

---

5 May 2023  
R Exchange 2023

---

# Enabling data science

Key considerations for effective R environments

Dr Uli Muellner – Managing Director Epi-interactive

---

# What we'll cover...

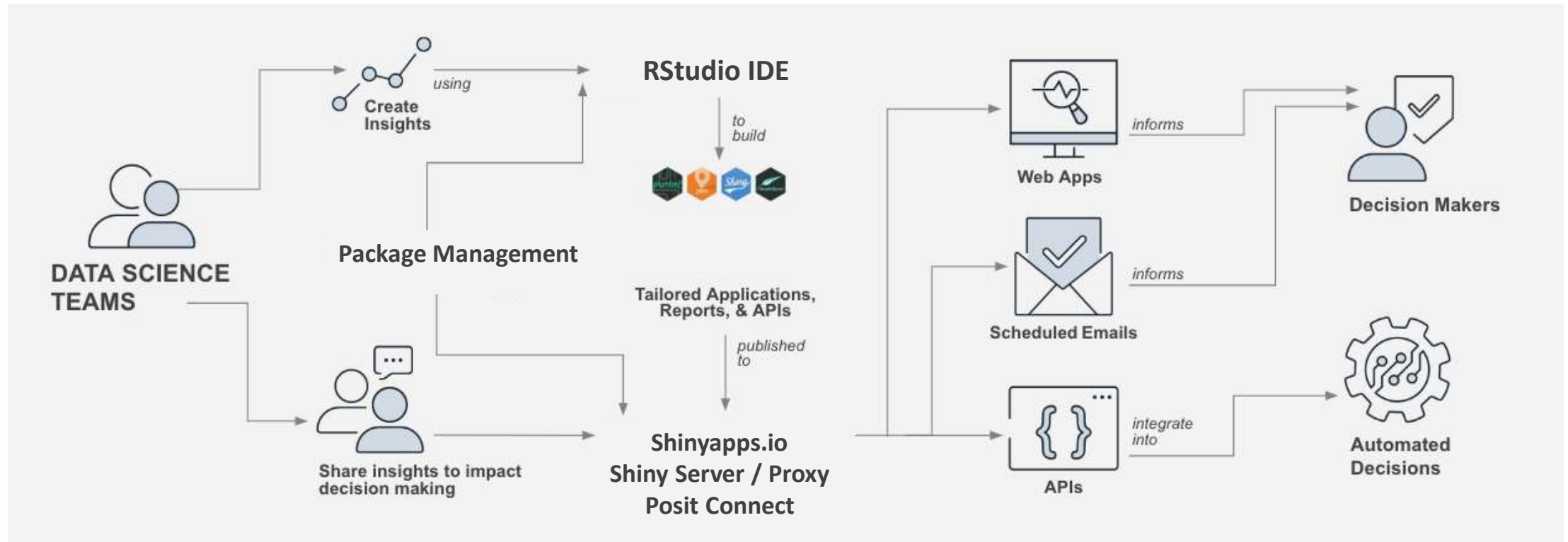
A bird's-eye view on R environments

Key considerations

1. Keep IT happy
2. Facilitate collaboration
3. Allow for scale



# R environments



---

Keep IT happy

---

# IT security

## A data scientist's "wish list"

- Fast access to new R packages and versions
- (Open) access to required data sources
- Easy publishing process

**KEY: Set up processes!**

## IT security

- Checking for vulnerabilities
- Security screening
- Restrict access to internal data sources, apply access controls
- Centralised controls

---

# Scanning of R Packages

- CRAN offers some protection!
- Scan installed packages (new / new versions)
- oysteR package for vulnerability scanning
- Check against OSS Index

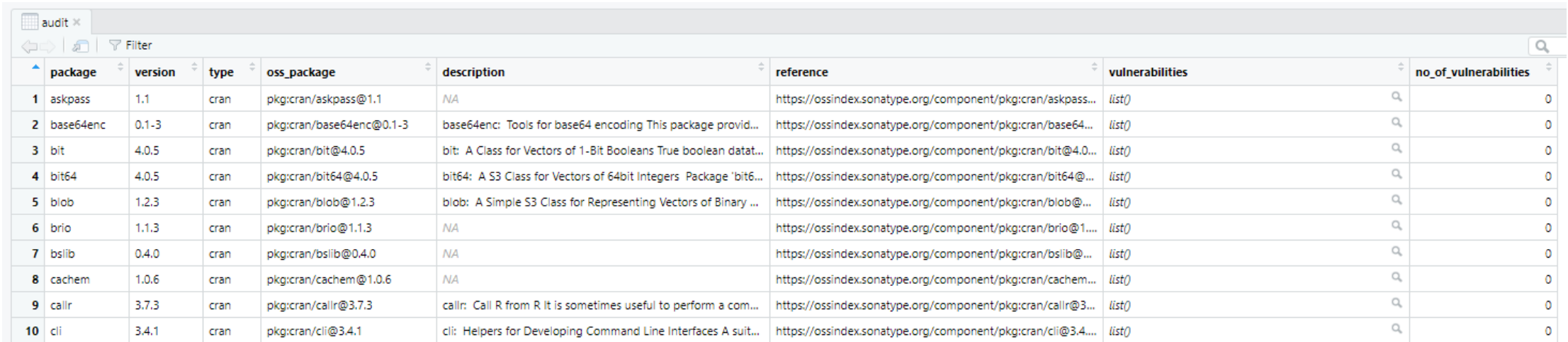
Add to pipeline:

- Configuration file with R packages on Git
- Auto-update of “approved” packages
- Add new packages after scanning



# Scanning of R Packages

Full audit:



The screenshot shows a web-based interface for an audit tool. At the top, there is a tab labeled 'audit' and a search bar. Below the search bar is a table with the following columns: package, version, type, oss\_package, description, reference, vulnerabilities, and no\_of\_vulnerabilities. The table lists 10 R packages from the CRAN repository, including askpass, base64enc, bit, bit64, blob, brio, bslib, cachem, callr, and cli. Each row includes a package name, its version, the type (all are 'cran'), the OSS package identifier, a brief description, a reference URL to the OSS index, a link to view vulnerabilities, and the count of vulnerabilities (all are 0).

package	version	type	oss_package	description	reference	vulnerabilities	no_of_vulnerabilities
1 askpass	1.1	cran	pkg:cran/askpass@1.1	NA	https://ossindex.sonatype.org/component/pkg:cran/askpass...	list()	0
2 base64enc	0.1-3	cran	pkg:cran/base64enc@0.1-3	base64enc: Tools for base64 encoding This package provid...	https://ossindex.sonatype.org/component/pkg:cran/base64...	list()	0
3 bit	4.0.5	cran	pkg:cran/bit@4.0.5	bit: A Class for Vectors of 1-Bit Booleans True boolean datat...	https://ossindex.sonatype.org/component/pkg:cran/bit@4.0...	list()	0
4 bit64	4.0.5	cran	pkg:cran/bit64@4.0.5	bit64: A S3 Class for Vectors of 64bit Integers Package 'bit6...	https://ossindex.sonatype.org/component/pkg:cran/bit64@...	list()	0
5 blob	1.2.3	cran	pkg:cran/blob@1.2.3	blob: A Simple S3 Class for Representing Vectors of Binary ...	https://ossindex.sonatype.org/component/pkg:cran/blob@...	list()	0
6 brio	1.1.3	cran	pkg:cran/brio@1.1.3	NA	https://ossindex.sonatype.org/component/pkg:cran/brio@1....	list()	0
7 bslib	0.4.0	cran	pkg:cran/bslib@0.4.0	NA	https://ossindex.sonatype.org/component/pkg:cran/bslib@...	list()	0
8 cachem	1.0.6	cran	pkg:cran/cachem@1.0.6	NA	https://ossindex.sonatype.org/component/pkg:cran/cachem...	list()	0
9 callr	3.7.3	cran	pkg:cran/callr@3.7.3	callr: Call R from R It is sometimes useful to perform a com...	https://ossindex.sonatype.org/component/pkg:cran/callr@3...	list()	0
10 cli	3.4.1	cran	pkg:cran/cli@3.4.1	cli: Helpers for Developing Command Line Interfaces A suit...	https://ossindex.sonatype.org/component/pkg:cran/cli@3.4....	list()	0

Vulnerabilities:



The screenshot shows a web-based interface for an outcomes tool. At the top, there is a tab labeled 'outcomes' and a search bar. Below the search bar is a table with the following columns: package, version, type, oss\_package, description, reference, cvss\_id, cvss\_title, cvss\_description, and cvss\_score. The table lists 1 vulnerability for the 'haven' package. The vulnerability details include the CVSS ID (sonatype-2021-1261), CVSS title (sonatype-2021-1261), CVSS description (1 vulnerability found), and CVSS score (1 non-CVE vulnerability found. To see more deta...

package	version	type	oss_package	description	reference	cvss_id	cvss_title	cvss_description	cvss_score
1 haven	2.5.1	cran	pkg:cran/haven@2.5.1	haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files Imp...	https://ossindex.sonatype.org/component/pkg:cran/haven...	sonatype-2021-1261	sonatype-2021-1261	1 vulnerability found	1 non-CVE vulnerability found. To see more deta...

---

Facilitate collaboration



---

# Why collaboration?

**Project planning** – assign tasks

**Project runtime** - working together on an R project

**Once work is done** - validation

**Over time** - resilience of project



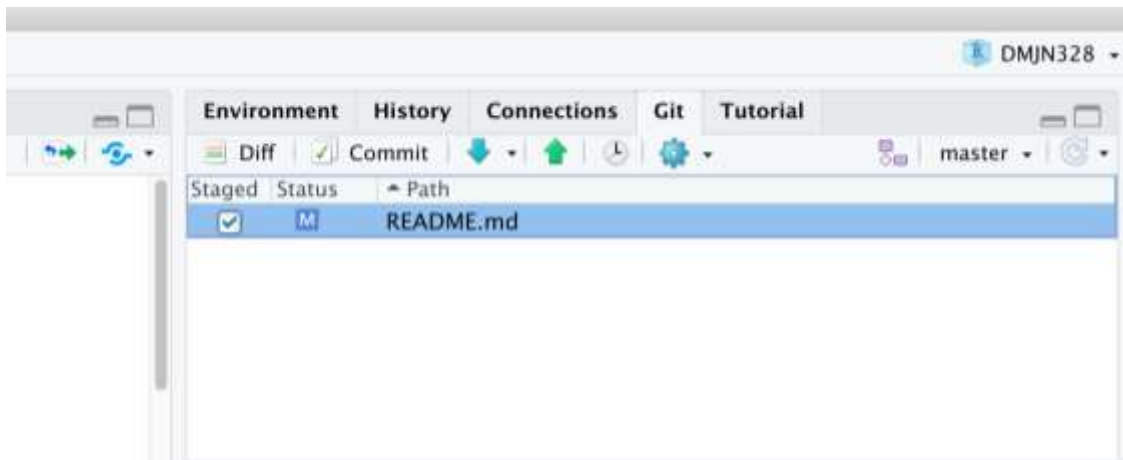
# How do we get there?

**Using Git is key!**

Plan > Share > Backup/Archive

The problem – most R users are not familiar with it!

RStudio comes with Git integration!



Azure Repos



---

# Other collaboration features

Posit Cloud / Workbench

- Project sharing

Reproducible environments

- Renv package
- Docker



```
renv::init()  
renv::snapshot()  
  > renv.lock  
renv::restore()
```

<https://rstudio.github.io/renv/articles/renv.html>

<https://rstudio.github.io/renv/articles/docker.html>

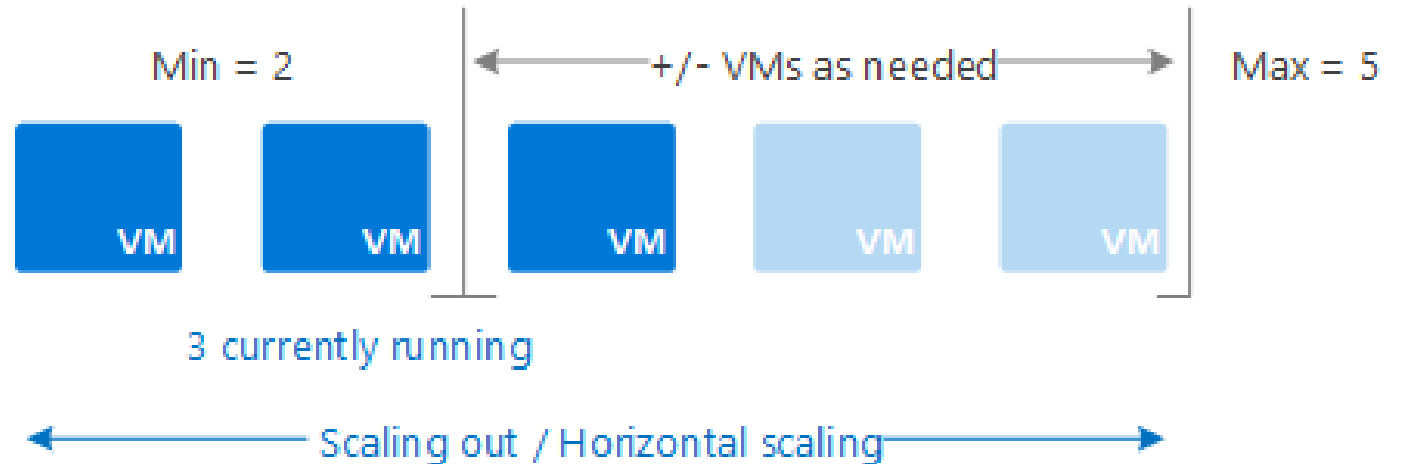
---

Scalability matters

# Scaling

- Variable workloads
- Variable R users and consumers
- Supporting projects

Auto-scaling through cloud services!

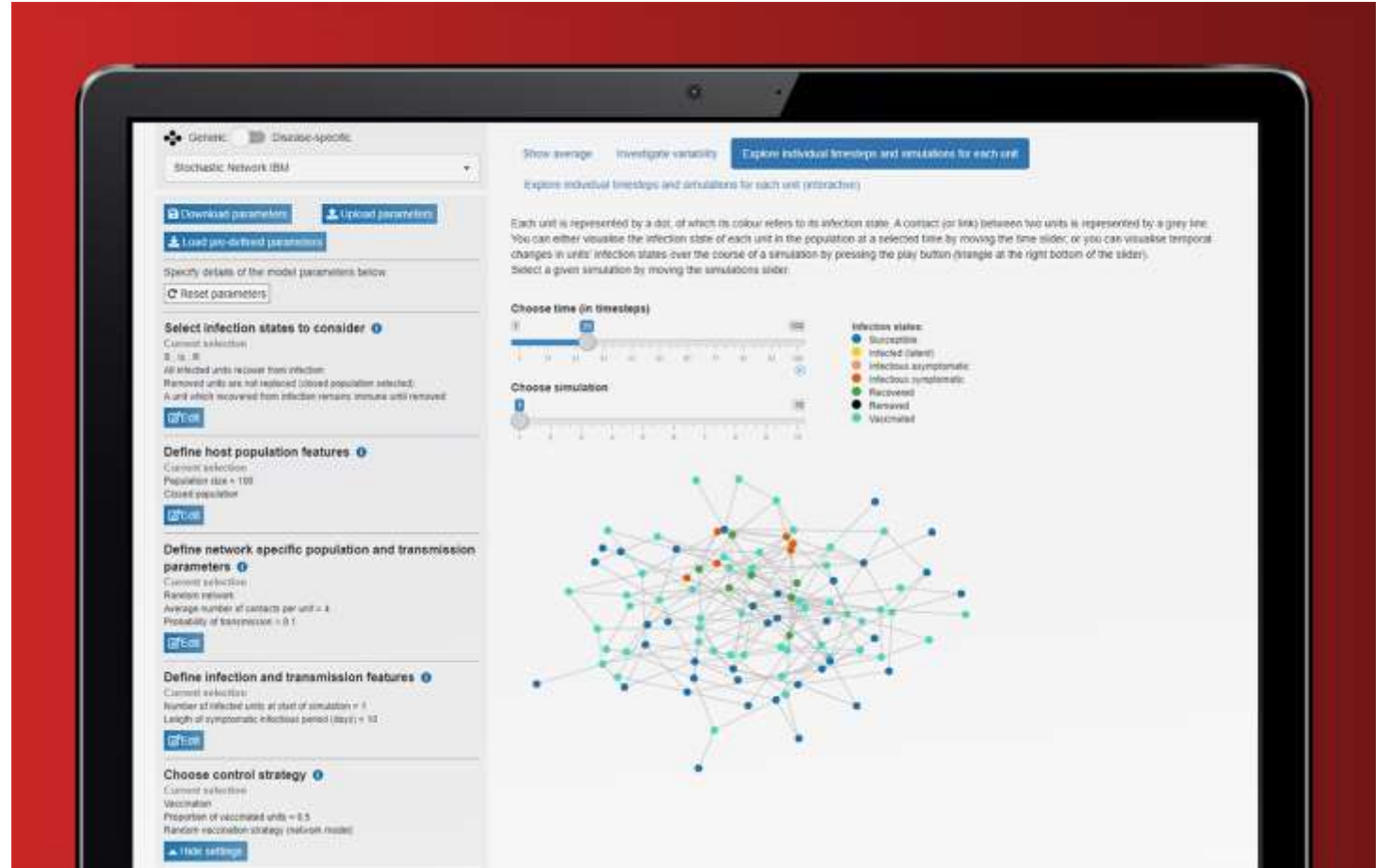


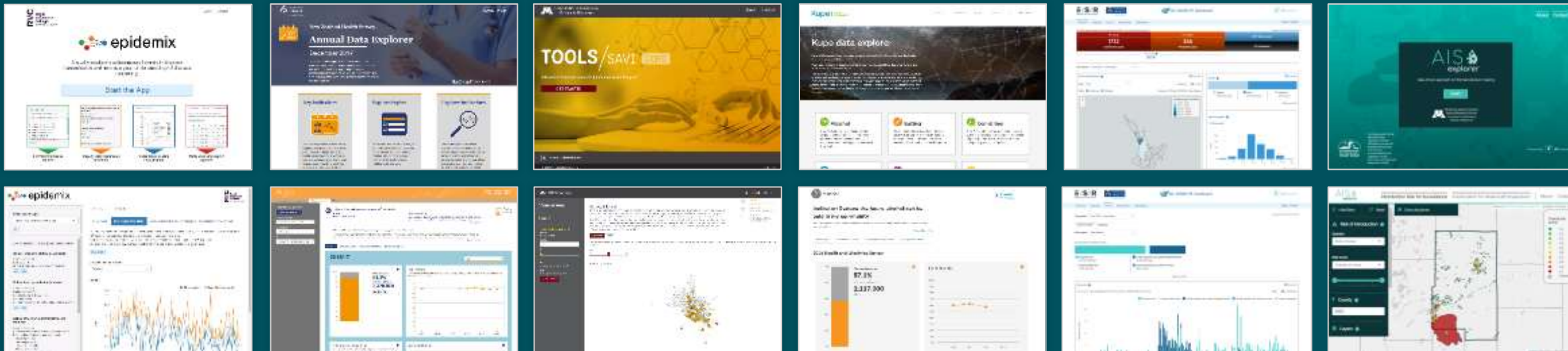
Source: <https://learn.microsoft.com/en-us/azure/azure-monitor/autoscale/autoscale-overview>

# Example: Epidemix

- Dedicated hosting
- R is single-threaded!
- Scaling on demand (AWS)

[epidemix.app](https://epidemix.app)





# Questions?