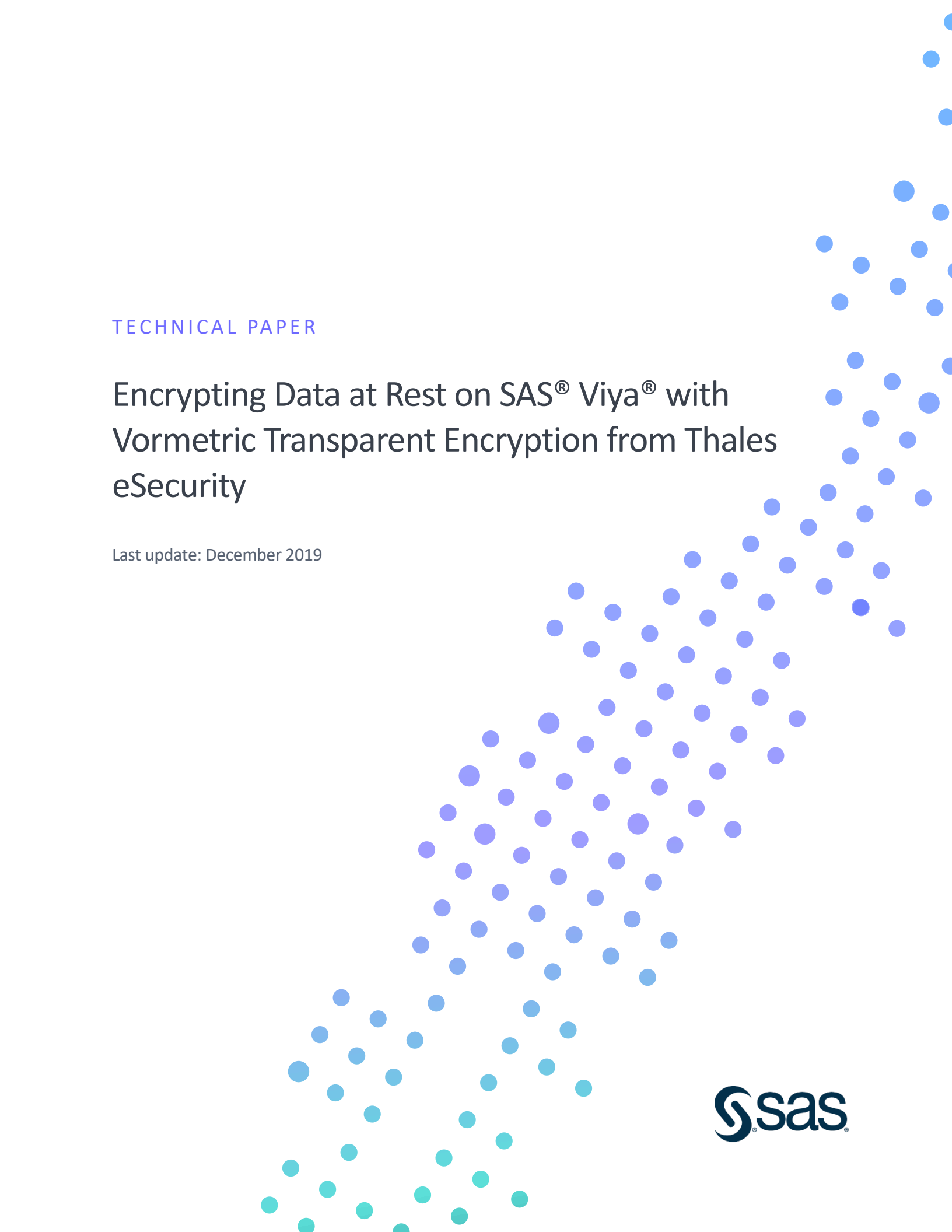
TECHNICAL PAPER

Encrypting Data at Rest on SAS® Viya® with Vormetric Transparent Encryption from Thales eSecurity

Last update: December 2019



Contents

Abstract

Highly regulated customers must meet specific regulatory requirements to secure their data at rest. Vormetric Transparent Encryption from Thales eSecurity is one solution whose technology complies with these requirements in order to encrypt data at rest, which is stored and processed by SAS Viya.

Introduction

With increasing scrutiny of the use of Personally Identifiable Information (PII) and strict industry standards for data security, the stakes have never been higher for governments, financial institutions, and other companies that handle highly sensitive data and that are subject to rigorous regulations. Although high-speed data analytics revolutionizes industries and empowers institutions to glean groundbreaking insights, it also underscores the need for care and utmost security when processing sensitive data. The breach of a single database can expose millions of full sets of PII, causing organizations to suffer major losses when their customers' data is stolen and exploited. That's when encryption becomes a powerful ally to analytics — even if a breach occurs, strong encryption can protect data with complex algorithms that render the data useless to anyone except the keyholder.

SAS is a long-time market leader in the analytics arena, and SAS Viya provides uniquely powerful and adaptable features for real-time business analytics on an enterprise scale in the cloud. Although cloud computing offers a complementary technology encompassing game-changing capabilities, many regulated companies, government entities, and banking institutions have delayed deployment of mission-critical applications, such as SAS, in cloud environments, opting to keep data on-premises in the interest of security. These organizations could benefit from the scalability and elasticity of the cloud, but their increasingly strict compliance and security requirements require secure and reliable encryption technology before they can make the transition to the cloud.

How can highly regulated customers harness the power of SAS while encrypting their sensitive data to meet the strictest security and compliance requirements? SAS Viya raises the bar on security by enabling customers to seamlessly integrate their SAS Viya deployments with Vormetric Transparent Encryption from Thales eSecurity. Vormetric, a leader in enterprise security for data at rest across physical, virtual, and cloud environments, delivers data-at-rest encryption with **centralized key management, privileged user access control** and **detailed data access audit logging** that helps organizations meet compliance reporting and best practice requirements for protecting data, whether it resides on premises or in the cloud. When Vormetric is added to SAS Viya, companies retain control over their data, security keys, and security policies.

Enterprise-Class Security for Enterprise-Class Analytics

For years, the key to success for any business intelligence solution has been the process known as extract, transform, and load (ETL). Selecting the right tool to bring data from disparate sources and transform it before loading into a target destination was the critical factor in building a data warehouse or a data mart to support an organization’s business intelligence projects. In fact, it became so important that the process became synonymous with the tool and the technology became known as ETL technology, which spawned many ETL tools.

Centralized Key Management

Centralized key management brings all facets of cryptographic key management (including hardware, software, and processes) into one physical and logical location. The solution handles fundamental activities such as lifecycle management, auditing, and security from one centralized source.

• Customers can control data by managing encryption keys and access policies from their local data center for their on-premises and cloud data, even in hybrid environment deployments.

Privileged User Access Control

Vormetric Transparent Encryption helps to prevent insider abuse of privileged user credentials by providing LDAP-based group access controls, which effectively limit access to privileged users within the organization. Although sensitive files can be encrypted, metadata can remain unencrypted.

• Data scientists or analysts can create and run their SAS jobs without seeing any protected data.

Detailed Data Access Audit Logging

Granular data access audit logs delivered by Vormetric Transparent Encryption are useful not only for compliance but also for the identification of unauthorized access attempts. Logs capture details about access in order to build baselines of authorized user access patterns for intelligent pattern analysis and alerting.

• Once Vormetric is applied to your SAS Viya deployment, you can plug-and-play by easily integrating Vormetric's audit log intelligence with any other security and information event management (SIEM) system of your choice.

• With data access logging, you can receive alerts about unusual data access activity. Logs specify when users and processes accessed data, under which policies, and if access requests were allowed or denied. The logs can reveal when a privileged user submits a command like “switch user” in order to attempt to imitate another user.

Vormetric: Enhancing Security and Data Control

Vormetric Encryption Architecture Overview

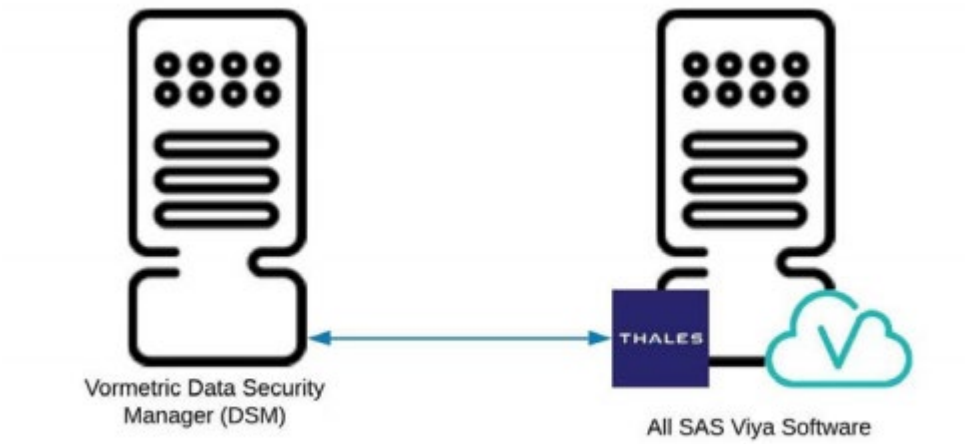
Vormetric can secure all data at rest, including logs, unstructured data, and databases for both physical and cloud-based deployments. All the customer needs to do is point Vormetric agents to the files, folders, libraries, or other areas that they want to encrypt, and then manage their policies from the Data Security Manager (DSM).

SAS Viya supports multiple deployment scenarios. The most common scenarios include a single machine deployment, a full deployment with a separate single-machine SAS Cloud Analytic Services (CAS), and a full deployment with a distributed CAS server. The following high-level diagrams show these three common deployment scenarios with the Vormetric agent that are rendered with simplified configuration options.

SAS Viya on a Single Machine

In a single-machine deployment of SAS Viya, all software is located on a single machine. In this scenario, the Vormetric software agent is installed on a single machine, and the agent communicates with the Vormetric DSM. In the diagrams below the Vormetric software agent is represented by the Thales logo.

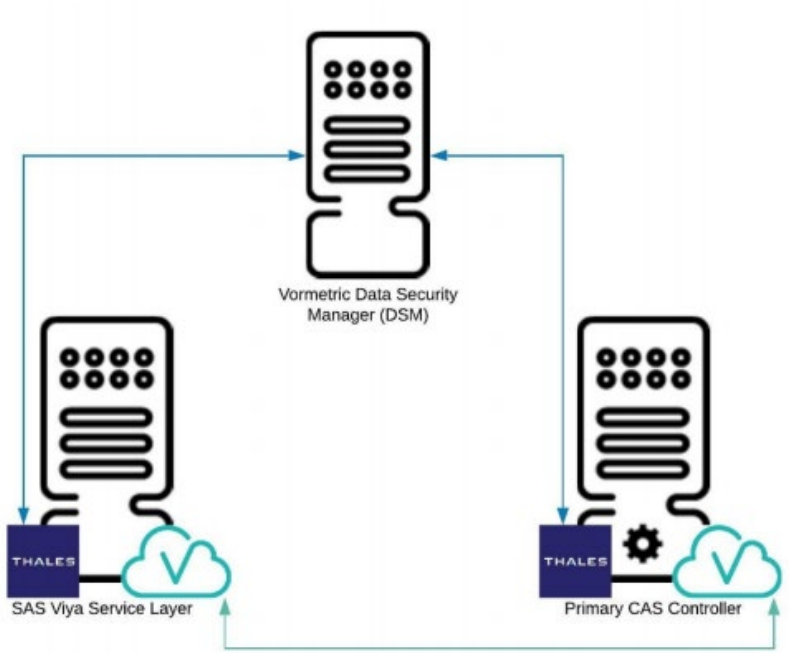
Figure 1. SAS Viya in a Single-Machine Deployment with Vormetric



SAS Viya with a Separate, Single-Machine CAS Server

In this deployment scenario, SAS Viya is spread across two machines. Services and applications exist on one machine, and the CAS Server and other components exist on a separate machine. The Vormetric software agent must be installed on the SAS service layer as well as on the CAS server.

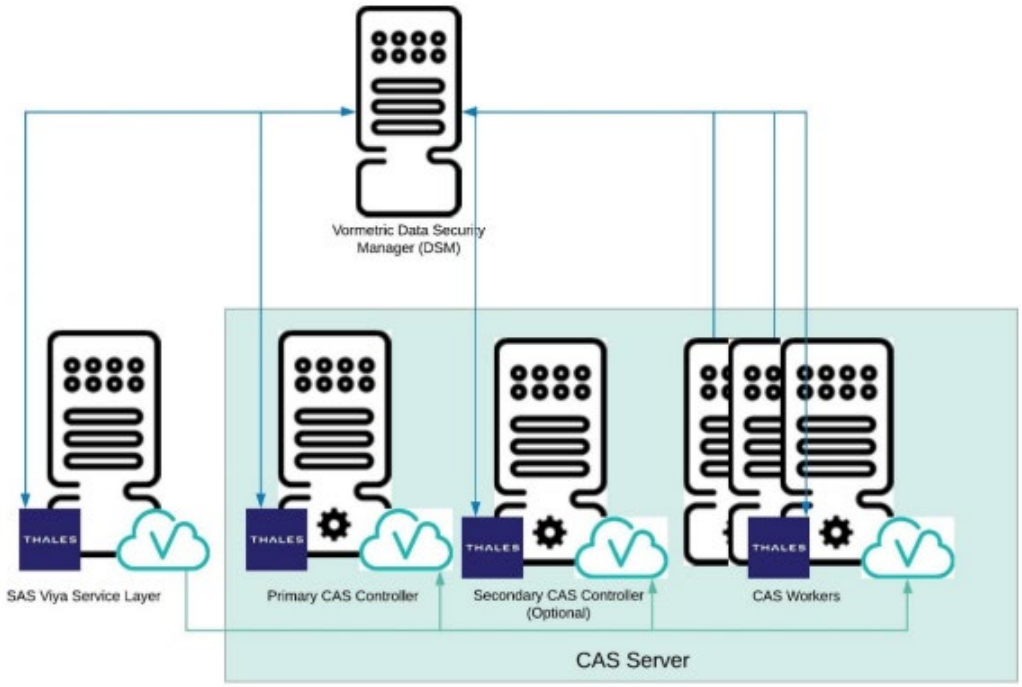
Figure 2. SAS Viya in a Multi-machine Deployment with the CAS Server on a Separate Machine



SAS Viya with a Distributed CAS Server

In this scenario, SAS Viya is deployed across multiple machines. The service layer and applications remain on the same machine, and the CAS server is deployed in a distributed configuration across multiple machines in order to take advantage of massively parallel processing (MPP). There are two CAS controllers and a group of CAS worker nodes. This type of deployment is available only on Linux and requires the use of a shared file system. The Vormetric software agent must be installed on each machine, including all of the worker nodes. Note that the diagram does not depict the agent on each worker node for simplicity's sake.

Figure 3. SAS Viya with a Distributed CAS Server



Deploying Vormetric Transparent Encryption with SAS Viya

The deployment of Vormetric Transparent Encryption with SAS Viya is accomplished via five primary steps. The following provides a high-level introduction. For more details about how Vormetric can be deployed, contact Thales eSecurity.

Deploy the Data Security Manager

Implementation starts with the deployment of the Vormetric Data Security Manager (DSM), which enables the customer to manage encryption, policies, keys, and security intelligence as a one-stop shop. The DSM enables administrators to specify data access policies, administer DSM users and logical domains, generate usage reports, register new hosts, access security logs, and manage third-party keys and digital certificates. Note that the DSM is available as either a hardware or virtual appliance.

Deploy the Vormetric Software Agents

The Vormetric software agents run on servers or virtual machines to control access to files, folders, and volumes, and to report activity to the DSM. The agent runs at the file system or volume level on a server, and is available for supported Windows, Linux, and UNIX operating systems. The agents perform the encryption, decryption, and access control activities locally on the operating system that is accessing the data at rest.

Create Encryption Keys and Transform Existing Data

Vormetric supports multiple key types and industry-standard encryption algorithms. Working with existing data requires that the data be transformed from clear-text to encrypted via a data transformation policy before any operational policies are applied. In greenfield environments that have no existing data, the transformation step can be skipped and an operational policy can be applied immediately.

Create Policies and GuardPoints

Vormetric uses the term “GuardPoint” to describe the resource that the customer intends to protect. Typical examples are a directory or file. Policies, which consist of rules, define the level of access for a given GuardPoint. Vormetric offers granular policy definition capabilities. Data access can be based on resource, users and groups, processes, and time of day with or without an action. For example, a policy could specify that Bob and the Accounting group can access resource `/opt/myGroup/myData` with Read-only permissions between 9 a.m. and 5 p.m. The same policy could allow Alice in the Admins group to have full (Read and Write) access to `/opt/myGroup/myData` while Pratima has Read-Only access to the metadata attributes of files in the same resource. Policies are evaluated and applied in order much like a firewall rule list. The first applicable rule set in a policy is applied and evaluation stops.

Tune Policies

New policies can be run in "Learn Mode," which permits, but logs, all actions. From the logs, a user can see how the policy is evaluated and what could have happened if the policy had been enforced. Once the appropriate access levels have been determined and the policies have been tuned, “Learn Mode” can be disabled and policies are enforced normally.

SAS and Vormetric Integration and Performance

To ensure synergy with minimal performance impact, SAS and Thales collaborated to integrate and test the combined solution of SAS Viya with Vormetric applied to SAS Visual Analytics and SAS Visual Investigator. These testing results represent the specific scenarios that were tested against. However, it is important to note that results can vary based on customer hardware, data size, and the specific deployment environment.

Testing of the SAS Viya and Vormetric integration showed the following results:

- The full suite of functional testing was performed and passed as expected for SAS Visual Analytics, SAS Visual Investigator, SAS Studio, and the middle-tier platform, which is the foundation of the SAS Viya application suite. The middle-tier platform includes core infrastructure components such as PostgreSQL, RabbitMQ, and Consul.

- The performance overhead of encrypting data at rest with Vormetric Transparent Encryption is minimal in SAS Viya 3.5, SAS Visual Analytics, and SAS Visual Investigator.

High-Level Performance Impact Summary

The following table shows a high-level summary of the performance impact for the CAS server (where the bulk of the analytics workload was executed), PostgreSQL database, SAS Visual Analytics, and SAS Visual Investigator.

Table 1. *High-Level Performance Summary*

Area	Performance Impact	Summary
SAS CAS Server CPU Usage	2.5–6.9%	The percentage of CPU usage during testing ranged from 2.5 to 6.9% for various operations using encrypted data. For comparison, CPU usage ranged from 2.0–5.9% for the same operations using unencrypted data.
PostgreSQL	1.5%	Minimal or no performance impact due to encryption.
SAS Visual Analytics	2%	Minimal or no performance impact due to encryption.

Encrypted Data at Rest with Minimal Performance Impact

Test results confirm that both functionality and peak performance are preserved when Vormetric Transparent Encryption is applied to data at rest in SAS Viya. During this proof of concept with Vormetric, a variety of tests were run to ensure that the integrity of SAS functionality and performance would be maintained when data at rest was encrypted by Vormetric Transparent Encryption. The following is a brief summary of the test results.

Cloud Analytic Services

Cloud Analytics Services (CAS) is the in-memory analytic server for SAS Viya and is where core analytics is performed. Functional and performance testing was performed to compare CAS behavior with non-encrypted data versus encrypted data. All functional testing completed without errors and performance testing using the recommended configuration showed minimal performance impacts. In the following sections, a summary of analyses and an interpretation of results is provided.

CAS Cache Directory

To understand these test results, it is important to understand the high-level operation of the CAS server. As a memory efficiency, CAS organizes in-memory data in blocks, and memory maps the blocks. The blocks are stored as temporary files in directories on the machine. The CAS controller also uses the cache directory to temporarily store uploaded files. The temporary nature of these files, combined with the frequent access that is required by CAS to operate efficiently, leads to the recommendation against encrypting the CAS cache directory as part of a data-at-rest protection strategy

Performance

Testing with the recommended configuration showed minimal impact on CAS performance. Testing scenarios

included the same common CAS operations that were performed against the two CAS deployment types: symmetric multi-processing (SMP) and massively parallel processing (MPP). A CAS SMP deployment exists on a single machine, and a CAS MPP distributed deployment spans multiple machines. Each of these deployment types was tested with both solid-state drives (SSD) and hard-disk drives (HDD) as well as three data set sizes (8GB, 12GB, 42GB). All tests were performed using SAS data sets in the proprietary sas7bdat and hdat formats.

Symmetric Multi-Processing (SMP) - CAS Deployment on a Single Machine

Figure 4. *Performance Impact of CAS-Encrypted Data in an SMP Deployment Using an SSD with Multiple Data Set*

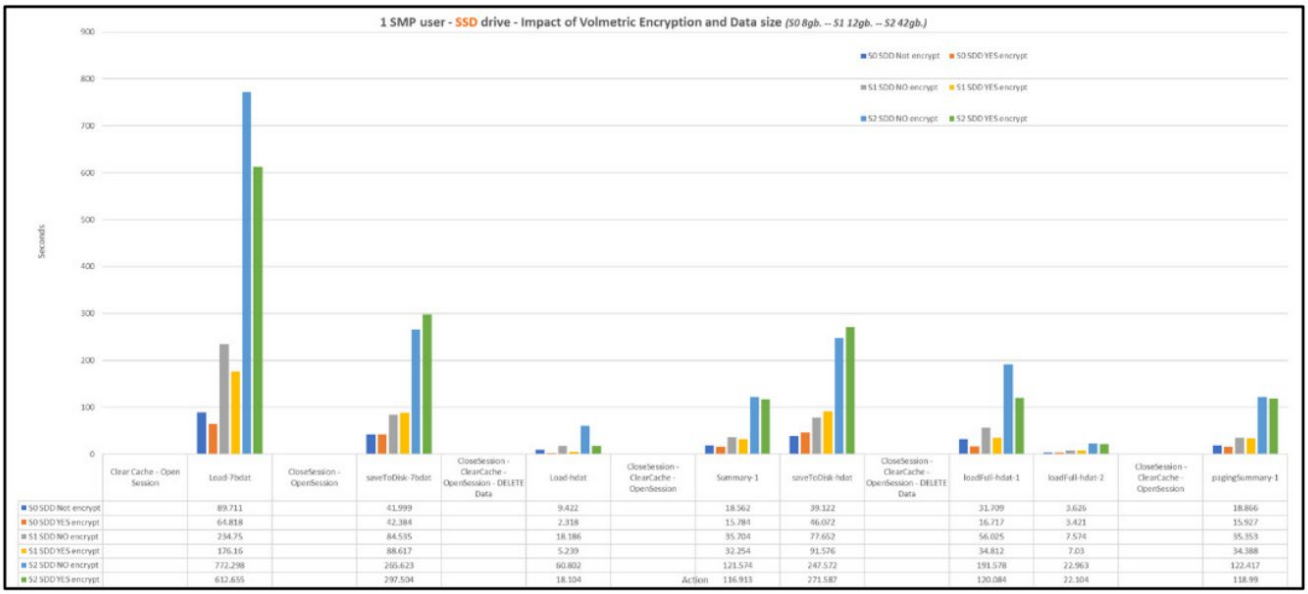
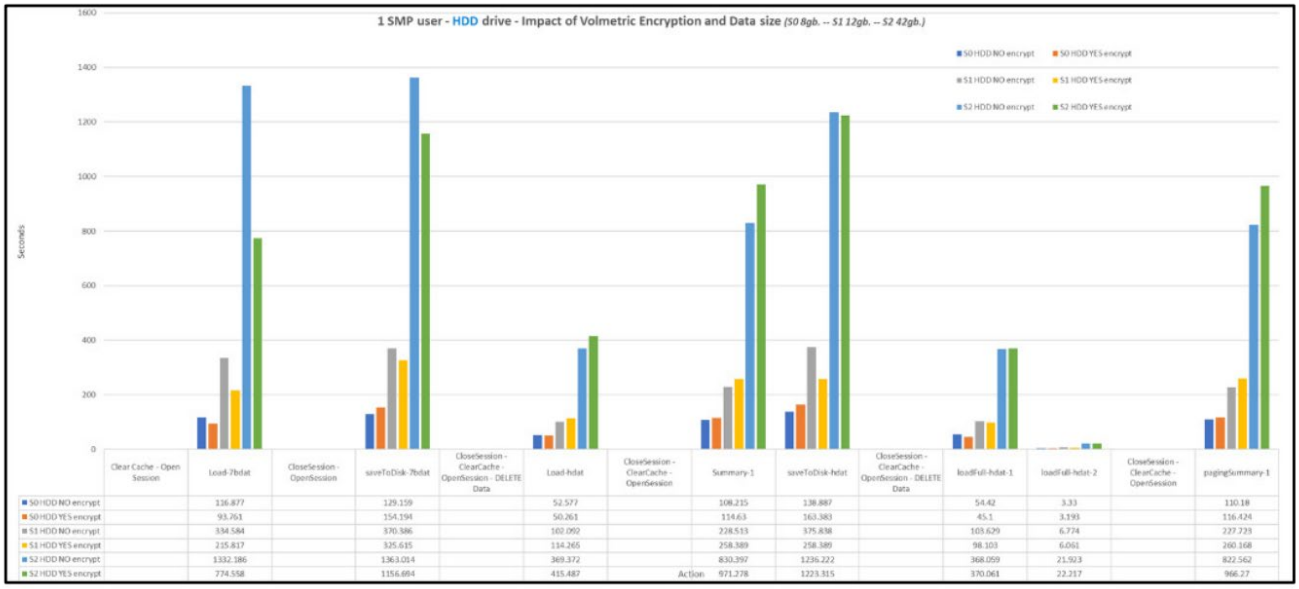
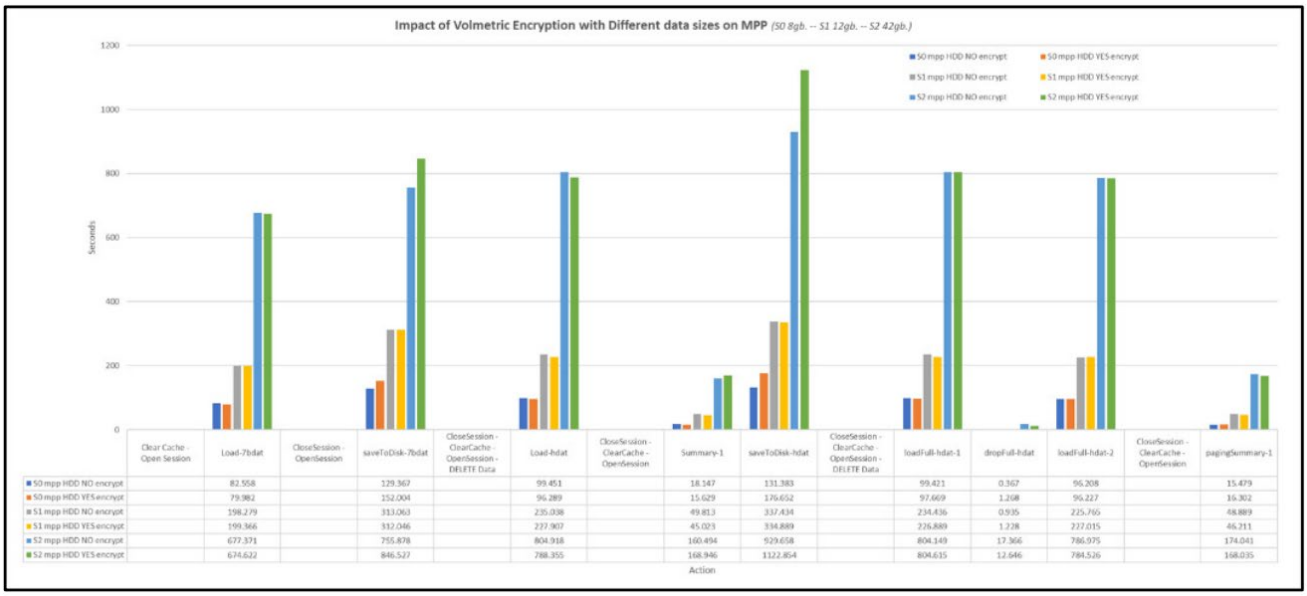


Figure 5. *Performance Impact of CAS-Encrypted Data in an SMP Deployment Using an HDD with Multiple Data Set Sizes*



Massively Parallel Processing (MPP) - Distributed CAS Deployment Across Multiple Machines

Figure 6. *Performance Impact of CAS-Encrypted Data in an MMP Deployment Using an HDD with Multiple Data Sizes*



Central Processing Unit

Central Processing Unit (CPU) performance was also measured during testing. As expected, there was a slight increase in cost of reading and writing encrypted data as the sizes of the data sets increased. Regardless, the overall cost was minimal from a performance perspective.

Figure 7. *Average CPU Usage Percentage When a Smaller Data Set Is Encrypted Versus Unencrypted*

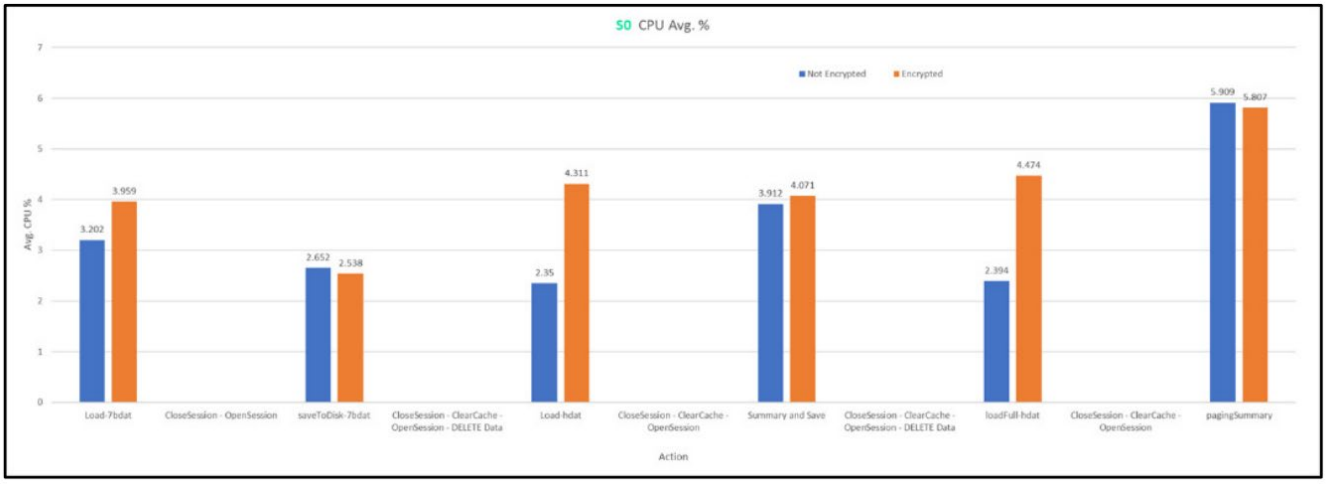
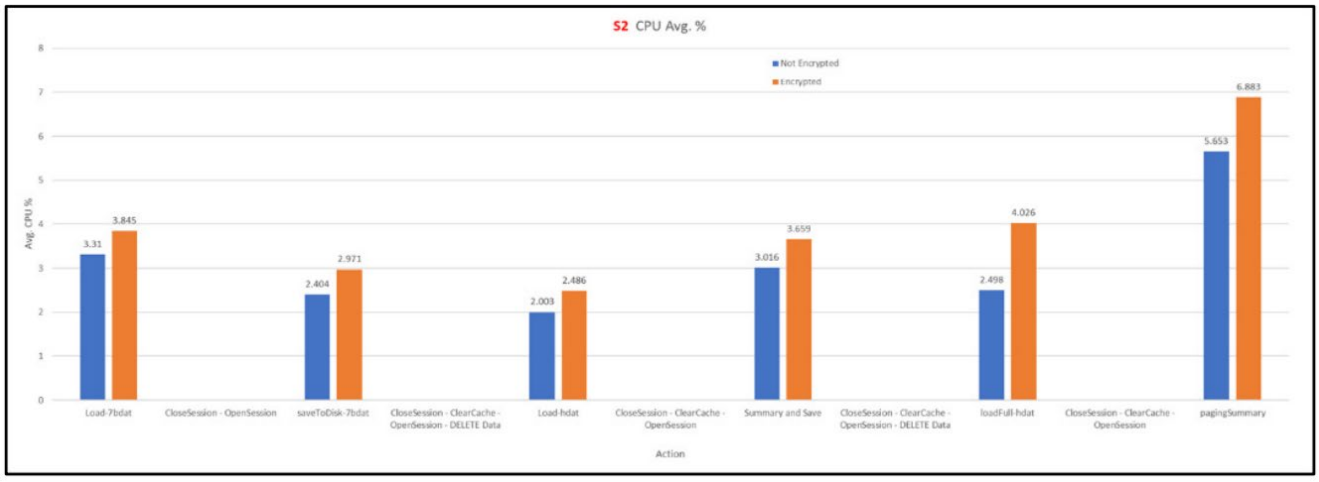


Figure 8. *Average CPU Usage Percentage When a Larger Data Set is Encrypted Versus Unencrypted*



PostgreSQL

Vormetric applied to the PostgreSQL database was tested by performing a series of operations, first with unencrypted data, and then with encrypted data. Performance test results showed a 1.5% performance impact when data was encrypted. The database was tested directly and indirectly, as it was also implicitly tested during the SAS Visual Analytics and SAS Visual Investigator testing.

SAS Visual Analytics

To measure the effect of encryption against performance, testing against SAS Visual Analytics was conducted three times:

1. SAS Visual Analytics was tested with unencrypted data.
2. Testing was conducted with encryption enabled for data (data-only encryption).
3. Testing was conducted for data and the `/opt/sas/viya/config` location (full encryption).

The following two figures compare the two sets of results measured against the first test as a control. After analysis, these results convey little to no impact on performance due to encryption.

Figure 9. *SAS Visual Analytics Test Results with No Encryption Versus Data-Only Encryption*

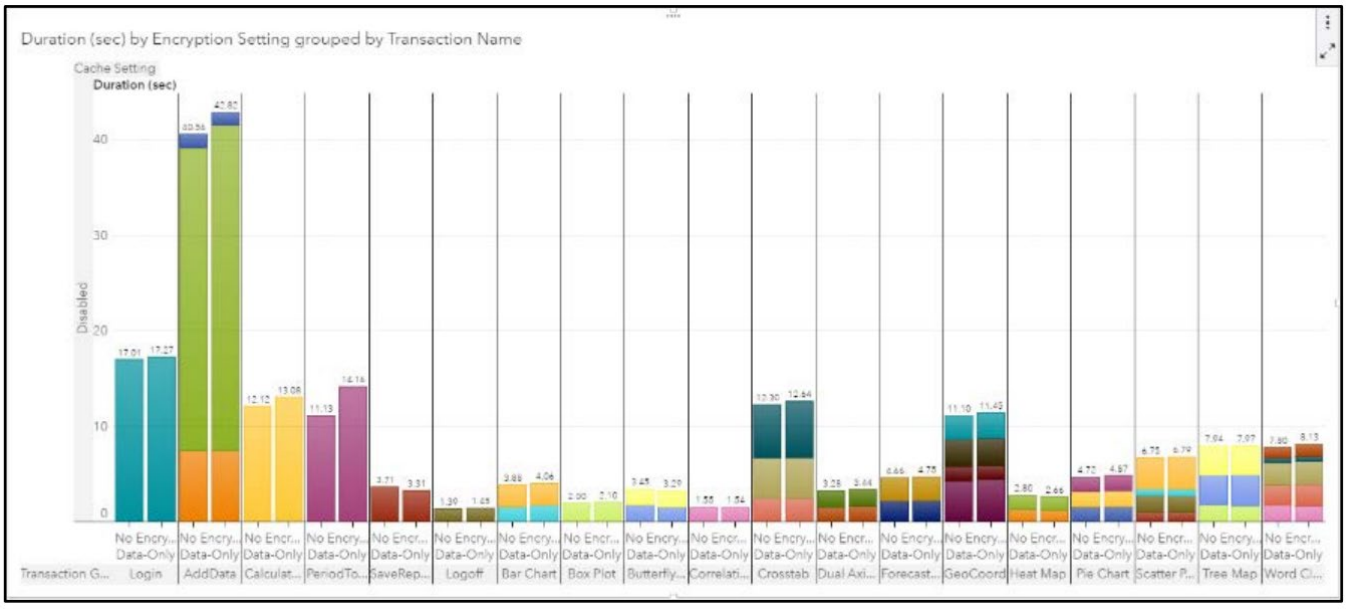
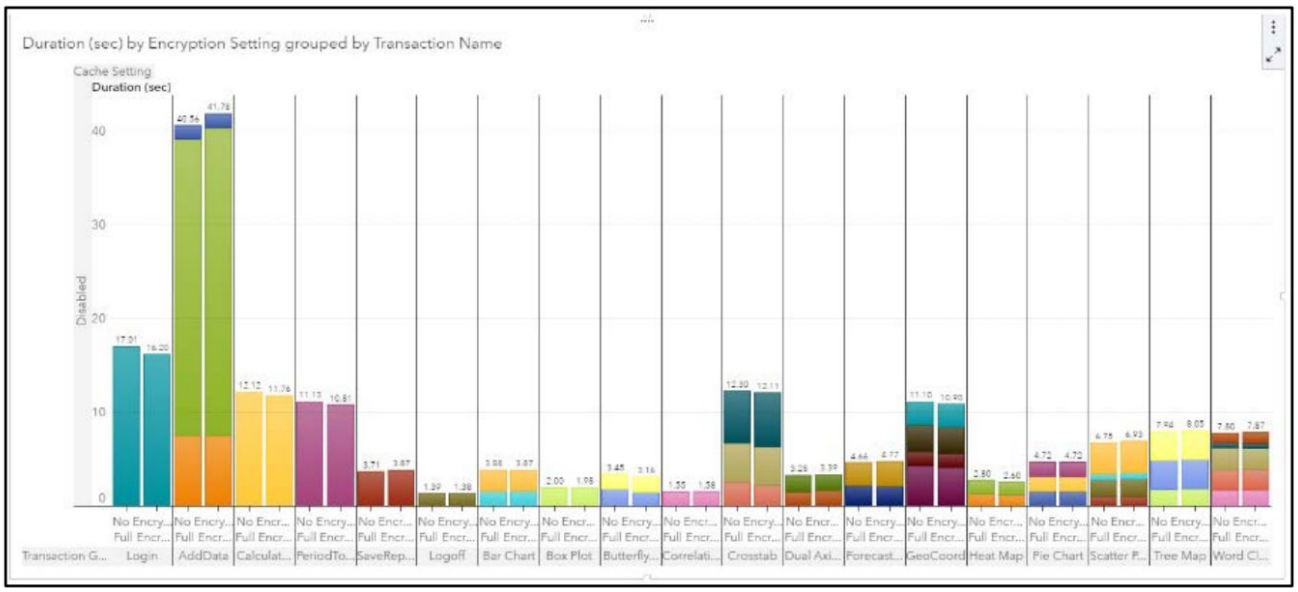


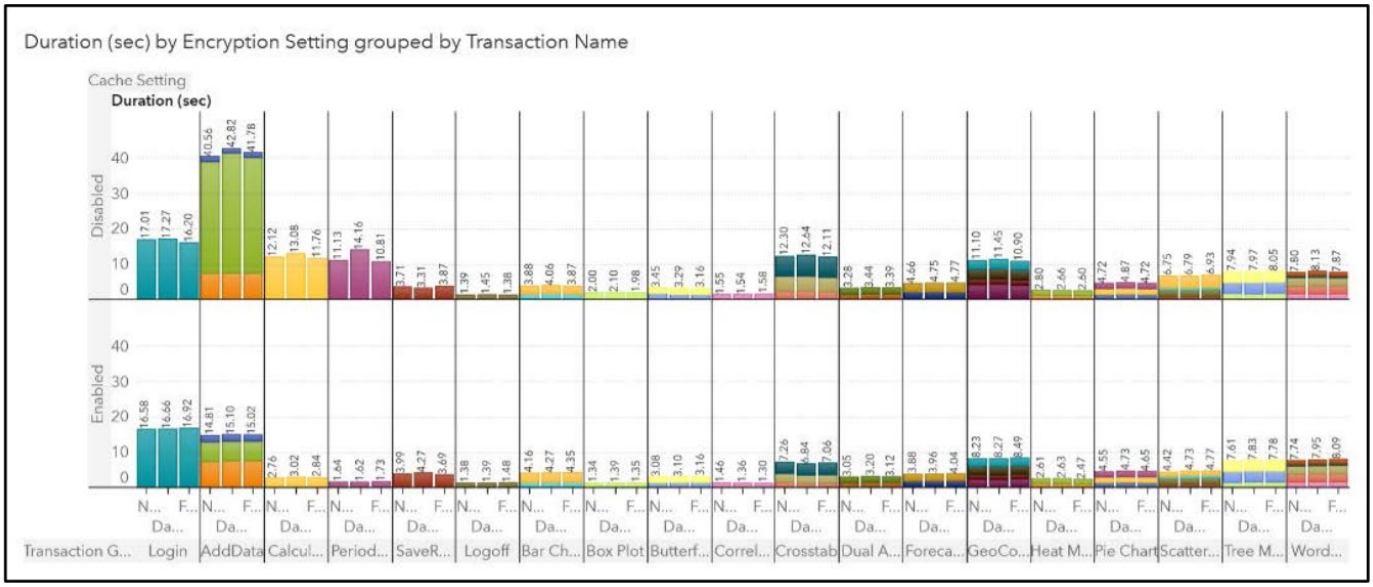
Figure 10. *SAS Visual Analytics Test Results with No Encryption Versus Full Encryption*



SAS Visual Analytics Results Cache

Because the SAS Visual Analytics Results Cache is integral to the operation of SAS Visual Analytics, it was also tested with and without encryption enabled. The following figure shows the results of a variety of tests performed with the Results Cache encrypted and unencrypted.

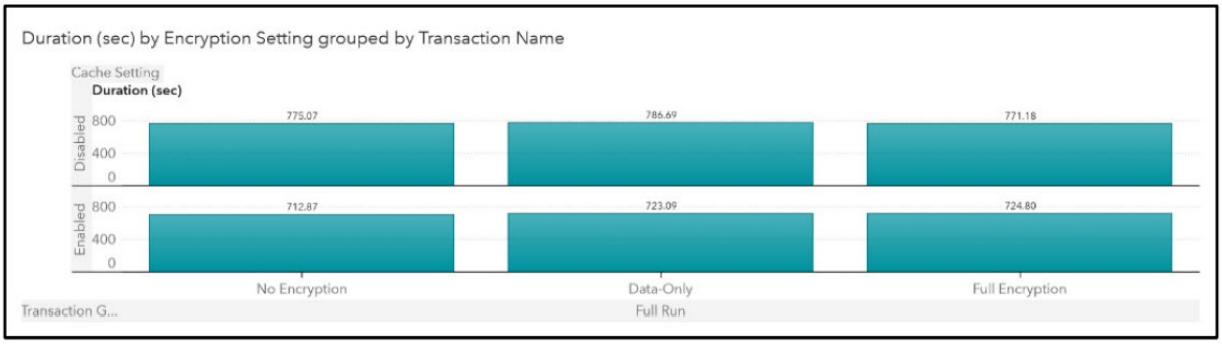
Figure 11. *Complete Performance Results for SAS Visual Analytics Result Cache Encrypted and Unencrypted*



When comparing these results, note that the top half of the table in Figure 11 shows the results using unencrypted data. The bottom half shows the results using encrypted data. Again, the results illustrate that minimal to no performance impact occurred due to encryption.

To simplify this comparison, the following figure provides an aggregated view of the preceding tests.

Figure 12. *Aggregated View of Performance Results for SAS Visual Analytics Result Cache Encrypted and Unencrypted*



SAS Visual Investigator

Similar to the test results obtained for Visual Analytics, the observed results for SAS Visual Investigator showed no significant degradation due to enabling Vormetric on the configuration paths. The increase in time was minimal. Most transactions used less than one second when unencrypted, and the average time increased by only a 12% fraction of a second when encrypted.

SAS Viya: Secure by Design

SAS Viya provides technical documentation on how to implement native encryption capabilities to ensure strong security across diverse deployment scenarios. For more information about security features in SAS Viya, please visit Introduction to Security Information and Encryption in SAS® Viya® 3.5: Data at Rest.

Learn More

To learn more about Vormetric Transparent Encryption, please visit Thales eSecurity.

Release Information Content Version: 1.0 December 2019

To contact your local SAS office, please visit: sas.com/offices