Domain 9: Data Security

Introduction

Data security is vital for organizational integrity, confidentiality, and customer trust, as well as regulatory compliance. With cloud services expanding rapidly and cyber threats growing more sophisticated, a robust approach to safeguarding information is imperative.

This domain addresses the complexities of data security in the cloud, covering essential strategies, tools, and practices for protecting data in transit and at rest. From data classification to encryption and access controls, it offers guidance on navigating the evolving data security landscape. Additionally, it provides insights into key concepts and technologies shaping the future of cloud data security, emphasizing the measures needed to prevent breaches and uphold data privacy.

Learning Objectives

The learning objectives for this domain aim to provide readers with knowledge on:

- Understanding data security fundamentals.
- Data classifications and states.
- Cloud storage types and their related security measures.
- Data security techniques, such as key management.
- Protecting various types of computing workloads.
- Posture management.
- Advanced data security concepts.

9.1 Primer on Cloud Storage

By categorizing data based on its type, sensitivity, and criticality, organizations can implement proper security methods per data type. Improper handling of data can lead to data breaches, compliance violations, and data loss. From a strategy perspective, understanding and implementing data classification practices help organizations align with operational and compliance strategies.

© Copyright 2024, Cloud Security Alliance. All rights reserved.

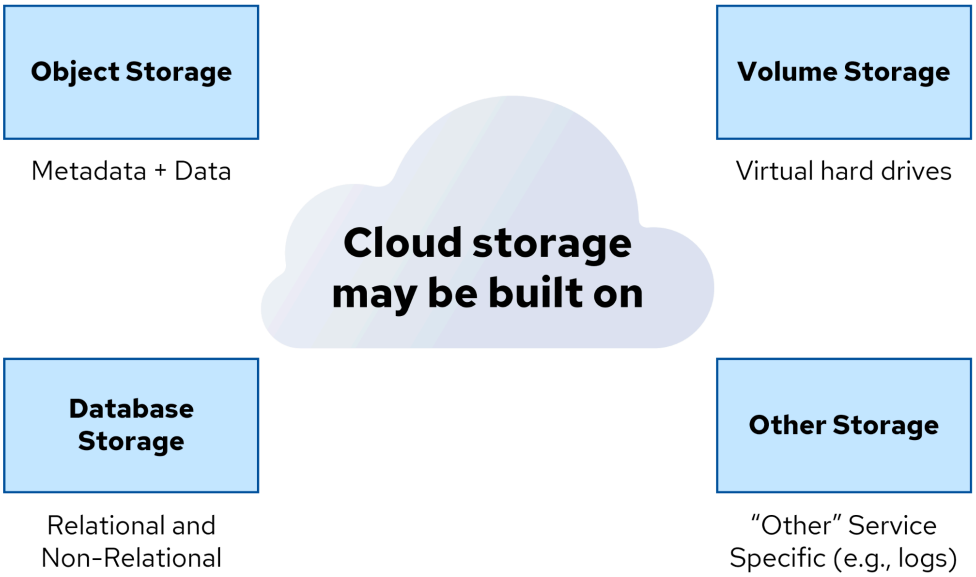


Figure 22: Types of Cloud Storage Solutions

9.1.1 Object Storage

Cloud object storage is a cloud-native data storage architecture where each file is represented as an object, including the data itself, metadata, and a unique identifier. Unlike block storage, it does not divide data into fixed-size blocks, and objects typically cannot be modified after creation. Cloud object storage is highly scalable and is the first choice for storing large amounts of unstructured data, such as media files, backups, and logs. In object storage, redundancy and availability are the responsibility of the cloud provider, while data governance, backups, and encryption are managed by the customer.

9.1.2 Volume/Block Storage

In Volume/Block storage, customers reserve a fixed block of storage and attach it to an existing workload (e.g., container, virtual machine (VM)), where it acts as a regular hard drive. Each block can be formatted, mounted to the operating system (OS), and managed independently. Block storage is known for its low latency, flexibility, and legacy support. However, customers are responsible for handling redundancy, availability, encryption, and backups.

9.1.3 Database Storage

Cloud providers offer managed services for both relational and non-relational databases.

Relational databases, or SQL databases, store data in structured tables with rows and columns. Cloud providers offer managed services for these databases, such as Amazon RDS, Google Cloud SQL, Microsoft Azure SQL Database, and Oracle Database services. These support engines like MySQL, Oracle, PostgreSQL, and SQL Server.

© Copyright 2024, Cloud Security Alliance. All rights reserved.

Non-relational databases, or NoSQL databases, store data in flexible formats like documents or key-value pairs. Examples include Amazon DynamoDB, Google Cloud Datastore, Oracle NoSQL Cloud DB, and Azure Cosmos DB. They are highly scalable and handle large amounts of unstructured data efficiently.

9.1.4 Other Types of Storage

PaaS storage refers to various cloud platform service-specific storage options. These can include logging services like Amazon CloudWatch, Google Cloud Logging, Oracle Events, and Azure Monitor, which store and analyze log data from applications and infrastructure. Message queues enable reliable communication between distributed application components, such as Amazon Simple Queue Service (SQS), Google Cloud Pub/Sub, and Azure Queue Storage. Oracle Cloud Infrastructure (OCI) Streaming and other PaaS storage services may include caches, in-memory databases, and more.

Cloud storage may also be offered as SaaS, such as Google Drive, Dropbox, Microsoft OneDrive, Box, and others. These services allow users to access and share files and resources through robust security, sharing, and collaboration through the Internet.

9.2 Data Security Tools and Techniques

Although technically, all information security is data security, the tools in this section form the core toolkit for focusing on the security of the data storage itself. Additional details on each of these tools will be provided in the remainder of the domain.

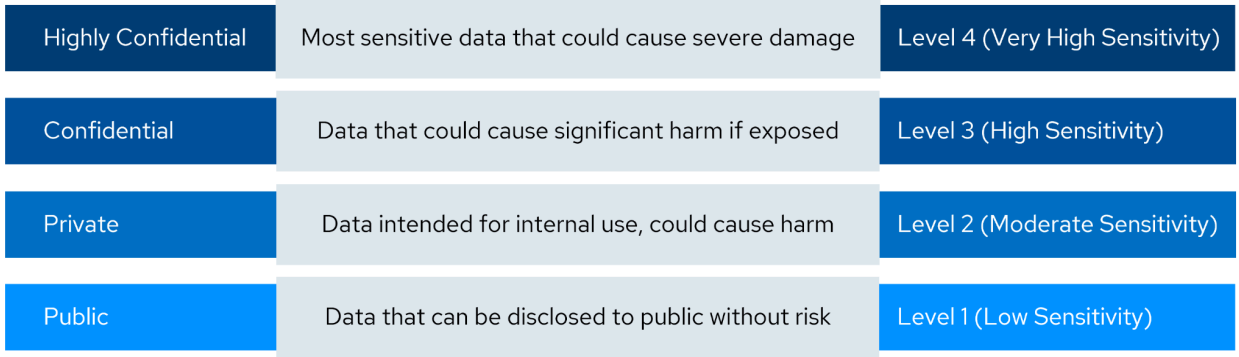
9.2.1 Data Classification

Data classification categorizes data based on type, sensitivity, and potential impact, crucial from both operational and compliance perspectives. Integrating data classification into organizational practices is essential for safeguarding data throughout its lifecycle. It helps prioritize asset protection and informs security and compliance strategies.

Continuous evaluation and adjustment are necessary as the organization evolves and regulatory requirements change. Combining a strong data classification strategy with clear ownership assignments enables swift incident response and advances data governance initiatives effectively.

© Copyright 2024, Cloud Security Alliance. All rights reserved.

Less controls and monitoring



More controls and monitoring

Figure 23: Data Classification Scale

9.2.2 Identity and Access Management

IAM systems govern entities' access to particular resources in the cloud environment when making API calls, or working within the service where the user and the data exist in the platform. This is different from access controls, which can also govern external access. In IaaS and PaaS, for example, access can be managed in a user-based IAM policy or in a resource policy attached to the storage.

9.2.3 Access Policies

Access policies govern resource access. They define the access and allowed actions (also known as permissions) for specific resources, and determine the network rules that regulate the traffic flow between resources. Both resource and network policies help enforce security boundaries.

9.2.4 Encryption and Key Management

Encryption protects data by converting it into unreadable ciphertext, decipherable only with the correct decryption keys. Key management systems securely store these keys, ensuring they remain separate from the CSP, either within their infrastructure or on an external Key Management Server (KMS). This dual approach ensures data confidentiality and integrity in the cloud.

© Copyright 2024, Cloud Security Alliance. All rights reserved.

9.2.5 Data Loss Prevention

Data Loss Prevention (DLP) ensures critical data like Intellectual Property (IP) and customer information stays within the organization. It identifies, monitors, and protects sensitive data, including in the cloud, by discovering, classifying, and enforcing security policies to prevent unauthorized sharing or exfiltration.

Cloud DLP faces challenges due to the vast data scale and distributed nature of the cloud, making comprehensive scanning difficult, especially for IaaS or PaaS. As a result, it's more commonly used for SaaS applications.

An effective cloud DLP strategy involves assessing the data landscape, prioritizing high-risk areas, and balancing cloud-native and third-party DLP solutions. A risk-based approach, alongside data minimization and robust access controls, helps manage cloud DLP challenges while safeguarding sensitive information across the enterprise cloud footprint.

9.3 Cloud Data Encryption at Rest

The image illustrates the different layers where data can be encrypted in the cloud, starting from the lowest layer (volume or object storage) and moving up to the application layer. As you move up the encryption layers, you gain more granular control and protection for your data, but it also involves more complexity in implementation and management. The appropriate encryption layer(s) should be chosen based on the sensitivity of the data, compliance requirements, performance needs, and the level of control and management required.

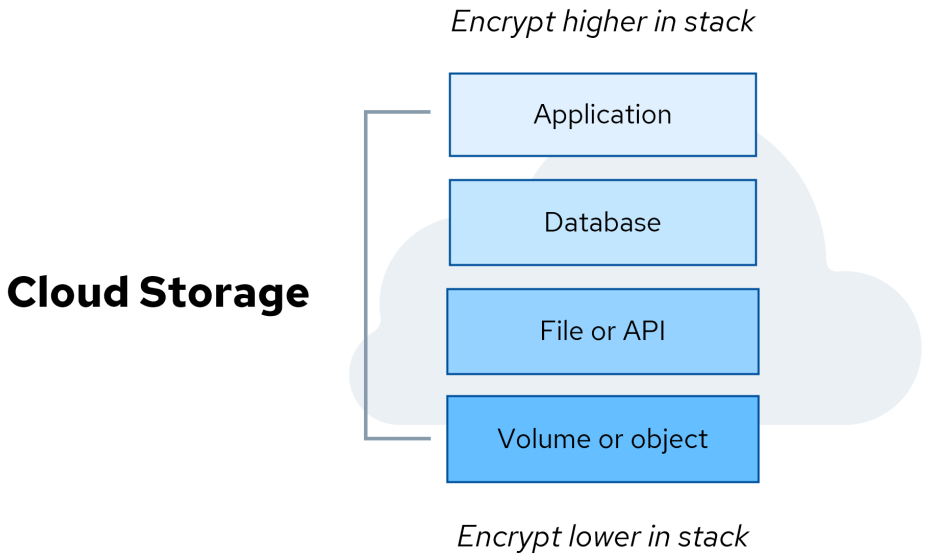


Figure 24: Cloud Data Encryption Layers

© Copyright 2024, Cloud Security Alliance. All rights reserved.

9.3.1 Application Level Encryption

Application level encryption means that the application layer encrypts new data items before sending them to the database for persistent storage and decrypting them when querying the database for information. This is done to secure specific sensitive data items like credit cards or personal sensitive information.

Key Features:

- Require changes to application components
- Affects performance and availability
- Protect data items even from database administrators
- Compliance: Meets regulatory and security standards

Use Cases:

- Protect data from any unauthorized access at the infrastructure level
- Backup and Archival: Securing backups, snapshots, and archival data
- Secure sensitive database items like credit cards

9.3.1.1 File/API Encryption

Encryption at the file or API level offers more granular protection than volume or object storage encryption. This layer allows for encrypting specific files or data accessed through APIs, providing more targeted security.

9.3.1.2 Database Encryption

Database encryption in the cloud involves securing data by encrypting the database that contains the data. Encryption can be implemented for the entire database (popular) or for specific tables or columns that contain sensitive information.

Key Features:

- Implemented by the database's internal tools (i.e., transparent database encryption (TDE)) or by encrypting the database storage
- Compliance: Meets regulatory and security standards
- Integrated Services: Works with other cloud services like access control and logging
- Comprehensive Protection: Encrypts both data and metadata

Use Cases:

- Backup and Archival: Securing backups, snapshots, and archival data
- Media Storage: Protecting media files and content
- Big Data and Analytics: Safeguarding data for analytics

© Copyright 2024, Cloud Security Alliance. All rights reserved.

- Secure sensitive database items like credit cards

9.3.1.3 Object Storage Encryption

Object storage encryption in the cloud secures data stored as objects in services like Amazon S3, Azure Blob Storage, and Google Cloud Storage.

Key Features:

- Automatic and Transparent: Handled seamlessly by the cloud provider without impacting performance
- Can be implemented per object or for all bucket/folder objects
- Compliance: Meets regulatory and security standards
- Integrated Services: Works with other cloud services like access control and logging
- Comprehensive Protection: Encrypts both data and metadata

Use Cases:

- Backup and Archival: Securing backups and archival data
- Media Storage: Protecting media files and content
- Big Data and Analytics: Safeguarding data for analytics
- Compliance-Sensitive Data: Meeting regulations like GDPR, HIPAA, or PCI DSS

Implementing object storage encryption ensures data security, regulatory compliance, and protection from unauthorized access, maintaining the confidentiality and integrity of cloud-stored data.

9.3.1.4 Volume Encryption in the Cloud

Volume storage encryption in the cloud secures data on virtual disks (volumes) from unauthorized access, commonly used in services like Amazon EBS, Azure Disk Storage, and Google Cloud Persistent Disks.

Key Features:

- Implemented either by an OS agent or by storage encryption services
- Data-at-Rest Protection: Encrypts stored data on the volume and copies of the volume like backups
- User Transparency: Seamless encryption/decryption with no user intervention, maintaining performance
- Automated Encryption: Easily enabled by default or through simple configuration
- Compliance and Security: Protect the data from access at the physical layer, as well as secure the snapshots of the data

Use Cases:

- Enterprise Applications: Encrypting volumes for databases and file systems
- Protects volume clones, snapshots, and backups

© Copyright 2024, Cloud Security Alliance. All rights reserved.

- Compliance-Sensitive Data: Securing data for General Data Protection Regulation (GDPR), Health Insurance Portability and Accountability Act (HIPAA), and Payment Card Industry Data Security Standard (PCI DSS) compliance
- Multi-Tenant Environments: Ensuring data isolation and security in shared cloud spaces

Implementing volume storage encryption enhances cloud data security, protecting against unauthorized access and ensuring compliance.

9.3.2 Cloud Data Key Management Strategies

The following are approaches to data key management that are available for cloud-based solutions.

9.3.2.1 Client-Side Encryption

Client-side encryption ensures that only encrypted data is stored by the cloud provider. Here, the customer encrypts their data before uploading it, preventing the cloud provider from accessing it in an unencrypted form. For example, an organization may encrypt sensitive files on-premises before uploading them to services like Amazon S3 or Google Cloud Storage. Since encrypted files can't be actively processed, this method is often used for data backup, archiving, or storing in a rarely accessed state.

9.3.2.2 Server-Side Encryption

Server-side encryption, offered by most cloud providers, encrypts data using keys stored by the provider. It's easy to set up and doesn't need specific customer configurations, making it attractive for organizations with less stringent security needs. The security relies on the provider's encryption protocols and key management.

9.3.2.3 Customer-Managed Encryption Keys

Provider key management systems allow customers to manage their encryption keys using the cloud provider's key management service (KMS) such as Amazon Web Services/Google Cloud Platform (AWS/GCP) or Key Vault in Microsoft Azure, while the cloud provider handles the encryption engine. This gives customers control over the keys' lifecycle—Creation, Rotation, Usage, and Deletion (CRUD). This ensures a clear separation of responsibilities and integrates with most cloud providers' services.

9.3.2.4 Customer-Provided Encryption Keys

In the context of customer-provided keys, also called Bring-Your Own-Keys (BYOK), the customer produces, controls, and stores a master key or key material used to create data encryption keys for use in the CSP environment. The data encryption keys are encrypted on boot with the master key to create a more flexible and secure key structure. This approach allows customers to achieve greater control over

© Copyright 2024, Cloud Security Alliance. All rights reserved.

encryption keys by bringing their own cryptographic materials and the ability to revoke the data encryption keys, if required, providing a higher level of security and control over their data.

This also imposes significant responsibility on the customer, however. They are now responsible for the ongoing operations of the key generation lifecycle and must ensure the availability and security of the master encryption keys. The customer must manage key lifecycle tasks such as rotation, storage, and secure handling. This includes maintaining the infrastructure and procedures to support these tasks, and ensuring that encryption keys are available when needed to avoid data access disruptions.

9.3.2.5 Custom Application Level Encryption

This approach uses hybrid scenarios and application-level encryption, where the customer manages both encryption and key management, adding control but also complexity. For example, using client-side encryption with the AWS Encryption Software Development Kit (SDK), data is encrypted before cloud transmission, and keys are managed client-side.

9.3.2 Data Encryption Recommendations

Having explored the fundamentals of data encryption in cloud environments, the following are recommended strategies for enhancing data encryption, each aimed at improving security, compliance, and overall data protection for your cloud-based operations:

9.3.2.1 Key Management Services (KMS)

To secure cloud applications and services, it is recommended that you utilize a KMS provided by your cloud provider. These services help manage cryptographic keys for the organization.

9.3.2.2 SaaS Considerations

If you're using SaaS, the provider KMS may be your only encryption option. SaaS often doesn't allow much customization, so you'll rely on the provider's tools for data protection. Some SaaS providers offer customer-managed keys by integrating their own SaaS with customer KMS keys if they use the same IaaS/PaaS providers.

9.3.2.3 Default Encryption

Default encryption can be sufficient for many needs. This method uses the cloud provider’s keys to encrypt data at rest and is usually included at no additional cost. It helps meet compliance requirements related to data protection, making it a convenient and secure choice for many users.

© Copyright 2024, Cloud Security Alliance. All rights reserved.

9.3.2.4 Different Keys for Services

Using different encryption keys for different services or deployments is a good practice. This approach enhances security by isolating the encryption domains and limiting a compromised key's potential impact.

9.3.2.5 IAM Policies on Keys

Apply IAM policies to your keys to enforce the principle of least privilege. Doing so ensures that only authorized users and services can use a particular key, and you can define what actions they can perform with it.

9.3.2.6 Alignment with Threat Models

Make sure your encryption strategies align with your threat models. For example, encrypting your database is less effective if an attacker can compromise application credentials or those of a Database Administrator (DBA). In such cases, the attacker can access or exfiltrate encrypted data through legitimate channels.

9.4 Data Security Posture Management

Data Security Posture Management (DSPM) is an emerging category of tools designed to focus on data-centered security. It includes data discovery and classification, which may also include DLP-like capabilities, to help you understand where you have data and its sensitivity. DSPM tools can then pull and evaluate all the overlapping access controls, IAM policies, resources, and network policies to assess and visualize WHO has access to the data and how. These tools offer suggestions and/or directly manage remediation or provide specific recommendations e.g. infrastructure as code (IaC) templates or policies.

The challenge in cloud data security is handling all the potential overlapping controls, all managed in different areas, that don't necessarily provide a complete view of your data use and exposure. DSPM is designed to fill that gap.

9.5 Object Storage Security

Object storage services like AWS S3 and Azure Blob Storage are crucial but pose data exposure risks. Misconfigurations and complex access settings, such as AWS's IAM roles and policies, increase this risk. Cloud providers offer tools to block public access at the deployment level, but this can limit necessary public access. Encrypting data with services like KMS adds security by keeping encryption keys separate from permission settings. Using Content Delivery Networks (CDNs) can enable safe public access to private storage. Continuous monitoring with tools like cloud security poster management (CSPM) and data security posture management (DSPM) is essential for maintaining security.

© Copyright 2024, Cloud Security Alliance. All rights reserved.

9.6 Data Security for Artificial Intelligence

As Artificial Intelligence (AI) technologies become more prevalent and integrated into critical business processes, ensuring the security and integrity of AI systems is paramount. Data security for AI involves implementing measures to protect AI systems, algorithms, and data assets from various security threats and vulnerabilities.

9.6.1 AI as a Service

In the AI as a Service(AIaaS) model, third-party providers offer AI capabilities through subscriptions, allowing organizations to integrate AI into their applications. Examples include Anthropic's Claude, OpenAI's ChatGPT, and Google Cloud's Vertex AI. Key considerations include:

- Clarifying data-deletion and retention policies
- Understanding data flow from company assets to services
- Evaluating the provider's AI security measures against adversarial attacks
- Understanding Service Level Agreements (SLAs), security practices, and regulatory compliance

These steps ensure the secure and reliable use of AI services.

© Copyright 2024, Cloud Security Alliance. All rights reserved.