



Updates on the CANFAR Science Platform

Brian Major
IVOA Interop, October 2022



CANFAR Science Platform Overview

A General Purpose Astronomy Platform

- **New Features and Software Containers benefit most projects and users**
- **Self-serve model (Users manage their groups, containers, data)**

Runs on Kubernetes on ~~Compute-Canada~~ Digital Research Alliance Canada (DRAC) infrastructure

- **Research Computing Group at University of Victoria (Jeff Albert, Ryan Taylor, Ryan Enge)**

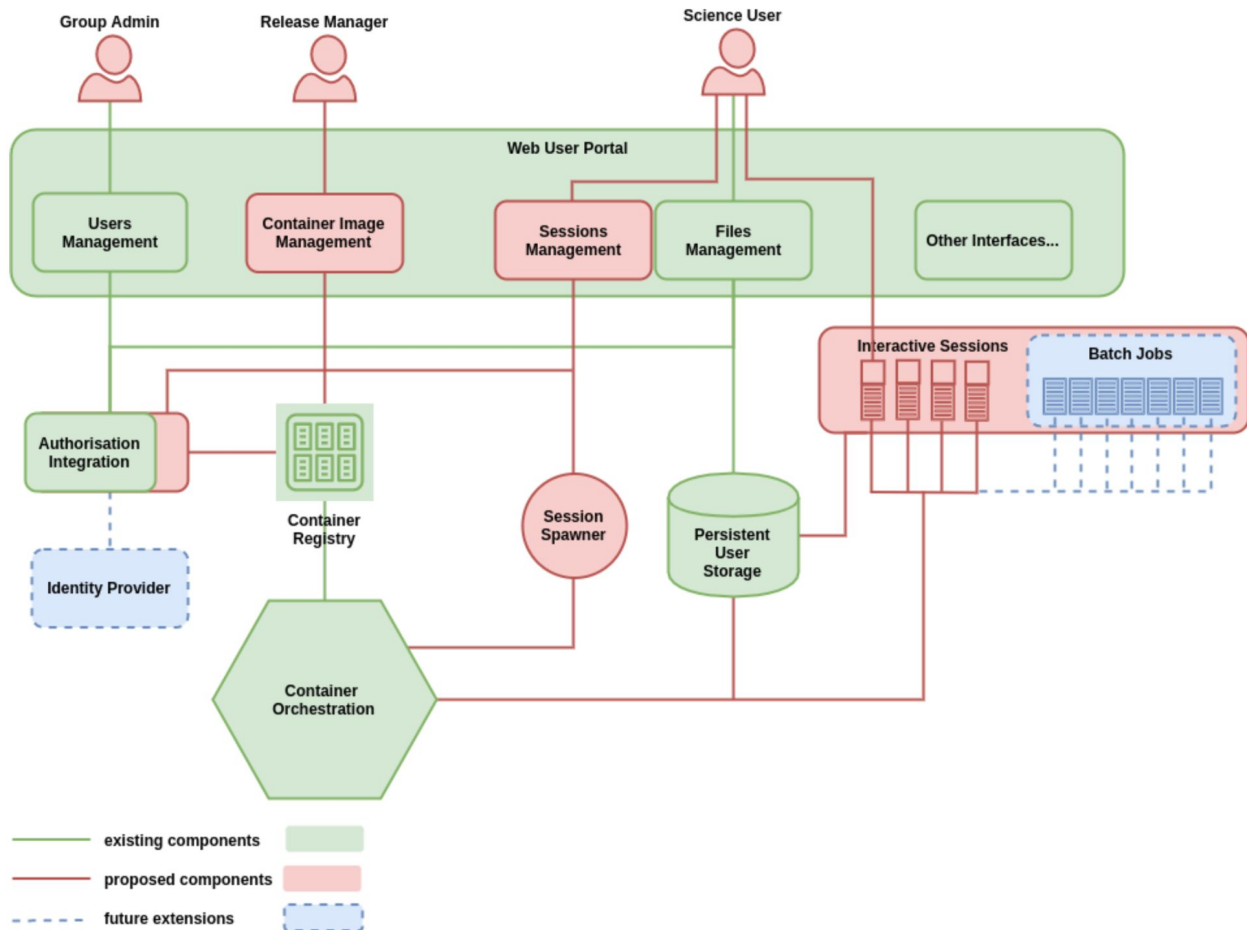
Rapid Growth

- **February: 120 Users, 20 Projects**
- **Today: 250 Users, 50 Projects**
- **Scale by adding nodes**
 - ~1200 CPU cores, 1xA100 GPU, 8xV100 GPU.
 - ~1.3Pb (2 VOSpaces: Object Store and File System)

Original Architecture Proposal (2020)

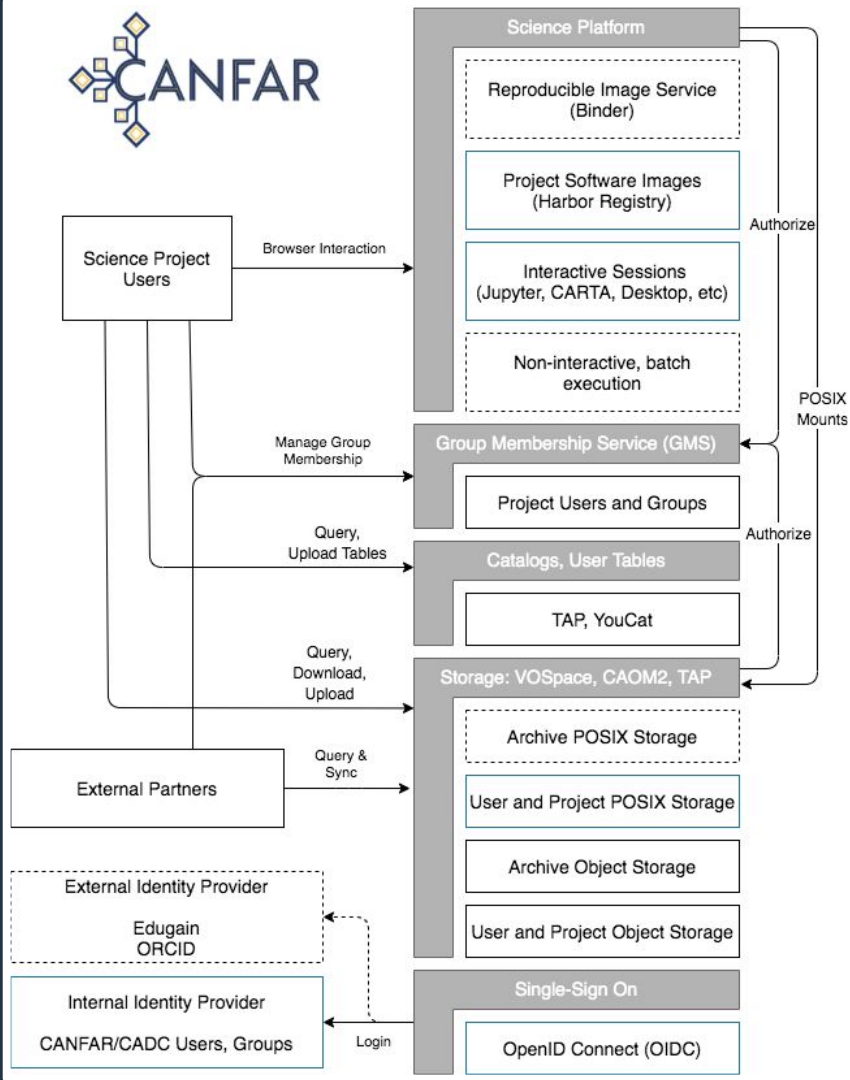
Closely resembles today's architecture.

Software Architecture



Key Components

- **Platform API**
 - "skaha"
 - launches containers (eg jupyter)
 - routes to containers
- **Image Registry**
 - technology: harbor
- **POSIX Mounted VOSpace**
 - "cavern"
- **Group Membership Service**
- **User-managed tables (YouCat)**



Simple Platform Use: Select and Run your Container

- Jupyter Notebooks
- CARTA
- Desktop
- Pluto Notebooks

Exploratory types:

- headless
- contributed
 - standard port
 - self aware of access URL

The screenshot displays the CANFAR Science Portal interface. At the top, there are navigation links for Documentation, Services, About, Open Source, Support, and a user profile for Brian Major. Below this is the 'Active Sessions' section, which shows three active sessions: 'carta1', 'desktop1', and 'notebook1'. Each session is represented by a small icon and a label. Below the active sessions is a 'Launch' section. A 'Session Type' dropdown menu is open, showing options for 'notebook', 'carta', and 'desktop'. The 'type' dropdown is currently set to 'notebook'. Other configuration options include 'container image' (images.canfar.net/highz-alma/notebook-astropy:0.1), 'memory' (16), and '# cores' (2). There are 'Launch' and 'Reset' buttons at the bottom of the form.

Jupyter Notebooks

- All containers run "as the user"
- All containers mount cavern VOSpace
 - /home
 - /projects
- Users' POSIX groups same as membership in GMS
- Permissions enforced on the file system via ACLs

The screenshot shows a Jupyter Notebook interface in a web browser. The browser tabs include CANFAR, Science Por, 1 - JupyterL, CARTA, and ARCADE. The address bar shows the URL: ws-uv.canfar.net/notebook/zieh596p/lab/tree/headless. The interface includes a menu bar (File, Edit, View, Run, Kernel, Git, Diagram, Tabs, Settings, Help, Share) and a memory indicator (Mem:137 MB).

The left sidebar shows a file browser for the directory / headless /. It contains two files: test-1a and test-1b, both last modified 3 months ago. The file test-1b is selected.

The main area displays a terminal window titled "Terminal 1" with the following output:

```
majorb ~ $
majorb ~ $
majorb ~ $ id
uid=20001(majorb) gid=20001(majorb) groups=20001(majorb),100(use
rs),30091(AdminTest),30093(ABC),30127(CANFAR-Staff),30272(Alinga
Test),30276(Adrian-TEST),30497(caom2TestGroupWrite),34127(CFHT-1
5AP10),34210(CFHT-15AP09),34241(CADC),34337(CAOM2),34347(CFHT-15
BP10),34381(CFHT-15BP09),34635(CFHT-16AP10),34637(CFHT-16AP09),3
4964(CFHT-16BP09),35030(W-CADC),35039(W-ALL),35091(CFHT-16BP10),
35124(cadcstats),35126(CFHT-17AP30),35130(cadc-dev),35131(CFHT-1
7AP99),35141(cfis-read),35207(CFHT-17AP98),35219(INAF-Group),353
50(CFHT-17AT10),35375(CFHT-17BP99),35383(CFHT-17BP97),35440(MAST
-RW),35544(CFHT-17BS02),35550(CFHT-17BT01),36002(arbutus-cloud-u
sers),36003(CADC-DEVELOPMENT),36227(jao-cadc),1025424273(skaha-u
sers),1240980498(rc-harbor),1477619040(ARCADE-Users),1623998838(
skaha-admins)
majorb ~ $
majorb ~ $ cd
majorb ~ $ pwd
/arc/home/majorb
majorb ~ $
majorb ~ $
majorb ~ $ cd /arc/projects
majorb projects $ ls
ALMA_Outflows          cfis                    LSST                    unions
antennae                chime_frb              mlao                    uvickbos
beta                    CIRADA                  new-earth               vertico
canucs                  dali_alma_data         NewHorizons
casa-data-repository   jwst-crds              ots
cfhtai                  k-pop                  signals
majorb projects $
```

Simple 1 0 Mem: 137.11 MB

Terminal 1

CARTA

- Now running version 3.0

The screenshot displays the CARTA web interface in a browser window. The browser tabs include Science Portal, ARCADE, CARTA v1.4, and JupyterLab. The address bar shows the URL: `ws-uv.canfar.net/carta/http/bjc3ggtg/?socketUrl=wss://ws-...`

The main interface features several panels:

- BHR71_cont_p1.image**: A central panel showing a radio astronomy image. The axes are labeled "Declination" (y-axis, 08:10 to 10:45:09:00:50) and "Right ascension" (x-axis, 46 to 26). A WCS header is visible: `WCS: (12:01:27.16, -65:08:14.8); Image: (501, 426); Value: 1.49326e-1 Jy/beam*`. The image shows a dark purple field with several bright spots.
- X Profile: Cursor**: A line plot showing the intensity profile along the X-axis. The x-axis is labeled "X coordinate" (0 to 400) and the y-axis is "Value (Jy/beam)" (0.00e+0 to 1.40e-1). A red vertical line indicates the cursor position at approximately X=210. The plot shows a prominent peak at the cursor position. Data: `(WCS: 12:01:27.16, Image: 501 px, 1.49326e-1)`
- Y Profile: Cursor**: A line plot showing the intensity profile along the Y-axis. The x-axis is labeled "Y coordinate" (0 to 400) and the y-axis is "Value (Jy/beam)" (0.00e+0 to 1.40e-1). A red vertical line indicates the cursor position at approximately Y=210. The plot shows a prominent peak at the cursor position. Data: `(WCS: -65:08:14.8, Image: 426 px, 1.49326e-1)`
- Render Configuration**: A panel for adjusting the image display. It includes a "Clip Percentile" dropdown set to "99.9%", a "Scaling" dropdown set to "Linear", and a "Color map" selector. A histogram shows the distribution of values (Jy/beam) from 0 to 0.15, with "Min" and "Max" markers. Other settings include "Invert color map" (unchecked), "Clip Min" (-0.0089254), and "Clip Max" (0.0753930).
- Image List**: A table listing the loaded images and their layers.

Image	Layers	Matchin
0 BHR71_cont_p1.image	R	XY

Desktop Session

Desktop a shell for displaying non-browser container apps.

Apps launched from within but run elsewhere in the cloud.

- all CASA versions
- TOPCAT
- CADC python tools
- Any container a user publishes!

Menu dynamically built from image registry

The screenshot displays a desktop environment with a web browser at the top showing a URL from ws-uv.canfar.net. Below the browser, there are several application windows. On the left, an application menu is open, listing various software categories and specific applications like CASA 3, 4, 5, GEMINI, and find-orb. In the center, a terminal window titled 'IPython: home/majorb (on casa-6-4-0-16-py3-8)' shows the output of a command, including 'Telemetry initialized' and 'CASA 6.4.0.16 -- Common Astronomy Software Applications [6.4.0.16]'. To the right, another terminal window displays a log file with a table of messages.

Priority	Origin	Message
INFO	::casa	Next telemetry data submission
INFO	::casa	
INFO	::casa	optional configuration file
INFO	::casa	
INFO	::casa	Checking Measures tables in
INFO	::casa	IERSeop2000 (version date

Advanced Platform Use

- **Projects have a software expert to build and maintain their containers**

```
> docker build -t images.canfar.net/JWST/canucs:1.1 -f Dockerfile .  
> docker login images.canfar.net  
> docker push images.canfar.net/jwst/canucs:1.1
```

- **Image then labelled with its "type" (notebook, carta, desktop-app, etc...)**
- **Project marked public or group-access only**
- **Becomes available on portal (or desktop) to those authorized**

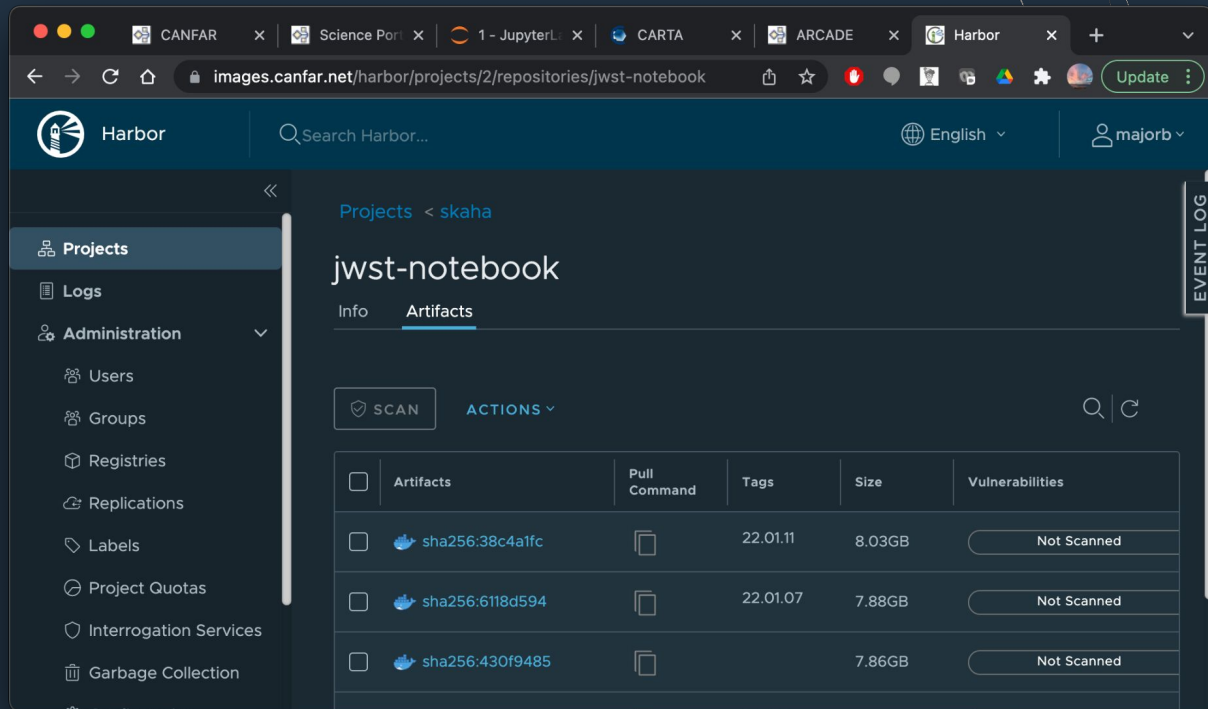
Image Registry

Users can contribute container software (images)

<https://goharbor.io>

Platform launches images in the registry. Protect with groups.

Login with OpenID Connect (CADC IDP)



The screenshot shows the Harbor web interface in a browser. The URL is `images.canfar.net/harbor/projects/2/repositories/jwst-notebook`. The page title is "jwst-notebook" under the "skaha" project. The "Artifacts" tab is selected, displaying a table of container images. The table has columns for Artifacts, Pull Command, Tags, Size, and Vulnerabilities. Three artifacts are listed, all with a "Not Scanned" status.

Artifacts	Pull Command	Tags	Size	Vulnerabilities
<input type="checkbox"/> sha256:38c4a1fc	<input type="checkbox"/>	22.01.11	8.03GB	Not Scanned
<input type="checkbox"/> sha256:6118d594	<input type="checkbox"/>	22.01.07	7.88GB	Not Scanned
<input type="checkbox"/> sha256:430f9485	<input type="checkbox"/>		7.86GB	Not Scanned

Efficient Resource Use in Kubernetes

Problem:

- Users select resource requirements on launch of interactive sessions.
- Resources (CPU, RAM) are reserved for lifetime of session (14 days).
- Often sit idle when user is done.

Ideally...

- Make resource specifications optional for users
- Dynamically modify session allocation reservations up and down based on actual use.

For now:

- Reduce expiry time on sessions so resources are freed earlier
- But, give users a "keep me alive" button that can be used indefinitely

In general:

- Want to hide technical details from users.

SKA SRCNet and Rubin IDAC Integration

Platform group currently a small team

- Need to be smart about planning work
- Aligned development
- Self-serve model and General purpose astronomy features help

Anticipate being an SKA Partner, would lead to a Canadian Science Regional Centre (SRC).

Agreement for Rubin IDAC-lite, and possibly full IDAC (Independent Data Access Centre).

Integration into CADC & CANFAR

- Want to maintain principles of IVOA, Multi-wavelength, common tools.
- A&A, Storage, and Platforms
- At what levels should integration occur?

Challenges ahead. Encourage working and learning together.

CADC and National Infrastructure

A fruitful arrangement...

Digital Research Alliance Canada (DRAC):

- Provide customized kubernetes
- Security locked down (eg no running containers as root)
- Can tear-down and rebuild in minutes
- Hoping to use this for platforms in other disciplines

CADC / CANFAR:

- Build Applications and Services on top
- Also can be rebuilt quickly
- Allows more focus on astronomy related work
- Perhaps handoff image registry and container launcher later

Roadmap - Experimental Github "Project"

The image shows a Jira backlog with three columns: Backlog, Next Up, and In progress. Each item in the backlog is represented by a card with a key, title, priority, and status.

Column	Item Key	Title	Priority	Status
Backlog (4)	science-platform #389	Onboarding tutorials	Medium	help wanted
	science-platform #390	Show current directory size in /arc	Medium	enhancement
	science-platform #291	Provide large data transfer service to CANFAR Science Platform storage	Large	enhancement
	science-platform #392	Group image selection list on portal	Small	
Next Up (1)	science-platform #388	Recursive /arc permission setting from UI	Medium	enhancement
In progress (4)	science-platform #322	feature request: show requested session resources on web GUI	Small	in progress
	science-platform #259	Show cluster health stats on the session launch page	Small	in progress
	science-platform #387	Python parallelization tool	Large	in progress
	science-platform #258	Provide warning when sessions die after 14 days	Small	enhancement

Notes and Discussion

- **Batch is difficult. Three types (from Rubin):**
 - Parallel notebook servers (python + jupyter container)
 - Parallel notebooks (python)
 - Traditional batch (any container)
- **YouCat and large catalogs**
 - requires that clients authenticate
- **Mount archival/caom2 data on containers?**
- **Multi-cloud and bursting**

THANK YOU

Brian Major • CADC • brian.major@nrc-cnrc.gc.ca

