SOLUTION BRIEF

Splunk Data Archival Using Microshard™ Technology

June 2021

Zack Link

Senior Security Engineer

SHARDSECURE®

Overview

Splunk is a popular data platform for ingesting, searching and visualizing data. As companies rely on this platform to ingest more data, and to keep it for longer periods of time, data archival becomes more of an issue. The user needs to balance the performance and cost of growing amounts of data. One popular strategy is to shift older, less frequently used data to slower, cheaper storage. Traditionally this was done with the concept of hot, warm, cold and frozen storage buckets. Hot and warm storage was used for the most recent and time critical data, while cold storage placed older data on slower, cheaper storage. Frozen storage was used for archival purposes and would need to be restored to be searched or used.

In Splunk Enterprise 7.2, Splunk released a new feature called SmartStore. SmartStore does away with the concept of hot/warm/cold/frozen storage and instead stores all data remotely, always in a thawed state (no frozen storage that needs to be re-ingested to be searched), while keeping relevant data cached on the Splunk cluster. Thus, all hot storage is kept locally, and any data that is frequently accessed from warm or cold storage is stored remotely and dynamically cached on the Splunk cluster. Frozen storage is no longer relevant. This practice enables companies to leverage inexpensive and scalable cloud storage for all Splunk data.

Splunk ingests all types of data, including security data which is often sensitive in nature. Thus, Splunk recommends encrypting remote data storage. This is not always an easy task, and the complexity level goes up depending upon the quantity of Splunk servers and the type of key management solution used. Per Splunk, uptime can be impacted if a server is offline when a periodic key rotation happens, or if a key request is throttled by the KMS platform.

Solution

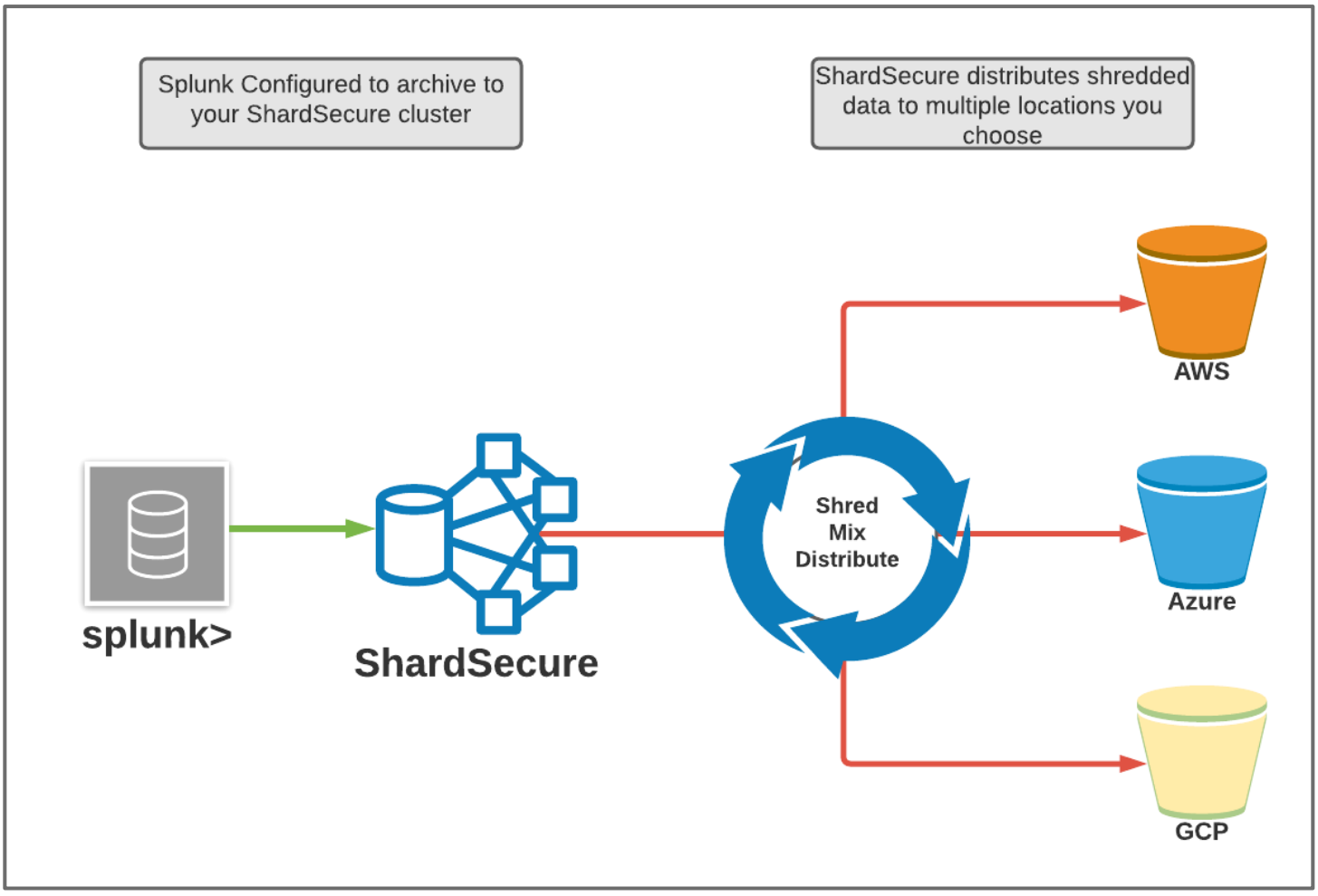
The ShardSecure cluster has an S3-compatible API for file storage and retrieval and is easy to configure in Splunk as the remote data storage location. There is no longer a need to configure and manage encryption through Splunk. The ShardSecure high availability cluster will manage all data at rest within your sensitive Splunk data. The data will be shredded, mixed and distributed to multiple storage locations of your choosing, ensuring the data is useless in case of a breach or data leak. Server-side encryption can also be enabled to transparently provide defense in depth, without the overhead of managing encryption keys. Since ShardSecure uses parallel I/O to store and retrieve the distributed data, in some cases, users will see an improvement in performance as well.

Advantages

1. Microsharded data remains incomplete, unreadable and unusable by an attacker in any storage location
2. No need to manage encryption or encryption keys
3. ShardSecure can operate in any major cloud environment (AWS, Azure, GCP, IBM, etc), including multi-cloud or hybrid-cloud environments, while Splunk currently only supports AWS and GCP, with no multi or hybrid cloud support

4. ShardSecure supports transferring microsharded data from one cloud to another, totally transparently to Splunk, with no downtime
5. ShardSecure will often perform faster than an S3 bucket, due to parallel I/O and the distributed file system utilized by the ShardSecure cluster, which writes fractional files to multiple locations
6. No third parties have access to your data, including ShardSecure. Even your cloud provider has no access to the underlying data

Diagram 1



SHARDSECURE®

+1 (800) 760 9445

info@shardsecure.com

@ShardSecure

101 Avenue of the Americas,
9th Floor

New York, NY 10013

United States of America