Privacera on AWS

Unified Data Access Governance Platform with Native AWS Integration

Whitepaper
December 2021

Content

1. Privacera Architecture Overview 3
   - How Privacera Works 3
   - Privacera and AWS 4

2. Key Benefit for Data Science and Analytics Teams 6
   - Democratize Data Access for Self-service Analytics 8

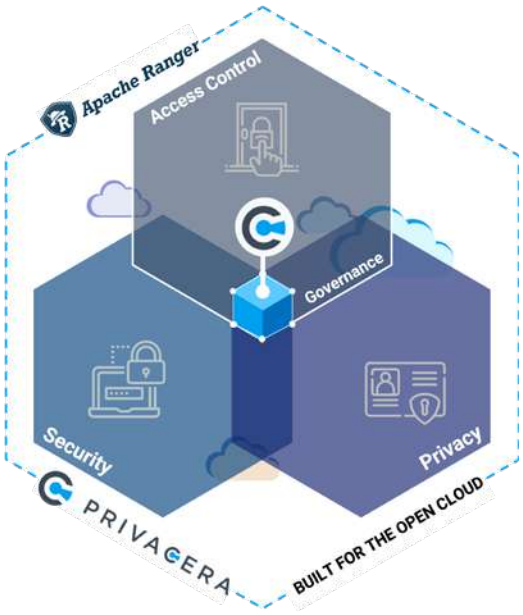
3. Key Benefit for Compliance & Privacy Teams 8
   - Unified Data Access Management 8

4. Key Benefits for Cloud and Data Management Teams 9
   - Cloud-native, Multi-vendor Support 9
   - Consolidated, Simplified Audit 9

5. A New Paradigm for Secure Data Democratization: Governed Data Sharing 11

6. Conclusion 13

1 Privacera Architecture Overview



How Privacera Works

Making data widely available while fully complying with regulations can be difficult, time-consuming, and inefficient. This challenge is magnified when dealing with the multiple facets of data governance across hybrid- and multi-cloud settings in an expanding data landscape.

Privacera supports the entire lifecycle of data access governance with a comprehensive unified platform that provides enterprises sensitive data discovery, fine-grained access control, distributed native policy enforcement, and dynamic data masking and encryption, all delivered through a single pane of glass.

Specifically, Privacera brings:

- Broad and deep support for existing technology
  - Native integrations that do not require moving or changing existing assets
  - Close partnerships with leading vendors in key areas:
    - Cloud Providers: AWS, Microsoft Azure, Google Cloud
    - Query Federation: Dremio, Starburst
    - Data Platforms: Databricks, Snowflake
- Proven scalability
  - Roots in the largest platforms in the Big Data world
  - Existing customers managing 100Ks users, 25+ petabytes and highly classified data
  - Head-to-head benchmarking shows 25x better performance and 75% lower TCO than other solutions
- Baked-in flexibility
  - Full support for the Apache Ranger ecosystem
  - Enterprises and partners can build native connectors leveraging the open framework
  - Extensible to any SQL sources via query federation
- Delegation and self-service
  - Access management delegation
  - Audit and data stewardship delegation
  - Self-service access request workflows

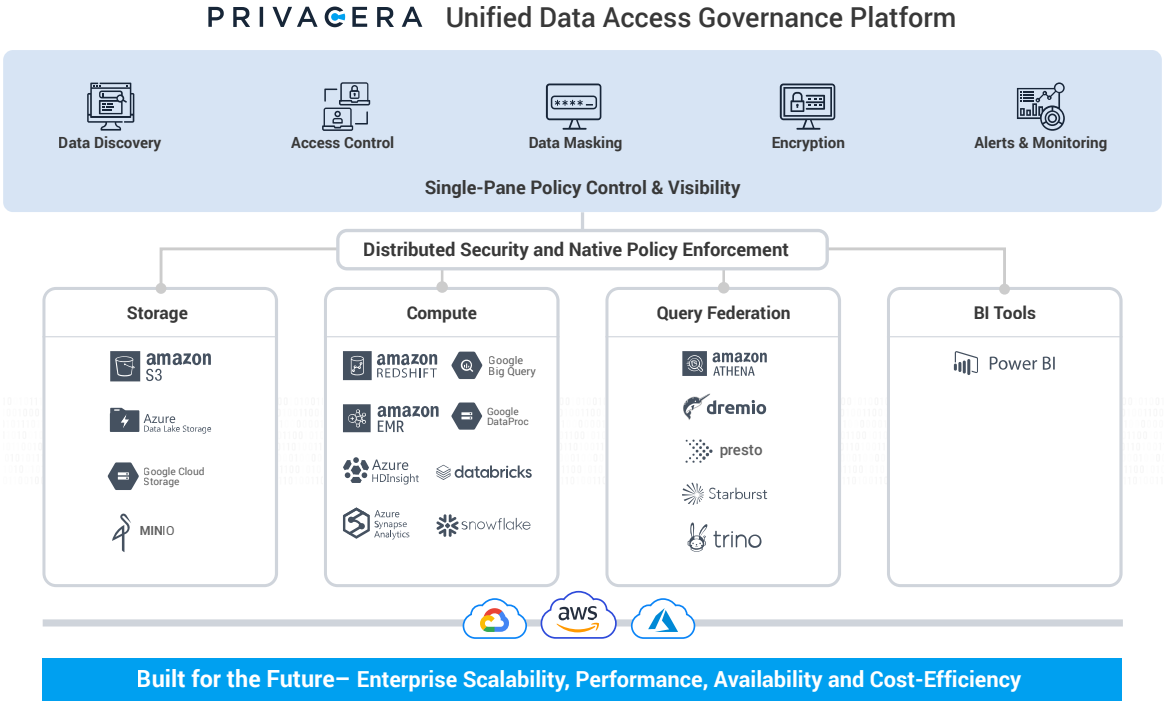


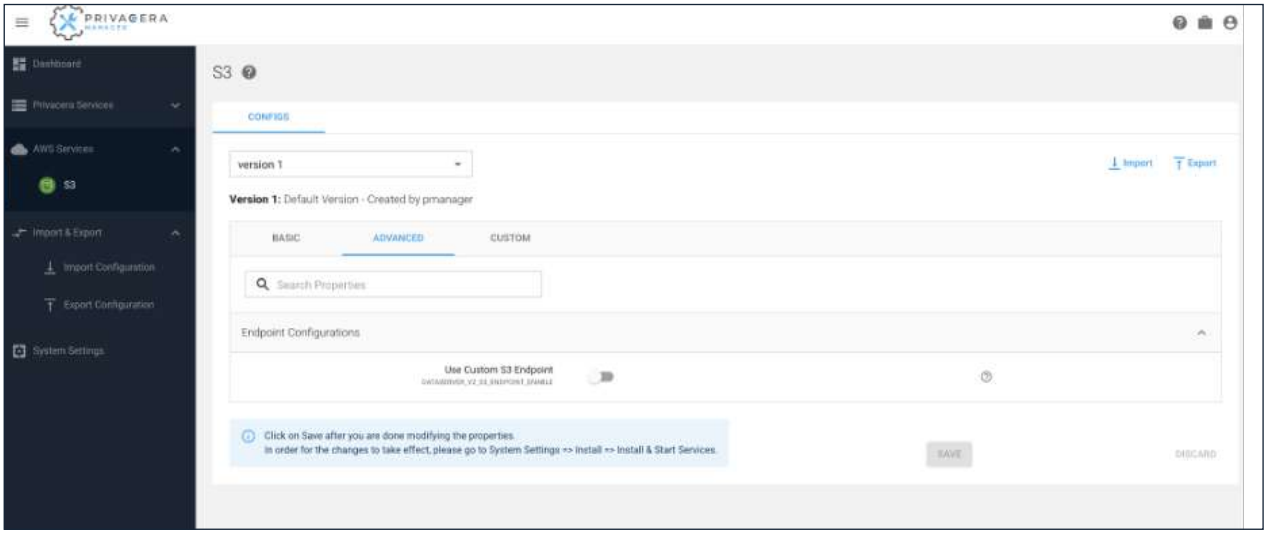
Exhibit: Privacera platform architecture

Privacera and AWS

Privacera is the first and only enterprise-ready cloud data access governance platform, providing universal, consistent data access control, governed data sharing, and compliance tools across today's modern data management landscape. In addition, it offers broad and deep support for existing technology through native integrations that don't require moving or changing existing assets, so it works with the tools and technology enterprises already have in place, like AWS's extensive ecosystem.

For example, Privacera Sensitive Data Discovery can scale linearly to identify and classify sensitive elements in data and works with data sources ranging from legacy databases to Amazon RDS, Amazon EMR, Amazon Athena, Amazon DynamoDB, and beyond. Sensitive Data Discovery is used by customers to scan huge volumes of data across diverse data sources. Discovery can scan tens of petabytes in a single scan and can incrementally scan individual files in near real-time as they land in an Amazon S3 data lake. Once classified, Privacera can report on data usage; take action to redact, quarantine, or anonymize data; or control access to critical data resources across AWS and beyond.

The Privacera platform is installed and configured via Privacera Manager's browser-based interface onto Docker or within AWS's highly scalable managed Elastic Kubernetes Service (EKS) infrastructure.



Privacera can be deployed directly into the enterprise's virtual private cloud (VPC) environments or accessed easily as a SaaS offering running natively in AWS. Privacera supports high availability (HA) and scalability in deployments and provides automation, including Cloud Formation templates (CFTs) for the standardized deployment of Privacera-secured EMR, native enforcement of access controls in Amazon Redshift, Amazon RDS Postgres, MySQL, and Oracle. Privacera also supports REST APIs and command-line tools to simplify and automate deployment and maintain native access—even for programmatic access to Amazon S3, Amazon Athena, Amazon Kinesis, and other services.

Amazon S3

Privacera-protected S3 requires no changes to query paths in existing code. That's lower risk than any other vendor can offer. In addition to basic path and object-based access control, S3 access can be managed by group, role, user, or attribute based on existing identity management tools, including Okta, Azure AD, Active Directory, and SAML, OAuth and LDAP providers. S3 access can also be controlled based on data classification tags applied by Sensitive Data Discovery or provided by a data catalog. The same access policies can be used — and audited — whether the S3 data is accessed from EMR, Hive, Boto3, or Trino (including Starburst). And, Privacera's encryption can encrypt or statically mask data directly *in situ*, allowing control of access to sensitive data at a column level.

Amazon EMR

Privacera-protected EMR can leverage AWS's tight native integration with Apache Ranger that was developed in consultation with Privacera, or use Privacera's extended plugin support for Spark, Hive and PrestoSQL– even when accessed through third-party tools. In addition, PrivaceraCloud provides an exclusive enterprise-ready,

fully-managed Apache Raner-based SaaS for both native EMR/Ranger integration and Privacera's plugins. As a result, there are no custom catalogs to maintain, no manual tagging, and no need to create separate policies for resources defined in Amazon Hive Catalog, AWS Glue, or just raw in Amazon S3.

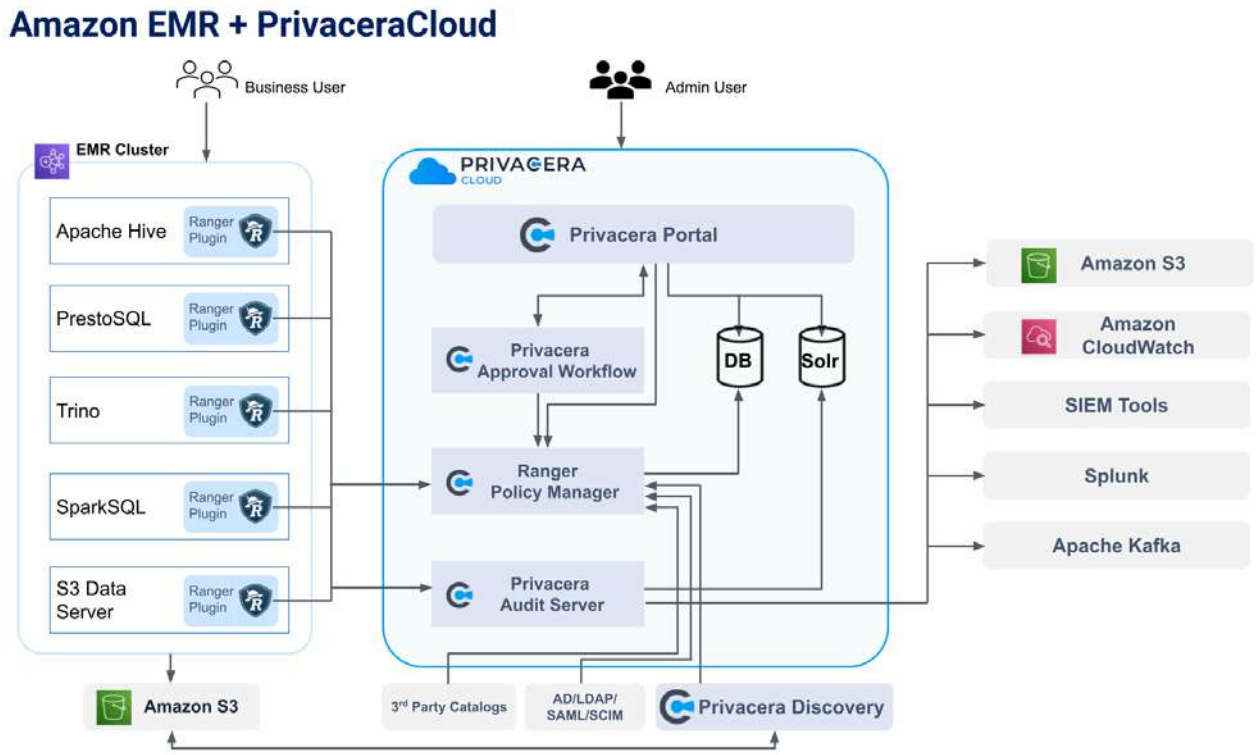
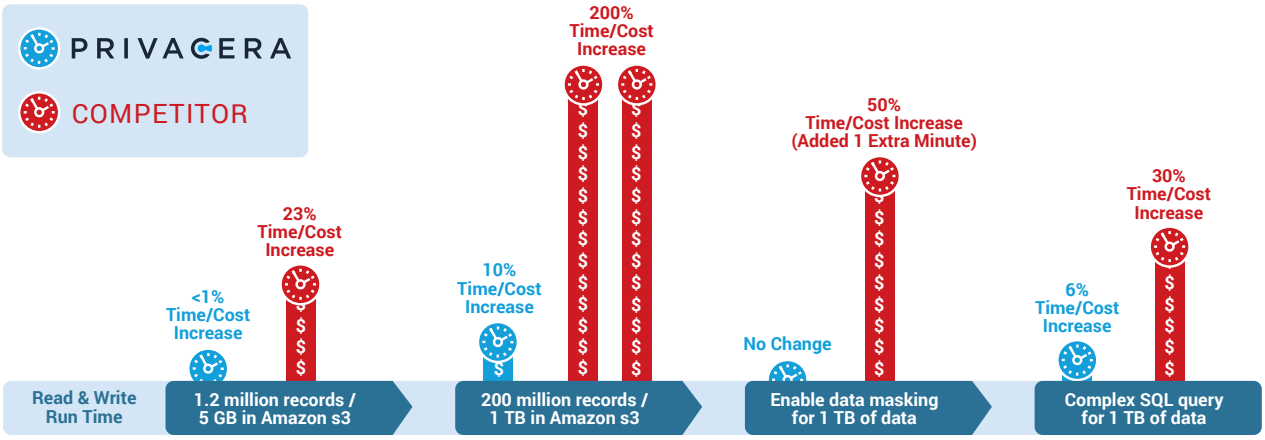


Exhibit: Amazon EMR and PrivaceraCloud architecture diagram

Amazon RDS for PostgreSQL, MySQL, and Oracle

Privacera's PolicySync engine leverages a native plugin for simple integration, extending the same policies for native access control directly in relational databases like Amazon Relational Database Service (RDS). This means that the same tag-based policies can provide access to data based on the same user attributes, roles, or groups available within Privacera. Direct access to the databases via the AWS console or JDBC is preserved, giving Privacera's access control a significant performance advantage at just **20%** of the cost and compute required by other tools, with no code changes needed for existing dashboards, reports, or analysis tools



> During our POC process, we conducted a performance benchmarking of the leading vendors. I was stunned by the results.

Sr. Manager
Data Engineering & Analytics

Fortune 100
Biotech Company

Exhibit: Performance comparison shared by Privacera customer

Other AWS Services

Privacera also protects Amazon Athena, Amazon Kinesis, and other services by enabling column-level access control and tag-based enforcement beyond what the native services provide. Privacera's Data Access Server can protect these and other services even when accessed via the AWS CLI command line or REST APIs. And all audits are tracked and aggregated within Privacera.

Customers have the flexibility to use QuickSight and other third-party tools for reporting and analysis, leveraging the same underlying controls provided by Privacera.

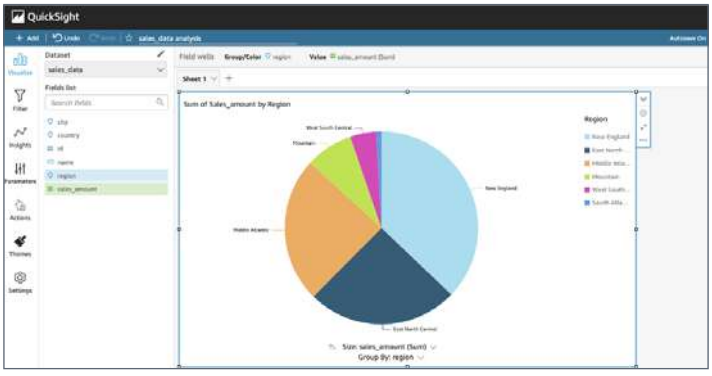


Exhibit: QuickSight interface

2 Key Benefit for Data Science and Analytics Teams

Democratize Data Access for Self-service Analytics

To put data to work and extract value from it through analytics, machine learning, or predictive modeling, having comprehensive data governance, access control, and security is essential to ensure that data is protected, compliant, and fit for consumption. Rahul Pathak, VP of AWS Analytics, said it best in his leadership session at AWS re:Invent 2021 that *"customers sometimes think that access control and governance is in conflict with velocity, it's NOT true. If you actually have good guardrails and protection on your data, you're able to set it free and allow people to experiment and iterate because you're confident that your data is well protected and only the right people have access to it."*

Data governance is a critical component that underpins the entire data analytics continuum in which data is consumed in purpose-built data stores, data lakes, and analytics services that are designed for machine learning and data science modeling. This resonates with Privacera's mission to help customers maximize the visibility and usability of data by enabling unified data governance, fine-grained access control, and enhanced data security through a single console across multiple cloud services. The comprehensive set of data security and governance capabilities that Privacera delivers sets data scientists and analysts free to innovate with trusted and faster access to data without compromising data privacy and compliance.

3 Key Benefit for Compliance & Privacy Teams

Unified Data Access Management

As many organizations migrate to multi-cloud and hybrid-cloud infrastructures, effective access control, governance, and security of an organization's sensitive data has never been more important. In these complex environments, data teams can become overburdened with managing unique access controls and policies across various cloud services and applications. According to Gartner, *"…to ensure operational control, enterprises want to unify administration and monitoring of their IT systems. They want to standardize policies, procedures, and processes and share some tools—especially those that enable cost governance and optimization—across multiple cloud providers."*

With tight integrations with major cloud providers, analytic services, popular IAM tools, and a wide variety of user authentication standards such as LDAP, SAML, OAuth, and Open ID, Privacera enables consistency in data compliance and streamlines processes to easily enable fine-grained access control from a single pane of glass. In addition, Privacera provides automated sensitive data discovery, tagging, and classification to help find where sensitive information resides and eliminate data and regulatory blind spots.

4 Key Benefits for Cloud and Data Management Teams

Cloud-native, Multi-vendor Support

With fine-grained access control, each data resource can be protected based on attributes, user roles or groups, or classification tags, enabling data with different access requirements to coexist in the same storage or analytics platform. With native support for EMR, Hive, S3, RDS, Redshift, and others, Privacera allows data administrators to create role-, attribute-, and tag-based policies to control data access at the file-, row-, and column-level¬ and implement dynamic data masking and filtering. These functions enable data scientists and analysts to query data rapidly while protecting against authorized access to sensitive data such as personally identifiable information. With the safeguards fine-grained access controls provide, data teams can:

- Alleviate the time-consuming process of manually requesting access from each separate data owner
- Empower data consumers with self-service access requests to support their analytics use cases
- Deliver high-quality insights and analytics faster with the ability to access any portion of data, no matter where it resides

Consolidated, Simplified Audit

Privacera simplifies regulatory reporting and auditing for compliance or forensics. While access logs for disparate data platforms can be challenging to aggregate, query, or summarize, Privacera provides a simple user interface that brings all these audits together in a common format, accessible through a single user interface.

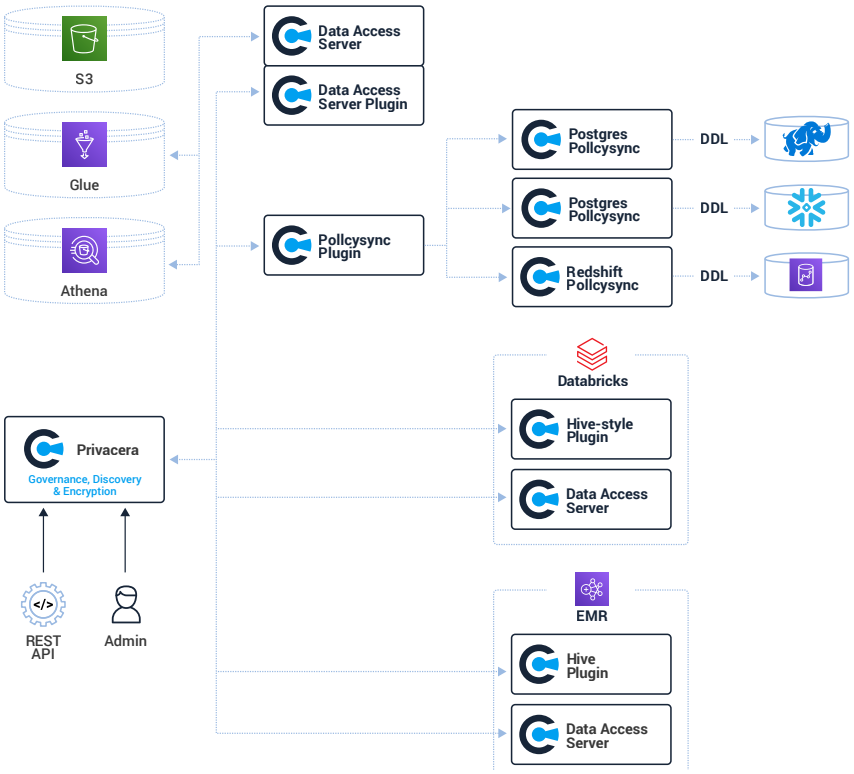


Exhibit: Privacera integrations

Privacera's centralized audit features consolidate detailed records for user data access attempts, authorization results, administrative actions, data access requests, and the complete history of all policies and when they were in effect on every Privacera-protected platform. Privacera's services provide a common set of APIs for extracting data and reports to simplify reporting requirements for both internal Compliance and Legal teams, such as for monitoring compliance with GDPR-related policies under Article 39.

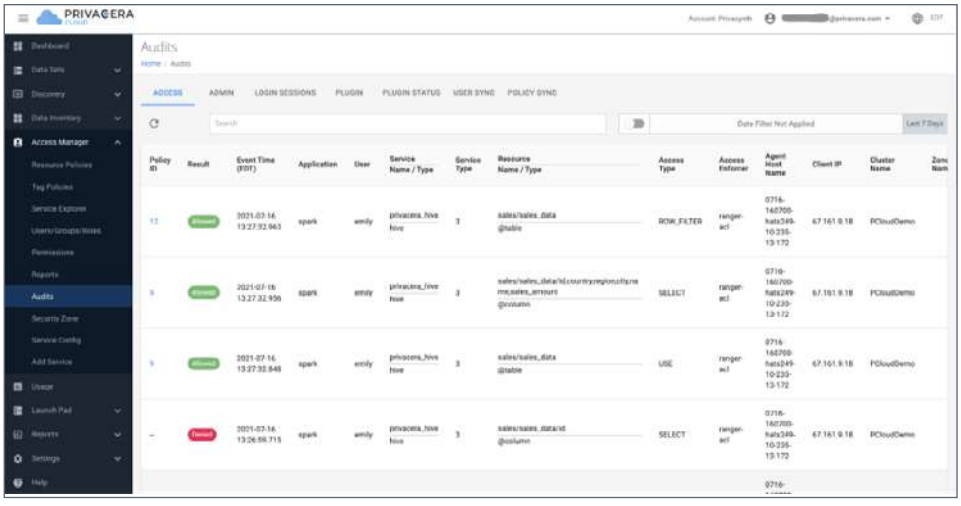


Exhibit: Audits of user access to different data sources

5 A New Approach for Secure Data Democratization

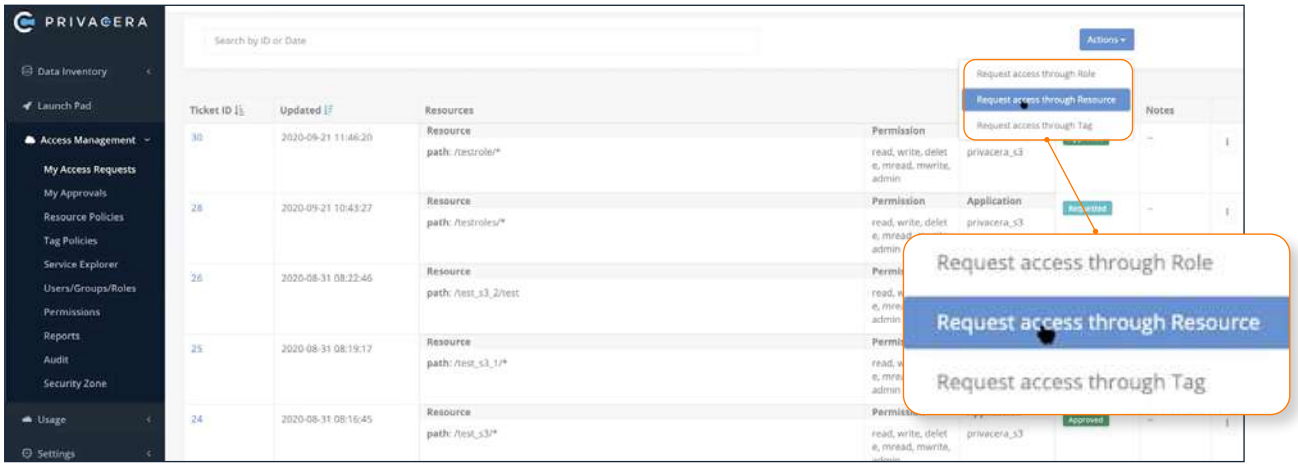
Governed Data Sharing

Enterprises cannot afford to have a bottleneck at the gateway to their data. Therefore, secure data democratization is critical to enable data scientists and analysts to make faster decisions, build more agile teams, and gain a more significant competitive advantage over companies with slower and manual access to data.

To drive latency out of the analytics process, PrivaceraCloud introduces a new data sharing approach, **Governed Data Sharing**, that aligns the components of data access governance with the objectives of the personas involved in the analytics process. Governed Data Sharing deploys a distributed data governance model that delivers an unmatched level of efficiency to accelerate analytical initiatives by grouping functional data—such as sales, marketing, and finance—into Data Domains. For example, a Marketing Data Domain may consist of data sets related to customer/CRM, campaigns, website traffic, or third-party demographics data stored in various cloud services – all of which may be shared together or in separate, curated data-sets. This enables faster access to authorized data without compromising data compliance and privacy.

Data teams using AWS data management features can further accelerate and customize their data access with Privacera. Privacera enables bulk access requests based on users' functional roles, projects, data sharing agreements, or engagements. In addition, data teams can request access to specific sets of data resources (e.g., "à la carte" data) or classifications, enabling:

- Improved cross-functional collaboration for project-based or time-bound assignments
- Decreased onboarding times for new data users
- Faster approvals for requested data access
- Automated manual access request processes



Example: Users can request access to data based on roles, resources, or tags applied to data

The use of Data Domains alleviates the operational burden on IT by placing data owners in direct contact with data consumers to manage access requests, thereby greatly improving collaboration, flexibility, and responsiveness. After IT organizes functional data into Data Domains, access policies are automatically applied to data sets inside them. Although Data Domains are created by IT, they are owned by the person or team with the most knowledge of that functional area. The "data owners" of a particular functional Data Domain are leaders in the relevant line of business, such as the VP of Sales, Marketing, or Finance. Data owners can create shared data sets within their Data Domains and make them discoverable by others in the company.

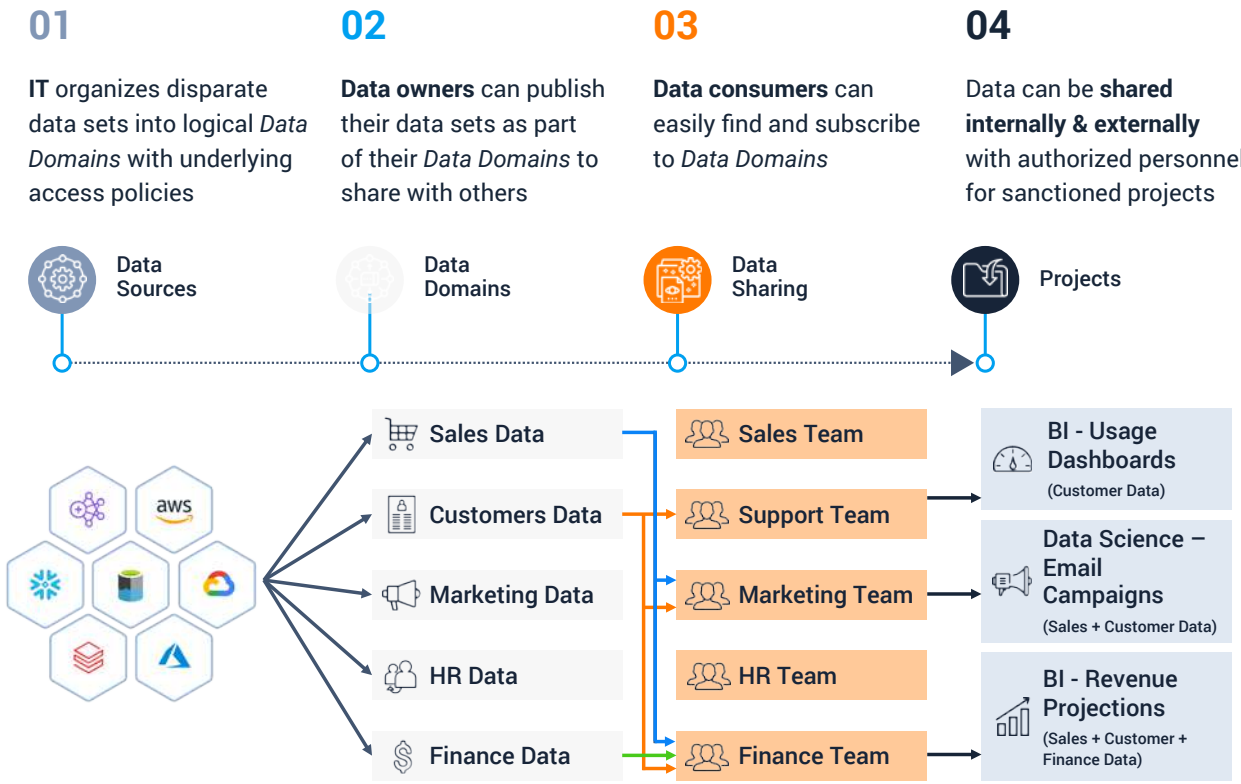
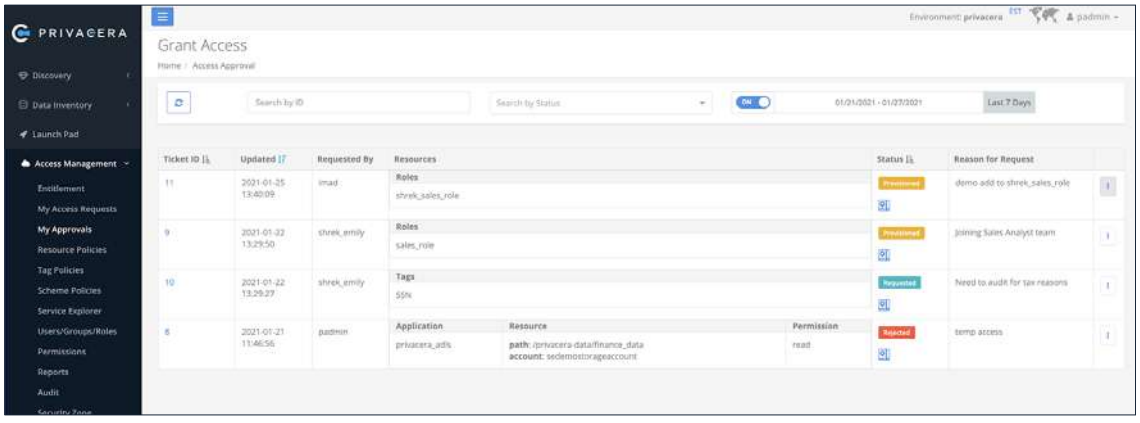


Exhibit: Framework for Governed Data Sharing



Example: Data administrators can grant access approvals with a single click

6 Conclusion

The Privacera Platform and PrivaceraCloud can easily integrate and support AWS and other leading cloud service providers to:

- Discover and classify sensitive data
- Apply fine-grained access controls
- Protect data via encryption, masking, and anonymization
- Enable consistent policies and audit across multi-cloud and hybrid cloud infrastructures– all from a centralized user interface

Privacera and PrivaceraCloud are available on the Amazon AWS Marketplace. With a simplified billing process, customers can more easily get started on PrivaceraCloud to gain end-to-end visibility of sensitive data across their AWS infrastructures. Additionally, customers receive confidence that PrivaceraCloud is a verified and proven technology, tested and validated by AWS.

To learn how PrivaceraCloud on AWS can secure and maximize the power of your enterprise cloud data, visit the AWS Marketplace, use our step-by-step guide to set up your account within minutes, or meet the Privacera team to get started.

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.