



OBJECT STORAGE 101

CLOUD-A

GEOFF SULLIVAN

LAST UPDATED 08.31.2015

TABLE OF CONTENTS

TABLE OF CONTENTS

WHY OBJECT STORAGE

WHAT IS OBJECT STORAGE?

BENEFITS OF OBJECT STORAGE

SCALABILITY

RELIABILITY

COST EFFECTIVE

WHAT PROBLEMS DOES OBJECT STORAGE SOLVE?

OBJECT STORAGE LIMITATIONS

USE CASES

CONTENT REPOSITORY

LARGE-FILE WORKFLOWS

INDUSTRY SPECIFIC LARGE-FILE WORKFLOW SOLUTION TRENDS

GENOMICS

FILM AND VIDEO

DOCUMENT SYNC & SHARE

SYNC & SHARE SOLUTION TRENDS

BIG DATA REPOSITORY

BIG DATA SOLUTION TRENDS

BACKUP, DISASTER RECOVERY AND ARCHIVAL

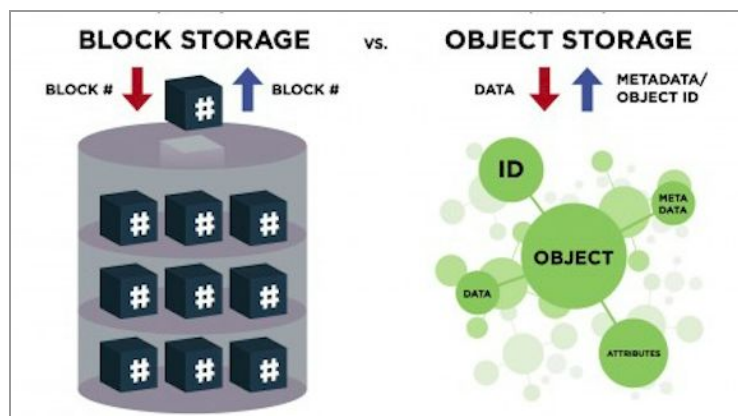
WHY OBJECT STORAGE

We have been seeing increasing demand for our object storage product, [Bulk Storage](#) over the past few months, which is exciting for us because we think that object storage is a phenomenal storage solution for many specific applications. We wanted to provide our users and followers with an overview of object storage technology and highlight some effective use cases. The biggest hurdle we see to adoption of our Bulk Storage product is an understanding of how exactly the technology can be used. As a result of this, a big portion of our job at Cloud-A is to educate our users about existing third party software solutions that integrate with object storage.

WHAT IS OBJECT STORAGE?

Object Storage is an extremely simple yet scalable and reliable file storage system that has been made popular by [Amazon's S3](#) product, the world's most recognized object storage deployment. In addition to S3, [OpenStack Swift](#) is an open source project that allows organizations to build out their own object storage solutions - Swift is what Cloud-A Bulk Storage is built with.

Rather than organizing files in a directory, object storage stores files in a flat organization called containers, and unique identifiers or "keys" are used to retrieve them. Object storage requires less metadata than traditional file systems do to store and access files by grouping metadata with each object. Object storage offers basic functionality, allowing users to store, retrieve, copy, and delete files and most object storage systems provide a [REST API](#) that allows users to programmatically integrate with containers and objects.



BENEFITS OF OBJECT STORAGE

SCALABILITY

The simplicity of object storage makes it incredibly scalable. Scaling an object storage cluster is as simple as adding additional nodes. In a public cloud setting, scalability is virtually infinite, as the public cloud operator owns the responsibility for capacity planning.

RELIABILITY

Object storage achieves reliability by replicating objects across multiple hard disks, nodes, and geographical locations. OpenStack Swift as an example offers a software management layer that provides 3-way replication natively, which ensures that hardware failures are handled promptly and gracefully to enterprise grade drives results in highly reliable storage systems.

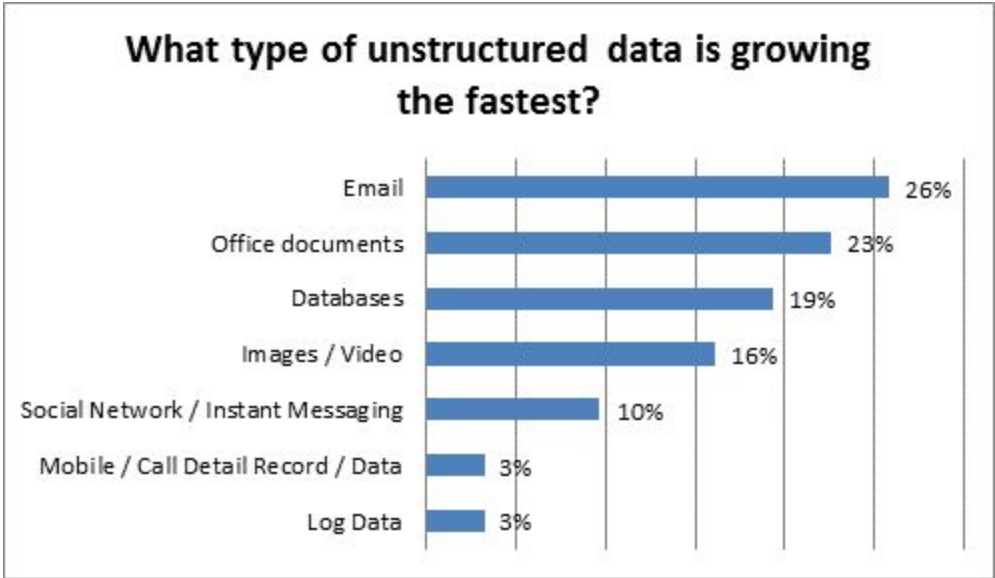
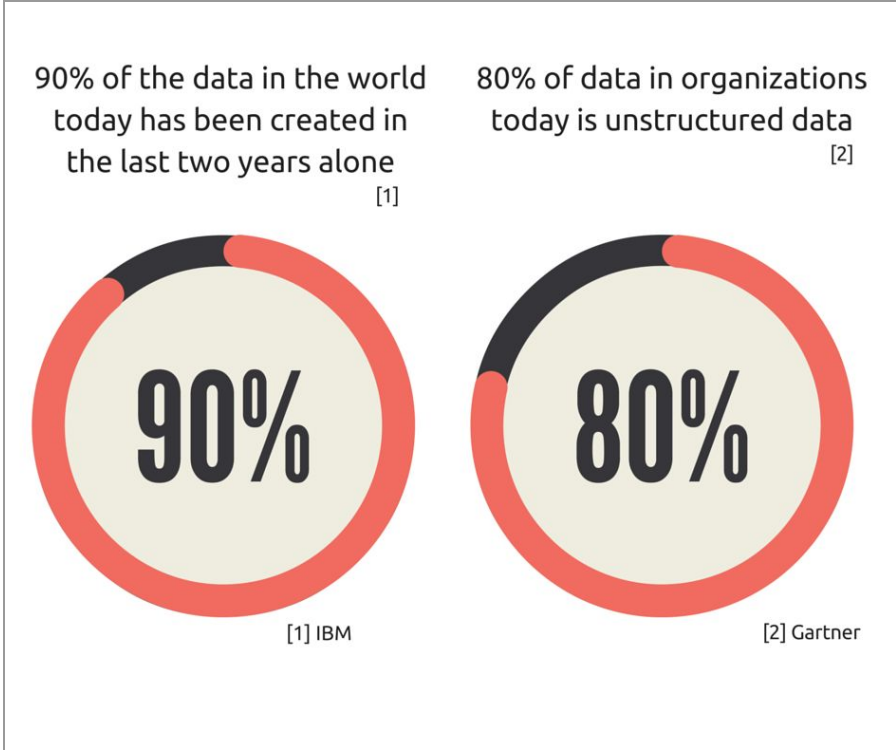
COST EFFECTIVE

Since object storage is “software defined,” it is the software management layer that provides all of the intelligence and redundancy to the object storage system. Because of this, object storage clusters can be built on commodity hardware, avoiding the cost of specialised appliances. This paired with the economies of scale that a cloud provider like Cloud-A achieves enables the price point of object storage to be much lower than traditional network attached storage (NAS) solutions.

WHAT PROBLEMS DOES OBJECT STORAGE SOLVE?

Object storage is an excellent solution for solving the problems associated with data growth. Mega trends like [Big Data](#), Big Media, Machine to Machine (M2M) communications and the Internet of Things (IoT) are creating more data more rapidly than the world has ever seen before, and these are all things that affect everything from the SMB space all the way up to the enterprise. As a result, storage systems must grow to handle the influx of data. Traditional block storage based systems ([SAN](#) and [NAS](#)) are prohibitively expensive and run into durability issues at the 100+ terabyte or multi-petabyte scale. Object storage excels at scale, avoiding all of the conflicts around provisioning management that traditional block storage systems run in to.

UNSTRUCTURED DATA GROWTH: LEADING RESEARCHERS ARE REPORTING MASSIVE GROWTH IN UNSTRUCTURED DATA IN ORGANIZATIONS OF ALL SIZES



SOURCE: Informatica, Symantec

OBJECT STORAGE LIMITATIONS

Object storage is built for unstructured data within cloud applications. The technology was not built to replace block storage for operating systems and databases, which are both structured data.



Object storage is eventually consistent, which means that in an effort to achieve high availability, if new updates are made to a given object, eventually all accesses to that object will return the last updated file. In the case of Cloud-A Bulk Storage, where each object is replicated across the cluster three times, any time that a file is updated, the latest version of that file will be made available only when the change has been made to all copies of the file in the cluster. This is why storing photos, video, and other unstructured data is an ideal use case for object storage systems; it does not need to be constantly altered. The downside to eventual consistency is that there is no guarantee that a read request returns the most recent version of the data.

Consistency is needed for real-time systems such as transactional databases that are constantly being written to, but provide limited scalability and reduced availability as a result of hardware failures. Scalability becomes even more difficult within a geographically distributed system. Strong consistency is a requirement, however, whenever a read request must return the most updated version of the data.

Lastly, and perhaps most importantly, organization's struggle to grasp exactly how to properly leverage object storage. Joe Arnold, founder of [SwiftStack](#) told [OSPod](#) that one of the struggles he sees with his customers, is getting them to look beyond the benefits of object storage and look towards the functional use cases within an organization so that they can leverage the technology to take advantage of these benefits.

USE CASES

CONTENT REPOSITORY

Object storage makes for an excellent solution to house content repositories for web & mobile applications. A single, web-addressable namespace is used rather than putting infrastructure knowledge in the application.

These applications access data via native HTTP RESTful APIs and developers leverage broad API support for programming languages including Java, C, C#, Python, PHP, Ruby, and others.

Recently at Cloud-A we have seen more and more users serving static files from Bulk Storage as an alternative storage backend for content management systems like Drupal or Wordpress.

LARGE-FILE WORKFLOWS

Several industries that store and archive large files like video production, life sciences and research have found object storage to be an extremely useful storage

solution. The scalability of object storage, both in public or private deployments, allows organizations to distribute their data across multiple regions for essentially unlimited amounts of capacity.

Most of these large-file use cases involve industry specific software that has been modernized to support cloud storage APIs like the OpenStack Swift API.

INDUSTRY SPECIFIC LARGE-FILE WORKFLOW SOLUTION TRENDS

GENOMICS

Genomics researchers are using massive amounts of machine generated life sciences data to provide insight to biological and chemical systems which will expand understanding, end suffering and drive funding for Genomics, Clinical and Pharmaceutical Research. This machine generated data is distributed and consumed by the petabyte, which has lead organizations to demand object storage to handle the scale and resiliency requirements for this precious life sciences data.

FILM AND VIDEO

New and emerging technology in the film and video production space has caused a massive influx of data in the industry. The large amounts of raw footage, and multiple iterations created with high end cameras utilizing technologies like high definition, ultra high definition, 4K and 8K can create hundreds of terabytes of data for even small film projects.

Inexpensive web connected surveillance systems, and video processing software is creating massive amounts of unstructured data for organizations to manage. Many organizations require long term retention policies of surveillance data to comply with various governing bodies for many industries.

DOCUMENT SYNC & SHARE

One of the largest consumers of Cloud-A Bulk Storage is third party sync and share solutions. Organizations are champing at the bit for DropBox-like solutions that are under their own control.

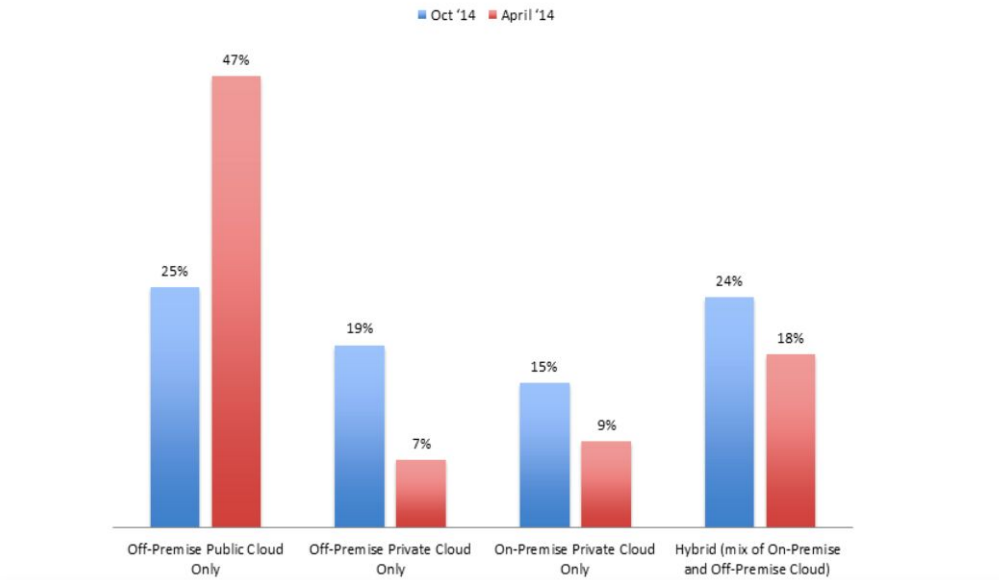
Object Storage provides an excellent storage backend for sync and share as it allows organizations to store unlimited documents and unstructured data, all accessed with no compatibility issues or data lock-in.

SYNC & SHARE SOLUTION TRENDS

A 2014 report from 451 research outlined data that showed that the adoption of on-premises enterprise sync & share solutions increased by 171% as a percentage of the whole market. Adoption of SaaS based sync & share solutions decreased by 46% as a percentage of the whole market

ChangeWave Research, October 2014 - EFSS Deployment Methods

How is File Sync & Share Deployed?



BIG DATA REPOSITORY

Object storage can provide a repository for incoming data, as well as the results of analysis and modeling in big data use cases.

Object storage provides simple storage management, allows organizations to avoid silos and sharding, and fast access to any data either as files or as native objects from Hadoop and other analytic tools.

Utilizing the Hadoop Swift Filesystem client (swiftsf), organizations can leverage OpenStack Swift deployments like Cloud-A Bulk Storage for their [Hadoop](#) workflows. swiftsf allows Hadoop applications like MapReduce and Pig to read and write data directly into OpenStack Swift containers, or in the case of Cloud-A, Bulk Storage containers. swiftsf separates the compute and storage resources of the Hadoop cluster, allowing each to have its own lifespan.

This makes OpenStack Swift clusters like Cloud-A Bulk Storage an ideal, long term repository for data that only needs to be processed by compute periodically. Users can keep their data in Bulk Storage, spin up Hadoop workers to crunch the data, and spin them back down after the job is complete to reduce the cost of the project. Additionally, data that already exists in Bulk Storage containers can be processed without moving the data over to the Hadoop's file system.

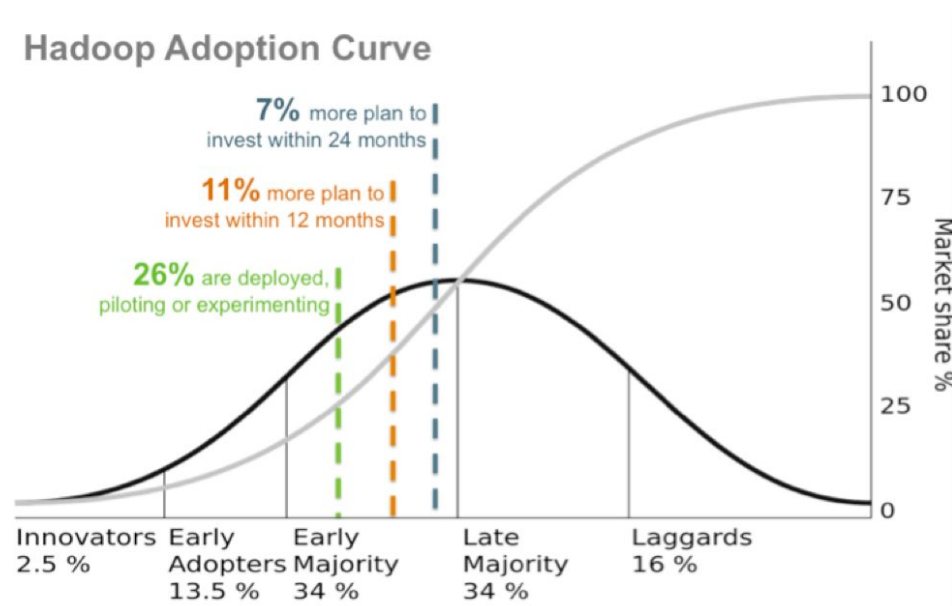
The Hadoop Swift filesystem library defines which Swift object hosts a container - and how to connect to it.

BIG DATA SOLUTION TRENDS

A [recent report](#) by Gartner published in May 2015 stated:

“Despite considerable hype and reported successes for early adopters, 54 percent of survey respondents report no plans to invest at this time, while only 18 percent have plans to invest in Hadoop over the next two years.”

When considering the [technology adoption lifecycle](#), that models adoption across five categories with corresponding market share %s: Innovators (2.5%), Early Adopters (13.5%), Early Majority (34%), Late Majority (34%), and Laggards (16%). Gartner shows that Hadoop is in the Early Majority of the market at the classic inflection point of its adoption curve (see below.)



SOURCE: Gartner

BACKUP, DISASTER RECOVERY AND ARCHIVAL

Another leading use case for Cloud-A Bulk Storage is backup, disaster recovery and archival storage. Object storage provides a resilient and easily scaled storage target for backups of VMs, files, and databases.

As a backup target, object storage can provide a much lower-cost solution than other on-premise storage options. Native data replication in object storage technology like OpenStack Swift ensures that backups remain available even if a server, site or network fails.

FOR MORE INFORMATION CONTACT: GEOFF SULLIVAN
GEOFF@CLOUDA.CA