Understanding Object Storage

Written by Ray Quattromini – Fortuna Data
24th May 2018

Object Storage & How It Works

Object storage is a method for storing millions of files and petabytes of data on storage nodes using a global namespace and file system. It is vastly more scalable than traditional file or block storage.

It is estimated within the next 7 years 80% of the worlds data will be stored on object storage.

Need additional disk capacity, just add another storage node. All object data is automatically distributed across the nodes, you can create object replicas or shard data (see below) across the nodes to aid recovery and increase performance.

Provides up to 75% cost savings compared to primary NAS filers whilst providing 100% business continuity.

How does object storage work?

Object storage requires 3 things to work.

1. The data and this could be anything

2. Metadata provides a customisable table and index of the object being stored

3. A globally unique identifier assigned to every object
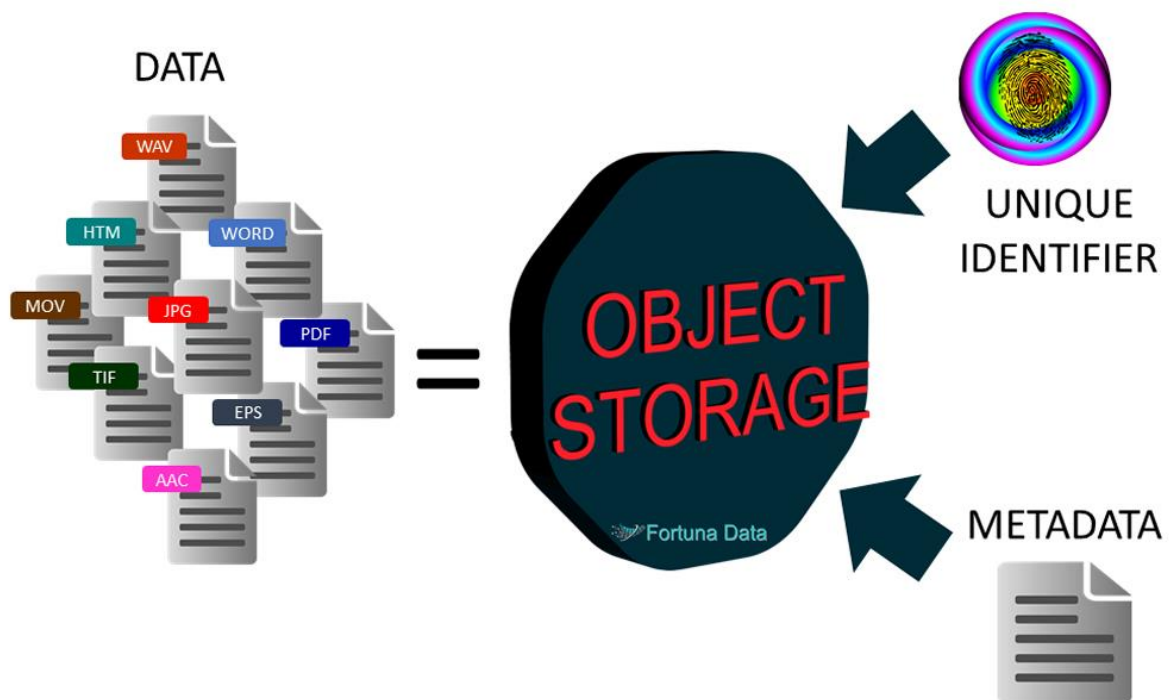
Object storage uses metadata to manage the millions/billions of objects it contains. Metadata is created by the administrator of the object store and it contains information relating to each type of object it is storing e.g. a photo, metadata information could be:

- GPS location	- Image size	- Created by	- Resolution
- Date taken	- Create duplicate copy	- Confidential	- Expiration date

The metadata could contain far more information relating to the object type than a file or block. Everything stored within your object storage is given a unique metadata identifier and every piece of data has the metadata and file combined into a single object.

An object could be any type of file. When it comes to object storage, metadata resides in the objects themselves. There is no need to build databases to associate metadata with the objects. Custom metadata can be created about an object file based on contents, dates, user information, permissions, etc. Attributes can be changed, added or deleted over time. Metadata is highly searchable and allows you to find files far faster than file or block storage. You can conduct searches that return a set of objects that meet specific criteria, such as what percentage of object are of a certain type or created by owner "x". This allows companies to extract insights from the big data they possess within their data and identify trends.

Files/Folders

Today, many businesses use files and folders to store information, typically a NAS. An operating system creates a file system that manages the data as a file hierarchy and this consists of:

1. Files
2. Folders
3. Drive letters or mount points

Files are retrieved using the operating system such as File Explorer. The data resides at a physical location i.e. \\server1\share1\drawings or \\10.10.1.10\share1\drawings.

Block Storage

Block storage is used primarily in Fibre Channel or iSCSI SANs and serves up data blocks to a host server on which the operating system or application can be loaded. Block storage handles data by evenly distributing the data block across the available storage system and giving each block a unique ID.

Object Storage Data Protection Methods

There are two types of data protection available depending on the object storage vendor. The first type is storage nodes using RAID as the backend for data protection, the second type is using software and JBOD storage. Our preferred method is software and JBOD for the reasons below:

Object storage is not ideally suited to using RAID and provides each drive with a unique ID. If a drive fails, the software automatically rebuilds just this drive and only the amount of disk space the object uses. This is unlike a traditional RAID that stripes data across all drives and requires all drives within the RAID set to rebuild a single drive. Therefore, rebuild times for large capacity disk drives is drastically reduced. With object storage the disk or SSD devices used to store the data do not have to be all the same capacity unlike RAID. Object storage takes care of where the object data resides add 10TB drives to your 4TB nodes and the data is evenly distributed across all nodes, although performance could be affected by having different brands, transfer rates and capacities of drives.

There is no need for IT to immediately replace a failed drive like they must do with RAID-based systems. Failed drives can be replaced during regular maintenance intervals or not at all (fail-in-place).

Object Management

The object storage software can manage, repair, retire, delete and deduplicate objects themselves based on information provided in metadata, making information life cycle management extremely

efficient. If you need to store a certain object for a longer, you can amend the metadata associated with this object and all objects with the same metadata are automatically updated.

Object storage could have object storage pools distributed globally. By distributing the object storage pools your data now doesn't reside at a single location and therefore by default you have a DR site setup. Each site acts as an independent DR site and you could have a geo cluster configuration for data protection on each site.

Object Storage Example

Imagine if you start with a single object storage pool based in London and this consists of 8 nodes and you create 2 replicas of each object. You open an office in Manchester you could put 4 of the nodes in the Manchester office and utilises the global name space. Using the traditional file system approach, a user in Manchester would need to remotely connect the file residing the London office and pull files back across the WAN, when updates are done the file is transmitted back across the WAN to London. By using object storage, the user in Manchester is able view the file locally and make any changes. This is where object storage is cleverer than file storage, if the user amends the object (file) it is changed locally and then the change is automatically replicated to the other storage nodes including the ones in London.

Global Namespace

Object storage uses a global namespace and is designed to reduce the number of mount points / shares in an organisation. Having a global namespace means your underlying storage layer could be distributed across geographic locations, the user would not know if the files are local or remote and could be located thousands of miles away.

The object (data) does not reside on a single storage platform and could be located anywhere in the world, instead it uses a virtual name space to create a http:\\ or https:\\ path or accessed using an API or use NFS/CIFS and mount the object store as a network share.
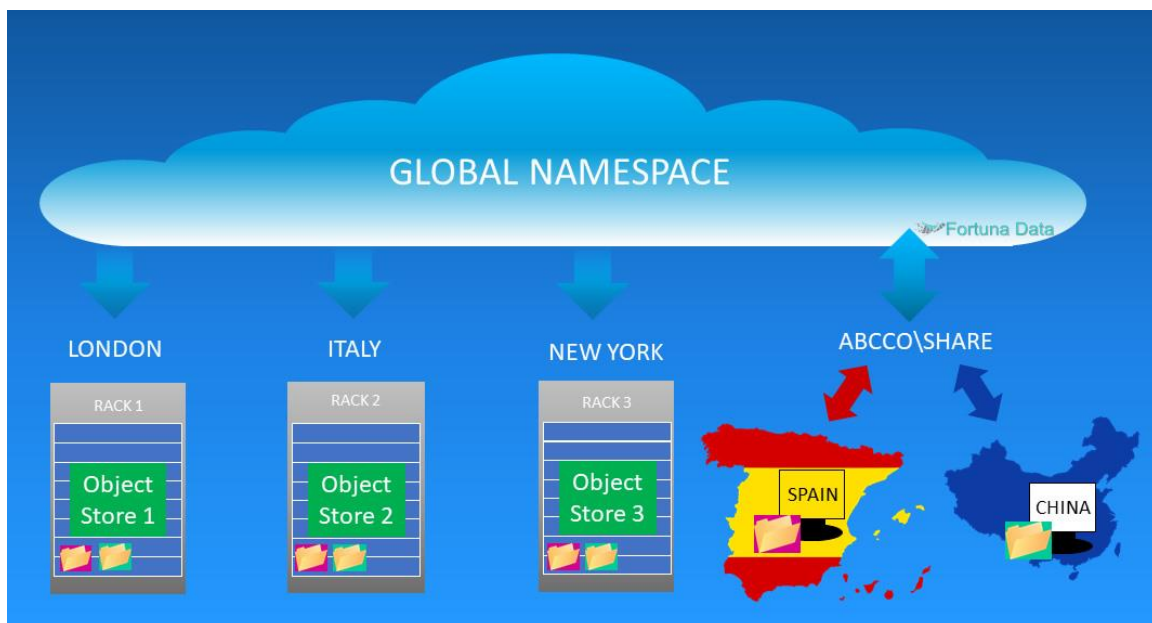
Disk Space Utilisation

Object storage provides far better disk space utilisation than block or file storage. Data is written in 512 bytes or 4k sectors. File systems write the data to fill the disk based on the file size. Our file size is 3k, this file fills 3k of sector 1 leaving 1k of free space. The next file is 2k this fills the sector 2 with 2k remaining and so this continues until the disk starts to be come full. We now have a disk with free space in the sectors and the next file is 3k, this 3k file is now written to the disk filling the remaining space in sector 1 & 2. As this continues, the disk becomes fragmented and slower as the disk head must move about over the disk platter to access the files. Every file that is written creates a parity bit so that in the event of disk failure the system understands how to rebuild a failed drive.
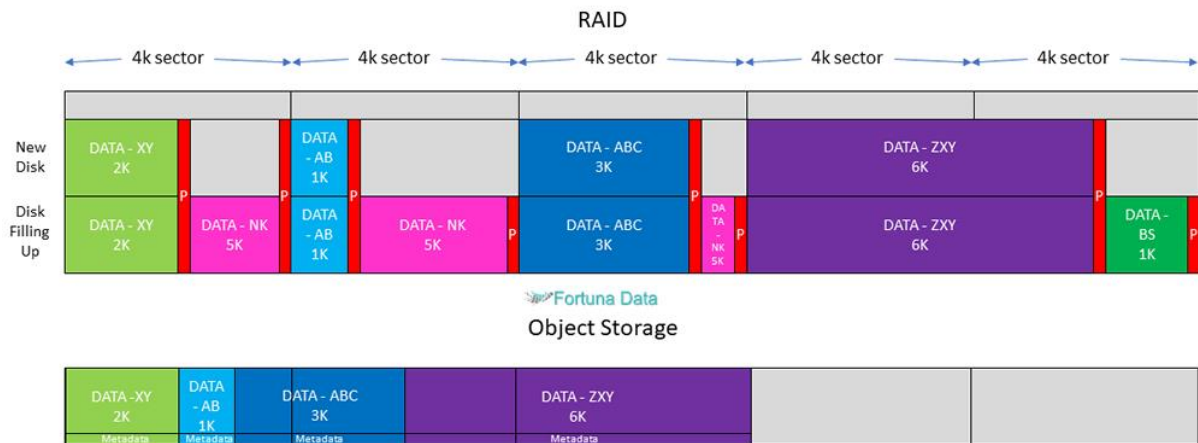
Object Storage Erasure Coding vs Replicas

There are primarily two types of object storage both provide data protection but work in different ways.

Erasure Coding

This is the algorithm that divides an object into chunks, a subset of which may be lost without causing object loss. Erasure coding is far more disk efficient than file/block or object storage using replicas.

Erasure coding provides a very high level of durability, specified up to 17 nines (99.999999999999999%), in support of the most demanding enterprise and service provider environments. This level of durability corresponds to an average annual expected loss of merely 0.000000000000001% of stored objects.

Erasure coding ensures data is far more secure than using RAID, by breaking the data in to manageable chunks e.g. 18 and distributing these chunks across your nodes, this is sometimes call BIT Spread, provides a high degree of data protection that means you would need to lose 5 data chunks to lose data, based on a single copy. Data consistency ensures that the data you read is the same data you wrote and is sometimes referred to as BIT Dynamics, this prevents data corruption at rest, by continually monitoring and checking the data for any loss and re-writing the data to ensure complete data integrity. Both these tasks are done transparently and automatically to ensure object data is always available and consistent.

Replicas

Unlike erasure coding a replica is a copy of the original object and these replicas are then replicated throughout the object store to provide data protection. Depending on how many replicas you need

determines how much disk space you require. Object storage using replicas is not as space or performance efficient as erasure coding, it also doesn't offer the same level of data protection.

Who is using Object Storage?

If you are using cloud storage there is a high probability that the service you are using albeit Google, Facebook, Dropbox, Apple, Microsoft or Amazon to store data then you are using object storage. Private Cloud is an ideal application for object storage as it scales for capacity, performance, reliability & redundancy. You could for example create a local private cloud in your datacentre and scale this out globally.
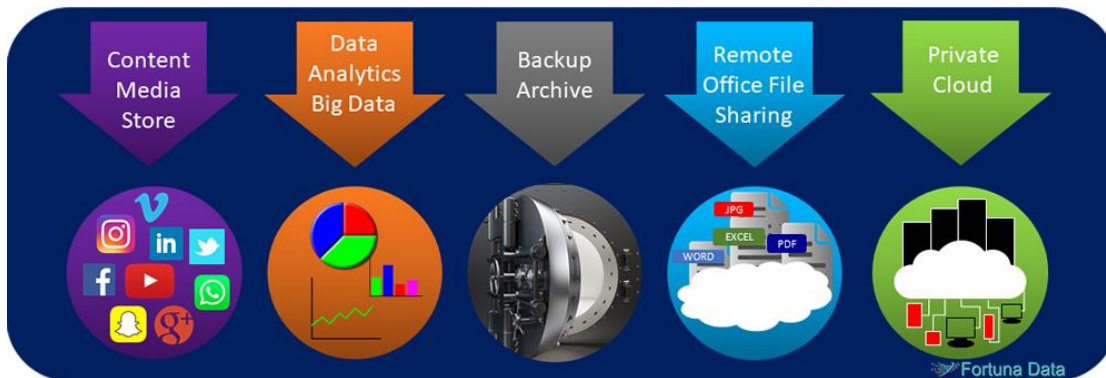

Data Growth

Object storage is ideal for managing data growth. Traditional storage has scaling limitations based on the number files/folders/capacity it can store and is limited by how many controllers and host connections it can support. Object storage grows by adding additional nodes, there is no data management or RAID configuration necessary. All you need to do is inform the object storage software that you have added an additional "X" node and the system takes care of the rest. Object storage can scale from TB's to PB's and grows as a single namespace across the entire workplace.
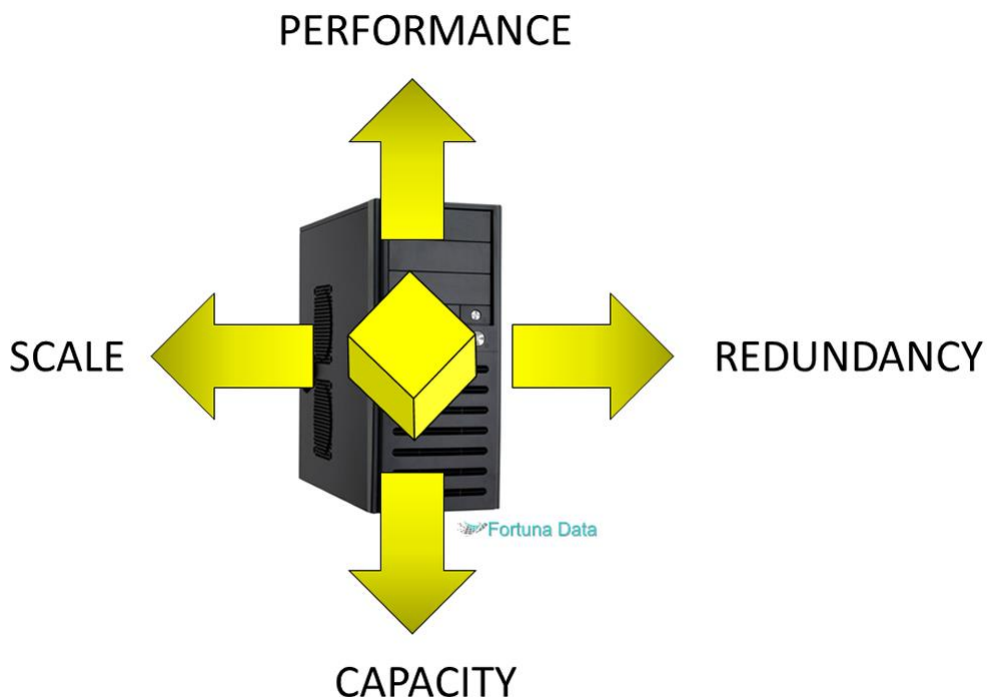
Optimal for unstructured data

Unstructured data is the biggest data management headache for companies today. They are being bombarded with all types of files from video, messaging, email, images, text etc. This type of data is normally stored on a NAS using a mapped share or drive letter. The problem with unstructured data is it is extremely difficult to sift through the information to try and find anything meaningful and useful as there is no structure. Object storage quickly and easily enables you to identify data created by an individual or application and allows you to classify the data for later use.
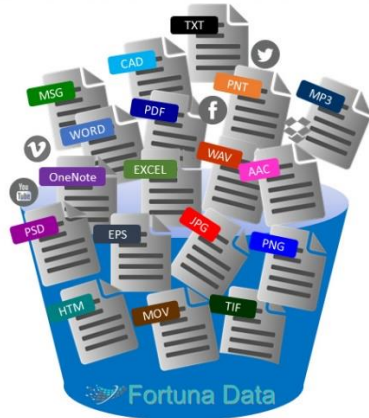

Flat Address Space

Unlike files in file systems, objects are stored in a flat structure. There is just a pool of objects — no folders, directories or hierarchies. You simply ask for a given object by presenting its object ID. Objects may be local or on a cloud server thousands of miles away, but since they are in a flat address space, they are retrieved the same. Each object has one unique object identifier (OID) that allows a server or end user to retrieve it without knowing the physical location of the data it contains. Unlike a file system there are no hierarchies or nested directories to search and retrieving data is faster. This flat address space approach easily scales from terabytes to petabytes and beyond.

Ideal Applications for Object Storage

- HPC Secondary storage layer where large datasets are not being changed
- Database dumps or log files that may be required in the event of system failure
- Web Content Store - By and large web pages tend to remain static with the occasional update
- Large-File Workflows – Media & Entertainment generate large volumes of information relating to i.e. a movie
- Document Sync & Share – Create your own Drobox type system in-house
- Backup & Disaster Recovery – Ideal as a backup target for protecting all types of backup data
- Private Cloud – Start with an in-house cloud and scale out
- CTV Surveillance - Relies on continually recording CCTV footage with zero frame loss or missing vital seconds of footage. This information needs to be viewable and retrievable in seconds with smooth playback of footage.
- Life sciences / Bioscience - generates a huge amount of information from sequencing a strand of DNA or running virtual simulations of a treatment for the flu virus to examining stem cells.
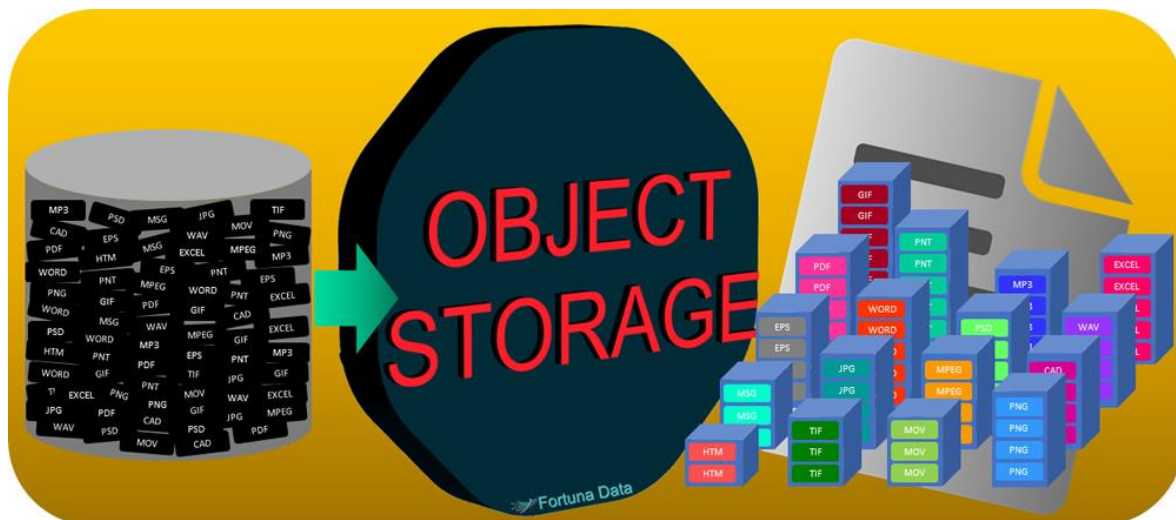
Where object storage isn't a good fit

Any application that requires very low latency or high IOPS such as a databases, transactional data or virtual machine host.

Conclusion

If you are trying to simplify your storage strategy, have a requirement to create a cloud solution, want to rationalise your storage vendors or reduce management and operational costs then object storage could be the answer to streamlining your IT infrastructure. We work with some of the world's leading object storage vendors to deliver a completely flexible architecture for protecting and securing your data.

If you would like to know more, please call us on 01256 331614 or email sales@fortunadata.com