databricks

Posit + Databricks

Improved productivity for your data teams

Rafi Kurlansik
James Blair

James Blair
Posit
Product Manager
Cloud Integrations

Rafi Kurlansik
Databricks
Product Specialist
Developer Experience, AI

6000+
global employees

$1.5B+
in revenue

$4B
in investment

Inventor of the **lakehouse**
&
Pioneer of **generative AI**

databricks
The data and AI company

Gartner-recognized Leader
Database Management Systems
Data Science and Machine Learning Platforms

Creator of

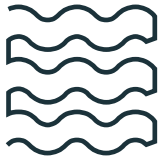
DELTA LAKE

mlflow™

APACHE Spark™

The winners in every industry will be data + AI companies

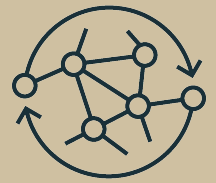
Data Lake



Machine Learning

Streaming

Generative AI



Data Science

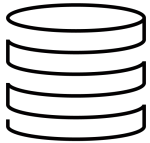
Most organizations struggle to realize this vision

Governance

Orchestration & ETL

BI

Data Warehouse



The Data Lakehouse

An open, unified foundation for all your data

Data Science & AI	ETL & Real-time Analytics	Orchestration	Data Warehousing
Databricks AI	Delta Live Tables	Workflows	Databricks SQL

Unified security, governance, and cataloging

Unity Catalog

Unified data storage for reliability and sharing

Delta Lake

Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)

2020
Databricks pioneered the lakehouse architecture

Today
74% of global enterprises have adopted lakehouse
MIT Technology Review Insights, 2023

Data Lakehouse

An open, unified foundation for all your data

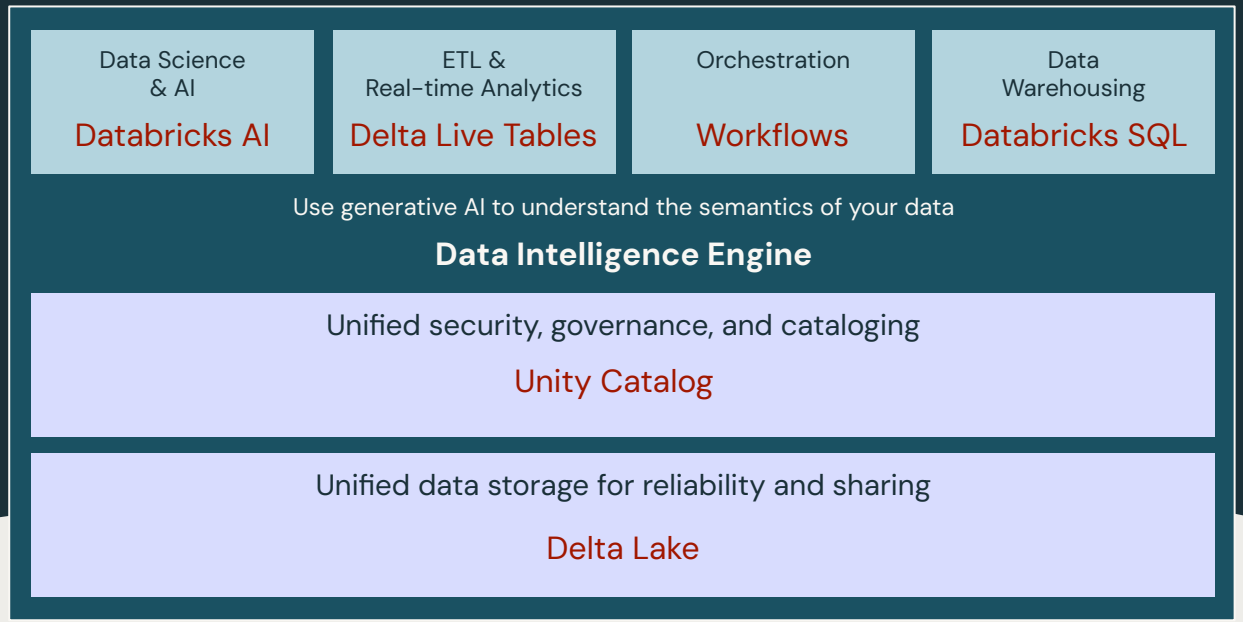
Generative AI

Easily scale and use data and AI

Data Intelligence Platform

Democratize data + AI across your entire organization

Databricks Data Intelligence Platform



Open Data Lake

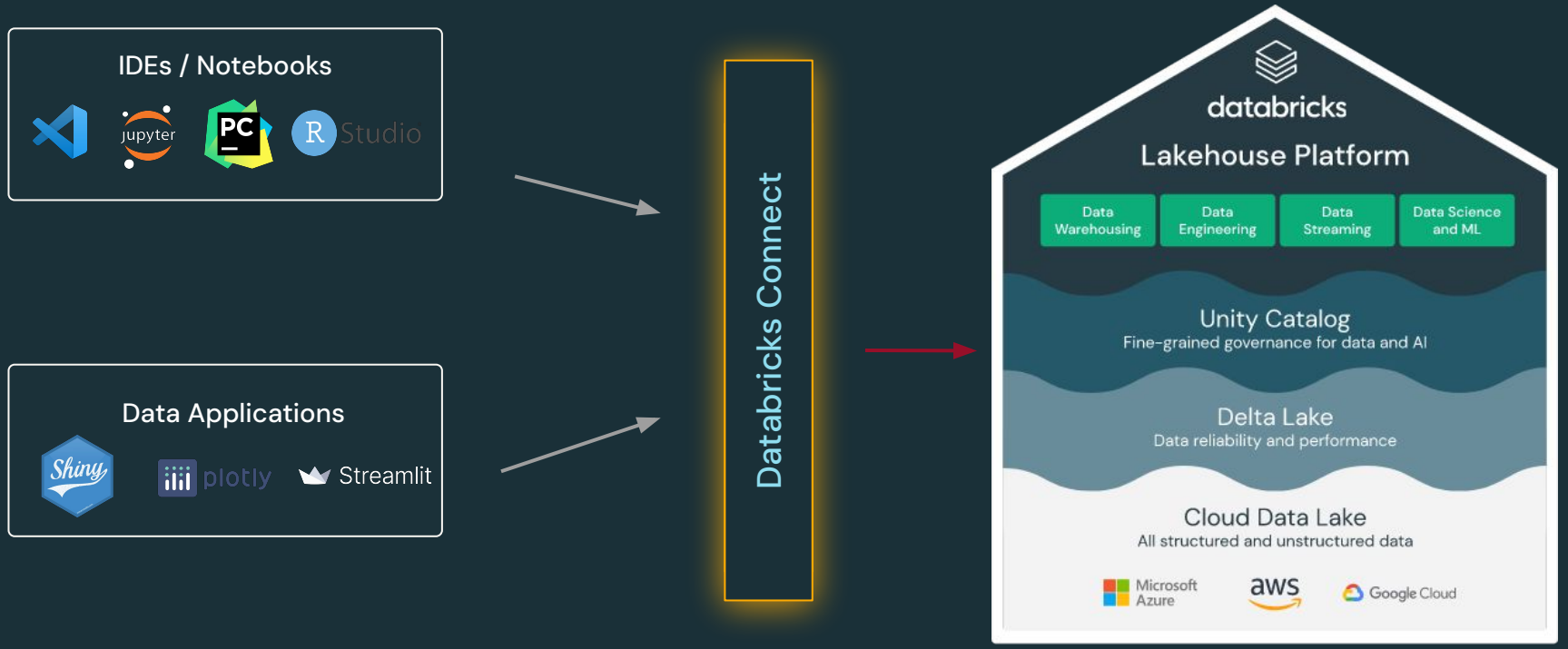
All Raw Data
(Logs, Texts, Audio, Video, Images)

How do we access the Data Intelligence Platform with Posit products?

Databricks Connect V2

Bringing remote connectivity to the Lakehouse

Python GA, R & Scala in preview



Databricks Extension for Visual Studio Code

GA

Simple Setup

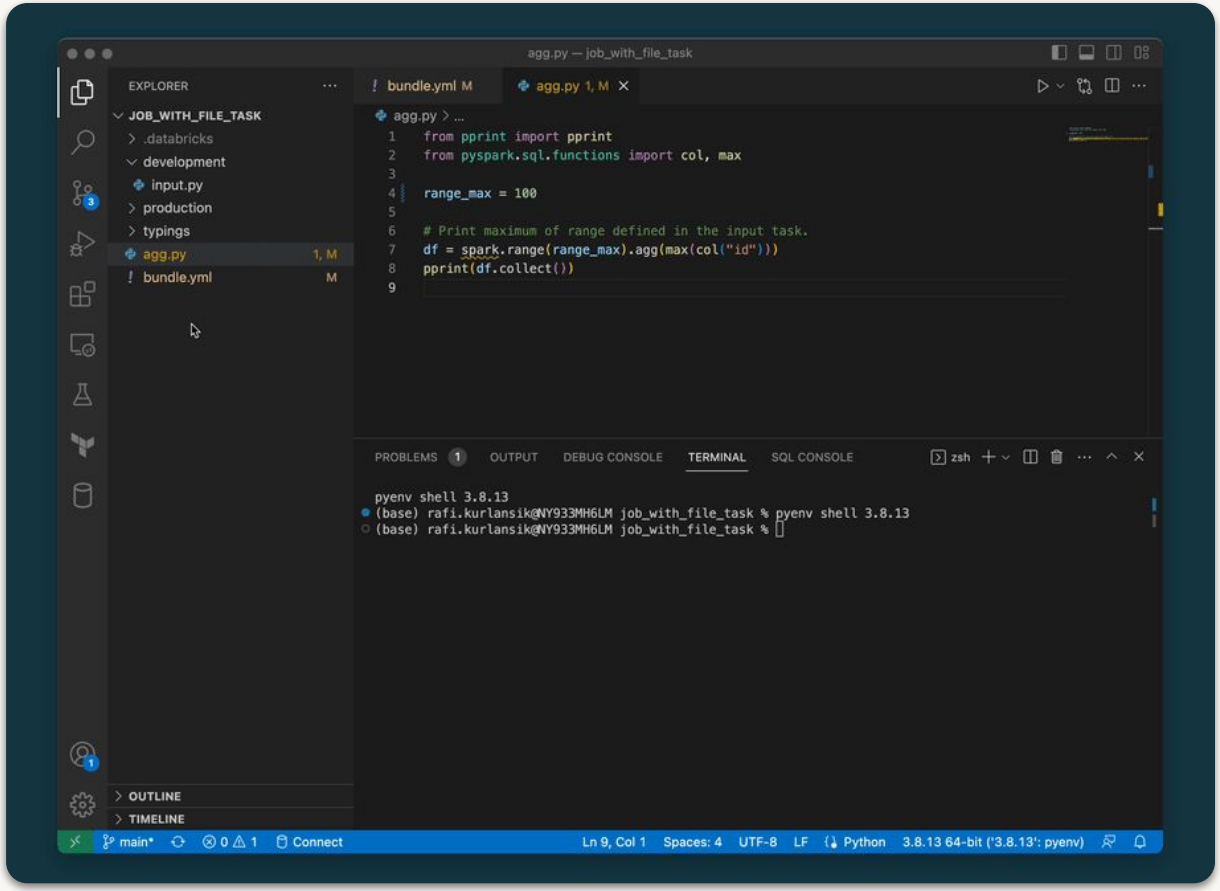
Find us on the VS Code Marketplace and get connected to compute in minutes

Native Experience

Write code using the productivity features you love from VS Code

Run on Databricks

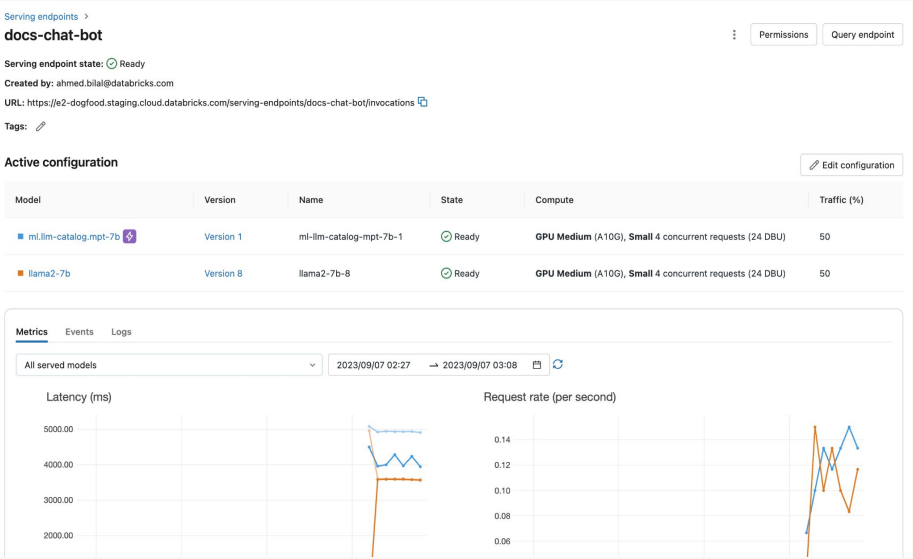
Execute batch workloads or start interactively debugging directly from your IDE



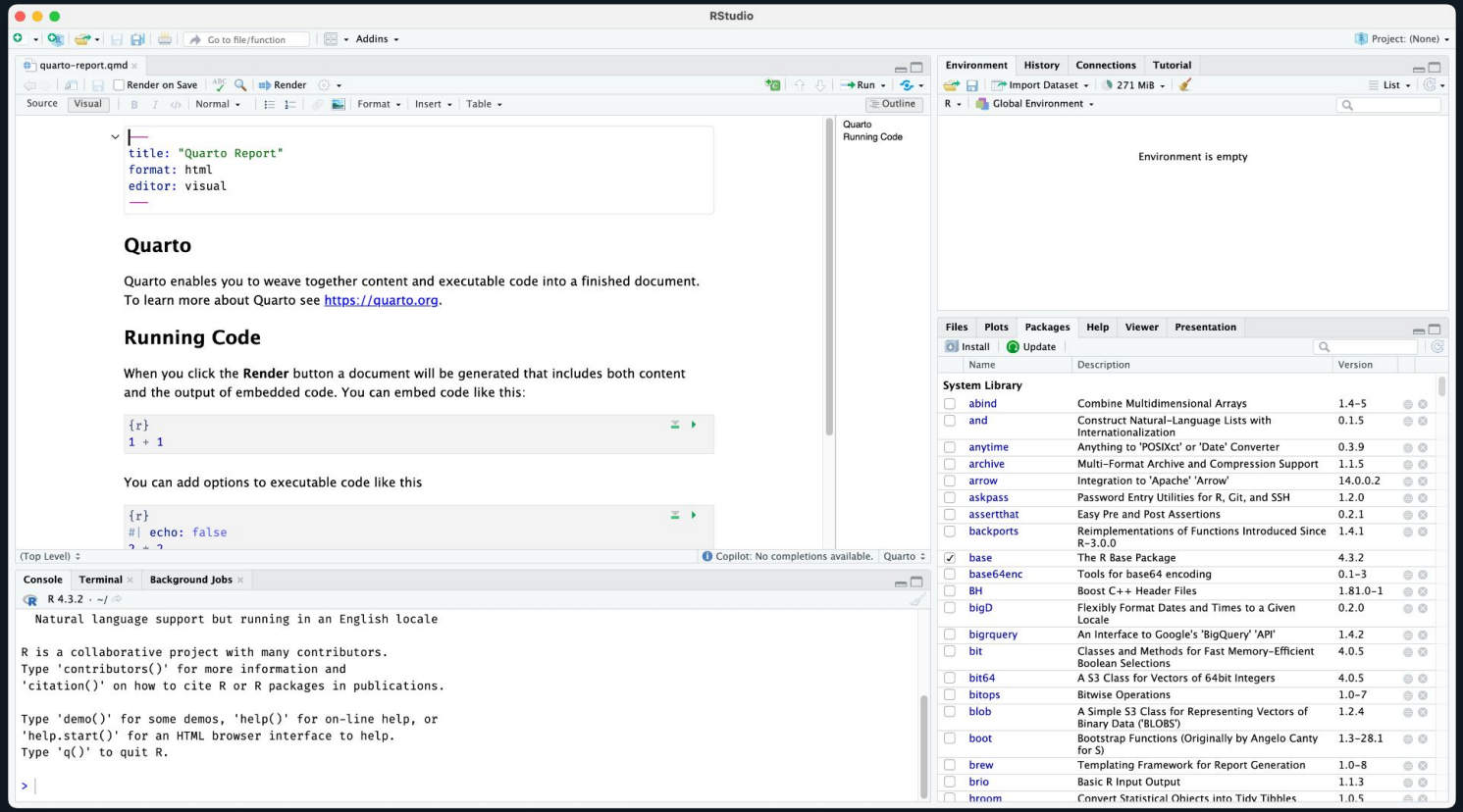
Databricks Model Serving

Real-time inference of Models, reducing latency and cost by up to 3-5x

- Fully managed, including GPU support, so you don't waste time managing infrastructure
- Optimized for specific LLMs to provide higher performance
- Automatic logging of Inference Tables into Delta tables and monitoring of performance

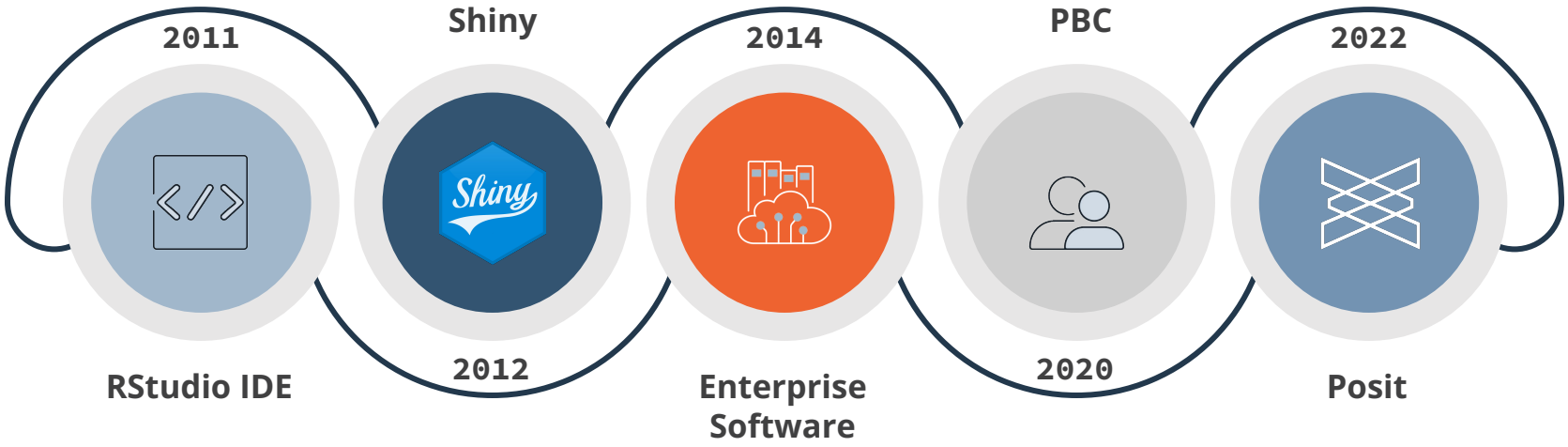



databricks
X
posit™

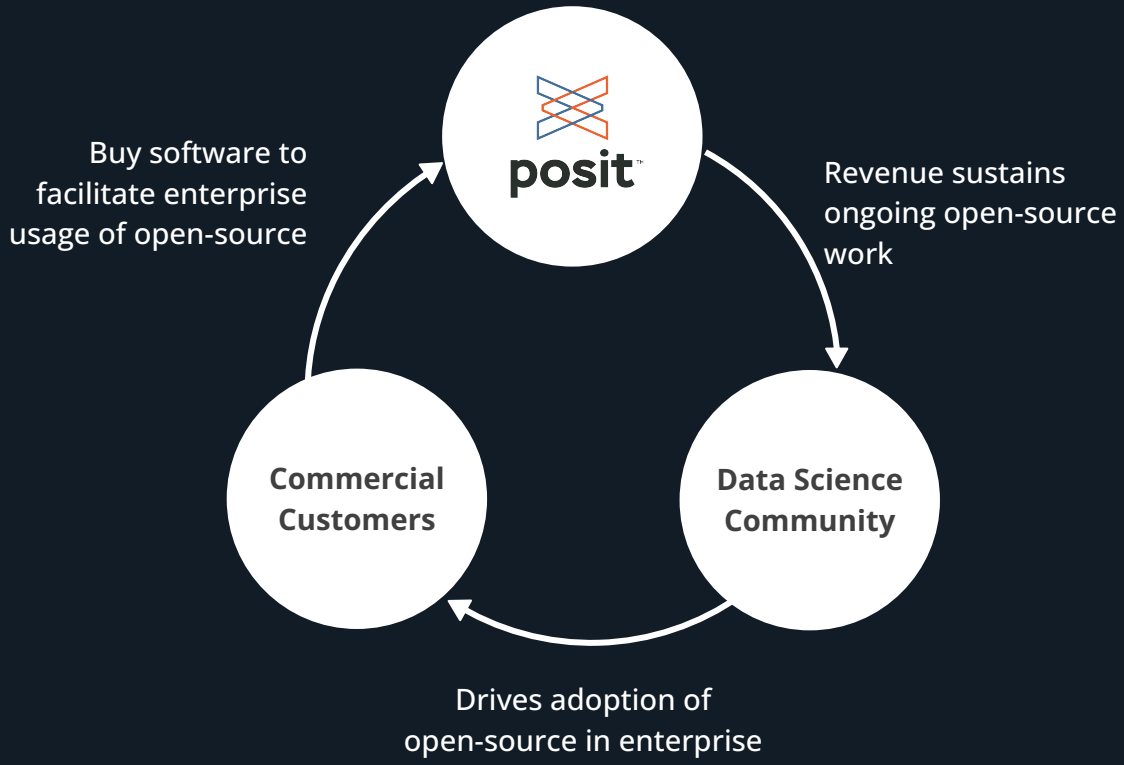

RStudio
quarto-report.qmd
Render on Save
Render
Run
Source
Visual
Normal
Format
Insert
Table
Outline
Quarto
Running Code
---
title: "Quarto Report"
format: html
editor: visual

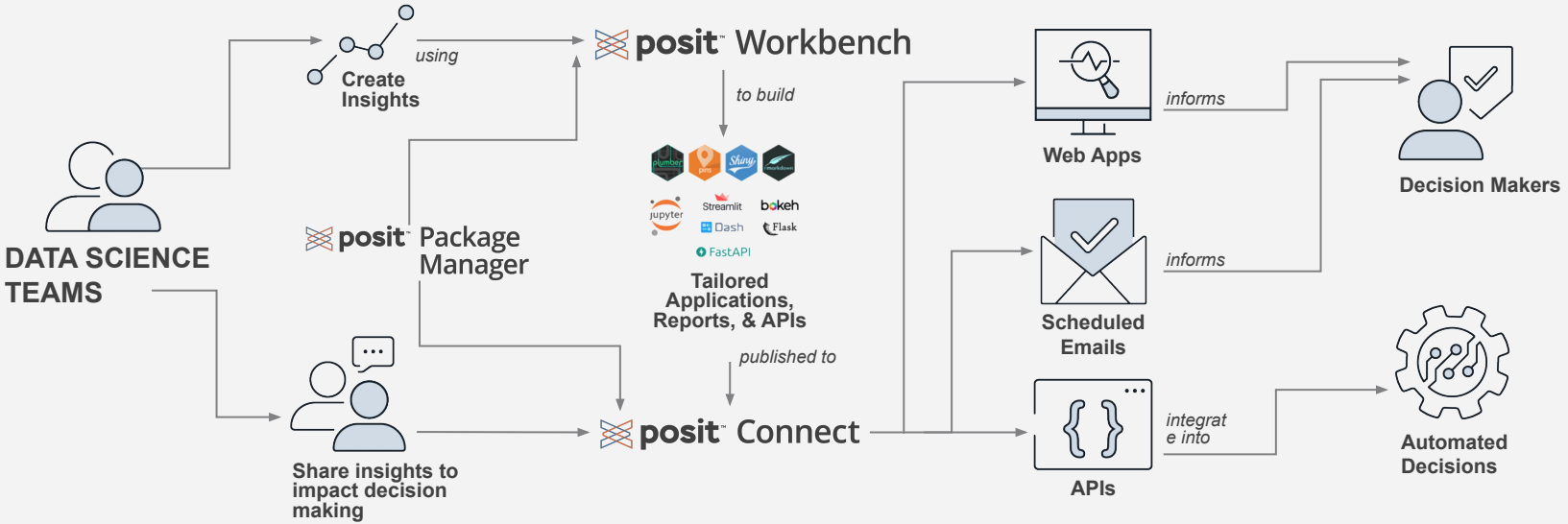
Quarto
Quarto enables you to weave together content and executable code into a finished document.
To learn more about Quarto see https://quarto.org.
Running Code
When you click the Render button a document will be generated that includes both content and the output of embedded code. You can embed code like this:
{r}
1 + 1
You can add options to executable code like this
{r}
#| echo: false
(Top Level)
Copilot: No completions available.
Quarto
Console
Terminal
Background Jobs
R 4.3.2 · ~/
Natural language support but running in an English locale
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
271 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Presentation
Install
Update
Name
Description
Version
System Library
abind
Combine Multidimensional Arrays
1.4-5
and
Construct Natural-Language Lists with Internationalization
0.1.5
anytime
Anything to 'POSIXct' or 'Date' Converter
0.3.9
archive
Multi-Format Archive and Compression Support
1.1.5
arrow
Integration to 'Apache' 'Arrow'
14.0.0.2
askpass
Password Entry Utilities for R, Git, and SSH
1.2.0
assertthat
Easy Pre and Post Assertions
0.2.1
backports
Reimplementations of Functions Introduced Since R-3.0.0
1.4.1
base
The R Base Package
4.3.2
base64enc
Tools for base64 encoding
0.1-3
BH
Boost C++ Header Files
1.81.0-1
bigD
Flexibly Format Dates and Times to a Given Locale
0.2.0
bigrquery
An Interface to Google's 'BigQuery' 'API'
1.4.2
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
bitops
Bitwise Operations
1.0-7
blob
A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')
1.2.4
boot
Bootstrap Functions (Originally by Angelo Canty for S)
1.3-28.1
brew
Templating Framework for Report Generation
1.0-8
brio
Basic R Input Output
1.1.3
broom
Convert Statistical Objects into Tidy Tibbles
1.0.5


tidymodels
reticulate
Shiny
webR
tidyverse
sparklyr
siuba
Vetiver
torch


2011
RStudio IDE
Shiny
2012
2014
Enterprise Software
PBC
2020
2022
Posit

Contribute ~50% of engineering resources to open-source projects




DATA SCIENCE TEAMS
Create Insights
using
posit Workbench
to build
posit Package Manager
Tailored Applications, Reports, & APIs
published to
posit Connect
Share insights to impact decision making
Web Apps
Scheduled Emails
APIs
informs
informs
integrat e into
Decision Makers
Automated Decisions

posit™ Team

posit™ Team

odbc sparklyr SDKs APIs

Data Science & AI	ETL & Real-time Analytics	Orchestration	Data Warehousing
Databricks AI	Delta Live Tables	Workflows	Databricks SQL

Use generative AI to understand the semantics of your data

Data Intelligence Engine

Unified security, governance, and cataloging

Unity Catalog

Unified data storage for reliability and sharing

Delta Lake

Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)


posit
Demo
pandas

posit

Demo

VS Code in Posit Workbench

- Databricks VS Code extension
- Databricks Connect V2

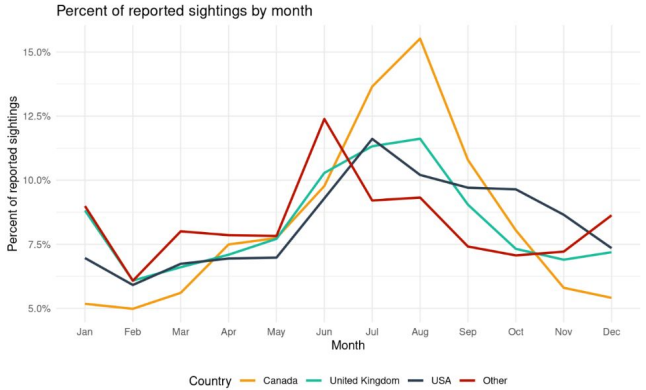
Databricks Model Serving via:

- Databricks Python SDK (GA)
- Foundation Models API (preview)

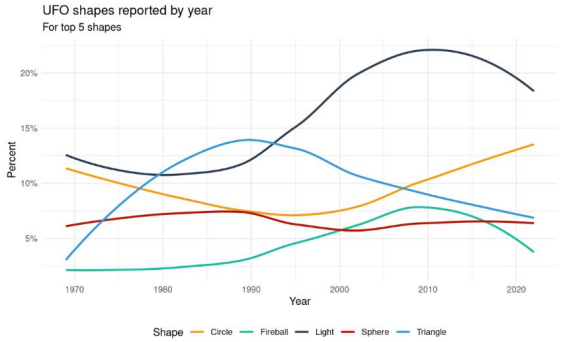
Demo

Reported UFO sightings

Month | Day | Hour



Plot | Data



Most frequently used words to describe sighting

	Word	Count	Percent
1	light/lights	57313	40.57
2	sky	29635	20.98
3	object	22697	16.07
4	bright	22630	16.02

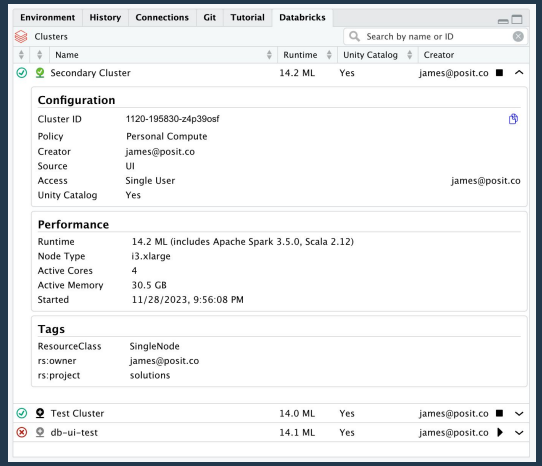
Sightings in 2022
4326

US sightings in 2022
3753

Most common shape
light

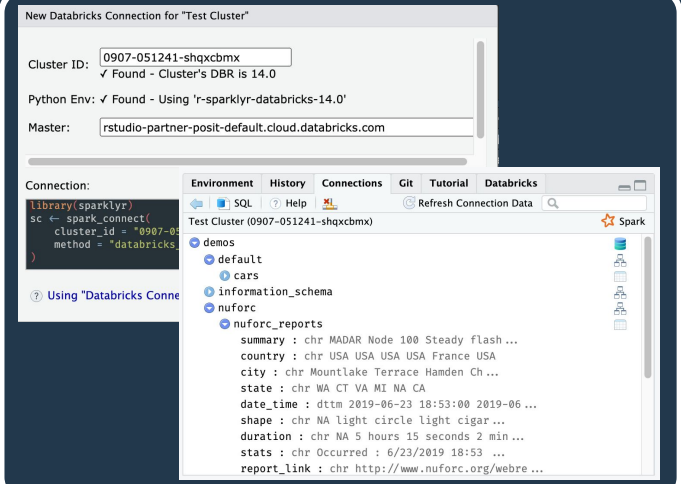
Databricks x Posit: **Now**

Seamless Databricks development in **RStudio** and **VSCode**



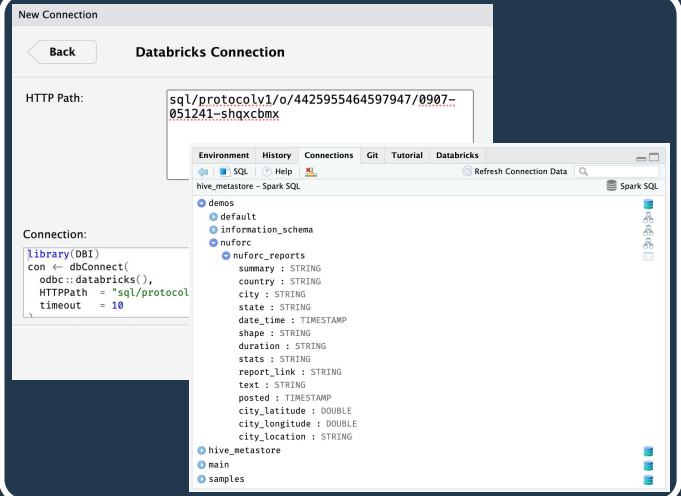
- Posit Workbench managed Databricks credentials - no need to supply PATs
- Manage and connect to Databricks clusters directly from the IDE
- Use managed credentials alongside Databricks tools and SDKs to interact with Databricks from RStudio and VS Code
- Enable Databricks VS Code extension to interact directly with Databricks

Making R and RStudio even  sparklyr



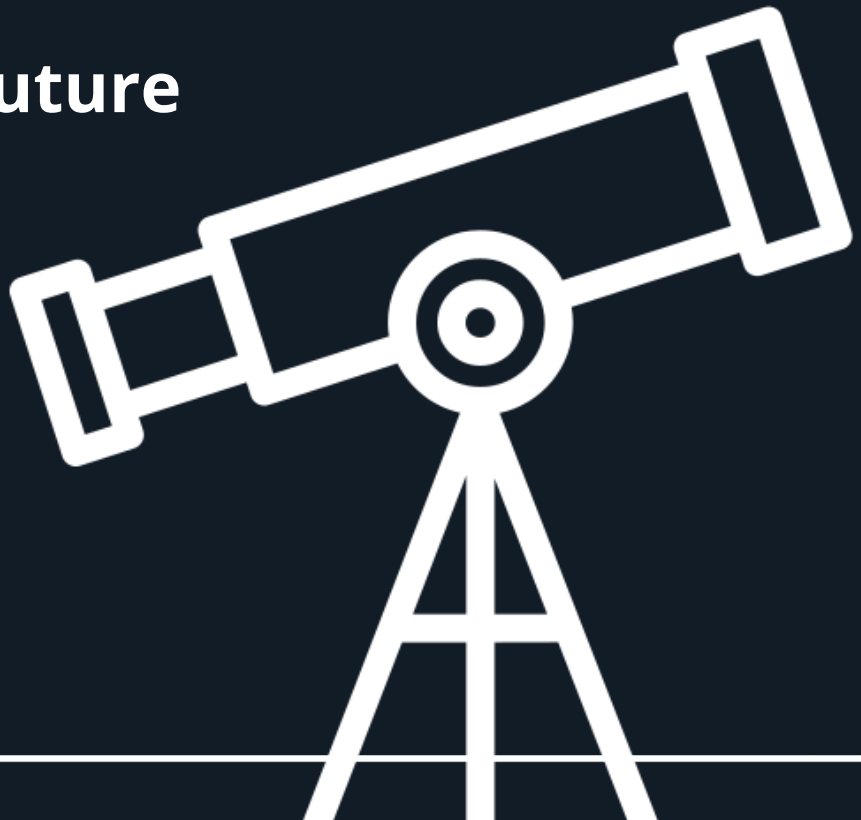
- Posit has updated the `sparklyr` package to support Databricks Connect v2
- This provides support for R on the latest Databricks runtimes (v13.0 and beyond!)
- `sparklyr` provides a fully interactive R experience via remote connection to Databricks
- Growing support for additional Spark functions and features

Simplified **odbc** connectivity



- Posit will begin delivering a Databricks specific ODBC driver
- The `odbc` R package includes a new `databricks()` function to simplify Databricks connections

Databricks x Posit: **Future**



posit™ Team

odbc sparklyr SDKs APIs

Data Science & AI	ETL & Real-time Analytics	Orchestration	Data Warehousing
Databricks AI	Delta Live Tables	Workflows	Databricks SQL

Use generative AI to understand the semantics of your data

Data Intelligence Engine

Unified security, governance, and cataloging

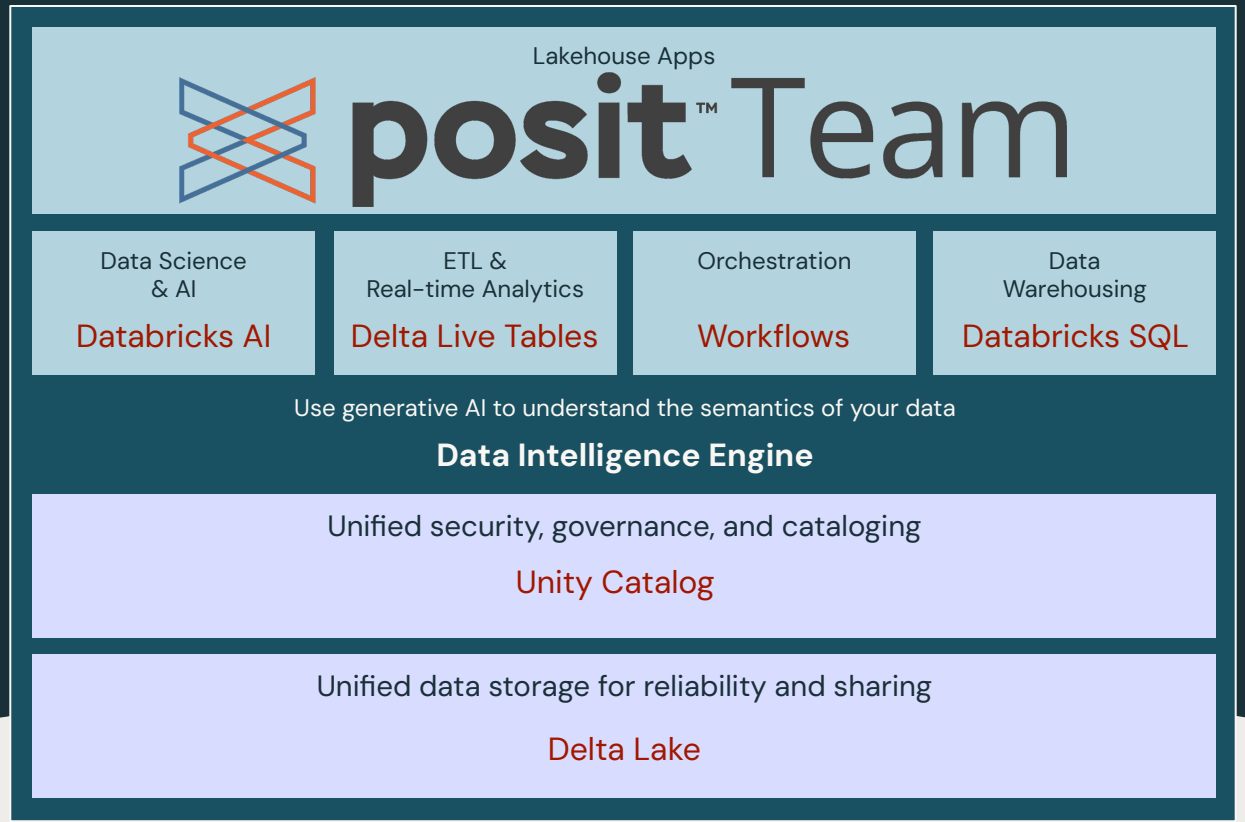
Unity Catalog

Unified data storage for reliability and sharing

Delta Lake

Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)



Open Data Lake

All Raw Data
(Logs, Texts, Audio, Video, Images)

Lakehouse Apps: Security Without Compromise

Lakehouse Apps addresses these challenges with a native, secure, no-compromise solution.

Lakehouse Apps run directly on a customer's Databricks instance, where they can integrate with the customer's data, use and extend Databricks services, and enable users to interact with a single sign-on experience - all with the same security, privacy, and compliance controls as Databricks. **Data never needs to leave the customer's instance**.

Lakehouse Apps can be instantly distributed and monetized to over 10,000 Databricks customers through the Databricks Marketplace. Customers can discover, install, secure, manage, and govern them as efficiently as Databricks native features. Compute resources used by apps are billed directly to the customer by Databricks. A customer's internal apps can also be directly installed into an instance.

Lakehouse Apps are secured, sandboxed, and governed. Every app runs within a secure, configurable sandbox. Customers can use Unity Catalog, Databricks' unified governance solution, to select which resources an app can access, control who can interact with the app, and automatically govern an app's activities.

Lakehouse Apps are built with the technology of your choice. Apps run on secure, auto-scale compute that runs containerized code that can be written in virtually any language, so developers are not limited to building in any specific framework. Applications that integrate with Databricks today can be easily converted into Apps.

Lakehouse Apps are fully Lakehouse-native. The Databricks platform includes a powerful set of scalable and cost-effective serverless services, including fast data warehousing, workflows and pipelines, and AI/LLM training and serving. Lakehouse applications can leverage any of these capabilities. Apps can also use catalogs and contribute metadata and lineage in Unity Catalog, integrate with the Lakehouse filesystem, and extend the lakehouse with custom functionality.


databricks
X
posit™

Resources

- GH Repository Link
- Sparklyr and Databricks
- Data Intelligence Platforms
- Lakehouse Apps

> databricks x posit™

Questions

<