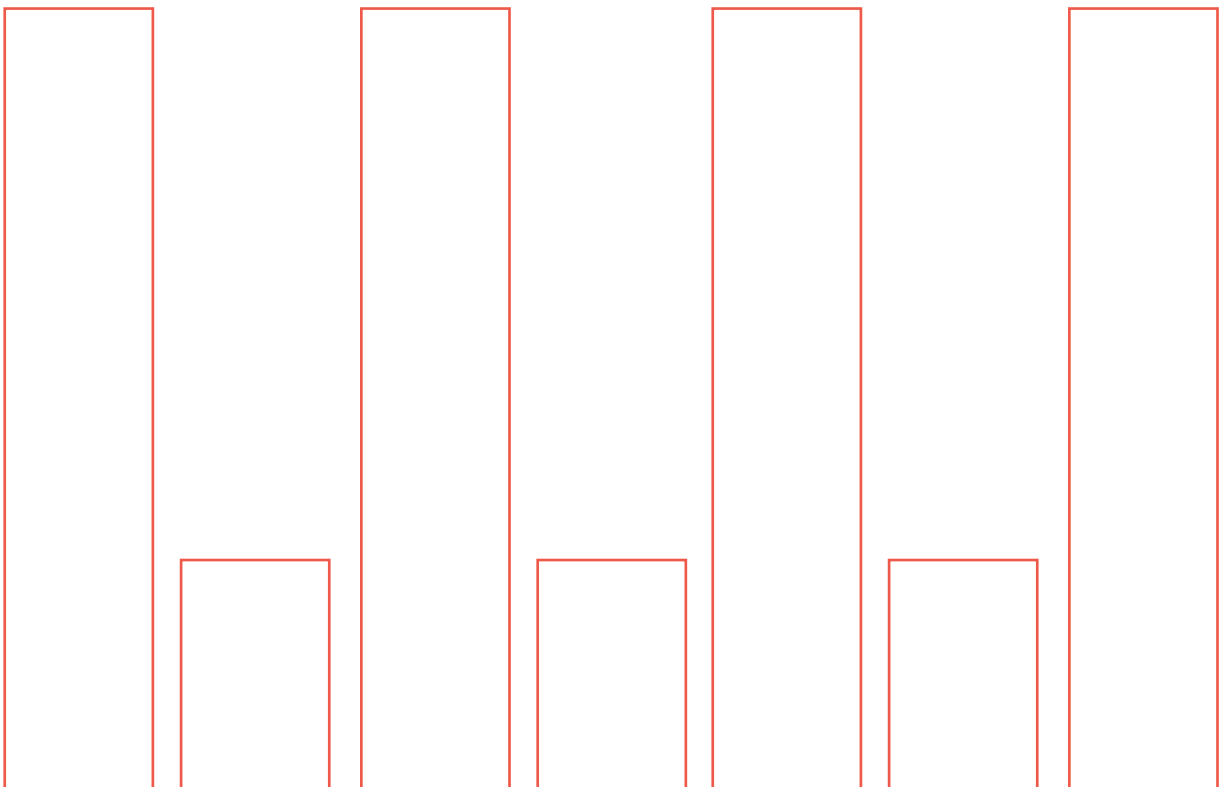
CMORG
CROSS MARKET OPERATIONAL RESILIENCE GROUP

Cloud-Hosted Data Vaulting

Good Practice Guidance

VERSION 1.0 | JANUARY 2025 | TLP CLEAR



Contents

Foreword: Co-Chairs of the CMORG Chief Information Officer Forum 2
Introduction 2
Examples of Financial Data Vaulting Applications 3
Financial Institution: 3
Third-Party Supplier: 3
Common to both: 3
Data Vaulting Good Practice 3
Define Data Vaulting Scope and Policy 3
Select Appropriate Vaulting Solutions 4
Automate Vaulting Processes 4
Implement Secure Transfer Mechanisms 4
Enforce Access Control and Auditing 4
Retention and Deletion Policies 5
Ensure Redundancy and Disaster Recovery 5
Compliance with Legal and Industry Standards 5
Foundation Principles 5
NIST Zero Trust Architecture 8
Annex A: Cloud-Hosted Data Vaulting Use Cases 9

Foreword: Co-Chairs of the CMORG Chief Information Officer Forum

Welcome to this good practice guidance for cloud-hosted data vaulting. The interconnected nature of the global marketplace has put data increasingly at risk, and there is a growing body of examples of organisations experiencing significant disruption as a result of cyber attacks compromising critical data or rendering infrastructure unavailable. The Cross Market Operational Resilience Group (CMORG) Chief Information Officers Forum-led (CIOF) has brought together subject matter experts and interested parties from across industry to share information on existing implementations and data vault designs with a specific focus on cloud-based solutions. We hope you find this best practice of use in shaping your organisation's pathway to increased operational resilience through robust recovery architecture.

Introduction

Cloud-hosted data vaulting refers to the process of secure cloud-based storing of isolated data using a virtual airgap that protects backups but allows for temporary network connections to enable necessary remote access. This is distinct from a cloud-based resolution tool.

A cloud-hosted model represents a credible approach to data vaulting being adopted by finance sector firms and, increasingly, a range of third parties. This is typically implemented as part of a layered recovery strategy for firms and other organisations to utilise during an incident to restore services. The key requirement for the data vault solution is to mitigate the impact of a cyber attack by providing a highly resilient solution to control the confidentiality, integrity, and availability of the critical data that a firm deposits and manages within the solution. A range of applicable use cases are captured at Annex A.

A cloud-hosted solution cannot provide protection against, or recovery from, all threats or risks, either accidental or malicious. However, there is a broad range of scenarios where it could assist in the recovery of critical services when 'traditional' recovery solutions, such as failover to a disaster recovery instance, are not effective. Logical segregation, coupled with immutability, can provide a faster recovery as well as greater assurance of data integrity.

This good practice guidance outlines the essential steps and recommendations for creating, maintaining, and securing immutable cloud-hosted data vaults to ensure the availability, integrity, and security of critical data. It takes the key outputs of the CMORG Cloud-Hosted Data Vault Reference Architecture[1] and outlines actionable practices for establishing an effective cloud-hosted data vaulting process that meets regulatory requirements and mitigates risks related to data loss, corruption, and

[1] Data Vaulting Reference Architecture | Cross Market Operational Resilience Group

unauthorized access.

Examples of Financial Data Vaulting Applications

Financial Institution:

Vaulting customer transaction data to meet regulatory requirements for data retention and disaster recovery.

Third-Party Supplier:

Single or multiple vaults of multiple customer transaction datasets for customer assurance and recovery purposes.

Common to both:

Ensuring backups are encrypted and stored appropriately (in geographically separate locations, for example), adhering to jurisdictional and regulatory compliance demands.

Data Vaulting Good Practice

Define Data Vaulting Scope and Policy

- **Collaborative in design:** A data vault solution should identify the needs, roles, and processes associated with managing and maintaining the vault service. It should also identify, understand and document service/object relationships and dependencies, as these will be required metadata for the vaulting process.
- **Threat-driven:** Firms face different risks and threats from actors. A data vault architecture should be informed by a threat evaluation specific to the firm. Doing so helps reduce wasted costs from vaulting items not at risk.
- **Identify critical data:** Not all data requires vaulting. Perform a risk assessment to determine which data sets are essential for operational resilience, compliance, or legal requirements.
- **Create a tailored vaulting policy:** A data vault needs to be deployed flexibly in terms of scale and location and support the unique requirements of individual firms. This could mean national, regional, or global deployments in one or more Cloud Service Provider (CSP) environments. A data vault solution will always be aligned to a single firm but can incorporate a multi-tenancy approach to facilitate separation requirements within a firm using multiple data vaults associated with a range of business services and information classifications.
- **Compliance and regulatory requirements:** Ensure vaulting practices align and adhere to regulatory requirements (e.g. Important Business Services (IBS) and Critical or Important Functions (CIF)).

- **Accountable Executive**: Consider assigning an accountable Executive for ownership of the vault recovery capability to ensure clear organisational decision making and representation.

- **Continuously refreshed and evaluated:** Firms are continually evolving, and a data vault solution should be continually managed in step with organisational development. It is not a 'one off' activity.

Select Appropriate Vaulting Solutions

- **Cloud vs. Physical vaulting:** Evaluate whether to use a cloud-based vault, an offsite physical facility, or a hybrid solution. Cloud vaulting offers scalability and accessibility, while physical vaults provide physically controlled environments for specific regulatory or security needs.
- **Encryption and compression:** To ensure firm data remains confidential, encryption must be used to protect data in transit and at rest at all times throughout the vaulted object's entire lifecycle within the data vault solution.
- **Service provider reliability:** A cloud vaulting solution requires a mechanism that can be deployed flexibly in terms of scale and location and support the unique requirements of individual firms.

Automate Vaulting Processes

- **Set regular backup intervals:** Automate regular vaulting intervals to occur at intervals aligned with Impact Tolerance Levels (ITOLs) and key business requirements. Not all data needs to be vaulted. Equally, not all vault-worthy data requires the same frequency of backup. Data must be in synchronisation to ensure that that the integrity of the business applications are met. Consideration should be given to how business processes are integrated, and data needs to be synchronised to similar points in time to ensure recovery.
- **Monitoring:** Key operational functions of the data vault will include operational monitoring, logging, and observability.
- **Record and guide restoration:** Produce plans and playbooks for reference in times of business as usual, and especially during a crisis restoration.

Implement Secure Transfer Mechanisms

- **Use secure data channels:** Data in transit between primary systems and the vault should be appropriately encrypted.
- **Integrity verification:** Ensure data integrity during and after transfer. This ensures that data has not been corrupted during the process.

Enforce Access Control and Auditing

- **Limit access:** Determine suitable personnel to act on authorisation requests for data access. Implement role-based access control (RBAC) policies, Multi-

Factor Authentication (MFA), and Quorum authentication to ensure users can only access data relevant to their job function.

- **Audit logs and tracking:** Maintain detailed logs and tracking of all access and modifications to the vault.
- **Assure immutability:** Ensure the data cannot be modified after it has been vaulted.

Retention and Deletion Policies

- **Set data retention schedules:** A data vault solution may have multiple data vaults with dedicated key management. This is created based on policies such as retention, data protection level, and business unit.
- **Automated deletion:** Configure automated systems to delete data after it exceeds its retention period. This ensures outdated information does not occupy valuable storage space or become a potential compliance liability.

Ensure Redundancy and Disaster Recovery

- **Geographic redundancy:** Store data in multiple geographic locations or use cloud services with region-based redundancy. This ensures business continuity even if one location experiences a disaster. However, assess data regulations and data sovereignty that will have implications for data storage and handling.
- **Recovery proving design:** Consider the recovery proving design as part of the end-to-end vault solution to ensure vault principles can be maintained whilst proving recovery capabilities (for example, being able to test egress without implicating the production network).
- **Test restoration processes**: Set restoration tests at regular intervals, aligned and defined by the objects being vaulted. Duplicate environments should be considered key to enable effective testing.
- **Recovery testing**: Test recovery plans that involve restoring data from the vault.

Compliance with Legal and Industry Standards

- **Adhere to legal obligations:** Follow regional and international mandates for the protection and retention of vaulted data.
- **Regulatory reporting:** Maintain records and ensure data vaulting processes meet regulatory standards.

Foundation Principles

The following foundation principles set out the core tenets for a data vault solution and inform the fundamentals of its operation and security. They represent collective good

practice developed through CMORG and can be tailored to specific organisational needs.

Number	Principle
1	All data vault solution(s) must be isolated from firm production environments and other data vault solutions.
2	The management capability must be isolated from production and not have any application integration to or from any other applications.
3	The credentials for the data vault solution must not be shared outside of the data vault.
4	Ingress of an object to the data vault must be a pull from the data vault ingress staging zone and egress of an object from the data vault must be a push from the data vault to the egress staging zone.
5	Objects in a data vault must be immutable.
6	The data vault solution must have the ability to have multiple logical data vaults, each with their own access control in alignment with role-based access controls (RBAC) and policies.
7	The data vault solution must have the ability to encrypt the objects at rest and be extensible to support different keys for each of the data vaults.
8	There must not be any direct network connections (e.g. TCP) open to the data vault zone from outside the data vault solution.
9	There must be no persistent and accessible compute in the data vault zone.
10	Objects in the data vault zone should be ingested using non-proprietary formats, allowing the ability to recover without any intermediary applications such as backup systems. The files egressed should be the same as at point of ingress – the exception being the metadata file, which may include details of activities, such as analytics, that have occurred in the data vault. To note: This may require the rebuild of the backup infrastructure before restoration.
11	The data vault solution must have the ability to run analytics against its objects to check integrity and for any anomalies without executing the object. Integrity checks must be done prior to securing the data, doing it post will not ensure recovery of the original data or the service that the data supported.

12	The data vault solution must have a management capability that tracks objects in each of the data vaults and holds metadata on those objects.
13	The data vault solution must be able to map objects to other objects and services to other services, using the object metadata to aid recovery tasks.
14	The data vault solution must have the ability to recover objects to a variety of targets to support a range of recovery scenarios.
15	Any changes or actions in the management capability must be audited, implement multi-factor authentication, and adhere to 'M of N' specifications as per section 2.5.3 of the CMORG Cloud-Hosted Data Vault Reference Architecture.
16	Access to a data vault solution must be from authorised interfaces only.
17	The data vault solution must facilitate the limiting of interfaces to enable firms to follow recommended good practice, such as browse-down (i.e. from 'high trust' devices with low risk of compromise).
18	The data vault solution must be architected with a high degree of resilience and data durability.
19	All data entering / exiting the data vault solution (in-transit) and all data stored within the data vault solution (at-rest) must be encrypted in line with the CMORG Cloud-Hosted Data Vault Reference Architecture Key Management Principles (as confidentiality is a critical priority of the data vault solution, it is vital that suitable encryption and key management practices are utilised).
20	The data vault solution will adhere to recognised cloud security frameworks, for example including the NCSC Cloud Security Principles.

Limitations: Given the potential for production services to be compromised or completely unavailable, it is strongly recommended that firms store reconstruction scripts, processes, and run books within the data vault too (scripts based on the different disaster scenarios where the data vault solution would be required). Cloud Vaulting, although scalable, is slow in the retrieval of data. Where the vault would be seen as the last option where the firm's other backup solutions have been compromised, the retrieval of the data would be limited to the size of the network pipe between the vault and the various systems that need to be recovered. Furthermore, if the data is stored as objects these would need to be transformed back to the original data form before it can be used.

NIST Zero Trust Architecture

The data vault solution will implement zero trust architecture principles as outlined in NIST Special Publication 800-207. In overview, these include:

1. All data sources and computing services are considered resources;
2. All communication is secured regardless of network location;
3. Access to individual enterprise resources is granted on a per-session basis; Access to resources is determined by dynamic policy—including the observable state of client identity, application/service, and the requesting asset—and may include other behavioural and environmental attributes;
4. The enterprise monitors and measures the integrity and security posture of all owned and associated assets;
5. All resource authentication and authorization are dynamic and strictly enforced before access is allowed; and
6. The enterprise collects as much information as possible about the current state of assets, network infrastructure, and communications and uses it to improve its security posture.

Annex A: Cloud-Hosted Data Vaulting Use Cases

Cloud-hosted data vault solutions are likely to be used in the following circumstances:

A) **Use Case 1: Data availability events.** The primary use case in this instance is a malware attack (usually ransomware) that has encrypted data and host systems, rendering the data unreadable and the system inoperable. It cannot be assumed that it is possible to unencrypt the data, even where a ransom is requested and paid, as some attacks are deliberately destructive. The ability to recover host images, business data, and other objects from the data vault solution may be the most effective mechanism of recovering systems and recent copies of data.

B) **Use Case 2: Data corruption events** defined as 'damage to, or errors in data that occur during writing, reading, storage, transmission, or processing and which introduce unintended changes to the original data'. Data corruption, when caused by malware, may result in data being overwritten with garbage code or data may be securely erased. Dependent on the specific impacts of the scenario, recovery from an alternate data source – such as a data vault solution – may be required. For this use case when data is "damaged" or has been manipulated having the data vaulted would not help, since the vaulted data would have backed up the "damaged" data. This is where one would need error detection and data integrity checks either via the application or via the backup product.

C) **Use Case 3: Data Integrity events**, where data is altered or added to causing unexpected results. Data alteration and addition can originate from an external or internal source and may take place over a long period, with altered data also being committed to conventional backups (and potentially the data vault solution). These events are less likely to cause service unavailability or to impact backup capabilities, as sustained access to the data is typically required, i.e. for the purposes of financial crime. While a data vault solution may form part of a recovery solution, it should not be considered of itself an archival solution; one implication of this data integrity use case is that firms need to balance the duration objects are held against the expected time between an attack occurring and it being detected and responded to.

D) **Use Case 4: All other versions of data**, including backups, in the production environment are either encrypted or unavailable. Backup attacks typically delete an organisation's backup infrastructure and storage snapshots before 'locking and encrypting file systems, preventing the recovery of backup data, thereby giving bad actors the leverage to coerce a company into paying ransom'[2]. Attackers may compromise backup management systems as part of an advanced longer-term attack, so that at the point that a firm becomes aware

[2] securityforum.org

of issues in production, they no longer have the mechanism to utilise traditional backups to recover service. Data may be corrupted before being moved offline and the backup infrastructure may also be locked and encrypted. Removing a firm's 'insurance policy' of backups is an increasingly common approach in sophisticated ransomware attacks as it significantly increases the likelihood that a firm has no option but to pay the ransom. Unless the data vault uses a completely different backup mechanism/tool it will be impacted also. The data would be intact but gaining access to it would the issue. One would need to rebuild the backup infrastructure solution to enable the restoration.