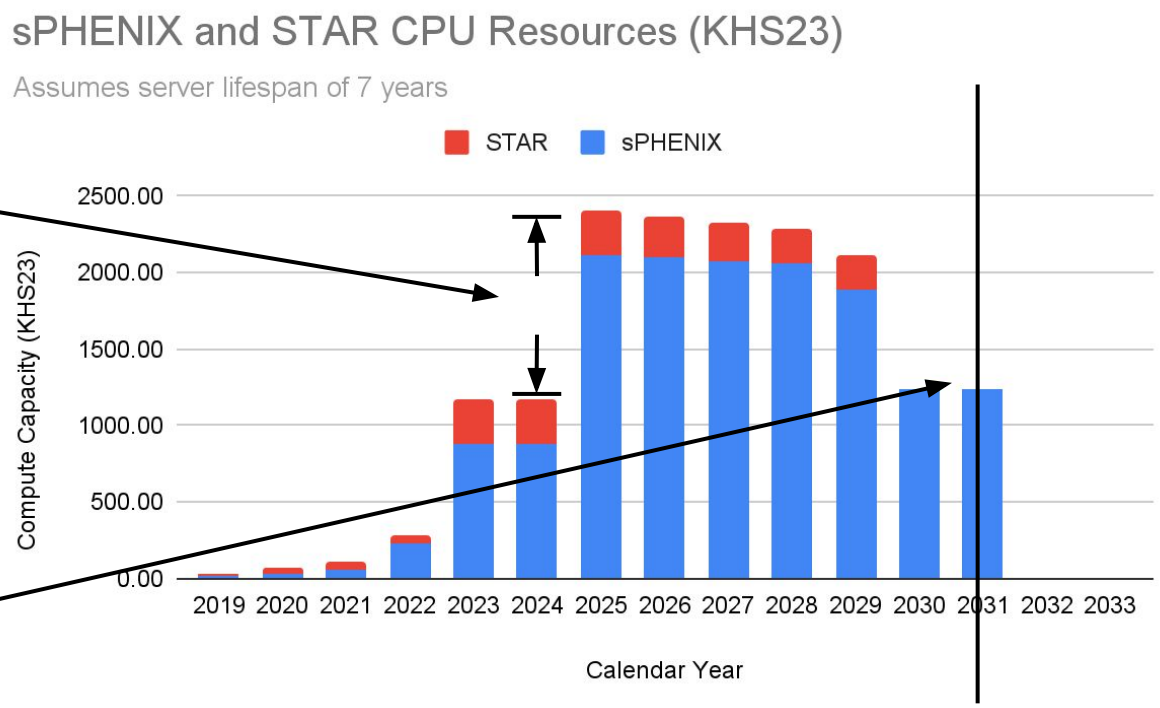
Data Preservation Hardware Resources

May 29, 2025

@BrookhavenLab

Evolution of CPU Resources

- Total CPU resources 2400kHS23
- ~50% of CPU resources purchased in FY25
- Resources split
  - 89% sPHENIX / 11% STAR
- 7 year life span assumed
- All CPU EOL by FY31
- Data preservation assumed to start 5 years after experiment end.

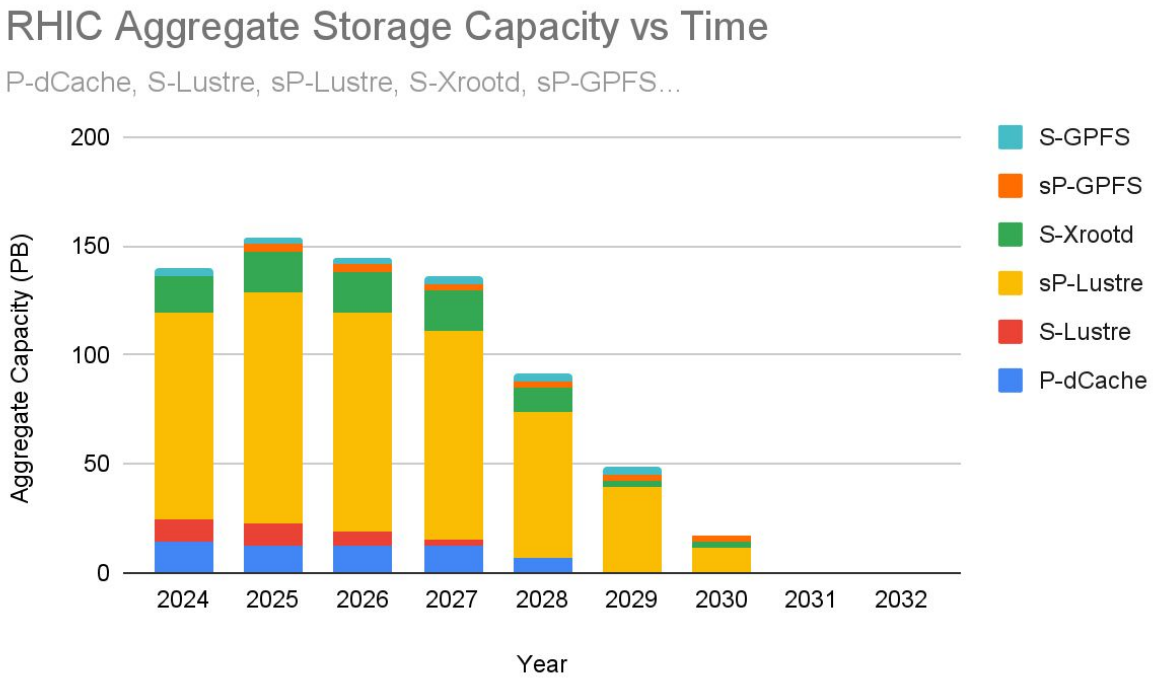


CPU Requirements After 2030

- Assumptions
  - Analysis resources required by STAR and sPHENIX assumed to be equivalent in absolute magnitude
  - Analysis resource requirements during data preservation period 50% of current level.
- Currently STAR analysis utilizes 20% of STAR CPU resources
  - Translates to 52.8 kHS23
- Total CPU requirement > 2030 - 52.8 kHS23
- CPU refresh cost 2031 = $348,000
  - @ $6600/kHS23 (current cost of CPU + network + integration)

Evolution of Disk Resources

- Multiple disk storage systems in use
  - Cost and performance vary by system type
- Total aggregate capacity 154PB
- 5 year life span for storage hardware



Disk Requirements After 2030

- Assume 30PB required for analysis beyond 2030
  - 10PB STAR
  - 20PB sPHENIX
- Refresh required in 2030/2031 time frame
- Cost for 30PB assuming “JBOD” cost:
  - $1.83M @ $61K/PB at current pricing
  - $1.414M @ $47.2/PB assuming 5% per year reduction in $/TB
  - Cost of network and metadata servers not included
- Two parts to “JBOD” costs, evolve at different rates
  - Server
  - Disk

Evolution of Tape Resources

- Expect new LTO generation every 5 years
  - LTO-10 expected this year
  - Typically migrate data to new media every 2 generations (repack)
- In 2030, expect LTO-11 available
  - Expect to repack data onto LTO-11 tapes as part of data preservation
- Expect $/TB for disk falls by 5%/year
  - Currently $5.45/TB dropping to $4.21/TB in 2023
- RHIC data volume
  - RAW - 563 PB
  - DST - 187 PB

Media Cost: Multiple Scenarios

	# Tapes	Cost
DST	2605	$790,209.00
2xDST	5210	$1,580,418.00
RAW + DST	10422	$3,161,370.66
2xRAW + DST	18240	$5,532,532.32
2xRAW+2xDST	20845	$6,322,741.32

Note RAW data volume ~ 2xDST volume

- RHIC IBM Library capacity ~32K cartridges, in Bldg 725
- RHIC Oracle Libraries ~ 70K cartridges - EOL risk, high tape drive cost, in 515

Tape Drive Cost

- 45 LTO-11 tape drives required to copy RAW data in one year assuming LTO-6/8 drive read performance @ 400MB/sec
- 15 LTO-11 tape drives required to copy DST data in one year
- Tape drives ~$5K
  - $225K LTO-11 for RAW
  - $75K LTO-11 for DST
- Additional tape drives required for concurrent data access
- Dual copy doubles tape drive cost*
  - May impact HPSS disk cache cost depending on implementation
- Doubling copy time halves tape drive cost