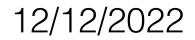
Cloud Computing and Containers



"The Angry Penguin", used under creative commons licence from Swantje Hess and Jannis Pohlmann.



Warwick RSE



Cloud Computing

- Cloud computing is a distributed computing paradigm based on you requesting resources that are provided "somehow" from a large pool of resources that the provider has available
- There are many could providers but the ones that you are most likely to encounter in academia are
 - Microsoft Azure
 - Amazon Various products, mostly EC2
 - Google Just called "Cloud"



Cloud Computing

- Can get most modern (up to cutting edge) hardware from one or other cloud provider
 - Some hardware that is unique to each provider e.g. Amazon Graviton2 CPUs
- "baseline" hardware is often comparable to HPC cluster systems
- Cloud providers often have high levels of segmentation
 - You have to chose the right "part" of their hardware offerings
 - Usually provide help e.g. https://azure.microsoft.com/en-us/ pricing/vm-selector/



Is the cloud right for me?

- Good starting point is to mentally replace the phrase "in the cloud" with "by renting a computer" and see if it still sounds like a good idea
- "I can't be bothered to sign up for the HPC system,
 I'll rent a computer" generally not a good idea
- "I want short term access to some unusual hardware, I'll rent it" not a bad idea
- Cloud providers often offer time to researchers as part of grants if you have it, use it



Types of Cloud

- Cloud systems offer a very wide range of services
- Bare metal Literally renting computers (more or less). You install everything remotely and run on it. Not recommended
- Blueprint/Image You get a system that starts up with a running operating system on it. Quite often technically a virtual machine
- Container You provide a container that is set up to run your problem and it just goes onto a cloud instance and runs
- Software as a Service (SaaS) A requested piece of software is literally run for you on the level of hardware that you request. Only available for specific software and not often for academic software



Comparison to Clusters

- Mostly physical clusters effectively work as a mixture of blueprint and SaaS
 - The operating system is there and working and usually you have a system to let you switch in and out software packages
- You can also run containers on clusters
 - Not the more popular Docker containers (directly)
 - Apptainer (formerly called Singularity)
 - You can convert docker images to Apptainer images



Containers

- If there is already a container that you have prepared or has been prepared for your program you may as well use it
- Making your own container is not really hard but generally isn't worth the effort unless you need very specific libraries that will not be on a normal cluster
- They do make it easier to move your work between clusters and cloud systems
- If you are already using containers on cloud you should be able to move easily to our clusters





- There are some costs to using containers
 - Effort of creating and maintaining them
 - 1-2% intrinsic speed penalty
 - Can substantial performance overheads from not being optimised for the machine you are running on
 - Might be some limits on containers on some clusters



- Pretty much everything about cloud computing has a cost associated with it
 - Uploading data (sometimes at least often this is free)
 - Running calculations
 - Storing data
 - Downloading data
- Be sure that you have costed all of the parts of it before using paid cloud provision



- Cloud compute pricing can be anywhere from very cheap to very expensive
- At the time of writing the median job on Sulis would cost about \$500 dollars to run on Amazon EC2
 - Price can be lower if you are willing to wait for prices to fall
 - Can be higher if you need the answer now and get unlucky



- Storage costs can be non trivial (e.g. from Amazon S3)
 - \$0.03 per GB per month to store data
 - \$0.09 per GB per download for downloading data
 - Additional costs for querying files in cloud storage
- If you filled your Sulis quota cost about \$60 per month to store data
- If you downloaded that data would cost you about \$180
- Make sure that you have enough budget to download data when you have generated it!



- Can all be cheaper if you are willing to play around with it
 - Cheaper compute if you wait for price fluctuations
 - Cheaper storage if you are careful enough about how often you need to access it and how quickly you need to access it when you need it
- A lot of the convenience of cloud systems goes away when you have to think like this



- Mostly academic cloud users are provided with cloud resources for free by a cloud provider as part of a research project
- Useful resource, well worth using
- Be careful about becoming reliant on a cloud workflow
- At the moment cloud providers are more expensive than HPC/HTC kit run by research institutions
 - Our personal opinion it always will be



On premises clouds

- Some universities (not Warwick, yet) offer cloud like "on demand" provision of computing resources
- Most of the problems associated with clouds go away at this point
 - If you have access to a resource like this and cloud type on demand access is what you want definitely use them before commercial providers

