# CANFAR Science Portal

## Overcoming Challenges to Enable Next-Generation Astronomical Research

Kevin Casteels · Canadian Astronomy Data Centre · Canadian Advanced Network for Astronomical Research · National Research Council of Canada

ADASS 2026
Perth

**161 → 268**
monthly active accounts in two years

**5**
accounts do half the computing

**21%**
of reserved CPU and memory is used

**68%**
of unused memory is in notebooks

**150 → 230**
interactive jobs in a typical hour

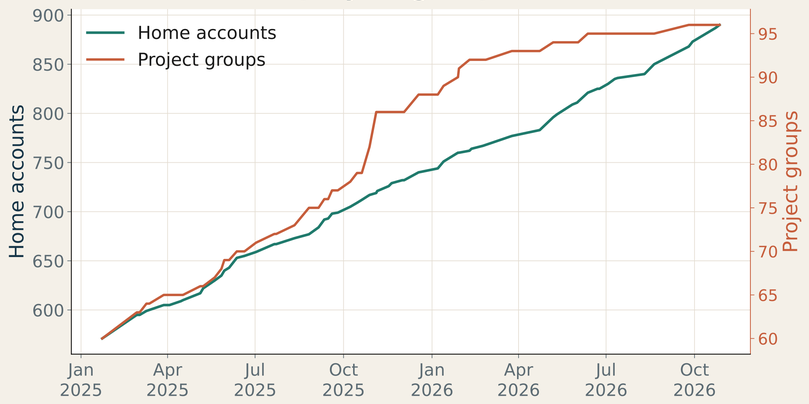
## 1 A primary hub

The CANFAR Science Portal is a web-based interface that allows astronomers to run different session types, including JupyterLab notebooks, CARTA, desktop sessions, and batch sessions. The system runs on Kubernetes using custom software developed by the CADC and others. It has been serving the Canadian astronomical community since 2022 and has seen significant and steady growth over that time.

During the past two years the user base has approximately doubled, and the number of active users has also steadily increased.

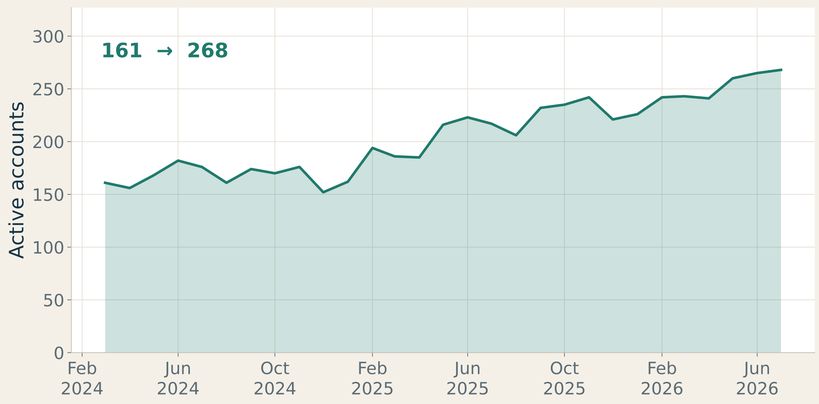
The user base is diverse, with most users running a couple of interactive sessions, and a few very heavy batch users who dominate the usage of compute resources. Balancing these different demands has been a challenge.

### Home accounts and project groups



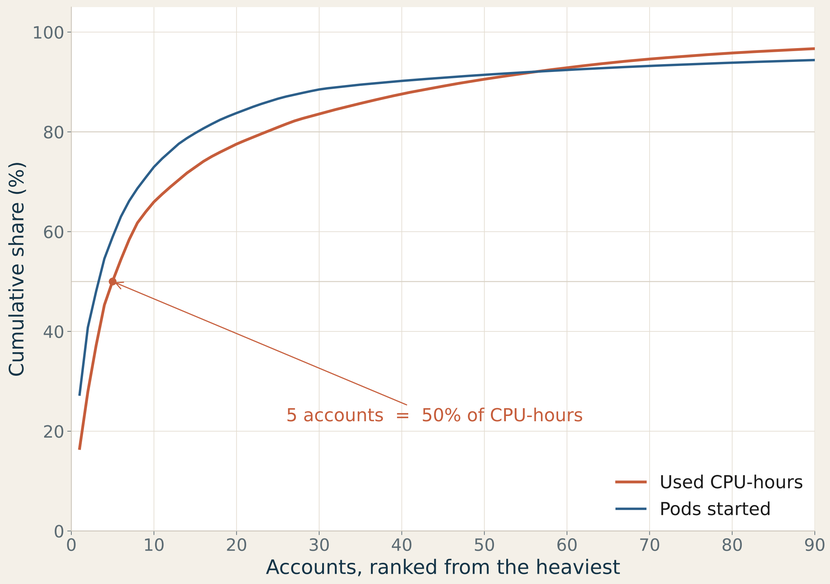

Registered home accounts and project groups.

### Accounts active each month



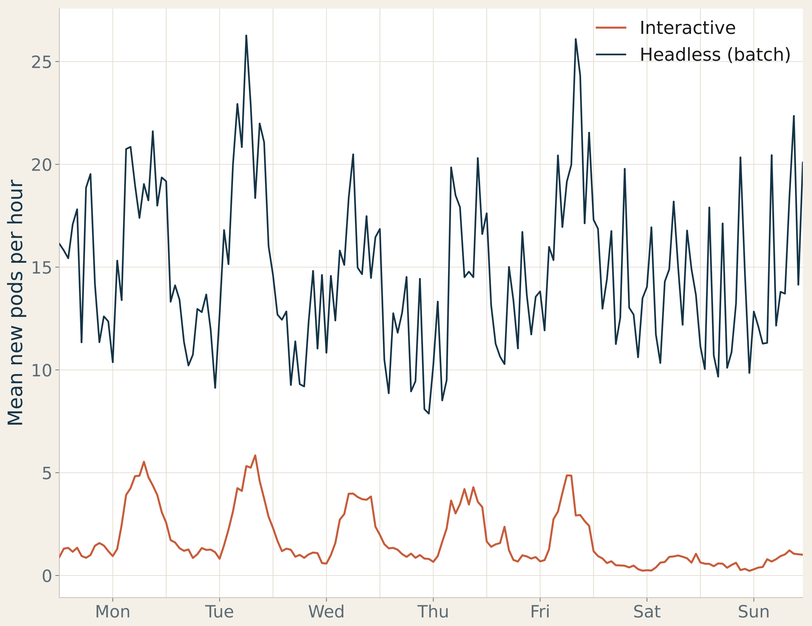

Complete months. An account counts if it has one hourly sample.

### Half the computing comes from five accounts




CPU-hours are cores actually used, summed over hourly samples.

### New sessions follow the workday. Batch starts do not.



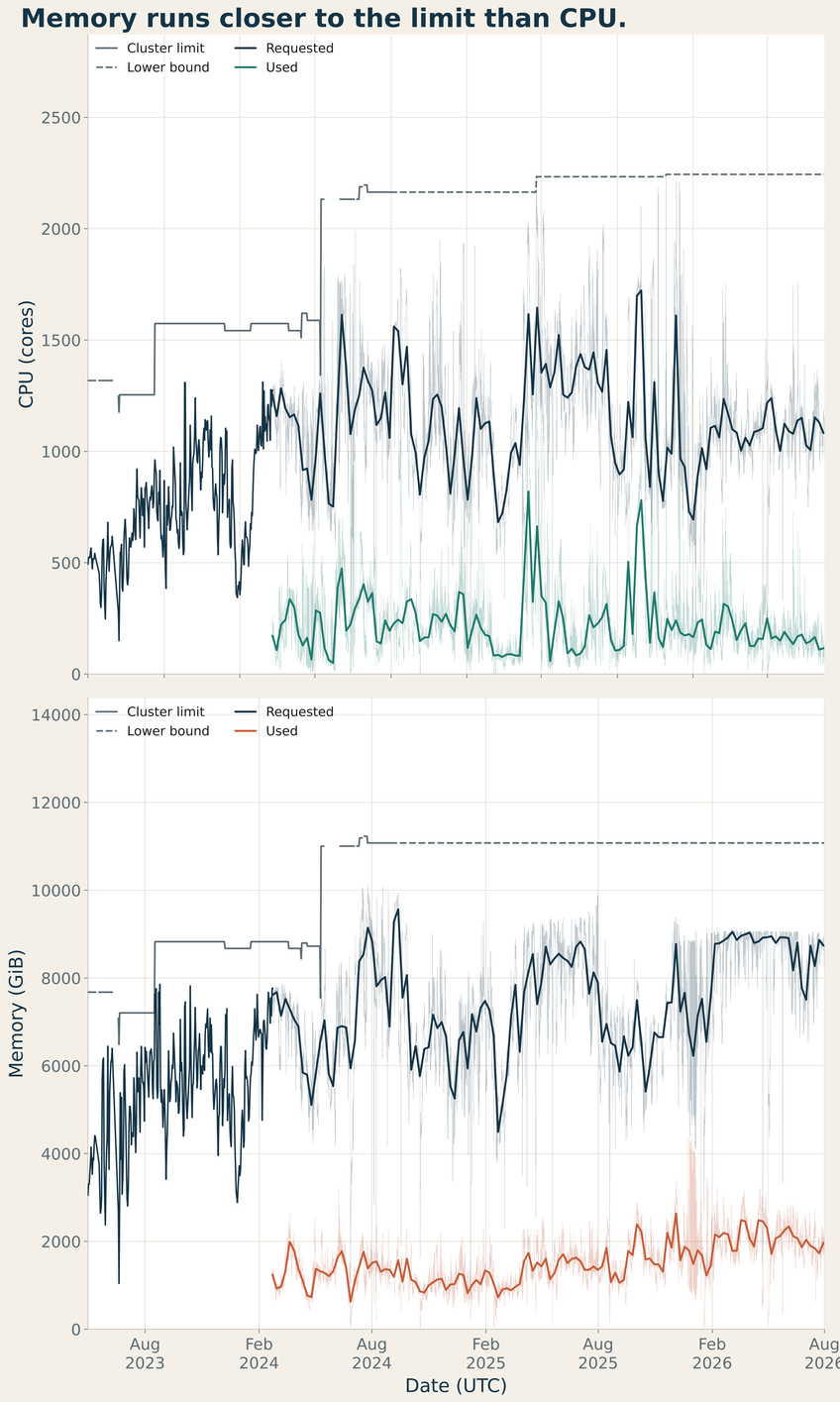

Hour the pod was first seen. Interactive includes notebooks.

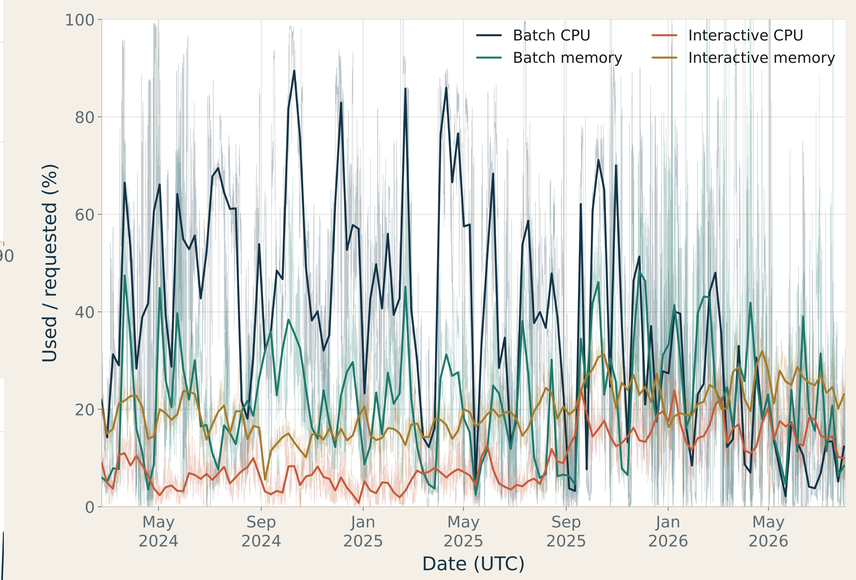
## 2 Where the cluster fills up

Users can specify the resources they need for their work, and those resources are then dedicated to that session. When the cluster runs out of resources, new sessions stay pending until space is freed up.

The limiting factor is often available RAM, with requested CPU usually well below the limit. Actual use is another matter. Batch sessions use more of the CPU they request than interactive sessions do. Interactive memory use is about 20%, with CPU lower, about 10%. Batch CPU use is about 45%, and batch memory use is about 23%.

### Memory runs closer to the limit than CPU.



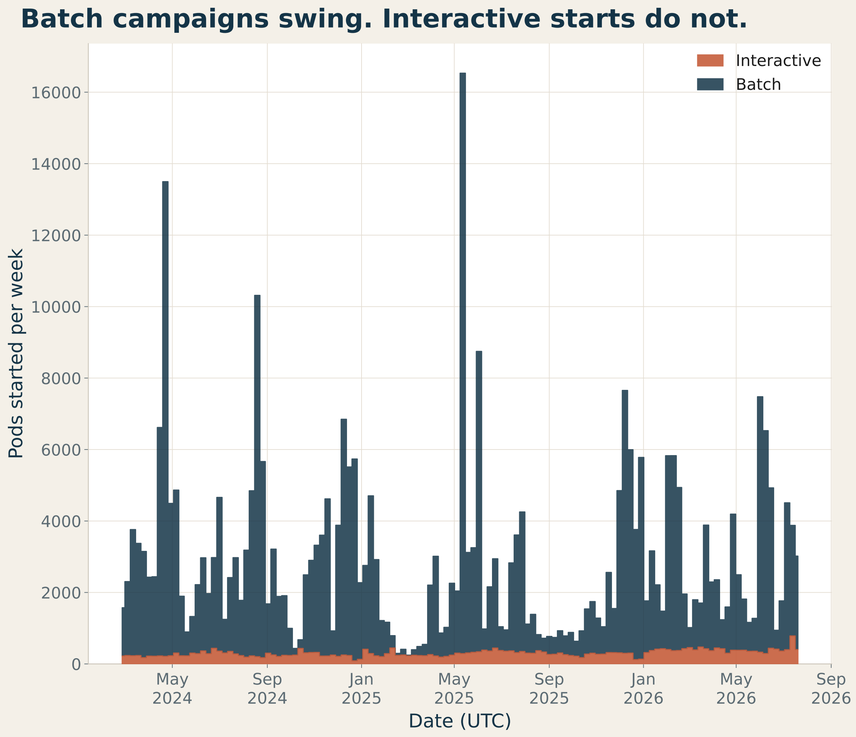

Dashed line is a lower bound after Aug 2024. Weekly binning.

## 3 What scaled, and what to copy

Over time, the number of interactive sessions has increased, along with the user base. Currently they sometimes dominate cluster resource usage, and have a higher priority than batch sessions.

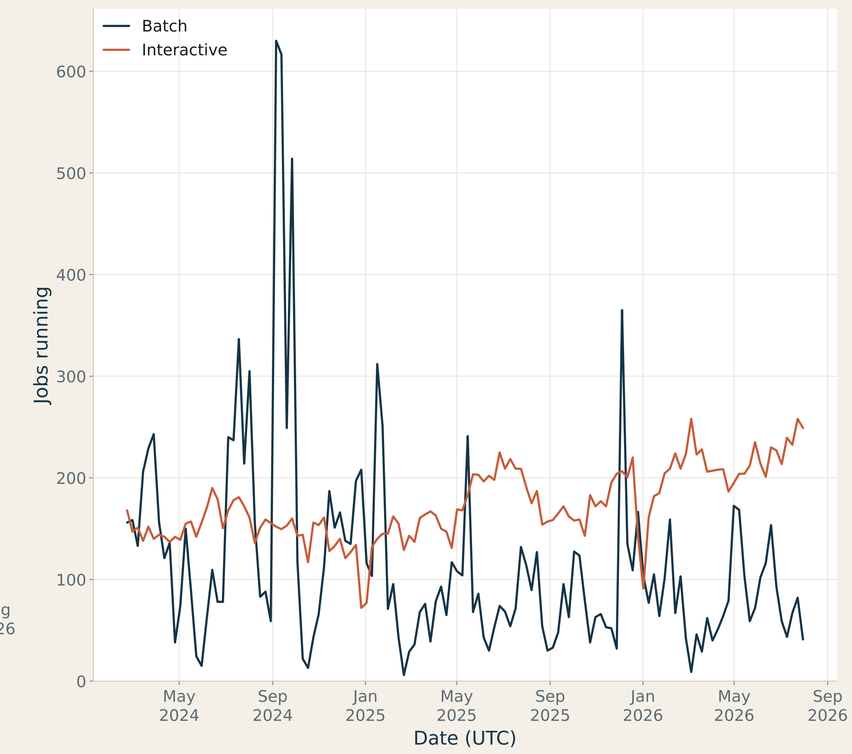
Batch sessions are often campaign-related and come in bursts, and are highly variable. When a campaign is in full swing, they have the potential to overwhelm the system, so managing that has been a challenge.

### Batch campaigns swing. Interactive starts do not.



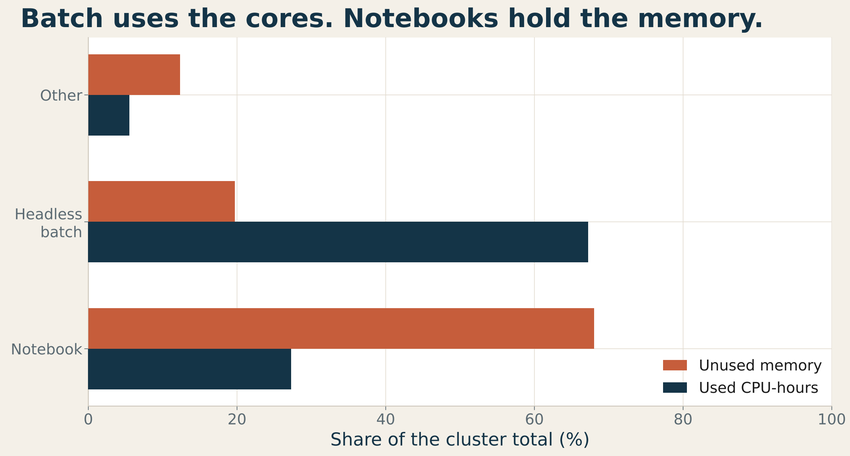

Pods first seen per week.

### Interactive jobs climbed. Batch comes in surges.




Weekly median binning.

### Batch uses the cores. Notebooks hold the memory.




Other is desktop, CARTA, and terminal. Unused memory is requested minus used.

### How resource management changed

**Initially**

The portal struggled with batch users overwhelming the system and leaving few resources for interactive sessions.

**Priority classes**

Interactive sessions were given higher priority than batch jobs, and sharing improved. The cluster was still overwhelmed by large workloads, especially when a batch user submitted many thousands of jobs.

**Kueue**

Jobs are queued before they reach Kubernetes, so large submissions no longer overload the system. Stability improved greatly. Sharing fairly between projects and groups is still difficult.

**Fair share**

Project and user quotas will give each the resources promised. Every user has different needs, so fair share is a target that keeps moving.

**Takeaways: Batch and interactive jobs compete, so a queuing system is essential. Jobs with flexible resource usage reduce waste. RAM is the limiter!**

Contact Kevin Casteels at kevin.casteels@nrc.gc.ca