

Cloud Data Security

Selecting the Right Cloud Crypto and Key Management Solution

Table of Contents

Cloud Data Security	1
Introduction	1
Cloud Service Provider Crypto and Key Management Services.....	2
CSP Key Management Services.....	3
Cloud HSM.....	3
Envelope Encryption	3
Bring Your Own Key.....	5
Google Cloud DLP.....	5
Issues and Challenges with CSP Crypto Services	6
What Is “Bring Your Own Key” Really Buying You?.....	8
Recommendations for Cloud Crypto Service Selection	10
Strategic Selection Criteria.....	11
Voltage Format-Preserving Encryption	13
Why Voltage SecureData Enterprise?	14
About the Author	17

Cloud Data Security

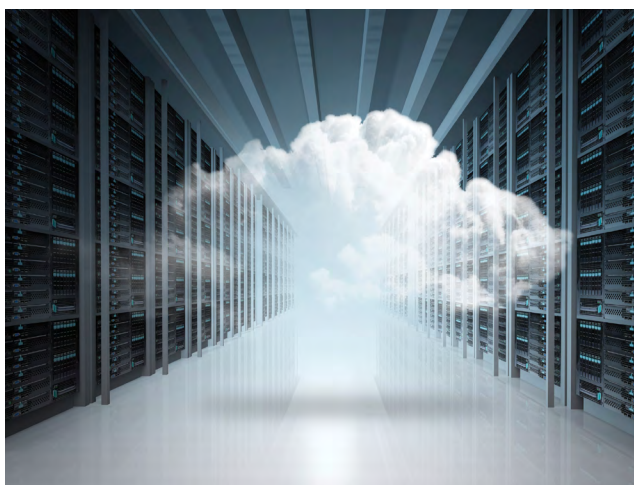
This position paper describes the various Cloud Service Provider (CSP) security offerings and then provides a framework for data protection, with a set of strategic selection criteria.

Introduction

Enterprises across industry segments are moving IT workloads and functions to the cloud, frequently ahead of any strategy or consistent capability to secure sensitive data. The advantages of cloud migration, such as scale, agility, and consumption-based pricing, are compelling and seem to outweigh the risks in the short term.

Most enterprise IT today is hybrid, with some workloads in the cloud and some hosted within the enterprise datacenter. Many are adopting a “cloud-first” or “cloud-only” approach for all new IT functions and business. Due to a combination of decentralized IT functions, frequent mergers and acquisitions, and shadow IT, most enterprises are multi-cloud, leveraging more than one cloud service provider (CSP).

Data security is rarely the first consideration for the selection of a CSP, but it is not the last one either as CSPs are also now ensuring that they address the data access controls and security needs of enterprises. The emergence of strict new data privacy regulations, such as GDPR and CCPA, is driving the need for CISOs to more effectively address data protection and data governance in complex and geographically-diverse hybrid IT ecosystems. The terms pseudonymization and anonymization are now common in the context of these privacy regulations when it comes to data protection and privacy. While pseudonymization of data still allows for some form of re-identification (even indirect and remote), anonymization of data cannot be re-identified. CISOs look to the CSPs for data security solutions to address these privacy requirements but struggle with the varied array of security models and services they offer.



Many enterprises are adopting a “cloud first” or “cloud only” approach for all new IT functions and business.

CSPs offer native key management, encryption, and Hardware Security Module (HSM) services. These security services have typically been added as a layer on top of their existing stacks; after-thoughts from a late recognition of their customers' increasing data security concerns, and are not feature-rich and often don't cater to both functional and non-functional needs of enterprises. As most enterprises are also multi-cloud, the challenges inherent in CSP security offerings include deficiencies in uniformity, homogeneity, coverage, customer control and ownership, functionality, scalability, performance, visibility, and more. On top of these, there are broader challenges with key management, and vendor lock-in.

In this position paper, we describe the various CSP security offerings and provide a framework for data protection with a set of strategic selection criteria. In relation to the “Big Three” CSPs—Amazon Web Services (AWS), Google Cloud Platform (GCP), and Microsoft Azure—we make an attempt to objectively reflect on what their data security services entail, based on discussions with CSPs and the published documentation made publicly available by the CSPs. It also outlines what enterprises should be aware of prior to consuming these services in the context of their belated yet increasing capabilities in the data-centric security space. It also needs to be noted that the information captured is a point-in-time assessment and is subject to change as the CSPs continue to enhance and expand their services. This position paper is intended towards any and all audience who deal with data security and cloud security, executives, hands-on IT and Security professionals, and anyone with a passion or interest in cybersecurity in general, across enterprises and service providers.

Disclaimer: This is a point in time assessment of these services and as cloud service providers continue to augment and enhance their service offerings, some of this information may be outdated at the time the paper is being reviewed.

Cloud Service Provider Crypto and Key Management Services

As enterprises transition from being just compliant to being secure, they must focus on data-centric security services that keep their sensitive data protected persistently—while at rest, in transit, and in use—rather than server-side or transparent encryption services across storage and databases, which offer very little actual security. The good news is that even the CSPs have realized the increasing need for data-centric security, and have started to offer new capabilities in this space.

CSPs offer two kinds of cloud cryptographic (crypto) services to enable the implementation of data-centric security:

- key broker and key management services (KMS), such as AWS KMS, GCP KMS, and Azure KeyVault; and
- cloud hardware security modules (HSMs), such as AWS CloudHSM.

This position paper also touches upon the Google Cloud DLP service, which is a composite service that detects sensitive data and applies policies on the detected data.

As most enterprises are also multi-cloud by default, the challenges inherent in CSP security offerings include lack of uniformity, homogeneity, coverage, customer control and ownership, functionality, scalability, performance, visibility, key management, and vendor lock-in.

CSP Key Management Services

Key broker and key management services typically expose an API for managing keys and secrets. The premise of key management or brokerage across all of the big three CSPs is the use of the Master Key and Working Key model. The Master Key, usually referred to as the Customer Master Key (CMK), never leaves the KMS application, and is not used to protect sensitive data in bulk. It is typically used to generate Working Keys and/or to encrypt Working Keys or other secrets, and thus serves as a Key Encryption Key (KEK). Working Keys are Data Encryption Keys (DEKs), and are used by applications to encrypt/decrypt actual sensitive data. AWS and GCP use symmetric (AES-256) CMKs, but Azure uses only asymmetric (RSA-2048, -3072, -4096) key pairs, storing the private keys in their KMS.

CMKs may either be software-managed or stored inside a FIPS 140-2-compliant HSM controlled by the CSP. There are different models of Master Key management in terms of customer control and visibility:

- **Customer-managed Master Key:** Customer can view key metadata and manage the key
- **CSP-managed Master Key:** Customer can view key metadata but cannot manage the key
- **CSP-owned Master Key:** Customer cannot view key metadata nor manage the key

Cloud HSM

Cloud HSM is a service through which keys are generated by, and stored within, FIPS 140-2-compliant HSMs that are hosted and managed by the CSP. This model allows higher throughput than the KMS-based model of encryption. These HSMs offer a subset of the PKCS#11 standard API specifications, which are exposed either directly or through the KMS interface to take advantage of the other cloud services integrations existing with the KMS.

An important caveat is that these CSP crypto services are available in specific physical locations, referred to as Regions. Even when these services are available, cross-region integrations and availability of keys across CSP regions are also not guaranteed. Some CSPs do not specify their level of FIPS 140-2 compliance.

Envelope Encryption

While these KMSs are used to generate, store, protect, and retrieve encryption keys, it is important to understand the mechanism of application-level data encryption implemented and supported at these CSPs. CSPs implement **envelope encryption**, which is the practice of encrypting plaintext data with a working key (a DEK), and then encrypting the DEK with a master key (the CMK). CSPs typically offer software development kits (SDKs) that are used by the application to perform envelope encryption.

The **encryption** process works like this:

- An application makes a request (using the SDK) to the KMS to generate a DEK, under a specific pre-provisioned CMK.

Note: *With certain CSPs, KMS requests go thru the internet by default.*

Customer Managed Keys (CMKs) may either be software-managed or stored inside a FIPS-compliant HSM controlled by the CSP. There are different models of Master Key management in terms of customer control and visibility.

- Across the 3 CSPs, the process for generation of the DEK varies. For AWS, the KMS uses the CMK to generate and encrypt a DEK when an application calls a specific method (*GenerateDataKey*) in KMS. Each key request results in the creation of a unique DEK that is created and protected under the same or multiple CMKs. However, in the cases of Azure and GCP, the DEK is generated locally but is protected by the KEK that is stored in their respective KMS services.

Note: One best practice that is common across these services is the use of unique DEKs for each data during a write operation, and hence negating the need for DEK rotation.

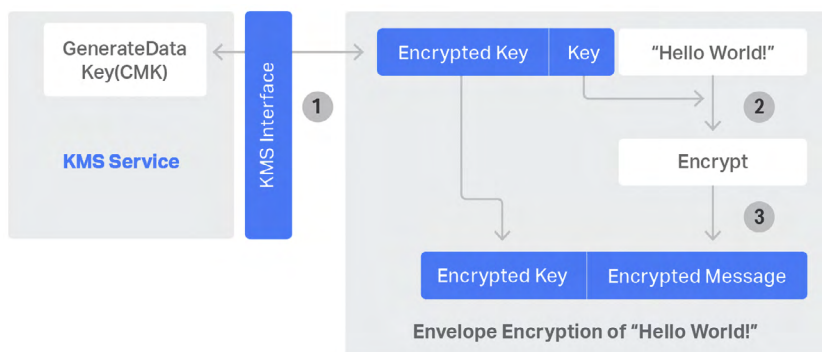


Figure 1. Illustrative Example—AWS Envelope Encryption

The mechanism of application level data encryption implemented and supported at CSPs is called envelope encryption and is the practice of encrypting cleartext data with a data key, and then encrypting the data key with another key.

- In case of AWS, the KMS returns both the plaintext and encrypted versions of the DEK to the application. However, there are options such as obtaining just the encrypted DEK by specifying a separate method (*GenerateDataKeyWithoutPlaintext*).
- The application uses the plaintext DEK to encrypt the sensitive data, and then typically deletes the plaintext DEK from memory.
- The encrypted DEK and the encrypted data are stored together.

Note: For CSPs that offer only asymmetric key pairs for DEK encryption, the application/SDK uses the public key of the pair to encrypt the DEK during storage.

The **decryption** process works like this:

- An application parses the envelope-encrypted message (typically via the SDK) to obtain the encrypted DEK and makes a request to the KMS to decrypt the DEK.
- The KMS uses the CMK/KEK to decrypt the DEK.

Note: For CSPs that offer only asymmetric key pairs for DEK encryption, the application/SDK sends the public key encrypted DEK to the KMS, where it is decrypted using the private key of the pair, and the cleartext DEK is returned to the application/SDK.

- The KMS returns the plaintext version of the DEK to the application.
- The application/SDK uses the plaintext DEK to decrypt the encrypted data.

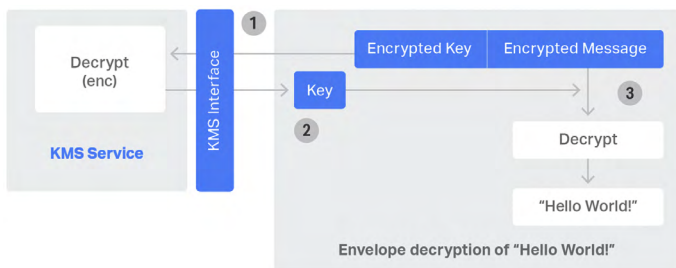


Figure 2. Illustrative Example—AWS Envelope Decryption

Bring Your Own Key

CSPs also allow customers to import their own key material. This “Bring Your Own Key” (BYOK) model lets customers generate the keys themselves (typically using on-premises HSMs) and upload them to the CSP KMS. Customers are usually required to download a certificate from the CSP, along with an import token. The symmetric keys generated by the customer are encrypted using the public key that is bound to the downloaded certificate. The encrypted symmetric key(s) plus the CSP token or a hash of the key material are then uploaded to the CSP KMS. The tokens/hashes are used for authentication and integrity purposes. In some BYOK implementations, the CSP requires padding and Base64-encoding encrypted key(s) prior to upload.

CSPs also allow customers to optionally import their own key material. This “Bring Your Own Key” (BYOK) model lets customers generate the keys themselves (typically using on-premises HSMs) and upload them to the CSP KMS.

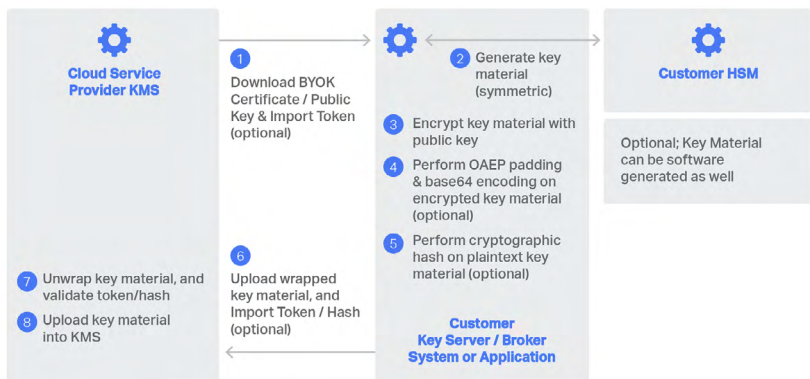


Figure 3. Illustrative Example—the “BYOK” model

Google Cloud DLP

Google Cloud Data Loss Prevention (DLP) provides APIs for sensitive data inspection, classification, and de-identification. It includes a number of built-in information type detectors, and allows definition of custom detectors. It offers de-identification techniques including redaction, masking, format preserving encryption, and date-shifting as optional actions to be taken on detected sensitive data within streams of data, structured text, files in storage repositories such as Google Cloud Storage and BigQuery, and even within images. The keys for data de-identification are either:

- generated by the application and embedded in the API request header in the clear;
- wrapped by a master key within the Google KMS; or
- generated by the Cloud DLP system once the API call is made.

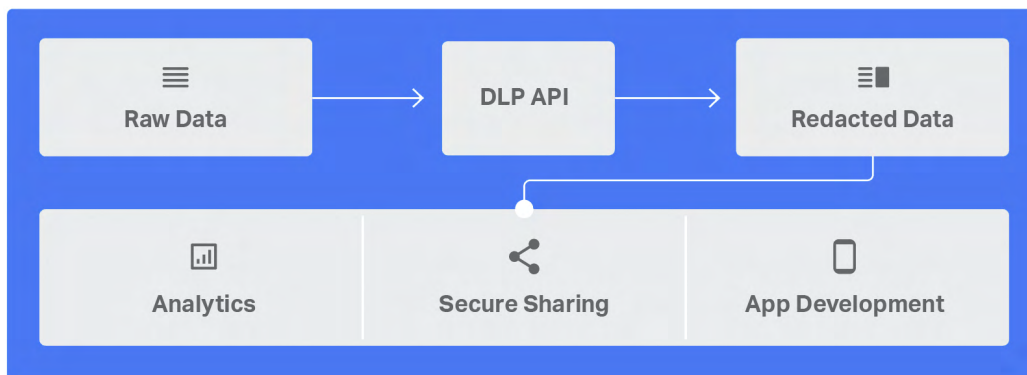


Figure 4. Illustrative Example—Google DLP

Google Cloud Data Loss Prevention (DLP) provides APIs for sensitive data inspection, classification, and de-identification. It includes a number of built-in information type detectors and allows definition of custom detectors.

Issues and Challenges with CSP Crypto Services

A number of issues and challenges around scale, availability, portability, performance, and security of these CSP crypto services should be considered.

- CSP lock-in**—Each CSP offers services that are available and functional within the confines of their cloud services. Enterprises typically use more than one CSP to implement their workloads, along with on-premises legacy implementations. The biggest challenge and concern for enterprises is that they cannot implement a single CSP-agnostic enterprise solution or service that can be applied across both on-premises and multi-cloud hosted workloads. Enterprises will face an immense challenge if they want to shift cloud workloads from one CSP to the other, often involving decryption and re-encryption of all of their data. Realizing that customers do not want to get locked into its KMS, Google has recently integrated with third parties like Fortanix (SDKMS) and Equinix (SmartKey) via its External Key Manager services. It refers to these capabilities as BYOKMS (Bring Your Own Key Management Service). *Such third-party KMS integrations are still coupled with, and hence rely on, the CSPs' systems and services, however.* There is no model that exists currently where the CSP's have worked together to figure out a cross-cloud and customer on-premises key re-use or key sharing mechanism, where data can seamlessly move between these entities while keeping data persistently protected and unprotected as and when needed for authorized users.
- Lack of total control over keys**—As previously described, CSPs either own or manage the master keys, which provide the root of trust used within the confines of their KMS or cloud HSMs even if the customer provides the key material. These keys and data encrypted by these keys are likely liable to subpoena, or other disclosure, or abuse. As discussed below, BYOK as a practice in the industry has created a false perception that customer ownership and control of keys is established when the fact is that, even if a customer generates and imports the keys into a CSP KMS or cloud HSM, it is the CSP that has direct or indirect control of the keys. There is no hard isolation when it comes to cloud infrastructure.

- **Lack of homogeneity**—With enterprises going global through mergers and acquisitions, they cannot afford to have regional services and silos. Even though CSPs have a global footprint, their crypto services are not always global. Not all regions may have the KMS and cloud HSM services available, which forces enterprises to deploy and migrate their applications and data to specific CSP regions. In addition, the CSPs' integrated crypto services are also fragmented, as crypto clients and SDKs built for one application may not be interoperable with other applications running within the CSP, even if they are using the same KMS under the same identity. The result is that customers encrypting data on one platform or in one CSP region cannot necessarily expect to be able to decrypt on another platform or in a different CSP region. Enterprises should conduct a thorough analysis and interoperability testing before embarking upon using these tools and going into design and development activities.
- **Lack of key management abstraction level**—Abstraction of key management is critical to ensuring developers do not spend too much time handling keys and needing to understand key management. Through these CSP crypto services, developers have access to the encryption keys and can use them as needed within their applications. CSPs offer developer SDKs to enable their application to handle crypto operations and key management, although a significant amount of metadata usage (such as key usage, key specification, encryption context, grant tokens) is required to create, request, encrypt, and decrypt keys. The lack of abstraction and giving developer access to physical keys, and providing them with the ability to manage those keys at the application code level, creates opportunities for exposure and breach.
- **Lack of encryption format choices and other data obfuscation options**—CSP crypto services SDKs all support 256-bit GCM-mode AES. As of early 2020, only Google Cloud DLP offers other encryption formats such as **Format-Preserving Encryption (FPE)** to enable business processes, retain application business rules and database schema, and allow secure analytics to be performed even after encryption of data. They also do not offer capabilities to **partially expose** certain data elements to enable business functions without the need to decrypt the data. CSPs also do not have a defined **tokenization** service offering, vault-based or vaultless, to meet various industry standards and regulations, such as PCI-DSS. *While Google Cloud DLP offers FPE, it does not clarify whether this is FF1 mode AES; FF3-mode AES, which is both vulnerable and has limitations; or a non-standard form of FPE not validated by NIST. It also makes no reference to supporting partial FPE or Unicode.*
- **Envelope encryption and Google Cloud DLP risks**—As explained above, envelope encryption involves the generation and protection of unique DEKs for each encryption operation, under the same or multiple Master Keys. Decryption involves decryption of the encrypted DEKs with the Master Keys prior to actual ciphertext decryption. Typical of distributed computing fallacies such as The Network is Reliable, The Network is Secure, and Latency is Zero, this CSP Encryption/DLP model introduces two key risks:
 - **Availability**—Each crypto or redaction operation at the application level depends on connectivity between the applications and KMS or Google Cloud DLP. Network glitches or service unavailability will introduce failures (protect/write, or decrypt/read).

Issues and challenges with CSP Crypto Services include a lack of control over keys, homogeneity worldwide, and key management abstraction.

- **Performance**—Every crypto or redaction operation requires a round trip to the KMS for DEK generation, encryption, and decryption. This introduces additional overhead and is subject to network latency, KMS or cloud HSM or Google Cloud DLP efficiency, and the overhead of multiple encryption operations (especially when asymmetric cryptography is used for DEK encryption/decryption). This does not scale for high-performance cloud workloads, bulk data processing, and delay-sensitive transactions.

Note: For CSPs that offer only asymmetric key pairs for DEK encryption, the application/SDK sends the public key encrypted DEK to the KMS, where it is decrypted using the private key of the pair, and the cleartext DEK is returned to the application/SDK.

Google Cloud DLP also has APIs that applications can invoke to detect sensitive data using pre-defined rules, and then redact that data. The same risks with availability and performance lie with this model.

What Is “Bring Your Own Key” Really Buying You?

When CSPs offer the “Bring Your Own Key” option, it creates a perception of increased security and control. But digging deeper into the BYOK model reveals that it is applied at varying tiers of the key hierarchy across CSPs, and that customers are not necessarily in control of the keys that actually protect data. In other words, BYOK is applied to Master Keys or Key Encryption Keys (KEKs), and customers almost never get to import Data Encryption Keys (DEKs) that are actually used for data encryption. Some CSPs have a key hierarchy where DEKs are at the lowest tier, with one or more KEKs layered above. One of the CSPs even offers BYOK only at the fourth tier of this hierarchy, meaning that the other three keys—the second-level KEK, first-level KEK, and the DEK—are owned by the CSP.

$DEK \leftarrow KEK1 \leftarrow KEK2 \dots \leftarrow KEKn$ (YOUR Key)

Regardless of the key hierarchy tier where BYOK is offered, CSPs still have control over all keys below: the perception created by CSPs and sold to enterprises is that the option to bring in their encryption keys provides control, when in fact it does not. BYOK can address concerns around key types and strengths (AES 256 vs. AES-128 or 3DES), key generation sources (HSM vs. software), and can be handy for enterprise audit processes. But when the keys are uploaded into the CSP KMS, there should be no perception that customers have total control of and exclusive access to those keys.

Issues and challenges with CSP Crypto Services include a lack of encryption format choices and risks associated with envelope encryption and Google Cloud DLP, such as service availability failures and failure to scale for high-performance cloud workloads, bulk data processing, and delay-sensitive transactions.

From a key rotation perspective, BYOK addresses enterprise policy requirements for Key Rotation. Since enterprises cannot rotate CSP-owned DEKs and KEKs, BYOK-KEKs are allowed to be rotated. This doesn't create any impact to the data that was encrypted by the DEKs down the key hierarchy, and this also satisfies audit and compliance requirements. For SaaS offerings such as Office 365, where Microsoft leverages Azure Key Vault based keys, the BYOK option is applied to the topmost key in the vault key hierarchy, and that is what gets rotated.

DEK ← KEK1 ← KEK2.... ← KEKm (YOUR Key, post rotation)

For **custom-developed applications** deployed across the CSPs Infrastructure as a Service (IaaS) and Platform as a Service (PaaS) offerings, these CSPs provide some common best practices around usage of DEKs. They encourage use of unique keys for each data element, whether it's generated by the KMS or Cloud HSM—which is the case with AWS, where every application request (GenerateDataKey) for DEKs will generate a new/unique key; or if it's locally generated by the application (in case of Azure and GCP). In either of these cases, it is also advised to encrypt the DEK with a KEK (stored securely in the KMS or Cloud HSM), and the encrypted DEK is stored with the encrypted data during a write operation—**Envelope Encryption** model explained before. The premise of this best practice is that in this model, there is no need for DEKs to be rotated.

However, generating unique keys for each data element, even for a specific data type, could result in a huge number of keys, and generating a key each time to perform a data encryption also adds additional overhead to the process, which impacts performance. Having developers generate and manage DEKs (itself a risk) at the application layer frequently means they use the same or shared keys across data types and environments. In such scenarios, the DEKs tend to persist, and thus are subject to compromise, necessitating rotation. With typical implementation of symmetric cryptographic services, a DEK rotation every n years could result in a burdensome requirement to re-encrypt all data. A periodic rotation at the KEK level avoids this situation and is easy to execute. However, this approach offers no assistance for enterprises with compromised DEKs.

What is "Bring Your Own Key" really buying you? The "BYOK" model reveals that it is applied at varying tiers of the key hierarchy across CSPs, and that customers are not necessarily in control of the keys that actually protect data.

Recommendations for Cloud Crypto Service Selection

Enterprises need to develop a strategy for cloud data security early in their cloud migration journey. They should implement data-centric security, preferably prior to the sensitive data being migrated to the cloud. CSPs do a good job of infrastructure hardening and in implementing security processes and policies, and they offer multiple tools to enable customers to secure their workloads. But it is very important to understand that CSPs are not responsible or liable for the security of the data that customers ingest into their services. As per the Shared Responsibility Model, Security and Compliance is a shared responsibility between the CSP and the customer. The separation of duties in this model is commonly referred to as Security “of” the Cloud versus Security “in” the Cloud.

CSPs are not responsible or liable for the security of the data that customers ingest into their services.

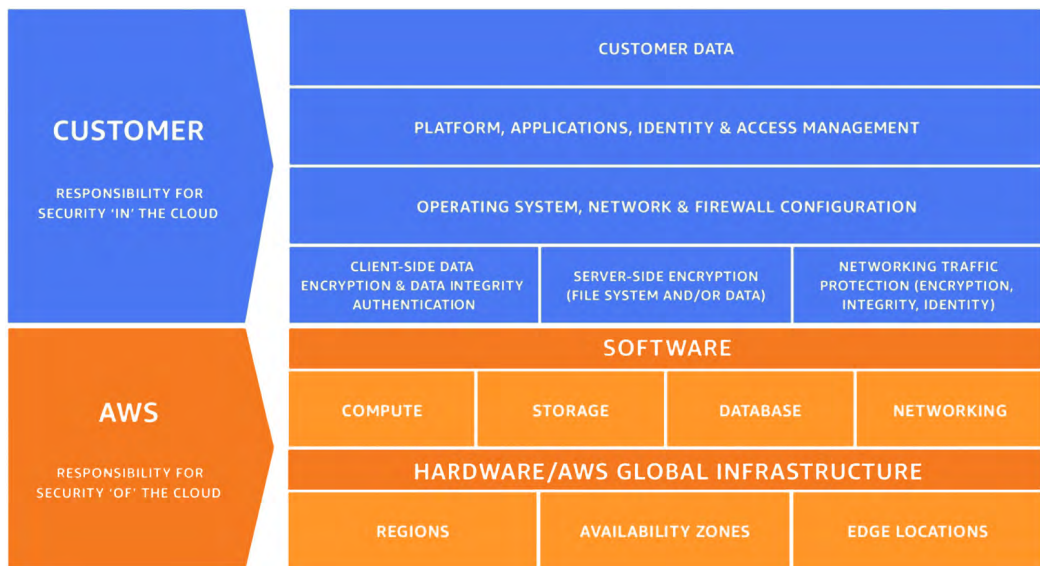


Figure 5. AWS Shared Responsibility Model

While the CSPs are responsible for the security “of” the cloud, where they are responsible for protecting the infrastructure that runs all of the services offered in the cloud. This infrastructure is composed of the hardware, software, networking, and facilities that run the cloud services. The customer is responsible for the security “in” the cloud, and hence has the responsibility to ensure the security of their data, while taking advantage of the scale, flexibility, features, and global reach of the cloud. A lack of this understanding has contributed to so many data breaches as witnessed in the recent past. Hence, an enterprise data security strategy is

crucial to ensure data is secure across all hosting locations (on-premises and in the cloud) and during migration, while at rest, in transit, or in use. Enterprises are wise to avoid locking themselves into these native CSP crypto services and to consider a CSP-agnostic, feature-rich, scalable, high performance, global solution.

Strategic Selection Criteria

There are eight criteria for the selection of enterprise-grade hybrid and multi-cloud crypto services:

1. **Avoid vendor lock-in and ensure portability and extensibility**—Choose a solution that can seamlessly apply to workloads hosted anywhere: on-premises or cloud, single or multiple CSPs, midrange or mainframe, Hadoop or cloud services. Migration of workloads from one platform to the other, or migration from on-premises infrastructure to cloud, or switching across CSPs, should not require changing crypto services and key management controls.
2. **Ensure adequate control of keys across all platforms**—Whether HSMs are on-premises or hosted by a provider as a service, ensure full control and visibility of the keys generated and stored in these HSMs, as well as where they are being used, and which applications and users have access to them.
3. **Drive toward state-less crypto services and key management**—Traditional and typical crypto services and key management infrastructures rely on key storage, token vaults, etc., and create challenges with operationalizing distributed architecture due to their state-ful nature. They rely on eventual consistency and replication across the network, and this problem aggravates in a global deployment scenario. Hence, Crypto and key management services should be stateless, enable distributed deployment without requiring any sort of replication, synchronization, and eventual consistency. This allows for these distributed services to be accessible to any workloads hosted at any location. Data can be protected at one location, stay persistently protected while being moved to another location, and unprotected from the other location, if needed. This is what we refer to as the “Protect Here, Access There” model.
4. **Strive for high availability and resiliency**—Crypto services must be highly available and resilient, as enterprise business-critical functions depend on them. Ensure that the solution does not fall for the fallacies of distributed computing, where a network glitch can cause an operation to fail or time out, causing severe impacts to the organization in terms of inconsistent data, customer experience, and service outage. The solution needs to be resilient to prevent and minimize any network or infrastructure outage.
5. **Strive for performance while not compromising security**—The solution must be high performance, minimizing crypto operation overhead as well as avoiding dependencies on network availability and latency. The system must be able to handle bulk operations as well as high-throughput, delay-sensitive transactions. Needless to say, the solution must not achieve high performance by taking design shortcuts that can lead to key exposure or other security vulnerabilities.

An enterprise data security strategy is crucial to ensure data is secure across all hosting locations (on-premises and CSPs) and during migration, while at rest, in transit, or in use.

- 6. Ensure adequate developer abstraction**—Application developers like simplicity when consuming crypto services: methods/functions to invoke the services that abstract the underlying cryptography and key management. They do not want to be required to build crypto and key management expertise.
- Ensuring that they never have direct access to encryption keys helps avoid mistakes that can expose keys. They want to do the right thing, so choose a solution that makes it easier for them to do so.
- 7. Invest in feature-rich crypto solutions**—Choose a solution that offers a variety of data protection formats that not only allow pseudonymization and anonymization of sensitive data, but also enable business processes, analytics workloads, etc., to operate on the data in its protected state.
- Voltage Format-Preserving Encryption by OpenText™** is a powerful data protection technology, and is currently becoming the de facto standard across the industry. Voltage FPE warrants a deeper examination, and the following section expands on Voltage FPE and its importance.
 - For PCI-DSS compliance, **tokenization** is the preferred method for credit card protection, so a data security solution or service must be able to tokenize credit card numbers. As explained earlier, a stateless tokenization solution is preferred to not only avoid consequences of token mapping table synchronization issue in a distributed deployment model, but also to enhance performance of the solution.
 - Privacy regulations such as GDPR and CCPA are driving data subjects' rights for erasure of personal data ("right to be forgotten"), and the anonymization of data is an effective way for enterprises to comply. Anonymized values cannot be converted back to the original data, and **Voltage Format-Preserving Hashing** by OpenText™ is a great way to achieve this.
- Note:** *Voltage SecureData Enterprise by OpenText™ is the only solution that offers Format Preserving Hash.*
- 8. Consider CSP Crypto Services and BYOK for appropriate use cases only**—When enterprises consume software as a service (SaaS) to store or process sensitive data, they face unique challenges. SaaS providers offer limited or no scope for custom development, so enterprises are typically restricted to very few options for data security. One option would be to leverage an encryption gateway solution, which acts as a web proxy to intercept and protect sensitive data prior to storing it in the SaaS cloud. The other option is to leverage the SaaS provider's encryption capabilities, if available. Some of these SaaS providers also offer BYOK. In those cases, adopting the BYOK model makes sense. Most SaaS providers implement platform-level encryption (database, storage) by default, and typically also offer field-level encryption services at additional cost. These allow selection of fields to protect, encryption algorithm, and key strength to be used, as well as whether to use provider generated keys or BYOK. Opting for such SaaS providers' data-centric crypto services certainly provides added security. One way to exercise some control in the key management space is by adding BYOK to the mix, which helps in meeting some compliance and audit requirements, even though the challenges discussed above continue to apply. In a nutshell, if you can't Bring Your Own Encryption (BYOE) to the cloud, at least Bring Your Own Key (BYOK).

For selection of enterprise-grade hybrid and multi-cloud crypto services, "follow the eight strategic selection criteria mentioned in the document"

Consider CSP Crypto Services and BYOK for appropriate use cases only. For example, when enterprises consume software in the cloud as a service (SaaS) and store or process sensitive data, leveraging the SaaS provider's crypto and key management services and any BYOK offering makes sense.

If you can't Bring Your Own Encryption (BYOE) to the cloud, at least Bring Your Own Key (BYOK).

Voltage Format-Preserving Encryption

Voltage Format-Preserving Encryption (FPE) refers to encrypting data in such a way that the output (the ciphertext) is in the same format as the input (the plaintext). “Format-preserving” implies that encrypting a 16-digit credit card number produces a ciphertext which is another 16-digit number; encrypting an English word produces a ciphertext comprising the same number of English characters; and so forth. These properties have several benefits and simplify data protection, especially for legacy applications, where it avoids major redesign and refactoring of applications and business processes:

- a. **Minimal or zero database schema impact**—Voltage FPE facilitates retrofitting encryption technology to existing devices or software where conventional encryption modes would not be feasible. In particular, database applications may not support changes to data length or format.
- b. **Minimal or zero data storage impact**—Since length preservation is an inherent property of Voltage FPE, enterprises do not have to worry about additional storage usage, unlike conventional (non format-preserving) encryption methods, which typically expand data.

Note: *Some exceptions do apply where the length of the output with some variants of Voltage FPE can be slightly longer than that of the input data.*
- c. **Analytics on protected data**—Format-preserved protected data elements such as credit card numbers, SSNs, etc., can still be used as index keys to facilitate statistical research, even across databases. With Voltage FPE, the same inputs to the algorithm will create the same ciphertext. This deterministic encryption preserves the referential integrity of the data and thereby the ability to glean valuable information from the protected dataset. Other crucial benefits of Secure Analytics enabled by the use of Voltage FPE is expanding the access to data across a broader set of analysts, and potential monetization of data sets, without compromising on security and privacy.
- d. **Cross-application dataflow preservation**—Voltage FPE lets protected data flow across applications without requiring changes to those applications to accept the protected data, an infeasible approach with conventional encryption methods, since applications require data of specific lengths and formats.
- e. **Using protected data without requiring decryption**—Voltage FPE can allow protection of only specified key portions of data elements, enabling use of the data in its protected state. For example, the “first six” digits of credit card numbers are used for charge routing, and the “last four” of SSNs is used for customer verification. If these are left in the clear, many applications in the data flow will not need access to the entire data element, and can perform required business functions without requiring any change to the applications, and not requiring to perform any decryption. Such partial encryption can facilitate functions such as sort and certain search use cases, such as “Starts with”, “Ends with”, etc., without requiring any decryption of the encrypted data.
- f. **Test data management**—Voltage FPE can also be used to obfuscate/scrub production data to populate test databases, enabling realistic test conditions based on production volume, variability, etc.

Voltage Format-Preserving Encryption (FPE) refers to encrypting data in such a way that the output (the ciphertext) is in the same format as the input (the cleartext). These properties have several benefits and simplify data protection. Unique values of Voltage FPE include enabling regulatory compliance, audit scope reduction, and secure analytics performed on protected data.

NIST Special Publication 800-38G, *Recommendation for Block Cipher Modes of Operation: Methods for Voltage Format-Preserving Encryption*, specifies two AES modes, FF1 and FF3, for format-preserving encryption. However, NIST has concluded that FF3 is no longer suitable as a general-purpose Voltage FPE method based on findings of cryptanalytic attacks on the FF3 algorithm. Few vendors have implemented Voltage FPE within their data security solutions other than Voltage SecureData Enterprise Security by OpenText™, who hold the patent for FF1-mode AES. Customers need to be cautious about other vendor implementations of Voltage FPE, ensuring that they are not using the vulnerable FF3 mode or a non-standard form of Voltage FPE not validated by NIST, or that, if using FF1, their vendor has obtained a license from Voltage for FF1 mode.

Voltage SecureData Enterprise by OpenText™ is the only enterprise-grade solution that fits the criteria for implementing crypto services for hybrid and multi-cloud implementations.

Voltage SecureData Enterprise is the only enterprise-grade solution that fits the criteria for implementing crypto services for hybrid and multi-cloud implementations.

Data element	Example data in the clear	Example protected data
Name	Härold Potter	5iW9VtS4ølwpQ
Tax ID	532-09-1847	821-90-7385
Credit Card Number	4210-9735-8310-4461	9328-0218-7219-4461
Date of Birth	09/04/1979	05/01/1998
GPS location	37° 46' 26.2992" N 122° 25' 52.6692" W	91° 52' 05.7217" N 731° 60' 21.6540" W

...

Figure 6. Voltage Format Preserving Encryption: Input-Output examples



Why Voltage SecureData Enterprise?


Voltage SecureData Enterprise is an industry leader in the data security space, where hundreds of enterprises rely on it to secure sensitive data at the application layer and establish the trust of their customers. Voltage SecureData Enterprise has been implemented across a wide variety of customer on-premises infrastructure and cloud hosting locations, providing ubiquitous, stateless, scalable, highly performant and highly resilient data security. Voltage SecureData Enterprise supports hybrid (mixed on-premises and cloud) and multi-cloud implementations, as well as multiple enterprise platforms such as midrange, IBM Z, Hadoop, and HPE NonStop. Voltage SecureData Enterprise is available both in a virtual appliance model that can be deployed using an IaaS model in public and private clouds/ on-premises, as well as in a containerized, micro-services model that can be deployed within Kubernetes clusters in a PaaS or CaaS (Container as a Service) model. When it comes to implementing cloud data security, Voltage SecureData Enterprise is the optimal choice, since it best addresses the issues and challenges of CSP native crypto services.

Voltage SecureData Enterprise offers a stateless, identity-based key management solution, even supporting few industry standard and popular FIPS 140-2 compliant, general-purpose HSM products for key derivation. It does not store keys, and hence is less susceptible to attacks. It offers client libraries for local crypto operations, as well as a REST interface for remote operations. It supports a range of data protection methods:

- Voltage Format-Preserving Encryption
- Voltage Embedded Format-Preserving Encryption (eFPE)
- Voltage Secure Stateless Tokenization (SST) by OpenText™
- Obviously protected output preventing false positives during scans, DLP and audits.
- Voltage Format-Preserving Hash (FPH) by OpenText™ for data anonymization and enabling searches for data encrypted using probabilistic encryption modes, or searches on old encrypted data post key rotation
- Standard AES encryption with 256-bit keys, for both structured and unstructured data.

Voltage SecureData Enterprise supports a flexible range of crypto methods offering both two-way and one-way transformations.

	First Name: Gunther Last Name: Robertson SSN: 934-72-2356 DOB: 08-07-1966		First Name: Jürgen Last Name: Klinsmann Checking Acct: 122105278 674301068
Voltage FPE	First Name: Uywjloq Last Name: Muwruwwbp SSN: 253-67-2356 DOB: 01-02-1972	First Name: KxyAçy Last Name: ĎwlämÜqßr Checking Acct #: 122105278 827572346	

	Credit Card
	4171 5678 8765 4321
SST	8736 5533 4678 9453
Partial SST	4171 5633 4678 4321
Obvious SST	4171 56AZ UYTZ 4321

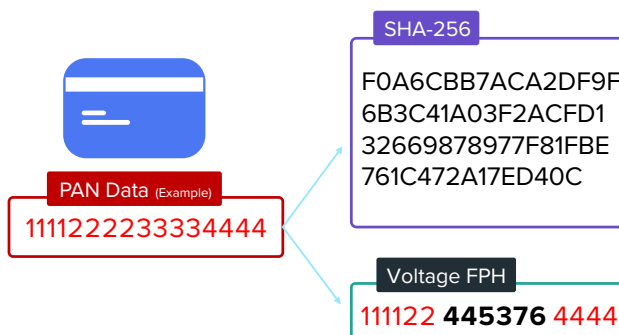


Figure 7. Voltage Crypto-methods: Input-Output examples

As illustrated, it also supports partial encryption to expose certain portions of data elements to enable business functions to operate on persistently encrypted data. It provides the right level of abstraction for developers, making it easy to add encryption without requiring extensive crypto knowledge.

Voltage SecureData Enterprise also introduces a variant of FPE called Embedded Format-Preserving Encryption (eFPE), which embeds key metadata within the ciphertext as part of the crypto operation. This provides significant benefits:

- **Zero Key Rotation Impact**—Even though Voltage SecureData Enterprise offers seamless key rotation at the CMK or KEK level for various cloud and on-premises deployment models, this feature even allows seamless rotation of keys at the DEK level. When a key is rotated, the data encrypted with the previous key need not be re-encrypted, as the solution detects the embedded key metadata from the ciphertext and determines which key to use for decryption.
- **Atomic Key Identification**—If a data element is re-encrypted after a key rotation, the new key metadata is automatically updated as part of the encryption operation.
Note: *If the metadata is stored elsewhere, a separate operation is required to update that metadata, which introduces the risk that this second operation might fail, or be omitted.*
- **Obviously Protected Output**—The embedded key metadata introduces additional characters into the ciphertext (beyond the input alphabet), and hence an encrypted credit card or Social Security number can be differentiated from actual values, preventing false positives during scans, DLP and audits.

Voltage SecureData Enterprise and Unicode—Normalization is normal in Unicode processing, and is a critical issue for format-preserving data protection, because if a protected value contains any normalizable characters, any normalization after encryption will destroy the ciphertext. Other solutions offering format-preserving data protection for Unicode generally try to avoid the issue by simply saying “Never normalize ciphertext”. Since one of the benefits of format preservation is the ability to pass data through other applications in its protected state, and normalization may occur at many points in the life of a data element, this is not a good solution. Voltage SecureData Enterprise is also the only solution in the industry to solve the **Unicode normalization** problem by providing Safe Unicode FPE. This support adds a new built-in format named PREDEFINED::UNICODE_BASE32K, or “Base32K” for short. Safe Unicode FPE provides a robust and elegant solution to the normalization problem, allowing Format-Preserving Encryption of Unicode data without risk.

Voltage SecureData Enterprise is the only solution in the industry to solve the Unicode normalization problem by providing Safe Unicode FPE. Safe Unicode FPE enables protection of structured data in any language, any region.

Voltage SecureData Enterprise is an industry leader in the data security space, where hundreds of enterprises rely on it when it comes to securing sensitive data at the application layer and establishing the trust of their customers.

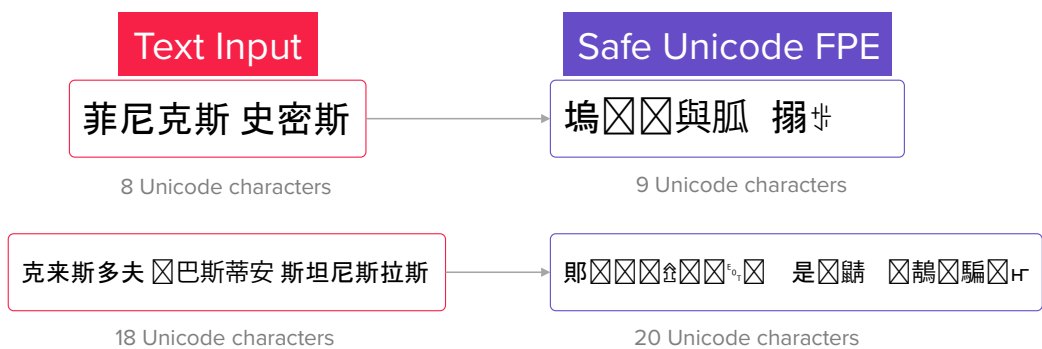


Figure 8. Voltage Safe-Unicode FPE

About the Author



Sid Dutta is an industry expert and leader in the cybersecurity space, specializing in cloud security and data privacy & protection. Sid comes from the enterprise side and has led cybersecurity risk remediation programs and services globally at various large corporations in the financial services sector for number of years. He was the Vice President & Global Head of Data Protection & Applied Cryptography at one of the largest global payment processors, responsible for enterprise PKI, Key Management, Encryption, Tokenization, and CASB. Prior to that, he also held the position of Director of Information Security at one of the largest and most respected global credit card brands, responsible for end-to-end delivery and operations of data privacy & protection globally. He has held several other leadership positions, driving enterprise CI-CD architecture & Tool Chain Automation, cloud security, application infrastructure uplifts, IT strategy.

He started his career in the contact center technologies space and spent several years implementing global network IVRs, CTI, Call Routing and Network Advanced Features, Call Recording, SIP Trunking, etc. Sid is currently employed with OpenText™, and heads product management for Data Security solutions, driving product strategies and roadmap, especially focusing on cloud enablement, integrations to cloud services, analytics, and Data Security-as-a-Service.

Sid is a Transformational, Visionary and Results Oriented leader, Industry Speaker, Cyber Security Expert with 20+ years of experience across Strategy, Architecture, Engineering, Quality, Program Management, Delivery and Operations, Customer Experience, Agile and DevSecOps, and Product Management.

Connect with Us

www.opentext.com



opentext™ | Cybersecurity

OpenText Cybersecurity provides comprehensive security solutions for companies and partners of all sizes. From prevention, detection and response to recovery, investigation and compliance, our unified end-to-end platform helps customers build cyber resilience via a holistic security portfolio. Powered by actionable insights from our real-time and contextual threat intelligence, OpenText Cybersecurity customers benefit from high efficacy products, a compliant experience and simplified security to help manage business risk.