OROCK TECHNOLOGIES

Five Key Considerations

WHEN MOVING FROM TAPE TO CLOUD STORAGE

How to secure and access data without paying cloud egress fees

x3

x2

x1

Create 3 copies of data

Keep data on 2 types of storage media

Store 1 of these offsite

Table of Contents

Management Summary:

Recording and storing information for later use is as old as humanity itself. From stone, to paper, to phonograph, to reel or cassette-and-capstan, to disk platters and solid-state drives, humans have evolved their methods of preserving text and images on solid surfaces over which they can later re-live the experience. In the enterprise storage market, more than half (58%) of data is backed up on disk. Nearly a quarter (21%) of worldwide data is kept on tape reels[1]. Despite digital storage advances, organizations continue to store to physical assets such as tape as a realistic part of a 3-2-1 storage and risk management discipline.

Close to 80 percent of overall data storage continues to be internally managed and backed up on physical equipment [1]. Tapes can tear and deteriorate, disks can skip, and the devices and drivers needed to make sense of what's stored on the tape require maintenance and gradually obsolesce. Rewinding, pausing, taking notes, splicing, and trying to pull files and data off video tapes is a legacy, nostalgic, but still necessary approach toward ultimate disaster scenarios. On the other hand, your organization's need to retrieve and re-experience the data shouldn't be hampered because of an inability to deliver real-time digital service and intelligence.

Storage and playback mechanisms have continued to evolve, becoming more digital and service-oriented. The cloud is making this easier, providing IT infrastructure and storage as-a-service. Both hot (active) and cold (inactive) data can be backed up automatically and these experiences can be revisited via software either in your data center or in the cloud. However, when moving off analog technology to digital mediums such as video and disk, or to server software and hybrid cloud strategies, the likelihood of unauthorized access over the internet, data theft, and on-device evil-maid attacks increase. This is especially a concern for organizations and agencies with high-target, highly sensitive data stores.

Moving to the cloud can also bring hidden costs in the form of cloud egress fees.

This white paper explores the pros and cons of storing data to tape, internal servers, the cloud, or a mix of all three. It proposes the characteristics of the right cloud provider, so that retrieving your information won't 'toll' your organization both financially, physically and reputationally.

Common Drivers for a Data Storage Migration:

Tape storage in warehouses can reduce equipment and air conditioning costs short term, but it leaves organizations unable to access informational assets in a timely fashion. Write it off as that third level of redundancy, protection from network-attached forms of data attack and misuse.

These are all reasons why a 3-2-1 storage migration approach is needed today [2]. Tape and disk are tried-and-true old-school methods, but should only represent part of your ultimate storage and disaster recovery plans. A third – or prioritized second tier – of storage is needed for digital transformation which enables business strategists and team members in the field to learn from experience by analyzing a cloud-based instance of that backed-up data.

Cloud storage still brings hesitancy from some organizations, even though some of their apps are already in the cloud via Office 365, Salesforce, Zoom, or point application service solutions. A small percentage of small businesses and organizations were born in the cloud, and operate at a largely virtual level, with their infrastructure residing off premises. For most enterprises, however, data is kept and processed on-site, including branch office servers for storage, and a near quarter share is still on tape, with some backup support from the cloud. Some of this virtual storage is public (collocated but segmented workloads) and some of it is in a dedicated private cloud.

While the tape storage market share remains constant at about 20-something percent, the cloud storage market is growing.

According to Allied Market Research, "The cloud storage market size was valued at $46.12 billion in 2019, and is projected to reach $222.25 billion by 2027, growing at a CAGR of 21.9% from 2020 to 2027," [3]. Meanwhile, Market Data Forecast, another business research and consulting firm, projected the cloud storage market will grow 37.2 billion to 106.7 billion by 2024 [4]. The Market Data Forecast report went on to say that 72 percent of organizations worldwide will migrate data from internal data centers to a public or private hosted cloud by 2022. Technology research firm Statista reported that, as of 2019, about 48 percent of all corporate data is stored in the cloud and that segment is continuing to grow [5].

This cloud data storage growth can be traced to the shared distributed resource and the use case advantages the cloud can bring.

Cloud Data Migration Use Cases:

- Tape Replacement— Here, the company is utilizing tape for backup and archiving, which is good practice from a disaster recovery perspective. But tape storage involves personnel time, data center floor space, licensing fees, physical tapes/storage media, cargo transportation, and the maintenance of tapes and tape drives. Tape storage can become inefficient and ineffective for organizations that can't wait for the latest data-driven analysis of emerging patterns or need access to historical data for more rapid compliance.

- Offsite Copy of Data— This involves an organization with a limited backup strategy. The two-tier approach is used for disaster recovery, or branch-office backup, and often involves disk or tape. But the inability to use the data suffers because of the logistical and 'transportational' challenges of distributed workforces accessing physical assets such as tapes and disks.

- Unwieldy Costs of Hyperscale Cloud Service Providers — In this scenario, the customer uses primary storage on-site and tiers off to AWS S3, AWS Glacier, or Azure Blob, but accumulates cloud egress fees that are unpredictable and difficult to budget for. Space agency NASA recently faced tens of millions of dollars in unexpected cloud egress fees just to access and analyze its own distributed data. Organizations would rather have their data "gaining interest" by providing business insight in real time without all the access fees [6]. In addition, both traditional tape backup and mainstream cloud storage paths introduce analytical obstacles, via data transcription delays and budgetary constraints, hindering data access patterns for when an organization might want its data back in 1 hour vs. 1 day vs. 1 week.

- Multicloud Strategy— Midsize to large organizations generally use four or more cloud partners. This often entails various cloud storage providers for redundancy and infrastructure and application modernization. In addition to the aforementioned cloud egress fees, cloud platform providers also hit you with application programming interface (API) charges, which, at scale, negatively impact storage cost optimization.

- Hybrid Cloud Backup Strategy— A hybrid cloud storage strategy involves an off-premises complement to on-prem storage. It entails spinning up a third backup site that bolsters security and disaster recovery posture and allows you to retrieve your data in storage without paying tolls – i.e., cloud egress fees. For your business, a 3-2-1 redundant storage approach might demand tape backup, but you also want a version of the data that you can look at when you need to, and not have to pay for per every viewing or analysis.

- Compute and Analytics Needed— If you are storing to tape or facing usurious egress fees, this can thwart both storage optimization and more rapid strategic intelligence. Getting information and business intelligence off physical storage appendages can be costly from a time-consumption perspective, involving playback, isolation, listening, and then loading or transcribing data. A better way is to harness machine learning for compute and analytics on your frequently accessed data.

- Support of Remote Users. The daily workday has changed, becoming less office-centric, more distributed, and your customer service personnel need to be able to securely access data and serve customers. This usually means the records-based data they can securely access without your organization paying the punitive cloud egress toll.

- Infrastructure Modernization. The cloud can support the extended private network, what Gartner calls the secure access service edge (SASE) [7], without analog storage equipment delays or other cost prohibitive data access restrictions. Infrastructure as a service should boil down to connectivity, access, and security. To deliver services, you need to provide network connectivity through servers, firewalls, the apps. Infrastructure modernization also means supporting different data types that distributed employees require such as textual data, image files, and videos.

These cloud migration drivers are complicated by security and privacy compliance challenges as industry and government regulations require organizations to protect, store, archive, and access data over many months or years, and report on it. There are X-rays and other medical images that need to be compared by authorized personnel, there are financial-transaction DCI OSS records, and there is business intelligence borne of Big Data and privacy-compliant data extraction and analysis. All this data must be protected yet accessed by authorized users when needed.

Different Types of Data to Be Stored

Data types include file, block and object, each with its fair share of challenges. Throughout these categories, you will hear about unstructured data, which is also a key consideration in an organization's cloud migration. Unstructured data is the type that cannot be easily managed and sorted in a traditional database. Unstructured data includes media files (video formats), maps, log files, and more. It's the kind of data that lacks the built-in relational information that enables it to easily organize in a traditional database table column.

File. File storage involves a data server typically built around a traditional hub-and-spoke network architecture, but can also involve remote and satellite locations and cloud infrastructure help. The server stores and serves files. It can scale out but the data to be stored must be hairpinned back to the appliance for central storage policy. Network attached storage implies it's part of the network (usually a file) and your employees can securely access it. When NAS is hosted in the cloud, you should not have to pay to access your own data, regardless if it is block, file, or object data.

File management systems and their tables for structured data are sufficient for some use scenarios, but when unstructured data elements such as audio or video files get backed up, they become difficult to search and find without investments in meta tag and analytical tools.

Block. Block storage separates data into smaller units within a table, an approach which is still necessary for frequent access and rewrites. However, block storage retrieval requires extra steps to reassemble the data from separate meta elements in the table, requires file names and file paths, and it can consume servers with the workload.

Object. Object storage is the caching of static unstructured data, with each object composed of three parts: data, metadata, and a global identifier. Object storage is a more efficient and effective way to manage unstructured data because it enables administrators to create metadata for each data object in the database, making the data findable, 'abstractable' and retrievable, versus having to build a separate database for the classification of the objects.

For organizations that stream data frequently, object is an easy, affordable, and scalable approach because they can embed these image and video object files on a website or other user web app without reliance on servers. Object storage is also more easily managed from a security perspective. Security administrators can make objects private and hidden from routine HTTP requests. Conversely, objects can be made public using access controls that don't require a server.

In terms of object storage limitations, stored objects are accessible but only in their native file types and they are not easily rewritable.

Various, Hybrid Approaches to Data Storage:

All this brings a hybrid cloud reality to storage, infrastructure, and application integration.

Hybrid is an apt word for the broad state of the data storage market, i.e., commercial and federal organizations using a mix and match of public and private solutions for their storage and access. A 3-2-1 hybrid cloud storage architecture has evolved for this landscape [2]. You can back up everything internally, on tape or on digital servers, but your organization can lose out on the wider business intelligence and performance when you don't use the cloud.

With many organizations operating a digital data center as well as a separate tape storage facility, 3-2-1 hybrid cloud storage equals a transitional reality, a combination of public and private, tape and cloud. You might still back up data to tape, which is a tried-and-true method of secondary or tertiary data protection, but this requires frequent rewinds to find data points and does not allow data to be easily and rapidly harnessed.

Cloud platform providers can facilitate and optimize storage and Big Data utilization through subscription relationships for infrastructure-as-a-service and storage-as-a-managed-service. Most cloud service providers make it easy to do business with them upfront, but then as part of the contract they charge you penalties when you want to access your own data in storage.

As part of the hybrid landscape, major cloud platform providers, such as Amazon, Microsoft, and Google, introduce:

- High cost of data egress
- Vendor lock-in (too expensive to move data)
- Complex billing. Pricing is not fixed and can get unpredictable depending on number of service requests from storage. In NASA's case, the federal agency could not have possibly predicted the tens of millions of dollars it would face based on its distributed computing and analytical needs.

10TB Data Egress

- AWS S3: Data Egress $900, Storage $1,150
- AZURE BLOB: Data Egress $870, Storage $1,040
- OROCK TECHNOLOGIES: Data Egress $0, Storage $1,050

30TB Data Egress

- AWS S3: Data Egress $2,600, Storage $1,150
- AZURE BLOB: Data Egress $2,530, Storage $1,040
- OROCK TECHNOLOGIES: Data Egress $0, Storage $1,050

Legend: Data Egress, Storage

There are also cloud vendors that let you access your data, but don't build in sufficient security. These hot cloud storage vendors don't offer high levels of data security or meet key government-grade security guidelines.

Five Steps for a Successful Cloud Storage Migration:

Increasingly data is moving off premises. It is too expensive to haul it back into your core network, especially due to network tariffs and cloud egress fees. Meanwhile, tape storage processes become cumbersome and costly when you need to access or analyze historical or trending strategic data.

To design a blueprint for today's hybrid cloud storage reality, organizations should:

1. Understand the limitations of local and magnetic tape storage. Storing data on-prem or on tape instead of in the cloud can limit an organization's scalability, flexibility, and quick-retrieval data analysis.

2. Don't let your data be contractually ransomed. When you move to the cloud, is your data (e.g., text, images, video) accessible when you need it? Understand what you'll get charged for your data – hint: watch out for egress fees. Data egress fees and other charges by large cloud service providers are high and unpredictable, resulting in billing overruns and de facto vendor lock-in.

3. Embrace the hybrid cloud reality. Most enterprises are toeing the water of hosted cloud management, with apps often representing the first step, for example, Microsoft 365 for office tools, Salesforce for data storage and analysis, and Amazon Web Services for additional data infrastructure. Find a cloud storage vendor that is just right in terms of security and strategic storage optimization.

4. Make sure your data is secure when you migrate storage to the cloud. Leaving the preservation of your backed-up data up to major cloud platforms' white-label equipment introduces risk and a lack of visibility. You want security built into the hardware, for example, through trusted brands such as Intel and Hewlett-Packard Enterprise (HPE), and the software, e.g., Red Hat containers.

5. Be able to analyze your data without having to pay small-print, unexpected fees. ORock recommends a secure 3-2-1 backup approach. You can still store data on premises, on tape, or to a separate facility, while ORock brings the secure but flexible open platform that empowers digital business intelligence and business innovation from storage, minus the cloud egress fees.

Migrating data storage to the cloud, keeping that data secure and compliant, having access to it, and analyzing it, without being penalized by cloud infrastructure-and-storage hotel fees can be a complex and costly endeavor. ORock can help:

- ORock's infrastructure and cloud solutions were built specifically for highly regulated industries and government agencies with a focus on security and compliance. ORock is authorized by the Federal Risk and Authorization Management Program (FedRAMP) for use by the federal government and Department of Defense, and is compliant with HIPAA, HITECH, and PCI DSS.

- Data stored in ORockCloud is instantly accessible and can be geo-replicated. It can be always accessible and immediately available data in hot or premium tiers. The costs are predictable and flat-rate data transport costs, without usurious fees.

- ORock's enterprise-class open source cloud avoids hardware and software vendor lock-in through software-defined storage. It's vendor-agnostic enterprise-class open source storage, empowering the integration of additional secure software components.

- ORock empowers organizations to optimize storage and maximize system agility and availability. Your resources can be securely orchestrated where cost and retrieval of information considerations are best for your business.

- Last, it's not a swipe and good-bye relationship. ORock brings the upfront and sustained expertise you need but won't get from major cloud platform providers.

Sources:

1. Tech Target: "How is tape backup doing in 2019?," 2019, https://searchdatabackup.techtarget.com/news/252459656/How-is-tape-backup-doing-in-2019
2. Hewlett-Packard Enterprise: "3-2-1 Data Backup Rule," 2020, https://h20195.www2.hpe.com/v2/GetDocument.aspx?docname=a00100041enw
3. Allied Market Research: "Cloud Storage Market Statistics; Global Opportunity Analysis and Industry Forecast, 2020–2027," April 2020, https://www.alliedmarketresearch.com/cloud-storage-market
4. Market Data Forecast: "Cloud Storage Market Research Report," 2019, https://www.marketdataforecast.com/market-reports/cloud-storage-market
5. Statista: "Cloud storage of corporate data in organizations worldwide 2015-2019," April 2020, https://www.statista.com/statistics/1062879/worldwide-cloud-storage-of-corporate-data/
6. The Register: "NASA to launch 247 petabytes of data into AWS – but forgot about eye-watering cloudy egress costs before lift-off," March 2020, https://www.theregister.com/2020/03/19/nasa_cloud_data_migration_mess/
7. Gartner: "Say Hello to SASE (Secure Access Service Edge)," December 2019, https://blogs.gartner.com/andrew-lerner/2019/12/23/say-hello-sase-secure-access-service-edge/.

About ORock Technologies:

ORock Technologies delivers hybrid cloud and IaaS solutions designed for secure, compliant data operations. ORock helps commercial organizations and government agencies protect their most sensitive data, control costs, and minimize vendor lock-in while enabling a range of IT modernization, application hosting, migration, and edge computing initiatives. ORock's private fiber optic backbone network and enterprise-grade open source cloud feature the latest HPE Gen10 secure hardware and a flat-rate OPEX billing model with no data egress fees. They support hybrid, private, and multicloud capabilities while providing superior security, performance, predictability, and control.

Learn more about ORock at www.orocktech.com.