

Crossing Bridges

Reporting on NYC Taxi data with
RStudio IDE and Databricks

James Blair & Rafi Kurlansik

posit
conf (2023)

FROM  **posit**
&  **databricks**

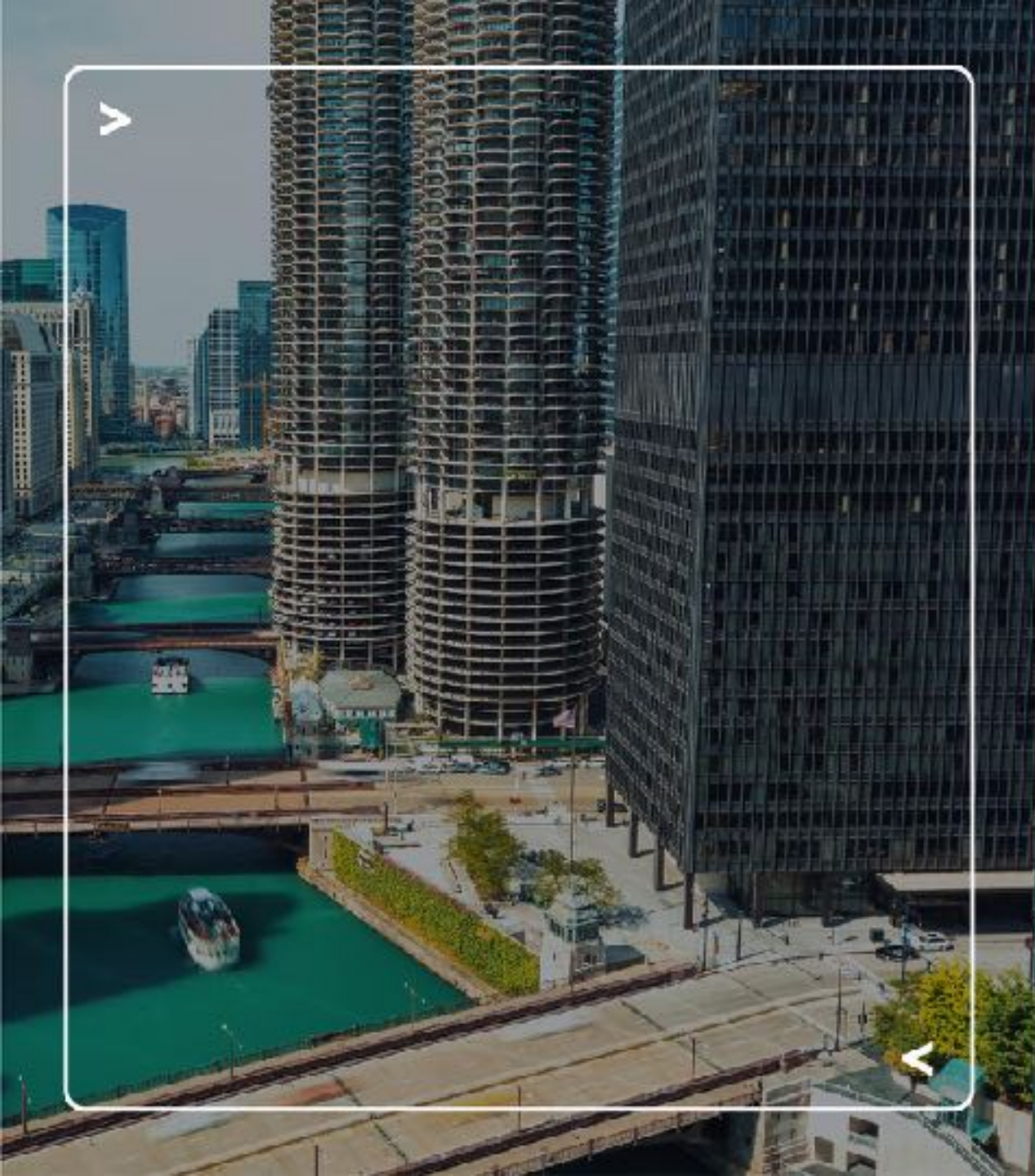




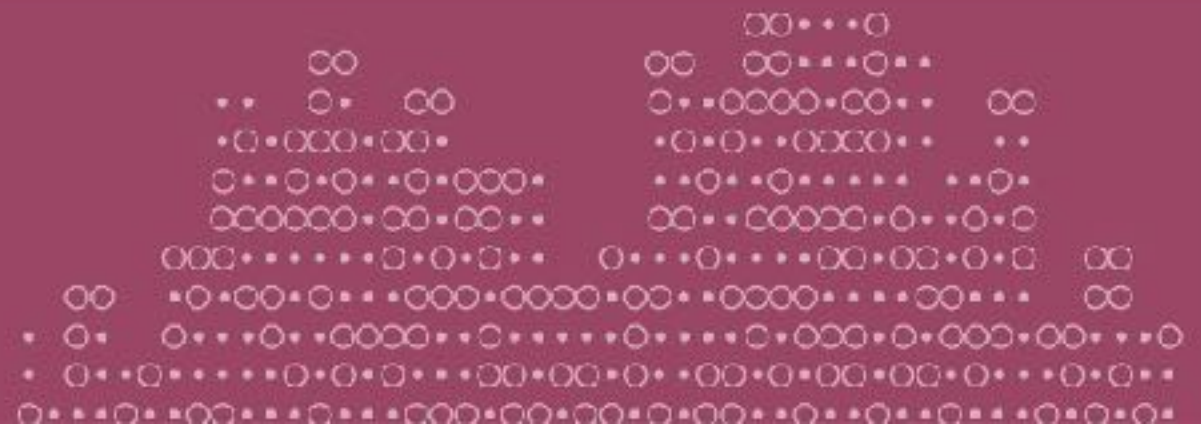
Outline

- Posit and Databricks overview
- Current and future integrations
- Demo
- Q&A





OVERVIEW



- > Posit is a **the open source data science company**. We build open source and professional tools to enable R and Python users to discover and share meaningful insights from data.

> Posit

posit Workbench

BUILD

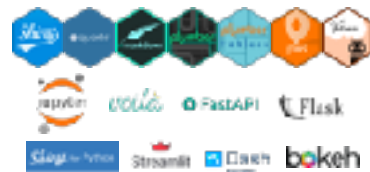
insights with your favorite open source code editors



posit Connect

SHARE

applications, notebooks, reports, visualizations, models, and APIs



posit Package Manager

CONTROL

open source packages and enable reproducibility



posit Academy

MASTER

data science skills with practice and mentorship



posit Cloud

DATA SCIENCE FROM YOUR BROWSER

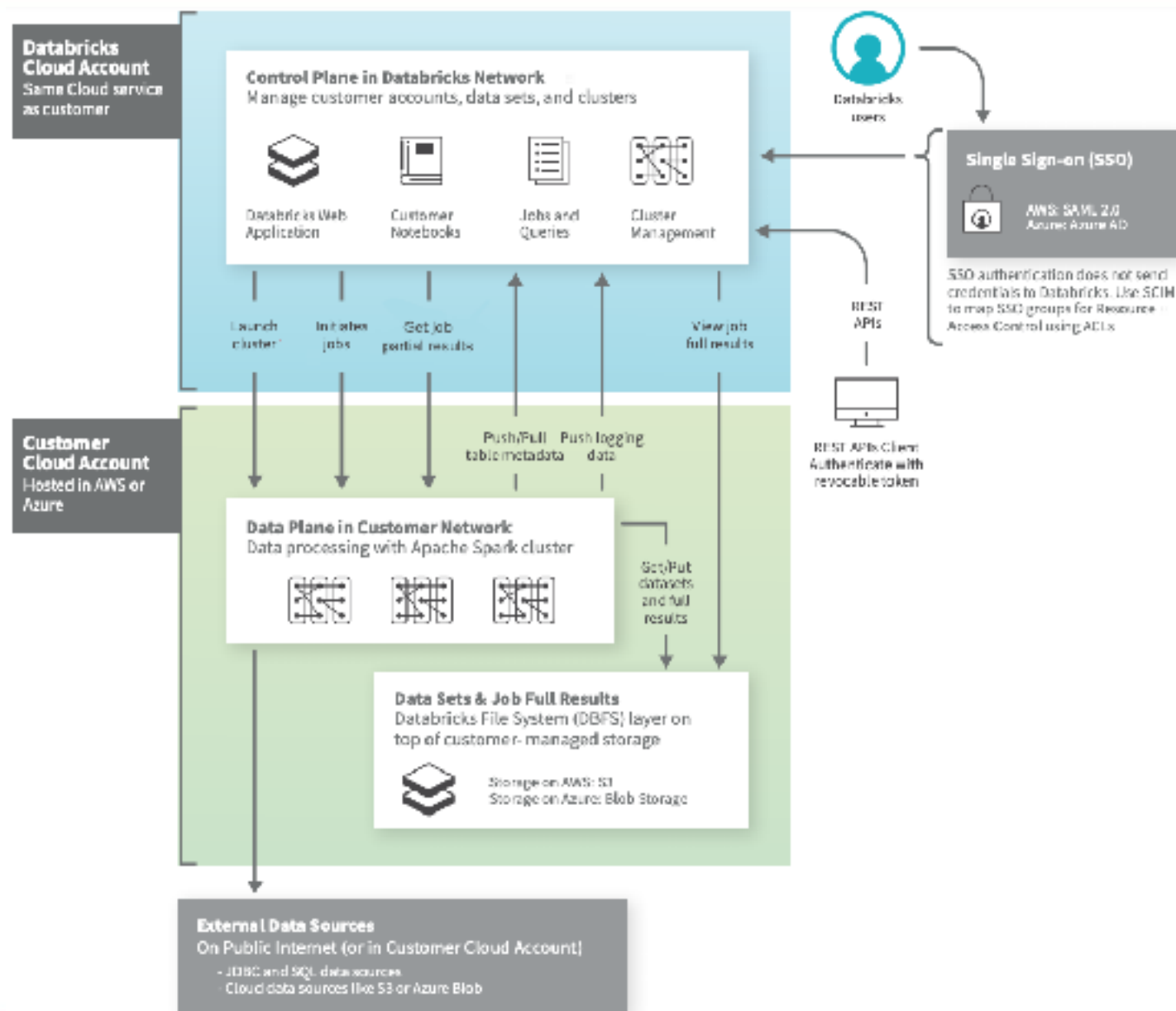
Do and share data projects in R and Python from the comfort of your web browser - no installation or configuration required



- > Databricks is a **unified, open analytics platform for building, deploying, sharing, and maintaining enterprise-grade data, analytics, and AI solutions at scale.** The Databricks Lakehouse Platform integrates with cloud storage and security in your cloud account, and manages and deploys cloud infrastructure on your behalf.



> Databricks

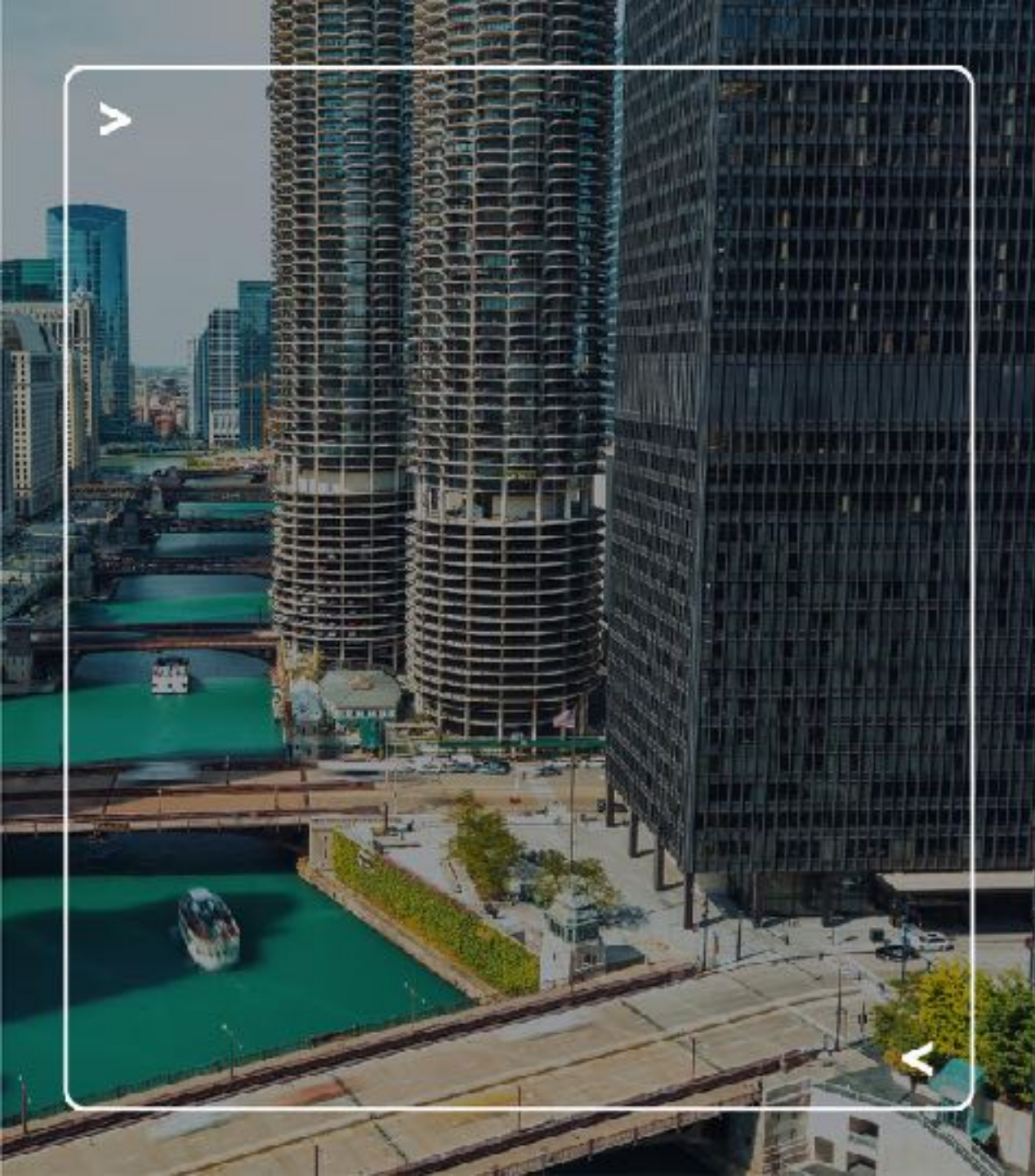


> Posit & Databricks

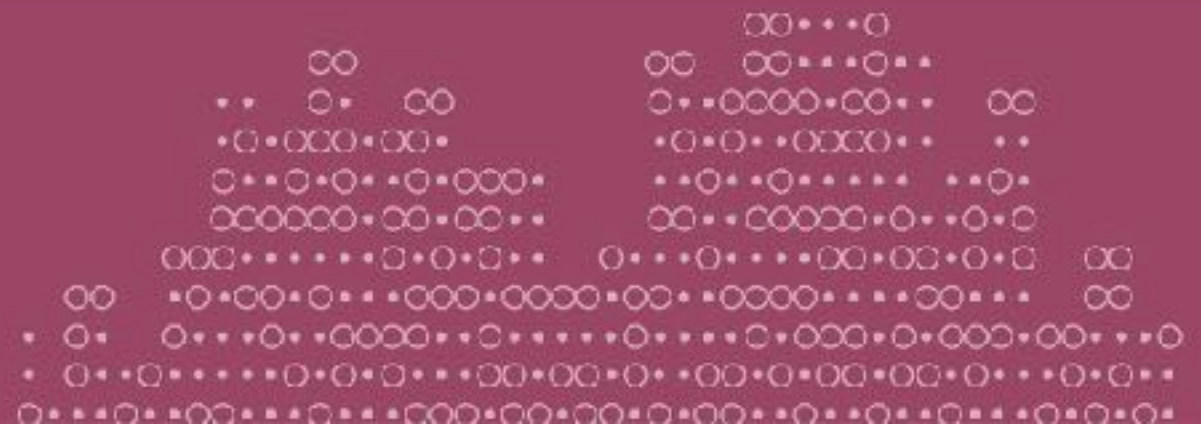
“In a partnership with Databricks, we are excited to offer R and Python users seamless access to their favorite authoring environments, such as RStudio, Jupyter, and VS Code. By bringing Posit Workbench into the Databricks Marketplace, we significantly lower the barrier for our joint customers, granting them secure data access while harnessing the security, reliability, scalability, and performance of the Databricks Lakehouse Platform.”

Tareef Kawaf
President
Posit





CURRENT AND FUTURE INTEGRATIONS



>



**Improved open
source experience
with sparklyr**

<

> **Improved open
source experience
with** sparklyr



> Improved open source experience with sparklyr



```
---  
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 the `databricks_connect` method:

```
{r}  
sc <- spark_connect(method = "databricks_connect")
```

Connect to Data

- Improved open source experience with sparklyr

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 the `databricks_connect` method:

```
{r}  
sc <- spark_connect(method = "databricks_connect")
```

Connect to Data

```
{r}  
taxi <- tbl(sc, in_catalog("samples", "nyctaxi", "trips"))  
count(taxi)  
head(taxi)
```



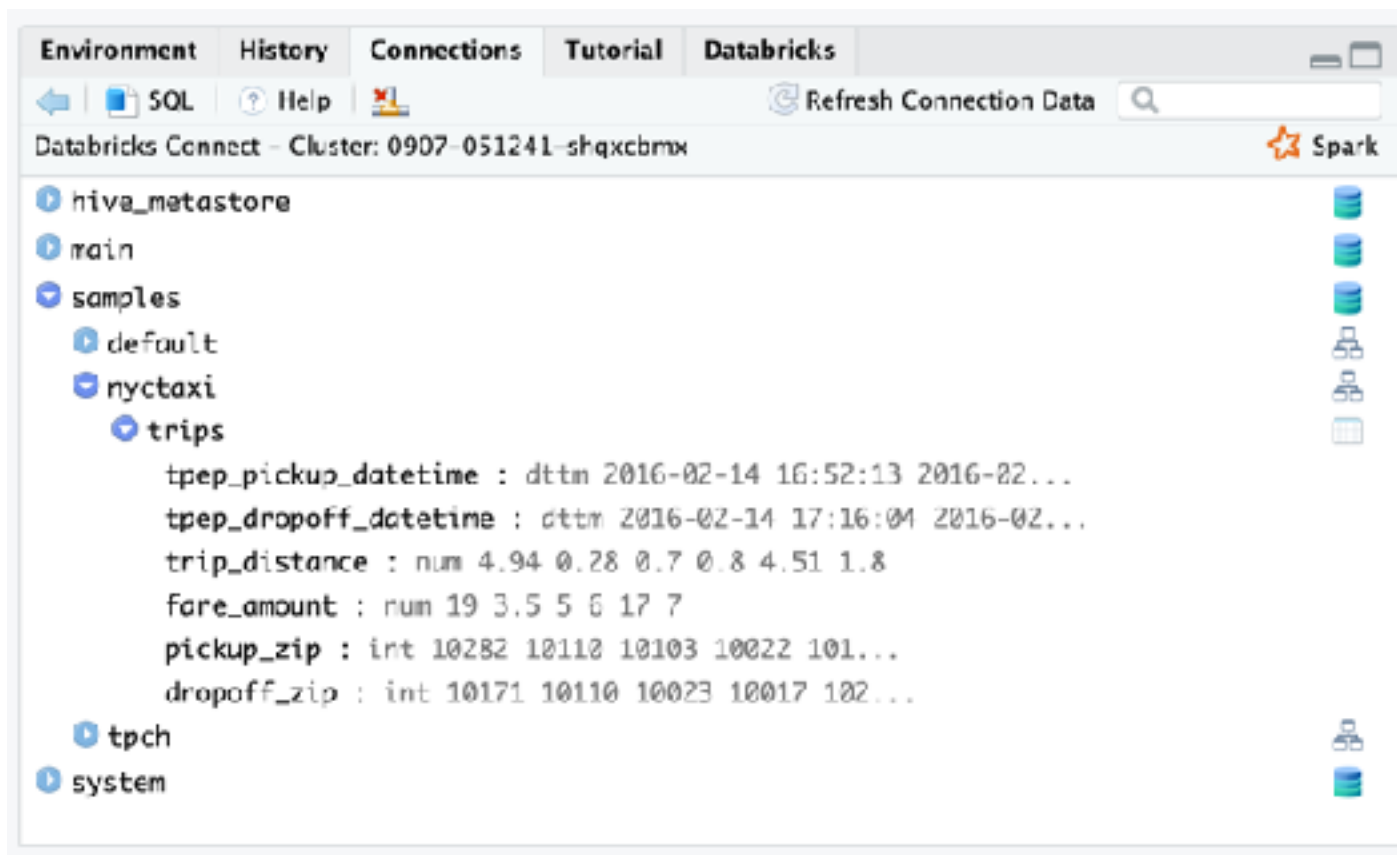
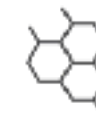
Source: SQL [?? x 6]

Database: spark_connection

tpep_pickup_datetime <S3: POSIXct>	tpep_dropoff_datetime <S3: POSIXct>	trip_distance <dbl>	fare_amount <dbl>
2016-02-14 16:52:13	2016-02-14 17:16:04	4.94	19.0
2016-02-04 18:44:19	2016-02-04 18:46:00	0.28	3.5
2016-02-17 17:13:57	2016-02-17 17:17:55	0.70	5.0
2016-02-18 10:36:07	2016-02-18 10:41:45	0.80	6.0
2016-02-22 14:14:41	2016-02-22 14:31:52	4.51	17.0
2016-02-05 06:45:02	2016-02-05 06:50:26	1.80	7.0

6 rows | 1-4 of 6 columns

> Improved open
source experience
with sparklyr





**Improved open
source experience
with sparklyr**



**Improved
Databricks
experience within
the RStudio IDE**



> **Improved
Databricks
experience within
the RStudio IDE**



> Improved
Databricks
experience within
the RStudio IDE



Environment	History	Connections	Tutorial	Databricks							
⇅	Name	⇅	Runtime	⇅	Unity Catalog	⇅	Active memory	Active cores	Creator		
✗	Secondary Cluster		14.0 ML		Yes				james@posit.co	▶	▼
✓	Test Cluster		14.0 ML		Yes		30.5 GB	4	james@posit.co	■	▼

> Improved
Databricks
experience within
the RStudio IDE



Environment	History	Connections	Tutorial	Databricks							
⇅	Name	⇅	Runtime	⇅	Unity Catalog	⇅	Active memory	Active cores	Creator		
✗	Secondary Cluster		14.0 ML		Yes				james@posit.co	▶	▼
✓	Test Cluster		14.0 ML		Yes		30.5 GB	4	james@posit.co	■	▼



**Improved open
source experience
with sparklyr**



**Improved
Databricks
experience within
the RStudio IDE**



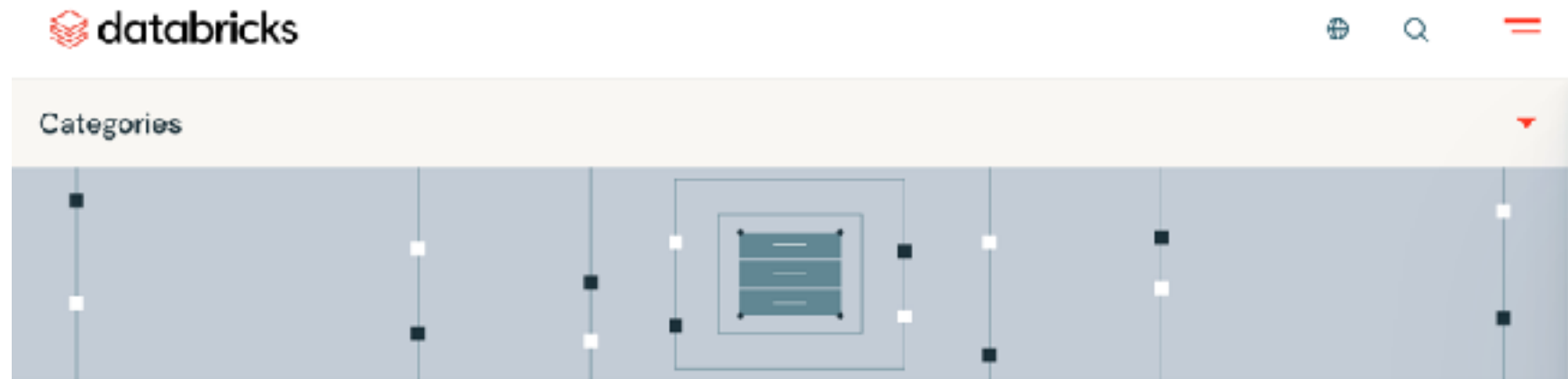
**Posit Workbench
available in
Lakehouse Apps**



➤ **Posit Workbench**
available in
Lakehouse Apps



➤ Posit Workbench
available in
Lakehouse Apps



Introducing Lakehouse Apps

Build Secure Native Data and AI Applications
Without Compromise



> **Posit Workbench
available in
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.

The Data Application Dilemma

Data and AI applications are among the fastest-growing software categories, and the growth in generative AI and large language models (LLMs) has accelerated that trend. Developers are developing innovative data and AI solutions every day, and development teams in companies are



> Posit Workbench available in Lakehouse Apps

"Developers use Retool to build and deploy custom business software with an easy-to-use graphical user interface that sits on top of any data stack. Retool's native integration with Databricks allows builders to create applications that get information into the hands of those that need it most — without compromising on speed, security, or reliability. Data engineers, data scientists, software developers and end users of data applications can now all benefit from our powerful offering listed on the Databricks Marketplace."

– **Mark Schaaf, COO, Retool**

"In a partnership with Databricks, we are excited to offer R and Python users seamless access to their favorite authoring environments, such as RStudio, Jupyter, and VS Code. By bringing Posit Workbench to the Databricks Marketplace, we significantly lower the barrier for our joint customers, granting them secure data access while harnessing the security, reliability, scalability, and performance of the Databricks Lakehouse Platform. As we move forward, we eagerly anticipate supporting over 10,000 Databricks customers with the advanced publishing capabilities that our loyal Posit Connect users currently enjoy."

– **Tareef Kawaf, President of Posit Software, PBC**

"Immuta's Data Security Platform enables organizations to unlock value from their cloud data. With Lakehouse Apps, we can bring our offering closer to the data managed in Databricks and enable our joint customers to get started faster. With the flexibility of Databricks Apps, customers can reuse existing code and skills, and build a closer integration with the Lakehouse platform."

– **Mo Plassnig, Chief Product Officer, Immuta**

"Plotly's Dash Enterprise is the leading platform for building production-grade data applications. With 3200 common organizations using Plotly & Dash and Databricks, Plotly is a major Databricks



➤ **Posit Workbench**
available in
Lakehouse Apps





**Improved open
source experience
with sparklyr**



**Improved
Databricks
experience within
the RStudio IDE**

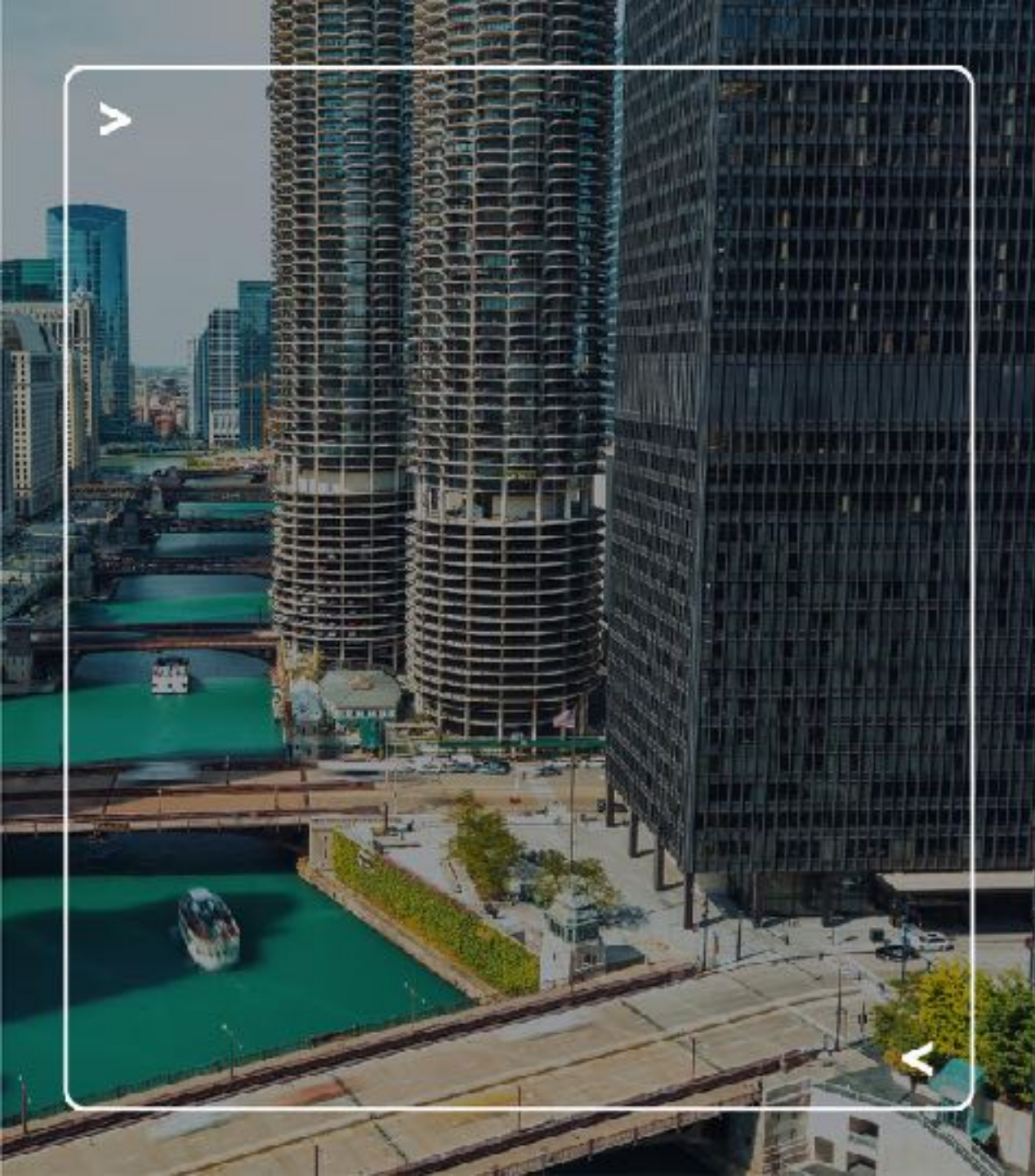


**Posit Workbench
available in
Lakehouse Apps**

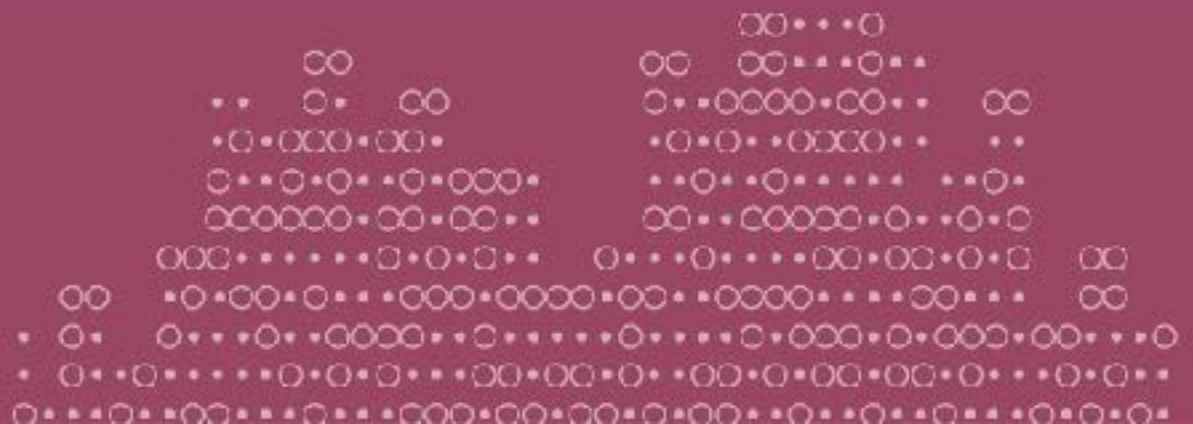


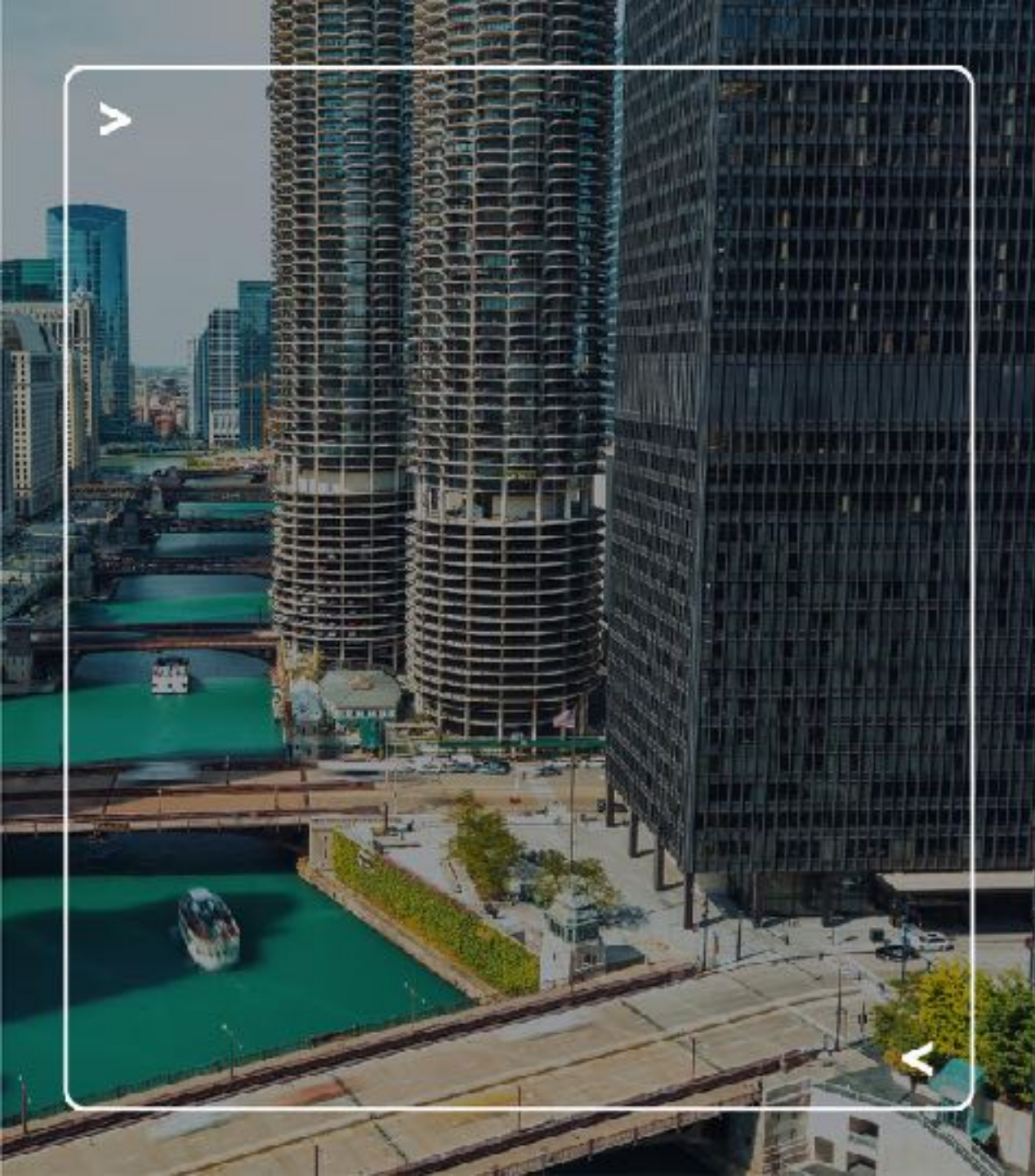
**Additional Posit
products on
Lakehouse Apps**



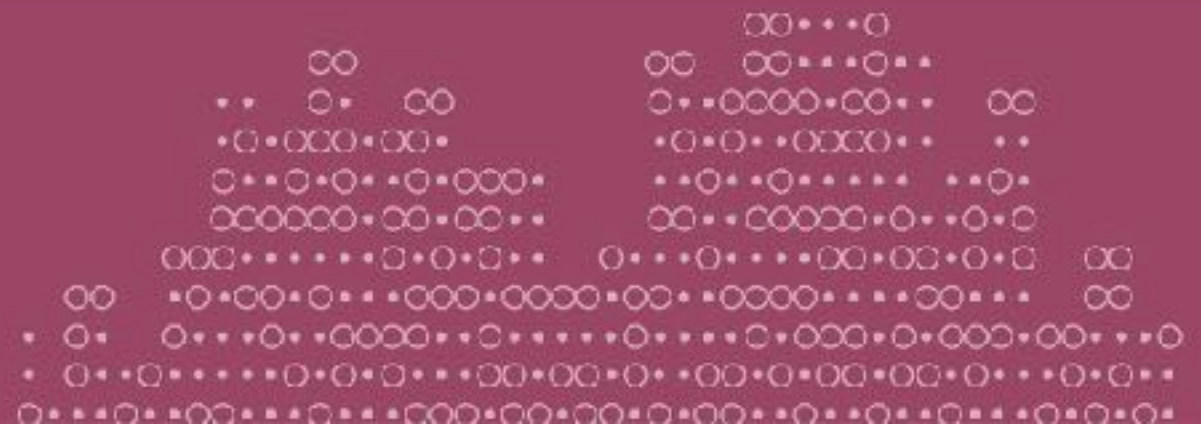


DEMO





Q&A



posit
conf(2023)

Thank you.