Balancing Research and Production:
Alces Flight's take on building up commercial compute

Cristin Merritt[1] and Wil Mayers[2]

[1]Alces Flight Limited

[2]Alces Flight Limited

1. Unknown/known, emerging/immediate, research/production

There are typically two complementary types of compute workload taking place in academic groups looking to spin-off into commercial entities:

1. Your known, production workflows which are showing promise for building up a company that can pioneer the way forward in science.
2. Your unknown, research concepts, ideas, and emerging workflows that can feed the cycle of improvement.

So how do you build up commercial compute without expending too much time and killing your budget in betting on the wrong resources?

For the past three years Alces Flight has been studying how public cloud is impacting High Performance Computing (HPC) as an emerging platform. Our exploration moved us to initially create a tool designed to launch an HPC environment straight in the cloud – but what we learned along the way has made us re-evaluate how emerging business can harness different resources in order to build a smart cycle of compute that can feed academics transitioning into company life. At the September, 2019 CompBioMed Conference we wish to share the insights made in:

- Honing characteristics of your workload to understand what constitutes production (paid commercial) work and what constitutes the research to feed improvements.
- Understanding how spreading your compute budget can help you not only survive getting started, but make strong future investments.
- Balancing both free and paid technologies and services to avoid the trap of believing you can do everything in-house.

2. Know thy commercial compute aims

Creating a commercially viable workflow or application is only part of the story of how a company comes into being – you also have to maintain and grow (or even replace) that which got you off the ground. So how do you characterise what is commercial and what is research feeding the future? Our experience has helped us set out some general guidelines of what to look for when reviewing your projects (Figure 1):
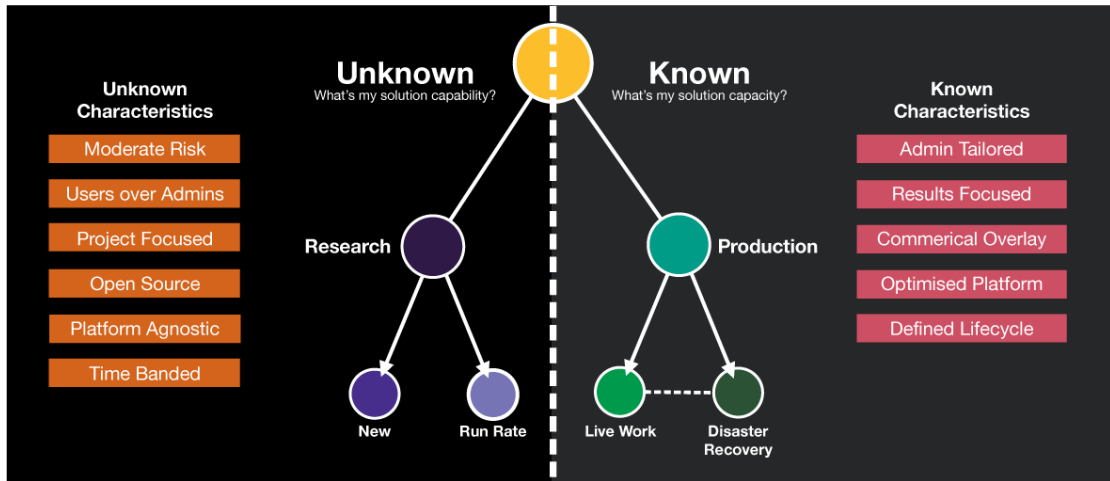
Figure 1 *Division of research and commercial compute with characteristics most common in each workflow.*

Ask yourself these questions to determine where your work is concentrated:

- How much risk would you assign to what you are working on?
- Are you the only user or are you involving a team to ensure your workflow is running smoothly?
- Is your work a project that has no financial outcome attached to it? Is it feeding a larger aim, or is the work itself the aim?
- Are you pursuing partners or other commercial groups to help spread the burden of maintaining and growing this work?
- Does this work have a defined end goal? Does it have a defined shelf life after which its relevance is reduced?

Once you see where your balance is you can then start looking at the platforms you have now and consider the future spend.

3. Spread your spend risk

It isn't uncommon in large corporate companies for there to be a spread- spend on the available resources in the market. Until recently, small companies rarely had the luxury of a large capital HPC spend to purchase an HPC cluster from vendors like HPE, Lenovo, DellEMC, Cray and Huawei. Public cloud has suddenly brought this concept to the small company level – but where do you start?

First, look at where the balance of your compute sits. If you are focused in short, well-defined projects then cloud is currently a stronger contender; but if your compute work is persistent and long-term then you will need to consider the financial impact of potentially owning or sharing in physical resource. Most mature companies are spread across all platforms and currently sit in spending Epoch 1 or 2, as illustrated in Figure 2. Those companies just starting out, however, might fall more completely into a stronger cloud category (Epoch 3) for research but share or own physical kit, if only at the workstation levels.
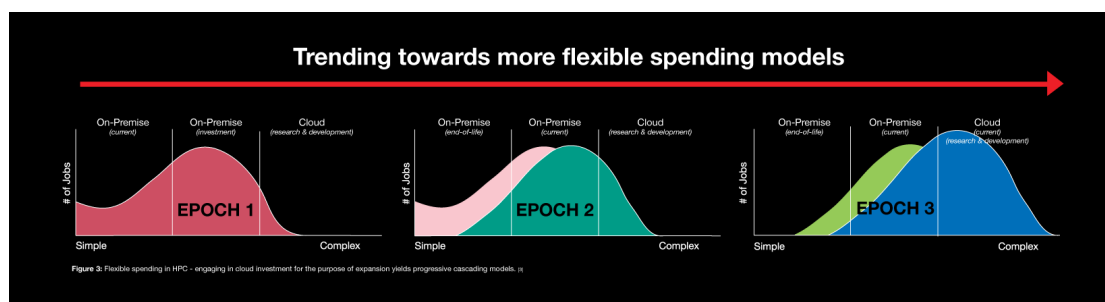
Figure 2 *Epoch-based spending model for compute resources – on-premises, progression (resource at end-of-life and current resource), and public cloud*

Secondly, once you have established where the compute you are looking to do best sits, you need to consider the "free" versus "paid" options available in getting your company into a cycle of consistent improvement.

4. Free isn't free.

It's quite easy to become mired in the detail of your work – but spending time is just as expensive, if not more, than spending money itself. Go back to your work balance sheet – if your research method is the focus of your work then carefully curating open source tools and resources are a good start to achieve this goal. Why? Because research isn't as mature or fixed (sometimes called 'hardened') as commercial workflows and applications – and far more open to change.

Once research becomes paid work, however, you should assess if it should be something you are managing or if partnering or farming out the management is a more sustainable solution. This is one of the key things we learned when studying cloud adoption. Since so much of the desired outcomes in cloud were research-based, our initial tool was very rarely put into Production. Why? Because the teams working on these projects wanted to be able to adjust and change things until they were ready. They were far more willing to engage and become more 'hands-off' as their workflow matured. This led the team at Alces Flight to split off our research into an open source project called *OpenFlightHPC* and develop management tools around mature workloads as part of our commercial endeavour.

5. Conclusion

The team at Alces Flight, through our project work in cloud HPC adoption, noted that there is a marked difference in workflows that are in commercial production and those that are research. Those who had a strong understanding of their aims could place their workflows where they operated most effectively in time/cost, and could make stronger decisions on where to invest in outside resource. Our work with cloud tools has evolved during our research, moving in two distinct directions – that of the experimental, open source model to bolster research within the HPC community and that of the commercial, managed tool-set focused on enabling proactive HPC for continued improvement. We are happy to share our insight into building commercial compute with the CompBioMed community and hope this short paper assists people on this journey.