SPOTLIGHT

Sponsored by: Pure Storage

The performance of a SIEM strategy is interwoven with the capabilities of its storage system. The right storage system will help the security team accomplish their work to their highest ability.

How the Right Storage Can Help Improve Enterprise SIEM Operations

March 2022

Written by: Michelle Abraham, Research Director, Security and Trust; and Eric Burgener, Research Vice President, Infrastructure Systems Group

Introduction

Security and information event management (SIEM) platforms ingest and store log data from other security tools to perform correlation and data analysis to alert the security team of items that could be issues for the organization. The amount of data to be ingested and stored is up to each organization, and the number of sources for relevant SIEM data is on the rise. Figure 1 shows the range of SIEM data ingestion in the United States, with 61.2% of enterprises requiring 1TB of storage or more per day. The number of sources for relevant SIEM data is also on the rise.

Given today's high data growth rates and the increasing importance of defending against ransomware and other attacks from bad actors, SIEM storage systems must meet a stringent set of requirements.

Storage requirements will vary based upon workloads, the compliance and regulatory environment, governance, and protection/recovery requirements. There is no doubt, however, that the ability of security teams to meet their objectives can be significantly impacted by the performance, availability, and scalability of the storage system used to capture and analyze relevant data and events. While the initial SIEM efforts of many enterprises may start by using direct attached storage (DAS), this approach poses challenges as the amount of SIEM data increases. Disaggregated scale-out storage architectures enable more efficient sharing of purchased storage capacity across different servers, include enterprise storage management capabilities that drive higher availability and increased efficiencies, and make storage administration and scaling storage capacity much easier and more cost-effective.

AT A GLANCE

WHAT'S IMPORTANT

Storage systems used to support SIEM systems must be able to accommodate modern data types and provide modern application support.

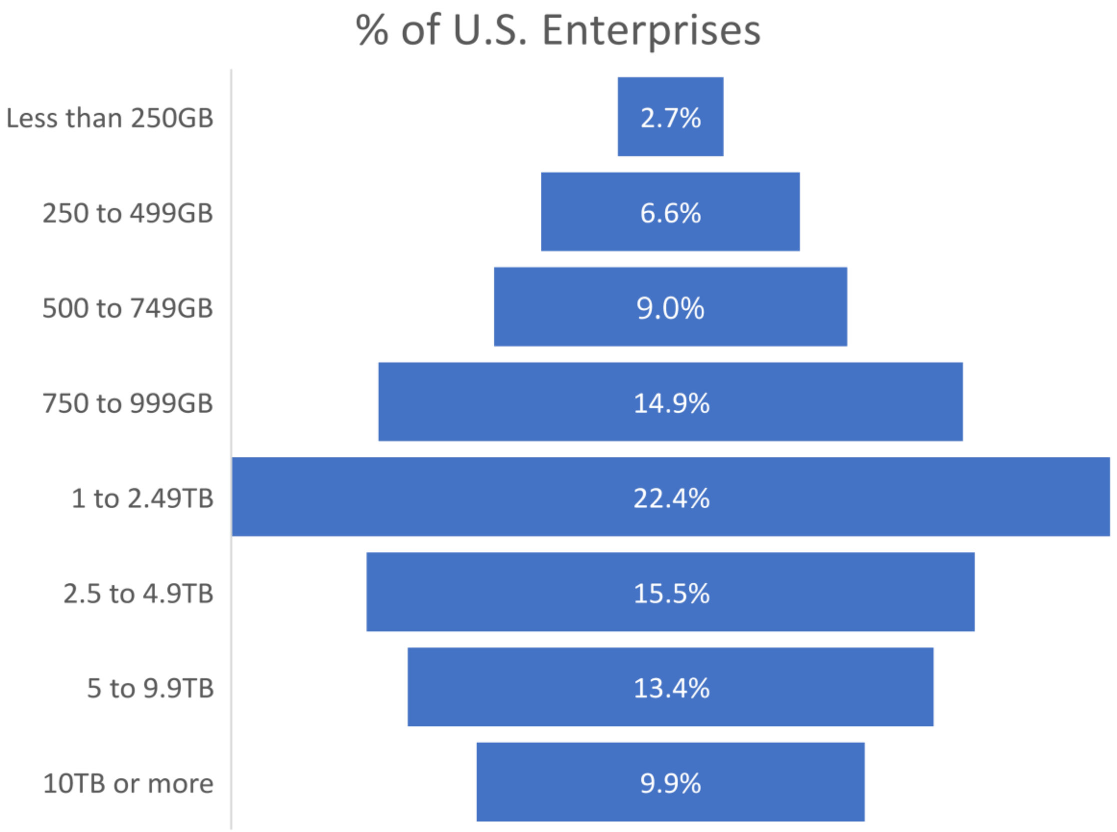
To meet these requirements, the storage on which SIEM runs must support unstructured (i.e., file- and/or object-based) storage as well as high performance, availability, and cost-effective scalability.

KEY TAKEAWAYS

Enterprise-class all-flash systems using disaggregated, scale-out architectures are a good fit for storage infrastructure supporting SIEM.

It is important to map storage system capabilities to site-specific SIEM requirements to make an optimal storage infrastructure purchase decision.

FIGURE 1: ***SIEM Data Ingested by Day in the United States***



n = 335

Source: IDC's Security SOC Tools Survey, November 2021

Some key capabilities to evaluate when purchasing a storage system that will host SIEM workloads include:

- High ingest performance to capture relevant data without impacting information and event collection capability
- Sufficient performance to enable real-time search, alerts, and correlation to provide comprehensive security protection, delivery of forensic evidence to authorities, and demonstration of compliance with applicable regulations (Also important is a system's ability to keep up with ingest requirements as data grows. All these tasks point to a need for linearly scalable performance in the storage system.)
- High availability to ensure that component failures and/or upgrades in the storage system do not impact an enterprise's ability to protect and/or recover its information assets

- Multi-petabyte capacity that can scale to collect the data needed from a growing number of sources (This enables enterprises to retain data over long periods to improve the accuracy of detection and correlation capabilities while also ensuring that an enterprise's SIEM capabilities will function even as data grows at an explosive rate. Administrators should evaluate this feature not only in terms of raw storage capacity but also by how many billions of files and/or objects can be supported. The more data the SIEM system possesses, the more information it has for its artificial intelligence and machine learning algorithms to differentiate between normal and abnormal behavior.)
- Unified unstructured storage capabilities (i.e., supporting file- and object-based data on the same storage platform) that make a system better suited to capture, store, protect, and analyze security telemetry since most of that data will be unstructured

The faster the search, the more efficient the security team can be when solving issues that have arisen. A storage system's metadata architecture can have a large impact on how quickly searches complete and correlations are discovered, with faster time to results driving better overall security. The quicker they find indicators of compromise, the less time an adversary has to breach the system. If an adversary is in, the security analyst wants to recognize and shut down their access as quickly as possible to provide less opportunity for exfiltrating data or detonating ransomware. The resulting productivity helps the security team improve its mean time to detect and mean time to respond, two common measures of security team proficiency.

Other storage system capabilities of interest impact security, availability, infrastructure efficiency, and cost. The storage array should support 256-bit encryption, giving administrators the option to use server-side or storage-side encryption for captured data, and in-line encryption within the storage should not impose noticeable performance impacts. For federal customers, FIPS certification may also be an issue. Role-based access control gives enterprises more options for providing better security when managing the storage.

Storage systems must be highly available to minimize any period where security operations are compromised. This means that systems must not only handle component failures transparently and feature redundant, hot-plug hardware but they must also allow upgrades, security patches, and standard maintenance operations to be performed nondisruptively. Capabilities like snapshots and replication that enable rapid recovery and protection against site disasters are also important.

Infrastructure efficiency addresses the challenges of capacity utilization, energy and floorspace consumption, and ease of expansion. Features like compression can double the amount of data storage versus raw storage consumption, while all-flash storage can significantly reduce both energy and floorspace consumption relative to more traditional architectures that leverage hard disk drives (HDDs). It is exceedingly difficult to scale DAS approaches without overprovisioning storage that cannot be easily shared, increasing the complexity of expansion and driving costs up.

Few (if any) of these capabilities are available with DAS approaches, and they are a major reason so many security-conscious organizations move to disaggregated, scale-out storage architectures as their SIEM data sets grow.

The need to effectively handle different data types also has enterprises evaluating object-based storage (instead of the block-based approach used by DAS). Object storage is better suited to handling a variety of different data types. It also provides rich metadata capabilities that aid in rapid search, is far more resilient and cost-effective than block-based DAS, and offers easy scalability into the multi-petabyte range and beyond. Many object platforms have a number of enterprise-class storage management capabilities that deliver benefits in SIEM environments. These capabilities include erasure coding (for very resilient, low-cost data protection), immutability, compression, snapshots, encryption, and replication (among others). Many SIEM application vendors have noted this trend and are moving to disaggregated object-based storage platforms built around scale-out architectures.

Benefits

The right storage system choice can have a major impact on the quality of security an enterprise can enforce. Higher performance, whether measured in terms of latency, I/O operations per second, or throughput speeds routine data capture and analysis and delivers results to protect or recover the business faster. High availability means that the SIEM environment always has access to the data to meet security objectives. Scalability means that the SIEM environment can scale as an enterprise grows, continuing to provide predictable performance against ever larger data sets. And storage infrastructure efficiency, enabled through capabilities like high performance and capacity density and data reduction, helps to drive SIEM infrastructure costs lower.

Considering Pure Storage FlashBlade

Pure Storage is an enterprise all-flash array (AFA) vendor with a broad portfolio of offerings that cover primary and secondary as well as structured and unstructured data workloads. As early as 2013, a group of innovators at Pure were evaluating the need to bring all-flash storage to data-centric, unstructured workloads. "Big data" was driving change across the industry. But infrastructure architectures were dominated by big, slow disks. The opportunity had arrived to rethink infrastructure for unstructured data with an all-flash architecture. This culminated with the 2016 introduction of FlashBlade. FlashBlade offered a multi-petabyte platform with high throughput that sped workloads with an I/O capability and management simplicity. Based on a scale-out all-flash design, FlashBlade is a unified fast file and object (UFFO) offering targeting workloads that use popular access methods like NFS, SMB, and S3. FlashBlade has been a product with significant traction. In 2021, the product hit a lifetime revenue milestone of $1 billion. Currently, it is deployed in 25% of the Fortune 100. The product has an installed base that includes (but is not limited to) over 200 customers that have spent more than $1 million on FlashBlade each.

FlashBlade has an architecture that is highly optimized around flash technologies. While these optimizations improve performance, increase media endurance in write-intensive enterprise environments, and make very efficient use of flash capacity and other system resources (compute, memory, bandwidth, and others), they also drive significant benefits around security:

- **Highly scalable, parallel architecture.** FlashBlade uses a parallel architecture where writes are highly distributed within the system depending on the size of the file or object being written. Read performance scales as data is spread across more FlashBlade storage devices. FlashBlade does not use off-the-shelf SSDs. It uses all-flash "blades," built by Pure Storage, which deliver better performance, storage density, capacity utilization, media endurance, and lower cost per gigabyte than available from many off-the-shelf devices. The storage operating system in FlashBlade, called Purity//FB, comprehensively manages all the flash media in a system directly, giving it an ability to optimize resource utilization at a global level (wear leveling, free space management). The metadata architecture design in FlashBlade is made to handle tens of billions of files and/or objects simultaneously.

  The system supports two levels of scalability. A single FlashBlade enclosure can accommodate up to 15 blades, each of which includes both performance and capacity resources. A single system can include up to 10 enclosures, all of which are managed under a single namespace. Each enclosure supports up to roughly 8PB of capacity, which allows them to store over 15PB when using FlashBlade's built-in compression. This results in a highly scalable system that produces predictable read and write performance in both small file and large file environments.

This design works well with security analytics applications. First, it allows ingest performance to be scaled simply and easily as needed to deliver several terabytes of bandwidth (in a 10-enclosure configuration). Cache-based architectures generally hit a write cliff at some point as additional data sources are added, so there are limitations to ingest performance. Low-latency FlashBlade write performance means that each enclosure can meet significant real-time data ingest requirements and easily scale by adding blades and/or enclosures as needed (all of which can be done nondisruptively). For workloads where real-time analytics are important, such as fraud detection, surveillance video analytics, malware signature recognition, and Internet of things (IoT) use cases, scalable ingest performance is critical.

Second, the design enables FlashBlade to support more consistent performance at scale and high degrees of data concurrency. For search operations, alerts, correlation, and queries, its predictable performance enables much larger data sets to be simultaneously accessed with high-performance capabilities that drive faster results, better insights, and more timely business decisions. The ability to handle large data sets in this manner is particularly important in PACS, genomics sequencing, real-time security, and electronic design automation (EDA) environments where FlashBlade is widely deployed. High concurrency also allows FlashBlade to simultaneously service varying workloads – ingest, search, and analysis – at the same time with predictable high performance, speeding time to results and making very efficient use of storage resources. This also allows in-line operations like compression and encryption to be performed without impacting application performance.

Third, it enables very high-speed data movement. Because of its fast restore capabilities, FlashBlade is often used in data protection environments that require low RPO/RTO or have to protect very large data sets from ransomware attacks. SafeMode snapshots, a built-in FlashBlade feature, enables the rapid creation of immutable (read-only) snapshots of backup data and associated metadata catalogs after a full backup has been performed. Data can be recovered directly from these snapshots, helping organizations bounce back after ransomware attacks and even rogue administrator activity.

- **Software-defined capabilities that assist security operations.** FlashBlade includes FIPS 140-2-compliant encryption that is used for data at rest and can also be used to ensure replicated data in-flight is protected. Immutable snapshots provide a defense against ransomware attacks, and these can be moved to offsite locations with replication for "air gap" protection. Role-based access control (RBAC) helps maintain security in administrative operations. In addition, FlashBlade supports management audit logs as well as the Kerberos network security protocol.

- **High availability.** When storage systems are used as the foundation for security operations and workloads, they are integral in protecting data and business operations from bad actors, but they can't perform if they aren't up and running. FlashBlade provides enterprise-class availability with transparent recovery from component failures (e.g., blades, power supplies, fans) and nondisruptive operations for component replacements, upgrades, and expansions.

- **Unified storage support.** Much of the SIEM data collected now and in the future will be unstructured. FlashBlade can store both file- and object-based data in its native format for high-performance access without semantic loss. This makes it an efficient platform for unstructured workload consolidation, and it enables easier sharing of data when analytics and other pipelines span multiple stages. FlashBlade's efficiency is also evident in its energy and floorspace consumption. A single enclosure (which supports over 1.5PB of effective capacity) takes up only 4U of rack space and consumes 1,800 watts nominal at full configuration.

Pure Storage has long competed against vendors that added SSDs to legacy architectures that were originally designed around HDDs (as well as those still using HDDs). By starting with a blank sheet of paper to design systems that are optimized for flash (not spinning disk) media, the vendor delivers very efficient systems that make the most of that media. The performance, availability, scalability, ease of use, and efficiency of resource utilization make FlashBlade an attractive platform for SIEM workloads and operations.

Challenges

Chief security officers and other security administrators may not have thought much about the importance of choosing the right storage architecture to provide the foundation for SIEM operations. However, it should be clear from the preceding discussion that there are very specific storage infrastructure capabilities that can significantly improve security for many enterprises. Getting security personnel to understand the value that different storage systems bring to the table for digitally transforming enterprises requires more awareness of just what those capabilities are and what value they can drive for customers. As security becomes ever more important, and data sets continue to grow at a rapid rate, it will be important for enterprise storage vendors like Pure Storage to help their customers make that connection.

Conclusion

Security is a paramount consideration for IT organizations today, and the quality of SEIM operations in those organizations is a key determinant of business success. Many enterprises start with DAS-based approaches to SIEM but quickly run into scalability challenges that result in complex, costly infrastructures that are difficult to manage and scale. As has happened in other operational areas of IT, many workloads initially placed on DAS architectures are moving to disaggregated, scale-out architectures that offer a variety of advantages in managing workloads depending on large, growing data sets. These newer architectures enable more efficient sharing of data across different workloads, maximize capacity utilization, and provide a number of storage management features not available with DAS that improve the performance, availability, scalability, and manageability of the storage that supports SIEM operations. They also allow compute and storage resources to be scaled independently, helping IT to achieve a more cost-effective allocation of those resources as needed over time.

> Security is a paramount consideration for IT organizations today, and the quality of SEIM operations in those organizations is a key determinant of business success.

Pure Storage offers FlashBlade, a disaggregated, all-flash, scale-out storage system that is widely deployed in enterprises across many workloads. The performance, availability, scalability, and manageability of FlashBlade, a unified, fast file, and object platform that simultaneously supports both file- and object-based storage, make it a good fit for storage infrastructure supporting SIEM operations for successful, growing companies. Its all-flash design and unique architecture make it an extremely compelling alternative to DAS and/or other HDD-based storage infrastructures that might be used as the foundation for SIEM operations. IDC believes the market for SIEM storage systems will continue to grow, and to the extent that Pure Storage can address the challenges described in this paper, the company has a significant opportunity for success.

About the Analysts

Michelle Abraham, *Research Director, Security and Trust*

Michelle Abraham is Research Director in IDC's Security and Trust Group responsible for the Cybersecurity Analytics, Intelligence, Response & Orchestration (AIRO) practice. Ms. Abraham's core research coverage includes the cybersecurity AIRO market, focusing on Security Information and Event Management (SIEM) platforms and device & application vulnerability management.

Eric Burgener, *Research Vice President, Infrastructure Systems Group*

Eric Burgener is Research Vice President within IDC's Enterprise Infrastructure Practice. Mr. Burgener's core research coverage includes Storage Systems, Software and Solutions, quarterly trackers, end-user research as well as advisory services and consulting programs. Based on his background covering enterprise storage, Mr. Burgener's research includes a particular emphasis on flash-optimized arrays, emerging persistent memory technologies, and software-defined storage. He is an active participant in the IT Buyers Research Program at IDC and blogs throughout the year on the topic of Infrastructure and Data Management.

IDC Custom Solutions

The content in this paper was adapted from existing IDC research published on www.idc.com.

IDC Research, Inc.
140 Kendrick Street
Building B
Needham, MA 02494, USA
T 508.872.8200
F 508.935.4015
Twitter @IDC
idc-insights-community.com
www.idc.com

This publication was produced by IDC Custom Solutions. The opinion, analysis, and research results presented herein are drawn from more detailed research and analysis independently conducted and published by IDC, unless specific vendor sponsorship is noted. IDC Custom Solutions makes IDC content available in a wide range of formats for distribution by various companies. A license to distribute IDC content does not imply endorsement of or opinion about the licensee.

External Publication of IDC Information and Data — Any IDC information that is to be used in advertising, press releases, or promotional materials requires prior written approval from the appropriate IDC Vice President or Country Manager. A draft of the proposed document should accompany any such request. IDC reserves the right to deny approval of external usage for any reason.