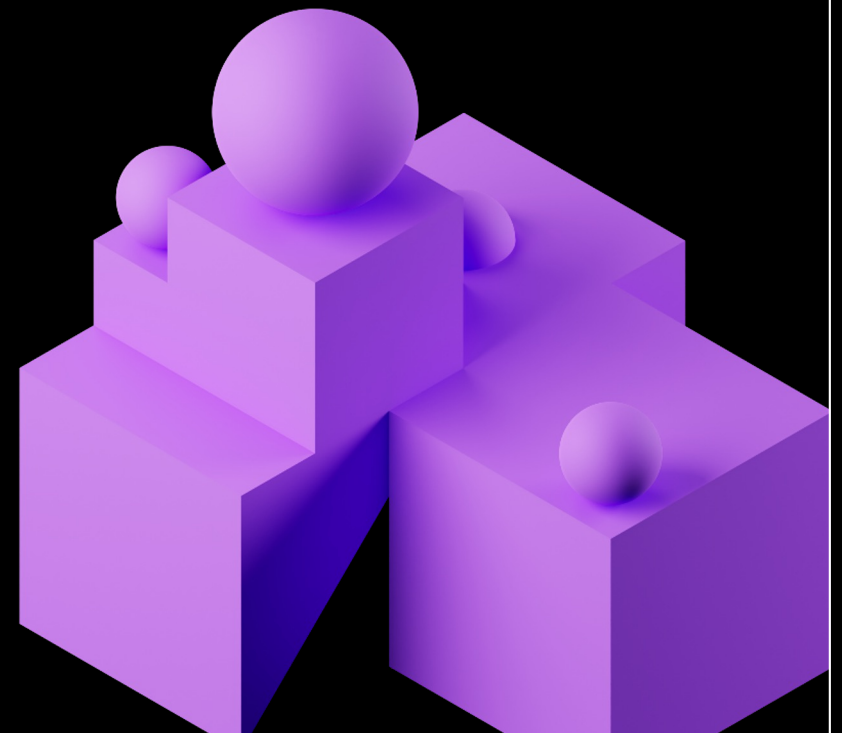


DATA+AI SUMMIT

BY  databricks

The Future of Data Access Control: Booz Allen Hamilton's Approach to Securing our Databricks Lakehouse with Immuta

Speaker: Jeffrey Hess



Databricks
2023

About Me



Jeffrey Hess

Lead Technologist
Booz Allen Hamilton



@jeffhessdata



Data Platforms and Applications Team



Data Design, Security, and Visualization



Bachelor's from The University of New Mexico



Cleveland, OH, USA



Golf, Cooking, Traveling



About Booz Allen Hamilton

Empower People to Change the World



Defense | Intelligence | Civil | Global/Commercial

Booz Allen Hamilton is a leading professional services company, providing a broad range of services and solutions in management, technology, consulting, and engineering

Founded in 1914

Booz | Allen | Hamilton®

Headquarters – McLean, VA, USA

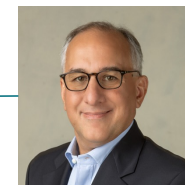
Employees – ~31,900

Revenue – \$9.3 Billion



**Horacio D.
Rozanski**

President and
Chief Executive
Officer

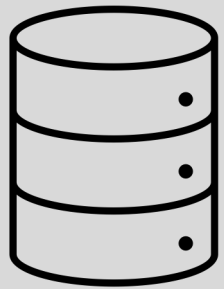


**Matt
Calderone**

Executive Vice
President and
Chief Financial
Officer



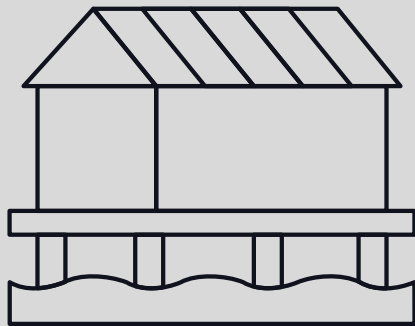
Migration from Oracle to Databricks



Oracle Data Warehouse



Databricks Data Lakehouse



1

Cost Reduction

2

Decrease Resource Needs

3

Increase Data Storage and Formats



The Immuta Data Security Platform delivers integrated sensitive data discovery, security, and monitoring, so organizations can simplify operations, improve security, and unlock more value from their data.





Discover Sensitive Data Across Data Sources

Discover and classify sensitive data with high confidence levels

Name	Type	Description	Actions
first_name	text	Patient first name	
Discovered ... Person Name X Discovered ... PII X			
last_name	text	Patient last name	
Discovered ... Person Name X Discovered ... PII X			



Scale & Automate Policies Using Dynamic ABAC

Write policies once, enforce everywhere with dynamic, native data security and privacy controls

Global Policy Builder

What's the name of this policy?

Mask PII

How should this protect the data?

Mask columns tagged Discovered ... ID Number X

using hashing for everyone except when user



Achieve Provable Compliance

Generate reports easily to prove compliance based on automatically logged data access

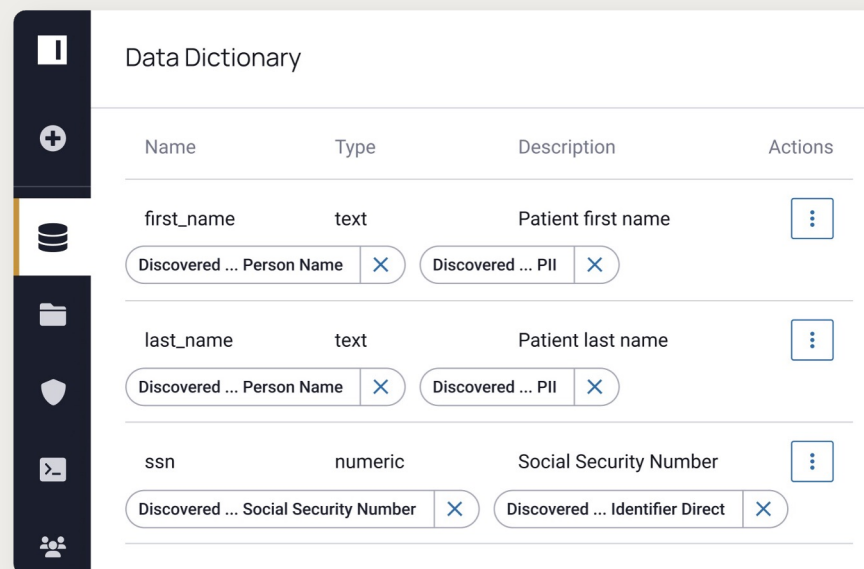
Audit		
Results can be filtered by Purpose, Blob ID, or Remote Query ID.		
Time	Data Source	Outcome
Project		
17 JUN 2022 01:25:29	Global Policy Applied	MHar
16 JUN 2022 08:39:10	Global Policy Applied	MHar
15 JUN 2022 09:12:43	Global Policy Applied	MHar
15 JUN 2022 12:11:01	Global Policy Applied	MHar



Data Discovery

 Discover & Classify

1. Connect to Any Data Source
2. Centralize Metadata
3. Apply Standard Tagging



The image shows a 'Data Dictionary' interface. On the left is a dark sidebar with icons for a document, a plus sign, a database cylinder, a folder, a shield, a code editor, and a group of people. The main area has a title 'Data Dictionary' and a table with columns: Name, Type, Description, and Actions. The table lists three fields: 'first_name' (text), 'last_name' (text), and 'ssn' (numeric). Each field has a 'Description' and one or more 'Discovered' tags with an 'X' icon to remove them. The 'first_name' and 'last_name' rows have two tags each ('Person Name' and 'PII'). The 'ssn' row has two tags ('Social Security Number' and 'Identifier Direct'). Each row also has a vertical ellipsis icon in the 'Actions' column.

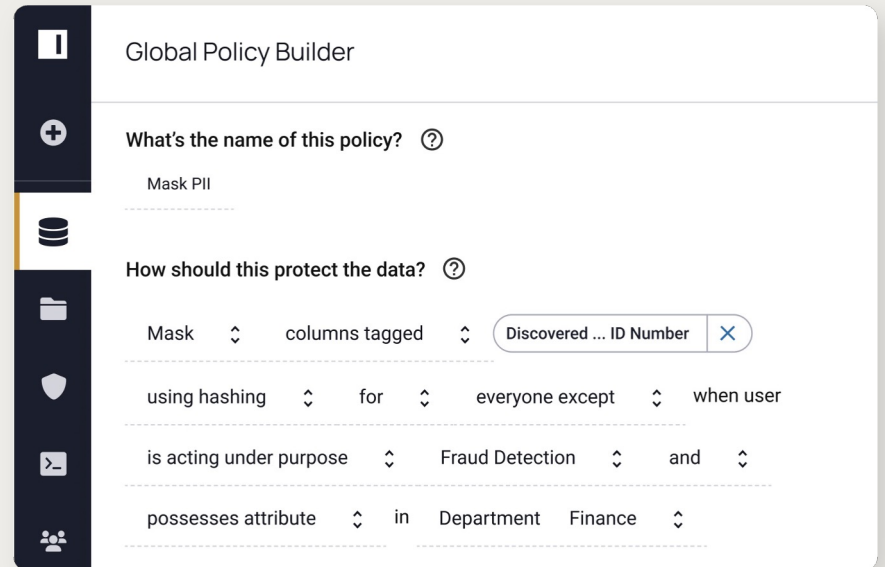
Name	Type	Description	Actions
first_name	text	Patient first name	
		Discovered ... Person Name	Discovered ... PII
last_name	text	Patient last name	
		Discovered ... Person Name	Discovered ... PII
ssn	numeric	Social Security Number	
		Discovered ... Social Security Number	Discovered ... Identifier Direct



Data Security

Create & Manage Policies

1. Author Cross Platform Global Policies
2. Uniform Row-, Column- and Cell-level Protection
3. Easy-to-Understand Policies for Any Role



The screenshot shows the 'Global Policy Builder' interface. It features a dark sidebar on the left with icons for a document, a plus sign, a database, a folder, a shield, a code editor, and a group of people. The main area is white and contains the following sections:

- Global Policy Builder**
- What's the name of this policy?** (with a help icon) - The input field contains 'Mask PII'.
- How should this protect the data?** (with a help icon) - This section contains a series of dropdown menus and text fields:
 - Mask (dropdown) columns tagged (dropdown) Discovered ... ID Number (text field with a close button)
 - using hashing (dropdown) for (dropdown) everyone except (dropdown) when user
 - is acting under purpose (dropdown) Fraud Detection (dropdown) and (dropdown)
 - possesses attribute (dropdown) in Department Finance (dropdown)



Data Auditing



Enforce & Audit

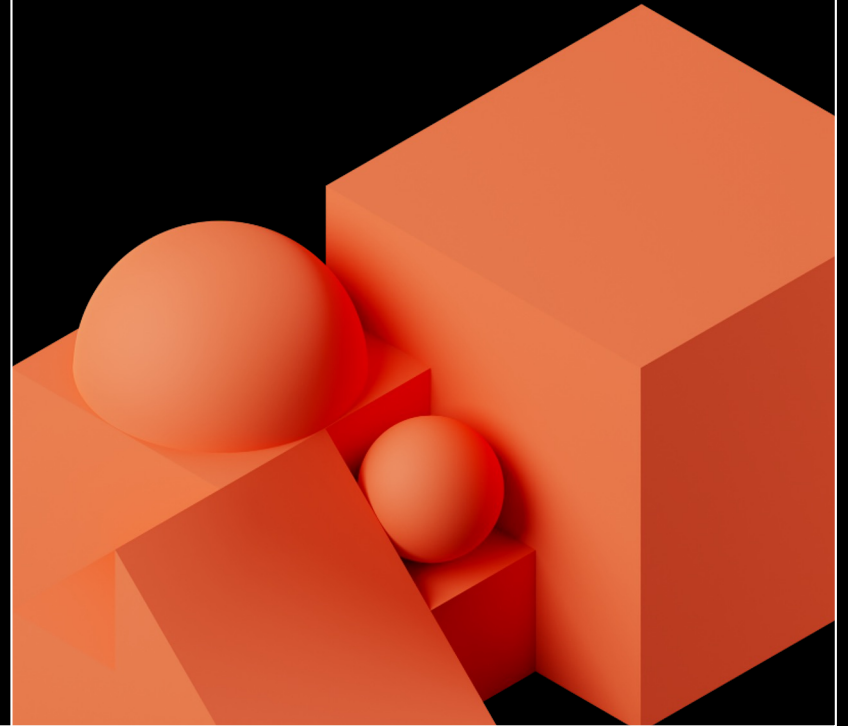
1. Transparent Enforcement at Query Time
2. Streamlined Data Request Workflows
3. Restrict and Log Access Based on Purpose / Intent
4. Unified Policy Logs

The screenshot displays the 'Audit' section of the Azure Policy console. On the left is a dark sidebar with navigation icons. The main area is divided into two panels. The left panel, titled 'Audit', contains filter options: 'Time' (with a date range from Tue Apr 12 2021 to Fri May 6 2022), 'Data Source', 'Outcome', 'Project', 'Record Type', and a checked checkbox for 'Global Policy Applied'. The right panel, titled 'Records', shows a table of audit results.

Timestamp	Record Type	User
17 JUN 2022 01:25:29	Global Policy Applied	MHar
16 JUN 2022 08:39:10	Global Policy Applied	MHar
15 JUN 2022 09:12:43	Global Policy Applied	MHar
15 JUN 2022 12:11:01	Global Policy Applied	MHar
14 JUN 2022 14:05:16	Global Policy Applied	MHar



ACCESS CONTROL



Role Based Access Control (RBAC) was created way back in the 90s as a way to easily bucket or group users to provide them access to data



Role Based Access Control



Role: Manager



Region	Rep	Profit
East	B Smith	50000
South	C Johnson	100000



Role: Sales



State	Sales	Profit
New York	30000	15000
Colorado	20000	10000



Role: HR



Employee	Emp ID	Office Location
W Davis	45432	Tampa, FL
P Anderson	98473	San Diego, CA

- Organizations are frequently changing
- Technology is constantly changing
- Role Explosion
- Administrative Bottlenecks

Attribute Based Access Control



Role: Manager
Security Clearance: Cleared
Level: 4
Location: USA
Team: East



Role: Sales
Security Clearance: Cleared
Level: 2
Location: USA
Team: East



Role: HR
Security Clearance: Not Cleared
Level: 1
Location: Canada
Team: West



State	Sales	Profit
New York	30000	15000
Colorado	20000	10000

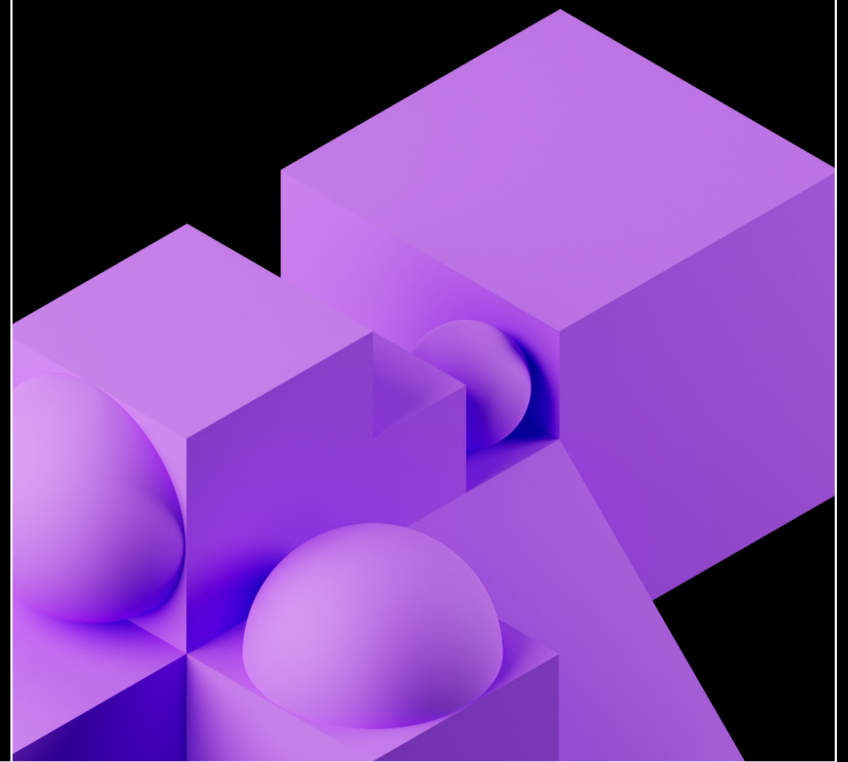
Policy:
Cleared Individuals
Level 2 and above
USA and Canada
East Region

Employee	Emp ID	Office Location
W Davis	45432	Tampa, FL
P Anderson	98473	San Diego, CA

Policy:
Cleared and Uncleared
All Levels
USA and Canada
West Region

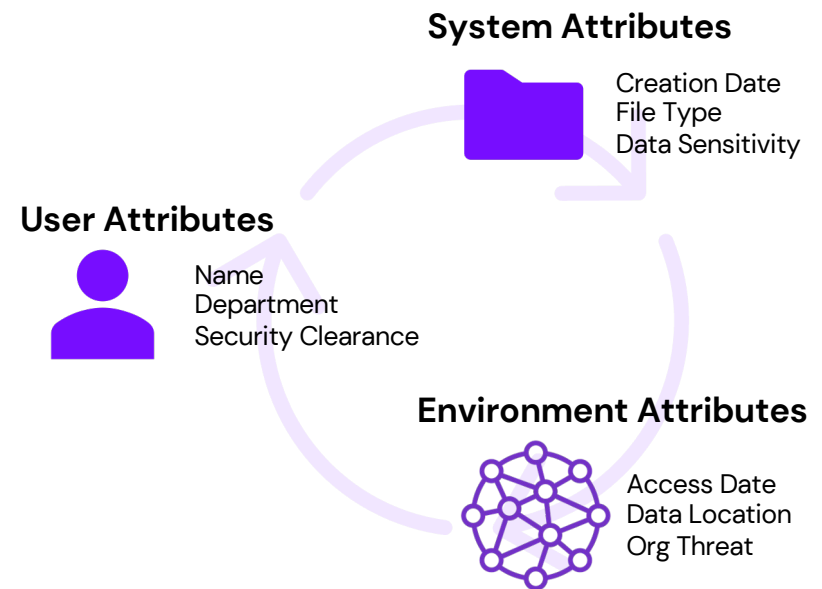


ATTRIBUTES AND HOW WE CREATE THEM

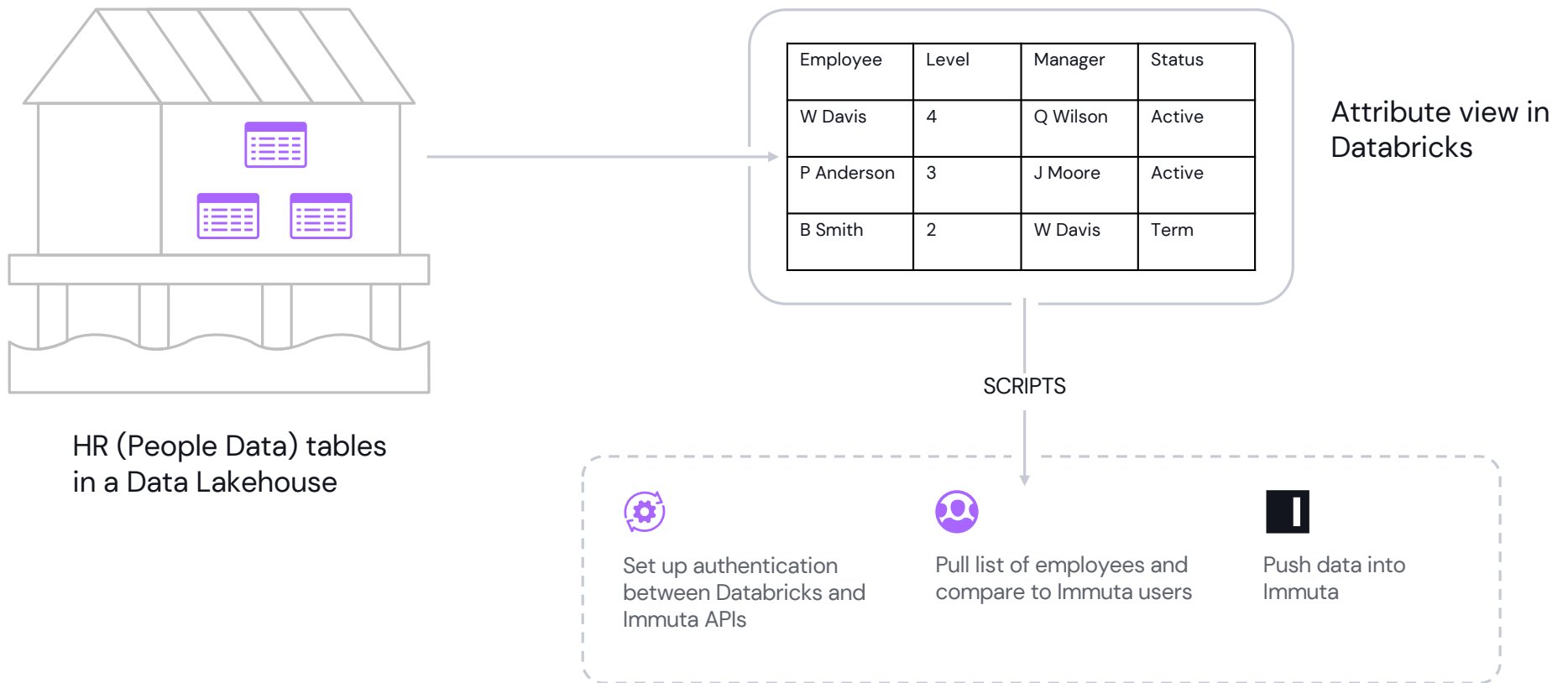


What are Attributes?

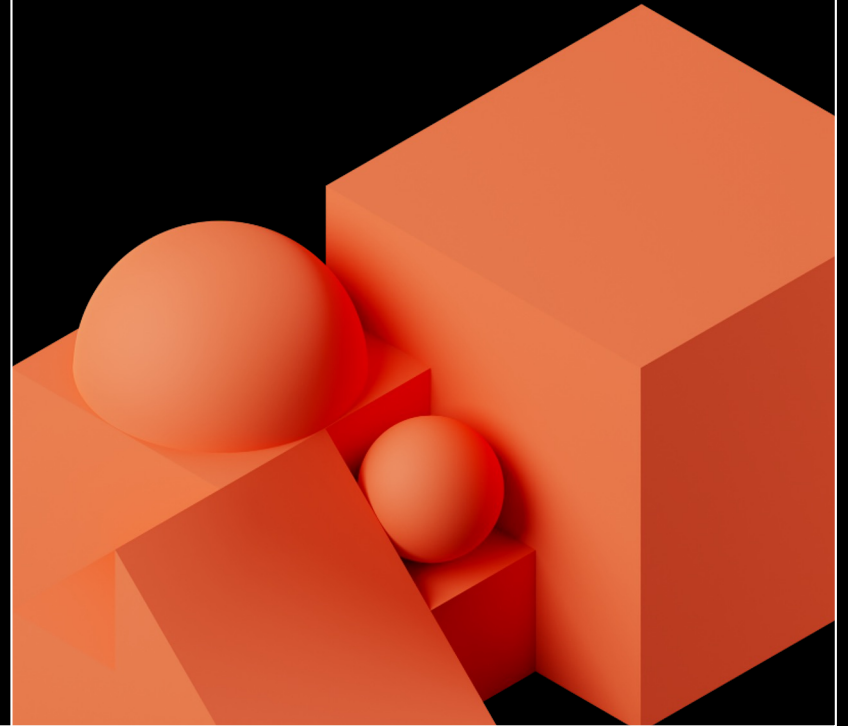
- System Users
- Objects Within the System
- Actions Taken by the Users
- The Environment Itself



Adding Custom Attributes to Immuta



DISCOVERING AND TAGGING DATA



Sensitive Data Discovery

Scan, Classify, and Tag

- Discover sensitive information from millions of **fields without manual effort.**
- Apply 60+ prebuilt classifiers alongside domain-specific, custom classifiers based on a desired confidence level **without worrying about false positives.**
- Enable different teams to inspect tags through workflows that certify data has been properly identified and tagged.
- Build your own sensitive data discovery algorithm for greater flexibility



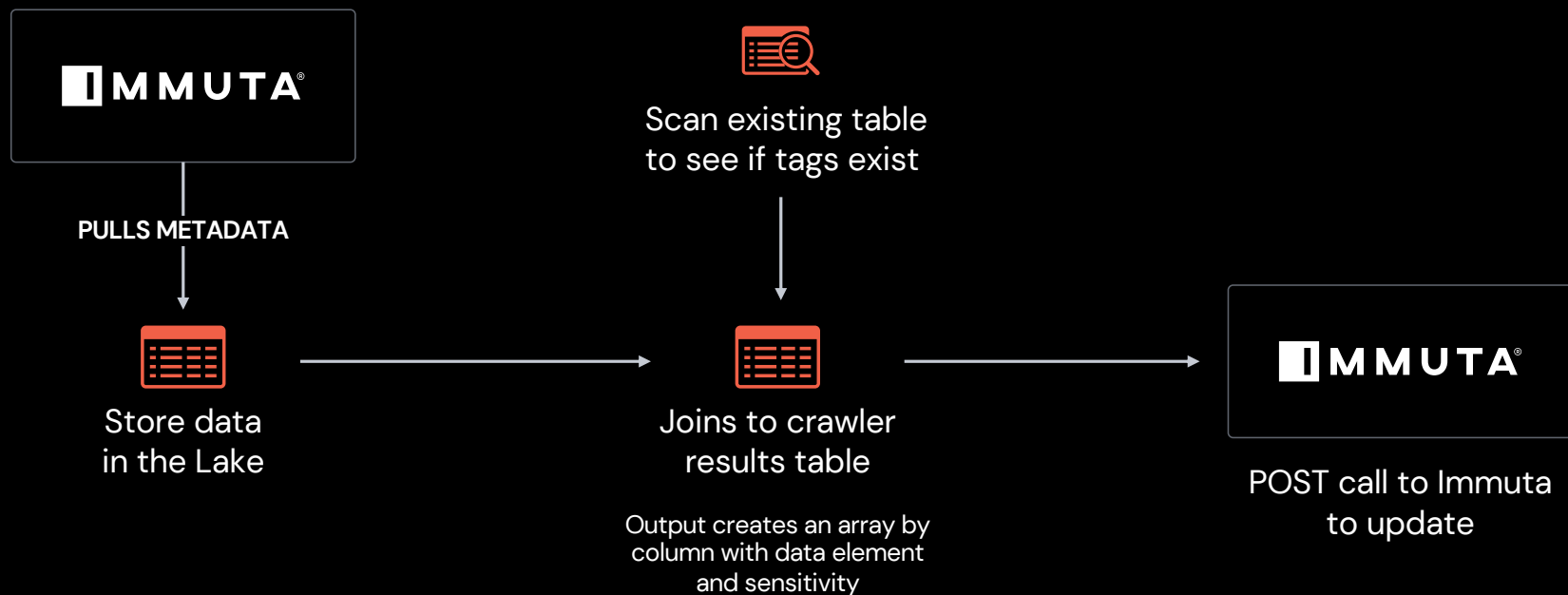
Data Crawlers

Crawler stats

- 20K+ tables
- Identified & tagged fields
 - 1700 sensitive
 - 180 critically sensitive



Apply Custom Tags via Immuta API



Tags must exist in Immuta first before they can be added via API



CONCLUSION

