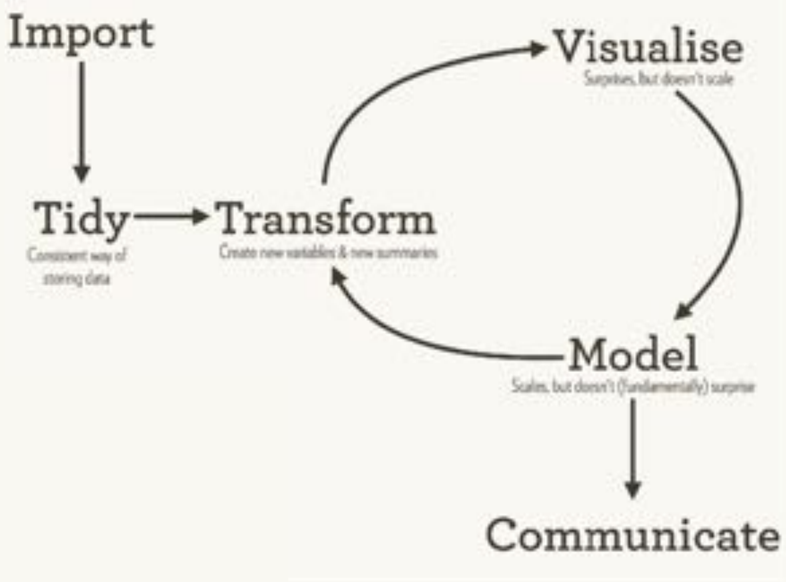
PROFESSIONAL R TOOLING AND INTEGRATION

ANALYTIC ADMINISTRATION FOR R

RStudio

DATA SCIENCE WORKFLOW

Program



Artifacts

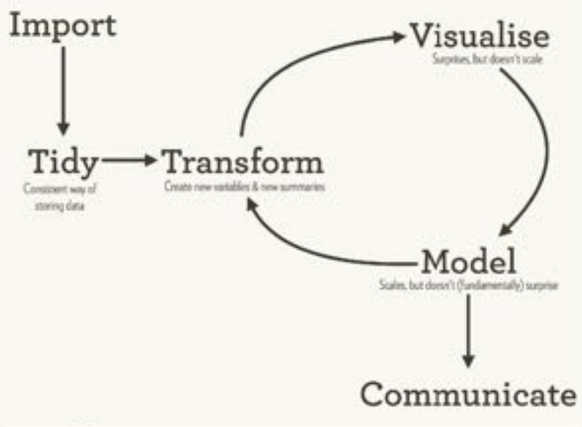
Reports
Dashboards
Presentations
Apps
API's
Data Products

DATA SCIENCE IN THE ENTERPRISE

Enterprise Environment

Data Sources

Program



Artifacts

Reports
Dashboards
Presentations
Apps
API's
Data Products

Deploy

Manage

Scale

Integration

BI

In-Database

Reporting

Web

Workflows

Security | Control | Access | Validation | API's

Programming Languages
Packages and dependencies
Operating Systems

ENTERPRISE TOOLING

Server based

Cloud or on-premises (or both)

IT managed

Problem:

IT doesn't know R

RStudio®

R IN THE ENTERPRISE

Legitimacy

Recognize R as an analytic standard

Competencies

Understand and manage R tooling

Adoption

Rely on integrated R based solutions

RStudio®

Legitimacy

Recognize R an Analytic Standard

BRING R THROUGH THE FRONT DOOR

Show the value of R

R Markdown reports, dashboards, and presentations

Shiny apps

Plumber API's

Other data products

Get official recognition

Get resources

RStudio®

SOME PERSONAL HISTORY

	Consumer Research (c. 2004)	Advertising Network (c. 2008)	Client Services (c. 2011)
R Users	Me	Team	Global
R Tools	Desktop	Server	Cluster
Relationship with IT	None	Support	Partnership
Approved By	Manager	Owner	Corporate Process
Groups Influenced	Department	Organization	Customers

GETTING STARTED

Build a sandbox / lab box

Proof of concept

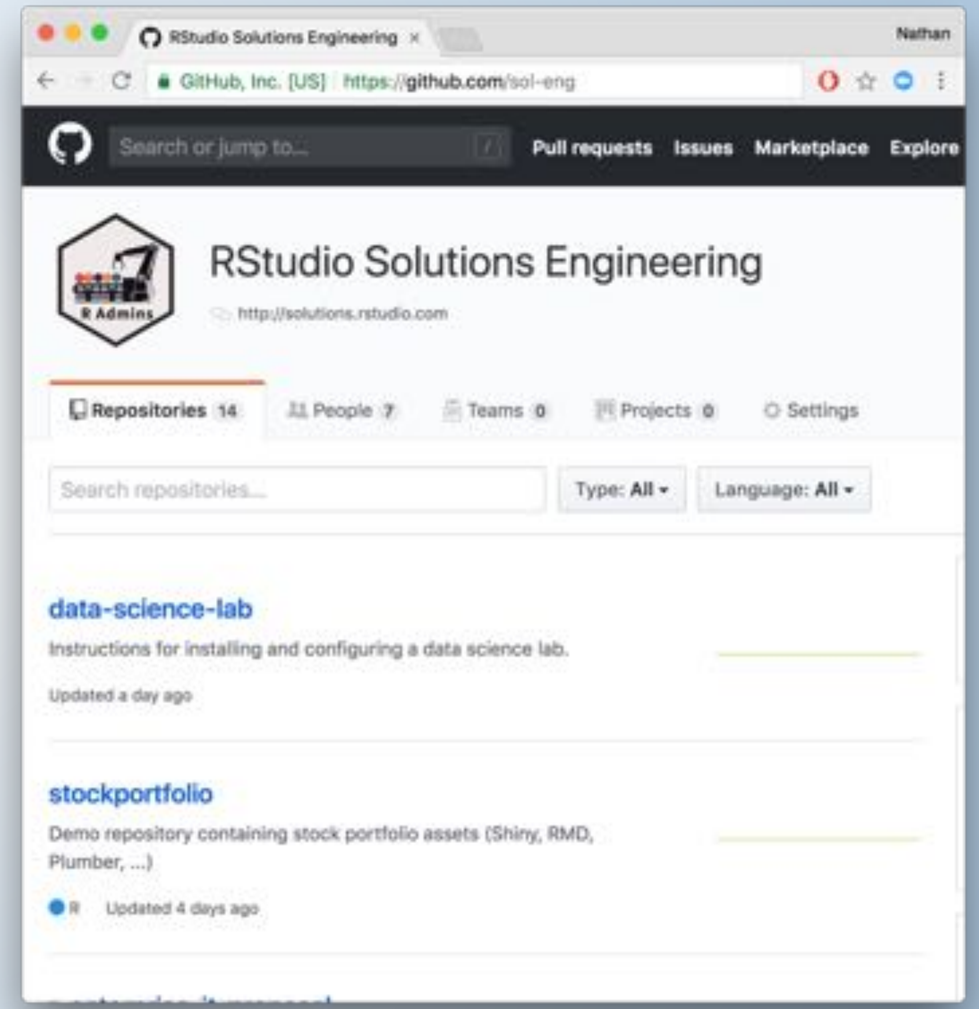
Show the value of R

R Markdown reports, dashboards, and presentations

Shiny apps

Plumber API's

Other data products



RStudio

Competencies

Understand and Manage R tooling

RSTUDIO

We build tools for doing data science

Committed to open source

We hire great people

We sell professional products

RSTUDIO PRO PRODUCTS HELP YOU…

Access Spark from R

Connect you to your databases

Deploy your Shiny applications

Automate and distribute your reports

Secure, scale, and support your environment

RStudio®

RSTUDIO SERVER PRO

Develop at scale

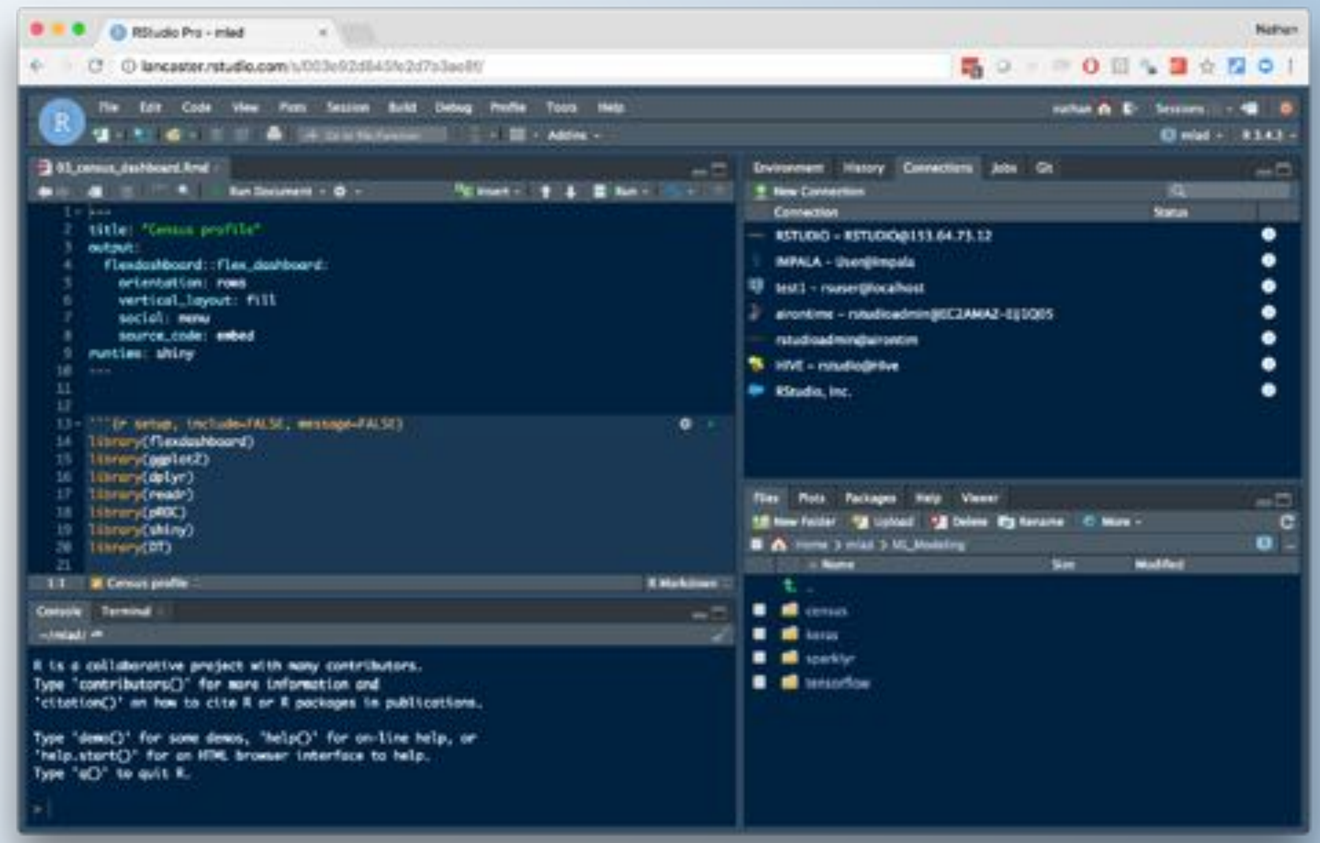
Runs securely in your environment

Horizontally scalable

Connects to your databases

Easy to administration

Supported



RStudio®

RSTUDIO CONNECT

Run your artifacts at scale

On premises solution

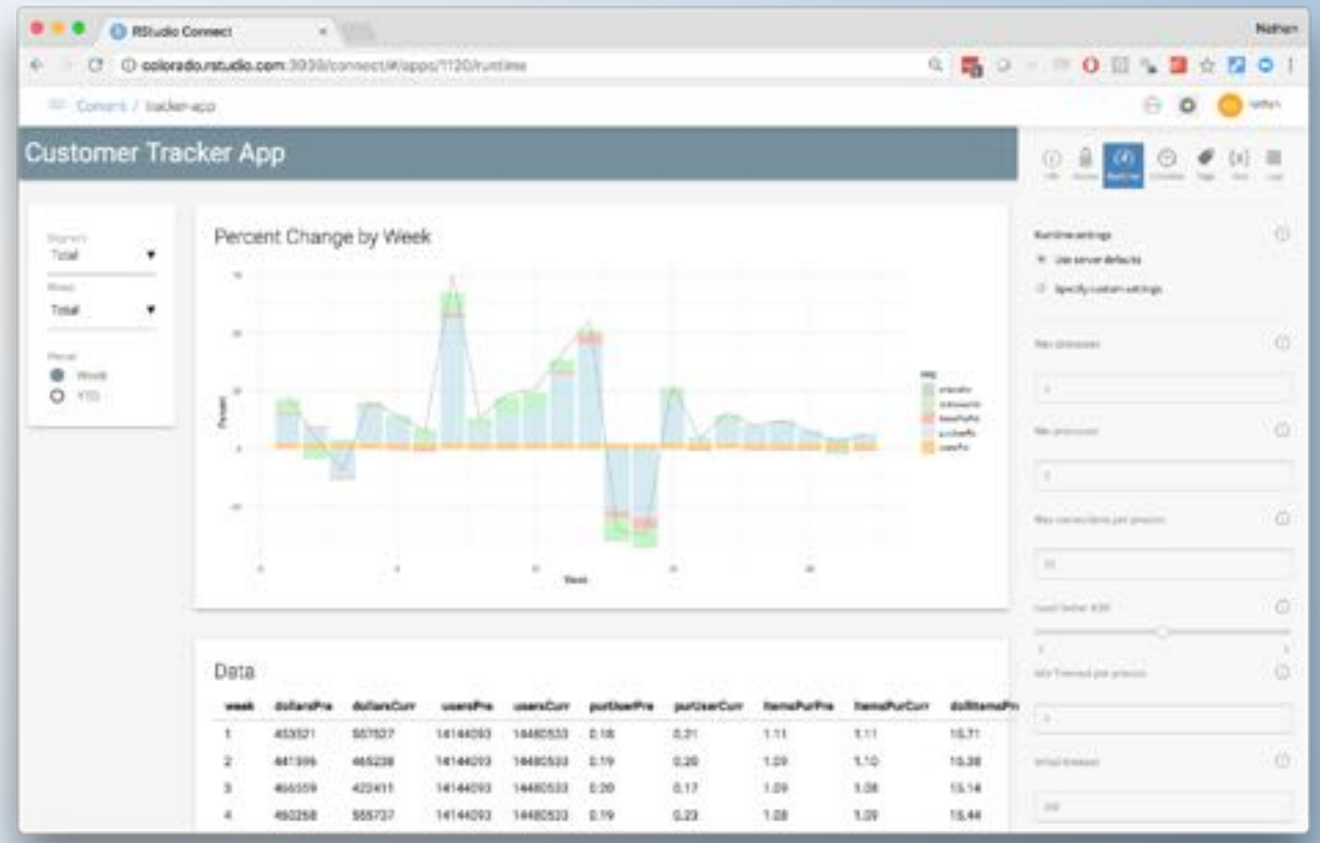
Push button publishing

Self managed content

Scheduled updates & distribution

Horizontally scalable

Secure and supported



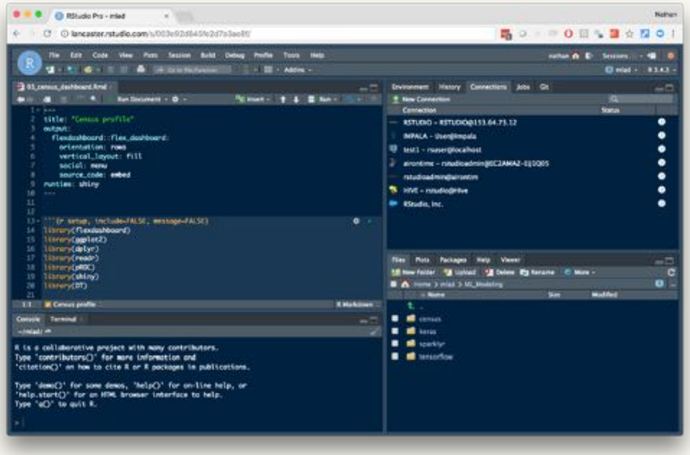
DATA SCIENCE LAB IN THE ENTERPRISE

Enterprise Environment

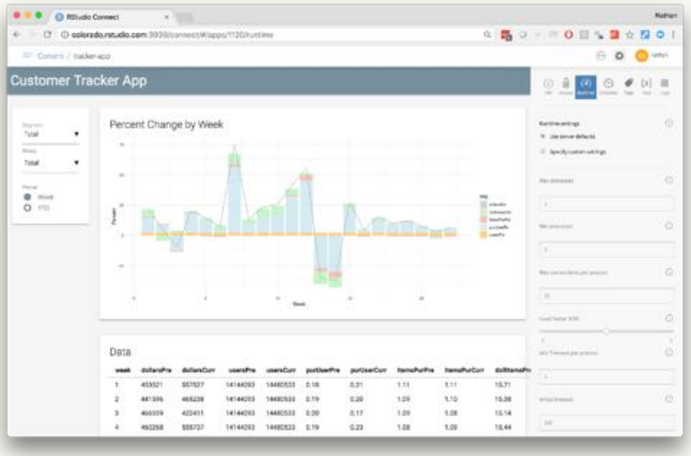
Data Sources

Data Science Lab

RStudio Server Pro
Build Artifacts



RStudio Connect
Run Artifacts



BI

In-Database

Reporting

Web

Workflows

API's

Programming Languages
Packages and dependencies
Operating Systems

Deploying Artifacts

DEPLOYMENT APPROACH

1. Build the bundle

Source Code

Data Files

Manifest

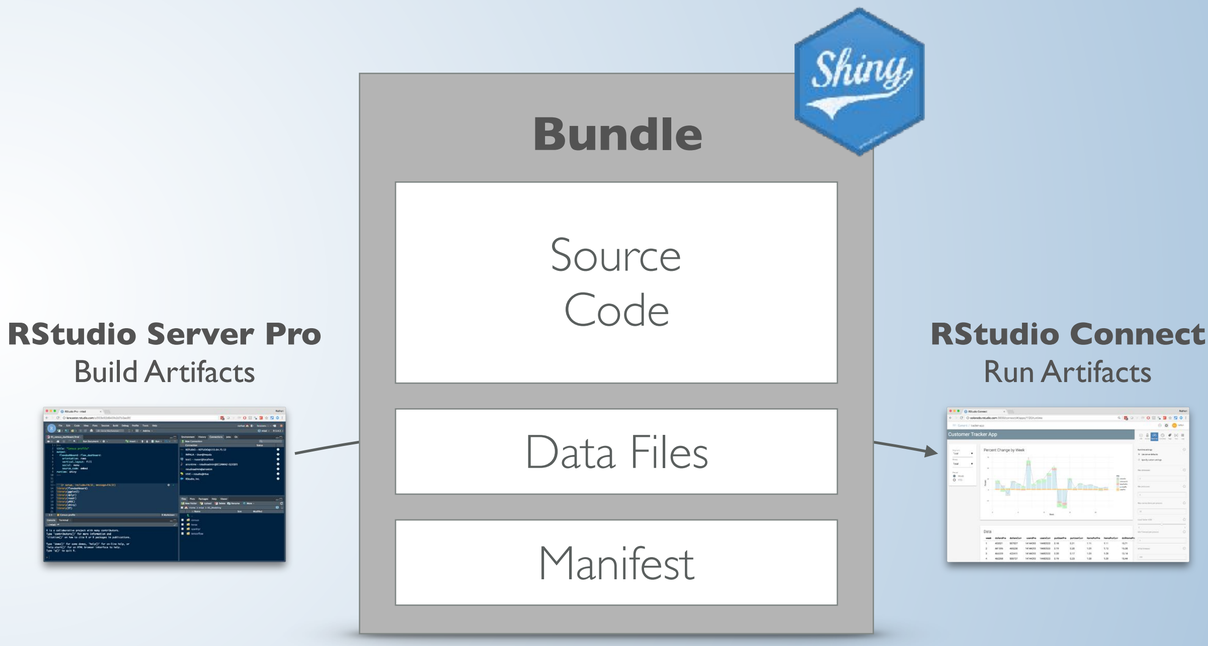
(Metadata and environment)

2. Push the bundle

3. Deploy the bundle

Parse the manifest

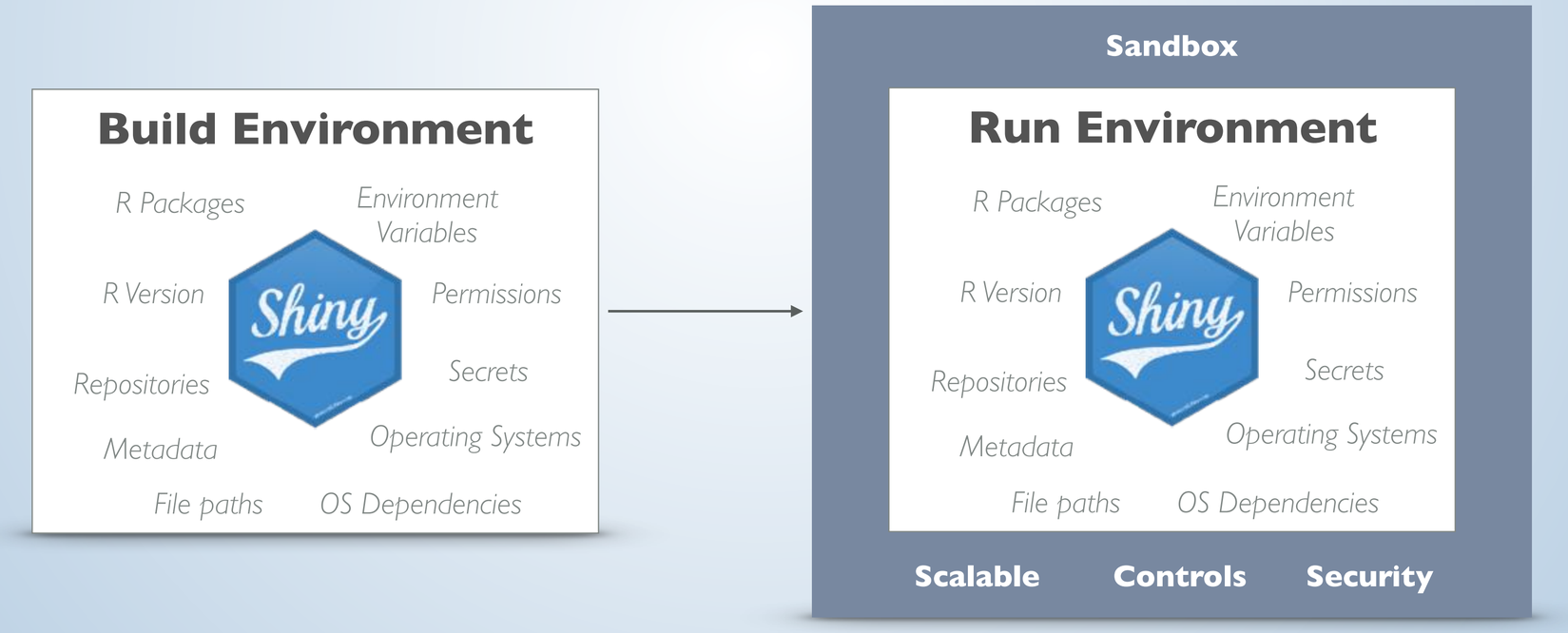
Packrat restore



RStudio

DEPLOYMENT GOALS

1. Reproduce the development environment
2. Productize the run environment



RStudio Server Pro

RStudio Connect

EXAMPLE: STOCK PORTFOLIO

Adoption

Rely on integrated R based solutions

CONNECT TO YOUR DATA

RStudio Data Connectors

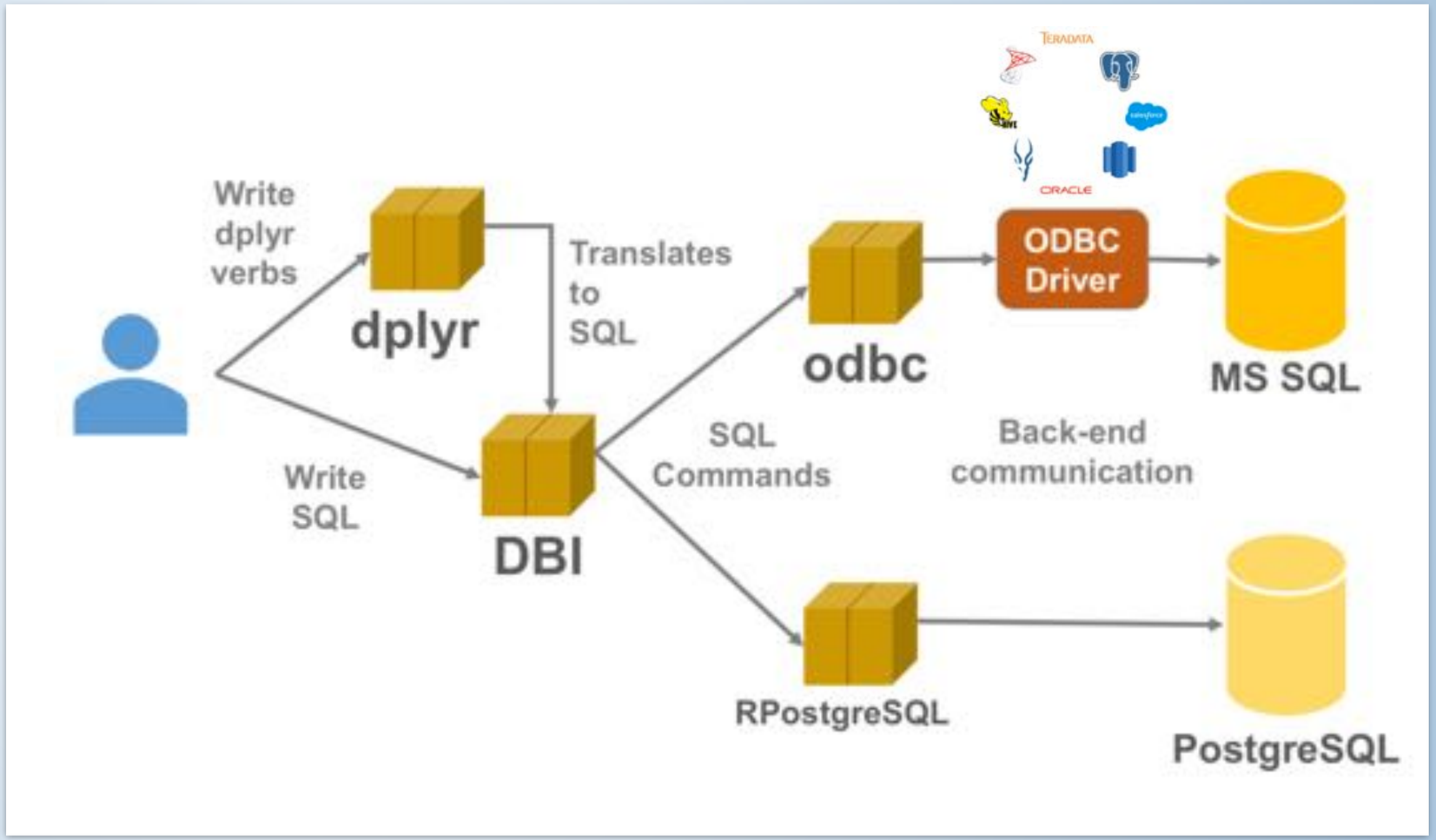
Work with all pro products

Fully supported

Easy to use

Easy to set up

Adheres to enterprise standards



RStudio

COMMUNICATE INSIGHTS

Shiny Apps

Interactive

Delivered via the web

R Markdown Reports

Snapshot in time

Deliver via email or web

	Shiny Apps	**R Markdown Reports**
User Experience	Interactive (Dynamic)	Snapshot (Batch)
Audience Delivery	Web (Pull audience)	Email or Web (Push to audience)

RStudio

AUTOMATE DATA PRODUCTS

Schedule execution

Monitor logs

Track render history

Access source versions

RStudio®

CALL R AS A WEB SERVICE

What is an REST based API?

Software that lets computer systems communicate

Humans use GUI's

Machines use API's

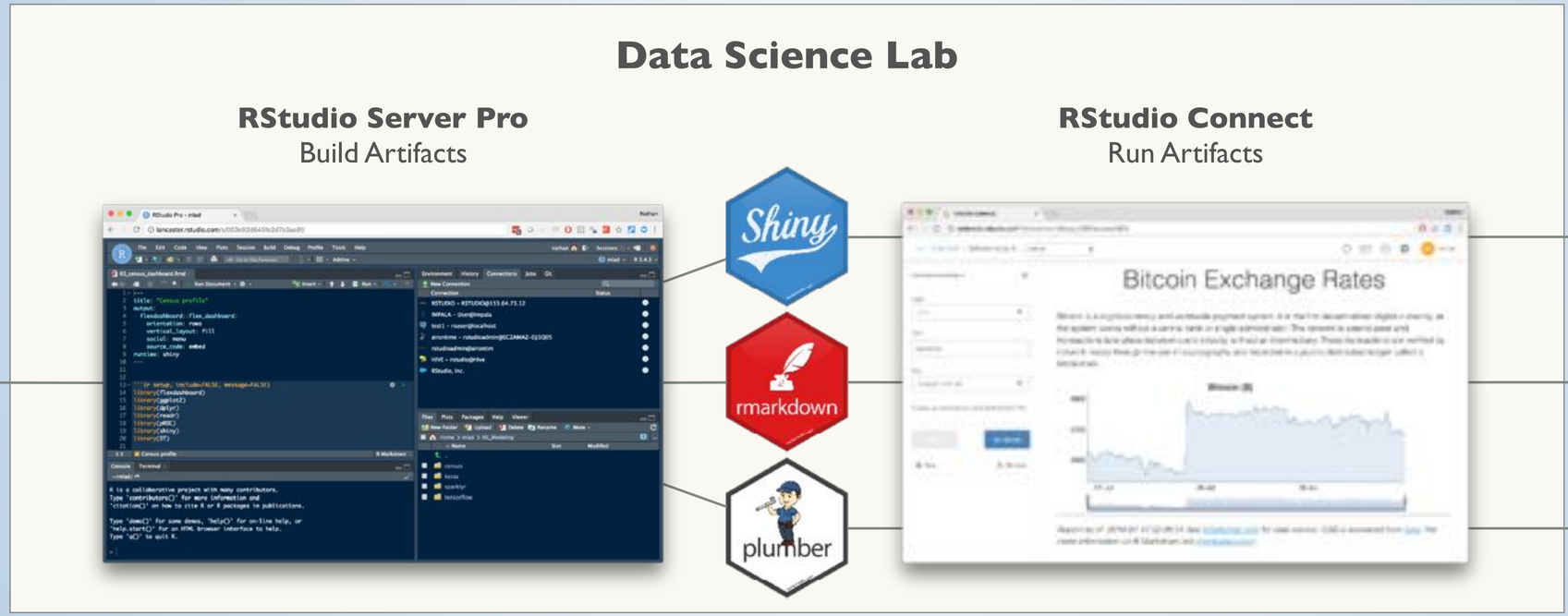
Why are they important?

Solve the handoff problem

Integrate R into workflows

RStudio

EXAMPLE: BITCOIN ARTIFACTS



Database

Integration

Webserver

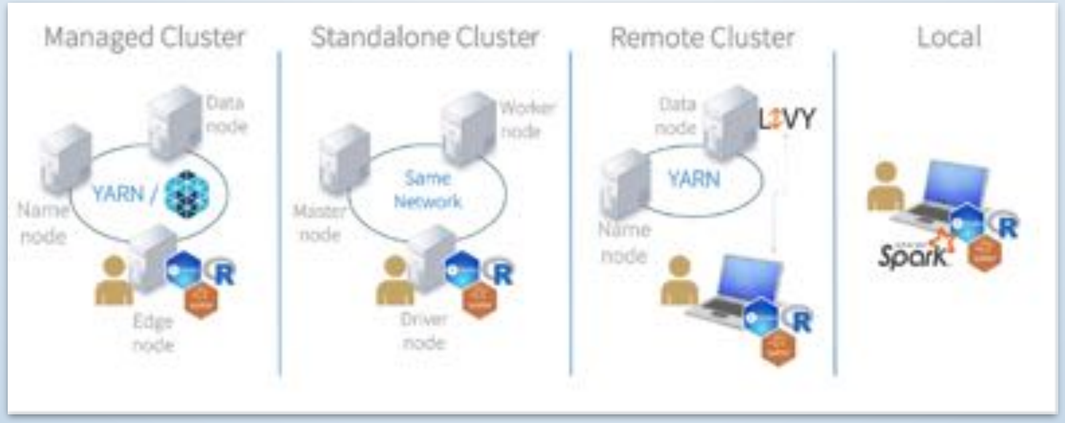
Email

REST API

SPARKLYR

Use the full power of R with Apache Spark

Integrated with the RStudio IDE
dplyr backend
Works with Hadoop



RStudio

PYTHON INTEROPERABILITY

Reticulate

Call Python from R

Translate between R and Python objects

Flexible binding to different version of Python

TensorFlow

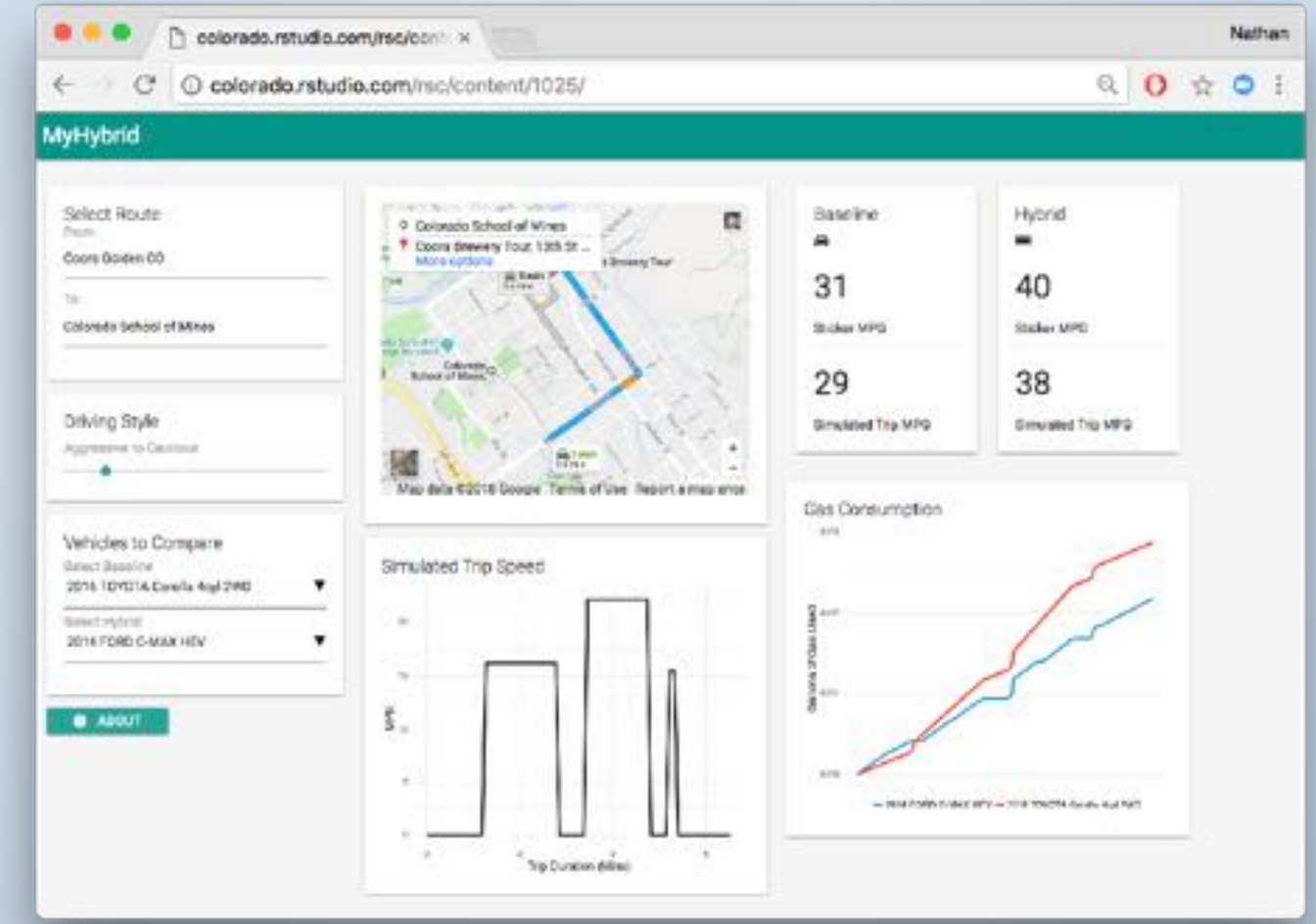
Open source software library for Machine Intelligence

Keras

High level interface for neural networks

RStudio

EXAMPLE: RETICULATED SHINY APP



SHARED INFRASTRUCTURE FOR DATA SCIENCE

A new data science runtime

Powered by Apache Arrow

RStudio®

The R Admin

THE R ADMIN

A **data scientist** (who knows R) and works with **IT/Ops** to:

1. Establish R as an analytic standard
2. Promote best practices for R tooling
3. Make R indispensable to the organization

RStudio®

BEST PRACTICES AND RECOMMENDATIONS

Use a Modern OS and Browser

Upgrade R yearly

Build R from source

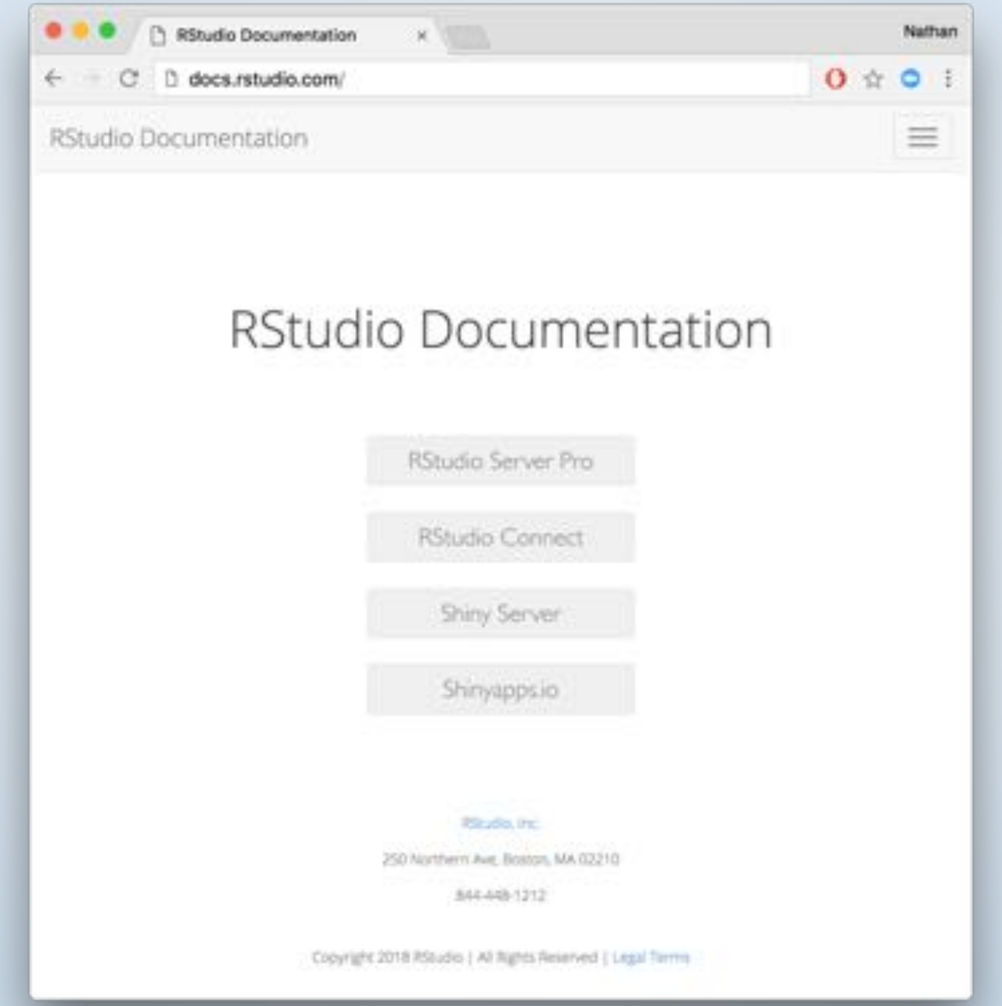
Size instance with plenty of RAM

RStudio®

THE R ADMIN HAS TO BE RAD!

They have a multiplier effect in their organization

DOCUMENTATION



RStudio®

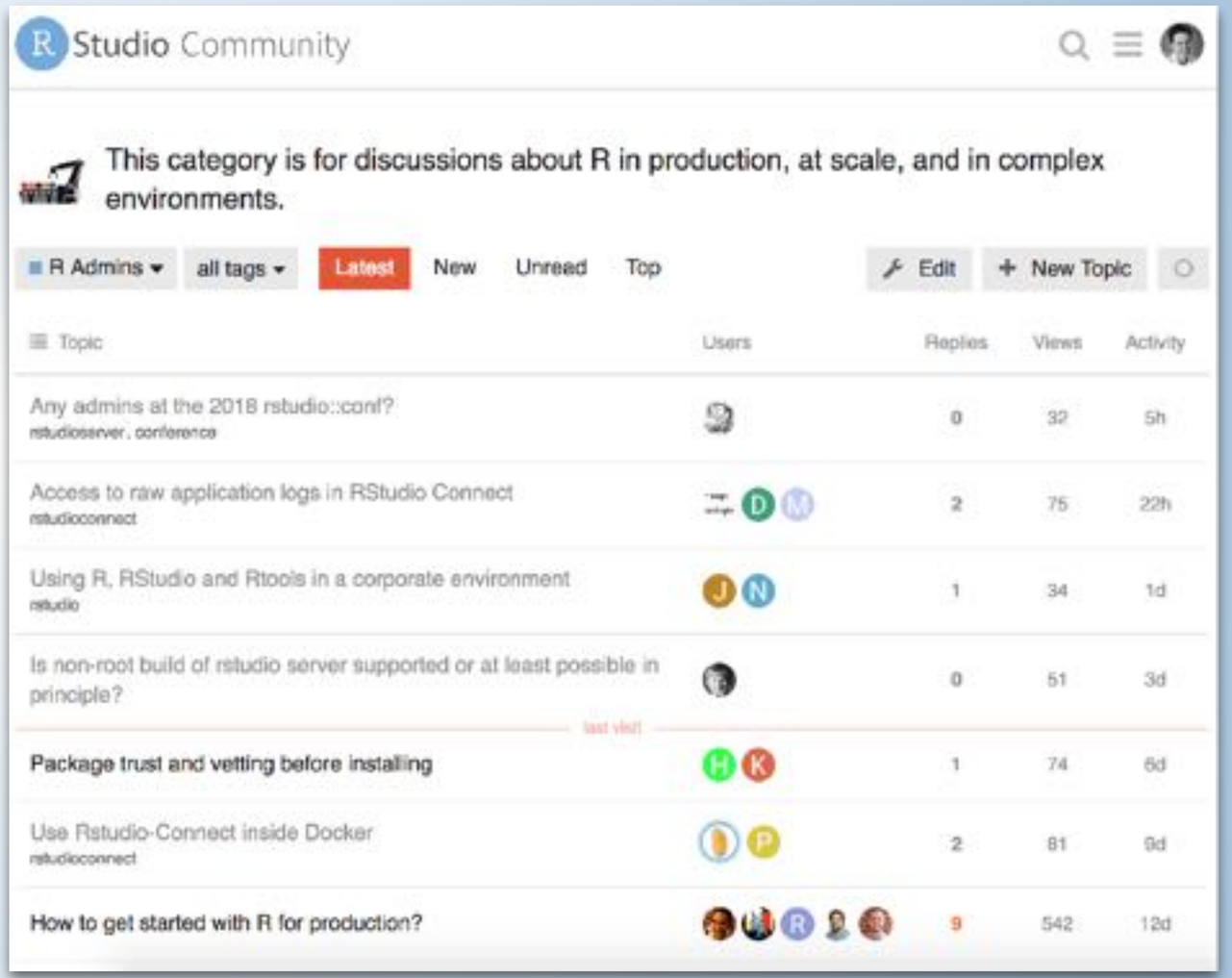
OTHER RESOURCES

community.rstudio.com

Developer Blog

R Views Blog

rstudio::conf
Jan 15-18, 2019
Austin, TX



SUMMARY

R for the Enterprise Roadmap	**R Admin** Call to action
Legitimacy Recognize R as an analytic standard	**Demonstrate the value of R** Build a sandbox / lab box Get resources
Competencies Understand and manage R processes	**Build a data science lab in production** Promote best practices
Adoption Rely on R based solutions	**Make R indispensable** Integrate R artifacts into other systems

THANK YOU

nathan@rstudio.com

@nwstephens

https://github.com/sol-eng/data-science-lab

RStudio