Hot Topics Coming to Digital, Data and Technology Department at the University of Bath

Storage, HPC and Cyber Updates
for Research Computing

Slide Show Topics

1.) CURRENT DATA STORAGE LANDSCAPE

2.) INFRASTRUCTURE JOURNEY

3.) WHAT THE FUTURE HOLDS

4.) HOW CAN YOU HELP

5.) CYBER SECURITY'S RECURRENT CHALLENGES

6.) BEST PRACTICE

7.) HPC AT A GLANCE

8.) WHAT TO CONSIDER WHEN CHOOSING

Current Data Storage Landscape

Who?	Storage	Purpose
👩‍🏫 Supervisor/Research Staff	1TB	Research Data
💼 Funded Project	+1TB per project	Project Data
📚 Library	Archive repository	Published papers

Infrastructure Journey

- We have moved from NetApp and CEPH into Nutanix.
  - Energy efficiencies, as they run solid state storage instead of hard drives
  - Nutanix is a Cloud based Computing Company
  - Several times faster than current storage
  - Tiered solution as a possibility and will be explored in the future.

What the future holds

- New Principal HPC Engineer starting in August
  - Review the Research Funding model.
  - Review the research storage policy as well
  - Review of archival options
- Working groups reviewing internal storage options

How will researcher concerns feed into our operations?

- Know how much Data Storage you would require at the beginning, and plan for it
- Know the criticality of your data (archived, backed up outside, etc)
- Give a sensible Data Holding time.
- Who owns the Data? End of Life plan

Cyber Security's Recurrent Challenges

At Bath, Cyber Security is a risk management function within the University

Recently, the main topics that keep coming up and slowing down projects are:

- **AI tools**: Deploying unvetted AI tools (including free versions of approved tools) to analyse study data or draft manuscripts presents a severe risk, potentially exposing sensitive participant information and surrendering intellectual property to public training models.
- **Free software**: Using "free" software is often a gamble. These programs and platforms are rarely your friend when it comes to long term security. Entrusting unpublished data, algorithms or sensitive materials to these software risks exposing this outward. Or even worst, installing vulnerabilities on your device without you noticing.
- **Relying on AI** for data synthesis or content generation exclusively directly threatens research integrity through the risk of "algorithmic hallucinations" and fabricated citations. This is turn results in non-compliance with Research Ethics guidelines.

Best Practice

- **AI tools**: Explore a subscription/paid for approved tool. Currently you can only check by raising a TopDesk ticket, and they would be available to advice if the proposed solution is not suitable.
- **Free software**: “If it looks like a duck, and it sounds like a duck, it is likely a duck.” Check with your IT teams to ensure they are happy to continue and not put the University and its reputation at risk.
- **Relying on AI** Remember AI is a helpful tool which output should be checked and not simply trusted. It should be a tool to assist with work, and not just replace all of it.

HPC at a Glance

- At Bath, high-performance computing is delivered as a portfolio rather than a single machine.

Facility	Positioning	Best suited to	Distinctive benefit
Nimbus	Cloud-based, university-wide HPC and AI platform	General HPC, engineering simulation, AI/ML, GPU computing, burst or variable demand	Elastic access to multiple CPU and GPU families with Pay-As-You-Go and Spot options
Anatra	On-premises specialist HPC/HTC platform	Computational chemistry, life sciences, licensed or cloud-constrained software, high-memory and interactive workflows	Curated partitions, specialist software, local high-memory/GPU nodes and discipline-aligned support
Isambard-3	GW4 shared HPC resource	CPU-based computations	ARM system, free allocation for GW4 partners

What to Consider When Choosing

Research requirement	Preferred starting point	Reason
Needs a flexible choice of CPU or GPU size	Nimbus	Multiple cloud instance families can be selected through partitions.
Large CFD or memory-bandwidth-heavy simulation	Nimbus HB-series	HBv2/HBv3 are designed for memory-bandwidth-intensive parallel workloads.
Compute-dense CPU job such as ORCA or benchmarking	Nimbus HC/Fsv2 or Anatra Chemistry	Choice depends on software/licensing, scale and local workflow requirements.
Modern AI training or GPU inference	Nimbus H100/A10	Flexible modern GPU options and scalable cloud capacity.
Gaussian/GaussView or chemistry-specific licensed workflow	Anatra	Specialist on-premises software environment and Chemistry partitions.
Very large-memory genomics or bioinformatics analysis	Anatra Life Sciences high-memory node	Dedicated 384-core node with 2.2 TB RAM.
Containerised reproducible bioinformatics workflow	Anatra or Nimbus	Anatra documents Apptainer/Anaconda workflows; Nimbus is suitable where broader scale is required.
Short-lived, checkpointable workload where cost is critical	Nimbus Spot	Discounted capacity, accepting possible eviction.

Any Questions?