

GIGAOM RESEARCH

Why Speed Matters for Cloud Storage File Transfers

Barb Goldworm

January 20, 2015

This report is underwritten by Signiant.

Table of Contents

Executive Summary.....	3
Introduction.....	4
Cloud and Cloud Storage Trends.....	4
Information Explosion into and out of the Cloud	5
Vertical Markets	6
Challenges and Limitations of Moving Large Amounts of Data	8
Limitations of Traditional Methods and Alternatives	8
When/Why Speed Matters.....	10
Speed-Sensitive Use Cases	10
Considerations.....	12
Key Questions.....	12
Key Solution Features to Consider	13
Key takeaways	15
About Barb Goldworm	16
About Signiant	16
About Gigaom Research	16

Executive Summary

To fully leverage the power of the cloud and cloud storage, businesses in many industries and of all sizes must move massive files into (and out of) the cloud, creating technical challenges for IT. Traditional methods of transferring data are no longer sufficient.

This report will help IT decision makers, IT managers, directors of cloud infrastructure, and those responsible for ensuring successful use of the cloud, understand the fundamental issues around transferring large amounts of information into and out of the cloud, learn when and why speed matters, evaluate alternative solutions, and establish best practices for cloud-storage large-data transfer.

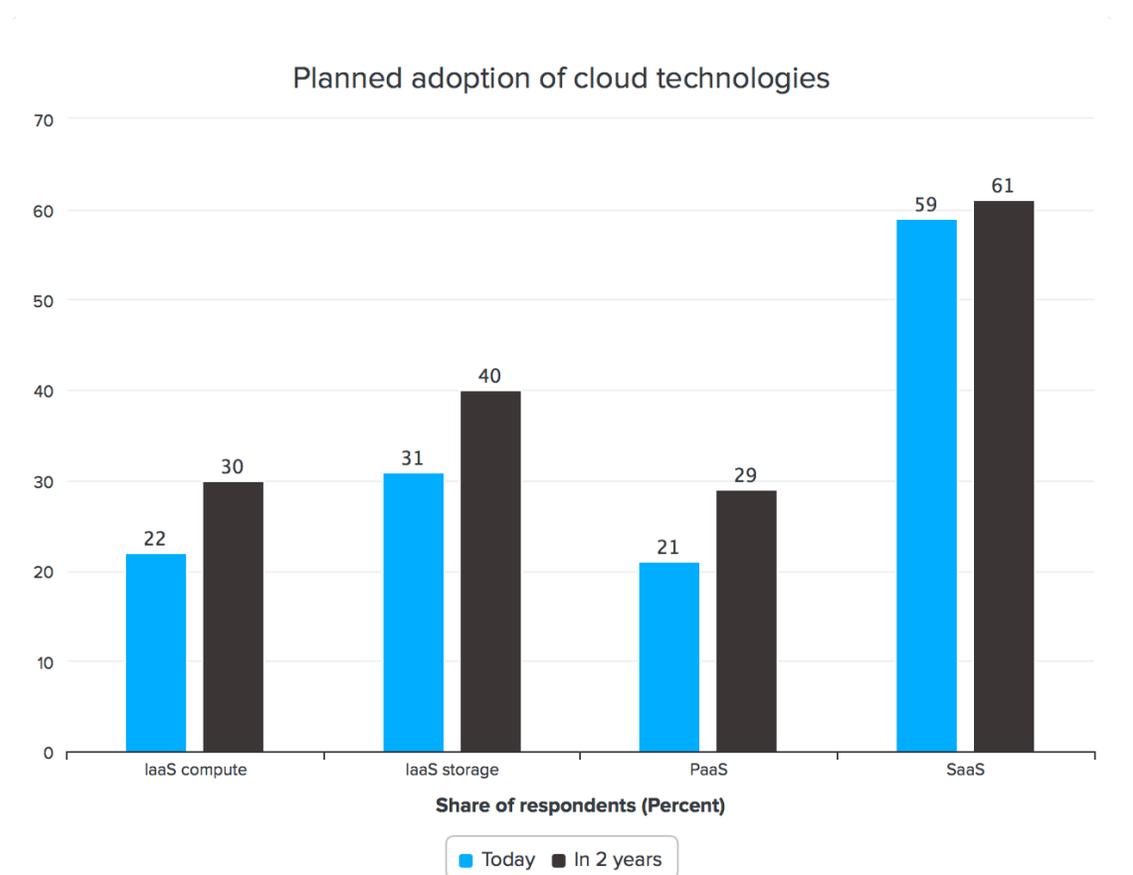
Key findings in this report include:

- Cloud and cloud storage offer big opportunities for organizations to change the way they do business. In many cases, companies now have access to computing capabilities previously unavailable to them due to the cloud pay-as-you-go model. However, without the right data at the right places in the right timeframe, these opportunities will be lost.
- Massive information growth and big data are driving requirements for large file transfer. File sizes in ranging from hundreds of gigabytes (GBs) up to petabytes (PBs) are becoming commonplace and critical across vertical markets (beyond media and entertainment to oil and gas, healthcare, pharmaceuticals, universities, satellite imaging, scientific analysis, and more) as well as horizontal business requirements such as big data analytics, using information as a strategic asset.
- Traditional methods of large file transfer, most based on 40-year-old technology, are no longer sufficient for today's needs. Innovation in data transfer must keep up with innovation in computing.
- Solutions are available today, based on a variety of approaches. Understanding the individual business needs first is important; then look for the solution that offers the best fit. Leveraging the experiences of those who already have solved this problem in traditional IT enterprise environments and the cloud can help smooth the road.

Introduction

Cloud computing has moved to the mainstream and is changing the way organizations compete and do business. Likewise, changes in storage technologies, the explosion of data, and the benefits of big data analytics are changing the way information benefits the business.

Planned Adoption of Cloud Technologies



Gigaom Research Enterprise IT Buyers Survey, 2014 n=502

Cloud and Cloud Storage Trends

The past few years have shown a significant increase in real adoption of the public cloud, as evidenced in the chart above. Enterprise use of cloud storage is now arguably mainstream, and is used primarily for application-specific needs, backup and recovery, disaster recover (DR), and archiving.

Chief concerns about cloud storage include security (far and away number one), reliability and availability, performance, regulatory issues, manageability, and cost. However, the benefits—flexibility,

elasticity, ease of use, simplification/elimination of on-premise infrastructure, and cost—outweigh the concerns at this point. One reason may be the tendency to overbuy storage to allow head room for growth—in other words, spending money before it is necessary. This makes storage an ideal candidate for the cloud pay-as-you-go model.

Today's most common form of cloud storage is object storage because it is massively scalable to billions of objects, overcoming the capacity limitations and management challenges of traditional file systems. It gained popularity in the cloud when companies like Amazon and Facebook used it to store the huge amounts of data they suddenly faced. Object storage is well suited for unstructured data, such as video, and applications such as Hadoop MapReduce for analytics. Use has increased significantly and will continue to grow, both in the cloud and on-premise storage.

Information Explosion into and out of the Cloud

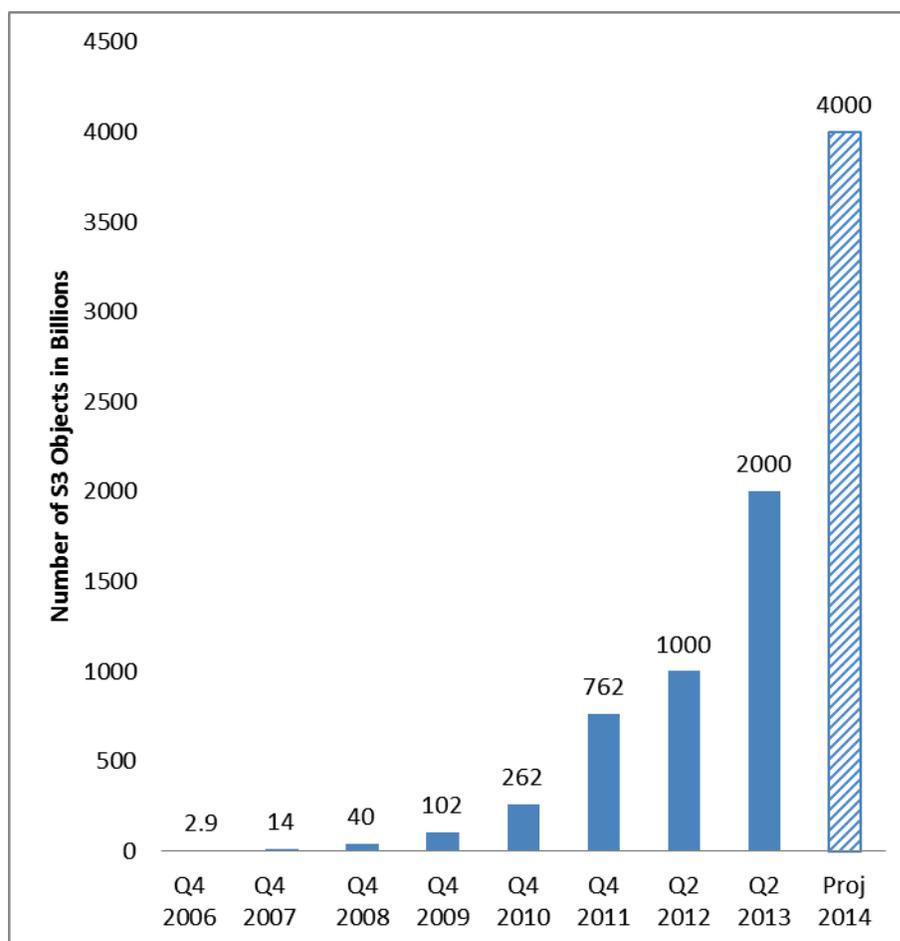
The past few years have seen an explosion of information. According to one source, by 2020, there will be as many as 6.6 zettabytes created, replicated, and consumed across the U.S., doubling approximately every three years. There will be a predicted 10 billion to 50 billion devices by 2020, and two-thirds of all mobile traffic is predicted to be video by 2017.

According to one source, enterprise IT growth rates range from 10 percent to 45 percent, with unstructured data growth estimated to represent more than 80 percent of storage. It is increasingly common to have datasets that are multiple petabytes. As an example, in the media industry, a complete digital movie at 4k resolution could be several petabytes today, with the industry now moving towards 8k resolution.

The plethora of new devices (smartphones, tablets, and cameras) generating more data, and faster than ever before conceived, and the migration of all media from analog to digital, are all having a huge impact. The additional emergence of the Internet of Things, with sensors of all kinds sending information for big data analytics, changes the game in terms of data transfer.

As one indication of the explosion of data into the cloud, on June 1, 2012 the number of objects stored in Amazon S3 object storage reached 1 trillion. Just 10 months later, Amazon announced that S3 had just reached 2 trillion. In November 2014, Amazon announced over 130 percent usage growth year-on-year in S3 data transfer. Amazon no longer releases the number of S3 objects, (probably since customers can now manage deletions themselves). The following chart shows a conservative projected growth of 100 percent.

Growth of Object Storage at Amazon S3



Source: Amazon Company Data

Vertical Markets

Media and entertainment was one of the first industries to push the limits in file sizes. It also pushed the limits in terms of time, due to a constant battle with deadlines. Broadcasters, studios, video surveillance, and gaming have also been dealing with extreme file sizes for a long time. Oil and gas, health care (particularly imaging), pharmaceuticals, universities, scientific analysis (e.g., genome sequencing), and satellite imaging are a few other examples where file sizes have created challenges.

The emergence of big data analytics has now broadened the field with competition pushing companies of all sizes in every industry to analyze whatever data they can to compete in their market. Today's global economy is also making sharing data at a large scale more important.

Companies are using Hadoop Map Reduce to make sense of all types of unstructured data. As these analytics workloads grow in the cloud, large file requirements are expanding to more companies. Some organizations are using the cloud pay-as-you-go model to access other types of compute-intensive horsepower, which may previously have been too expensive and thus unavailable to them, such as uploading video to be transcoded or access to supercomputers.

Big data combined with big compute can create big opportunities, but if organizations can't get the data in and out of the cloud, the cloud can't solve the business problem.

Challenges and Limitations of Moving Large Amounts of Data

Organizations in industries like media and sciences have been struggling for years with traditional methods of moving large files within the enterprise environment.

Most file transfer solutions and even newer cloud file sharing solutions were designed to handle small files such as office documents, and to run on standard networks using transmission control protocol (TCP). Attempting to move large documents using these methods, particularly over long distances, can be painfully slow, if not impossible.

Many organizations now look to move petabytes of data. Not surprisingly, this creates challenges in a world where anything over 10GB is hard, files over 100GB can be too slow to be practical using traditional methods, and larger than that may fail completely. Additionally, most file transfer mechanisms were not designed to address the security, management control, or compliance requirements needed for enterprise IT. And they were not designed as cloud applications or services.

When it comes to speed, latency and bandwidth are the enemies of moving large files. These two factors, along with the error rate (lossiness) on the connection, are the biggest factors in network slow-downs. TCP has built-in functionality for things like error handling and flow control, but was designed for small packets, and its “chattiness” adds significant overhead, particularly when sending large files over long distances.

Finally, given that security is generally the top concern in cloud storage, encryption may be required, for data in transit and at rest. Many traditional solutions do not offer encryption, but for those that do, look closely at the overhead they add to transfer times.

Limitations of Traditional Methods and Alternatives

FTP, which is more than 40 years old, is no longer fast enough for today’s massive file sizes. Simply put, in high-latency networks with high bandwidth, FTP can’t fill the network pipe. Newer accelerated protocols, designed for high latency high bandwidth networks, accomplish these three things differently. FTP also lacks advanced functionality such as reliability, notification, manageability, and auto-scaling.

HTTP can also be used to move files, but it was designed for browser-based activity: loading and displaying web pages, which are much smaller than the large files being transmitted today. Various

approaches such as Dropbox and Box have emerged for ad hoc file sharing via the cloud, and have moved from consumer to corporate use. These solutions, while easy to use, are still slow, and have raised IT concerns about security, control, manageability, audit, and compliance. The use of unapproved cloud services for file storage puts data in personal accounts, with no enterprise visibility or control over access to that data.

Content delivery networks (CDNs) address the challenges of latency when distributing the same content to large numbers of users, by caching at the edge of the network, close to the users. This approach does reduce latency for outbound data to those users. However, CDNs do not help when sending unique content or moving data into the cloud, so a CDN solution has limited use for cases such as those involving transferring large datasets for big data analytics. Lastly, shipping hard drives requires a wide range of individuals to handle your assets—increasing the overall security risk.

When/Why Speed Matters

Speed-Sensitive Use Cases

The types and sizes of files discussed here require a level of transmission speed that is hundreds of times faster than has historically been required, making things that were previously impossible (other than by shipping disks) now possible.

The ability to use the cloud for resource-intensive computing tasks that previously required major capital expenditures has unlocked a huge wave of business opportunities. Organizations now can exploit cloud supercomputer power and big data analytics and use the results as a serious competitive advantage. For example:

- In media and entertainment, organizations not only store assets in the cloud, but also use cloud-based transcoding and distribution to help reduce project timelines, meet contract deadlines, and realize new revenue streams.
- Life sciences' firms are using the cloud for genome sequencing and drug discovery to analyze data much faster than they could have just a few years ago, allowing them to bring new products and insights to market faster than ever.

A few real-world examples can provide more depth on when speed matters:

A major distributor of video assets doing large-scale data ingest from many sources

In a major metro area, a large distributor of video assets has massive storage requirements but no room for more storage onsite, due to limited real estate. As a result, it must use onsite storage more efficiently and uses cloud storage wherever possible. There are 600 providers creating assets for the station that must be loaded quickly to stay on their very tight schedule. Adding one high-speed fixed pipe doesn't help, since the data comes from all over. The distributor regularly transmits multi-GB files (a 60 minute HD quality video can range from 50 to 400 GB depending on encoding.).

A medical company transferring 100 GB files, resorting to trucks and HDDs

A medical company, with a 100-megabit network, needs to send large numbers of files in the hundreds of GBs per day. Transmitting during the day over its network bogs down the whole network. The company has tried moving its data overnight, but it doesn't have enough hours available to complete the transfers.

As a result, it resorted to shipping hard drives. Transferring to the cloud, using high-speed file transfer, eliminated the overnight window restriction and the need for trucks.

Massive ingest of imagery or location-based data for analytics

As an example of innovation, companies with large deployments of cameras are finding creative uses for their imagery (e.g., satellite, surveillance), making it available for big data analytics to help predict consumer behavior. The usefulness of the data is very time-sensitive. There are other opportunities in many industries, for example, looking at location-based data to optimize physical package delivery. Without high-speed massive cloud ingest capabilities, this type of opportunity didn't exist previously.

Considerations

Key Questions

Many things must be considered when evaluating requirements and solutions for large file transfer in and out of the cloud.

- How does the data need to be sent?
- From where to where? How many users and locations will be involved?
- What clouds must be supported? What geographical areas?
- What devices and operating environments must be supported?
- Are there requirements for one-to-one, one-to-many, many-to-one?
- Are there requirements for person-to-person transfers?
- Ad hoc versus scheduled?
- Manual transfers or automated into applications?
- What type of interfaces are required for your company's needs (GUI, API, SDKs, etc.)? Consider the level of complexity, ease of use, and/or ease of integration.

In some cases it makes sense to consider a dedicated connection from your data center (or colocation) to the cloud. This works best for regular amounts of data transfer (not elastic or on demand). Examples include AWS Direct Connect and Azure Express Route.

Consider software versus hardware WAN acceleration. Hardware works well in use cases where the data always goes between the same locations. It is also a good fit for data with a lot of redundancy. Software offers flexibility when there are many or changing sources or destinations.

Consider open-source options. The best known, Tsunami, is a hybrid UDP/TCP protocol with command line interface. The current free version was released in 2009. It is command line only, not integrated with Amazon S3, has no advanced features like encryption or an SDK or multi-threading and has no commercial support or active forums.

There is sometimes the need for advanced features such as full WAN acceleration or centralized management. If these factors aren't critical to a project, there are a number of developer and GUI tools that use capabilities like an HTTP-based multi-part API to move files.

Are you looking for a turnkey solution that handles auto-scaling, load balancing, managed servers, and automatic updates? Or do you want to implement a solution that is more DIY, requiring your staff to buy/provision servers to handle the transfer, buy and manage file transfer software licenses, handle the scale-up and scale-down, and potentially write code to automate and manage the system and maintain software updates?

Cost is always an issue, so when looking at costs, also consider the level of effort to implement and manage the solution on an ongoing basis.

Key Solution Features to Consider

When evaluating solutions, the following list is a good starting point for features to consider in your selection criteria. Customize and prioritize the list based on your specific business requirements.

Cloud storage support. What clouds and cloud storage do you need to support (e.g., Amazon S3, Microsoft Azure, Google Cloud Platform)? Even if multiple clouds are not an immediate requirement, look at what clouds are supported (current and planned) to provide flexibility across geographic areas and to avoid vendor lock-in down the road.

Storage types and file sized supported. What types of storage (file, block, object storage) are supported and what are the file size limits, if any?

Devices and platforms supported. What devices are supported: Mac, Windows, Linux, iOS, Android, and what developer APIs are available?

Actual file transfer speed. File transfer speed is critical for large files and will be a result of the following features, but should be tested against your actual types of data.

Base transfer protocol. UDP versus TCP to minimize latency (TCP only for control).

Protocol acceleration. Software- or hardware-based WAN acceleration that speeds up the transfer into the cloud infrastructure and provides flow-control, congestion control, reliability mechanisms, bandwidth optimization, and error correction, and handles high latency and lossy networks. Cost/ benefit

is dependent on the bandwidth and latency of links used; at least megabits or gigabits and long distance can see up to 200x faster.

Multi-part upload capability. This can also help address latency for long-distance connections in the cloud by breaking the data into chunks and using parallel transfers.

Reliability and availability. Features to consider should include redundancy, automated restart capability with checkpoint restart, and notification of successes and/or failures.

Scalability and auto-scaling. How well does the solution scale? Does it provide load balancing and auto-scaling automatically, or are you responsible for scaling operations?

Security. Security features including encryption, audit trails, and any regulatory compliance features required for your industry that should be reviewed with your security team.

Management visibility and control. If using cloud object storage, direct visibility and control will aid in overall management.

Software delivery. Is the solution delivered as SaaS or does it require work by your staff for setup and automated operations. Which is the best fit for your requirements and availability of the necessary skilled resources?

Key takeaways

- Cloud and cloud storage offer big opportunities for organizations to change the way they do business. In many cases, companies now have access to computing capabilities previously unavailable to them due to the cloud pay-as-you-go model. However, without the right data at the right places in the right timeframe, these opportunities will be lost.
- Massive information growth and big data are driving requirements for large file transfer into and out of the cloud. File sizes in the hundreds of GBs, TBs, and PBs are becoming commonplace and critical across vertical markets as well as horizontal business requirements such as big data analytics, using information as a strategic asset.
- Traditional methods of large file transfer, most based on forty-year-old technology, are no longer sufficient for today's needs. Innovation in data transfer must keep up with innovation in computing.
- Solutions are available today, based on a variety of approaches. It is important to understand the individual business needs first, and then look for the solution that offers the best fit. Leveraging the experiences of those who already have solved this problem in traditional IT enterprise environments and the cloud can help smooth the road.

About Barb Goldworm

Barb Goldworm is founder, president and chief analyst of FOCUS, LLC, and is a well-recognized industry expert focused on emerging and game-changing technologies including cloud computing, virtualization, software defined data centers, systems management, and storage technologies, with over thirty years' experience in software architecture, development, marketing, sales, and industry analysis. She has spent the last fifteen years keynoting, presenting, teaching, and writing for industry events, webinars, publications and other industry channels. Her past experience includes roles in engineering, marketing, and executive management with IBM, StorageTek, Novell, Enterprise Management Associates (EMA), and numerous startups.

Barb had also been content chair, advisory board member, and keynote speaker for industry events such as Interop, Cloud Connect, Enterprise Cloud Summit, Data Decisions, Storage Decisions, Server Blade Summit, and Comdex. In addition to publishing hundreds of research reports, white papers, and regular columns (NetworkWorld, Computerworld, Storage Networking World Online, TechTarget, and others), Goldworm is also the author of the Wiley book, "Blade Servers and Virtualization" available on Amazon. In addition to authoring Gigaom Research reports, she routinely acts as moderator for Gigaom Research webinars.

About Signiant

Signiant's intelligent file movement software helps the world's top content creators and distributors ensure fast, secure delivery of large files over public and private networks. Built on Signiant's patented technology, the company's on-premises software and SaaS solutions move petabytes of high-value data every day between users, applications and systems with proven ease. Signiant Flight is the only SaaS solution on the market that accelerates the movement of large files and data sets into and out of cloud storage. Automatically maintained and updated for you, Signiant Flight is easy to deploy, scales as needed, and works with Amazon S3 and other leading cloud platforms.

Learn more at: www.signiant.com

About Gigaom Research

Gigaom Research gives you insider access to expert industry insights on emerging markets. Focused on delivering highly relevant and timely research to the people who need it most, our analysis, reports, and

original research come from the most respected voices in the industry. Whether you're beginning to learn about a new market or are an industry insider, Gigaom Research addresses the need for relevant, illuminating insights into the industry's most dynamic markets.

Visit us at: research.gigaom.com.

© 2014 Giga Omni Media, Inc. All Rights Reserved.

This publication may be used only as expressly permitted by license from Gigaom and may not be accessed, used, copied, distributed, published, sold, publicly displayed, or otherwise exploited without the express prior written permission of Gigaom. For licensing information, please [contact us](#).