The Faculty of Science Research Cloud Procedures and Pricing
May, 2018

This document outlines the pricing structure for cloud resources as well as procedures related to purchasing and maintaining them.

The research cloud is intended to replace the need for individual server machines or clusters used by researchers. The virtualization of these physical machines offers several advantages over physical, researcher owned machines which add up to give significant savings in time and a large improvement in convenience, specifically:

	Stand-alone Machine	**Research Cloud**
Purchase of hardware	Researcher must come up with specifications, request quotes, select and pay the winning bid, wait for delivery and finally arrange for the machine to be installed and connected to power and network.	Researcher selects resources from the list, pays and, depending on size of the order and availability in the cloud, some or all these resources will be immediately available.
Configuration of hardware	Changing a machine's hardware configuration requires physically removing memory from one machine and installing it in another. It is not possible to split each physical core into a separate machine.	A machine's hardware can be reconfigured through a web page. Memory can be added or removed and cores can be split into individual machines if desired.
Software Installation	The researcher has to install an operating system from scratch on every physical machine followed by installation of any software applications. Finally both the operating system and any additional software must be configured e.g. adding user accounts, setting up disk exports, configuring firewalls etc.	Several basic images with pre-installed operating systems will be available. The researcher installs software and configures the machine once and can then clone the image as many times as needed.
Resource Sharing	When not in use by the researcher the hardware lies idle. Network connections are typically limited to 1Gb speeds.	Idle resources can run a different researcher's virtual machine and their idle resources can run yours. Links are 10Gb which works out as ~0.5-1Gb per virtual machine if all are at maximum usage but this is

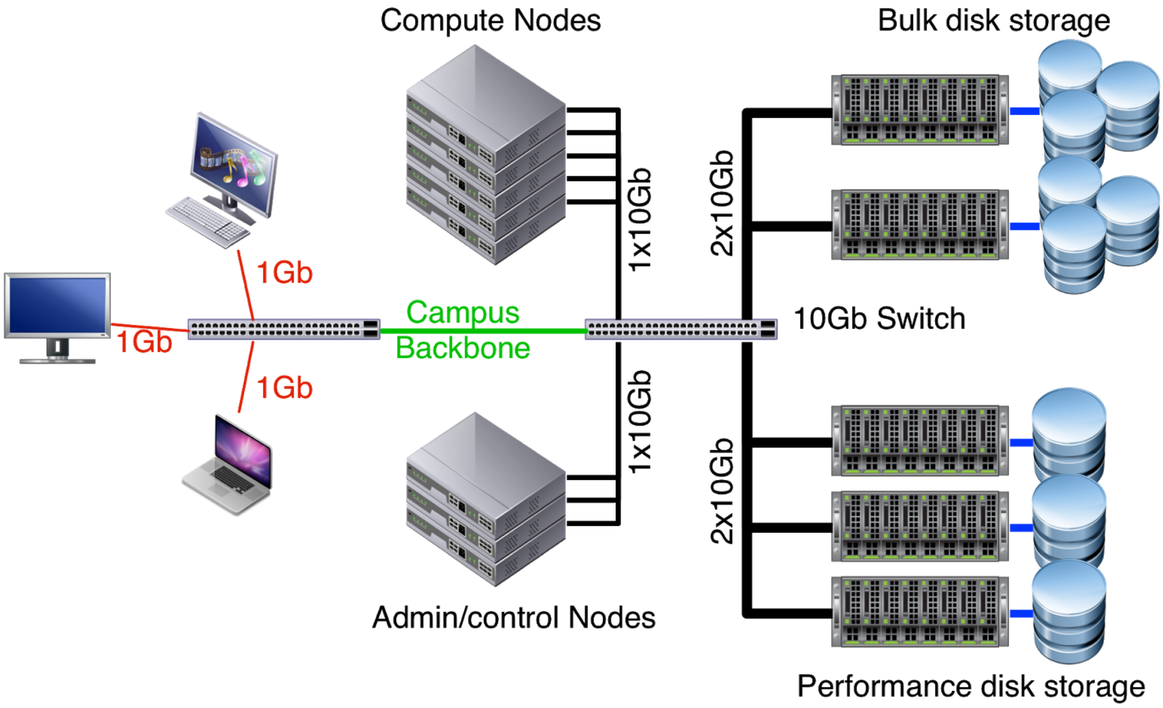
		unlikely and so you get a lot more.

Cloud Resources and Prices

The research cloud contains four types of resource which are available for purchase:

- CPU units
- Performance disk storage units
- Bulk disk storage units
- GPU slots

These are each described in the following sections and their connections are shown in the diagram below.



The minimum resources to create a functioning machine in the cloud is one CPU unit and performance disk storage unit (100GB) where the machine image is stored.

Please note that all pricing below use actual GB/TB units and will be the amount of usable storage the researcher receives. For comparison a ‘6TB’ hard drive only provides 5.45TB of actual storage and this is further reduced by another 10-20% by the redundant nature of a RAID array which provides insurance against 1-2 drive failures.

CPU Units (2 virtual cores+15GB memory for $550) [Intel E5-2650v3, DDR4/2133MHz]

A CPU unit consists of one physical CPU core and 15GB of memory. This will give a quota of two virtual cores and 15GB of memory. Each virtual machine must have at least one virtual core so a single unit can be used to create two virtual machines each with a single core and 7.5GB memory.

All CPU units go into a single quota pool so that multiple CPU units can be combined to create virtual machines with larger numbers of cores and greater memory. The only limit to this is the physical hardware which has 24 physical cores and 360GB of memory. However allocating a virtual machine close to this limit would be very challenging since it would require an almost empty physical compute node. For this reason we currently have a maximum size of 10 virtual cores and 128GB memory. If this is insufficient for your needs contact your departmental ITOC representative and we can see what we can do.

GPU Slots (1 free with each 6 CPU units or 1 Slot for $450 with purchase of 2 CPU units)

Every 6th CPU unit you purchase from now on comes with a FREE GPU slot which will allow you to install a GPU in Cirrus for use in a VM. For the moment you will need to purchase the GPU card yourself and currently we support NVidia retail cards (1080Ti, TITAN etc.) as well as the NVidia Tesla cards. SRIT is also testing an AMD GPU so these may soon be supported too.

If you need more that one GPU slot per 6 CPU units then you can purchase additional slots at $450/slot up to a maximum of one GPU slot per 2 CPU units which are purchased at the same time. This additional cost is because, to provide these extra slots, we have to purchase machines with fewer CPU cores which have a higher overall cost per core. This also means that CPU units can only count towards either paid OR free slots, not both i.e. if you purchase 6 CPU units you can either have one free GPU slot or pay for 3 GPU slots, you cannot pay for 2 GPUs and get the third free (sorry!).

Lastly SRIT are developing a test-bed facility on Cirrus where users can try out GPU cards for free for a limited time so that you can gauge their suitability. If you are interested in this then please contact SRIT.

Performance Disk Storage Units (100GB for $80)

Performance disk storage consists of high speed (7.2k RPM) disk drives which are tiered using SSDs so that recently accessed data can be accessed again at far greater speed. The data is also replicated between multiple servers so that network bandwidth is spread over multiple machines to reduce the chance of a bottleneck.

All virtual machine images, i.e. the root filesystem and boot disk, live in this storage to ensure a rapid boot up of new virtual machines. This means that a minimum of one unit, which is 100GB, is required to run any virtual machines on the cluster.

The other use of performance disk storage is for data which is either frequently accessed or will be accessed simultaneously by many different machines in the cluster. Examples of this include a dataset which will be read by many different virtual machines at once, a block of configuration data which may need to be copied to many virtual machines when they run a job, database files which may either be copied locally or frequently accessed by a server machine etc.

Performance disk storage can also be exported to machines outside the cluster. This will require a virtual machine in the cluster to mount the storage and then re-export it using a standard network disk protocol. However the bandwidth limitation of the Campus backbone is far below what performance disk storage can provide so you may get exactly the same performance to your external machines using cheaper, bulk disk storage.

[NOTE: This pricing assumes a flat 10x5.45TB per server and triplicate storage]

Bulk Disk Storage Units (1TB for $125)

Bulk disk storage sits on servers which contain large numbers of enterprise-class hard disks arranged in RAID arrays to ensure reliability. This provides very cheap bulk disk storage but with a lower bandwidth and higher latency than the performance disk storage.

This storage is intended to be used for keeping large datasets which will not require high bandwidth access. It can also be used to provide external storage to machines outside the cluster by mounting the storage on a local virtual machine and then re-exporting it.

One other feature unique to the bulk storage is that the storage can be backed up on a regular schedule using a feature of the underlying ZFS file system called snapshots. Initially three types of configuration will be available for storage:

- **scratch:** no snapshots and no duplication. Data will be lost if there is catastrophic failure of the underlying disk array (three or more disks fail at once). Every 1TB of purchased storage gives 1TB of available quota.
- **work:** a snapshot is taken every 2 hours for a period of 48 hours and every week for a period of a month but all the snapshots are stored on the same server. This will allow recovery of accidentally deleted files but data will still be lost if there is catastrophic failure of the underlying disk array. To provide storage for the snapshots every 2TB of purchased storage gives 1TB of available quota. This is enough to allow the entire quota to be rewritten once within the period of a month.
- **home:** snapshots on the same schedule as “work” plus the entire drive and all snapshots are duplicated to a second server. Loss of data will only occur if both servers suffer catastrophic, failure of three disks each at the same time. This is by far the safest storage option but to provide the storage for the snapshots and duplication every 4TB of purchased storage will give 1TB of quota.

Snapshots containing the older version of the data are accessible from the same mount point as the volume itself so recovering an old, deleted file is as simple as navigating a directory structure and copying the file back into the main volume. In the event of a very significant data loss it is possible to revert the entire volume to any given snapshot with administrator intervention.

The additional capacity required to store the snapshots is an estimate because it depends on the frequency with which data is overwritten. A volume which is close to full and/or which regularly has large portions overwritten may exceed this estimate in which case it may become necessary to either reduce the snapshot schedule, reduce the usable quota or purchase additional storage. The reverse is also possible for a volume where the data change rarely and so snapshot storage use is well below the estimate.

The snapshot schedule is configurable should researchers have specific needs not covered by these options. However it should be noted that duplication between servers is only possible with snapshots enabled and that non-standard snapshot schedules may require a different ratio of storage purchased to available quota in order to provide enough storage for the snapshots.

Hardware Lifetime

All the hardware in the cloud has an underlying three stage life-cycle:

- **3 years:** the hardware is under warranty and any hardware failures will be replaced.
- **3-5 years:** guaranteed “best effort” support where we cannibalize parts from some machines to keep others up and running.
- **After 5 years:** “best effort” support continues until either nobody is using them or they are too costly to maintain at which point they are retired and removed from the cloud.

This is designed to provide the same benefits as owning your own hardware but without many of the disadvantages. A direct comparison is given below:

	Stand-alone Machine	**Research Cloud**
Failure in first 3 years (inside warranty)	Researcher has to arrange vendor replacement themselves which may include diagnosing the problem and shipping parts depending on the warranty details. Resources are unavailable until replaced/repaired.	Cloud support identify the problem and arrange for vendor replacement. For most cases the redundant cloud design and spare resources will mean that there is little to no effect to a researcher's resources while the system is awaiting repairs.
Failure after 3-5 years	Researcher is responsible for diagnosing the problem and purchasing replacement parts. Without redundant parts if the failure is serious the affected resources may be too expensive to replace and so are completely lost.	Cloud support identify the problem and for cheap repairs the cost may be covered by the Faculty. For more expensive repairs previously failed machines can be cannibalized for parts. There may be a slight degradation in resources over time (depending on the nature of the failures) but it will not be all-or-nothing.
After 5 years	Researcher runs the hardware until it dies, support is too expensive (in time or money) or it is no longer used.	Cloud support run the hardware until it dies, support is too expensive (in time or money) or it is no longer used.

Storage Failures	Researcher is responsible for ensuring that their data is safe against hardware failure by ensuring that there are redundant copies of important data, monitoring their own hardware for failures and responding rapidly to replace failed hardware...or they live with a significant risk of data loss on failure!	All data is stored redundantly and the hardware is monitored and rapidly replaced by cloud support. The high level of redundancy (at least 3 disks must fail before any data is lost) means there is plenty of warning of failures so appropriate action can be taken before data is lost in all most the most unlikely of scenarios.

In addition to all the above if researchers are still reliant on old cloud hardware when a decision to shut it down is made then the Faculty of Science may decide to replace the affected researchers' equipment with modern resources that provide an equivalent computing power (if Moore's law still holds one modern machine would replace ~8 5-year old machines and more if the machines are older). Such a decision would require a compelling case for the critical need for the resources in question.

CPU Quota Buckets

The power of a CPU core and the amount of memory required to run machines tends to increase with time. To ensure fairness when purchasing resources we will split the CPU quotas into "buckets" whenever there is a significant improvement in CPU and/or memory specifications. This will ensure that those purchasing new resources will have access to new CPUs while still ensuring that older resources are available to those who purchased them.

The frequency of buckets will depend on how rapidly CPU and memory technology advances. For example there is a recent trend where the number of cores per CPU increases while the performance per core does not change much. In such a case the price per CPU unit would drop but they would remain in the same quota bucket.

Best Practices

To maximize the benefits of the research cloud the following are recommended as best practices:

- Shutdown virtual machines which are not in use. The cloud monitors the average number of machines in use and, the lower this average is, the larger we can make everyone's quotas in relation to the resources purchased. So by shutting down machines which are not in use it will help increase your quota.
- Enable automatic software updates in your operating system (or do not disable them if you use a provided system image!). These are configured to only perform security updates and should not change major package versions significantly. If your machine gets hacked you can lose research data and spend a lot of time getting the associated mess sorted out.

- Only use external IP addresses for machines which really need them. IP addresses are a limited resource and machines which only have internal IP addresses are more secure.

Future Features

Once the cluster is rolled out and providing solid, basic functionality several features are planned to be implemented:

- **Better Resource Sharing:** Support for resource scheduling in clouds is still in its infancy. The current model does not provide motivation to shut down machines when not in use to free up resources which others might use. Compute Canada is also aware of the need for this and as the technology develops options for better sharing may be added. Whether to take part in such a future scheme would be at the discretion of the researcher.
- **Specialized Machines:** If there is demand for particular machine configurations outside the normal balance of CPU and memory, e.g. you need 64GB per core, then we may add support for this in the future. If this applies to you please contact your departmental ITOC representative and let them know what you need!

The research cloud is a research computing resource. We actively encourage researchers to find new and better ways of using the technologies it employs to further their research goals. If you have an idea to improve the cloud, or need something which the cloud does not currently provide please get in touch with your department ITOC representative!