PRIVITAR

Cloud Data Privacy 101

Keeping Sensitive Data Safe and Usable for Analytics in the Cloud

www.privitar.com

Forward-looking and disruptive organizations are increasingly adopting modern data architectures to democratize data for analytics and machine learning

Many are turning to AWS to store, catalog, manage and analyze that data. They are taking advantage of the cloud's elastic scalability and leveraging powerful cloud analytics and machine learning technologies to innovate and accelerate data-driven insights.

While centralizing data in the cloud makes it more readily available by enabling easier data access, doing so is not without some common challenges. Migrating sensitive data to the cloud can create new risks, as can provisioning that data for analytics, and widening access to it. It also increases the privacy risks and the points of exposure in the event of a potential breach. When you combine that with the rapid development of sweeping privacy regulations like the GDPR, the CCPA, and the Brazilian LGPD, many organizations, especially those that handle personal or otherwise sensitive information in large quantities (e.g. financial services, insurance, healthcare, telecommunications) place such strong restrictions on the use of sensitive data that it cannot be leveraged for analytical, AI, and ML services. As a result, the value from data often goes unrealized because personal or sensitive data assets are unable to be easily, broadly, and efficiently accessed by business analysts and data scientists.

Safe, usable data is one of the most valuable assets for any organization. When used well, it can be leveraged to gain valuable insights, support data-driven decisions, and fuel everything from identifying and capitalizing on new revenue opportunities to personalized customer experiences.

As organizations continue to build out data pipelines and analytics infrastructures in AWS, they need to keep their data safe while maintaining its analytical utility. Solving their privacy concerns is critical to enabling analytics and data-driven business.

Organizations scaling their use of data in the cloud must leverage automated data privacy tools to ensure data privacy is built into the architecture by design. This will reduce some of the challenges of moving to the cloud. Data can be stored there safely by applying advanced privacy protections to the data itself, significantly reducing the risk inherent in using sensitive information, while retaining the analytical quality and resolution of the data.

Here are five key considerations to ensure your data stays safe, accessible, and usable in the cloud.

1. Take a Layered Approach to Data Protection

Embracing both security and data privacy strategies is the best line of defense to ensure that sensitive data remains safe in the cloud. Traditional security measures such as encryption and attribute-based access control are necessary but not sufficient, as exposure to sensitive data assets increases in the cloud. While established security technologies can prevent unauthorized access to sensitive data and reduce the likelihood of data leakage, they provide no protection when data is in use or once it leaks. Doing this requires protection of the data itself that travels with the data. It requires data privacy, which controls what a user can learn and prevents exposure of individuals, and thereby reduces or eliminates the consequences of a data breach or misuse, whether inadvertent or malicious.

Data privacy complements and strengthens data security, offering a different kind of protection that ensures your data is safe while being used, regardless of where it is stored.

2. Know Your Responsibilities

Preserving privacy means protecting your data subjects. The AWS Shared Responsibility Model defines the split of responsibilities that protect data. While AWS provides an extremely secure cloud platform, you must still take steps to ensure that the privacy of your data is protected.

This is not because AWS are avoiding responsibility. Rather, it relates to obligations set out under GDPR and other data protection regulations. Cloud platform providers are the "data processors" and cloud customers are the "data controllers." This means that you remain responsible for processing your organization's personal data, even if the processing takes place in the cloud. You must ensure that your data is sufficiently safeguarded. You must also understand the data protection laws for each country's data that you're working with, or face significant fines if you fail to comply. You are responsible for protecting your data in the cloud.

3. Build in Privacy by Design

Data privacy must be a key ingredient of your cloud strategy, and it is critical it is embedded into your system design process, not an afterthought. Privacy should be considered from the outset – from the point of data collection all the way through to data usage and data deletion. Privacy then needs to be enforced by building it into a process and a pipeline that makes it easier to do the right thing, because the right thing happens automatically.

The organizations that leverage data driven-insights most successfully view data privacy as intrinsic to the data lifecycle. They take a policy-driven approach that encompasses the entire data management organization.

4. Embrace Automation

Automated data provisioning with privacy protections built in accelerates your ability to access the right data and unlock its potential. It takes a long time to provision data when every step in the process is subjective, made by humans making different decisions each time, and each instance of provisioning requires the same repetitive steps taken the first time. Automating the process provides consistency, transparency, clear auditability, and reduces human error. Automation makes it easier to get the right data into the hands of data consumers, enabling faster time to data-driven insights.

Enterprises scaling their use of data in the cloud need systematic and automated data privacy tools, which remove the friction between data sources and users while still enforcing data privacy.

5. Take a Robust Approach to Data Privacy

Data is safest when it is de-identified, meaning a person can no longer positively or reasonably be identified even when combining different datasets together. There are a range of de-identification techniques and data privacy tools that can be used to minimize the risk of identifying individuals in data sets, while maintaining usability of that data for the business.

Which techniques and tools are right for you? Here are a few questions to help you weigh-up what may be important:

- What kind of data do you need to protect? Operational data or analytical data?
- Is your data to be used for internal consumption only, or do you need to share it with (multiple) third parties?
- How do you plan to manage your de-identified dataset after it has been used?
- Are you planning on creating your data privacy policies centrally, or in a distributed pattern?

- > Are you required to audit your data and trace it if there was a breach?
- > How are you planning on keeping your de-identified data highly usable?

Techniques such as encryption can protect your data, but they also destroy the usability of that data. De-identification techniques can protect your data without sacrificing utility.

Data privacy tools can help you build safe data pipelines, enabling you to protect sensitive data while leveraging data-driven insights.

How Privitar Can Help

Data-driven, cloud-first organizations rely on Privitar to realize the promise of one of their most valuable assets – safe, usable data.

The Privitar Data Privacy Platform™ seamlessly integrates with AWS to protect and manage sensitive data while optimizing its utility for analytic applications in services including Redshift, Athena and QuickSight. This enables organizations to use data to gain timely insights and support data-driven decisions that lead to better products, services and customer experiences, increased revenue and profits, and decreased time-to-market with improved outcomes.

Privitar has the most flexible data privacy engine in the world, capable of supporting batch, data flow and on-demand processing of data across cloud, hybrid cloud, multi-cloud and on-premise environments. So whether you want to protect data as part of your data flow prior to ingesting into your AWS data lake or produce purpose-limited datasets in batch, Privitar has you covered.

Plus, with native and API integrations to the full range of cloud and on-prem big data standards, and a commitment to remain vendor neutral going forward, you can rest easy knowing you can evolve your data pipeline and analytics environments without risking support or vendor lock.

Privitar protects sensitive data for analytics:

- > Reduce privacy risk and meet compliance goals
- > Maximize analytical value of sensitive data
- > Increase the agility of sensitive data by accelerating and democratizing its use

Together this protects organizations from the fallout of any data breach, enables faster and better business decisions, better customer engagement, enhanced innovation and faster time to market. Only Privitar can help achieve privacy, utility and speed of sensitive data together.

aws partner network

Advanced
Technology Partner

Data & Analytics Competency

Security Competency

Contact us:

e: info@privitar.com
t: UK +44 203 282 7136
US +1 857 347 4456

@PrivitarGlobal

www.privitar.com

Find Privitar on the AWS Marketplace:

aws marketplace