Data Classification policy

Data are classified in cloud on the basis of:

* **Low Risk:** 
  + Data for public use
  + Loss of confidentiality, integrity, or availability of the data or system has no significant impact.
* **Medium Risk**
  + Personally Identifiable data
  + Data is not general available to the public
  + Loss of the data or system could have a mildly adverse impact
* **High Risk**
  + Sensitive (Personally Identifiable) data
  + Loss of the data or system could have a significant adverse impact on our mission, safety, finances or reputation.

Data are classified on the basis of critical and non-critical data. The three main types of Data Classifications are:

* **Content**-based classification inspects and interprets files looking for sensitive information.
* **Context**-based classification looks at application, location, or creator among other variables as indirect indicators of sensitive information
* **User**-based classification depends on a manual, end-user selection of each document. User-based classification relies on user knowledge and discretion at creation, edit, review, or dissemination to flag sensitive documents.

1. **Government Classification:** Higher classifications protect information that might endanger national security.
2. **Typical Classification Level:** 
   * **Top Secret**
   * Top Secret is the highest level of classified information
   * Information is further compartmented so that specific access using a code word after top secret is a legal way to hide collective and important information.
   * **Secret**
   * Secret material would cause "serious damage" to national security if it were publicly available.
   * **Confidential**
   * Confidential material would cause "damage" or be prejudicial to security if publicly available.
   * **Restricted**
   * Restricted material would cause "undesirable effects" if publicly available. Some countries do not have such a classification in public sectors, such as commercial industries.
   * **Official**
   * This includes a diverse range of information, of varying sensitivities, and with differing consequences resulting from compromise or loss. OFFICIAL information must be secured against a [threat model](https://en.wikipedia.org/wiki/Threat_model) that is broadly similar to that faced by a large private company.
   * **Unclassified**
   * Unclassified is technically not a classification level, but this is a feature of some classification schemes, used for government documents that do not merit a particular classification or which have been declassified. This is because the information is low-impact, and therefore does not require any special protection, such as vetting of personnel.
   * **Clearance**
   * Clearance is a general classification, that comprises a variety of rules controlling the level of permission required to view some classified information, and how it must be stored, transmitted, and destroyed.

**Data Classification Process**

**PLAN**: Identify data assets, a data custodian to deploy the classification program, and develop protection profiles.

**DO**: After data classification policies are agreed upon, deploy the program and implement enforcement technologies as needed for confidential data.

**CHECK:** Check and validate reports to ensure that the tools and methods being used are effectively addressing the classification policies.

**ACT:** Review the status of data access and review files and data that require revision using a reclassification and revision methodology to adopt changes and to address new risks.

* Establish a matrix of classification types and determine eligibility of each element for use in a cloud setting, along with any required safeguards that inform eligibility, e.g., the absence of public file sharing.
* Determine authorized users of the data and permissible actions, such as access, delete, and storage constraints by time, date, geography, and device.
* Determine response and remediation to actions inconsistent with policies created.
* Implement secure key management: Encryption keys must be stored securely, with strict access control, for example, by using a key management service such as AWS KMS.
* Keep all users away from directly accessing sensitive data and systems under normal operational circumstances. For example, provide a dashboard instead of direct access to a data store to run queries.
* Enforce encryption at rest: Enforce your encryption requirements based on the latest standards and recommendations to help protect your data at rest.
* Automate data at rest protection: Use automated tools to validate and enforce data at rest protection continuously, for example, verify that there are only encrypted storage resources.
* Enforce access control: Enforce access control with least privileges and mechanisms; including backups, isolation, and versioning, to help protect your data at rest. Prevent operators from granting public access to your data.
* Define data lifecycle management: Your defined lifecycle strategy should be based on sensitivity level, as well as legal and organization requirements. Aspects including the duration you retain data for, data destruction, data access management, data transformation, and data sharing should be considered.
* **Identify the data within your workload:** Consider discovering data using [Amazon Macie](https://wa.aws.amazon.com/wat.concept.macie.en.html) to recognize sensitive data such as personally identifiable information (PII) or intellectual property.
* **Define data protection controls**
* **Automate identification and classification**
* **Define data lifecycle management**
* **Identify data types**: Identify the types of data that you are storing or processing in your [workload](https://wa.aws.amazon.com/wat.concept.workload.en.html). That data could be text, images, binary databases, etc.

**Data Classification**

* Gain visibility into where sensitive data is located, who is accessing it, and how they are using it by leveraging Amazon Web Services.
* When storing sensitive personal information in the cloud, it is imperative to choose a provider that can facilitate compliance to privacy regulations and has a global presence in the various regions needed to support these important regulatory requirements.

Types of control in Data Classification:

**Encryption:**

* File/folder encryption—File and folder encryption relies on applying a policy that dictates what to encrypt and who can access it.
* Full-disk encryption for cloud workload storage volumes—Full-disk encryption can help solve the problem of data exposure within virtual machines, but key management is a major concern.
* Specialized encryption (database, email)—Specific encryption for database columns or tables, as well as email stores, can be implemented in the cloud too.
* Cloud-native storage encryption—For specialized storage options like Amazon S3 buckets, encryption is easiest to implement through built-in AWS configuration options that allow for selection of encryption keys and access controls.

**Data access**—Enable data access controls that align with least-privilege business use cases.

**Legal access —**Determine whether the data will be accessible to legal counsel (for electronic discovery, for example).

**Transformation:**

**Integrity:** Data integrity will be handled through SLAs to ensure no corruption or data loss occurs.

**Sensitivity:** This classification largely depends on how the data is being sent to the cloud and processed. At one stage, it may be considered sensitive data, whereas at another it may be obfuscated or not have any recognizable qualities as personal or sensitive data.

**Attribution**: This will depend on the applications in use and the manner of storage.