Capital Group

POV via. Ahead

RSAC Innovation Sandbox 2024 FINALIST

AICPA SOC

BEDROCK SECURITY

Table of contents

ℹ️ In this report, each **underlined** segment is a deep link into Bedrock that can be clicked.

Try it out!

BEDROCK SECURITY

Current state of data security at Capital Group

- Incomplete visibility into critical data across all data platforms
  - *"We don't know what the data looks like. We don't know where our data is." - Khanh*
  - ↳ Data policy violations, leaks or exposures may be unnoticed
- Manual data classification and tagging lacks accuracy (Resource Constraints)
  - *"[User] might say this is non-sensitive, but then six months later it's sensitive, they don't go back and change that data tagging" - Khanh*
  - ↳ Unable to write accurate data policy/controls to keep data safe w/ growth and GenAI
- Limited visibility into data access, usage, and movement (Data Management)
  - "What kind of sharing policies do we have, but also show me all the data that's associated with this particular identity or line of business." - *Khanh*
  - ↳ Cannot track exposure of data as access levels change or as data is moved

The S.A.F.E approach to securing data

Scalable Data should be secured without gaps in multi PB environments such as CG's Caspian data lake

Accurate Identification of sensitive data and risks must be reliable for both regulatory data (e.g. NPI) and proprietary business data (e.g. financial trades)

Fast Data scanning and risk assessment techniques must deliver rapid time to value and capture exposures / control violations in near real-time

Efficient Data security initiatives should not result in costly infrastructure bills

S.A.F.E performance with Bedrock in Ahead PoV

Scalable — **600TB scanned daily for 1 week**

- AWS, Azure, Snowflake, Databricks & M365

Accurate — **Regulated data and proprietary data identified**

- PII/NPI, PCI, Financial Transactions Records and Invoices
- Transactions and Invoices were auto-discovered by Bedrock's ML

Fast — **Under 9 hours** required to scan 600TB

Efficient — **Order of magnitude** more cost efficient scanning

Key out of box findings: Data discovery

		Capital Group's Custom Data Learned by Bedrock								Regulated Data	
		Secrets / Keys	Analytics Data	Developer Data	Financial Data	HR Data	Sales Data	Security Data	Logs	PII/PI/NPI†	PCI†
Infrastructure / Staging	AWS S3			33	7‡				8	43	40
	AWS DBs*									3	3
	AWS EFS			1						1	1
	Azure Blob									3	3
	Snowflake									1	1
Corp	OneDrive		19	19	1	9	6	2		93	
	Sharepoint	1	3		4	2	4			227	

Colors represent relative risk for data in a given location

* Databases include AWS RDS and Unmanaged SQL

† Identified data is likely synthetic / simulated

BEDROCK SECURITY

Key out of box findings: Custom data discovered

Bedrock's ML automatically categorizes and labels the data it scans. Below are examples of data custom to the Ahead PoV environment that Bedrock identified, including simulated financial transactions.

Data category	Examples of custom data types learned by Bedrock
Secrets / Keys	AWS Credentials, Database Credentials
Financial Data	Invoice Details, Purchase Orders, Tax Documents, Synthetic Financial Transactions
HR Data	Health Authorization, Resume, Onboarding Documents, Offshore Staffing, Termination Notices
Sales Data	Carrier Tracking, Dell Server Quotes, Production Orders

Key out of box findings: Sensitive data access

- Identified 402 identities w/ sensitive access to PII/NPI, PCI and Transactions data
  - ↳ Decommission 2 stale, high impact human identities, inactive for over 6 months
  - ↳ Investigate 206 unknown identities from domain @dccplbng for sensitive access
  - ↳ Revoke PII shared with personal @gmail accounts: Uma and Sushma

	Infrastructure (e.g. AWS, Azure, Snowflake)	Corporate (e.g. OneDrive, Sharepoint)
Human	**17 (2 stale)**	**152 (@dhruvts.com)**
Non-human	**24**	**206 (@dccplbng)**
Third-party	**1 (Databricks)**	**2 (@gmail.com)**

Key out of box findings: Sensitive data leakage

- Sensitive database backups show up in 13 personal OneDrives / Sharepoints
  - ↳ Delete database backups
- PII residing in non-production assets (created by John Soto)
  - ↳ Quarantine PII in high PII volume assets using built-in remediations
- Data from service database i-04a823e99f856141b leaking outside of AWS
  - ↳ Delete data copied into Azure Blob Store

Roadmap

Integrations

Q4	Wiz (integration already approved)
Q4	Enhancements to Databricks e.g. unity catalog
Y25	On-premise sources based on roadmap agreement

Platform

Q3	Data catalog and Data Bill of Materials; (shipped)
Q3	Chained entitlement visualization; (shipped 2nd generation)
Q4	Least-privilege recommendations
Y25	Investigation and remediation copilot, impact analysis

Beyond DSPM: Stakeholder value, present and future

- **Data context for security efforts**
  - SIEM and CNAPP alert correlation with sensitive data
  - Improved classification for DLP
  - Accelerated least privilege initiatives: reduce sensitive entitlements
- **Enterprise data management**
  - Automated data labeling and metadata lake
  - Data retention, deduplication, minimization
- **Generative AI**
  - Data bill of materials for model training
- **Intellectual property tracking**
  - Tracking specific sensitive data content and trust boundary management

Appendix

Metrics and screenshots

RSAC Innovation Sandbox 2024 FINALIST

AICPA SOC

BEDROCK SECURITY

S.A.F.E performance details in Ahead PoV

Bedrock's adaptive sampling scanned ~600 TB of data daily for 1 week. The **total time spent scanning was less than 9 hours** and the **infrastructure cost to Ahead was less than $1,000**.

Workload	Data Volume	Bedrock Performance
AWS (S3, EFS, RDS)	554 TB	2.23 Hours
Unmanaged SQL Server	4.1 GB	10 Minutes
Databricks on S3	250 MB	2 Minutes
Snowflake	67 GB	3 Minutes
Azure Blob Store	15 TB	9 Hours (Factoring in retry)
OneDrive & Sharepoint	10.6 GB	8.13 Hours

Secrets / keys in Sharepoint

PRODUCTION capitalgroup

Search... ⌘K

NAVIGATION

- Inventory
- Data Dashboard
- Findings
- Policies
- Data Catalog
- Reports

SETTINGS

- Users
- Integrations
- Authentication
- Data Type Manager
- Configurations

Feature Flags

SharePoint:SharePoint/CG - Ahead

Overview | Contents | Data Exposure | Access Activity | Similar Data | Visualizations | Unresolved Findings

Search for Datasets

Export

More filters

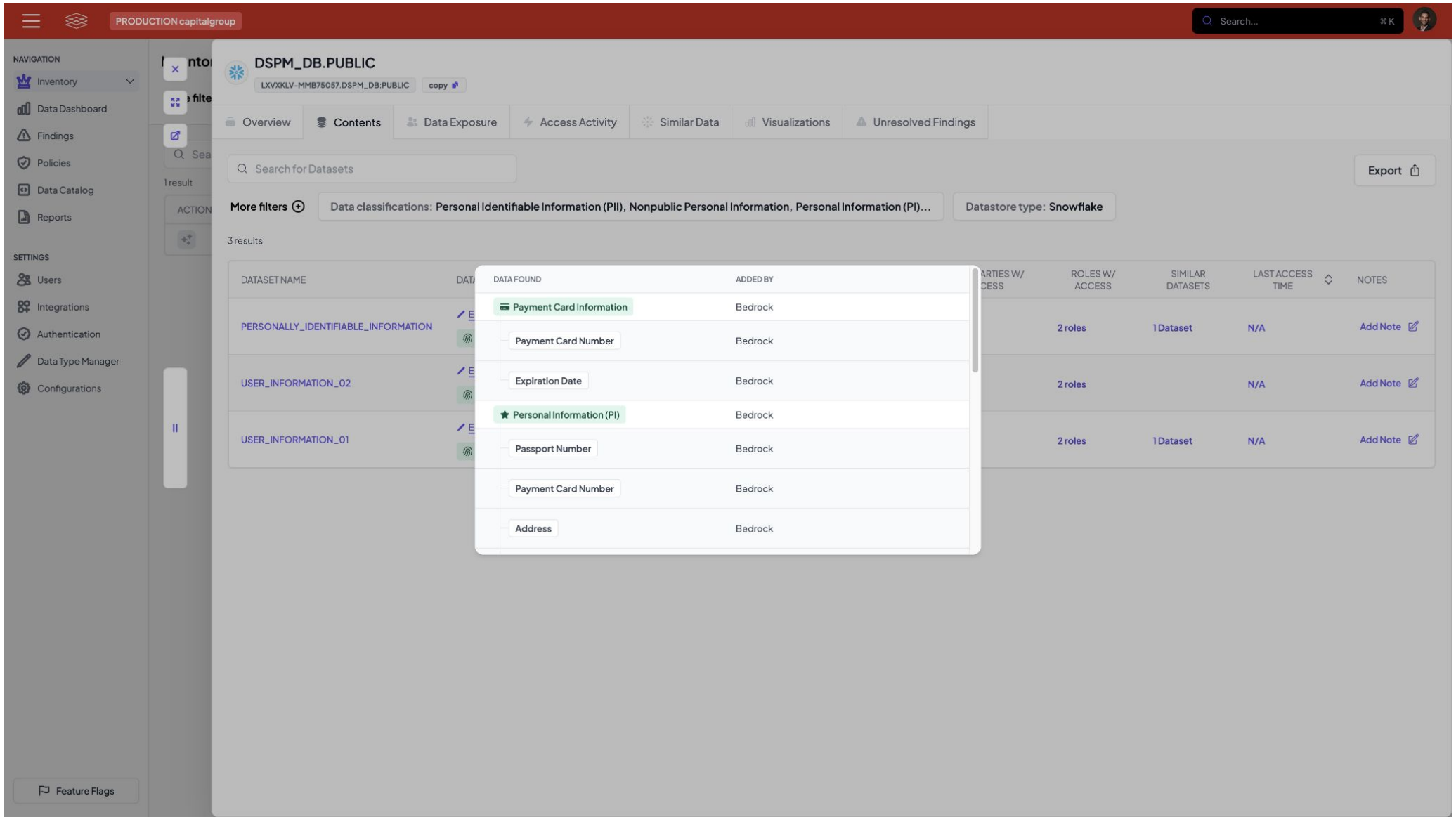
Data classifications: Access

1 result

DATASET NAME	DATA CATEGORY	USERS W/ ACCESS	THIRD PARTIES W/ ACCESS	ROLES W/ ACCESS	SIMILAR DATASETS	LAST ACCESS TIME	NOTES
Documents/Cloud Infrastructure/dspm aws creds cons.zip	Edit Access		1 third party			N/A	Add Note

BEDROCK SECURITY

PII/PI/NPI/PCI in Snowflake



BEDROCK SECURITY

High impact human identities, inactive for >6mos

PRODUCTION capitalgroup

Search... ⌘K

NAVIGATION

- Inventory
- Data Dashboard
- Findings
- Policies
- Data Catalog
- Reports

SETTINGS

- Users
- Integrations
- Authentication
- Data Type Manager
- Configurations

Feature Flags

Inventory

More filters | Entity: Principal | Time since last usage: Greater than 26 Weeks | Principal types: User | Accessible data categories: *No Selection*

Search 2 Principals

Export

2 results

ACTIONS	NAME	IMPACT	ACCESSIBLE DATA CATEGORIES	IDENTITY ORIGIN	PRINCIPAL TYPE	LAST ACTIVITY
	aws-avtr-user	●●●●	Engineering Data, Synthetic Data, +6	AWS	USER	9:01am, Feb 21st 2024
	shiva.p	●●●●	Synthetic Data, Personal Identifiable Information (PII), +3	AWS	USER	12:09am, Jul 7th 2023

BEDROCK SECURITY

Data shared externally w/ Uma’s @gmail

PRODUCTION capitalgroup

Search... ⌘K

NAVIGATION

- Inventory
- Data Dashboard
- Findings
- Policies
- Data Catalog
- Reports

SETTINGS

- Users
- Integrations
- Authentication
- Data Type Manager
- Configurations

Feature Flags

Uma Shankar Reddy A

ROLE

Overview | Role Accessibility | Roles | Accessible Data | Principal Activity | Unresolved Findings | Cloud Provider Tags | Visualizations

Accessible Data

Showing the datastores that are accessible by this principal

Export

Search 1 Datastore

More filters | Data classifications: *No Selection* | Account name: *No Selection*

1 results

DATASTORE NAME	DATASTORE TYPE	IMPACT	DATA CATEGORY	ROLE(S) USED TO ACCESS	DATASET COUNT	ACCOUNT ENVIRONMENT	ACCESS TYPE	RELATIONSHIP TO DATA STORE	ACCESS TO DATASETS
SharePoint:SharePoint/Global Dhruv Team	SharePoint	● ● ● ●	Personal Identifiable Information (PII) +2	0 Roles	238 Datasets	Staging	Admin, Write, Read	-	Partial

- Personal Identifiable Information (PII)
  - Name
  - Email Address
- Nonpublic Personal Information
  - Name
- Personal Information (PI)
  - Name
  - Email Address

BEDROCK SECURITY

Data shared externally w/ Sushma’s @gmail

PRODUCTION capitalgroup

Search... ⌘K

NAVIGATION

- Inventory
- Data Dashboard
- Findings
- Policies
- Data Catalog
- Reports

SETTINGS

- Users
- Integrations
- Authentication
- Data Type Manager
- Configurations

Feature Flags

sushma BN

ROLE

Overview | Role Accessibility | Roles | Accessible Data | Principal Activity | Unresolved Findings | Cloud Provider Tags | Visualizations

Accessible Data

Export

Showing the datastores that are accessible by this principal

Search 1 Datastore

More filters | Data classifications: *No Selection* | Account name: *No Selection*

1 results

DATASTORE NAME	DATASTORE TYPE	IMPACT	DATA CATEGORY	ROLE(S) USED TO ACCESS	DATASET COUNT	ACCOUNT ENVIRONMENT	ACCESS TYPE	RELATIONSHIP TO DATA STORE	ACCESS TO DATASETS
OneDrive:OneDrive/Jyothi Swarup Reddy	M365	● ● ● ●	Personal Identifiable Information (PII) +2	0 Roles	1 Dataset	Staging	Admin, Write, Read	-	Partial

- Personal Identifiable Information (PII)
  - Name
  - Email Address
- Nonpublic Personal Information
  - Name
- Personal Information (PI)
  - Email Address
  - Name

BEDROCK SECURITY

Sensitive DB backups leaking to OneDrive/Sharepoint

PRODUCTION capitalgroup

Search... ⌘K

NAVIGATION

- Inventory
- Data Dashboard
- Findings
- Policies
- Data Catalog
- Reports

SETTINGS

- Users
- Integrations
- Authentication
- Data Type Manager
- Configurations

Feature Flags

Policies > Sensitive database backups should only stay in AWS

Sensitive database backups should only stay in AWS

ACTIVE

Edit Policy | Delete Policy | Pause Policy

Overview | Unresolved Findings | Entities Affected

Entities Affected

Showing all entities that are involved in the findings created by this policy

Export

Search entity names

More filters

13 ENTITIES SHOWN

ENTITY NAME	DATA OWNER	ENTITY CLASSIFICATION	ACCESSIBLE DATA CATEGORIES
OneDrive:OneDrive/Akram Pasha	-	M365	Personal Identifiable Information (PII), Nonpublic Personal Information, +1
OneDrive:OneDrive/Arpitha SA	-	M365	HR Data, Personal Identifiable Information (PII), +2
OneDrive:OneDrive/Ananthan	-	M365	Security Data, Personal Identifiable Information (PII), +2
SharePoint:SharePoint/GKB Vision	-	SharePoint	Engineering Data, Personal Identifiable Information (PII), +2
SharePoint:SharePoint/Cytiva	-	SharePoint	Engineering Data, Personal Identifiable Information (PII), +2
SharePoint:SharePoint/Dhruv BI Quick Start Solution	-	SharePoint	Personal Identifiable Information (PII), Nonpublic Personal Information, +1
SharePoint:SharePoint/Dixon AMS Internal	-	SharePoint	Personal Identifiable Information (PII), Nonpublic Personal Information
SharePoint:SharePoint/Dhruv Rebranding	-	SharePoint	Personal Identifiable Information (PII), Nonpublic Personal Information, +1
SharePoint:SharePoint/Pacbio Internal	-	SharePoint	Personal Identifiable Information (PII), Nonpublic Personal Information, +1
OneDrive:OneDrive/Dennis Mosley	-	M365	Analytics Data, Engineering Data, +4
OneDrive:OneDrive/Gagan C	-	M365	Personal Identifiable Information (PII), Nonpublic Personal Information, +1
OneDrive:OneDrive/GANSHEKAR RAAJ H	-	M365	Personal Identifiable Information (PII), Nonpublic Personal Information, +1
OneDrive:OneDrive/Harish N	-	M365	Analytics Data, Personal Identifiable Information (PII), +2

BEDROCK SECURITY

Data shared externally w/ Sushma’s @gmail

PRODUCTION capitalgroup

Search... ⌘K

NAVIGATION

- Inventory
- Data Dashboard
- Findings
- Policies
- Data Catalog
- Reports

SETTINGS

- Users
- Integrations
- Authentication
- Data Type Manager
- Configurations

Feature Flags

sushma BN

ROLE

Overview | Role Accessibility | Roles | Accessible Data | Principal Activity | Unresolved Findings | Cloud Provider Tags | Visualizations

Accessible Data

Export

Showing the datastores that are accessible by this principal

Search 1 Datastore

More filters | Data classifications: *No Selection* | Account name: *No Selection*

1 results

DATASTORE NAME	DATASTORE TYPE	IMPACT	DATA CATEGORY	ROLE(S) USED TO ACCESS	DATASET COUNT	ACCOUNT ENVIRONMENT	ACCESS TYPE	RELATIONSHIP TO DATA STORE	ACCESS TO DATASETS
OneDrive:OneDrive/Jyothi Swarup Reddy	M365	● ● ● ●	Personal Identifiable Information (PII) +2	0 Roles	1 Dataset	Staging	Admin, Write, Read	-	Partial

- Personal Identifiable Information (PII)
  - Name
  - Email Address
- Nonpublic Personal Information
  - Name
- Personal Information (PI)
  - Email Address
  - Name

BEDROCK SECURITY

Remediate sensitive data outside of production

PRODUCTION capitalgroup

Search... ⌘K

NAVIGATION
- Inventory
- Data Dashboard
- Findings
- Policies
- Data Catalog
- Reports

SETTINGS
- Users
- Integrations
- Authentication
- Data Type Manager
- Configurations

Feature Flags

user-registration-data-usw-1 - Test - Sensitive Data should not reside in non-sensitive area

Oct 1st 2024, 5:48pm (1d ago) ● Unresolved

Resolve Finding | Open Ticket | Ignore finding

Overview | Impact Summary | Data Exposure | Access Activity | Compliance | Remediation

Isolate S3 Bucket

Isolate the bucket using the following Cloud Formation Template

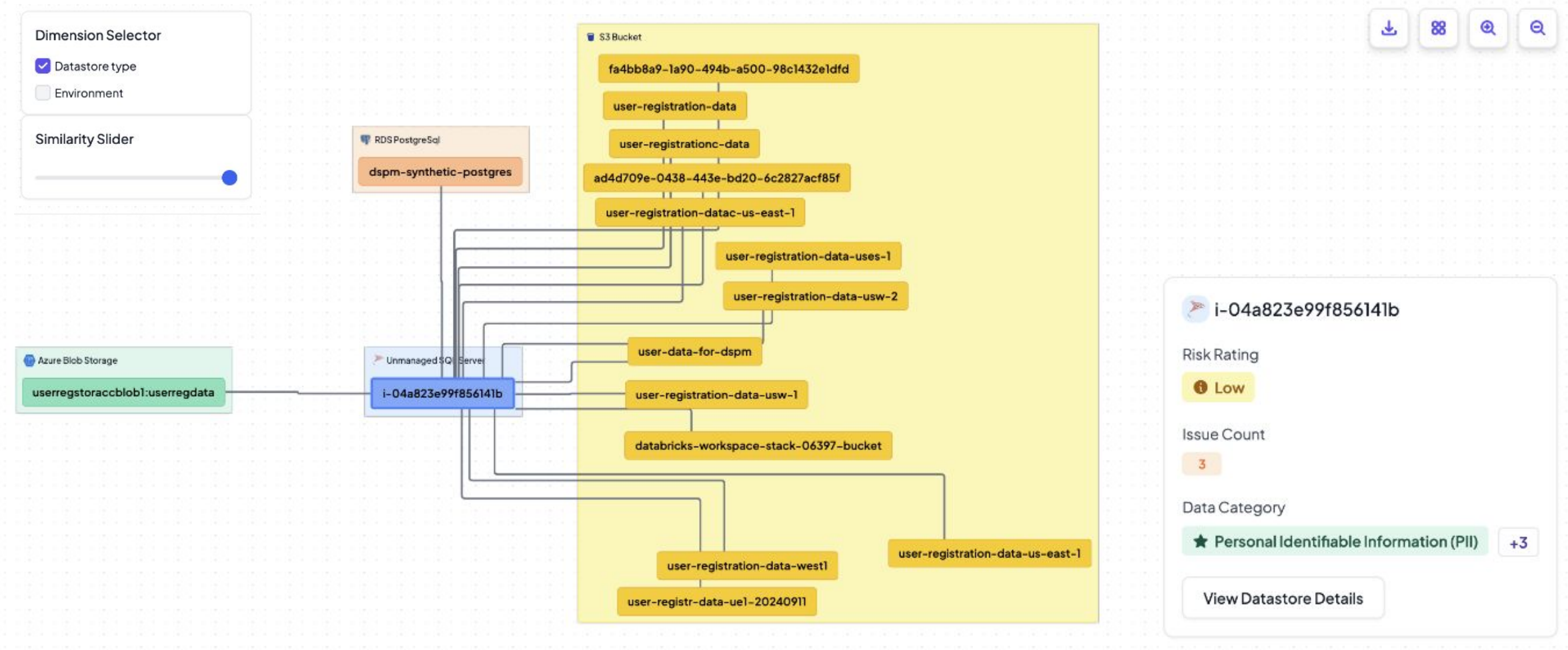
Please add the administrators in the cloud formation template or do the following steps:

1. **Create a JSON file** (e.g., `bucket_policy.json`) containing the bucket policy. Replace `administrator_arns`
with the ARNs of the IAM users or groups you want to grant access to.

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "QuarantineBucket",
      "Effect": "Deny",
      "Principal": "*",
      "Action": "s3:*",
      "Resource": [ .
        "arn:aws:s3:::user-registration-data-usw-1",
        "arn:aws:s3:::user-registration-data-usw-1/*"
      ],
      "Condition": {
        "StringNotLike": {
          "aws:userId": [
            "administrator_arn",
            "administrator_arn2"
            // ...additional administrator ARNs, if any
          ]
        }
      }
    }
  ]
}
```

BEDROCK SECURITY

Data leaving core service DB into other services



BEDROCK SECURITY