Datatility

A case study in making the Cloud a big data opportunity by utilizing the IBM Cloud Object Storage solution

Overview

The need
Datatility needed a cost-effective, highly resilient and durable, multitenant and multifunctional cloud-storage solution

The solution
Datatility chose to implement the IBM® Cloud Object Storage (COS) solution which runs on industry-standard hardware, offers high data resilience and durability, and requires less storage capacity to help with cost-savings

The benefit
Datatility enhanced their cloud environment in three ways –
(i) Unlimited scalability by expanding capacity through the addition of IBM® Cloud Object Storage Slicestor® nodes,
(ii) Improved performance by adding IBM® Cloud Object Storage Accesser® nodes, and
(iii) Increased resiliency cost-effectively by increasing the distribution of data around the country and/or around the world

Datatility is a leading provider of data center services known for delivering innovative solutions that keep businesses running smoothly and without interruption. Founded in 2003, Datatility's operations are based within the Equinix data facility in Ashburn, Va., and consist of a dedicated team of network and storage engineers. Datatility's solutions address a wide range of data storage and network support needs, including storage, backup as a service, disaster recovery and network services.

From data explosion to Cloud storage demand

In recent years, Datatility has seen demand escalate swiftly for tiered storage solutions in the cloud. The result: data explosion across all their verticals and, particularly, in the financial services, media, legal, healthcare and government sectors. Realizing that a large percentage of organizations' unstructured data residing in their primary storage could be migrated to lower-cost cloud storage, Datatility CEO Jan Rosenberg set the addition of a multifunctional cloud-storage solution as a high priority business goal.

Datatility realized quickly that object-based storage "rather than file-based storage" would better fit its demanding storage requirements. The self-managing, self-healing nature of object-based storage would allow clients to store and protect data on cost-optimized SATA disk drives.

Solution components

Software

IBM® Cloud Object Storage Manager
IBM® Cloud Object Storage Accesser®
IBM® Cloud Object Storage Slicetor®
QStar Archive Manager

Identifying the right storage technology

To develop its cloud storage solution, Datatility in mid-2012 began an extensive review of storage technologies and providers. To meet Datatility's needs, the storage technology had to meet several key requirements, including:

- Reduced data storage costs by eliminating RAID and replication overhead
- High performance, reliability and security
- World-class scalability and easy expandability
- Built-in data protection and integrity checks
- Automatic dispersal to multiple sites

Datatility realized quickly that object-based storage "rather than file-based storage" would better fit its demanding storage requirements. The self-managing, self-healing nature of object-based storage would allow clients to store and protect data on cost-optimized SATA disk drives

Identifying the right technology partners

Datatility started with a short list of six storage technology providers that was reduced to two after initial discussions: a large, brand-name storage provider and IBM. These two companies and their products shared several critical attributes, from product features and company pedigree, to an ability to provide a total solution while combining software and hardware into one product.

The selection decision centered on cost and resiliency of storage. IBM COS uses erasure coding, a type of forward error correction, that offers far higher data resilience than RAID. IBM COS also requires far less storage capacity than traditional object storage solutions, which rely on replication to provide high availability. With these two critical technology distinctions, the IBM COS technology more closely aligned with the cost-saving and built-in data needs of Datatility's original solution requirements.

Expanding connectivity with QStar

With user data emanating from a variety of applications, Datatility sought to augment IBM COS System's native RESTful interfaces to increase user connectivity. IBM recommended QStar's Archive Manager to Datatility as a software gateway to provide both CIFS and NFS interfaces to client applications.

The QStar gateway provides a standard file system to external users and their data, yet stores files in the form of objects in the IBM COS System environment through IBM COS Accesser® nodes (access nodes) and IBM COS Slicestor® nodes (storage nodes). This blended solution allows for the file-based approach users prefer on the front end, while utilizing an object-based approach on the storage side to help maintain cost, security and reliability benefits.

The information dispersal advantages

Projecting that users would store roughly 50TB per month, Datatility determined that its cloud storage solution also needed a simple data-growth model for users, as well as a way to keep upfront capital expenditures as low as possible. IBM COS employs Information Dispersal Algorithms to virtualize, slice and disperse the data to separate storage nodes (Slicestors) rather than replicating objects to multiple sites. In Datatility's case, a replication approach across three sites would require 1.35PB of raw disk storage. IBM COS's information dispersal helps Datatility's solution provide much higher data reliability, availability and security than is achievable using traditional RAID and replication protection schemes with far less raw disk storage requirements of just 770TB.

Datatility's dispersed storage architecture

IBM COS Dispersed Storage architecture allows customers to select the level of data protection they desire, which in turn determines the raw storage required. After discussions with the IBM COS infrastructure team, Datatility chose a 12/7/8 configuration. As the diagram in Figure 1 illustrates, this requires 12 Slicestors distributed across three sites (Ashburn, Chicago and New York), or four Slicestors per site, equivalent to a width of 12. Next, Datatility needed to determine its redundancy levels. The value of 7 indicates that data can be read even if only 7 of the 12 storage nodes are available. Conversely, even if 5 storage nodes are unavailable, the data remains accessible. The value of 8 refers to the write threshold and indicates that data can still be written to the IBM COS System as long as 8 of the 12 storage nodes are available. This configuration was chosen so that a whole data center PLUS one additional storage node could be unavailable while still enabling the user to read their data.

The 12/7/8 configuration requires only 770TB of raw disk storage capacity to protect the 450TB of usable storage capacity (expansion factor of 1.71) and returns ten nines (99.99999999 percent) of data reliability, seven nines (99.99999 percent) of read availability and eight nines (99.999999 percent) of write availability. If required, the IBM COS dispersed storage environment can be configured to be 100 million times more reliable than RAID.

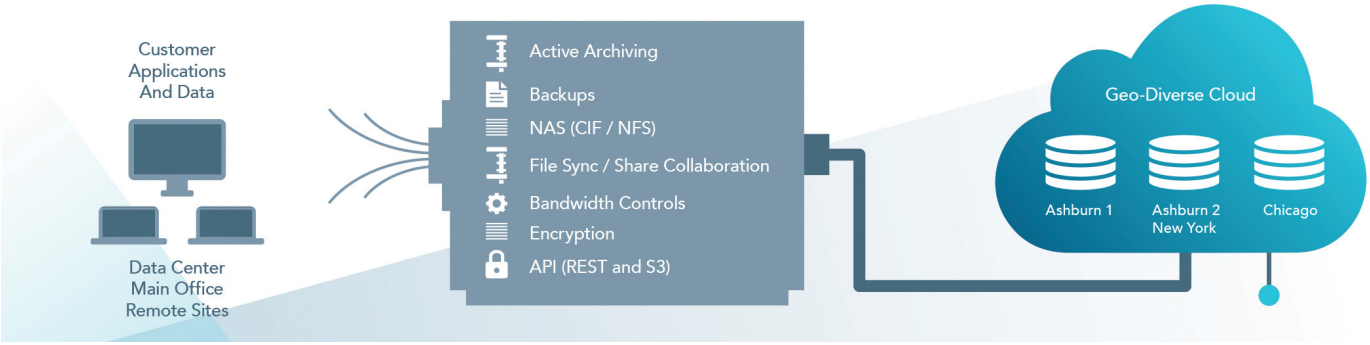

Figure 1: This diagram shows 12 Slicestor nodes dispersed across 3 geographically separate data center locations and a 12/7/8 configuration, which allows an entire site (for example, in this graphic, Chicago) to go down while still providing the users read and write access to their data from the other two remaining sites

Hydra: The big data storage solution goes to market

Datatility's Multifunctional Cloud Storage Platform, called Hydra Cloud, launched in September 2013. This versatile, world-class and enterprise-grade storage solution for big data was named "Hydra" after the mythical nine-headed beast that grew back two heads when one was cut off. Many of Datatility's initial customers will be Equinix users looking to add cost-effective, cloud-based capacity to their current storage infrastructure. They include financial services, media, legal, and healthcare companies, as well as government agencies.

Figure 2: Datatility Hydra: A multitenant, multifunction cloud storage platform

Multi-tenancy capability

Another key element of the Datatility solution is a multi-tenancy capability. IBM COS creates independent "vaults" for each user, which segregates the storage environment to help ensure that users have access to only their data. The QStar Archive Manager gateway is configured to match these vaults, exporting them individually via a network file system, or NFS, as multiple network mount points.

Datatility has chosen to install Archive Manager on a CentOS Linux platform. It provides high resiliency by using multiple instances of the gateway software in a high availability mode.

Always ready for expansion

Datatility's Hydra Cloud storage capacity with IBM COS and QStar is designed for 1 petabyte of storage, which can be increased to support customer requirements. The Datatility Cloud environment can be scaled in three ways. Capacity can expand through the addition of IBM COS Slicestor® storage nodes. Performance can improve by adding IBM COS Accesser® nodes. And data protection can be enhanced by adding data centers and increasing the distribution of data around the country and/or around the world.

About QStar® Technologies

Founded in 1987, QStar Technologies, Inc. is a leading global provider of enterprise-class archive and data management software solutions. QStar software is a key element to creating a robust vendor-neutral active archive as well as delivering secure, cost-effective and reliable protection of valuable digital assets.

About IBM Cloud Object Storage

IBM Cloud Object Storage provides organizations the flexibility, scale and simplicity required to store, manage and access today's rapidly growing unstructured data in a hybrid cloud environment. Relied upon by some of the world's largest repositories, our proven solutions turn storage challenges into business advantage by reducing storage costs while reliably supporting both traditional and emerging cloud-born workloads for enterprise mobile, social, analytics and cognitive computing. IBM Cloud Object Storage is built on technology from object storage leader Cleversafe, acquired by IBM in 2015.

For more information

Please call: 312-423-6640 or email: sales@cleversafe.com or visit the Cleversafe website: www.cleversafe.com.

To learn more about IBM Cloud Object Storage, please visit http://www.ibm.com/cloud-computing/infrastructure/object-storage/

IBM Corporation
Route 100
Somers, NY 10589

Produced in the United States of America
July 2016

Please Recycle

TSC03375-USEN-02