

Matching Tools to Titans

Tailoring Posit Workbench for Every Cloud



James Blair
Product Manager
Cloud Integrations
Posit, PBC

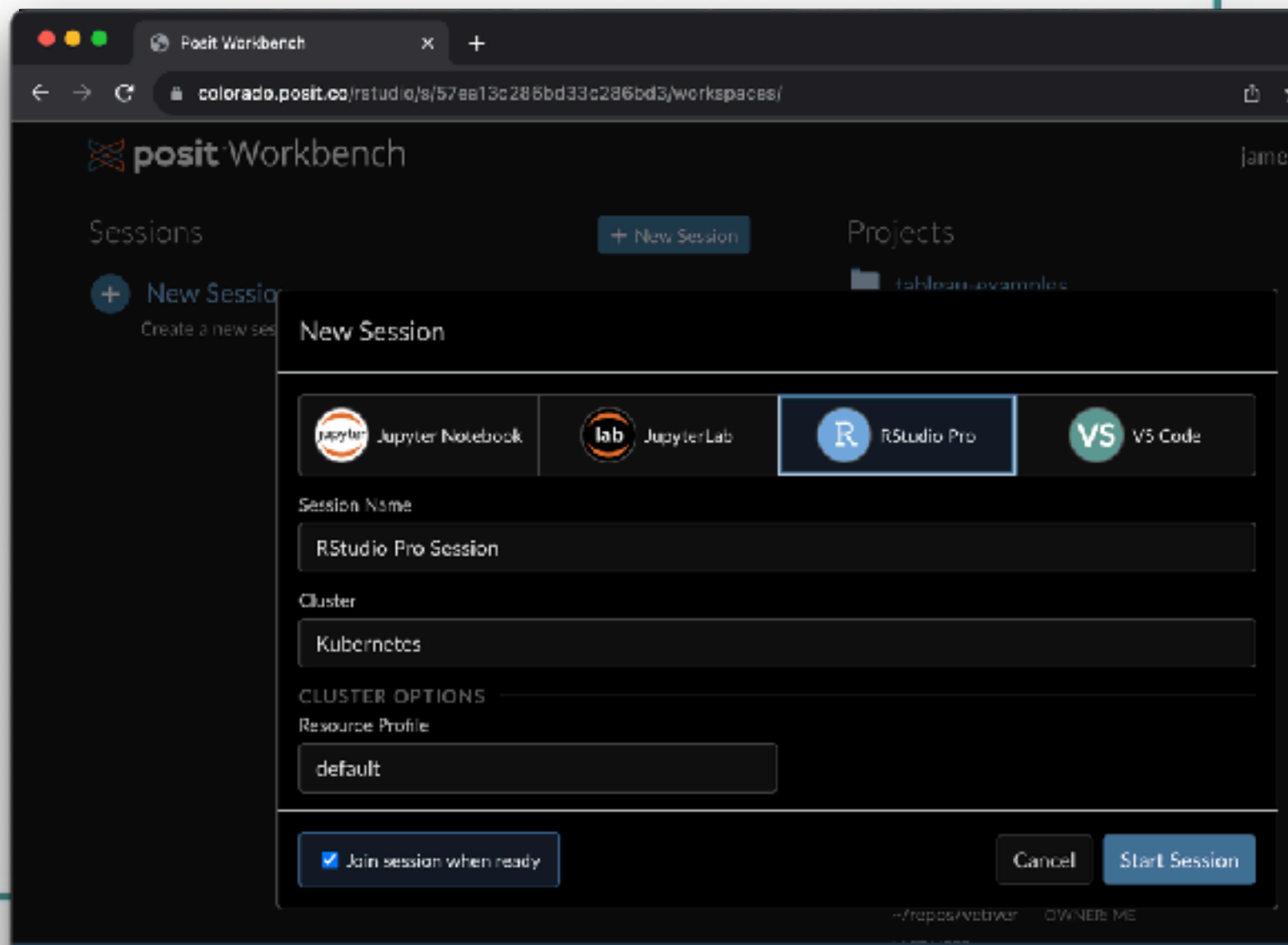
FROM  **posit**

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

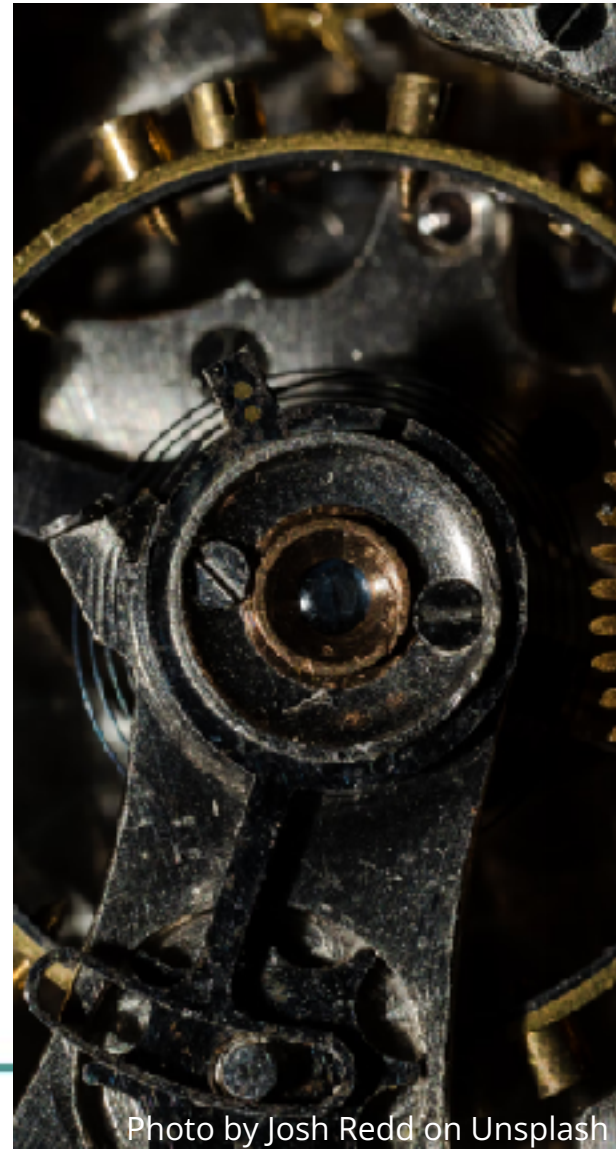
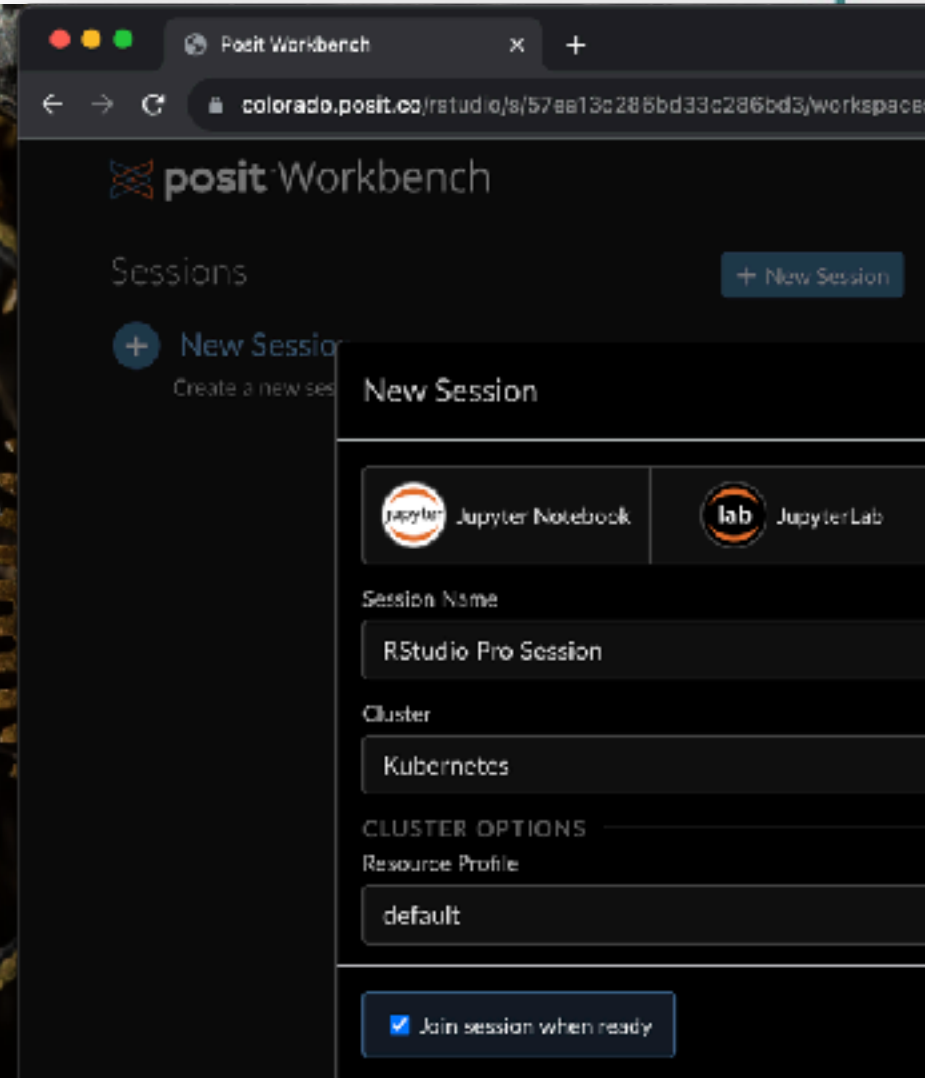


Photo by Josh Redd on Unsplash





Guide ▾

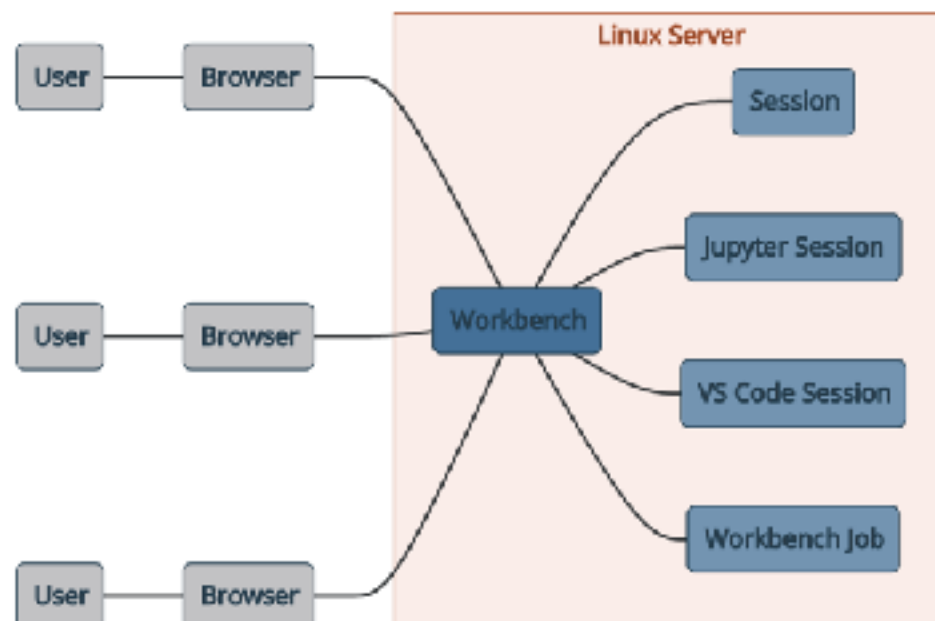
[Write Effective Code](#) >[Connect to Data Sources and Systems](#) >[Manage Packages and Reproduce Environments](#) >[Secure Access](#) >[Implement Operational Patterns](#) >[Choose the Right Architecture](#) ▾[Configuring Python within Posit products](#)[Running Posit Products in Containers](#)[Using 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:

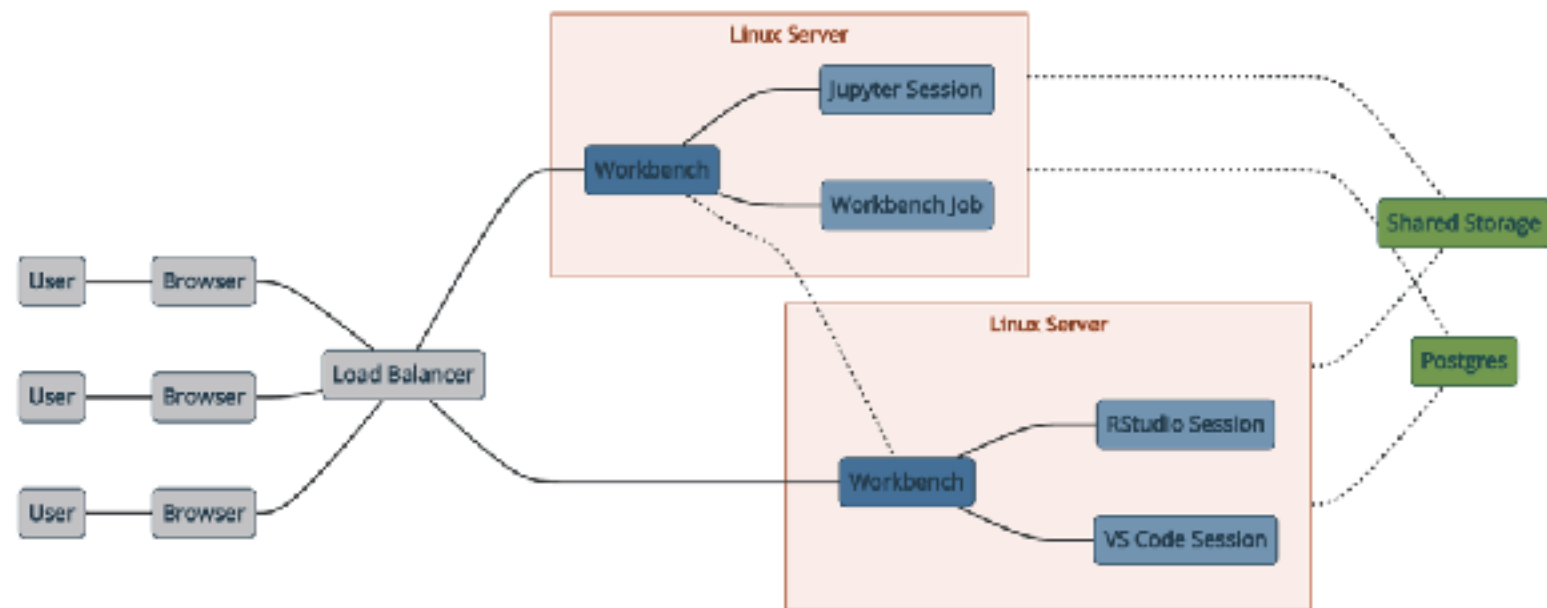
- Access to RStudio, Jupyter Notebook, JupyterLab and VS Code development IDEs
- Multiple concurrent sessions per user
- Use of multiple versions of R and Python



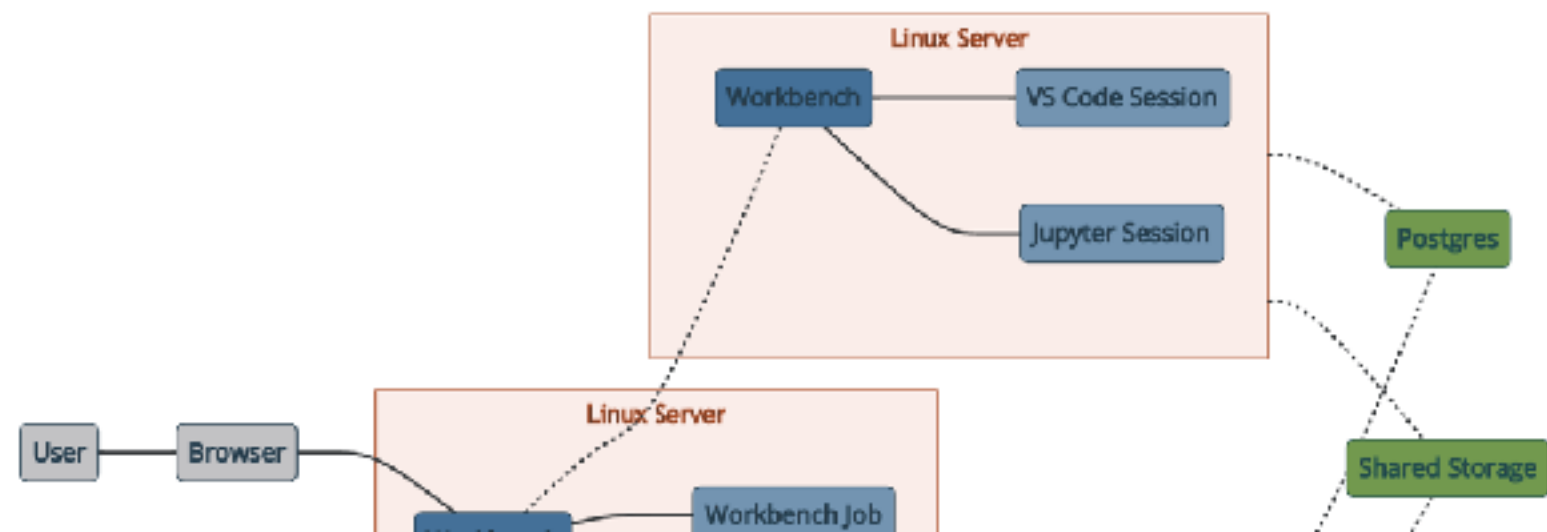
On this page

[Using Workbench on a single server](#)[Using Workbench as a cluster](#)[Using Workbench with an external resource manager](#)[Additional Resources](#)

External Load Balancer

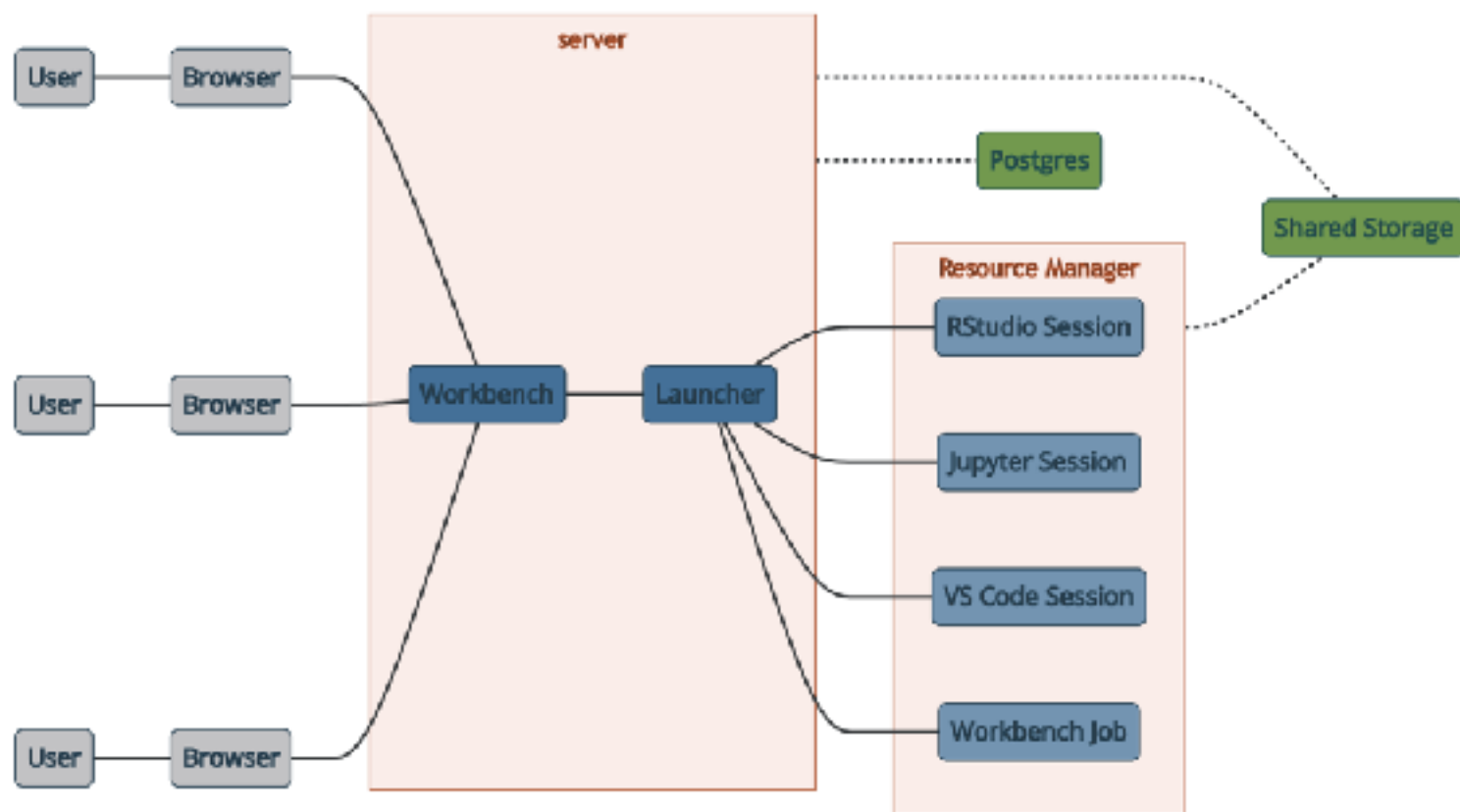


Single Node Routing



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

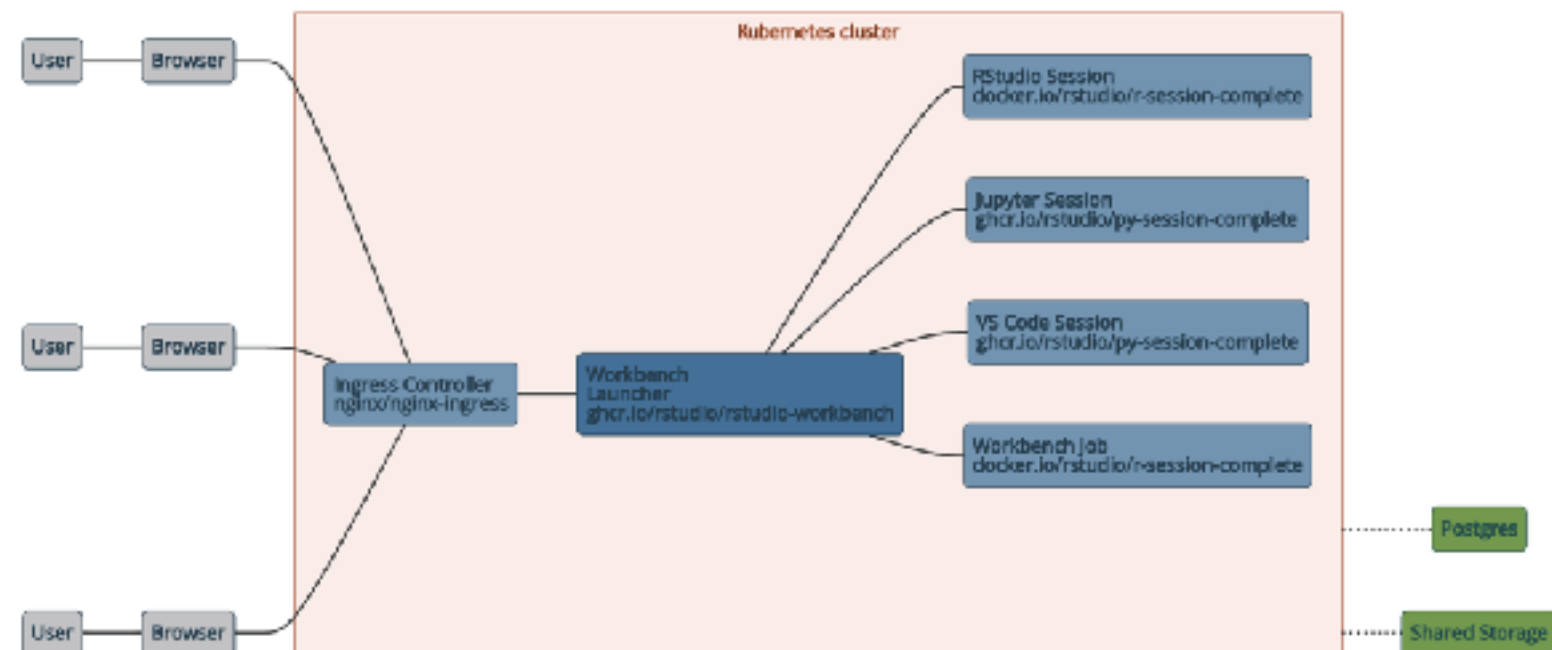
In this configuration, Workbench is installed entirely inside a Kubernetes cluster and enables:

- 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

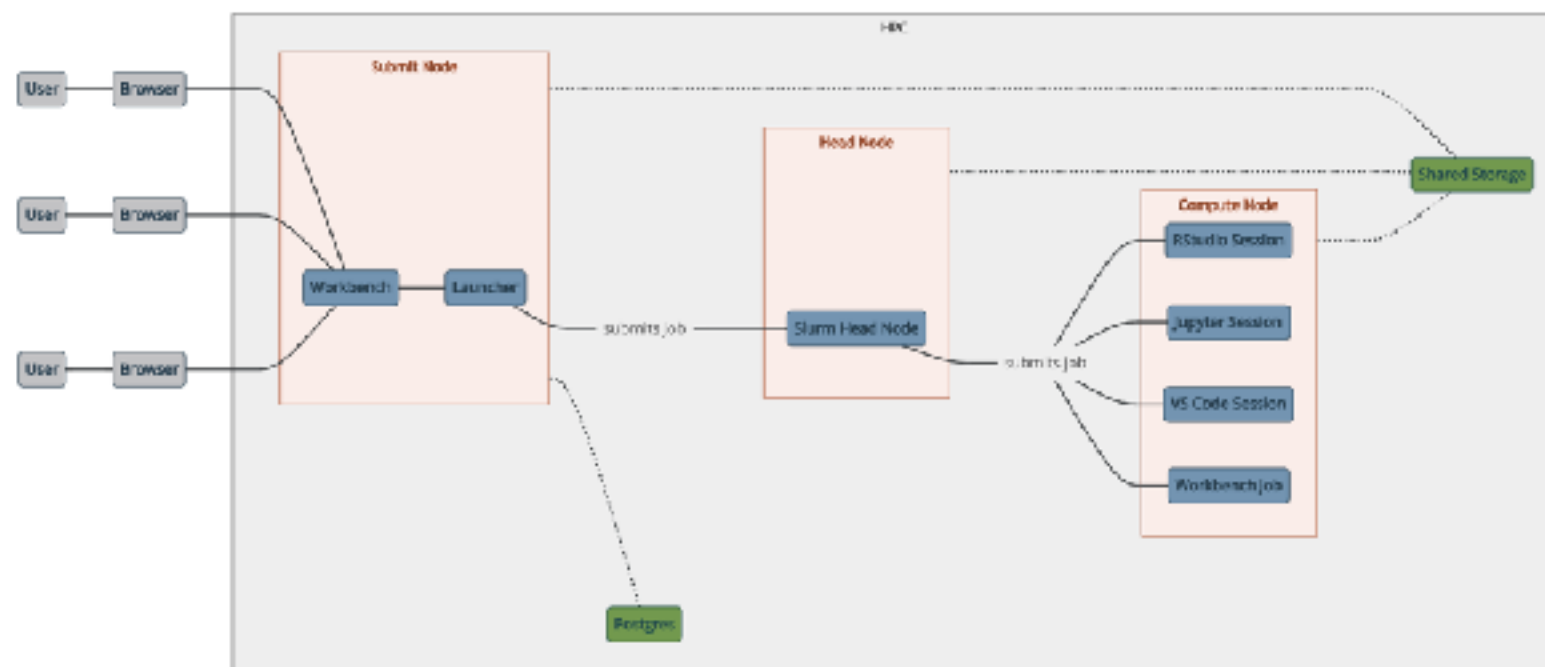


Additional Kubernetes Resources

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](#)

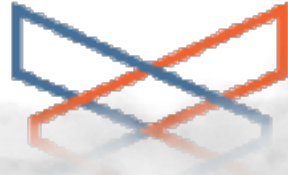




>



>





Domain: Example-Domain | Domains

us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/studio/d-cvzx1awvv14c

CloudWatch

Amazon SageMaker

- Getting started
- Studio
- Studio Lab
- Canvas
- RStudio
- TensorBoard
- Profiler

▼ Admin configurations

- Domains
- Role manager
- Images
- Lifecycle configurations

SageMaker dashboard

Search

▼ JumpStart

Example-Domain

Domain details

Configure and manage the domain.

User profiles | Space management | Environment | Domain settings

User profiles [Info](#)

A user profile represents a single user within a domain. It is the main way to reference a user for the purpose of other user-oriented features.

Search users

Name	Modified on	Created on
james	Aug 08, 2023 05:01 UTC	Aug 08, 2023 05:01 UTC

Personal apps

- Studio
- RStudio**
- Canvas
- TensorBoard
- Profiler

Collaborative

- Spaces

Launch

<https://us-east-1.console.aws.amazon.com/sagemaker/home?region=us-east-1#/studio/rstudio/open/d-cvzx1awvv14c/james> | Web Services, Inc. or its affiliates. | Privacy | Terms | Cookie preferences



Posit Workbench

d-cvzxfawv14c.studio.us-east-1.sagemaker.aws/rstudio/default/s/57ea13c286bd33c286bd3/workspaces/

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

Posit Workbench 2023.03.2-151pro2, "Cherry Blossom" (09192166)

© 2009 - 2023 Posit Software, PBC



Workstations – Cloud Workstations

console.cloud.google.com/workstations/list?project=white-cider-340618

Google Cloud James GCP Experiments Search (/) for resources, docs, products, and more Search

Cloud Workstations

Overview Workstations Workstation configurations Cluster management

Workstations CREATE REFRESH

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, 3:53:06 PM

Configuration demo-config

START

SHOW MORE

All workstations

Filter Enter property name or value

	Status	Name	Location	Workstation cluster	Configuration	Machine
<input type="checkbox"/>	Stopped	posit-wb	us-east1	test-cluster	posit-workbench	e2-stand

41



Posit Workbench

80-posit-workbench-example.cluster-dcprairrfagcux3fymnott7x6.cloudworkstations.dev/a/57ea13c286bd33c286bd3/w...

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

Posit Workbench 2023.03.2-154-prod, "Cherry Blossom" (09192466)

© 2009 - 2023 Posit Software, PBC



RStudio Pro

Not Secure 35.171.27.105:8787/sj0da29cfc78a31bfe742ba7/launcher=1

File Edit Code View Plots Session Build Debug Profile Tools Help

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(dbp1yr)
library(dbp1ot)
```

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):

Console Terminal Background Jobs Workbench Jobs

R 4.3.1 ~ /

Environment History Connections Tutorial Databricks

Databricks Connect - Cluster: 6907-051241-shgxcbm

- hive_metastore
- main
- samples
- system

Files Plots Packages Help Viewer Presentation

Install Update

Name	Description	Version
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
cli	Helpers for Developing Command	3.6.1



RStudio Pro

Not Secure 35.171.27.105:8787/s/0da29cfc78a31bfe742ba7/launcher=1

File Edit Code View Plots Session Build Debug Profile Tools Help

James Sessions - RStudio Pro Session

Render on Save Render Run

Source Visual B J Normal Format Insert Table Outline

```
---
title: "Databricks Test"
format: html
editor: visual
---
```

Connect to Databricks

```
{r_setup}
library(sparklyr)
library(tidyverse)
library(dbp1yr)
library(dbp1ot)
```

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) :

Console Terminal Background Jobs Workbench Jobs

R 4.3.1 - ~/

Environment History Connections Tutorial Databricks

Databricks Connect - Cluster: 6907-051241-shgxcbmh Spark

- hive_metastore
- main
- samples
- system

Install Update

Name	Description	Version
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
cli	Helpers for Developing Command	3.6.1



RStudio Pro

Not Secure | 35.171.27.105:8787/s/0da29cfc78a31bfe742ba7/launcher=1

File Edit Code View Plots Session Build Debug Profile Tools Help

James | Sessions - RStudio Pro Session

Render on Save | Render | Run | Addins

Source Visual | B | J | C | Normal | Format | Insert | Table | Outline

Connect to Databricks...
Connect to Data

```
---
title: "Databricks Test"
format: html
editor: visual
---
```

Connect to Databricks

```
{r_setup}
library(sparklyr)
library(tidyverse)
library(dbp1yr)
library(dbp1ot)
```

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) :

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
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
cli	Helpers for Developing Command	3.6.1



Introducing Lakehouse Apps

Build Secure Native Data and AI Applications Without Compromise



by [Matei Zaharia](#), [Shanku Niyogi](#) and [Prem Prakash](#)

June 21, 2023 in [Platform Blog](#)

Share this post



Lakehouse Apps is a new way to build native applications for Databricks. Lakehouse Apps will offer the most secure way to build, distribute, and run innovative data and AI applications directly on [the Databricks Lakehouse Platform](#), next to the customer's data, with the full security and governance capabilities of Databricks.

For developers building data and AI solutions, distributing Lakehouse Apps through the Databricks Marketplace will dramatically reduce time to adoption, with access to over 10,000 Databricks customers. For customers, Lakehouse Apps will be the most secure way to run applications that unlock the full value of data in their Lakehouse, leverage Databricks-native services, and extend Databricks with new capabilities.



Browser address bar: <https://snowflakecomputing.com/>

Application: RStudio

Code Editor:


```
1 # ...
2 title: "snowflake-report"
3 output: html_document
4 date: "2013-09-15"
5 # ...
6 ## [knitr, knitr::knit]
7 knitr::opts_chunk$set(echo = TRUE)
8 library(cdyverse)
9 # ...
10
11
12 ## R Markdown
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
```

Environment:

- BANK_STOCKS_DB - jrgmcdm@snowflake
- PROGBYTE_TASTY_BYTES
- PROGBYTE_TASTY_BYTES_SETUP_S
- PROGBYTE_TASTY_BYTES_VT
- SLAMP_3B
- PROD1_3B
- QUANT_RESEARCH_DATASET
- SNOWFLAKE
- TUTORIAL_DB

Table: PROGBYTE_TASTY_BYTES

[1] "CUSTOMER_LOYALTY_METRICS_V"	"LOCATION_DETAIL_V"	"ORDERS_V"
[2] "SHIFT_SALES"	"CUSTOMER_LOYALTY_METRICS_V"	"LOCATION_DETAIL_V"
[3] "ORDERS_V"	"APPLICABLE_ROLES"	"CLASSES"
[4] "CLASS_INSTANCES"	"CLASS_INSTANCE_FUNCTIONS"	"CLASS_INSTANCE_PRODUCTS"
[5] "COLUMNS"	"DATABASES"	"ENABLED_ROLES"
[6] "EVENT_TABLES"	"EXTERNAL_TABLES"	"FILE_FORMATS"
[7] "FUNCTIONS"	"INFORMATION_SCHEMA_CATALOG_NAME"	"LOAD_HISTORY"
[8] "OBJECT_PRIVILEGES"	"PACKAGES"	"FILES"
[9] "PROCEDURES"	"REFERENTIAL_CONSTRAINTS"	"REPLICATION_DATABASES"
[10] "REP_SCATTER_GROUPS"	"SCHEMAS"	"SEQUENCES"
[11] "TABLES"	"STATES"	"STREAMS"

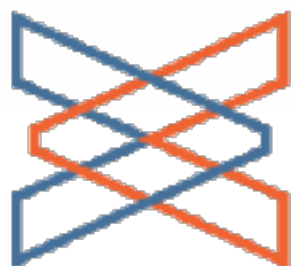


The screenshot shows the Databricks interface with the 'Tutorial' tab selected. The SQL editor contains the command 'SHOW TABLES;'. The output displays a list of tables in the 'TUTORIAL_DB' database:

```

SHOW TABLES;

TABLE
PROFITBYTE_TASTY_BYTES
PROFITBYTE_TASTY_BYTES_SETUP_S
PROFITBYTE_TASTY_BYTES_V2
ELAND_36
PDS11_36
QUANT_RESEARCH_DATASET
SHOWPLAGE
TUTORIAL_DB
  
```

Posit Cloud

posit.cloud

Log In

Sign Up

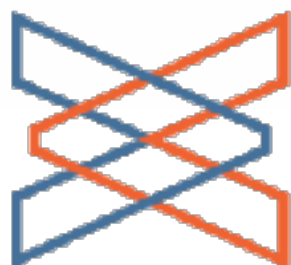
Friction free data science

Posit Cloud (formerly RStudio Cloud) lets you access Posit's powerful set of data science tools right in your browser – no installation or complex configuration required.

GET STARTED

ALREADY A USER? LOG IN

If you already have a shinyapps.io account, you can [log in](#) using your existing credentials.



Posit Cloud

posit.cloud/spaces/50579/content/all?sort=name_asc

Posit Cloud

Scratch Pad James Blair

Content Data Members Usage About

James Blair

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

All Content

Your Content

Templates

Archive

Trash

New List

All Content (1)

TYPE * ACCESS * SORT A Z

Sample Project

RStudio Project James Blair Private Created Sep 19, 2023 11:50 PM

posit Cloud

Terms Status

f

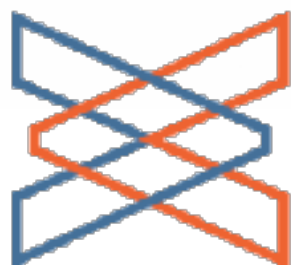
t

in

o

o

© 2023 Posit Software, POC



Posit Cloud

Scratch Pad / Sample Project

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

File Edit Code View Plots Session Build Debug Profile Tools Help

R 4.3.1 - /cloud/project/

Console

Terminal

Background Jobs

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

126 MB

List

R

Global Environment

Environment is empty

Files

Plots

Packages

Help

Viewer

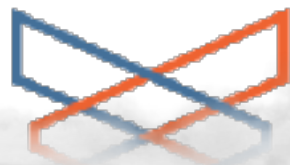
Presentation

Cloud > project

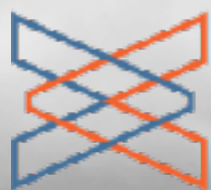
..

.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.