



**What is Object Storage and what can you use it for?**

*Prepared by: George Crump, Lead Analyst  
Prepared: December 2015*

For an increasing number of organizations, the traditional file server may have outlived its usefulness. File servers were designed in an era where most employees were in a single location; road warriors were a rare breed and the only files being created were user generated office productivity files. Now many organizations have employees all over the world; road warriors are commonplace and machines (sensors) generate far more data than users ever will. Data is now *the* product for many organizations and the standard file server is ill-equipped to meet the requirements of the modern data center.

### **What is Object Storage?**

For decades, most file data was stored in a POSIX (Portable Operating System Interface) file system. POSIX is a family of standards specified by the IEEE Computer Society. It is a means of organizing the way data is written to a storage system. Within it, data is organized by a series of volumes, directories (also known as folders) and sub-directories. The data about data, known as metadata, that POSIX stores is relatively simple and includes information like date created, date changed, hierarchy, and archive status.

Object storage is also a technology that organizes the way data is written to a storage system. In it, data is written into self-contained entities called objects. Think of an object as a file. Unlike a POSIX file system, an object storage system gives each object a unique ID, which is managed in a flat index. When a user or application needs access to a file, they provide the object storage system with the unique ID. This flat index provides greater scalability, enabling an object storage system to support faster access to a much higher quantity of objects or files as compared to most POSIX based NAS counterparts.

The flat organizational structure also enables object storage to provide a much richer metadata component for the object. Most object storage systems, in addition to date created and date modified information, can store expiration dates, object protection requirements as well as descriptive tagging about the objects they store.

Architecturally most object storage systems are software defined scale-out architectures that can leverage commodity servers and storage. With this type of design, IT planners can add additional storage nodes as their capacity demands grow. These systems are ideal for a use case where a lot of data needs to be stored for a long period. That data can be data that is large per object like video, images, and audio. It can also be data that is very small per object but contains billions of objects like sensor data or data from Internet of Things (IoT) devices.

### **When Should Object Storage be Used?**

The case for object storage grows stronger when the number of objects (or files) increases; the speed of ingesting or access decreases and the importance of long-term data durability, becomes more critical. Its advantages also become more obvious as the software that interfaces with the system leverages its metadata capabilities to provide a more rich categorization of the data being stored on it. One of the strengths of object storage is its ability to solve a variety of storage management challenges.

### **File Sync and Share, and Distribution**

A key challenge facing a broad cross-section of data centers is what to do about file sync and share (FSS) made popular by consumer grade solutions like Dropbox. Users like the idea of having their data available to them anywhere and on any device as well as the ability to share that data with colleagues and business partners. The problem is the consumer solutions are often not secure and are almost always outside of the purview of IT. With consumer FSS

solutions, data may never be stored on a corporate file server, making it impossible for IT to protect it.

Fortunately, there are several software solutions available that can work with object storage systems to provide an on-premises alternative. These solutions can synchronize data between a user's various devices as well as provide the ability for a user to share data with a colleague. These solutions also provide IT oversight to control what data can be shared, with whom and for how long.

While some modern NAS solutions have the ability to scale to support the raw capacity as well as the number of files that a single enterprise will manage in an FFS solution an FSS solution can leverage object storage's unique capabilities to maintain a rich metadata history so that important attributes and policies can be tracked. For example, an FSS solution could embed sharing permissions, expiration dates, number of download limits as well as number of copy limits. Also because object storage tends to use off-the-shelf hardware and higher capacity hard disk the cost of the FSS solution is often less expensive. Finally FSS with object is typically just a first step for an organization's use of the object storage. It paves the way for some of the other use cases described below.

### **Object Storage as an Archive Repository**

If the total capacity of the data center is categorized, as a rule of thumb most data centers will find that less than 25% of the capacity is active and requires high-performance storage. In this active data set, more performance typically equals higher productivity or greater revenue generating opportunities. This data is best served by an all-flash or hybrid array primary storage array. The remaining 75% of data is made up of user file data, machine-generated or sensor data, backup data, copy data, and long-term archival data.

The object storage system is an ideal archive storage area, and archive especially with the simplification that object storage can bring to the process, is something that data centers should explore. The problem is that many data centers today count on their disk backup appliance and backup software as a pseudo-archive, retaining backup data for a long period and never removing it from primary storage. Using backup as archive means that backups of primary data will continue even if most of that data is unchanged. It also means that backup catalogs and disk backup appliance capacity requirements will increase.

Object storage can complement a disk backup appliance. Adding object storage as an archive tier and moving inactive data to that tier can reduce primary storage capacities significantly. The archive process also reduces the load on the backup process. It now has less data to backup and less data to store and manage. Object storage is an ideal archive tier because it can provide reasonable response times while being far more cost effective than a disk backup appliance.

To move this data to the object storage system, the customer can leverage a cloud gateway, which translates between NFS/CIFS and object storage, essentially turning the object storage system into a network share. Additionally, some backup software applications can natively leverage object storage as a means to move old backup sets out of the primary backup storage repository.

The archiving or backup software that drives the identification of archivable data and its movement to the object store can leverage object storage's metadata to track expiration data, number of copies to be maintained and whether or not the file is allowed to be modified after being stored in the archive.

## **Privatizing the Public Cloud**

A third area to consider is using object storage as a means to balance the use of the public cloud. The upfront costs of public cloud storage are very appealing, but the longer-term periodic costs do add up as capacity and retention requirements increase. As a result, many companies that began cloud initiatives are now looking for a way out or looking to create a hybrid approach to cloud usage where some of the data is on-premises. The problem is their applications are now written to support cloud storage that is typically object based or at least need to be to support a hybrid model.

In this use case, object storage allows the organizations to move data back and forth between the cloud and on-premises without having to change their applications. Plus object storage systems provide many, if not most, of the same benefits of cloud storage; cost effective capacity, pay-as-you-grow scaling, and high data durability. To that, they add lower latent access of data and increased control over data security.

It is important to note that moving from a public cloud storage model only to a hybrid mode does not return the data center to square one. By leveraging a storage infrastructure similar to the cloud provider, object storage, the data center should see gains in operational efficiency. In addition, the object storage system through replication or data dispersion can easily replicate data to a second or even multiple facilities addressing both disaster recovery and data distribution needs. Data centers can leverage their private cloud for FSS and other solutions and the public cloud as a deep archive. They may also find that the operating cost of the object storage system is so compelling that they shift to a private model only.

## **Internet of Things Data**

Sensors, connected to the Internet, have the potential to change the way businesses create products or manage product lifecycles. These sensors can provide continuous monitoring and data capture of the condition of everything from humans to livestock to household appliances. Organizations can use this data to improve the efficiency of the product or service they offer. By comparing how this captured data changes over time, they can improve the product or service.

IoT devices are capturing very small amounts of data, continuously creating billions if not trillions of files or objects. The capability to compare this data across various time spans or market conditions requires that it all be accessible for a long time and that it be of the same quality.

All of these aspects, high capacity, high file count, and data insurance, make object storage ideal for the storing of sensor data. IoT data also does not typically require high ingest rates since small data sets are trickling in. An object storage system can store all this data in a very cost effective manner and leverage its built-in data checking to ensure data integrity. The addition of more sensors or more detail per sensor increases capacity requirements. Object storage, thanks to its scale-out design, can easily keep up with increases in capacity or retention requirements as well as increased file size due to higher detail per sensor.

The rich metadata of object storage can be used to tag this sensor data with the serial number of the device that created the data, the location or the product that it is on. Also, retention times for each object can be stored in metadata, ensuring that data governance policies are upheld.

## **Video Surveillance**

In addition to IoT, video surveillance data is becoming common, impacting many data centers. Video surveillance has grown rapidly since the introduction of wireless IP-based cameras that

allow their rapid deployment almost anywhere. This data is used by government agencies and private organizations to protect facilities and personnel. In the past, this data was not retained for longer than 30 days, but recent changes to state and federal laws have increased the length of time that agencies and organizations are required to keep it.

Also, some organizations are using video surveillance data for more than just security and protection. For these organizations, video surveillance data is another type of IoT data. Organizations are using video data with facial recognition software, for example, to enhance marketing data.

Once again, object storage is ideal for this type of data, and it highlights the flexibility of object storage technology. In contrast to IoT data, video surveillance data has a much smaller object storage count, but a much higher capacity footprint per file. While data authenticity is important, data retention is not as critical. Object storage can be a low-cost landing area or secondary storage location for this data. Video management software can leverage object storage's rich metadata to tag files with important details like camera location, retention requirement and data protection requirement. Also, object storage's metadata can be used to set the file to a read-only status to ensure a chain of custody on the file.

Once again, object storage's ability to provide cost effective but reliable data storage is ideal for this data set. Similar to the other use cases storage response time needs to be acceptable but does not fall into the high-performance category. The enterprise (FSS) solutions can also leverage object storage metadata to set more than just retention times. The software, for example, can also use the object storage's metadata to set sharing restrictions and allowable sharing times.

## **Conclusion**

Almost every data center can benefit in some way from object storage, but it is a complimentary storage technology. There will likely always be a need for high-performance primary storage, NAS storage, and data protection storage. Object storage can be used to offload those storage tiers with a technology that is designed to cost effectively store billions, if not trillions, of files while ensuring the long-term data integrity. More importantly Object Storage provides simplified managing and scaling of capacity while keeping costs under control thanks to its use of commodity x86 hardware and high capacity disk drives.

*This document was developed with [IBM](#) funding. Although the document may utilize publicly available material from various vendors, including IBM, it does not necessarily reflect the positions of such vendors on the issues addressed in this document.*