

The Best of Both Worlds

RStudio and Qubole Come Together

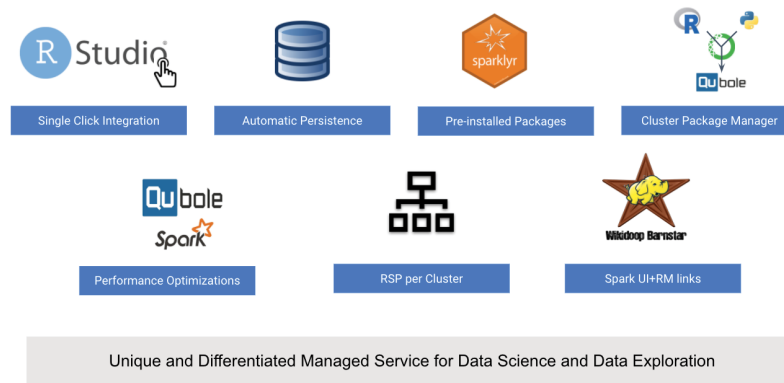
The integration of both platforms accelerates data science and scientific research now with the single click access to large datasets with IDE environments of data scientists.

Data scientists use RStudio for statistical computing, graphics, and machine learning. With vast amounts of enterprise data becoming accessible, the size of data to be processed becomes large enough, to feel the need to leverage the power of cluster computational frameworks like Apache Spark which have seamless integration with R using libraries like SparkR and Sparklyr. Data scientists can now easily crunch large datasets and derive actionable insights for their projects closer to the real-world scenarios in significantly shorter time with RStudio Server Pro and Qubole Open Data Lake Platform Integration. It combines the power of the RStudio Server Pro with enhanced Apache Spark capabilities powered by Qubole.

Users now benefit with following for having RStudio Server Pro with Qubole:

- Simplified Access to Large Datasets for R Users**
 Access real datasets for your R-based AI/ML projects and deliver real-world scenario results faster
- Increased Productivity of R Users**
 Keep using familiar tools and languages to run and execute jobs on Qubole skipping a learning curve for new ML tools
- Get TCO Optimized scalable infrastructure for R-based projects**
 Have clear ROI insights and cost savings of their data science projects without disrupting existing workflows

Integration provides a unique and differentiated managed service with following out-of-box features:



Single Click Integration

Provides RStudio Server Pro experience with Qubole Managed Spark clusters. No other platform provides this user experience.

Automatic Persistence

Persists all the files and resources created in the home folder of the user to cloud storage and restores the state on cluster restart automatically and transparently.

Pre-Installed Packages

Offers various pre-packaged libraries to start data science journey with R at enterprise level and a convenient qubole method in sparklyr to start spark sessions on Qubole clusters.

Cluster Package Manager

Provides a unified package management solution available in Qubole for Spark clusters using Conda, PyPi, CRAN & RSPM repositories. Users can define cluster-wide R & Python dependencies for Spark applications and have a centralized location for all repositories.

Performance Optimizations

Automatically scale up Qubole cluster when the sparklyr application needs more resources and downscale as cluster resources are not in use. Beyond the scalability and access to managed Apache Spark, R users in Qubole can now leverage Qubole's optimizations to Apache Spark to speed up the time to value.

Zero License Management and On-Demand License

Have on-demand access to RStudio on ephemeral Spark clusters with zero license management. Users can also have on-demand licenses via Qubole to enable RStudio Server Pro within the Qubole Data Lake Platform.

[To get started, click here for Free Trial](#)

About Qubole

Qubole is the open data lake company that provides a simple and secure data lake platform for machine learning, streaming, and ad-hoc analytics. No other platform provides the openness and data workload flexibility of Qubole while radically accelerating data lake adoption, reducing time to value, and lowering cloud data lake costs by 50 percent. Qubole is trusted by leading brands such as Expedia, Disney, Oracle, Gannett and Adobe to spur innovation and to transform their businesses for the era of big data. For more information visit us at www.qubole.com