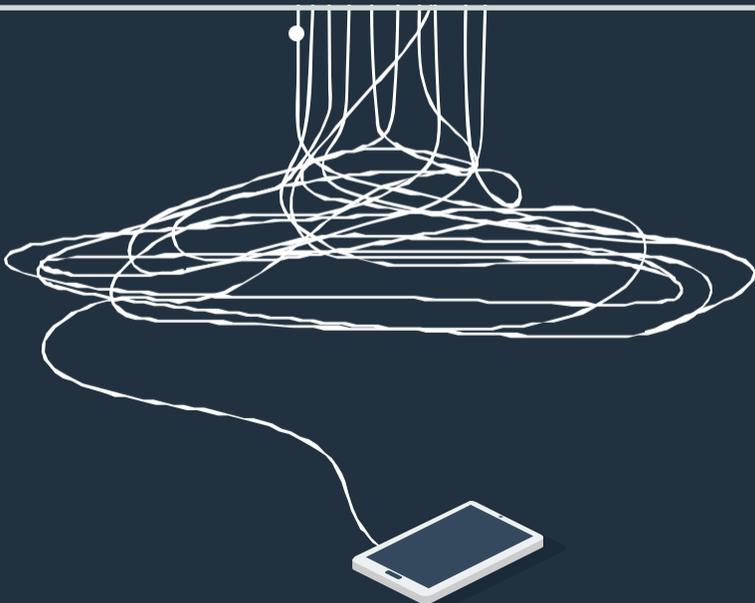# THE CASE FOR CLOUD OBJECT STORAGE

Shawn Sabanayagam >> Tributary Systems >> Chairman and CEO
Glenn Garrahan >>   Director of HPE Business for Tributary Systems

> "I've looked at clouds from both sides now
> From up and down and still somehow
> It's cloud's illusions I recall
> I really don't know clouds at all"
>
> — Joni Mitchell, "Both Sides, Now" 1967

Well, Ms. Mitchell certainly wasn't thinking of Enterprise Cloud Object Storage when she wrote "Both Sides, Now" a half-century ago, but there is some small truth in these lyrics when applied to NonStop and cloud based backup.

The Evolution of Storage:

The first use of tape for storing computer data was in 1951, in conjunction with the UNIVAC 1. The media was a one-half inch wide thin metallic strip, with a recording density of 128 characters per inch on eight tracks. Cartridge style media began to gain popularity in the 1980's, and it wasn't long before this media type displaced the open reel-to-reel units, due to the ease of use, simplicity of archiving cartridge based media, and reduced possibility of media damage when manually reloading large reels of loose tape. Cartridges have been available in several formats, modern types include DAT DDC, DLT, and of course the most ubiquitous format of all, Linear Tape Open, or LTO. Along with cartridge style media came tape automation, including massive tape libraries and silos such as those manufactured by StorageTek (STK) and IBM. With the use of robotics, these silos could hold thousands of tape cartridges, and required host based media manager software to perform the backup/restore functionality.

IBM 2401 Reel to Reel Tape (circa 1964)

The first hard disk drive, the IBM 350, was developed in the early 1950's, and the first production drive was shipped in 1956. The size of a small refrigerator, this unit had a capacity of approximately 3.75 megabytes. The groundbreaking "Winchester" drives were introduced by IBM in 1973. This drive technology and its derivatives, using low mass/low load heads and lubricated platters, remained state-of-the-art for over 40 years, and was emulated by most disk drive manufacturers. Of course, every succeeding generation of new disk drives endeavors to replace large, cumbersome, sensitive and costly devices with smaller, cheaper more robust and more reliable units, and always with greater storage capacity. That's why today, Enterprise Class 3.5" HDD's are available in capacities of up to 12 TB, at a cost of about $0.04 per gigabyte.
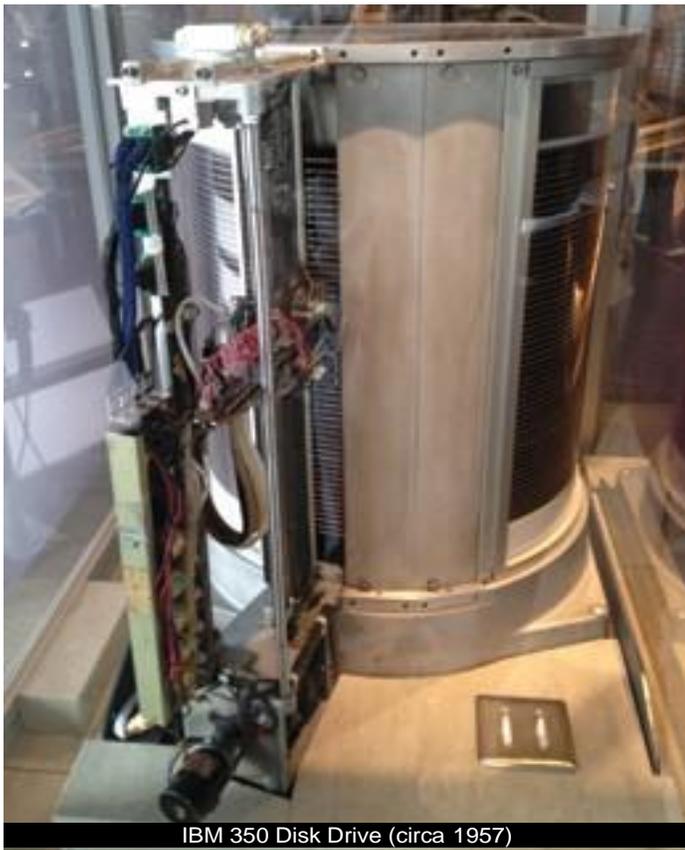
IBM 350 Disk Drive (circa 1957)

The latest offshoot of disk drive technology is the Solid State Disk (SSD). While the basic technology goes back to the 1950's, with magnetic core memory and charged capacitor read-only storage, the development of SSD's was hindered by the rapidly improving and lower cost electro-

mechanical HDD's. Modern SSD's have specific advantages over HDD's, such as the lack of moving components, and in the case of NonStop, enhanced I/O performance, especially random reads and writes. However, the cost per gigabyte for SSD's is about $0.24, or roughly 6x the cost of HDD storage.

The Data Explosion:

It's no secret, the amount of digital information continues to grow, and the rate of growth is rapidly accelerating! In fact, IDC calculated the total amount of digital information created and replicated at 4.4 zettabytes (one zettabyte is equal to 1000 exabytes), way back in 2013. They figure this number will grow to something like 44 zettabytes by 2020, only three years from now. Another starling fact, 90% of all existent data was created in the last 2 years!

And while most of this data is created by individuals, enterprises handle about 85% of this information at some point in its lifecycle. This means organizations take on the responsibility for architecting, delivering and maintaining information technology systems and data storage systems to meet this demand.

So massive amounts of data must and are being archived!

But with massive data archival comes multiple challenges for traditional storage systems. As the petabyte-scale is reached:

• Data integrity suffers when system size is 10 billion times larger than the bit error rate of a hard drive

• Data availability suffers when hundreds of drives fail every day and require lengthy rebuilds

• Data security suffers with millions of devices and multiple copies in multiple locations

• Costs spiral out of control as it becomes necessary to add more and more hardware to meet archival requirements

Fortunately, there are two immediately accessible technologies, data deduplication and object storage, which massively reduce data redundancy and overheads. However, data deduplication, by its very design, has several issues which must be considered prior to new deployments:

With the rise of cloud, mobility, IoT, social and analytics, the data explosion is accelerating.

New, mostly unstructured data sources emerge constantly, creating an expanding data ecosystem for every organization.

75 billion
Internet-connected devices by 2020[2]

90%
of all data was created in the last 2 years[1]

80%
Unstructured Data[3]

Projected Exabytes

Sources:
1. Science Daily, Big Data, for better or worse: 90% of world's data generated over last two year, 2013
2. Business Insider, Morgan Stanley: 75 Billion Devices Will Be Connected to The Internet of Things by 2020, 2013
3. Digital Universe of Opportunities: Rich Data & The Increasing Value of the Internet of Things. EMC Digital Universe with Research & Analysis by IDC, April 2014

- Backup windows increase, especially with dedup VTL appliances, due to front end data processing

- Rehydrating massive amounts of deduped data for restores may become unwieldy

- As the cost of disk drive storage decreases, the case for deduplication of data becomes less significant. And the cost to store a terabyte of data continues to spiral downward…..

- Certain data sets are conducive to dedup; incremental changes from backup to backup favor this technology, others, such as massive imaging data, do not

Yet, data storage, by being a long-term strategy and investment, has made it difficult and usually expensive for CIOs to embrace newer technologies and systems, especially at a time when enterprises are witnessing unprecedented growth data generation and retention. Consequently, many enterprises still use legacy applications running NonStop mission-critical servers with legacy data backup devices. Nonetheless, it is important for players in these sectors to reduce their 'long-term' data backup and archival expense by leveraging low cost advanced technologies available today, such as Cloud Object Storage technology, which has the capability to economically accommodate

Petabyte-scale data backup and long-term retention.

NonStop and Cloud Object Storage:

Object Storage architecture manages data as objects, unlike traditional file systems which manage data as a file hierarchy or block storage which manages data as blocks within sectors and tracks. Each object includes the data itself, some amount of metadata, and a globally unique identifier. Object storage can be implemented at multiple levels, including the object storage device level, the system level, and/or the interface level. The advantage of object storage is it enables capabilities not addressed by older storage architectures. Examples may include interfaces that are directly programmable by the application, a namespace that can span multiple instances of physical hardware, and data management functions like data replication and data distribution at object-level granularity.

There's no doubt, the future of archiving is object storage. Object storage is rapidly replacing on premise tape, disk, and dedup disk technology as an archival or backup methodology. Massive amounts of unstructured data may be retained

efficiently by object storage as it is ideal for purposes such as archiving medical imaging, photos, songs, videos, etc.

This is why Amazon S3, Google Cloud, MS Azure and all other public clouds use object storage. Its flexibility, scalability, and cost are all substantive advantages realized when retaining huge amounts of unstructured data in the cloud. Object storage (outside of back up and archive) can also be flexible with users being able to access data from anywhere.

For NonStop users in particular, there are definite concerns when contemplating a move from legacy tape or disk archiving methodologies to Cloud Object Storage, Generally these concerns would fall into one of four categories: Scalability, Security, Performance and Cost:

Scalability:

- Dispersed storage technology, employing Information Dispersal Algorithms, available with Cloud Object Storage, provides massive scalability with significantly reduced administrative overhead.

- Cloud Object Storage can grow easily from terabytes to petabytes to exabytes, and may be implemented on premise, or in public, private or hybrid clouds.

- These advanced scalability capabilities are ideally suited to rapidly growing data backup environments.

Security:

- Cloud Object Storage is a highly secure object storage archival technology that has been in the marketplace for 13 years.

- The use of Information Dispersal Algorithms (IDA), otherwise known as Erasure Coding, in addition to AES 256 bit encryption, greatly enhances data security.

  - IDA's separate data into unrecognizable "slices" that are then distributed via network connection to storage nodes locally or across datacenters. Think of it this way, if you store a classic Ferrari in a single garage, a thief can break in, hotwire the car, and make off with it. If you disassemble the Ferrari and store the components in multiple garages, it's very difficult, if not impossible, to steal and then reconstruct the vehicle.

  - IDA eliminates the need for data replication.

  - No complete copy of the data resides in any single storage node, and only a subset of nodes needs to be available in order to fully retrieve the data.

Performance:

- With an appropriate and compatible front end device, ingesting data from any backup application including NonStop, TSM, Commvault, NetBackup, Veeam can be optimized, this will allow rapid data ingestion and caching, reducing backup windows while streaming the data policy based pools to Cloud Object Storage at the back end.

  - With the use of FlashSystem for the cache layer, backup data to Cloud Object Storage at rates of up to 10.4 GB/sec or 37.4 TB/hour per node, and restore data at rates of up to 9.6 GB/sec or 34.6 TB/hour are possible.

Cost:

- Cloud Object Storage typically delivers significantly lower total cost of ownership for storage systems at the multiple petabyte level, reducing or in many cases eliminating the need for data replication and the need for multiple copies

- Cloud Object Storage is 55-60% of the cost per GB of archived data when compared to any dedup VTL

- Cloud Object Storage may be purchased as a service (Storage as a Service, SaaS), thereby eliminating the need to procure and maintain backup hardware on premise or in remote DR sites. Storage capacity can be varied in the short term to deal with peak periods, and increased over the long term as a natural function of data growth.

Forged by Power and Partnership:

For NonStop customers, a great answer is Tributary System's Cloud Object Storage Solution. Employing advanced COS technology coupled with Tributary's proven Storage Director® as the "front end", NonStop customers can transparently take advantage of IBM's Cloud Object Storage® (COS), formerly known as Cleversafe®, without any changes to their NonStop applications.

Storage Director, is a policy-based, tiered, and virtualized software product especially designed

for backup which can be seamlessly integrated with any media, including tapes, disk drives, virtual environments and NonStop or other proprietary environments, plus open systems. Storage Director can group data into different pools and apply different protection policies at different times across any storage medium simultaneously. Tributary Systems has gained a massive strategic edge as it has entered into a synergistic partnership with IBM to revolutionize enterprise cloud data backup/restore, archive and DR. Combining the capabilities of Storage Director while endorsing long-term archival to Object Storage is where Tributary sees the data backup and retention market evolving. Tributary claims to be the only company in the marketplace that can backup all NonStop mission-critical servers using a single solution. In addition to Storage Director's AES 256-bit encryption, data is also erasure-coded in the storage tier. From a performance standpoint, Tributary's solution can ingest data at a rate of 12TB per hour and restore at about 8.5TB per hour. Should a flash storage be used in the cache layer, the ingestion rate goes up to 37TB per hour and restores at 35TB per hour; these are metrics that are unmatched in the market. Thus, Tributary's

IP, when combined with the leading Cloud Object Storage solution imparts exclusive cutting-edge data storage and management capabilities that can be well extended beyond public cloud models— into hybrid and on- premise environments—and also offers double-layered security for NonStop clients. Finally, the Storage Director and IBM Cloud Object Storage solution can be implemented with a monthly fixed cost model unlike all public cloud providers such as AWS who charge customers for accessing their data through additional fees for "puts and gets." Employing a cloud backup solution where costs vary widely from month to month, based on access to their own data, is challenging for most enterprise customers.

Conclusion:

Massive data growth coupled with the continual need to reduce cost, and concurrently improving backup and restore performance may seem like incompatible goals for the NonStop professional. But really, with judicious use of the most modern Cloud Object Storage solutions available today, perhaps we can look at clouds from both sides now; data retention capacities and performance go up while costs go down. That's no illusion at all… ⟳

Shawn Sabanayagam is Chairman and CEO of Tributary Systems. Shawn is intimately involved in the information technology and services industry, particularly in Storage, Go-to-market Strategy, IT Strategy, Professional Services, and Management.

Glenn Garrahan is Director of HPE Business for Tributary Systems. Prior to this, Glenn spent 17 years as a Product and Program Manager with HPE NonStop.