The Data Center Storage Dilemma

Scalability, Durability, Density, Serviceability, and Affordability

Never before has there been such mind boggling digital data growth as there is today. IT organizations are awash in digital data creation with signs that it's growing not abating. More than 90%[1] of all data every created by humankind has been created in the past 2 years. Gartner[2] predicts that in just 4 more years, the year 2020, humankind will generate more than 50 zettabytes (that's 50 billion terabytes) of new data. How is this possible? Where is all this data coming from?

Most of that new data is unstructured coming from machine logs, sensors, media, analytics, business applications, archives, docs, public web, and social media. Machine generated data is growing at more than 15 times[3] the rate of people generated data. Unstructured data in general is growing at approximately 3 times the rate of structured (databases, database applications, email, etc.) data per IDC[4]. In addition files keep getting bigger. Take the example of visual media (film, video, and CCTV). A single hour of HD 1080P recording creates 179GB of data. When the resolution increases to 4K, which is becoming more common, that hour of recording goes up more than 3800% to 6.88 terabytes. By 2020 that resolution will be a minimum of 8K. That same one-hour recording will more than 48000% larger than the 1080P recording at the whopping size of 86 terabytes.

Although file and data growth in general are rapidly and exponentially growing, IT budgets are not keeping up. Storage cost decreases has slowed. Storage media cost per TB has not been declining at the same rate as recent history. Storage system capacity and performance scalability to manage this type of data has generally not been keeping up, meaning more of them and supporting infrastructure, is required to handle that data. To summarize this: unstructured data is growing exponentially; storage costs not declining as fast so not keeping up; storage systems and IT budgets are not growing enough to make up the difference. This is creating a huge gap between what is being created and what can be stored affordably. In other words a lot of unstructured data is discarded or thrown away. That data has value. Potentially, it has enormous value. This is what is called the data center storage dilemma.

Legacy Storage

The most common rebuttal is: …that's why there's public cloud storage. Except public cloud storage will cost more, especially after a year or 2 than just about any on-prem storage including Enterprise class. Reasons to utilize cloud storage include not wanting to be responsible for managing storage, storage infrastructure, the data center, and data center personnel. Other reasons include utilizing it as a target for backup and or archival (primarily longevity), and to simplify business continuity as well as disaster recovery. Whatever the rationale, closing the data creation storage affordability gap is not one of them.

Overcome By Unimagined Growth

Traditional on-prem storage systems were simply not designed for the massive unprecedented unstructured data growth. Most of them have been and still are ALUA or active-active controller architectures. That limits their scalability in both capacity and performance. Data durability is limited by parity RAID. Large capacity drives (the prevalent drive for unstructured data growth) no longer take minutes, hours, or even days to rebuild. No, these systems can take weeks, even months for large HDDs rebuilt as a background task, increasing the risk of more and multiple drive failures in that RAID group. When that happens, data has to be restored from snapshots or backups. But since most data protection has relatively long recovery point objectives (RPO), there will be data loss.

[1] SINTEF (largest independent scientific research firm in Scandinavia) first reported in 2013 and repeated every year since.
[2] Gartner made this prediction at their 2016 Gartner Symposium ITXPO™
[3] IDC Digital Universe of Things By 2020
[4] IDC Worldwide File and Object-Based Storage Forecast 2016-2020

Yet Another Refresh

Then there is the tech refresh problem. Every 3 to 5 years those storage systems have to be refreshed or upgraded. Doing so requires a data migration. Tech refresh data migrations average approximately 9 months (and increasing) with more than 84% failing to meet time and/or budget objectives[5]. There is additional risk of data loss, failed server remediation, and outages. Costs are rarely less than 30%[6] of the new system and frequently more.

Limited Scalability

The scale-out variations of traditional storage systems increase capacity and performance considerably but still don't scale to the levels now required by that mind-boggling unstructured data growth. And they're still limited by parity RAID, limited RPOs, and tech refresh issues. Some can mitigate tech refresh issues, but the problem does not go away.

Object Storage

This is why so many organizations have been moving to or considering object storage systems. Object storage scales massively to handle that unstructured data growth; utilizes commodity off-the-shelf server hardware (COTS) and server side storage drives to lower storage costs; solves the RAID problem through implementation of erasure coding across all nodes; and eliminates tech refresh as an issue by never having to migrate the data because of the way erasure coding or multi-copy mirroring works. In the case of erasure coding it is as simple as adding a new node and retiring an old until all are replaced or upgraded. No data need be migrated because erasure coding breaks down into shards before it is written. It then spreads those shards across multiple drives and nodes. The total number of shards is called the 'width'. The number of shards required to read the data is called the 'breadth' and is a subset of the width. To protect against 3 concurrent drive and or node failures, a typical width/breadth is 12/9. For protection against 6 concurrent drive or node failures it is commonly 16/10. This greatly simplifies tech refresh data migration. As old object storage nodes are removed the system detects missing erasure-coded shards and recreates them on the new nodes automatically. They are technically not migrated, but rather recreated.

Multi-copy mirroring (frequently referred to as triple mirroring), has a similar process. When a node is retired its data, any data that was on that retired node is automatically copied from one of the other copies to new nodes from another copy.

The Consistency Conundrum

Object storage introduces other issues such as eventual consistency and ineffective or inefficient storage density (measured as capacity per rack unit). Most object storage systems guarantee eventual consistency. The problem occurs when an application or individual reads old data from a node on the object storage system that has not been updated with changes that occurred on a different node. That causes errors. Some object storage providers have solved that problem by not allowing reads of the old data. The bigger problem is the ineffective or inefficient storage density.

Fighting Data Center Sprawl

Data center floor space is not cheap today ranging from $12,000 to $24,000 per rack[7] for cloud or co-los and closer to $18,000 per rack for on-premises.[8] Low-density storage consumes more rack space. Object storage systems utilizing COTS are supposed to be simple and inexpensive. They are after all utilizing commodity server and storage media hardware. But their cost escalates quickly when the rack cost and supporting infrastructure costs are added back in. More servers equates into more infrastructure including more cables, transceivers, conduit, switch ports, leaf/top-of-rack switches, core switches, power, cooling, etc. Those costs add up because of this density inefficiency.

[5] Bloor Research Data Migration Report, 2011, Data Migration 2011, Data Migration Market 2013
[6] Based on Dragon Slayer Consulting conversations with EMC, NetApp, HDS professional services organizations
[7] Approximates based on power density, maintenance, caged space, cross connects, bandwidth, VPNs, labor, etc.
[8] Approximates based on ongoing operations including cost to build, maintenance, labor, operations, etc.

Density Is Heavy

Some object storage systems have opted for high-density HDD drawers. These drawers range from 48 to 90 HDDs of the 3.5" form factor. Utilizing 10TB high capacity drives enables up to 900TB in a 4U form factor or 22.5TB per RU. That mitigates the density inefficiency factor, but introduces some other costs. These high-density drawers service the HDDs from the top not the front of the drawer. To replace a failed HDD requires pulling the drawer out of the rack and opening the top. These drawers weigh more than 200 pounds (more than 90 KG) and frequently more than 300 pounds (more than 136 KG). Most administrators are not going to be able to hold up the drawer with one arm while replacing a drive. And do not expect the sliding rails to support that weight. That's just begging for a nasty accident. To replace a drive in reality demands a hydraulic lift (not inexpensive) and a ladder (to access the drawer from above) making the replacement process much slower. And per "Murphy's law" the hydraulic lift and/or ladder are never where they're needed slowing the process down a bit further.

These problems and issues are what the Active Archive System™ aims to address.

Western Digital Active Archive System

The Active Archive System is an object storage system. It is designed from the ground up to specifically solve the problems of storage scalability, data durability, tech refresh/data migration, immediate consistency, efficient and effective density, as well as affordability.

Scalability

The Active Archive System scalability is currently only limited by testing. Today it supports up to 5.88PBs raw per 42U rack utilizing Western Digital HelioSeal™ 10TB HDDs, and up to 35.28 petabytes raw in 6 racks.

Durability

It delivers exceptional durability via its BitSpread™ erasure coding. That erasure coding is both width and breadth selectable. That selectability empowers the admin to increase concurrent HDD failure tolerance without losing data. It also enables the storage admin to minimize required capacity overprovisioning to meet that tolerance. The bit spread erasure code knows how the data is distributed across the hierarchy strengthening that data durability even more. It also makes tech refresh and data migration a "non-problem". Adding drives or compute modules is an online non-disruptive event. So is retiring old ones. Any data that is removed is automatically recreated on the new equipment via the erasure coding. No data migration or project is required.

The Active Archive System provides immediate consistency at all times. It doesn't allow read of old non-updated objects or files…period. In other words, it does not send the acknowledgement (ACK) back until all of the shards have been written.

Density and Serviceability

But where the Active Archive System uniquely shines is efficient and effective density. It has the highest density enclosures on the market supporting up to 98 (3.5") 8 or 10 TB WD HelioSeal™ HDDs for a max total of 980TB raw per 4U. Unlike any other high-density enclosure, the Active Archive System allows HDD access from the front. It utilizes 14 HDD sleds that pull out from the front of the enclosure. There are 7 sleds per high-density enclosure. No hydraulic lifts or ladders are ever required to replace a drive. It's simple, fast, and easily serviceable.

Affordability

That leaves the affordability problem. The Active Archive System is architected with affordability in mind. It always leverages the highest density WD HelioSeal HDDs and is engineered specifically to take advantage of the WD HelioSeal HDD characteristics for optimal performance. To minimize entry costs, an Active Archive System can start as low as 672TB. Scaling is a simple seamless online operation. Management is intuitive and highly automated. And it's engineered with easy low cost serviceability in mind.

Conclusion

The Active Archive System is built to solve the mind boggling unstructured data growth of today and tomorrow. It does so while solving the problems that come with legacy scale-up, legacy scale-out, and object storage systems. And just as importantly it narrows the affordability gap between creating and storing unstructured data.

For more information

Web: Active Archive System

About the author: Marc Staimer, as President of the 18-year-old Dragon Slayer Consulting in Beaverton, OR, is well known for his in-depth and keen understanding of user problems, especially with storage, networking, applications, and virtualization. Marc has published thousands of technology articles and tips from the user perspective for internationally renowned online trades including many of TechTarget's Searchxxx.com websites as well as Network Computing. Marc has additionally delivered hundreds of white papers, webinars, and seminars to many well-known industry giants such as: Brocade, Cisco, DELL, EMC, Emulex (Avago), HDS, HP, LSI (Avago), Mellanox, NEC, NetApp, Oracle, QLogic, SanDisk; as well as smaller, less well-known vendors/startups including: Asigra, Clustrix, Condusiv, DH2i, Diablo, FalconStor, Gridstore, Nexenta, Neuxpower, NetEx, NoviFlow, Permabit, Qumulo, StorONE, Tegile, and many more. His speaking engagements are always well attended, often standing room only, because of the pragmatic, immediately useful information provided. Marc can be reached at marcstaimer@me.com, (503)-579-3763, in Beaverton OR, 97007.