Fortifying Data Access and Security Controls for the Public Sector

Whitepaper
August 2021

The United States federal government has some of the largest amounts of data of any entity in the world. Data is one of its greatest assets and, by the same token, one of its biggest liabilities.

Much of this data is extremely sensitive. There are endless possibilities for harm if it gets into the wrong hands. Additionally, the public sector is highly regulated with constant audits. Non-compliance results in costly penalties and potential public security issues.

The dichotomy between analyzing this data to fulfill mission critical objectives and the inherent risks of accessing it leads to conflicting objectives throughout the federal space indicative of those throughout the public sector. On the one hand, data democratization is desired to put data in the hands of as many users as possible for advanced analytics, which partly accounts for this industry's dominant theme of migrating to the cloud to improve data access.

On the other hand, the enormous amounts of data and numbers of regulations involved are constantly increasing the need for data security and access controls. This concern is magnified across modern hybrid- and multi-cloud architecture in which consistently applying policies across settings and tools is a challenge.

Balancing these opposing interests so that they're both met—data is readily accessible via the cloud and equally secure while doing so—requires the combination of centralized data access governance and decentralized enforcement in which policies are created once, yet consistently implemented and enforced in any source or setting.

This approach hinges on a lightweight architectural footprint that won't compromise the performance of storage and compute systems. It leverages robust security plug-ins for enhanced authorization measures. It's also fully automated, cloud ready, and substantially more cost-effective than traditional perimeter defenses are.

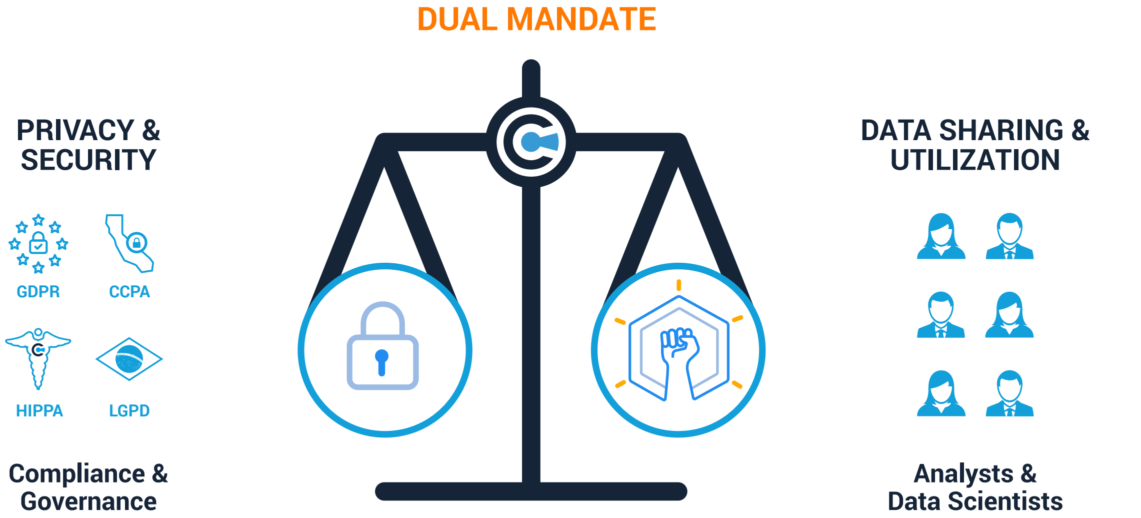
It's the key to resolving the government's opposing interests, delivering data democratization and secure access control.

Contradictory Mandates

The federal government's conflicting data mandates exist throughout the public sector in general. One demand is to democratize data access for a broadening user base seeking insights by exploiting Artificial Intelligence and machine learning. AI and ML are economical and accessible tools for improving mission effectiveness for government agencies—especially when aspects of these approaches involve the cloud. Public sector agencies are universally moving to the cloud to meet this demand and maximize the value of their data.

But data security and access control objectives of the other mandate for protecting those assets directly contradict the goals of data democratization. Controlling data access and meeting regulatory compliance standards are the top priorities for this mandate. Those responsible for it typically do so with overly cautious measures to secure, control, and authenticate access—which is time consuming and limits data's use, effectiveness, and accessibility.

CLOUD MIGRATION WITHOUT COMPROMISING SECURITY



In the federal space, data democratization is realized through a centralized data platform that's a single place where all users access data and accompanying resources (like compute) for any application. It's designed to eliminate silos and make it easy for analytics users to pull data securely. Almost every federal agency has one; for example, in the Department of Defense (DoD), it's called Advana[1].

However, many data security and control requirements contradict the ends of the central platform concept because of the sensitive nature of the data. The Centers of Medicare and Medicaid Services, for example, must preserve personally identifiable information (PII) and

[1] Meet Advana: How the Department of Defense Solved its Data Interoperability Challenges; https://governmenttechnologyinsider.com/meet-advana-how-the-department-of-defense-solved-its-data-interoperability-challenges/, April 2021

personal health information (PHI). The DoD has highly sensitive data. Thus, no matter how much federal entities want to migrate to the cloud for a unified platform, security and compliance mandates are a clear conflict of interest in doing so.

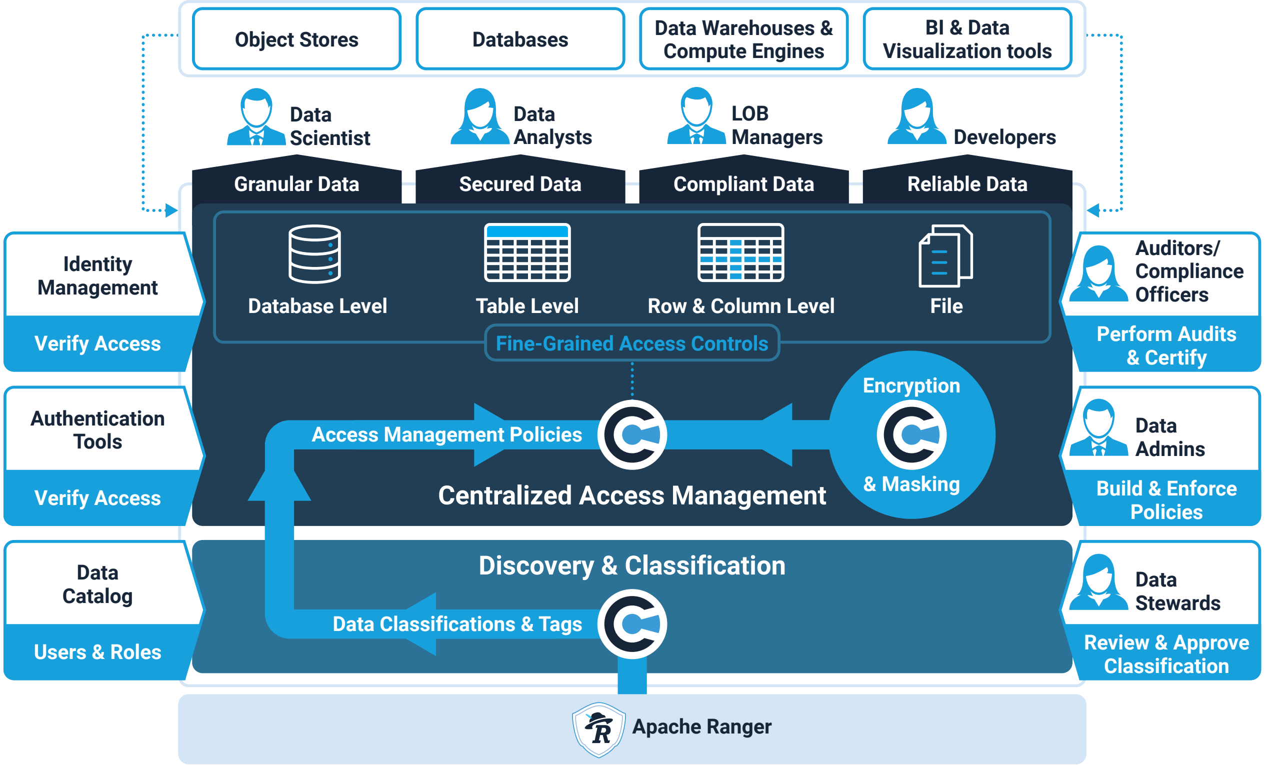
Federal Data Security and Control Challenges

Regulatory compliance, data security, and data access are so important to federal deployments because of their immense data amounts, numbers of sources, and repositories they must protect. Federal organizations are the original collectors of big data since their missions began. Consider, at the very least, the voluminous quantities of their personnel, client, and constituent data—like pension plans and life insurance for veterans. Say someone enlists in the Department of Defense (DoD) at 18, stays with the department for 30 years and has a family consistently treated in DoD health agencies. At some point Veterans Affairs takes over and stores that person's medical data for life. With the average life expectancy spanning 100 years and the likelihood that family members may also enlist in the DoD, the VA's data amounts are considerable. Many think the greatest need for data is feeding AI and ML models (courtesy of data democratization), but the bigger worry is improper data access without sufficient security controls.

Feds must also contend with both malicious and non-malicious data security issues—meaning malefactors and human mistakes. The latter requires constant vigilance because the combination of sheer numbers of repositories, the continuously expanding volume of data, and the number of users trying to access that data is hard to safeguard even without nefarious intentions. All that sensitive data makes federal agencies huge targets for malefactors looking to make this data public or sell it. For example, a U.S. Customs and Border Protection vendor was breached in 2019, compromising traveler photographs and license plates of approximately 100,000 people. Moreover, traditional data security approaches are based on insufficient perimeter defenses that aren't designed for hybrid- and multi-cloud deployments—as all the Virtual Private Network (VPN) data breaches demonstrate. A far better approach is to complement perimeter security with securing IT assets at the repository, database, table or even down to row and column levels.

Resolving the Conflict

Privacera's centralized access governance and data privacy framework uses the approach of securing data assets at the data level to resolve the government's conflicting data interests in several ways. The first is by a flexible architecture designed for cloud services that resonate with this space because many federal entities are coming from Cloudera, are familiar with Apache Ranger, and have existing Ranger policies. The lightweight nature of this architecture is perfect to layer into federal compute platforms because it's mostly plug-in based. These plug-ins swiftly authenticate users to support the performance of thousands of users simultaneously accessing and querying data while these unified platforms continue to run natively.



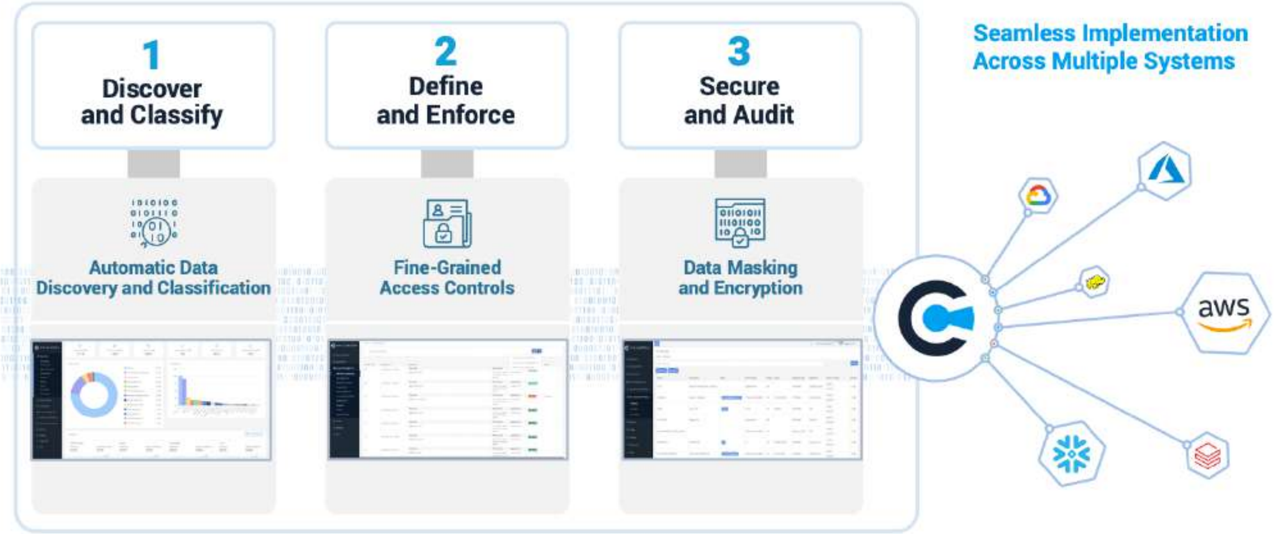
Privacera scales to manage hundreds of thousands of users and more than 25 petabytes of data; it's also highly extensible and built on open standards. Moreover, it fortifies authentication, access controls, and data governance to enhance security within individual sources, operating at the data (instead of the perimeter) level. This is key for satisfying the dual mandate of data democratization with a light architectural footprint in sources that supports performance needs while delivering the security for dependable access controls and regulatory compliance.

Cost Effective Compliance

With other approaches, scaling to meet governmental challenges of expansive repositories and data quantities is cost prohibitive. Such scalability should be affordable enough to deploy across all systems while sustaining performance requirements. Traditional approaches involving customized cross-domain security solutions aren't scalable because they cost upwards of $10 million; many also result in silos. Privacera is a more cost-effective solution via its automation, which is founded on building policies once in a centralized solution and automatically applying them into any source. For example, platforms like Databricks, S3, and

Snowflake have unique ways of administering security. With Privacera, users can efficiently enforce the same policy into each of them according to their mechanism, versus building security for each one.

Additionally, regulatory requirements in the federal space are unparalleled. The surplus of data systems and regulations means audits—requiring reports and justification of actions—are always occurring. Privacera significantly reduces the time and cost of audits by simplifying compliance measures. It creates log files every time anyone attempts to access data across its central platform—whether they were given access to requested data or not. These auto-generated reports document the applicable policies for approval or denial of access, giving administrators, for instance, real-time visibility into their clusters. Case in point is pre-built reports of top 20 users attempting data access and customizable reports to scrutinize this information so auditors can review it at whatever granularity required to determine if credible governance and access policies were used.


1
Discover and Classify
Automatic Data Discovery and Classification
2
Define and Enforce
Fine-Grained Access Controls
3
Secure and Audit
Data Masking and Encryption
Seamless Implementation Across Multiple Systems
aws

CASE STUDY:

Executive Department for National Security Supports Large Analytics Community with Secure Data Democratization

An executive department for national security in the U.S. has a centralized public cloud platform for all authorized users to pull and query data from for advanced analytics. However, the repository still needed a scalable, performant, cost-effective authorization system to support operations. Protecting its hundreds of thousands of users (not including external agencies) would've been nightmarish with an individual security system approach.

Privacera handily resolved these issues for near-impenetrable security. Its support for Apache Ranger allowed the U.S. executive department to extend existing Ranger policies to the cloud and their analytics suite. Privacera's architecture was lightweight enough to present minimal technical overhead. It enabled the agency to open their centralized cloud platform up to 100,000s of users to provide self-service analytics for several machine learning, AI and analytics projects. Most of all, Privacera's plug-in methodology enforced security policies across all applications on the cloud platform.

The Best Choice

The public sector's data security issues are characterized by massive quantities and sources of data that either are currently or soon will be accessible via the cloud. Privacera facilitates controlled data access by protecting these resources with a lightweight architecture that doesn't compromise performance and robust plug-in based security that protects data within sources as needed. It's the most cost effective, sustainable solution for meeting regulations and national security issues.

About Privacera

At the intersection of governance, privacy, and security, Privacera's unified data access governance platform maximizes the value of data by providing secure data access control and governance across hybrid- and multi-cloud environments. The hybrid platform centralizes access and natively enforces policies across multiple cloud services—AWS, Azure, Google Cloud, Databricks, Snowflake, Starburst and more—to democratize trusted data enterprise-wide without compromising compliance with regulations such as GDPR, CCPA, LGPD, or HIPAA. Trusted by Fortune 500 customers across finance, insurance, retail, healthcare, media, public and the federal sector, Privacera is the industry's leading data access governance platform that delivers unmatched scalability, elasticity, and performance.

Headquartered in Fremont, California, Privacera was founded in 2016 to manage cloud data privacy and security by the creators of Apache Ranger™ and Apache Atlas™.

Visit www.privacera.com or follow @Privacera on LinkedIn and Twitter.