



## Giving your scientific computing environment (SCE) a voice

Experiences and learnings leveraging operational data from our SCE and Posit products to help us serve our users better

*Name, Project Role*

James Black, SCE product owner

Vijita Malhotra, API engineer

Krystian Igras, dashboard engineer

## Problem #1



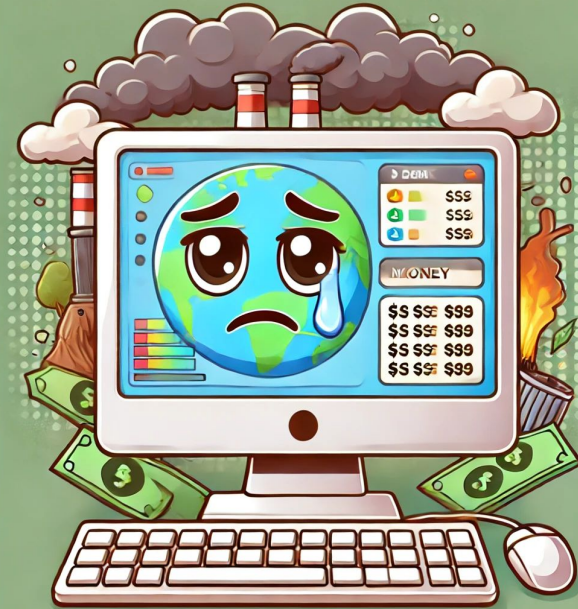
### #1 Building our SCE

We know what users say  
Can we leverage what users do?

### #2 Optimising use

Compute is seen as free  
We have >1,200 users  
Compute is not free

## Problem #2



# What can we ask operational data?



What's the split  
of active  
containers over R  
versions?

---



**Do our  
interactive  
containers  
mostly sit idle?**

---



How many of our  
Connect  
deployments  
failed, or haven't  
been looked at in  
more than a  
year?

---



What teams are  
using our  
platform, and  
what's the split  
on daily, monthly  
and ever use?

---



# Meet Ocean & Apollo



# Meet Ocean and Apollo: 2 UIs, one SCE

Advanced Analytics/RWE  
*Data scientists*



Clinical Trial Analysis  
*Statistical Programmers*



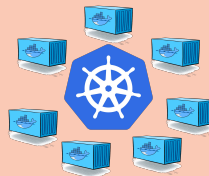
repos



packages



apps/content



interactive



homepod



batch

# Types of data readily accessible

## Containers

Which image  
CPU use  
Creation/Deletion



## Homepod

Space used  
IO?



## Users

Name  
Department  
Manager or Contributor



## Package manager

Downloads



## Git repos

Packages (renv.lock)  
Languages



## Connect

Last accessed  
Times accessed  
Published  
Modified



App info

SCE

- Containers
- Container trends
- Heavy users
- Users
- Connect
- Popular content
- User
- Your Stats

Apollo & Ocean metrics

This app contains metric data collected by ROPYD (Jeeva's team) from the use of Apollo, via the data science module in platform.apollo.roche.com, and workbench.ocean.roche.com.

For information on our roadmap and architecture log, please visit [go.roche.com/workbench](https://go.roche.com/workbench).

To ask a question or provide a comment about this dashboard, please send an email to [service\\_desk+datascience-general-275-apollo-ocean-metrics-277454-issue@code.roche.com](mailto:service_desk+datascience-general-275-apollo-ocean-metrics-277454-issue@code.roche.com).



Available data

Metric	All	Filtered
<b>Workbench</b>		
Running containers	2,002	1,979
Total containers	8,917	8,256
Container users	1,764	1,755
EU Home drives (GB)	8,165	8,126
US Home drives (GB)	2,690	2,673
<b>Connect</b>		
Ever Connect users	2,782	2,255
Connect developers	620	620
Connect apps	5,013	5,013

Available data range



Departments

All users

Which timespan would like to view

- All data
- Last 30 days
- Last 6 months
- Last 12 months

SCE specific filters

Which compute namespace(s):

- Both
- Ocean
- Apollo

Should we include the team building the platform?

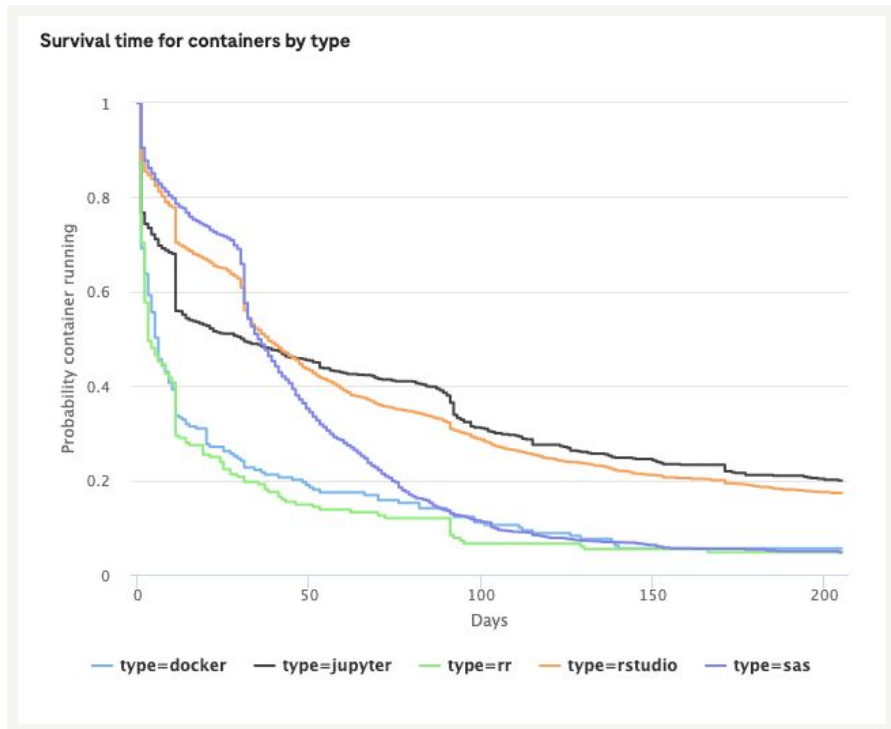
- Exclude
- Include

# What can we ask operational data?



What's the split of active containers over R versions?

Tag	Containers	People
R4.3.3-16apr2024	364	292
R4.3.1-11oct2023	293	247
R4.2.2-19may2023	51	46
r4.2.1-21oct2022	29	26
r4.1.3-27apr2022	24	24
R4.1.3-08apr2022	19	18
R4.2.2-21feb2023	13	12
r4.1.1-rstudio1.4-30aug20	6	3
r4.2.1-21oct2022	5	3
r4.0.4-rstudio1.4-09apr20	4	4
r4.1.0-rstudio1.4-06jul20	4	4
r4.1.1-rstudio1.4-30aug20	4	4
r4.1.2-rstudio382-12jan20	4	4
r4.0.4-rstudio1.4-09apr20	4	4
R4.2.1-17oct2022	2	1
r4.1.0-rstudio1.4-06jul20	2	2
r4.1.3-27apr2022	2	2
r4.0.0-rstudio1.3	1	1
r4.1.2-rstudio382-12jan20	1	1

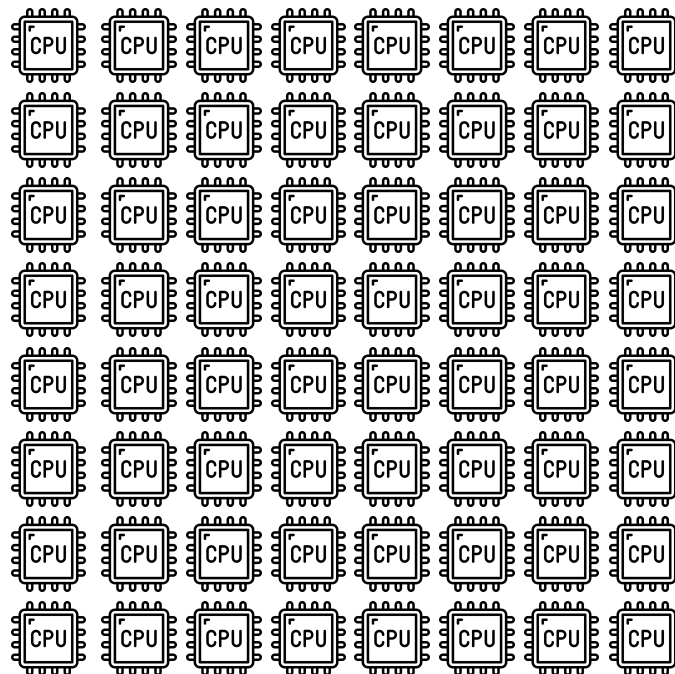
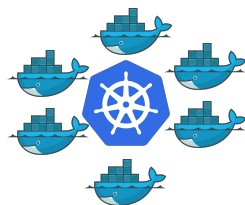




# What can we ask operational data?



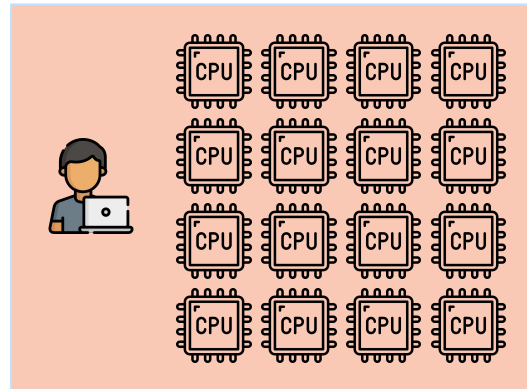
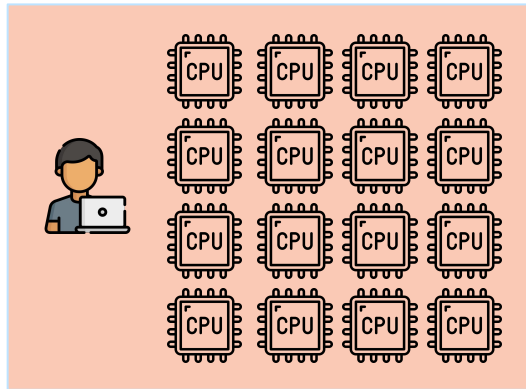
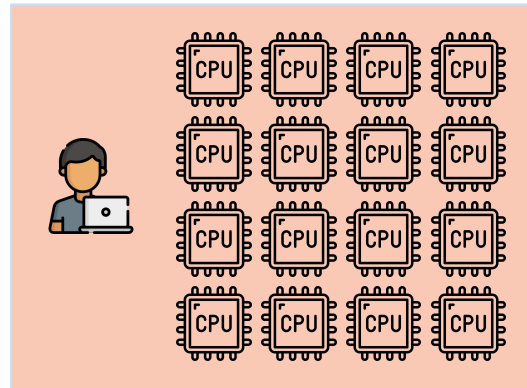
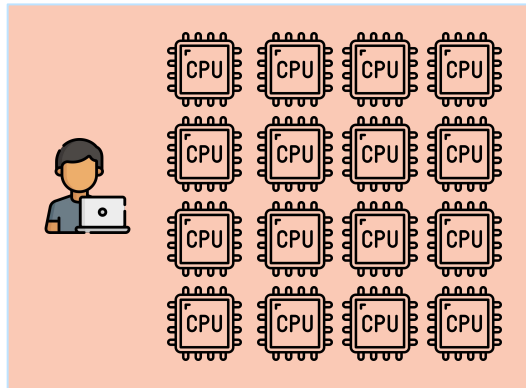
Do our interactive containers mostly sit idle?



# What can we ask operational data?



Do our interactive containers mostly sit idle?

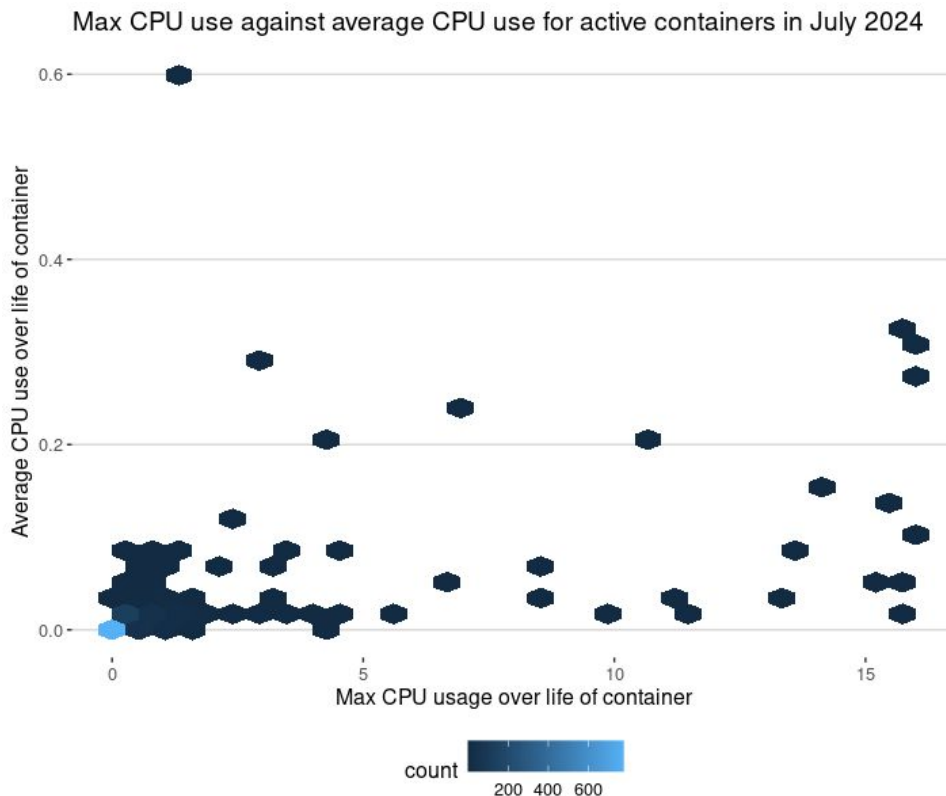


~\$5 a day per container

# What can we ask operational data?



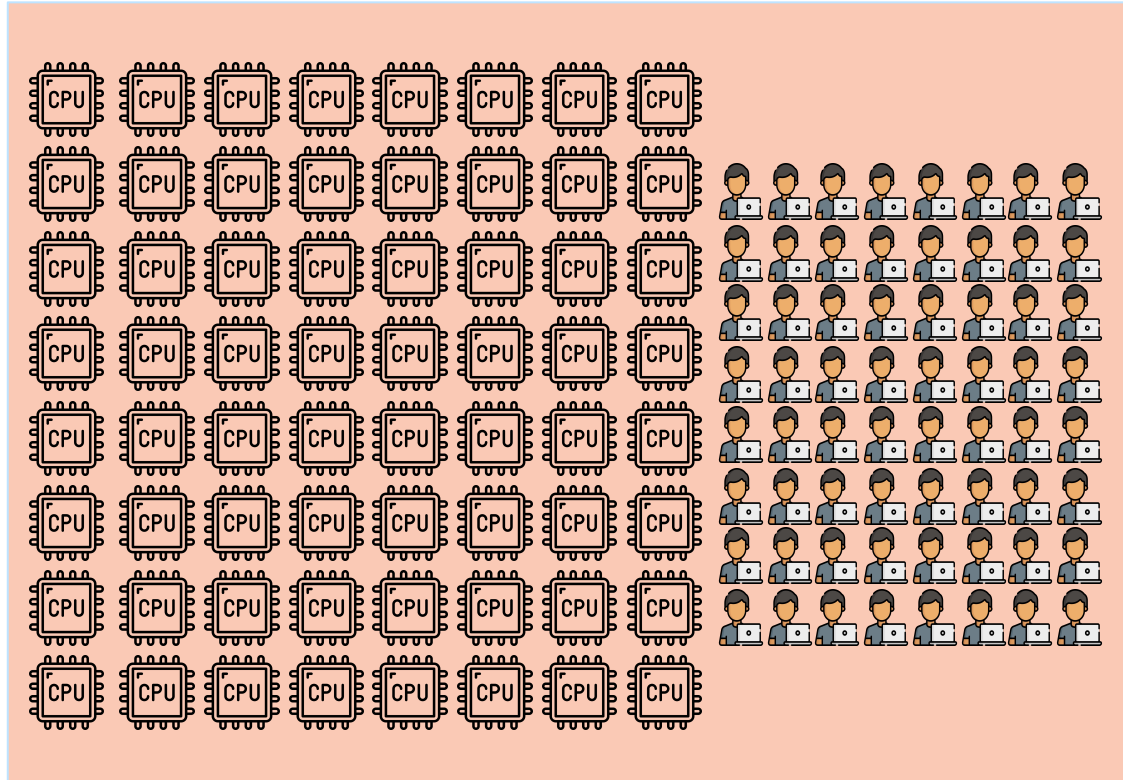
Do our interactive containers mostly sit idle?



# What can we ask operational data?



Do our interactive containers mostly sit idle?



~~\$5 a day per container~~ → ~30c a day per container

# What can we ask operational data?



How many of our Connect deployments failed, or haven't been looked at in more than a year?

---

707

Items had **never been successfully deployed**



Automate deletion if not fixed in 30 days

502

Items had **not been viewed in the last year**



Can we delete?

Inform users

# What can we ask operational data?



What teams are using our platform, and what's the split on daily, monthly and ever use?

## Frequency of interactive container use by department

Subset of departments manually selected

Department	Daily	Weekly	Monthly	Less
Affiliates	27%	7%	27%	40%
Access	11%	26%	26%	37%
Data & Statistical Sciences	13%	25%	40%	22%
Safety	33%	22%	33%	11%
PHC imaging	19%	29%	14%	38%
US Medical Affairs	19%	22%	36%	22%

All container log data from 2024-01-01 to 2024-08-08 used. Frequency calculated as days active / days total where CPU load suggested person active in a container. Frequency categories are Daily >0.5, Weekly >0.23, Monthly >0.04.

 **Can we build a feedback loop?**

The icon depicts a blue speech bubble with a white outline, positioned inside a circular loop. The loop is formed by two curved arrows: a red one at the top and a yellow one at the bottom, both pointing clockwise.

# Communicating back to users

**2,971**

Total CPU hours in July by a representative 'simulation running' statistician

**\$~62**

Total cost, based on 2c a CPU hour\*

**~0.13kg CO<sub>2</sub>**

Carbon footprint for this users CPU use

**9,000kg CO<sub>2</sub>**

Carbon footprint of my flight to posit::conf



*\*Not how we are actually billed, but illustrative  
Example methodology for CO<sub>2</sub>e calculations [aether.green/docs/methodologies/](https://aether.green/docs/methodologies/)*



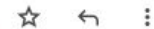
[SCE bot] Hey James, I've got some details on your SCE use

Inbox x Misc x



**SCE bot** <scebot@apollo.roche.com>  
to me

18:54 (1 minute ago)



Hello James!

You have 3 interactive containers that sit idle 93% of the time. Idle containers continue to reserve a slot on our workernodes, by decreasing from 3 to 1 container would free up 2 slots on our cluster, potentially saving **£1000** a month.

You have 75 items deployed to Connect. You have 2 items that have never been successfully deployed and 2 items that have not been viewed in 1 year. The following items are tagged for deletion if they remain un-accessed

Stale Content

<a href="#">Apollo and Ocean SCE metrics</a>	2023-12-27
<a href="#">Apollo &amp; Ocean metrics</a>	2023-12-25

Defective Content

<a href="#">flatiron-edm</a>	2020-05-07
<a href="#">PHC-23-DRAFT</a>	2019-10-18



**Where are now?**

# Where are we now?



Our dashboard provides transparent metrics on already existing operational data

---



We can **objectively answer many questions we have about SCE use**

---



Soon we'll feed back to users how their usage could be modified to lower costs

---



Many of these data sources are common across companies!

---

**Doing now what patients need next**