AWS Whitepaper

Data Classification

Data Classification: AWS Whitepaper

Table of Contents

Data Classification

Publication date: **August 3, 2022** (*Document revisions*)

Abstract

This paper provides insight into data classification categories for public and private organizations to consider when moving data to the cloud. It outlines an example of a process through which customers can build data classification programs, shares examples of data and the corresponding category it may fall into, and outlines practices and classification models currently implemented by early adopters and organizations migrated to the cloud, along with data classification and privacy considerations. It also examines how implementation of a data classification program can help simplify cloud adoption and management, and recommends that customers use internationally recognized standards and frameworks when developing their own data classification rules.

Are you Well-Architected?

The AWS Well-Architected Framework helps you understand the pros and cons of the decisions you make when building systems in the cloud. The six pillars of the Framework allow you to learn architectural best practices for designing and operating reliable, secure, efficient, cost-effective, and sustainable systems. Using the AWS Well-Architected Tool, available at no charge in the AWS Management Console, you can review your workloads against these best practices by answering a set of questions for each pillar.

For more expert guidance and best practices for your cloud architecture—reference architecture deployments, diagrams, and whitepapers—refer to the AWS Architecture Center.

Data classification overview

Data classification is a foundational step in cybersecurity risk management. It involves identifying the types of data that are being processed and stored in an information system owned or operated by an organization. It also involves making a determination on the sensitivity of the data and the likely impact should the data face compromise, loss, or misuse.

To ensure effective risk management, organizations should consider classifying data by working backward from the contextual use of the data, and creating a categorization scheme that takes into account whether a given use case results in significant impact to an organization's operations (for example, if data is confidential, it needs to have integrity, and/or be available).

As used in this document, the term "classification" implies a holistic approach inclusive of taxonomy, schemes, and categorization of data for confidentiality, integrity, and availability.

Data classification value

Data classification has been used for decades to help organizations make determinations for safeguarding sensitive or critical data with appropriate levels of protection. Regardless of whether data is processed or stored in on premises systems or in the cloud, data classification may be a suggested starting point for determining the appropriate level of controls for the confidentiality, integrity, and availability of data based on risk to the organization.

For example, data that is considered confidential should be treated with a higher standard of care than data consumed by the general public. Data classification allows organizations to evaluate data based on sensitivity and business impact, which then helps the organization assess risks associated with different types of data.

Standards organizations, such as the International Standards Organization (ISO) and the National Institute of Standards and Technology (NIST), recommend data classification schemes so information can be effectively managed and secured according to its relative risk and criticality, advising against practices that treat all data equally. According to these organizations, each data classification level should be associated with a recommended baseline set of security controls that help provide protection against vulnerabilities, threats, and risks commensurate with the designated protection level.

It is important to note the risks with over-classifying data. Sometimes organizations err by broadly classifying large disparate sets of data at the same sensitivity level. This over-classification can

incur unwarranted expenses by putting into place costly controls that can additionally impact business operations. This approach can also divert attention to less critical datasets and limit business use of the data through unnecessary compliance requirements due to over-classification.

Data classification process

Customers often seek tangible recommendations when it comes to establishing data classification policies. These steps help not only in the development phase, but can be used as measures when reassessing if datasets are in the appropriate tier with corresponding protections.

The following paragraphs recommend a step-by-step approach, based on guidance supported by international standards that customers can consider when developing data classification policies:

1. **Establishing a data catalog** — Conducting an inventory of the various data types that exist in the organization, how they are used, and whether any of it is governed by a compliance regulation or policy. Once the inventory is complete, group the data types into one of the data classification levels the organization has adopted. AWS Glue Data Catalog lets you store, annotate, and share metadata in the AWS Cloud while providing comprehensive audit and governance capabilities, with schema change tracking and data access controls.
2. **Assessing business critical functions and conduct an impact assessment** — An important aspect in determining the appropriate level of security for data sets is to understand the criticality of that data to the business. Following an assessment of business-critical functions, customers can conduct an impact assessment for each data type.
3. **Labeling information** — Undergo a quality assurance assessment to ensure that assets and data sets are appropriately labeled in their respective classification buckets. Additionally, it may be necessary to create secondary labels for data sub-types to differentiate particular sets of data within a tier based on privacy or security compliance requirements. Using services such as Amazon SageMaker and AWS Glue provide insight, and can support data labeling activities.
4. **Handling of assets** — When data sets are assigned a classification tier, data is handled according to the handling guidelines appropriate for that level, which include specific security controls. These handling procedures should be documented but also adjust as technology changes. (Refer to *Customer considerations for implementing data classification schemes* later in this document for additional information on data handling.)
5. **Continuous monitoring** — Continue to monitor the security, usage and access patterns of systems and data. This can be done through automated (preferred) or manual processes to identify external threats, maintain normal system operations, install updates, and track changes to the environment.

For guidance on how this process can be supported by AWS services, refer to the *Leveraging the AWS Cloud to support data classification* section of this document.

Data classification and privacy considerations

Data classification is particularly important as new global privacy laws and regulations provide consumers with rights to access, deletion, and other controls over personal data.

At the time of this writing, according to the United Nations Conference on Trade and Development (UNCTAD) 71% of the world's countries have data protection and privacy legislation in place while 9% have a draft legislation in progress.

For example, under the European Union's General Data Protection Regulation (GDPR), certain organizations are required to respond to certain consumer requests within a month of receipt. Similarly, acts such as the California Consumer Protection Act (CCPA) and Health Insurance Portability and Accountability Act (HIPAA) gives patients and consumers the right to control how their Personally Identifiable Information (PII) and Protected Health Information (PHI) is handled.

To respond appropriately, organizations must generally verify a requester's identity, locate the requestor's personal data, ensure the data returned only contains the requestor's personal data, and possibly refuse a request if it's inconsistent with applicable law.

Organizations that adopt strong data classification policies are better positioned to provide timely responses to these requests. A data classification framework along with proper tagging and labeling will help protect this personal data. Secondary labels can be used within a classification tier to assist with the tagging and discovery of relevant privacy data. This allows an organization to quickly address issues as they arise. Such additional mechanisms also aid in traceability and access monitoring of sensitive data sets.

Data classification models and schemes

Classification models and schemes can be divided into government classification schemes, and commercial classification schemes. Government classification schemes provide a set standard based on laws, policies, and executive directives. Commercial classification schemes, on the other hand, are less standardized and depend on the respective organizational need for protection of data with varying levels of sensitivity, as well as the need to meet compliance and regulatory requirements.

The city of Washington, D.C. implemented a new data policy in 2017 focused on being more transparent, while still protecting sensitive data. While Washington D.C. implemented a five-tier model, these tiers can align with other widely-adopted three-tier classification schemes used in cloud accreditation regimes.

- **Level 0 — Open Data**. Data readily available to the public on open government websites and datasets.
- **Level 1 — Public Data, Not Proactively Released**. Data not protected from public disclosure or subject to withholding under any law, regulation, or contract. Publication of the data on the public Internet would have the potential to jeopardize the safety, privacy, or security of anyone identified in the information.
- **Level 2 — For District Government Use**. Data that is not highly sensitive and may be distributed within the government without restriction by law, regulation, or contract. It is primarily daily government business operations data.
- **Level 3 — Confidential**. Data protected from disclosure by law, regulation, or contract and that is either highly sensitive or is lawfully, or contractually restricted from disclosure to other public bodies. This includes privacy-related data (such as PII, PHI, payment card industry data security standard (PCI DSS), federal tax information (FTI), and so on.
- **Level 4 — Restricted Confidential**. Data that unauthorized disclosure could potentially cause major damage or injury, including death to those identified in the information, or otherwise significantly impair the ability of the agency to perform its statutory functions.

U.S. national classification scheme

The U.S. government uses a three-tier classification scheme for national security information as described in Executive Order 135261. This scheme is focused on handling instructions based on potential impact to national security if it is disclosed (confidentiality).

1. **Confidential** — Information where unauthorized disclosure reasonably could be expected to cause damage to national security.
2. **Secret** — Information where unauthorized disclosure reasonably could be expected to cause serious damage to national security.
3. **Top Secret** — Information where unauthorized disclosure reasonably could be expected to cause exceptionally grave damage to national security.

Within these classification tiers, there are also secondary labels that can be applied that give origination information and can modify the handling instructions. The U.S. also uses the term *unclassified data* to refer to any data that is not classified under the three classification levels. Even with unclassified data, there is the potential use of secondary labels for sensitive information, such as For Official Use Only (FOUO) and Controlled Unclassified Information (CUI) that restrict disclosure to the public or unauthorized personnel.

U.S. information categorization scheme

Due to the targeted focus of the U.S. classification system and to address additional risks to information beyond confidentiality, NIST developed a three-tiered categorization scheme based on the potential impact to the confidentiality, integrity, and availability of information and information systems applicable to an organization's mission. Most of the data processed and stored by public sector organizations can be categorized into the following:

- **Low** — Limited adverse effect on organization operations, organization assets, or individuals.
- **Moderate** — Serious adverse effect on organization operations, organization assets, or individuals.
- **High** — Severe or catastrophic adverse effect on organization operations, organization assets, or individuals.

According to Fiscal Year 2015 data, U.S. federal departments and agencies categorized 88 percent of their systems into the low and moderate categories. AWS has Regions and services that are accredited to support these types of data categories and classifications.

United Kingdom (UK) government

In 2014, the UK simplified its data classification scheme by reducing the levels from six to three. They are:

1. **Official** — Routine business operations and services, some of which could have damaging consequences if lost, stolen, or published in the media, but none of which is subject to a heightened threat profile.
2. **Secret** — Very sensitive information that justifies heightened protective measures to defend against determined and highly capable threat actors (e.g., compromise could significantly damage military capabilities, international relations, or the investigation of serious organized crime).
3. **Top Secret** — Most sensitive information requiring the highest levels of protection from the most serious threats (such as compromise could cause widespread loss of life or could threaten the security or economic well-being of the country or friendly nations).

According to a cabinet office core briefing in 2013, the UK government categorized approximately 90 percent of its data as *Official*, which serves as the basic level of data classification.. The UK uses a flexible, de-centralized accreditation approach where individual agencies determine the cloud services suitable for *Official* data based on a cloud service provider's (CSP's) security assurance against 14 cloud security principles. Most UK government agencies have determined that it is appropriate to use reputable, hyper-scale CSPs when running workloads with *Official* data.

Commercial data classification scheme

In contrast to government classification schemes that provide a set of standards based on laws, policies, and executive directives, classification schemes used in commercial and non-government organizations are more individual to the organization and the sensitivity of the data. A commercial classification scheme can range from a simple two-tiered approach with public and confidential data to a more granular approach (refer to the following table).

There is no single formula for creating a commercial data classification scheme. Organizations should consider the individual need for protection of proprietary, business, or user data with

varying levels of sensitivity, the need to meet compliance and regulatory requirements, and the possibility to align with cloud security best practices when creating a scheme and the process for classification. The scheme should enable categorization of organizational data based on criticality and sensitivity in order to help determining appropriate protection and retention controls.

For example, under certain conditions GDPR grants rights such as the *Right to be Forgotten*, the *Right to Know*, and the *Right to Data Portability*. To implement these rights organizations may seek to understand their data, especially how it is categorized and where it lives.

The GDPR itself considers different categories of data: personal data, special categories of personal data, publicly available data (that contains personal data), and non-personal data. Non-personal data is not covered by GDPR, while certain special categories of personal data (such as health data) are considered very sensitive and require more protection. Therefore, organizations should consider implementing a classification scheme and process to help comply with regulatory obligations, and help prevent mishandling of data. Additional background on GDPR is provided in AWS whitepaper Navigating GDPR Compliance on AWS.

Table 1 — Five-tiered commercial data classification approach according to the book CISSP Security Management and Practices

Classification	Description
Sensitive	Data that is to have the most limited access and requires a high degree of integrity. This is typically data that will do the most damage to the organization should it be disclosed.
Confidential	Data that might be less restrictive within the company but might cause damage if disclosed.
Private	Private data is usually compartmental data that might not do the company damage but must be keep private for other reasons. Human resources data is one example of data that can be classified as private.
Proprietary	Proprietary data is data that is disclosed outside the company on a limited basis or contains information that could reduce the

Classification	Description
	company's competitive advantage, such as the technical specifications of a new product.
Public	Public data is the least sensitive data used by the company and would cause the least harm if disclosed. This could be anything from data used for marketing to the number of employees in the company.

Industry-specific approaches

This section identifies industry-specific examples for data classification, which may include sector-specific requirements. As mentioned earlier, different data types (such as government, financial, and healthcare data) may require additional considerations for tiers and secondary labels to address different handling procedures. Regardless of data belonging to public or commercial entities, customers must conduct the due diligence of adhering to local compliance and regulatory requirements.

The following chart contains examples of data classification schemes in practice today, descriptions of what can be included in that category based on tier, and examples of workload types for a particular tier.

Example 1

Table 2 — Data classification – public sector

Data classification	Examples of workloads
Tier 3 – Government confidential and above-sensitive data	-National security and defense information -Government intelligence information -Law enforcement information -Government program monitoring or oversight investigations information

Data classification	Examples of workloads
Tier 2 – Restricted	-Personally, identifying information about individuals -Human Resources Management -Personal profile information -Aggregated financial or market data
Tier 1 – Public data	-Marketing or promotional information -Information related to other general government administrative or program activities -Intra-agency workplace policy development and management

Example 2

Table 3 — Data classification – enterprises

Data classification	Examples
Tier 3 – Highly Strategic	Highly sensitive trade secret and material confidential business information (such as certain pricing, merger and acquisition information, marketing plan, proprietary processes, marketing plans, new product designs, inventions prior to a patent applicati on or held as trade secret) the public disclosur e of which could be expected to cause severe or catastrophic legal, financial or reputational damage.

Data classification	Examples
Tier 2 – Restricted	-Most material and non-material business data (such as email, sales and marketing account data, signed contracts, receipts) -Information required by law to be protected from unauthorized disclosure -Employee HR records (including employee disciplinary reports)
Tier 1 – Protected data	-CRM systems -Vendor bank account numbers and payment instructions -Information that is available only to a specific group of the company's employees for the purpose of conducting business -Information for internal use only

AWS recommendations

In most cases, AWS recommends starting with a three-tiered data classification approach (refer to the following table), which has been shown by public and commercial organizations that have adopted the AWS cloud, to sufficiently meet their data classification needs and requirements. As an example, the table below includes three tiers, and a naming convention for each tier. For organizations that have more complex data environments or varied data types, secondary labeling is helpful without adding complexity with more tiers. AWS recommends using the minimal number of tiers that make sense for the organization.

Table 4 — Three-tiered data classification approach

Data classification	System security categorization	Cloud deployment model options
Unclassified	Low to High	Accredited public cloud
Official	Moderate to High	Accredited public cloud
Secret and above	Moderate to High	Accredited private/hybrid/community cloud/public cloud

Data residency consideration — AWS encourages customers to assess their data classification approach and hone in on which data needs to stay within their country or region, and why. By doing so, customers may find that their data, potentially even sensitive and critical data, may be stored and/or replicated elsewhere if there is no particular law or policy requiring geographic restrictions. The ability to failover to another region can help further reduce risk of loss in the event of a disaster and provide access to technologies and capabilities that may not be available in their area. Learn more in the AWS Data Residency whitepaper.

The NIST data classification scheme is widely recognized as adequate classification scheme in sector-specific, national, and international certifications. In fact, governments such as the Philippines and Indonesia are evaluating and adopting data classification schemes that apply similar principles as the US (like NIST) and UK models. However, organizations are best positioned to develop their own classification schemes based on organizational and risk management needs. Organizations seeking to move away from complex, burdensome tiered schemes can run risk impact assessments to evaluate whether a simpler scheme, such as the three-tiered model, would more effectively meet their management and classification needs.

Organizations should select the appropriate cloud deployment model according to their specific needs, the type of data they handle, and assessed risk. Depending on the classification of the data, they will need to apply the relevant security controls (such as encryption) within their cloud environment.

When assessing risk and determining security controls, it is important to understand how cloud services differ from on-premises systems, alternate controls to consider as compared to traditional IT implementation, and the differences in implementation of controls (such as the shared responsibility model).

When organizations have fully evaluated the commercial cloud with its numerous security benefits (such as the potential for improved availability, resiliency, visibility and automation, and a continually audited infrastructure), they may find that the vast majority of their workloads can be deployed in the cloud with due regard to a data classification scheme, similar to what the US and UK governments have done.

Globally, public sector organizations are increasingly using the native security benefits of the commercial cloud, and taking steps to help them meet their security and compliance requirements through appropriate data classification and implementation of security controls.

When organizations have fully evaluated the commercial cloud with its numerous security benefits, they may find that the vast majority of their workloads can be deployed in the cloud with due regard to a data classification scheme, similar to what the US and UK governments have done.

Customer considerations for implementing data classification schemes

In addition to implementing a data classification scheme, it is equally important to determine data handling roles. ISO, NIST, and other standards place the responsibility of data classification on data owners, as they are the best positioned to determine the value, use, sensitivity, and criticality of their own data.

Risk management obligations vary depending on the role of the parties that handle the data. In other words, data owners (such as controllers who generate and control content like agencies and ministries) and non-data owners (such as processors that handle data in order to provision services) should be subject to requirements appropriate for the roles they play. In the context of public sector data classification, agencies or ministries work as the data owner and are responsible for classifying their data and determining the security accreditation that they expect their CSP to meet.

It is important to note that organizations applying a blanket high classification level to all data (despite its true risk posture) do not reflect a risk-based, outcome-focused approach to security. Protecting data classified at higher levels requires a higher standard of care, which translates into the organization spending increased resources on securing, monitoring, measuring, remediating, and reporting risks. It is impractical to commit the significant resources required to securely manage higher impact data for data that does not meet the requisite thresholds.

Also, the additional controls placed on data at the lower classification levels can negatively affect the availability, completeness or timeliness of that data to the general workforce, customers, and/or constituents. Where risks can be managed so that data is handled at a lower classification level, organizations will experience the most flexibility around how they use that data.

Newer considerations in data classification

Whether the journey to the cloud is new or established, it's critical to establish data classification rules. Similar to reviewing existing security practices and establishing better policies based on newer threats, considerations of how to help protect data are highlighted in this section as an example of what customers should consider when revisiting existing data classification policies. Most recently, conversations in industry consortiums have raised the following points:

- **Data is scattered everywhere** —The ubiquitous use of modern technology and reliance on information in enterprises across all sectors means massive volumes of data are stored, processed, and are in transit across numerous systems, devices, and end users. This can pose challenges for enterprises that are responsible for managing and securing large volumes of data.
- **Intra- and inter-organizational dependencies** —The ever-increasing need to collaborate and share information within an organization and across organizations within the same sector or with similar missions (such as hospital and health care networks).
- **End user knowledge** — Models that rely on end users to identify and classify data, such as those for machine learning (ML) processes, can be error-prone and incomplete. End users may lack the training or awareness of risks to categorize and manage data effectively.
- **Data classifiers and tagging** — There may be a lack of common definitions and understanding of classifiers, along with a lack of applicable standards in a few industries or persistence of labeling.
- **Context** — Context matters. The actual sensitivity and criticality of information depends greatly on other factors, such as how it is used and with whom; more so than what the information is necessarily about.

While these challenges may not seem new, they are factors worth considering as organizations develop and implement data classification.

Using AWS Cloud to support data classification

Cloud computing can offer customers the ability to secure their workloads. Organizations in highly regulated industries, public sector, enterprises, small and medium sized businesses, or startups, can work to meet their data classification policies and requirements in the cloud. Cloud service providers (CSPs), such as AWS, provide a standardized, utility-based service that is self-provisioned by customers. AWS does not have visibility into the type of data customers run in the cloud, which means AWS does not distinguish, for example, personal data from other customer data when providing cloud services. It is the customer's responsibility to classify their data and implement appropriate controls within their cloud environment (for example, encryption). However, the security controls CSPs implement within their infrastructure and their service offerings can be used by customers to help meet the most sensitive data requirements.

AWS Cloud security and compliance

AWS services offer the same high level of security to all customers, regardless of the type of content being stored. These services are then queued for certification against international security and compliance "gold" standards, which translates to customers benefiting from elevated levels of protection for customer data processed and stored in the cloud.

The risk events and threat vectors of greatest concern are largely accounted for through foundational cyber hygiene disciplines (such as patching and configuring systems), which CSPs can demonstrate through widely adopted, internationally-recognized security certifications and assurance programs such as ISO 27001, Payment Card Industry Data Security Standard (PCI DSS), and Service Organization Controls (SOC).

ISO 27001/27002 is a widely-adopted global security standard that sets out requirements and best practices for a systematic approach to managing company and customer information that's based on periodic risk assessments appropriate to ever-changing threat scenarios.

The Payment Card Industry Data Security Standard (also known as PCI DSS) is a proprietary information security standard administered by the PCI Security Standards Council, which was founded by American Express, Discover Financial Services, JCB International, MasterCard Worldwide and Visa Inc. PCI DSS applies to entities that store, process or transmit card data.

Service Organization Controls reports (SOC 1, 2, 3) are intended to meet a broad range of financial auditing requirements for U.S. and international auditing bodies. The audit for this report is conducted in accordance with the International Standards for Assurance Engagements No. 3402

(ISAE 3402) and the American Institute of Certified Public Accountants (AICPA): AT 801 (formerly SSAE 16).

Security and compliance reports such as SOC 1, PCI, FedRAMP are available to customers through AWS Artifact, a self-service portal for on-demand access to AWS' security and compliance reports. You can use those documents to validate the implementation and operating effectiveness of AWS security controls. Those documents can also be used as guidelines to evaluate and assess the effectiveness of your company's internal controls. AWS customers are responsible for developing or obtaining documents that demonstrate the security and compliance of their workloads in the AWS Cloud. For more information, refer to the Shared Responsibility Model.

In evaluating CSPs, organizations should leverage these existing CSP certifications so that they can appropriately determine whether a CSP (and services within the CSP's offerings) can support their data classification requirements. AWS encourages organizations to implement a policy identifying which existing national, international, or sector-specific cloud certifications and attestations are acceptable for each level in the data classification scheme to streamline accreditation and accelerate migrating workloads to the cloud.

AWS Well-Architected Framework and data protection best practices

The AWS Well-Architected Framework helps you understand trade-offs for decisions you make while building workloads on AWS. The security pillar provides guidance to help you apply best practices and current recommendations in the design, delivery, and maintenance of secure AWS workloads.

Two of the design principles given focus on data protection include:

- **Protect data in transit and at rest** — Classify your data into sensitivity levels and use mechanisms, such as encryption, tokenization, and access control where appropriate.
- **Keep people away from data** — Use mechanisms and tools to reduce or eliminate the need for direct access or manual processing of data. This reduces the risk of mishandling or modification and human error when handling sensitive data.

In regard to data classification, the framework provides these additional recommendations:

- **Identify the data within your workload** — Understand the type and classification of data your workload is processing, the associated business processes, data owner, applicable legal and

compliance requirements, where it's stored, and the resulting controls that are needed to be enforced.

- **Define data protection controls** — Using resource tags, separate AWS accounts per sensitivity (and potentially also per caveat / enclave / community of interest), Identity and Access Management (IAM) policies, Organizations Service Control Policies (SCPs), AWS Key Management Service (AWS KMS), and AWS CloudHSM, organizations can define and implement policies for data classification and protection.
- **Define data lifecycle management** — Have a defined lifecycle strategy based on sensitivity level and legal and organizational requirements., Consider the duration for which your organization has to retain data, data destruction processes, data access management, data transformation, and data sharing.
- **Automate identification and classification** — Automating the identification and classification of data, as opposed to directing access from an individual or team, reduces the risk of human error / exposure and helps implement the correct controls.

For more in-depth guidance, refer to Data Classification.

AWS services and features

AWS offers several services and features that can facilitate an organization's implementation of a data classification scheme. For example, Amazon Macie can help customers inventory and classify sensitive and business-critical data stored in AWS. Amazon Macie uses ML to automate the process of discovering, classifying, labeling, and applying protection rules to data. This helps customers better understand where sensitive information is stored and how it's being accessed, including user authentications and access patterns.

Another important feature supporting data classification and protection is AWS resource tagging. By assigning metadata to your AWS resources in the form of tags (with each tag being a label consisting of a user-defined key and value), you can manage, identify, organize, search for, and filter resources. Security tags can contain information on confidentiality, identifying specific data confidential level a resource supports, or compliance, like an identifier for workloads that must adhere to specific compliance requirements.

Other AWS services and features that can support data classification include, but are not limited to:

- AWS Identity and Access Management (AWS IAM) for managing user credentials, setting permissions, and authorizing access.
- AWS Organizations helps you centrally govern your environment with automated account creation, account grouping to reflect your business needs, and policies to enforce governance. Policies can include required actions such as tagging of resources.
- AWS Glue to store data and discover associated metadata like table definition and schema, in the AWS Glue Data Catalog. Once cataloged, your data is immediately searchable and available for ETL.
- Amazon Neptune, fully managed graph database, can give you insights into the relationships between different data sets. This can include identification and traceability of sensitive data through metadata analysis.
- AWS KMS or AWS CloudHSM for encryption; key management with AWS-generated keys, or bring your own key (BYOK) with Federal Information Processing Standards (FIPS) 140-2 validation.
- AWS CloudTrail for extensive logging to track who, what, and when data was created, accessed, copied/ moved, modified, and deleted.
- AWS Systems Manager to view and manage service operations such as patching, along with AWS Inspector to conduct vulnerability scans.
- Amazon GuardDuty for intelligent threat detection, supporting nearly continuous monitoring requirements.
- AWS Config to manage configuration changes and implement governance rules.
- AWS Web Application Firewall (AWS WAF) and AWS Shield to help protect web applications from common attack vectors (such as SQL injection, cross-site scripting, and DDoS).

For the entire list of AWS security services, refer to Security, Identity, and Compliance on AWS.

Conclusion

This paper covered data classification categories in governmental and commercial/enterprise sectors, privacy considerations, areas to consider for implementing classification schemes, and factors to consider to review, develop and enhance existing policies. This paper also provides inputs and best practices on how customers can use the AWS Cloud to support their classification needs and work to meet data sensitivity requirements.

Contributors

Contributors to this document include:

- Momena Cheema, Cloud Security Strategist, Amazon Web Services
- Min Hyun, Global Lead for Growth Strategies, Amazon Web Services
- Tim Anderson, Senior Security Advisor, Amazon Web Services
- Kiran Lakkireddy, Senior Solutions Architect, Amazon Web Services
- Neil DCruz, Startup Solutions Architect, Amazon Web Services
- Alexander Barge, Startup Solutions Architect, Amazon Web Services

Document revisions

To be notified about updates to this whitepaper, subscribe to the RSS feed.

Change	Description	Date
Whitepaper updated	Updated guidance, added commercial data classification schemes, and cloud security best practices.	August 3, 2022
Whitepaper updated	Minor formatting updates.	April 1, 2022
Whitepaper updated	Updated to reflect latest services and technologies.	March 3, 2020
Initial publication	Whitepaper published.	June 1, 2018

Notices

Customers are responsible for making their own independent assessment of the information in this document. This document: (a) is for informational purposes only, (b) represents current AWS product offerings and practices, which are subject to change without notice, and (c) does not create any commitments or assurances from AWS and its affiliates, suppliers or licensors. AWS products or services are provided “as is” without warranties, representations, or conditions of any kind, whether express or implied. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

AWS Glossary

For the latest AWS terminology, see the AWS glossary in the *AWS Glossary Reference*.