Investigating the power and carbon cost of disk storage

Tom Byrne
STFC Scientific Computing Storage architect
Storage Infrastructure Group Leader

Introduction

- We run two CephFS clusters in SCD, **Deneb** and **Arided**
  - Deneb uses 8TB hard disk drives, which provides a balance of affordable capacity with ‘enough’ performance for general-purpose, bulk use
  - Arided uses flash devices, providing performance for more intensive usage patterns, albeit at a much higher cost
- We’re always cautious about increasing HDD size on Deneb
  - Larger HDDs generally provide less performance-per-TB
  - **Our cautiousness places a limit on cost, rack density and power consumption of our HDD based CephFS storage**

Cluster	Description
Deneb	HDD based CephFS storage, supports CLF, RFI, SCD and some IRIS users
Arided	Flash based CephFS storage for STFC cloud users, user provisioned shares via manilla

Science and Technology Facilities Council

In 2024, we procured three types of hardware for our two STFC Cloud CephFS clusters with funding from IRIS, UKSRC (SKA) and STFC

Specification	Nodes	Data devices*	Capacity	Comparative cost**	Funded by
Regular HDD (Deneb)	10	8TB HDD	2.0PB	1	IRIS
Dense HDD (Deneb)	5	22TB HDD	2.7PB	0.6	UKSRC
Dense Flash (Arided)	6	15TB TLC NVMe	2.1PB	2.4	UKSRC & STFC

‘Regular HDD’ was a normal refresh for our Deneb cluster, replacing aging out hardware and increasing capacity

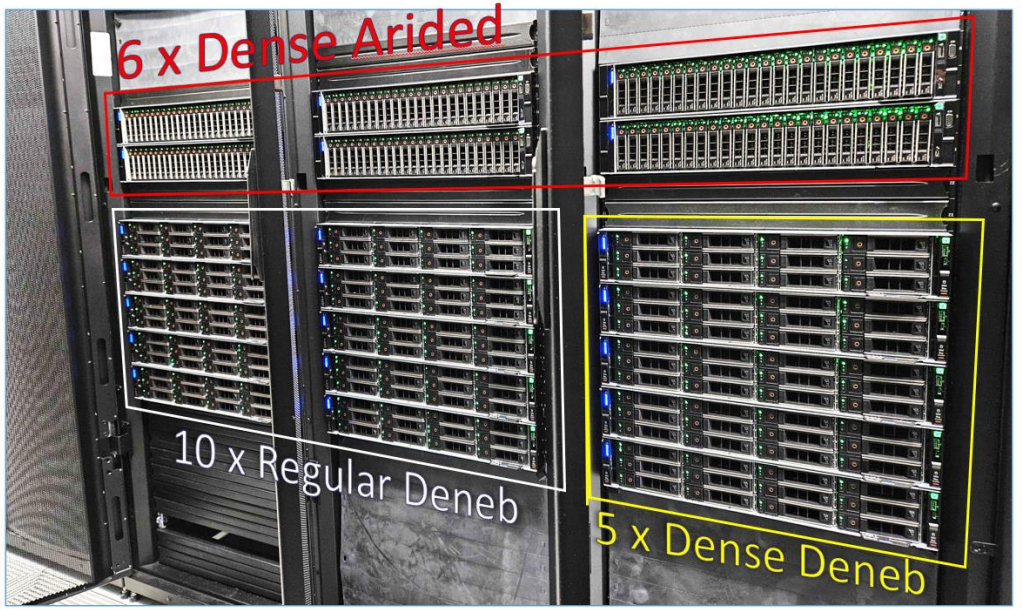
The ‘Dense’ specifications are significantly denser compared to the existing CephFS clusters hardware, and were bought to evaluate suitability of these denser formats for STFC use-cases

* 24 data devices in all nodes
** Comparative cost per unit of usable capacity compared to the Regular HDD spec, for server hardware cost only

I set out to evaluate the suitability of the current Deneb hardware vs…

1. **a denser HDD solution** – can bigger HDDs provide enough performance while being meaningfully cheaper, smaller, more power efficient?
2. **a dense flash based solution** – can we quantify the benefits of moving away from HDD?

Also, a good opportunity to try and understand the power efficiency and carbon cost of running storage…

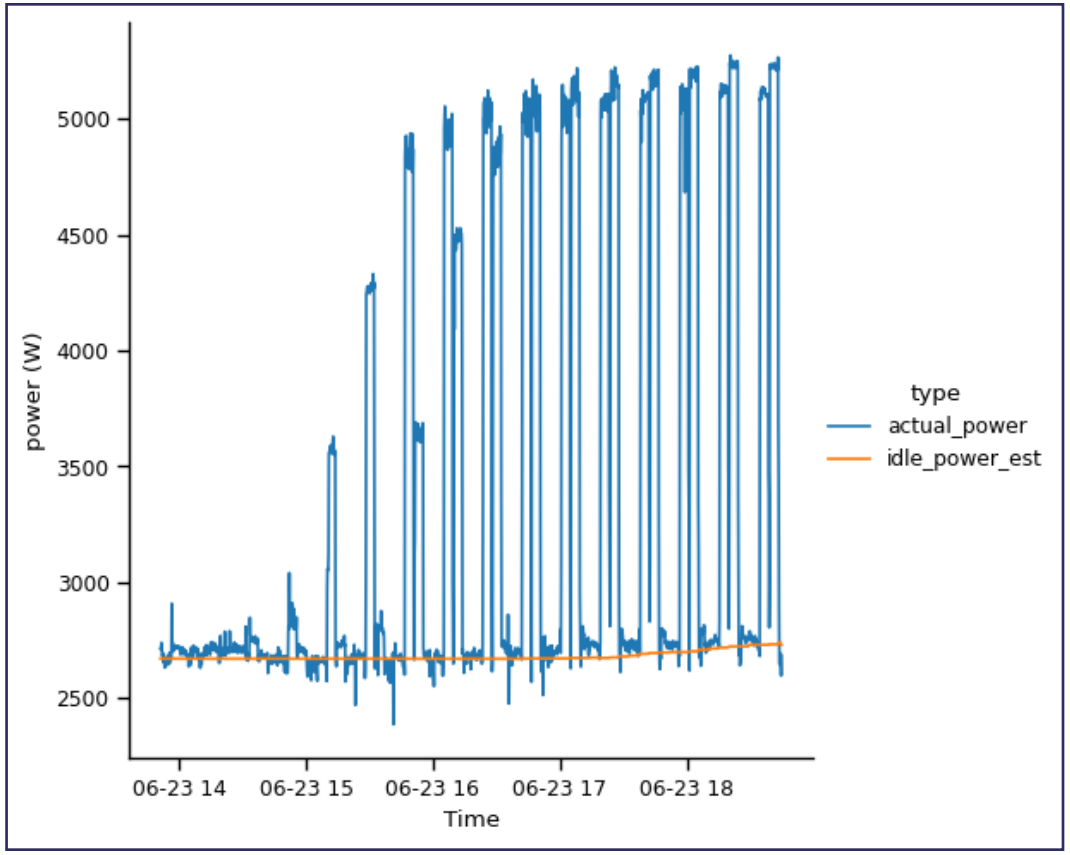


The hardware installed in R89 HPD

Science and Technology Facilities Council

Testing

- Sweeps of number of benchmarking clients to find performance limits
  - Small and large block sizes to identify **IO operations per second (IOPS) limits** and **bandwidth limits**
  - 50:50 mixed reads and writes used in all cases
- Collect power consumption stats from hardware during testing
- **Normalise performance and power usage against usable storage provided to allow comparison**



IPMI measured power usage and estimated idle power during a benchmarking sweep of the dense flash hardware

IO type	Use-cases
Bandwidth	Whole file upload/download, WAN transfers
IOPS	Direct file access by analysis frameworks, database and block device use-cases

Science and Technology Facilities Council

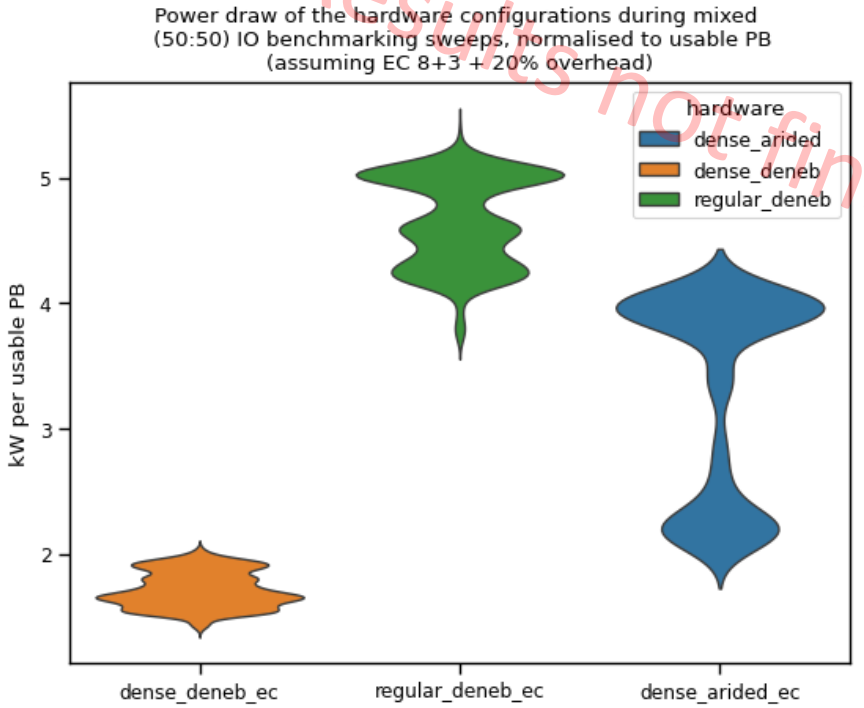
Quantifying power usage

Power efficiency of storage

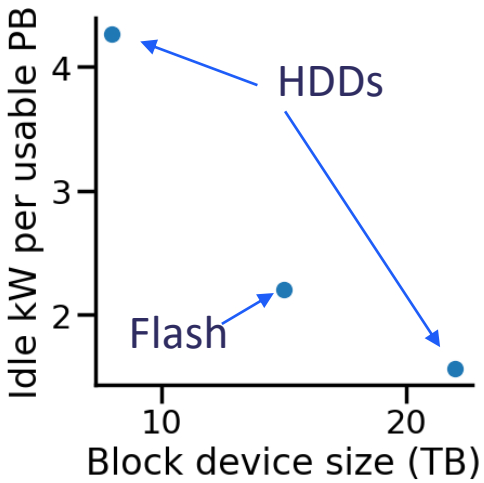
	Power usage (kW/PB usable)	% power increase seen
Dense HDD	1.6 – 1.9	21%
Dense Flash	2.2 – 4.0	80%
Regular HDD	4.3 – 5.0	19%

observations made during a series of mixed IO workload sweeps

- At a first glance, we see a reasonably sensible relationship between idle power consumption and storage density
- The dense flash is reassuringly competitive at ‘idle’ power usages
  - Half the idle power consumption of the regular HDD solution!
  - Larger flash devices potentially more power efficient at idle than the densest HDD solutions available
- The flash solution consumes significantly more *extra* power under load than HDD based solutions
- **Why does flash use so much more power under load, is it less efficient or is something else going on?**
  - Let’s quantify the power required to perform IO



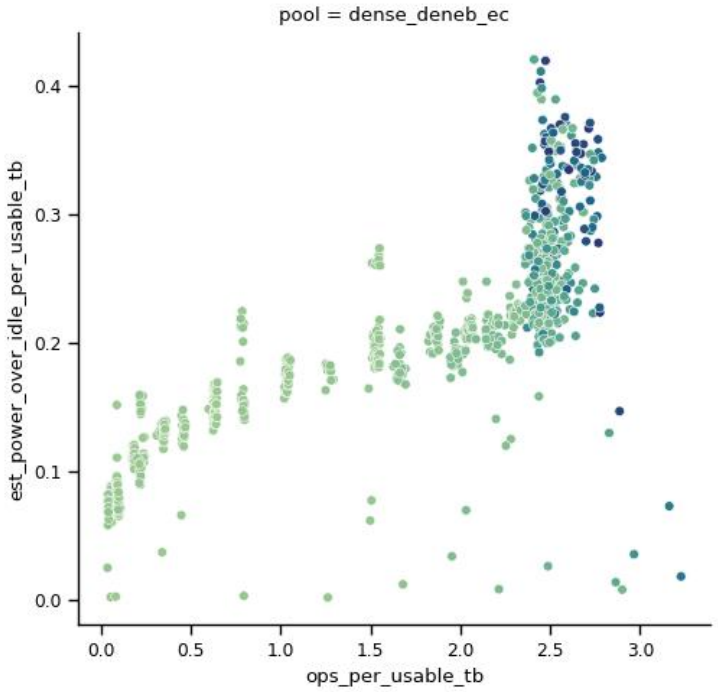
Results not final!



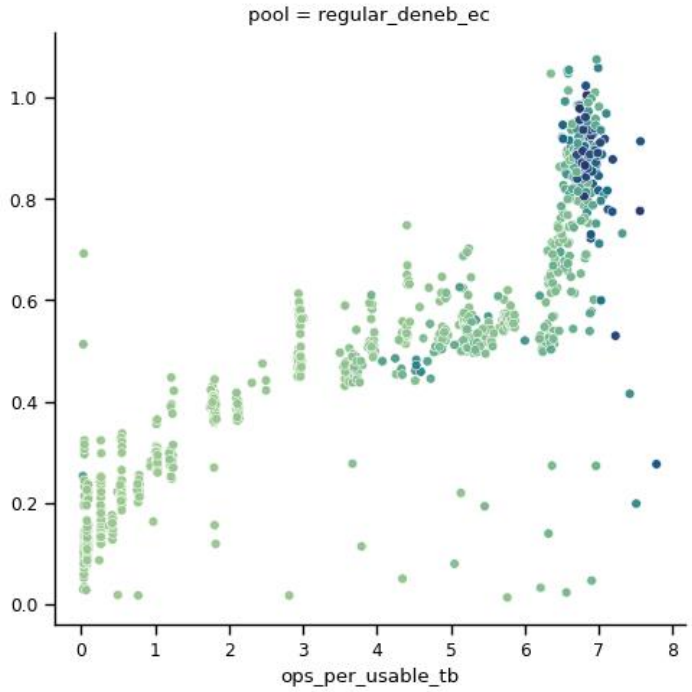
Results not final!

Estimated power above idle (watts per usable TB) against client read IO (IOPS per usable TB)
50:50 r/w mixed 32KB random IO operations
exponential sweep (1 - 4096 client pairs)

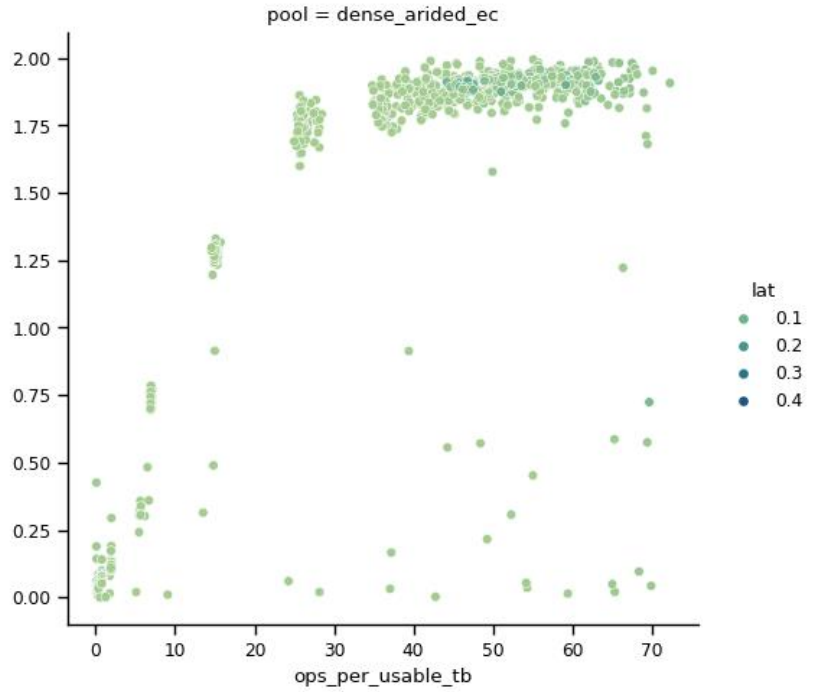
Dense HDD



Regular HDD



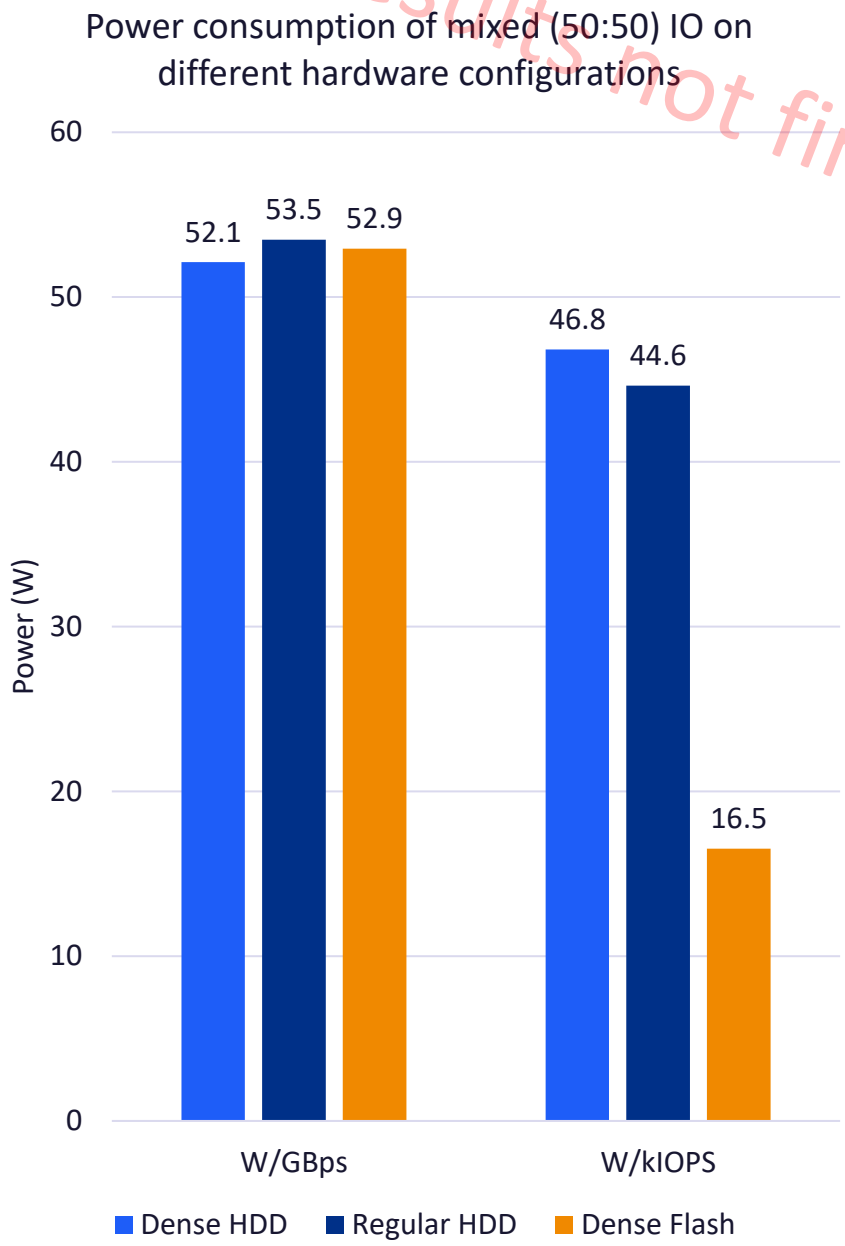
Dense Flash



Science and Technology Facilities Council

Power usage of performance

- The relationship between power used and performance achieved is interestingly non-linear
  - Let's just look at the max for now
- The 3 hardware configurations are well matched for streaming power consumption at full load
  - The difference in power consumption under load is due to the performance envelope rather than any efficiency differences
- The flash solution is substantially more energy efficient for small IO workloads



Science and Technology Facilities Council

The carbon cost of running storage

Results not final!

- Using 2024 UK figures for carbon intensity of electricity, we can estimate the carbon costs of running these ceph storage types
- In most applications, the power used by the data ‘at-rest’ will be the dominant contributor to scope 2 carbon emissions
- Need to consider the other indirect emissions (scope 3) too
  - Embodied carbon of flash vs HDD is an evolving topic**

	Tonnes of CO2 per usable PB per year (at rest)	kgCO2 per PB transferred	kgCO2 per 10^9 IO ops	Tonnes of CO2 per usable PB per year (at max load)
Dense HDD	2.8	3.0	2.7	3.4
Dense Flash	4.0	3.0	0.95	7.3
Regular HDD	7.7	3.1	2.6	9.1

For reference – the average UK car emits 1.7 tonnes of CO2 per year*

All CO2 figures derived using the 2024 UK kg CO2e electricity conversion factor of 0.207 kgCO2e per kWh
https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2024

* https://www.nimblefins.co.uk/average-co2-emissions-car-uk

** https://blog.purestorage.com/perspectives/how-does-the-embodied-carbon-dioxide-equivalent-of-flash-compare-to-hdds/

Thanks for listening

Questions?

Science and Technology Facilities Council