FUTUREWEI Technologies

Whitepaper

Cloud Data Protection Market Report

V1.0	March 2022

Trademarks and Permissions

Notice

The purchased products, services, and features are stipulated by the contract made between Huawei and the customer. All or part of the products, services, and features described in this document may not be within the purchase scope or the usage scope. Unless otherwise specified in the contract, all statements, information, and recommendations in this document are provided "AS IS" without warranties, guarantees, or representations of any kind, either express or implied.

The information in this document is subject to change without notice. Every effort has been made in the preparation of this document to ensure accuracy of the contents, but all statements, information, and recommendations in this document do not constitute a warranty of any kind, express or implied.

FUTUREWEI® TECHNOLOGIES, INC.

Boston Research Center

Address: 111 Speen Street, Suite 114
Framingham, MA 01701
United States of America

Website: http://www.futurewei.com/

Contents

1 Executive Summary 5
2 Data Protection 6
2.1 Overview 6
2.2 Data Protection Strategy 7
2.2.1 Components 7
2.2.2 Procedures 9
2.2.3 Technologies 9
2.3 Trends 10
3 Cloud Data Protection 12
3.1 Cloud Adoption 12
3.1.1 Market Trends 12
3.1.2 Hybrid Multicloud 13
3.2 Cloud Data Protection 14
3.2.1 Cloud Storage Protection 14
3.2.2 Challenges 15
3.2.3 Principles 16
3.2.4 Strategic Changes 17
3.3 PBBA in Transition 18
3.3.1 Market Trends 18
3.3.2 Hyperconverged Data Platform 18
3.3.3 Cloud Data Protection Use Cases 19
4 Solutions 20
4.1 Overview 20
4.1.1 Considerations 20
4.1.2 Legacy Solutions 20
4.2 New Solution 20
4.2.1 Architecture 20
4.2.2 Multicloud Data Protection 21
4.2.3 Benefits 21
4.3 Desired Features 22
4.3.1 Traditional Data Protection 22
4.3.2 Ecosystem 22

4.3.3 Cloud 22
4.3.4 Data Security 22
4.3.5 Data Privacy 23
4.3.6 Data Management 23
4.3.7 Data Reuse 23
5 Conclusions 25
6 Bibliography 26

1 EXECUTIVE SUMMARY

Today, hybrid and multicloud are becoming the new normal for enterprises. As organizations increasingly rely on more than one cloud provider to take advantages of specific features or pricing, the need to adopt a multi-cloud, hybrid data protection strategy will continue to grow as well.

There are several major trends we observed for the next few years:

1. Data protection spans three broad categories: traditional data protection (such as backup and restore), data security, and data privacy. Any data protection must include practices and technologies from all the categories.
2. A data protection strategy is the organized process that defines all the measures implemented for the purpose of protecting data in an organization. A successful data protection strategy should include core elements such as data lifecycle management, data backup and recovery, data access management, data storage management, standards and regulatory compliance, CIA (Confidentiality, Integrity and Availability), policies and procedures, etc.
3. Hybrid and multicloud are becoming the new normal for enterprises. Recent research shows, 89% of organizations have a multi-cloud strategy, and most (80%) are taking a hybrid approach, combining the use of both public and private clouds.
4. Public cloud providers operate under a shared responsibility model. While they are responsible for the protection and availability of the cloud, it is still organizations' responsibility to protect resources in the cloud at the same level as on premises.
5. The biggest challenges for multicloud data protection come from data fragmentation, regulatory compliance, security inconsistency, and performance. Data protection strategies need to be enhanced in areas of compliance, security management, storage management and data lifecycle management, etc. These enhancements can be realized by data protection appliance with storage differentiators.
6. Hyperconverged data platform offers all-in-one solution of multi-purposes, multi-data services including multi-cloud capabilities. We discussed popular use cases with native cloud integration for backup & recovery, tiering, archiving, and replication to public cloud services.
7. A data protection solution is proposed for hybrid and multicloud environments. Hyperconverged data platforms are deployed both on premise and in cloud to:
    - Protect physical servers, virtual machines and cloud-hosted servers.
    - Implement data protection policies across public and private clouds and on-premise infrastructure with simplicity.
    - Discover, protect, and manage all of your data and applications on multiple clouds from a single management pane.
8. Hyperconverged data platform provides multiple data services in one platform, including features for traditional data protection, ecosystems, data security, data privacy, and add-on services.

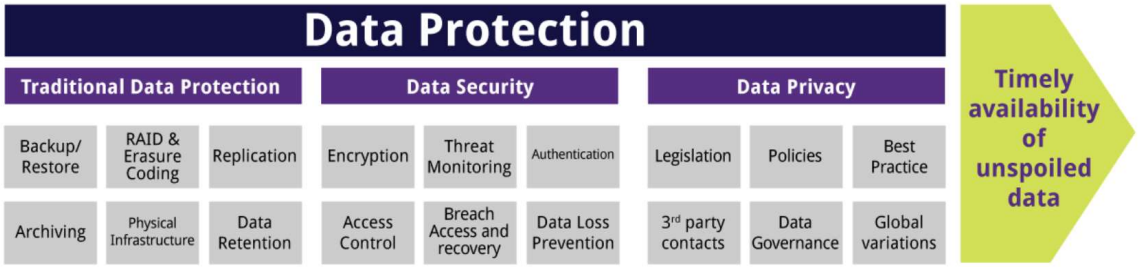
2 DATA PROTECTION

Across businesses of all sizes, data protection is becoming increasingly challenging. The exponential growth of data, the distribution of mission-critical workloads across edge, core and cloud environments and the need to protect newer workloads such as containers, cloud-native and SaaS applications, is bringing many data protection technologies to their breaking point and putting many businesses at risk.

2.1 OVERVIEW

Data protection is the processes of safeguarding important data from corruption, compromise or loss and providing the capability to restore the data to a functional state should something happen to render the data inaccessible or unusable[1].

Data protection spans three broad categories, namely, traditional data protection (such as backup and restore copies), data security, and data privacy as shown in the Figure below. The processes and technologies used to protect and secure data can be considered as data protection mechanisms and practices to achieve the overall goal of continual availability, and immutability, of critical business data.



Data protection is defined by the European Union (EU) in a very different way and is often used where other regions may use the term data privacy.

Data privacy concerns the proper handling of sensitive data (including personal data and other confidential data, such as certain financial data and intellectual property data) to meet regulatory requirements as well as protecting data confidentiality and immutability. Organizations have to meet legal responsibilities about how they collect, store, and process personal data, and non-compliance could lead to a huge fine.

Data security is an important element in protecting the data from external and internal threats, as well as determining what data can be shared and with whom.

Storage security is a specialty area of security that is concerned with securing data storage systems, ecosystems, and the data that resides on these systems [7]. Storage security is mainly focused on the physical, technical and administrative controls, as well as the preventive, detective and corrective controls associated with storage systems and infrastructure.

2.2 Data Protection Strategy

A data protection strategy is an organized effort that includes all the measures implemented for the purpose of protecting data in the organization. A data protection strategy can help organizations standardize the security of sensitive data and corporate information, ensuring the privacy of customers and employees and the security of trade secrets.

2.2.1 Components

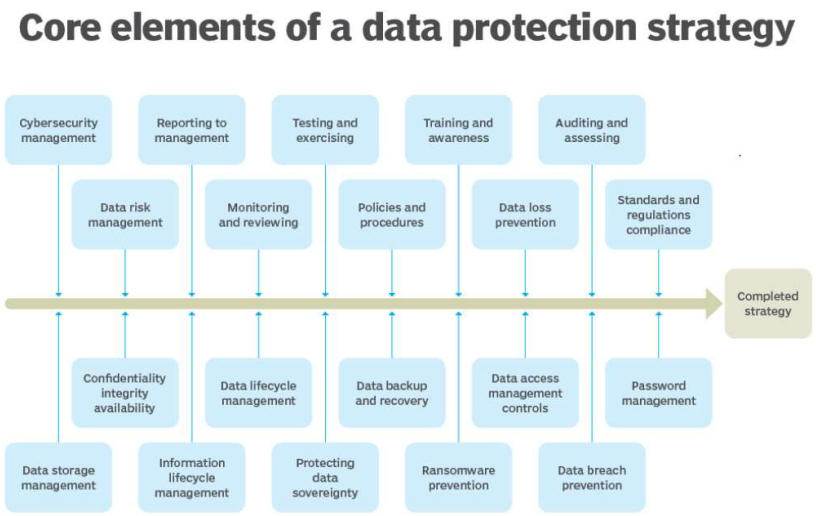
A successful data protection strategy will typically include the following components[2].

1. **Data Lifecycle Management**
   Data lifecycle management is a framework that standardizes data processes in the organization, from data creation, through storage, archiving, and until its final deletion. Data lifecycle management is a core component of a solid data protection strategy.
2. **Data Risk Management**
   In order to properly protect data, the organization must first identify and assess all risks and threats that may affect the data. A data protection strategy needs to take these risks and threats into account and include measures designed to minimize and mitigate these risks.
3. **Data Backup and Recovery**
   Backup and recovery measures are a critical component of data protection strategies. Once data is created, it should be backed up to ensure it can be recovered in case of failure.
   A data protection strategy should define which types of data should be backed up, how data should be recovered when disaster occurs, and which storage mediums should be used. All of these measures should also be included in business continuity and disaster recovery (BCDR) initiatives.
4. **Data Access Management Controls**
   Access controls are a crucial aspect that should be properly implemented and maintained. These controls ensure that only authorized users gain access to company data and systems, and can prevent unauthorized access, use, or transfer of data. This aspect is often examined and assessed by external auditors.
5. **Data Storage Management**
   Data storage management includes tasks related to securely moving production data into data stores, either on premises or in external cloud environments. These may be data stores intended for frequent, high performance access, or archival storage intended for infrequent access.
6. **Data Breach Prevention**
   Data breach prevention measures are implemented for the purpose of preventing unauthorized access to data. The goal is to prevent external malicious actors or internal threats from gaining unauthorized access to data and systems.
   Cyber security measures are put in place for the purpose of preventing attacks on internal networks, network perimeters, data-in-transit, and data-at-rest. Typically, these measures include data encryption, implementation of antivirus software, protection against ransomware, perimeter security hardware and software, and access management software.

7. **Confidentiality, Integrity and Availability**
   Confidentiality, integrity and availability (CIA), also known as the CIA triad, are key components that must be maintained in order to ensure the protection of data. Confidentiality is a set of rules that limits access to information, integrity is the assurance that the information is trustworthy and accurate, and availability is a guarantee of reliable access to the information by authorized people.
8. **Data Protection Policies and Procedures**
   Policies define what data protection activities the organization uses and procedures define how these activities are implemented. Policies and procedures are both essential components of a data management program, and must be well-documented, because they are usually examined during an audit process.
9. **Standards and Regulatory Compliance**
   Industry standards help organizations maintain adequate and current data protection. Regulatory compliance agencies define measures designed to protect data, which organizations are obliged by law to comply with. Each regulation is relevant to certain businesses, industries, and locations.
   Perhaps the most commonly known regulation is the General Data Protection Regulation (GDPR) of the European Union (EU), which protects the private data of individual EU citizens. However, there are many other regulations and industry standards that organizations are required to comply with. A data protection strategy must account for all relevant regulations and define how the organization maintains compliance.
10. **Monitoring and Reviewing**
    Monitoring and reviewing processes help organizations gain visibility into data activities, risks and controls, helping improve protection and respond to threats and anomalies. Monitoring and reviews may also be necessary to meet compliance requirements.
    Ongoing monitoring provides visibility into all aspects of the data lifecycle, including data creation, storage, transmission, archiving, and destruction. These activities offer essential evidence for internal and external auditors that examine controls set in place for data protection and management.

Other components include data loss prevention, data sovereignty protection, information lifecycle management, cybersecurity management, ransomware protection, password management, reporting to management, testing and exercising, training and awareness, auditing and assessing [8].

Once a strategy has been established, its development and implementation will occur based on the company's business needs, available funding, staffing and senior management's commitment to a secure data infrastructure.



2.2.2 Procedures

Enterprise data protection refers to the process of delivering, managing and monitoring security across all data repositories and objects within an organization. It is a broad term that is inclusive of several tools, policies, techniques and frameworks to ensure the safety of data.

Data protection strategies typically involve multi-step processes that define how security measures are implemented and maintained. The goal is to minimize the footprint of sensitive data and secure business-critical and regulated data.

1. Audit data: make a thorough audit of your current security systems to find vulnerabilities, and classify all data to determine the sensitive data location and access levels.
2. Assess internal and external risks.
3. Define a data protect policy: based on the analysis of data assets and most relevant threats, develop a data protection policy that determines the tolerance of risk for every data category.
4. Define security strategy: take measures to prevent threat from accessing data, ensure productivity and availability.
5. Define a compliance strategy: consider the regulations affecting data protection, such as GDPR, FTCA, and HIPAA.

2.2.3 Technologies

The data storage industry looks at data protection mainly from a technology viewpoint to keep data secure and available. Here are some of the most commonly used practices and technologies for data protection strategy:

1. **Data discovery & classification**—a first step in data protection, this involves discovering which data sets exist in the organization, which of them are business critical and which contains sensitive data that might be subject to compliance regulations.

2. **Data loss prevention (DLP)**—a set of strategies and tools that you can use to prevent data from being stolen, lost, or accidentally deleted. Data loss prevention solutions often include several tools to protect against and recover from data loss.
3. **Storage with built-in data protection**—modern storage equipment provides built-in disk clustering and redundancy.
4. **Backup**—creates copies of data and stores them separately, making it possible to restore the data later in case of loss or modification. Backups are a critical strategy for ensuring business continuity when original data is lost, destroyed, or damaged, either accidentally or maliciously.
5. **Snapshots**—a snapshot is similar to a backup, but it is a complete image of a protected system, including data and system files. A snapshot can be used to restore an entire system to a specific point in time.
6. **Replication**—a technique for copying data on an ongoing basis from a protected system to another location. This provides a living, up-to-date copy of the data, allowing not only recovery but also immediate failover to the copy if the primary system goes down.
7. **Firewalls**—utilities that enable you to monitor and filter network traffic. You can use firewalls to ensure that only authorized users are allowed to access or transfer data.
8. **Authentication and authorization**—controls that help you verify credentials and assure that user privileges are applied correctly. These measures are typically used as part of an identity and access management (IAM) solution and in combination with role-based access controls (RBAC).
9. **Encryption**—alters data content according to an algorithm that can only be reversed with the right encryption key. Encryption protects your data from unauthorized access even if data is stolen by making it unreadable.
10. **Endpoint protection**—protects gateways to your network, including ports, routers, and connected devices. Endpoint protection software typically enables you to monitor your network perimeter and to filter traffic as needed.
11. **Data erasure**—limits liability by deleting data that is no longer needed. This can be done after data is processed and analyzed or periodically when data is no longer relevant. Erasing unnecessary data is a requirement of many compliance regulations, such as GDPR.
12. **Disaster recovery**—a set of practices and technologies that determine how an organization deals with a disaster, such as a cyber attack, natural disaster, or large-scale equipment failure. The disaster recovery process typically involves setting up a remote disaster recovery site with copies of protected systems, and switching operations to those systems in case of disaster.

2.3 Trends

Organizations must constantly monitor the changing data landscape and be alert to new challenges and technologies. They must be aware of ever-evolving privacy regulations and security threats, which may appear from anywhere around the globe. At the same time, they also need qualitative improvements, e.g., reducing data loss, reducing down time, improving reliability.

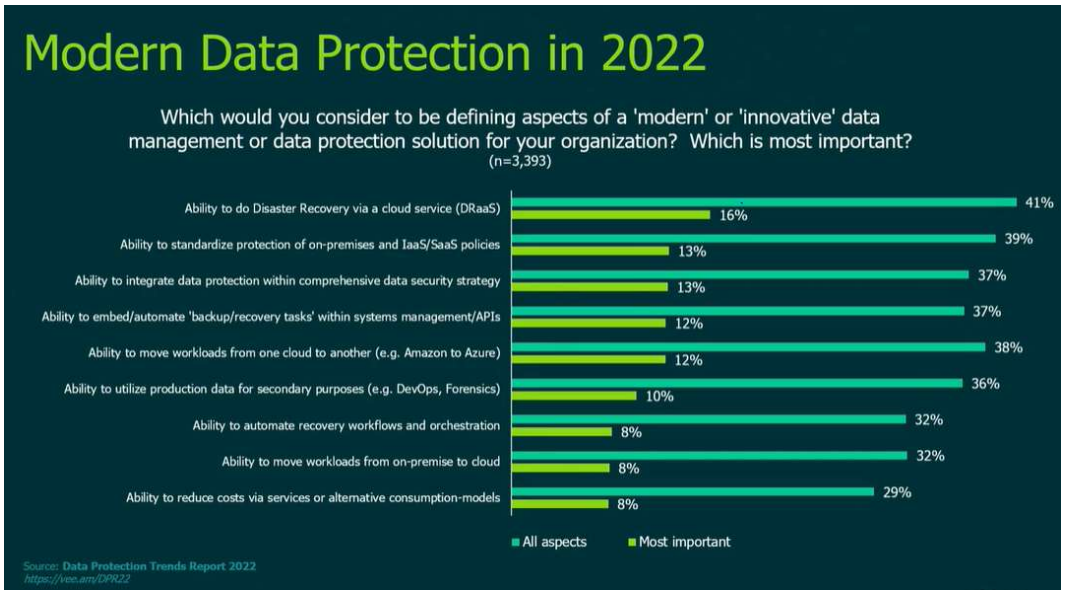
Here are the emerging trends that will shape the way companies approach data protection and management in the year ahead [3].

First, multicloud and hybrid cloud becomes the first concern of data protection. In addition to cloud disaster recovery and data protection for IaaS/PaaS applications, organizations also need the ability to move workloads between clouds, which means a third-party solution is needed for this requirement.

Second, hyberconvergence: organizations want the data protection solutions to be integrated into the security strategy and maintained by the same management system.

The third trend is the automation of workflows and orchestration, and data reuse.

Moving workloads from on-premise to cloud, and alternative consumption models still remains in trend, but not as hot as several years ago.



In the next section, we will discuss the first trend – hybrid and multicloud – for data protection.

3 CLOUD DATA PROTECTION

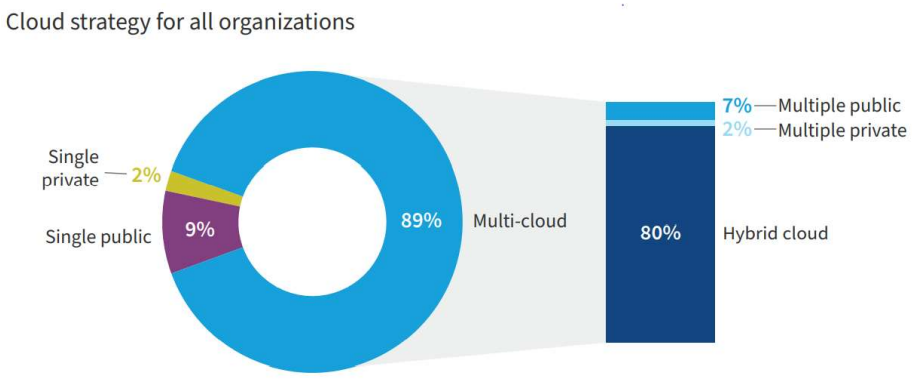
Data protection becomes more complex as organizations adopt hybrid and multicloud environments. In this chapter, we first probe into the market trends of cloud adoption. Then we discuss the challenges and strategy changes for cloud data protection. At last, we discuss PBBA market trends and use cases in hybrid multicloud environments.

3.1 CLOUD ADOPTION

3.1.1 Market Trends

According to ESG reports, 49% of enterprises surveyed now run production applications as their primary public cloud use case, closely followed by business intelligence queries and test/dev. Data protection fell to the number four spot in this survey, with just 40% of respondents saying they were currently using the cloud for backup or archival. This change points to a dramatic increase in the use of the cloud as an application workload destination.

A Flexera survey [4] shows, nearly all organizations have employed multi-cloud. 89% of organizations have a multi-cloud strategy, and most (80%) are taking a hybrid approach, combining the use of both public and private clouds. Private cloud plays an important role in this hybrid approach.

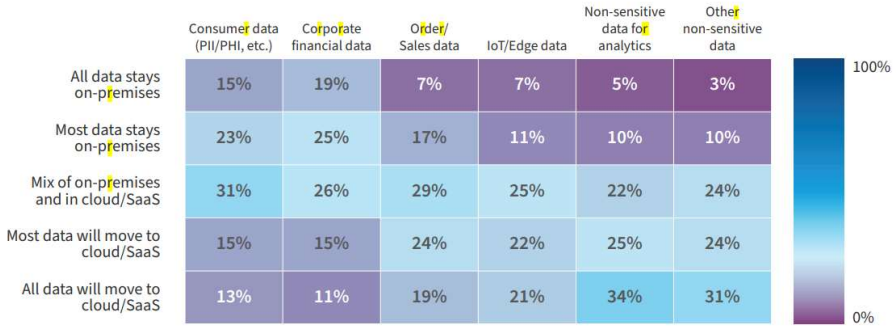


Enterprises often end up with a multi-cloud strategy designed to get the best price performance from each workload. Today, a hybrid, multi-cloud IT environment is the norm, not the exception.

SMBs (small to midsized businesses) are moving quickly to the public cloud, with 63% of workloads will reside in a public cloud within the next twelve months.

In the past, some organizations hesitated to put certain data in public clouds. The new survey found that many respondents are reconsidering. More than half of them said they will consider moving at least some of the sensitive consumer data or corporate financial data to the cloud.

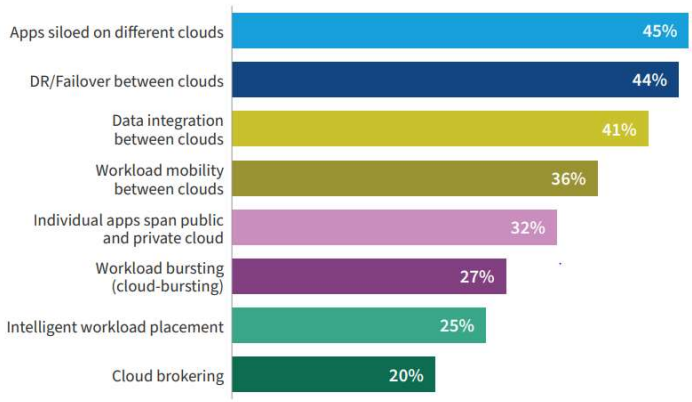
Data in the cloud



Organizations continue to increase their cloud spend, although the rate of growth is slowing. The public cloud spend was over budget by an average of 13 percent, down from 24 percent YoY. This could be an indication that some organizations have already deployed high priority applications to the cloud and will continue to expand their cloud footprint but **at a less accelerated rate**.

Applications are often siloed on different clouds: individual applications are not spanning clouds, as the following figure shows. DR/Failover saw a ten percent gain YoY and is now at 44 percent [4].

Use of multi-cloud architectures by all organizations



Container use becomes increasingly mainstream. The use of Docker and Kubernetes continues to be considerable. At the same time, container-as-a-service (CaaS) offerings from the public cloud providers continue to gain traction with customers; this year, AWS® Elastic Container Service® (ECS) and AWS® Elastic Kubernetes Service® (EKS) surpassed Docker for the top spot. However, Docker and Kubernetes remain near the top of the list of container tools, each used by over 40 percent of respondents.

3.1.2 Hybrid Multicloud

From previous section, 89% of organizations have a multi-cloud strategy, and most (80%) are taking a hybrid approach. The hybrid multi-cloud environment seems to be

most popular among all the organizations. Organizations choose the services from each cloud provider according to costs, technical requirements, geographic availability, and other factors.

The advantages of a hybrid multicloud strategy include:

- Avoiding vendor lock-in: Cloud vendor lock-in can be avoided by migrating data from public cloud to public cloud to optimize application service quality. Keep in mind that workload portability across heterogeneous clouds can be expensive given the transfer costs posed by public cloud providers.
- Increased redundancy: multicloud environments ensure that organizations always have compute resources and data storage available.
- Flexibility: different applications have different workload needs. For an individual application, a suitable cloud provider can be used.
- Security: organization can store and maintain sensitive data privately to satisfy data security and privacy obligations.
- Cost-performance optimization: achieve best performance within the budget.

From recent reports[4], the top three cloud challenges are security (85%), lack of resources/expertise (83%) and managing cloud expand (81%). Other challenges, such as Governance (77%) and Compliance (76%) are also significant obstacles to overcome, as cloud estates expand.

3.2 Cloud Data Protection

Cloud data protection is a data protection model that focuses on protecting data that is stored, manipulated, and managed in a cloud environment. It requires many data policies, strategies, and solutions to work together. Cloud data protection practices have become key aspects of data security as companies increase the amount of data stored in the cloud.

3.2.1 Cloud Storage Protection

Both proprietary and standards-based, cloud storage offerings commonly provide data protection capabilities such as copy capabilities (e.g., mirror some or all the storage on a system), backup and recovery capabilities, long-term retention capabilities (e.g., archives), and multi-system synchronization capabilities.

ISO/IEC 27040 provides the following guidance for cloud storage [7]:

- Transport security, such as IPsec or Transport Layer Security (TLS), should be used for all transactions.
- Data at rest encryption and appropriate key management processes should be used to prevent access by unauthorized parties (e.g., cloud service provider personnel, other tenants, adversaries, etc.) when sensitive data is stored in a third-party cloud environment
- User registrations should be handled securely, and strong password authentication should be used to protect access to data

- Access controls that guard against unauthorized access from other tenants while providing appropriate access privileges to users permitted to access the data should be used
- Sanitization capabilities should be used to clear sensitive data from the cloud computing storage ISO/IEC 27040 provides additional, specific guidance for SNIA Cloud Data Management Interface (CDMI)18 implementations and use.

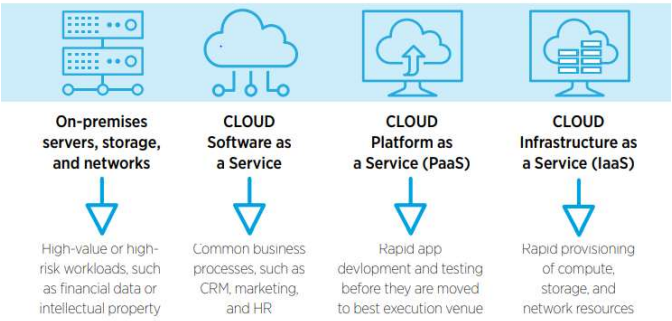
3.2.2 Challenges

Public cloud providers operate under a shared responsibility model, which means while they are responsible for the protection and availability of the cloud, it is still your responsibility to protect your resources in the cloud. The bottom line is: Everything you previously protect on-premises, you need to protect in the cloud. You are still responsible for backup, RPO/RTO, access control, and other data protection.

Data protection becomes more complex as companies adopt cloud services and storage, especially when doing so with multiple vendor solutions.

1. Data fragmentation

The biggest challenge to data management and protection in hybrid and multicloud environments is data fragmentation. Today, a preferred architecture has evolved in most organizations that looks like this: [5]



Each cloud infrastructure, as well as on-premises, is siloed. These siloed platforms each have their own tools and processes for data management and protection. The way you protect data on AWS® is drastically different from how you protect data on Microsoft Azure®, for instance, in terms of operator experience, policy model, capabilities, and limitations. Some of the processes can be rudimentary, and involve manual tasks. It can make governance, security, and compliance much more complex and inefficient.

2. Regulatory compliance

Whether your workloads are in the public cloud, a private cloud, a hybrid infrastructure or a multicloud with several cloud providers, you need to comply with data regulations and ensure the security of your data. Non-compliance with data regulations and a subsequent breach can lead to monetary losses and brand damage.

Data sovereignty refers to digital data that is subject to the laws of the country in which it is located. The increasing adoption of cloud data services and a perceived lack of security has led many countries to introduce new legislations that require data to be limited within the originating countries. Ensuring that data exists only in the host country can be complex and often relies on SLA provided by the Cloud Service Provider. Data protection should be designed in a way that clearly defines data location at all times.

3. Security Inconsistency

If companies are using multiple cloud providers or hybrid infrastructures, security may be inconsistent. There may be multiple vendors involved in a single step. Cloud infrastructure is entirely controlled by the vendors. This means organizations must rely on vendors to ensure that physical infrastructure, networks, and data centers are secure. Ensuring that data protection strategy dovetails with each cloud provider is critical.

4. Performance

While those cloud-born applications still need the same data protection workflows as when they were on-premises, the native protection services offered by public cloud providers often can't deliver application-consistent restores or meet stringent service level agreements (SLAs), and may not deliver the recovery point objectives and recovery time objectives (RPO/RTO) to sustain business functionality when an outage or data loss occurs.

3.2.3 Principles

What must enterprises consider when deciding on data protection for cloud workloads?

- Ubiquity—the support for every major public cloud provider as well as for on-premises workloads is essential. A single, comprehensive solution is preferred for simplicity's sake—eliminating the need for multiple point products makes both backup admins and line of business users more efficient.
- Comprehensive - a comprehensive data protection solution should offer near real-time replication to help decrease RPO and RTO, yielding maximum uptime and minimum data loss across the enterprise.
- Efficiency - Data protection utilizes three public cloud consumables, namely CPU resources, block storage and object storage. An analysis of different multi-cloud offerings should consider how each candidate utilizes these resources, not only at current usage levels, but also as the number of workloads and projected storage utilization scale over time with changing business demands. Efficient deduplication reduces the amount of data moved from site-to-site and reduces resources needed to store that data. In a multi-cloud world, less data means lower costs and faster replications or restores.
- Scalability - Data growth means that both the number of workloads and the amount of storage are constantly growing. For that matter, enterprises must ensure that the chosen solution provides the scalability that supports tens or hundreds of terabytes or more without breaking the budget.
- Simplicity - While simplicity begins with deploying a single solution, that single, comprehensive solution must offer additional functionality and features.

- Cost - Since pricing for cloud storage—both object and file—varies by provider over time, data protection solutions should offer a tiered architecture that enables data and workloads to move to one or more clouds for archival or disaster recovery, and should easily (and cost effectively) be migrated back to on-premises when desired.

3.2.4 Strategic Changes

As nearly all organizations have deployed multi-cloud and hybrid infrastructure, data protection strategies need to be changed as well. Here are examples of components of a data protection strategy that will be affected by cloud involvement:

- A data protection strategy needs to accommodate physical servers, virtual machines (within a virtualization host) and multiple cloud-hosted servers (within MSP or a hyper scale cloud) options. Enterprise IT organizations are hybrid, multicloud environments, with approximately 50/50 between on-premise servers and cloud-hosted servers.

- Regulatory compliance & Data sovereignty
  Data are subject to the laws and governance structures within the nation it is collected. For global organizations dealing with data from other countries, it's important to have measures in place to protect data sovereignty so it remains undamaged by internal or external attacks.
  When much of the data resides in the cloud (a globally distributed data infrastructure), the organizations must keep track of sovereignty issues in different jurisdictions. Data protection providers will have to work more closely with their customers to manage sovereignty and compliance with varying rules.

- Cybersecurity Management
  Cybersecurity management is an organization’s strategic-level capability to protect information resources in a complex and evolving threat landscape. Hybrid multicloud environment increases exposure to security threats, both on-premises and in clouds.

- Data storage management
  Data storage management includes tasks related to securely moving production data into data stores. When data stores include cloud, the tasks need to consider a lot of issues (cost, location, security measures, etc) which makes data storage management more complex. Data mobility is also an important issue to consider in multi-cloud environment.

- Data Lifecycle Management (DLM)
  DLM are based on specific conditions defined in advance by the user to organize data into different tiers and to automate data migration from one tier to the other. For example, more recent data is stored on faster storage media while older data is stored on less efficient media.

Cloud can participate in any stage from creation to deletion, which affect the techniques used to achieve the goal of CIA (Confidentiality, Integrity, Availability).

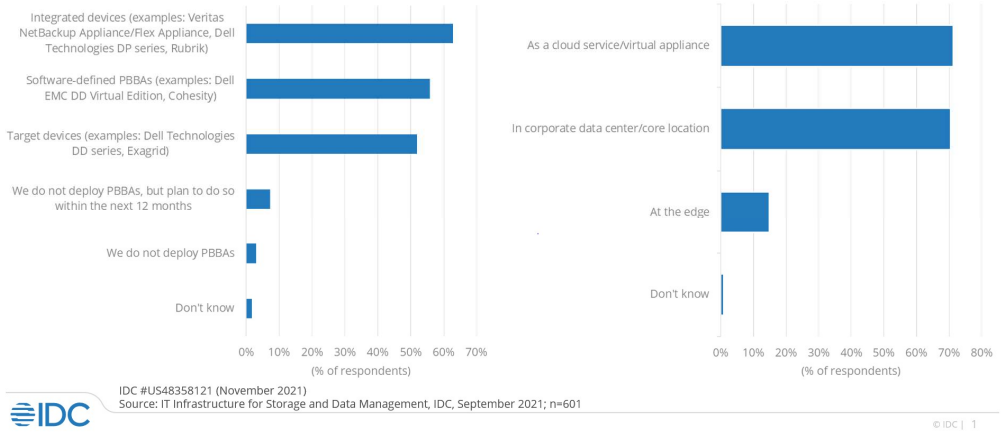
- Data Backup and Recovery / Disaster Recovery
  Traditional storage protection techniques, such as backup & recovery, replication, and archiving, should consider cloud both as source and destination. We will discuss the cases later.

3.3 PBBA IN TRANSITION

Purpose-built backup appliances (PBBAs) have been foundational components to many data protection infrastructure for more than a decade. Over the past years, we have experienced tremendous innovations in data protection technology, with each innovation made huge advantages in PBBA market.

3.3.1 Market Trends

Hybrid multicloud adoption is the most important trend shaping the PBBA market in recent years. IDC survey shows the market appears to be in the midst of a transition, with growth coming from integrated appliances and software-defined (virtual) instances in cloud environments [6].



From the figure above, integrated appliances surpass target devices in 2021. Then, virtual appliances as a cloud service are used at the same percentage (70%) as physical appliances in data centers. Cloud data protection is not an expectation. It is happening fast to enterprises.

3.3.2 Hyperconverged Data Platform

Hyperconverged data platform offers all-in-one solution of multi-purposes, multi-data services including multi-cloud capabilities.

Hyperconverged data platform solves problems caused by **data fragmentation**. It consolidates all secondary storage that spans from the core to edge to cloud. It also

eliminates complexity with a unified platform for end-to-end solution and unified management platform.

Hyperconverged data platform can be used to protect physical servers, virtual machines (within a virtualization host) and multiple cloud-hosted servers (within MSP or a hyper scale cloud). This platform applies unified policies and procedures to data on premise or in cloud, thus provent **security inconsistency** problems.

Regulatory compliance becomes more complex in hybrid and multicloud environments. For example, data protection should be designed in a way that clearly defines data location at all times for the purpose of data sovereignty. Hyperconverged data platform need to provide enhanced metadata indexing (data catalogs) with information such as file name, location, owner, dates, ACL, etc. Other capabilities, such as encryption, secure erasure, data masking, global search, should also be provided for compliance purpose.

Ensure fast recovery point and near-instantaneous recovery time is an objective for hyperconverged data platform. Data protection is further enhanced through an indexing engine that rapidly indexes all the data being backed up, and all associated metadata. This has the benefit of easily mining backup data with a simple text-based search and restore.

Hyperconverged data platform extends to all secondary storage use cases, starting with data protection and extending to data management, dev/test, data analysis and even file/object services. These extended use cases may create substantial additional value for organizations and take full advantage of PBBAs.

3.3.3 Cloud Data Protection Use Cases

In this section, we discuss some popular practices that are developed for data protection in multicloud, hybrid deployment environment.

For Cloud-Native Applications:

- **Cloud-Native backup & recovery** is an agentless solution minimizing operational overhead for AWS®, Azure® and other public cloud security and management.
- **Cloud-Native archival** archives cloud-native data to public cloud provider's storage.

For Hybrid Cloud Applications:

- **Backup to cloud:** use public cloud as the backup target.
- **Recovery to cloud** uses the cloud to recover applications from on-prem or cloud.
- **Replication** on-prem to cloud or cross cloud: replicating the latest data to an appliance in the cloud to provide DR and off-site data protection.
- **Tiering**: offload cold data into cost-effective cloud storage. It decreases the amount of local storage while keeping the performance of on-premise storage.
- **Archival**: send application data to the cloud for long-term retention
- **Migrating** test/dev to cloud: migrate existing on-prem applications to the cloud for test or development tasks.

4 SOLUTIONS

In this chapter, we propose a data protection solution for hybrid and multicloud environment. Then we list the features that are desired for the solutions.

4.1 OVERVIEW

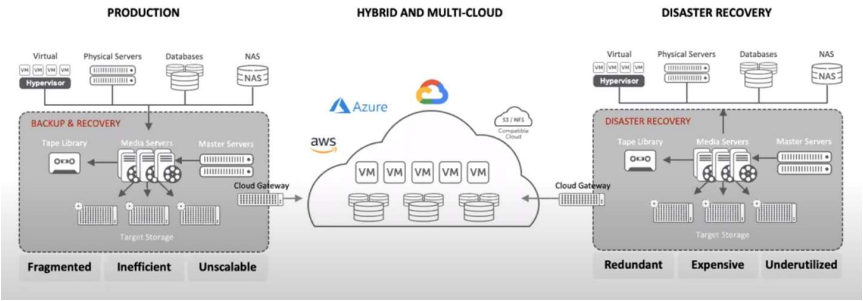
4.1.1 Considerations

Data is growing exponentially and fragmented and duplicated across data centers, edges, and public clouds. Data protection could be several different software or hardware-based solutions for VMs, containers, physical servers, and cloud-based data. These solutions may come from different vendors and medias. These siloed point solutions caused a lot of problem such as resource waste, complex architecture and management, and low performance. We need to come up with a comprehensive solution to consolidating all data and workload under a single platform.

4.1.2 Legacy Solutions

Today's legacy data protection solutions are usually complex software and hardware combinations on production sites and duplicated sets on DR sites. With public clouds, the production environments are replicated to cloud with increased complexity and cost.

As the figure shows [3]: each site typically has backup software, target storage, master servers, media servers, and proxies, replication. Most solutions have cloud gateways and cloud storage is used as a backup target.



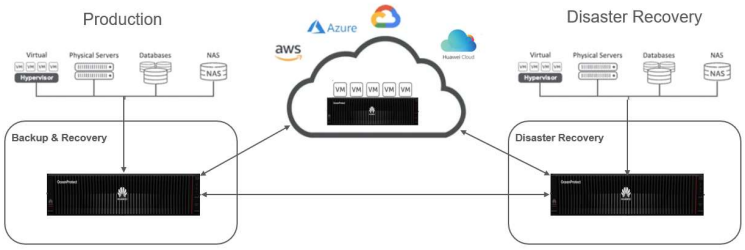
4.2 NEW SOLUTION

4.2.1 Architecture

Our new solution starts with the hyperconverged data platform (integrated appliance) with storage consolidation and pay-as-you-grow scalability. It integrates backup software, target storage, master servers, media servers, proxy servers, replication, and deduplication into one platform. This simplifies architecture, increases efficiency and reduces cost. This platform should be software defined so that it could be deployment as virtual appliance on cloud as well.

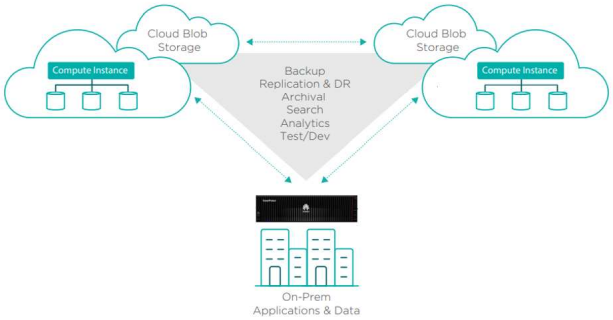
It needs to be managed by a single management interface in order to drive operational efficiency. Scalability, performance and NDU are other requirements for this hyperconverged data platform.

Native multicloud and hybrid deployment need to be supported for organizations to implement cloud data protection strategies.



4.2.2 Multicloud Data Protection

Organization can use multiple public cloud for different workloads. To protect applications in clouds, they need to deploy a **software instance/virtual appliance** in the public cloud to orchestrate all critical data management and protection functions. The following figure shows hybrid deployment with two public clouds.



This solution allows you to implement data protection policies across public and private clouds and on-premises infrastructure with ease. It allows you to discover, protect, and manage all your data and applications on multiple clouds from a single management pane. It can work in hybrid and multicloud environments and integrate with AWS®, Azure®, and Google Cloud Platform® to protect workloads.

4.2.3 Benefits

The new solution provides the following benefits for cloud data protection:

- End-to-end protection
- Reduced complexity and cost with hyperconverged data platform to replace all the hardware platforms and software.
- Storage efficiency by managing copies and data reduction features.
- Leverage public cloud services, such as analytic services.

- Performance improvement provided by comprehensive data platform with defined RPO and RTO.
- Multi-Purpose with add-on services.

4.3 Desired Features

Hyperconverged data platform provides multiple data services in one platform. In this section, we list a set of features for data protection that is deployed in hybrid and multicloud environments.

4.3.1 Traditional Data Protection

- Backup & recovery: Incremental backups - such as CBT (changed block tracking) based backups; parallelized backup; instant mass recovery at scale; incremental restore for standby for rapid recovery of large datasets; file-level restore.
- Deduplication & compression: Global variable-length deduplication.
- RAID & Erasure coding.
- Snapshot, CDP.
- Replication/DR: synchronous & asynchronous replication, 3DC.
- Data retention: policy-based; short to medium-term (less than 10 years) and long-term.
- Archiving: to tape and to cloud.
- Instant access via instant mount.
- Scale-out ingest, such as ingesting large VMDKs in parallel.
- Air-gapped vault.
- Immutable file system: Datalock, WORM.

4.3.2 Ecosystem

- OS: Hyper-v®, vSphere®, KVM®, Redhat®, solaris, linux, windows.
- Apps: Oracle®, Exchange®, SQL, SAP HANA®, NoSQL, mongoDB®, Office 365, Hadoop and distributed databases, SaaS-based applications.
- New workloads: containerized applications, Kubernetes, OpenStack®, VMware Tanzu®, OpenShift®.
- Storage: major primary storage vendors.
- Public cloud IaaS/PaaS backup.

4.3.3 Cloud

- Hybrid cloud: Cloud backup, Cloud recovery, Cloud tiering, Cloud replication, Cloud archival.
- Cloud-native: cloud-native backup & recovery, cloud-native archival.
- Cloud platforms: Google® Cloud Platform (GCP), Microsoft® Azure® and Amazon® Web Services (AWS).

4.3.4 Data Security

- Authentication & Authorization: Granular RBAC, multi-factor authentication.

- Encryption: at-rest (SED, software encryption), in-transit, encryption dedup, KMS, kmip.
- Data disposition/Sanitization: crypto erasure.
- Transport security: IPSec, TLS.
- Ransomware protection: Immutable copies, ransomware detection, recovery vault, sandbox, advanced technologies (machine learning) for detection and protection.
- Third-party antivirus scan.
- Data masking: obfuscate sensitive data.
- Data loss prevention (DLP): inspect data in transit, at rest on-premise or in cloud
- Monitoring: monitor data access and activities to identify excessive, inappropriate and unused privileges. This help organizations gain visibility into all aspects of the data lifecycle. These activities offer essential evidence for internal and external auditors that examine controls set in place for data protection and management.
- Security analysis: AI and machine learning technology can be used to analyze data access behavior and detect and alert on abnormal and potential risky activities.
- Alert prioritization: look acoss the stream of security events and prioritize.

4.3.5 Data Privacy

- Data discovery and classification: reveals the location, volume and context of data on-premise and in the cloud.
- Regulatory compliance.
- Risk assessment and management.
- Data retention: policy-based.
- Data sovereignty: metadata with data location and other information.
- Audit log: protocol access log, management log.
- Incident report: automation.

4.3.6 Data Management

- Centralized data management: unified management pane for multiple cloud platforms.
- Metadata indexing: includes file name, size, location, owner, dates, ACL, etc.
- Data discovery and classification reveals the location, volume, and context of data on-premises and in the cloud.
- Automated Orchestration: Advanced orchestration tools allow the automation of the entire recovery process (including disaster recovery).
- Load balancing: software load balancer delivers elastic application services between cloud platforms.
- Migration: enables workloads to be moved more efficiently across various cloud services.

4.3.7 Data Reuse

New use cases are constantly emerging. These may extend to all secondary storage use cases, starting with data protection and extending to data management, dev/test, data

analysis and even file/object services. These extended use cases may create substantial additional value for organizations and take full advantage of the related products.

- Global search, wild card search, across multiple workloads.
- App run: such as Cohesity® Marketplace (download and run apps directly on the platform, from Cohesity® and 3rd party ecosystem, SDK for easy custom app development).
- Consumption and compliance analytics: provide insights into data management, compliance reporting, and capacity planning across cloud environments.
- Data analytics: deploy third-party applications, such as Splunk Enterprise Analytics®, SentinelOne® and Clam® anti-virus, plus Imanis® for Hadoop and NoSQL database protection.
- Dev/test environment: live mount, sandbox.

5 CONCLUSIONS

We analyzed multicloud market. Nearly all organizations have employed multi-cloud. 89% of organizations have a multi-cloud strategy, and most (80%) are taking a hybrid approach, combining the use of both public and private clouds.

Data protection becomes more complex as organizations adopt hybrid and multicloud environments. The biggest challenges come from data fragmentation, regulatory compliance and protection inconsistency. Data protection strategies need to be enhanced to adapt to this trend.

We discuss trends of the PBBA (Purpose-built backup appliance) and popular use cases that are developed for in multicloud, hybrid deployment environment.

We presented a solution for hybrid multicloud data protection, with hyperconverged data platform deployed both on premise and in clouds. We proposed a list of features that need to be implemented for hyperconverged data platform, including features for traditional data protection, data security, data privacy, and add-on services.

6 BIBLIOGRAPHY

[1] SNIA, [Online]. Available: https://www.snia.org/education/what-is-data-protection.

[2] [Online]. Available: https://cloudian.com/guides/data-protection/data-protection-strategy-10-components-of-an-effective-strategy/.

[3] Veeam, "2022 Data Protection Trends Report", https://vee.am/DPR22.

[4] Flexera, "State of the Cloud Report".

[5] Rubrik, "Protecting Hybrid and Multicloud Data".

[6] IDC, "Purpose-Built Backup Appliances: 3Q21 Trends".

[7] SNIA, "Storage Security: Data Protection".

[8] [Online]. Available: https://www.techtarget.com/searchdatabackup/tip/20-keys-to-a-successful-enterprise-data-protection-strategy.

[9] Cohesity, "Cohesity Data Protection White Paper".