SHARDSECURE

Solution Brief

Protection for Machine Learning Datasets

Learn how ShardSecure's technology protects high-value machine learning datasets against common threats like exposure, tampering, and ransomware attacks

Controlling data access and integrity for machine learning datasets

Machine learning datasets are highly valuable to businesses. They serve as the foundation for training models, allowing companies to create accurate and effective machine learning algorithms. With high-quality datasets, businesses can build robust AI models that accomplish everything from better customer targeting and fraud detection to reliable medical diagnoses and compelling artwork. Preparing and choosing the proper data is a crucial part of this process, and the right dataset can determine the success or failure of a project.

Machine learning datasets are also highly sensitive to data tampering and leaks. In what's known as adversarial machine learning or data poisoning, a malicious third-party can change the outcome of an AI model simply by changing or replacing sets of training data. This kind of data manipulation is on the rise, and it can become thoroughly ingrained in a model, ultimately ruining its outputs.

Given their importance and their susceptibility to tampering, machine learning datasets pose a major security challenge to organizations. ShardSecure offers a way to protect these datasets from unauthorized access with agentless file-level protection, robust data resilience, and easy integration.

Machine learning dataset protection with ShardSecure

Historically, organizations have stored their ML datasets within cloud storage providers like AWS S3, Azure Blob, and Google Cloud Platform. But the shared responsibility model, the likelihood of human error, and the undeniable fact of cloud admin access to machine learning datasets all mean that companies must add security measures to keep their datasets safe.

ShardSecure offers an innovative software solution to protect machine learning datasets in the cloud locations of your choice. Our platform provides advanced data security for machine learning datasets against malicious actors and unauthorized third-party access. With our technology, organizations can secure their datasets from internal and external threats without the complexities and overhead of traditional security solutions.

Separation of duties for confidentiality and compliance

ShardSecure's platform allows companies to control exactly who has access to their machine learning datasets in on-prem, cloud, and hybrid- and multi-cloud environments. An innovative approach to file-level encryption, our technology separates infrastructure administrators and cloud service providers from sensitive data.

Our approach offers major benefits for compliance and regulations like the GDPR and SOC 2. It also supports strong data confidentiality and privacy. By protecting data before

it's stored on storage locations, performing regular integrity checks, and offering features like object locking and immutable storage, we ensure that your valuable machine learning datasets remain unintelligible to unauthorized users.

Easy, agentless integration and management

The ShardSecure Platform provides an easy and transparent implementation with no need to change user behaviors or data flows. Without the overhead and complexity of traditional data security solutions, it is vendor-agnostic and appears to other applications as a storage location for machine learning datasets. It also allows for instant data access and fast data migration with just a few clicks.

Robust data integrity and high availability

Unlike traditional solutions, which often focus on data confidentiality alone, ShardSecure ensures high availability for machine learning datasets. Each instance of ShardSecure is a virtual cluster that can run on-premises or in the cloud, and customers can configure two or more virtual clusters for failover. The result is high availability in the storage locations of your choosing.

ShardSecure also provides strong data integrity with its self-healing data feature. If your machine learning datasets are lost, deleted, or tampered with in an adversarial machine learning attack, our platform can reconstruct the affected datasets automatically and transparently, returning them to their original state without costly downtime or disruption to users. This helps keep your mission critical datasets available and accurate.

Each instance of the ShardSecure platform is a virtual cluster that may be deployed on-site or in the cloud. The S3-compatible API and iSCSI interface make it simple for most applications to migrate to ShardSecure with minimal configuration changes. ShardSecure's native multi-cloud and hybrid-cloud support provides a single interface to manage storage locations and move data as needed without impacting application performance.

As a result, ShardSecure has a minimal impact on operations teams. Our solution works in the background as a transparent, zero-downtime event, and data protection is achieved without expending significant resources on running and maintaining complex systems.

Learn more...

ShardSecure integrates seamlessly with your existing security controls and cloud storage providers for ease of deployment. In addition to protection for machine learning datasets, we provide agentless file-level protection, cloud resource optimization, strong data resilience, and native cloud-based ransomware protection.

To learn more about our technology, follow us on social media or visit us online.

@ShardSecure
@ShardSecure
@ShardSecure

101 Avenue of the Americas
9th Floor, New York, NY 10013
United States of America

info@shardsecure.com