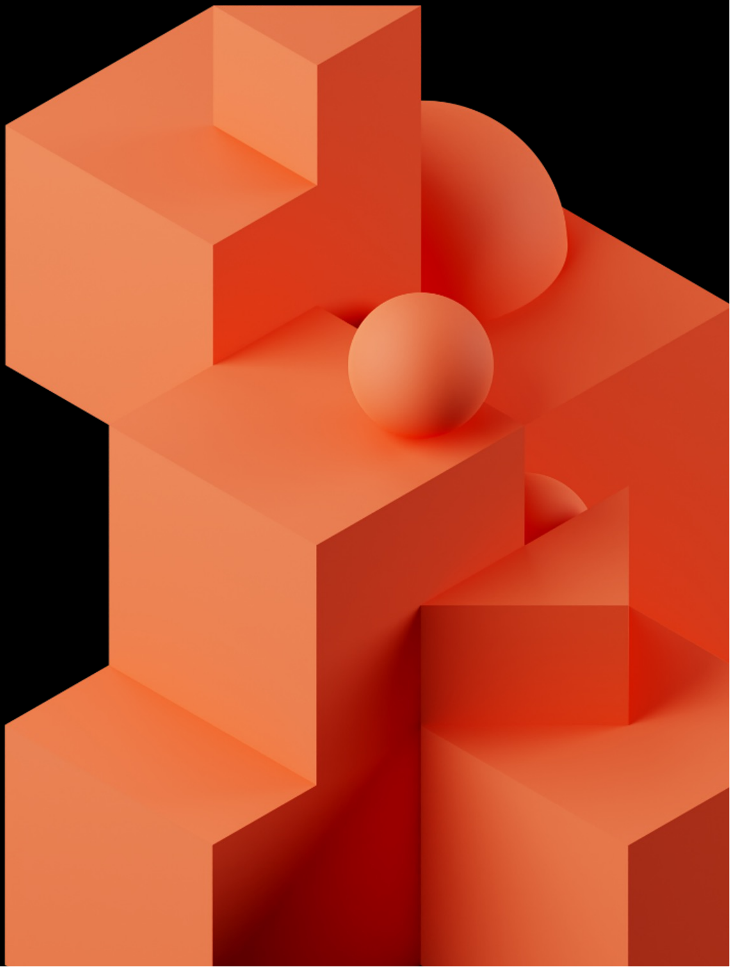
DATA+AI SUMMIT
BY databricks

Lakehouse Architecture to Advance Security Analytics at the Department of State

Tim Ahrens
Brendan Barsness

Databricks
2023



Agenda

1 What is the Department of State’s Center for Analytics (CfA)

2 The Challenge

3 Technical
The Response and Recommendations

4 Organizational
The Response and Recommendations

Center for Analytics

Who We Are

CfA is the Department of State's enterprise data management and analytics capability.

Led by the Chief Data Officer, we transform data into bold insights that help make better management and foreign policy decisions.

Who We Support

We empower employees across every bureau and over 200 missions, from working-level to the Secretary.

What is State Department M/SS CfA?

ANALYTICS

Accelerate Decisions through Analytics

DATA CULTURE

Cultivate a Data Culture

DATA MANAGEMENT

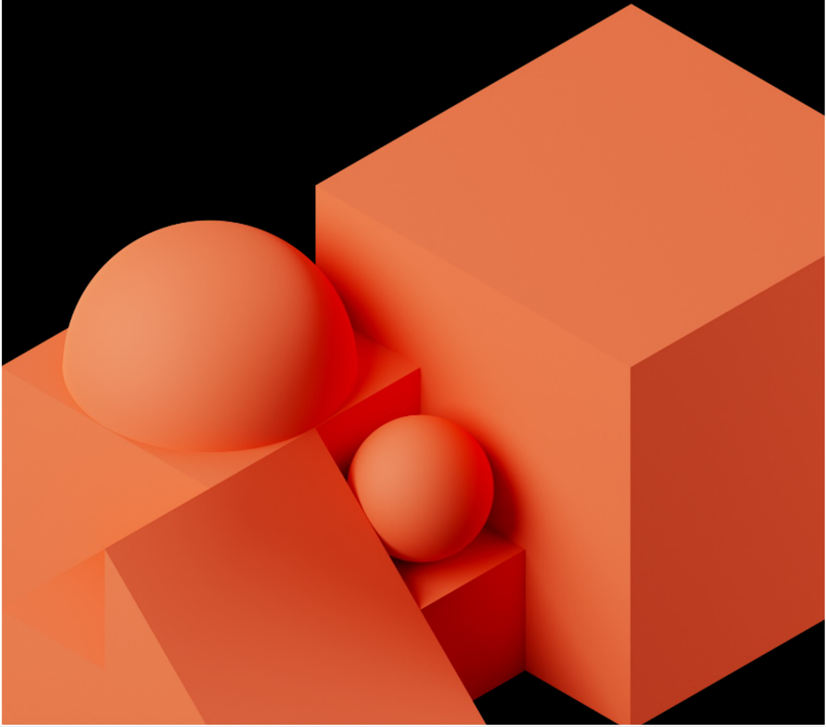
Establish Mission-Driven Data Management

DATA GOVERNANCE

Enhance Enterprise Data Governance

Analytics	Enterprise Engagement & Communications	Enterprise Data Management	Technology
Empower the Department's global workforce to utilize data by providing easy access to the Department's data assets, modern analytics tools, and customer service to enable their use.	Recruit, train, and incentivize a workforce and workplace where data is routinely sought, valued, and fluently utilized for decision-making at all levels and geographies.	Implement technology solutions to effectively create, collect, store, protect, and share data across the Department, the interagency, and with the public.	Enable oversight and coordination of Department data through effective stewardship, policies, process controls, and investment decisions that appropriately value data.

The Challenge



Requirement

M-21-31 requirements are substantial and prescriptive

OMB 21-31* Requirements

M-21-31 defines event logging requirements to support the detection, investigation, and remediation of cyber incidents on federal information systems.

1. **Retain a prescriptive list of required log types, fields, formats for systems** and other IP addressable assets
2. Significant retention duration **(Up to 2.5 years for many log types)**
3. **Central visibility for Security Operations Center (SOC)**

*M-21-31 Improving the Federal Government's Investigative and Remediation Capabilities Related to Cybersecurity Incidents

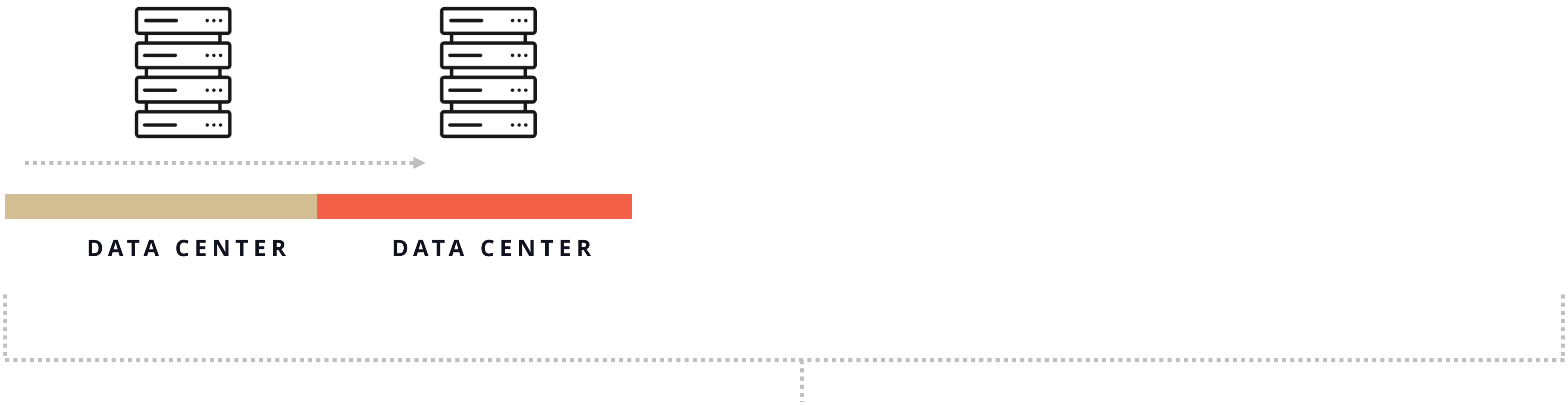
The Challenge

DATA CENTER

Systems must meet all logging requirements for breadth or duration of retention

TB/day-scale data is expensive to store, especially On Prem or in SIEM tools, which can lead to a massive duplication of stored data. Expensive to exfil between clouds

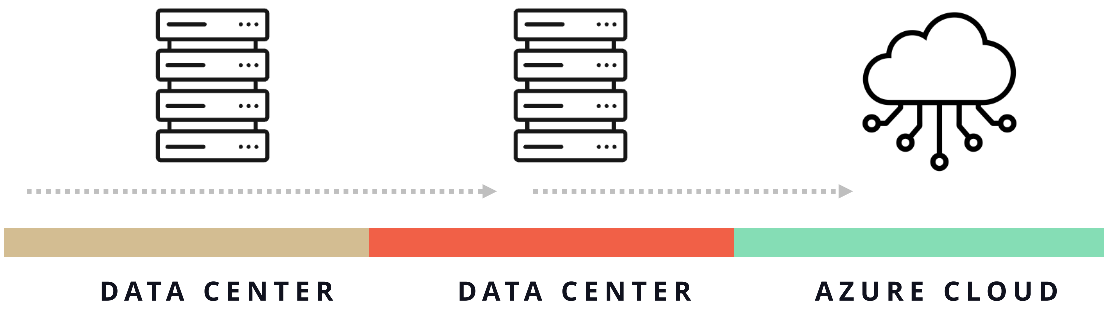
The Challenge



Systems must meet all logging requirements for breadth or duration of retention

TB/day-scale data is expensive to store, especially On Prem or in SIEM tools, which can lead to a massive duplication of stored data. Expensive to exfil between clouds

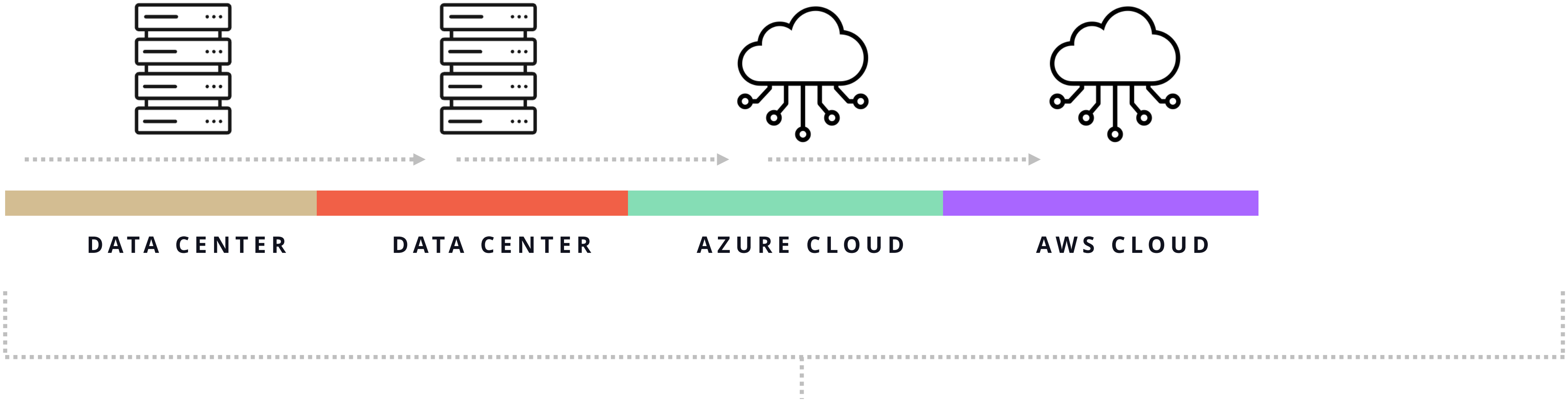
The Challenge



Systems must meet all logging requirements for breadth or duration of retention

TB/day-scale data is expensive to store, especially On Prem or in SIEM tools, which can lead to a massive duplication of stored data. Expensive to exfil between clouds

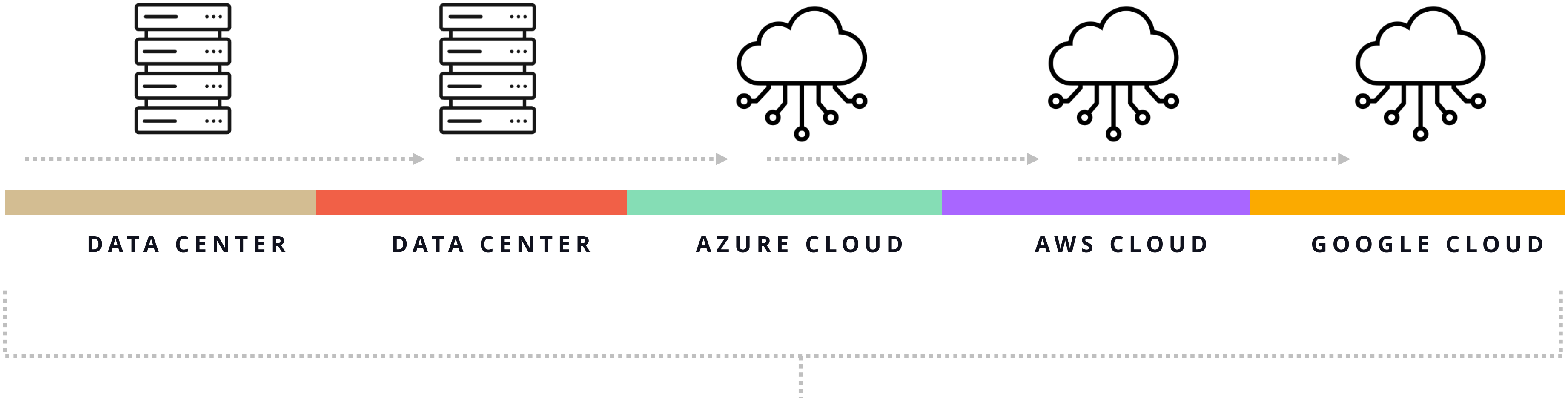
The Challenge



Systems must meet all logging requirements for breadth or duration of retention

TB/day-scale data is expensive to store, especially On Prem or in SIEM tools, which can lead to a massive duplication of stored data. Expensive to exfil between clouds

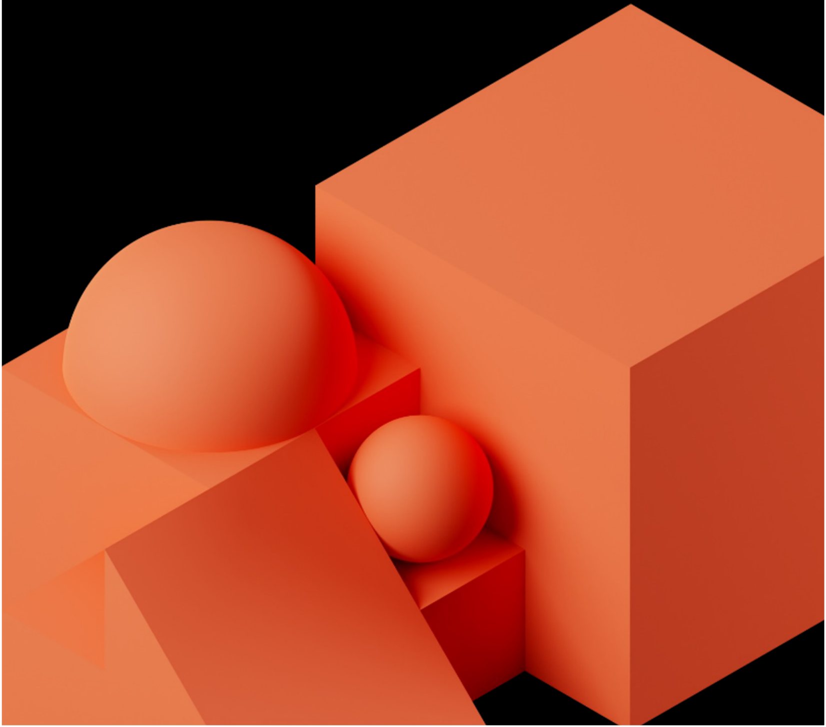
The Challenge



Systems must meet all logging requirements for breadth or duration of retention

TB/day-scale data is expensive to store, especially On Prem or in SIEM tools, which can lead to a massive duplication of stored data. Expensive to exfil between clouds

Technical



The Response: Technical

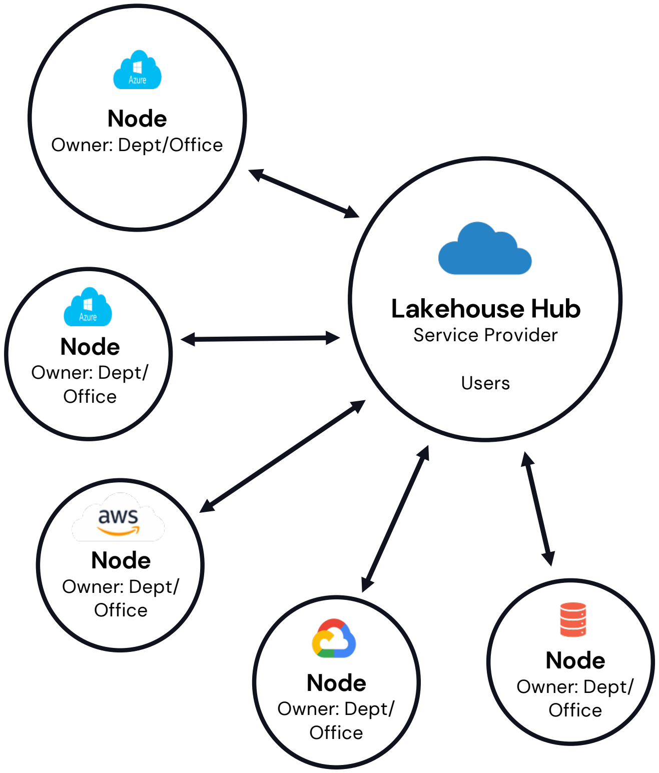
Enterprise Lakehouse effort will help system owners and the cybersecurity community respond

What is the Lakehouse?

Distributing data processing to each cloud or data "node" allows the Department to create direct connections to a centralized analytics cloud without migrating raw data across boundaries, enabling a cloud-agnostic storage solution.

Centralized Analytics Platform with Distributed Computing

Protocols for securely sharing data from individual nodes allow a centralized analytics platform to compile, search, visualize, and perform advanced analytics on distributed data from nodes, with compute handled locally.



Benefits

- Supports multi-cloud & multi-region
- Supports long-term data retention
- Supports petabyte scale
- Built in AI/ML to support batch processing and near real-time advanced analytics

- Reduce egress transfer costs
- Reduce data duplication
- Improve visibility across the enterprise
- Leverage low-cost cloud storage
- Central data governance model

Recommendations

Technology

Data Access Management

- Unity Catalog in Azure Gov't
- Implement dataset/ domain tagging
- Leverage RBAC/ABAC
- Establish multi-layer security
- Ensure maintainability

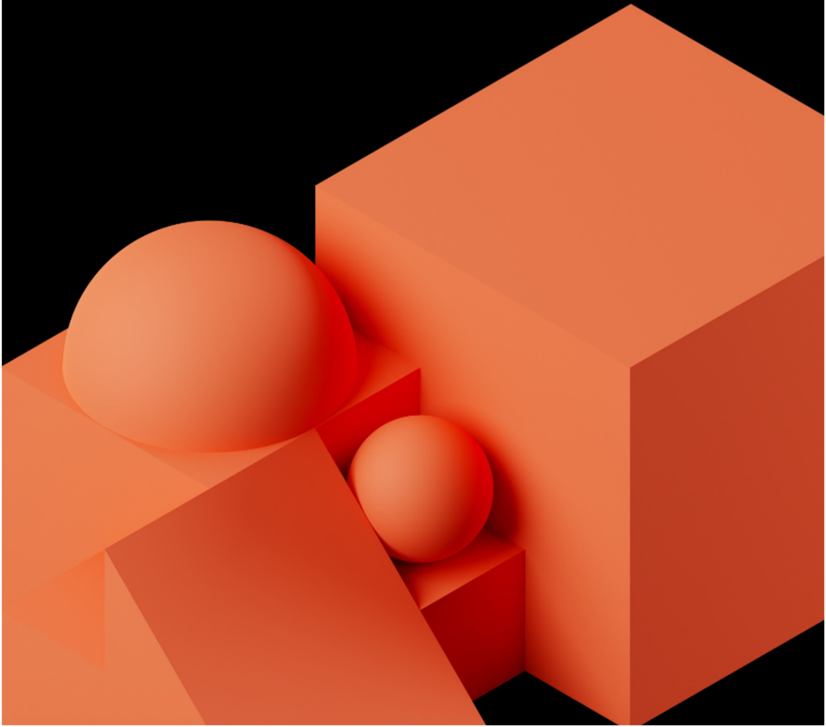
Data Processing

- Develop fungible/reusable parsers
- Exercise Databricks Autoloader
  - Ad-hoc queries
  - Scheduled queries
  - Continuous queries
- Delta optimizations
- Limit Silver and Gold tables

Federated Queries

- Develop intelligent queries
- Implement query guardrails
- Balance agency and flexibility while supporting "citizen" operators
- Integrate BI tools
- Build common pre-canned queries

Organizational

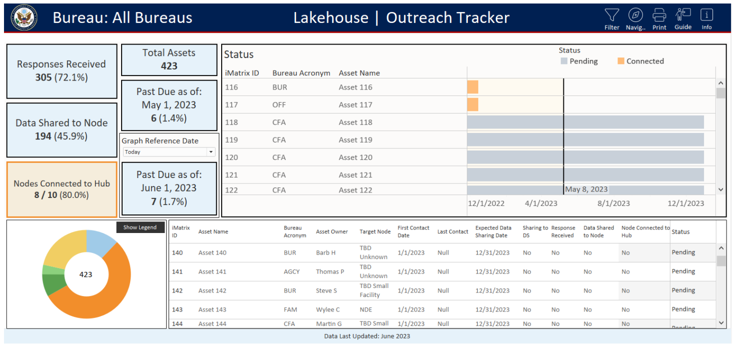


UNCLASSIFIED

The Response: Organizational

Lakehouse is supported by stakeholder governance and policy to ensure participation, implementation, and security

- Lakehouse Oversight Group
  - Reviews and authorizes new use cases, changes to access control policy, and new data connections
- Policy enablement to support participation
  - '*Thou shalt share*' CIO Action Memo; '*How To*' policy
  - Cyber Ops to reduce barriers to participating
- Tracking and reporting to IT leadership
  - Live reporting will show systems and nodes connected
  - Integrated with tracking of M-21-31 progress by system, by data element
- System owner incentives once fully operational
  - Compliant systems receive incident response common controls and support moving to Continuous Authorization



Bureau: All Bureaus

Lakehouse | Outreach Tracker

Responses Received
305 (72.1%)

Total Assets
423

Past Due as of:
May 1, 2023
6 (1.4%)

Data Shared to Node
194 (45.9%)

Graph Reference Date
Today

Nodes Connected to Hub
8 / 10 (80.0%)

Past Due as of:
June 1, 2023
7 (1.7%)

Status

iMatrix ID	Bureau Acronym	Asset Name
116	BUR	Asset 116
117	OFF	Asset 117
118	CFA	Asset 118
119	CFA	Asset 119
120	CFA	Asset 120
121	CFA	Asset 121
122	CFA	Asset 122

iMatrix ID	Asset Name	Bureau Acronym	Asset Owner	Target Node	First Contact Date	Last Contact	Expected Data Sharing Date	Sharing to DS	Response Received	Data Shared to Node	Node Connected to Hub	Status
140	Asset 140	BUR	Barb H	TBD Unknown	1/1/2023	Null	12/31/2023	No	No	No	No	Pending
141	Asset 141	AGCY	Thomas P	TBD Unknown	1/1/2023	Null	12/31/2023	No	No	No	No	Pending
142	Asset 142	BUR	Steve S	TBD Small Facility	1/1/2023	Null	12/31/2023	No	No	No	No	Pending
143	Asset 143	FAM	Wylee C	NDE	1/1/2023	Null	12/31/2023	No	No	No	No	Pending
144	Asset 144	CFA	Martin G	TBD Small	1/1/2023	Null	12/31/2023	No	No	No	No	Pending

Data Last Updated: June 2023

UNCLASSIFIED

Recommendations

Organizational

Governance

- Establish clear and delineated responsibilities
- Ensure executive support (political buy-in) to succeed
- Executive oversight group, aka the LOG
- Working group to support collective development

Policy

- Update policies as necessary to support federal requirements
- Establish guidance for system owners/data providers

Outreach

- Find willing participants
- Quick wins
- Prioritize systems/orgs based on requirements
- Sufficient and compelling use cases
- Engage your stakeholders

Thank You

This publication contains general information only, and none of the member firms of Deloitte Touche Tohmatsu Limited, its member firms, or their related entities (collective, the "Deloitte Network") is, by means of this publication, rendering professional advice or services. Before making any decision or taking any action that may affect your business, you should consult a qualified professional adviser. No entity in the Deloitte Network shall be responsible for any loss whatsoever sustained by any person who relies on this publication. As used in this document, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte USA LLP, Deloitte LLP and their respective subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.