

Designing and Building Secure Data Environment

Services for handling sensitive data in a research context

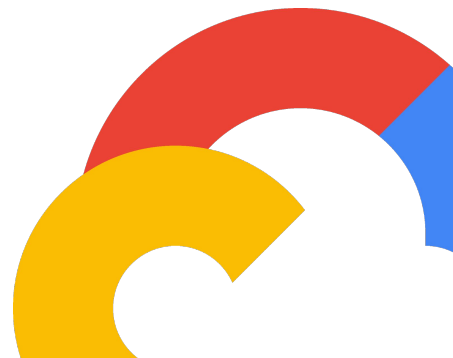


**Hariprasad
Radhakrishnan (Hari)**
Customer Engineer
Healthcare and Research
Google Cloud



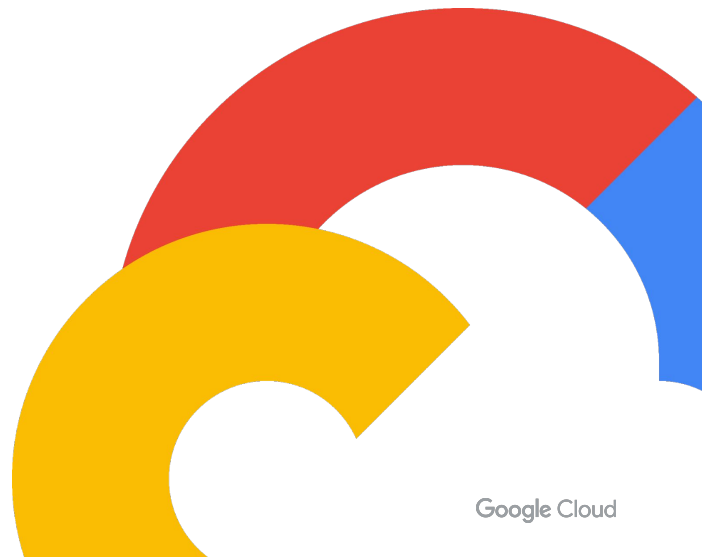
5 years @ Google

Work with Research, Higher Ed and Healthcare organizations across EMEA. Passionate about all things Life Sciences, Healthcare, genomics and fundamental research.



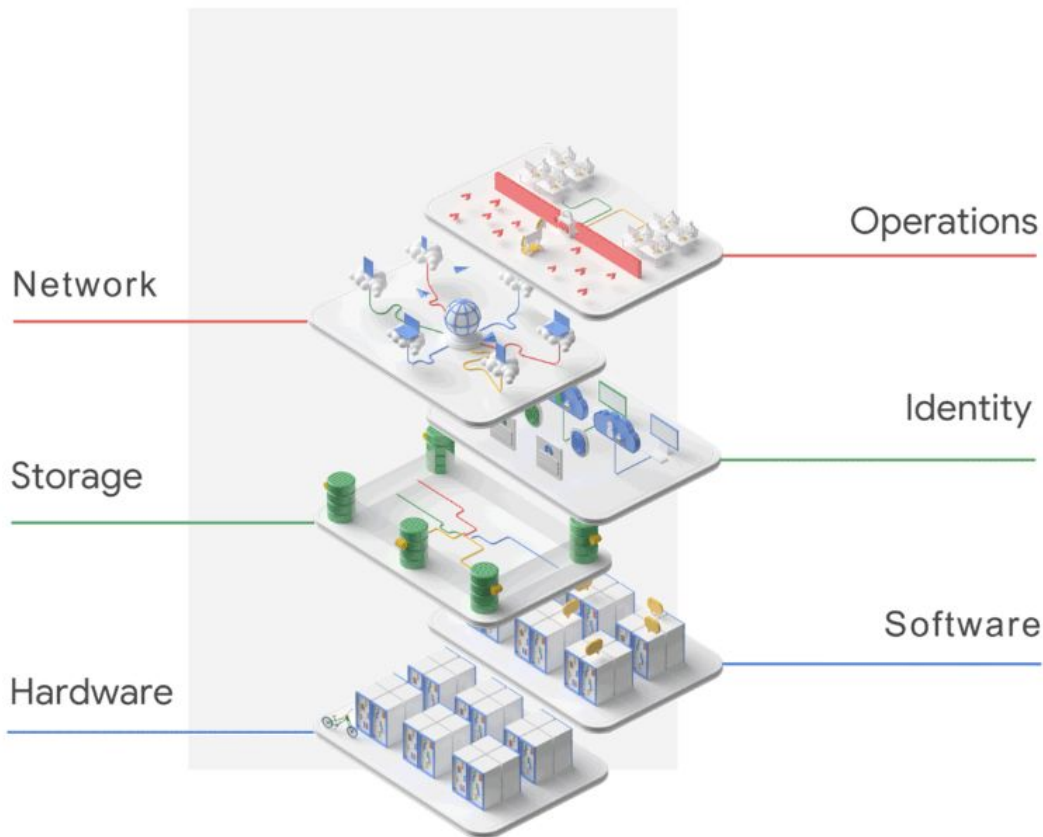


A Global High-performance
infrastructure for **cloud**
computing, data analytics &
machine learning.



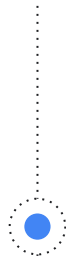
Security
that's **built in**

Defense in depth
by default and at
scale



Contents

- 01** **Introductions**
- 02** **Five safes, use cases**
- 03** **Secure Data environment**
- 04** **GCP Building blocks**
- 05** **Blueprints & Deployable Solutions**
- 06** **Q&A**

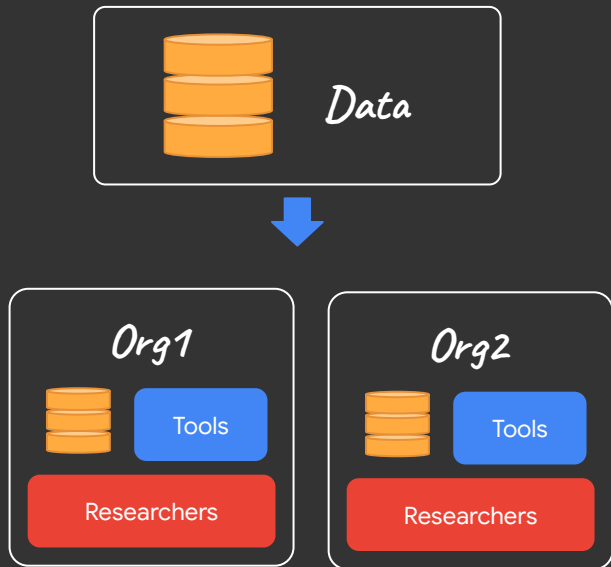


Data Protection Challenges

Key concerns when it comes to sensitive data

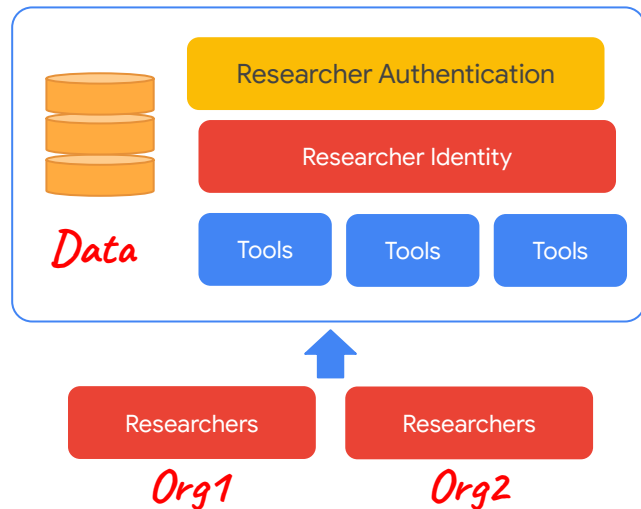
- How do I protect sensitive data or my IP?
- How do I stay compliant with data protection regulations?
- How do I collaborate with other companies processing their sensitive data?
- How do I protect my clients' and users' data?

When Data Leaves your system



Varying security, auditability, least control

When Users come to your system for Data



Consistent experience, shared tools, security, auditability and control

Secure Data environment - Things to consider



Safe

- | Data
- | People
- | Projects
- | Outputs
- | Setting
- | Computing*

Identify sensitive or PII Data

Is there a real need to share sensitive information

Can we get things done by using anonymised data instead - re-identification risks

Who is the custodian of the data and what Data Governance needs to go in place

Who needs to access the data and for how long, who should not have access

Is there a need to have data exfiltration controls

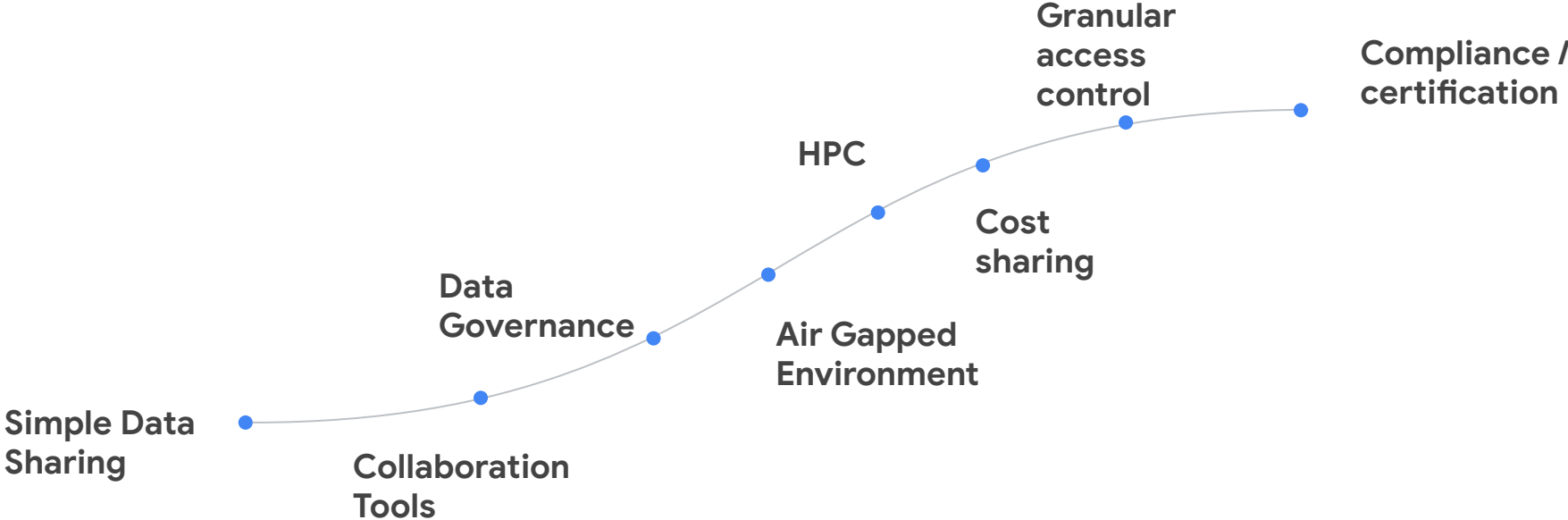
What Data is being extracted out of the system

Ethical Principles

Complying with Law, transparency and auditability

Security

Need vs Acceptable Complexity vs Cost



Use cases we see for SDEs / TREs

- Government National Health Initiatives requiring SDE / TRE service on Google Cloud.
- Bio-Bank initiatives requiring secure access to genetic data to researchers
- Pharma looking for SDE / TRE services for their external collaboration with Academia.
- Research organisations hosting sensitive data for cross academia, industry, research collaboration projects.

Example: Cancer Gateway data on GCP

National Cancer Institute

ISB Cancer Genomics Cloud

Industry: Healthcare & Life Sciences

Country: United States

<https://github.com/isb-cgc/>

<https://isb-cgc.appspot.com/>



All Ref to: ISB-CGC - Cancer Gateway in the Cloud

ISB-CGC: Gateway to cancer research in Google cloud



Data

- Upload your data to cloud
- Access 15+ data types from 25 projects
- Integrated image viewers



Tools

- Choice of workflow technology
- Web-based tools
- Google BigQuery
- Fully customizable Google VMs



Analysis

- Interactive notebooks from Jupyter Lab, R-Studio, or Google Colab
- Pre-designed bio-stats notebooks
- Plot-based visual analysis



Results

- Export charts from notebooks for publication

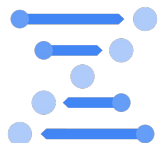
More information at: <https://isb-cgc.appspot.com/>

Existing SDE / TRE on GCP

National Health Organisation Patient Genomics and Health Data

Industry: Healthcare & Life Sciences

Country: Nordics



(GWAS)

National Health Data - SDE / TRE for 500K Genetic and phenotype data

Air gapped environment for National population health data comprising of both Genetic data and patient health information

Ability to run large scale Genomic pipelines

Access to bioinformatic tools through a Vdi solution

Data governance and approval workflows for extracting summary data

Access to pharmaceutical industry partners

Exfiltration controls using perimeter security



Existing SDE / TRE on GCP

UK Consortium of Research & Universities with Pharma partners

Industry: Healthcare & Life Sciences

Country: United Kingdom



Project SDE / TRE hosting 50,000 patient cohort data to investigate genetic contribution to diseases.

Jointly analyse genetic data and electronic health record data inside a certified SDE / TRE on Google Cloud

Workspaces for different groups of scientists working on sensitive data

Ability to share data between workspaces by researchers when needed

Access to bioinformatic tools through a virtual desktop solution

Ability for researchers to run High performance computing workloads within their workspace

IGV Browser support for analysing genetic variants

Exfiltration controls using perimeter security



Existing SDE / TRE on GCP

Big Pharma - Europe

Industry: Healthcare & Life Sciences

Country: EMEA



Open Health
Imaging Foundation



SDE / TRE for biomedical Images and FHIR

SDE / TRE to support ingestion of biomedical and radiology images along with phenotype information in FHIR format.

Ability to natively store DICOM and FHIR standard data.

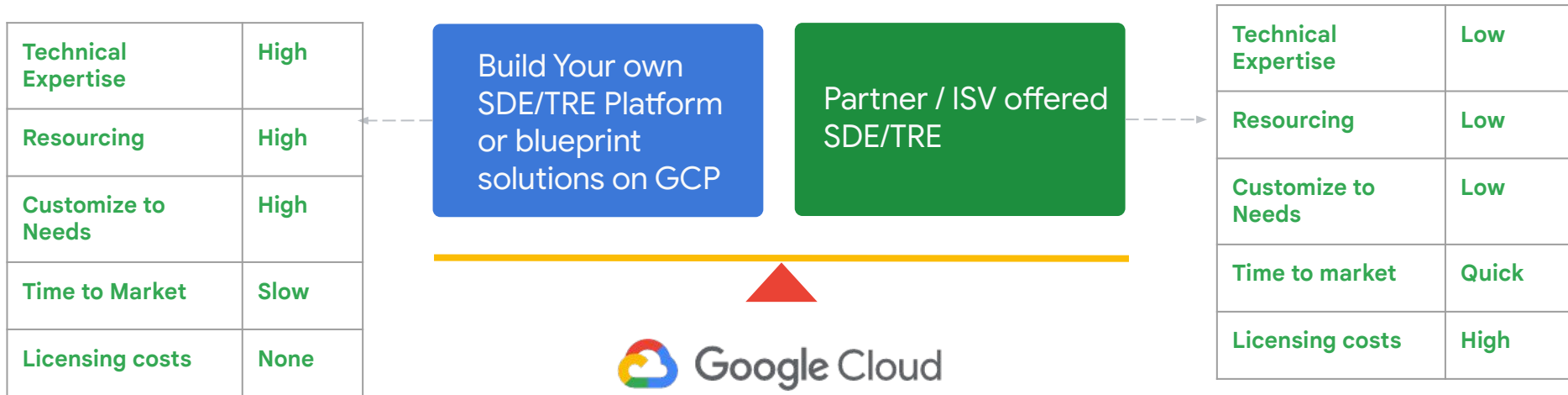
Data harmonization workflows with PII data scans and redaction.

Support for DICOM viewer and annotation tools for Researchers using the SDE / TRE

Support for Workspaces for different pools of researchers and clinical teams

Ability to collaborate and securely share data in SDE / TRE with external collaborators

Choosing your approach



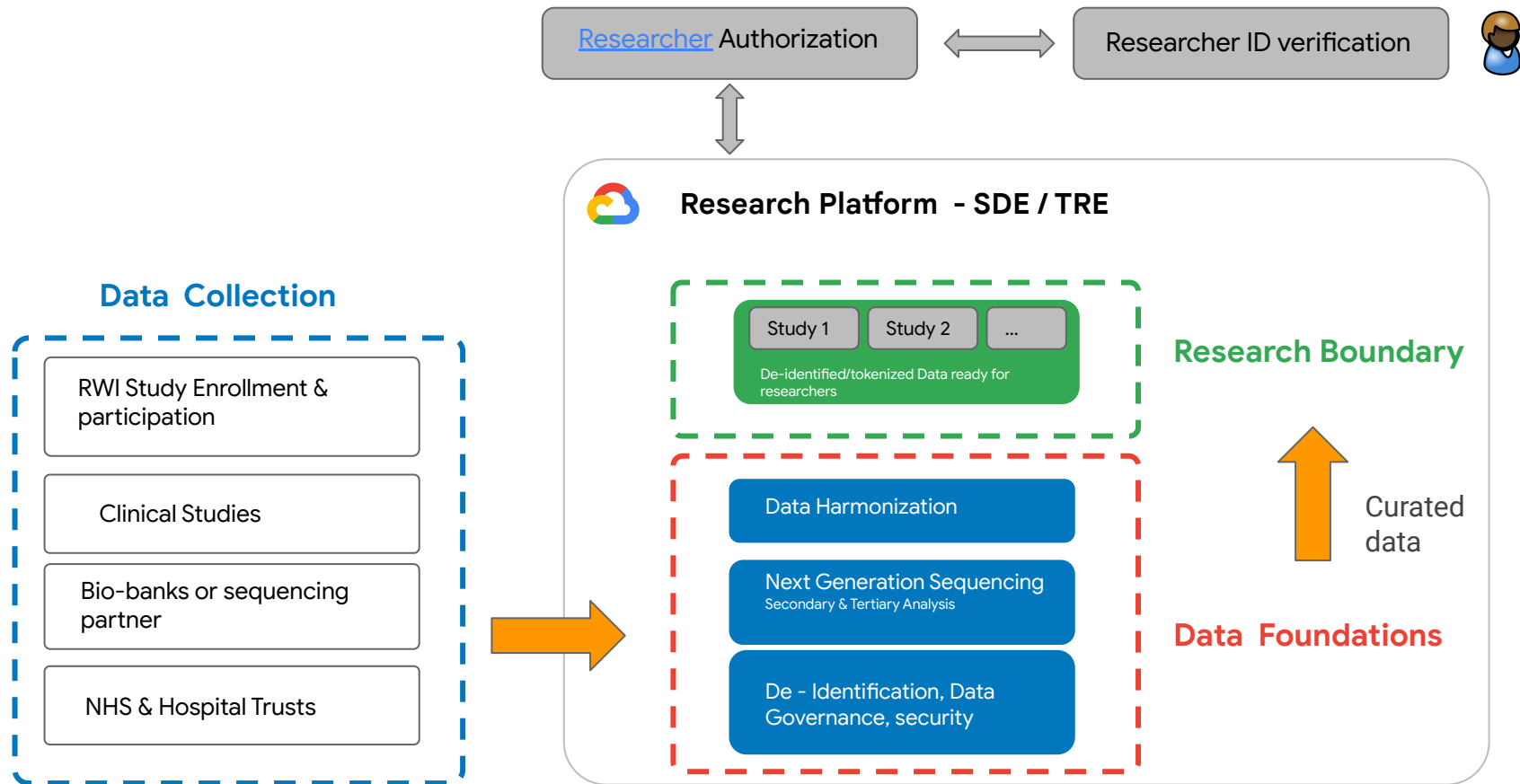
Building blocks of Secure Data Environment (SDE / TRE)

**Data Collection
and Storage**

**Data
Foundations**

**Data Sharing or
Research Boundary**

Secure Data Environment (SDE / TRE)



1

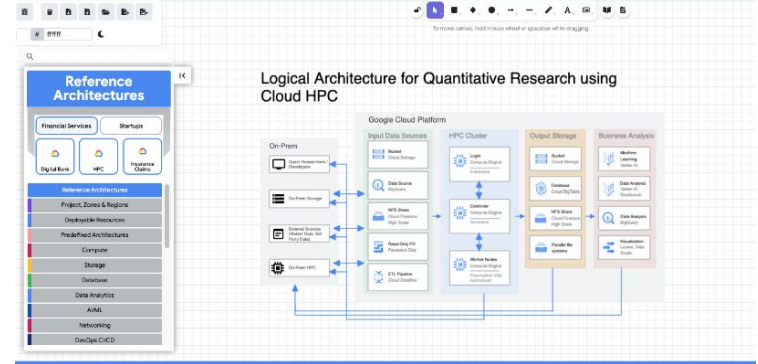
GCP Building Blocks

Services to handle and protect sensitive data



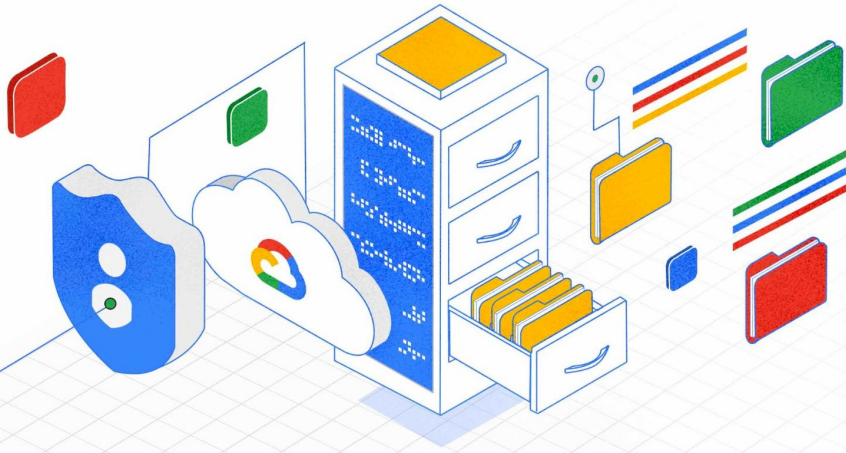
2

Blue prints



GCP Building Blocks

Services to handle and protect sensitive data





Sensitive Data Protection

Discover-Govern-Protect-Report

Sensitive Data Protection

Sensitive Data Protection provides resources to help you discover, govern, protect, and report on sensitive data across your ecosystem.

Low Operational Overhead

Sensitive Data Protection










[OVERVIEW](#) [DISCOVERY](#) [INSPECTION](#) [RISK ANALYSIS](#) [CONFIGURATION](#) [SUBSCRIPTIONS](#)

Sensitive Data Protection

Sensitive Data Protection provides resources to help you discover, govern, protect, and report on sensitive data across your ecosystem.
















Learn about your data

Find, classify, and understand the risks to your sensitive data in Google Cloud and beyond.

Service	Purpose
 Discovery	Get continuous visibility into all your sensitive data.  
 Deep inspection	Inspect your data in storage systems exhaustively and investigate individual findings.  
 Risk analysis	Assess data for privacy and re-identification risk.  




Protect your data

Prevent and remediate attacks on your sensitive data.

Service	Purpose
 Content de-identification	Transform and derisk sensitive data findings.  
 Data de-identification at query time	De-identify data while querying using a remote function.  
 Cloud Storage de-identification	Create de-identified copies of Cloud Storage data.  
 Chat-log redaction for Dialogflow and Contact Center AI	Redact sensitive data from unstructured chat logs.  
 Chronicle integration	Publish sensitive data intelligence into Chronicle  

Build privacy-aware applications

Use APIs to discover, inspect, and protect sensitive data in your own workloads.

Service	Purpose
 Cloud DLP API	Inspect and de-identify data in custom workloads.  

Sensitive Data Protection










[OVERVIEW](#) [DISCOVERY](#) [INSPECTION](#) [RISK ANALYSIS](#) [CONFIGURATION](#) [SUBSCRIPTIONS](#)

Sensitive Data Protection

Sensitive Data Protection provides resources to help you discover, govern, protect, and report on sensitive data across your ecosystem.


















Learn about your data

Find, classify, and understand the risks to your sensitive data in Google Cloud and beyond.

Service	Purpose
 Discovery	Get continuous visibility into all your sensitive data.  
 Deep inspection	Inspect your data in storage systems exhaustively and investigate individual findings.  
 Risk analysis	Assess data for privacy and re-identification risk.  

Protect your data

Prevent and remediate attacks on your sensitive data.

Service	Purpose
 Content de-identification	Transform and derisk sensitive data findings.  
 Data de-identification at query time	De-identify data while querying using a remote function.  
 Cloud Storage de-identification	Create de-identified copies of Cloud Storage data.  
 Chat-log redaction for Dialogflow and Contact Center AI 	Redact sensitive data from unstructured chat logs.  
 Chronicle integration 	Publish sensitive data intelligence into Chronicle  

Build privacy-aware applications

Use APIs to discover, inspect, and protect sensitive data in your own workloads.

Service	Purpose
 Cloud DLP API	Inspect and de-identify data in custom workloads.  

Discover your sensitive data

Sensitive Data Discovery, a service that continuously profiles your sensitive data so that you can understand and protect it.

Risk Analysis

Assess data for privacy and re-identification risk.

Risk analyses can help you see how the size, shape, and distribution of data can increase re-identification risk.

Protect data by de-identifying it

De-identification is the process of removing identifying information from data. Sensitive Data Protection can de-identify sensitive data in text content, including text stored in container structures such as tables.

Running Discovery scans for sensitive data

← Create scan configuration

- Select a discovery type**
 - BigQuery**
Create data profiles of BigQuery tables.
 - Cloud SQL**
Create data profiles of Cloud SQL tables.
 - Secrets/credentials vulnerabilities**
Scan resource metadata for secrets and credentials, and report any findings as vulnerabilities to Security Command Center. [Learn more](#)

CONTINUE
- Select scope**

Specify whether to scan the entire organization, a specific folder, or a project. Each organization, folder, or project can only have one configuration.
- Manage Schedules (Optional)**

Create a new schedule to specify the frequency and conditions for profiling specific subsets of data. For example, you can create a schedule that excludes a specific dataset from profiling operations.
- Select inspection template**

Use inspection templates to ensure consistency across scan configurations
- Add actions**

Manage scan output
- Set location to store co**

Set geographic location to store this configuration. This location will be used for all subsequent scan configurations.
- Review and create**

Review the above information before creating the scan configuration.

CREATE **CANCEL**

Data profile

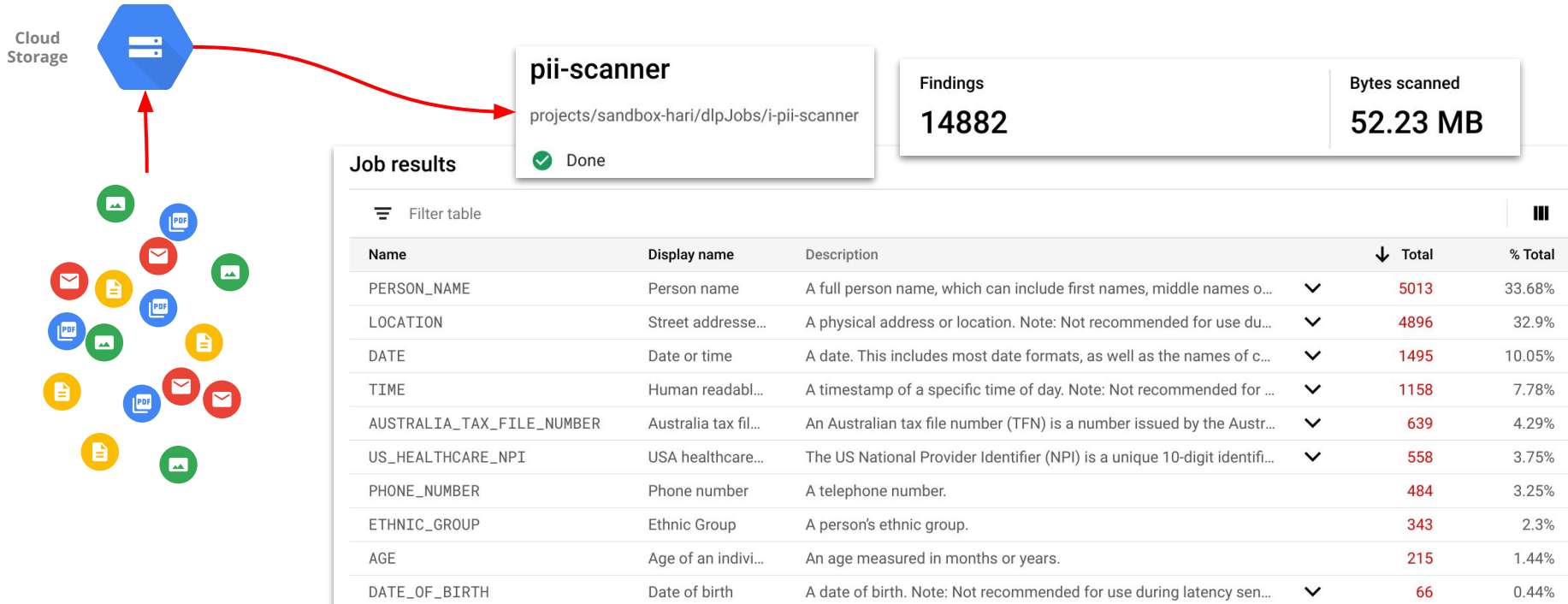
Status	✓ Done
Data risk ?	🔴 High
Sensitivity ?	High
Predicted InfoTypes ?	—
Other InfoTypes ?	▼ 3 InfoTypes GENDER (sensitivity: Moderate) PERSON_NAME (sensitivity: Moderate) UK_NATIONAL_HEALTH_SERVICE_NUMBER (sensitivity: High)
Profile last generated ?	Jun 27, 2023, 8:10:42 PM

Column Profiles for this Table

Filter Enter property name or value

Field ID ?	Data risk ? ↓	Sensitivity ?	Predicted infoType ?	Other infoTypes (Estimated prevalence) ?	Data type ?	Policy tags ?
✓ Links	🔴 High	High	—	PERSON_NAME (3%) UK_NATIONAL_HEALTH_SERVICE_NUMBER (1%)	TYPE_STRING	No
✓ Name	🟡 Moderate	Moderate	—	PERSON_NAME (5%)	TYPE_STRING	No
✓ Feature	🟡 Moderate	Moderate	—	PERSON_NAME (2%)	TYPE_STRING	No
✓ Description	🟡 Moderate	Moderate	—	PERSON_NAME (10%) GENDER (1%)	TYPE_STRING	No
✓ Price	🟡 Low	Low	—		TYPE_STRING	No
✓ SKU	🟡 Low	Low	—		TYPE_STRING	No
✓ Component	🟡 Low	Low	—		TYPE_STRING	No

Scan documents





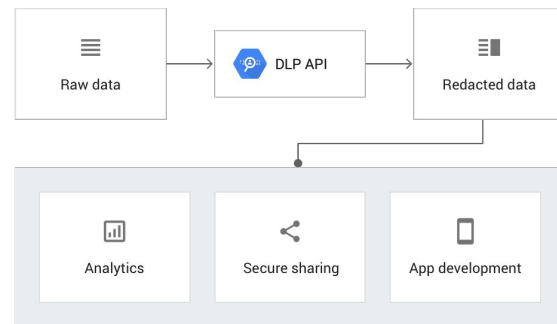
Data Loss Prevention

Identifying, redacting and cataloging sensitive PII



Data Loss Prevention (DLP) API

Cloud DLP provides access to a powerful sensitive data inspection, classification, and de-identification platform.



ID	Job Title	Phone	Comments
359740	Senior Engineer	307-964-0673	Please email them at jane@imadethisup.com
981587	VP, Engineer	713-910-6787	none
394091	Lawyer	692-398-4146	Updated phone to: 692-398-4146
986941	Senior Ops Manager	294-967-5508	none
490456	Junior Ops Manager	791-954-3281	Tried to verify account with their SSN 222-44-5555

De-ID: Masking and Tokenization

Input → “This is my phone number: (858)867-5309”

Partial Masking

Output → “This is my phone number: (858)XXX-XXXX”

Hashing or Tokenizing



Output → “This is my phone number: ga+32mx32s2as8cw38AEfknsFthc”

Format Preserving Encryption or Tokenization



Output → “This is my phone number: (431)582-6528”

De-ID Healthcare data

Google Cloud Healthcare De-ID Demo

Type text to: to inspect results.

AutoSubmit:

```
PATIENT: Barrande, Jameis  
MRN: 034246802  
DATE OF OPERATION: 12/03/2016  
SURGEON: John Palasides, MD.
```

```
PATIENT: Barrande, Jameis  
MRN: 034246802  
DATE OF OPERATION: 12/03/2016  
SURGEON: John Palasides, MD.
```

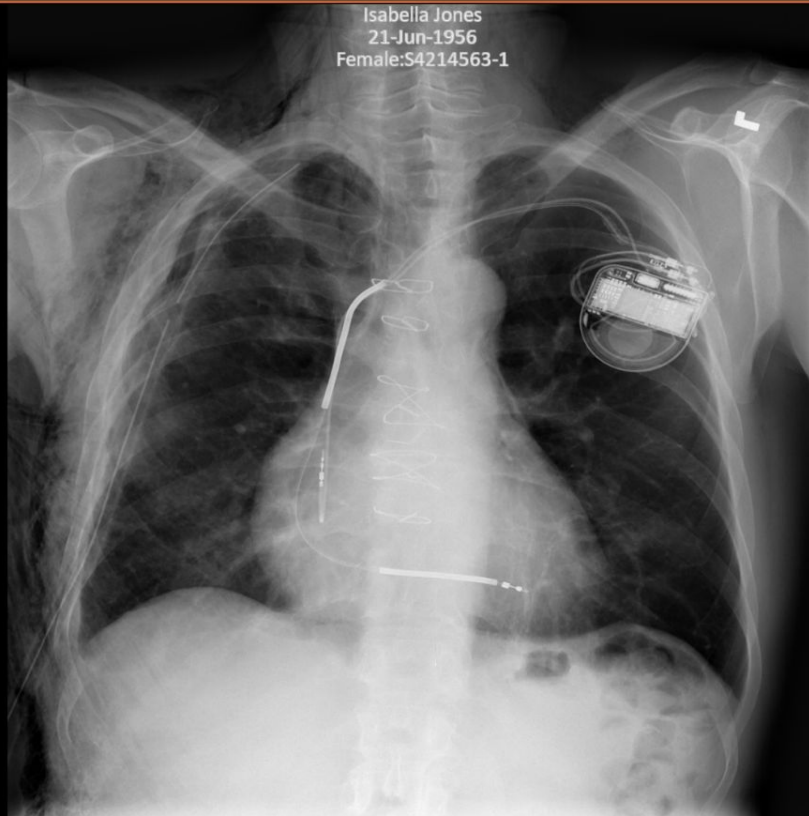
```
PATIENT: [LAST NAME #5], [FIRST  
NAME #1]  
[LAST NAME #4]: [MRN]  
DATE OF OPERATION: [DATE]  
SURGEON: [FIRST NAME #2] [LAST  
NAME #3], [STATE #1].
```

G1

De Identification for images

Julia Jones
MRN: S4214563-1
Age:
Sex: F

Isabella Jones
21-Jun-1956
Female:S4214563-1



Desc:
ACC #:
Study Date: 08-Sep-2005
IM Time:
Zoom Factor: 0.60
Lossy 14:1

1/1
IM #:
SE #:
WL : W:256 L:128

Inst:

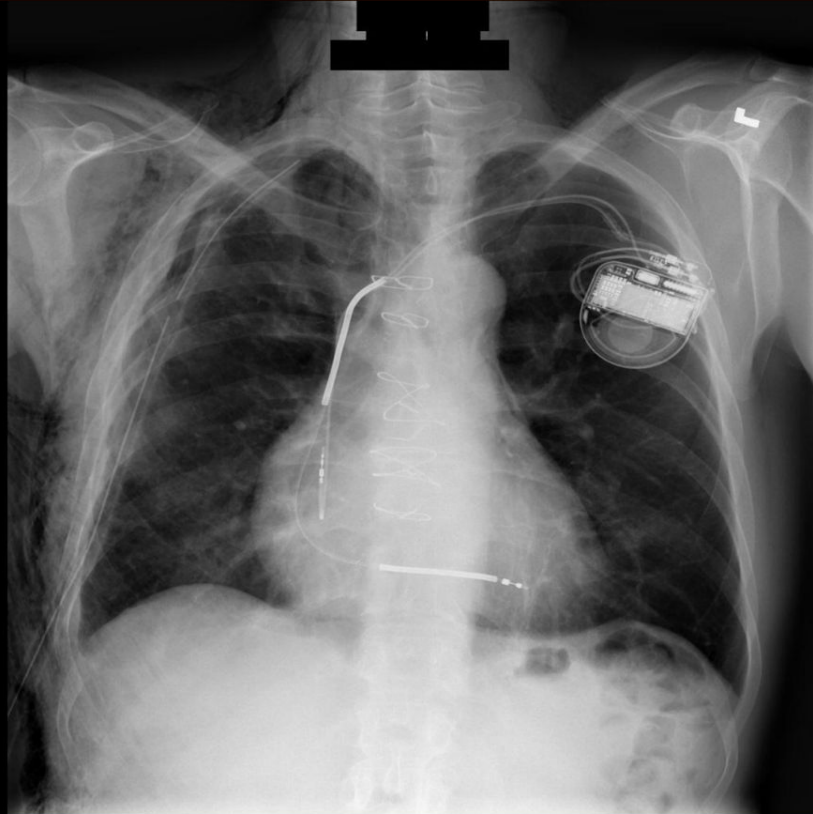
The first image shows an x-ray image with sample PII and PHI data appearing in both metadata and burned-in text.

De Identification for images

MRN: null

Age:

Sex:



Desc:

ACC #:

Study Date:

IM Time:

Zoom Factor: 0.60

Lossy 34:1

1/1

IM #:

SE #:

WL : W:256 L:128

Inst:

The second image shows the same x-ray image with all sample PII and PHI metadata removed or obscured.



Data Fusion - De Identification - ETL Pipelines

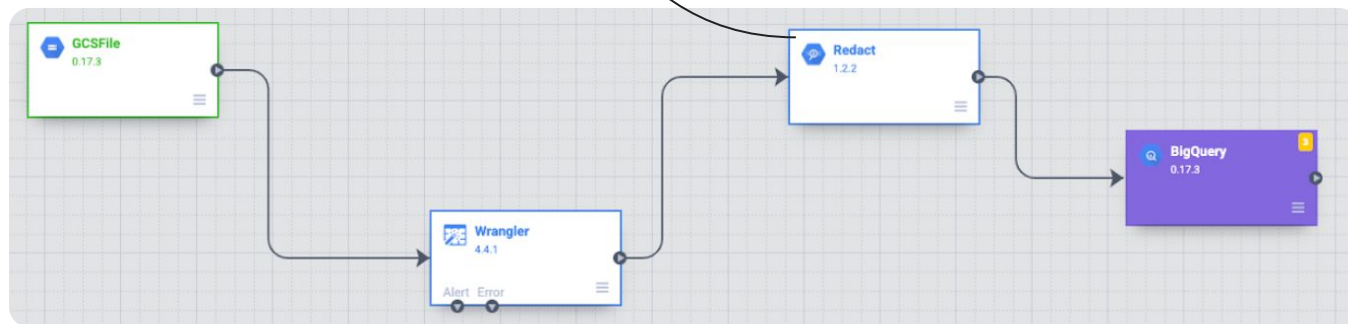
Cloud Data Fusion is a fully managed, code-free data integration service that helps efficiently build and manage ETL/ELT data pipelines.

Matching

Fields to Transform *

Apply **One-way Hash** on **Everything** within **provider_id**

Apply **Masking** on **Everything** within **street_address**



Source

Transformation

PII Data Redaction

Target

Google Cloud



Data Fusion - Source to Target

provider_id	agency_name	street_address	city	state	zip_code	total_episodes_r	distinct_users_n	total_hha_charg
337290	AMERICARE CF 5923 STRICKLA	BROOKLYN		NY	11234	3148	2310	14112445
58217	BRADBOURNE 16029 ARROW	IRWINDALE		CA	91706	620	297	1786721
58419	INSIGHT HEALT 500 S KRAEMEI	BREA		CA	92821	19	15	127300
108167	D & N HOME HE 1140 W 50 ST S	HIALEAH		FL	33012	208	63	1399981
109023	PSN HEALTH C. 17670 NW 78 A	MIAMI GARDEN		FL	33015	59	18	460554
109029	RAINBOW HOM 15165 NW 77 A	MIAMI LAKES		FL	33014	127	54	554558
109181	MARTINS FLOR 6501 NW 36ST,	VIRGINIA GARE		FL	33166	297	77	1586487
109383	LIFETIME HOMI 8785 SW 165TH	MIAMI		FL	33193	218	134	1002418
109437	SUPREME HOM 8910 MIRAMAR	MIRAMAR		FL	33025	100		
109589	ADVANCE HOM 9835 SW 72ND	MIAMI		FL	33173	165		
109694	POTENTIAL HO 9560 SW 107 A	MIAMI		FL	33176	76		
178089	MERCY HOME 2102 EAST 21S	WICHITA		KS	67214	70		
248048	BARNABAS HE, 223 WASHINGT	BRAINERD		MN	56401	47		
368049	OMNICARE HO/ 24800 CHAGRIN	BEACHWOOD		OH	44122	19		



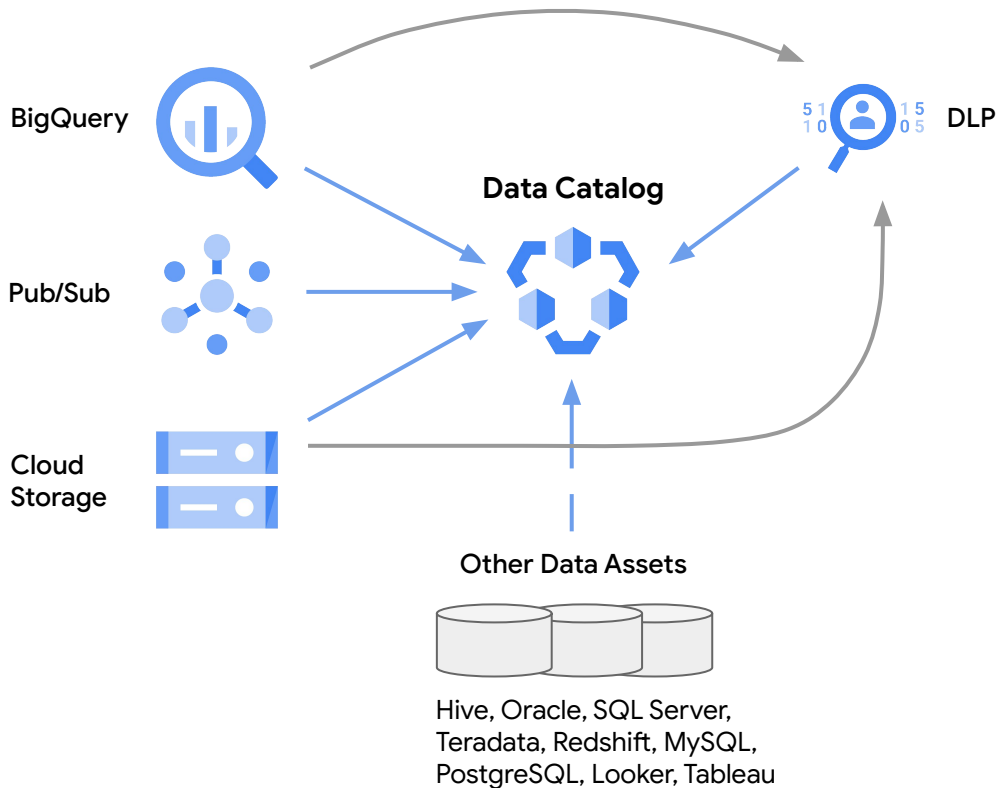
Hashing or Tokenizing

Masking

Row	provider_id	agency_name	street_address	city
1	IFTGFGYM4NrkSFKatwpCddis9rUXeyLrIDhZK/9njE=	CUIDADO CASERO EL GIGANTE	xx	ADJUNTAS
2	aZ2srYIPPI3Bcp2lp9NHqrQ21Zflk/6oMY5dj01/fZU=	DD HOME CARE SAN FRANCISCO DE ASIS, INC	xx	AGUADILLA
3	ZUwJS/VgYVl0Q0yB5laG7Z5VmPY62PnRrgd06oDH0cQ=	ST LUKES HCP	xx	AGUADILLA
4	f021vitE1ui/b80DTor8uvrA0Rs5B82HgSrYPho9FCk=	ARECIBO MEDICAL HCP INC	xx	ARECIBO
5	Fj3p1ixSbGG35cd6l+1zZRzJJaWt/ryKQoLsmiTMXQc=	ST LUKES HCP	xx	ARECIBO
6	d+fXOui4jvZbHqUz6ajyxtU3V04ABoySCH4cn4rHc=	ST LUKES HCP	xx	LADES
7	7hmcaxKjHyEI29FMegMQBYfKm5HYdZF2BHj8rpluBo=	CORPORACION DE LAS VEGAS, INC	xx	MANATI
8	aZhdmA/is7blRm075ggykYRGfHGFRdwGWH8TVVBUVE=	LUZ DE ESPERANZA HOME CARE INC	xx	ALT DE LUCH
9	2y00NaSJELBT5W1scWaEe8L3w54rKkcc0/8hbvsCb8=	ST LUKES HCP	xxxxxxxxxxxxxxxx	MAYAGUEZ
10	mUKleYIBPFCgLX13iilK9NSNGjFlre1T8EuXQZwrvQ=	FIRST HOME CARE CENTER INC.	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	MAYAGUEZ
11	3HZF1kAWRz/L5YTF5X7xkLrv/d7d1QzWM5hMP7tptfc=	CUIDADO CASERO MAYAGUEZ INC	xx	MAYAGUEZ
12	q4NJEWdzJjSY2FpCll1Bg3j5kWJDJOl2LJ5wOHpww1k=	VISITING HEALTH SERVICES PROGRAM HOSPITAL DAMAS	xxxxxxxxxxxxxxxxxxxxxxxx	PONCE
13	Jv/YYe4uNI08AunCVGDxpAcc1Qf0nLyDIST248FFUM=	ATENCION MEDICA EN EL HOGAR INC	xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	CAGUAS

Auto Tag sensitive data - Dataplex - Data Catalog

1. Auto-syncs 'technical metadata' from GCP data assets in near real-time
2. **Auto-tags PII data** through DLP integration
3. Supports non-GCP data assets through open-source connectors

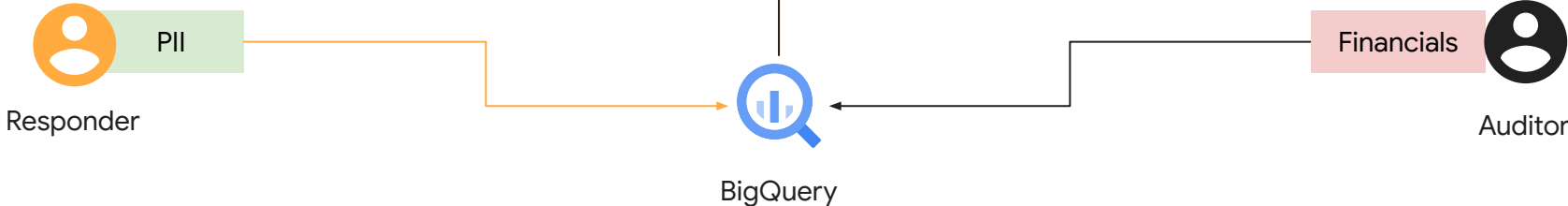


BigQuery - Using Metadata tags to restrict access



Data Catalog

IncidentId	IncidentType	PhoneNum	Location	\$Amount
234698	Mooring	510-45-6789	40.44N, 73.59W	\$10,000
089145	CocInspection	405-94-7201	37.46N, 122.25W	\$25,000,000





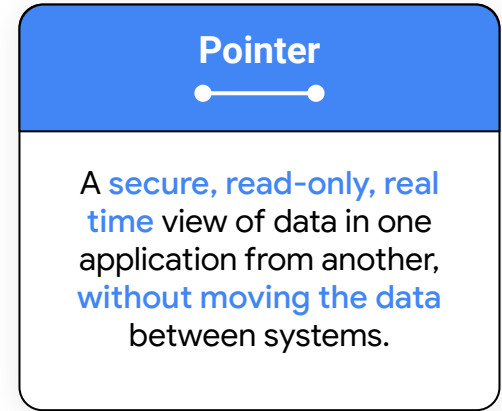
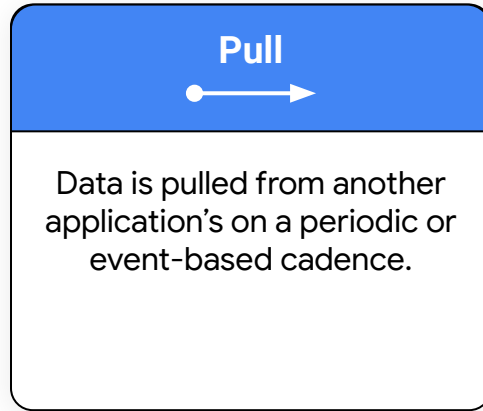
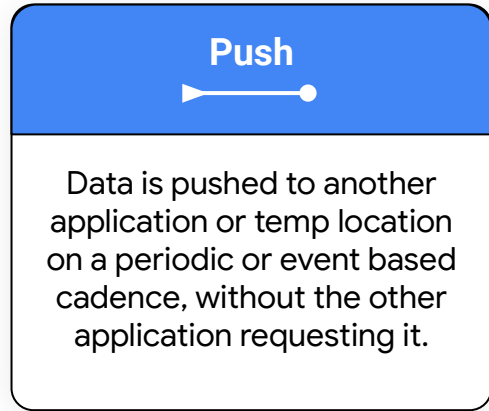
Analytics Hub

Data Clean Rooms

Privacy-centric data sharing and analysis



The Evolution of Data Sharing



← Traditional Model →

Data is copied, adding network and storage costs
Asynchronous feeds: Incremental data is complex
Set-up and governance is shared and can be fragile
No visibility of data usage once shared

← Next Gen Model →

Data remains in place, reducing costs
Updates are available in realtime
User friendly Publish/Subscribe model
Usage Metrics automatically available



Analytics Hub

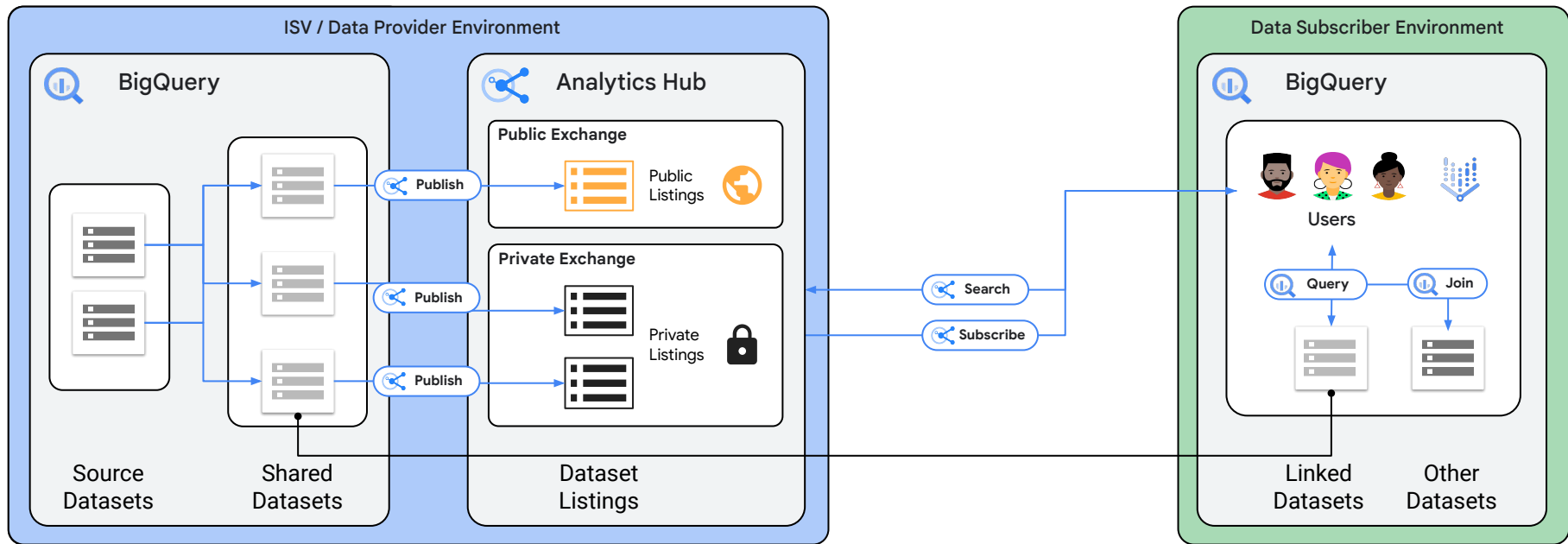
Simple, secure,
zero-copy data
and analytics
assets sharing
via BigQuery

Analytics Hub Ecosystem

250+ data providers | 2,900+
listings | Over 350Pb Shared p/w



Data Sharing Architecture



Shared Datasets are collections of tables and views defined by the **Data Provider** and are the unit of cross-project / cross-org sharing.

Exchanges are collections of **Shared Dataset Listings**. Exchange visibility can be Private, Public, Internal, or Restricted. Permissions are set by the Exchange Admin.

Data Subscribers can search across **Exchanges** and Subscribe to **Dataset Listings** that they are authorized to.

Data Subscribers get an opaque, read-only **Linked Dataset** inside their project that they can query and combine with their own datasets.

Providers pay for storage

No cost for Exchanges & Listings

Subscribers pay for queries

End User (Researcher) Searches and requests access to subscribe to the data.

Analytics Hub
Find and use public, private, and internally shared data sources

Search for listings

Sort by: Relevance

Filters: Clear

Listings

- Public
- Private
- Within my org

Categories (1)

Search

- Advertising & Marketing
- Climate & Environment
- Commerce
- Demographics
- Economics
- Education

+ more

Location

Search

- US (multiple regions in United States)
- EU (multiple)

Results

<p>Open Market Operations The Peoples Bank of China</p> <p>Open market operations conducted by the People's B...</p> <p>Financial Commerce</p>	<p>USASpending Full Database USAspendinggov</p> <p>USAspending.gov includes data on all spending by t...</p> <p>Financial Commerce</p>	<p>Authorized Digital Sellers Phantom Labs</p> <p>Access records from ads.txt, app-ads.txt, and sell...</p> <p>Advertising & Ma... Commerce</p>	<p>Global Insider Model Scores 2iQ Research</p> <p>2iQ leverages over a decade of experience working ...</p> <p>Financial Commerce</p>
<p>Global Insider Transaction Data 2iQ Research</p> <p>Gain access to an accurate and complete global ins...</p> <p>Financial Commerce</p>	<p>Shipping FactSet Research Systems Inc</p> <p>Better understand the increasingly complex global supply chains of 400,000 companies with access t...</p> <p>Commerce</p>	<p>Supply Chain Relationships FactSet Research Systems Inc</p> <p>FactSet Reverse Supply Chain Relationships data is ...</p> <p>Commerce</p>	<p>Google Trends Google Trends</p> <p>Daily top 25 and top 25 rising Search terms across the globe, including the United States</p> <p>Advertising & Ma... Commerce</p>

1 - 10 of 10



Acme Trends Data

US and International Trends Data

+ ADD DATASET TO PROJECT

REQUEST ACCESS

Trends Dataset Access Request

Thank you for your interest in Acme Trends Data via Google Analytics Hub.

Please complete the short form below. A representative will follow up within 1 working day.

* Indicates required question

Company Name *

BigCorp ic.

Your Name *

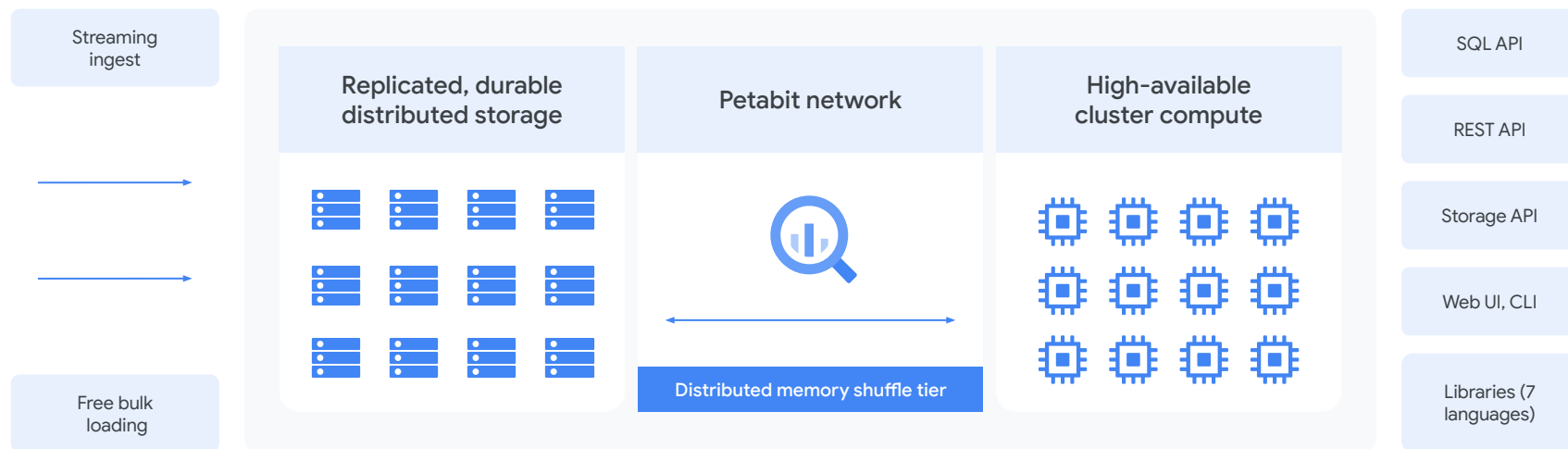
Jane Smith

Business Email *

jane.smith@bigcorp.com

Building on BigQuery

Decoupled storage and compute for maximum flexibility



BigQuery sees more than **6000+ organizations** sharing over **275+ PB's** of data per week.



A data clean room is a secure collaboration environment which allows two or more participants to leverage data assets for specific, mutually agreed upon uses, while guaranteeing the enforcement of strict data access limitations, e.g., not revealing or exposing the personal data of their customers to other parties.

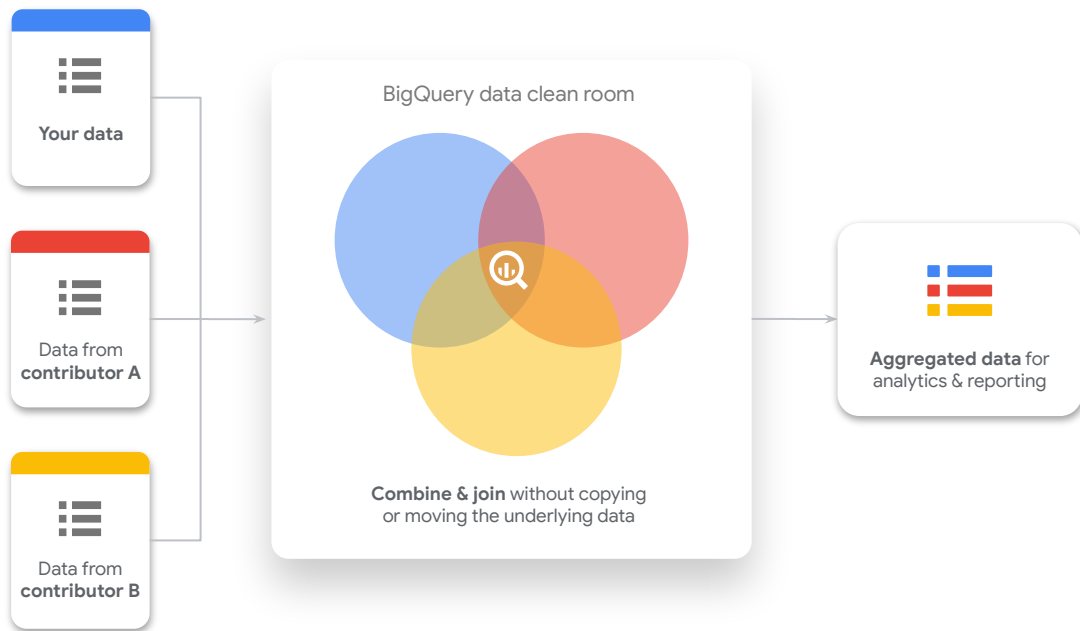
Jeffrey Bustos

VP. Measurement, Addressability & Data Center at IAB

BigQuery - Data clean rooms

Combine, analyze, and securely share sensitive data in a privacy-centric way.

- Create a **low-trust environment** for you and your partners to collaborate without copying or moving the underlying data.
- Perform **privacy-enhancing transformations** in BigQuery SQL interfaces.
- Monitor usage to detect **privacy threats** on shared data.



Data Clean Room - Walkthrough

The screenshot displays the Google Cloud Analytics Hub interface. At the top, there is a navigation bar with the Google Cloud logo, a dropdown menu showing 'dai-storage', and a search bar with the text 'Search for resources, docs, products, and more'. Below the navigation bar, the main content area is titled 'Analytics hub' and includes three action buttons: 'CREATE EXCHANGE', 'CREATE CLEAN ROOM', and 'SEARCH LISTINGS'. A descriptive paragraph explains that Analytics Hub allows users to publish, discover, and subscribe to share BigQuery datasets. Below this, a table lists various data clean room listings. The table has columns for 'Display name', 'Project', 'Region', 'Primary contact', 'Listings', and 'Type'. The 'Display name' column is sorted in ascending order. The table contains six rows of data, each with a vertical ellipsis icon on the right side.

Analytics Hub provides an easy way to publish, discover and subscribe to share BigQuery datasets between users in your organization or other organizations. Once you create an exchange, you can invite others to publish or subscribe to data in the exchange. [Learn more about analytics hub](#)

Filter table

Display name ↑	Project	Region	Primary contact	Listings	Type	
AH exchange	project-id-872683746472829292783	US	shobhitgu@google.com	10	Exchange	⋮
Listing2	project-storage	US	bwelcker@google.com	5	Data clean room	⋮
Sales	dai-storage	EU	nikhilga@google.com	20	Data clean room	⋮
Marketing	tukan-pan-waw	US	ipraveen@google.com	100	Exchange	⋮
Data commons	dai storage	US	varunchandra@google.com	18	Exchange	⋮
Cloud public datasets	dai storage	US	shobhitgu@google.com	27	Exchange	⋮

Create data clean room

The screenshot shows the Google Cloud Analytics Hub interface. On the left, there is a sidebar with navigation options and a table of data exchanges. The main area displays the 'Create data clean room' configuration panel.

Analytics hub

Analytics Hub provides an easy way to publish, discover and subscribe to share BigQuery datasets between users in your organization or other organizations. Once you create an exchange, you can invite others to publish or subscribe to data in the exchange. [Learn more about analytics hub](#)

Filter table

Display name ↑	Project	Region	Primary contact
AH exchange	project-id-872683746472829292783	US	shobhitgu@google.com
Listing2	project-storage	US	bwelcker@google.com
Sales	dau-i-storage	EU	nikhilga@google.com
Marketing	tukan-pan-waw	US	ipraveen@google.com
Data commons	dau-i storage	US	varunchandra@google.com
Cloud public datasets	dau-i storage	US	shobhitgu@google.com

Create data clean room

Data clean room is a specialised solution for privacy centric data sharing. Tell users how it is different from normal data exchanges. [Learn more](#)

1 Clean room configuration

Project *
dau-i-project [BROWSE](#)

Region *
US (multiple region in United States) ▼ ⓘ

Display name*
Advertising data clean room ⓘ

Primary contact*
shobhitgu@google.com ⓘ

Upload icon
shobhitgu@google.com [BROWSE](#)

Description*
Lorem ipsum

This will be visible to the subscribers while finding this data clean room

[CREATE CLEAN ROOM](#) [CANCEL](#)

2 Clean room permissions

Set clean room permissions

The image shows a screenshot of the Google Cloud console. On the left, the 'Analytics hub' page is visible, featuring a table of data exchanges. On the right, the 'Create data clean room' configuration page is shown, with two steps: 'Clean room configuration' and 'Clean room permissions (optional)'. The permissions step includes input fields for 'Clean room owners', 'Data contributors', and 'Subscribers', each with a help icon. At the bottom of the permissions section are 'SET PERMISSIONS' and 'SKIP' buttons.

Analytics Hub Exchanges Table:

Display name	Project	Region	Primary contact
AH exchange	project-id-872683746472829292783	US	shobhitgu@google.com
Listing2	project-storage	US	bwelcker@google.com
Sales	dau-storage	EU	nikhilga@google.com
Marketing	tukan-pan-waw	US	ipraveen@google.com
Data commons	dau storage	US	varunchandra@google.com
Cloud public datasets	dau storage	US	shobhitgu@google.com

Create data clean room

Data clean room is a specialised solution for privacy centric data sharing. Tell users how it is different from normal data exchanges. [Learn more](#)

- Clean room configuration
- Clean room permissions (optional)

Clean room owners *
nikhilga@google.com, muntasir@google.com

Creator of clean room is owner by default.

Data contributors
shobhitgu@google.com

Please add yourself as well if you want to have publisher privileges.

Subscribers
varunchandra@google.com

Please add yourself as well if you want to have subscriber privileges.

SET PERMISSIONS **SKIP**

Clean room created

Google Cloud | dau-storage | Search for resources, docs, products, and more | Search | [Icons]

Analytics hub | CREATE EXCHANGE | CREATE CLEAN ROOM ? | SEARCH LISTINGS

Analytics Hub provides an easy way to publish, discover and subscribe to share BigQuery datasets between users in your organization or other organizations. Once you create an exchange, you can invite others to publish or subscribe to data in the exchange. [Learn more about analytics hub](#)

Filter table [?] [Menu]

Display name ↑	Project	Region	Primary contact	Listings	Type	
AH exchange	project-id-872683746472829292783	US	shobhitgu@google.com	10	Exchange	⋮
Listing2	project-storage	US	bwelcker@google.com	5	Data clean room	⋮
Sales	dau-storage	EU	nikhilga@google.com	20	Data clean room	⋮
Marketing	tukan-pan-waw	US	ipraveen@google.com	100	Exchange	⋮
Data commons	dau storage	US	varunchandra@google.com	18	Exchange	⋮
Cloud public datasets	dau storage	US	shobhitgu@google.com	27	Exchange	⋮

New data clean room created [X]

<1

First time view - Add your data

Google Cloud | dau-storage | Search for resources, docs, products, and more

Advertising - Data clean room | ADD DATA

LISTINGS | SUBSCRIPTIONS | USAGE STATISTICS | DETAILS

This is data clean room where publishers can collaborate and add listings. Clean rooms has fine grained access controls which means that very specific roles can have publishing and subscription access. All the added data/listings need to have analysis rules associated with them. [Learn more](#)

Only data contributors have permission to publish listings into clean room. You are not listed as data contributor. [LEARN MORE](#)

Filter table

Display name ↑ | Source project | Shared dataset | Publisher | Subscribers | Analysis rules

There are no listings in this data clean room. Share link of the clean room with data contributors to add data. Only publishers can publish listings.

[COPY DATA CLEAN ROOM LINK](#)

← Add data

Choose an existing BigQuery dataset and configure it for use within this clean room.

1 Configure Data
Select dataset to add into the clean room

2 Review
Review before adding dataset

NEXT CANCEL

Display name *

Dataset *

Primary contact *

Description

Data Egress Controls

Disable copy and export of shared data

Disable copy and export of query results

Published listings in a clean room

The screenshot shows the Google Cloud console interface for an 'Advertising - Data clean room'. At the top, there's a navigation bar with the Google Cloud logo, a dropdown menu showing 'daii-storage', and a search bar. Below this, the breadcrumb navigation shows 'Advertising - Data clean room' with an 'ADD DATA' button. The main content area has tabs for 'LISTINGS', 'SUBSCRIPTIONS', 'USAGE STATISTICS', and 'DETAILS'. A descriptive paragraph explains that this is a data clean room with fine-grained access controls. Below the text is a notification box stating that only data contributors have permission to publish/edit listings, and the user is not listed as a publisher. A table below the notification lists the published listings with columns for 'Display name', 'Source project', 'Shared dataset', 'Publisher', and 'Subscribers'. The table contains six rows of data, each with a status icon (checkmark or warning) and a link to the listing.

Google Cloud daii-storage Search for resources, docs, products, and more Search

← Advertising - Data clean room ADD DATA

LISTINGS SUBSCRIPTIONS USAGE STATISTICS DETAILS

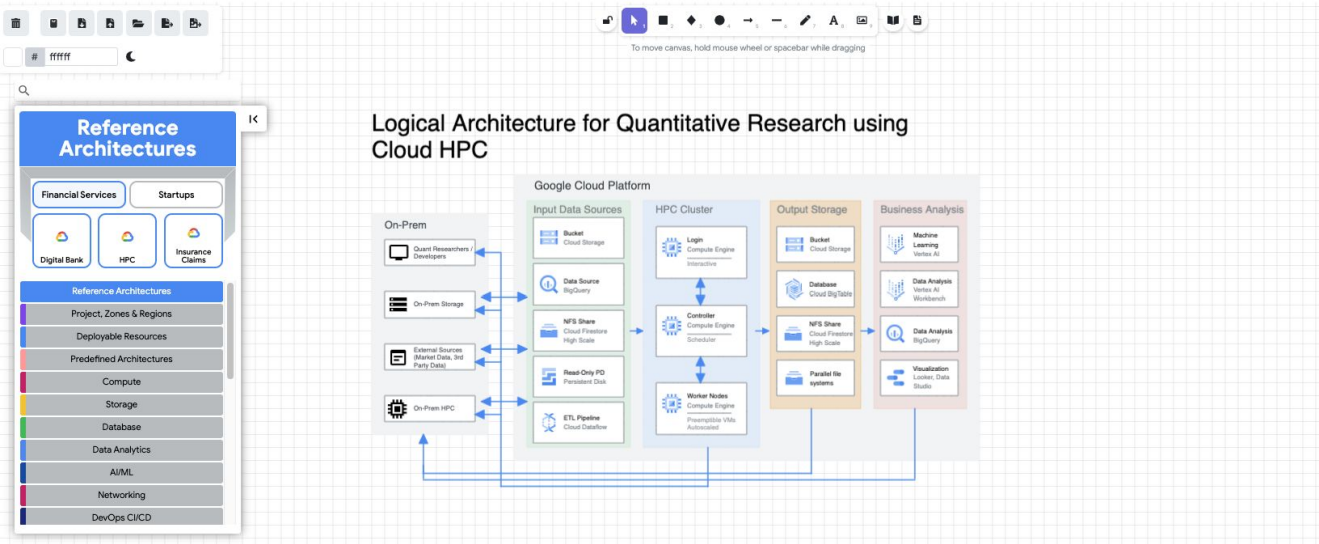
This is data clean room where publishers can collaborate and add listings. Clean rooms has fine grained access controls which means that very specific roles can have publishing and subscription access. All the added data/listings need to have analysis rules associated with them. [Learn more](#)


! Only data contributors have permission to publish/edit listings into clean room. You are not listed as publisher. [LEARN MORE](#)

Filter table

Display name ↑	Source project	Shared dataset	Publisher	Subscribers
✓ Covid dataset	project-id-8726832783	romanomike_source	shobhitgu@google.com	10
! NY bike trips	project-storage	public_dataset	bwelcker@google.com	5
✓ Africa temperature	daii-storage	shobhitgu_dataset	nikhilga@google.com	20
⊖ Sustainability data	tukan-pan-waw	romanomike_source	ipraveen@google.com	100
✓ Data commons	daii storage	romanomike_source	varunchandra@google.com	18
✓ World demographics data	daii storage	romanomike_source	shobhitgu@google.com	27

Blue prints





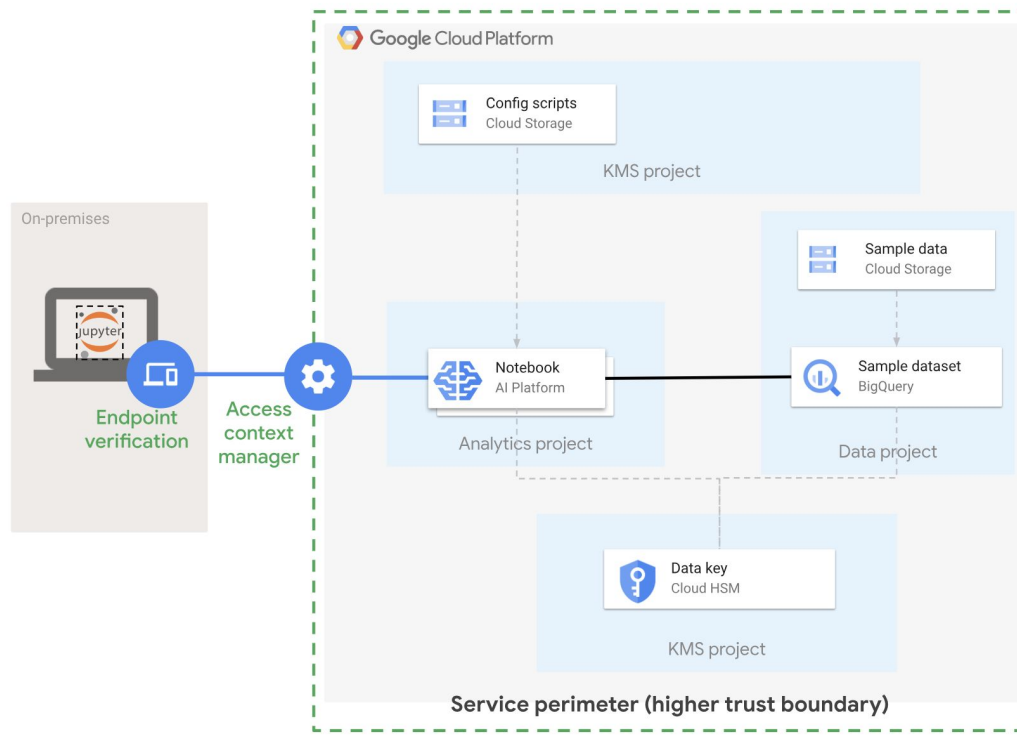
Secure Notebook and Perimeter controls

Blueprint

Protecting confidential data in Vertex AI Workbench user-managed notebooks

The architecture also creates security controls that help you to do the following:

- Mitigate the risk of data exfiltration to a device that is used by data scientists in your enterprise.
- Protect your notebooks instances from external network traffic.
- Limit access to the VM that hosts the notebook instances.



[GitHub repository](#)



Secure Data Warehouse

Blueprint

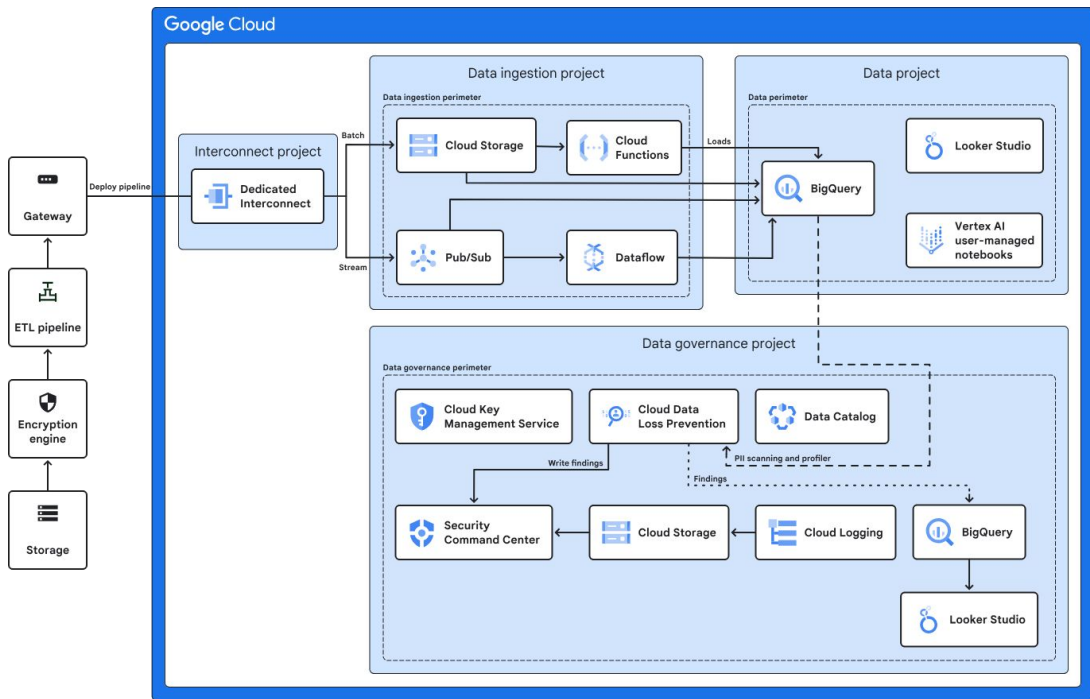
Secure Data Warehouse

If your Project includes confidential data, you must take measures to preserve the security, confidentiality, integrity, and availability of the business data while it is stored, while it is in transit, or while it is being analyzed.

The blueprint helps


- Configure controls that help secure access to confidential data.
- Configure controls that help secure the data pipeline.
- Configure an appropriate separation of duties for different personas.
- Set up templates to find and de-identify confidential data.
- Set up appropriate security controls and logging to help protect confidential data.
- Use data classification and policy tags to restrict access to specific columns in the data warehouse.

Secure Data Warehouse Blueprint



LEARN Tutorial ⓘ ✕

Deploy a secured data warehouse to store confidential data



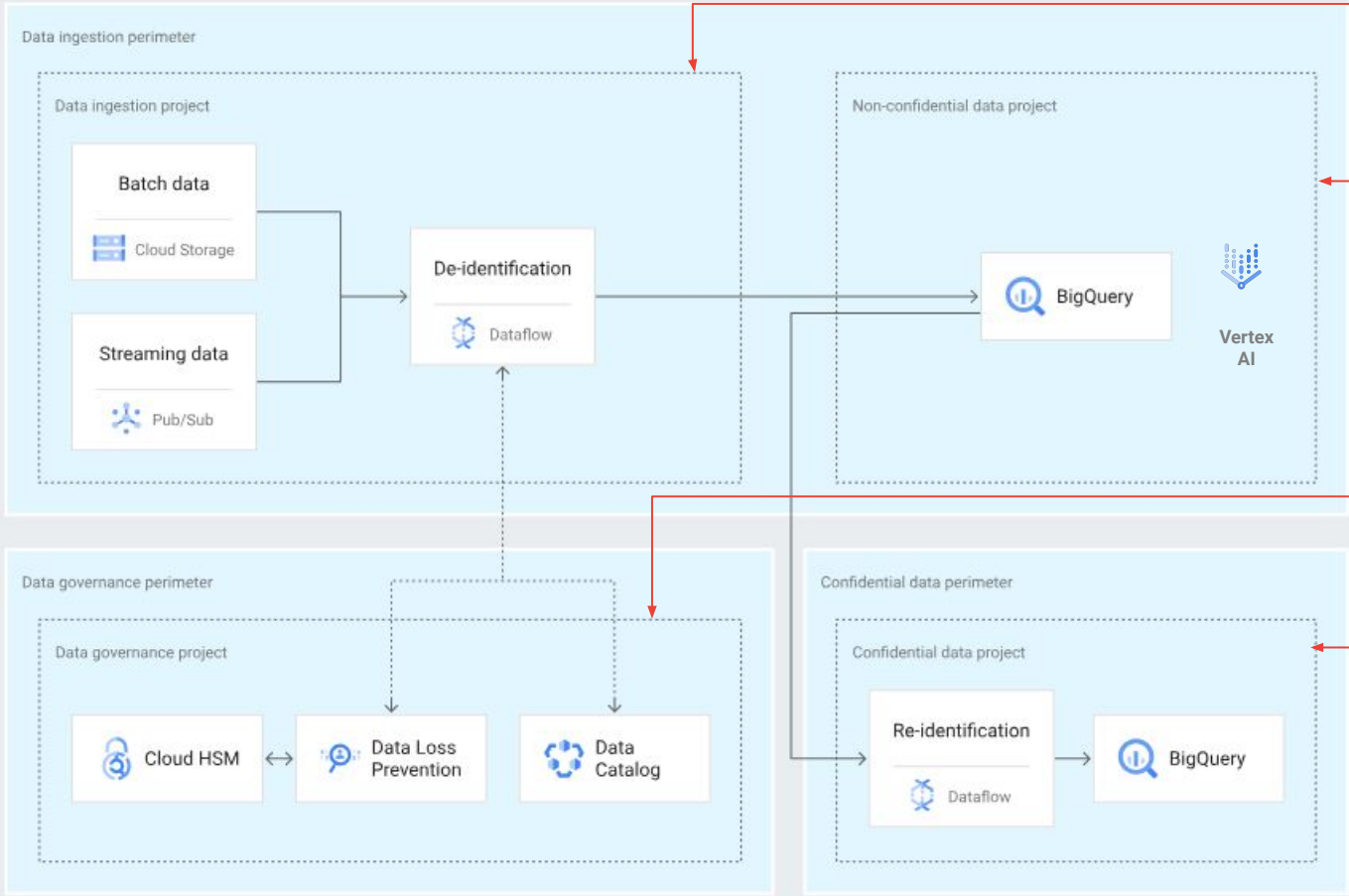
Learn how to use Terraform to deploy a demo of a [secured BigQuery data warehouse](#) that can store confidential data. This tutorial describes how to do the following:

1. Set up your environment (projects, service accounts, groups, and so on).
2. Deploy the Terraform code required to create a BigQuery data warehouse.
3. View the security controls in the deployed environment.
4. Clean up to avoid billing charges.

Before you start

1. Verify that your user identity has the `iam.serviceAccountUser` and `iam.serviceAccountTokenCreator` roles for your organization's development folder, as described in [Organization structure](#). If you do not have a folder that you use for demos.

START



Data Ingestion Project

1

Non-confidential Data Project

2

Data Governance Project

3

Confidential Data Project

4

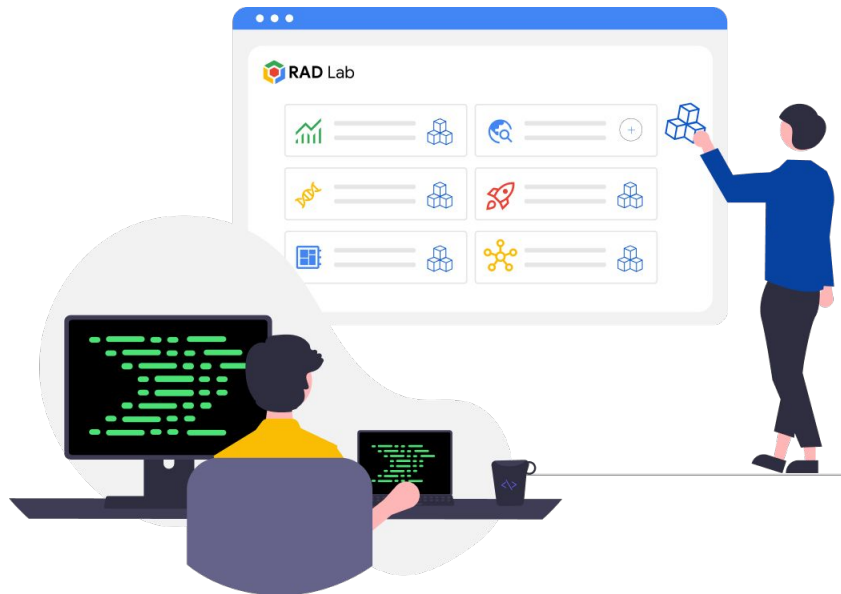


RAD Lab

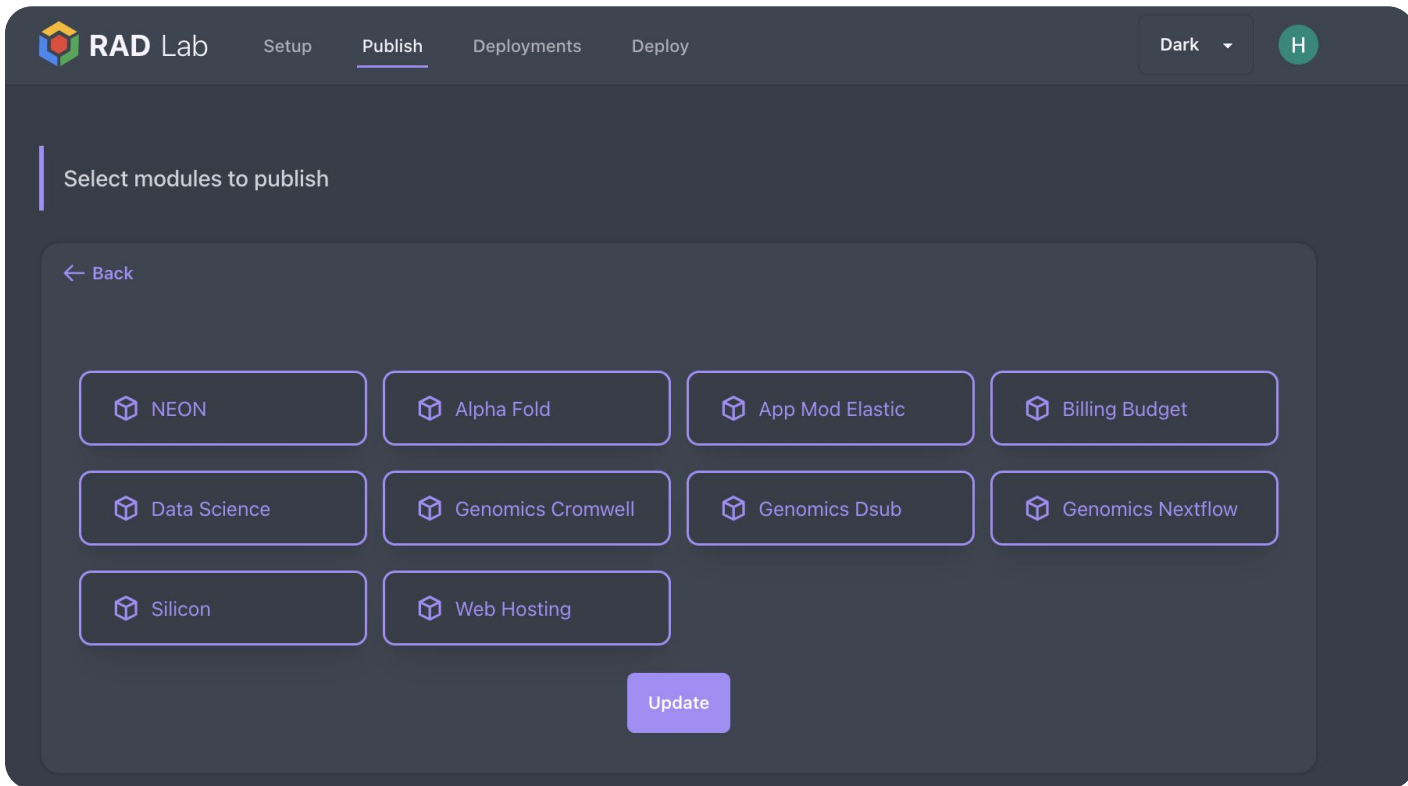
Deploying Secure Data Warehouse Solution



Cloud admins configure
RAD Lab UI for their
organization to **enable**
self-service cloud
deploys



Researchers, analysts, and
other professionals select a
customized cloud
environment for their work
and **deploy with a few**
clicks



The screenshot shows the 'Publish' page in the RAD Lab UI. The top navigation bar includes the RAD Lab logo, the text 'RAD Lab', and menu items for 'Setup', 'Publish' (which is underlined), 'Deployments', and 'Deploy'. On the right side of the navigation bar, there is a 'Dark' theme selector with a dropdown arrow and a circular profile icon containing the letter 'H'. Below the navigation bar, the main content area has a heading 'Select modules to publish' followed by a '← Back' link. The modules are displayed in a grid of 12 items, each with a cube icon and a label: NEON, Alpha Fold, App Mod Elastic, Billing Budget, Data Science, Genomics Cromwell, Genomics Dsub, Genomics Nextflow, Silicon, and Web Hosting. At the bottom center of the grid is a purple 'Update' button.

Setup Publish Deployments Deploy

Dark ▾ H

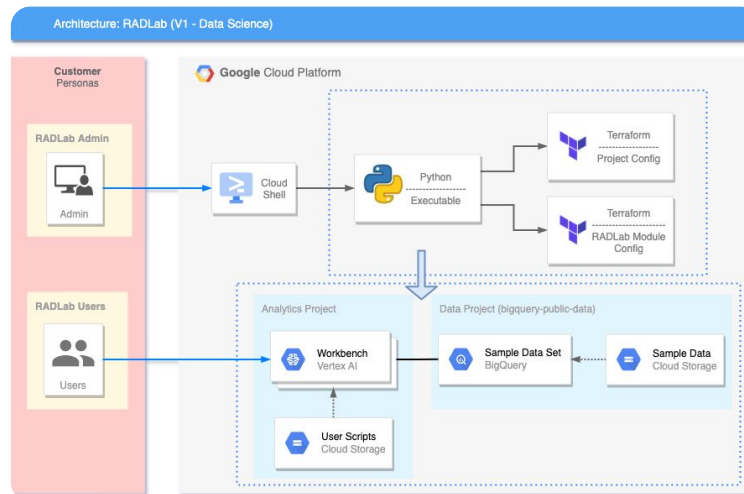
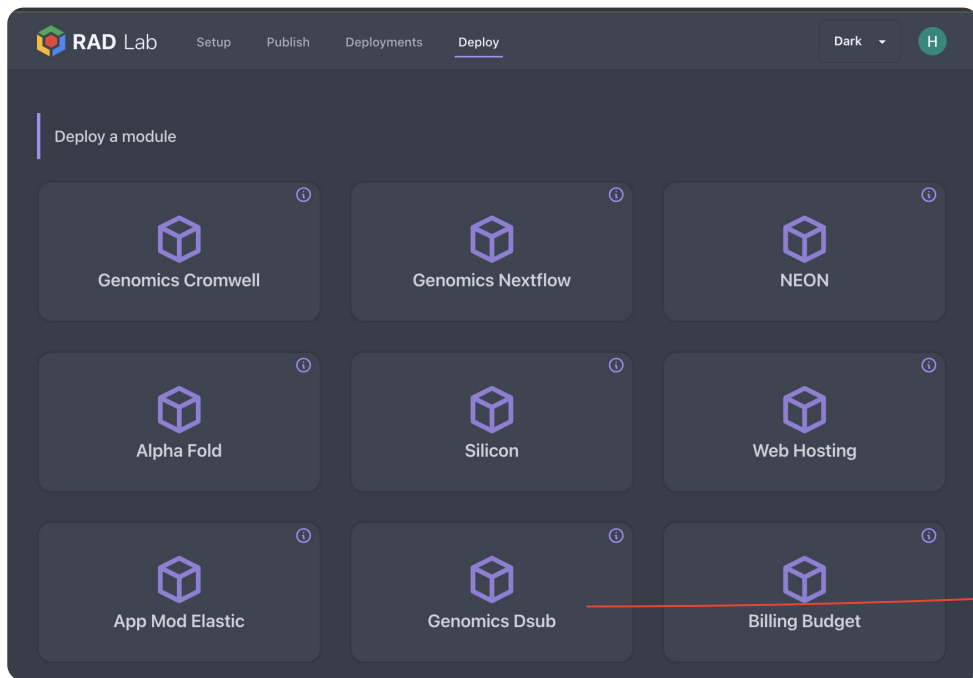
Select modules to publish

← Back

- NEON
- Alpha Fold
- App Mod Elastic
- Billing Budget
- Data Science
- Genomics Cromwell
- Genomics Dsub
- Genomics Nextflow
- Silicon
- Web Hosting

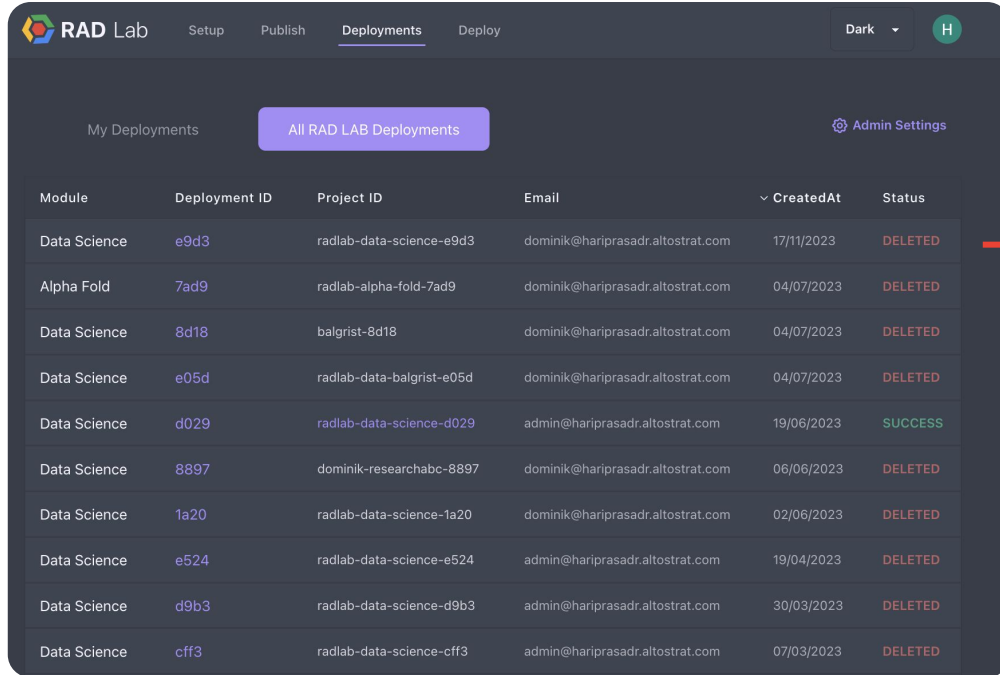
Update

RAD Lab UI - self service model for researchers



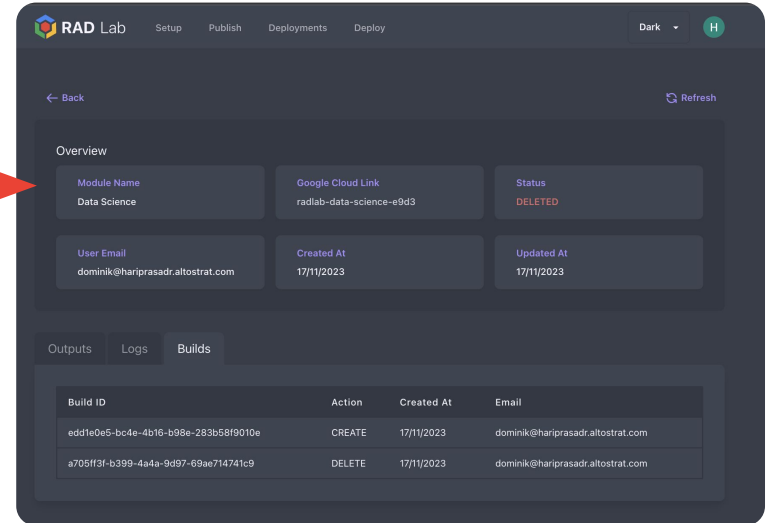
*Click to
deploy*

Admin - Monitor and Delete deployments



The screenshot shows the 'Deployments' page in the RAD Lab Admin interface. It features a navigation bar with 'Setup', 'Publish', 'Deployments', and 'Deploy' tabs. A 'Dark' theme toggle and a user profile icon 'H' are also present. Below the navigation, there are buttons for 'My Deployments' and 'All RAD LAB Deployments', along with an 'Admin Settings' link. The main content is a table listing various deployments with columns for Module, Deployment ID, Project ID, Email, CreatedAt, and Status.

Module	Deployment ID	Project ID	Email	CreatedAt	Status
Data Science	e9d3	radlab-data-science-e9d3	dominik@hariprasadr.altostrat.com	17/11/2023	DELETED
Alpha Fold	7ad9	radlab-alpha-fold-7ad9	dominik@hariprasadr.altostrat.com	04/07/2023	DELETED
Data Science	8d18	balgrist-8d18	dominik@hariprasadr.altostrat.com	04/07/2023	DELETED
Data Science	e05d	radlab-data-balgrist-e05d	dominik@hariprasadr.altostrat.com	04/07/2023	DELETED
Data Science	d029	radlab-data-science-d029	admin@hariprasadr.altostrat.com	19/06/2023	SUCCESS
Data Science	8897	dominik-researchabc-8897	dominik@hariprasadr.altostrat.com	06/06/2023	DELETED
Data Science	1a20	radlab-data-science-1a20	dominik@hariprasadr.altostrat.com	02/06/2023	DELETED
Data Science	e524	radlab-data-science-e524	admin@hariprasadr.altostrat.com	19/04/2023	DELETED
Data Science	d9b3	radlab-data-science-d9b3	admin@hariprasadr.altostrat.com	30/03/2023	DELETED
Data Science	cff3	radlab-data-science-cff3	admin@hariprasadr.altostrat.com	07/03/2023	DELETED




The screenshot shows the 'Details' page for a specific deployment in the RAD Lab Admin interface. It includes a 'Back' button and a 'Refresh' icon. The 'Overview' section displays key information in a grid: Module Name (Data Science), Google Cloud Link (radlab-data-science-e9d3), Status (DELETED), User Email (dominik@hariprasadr.altostrat.com), Created At (17/11/2023), and Updated At (17/11/2023). Below this, there are tabs for 'Outputs', 'Logs', and 'Builds'. The 'Builds' tab is active, showing a table of build records.

Build ID	Action	Created At	Email
edd1e0e5-bc4e-4b16-b98e-283b58f9010e	CREATE	17/11/2023	dominik@hariprasadr.altostrat.com
a705ff3f-b399-4a4a-9d97-69ae714741c9	DELETE	17/11/2023	dominik@hariprasadr.altostrat.com

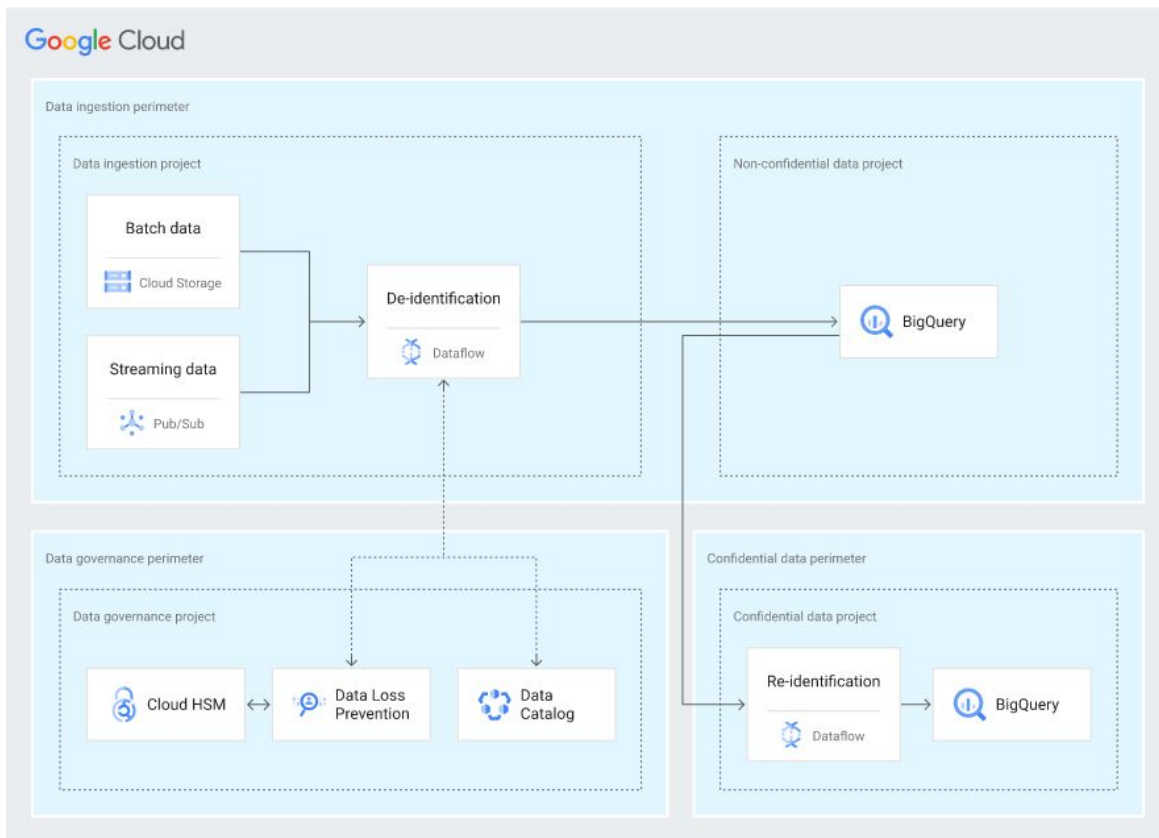
Secure data warehouse



Quick deploy using RAD Lab
as a self serve module

 OPEN IN GOOGLE CLOUD SHELL

<https://github.com/GoogleCloudPlatform/rad-lab>





Thank you.