privacera

DATA DEMOCRATIZATION:

# 5 Questions to Ask Your Data Access Platform Vendor

WHITEPAPER

privacera

# contents

**Your company is on a mission: to transform itself and improve profitability and competitive positioning in the industry.** Migrating enterprise data and analytic workloads to the cloud is a critical component of this mission; one that not only lifts the burden of operating and maintaining the data infrastructure from your IT team, but also enables your company to respond faster to changes in market dynamics.

But with the advantages of cloud migration comes a difficult dual mandate you must balance: providing data scientists and analysts access to increasingly diverse data to derive valuable business insights, while ensuring proper access controls are in place.

To help navigate these challenges and help your company make cloud migration as seamless as possible, you must implement a centralized platform for scalable data access management. Understanding your company's unique requirements and needs, this paper will provide five questions to help you evaluate potential data access control platforms and assess the best solution for your enterprise.

1

## Is Your Current Platform Built to Govern Users' Access to Data?

This is a critical question to ask your data security platform vendor, as a number of tools currently marketed as data access control solutions were originally built for another purpose. For example, some data access control solutions were originally built for another purpose – for example, data virtualization platforms originally developed to provide data analysts and data scientists access to data from a number of sources. Whether a solution was built specifically to define and administer data access control policies or not has several important implications.

First, if a virtualization platform is repositioned to fulfill a market need, or a product's additional features are bolted on, it is very difficult to match the rich functionality and capabilities such as fine grained access control, column masking or row filtering available in a platform built specifically for the purpose of defining and administering access control policies.

Second, data virtualization-based solutions connect to different data sources to access data from a common logical data access point. Virtualization products act as an additional proxy middle tier between analytics services and data storage and perform extensive data processing. When used as a data access control solution, all users' requests to access data are routed through the virtualized data access point, which negatively impacts the system's performance, as all queries are required to go through the virtualized data layer. In addition, metadata must be recreated in the new solution, creating substantial overhead for IT teams who must then rewrite client applications in the new virtualization platform.

Overhead is a significant challenge for cloud environments, as the volume of data rises rapidly and user queries against data cannot be anticipated. To effectively manage access to data in the cloud, companies need a platform with a native cloud integration that does not disrupt data being accessed.

### BEST PRACTICE

Perform due diligence with your IT team to determine the origin of the platform you are considering to manage user access to data.

Thoroughly investigate the platform's performance during its evaluation by simulating data volume and user request scenarios.

2

## Is Your Solution Based on Open Source Technology?

If your company is investing IT resources in open source technologies, it is important to consider if the data access control platform vendor is aligned with open source philosophy. Solutions based on open source technology offer many benefits, including:

- Eliminating vendor lock-in to source the solution exclusively from one software company
- Leveraging resources and expertise of the global developer community that is difficult for even the largest commercial software company to match
- Access to continuous innovation, driven by requirements from leading companies
- Attracting and retaining software talent with contributions to open source projects

The effectiveness of open source technology in safeguarding against unauthorized access to enterprise data has been well documented. In 2014, Apache Ranger started as an incubator open source project to authorize, audit, and encrypt data in Hadoop-based data lakes. Since then, it has been widely adopted as an extensible, robust, data access control framework used by hundreds of enterprises globally.

Open source solutions like Ranger are more extensible than proprietary counterparts, making integration with third-party products – such as data catalogs, reporting, and custom applications – much easier.

### BEST PRACTICE

If your company is in the midst of migrating data from an on-premises data lake to the cloud, or you are interested in building an integrated ecosystem, consider a centralized platform built on open source technology to manage data access across various services in the cloud to enable secure data sharing among data consumers while complying with privacy regulations.

3

## Does Your Solution Require Customers to Rewrite Data Access Policies When Migrating to the Cloud?

If your company has invested significant time, effort, and resources in building its unique access control policies, it is unreasonable to abandon the data access policies built for your on-premises data repositories and recreate them in the cloud; however, some data access platform vendors may lead you to believe this is required. As these access policies are crucial to your company's business operations, tools that require your company to recreate data policies in the cloud not only forces you to abandon intellectual property, but also stifles how fast you can onboard users. Without robust data access policies in place, your company cannot give users access to data, which could result in missing critical deadlines. Because a majority of companies are implementing hybrid architectures, it is crucial for on-premises and cloud-based data access policies to be consistent and in sync.

Using a data access platform based on virtualization technology for your cloud environment will also force your data analysts and scientists to rewrite all their queries from scratch to point to the virtualized layer. This is an unnecessary burden on the data consumers that significantly delays when they can start finding insights from the data.

### BEST PRACTICE

Understand how your company can leverage access policies built on-premises for cloud deployment. If the data access platform you are evaluating requires you to re-register existing tables, as well as any new data source you may add in the future, consider it a red flag. Investigate how the data access control platform ensures secure access to data, so data analysts and data scientists do not have to rewrite policies or queries for use in the cloud.

4

## Is Your Solution Highly Scalable and Extensible, and Can You Demonstrate it?

Your company likely has petabytes of data in its cloud environment. If you are streaming data from sensors or IoT devices into the cloud, data volume will only continue to grow exponentially. According to the International Data Corporation (IDC), by 2025, 49% of the world's data will reside in public cloud environments; therefore, it is imperative your data access management platform accommodates, not only current data volume, but also future data growth, to avoid impeding system performance.

When migrating data to the cloud, the majority of data is stored in object stores like S3 and ADLS. Applications read this data using languages such as SQL, Python, Java, and others. When implementing access controls, ensure that the platform comprehensively secures all your data with a scalable architecture that supports all your current, as well as future data needs.

### BEST PRACTICE

Ask about the scalability of the data access management platform your company evaluates; understanding this will help uncover architectural details and determine if the architecture is extensible enough to support significant data growth. Also identify if the solution is deployed in production environments with large amounts of data (petabytes or more).

Inquire whether the proposed data access solution is limited to only securing tabular data and what would be the mechanism to provide coverage for data sources that might be added to the data infrastructure in the future.

5

## Does Your Solution Centrally Manage Access Across Your 5 Cloud Services?

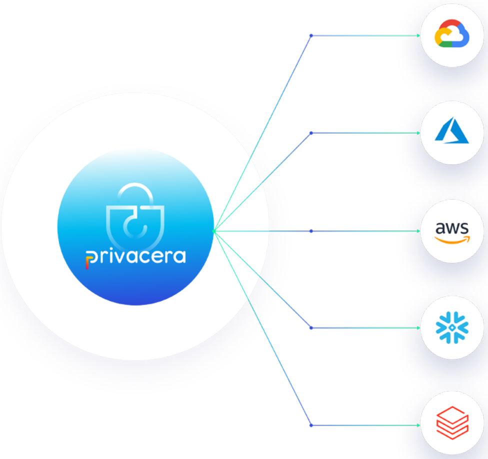
Cloud represents a new world, even for the largest and most sophisticated companies in the world. This world exists beyond a company's firewall and security perimeter. In this world, your cloud service provider operates the servers in which your data is stored, and the allocation of IT resources is dynamically managed. Cloud is also a complex world for data infrastructure and platform teams to navigate. Each of the public cloud providers offer a unique mechanism for administering data access.

To make matters more complex, cloud-native services operating on public cloud infrastructure also have unique data sharing capabilities. For example, if your company uses AWS for its public cloud provider, S3 as its transactional data storage layer, Redshift or Snowflake as its data warehouse, and Databricks as its analytics platform, the result is four or more unique mechanisms defining and administering data access policies across your cloud services. This means that an administrator needs to familiarize herself with various mechanisms for administering access control policies across the cloud environment and navigate to multiple interfaces to do that. This not only adversely impacts the productivity of platform administrators to onboard new users but also raises the probability of making costly security errors.

Due to the complex nature of public cloud infrastructure, it is critical your data access platform provides a centralized interface – with pre-built connectors to leading cloud services – that can be used by your platform administrators to implement consistent access control policies from a single location.

### BEST PRACTICE

Ask for a comprehensive list of cloud services for which the vendor's platform offers pre-built integration or connectors. This will provide a faster onramp and reduce the time to onboard users to those services. More importantly, investigate how long it will take the vendor to build a connector for a new service that your company may subscribe to in the future.



Fortune 500 enterprises trust Privacera for their universal data security, access control, and governance. Discover how to streamline data security governance with Privacera.

## Take a unified approach to data access, privacy, and security with Privacera.

REQUEST A DEMO →    CONTACT US →

Privacera, based in Fremont, CA, was founded in 2016 by the creators of Apache Ranger™ and Apache Atlas. Delivering trusted and timely access to data consumers, Privacera provides data privacy, security, and governance through its SaaS-based unified data security platform. Privacera's latest innovation, Privacera AI Governance (PAIG), is the industry's first AI data security governance solution. Privacera serves Fortune 500 clients across finance, insurance, life sciences, retail, media, consumer, and government entities. The company achieved AWS Data and Analytics Competency Status, and partners with and supports leading data sources, including AWS, Snowflake, Databricks, Azure and Google. Privacera is recognized as a leader in the 2024 GigaOm Radar for Data Access Governance. Learn more at Privacera.com.

privacera