

Kempner Institute Compute Governance Guidelines

Revised: 9/28/2023

Approved: 9/29/2023

Overview

The Kempner Institute will build and maintain one of the largest academic computational clusters in the world supported by a team of engineers, engineering fellows and dedicated research computing staff. The cluster will be state of the art and structured to support a diverse portfolio of research in intelligence.

This document outlines policies governing use of and access to this equipment, access to the engineers and other support staff, contributions of hardware to the cluster, costs associated with use of the cluster, and fair share principles. *This is a working set of guidelines which will be revised as appropriate to better serve the needs of the Kempner Institute community and to help advance its scientific mission.*

Access to the cluster

The Kempner Institute cluster will be housed with other [FASRC](#) equipment. Access to the cluster is determined solely by the Kempner Institute. The cluster may only be used for research that is in alignment with the Kempner Institutes mission. Access is granted to:

- Faculty Directors and Institute Investigators and members of their labs/research groups.
- Associate Faculty and members of their labs/research groups working on Kempner-related projects.
- Visiting Scientists
- Kempner Research Fellows
- Kempner Graduate Students
- Kempner Institute Grant recipients and members of their lab working on their funded research project who have requested and been granted access to the cluster. Access to the cluster will be restricted to the award period and may only be used to advance the specific work of the funded proposal.
- Kempner Affiliate Faculty who have requested and been granted access to the compute cluster for specific projects or research activities.

General principles

Alignment with Mission

Access to the Kempner Institute cluster is based on an ongoing and direct relationship with the Kempner Institute. Access to the cluster includes both the ability to run jobs on the cluster, and to work with engineers and research computing support staff employed by the Kempner Institute. Access to and use of the cluster and the engineering support will be solely to facilitate research that is directly related to the Kempner Institute's mission and strategic initiatives.

Using the cluster for research objectives that are not related to the Kempner Institute's mission may result in the loss of access to the cluster or a significant reduction in fairshare allocation.

Good citizenship on the Cluster

The goal of the cluster is to support research across a wide community of users and projects. We do not wish to limit submission of jobs on the basis of size and encourage users to use capacity when it is available. As a result, users are able to run jobs that are very large when there is capacity.

However we need to balance the goal of supporting large projects with the understanding this is a shared resource. Users should always work to make sure that code maximizes the GPUs requested and runs as efficiently as possible. However, even with efficient code, users submitting large jobs may monopolize resources on the cluster and may prevent other members of the community from being able to engage in research. Thus, we rely on users to act in good faith and use consideration when submitting jobs.

Users who submit large jobs (i.e. jobs that use more than 25% of cluster resources) which they expect will run for a long period of time (e.g. 4 or more hours) should:

- Use the [Kempner requeue](#) which allows jobs to be preempted by other jobs and then restarted, thus allowing large jobs to run without disrupting access for the rest of the community.
- If not using the requeue, only submit jobs when there is significant excess capacity and actively monitor the cluster (e.g. hourly) to ensure that there are still resources available for other users.
- Use a run-time limit to prevent large jobs from running for a long period of time unexpectedly.
- Consider running the job with a smaller number of GPUs over a longer time-period.

If none of the above strategies will work, users should submit a reservation request ([described below](#)) so that large projects can be planned and communicated to the community, limiting disruptions for other users.

Users who monopolize cluster resources may have jobs canceled without notice, have fairshare significantly reduced, and in serious or repeated circumstances, may lose access to the cluster.

Graduate Student Access

Kempner graduate students have access to the cluster to support their own research. If a Kempner graduate student attempts to run jobs on the cluster for others in order to circumvent the access categories described above, or uses the cluster for purposes other than their own research, the student will lose the ability to submit jobs via fairshare and will only be able to access the cluster at the same level as unaffiliated students.

Fairshare

Use of the Kempner cluster for approved users is governed by fairshare. The algorithm prioritizes a balanced allocation of resources, aimed at facilitating the timely completion of tasks from various user groups (see [FASRC fairshare accounting](#) for a more detailed overview). This means that jobs, particularly those that are resource intensive or are being run in labs with high recent usage, may not run immediately or on demand. Fairshare guarantees access to resources averaged over a period of several months. At any given moment in a time a user or lab may be using a greater or lesser amount of their available share.

Fairshare is allocated based on the user's affiliation category below. Each category receives a share of the available computational resources across each category and then subdivides those resources to members within each category. Resources might occasionally be reduced due to maintenance or approved reservation requests (described below).

Current allocations:

- 80% - Institute Investigators, Associate Faculty, and Affiliate Faculty (Affiliate access is for specific research projects only) and their labs
- 12% - Kempner Graduate Student Fellows
- 8% - Kempner Research Fellows

Members within each group share computation resources equally (e.g. each of the 8 research fellows are allocated 1% of available resources on the cluster).

As the computing needs of the Kempner community and the computing power within the cluster grows, fairshare levels will be reassessed by the Kempner Cluster Committee and changes announced prospectively.

However, if users believe that they have insufficient access or that the wait time to run jobs is longer than it should be, we encourage them to contact Kempner and FASRC to help identify potential bottlenecks and troubleshoot solutions.

Reservation Requests

There may be times in which the Kempner Institute wishes to allow users to request access to the cluster, reserve GPUs on the cluster for specific projects related to the Kempner's strategic initiatives, specialized educational activities, to meet specific deadlines, or to allow jobs to run beyond the cluster time limit.

Requests for reservations should be made using our [request form](#) available here and on the Kempner Institute website. Reservation requests are forwarded to and evaluated by the Kempner Cluster Committee.

A decision made by the committee is final although the committee may, at its sole discretion, provide feedback on the request. If a request is granted, the committee may also make modifications to the requested resources (e.g. date, run length, GPUs, etc.).

Courses

The Kempner Institute cluster is not intended to replace the obligation of any school or program to provide the resources necessary to pursue its own academic mission, including in the teaching of undergraduate and graduate students.

Courses that have been identified as "Kempner Institute" courses— for example boot camps— may request a time-limited reservation for the use of students enrolled in the course to complete course-related projects. Any special access granted will immediately terminate upon the end of the course.

Access should be requested well in advance in the event that cluster access is not granted so that alternative arrangements might be made.

If granted, the Kempner Institute solely provides an approval that allows cluster access for specified users. Course directors will need to work directly with FASRC to ensure that support is sufficient and available for the course duration and that students have appropriate access established. At no time will the Kempner Institute pay for the operating expenses related to use of the cluster.

Requests for course reservations should be made using our [request form](#) available on the Kempner Institute website. Course-related reservation requests are forwarded to and evaluated by the Chair of the Kempner Cluster Committee and the Kempner Assistant Director of Education.

Overhead, operating and data storage expenses

As is the current practice with users of other computational resources hosted by the FASRC, operating expenses for cluster use will be charged as per existing FASRC processes. Dictated by MOU at the school level, operating expenses are typically charged at the school level based on the primary academic appointment of the PI/user and charged based on resources used. More information on this can be found on the FASRC website under "[Offerings \(Tiers of Service\)](#)".

At no time will the Kempner institute pay for data storage expenses on behalf of faculty, fellows or students using the cluster. Data storage expenses may be charged as a direct cost on Kempner Institute research funding for those that have such funding.

For those very few people who are solely affiliated with Harvard University via the Kempner Institute (i.e., Kempner Research Fellows, Kempner Visiting Scientists, and Kempner Institute computational staff) and who have no other academic home, the Kempner will pay for cluster operating expenses and storage as required.

Engineer & Research Computing Support

Engineers are available to support Kempner Institute faculty, fellows and students. They will: participate in the development, troubleshooting and review of drafted code; provide support to projects related to the Kempner's strategic initiatives; and advise on scalability, code management, documentation and best practices for open science.

Affiliate Faculty will have access to Kempner Engineers and research cluster (RC) support staff only if they have contributed hardware to the Kempner Institute cluster or if they have provided salary support for Kempner Institute Engineers via grants or other sources of research funding. Those who are unaffiliated with the Kempner Institute will not have access to the Kempner Institute engineers or RC support staff.

Research computing staff funded by the Kempner will have two primary roles: they will serve as a partner for the Kempner community to ensure the optimal use of the cluster and will be responsible for on-site management of the cluster at the MGHPCC. They will provide ongoing review of jobs run on the cluster to identify those that use the cluster inefficiently. They will also

be available to provide support for questions related to scheduling including estimating resources required for scheduling.

Faculty Contribution to the Kempner Institute Cluster

Faculty investment in the Kempner computational infrastructure not only accelerates each faculty member's own research but also fosters more ambitious research initiatives for the institute as a whole, helping to ensure that the Kempner remains at the forefront of academic computational capabilities.

When considering the benefit of contributing to the cluster versus pursuing an independent purchase, Faculty contributors benefit in 5 major ways:

1. The Kempner is buying at scale, thus it gets significantly better pricing per GPU relative to more modest purchases that might be made by individual faculty.
2. Faculty are able to obtain fairshare equivalent that includes fractional computing power which maximizes their funding. For example, one might receive a Kempner fairshare equivalent of 4.5 GPUs where, if purchased independently, one might only purchase 4 GPUs.
3. The Kempner Institute will provide for the installation, ensure adherence to the warranty, and ensure maintenance support for equipment.
4. Faculty who contribute receive access to the rest of the cluster, allowing them to occasionally exceed the computing capacity of their contributions and if any of the equipment fails or is taken down for maintenance, this still allows access to the rest of the cluster for research.
5. Faculty contributors get access to Kempner employed software engineers who are deeply familiar with its hardware and its science.

However, in accepting contributions to the cluster we must balance the desire to add to the total resources available to the Kempner community with the need for effective planning, efficient installation, and compatibility with existing resources.

Thus, any equipment added to the cluster must follow the guidelines below. These terms must be acknowledged as part of the Contribution Request form (template below for reference).

It is possible that in any given year the Kempner may decide to limit or forgo faculty contribution to the cluster or limit or forgo compute purchases entirely. The Kempner cannot promise a specific date whereby equipment will be ordered, received, installed, and open for use. However, reasonable effort will be made to ensure open communication about timelines so that faculty contributors can make appropriate plans for research.

Terms for Faculty Contribution to the Kempner Cluster

1. Only Kempner Faculty (Institute Investigators, Associate Faculty, Affiliate Faculty, and research award recipients) may request to contribute funds toward the annual purchase of equipment for the Kempner Institute.
2. Individual contribution of equipment, as opposed to funds, will not be accepted.
3. Faculty must demonstrate that research to be undertaken on the cluster is in alignment with the Kempner Institute's research mission.

4. Faculty will be credited an increase in their fairshare on the Kempner cluster proportional to the funds they contributed against the total cost of the Kempner's annual purchase of equipment.
5. This increase in fairshare on the cluster is normalized by computational power. For example, a 5% financial contribution will result in a credit equal to 5% of the computational power of the resources purchased in that year.
6. As described in the Kempner Compute Governance guidelines, all use of the Kempner Institute cluster is governed by fairshare. Priority access to resources, by definition, prevents other users from accessing GPUs. Thus, contribution of funds will not result in any priority access to GPUs. If urgent or dedicated access is needed faculty may make reservation requests as per Kempner Compute Governance guidelines.
7. The increase in credited fairshare will remain in place for as long as the equipment purchased is operational. Generally, equipment will be warranted for 5 years and thus is expected to last 5 years. As a result, the incremental increase in credited fairshare will be expected to last for that time period.
8. However, if any portion of the hardware purchase fails or is otherwise decommissioned the fairshare credit will be reduced by a commensurate amount. For example, if faculty contributed 5% towards a purchase of 144 GPUs and 8 failed after warranty, faculty would have access to 5% of the remaining 136 GPUs as part of their fairshare allocation.
9. Kempner Institute has sole discretion to decide if and when any equipment will be decommissioned and removed from the cluster at any time. Faculty contributors will be notified as soon as reasonably possible if and when this determination is made.
10. Once installed in the cluster, the Kempner Institute will ensure maintenance of the equipment.
11. If the faculty member leaves Harvard prior to the end of the useful life of the equipment, the equipment will be retained by the Kempner Institute for general cluster use and no funds will be provided in recompense.
12. For purchases made using funds **internal** to the Kempner Institute (e.g. Kempner start-up funds, research grants, Kempner faculty research funding):
 - The Kempner Institute will accept costs related to rack and row installation as is customary for Kempner Institute purchased equipment.
 - Operating expenses related to use of the cluster will be charged to schools based on use, as is customary with all computing equipment managed by FASRC.
13. For purchases made using funds **external** to the Kempner Institute:
 - Funding that limits use or users of the equipment will not be accepted.
 - Funding from a restricted source (grants, foundation, or donor funding) must include approval from a department representative (e.g. Office of Sponsored Programs officer, or Finance Manager or Director, etc.) who can confirm that the use of the funds is an allowable expense per the terms of the funding entity.
 - Operating expenses related to use of the cluster will be charged to schools based on use, as is customary with all computing equipment managed by FASRC.
 - Costs associated with facility expenses to build and maintain hardware at the rack and row level are typically charged back to the faculty member's school as is customary with investigator contributed equipment within the FASRC. Understanding that the Kempner compute cluster is new and evolving, the Kempner will pay these costs for faculty who provide funds in AYXX for equipment arriving in AYXX.

Faculty Contribution Request Form

Anticipated Total Purchase: \$XX.XXM, XXX GPUs (XX nodes), MM/YY anticipated installation date, YYYY anticipated end of warranty period

Name:
Department / School:
Website:

Amount to be contributed:
Source of Funding:
33-Digit Account String:

Describe research projects to be undertaken on the cluster (250 words):

Signature: _____

My signature confirms that I have read the terms outlined above and acknowledge that they are binding.

RESTRICTED FUNDING ONLY:

Name of approver:

Signature: _____

My signature confirms that I have read the terms outlined above and that this is an allowable use of funds as per the funding entity guidelines or requirements