REDIVIS

data. (re) imagined.

XLDB • 30 April 2018

Ian Mathews (co-founder) • ian@redivis.com

Goals & Vision

We imagine a world where policy and scientifically relevant data are readily available and interpretable for a wide range of individuals.

- Break down barriers in data driven research
  - Reduce time-to-science
- Promote and facilitate collaboration
  - Crowd-source data curation
- Facilitate communication of data science
- Reduce fundamental tension between data availability and security

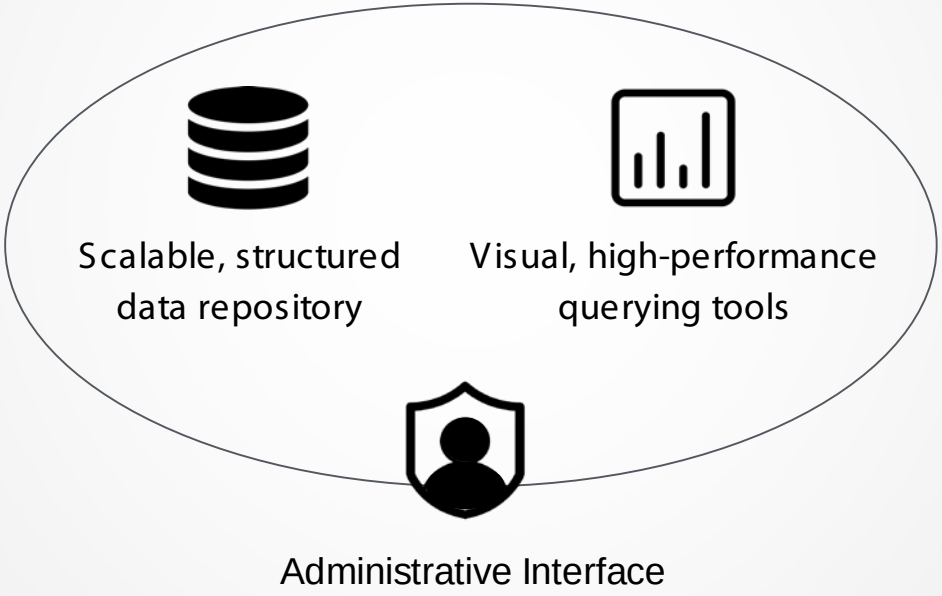
Stanford MEDICINE | Center for Population Health Sciences

> “Provide a central hub of data assets to facilitate transdisciplinary population health projects and collaborations across the university”

Using Redivis to manage access for:

- Over **120** active researchers
- Utilizing **93** source datasets
- Across **215** projects
- Containing **726** billion records

Redivis is a **cloud-based data platform** that provides in-browser tools for **administrators to host and manage** access to their data resources and for **researchers to discover and query datasets**



REDIVIS

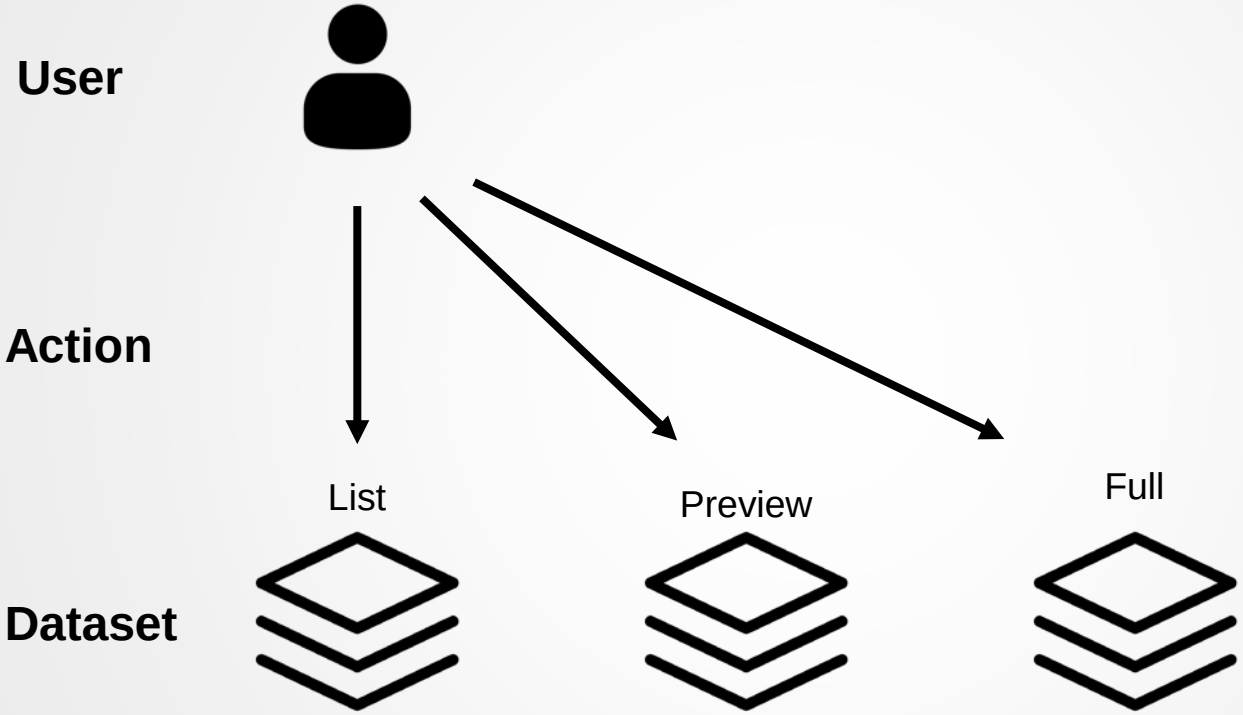
Two biggest challenges

- Access control and administration
  - Streamline user experience; remove “security theater”
  - Grant administrators ability to fully enforce data usage restrictions and audit behavior
- User experience in working with big data technologies
  - Remove technical barriers for novice researchers
  - Decouple computer science from medical science

Access control

You can’t democratize data without addressing access

Access control



Access paradigms

Direct approval

- Defined on a user + dataset + action
- Needs to be approved for every combination specified

Requirement

- Applies across a permission group
- User only fills out once
- E.g. HIPAA training

Study

- Combines multiple users + datasets into one authorized workspace
- E.g. IRBs

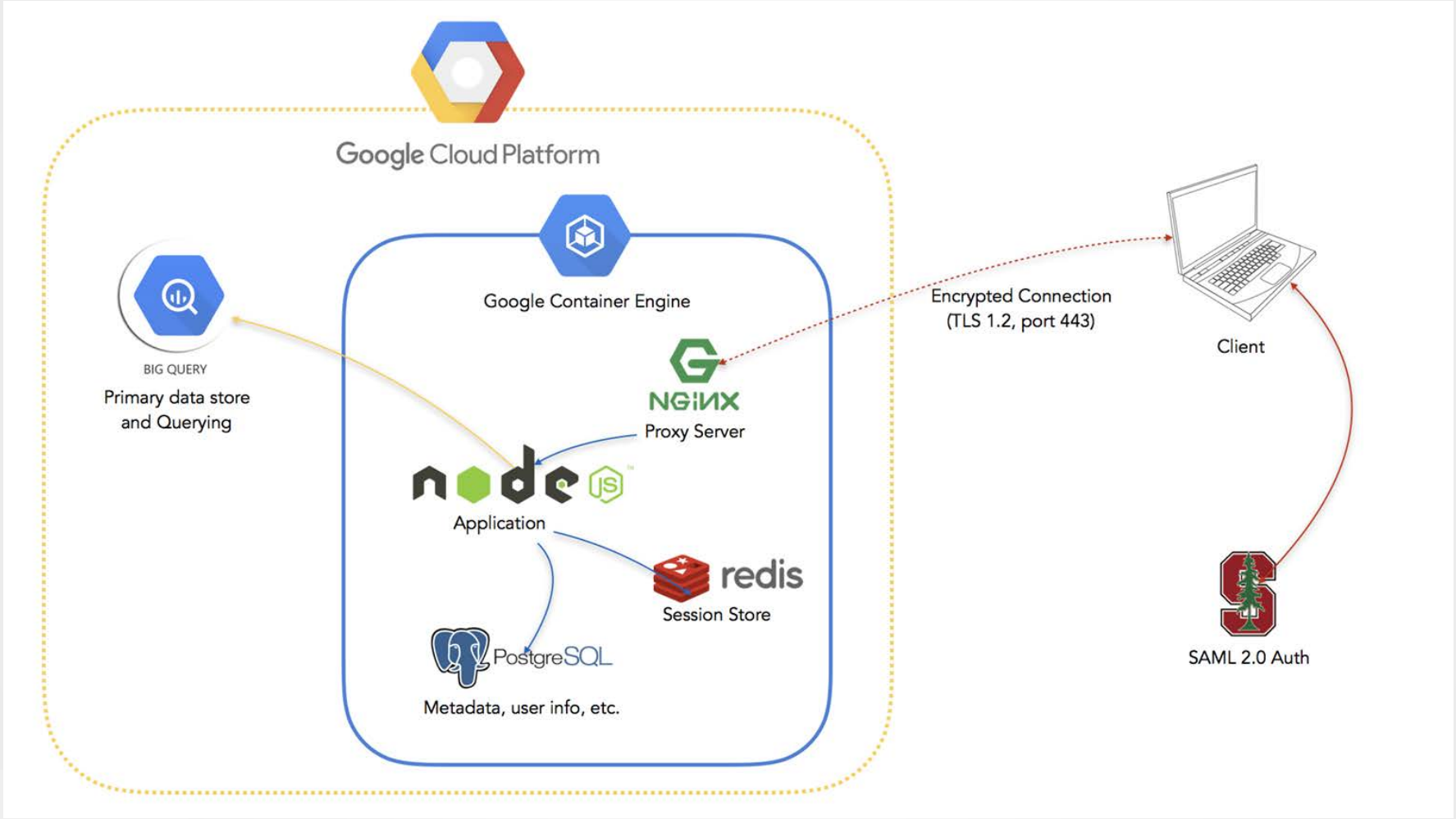
User experience

- Promote discovery & exploration
- Facilitate understanding of new datasets
- Reduce barrier to entry in working with large datasets
- Aid in documentation and reproducibility

Demo

Let's generate a cohort of postpartum depression candidates from insurance claims data

https://redivis.com/StanfordPHS


Google Cloud Platform
Google Container Engine
BIG QUERY
Primary data store
and Querying
NGINX
Proxy Server
Application
redis
Session Store
PostgreSQL
Metadata, user info, etc.
Encrypted Connection
(TLS 1.2, port 443)
Client
SAML 2.0 Auth

Next steps

May 2018

Improved querying interfaces

- Support greater complexity while reducing overhead
- Fully expose SQL interface for more technical users

Summer 2018

Integration with analytical tools

- Connect to open-source toolkits (e.g., Jupyter labs)
- Develop and connect with data visualization ecosystem to facilitate communication to a broad audience

Fall 2018

Collaboration and crowd-sourced curation

- Build community around datasets, allow users to share knowledge
- Explore mechanisms for user contributions to datasets (e.g., Github)

REDIVIS

Q & A

REDIVIS

data. (re) imagined.

XLDB • 30 April 2018

Ian Mathews (co-founder) • ian@redivis.com