Matching Tools to Titans

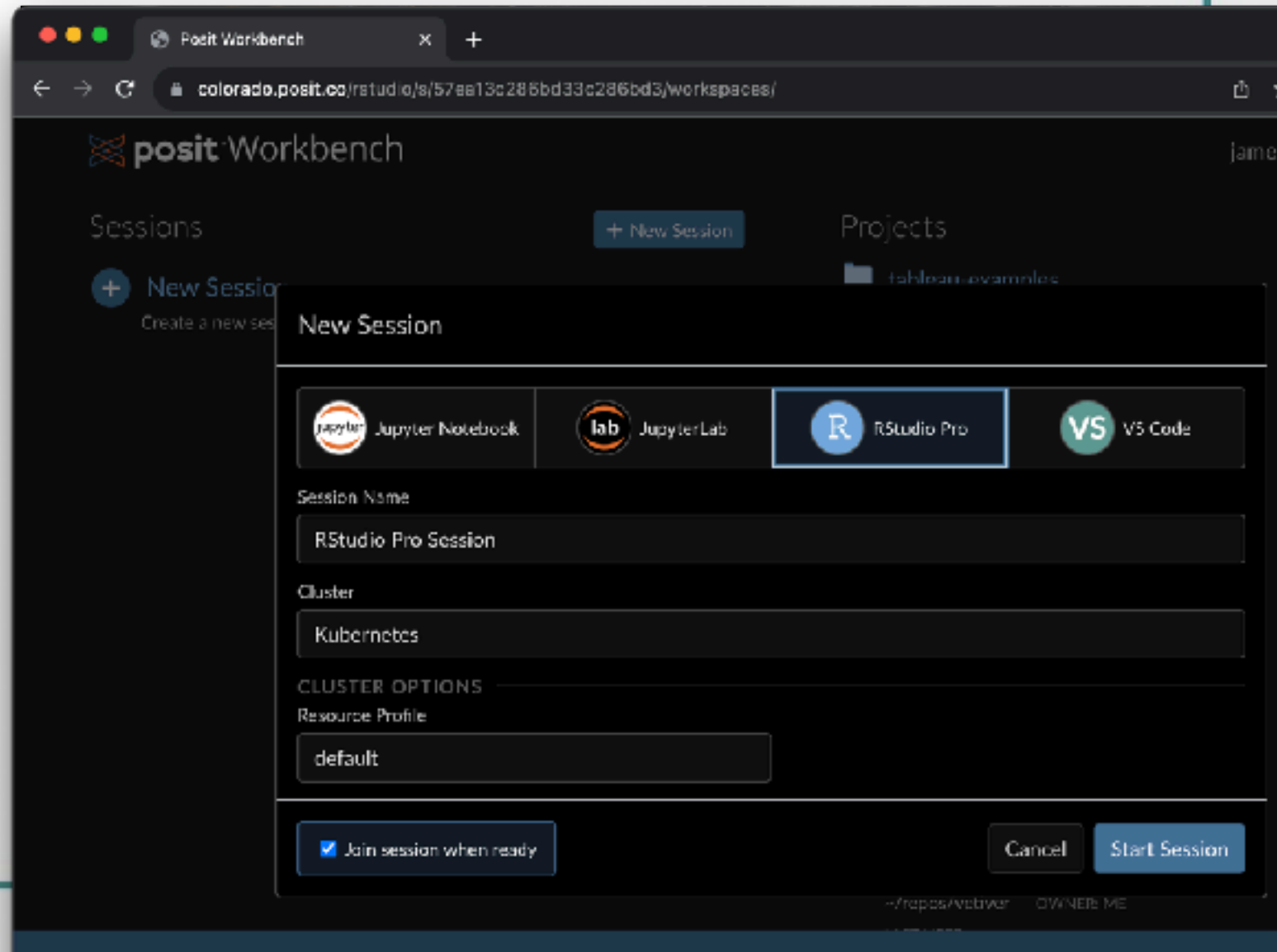
Tailoring Posit Workbench for Every Cloud

James Blair
Product Manager
Cloud Integrations
Posit, PBC

FROM posit

posit conf(2023)

At Posit, we work hard to ensure that Posit Workbench meets the needs of the modern Data Science developer



Developers don’t always see the supporting infrastructure

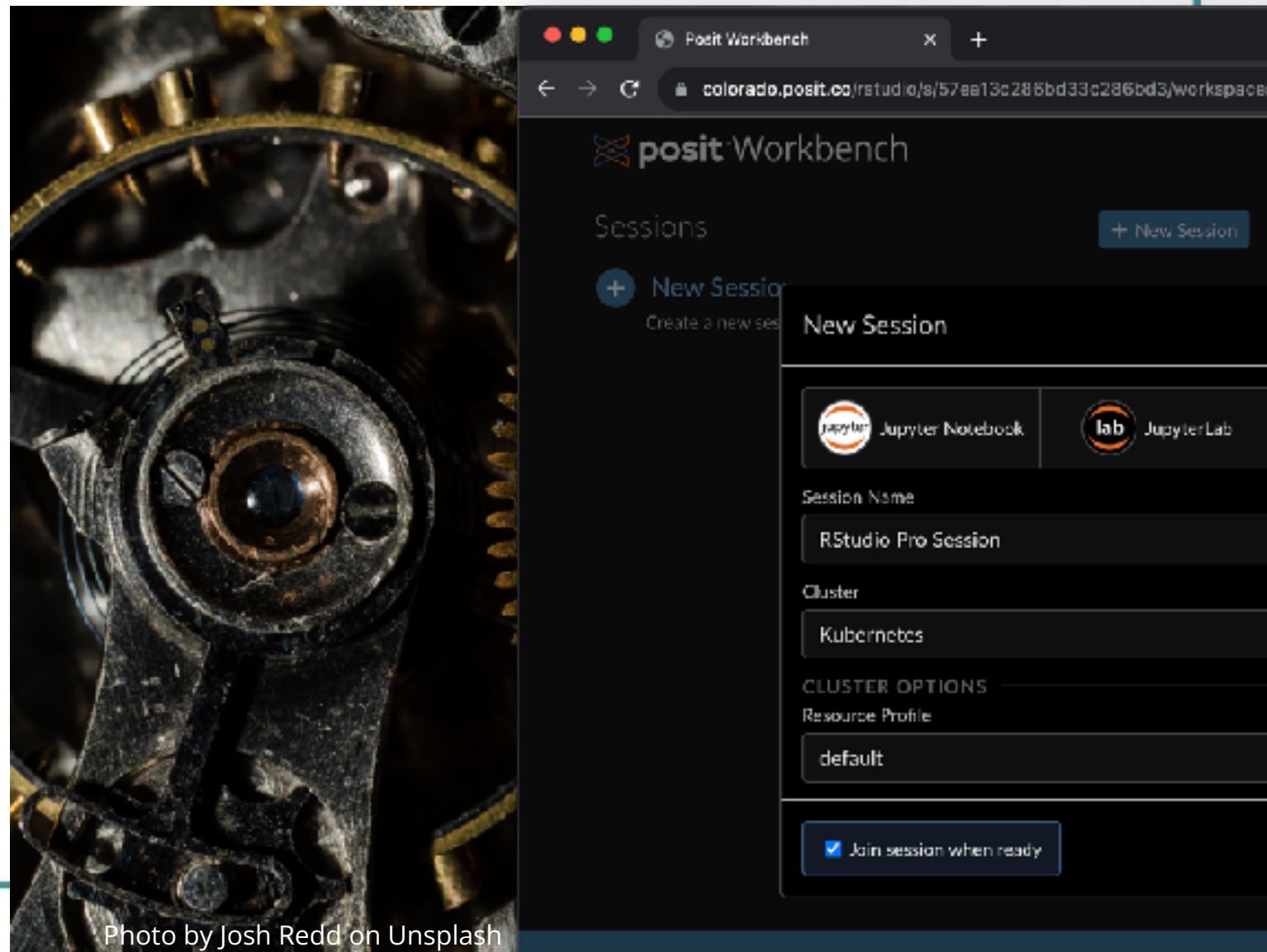


Photo by Josh Redd on Unsplash

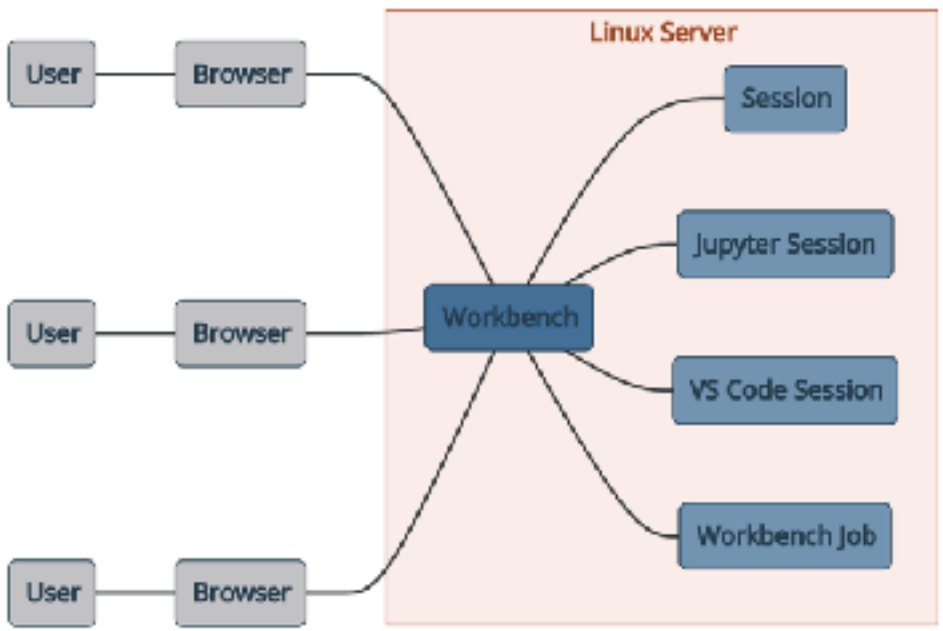
Guide
Write Effective Code
Connect to Data Sources and Systems
Manage Packages and Reproduce Environments
Secure Access
Implement Operational Patterns
Choose the Right Architecture
Configuring Python within Posit products
Running Posit Products in Containers
Using AWS managed File Systems with Workbench
Architectures
Connect Architectures
Posit Team Architectures
Package Manager Architectures
Workbench Architectures
Launcher
Sagemaker

Workbench Architectures

Using Workbench on a single server

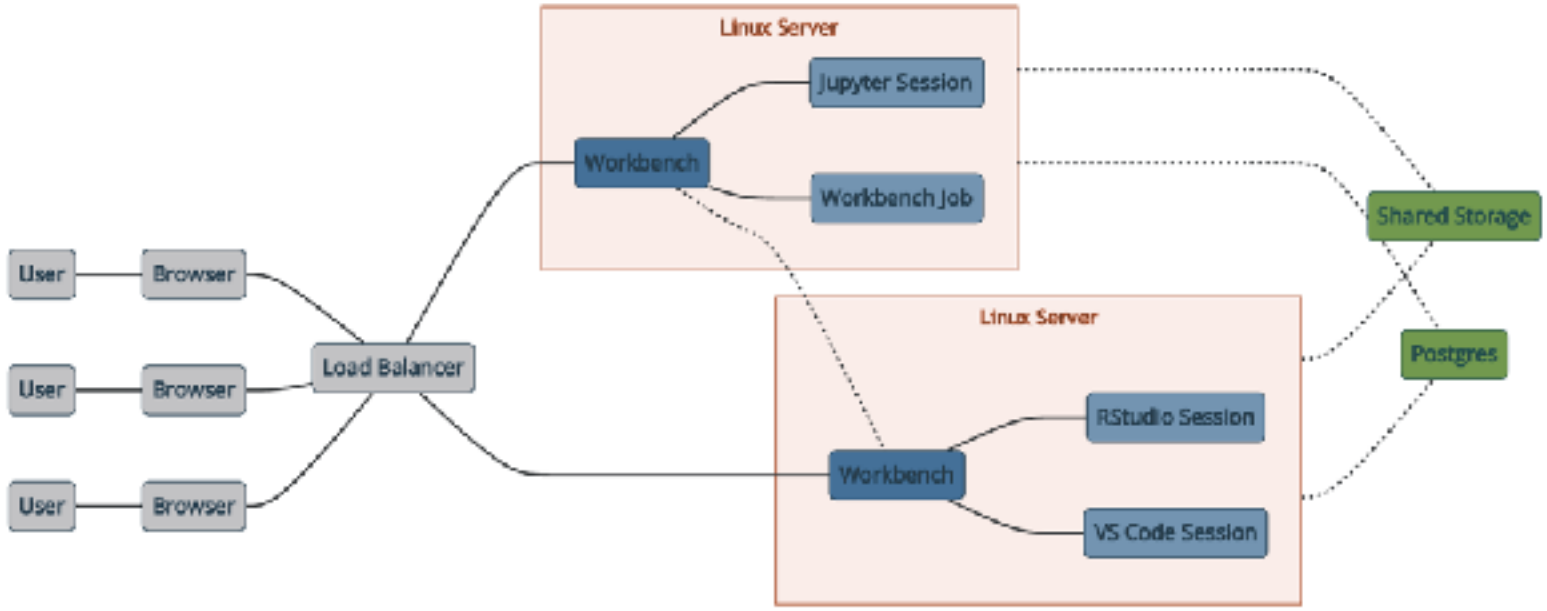
In this configuration, Workbench is installed on a single Linux server and enables:

- Access to RStudio, Jupyter Notebook, JupyterLab and VS Code development IDEs
- Multiple concurrent sessions per user
- Use of multiple versions of R and Python

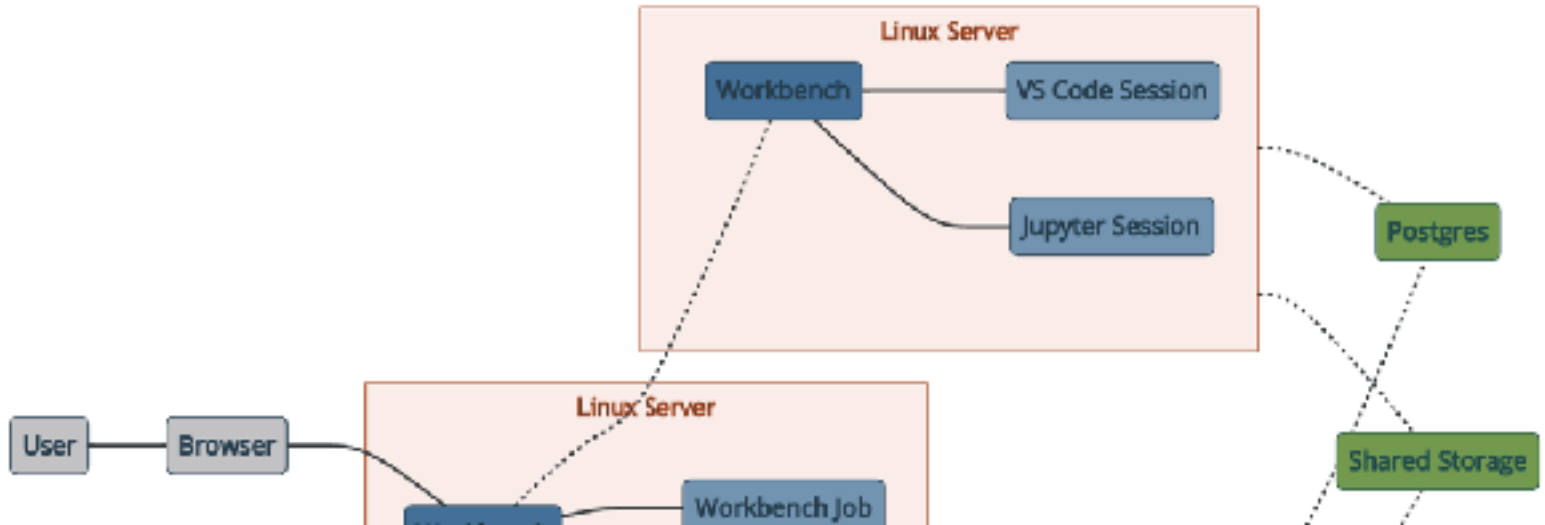


On this page
Using Workbench on a single server
Using Workbench as a cluster
Using Workbench with an external resource manager
Additional Resources

External Load Balancer

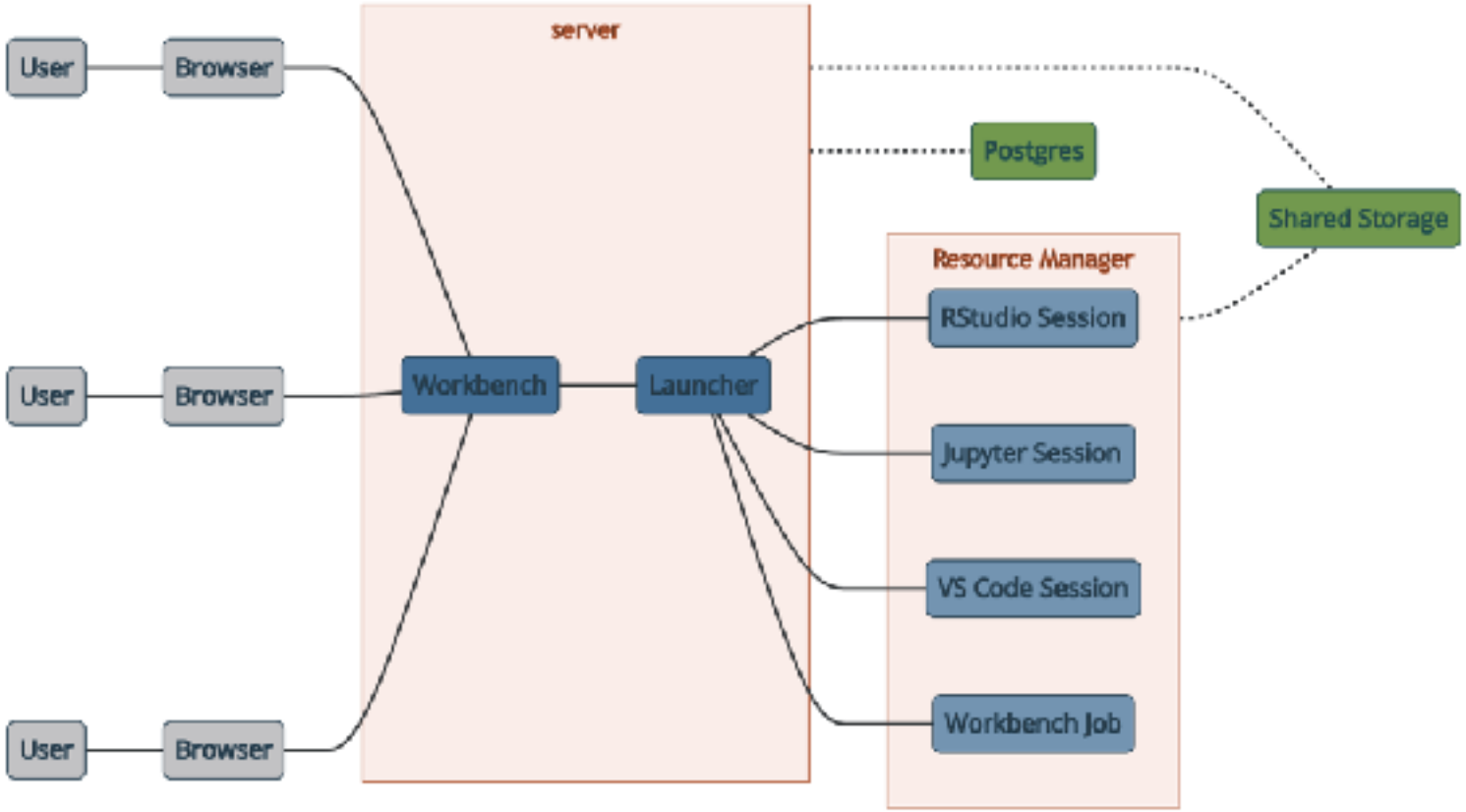


Single Node Routing



Requirements to support this architecture:

- Users' home directories must be stored on an external shared file server (typically an NFS server)
- It is strongly recommended that session metadata be stored on an external PostgreSQL database server



Using Workbench entirely in Kubernetes

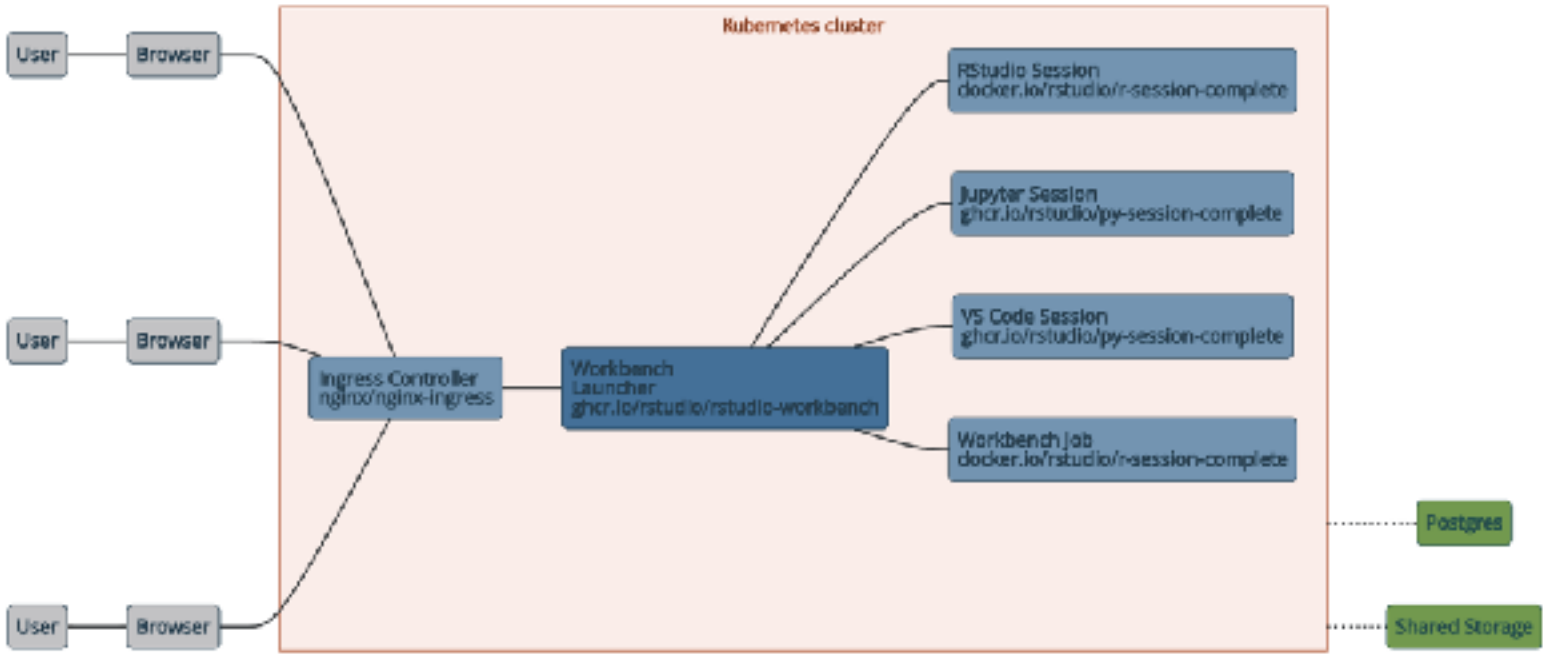
In this configuration, Workbench is installed entirely inside a Kubernetes cluster and enables:

- User sessions and jobs run in isolated pods, potentially from different base images
- The entire installation is managed in Kubernetes with tools like Helm
- Optional replicas for high availability
- Access to RStudio, Jupyter Notebook, JupyterLab and VS Code development IDEs
- Multiple concurrent sessions per user
- Use of multiple versions of R and Python

- Multiple concurrent sessions per user
- Use of multiple versions of R and Python

Requirements to support this architecture:

- Users' home directories must be stored on an external shared file server (typically an NFS server)
- Session metadata must be stored on an external PostgreSQL database server

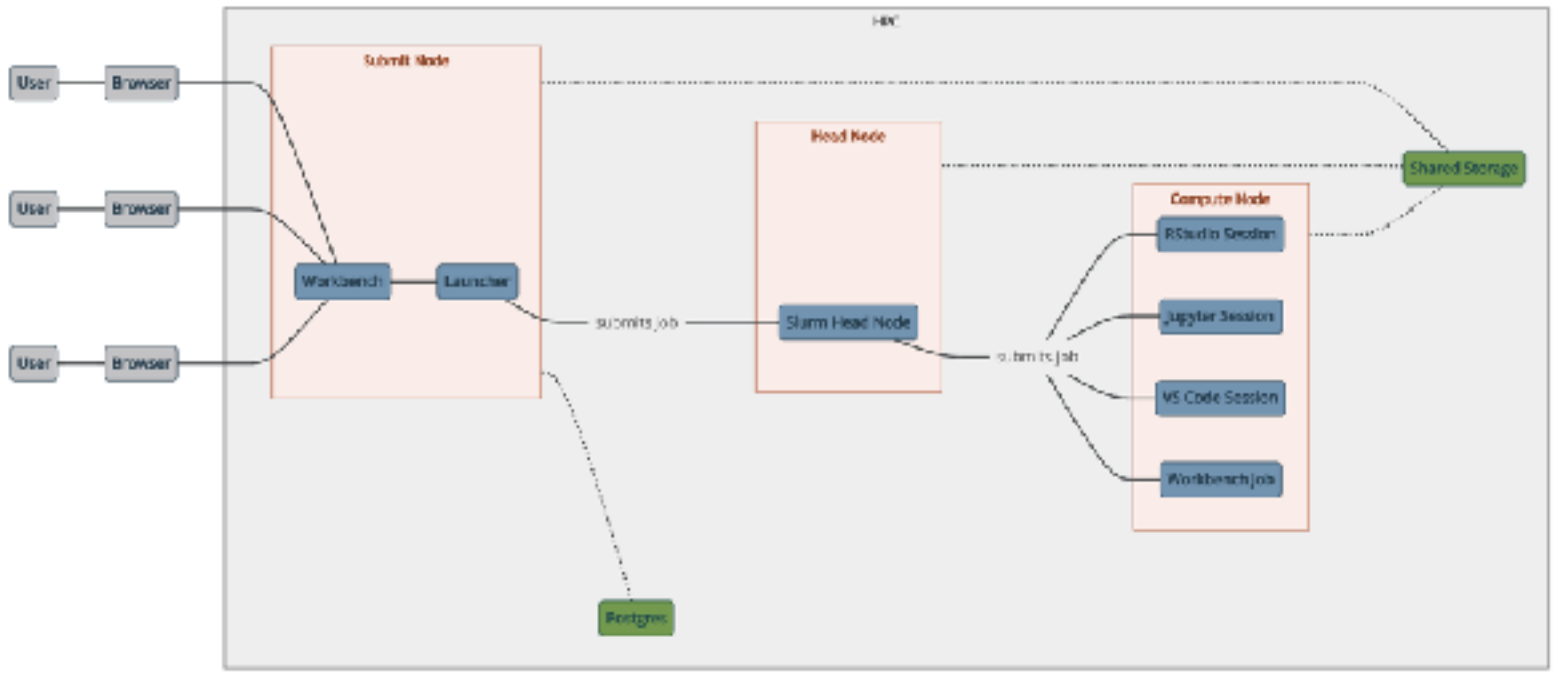


Additional Kubernetes Resources

Ready to get started with Workbench and Kubernetes? View the documentation for Integrating Workbench with Kubernetes

Want to use custom Docker images with Kubernetes? View the guide for Using Docker images with Workbench, Launcher, and Kubernetes

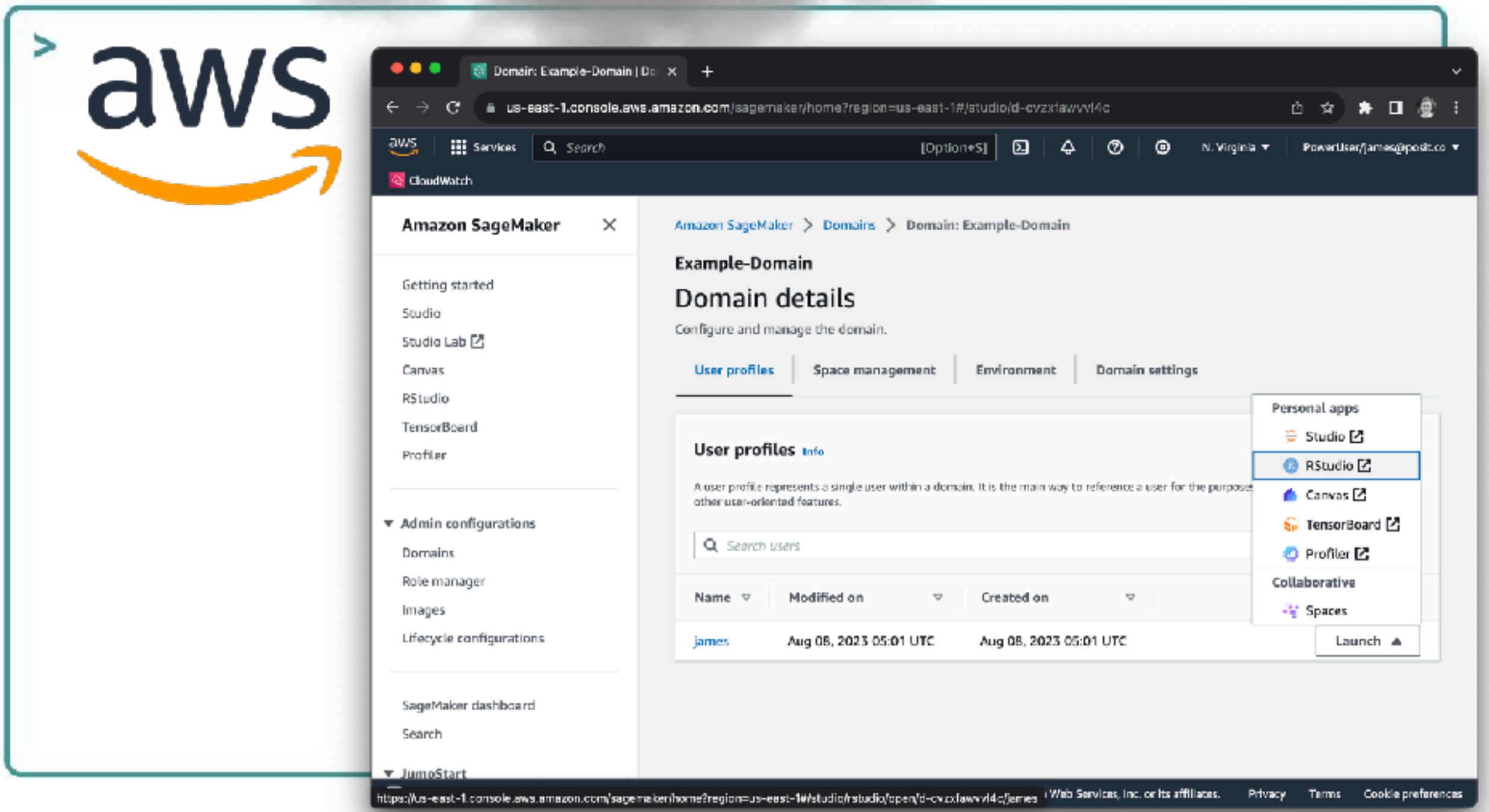
- Users must exist on both Workbench servers and the Slurm cluster node, for example by pointing to the same authentication provider
- The use of an external PostgreSQL database server is necessary when using multiple Workbench servers

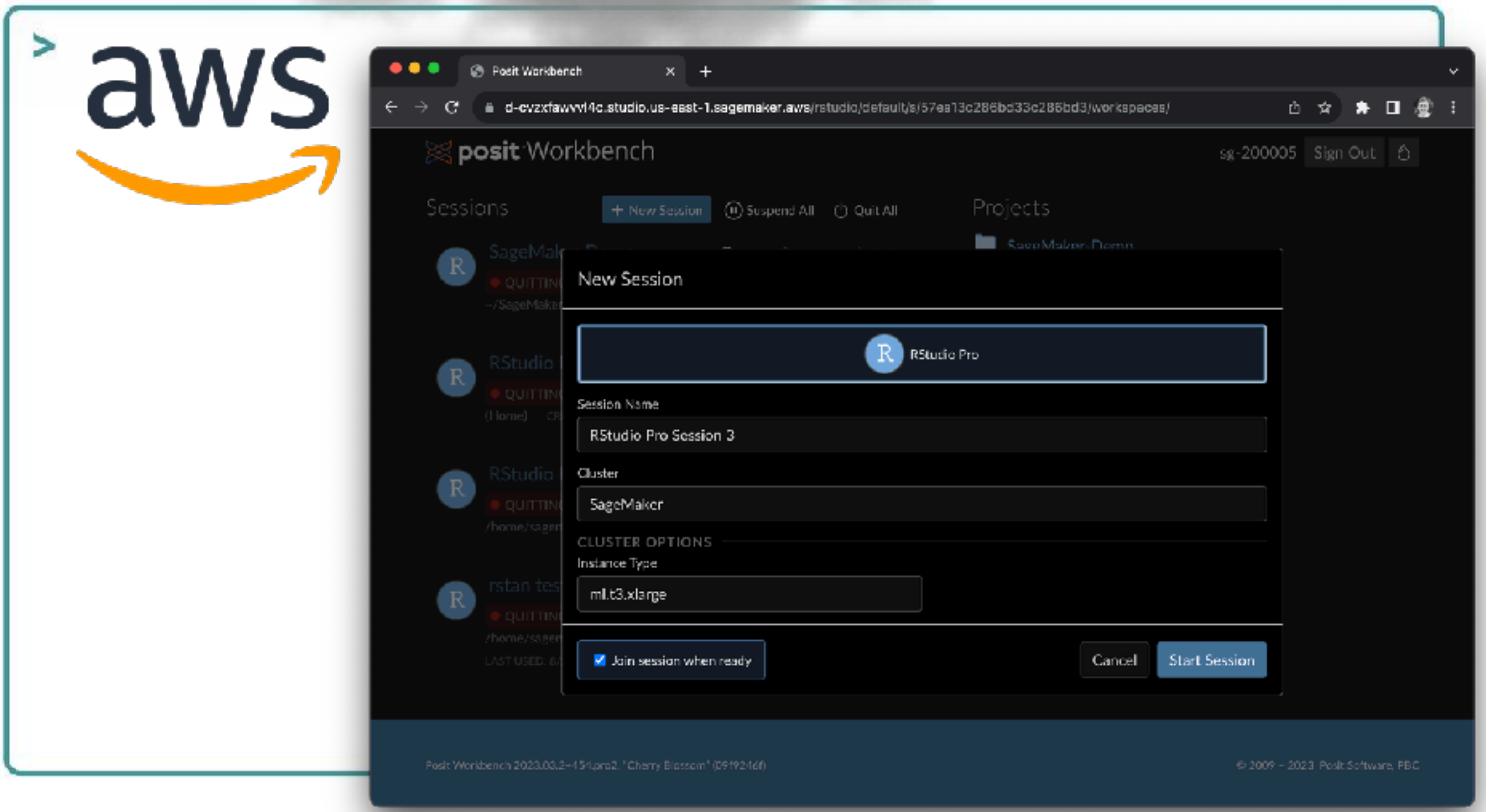


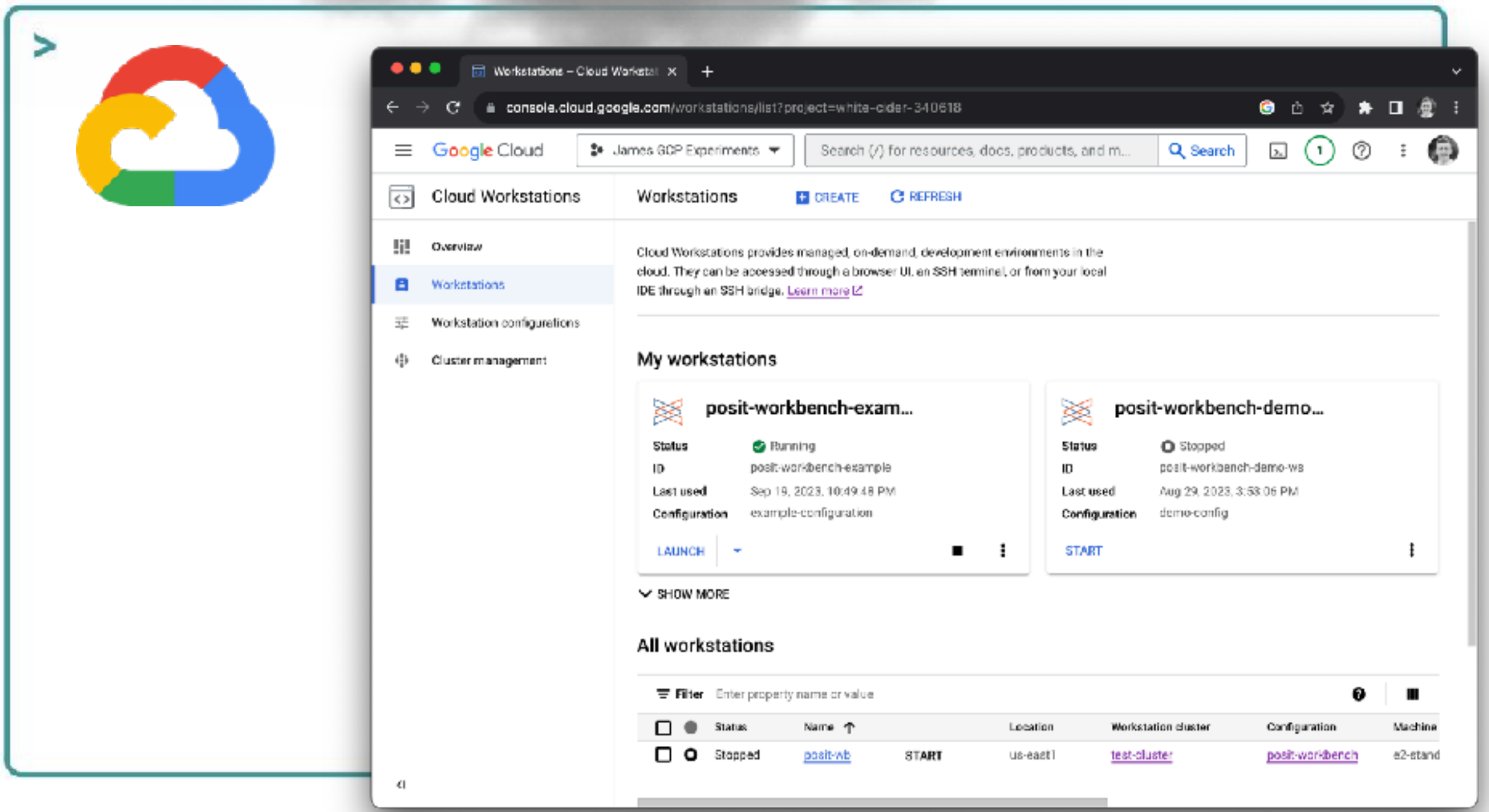
Additional Slurm Resources

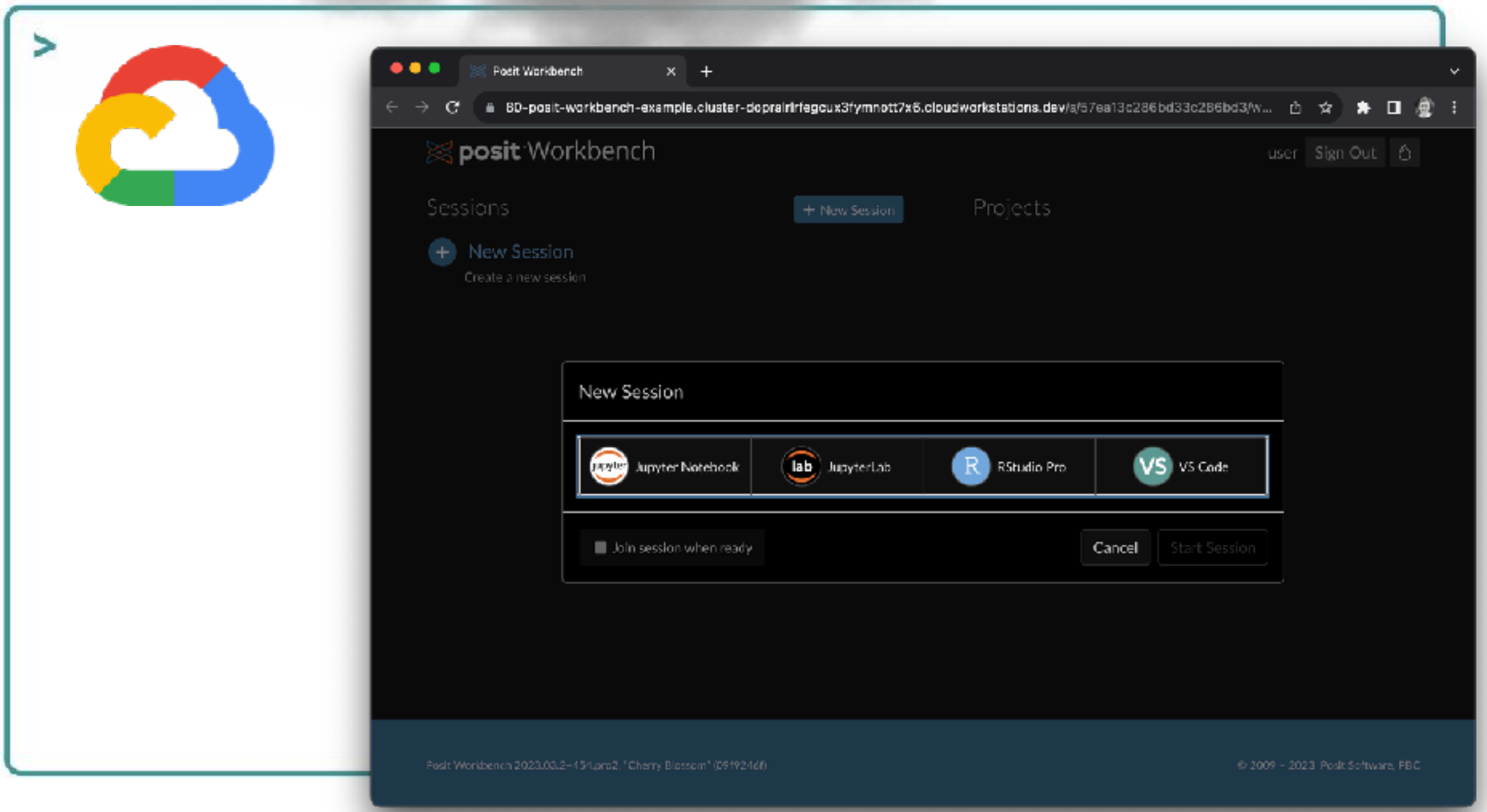
Ready to get started with Workbench and Slurm? Proceed to the documentation for Integrating

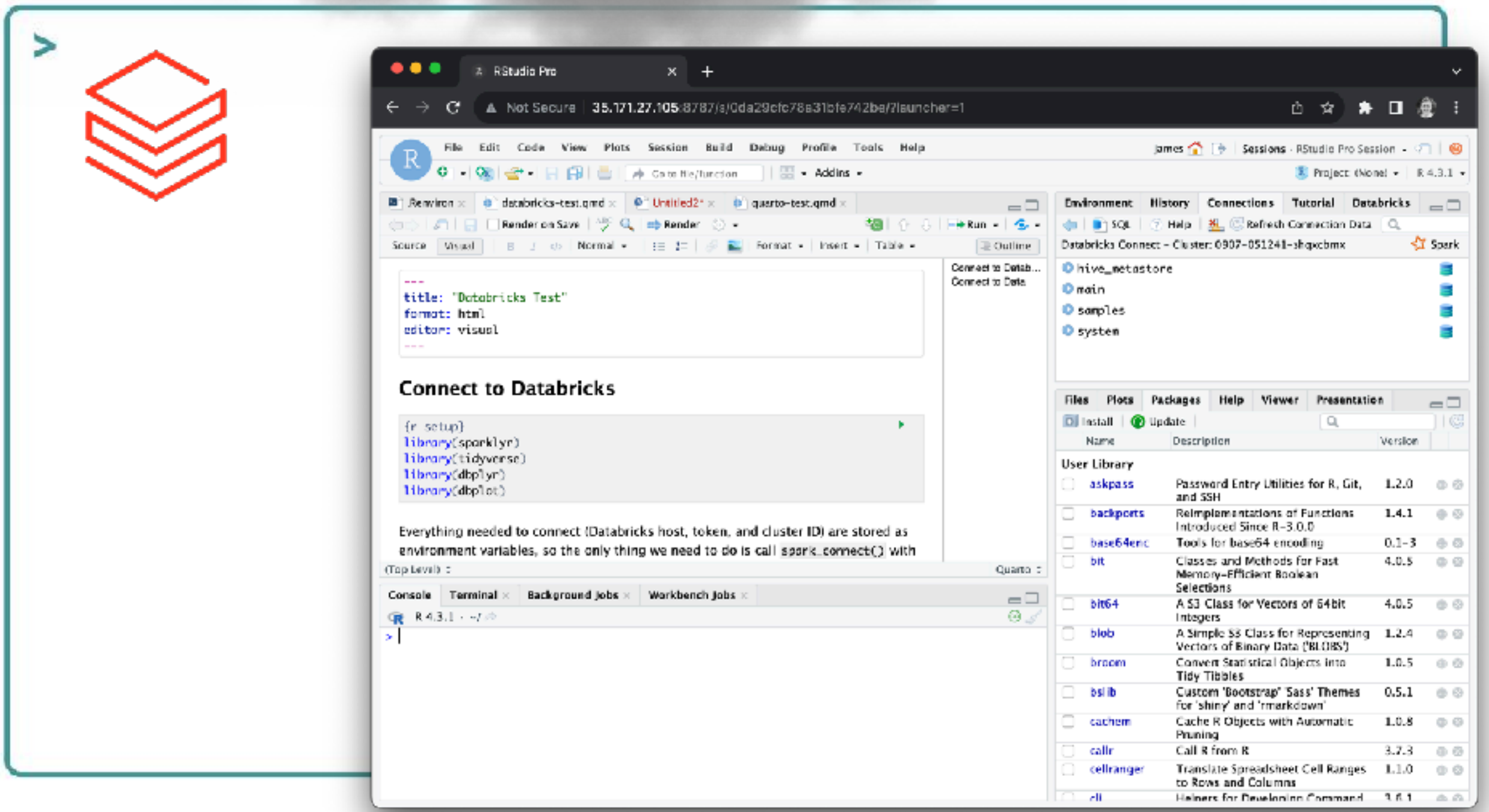

aws


Domain: Example-Domain | Do
us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/studio/d-cvzxfawvvl4c
Services
Search
[Option+S]
N. Virginia
CloudWatch
Amazon SageMaker
Getting started
Studio
Studio Lab
Canvas
RStudio
TensorBoard
Profiler
Admin configurations
Domains
Role manager
Images
Lifecycle configurations
SageMaker dashboard
Search
JumpStart
Amazon SageMaker > Domains > Domain: Example-Domain
Example-Domain
Domain details
Configure and manage the domain.
User profiles
Space management
Environment
Domain settings
User profiles Info
Search users
Name
Modified on
Created on
james
Aug 08, 2023 05:01 UTC
Aug 08, 2023 05:01 UTC
Launch
Personal apps
Studio
RStudio
Canvas
TensorBoard
Profiler
Collaborative
Spaces
Web Services, Inc. or its affiliates.
Privacy
Terms
Cookie preferences

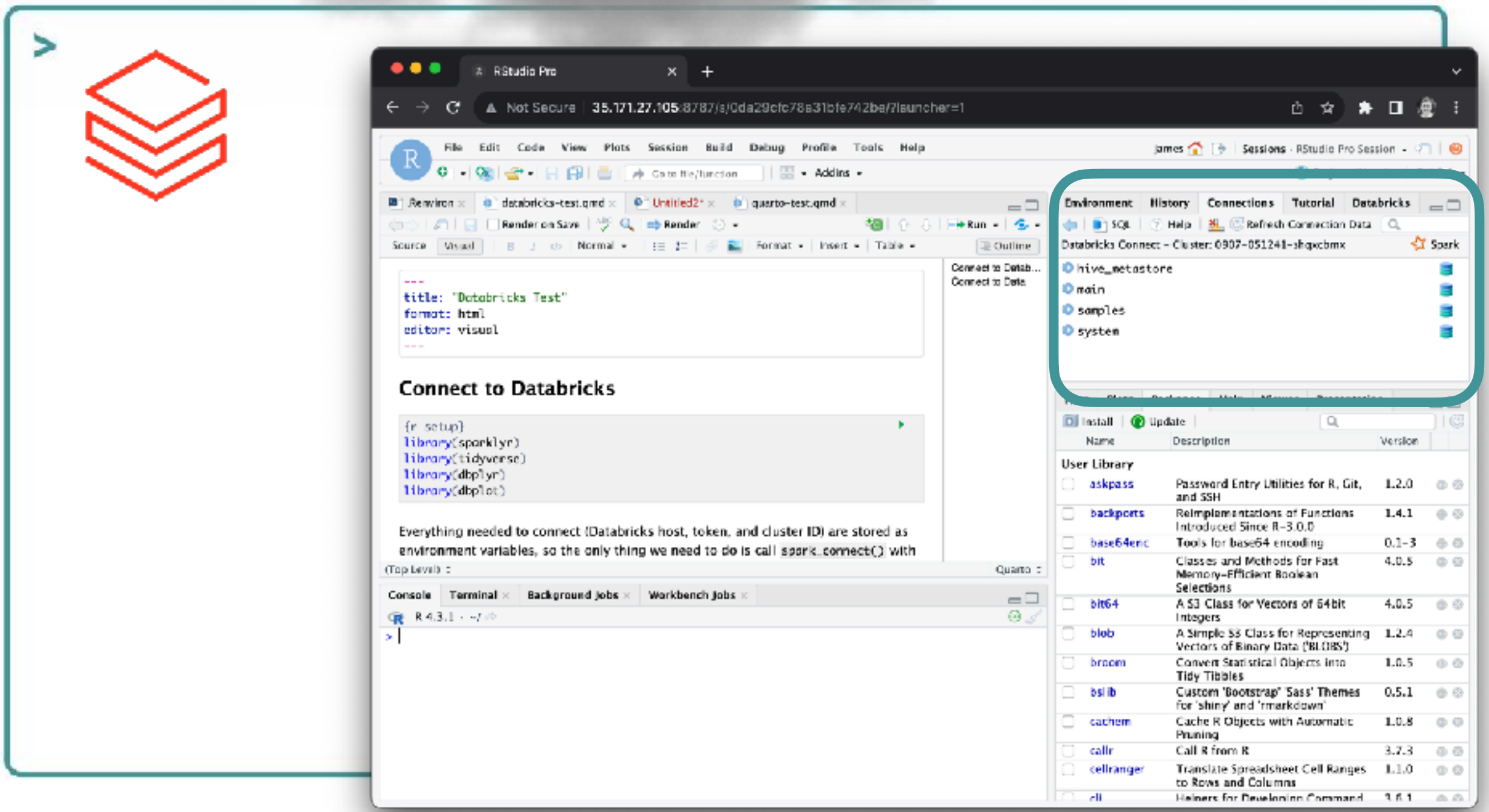

Posit Workbench
d-cvzxfawvvl4c.studio.us-east-1.sagemaker.aws/rstudio/default/s/57ea13c286bd33c286bd3/workspaces/
posit Workbench
sg-200005
Sign Out
Sessions
+ New Session
Suspend All
Quit All
Projects
New Session
RStudio Pro
Session Name
RStudio Pro Session 3
Cluster
SageMaker
CLUSTER OPTIONS
Instance Type
ml.t3.xlarge
Join session when ready
Cancel
Start Session


Workstations – Cloud Workstat
console.cloud.google.com/workstations/list?project=white-cider-340618
Google Cloud
James GCP Experiments
Search (/) for resources, docs, products, and m...
Search
Cloud Workstations
Workstations
CREATE
REFRESH
Overview
Workstations
Workstation configurations
Cluster management
Cloud Workstations provides managed, on-demand, development environments in the cloud. They can be accessed through a browser UI, an SSH terminal, or from your local IDE through an SSH bridge. Learn more
My workstations
posit-workbench-exam...
Status
Running
ID
posit-workbench-example
Last used
Sep 19, 2023, 10:49:48 PM
Configuration
example-configuration
LAUNCH
posit-workbench-demo...
Status
Stopped
ID
posit-workbench-demo-ws
Last used
Aug 29, 2023, 3:58:06 PM
Configuration
demo-config
START
SHOW MORE
All workstations
Filter
Enter property name or value
Status
Name
Location
Workstation cluster
Configuration
Machine
Stopped
posit-ab
START
us-east1
test-cluster
posit-workbench
e2-stand

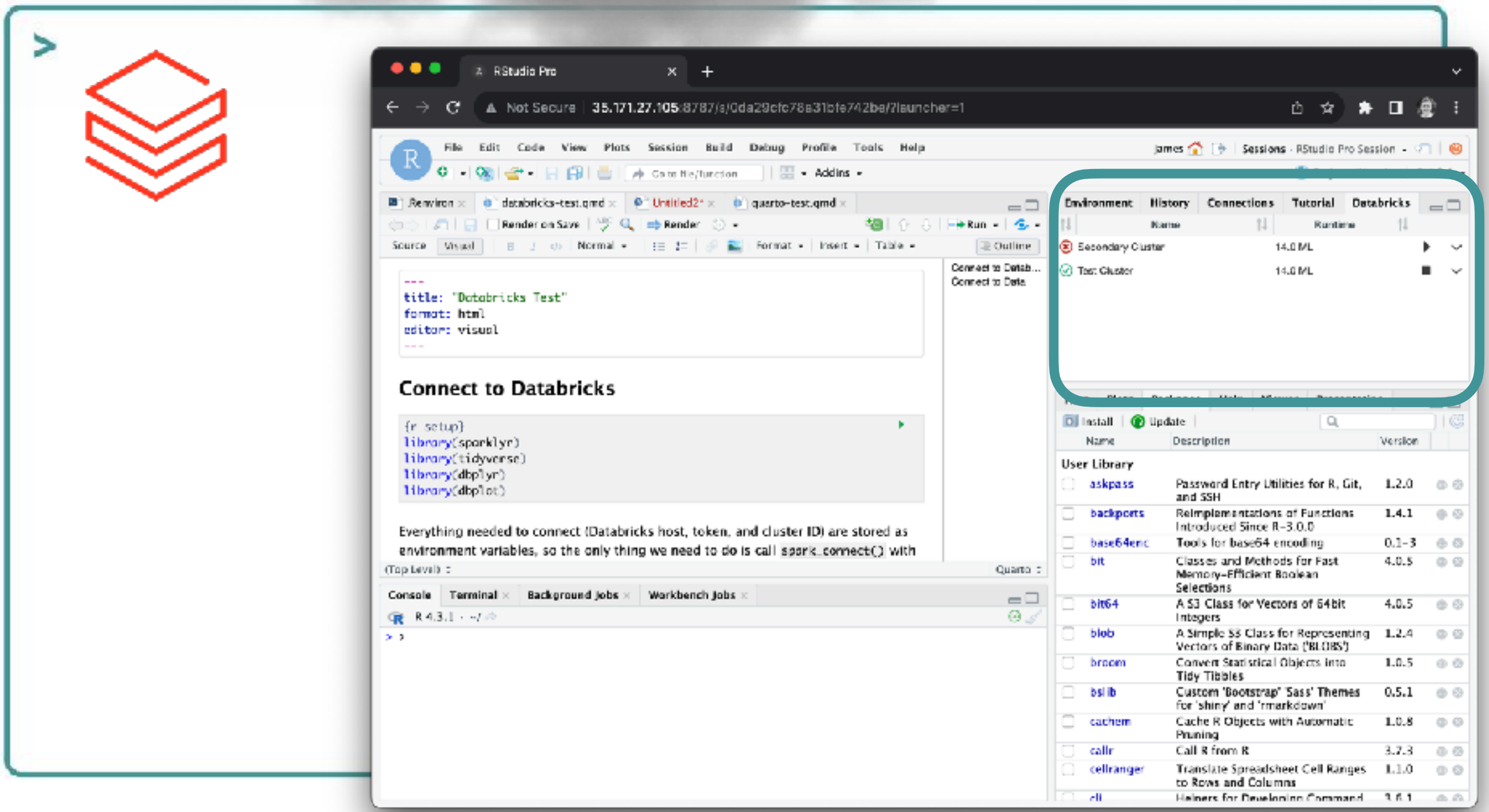

Posit Workbench
posit Workbench
user
Sign Out
Sessions
+ New Session
Projects
New Session
Create a new session
New Session
Jupyter Notebook
JupyterLab
RStudio Pro
VS Code
Join session when ready
Cancel
Start Session


RStudio Pro
Not Secure
35.171.27.105:8787/s/0da29cfc78a31bfe742be/?launcher=1
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Rewrion
databricks-test.qmd
Untitled2*
quarto-test.qmd
Render on Save
Render
Run
Source
Visual
Normal
Format
Insert
Table
Outline
---
title: "Databricks Test"
format: html
editor: visual

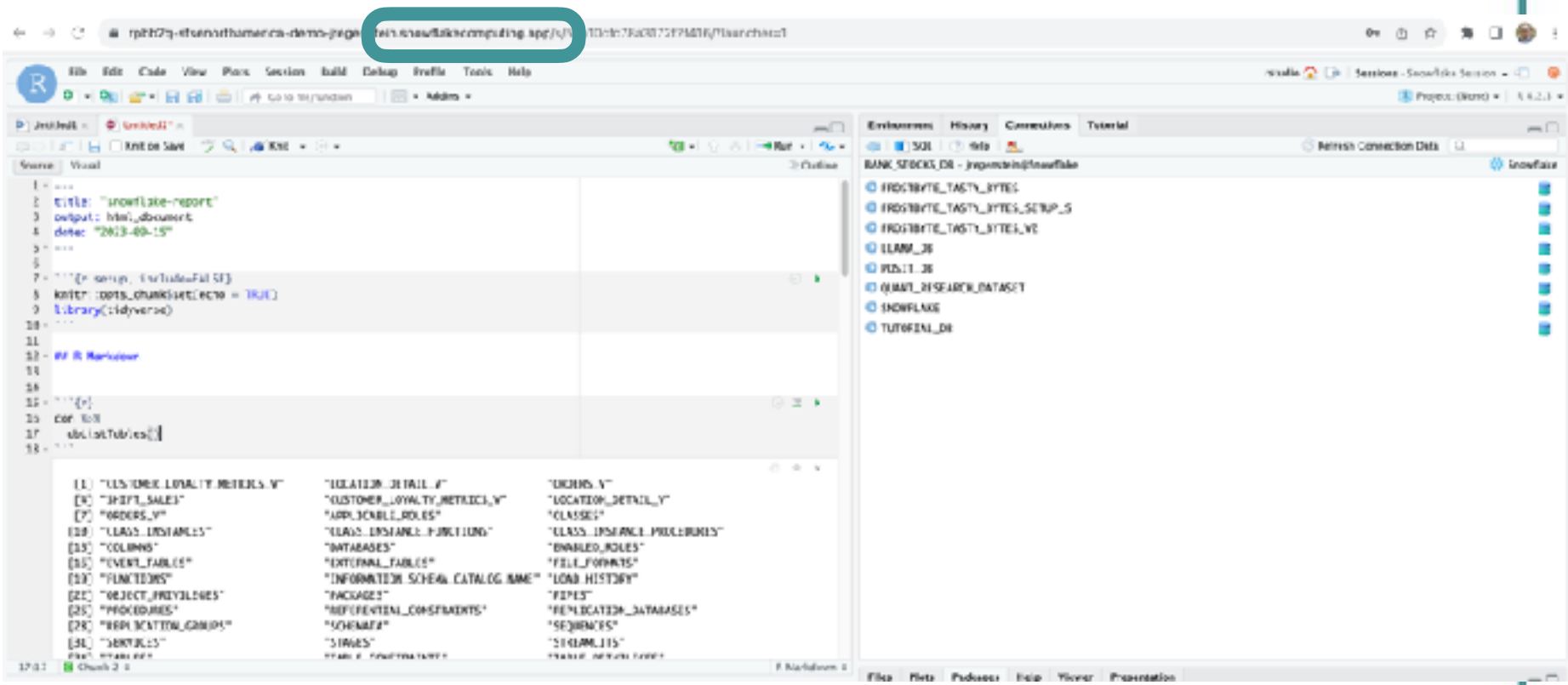
Connect to Databricks
{r setup}
library(sparklyr)
library(tidyverse)
library(dbplyr)
library(dbplot)
Everything needed to connect (Databricks host, token, and cluster ID) are stored as environment variables, so the only thing we need to do is call spark_connect() with
(Top Level)
Quarto
Console
Terminal
Background Jobs
Workbench Jobs
R 4.3.1 · ~/
>
Environment
History
Connections
Tutorial
Databricks
SQL
Help
Refresh Connection Data
Databricks Connect - Cluster: 0907-051241-shqxcbmx
Spark
hive_metastore
main
samples
system
Files
Plots
Packages
Help
Viewer
Presentation
Install
Update
Name
Description
Version
User Library
askpass
Password Entry Utilities for R, Git, and SSH
1.2.0
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.4.1
base64enc
Tools for base64 encoding
0.1-3
bit
Classes and Methods for Fast Memory-Efficient Boolean Selections
4.0.5
bit64
A S3 Class for Vectors of 64bit Integers
4.0.5
blob
A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')
1.2.4
broom
Convert Statistical Objects into Tidy Tibbles
1.0.5
bslib
Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'
0.5.1
cachem
Cache R Objects with Automatic Pruning
1.0.8
callr
Call R from R
3.7.3
cellranger
Translate Spreadsheet Cell Ranges to Rows and Columns
1.1.0


RStudio Pro
Not Secure
35.171.27.105:8787/s/0da29cfc78a31bfe742be/?launcher=1
File Edit Code View Plots Session Build Debug Profile Tools Help
james
Sessions - RStudio Pro Session
Addins
Rewrion
databricks-test.qmd
Untitled2*
quarto-test.qmd
Render on Save
Render
Run
Source
Visual
Normal
Format
Insert
Table
Outline
---
title: "Databricks Test"
format: html
editor: visual

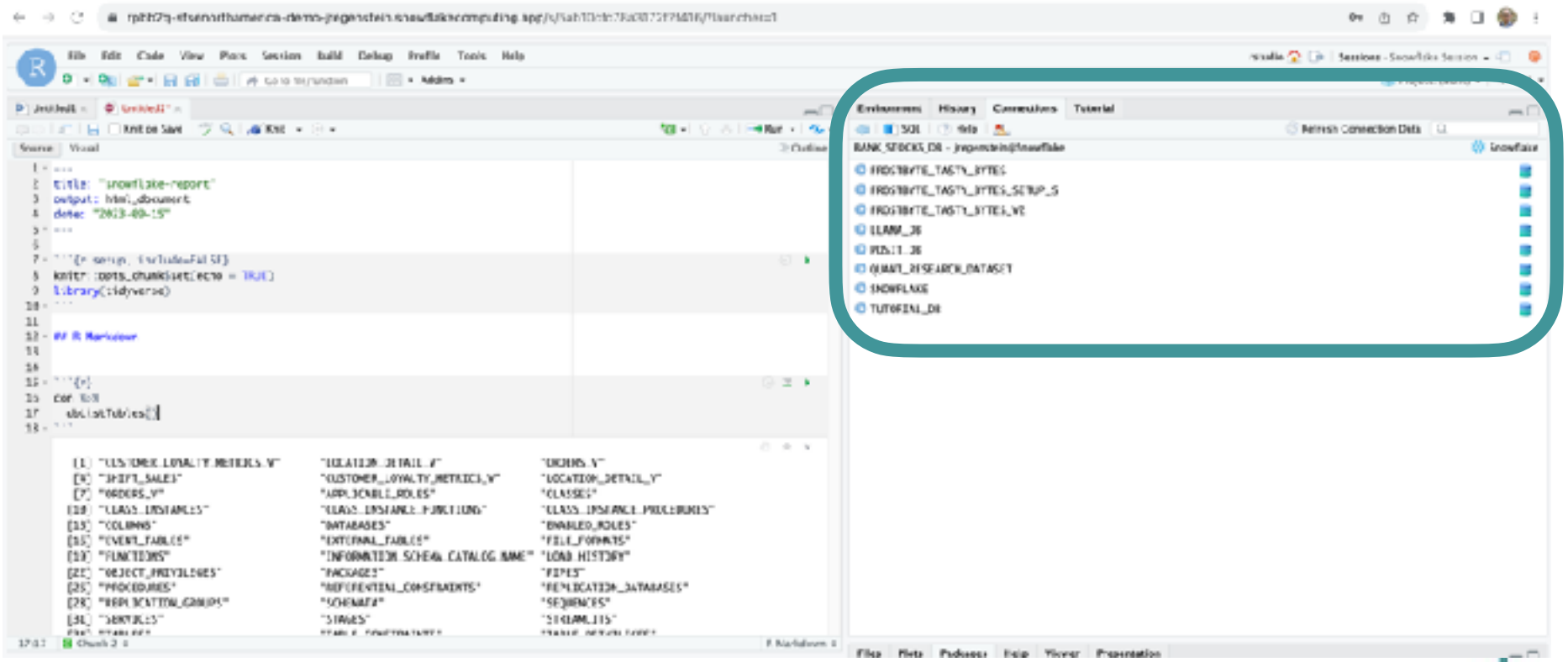
Connect to Databricks
{r setup}
library(sparklyr)
library(tidyverse)
library(dbplyr)
library(dbplot)
Everything needed to connect (Databricks host, token, and cluster ID) are stored as
environment variables, so the only thing we need to do is call spark_connect() with
(Top Level)
Quarto
Console
Terminal
Background Jobs
Workbench Jobs
R 4.3.1 · ~/
Environment
History
Connections
Tutorial
Databricks
SQL
Help
Refresh Connection Data
Databricks Connect - Cluster: 0907-051241-shqxcbmx
Spark
hive_metastore
main
samples
system
Install
Update
Name
Description
Version
User Library
askpass
Password Entry Utilities for R, Git, and SSH
1.2.0
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.4.1
base64enc
Tools for base64 encoding
0.1-3
bit
Classes and Methods for Fast Memory-Efficient Boolean Selections
4.0.5
bit64
A S3 Class for Vectors of 64bit Integers
4.0.5
blob
A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')
1.2.4
broom
Convert Statistical Objects into Tidy Tibbles
1.0.5
bslib
Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'
0.5.1
cachem
Cache R Objects with Automatic Pruning
1.0.8
callr
Call R from R
3.7.3
cellranger
Translate Spreadsheet Cell Ranges to Rows and Columns
1.1.0

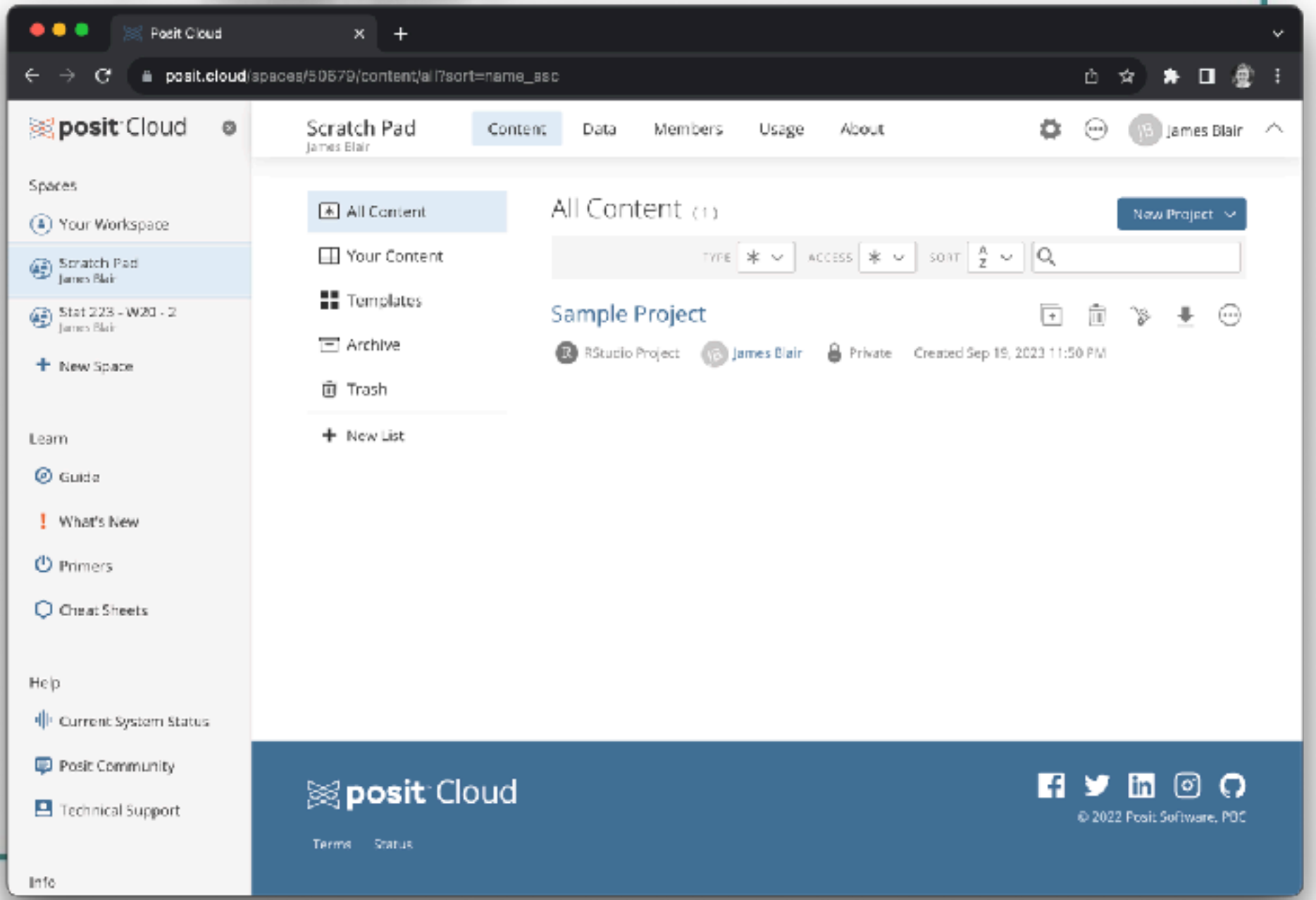

RStudio Pro
Not Secure
35.171.27.105:8787/s/0da29cfc78a31bfe742ba/?launcher=1
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
databricks-test.qmd
Untitled2*
quarto-test.qmd
Render on Save
Render
Run
Source
Visual
Normal
Format
Insert
Table
Outline
---
title: "Databricks Test"
format: html
editor: visual

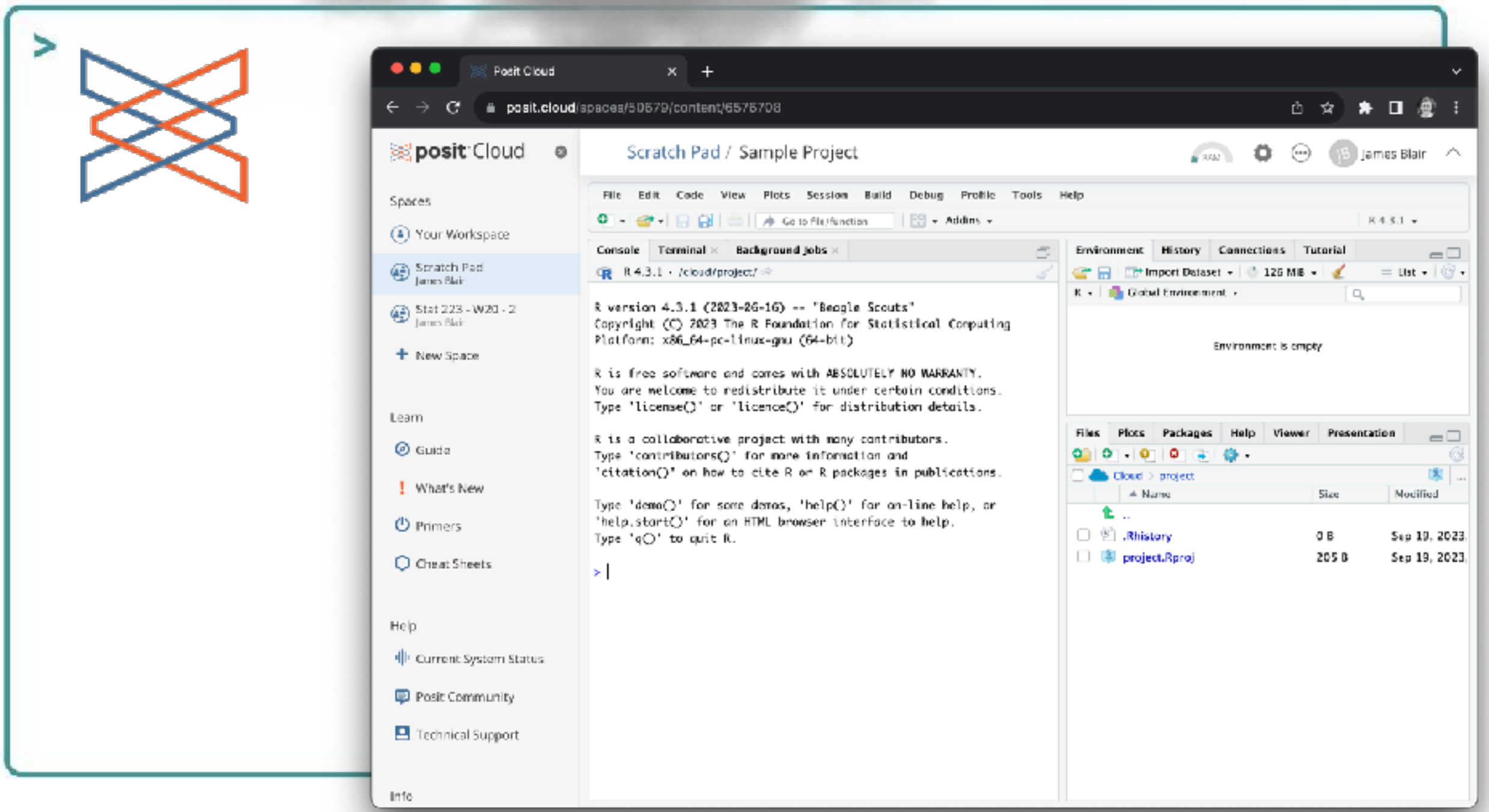
Connect to Databricks
{r setup}
library(sparklyr)
library(tidyverse)
library(dbplyr)
library(dbplot)
Everything needed to connect (Databricks host, token, and cluster ID) are stored as environment variables, so the only thing we need to do is call spark_connect() with
Connect to Datab...
Connect to Data
(Top Level)
Quarto
Console
Terminal
Background Jobs
Workbench Jobs
R 4.3.1 · ~/
> >
Environment
History
Connections
Tutorial
Databricks
Name
Runtime
Secondary Cluster
14.0 ML
Test Cluster
14.0 ML
Install
Update
Name
Description
Version
User Library
askpass
Password Entry Utilities for R, Git, and SSH
1.2.0
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.4.1
base64enc
Tools for base64 encoding
0.1-3
bit
Classes and Methods for Fast Memory-Efficient Boolean Selections
4.0.5
bit64
A S3 Class for Vectors of 64bit Integers
4.0.5
blob
A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')
1.2.4
broom
Convert Statistical Objects into Tidy Tibbles
1.0.5
bslib
Custom 'Bootstrap' 'Sass' Themes for 'shiny' and 'rmarkdown'
0.5.1
cachem
Cache R Objects with Automatic Pruning
1.0.8
callr
Call R from R
3.7.3
cellranger
Translate Spreadsheet Cell Ranges to Rows and Columns
1.1.0


File Edit Code View Plots Session Build Debug Profile Tools Help
Environment History Connections Tutorial
Knit on Save
Source Visual
---
title: "snowflake-report"
output: html_document
date: "2023-09-15"

```{r setup, includeFalse}
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse)
```
R Markdown
```{r}
con %>%
dbListTables()
```
[1] "CUSTOMER_LOYALTY_METRICS_V" "LOCATION_DETAIL_V" "ORDERS_V"
[4] "SHIFT_SALES" "CUSTOMER_LOYALTY_METRICS_V" "LOCATION_DETAIL_V"
[7] "ORDERS_V" "APPLICABLE_ROLES" "CLASSES"
[10] "CLASS_INSTANCES" "CLASS_INSTANCE_FUNCTIONS" "CLASS_INSTANCE_PROCEDURES"
[13] "COLUMNS" "DATABASES" "ENABLED_ROLES"
[16] "EVENT_TABLES" "EXTERNAL_TABLES" "FILE_FORMATS"
[19] "FUNCTIONS" "INFORMATION_SCHEMA_CATALOG_NAME" "LOAD_HISTORY"
[22] "OBJECT_PRIVILEGES" "PACKAGES" "PIPES"
[25] "PROCEDURES" "REFERENTIAL_CONSTRAINTS" "REPLICATION_DATABASES"
[28] "REPLICATION_GROUPS" "SCHEMATA" "SEQUENCES"
[31] "SERVICES" "STAGES" "STREAMLITS"


Environment
History
Connections
Tutorial
BANK_STOCKS_DB - jregenstein@snowflake
SNOWFLAKE
TUTORIAL_DB


Posit Cloud
posit.cloud/spaces/50679/content/all?sort=name_asc
posit Cloud
Spaces
Your Workspace
Scratch Pad
James Blair
Stat 223 - W20 - 2
James Blair
New Space
Learn
Guide
What's New
Primers
Cheat Sheets
Help
Current System Status
Posit Community
Technical Support
Info
Scratch Pad
James Blair
Content
Data
Members
Usage
About
James Blair
All Content
Your Content
Templates
Archive
Trash
New List
All Content (1)
New Project
TYPE
ACCESS
SORT
Sample Project
RStudio Project
James Blair
Private
Created Sep 16, 2023 11:50 PM
posit Cloud
Terms
Status
© 2022 Posit Software, PBC


Posit Cloud
posit.cloud/spaces/50679/content/6575708
posit Cloud
Spaces
Your Workspace
Scratch Pad
James Blair
Stat 223 - W20 - 2
James Blair
New Space
Learn
Guide
What's New
Primers
Cheat Sheets
Help
Current System Status
Posit Community
Technical Support
Info
Scratch Pad / Sample Project
James Blair
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Console Terminal Background Jobs
R 4.3.1 · /cloud/project/
R version 4.3.1 (2023-06-16) -- "Beagle Scouts"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
Environment History Connections Tutorial
Import Dataset
125 MiB
List
Global Environment
Environment is empty
Files Plots Packages Help Viewer Presentation
Cloud > project
Name
Size
Modified
.Rhistory
0 B
Sep 19, 2023,
project.Rproj
205 B
Sep 19, 2023,


aws
posit™ Team

If you didn’t see your preferred cloud solution here, please reach out:
james@posit.co

Resources

- AWS
- GCP
- Databricks
- Snowflake (coming soon)
- Posit.cloud
- Slides

posit
conf(2023)

Thank you.