Cloud use case 3: HPC

Presenter

Dr Stefano Angioni

Research Computing Manager

University of Bath

UNIVERSITY OF BATH

University of Bath

Public research university with a strong focus on technology and science

One of the top ten universities in the UK

20,000 students from 147 nationalities

UNIVERSITY OF BATH

On-premise HPC challenges

Replacement every 6-7 years

Fixed and aging system

Long queues slowing down research

Research needs evolve at pace

UNIVERSITY OF BATH

Bath’s ambitious research agenda

Cutting-edge computational facilities

Flexible and scalable resources for changing needs

Grow and diverse user base

Increase capabilities and productivity

UNIVERSITY OF BATH

Turbocharging research with Cloud

Highly flexible and on-demand computing

Computational environments which can evolve

Modern and cutting-edge

Cost-optimised

Increased cost transparency

UNIVERSITY OF BATH

Research Computing Portfolio

Cloud based

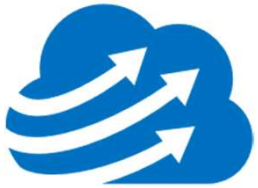
1. Teaching HPC environment
2. Research HPC – phase 1 Janus
3. Research HPC – phase 2 Nimbus

On-premise

4. Restricted Software HTC – Anatra
5. GW4 Tier 2 HPC – Isambard

UNIVERSITY OF BATH

Cloud solution



CycleCloud

Research HPC environments

Phase 1	Lightweight Cloud environment for selected user groups
Phase 2	Functional and resilient platform mirroring Balena Cost control and full software stack
Phase 3	Non-traditional HPC groups, optimising and tuning the environment

UNIVERSITY OF BATH

Janus

Is our phase 1 HPC system in Azure.

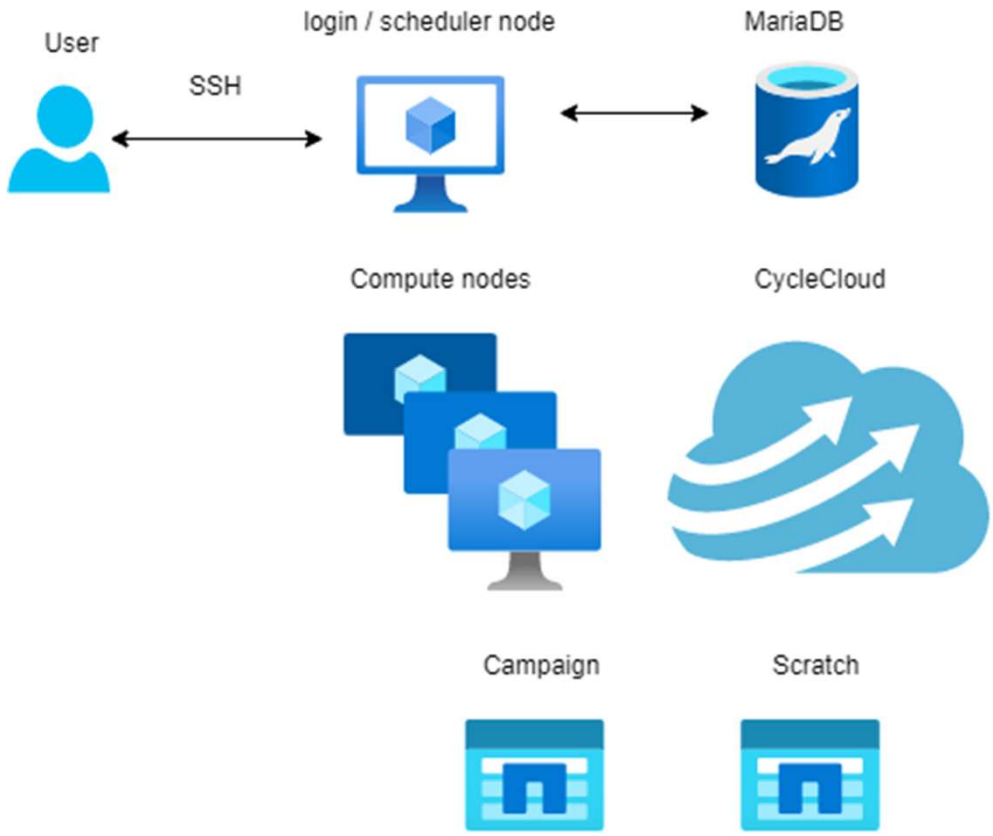
Nodes are not physical, but virtual machines. Nearly 800 compute nodes, but we are only charged if they are used ☺

Cloud resources are orchestrated using Azure CycleCloud.

Users can access Janus using SSH.

The Job scheduling software is Slurm.

Only one login / scheduler node ☹



UNIVERSITY OF BATH

Nimbus

Development guided from learnings from Janus

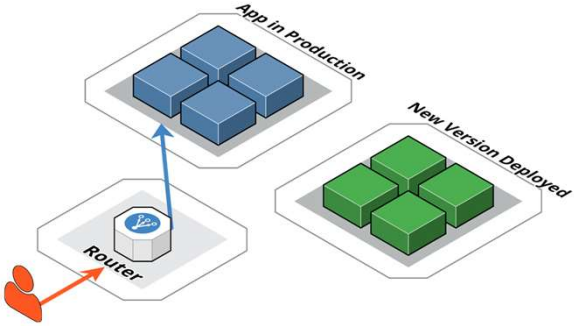
Improved resilience

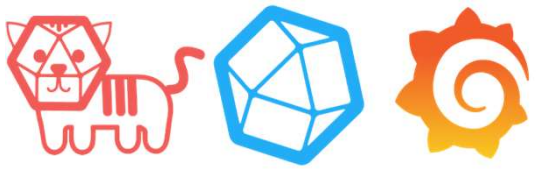
New lifecycle model for max availability

New cluster monitoring with TIG stack

New cost management tools and cost controls

New storage options



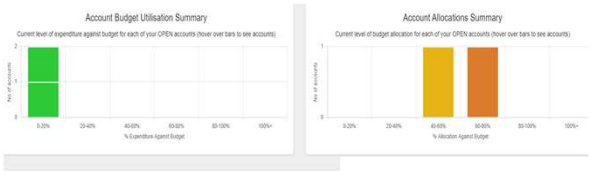


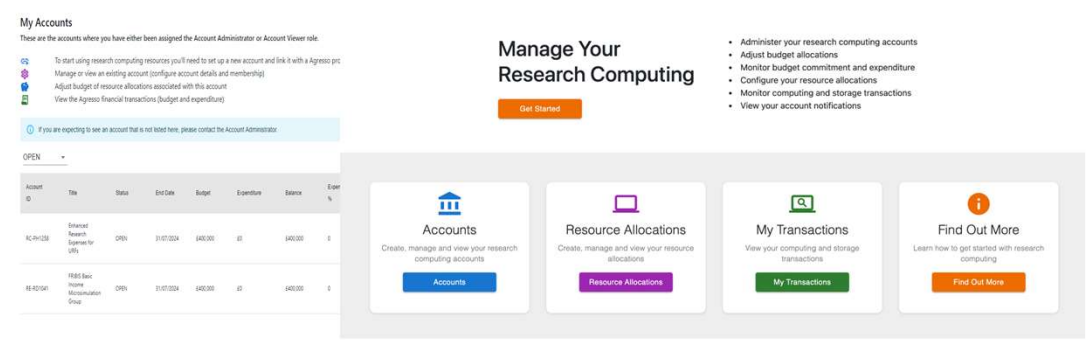
UNIVERSITY OF BATH

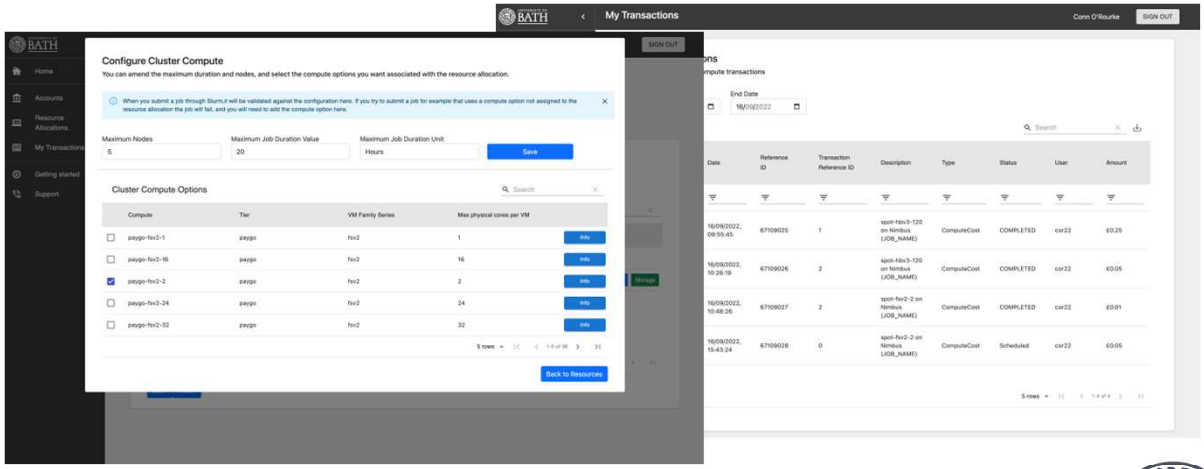
Accounting Portal

The Accounting portal allows users

- to manage / view the budgets associated with their accounts, and
- to allocate resources (compute & storage) to their accounts





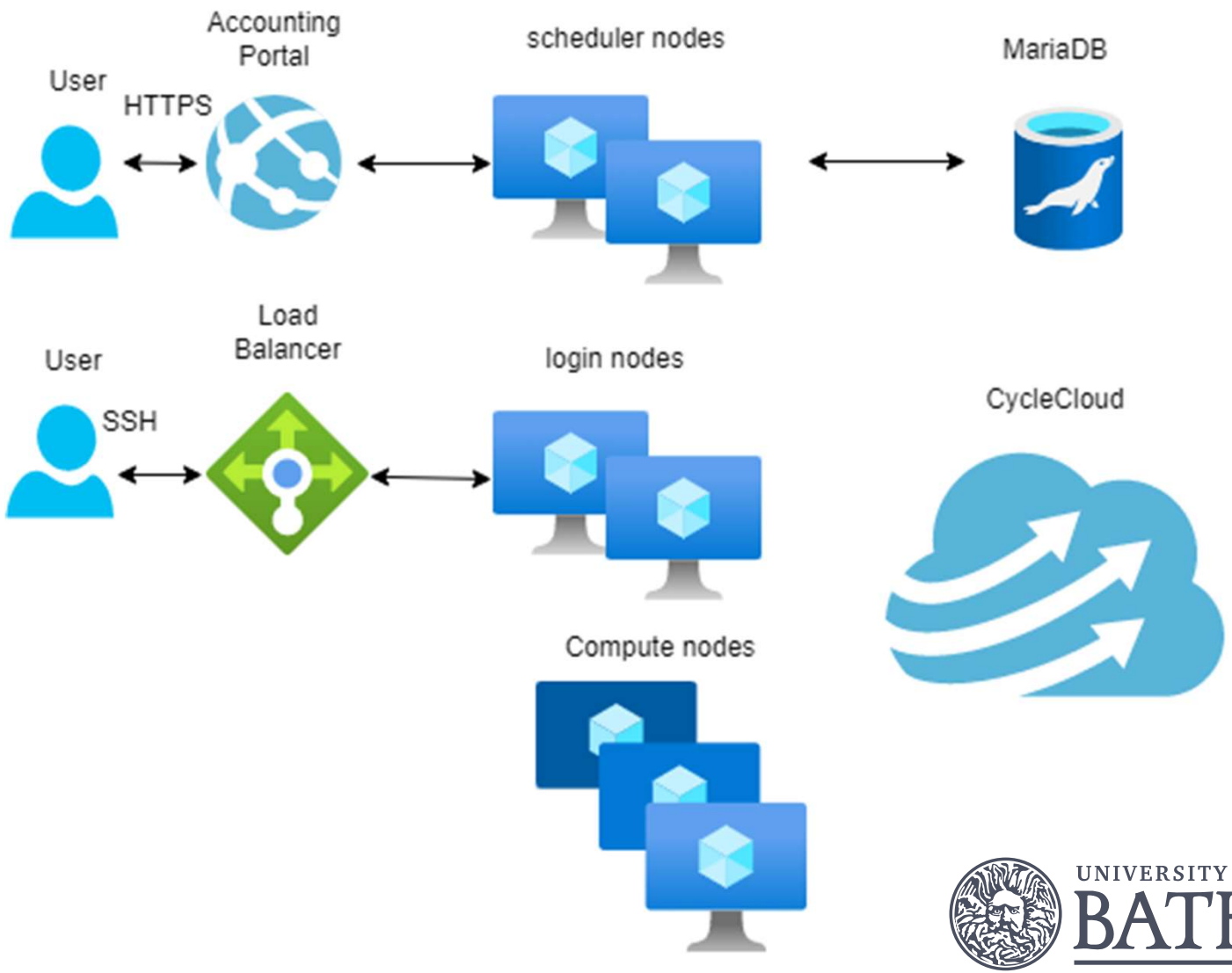


UNIVERSITY OF BATH

High Availability

Login and scheduling functionality are distinct and now there are several nodes of each kind (“high availability”) ☺

Users can still access the phase 2 system using SSH.



UNIVERSITY OF BATH

Monitoring

Grafana dashboards provide an insight into the status of the system.

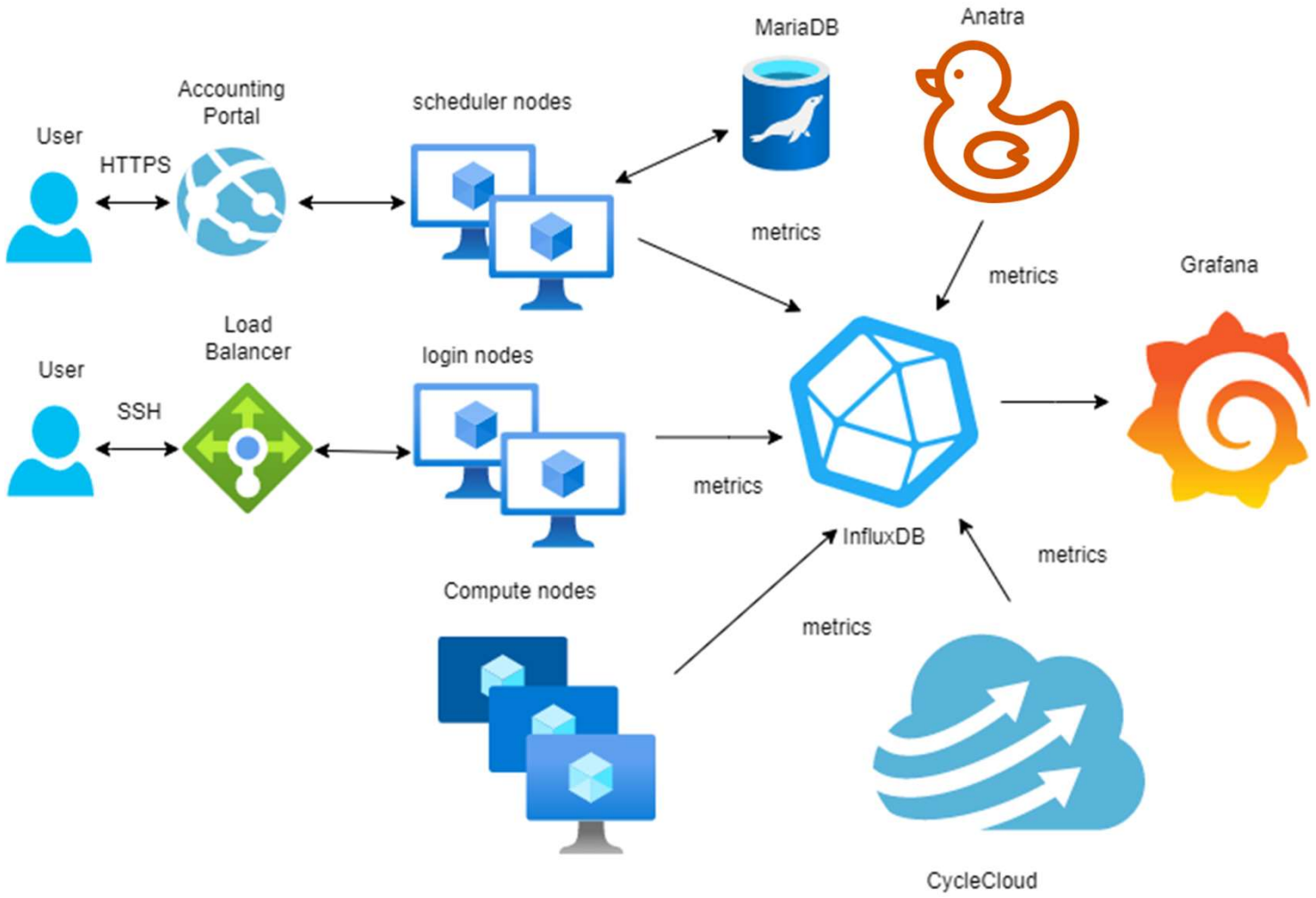
UNIVERSITY OF BATH

Metrics

Metrics are collected and displayed using the TIG stack (Telegraf, InfluxDB and Grafana) from all nodes in the system.

Initially only sysadmin access, but we might allow access to all users in future.

We can also collect metrics from our Antra system using this same approach ☺



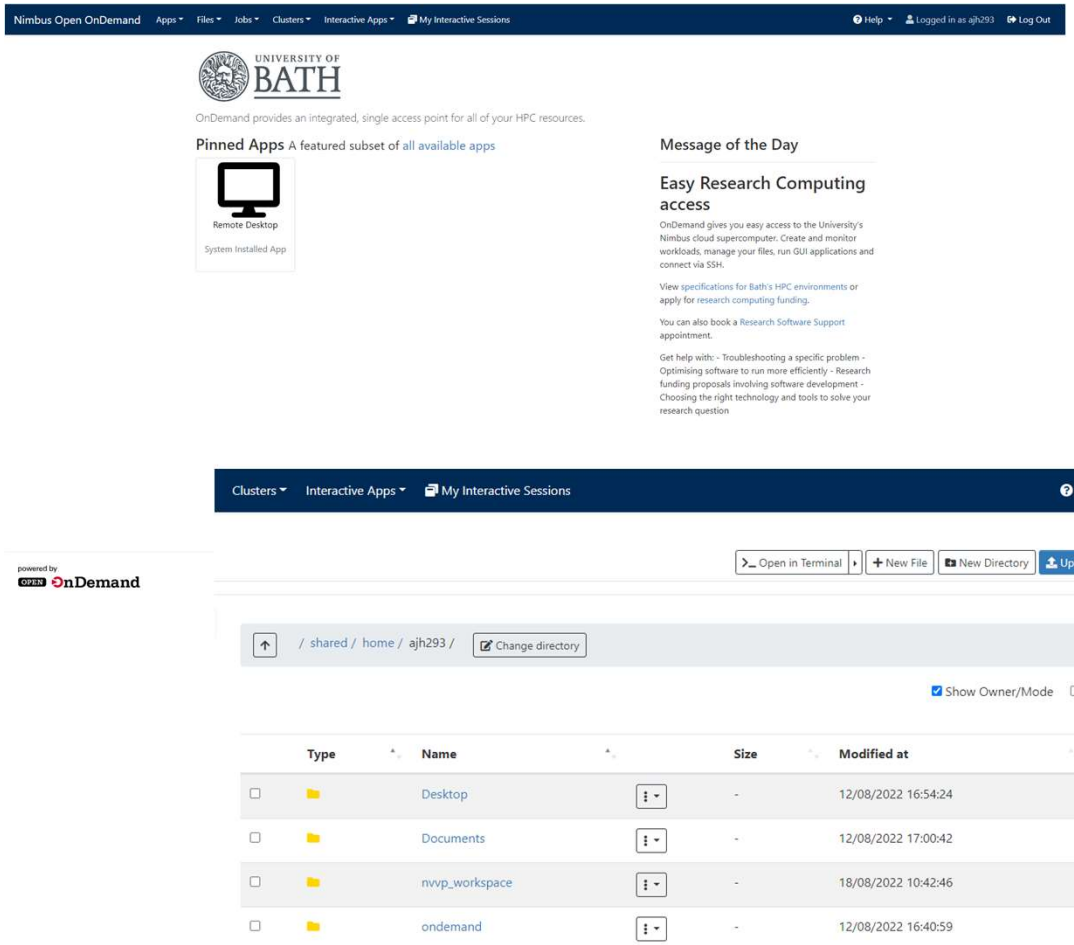
UNIVERSITY OF BATH

OpenOnDemand

Nimbus access from any device.

Create, save, submit, and monitor jobs.

Direct access to graphical applications on compute nodes



UNIVERSITY OF BATH

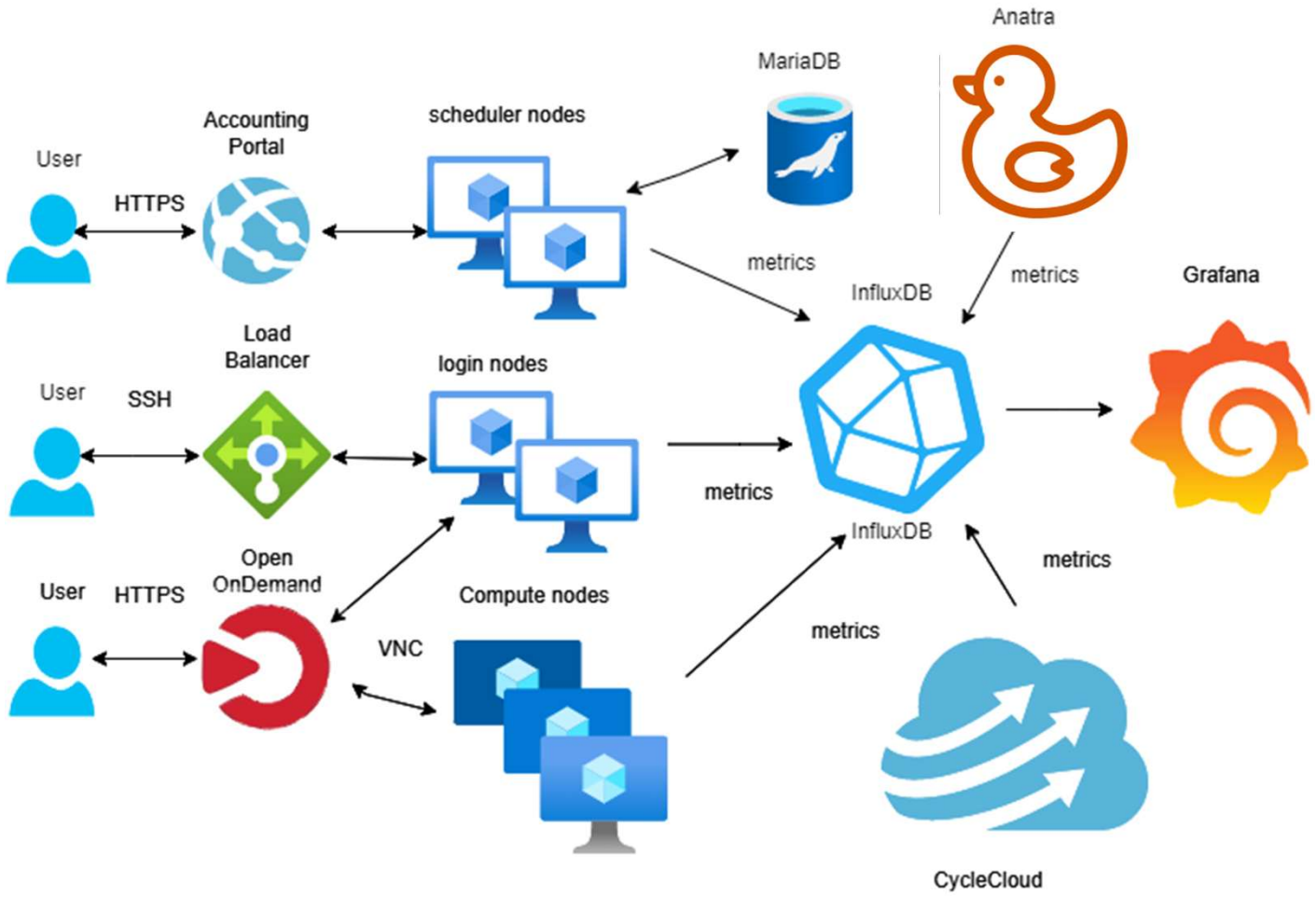
Open OnDemand

Open OnDemand (OOD) offers a unified access approach through a web browser (works on mobiles too ☺).

Can still access the cluster through SSH ☺

Can run remote visualization sessions though OOD.

Could also offer access to metrics through OOD.



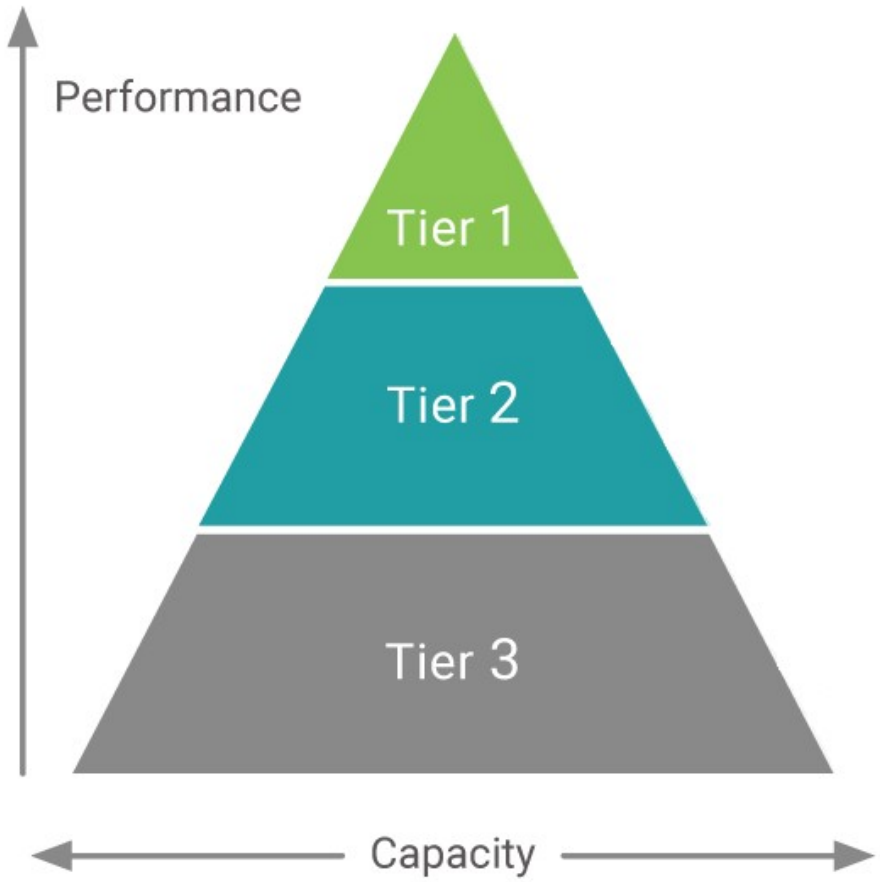
UNIVERSITY OF BATH

Storage Tiering

The price of storage in the cloud increases with capacity

To keep costs low, we must use only a smaller amount of high-performance storage and a larger amount of low-performance storage (“storage tiering”).

Make best use of storage moving data into tier 1 storage (and out after) only when high performance is required.



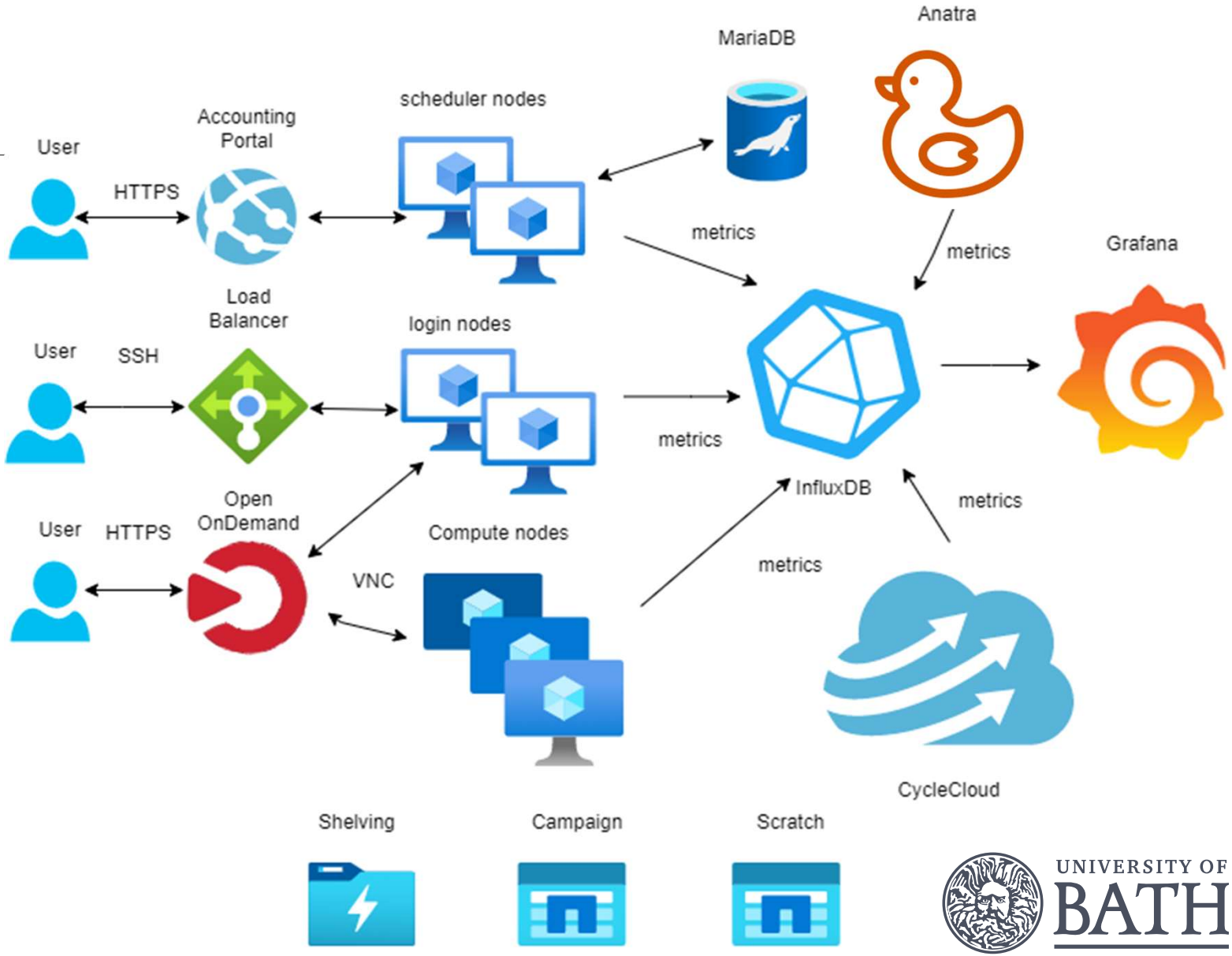
UNIVERSITY OF BATH

Storage

Tier 1 / scratch

Tier 2 / campaign

Tier 3 / shelving

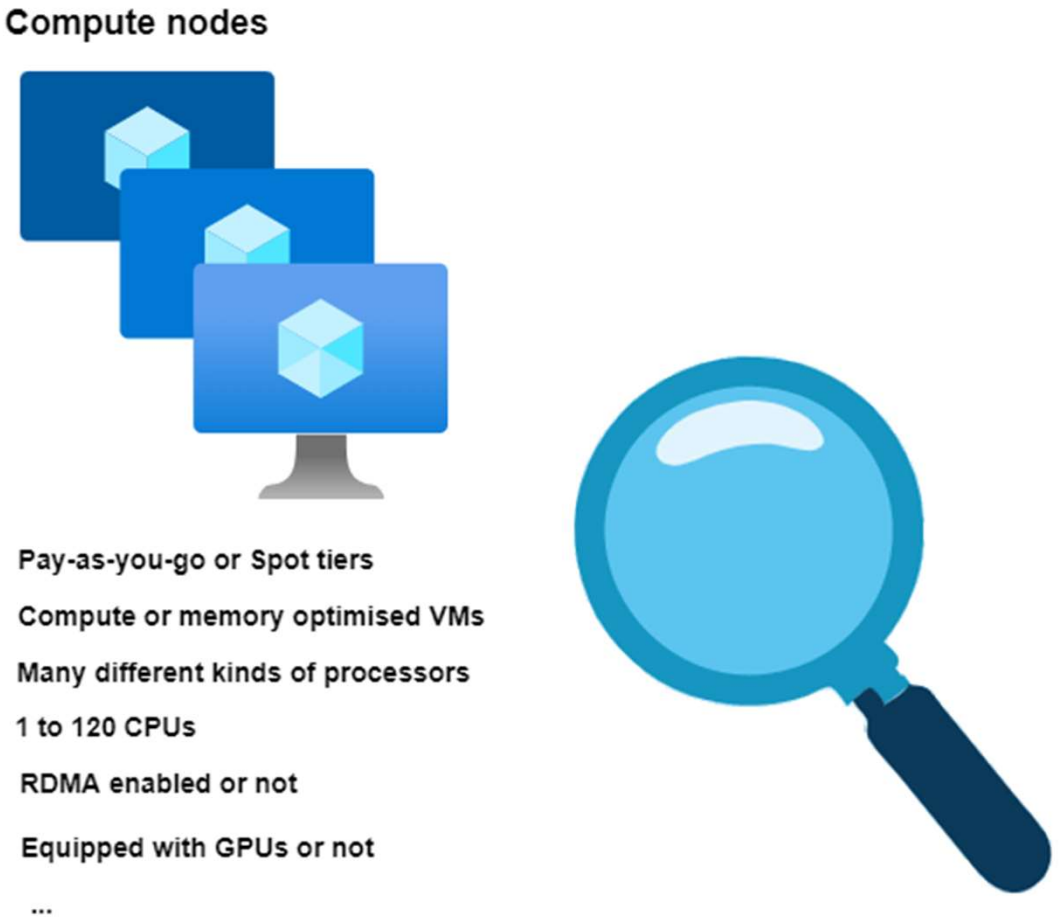


UNIVERSITY OF BATH

Compute

There are a lot of options for compute (many more than displayed here ☺)

“Spot” instances are spare compute capacity in the cloud available at highly discounted rates, so much cheaper than pay-as-you-go instances.

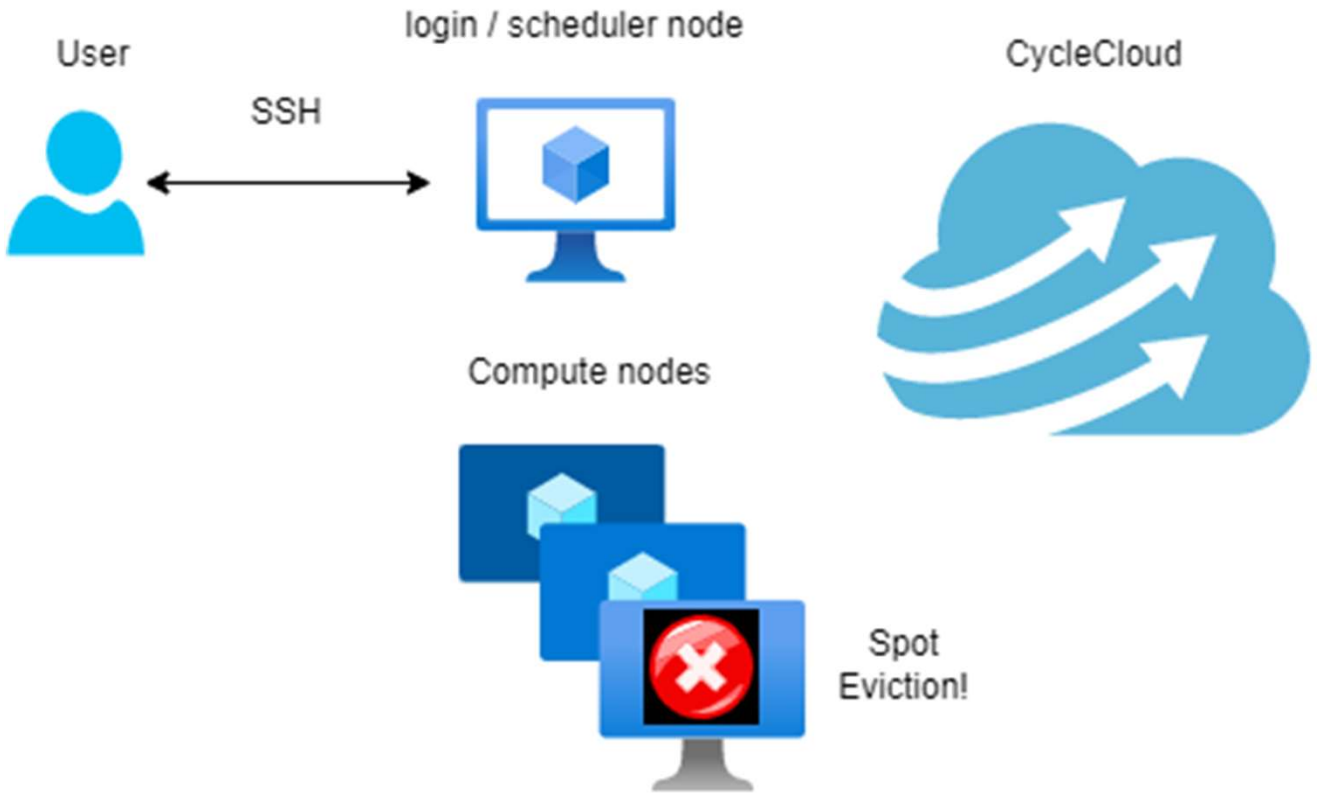


UNIVERSITY OF BATH

Spot evictions

We notify users when a job gets evicted, and track eviction rates.

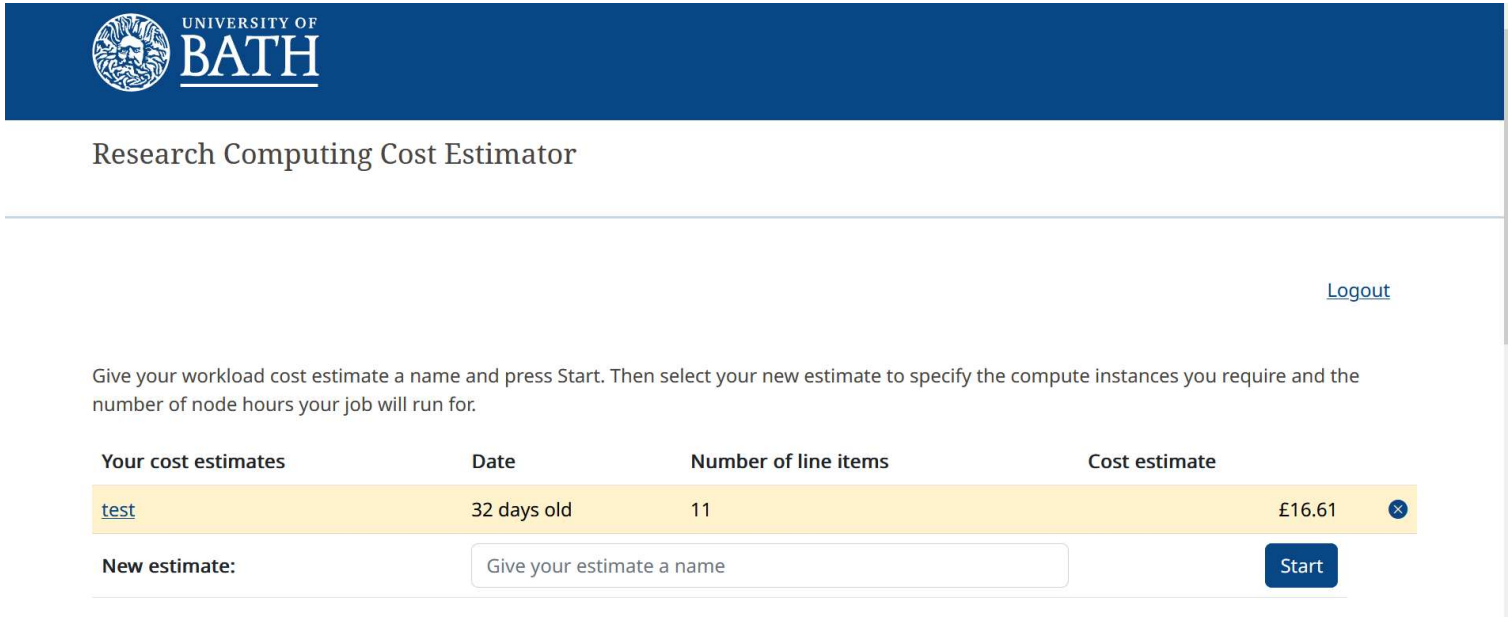
Ideally, for high eviction rates, you want to use spot instances only when your application performs checkpointing.



UNIVERSITY OF BATH

Cost Estimation Tool

Users estimate their compute costs by selecting the compute instances required and specifying how many node hours are needed.



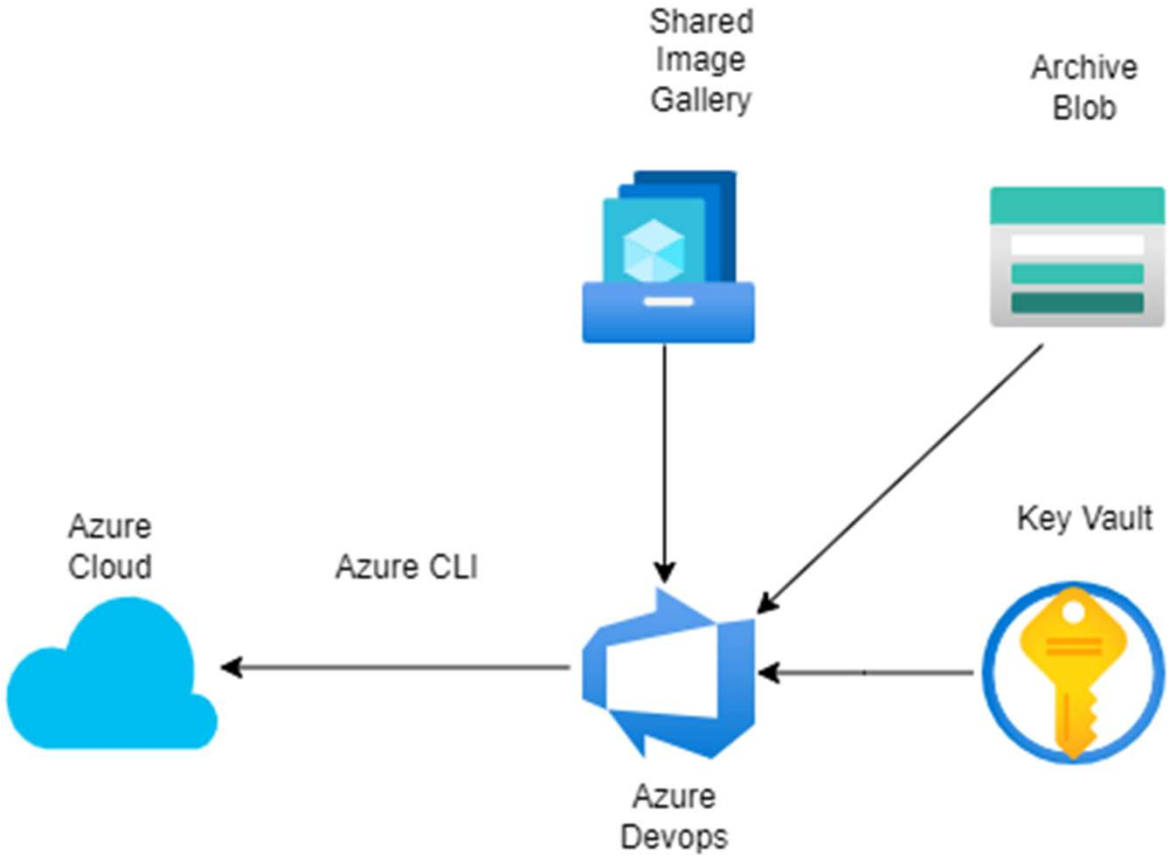
UNIVERSITY OF BATH

DevOps

The phase 2 system is provisioned all through software (Infrastructure as Code). Deployments are performed using Azure DevOps.

We are creating our own virtual machine images that we then load on to the nodes. Creating a new image can take up to 30 minutes.

Except images, we can create Janus or destroy it in less than one hour ☺



UNIVERSITY OF BATH

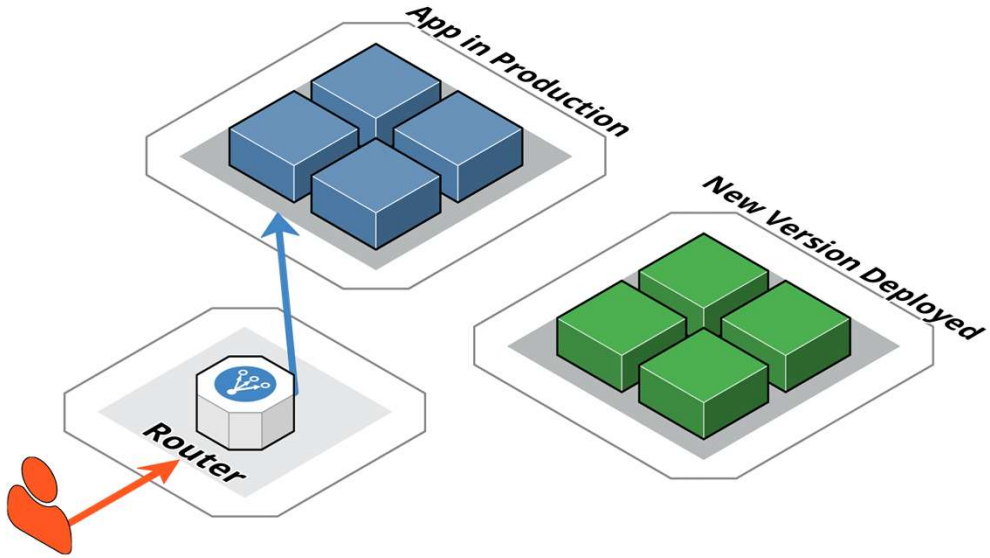
Environment lifecycle model

High velocity of change in cloud

Adopted a Blue/Green deployment model

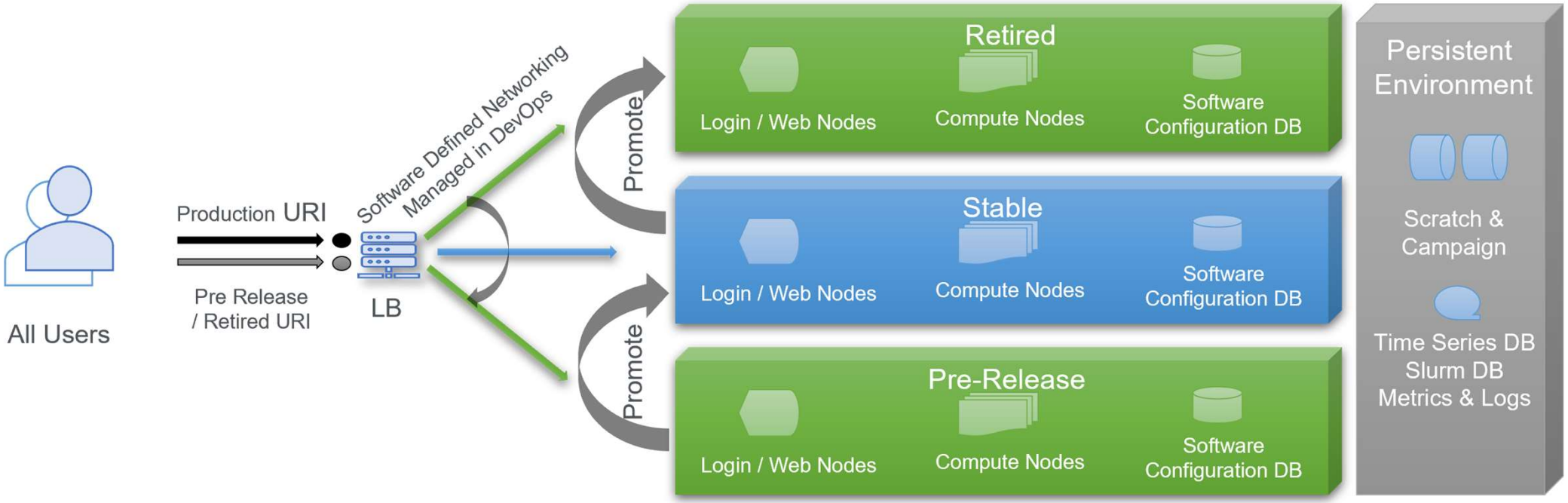
New features and patching can be implemented in code and deployed to keep environment secure

Minimum disruption to downtime and ability to roll-back



UNIVERSITY OF BATH

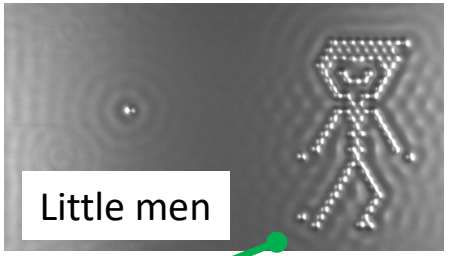
Environment lifecycle model



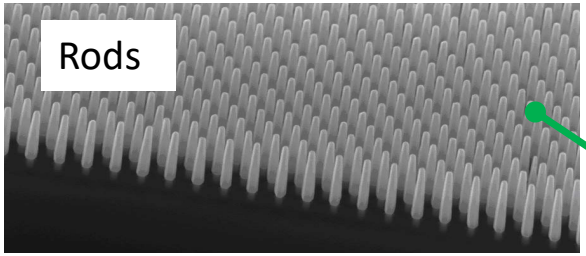
UNIVERSITY OF BATH

Nanoscience and 2D materials

A boy and his atom

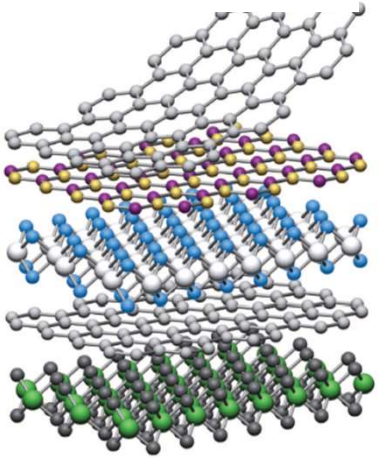


https://www.youtube.com/watch?v=oSCX78-8-q0

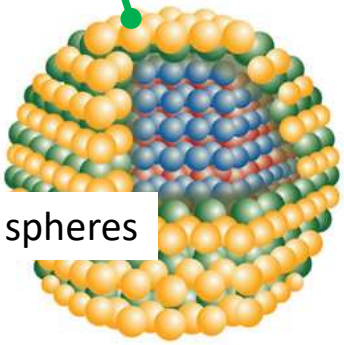


Example from Prof Daniel Wolverson, UoB, funded by OCRE

Quantum wells;
Layered 2D materials

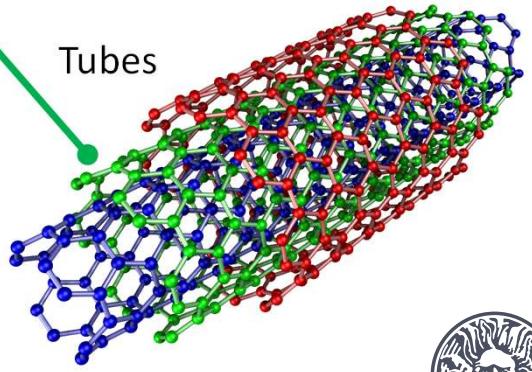


https://www.nature.com/articles/nature12385



https://www.photonics.com/Articles/Quantum_Dots_Small_Structures_Poised_to_Break/a22350

Tubes

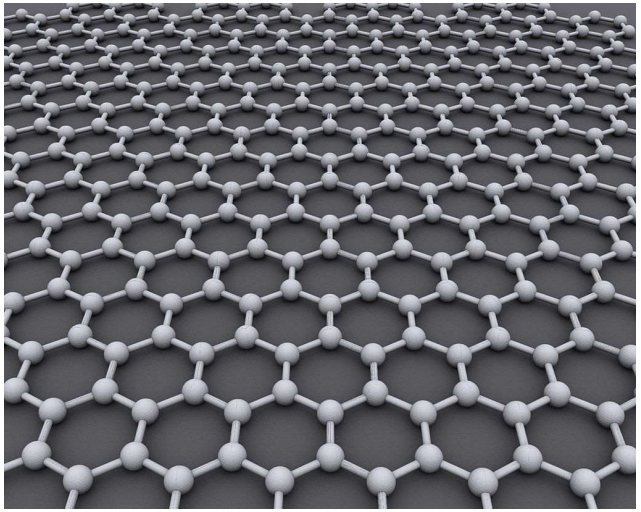


https://en.wikipedia.org/wiki/Carbon_nanotube#/media/File:Multi-walled_Carbon_Nanotube.png

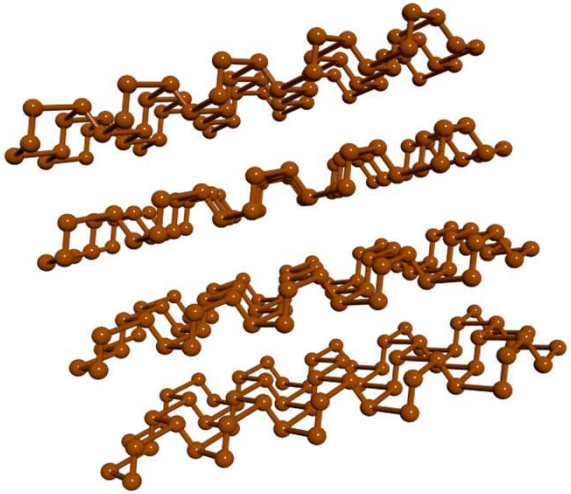
All nanoscience is contained in a bag of liquorice allsorts!

UNIVERSITY OF BATH

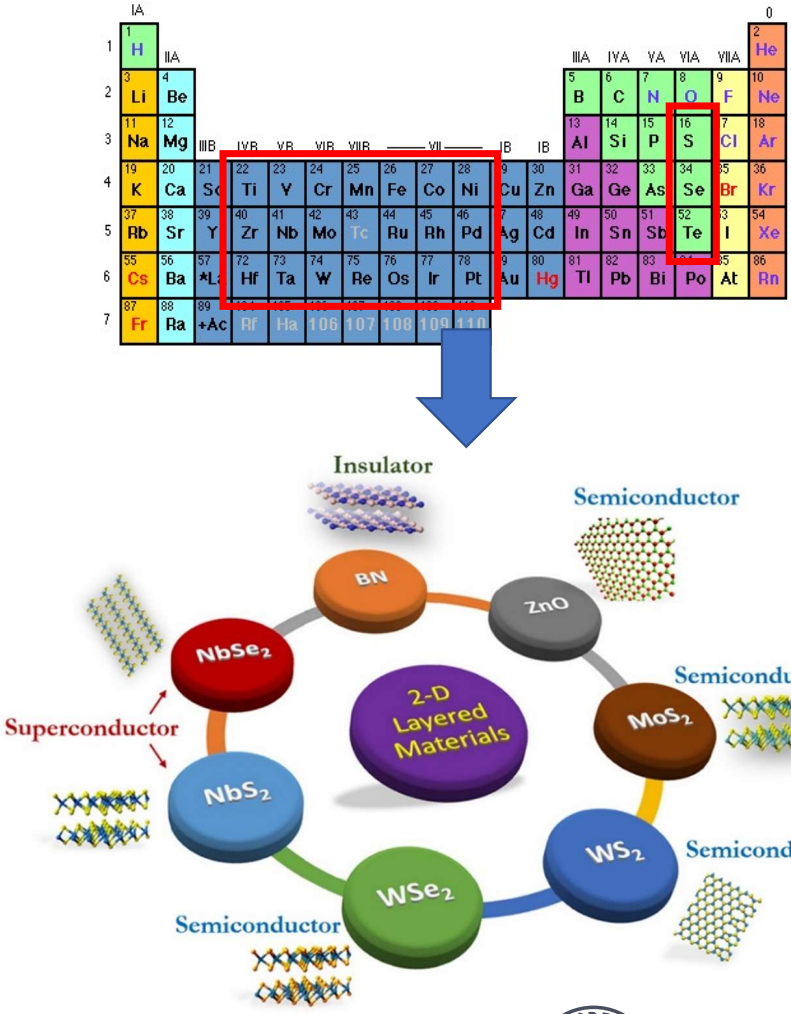
Isn't nanoscience all about graphene?



https://en.wikipedia.org/wiki/Graphene#/media/File:Graphen.jpg



https://en.wikipedia.org/wiki/Phosphorene#/media/File:Orthorhombic_bulk_black_phosphorus.png

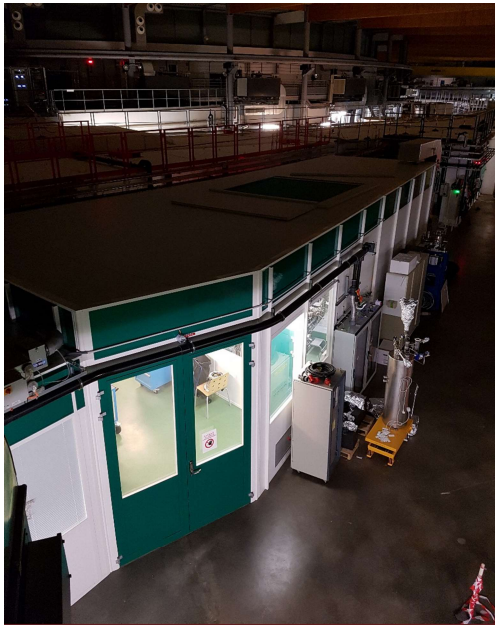


- Graphene is unique, but:
- 40 other natural layered MX2 compounds;
- Thousands more we could make if we wish.

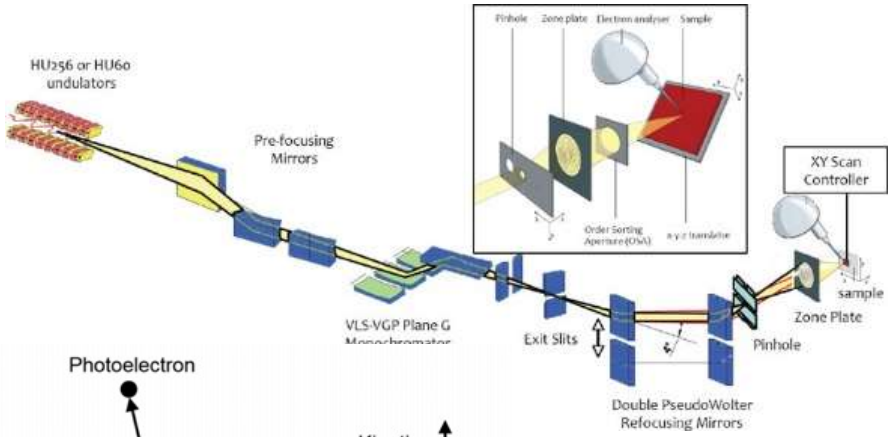
Ruitao Lv et al., Nano Today (2015) **10** 559-592

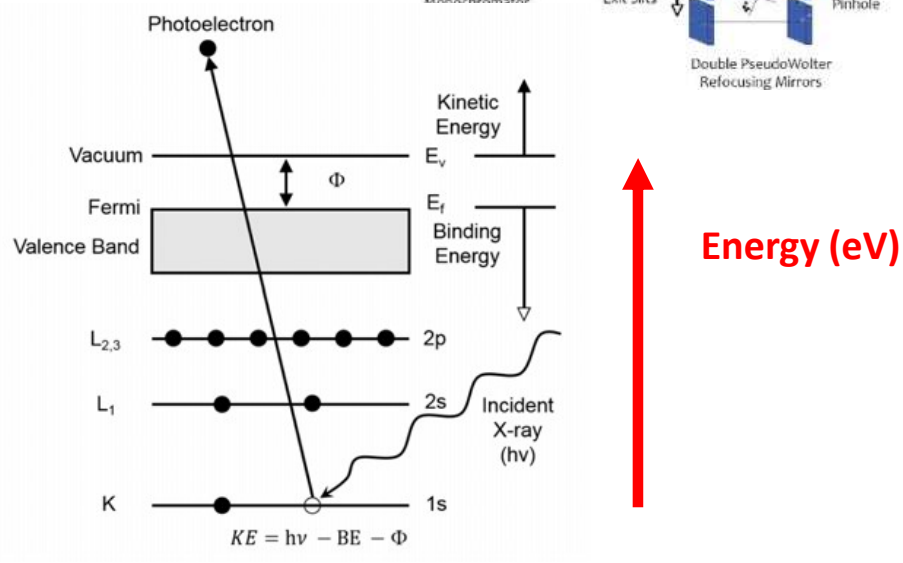
UNIVERSITY OF BATH

How are these materials studied?



X-ray photoemission beamline:





UNIVERSITY OF BATH

What tools do we need?

Density functional theory (DFT) codes (Walter Kohn, Nobel Prize in Chemistry 1998)

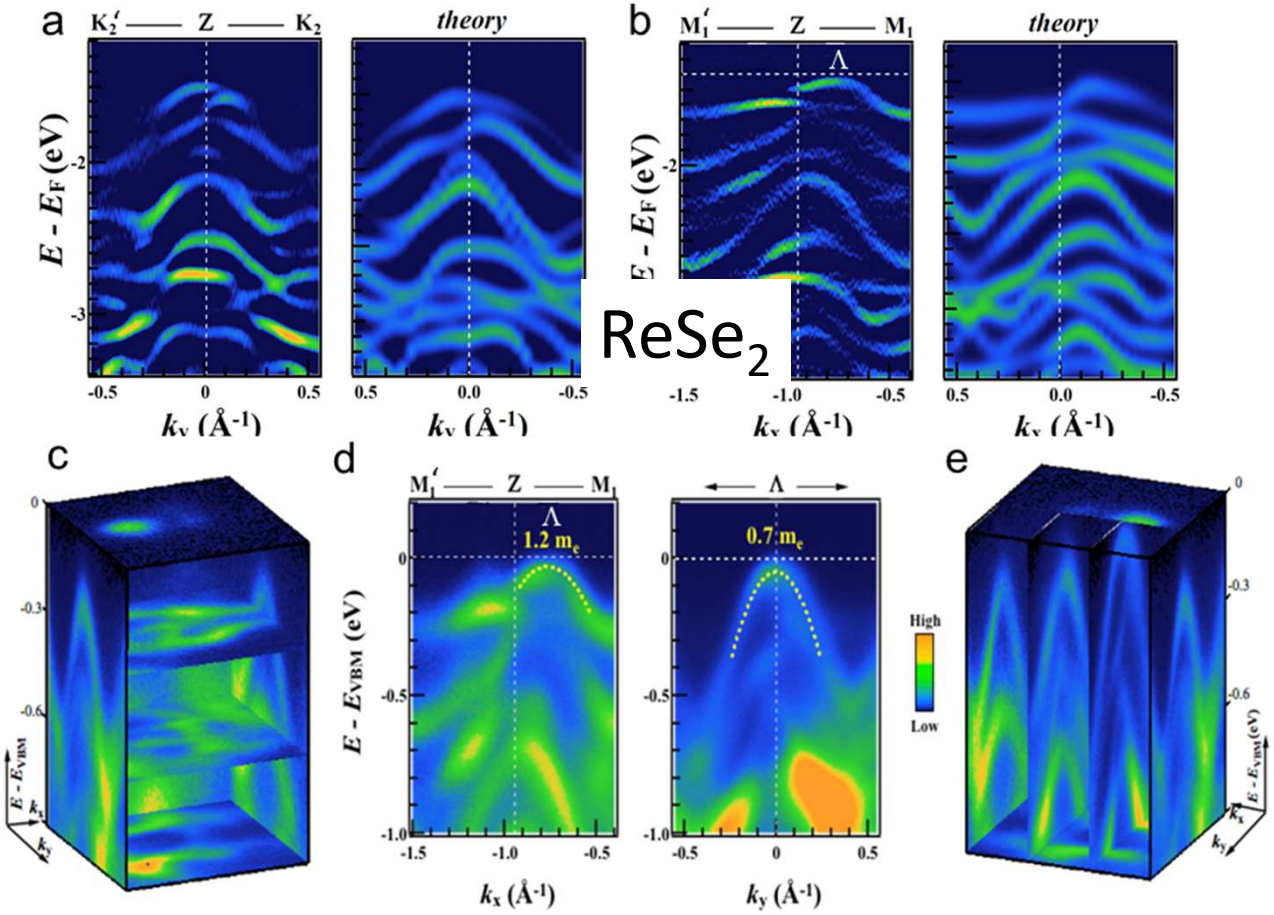
- **CASTEP, VASP, QUANTUM ESPRESSO** and many others.
- Plus post-processing, extensions and visualisation codes (Wannier90, yambo)
- We mainly use QE: GNU licence, open source, and we do modify source code.

- Libraries typically: BLAS, LAPACK, FFTW, OpenMPI, MPICH, HDF5.

- Qauntum Espresso code is very easy to install on desktops / small servers for training, so is well-suited for the early stages of undergraduate and Masters projects. Cloud resources allow one to escalate the calculations to a “production” level fast and flexibly.

UNIVERSITY OF BATH

What's the role of HPC computing?



We generate large and expensive datasets on highly topical materials.

These deserve modelling to the best of our ability to extract maximum information and understanding.

We need to work quickly.

It's appropriate to do some of this within an experimental group; postgrads can easily be involved in both theory and experiment.

We can even involve undergraduates at an appropriate level; HPC is not necessary for training them but is needed for real problems.

UNIVERSITY OF BATH

Positive user feedback

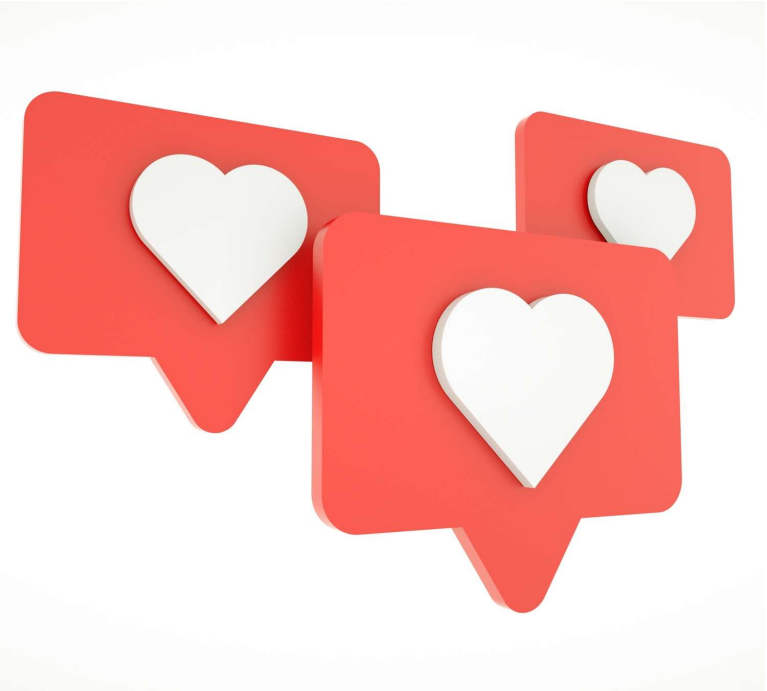
Larger core numbers delivering longer execution times, beyond default six hour limit on Balena

Big diversity in compute instances and storage options

Faster results and shorter queuing times

Choice of instance to better fit workload, exact core numbers matter a lot to Quantum Espresso

Great technical support from the Research Computing team – on what was my fault / the cluster's fault



UNIVERSITY OF BATH

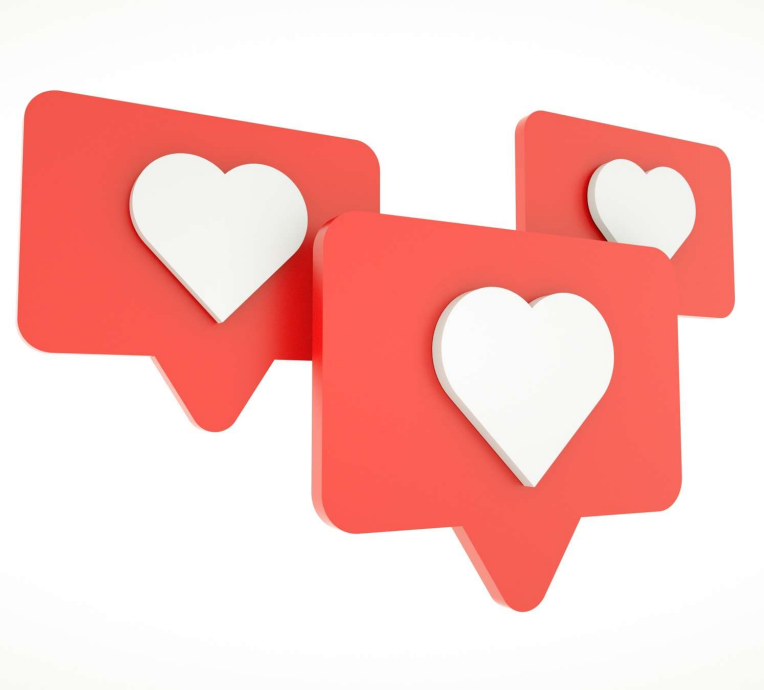
Positive user feedback

Seamless transfer from old on-premise environment to cloud (connection same, scripts virtually identical)

Cutting edge environments is leading to more ambitious computational plans in research proposals

Software running well and as expected

Monitoring jobs is easy and user-friendly

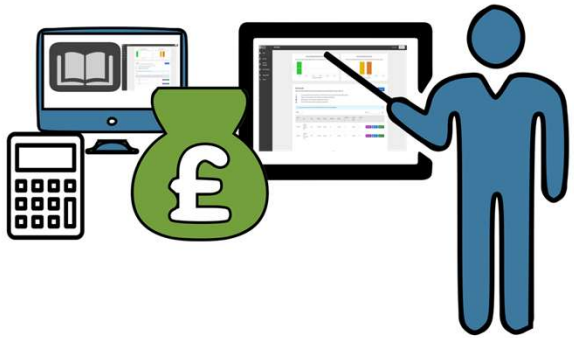


UNIVERSITY OF BATH

Lessons learnt

Advice

- Take your users on the journey with you
- Involve users in decision making and get them excited
- Keep monitoring your costs
- Allow for 24 hour to 5 day delays in Azure's cost reporting
- Directly map true costs to Slurm workloads



UNIVERSITY OF BATH

Lessons learnt

Advice

Monitor and review the impacts from evictions

Nothing is static in the cloud, things keep changing, some for the better

Get to know how your existing system works, storage performance needs, data churn

UNIVERSITY OF BATH

Challenges

Continue to provide internal funding!

Difficult to hire HPC engineers with Cloud experience

Users dealing with larger-scale parallelisation

Eviction events was confusing at first

High demand for the best partitions

Understanding the best compute instance

Cost-optimising

Keeping up with changes in Azure

UNIVERSITY OF BATH