FUTUREWEI Technologies

Whitepaper

Backup & Archiving Market Report

V1.0	Feb 2022

Copyright © 2021, Futurewei® Technologies, Inc. All rights reserved.

No part of this document may be reproduced or transmitted in any form or by any means without prior written consent of Futurewei® Technologies.

Trademarks and Permissions

FUTUREWEI Technologies and other Futurewei® trademarks are trademarks of Futurewei® Technologies. Huawei trademarks are trademarks of Huawei Technologies Co., Ltd. All other trademarks and trade names mentioned in this document are the property of their respective holders.

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

As the primary storage market leaders are enjoying the steady, huge market size due to the growth of the data, the secondary storage market (refers to storage systems for data protection, backup, archiving and so on) has seen a significant market expansion. There are several major trends we observed for the next few years:

1. The secondary storage market covers i. backup & recovery market ii. archiving market, iii. object storage market and others. The on-premises offerings for the secondary storage market are developing toward comprehensive integrated appliances, i.e., hyper-converged data platforms that offers all-in-one solution of multi-purposes, multi-data services including multi-cloud capabilities.
2. The global data backup and recovery market is projected to reach US$19.96 Billion by 2028, from US$10.05 billion in 2020, at a CAGR of 9.0% during 2021-2027. The Purpose-Built Backup Appliance (PBBA) market will account for 1/3 of the total backup and recovery market.
3. The global enterprise information archiving (EIA) market was valued at USD 5.93 billion in 2020, and it is expected to reach a value of USD 12.49 billion by 2026, with a CAGR of 13.51% over the forecast period (2021-2026).
4. With the secondary storage market is evolving toward hyper-converged data platforms to unlock data value, our recommendations for storage manufacturers are to i. enhance ecosystems (Application/backup ISVs, OS, Hypervisors, Application platforms, etc.), ii. expand add-on data services and iii. support hybrid cloud integrations to transform the current product or solution offerings.
5. Add-on services in data protection, dev-test environments, data analytics and other areas enable extended modern data uses that can create substantial additional value for organizations.
6. New vendors like Cohesity® and Rubrik® with their hyper-converged data platforms are gaining customers (especially in small/midsize enterprises) in PBBA market.
7. Hybrid deployment mode across on-premises data center, cloud and edge is adapted by more enterprise customers. For security, compliance, user experience, data service and other considerations, on-premises will still play a dominant role in any modern data-driven businesses. Backup combined with archiving will become a common solution.

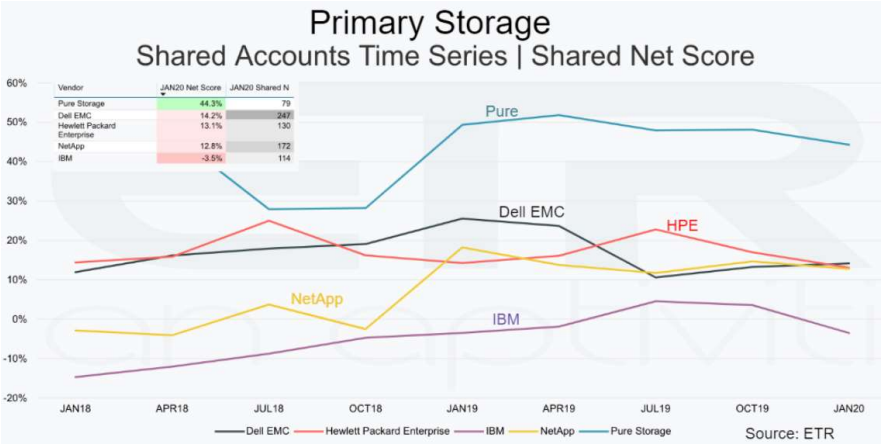
2 MARKET INSIGHT

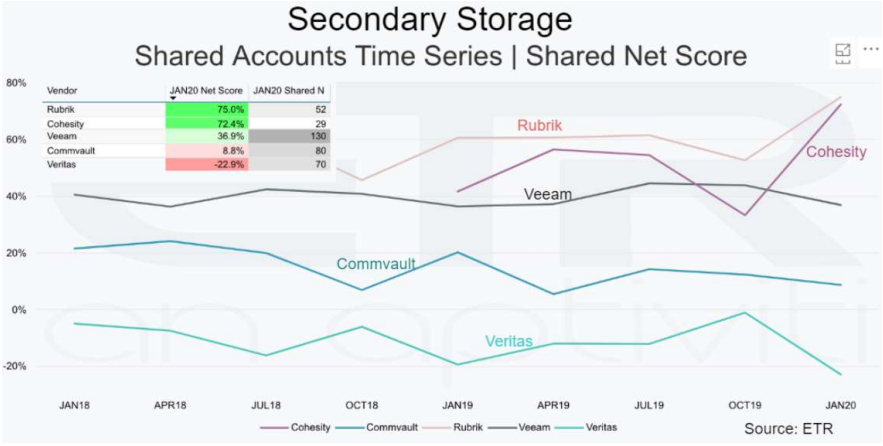
2.1 SECONDARY STORAGE MARKET

2.1.1 Overview

According to IDC, enterprises data is expected to grow to 93 Zettabytes by 2025. About 80% of that data is secondary data in the form of files, objects, and backups. Primary storage typically only comprises 20% of overall capacity. The remaining 80% consists of secondary storage that covers all your secondary use cases, including backup, archiving, file and object storage, dev/test, analytics, private and public cloud.

As the primary storage market leaders stablized from a spending perspective, the secondary storage market (refers to storage systems for data protection, backup and so on) has seen a big total addressable market (TAM) expansion, with companies like Cohesity® and Rubrik® expanding the notion of data protection from backup into data management, analytics and security.





Secondary storage infrastructure is fragmented across a patchwork of point appliances, including deduplication appliances, backup servers, cloud gateways, NAS, and data

lakes. This approach is complex: each needs to be provisioned, configured, managed and updated through its own proprietary UI and processes. Data must be copied and stored across the silos, with enterprises keeping an average of 10 to 15 redundant copies of data across silos. Enterprises need a simpler, more efficient way to manage their ever-increasing volumes of data.

2.1.2 Cloud Adoption

The number of organizations committed to or interested in a hybrid cloud solutions has increased from 81% to 93% since 2017. 89% of organizations still expect to have a meaning on-premises footprint in three years.

Over the past 10 years, storage evolved around two very different directions. Traditional enterprise storage, such as Dell® and HPE®, focused on providing standardized file interfaces (NFS and SMB), on 'scale-up' hardware, and snapshots for resiliency. On the other hand, cloud storage, developed by hyperscale companies like Google and Amazon, focused on delivering scale-out solutions on commodity hardware, strong resiliency to hardware failures, but relying on proprietary protocols and APIs for data access.

Cloud solutions is increasing rapidly for its agility, scalability and lower costs. The price for archiving could be as low as 0.1 cents per GB per month. Therefore, the vendors in the market are increasingly offering cloud-based solutions to manage data.

Cloud also provides advanced services, such as machine learning, analytics, and data pipelines, so that customers can use "in place" with their data in cloud.

Cloud solutions are one of the most comprehensive tools against cyber-attacks and data breaches. However, if mismanaged, it may pave the way for the attackers to seamlessly enter the database stored on the backup server and use it against the user. Hence, the privacy and security issues act as a major hindrance to the adoption of cloud backup solutions.

What enterprises need today is the best of both solutions. They need to support standardized interfaces like NFS and SMB protocols, to interoperate with existing applications. They need to save critical data "in house" for security and compliance reasons. But they have also need to move to software-defined, web-scale solutions on commodity hardware, just like cloud storage. Web-scale provides multiple advantages like 'pay-as-you-grow' consumption, always-on availability, simpler management, and lower costs.

2.2 Industry Trends

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].

2.2.1 Hybrid Cloud

Major enterprise storage vendors provide hybrid cloud solutions for users to take advantage of both enterprise storage and cloud. The data can be saved in local storage or cloud or both for different purposes. A virtual appliance can be run in cloud to help data mobility.

Major storage vendors such as Dell® supports copy data to cloud for different uses cases:

- Tiering supports cold data bursting to cloud. Cold data are automatically stored in the cloud. Once they become hot, they will be tiered back to private appliances.
- Cloud archiving enables long-term archival to the cloud, providing a more manageable alternative to tape.
- Cloud replication provides replication with virtual appliances running in the cloud.
- Recovery in the public cloud: Today, leading backup vendors support restoring backup data to servers in the public cloud. The backup data can also be used for test/development purposes in the public cloud.

Enterprises utilizes multiple public cloud services from different cloud provider. For example, backup solutions may support cloud integration for tiering, archival, and replication to public cloud services such as Google Cloud Storage Nearline, Microsoft Azure and Amazon S3/Glacier. Secondary storage vendors may need to support multi-cloud solutions for enterprises.

2.2.2 Comprehensive Appliances

Second, secondary storage infrastructure is fragmented across different point appliances and clouds, including deduplication appliances, backup servers, archiving appliance, cloud gateways, etc. The complexity also comes from multiple software and management tools. To reduce the complexity and improve efficiency, secondary storage vendors address this issue with comprehensive appliance for all the use cases. Cohesity® is among the vendors that first came up with the concept.

A comprehensive secondary storage appliance is intended to provide following benefits:

- Eliminate complexity with a unified platform for end-to-end solution and unified management platform.
- Reduce TCO by consolidating multiple solutions across backup, replication, and archival and manage copies of their data.
- Increase storage efficiency with deduplication and other data reduction techniques.
- Integrate seamlessly with all the leading cloud providers to leverage economics of public cloud services.
- Provide a platform for data reuse such as data mining and analytics.

2.3 What's Next

The secondary storage market covers all secondary use cases, including backup, archiving, file and object storage, dev/test, analytics, private and public cloud. In this report, we will focus on backup and archiving markets. Although these markets are more mature than other sections, they seem to be in midst of a transition with new technology boom.

3 BACKUP

In this chapter we first summarize market trends, and then discuss some common use cases for backup appliances. At last, we discuss the evolution of PBBAs and give out some recommendations.

3.1 MARKET TRENDS

3.1.1 Overview

The global market for data backup and Recovery is projected to reach US$19.96 Billion by 2028, from US$10.05 billion in 2020, at a CAGR of 9.0% during 2021-2027[3]. The key drivers include massive data boom, a shift to cloud storage and computing, focus on data protection and regulatory compliance, and other related technological advancements.

Data backup and recovery software is used mainly in the type of on-premises. The total value of PBBA (purpose built backup appliance) market in 2020 was $4.33 billion, about one third of the total backup hardware and software infrastructure by IDC. Cloud-based type is estimated to grow in a higher rate of over 18% than 7.2% of on-premises in 2019-2025.

The BFSI industry accounted for the largest market share of ~20% in the year 2018, attributed to the rapid adoption of digitalization in the sector, for payments, retailing, data storage, and other end uses.

USA is the largest region in the world in the past few years and it will keep increasing in the next few years. USA market took up about 31% the global market in 2018, while Europe was about 23%. Asia Pacific is forecasted to have the highest CAGR of 10.4% during 2021-2027[3].

Veeam®, Veritas Technologies®, Commvault®, Acronis®, Netapp®, Dell® etc. are the key suppliers in the global Data Backup and Recovery Software market. Top 10 took up above 55% of the global market in 2018.

3.1.2 Cloud Backup Market

The global Cloud Backup market size is projected to reach USD 4.2 billion by 2026, from USD 1.8 billion in 2019, at a CAGR of 12.5% during 2021-2026[4]. According to Mordor Intelligence, the cloud backup market is expected to grow with CAGR of approximately 24% during 2021-2026. Other researches also pointed out cloud backup market will increase at a much higher pace in the next several years.

The growth of the cloud backup market is driven mainly by:

- Advantages of easy management and monitoring, real-time backup and recovery, simple integration of cloud backup with other enterprise applications, data deduplication, and customer support.

- Rapid growth of volume and diversity of data generated in the organization: Business-critical data cannot be jeopardized, and it necessitates continuous backup and on-demand accessibility.
- In response to the growing use of cloud-based technologies & services such as SaaS.
- Great efficiency and reduced costs when compared to traditional on-premise backup systems.

Cloud backup solutions are one of the most comprehensive tools against cyber-attacks and data breaches. However, if mismanaged, it may pave the way for the attackers to seamlessly enter the database stored on the backup server and use it against the user. Hence, the privacy and security issues act as a major hindrance to the adoption of cloud backup solutions.

3.1.3 Focus Areas

The enterprise backup and archiving software market underwent significant transformation in the past several years. Customers mainly focused on following areas[1]:

- Centralized management: Workloads are distributed across the data center, public cloud and the edge. Leading backup vendors are offering a management platform that can be deployed either in the main data center or as a service hosted in the public cloud.
- Ransomware resilience, detection and remediation: While most vendors support the creation of immutable second copies of backup through write once, read many (WORM)-enabled storage, others such as IBM® and Rubrik® aim to make the primary backup repository more resilient by supporting immutable snapshots. Leading vendors are building capabilities to detect ransomware attacks by tracking large changes to file system data, and through other means, by partnering with security vendors or by developing these capabilities in-house. Most vendors also aim to simplify the ransomware recovery process through creation of an isolated test environment, and provide a clean backup copy to recover specific files. Such efforts remain largely a work in progress.
- Support for public cloud IaaS and PaaS backup: Leading on-premises backup vendors increased their investment toward building capabilities to protect cloud-native workloads, particularly VMs and applications hosted in AWS, Microsoft Azure and Google Cloud Platform. Some backup vendors are also supporting backup of DBaaS products such as Amazon RDS, Amazon Aurora and Microsoft Azure SQL.
- Support for SaaS-based applications: In the past two years, I&O leaders have begun to include SaaS applications such as Microsoft Office 365, Google G Suite and Salesforce as a part of their backup strategy.
- Tiering to the public cloud: The most commonly supported public cloud storage targets are Amazon Simple Storage Service (Amazon S3) and Azure Blob storage.
- Recovery in the public cloud: Today, leading backup vendors support restoring backup data to servers in the public cloud. The backup data can also be used for test/development purposes in the public cloud.

- NoSQL database backup: Established vendors such as Commvault, Dell Technologies® and Veritas Technologies® have started addressing these backup requirements by building such capabilities natively into the backup platform. Vendors such as Rubrik® and Cohesity® have made strategic acquisitions in this space.
- Instant recovery of databases and virtual machines: A majority of vendors support instant recovery of VMs by mounting the backed-up VM directly on the production host via NFS. VMs can thus become instantly available, while the actual recovery process can be initiated in the background. Similarly, vendors such as Cohesity® and Rubrik® offer instant recovery of databases such as Microsoft SQL and Oracle.
- Container backup: Leading vendors announced support for container backup either by building these capabilities natively into their existing platform or through acquisitions.
- Subscription licensing: Enterprises that are migrating to the public cloud find the subscription-based model a simpler way to procure backup solutions. While subscription-based licensing is not necessarily less expensive compared to perpetual licensing, it is more predictable and easier to manage.

3.2 USE CASES

3.2.1 Recovery Vault

Backup 3-2-1 rule requires 3 data copies (or 3 backup copies), two media types, and one offsite copy (vault, cloud, another site). Recovery vault provides multiple layers of protection against cyber attacks even from an insider threat. It removes critical data away from attack surface with:

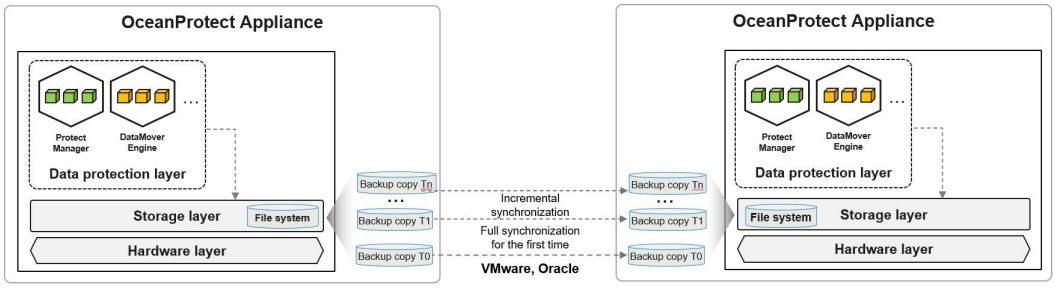
- Physically isolated within a protected part of the data centers;
- Separate security credentials;
- MFA authentications for access;
- Removal of management interface which could be compromised;
- Automated air gap for network isolation.

The recovery process synchronizes the data between production site and vault to create immutable copies with locked retention. If a cyber attack happens, it will quickly identify a clean copy of the data and recover the system.

Major vendors such as Dell®, Cohesity®, Veeam®, all provide similar vault solutions to customers.

3.2.2 Backup with Replication

With replication, backup data can be copied to a remote site or even cloud. Replication of data happens at the storage layer for remote data protection. It facilitates rapid recovery from remote site or cloud. This solution satisfies the needs in many scenarios, such service migration, service expansion and service DR.



3.2.3 Hybrid Cloud

Hybrid cloud solutions have the ability to replicate and retrieve backup data from and to at least one cloud storage provider. A secondary storage device used as a cloud gateway allows enterprises access cloud more efficiently.

3.2.3.1 Tiering to Public Cloud

Most vendors support tiering backup data to the public cloud. This reduces on-premises backup storage cost. The most commonly supported public cloud storage targets are Amazon Simple Storage Service (Amazon S3) and Azure Blob storage.

If the on-premises data and catalog are lost, then an instance of the backup software can be reinstalled in the cloud and data can be restored.

Some vendors also integrate with the life cycle policies of cloud providers (for example, data migration from AWS S3 to Glacier, or Azure Blob to Azure Archive Blob storage)

3.2.3.2 Cloud-based Recovery

Today, leading backup vendors support restoring backup data to servers in the public cloud. An instance of the backup software can be installed in the public cloud, and backup data can be restored to a compute instance in the public cloud. This provides quick operational recovery if the on-premises environment is not available.

3.2.3.3 Replication to Cloud

With cloud replication, backup data can be replicated to clouds for remote data protection. It facilitates rapid recovery from remote site or cloud. This solution satisfies the needs in many scenarios, such service migration, service expansion and service DR.

3.2.3.4 Cloud Archiving

Cohesity etc. provides cloud archiving in different modes. The appliances can backup data and copy to archival, or act as a archiving gateway by streaming data directly into lower-cost media and storing index and metadata locally for search and recovery.

3.2.4 Backup With Long-Term Data Retention

Many organizations are still using tape for long-term data retention. Tape is slow and can’t be accessed quickly, which is now a necessity for many use cases, such as e-discovery, regulatory compliance, product development and customer service.

Modern backup appliances normally support long-term data archiving to various locations and various medias, such as tape, local object storage, and cloud.

3.2.5 Add-On Services

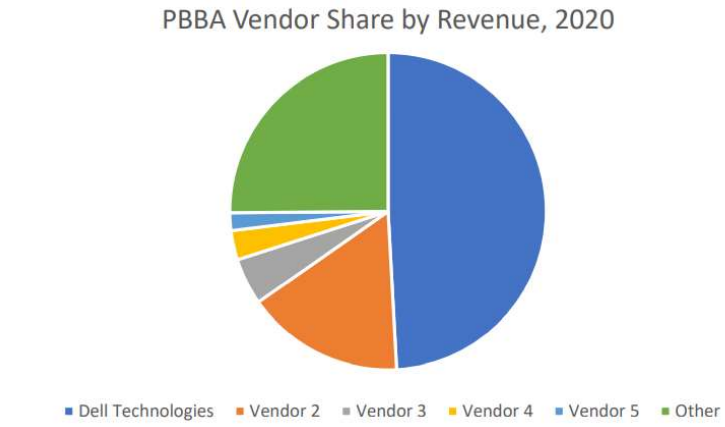
New use cases are constantly emerging. These may extend to all secondary storage use cases, starting with data protection and extending to data management, dev/test, data analysis and even file/object services. These extended use cases may create substantial additional value for organizations and take full advantage of the related products. We will discuss them in the later sections.

3.3 Key Market Players

Key vendors in this area include Veeam®, Commvault®, Rubrik®, Cohesity®, Dell®, Veritas®, IBM®, Arcserve®, Druva®, Acronis®, etc.

For PBBA market, currently Dell® enjoyed 47% market share in 2020, including target and integrated devices plus associated software[8].

Some new vendors are gaining spending velocity, although the current market shares are still small. We will discuss Cohesity® and Rubrik® as examples.



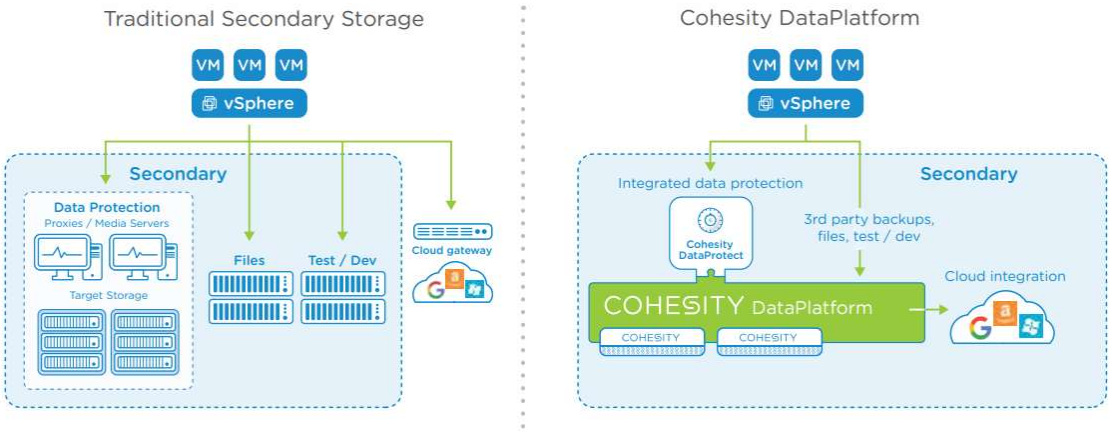
Source: IDC's Worldwide Quarterly Purpose-Built Backup Appliance Tracker, March 2021

3.3.1 Key Player - Cohesity®

ETR "net score" (spending velocity, whether a customer is planning to spend more on a particular vendor) shows Cohesity® garners much of the venture capital money and really expanded the data protection from backup into data management, analytics and security.

Cohesity® provides the hyper-converged platform that eliminates the complexity of traditional data protection solutions by unifying your end-to-end data protection infrastructure – including target storage, backup, replication, disaster recovery, and cloud tiering. This solution consolidates all secondary storage on a single web-scale platform that spans from the core to edge to cloud.

Cohesity® DataPlatform provides scale-out, globally deduped, highly available storage to consolidate all your secondary data, including backups, files, and test / dev copies. Cohesity® also provides Cohesity® DataProtect, a complete backup and recovery solution fully converged with Cohesity® DataPlatform. It simplifies backup infrastructure and eliminates the need to run separate backup software, proxies, media servers, and replication[5].



3.3.2 Key Player – Rubrik®

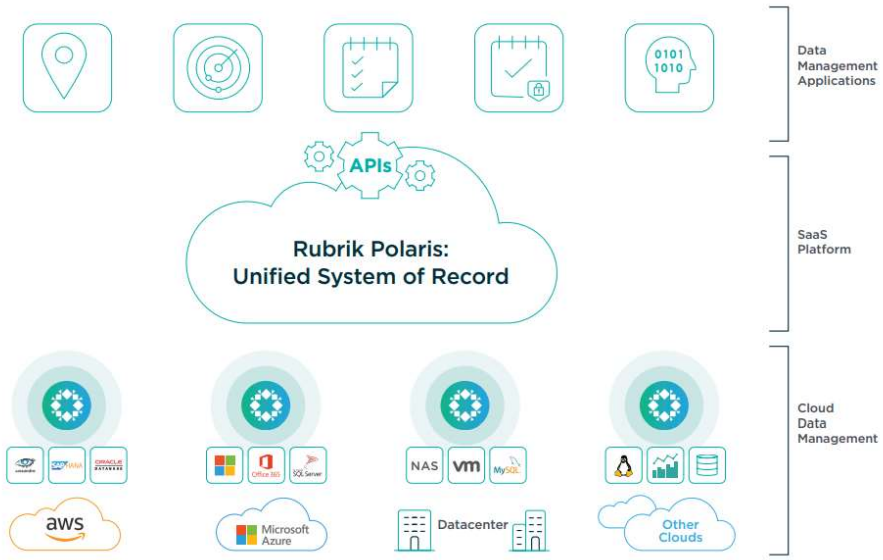
Rubrik® is a cloud data management company. It delivers a single software platform for complete enterprise data manage across data center to cloud[7]. Its product portfolio mainly consists of:

- Rubrik® Cloud Data Management® (RCDM): its core backup software platform for backup and recovery, archival, DR on-premises, at the edge and in the cloud;
- Polaris: a SaaS-based platform that provides centralized visibility and management, and leverages metadata to provide ransomware assessment, recovery and data classification;
- Mosaic: automates application-consistency and recovery for NoSQL workloads.

Rubrik®'s strengths are in its centralized monitoring and management (Polaris), operational simplicity and automation, easy cloud mobility, and comprehensive support for all major relational databases and NoSql databases.

RCDM is software-defined: it consolidates different hardware and software components into a single software fabric. Enterprises can run Rubrik® anywhere via plug-and-play appliances on-premises, as software on third-party hardware, or as software in the cloud.

Rubrik® provides freedom of choice, allowing enterprises to operate both on-premises and in any cloud while avoiding vendor lock-in.



3.3.3 Customer Considerations

Small/midsize enterprises have several basic requirements before they purchase appliances for backup and data protection:

- Plug & Play: function effectively out of box.
- Scalability: expand and contact with the changing needs.
- Manageability: centralized management via a central console.

- Add-on Services: new features such as archiving, e-discovery from a single hub.

Integrated backup appliances have been mainstream for enterprises, especially small/midsize enterprises. Small/midsize enterprises present some of the same challenges as edge environment, including limited IT resources to deal with the growth in data. In addition, they need flexibility. By bundling together both hardware and software, enterprises get the backup solution that eliminates the need to spend a long time on configuring and deploying solutions to their environments.

In addition, vendors such Cohesity® directly address the new use cases including copy management, archiving, test/dev, data analytics and even file services. These add-on services help enterprises to use the system for more than just data protection.

3.4 EVOLUTION OF PBBAS

Over the past ten years, we have experienced tremendous innovations in data protection technology, with each innovation made huge advantages in backup and recovery market.

Purpose-built backup appliances (PBBAs) have been foundational components to many backup and recovery infrastructure for more than a decade. Recent surveys show the market appears to be in the midst of a transition, from target devices to integrated appliances to comprehensive integrated appliances (HyperConverged Data Platform).

3.4.1 Integrated Appliances

Target PBBAs are disk arrays with specific functionalities (deduplication, encryption, etc.) to house and manage backup data sets. Target devices do not include backup software. Integrated appliances are like target devices in configuration and capacity, with the addition of installed backup software:

- Target devices can be used with any backup software, and can be added to an existing environment seamlessly. Organizations wishing to use existing software will likely to choose a target device.
- Integrated appliances have backup software preinstalled. So the deployment is rapid and simple. Organizations seeking simplest implementation will likely to choose an integrated appliance.

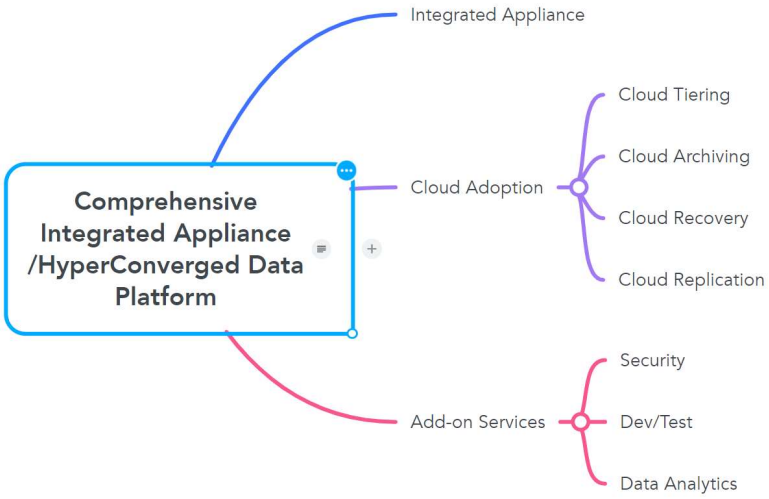
IDC survey shows the market growth is coming from integrated appliances[6].

3.4.2 HyperConverged Data Platforms

HyperConverged data platforms, or comprehensive integrated appliances, are like integrated appliances, with addition of hybrid cloud support and add-on services. They also consolidate multiple solutions across backup, replication, and archival. Hybrid cloud services include (but not limit to) cloud tiering, cloud archiving, cloud recovery and cloud replication. Add-on services include security services, dev/test environments, data analytics, and more. We will discuss add-on services with more details in a later chapter.

A hyperconverged data platform is intended to provide following benefits:

- Eliminate complexity with a unified platform for end-to-end solution and unified management platform.
- Reduce TCO by consolidating **multiple solutions across backup, replication, and archival** and manage copies of their data.
- Increase storage efficiency with deduplication and other data reduction techniques.
- Integrate seamlessly with all the leading cloud providers to leverage economics of public cloud services.
- Provide a platform for data reuse such as data mining and analytics.



3.5 RECOMMENDATIONS

As analyzed in previous sections, PBBA market is moving toward HyperConverged data platforms, we have several recommendations for current integrated appliances.

3.5.1 Enhance ecosystem support

A broad ecosystem support is essential for the current PBBA market. This includes:

- Software support: Oracle, SQL, SAP HANA, NoSQL, SaaS-based applications, etc;
- A broad range of operating system support;
- Different hypervisors;
- Public cloud support;
- IaaS/PaaS backup;
- Integration with major backup softwares for target devices.
- Major primary storage products integration
- Kubernetes, OpenStack, VMware Tanzu, OpenShift

3.5.2 Enhance hybrid cloud support

As public cloud adoption is common in all enterprises, PBBAs need to adopt hybrid cloud solutions such as:

- Cloud backup, recovery, archival, replication;
- Broad public cloud support: EC2, GCE, AWS, RDS, Azure

3.5.3 Expand add-on services

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 PBBAs. We recommend that more effort is needed in this developing area.

4 ARCHIVING

Enterprise information archiving (EIA) solutions are designed for archiving data sources to a centralized platform to

- satisfy information governance requirements, including regulatory and/or corporate governance and privacy;
- improve data accessibility;
- surface new data insights;
- and gain operational efficiencies.

There are several core capabilities of this market. They include archiving digital communication content, such as email, workstream collaboration, instant messaging (IM) and SMS; classifying data and enabling retention management of archive content; creating a searchable index of content; and providing basic tools for e-discovery and supervision.

Archiving and backup are often used interchangeably but they have different meanings:

- Backup protects both your active and inactive data (all of your production data). You can back up your information via tape, disk or the cloud. Backup is a copy of production information. Your data still resides on the production storage systems themselves. That means if your backup system faces a major data loss (due to a security breach, disaster, infrastructure failure, etc.), you could continue normal operations. Your production data won't be impacted, though you would be operating at an increased risk. Backup applications may be used to protect application and OS files, in addition to individual data objects—though it's optimized for larger scale recoveries. It's best for recovering applications or complete systems.
- Archive solutions are often used to retain inactive or older data for extended periods of time. Archives are optimized for low-cost, long-term storage. Archives hold production data, meaning a loss or corruption of an archive system will likely result in the permanent loss of production information. Keep in mind that this data will likely be older or less used, but it could also be the only copy. Designed to store individual data objects such as email messages, files and databases, along with their metadata.

4.1 MARKET INSIGHT

4.1.1 Overview

The global enterprise information archiving (EIA) market was valued at USD 5.93 billion in 2020, and it is expected to reach a value of USD 12.49 billion by 2026, with a CAGR of 13.51% over the forecast period (2021-2026) [2].

The growth of enterprise information archiving solutions is attributed to several factors:

- The rising generation of data across organizations: According to Cisco, the volume of Big Data in data center storage is anticipated to increase to 403 exabytes by 2021 from 179 exabytes in 2019.

- The requirement of reducing storage costs for EIA.
- Government mandates for audit and investigation: several regulatory standards have been designed and developed by governments for company compliance. Regulations often mandate that electronic data such as email and social media needs to be stored in a secure, tamper-proof format.
- Enterprises across the globe are adopting technical solutions as part of their digitalization process.

With the outbreak of COVID-19, the market is likely to grow as the total volume of data in the world is growing at a fast pace. According to Seagate Technology PLC, the volume of data is expected to reach 149 zettabytes by 2024 from 26 zettabytes in 2017.

4.1.2 Key Market Trends

BFSI industry held the highest market growth rate of 14.77% over the year 2021 to 2026. Also it held the largest share of 23% in 2020. The moving toward digitization changes the market landscape, new regulations and consumer behaviors.

Large enterprises held a larger market size of 65.02% in 2020. The small and mid-range enterprises is anticipated to have the fastest growth rate of 14.37% till 2026. The adoption of EIA solutions among large enterprises is higher, owing to the increased generation of data, attributing to the widespread geographical presence and customer base. There is an increasing demand for enterprise information archiving solutions. These solutions optimize the storage information resources, lower the risks, improve enterprise efficiency, maintain transparency of the enterprise, and reduce the risks of DR.

North America held the highest market share of 31.94% in 2020. This region is an early adopter of the latest technologies, such as AI, cloud, and mobile technologies within traditional enterprise information archiving solutions. Moreover, the region has a stronghold of the EIA vendors such as Google®, Microsoft®, IBM®, Dell®, Veritas®, Barracuda Networks®, Proofpoint®, Smarsh®, etc.

Asia Pacific to grow at the highest CAGR of 15.32% over 2021-2026: Major APAC countries, such as China, Australia, India, and New Zealand are expected to record high growth rates. Moreover, large companies, such as NetApp, Dell, Cisco, and IBM are expanding their cloud business rapidly in the region due to the availability of a robust IT infrastructure.

4.2 Key Market Players

Key vendors in this area include Barracuda Networks®, Google® Vault, Microsoft®, ZL Technologies®, Veritas®, Archive 360®, Mimecast Services®, Smarsh®, Proofpoint®, Micro Focus®.

Cloud vendors take a big part in EIA market. They provide highly available, affordable and protected solutions for enterprises.

4.3 CLOUD DEPLOYMENT

Archiving has 3 deployment modes: On-premise, cloud and hybrid. Fow now, on-premise segment held the major market share, with 42.94 in 2020. The cloud segment is anticipated to have the fastest growth of 15.54% over the period till 2026 [2].

Cloud deployment is increasing rapidly for its agility, scalability and lower costs. The price could be as low as 0.1 cents per GB per month. Therefore, the players in the market are increasingly offering cloud-based solutions to manage data.

Archiving data directly to the cloud in its original format allows data to be analyzed "in place" by advanced cloud services such as machine learning, analytics, and data pipelines. A fast indexing engine is needed to access and retrieval of data from public cloud and makes data more useful for teams to gain meaningful insights.

Public clouds have a vast customer base that includes enterprises and educational institutes.

Hybrid deployment archives data to multiple targets through the same UI – public clouds, private clouds, any object storage, and tape libraries. More enterprise vendors provide this deployment mode for customers when they need part of the data on-site for security or other considerations.

Some vendors use global deduplication to optimize both data transfer and long term cloud storage costs. Along with compression, only the necessary data is sent to cloud for archival to further reduce data transfer costs and improve archival and retrieval costs.

4.4 BACKUP WITH ARCHIVING

We already discussed backup and archiving as isolated solutions in previous sections. Cohesity® and many other vendors support backup and archiving within the integrated appliances. They support archival to public cloud, S3-compatible object storage, NFS, and tape.

The appliances can backup data and copy to archival, or act as a archiving gateway by streaming data directly into lower-cost media and storing index and metadata locally for search and recovery. In the first case, the appliances still keep a copy of data locally.

The advantages of backup with archiving are:

- Backup efficiency gains with faster backups, low media cost, and reduced compliance risk.
- Reduced complexity by using comprehensive appliance
- Unified management improves operational efficiency.

Some customers, already have archiving software working, or they need a different vendor for archiving. They may be reluctant to uses backup with archiving solution as a replacement.

5 ADD-ON SERVICES

Viewing integrated backup appliances strictly in the context of "backup and recovery" is a mindset that organizations must overcome. 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.

5.1 DATA PROTECTION

5.1.1 Ransomware Protection

The recent increase in the number of ransomware attacks has resulted in vendors taking concrete steps toward providing ransomware detection and remediation as well as a resilient backup infrastructure.

Most vendors use following techniques to achieve ransomware protection:

- Immutable second copies of backup through write once, read many (WORM)-enabled storage,
- Immutable snapshots to make the primary backup repository more resilient.
- Capabilities to detect ransomware attacks by tracking large changes to file system data, and through other means, by partnering with security vendors or by developing these capabilities in-house.
- Recovery vault, as mentioned in previous section.
- Creation of an isolated test environment, and provide a clean backup copy to recover specific files.

Many vendors made a lot of effort for advanced ransomware protection. For example, Rubrik® Radar applies machine learning algorithms against application metadata to alert anomalous activity and reduce the time spent on monitoring.

Dell PowerProtect® integrates CyberSense® which adds an intelligent layer of protection to help find data corruption when an attack penetrates the data center. This approach provides full content indexing and uses machine learning (ML) to analyze over 100 content-based statistics and detect signs of corruption due to ransomware.

5.1.2 Antivirus Scan

Most storage vendors provide antivirus solutions with 3rd party security vendors. These solutions are designed to scan storage system during end user access, or based on manually scheduled policies from a central antivirus servers.

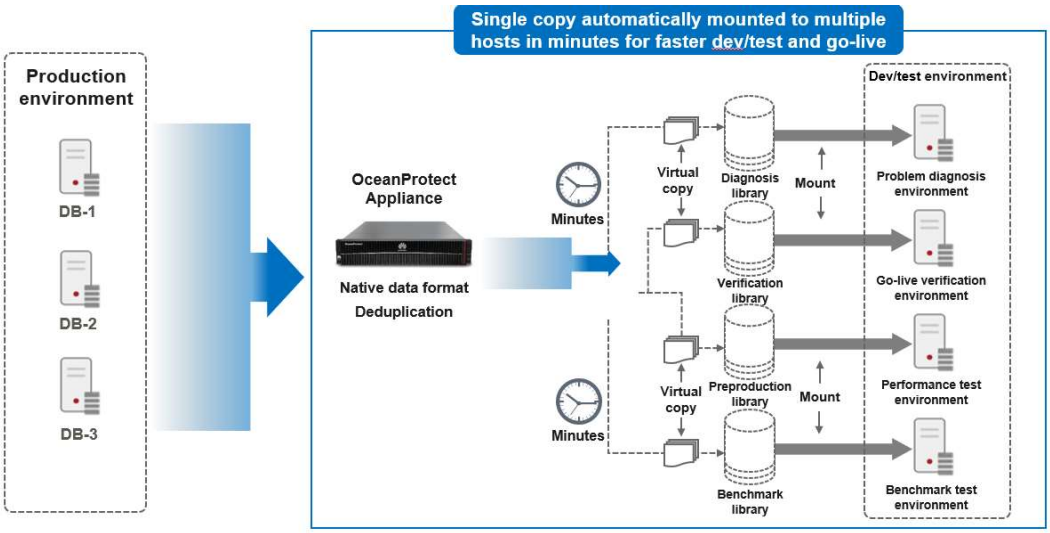
5.1.3 Data Anonymization

Data anonymization can help enterprises to protect data such as telephone numbers, national identification numbers, email addresses, file names and other places. Also, it is a

good way for crypto erasure, if the link between the real information and the mask is removed.

5.2 Development and Test Environment

According to Gartner, more than 30% of users use backup data for development and testing. Many vendors (Huawei®, Rubrik®, etc) provide live mount capabilities for test/dev work flow with backup data. So customers can use datasets with the latest production data, at the same time the datasets sit on non-production resources (on premises or cloud), avoiding interference and performance downsides.



Veeam® provides on-demand sandbox for test by starting one or more VMs from backups, replicas or snapshots to create an isolated environment.

5.3 Data Intelligence

5.3.1 Data Classification

Different data has different requirement for security. A one-size-fits-all approach will create areas of too much security and others too little, thus increasing the risk for enterprises. Enterprises spend a lot of time and money on finding sensitive data for protection. Many vendors tried to use advanced classification techniques to automatically scan and identify all kinds of sensitive data and locations. Classification is critical to organize data with different compliance. For example, Rubrik® Sonar uses machine leaning to automatic data classification with policy templates.

5.3.2 Data Analytics

Many vendors allow applications and third parties to exploit secondary data with data intelligence, without making copies or run in a separate location. The solutions allow data analytics on structured or unstructured data, provide index and global search, as well as anti-virus operations. For example, Cohesity® Marketplace allows customers to deploy

third-party applications, including Splunk Enterprise Analytics, SentinelOne and Clam anti-virus, plus Imanis for Hadoop and NoSQL database protection.

6 CONCLUSIONS

We analyzed the secondary storage market, focusing on backup and archiving markets. The secondary storage market has seen a big addressable market expansion, driven by data boom, data protection and regulatory compliance.

Enterprises of all sizes began embracing hybrid cloud strategies that are becoming more complex and structured. Comprehensive integrated appliances, with hybrid cloud solutions and add-on services, are gaining customer attentions, especially in small and medium enterprise markets.

We summarized add-on services in data protection, dev/test environments, and data analytics. These extended use cases create substantial additional value for organizations and take full advantage of the products. We think this is a developing area that needs our attention.

7 BIBLIOGRAPHY

[1] Gartner, "Magic Quadrant for Enterprise Backup and Recoery Software Solutions".

[2] Mordor Intelligence, "Global Enterprise Information Archiving (EIA) Market (2021 - 2026)".

[3] Report and Data, "Data Backup and Recovery Market By Type, By Deployment Type, By Organization Size, By Component, By Industry Vertical, and Segment Forcasts, 2020-2028".

[4] QYResearch, "Global Cloud Backup Market Size, Status and Forecast 2020-2026.

[5] [Online]. Available: https://www.cohesity.com/resource-assets/white-paper/Cohesity_Data_Protection_White_Paper.pdf

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

[7] [Online]. Available: https://www.rubrik.com/content/dam/rubrik/en/resources/white-paper/Definitive-guide-rubrik-cloud-data-management.pdf

[8] [Online]. Available: https://www.delltechnologies.com/asset/en-id/products/data-protection/industry-market/idc-purpose-built-backup-appliances-market-results.pdf