Open source & enterprise-ready professional software for data science

RStudio Cloud & RStudio Workbench: *Academic Teaching/Research IT Overview*

If you're reading this your institution likely has the infrastructure resources to host an RStudio Workbench server but what that entails is not always initially grasped. This document will point out the human resources needed in addition to the infrastructure and how that might compare to simply logging into RStudio Cloud.

RStudio Workbench / RStudio Cloud Overview

	RStudio Workbench	RStudio Cloud
Server Provisioning	Your responsibility	Included
Scaling	Your responsibility	Included
System Updates	Your responsibility	Included
Install and Update R and Python	Your responsibility	Included
Install and Update Required System and R packages	Your responsibility	Included
Course Spaces and Assignment Projects	Not available	Included
Usage Metrics	Not available	Included (per account, per course space, per student, per project)
Technical Support	Limited (software only)	Full (software and infrastructure)

RStudio Workbench

RStudio Workbench is software that is installed and maintained internally by the purchasing organization. For large, academic workloads, this effort typically involves the following:

Provisioning the server environment

RStudio Workbench must be installed in a Linux environment. This can either be a bare metal server or a cloud VM hosted by a cloud provider. RStudio Workbench can optionally be configured in a cluster, with multiple nodes. It can also be configured to interact with a resource manager like Kubernetes to provide on-demand resources for R-sessions. Each of these components would need to be managed and maintained by IT.

Wait! What is Kubernetes and how is it related to RStudio?

Kubernetes is an open-source container-orchestration system for automating computer application deployment, scaling, and management. RStudio Workbench combined with Kubernetes allows you to use a container-based approach to help your development environment scale out as your data science users (I.e students/instructors/researchers) and computational needs grow. Coupled with an elastic infrastructure, this functionality helps solve the problem of manually scaling RStudio Workbench along with a growing user base.

Installing RStudio Workbench and associated dependencies

Once the environment is provisioned, RStudio Workbench needs to be installed along with relevant versions of R and, optionally, Python. There are additional system dependencies that may need to be managed and installed depending on the intended usage of R and Python. R and Python ecosystems both heavily rely on community-developed packages for enhanced functionality. These packages often introduce further system dependencies that will need to be managed and maintained over time.

Maintaining the RStudio Workbench environment

Once installed, the RStudio Workbench environment needs to be maintained, which means keeping the underlying operating system up to date and managing the necessary system dependencies.

Keeping RStudio Workbench Updated

RStudio Workbench receives periodic updates with bug fixes and new functionality. It is the responsibility of IT to keep RStudio Workbench updated.

Keeping RStudio Workbench Dependencies Updated

R and Python receive periodic updates as well. It is the responsibility of IT to ensure the latest R and Python versions are available on the server.

RStudio Cloud
Much simpler to jump in than everything else you just read and requires virtually no maintenance.

RStudio Cloud is a hosted solution that provides users with access to the familiar RStudio IDE plus capabilities that make teaching easier which you would not find in RStudio Workbench. An example would be the ability to set up a Workspace for each course akin to a virtual classroom. This solution is entirely managed by RStudio. Therefore, all the above tasks are managed by RStudio. In addition, RStudio Cloud leverages Kubernetes so that users can define their resource requirements on demand.

Sounds great but how?

RStudio Cloud makes use of virtualized servers called containers to provide isolated environments for running individual projects. Under the hood, these containers are managed by Kubernetes, a best-in-class container orchestration platform. Using Kubernetes allows the RStudio Cloud service to quickly and easily auto-scale to meet the demands of our customer's workloads. Whether you're running a class of 50 or 500 students, Kubernetes helps make sure RStudio Cloud always has the capacity to run your project at a moment's notice.

Users can specify their CPU and memory requirements as they work on different projects. Furthermore, RStudio Cloud provides access to pre-built R package binaries, reducing the installation time for R packages and significantly reducing the need to worry about underlying system dependencies. Issues with RStudio Cloud would be resolved via RStudio Support rather than relying on internal IT.

Even though RStudio Cloud is a hosted SaaS offering you can integrate SSO for a seamless authentication experience.

About RStudio

RStudio® makes data analysis with R and Python easier and provides powerful tools for sharing reports, dashboards, interactive Shiny® applications and Jupyter Notebooks with your entire enterprise.