

# The movement towards HPC inclusivity:

## Achieving on-demand public cloud accessibility of High Performance Computing (HPC) in ephemeral projects



Wil Mayers, Mark Titorenko, Steve Norledge and Cristin Merritt  
Alces Flight Limited, Bicester UK OX26 4PP



### Overview

The Alces team, through our work with the open-source Alces Gridware project, created Alces Flight Compute: a fully-featured, scalable HPC environment for research and scientific computing. A free version was created for cloud and hybrid use to understand how researchers would approach and consume on-demand HPC resource regardless of platform.

Over the past three years we have reached the following findings:

- Acquisition of public cloud over traditional HPC will work in epochs, with the potential for acquisition to slowly migrate towards a strong favouritism for working in public cloud.
- The strength of public cloud currently lies in auto-scaling HPC clusters hosting ephemeral (temporary) cloud workloads that are embarrassingly parallel.
- Clients focussed on hybrid solutions primarily utilise public cloud as a resource which might be cost prohibitive in on-premises design. This balance of utilising public cloud to research 'unknown outcomes' is allowing solutions to evolve more intelligently and without requiring an immediate increase of on-premises footprint.
- Creation of a consistent environment across platforms optimises the research outcome.

In coming to an understanding over the past three years that public cloud has a role to play, and that role that is not based strictly in on-demand consumption, we are transitioning this project to open-source (OpenFlightHPC) to address issues in procurement, persistent workloads, and partnerships with on-premises technology.

### Methodology Applied

In each project the following criteria were set:

- The Community Edition of Alces Flight Compute was used.
- In the case of cloud, only public cloud-compute time was charged.
- In the case of Alces Flight on hardware, costs were included as part of integration.
- Focus was on ephemeral, or temporary, projects.
- In most cases, anonymized data was set-up within a single user environment.
- Additional time measurements were made for application optimisation, scalability and performance.
- Public cloud testing used Amazon Web Services (AWS), focusing on the Ireland and UK regions.

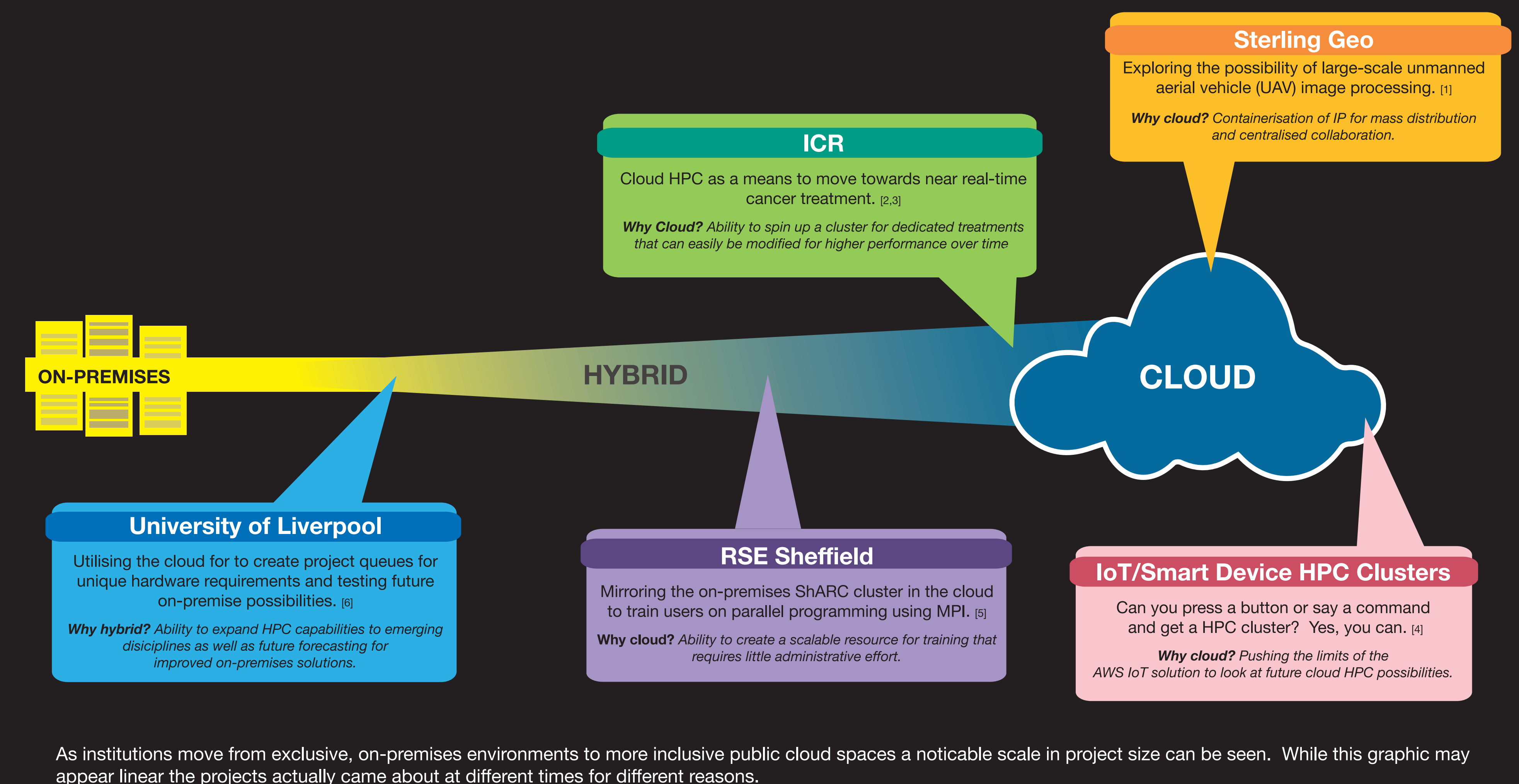
### Results and Roadmap

The following trends were observed in our projects on how researchers were consuming public cloud HPC resource:

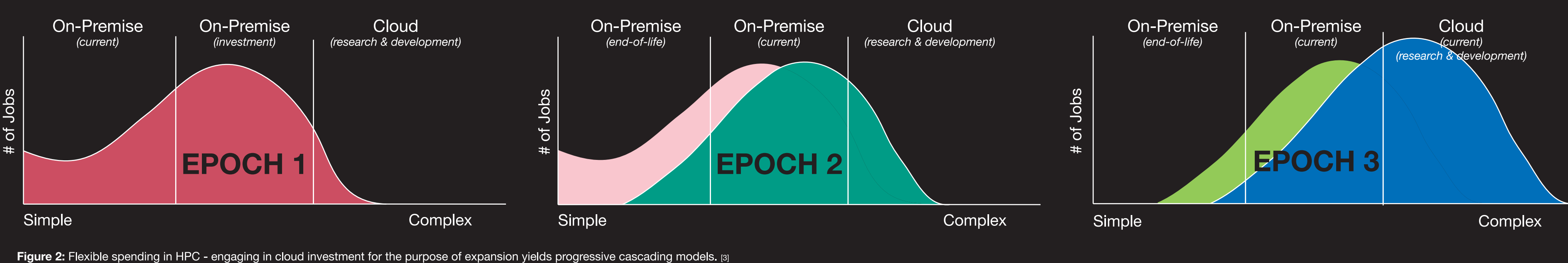
- Projects towards the left (on-premises) side tend towards future planning and mirroring of an on-premises design.
- Those within the hybrid spectrum play closely to the strengths of cloud.
- Embarrassingly parallel workloads utilising autoscaling were heavily favoured.
- On the right (cloud) side ease-of-use was placed as the primary design objective as the intent of these projects is to engage and enable as many users as possible.

These trends have moved our work from a focus on application-based, single-user, on-demand HPC clusters towards breaking our environment into open-source components under the name of OpenFlightHPC, launching November, 2019.

### Trending towards Inclusivity



### Trending towards more flexible spending models



Spending on public cloud projects denoted a changing spread in spending over a simple shift in spending. How each project chose to spend in public cloud was based on:

- How they traditionally procure HPC resource
- The skill sets on public cloud they had developed
- The level of 'newness' that their project/idea is in the field of HPC

### How Inclusive is your HPC solution?

**Inclusive:** adjective (/ɪnˈkluːsɪv/) - Including everything (and possibly everyone). Is your solution inclusive? Are you working to invite even more users and technologies into HPC? Here's four questions you need to answer to find out:

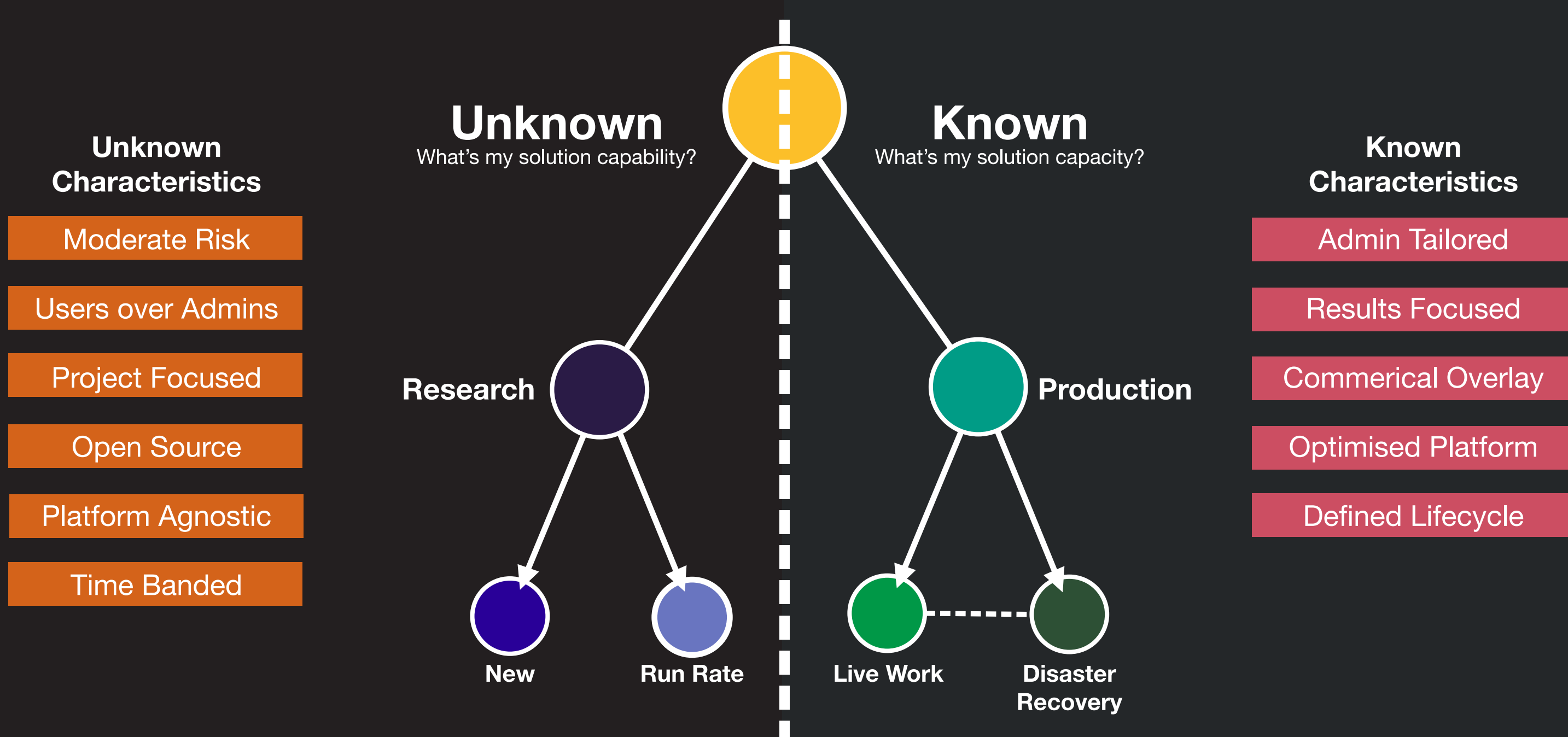
How easy is it to access your solution?

How long does it take for your users to learn and optimise?

How much experimentation is encouraged?

Are you able to be flexible with your resource provision?

### The Balance of Unknown and Known Elements in a HPC Solution (and Where Cloud Fits)



The introduction of public cloud into HPC has opened up the opportunity for discovery in areas previously blocked primarily by the ability to invest due to the high risk nature of the work being proposed. By allowing unknown elements of HPC to be explored at the ephemeral level users are empowered to gather the data necessary to feed either a continued use of public cloud or a transition to currently managed solutions.

### References

#### IMAGES

- [1] Mayers, Wil; Norledge, Steve; and Merritt, Cristin. *Unknown (Research) and Known (Production) Workloads in HPC and their general characteristics*. (2019) Image used with permission of Alces Flight R&D.
- [2] Mayers, Wil and Merritt, Cristin. *Epoch-based spending model in on-premises and cloud HPC*. (2018, updated 2019) Image used with permission of Alces Flight R&D.

#### PAPERS

- [1] **Breaking Boundaries, Saving Lives - Sterling Geo reimagines UAV imaging in public cloud.** Phil Cooper, Cristin Merritt, Wil Mayers. *Sterling Geo and Alces Flight Limited*
- [2] **Towards real-time Photon Monte Carlo Dose Calculation in the Cloud.** Peter Ziegenhein, Igor N. Kozin, Cornelis Ph. Kamerling and Uwe Oelfke *Joint Department of Physics at The Institute of Cancer Research and the Royal Marsden NHS Foundation Trust, London, UK SM2 5NG*
- [3] **Near Real-Time: How Cloud is Shaping Cancer Treatment**, Wil Mayers, *Alces Flight Limited*
- [4] **Real-World High Performance and High Throughput Computing on AWS**, Matthew Berryman and Adrian White *University of Wollongong and Amazon Web Services*
- [5] **Bespoke High Performance Computing Clusters in the Cloud with Alces Flight**, Mike Croucher, *University of Sheffield*
- [6] **The Liverpool Barkla Cluster: Exploring HPC Hybridisation**, Cliff Addison, *University of Liverpool Advanced Compute Facilities*