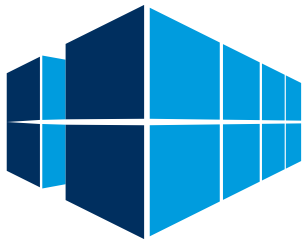
University of BRISTOL

Update on Isambard-AI and Isambard-3

Thomas Green

Bristol Centre for Supercomputing



BriCS

Bristol Centre for Supercomputing

Dirac Day
2024-12-12

Acknowledgements

Isambard-AI is funded by the UK Government's Department of Science, Innovation and Technology (DSIT) via UKRI / STFC.

Isambard 3 is funded by UKRI and EPSRC.

The project has only been possible thanks to very talented and extremely hardworking teams:

- BriCS team
- University of Bristol
- National Composites Centre
- HPE
- NVIDIA
- Contour
- Oakland

https://docs.isambard.ac.uk/acknowledge/

Isambard Facilities

Two different supercomputers, funded differently:

- Isambard-AI: UK AI Research Resource (DSIT via UKRI/STFC)
- Isambard 3: GW4 alliance Tier 2 HPC (UKRI/EPSRC)

Practically very similar:

- Physically co-located in Modular Data Centres
- User experience & sys admin approach

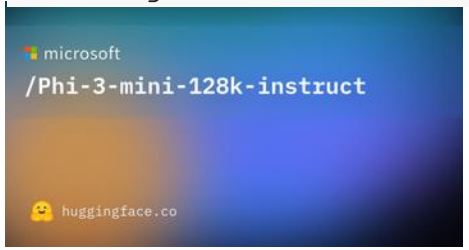
https://docs.isambard.ac.uk/acknowledge/

(ARM based Isambard 1, 2) Isambard 3 in 2023/4 (no data centre and Isambard PI Simon MS with a non-dedicated GW4 team)

Isambard-AI procurement

Isambard hiring started

"Who is Isambard?" running on Isambard-AI



Service configuration and hardening for AI users

2016–July 2023 | Aug. 2023 | Sep.–Oct. 2023 | Nov. 2023 | Dec. 2023 | Mar. 2024 | May 2024 | Aug-Sept 2024

UK govt feedback on Bristol's Isambard-AI proposal

AI Safety Summit

Modular data centre (POD) and Isambard-AI phase 1 installed + a new team in ~3 months

Onboarding Early Users via co-design and technical readiness projects

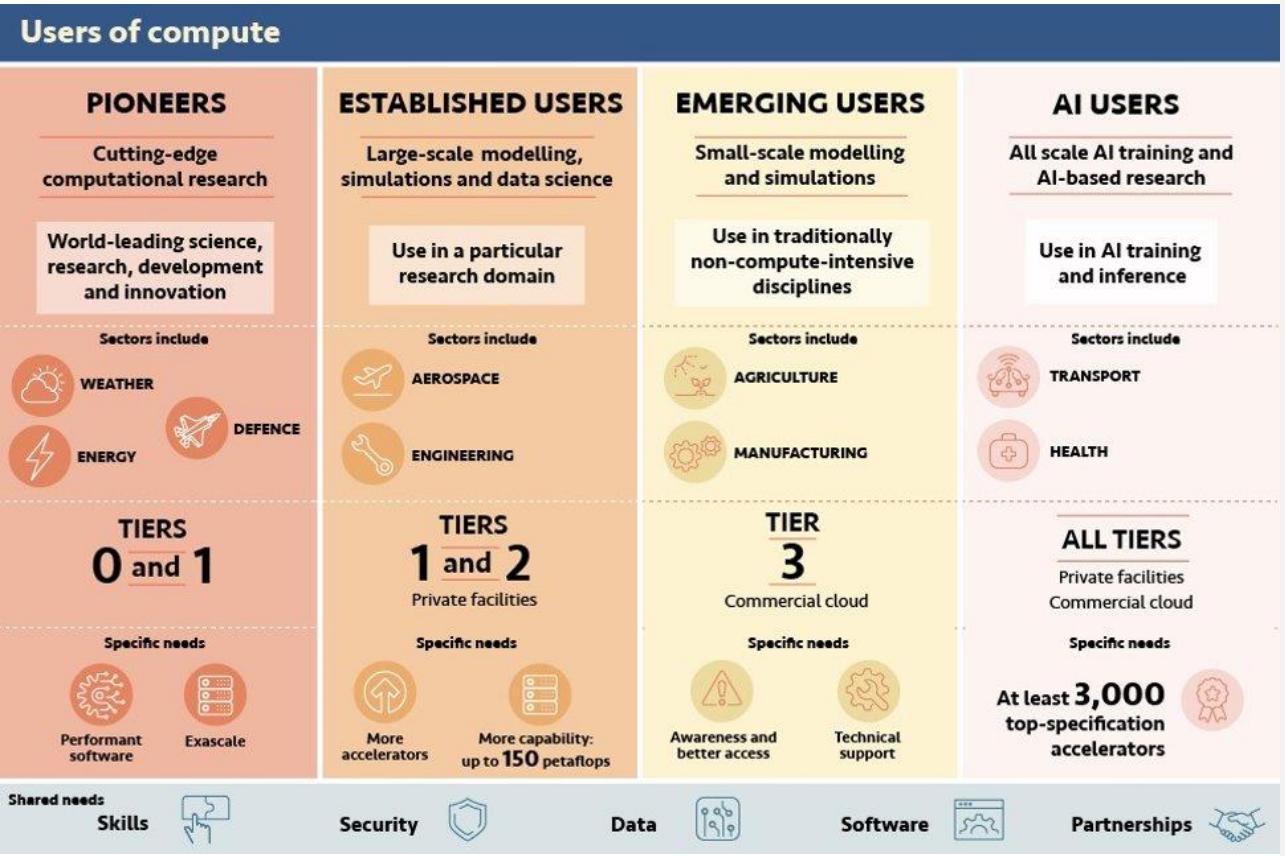
Isambard-AI

A UK national AI research resource

Independent report

Independent Review of The Future of Compute: Final report and recommendations

Updated 6 March 2023



Recommendation 6: Immediately and significantly increase compute capacity for AI research

The AI community has immediate requirements for large-scale accelerator-driven compute to remain internationally competitive and deliver on the UK's ambitions to be an AI superpower. Provision of compute for AI as a first-class use case should also be sustained and provided through future facilities, from exascale through to local clusters.

> **6a) Establish a UK AI Research Resource by summer 2023.**

The government should establish a UK AI Research Resource for immediate use by academic and commercial users within the AI community. It should provide significant accelerator capacity of at least 3,000 top-spec AI accelerators, sufficient to support exploratory compute for every UK AI researcher as well as large-scale training runs, and provide access to a wide range of key datasets and skilled staff to support its use. This should be complementary to existing investments and upgrades in accelerator-driven compute.


December 2023


Isambard 03


May 2024
JCB

Nvidia Grace-Hopper Superchip

- CPU and GPU on a single Superchip
- 72 Neoverse V2 cores
- CPU Memory LPDDR5X
- GPU Memory HBM based
- Further information from:

https://resources.nvidia.com/en-us-grace-cpu/grace-hopper-superchip

Green500

2nd in June, 4th in November

Number 155 in the Top500

Green500 Data

Rank	TOP500 Rank	System	Cores	Rmax (PFlop/s)	Power (kW)	Energy Efficiency (GFlops/watts)
1	189	**JEDI** - BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, ParTec/EVIDEN EuroHPC/FZJ Germany	19,584	4.50	67	72.733
2	128	**Isambard-AI phase 1** - HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11, HPE University of Bristol United Kingdom	34,272	7.42	117	68.835


May 2024
JCB


November 2024


November 2024


November 2024
KING LIFTING
Rocket

Isambard-AI Phase 1 *versus* Phase 2

Currently in "technical preparatory access"

- 42 super-nodes
- 168 GH superchips
  - 12,096 Grace Arm CPU cores
  - **168** Hopper GPUs

- 21.5 TB CPU memory
- 16.1 TB high bandwidth GPU memory
- 37.6 TB total memory

- ...plus ~1 PB all-flash storage

x30!

Arriving early next year

- 1320 super-nodes
- 5,280 GH superchips
  - 380,160 Grace Arm CPU cores
  - **5,280** Hopper GPUs

- 675 TB CPU memory
- 506 TB high bandwidth GPU memory
- 1.18 PB total memory

- ...plus ~27 PB all-flash storage
  - (~20 PB POSIX, ~7 PB object)

Isambard 3

A UK national Tier 2 service

Isambard 3

- Isambard 3 is not Isambard-AI
- Successor to Isambard 2

- ‘Traditional’ Tier-2 HPC
  - Largely CPU resource
  - Anticipate mostly batch

- Housed in Modular Data Centre alongside Isambard-AI
- Sys Admin approach very similar to Isambard-AI

- 384 nodes NVIDIA Grace CPU Superchips:
  - 2x72 cores; 2x120 GB LPDDR5
  - Slingshot interconnect
- MACS nodes
  - AMD Milan, Genoa, Bergamo
  - Intel SPR DDR5, HBM
  - AMD Milan + GPUs
    - 4 x AMD Instinct MI100
    - 4 x NVIDIA A100 SXM4
    - 4 x NVIDIA H100 PCIe

https://docs.isambard.ac.uk/specs/

Nvidia Grace CPU Superchip

- 144 Neoverse V2 cores
- Memory LPDDR5X
- 2 CPUs on a single Superchip
- Further information from:

https://resources.nvidia.com/en-us-grace-cpu/nvidia-grace-cpu-superchip

Air-cooled racks

User interfaces

What will you see?

Self-service Portal

portal.isambard.ac.uk

Individual

Project PI role
- Invited by BriCS on award
- Legally responsible
- Invites/manages users

Project researcher role
- Invited by PI

MyAccessID

1. portal.isambard.ac.uk
   - Accept policies, set user-name
2. Accept invite (per project)
3. Generate SSH key pair (once)
4. Authenticate via clifton app (daily)

https://github.com/isambard-sc/clifton

Projects on Isambard-AI & 3

- Projects are allocations of resource
  - Finite in time
- Individual has user account per project
  - myname.project1, myname.project2
- Projects cannot see each other's files
  - Additional separations for sensitive projects
- Data removed at project expiry

```
Successfully authenticated as myname@example.com (myname) and downloa
for projects:
 - project1
 - project2
 - project3


Certificate file written to /home/user/.ssh/file.pub
Certificate valid for 11 hours and 59 minutes.
You may now want to run `clifton ssh-config write` to configure your SSH config
$
```

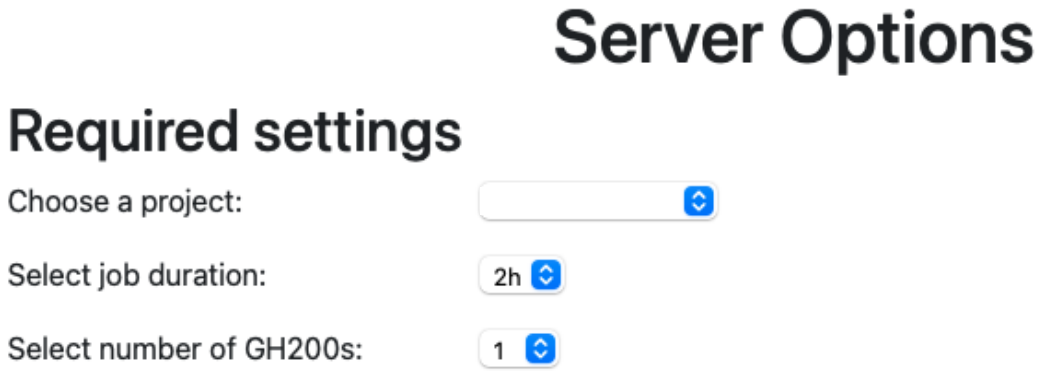
Jupyter notebooks

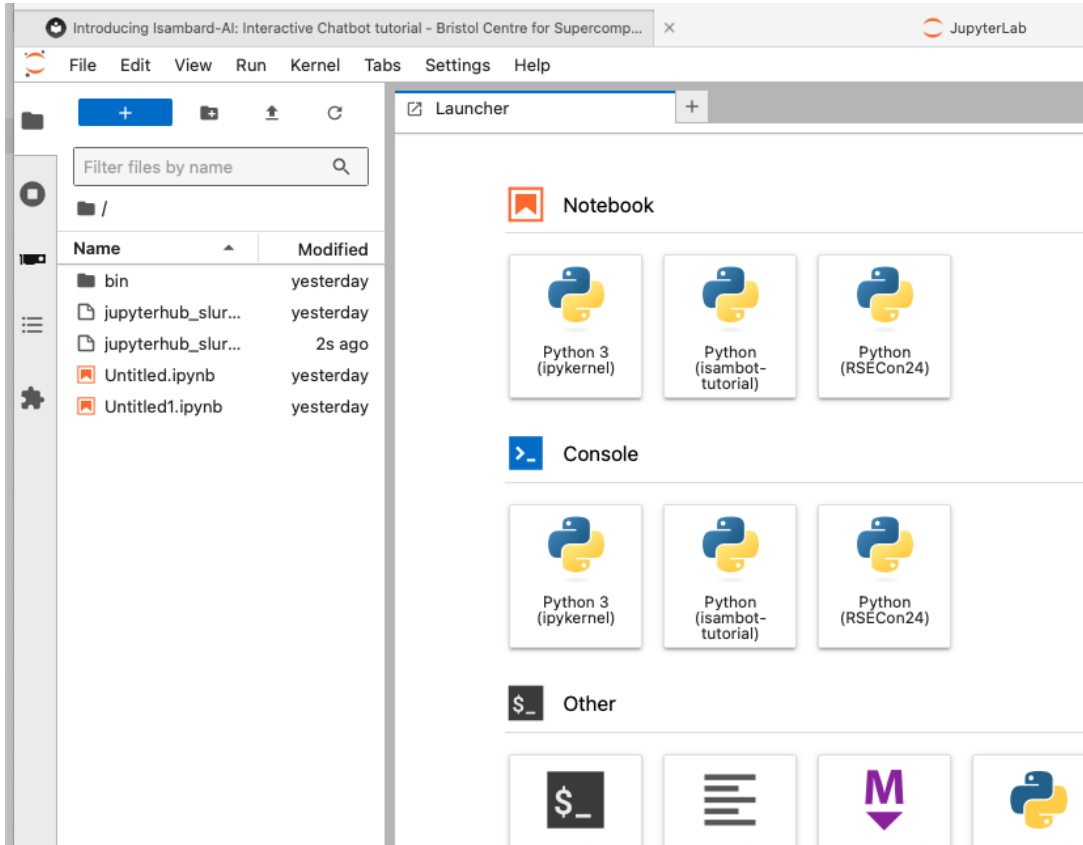
Interactive access to compute- up to 4xGH200

1. apps.isambard.ac.uk/jupyter
2. Choose project & GPUs
3. Do research

- Also VSCode
- Extensible solution for future development

Presently pre-production





Self-service software stacks

Scalable solution for everyone (both users and sysadmins)?

Everyone's needs are different- supercomputing is bespoke!

- Applications (versions, branches, forks)
- Dependencies (good, bad, ugly)
- Skillsets (beginner, improver, expert)

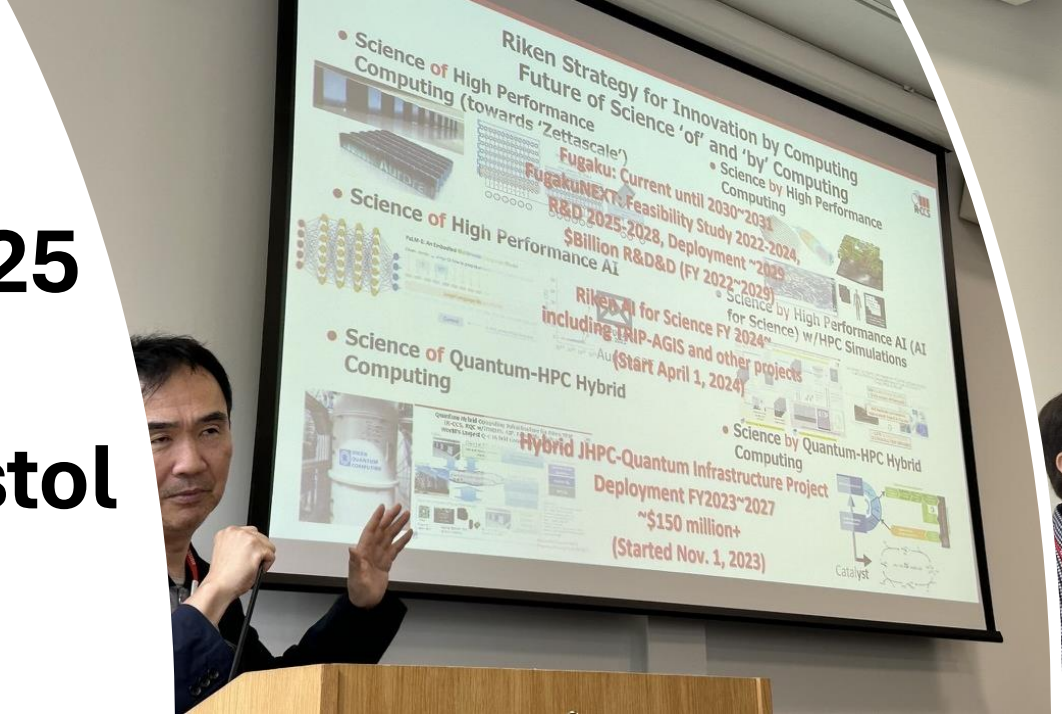
If you can install it on your laptop, why not Isambard-AI or 3?

- Spack, conda, containers etc exist to make things easy
- Users can use (and extend) their own expertise

Isambard Day 2025

11-12 March, Bristol

docs.isambard.ac.uk/events/isambard_day_2025



Thank you

Thomas Green
thomas.green@bristol.ac.uk

brics-enquiries@bristol.ac.uk

Access Calls

- Isambard 3 in UKRI ”Access to HPC”
- Isambard-AI still to be announced.
  - 8-week technical readiness projects:

brics-enquiries@bristol.ac.uk

University of BRISTOL

BriCS
Bristol Centre for Supercomputing