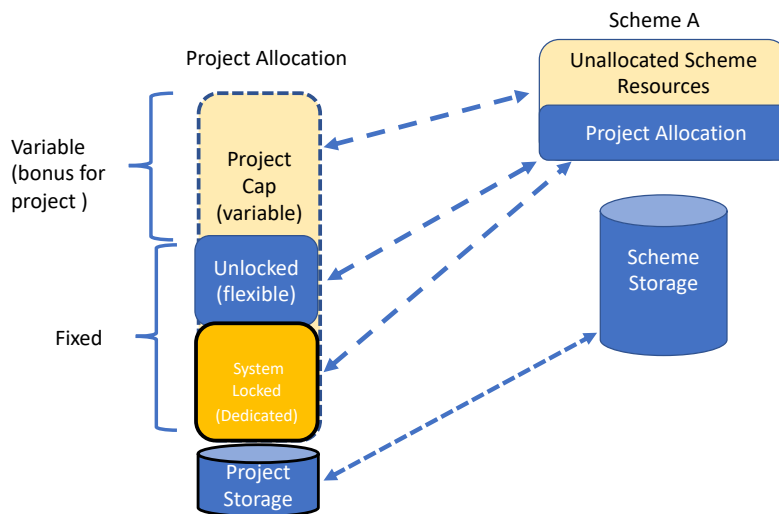# Lead CI Allocations

**Merit based Projects – no institutional backing**

For projects that receive a merit allocation for HPC, the basic allocation process remains the same for CI. SU allocations are assigned to the project for HPC time and are used as jobs are submitted by members of the project. Within a project however, it is proposed that the lead CI will now be able to quota user so that an individual user is limited in the amount of HPC SU that they can use from the project allocation. The default is an unlimited quota potentially allowing an individual user to use all the HPC SU resource as is currently the case. When a quota is applied it is absolute and is a hard limit on that user.

Institutionally backed projects

For projects that are supported by an institutional scheme, with a merit allocation, the following diagram attempts to provide a broad overview showing the flow of SU allocations from resources pools to services. The bidirectional dashed lines are "two way" meaning that allocations can be given out and pulled back if the resource has not been used.

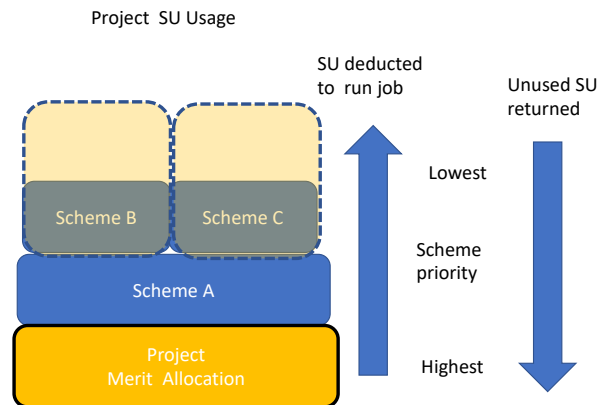Under the proposed changes, the Lead CI will be able to choose where the SU for that project is spent.
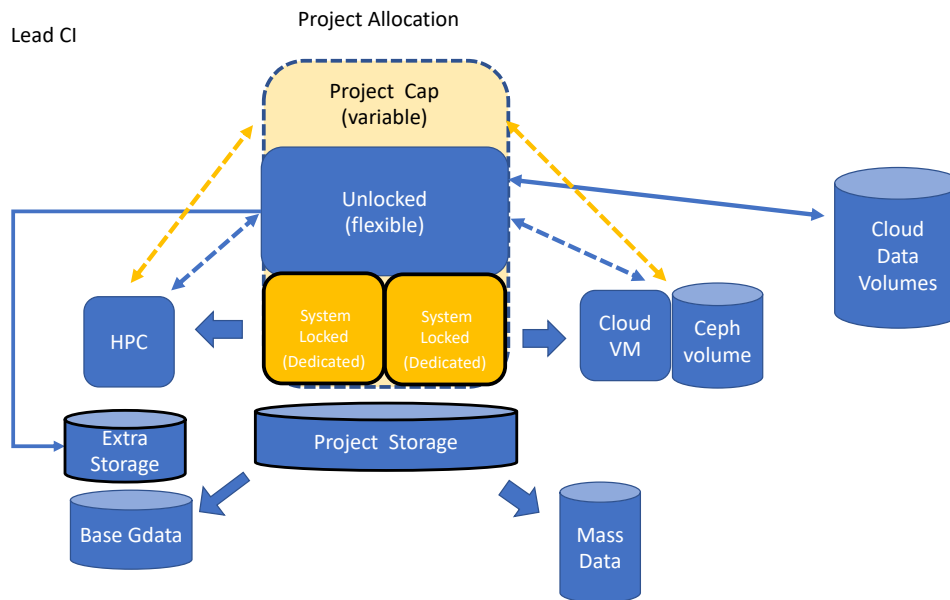
Dedicated allocations such as merit allocations made for HPC access must be used for HPC as reflected in the diagram above.

Project allocations made by the institution can now be used by the Lead CI to be used for access to Cloud resources, HPC resources or storage.

HPC Allocations

In order for a HPC job to run, the job scheduler needs to confirm that the project has sufficient SU resources to run the job. It does this by checking the SU balance within the project. If there are sufficient resources, the job can commence.

Project SU Usage

SU deducted to run job

Unused SU returned

Scheme priority: Lowest → Highest

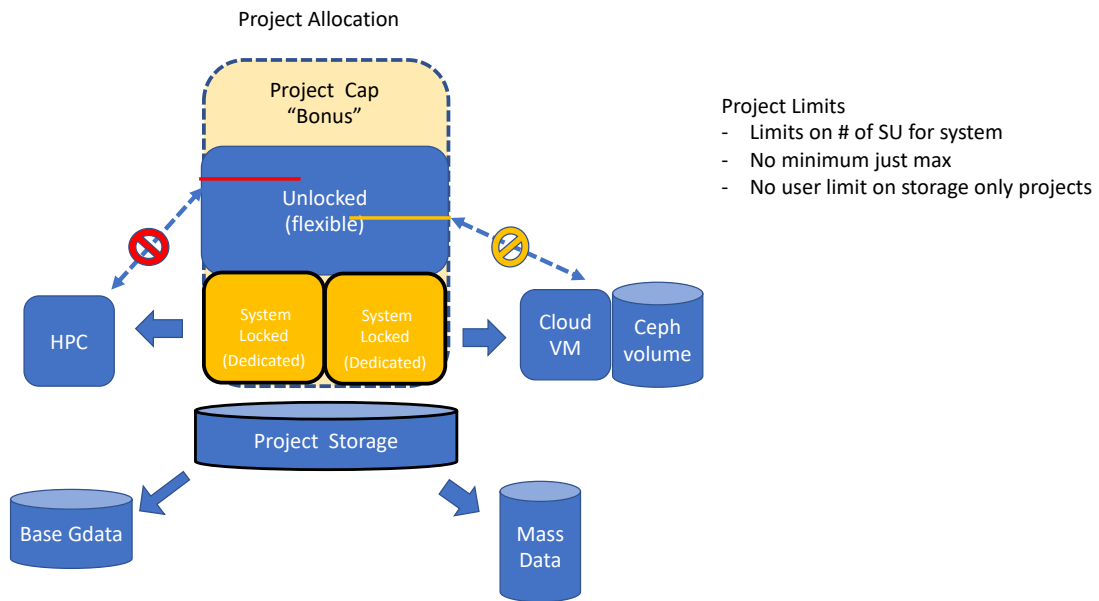Scheme B · Scheme C · Scheme A · Project Merit Allocation

The balance for the job is then deducted from the allocations in priority order of the schemes. If the job does not use all the resources, they are returned to the eligible schemes that were used to draw resources for that job but in reverse priority order. This ensures that a project's merit allocation will be used first before utilising any of the institutional schemes.

If there are no available SU within the project allocation, the project will confirm if the project has been allocated any SU under the "project cap". If the project has been granted access to some additional time under the institutional scheme, the scheduler will check if the scheme has a non-zero balance and has the resources available for the job.

This process will continue until, the project reaches its project cap or the scheme pool of unallocated scheme resources has been exhausted. Once this occurs, the project will no longer be able to run HPC jobs.
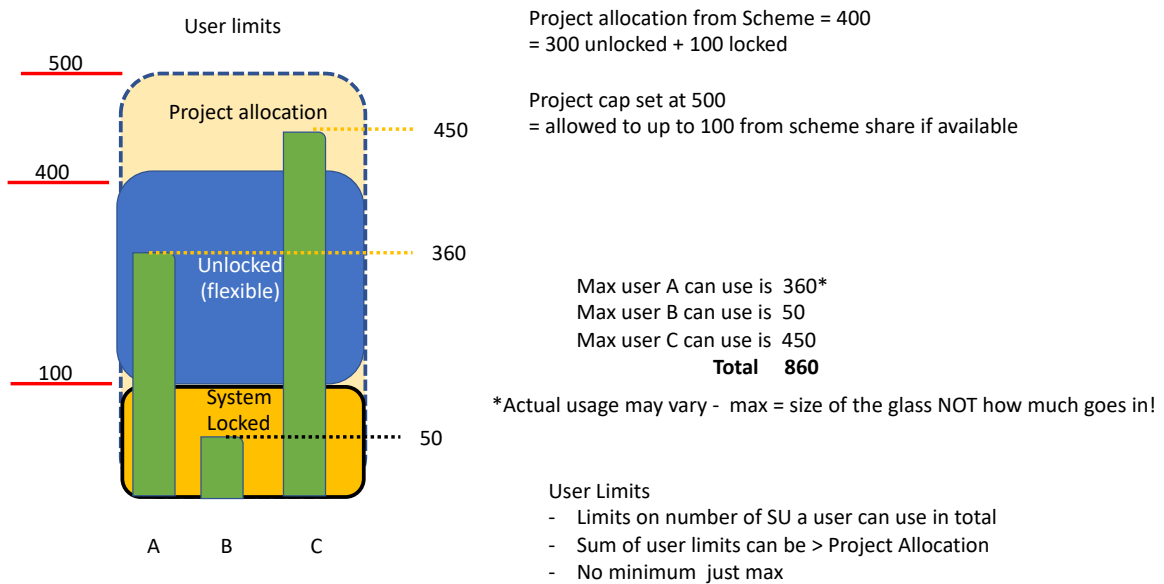
Project Quotas
The Lead CI will be able to limit the amount of their projects that can be used on a particular system. In other words, a limit can be placed on how much of the allocation can be used for HPC vs how much of the projects SU resource could be used on Cloud compute.

Project Allocation

Project Cap
"Bonus"

Unlocked
(flexible)

HPC

System
Locked
(Dedicated)

System
Locked
(Dedicated)

Cloud
VM

Ceph
volume

Project Storage

Base Gdata

Mass
Data

Project Limits
- Limits on # of SU for system
- No minimum just max
- No user limit on storage only projects

Once a system reaches the quota in SU that has been set, no more jobs can be run on that system even if more SU are available within the project allocation.

User quotas

It is proposed to allow the lead CI to set a user-based quota. This quota sets the maximum SU resource that can be used by that researcher within a particular project.

User limits

500

450

Project allocation

400

360

Unlocked
(flexible)

100

System
Locked

50

A   B   C

Project allocation from Scheme = 400
= 300 unlocked + 100 locked

Project cap set at 500
= allowed to up to 100 from scheme share if available

Max user A can use is 360*
Max user B can use is 50
Max user C can use is 450
       Total   860

*Actual usage may vary - max = size of the glass NOT how much goes in!

User Limits
- Limits on number of SU a user can use in total
- Sum of user limits can be > Project Allocation
- No minimum just max

The quota is the "potential maximum", it does not guarantee a user have access to this amount of resource. Once a user reaches this limit, they can no longer use SU from that project to run jobs. At any point in time the Lead CI may change this limit to increase the resource that might be available to a user. A Lead CI can reduce the limit to below what a user has already used however the only effect this will have is to block the user from starting any more jobs it will not stop running jobs.

Cloud Allocations

At NCI the accounting of cloud usage has been in arrears. With the proposed new accounting system, the model changes slightly to be like a HPC job with the SU deducted in advance and held when a VM starts.

Given that the length of time for a VM will run is indeterminate at the time it is initiated, the accounting system deducts two hours of run time at the start of the VM from the project allocation and holds the allocation. If the VM runs for 30 minutes and is then deleted, the remainder of the SU is recredited back to the project schemes.