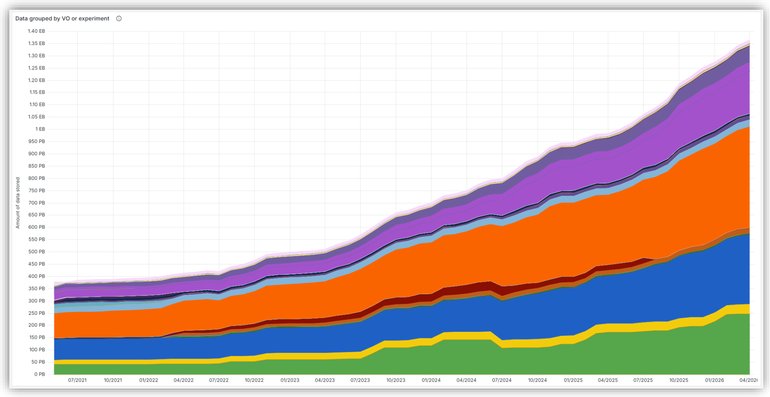
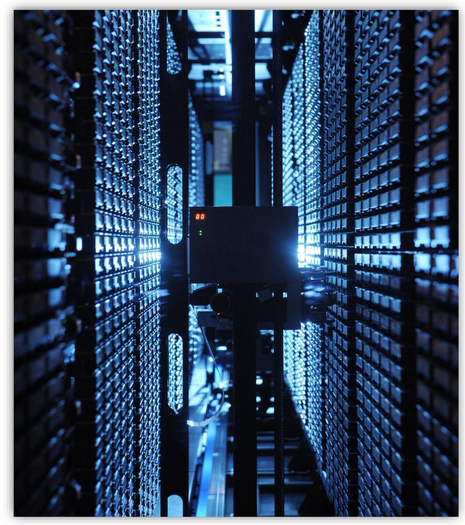
The Green Frontier of Big Data: Environmental Efficiency of Tape Storage at CERN

Vladimír Bahyl
IT Department

Summary

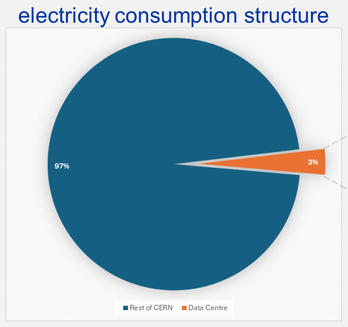
- **CERN Tape Archive** holds over **1.4 EB** of precious physics experiment data.
  - Generates around 300 mt CO_2e per year (80 petrol cars driving 15000 km per year).
  - Consumes around 18 kW (around 15 W per PB).
- **Advantages of** using **tape** technology for data archiving:
  - Energy Comparison: **Power savings** of tape when idle versus constantly spinning disk arrays.
  - Operational **Longevity**: Material lifecycle of 10+ years for tape libraries and 7+ year for tape media reduces electronic waste compared to the usual 5-year server refresh cycles.
  - **Reliability** Metrics: Superior uncorrectable bit-error rates (10 000x more reliable than typical HDD) offer multi-decadal data integrity.
- **Carbon Footprint**:
  - Dominated by the usage pattern and the capacity.
  - Adding capacity will generate more CO_2e (in manufacturing), but not power consumption.
  - Increasing throughput of the archive increase both.
- Modern tape technology indispensable, **eco-friendly** and **cost-effective** backbone for any "Big Science" at the exabyte scale.



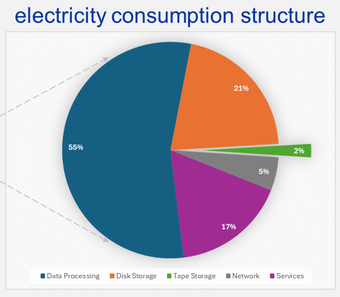


Electricity Consumption Structure

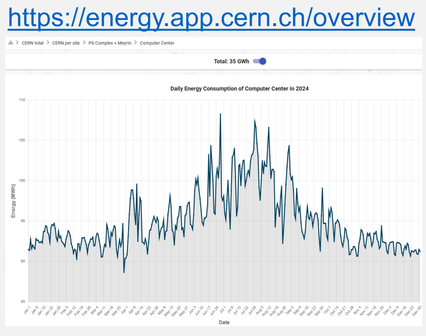
CERN electricity consumption structure



CERN Data Centre electricity consumption structure



https://energy.app.cern.ch/overview

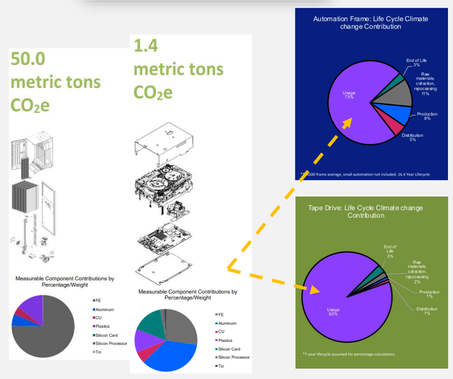


Tape Infrastructure

CERN Tape Archive

- Archive of the physics data
- Provisioned capacity: ~1.63 EB
- Libraries:
  - 4 x IBM TS4500, 3 x Spectra Logic TFinity
- Drives:
  - 40 x IBM TS1170, 46 x IBM TS1160
  - 88 x LTO-9, 10 x LTO-8
- Media:
  - 450 PB on 3592JF, 150 PB on 3592JE, 210 PB on 3592JD
  - 758 PB on LTO-9, 17 PB on LTO-8, 40 PB on LTO-7M

IBM CO2e Report



Environmental footprint

Best case:

- All data would fit into 2 IBM TS4500 tape libraries, each with 15000 slots (IBM TS1170 drives and 3592JF cartridges)

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	2
Number of TS4500 S-frames:	15
Number of TS1100 tape drives:	48
Number of LTO Full High tape drives:	0
Enter number of Cartridges:	15000
Enter type of media:	3592 Media
The total CO_2e contribution for a TS4500 configuration as listed, with all shipping to Europe is:	530 mt
Based on the TS4500 having a 9-year life cycle.	

Reality more complex:

- Throughput requirement – at least 40 GB/s
- Cost optimisation = using cheaper media
- Multiple generations of tape drive technologies
- Protection against library failures / outages
- Avoiding vendor lock-in
- Configuration not constant / Variable growth

IBMLIB1

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	2
Number of TS4500 S-frames:	15
Number of TS1100 tape drives:	0
Number of LTO Full High tape drives:	38
Enter number of Cartridges:	17500
Enter type of media:	LTO Media
The total CO2E contribution for a TS4500 configuration as listed, with all shipping to Europe is:	450 mt
Based on the TS4500 having a 9 year life cycle.	

IBMLIB2

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	3
Number of TS4500 S-frames:	14
Number of TS1100 tape drives:	0
Number of LTO Full High tape drives:	10
Enter number of Cartridges:	1014
Enter type of media:	LTO Media
The total CO2E contribution for a TS4500 configuration as listed, with all shipping to Europe is:	296 mt
Based on the TS4500 having a 9 year life cycle.	

IBMLIB3

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	2
Number of TS4500 S-frames:	15
Number of TS1100 tape drives:	48
Number of LTO Full High tape drives:	0
Enter number of Cartridges:	15394
Enter type of media:	3592 Media
The total CO2E contribution for a TS4500 configuration as listed, with all shipping to Europe is:	534 mt
Based on the TS4500 having a 9 year life cycle.	

IBMLIB4

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	2
Number of TS4500 S-frames:	15
Number of TS1100 tape drives:	48
Number of LTO Full High tape drives:	0
Enter number of Cartridges:	13639
Enter type of media:	3592 Media
The total CO2E contribution for a TS4500 configuration as listed, with all shipping to Europe is:	520 mt
Based on the TS4500 having a 9 year life cycle.	

SPECTRALIB1

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	1
Number of TS4500 S-frames:	12
Number of LTO Full High tape drives:	28
Enter number of Cartridges:	10571
Enter type of media:	LTO Media
The total CO2E contribution for a TS4500 configuration as listed, with all shipping to Europe is:	332
Based on the TS4500 having a 9year life cycle.	

SPECTRALIB2

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	1
Number of TS4500 S-frames:	12
Number of LTO Full High tape drives:	26
Enter number of Cartridges:	11295
Enter type of media:	LTO Media
The total CO2E contribution for a TS4500 configuration as listed, with all shipping to Europe is:	334 mt
Based on the TS4500 having a 9year life cycle.	

SPECTRALIB3

Tape solution being chosen:	TS4500
Geographic region of deployment:	Europe
TS4500 Library frame	1
Number of TS4500 Drive frames:	1
Number of TS4500 S-frames:	12
Number of LTO Full High tape drives:	6
Enter number of Cartridges:	810
Enter type of media:	LTO Media
The total CO2E contribution for a TS4500 configuration as listed, with all shipping to Europe is:	232
Based on the TS4500 having a 9year life cycle.	

- Summary:
  - IBM: 450 + 296 + 534 + 520 = 1800 mt / 9 years
  - Spectra Logic: 332 + 334 + 232 = 898 mt / 9 years
  - Total: 2698 mt / 9 years = 300 mt / year
- 300 mt / 1400 PB = 0.25 mt of CO_2e per PB per year
- 300 mt of CO_2e = ~80 petrol cars driving 15000 km per year
  Source: https://www.epa.gov/energy/greenhouse-gas-equivalencies-calculator

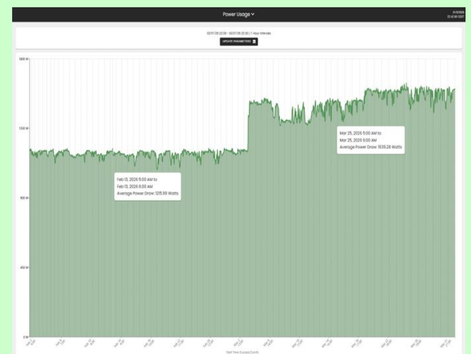
- Comparison: *The Dirty Secret of SSDs: Embodied Carbon* (https://arxiv.org/pdf/2207.10793)
  - Table 1 – 1 TB HDD generates 100 Kg of CO_2e per 5 years = 20 Kg per year
    - This includes server hardware
  - With ~100000 HDDs at CERN, that would be around 2000 mt of CO_2e per year

Energy consumption

- IBM
  - Considering maxiumum values from: https://www.ibm.com/docs/en/ts4500-tape-library?topic=planning-power-consumption-cooling-requirements
  - LTO: IBMLIB1 : 1925 W, IBMLIB2 : 939 W
  - 3592: IBMLIB3 : 3543 W, IBMLIB4 : 3543 W
  - Total IBM (maximum continuous): 9950 W
    Maximum continuous power is consumed when the drives are actively reading or writing to the tape and the cooling fan rotates at normal speed.
- Spectra Logic
  - Considering maximum values as reported by the libraries:
  - LTO: SPECTRALIB1: 3200 W, SPECTRALIB2: 3200 W
  - SPECTRALIB3: 1600 W
  - Total Spectra Logic (estimation): 8000 W
- Total: ~18 kW per 1.2 EB = ~15 W per PB

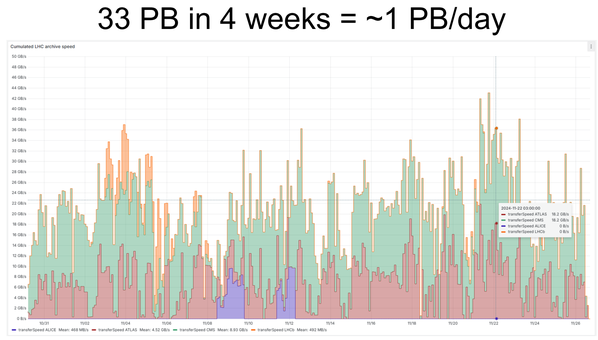
IBMLIB3

Frame model, feature codes, and drives	Quantity	Per unit	Max total
L25	1,00	130,00	130,00
FC 1460 (redundant accessor power/network)	1,00	5,00	5,00
FC 1531, 1532, 1533, and 1534 (one of each FC per frame with drives)	12,00	3,00	36,00
D25	2,00	11,00	22,00
FC 1450 (TS4500 FCA including one LCC and two ac/dc power supplies)	2,00	38,00	76,00
S25	15,00	11,00	165,00
FC 1442 (HA kit with second accessor; for Dx5 or Sx5)	1,00	85,00	85,00
LTO drives	0,00	37,00	0,00
TS1160 drives	48,00	63,00	3024,00
Total library power consumption (watts)			3543,00

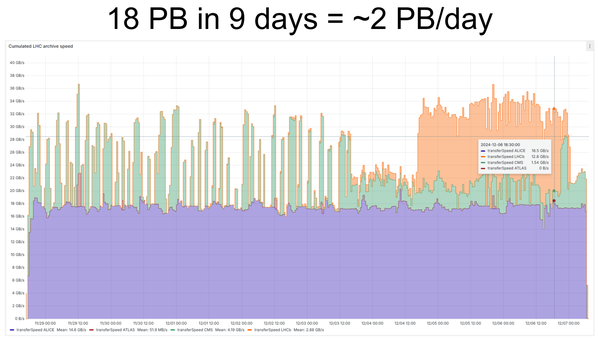


Usage Dynamics

33 PB in 4 weeks = ~1 PB/day



18 PB in 9 days = ~2 PB/day



High Storage Density | Long Archival Lifespan | Reliability | Superior Energy Efficiency (power-off state)
Air-gapped Security | Low Cost-per-Terabyte | Longevity | Massive scalability for massive datasets