

Building the Business Case for Research Computing Investment

Manveer Munde



Red Oak Consulting
Expert advice, exceptional delivery



Contents

- 1 Review of sources of funding for HPC at UK universities
- 2 Total cost of ownership (TCO)
- 3 Demonstrating the value of HPC
- 4 Strategies to deal with the current climate
- 5 Wrapping up



Review of sources of funding for HPC at UK universities



In the UK there generally four sources for funding HPC

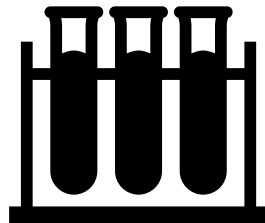
1 Central funding

- Provided by University central budget.
- Indirect cost recovery via overheads.



2 Faculty funding

- Funding taken from faculty-controlled allocation.
- Indirect cost recovery from faculty funds.



3 Large capital item grant

- University, research council or other external funders.
- Direct cost recovery from funder.



4 Research grants

- UK research councils cover 80% of full economic cost (FEC).
- Complexities of claiming VAT further reduce the above.
- Direct cost recovery via grants.



Each funding source has advantages and disadvantages

1 Central funding

- **Advantage:** stable and predictable funding so easy to plan refresh cycles.
- **Disadvantage:** competes with other central priorities.

2 Faculty funding

- **Advantage:** encourages cross-faculty collaboration if pooled; diversifies funding sources.
- **Disadvantage:** disputes over “fair share” contributions.

3 Large capital item grant

- **Advantage:** can enable acquisition of cutting-edge systems which can raise institutional profile (e.g., Tier 2 facilities).
- **Disadvantage:** episodic and irregular funding; operational overheads generally excluded.

4 Research grants

- **Advantage:** the most agile in responding to new needs; diversifies funding sources.
- **Disadvantage:** research councils cover only 80% of FEC, need another funding source to make up the rest.

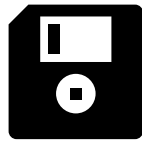


There are various options for direct cost recovery via grants but recovering costs for cluster nodes is complicated

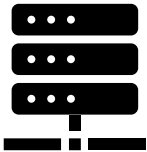
We have observed various approaches to recovering costs via grants for HPC



Recovering costs for people
eg RSEs, project managers



Data storage
(beyond free quota)



Cluster nodes



Priority access
on cluster



Charging for
compute time

Cost recovery for cluster nodes has caused friction with UKRI

- We are aware of cases where UKRI auditors have flagged providing open access to such servers purchased via grants.
- This creates a dilemma:

Allow only private access, with poorly utilised infrastructure and a higher effective cost per core hour.

Ignore the auditors to some extent and risk penalties but have a more efficiently utilised system.

- There is an ongoing dialogue around this matter.



Central funding almost always comes into play in one way or another

- Operational costs usually end up being covered centrally (albeit indirectly/partially).
- Faculty budgets are allocated from central funding.
- Providing free at point of use facilities may necessitate some central funding.
- In our experience, centrally funded models have tended to be the most successful and sustainable.



Total cost of ownership (TCO)



How much does your current HPC really cost?

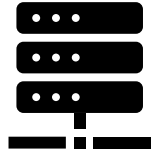
1 Hardware purchase

- The project to buy, install and commission it
- Actual hardware
- Warranty costs
- Support costs (vendor)
- Decommissioning costs



2 Data centre costs (including h/w installation costs)

- Infrastructure costs to enable installation.
- Datacentre running costs, eg staff, property taxes, building insurance.
- Space charge in collocation facility.



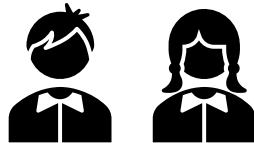
3 Electricity cost (including suitable PUE)



4 Hardware feeding and watering



5 Staff support costs

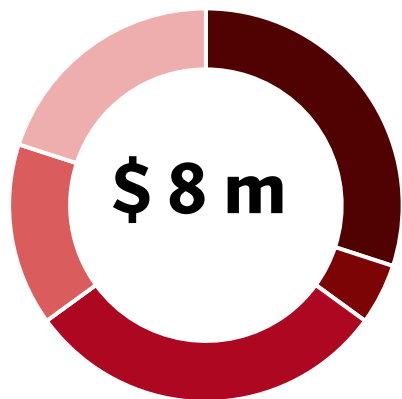


6... (About 20 other costs, including insurance, security, project duration, training, ...)



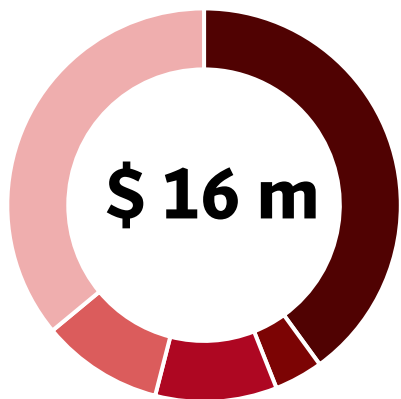
Lifetime on-premises costs for different sites

Site 1: University



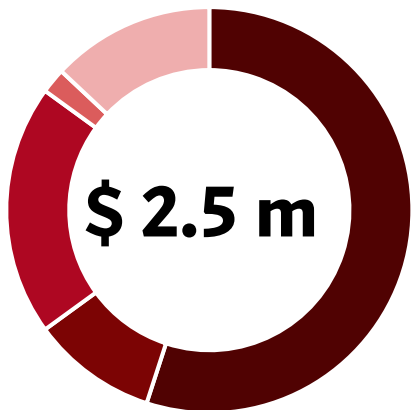
- Hardware
- Procurement costs
- Electricity
- Data Centre
- Other

Site 2: Commercial



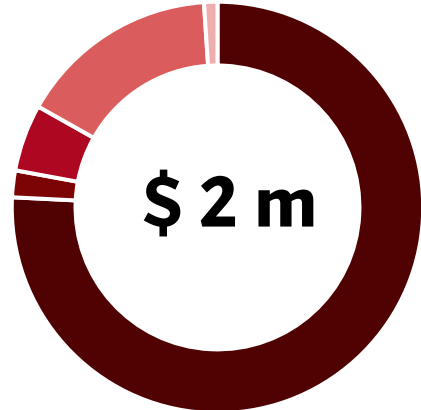
- Hardware
- Procurement costs
- Electricity
- Data Centre
- Other

Site 3: Commercial



- Hardware
- Procurement costs
- Electricity
- Data Centre
- Other

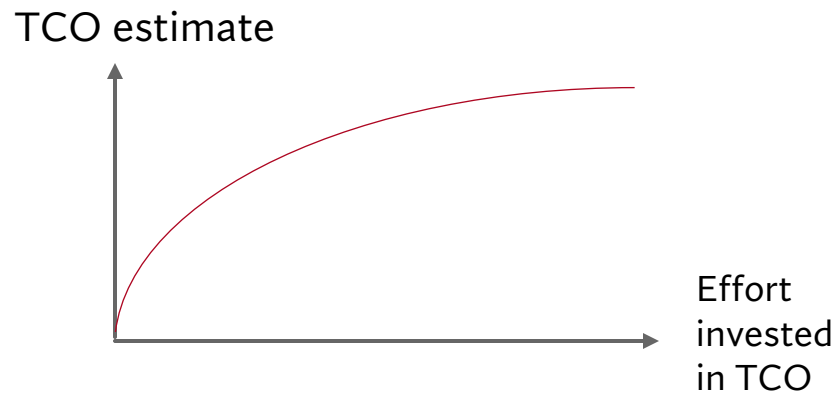
Site 4: Commercial



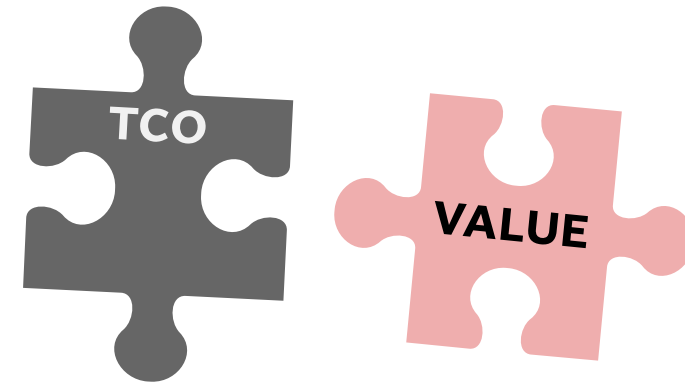
- Hardware
- Procurement costs
- Electricity
- Data Centre
- Other

Your TCO model does not need to be perfect and is only one piece of the puzzle

- There comes a point where additional refinements lead to diminishing returns.
- TCO does by itself does not provide any reason to invest in HPC.



You're aiming to get a reasonable estimate of the amount of funding needed.



Your aim is to convince management that the value of a HPC service exceeds TCO.



Demonstrating the value of HPC



Return on Investment from HPC is both tangible and intangible

Various forms of data can be considered when estimating RoI for HPC

Tangible: provide quantitative data

Research grant income
related to HPC

Research publications
which use HPC

Intangible: can provide good stories/case studies

Expanded social networks

Patents

Spin out companies

Lessons learnt



Telemetry is key: most universities we have worked with collect too little information to provide solid RoI data to management

- Data needs to be collected and shared often with management to affirm the value of HPC.
- Tracking grants and publications is a good place to start.
- Outside of the UK, it is not uncommon to require this data for continued system access.
- Good stories for less tangible benefits also sell the value of HPC.

Our experience suggests that for universities grant income related to HPC can be easily demonstrated to be at least 5x capital costs



Strategies to deal with the current climate



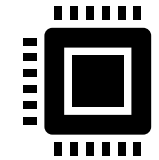
Hardware price increases have significantly decreased the capacity most universities can purchase



AI focus in DRI has led to increase in GPU costs with knock on effects: currently around £40-80k per card.



Memory is most severely affected, forming up to 80% of capital costs in some procurements: up 4-8x since 2025. Is being purchased years in advance.



CPUs price increases due to their importance in agentic workloads (though cost per core hour consistent): up ~2x since 2025.

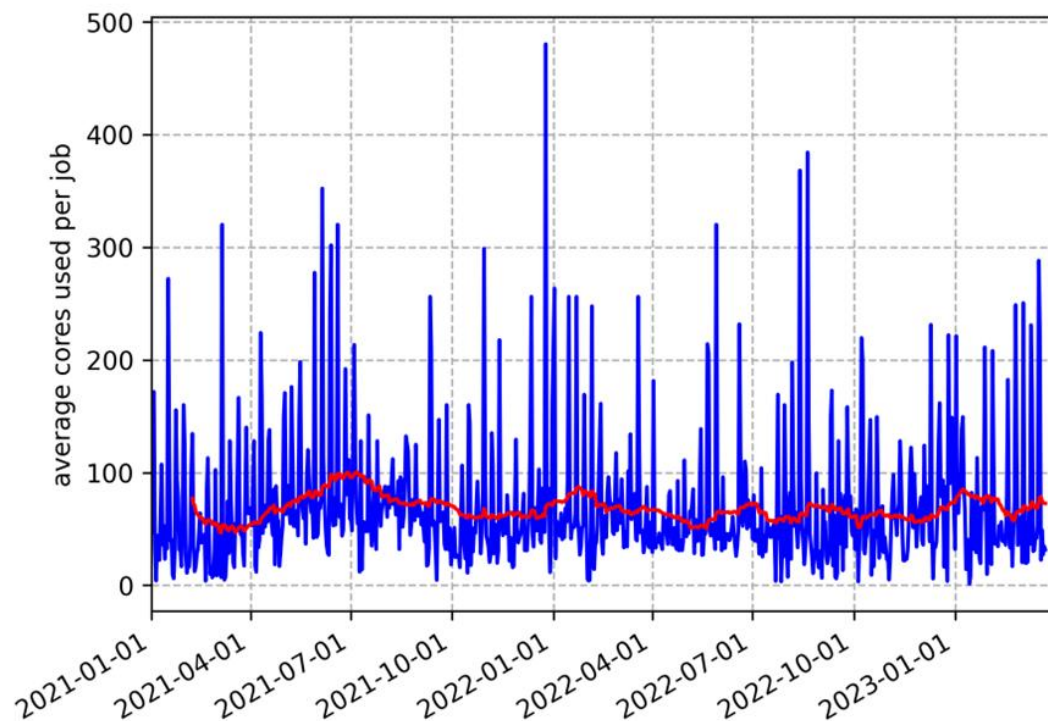


Vendors are not willing to invest significant time in smaller HPC procurements. Quotes valid for days/weeks rather than months.

Increased costs are likely to continue at least until 2028.



How much capacity do you actually need?

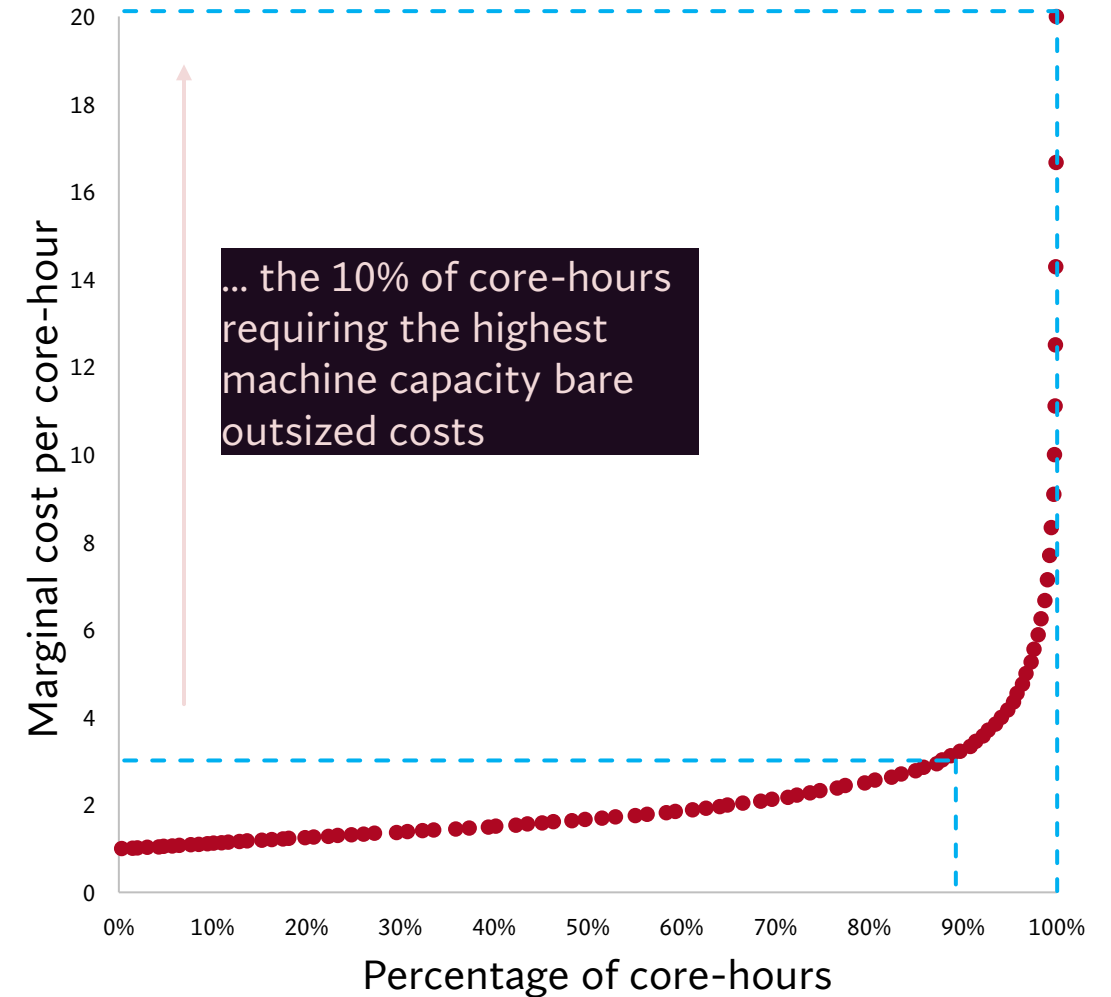
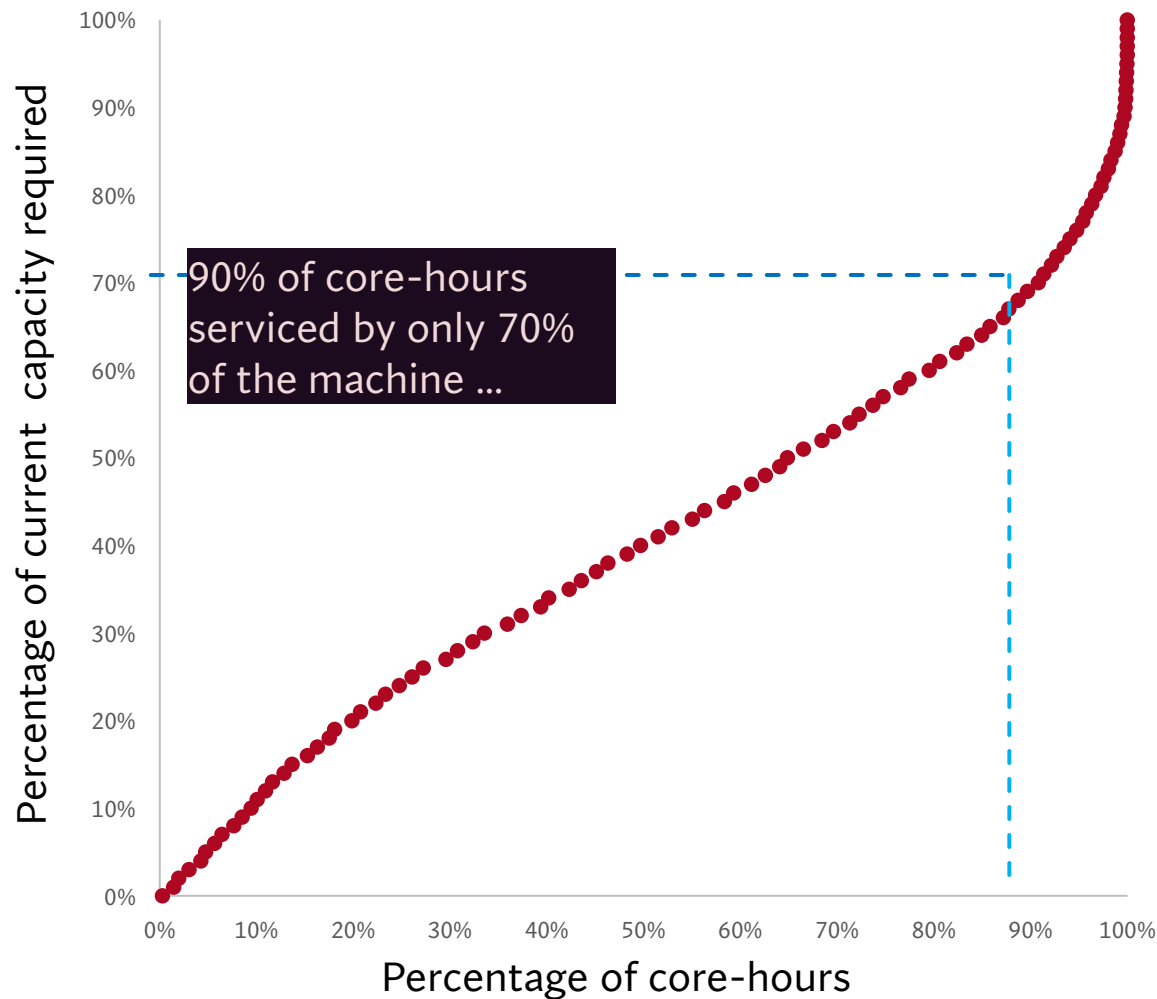


Bins [in MB/core]	Jobs count (%)
0 – 2,000	329,542 (98)
2,001 – 4,000	7,297 (2)
4,001 – 8,000	490 (tr)
8,001 – 16,000	106 (tr)
16,001 – 32,000	12 (tr)
32,001 – 64,000	4 (tr)
64,001 – 128,000	9 (tr)
128,001 – 256,000	3 (tr)



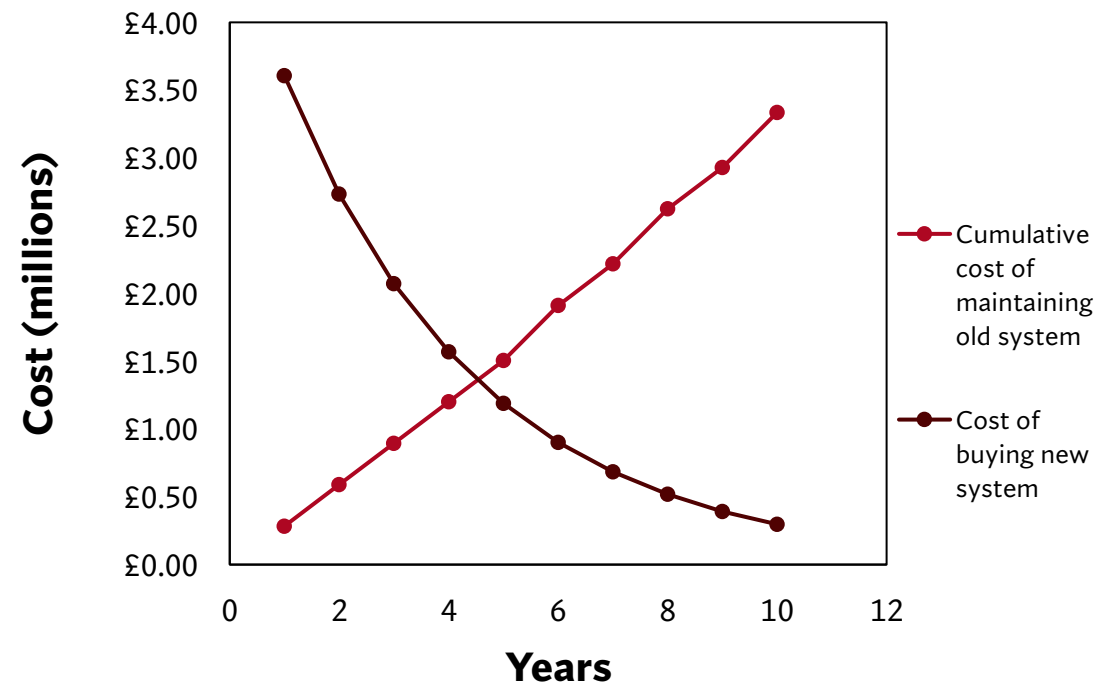
Case for hybrid: when few jobs require significant system capacity and drive up on-prem provision, their relative cost is outsized

A real Red Oak case study...



Do continue to upgrade to maintain capability and competitiveness but keep things affordable and flexible

- Make good use of scheduler logs to analyse capacity needs to avoid oversizing.
- Do take into account capacity growth: a well run HPC system rarely gets smaller.
- Consider a hybrid approach to make up missing capacity when needed.
- Be flexible with technology choices and consider 'agile' small upgrades rather large overhauls.
- Keep things simple for vendors to increase the likelihood of bids.



Maintaining an old system becomes more expensive than purchasing and maintaining a new one.



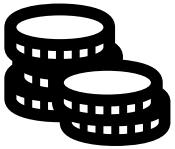
Wrapping up



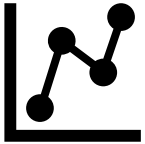
Continuously collate and update management information to ensure adequate funding for HPC



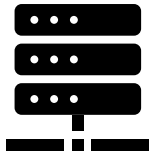
Calculate the TCO of the HPC service



Track RoI to support business case for central funding. Demonstrate that RoI trumps TCO.



Analyse logs for right sizing a new system



Aim to have up to date infrastructure, even at low capacity.





Red Oak Consulting

Expert advice, exceptional delivery

www.redoakconsulting.co.uk

Scan me



Return on Investment from HPC is both tangible and intangible

Research grant income related to HPC

- Count all income from projects dependent on HPC.
- Count line items which specifically mention HPC.



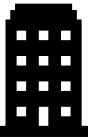
Research publications which use HPC

- Count all papers which cite or use HPC.
- Need to find ways to enforce citation or internally record HPC usage.



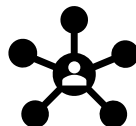
Spin out companies using HPC

- Can count some proportion of investment won.
- No sure way to predict future value generated though methods exist.



Expanded social networks

- A reputable HPC service can lead to collaborations and research grants.
- A greater social network leads to more influence.



Patents supported by HPC

- Income from selling/royalties.
- Further knock on effects from products making use of patent are difficult to predict.



Lessons learnt

- HPC projects can go on to influence organisational policy.
- Difficult to estimate how policy changes affect organisational value.

