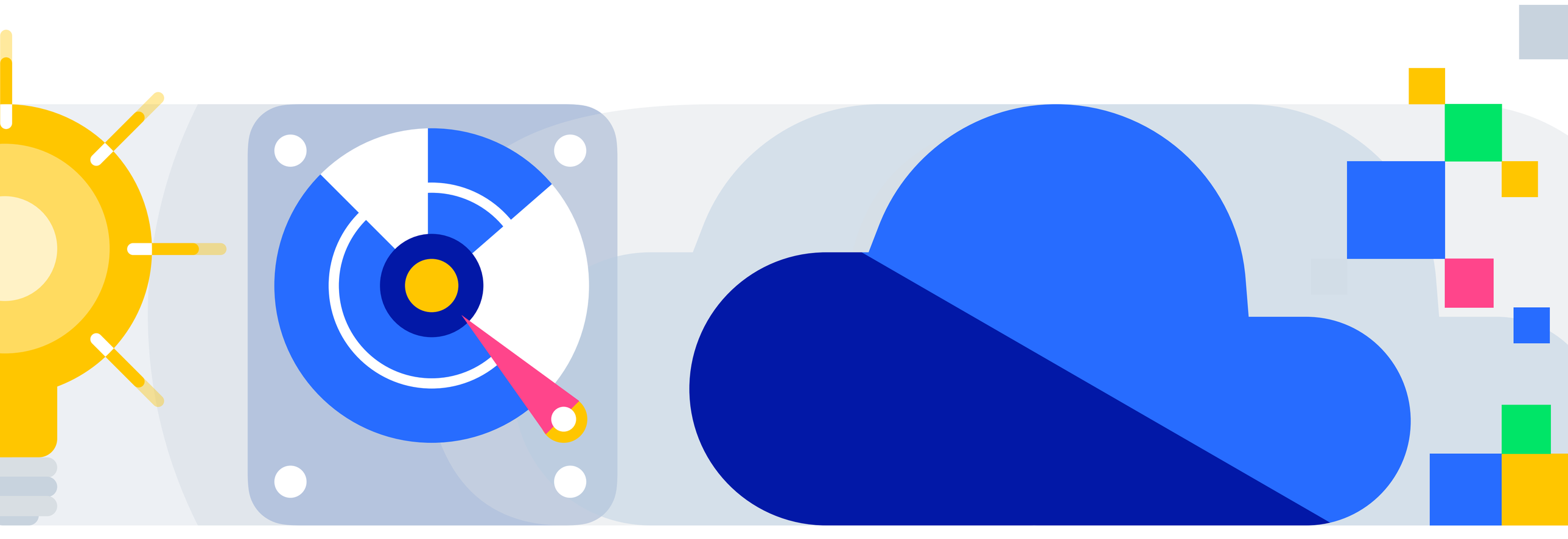


The 5 Phases of Cloud Storage Evolution

Drivers Forcing Technology Change and What it Means for Developers Today

STORJ

www.storj.io

Table of Contents

Introduction

The story of the cloud as we know it today starts in the early 60's when innovators had the vision of connected computer networks. Cloud ideation rapidly turned into cloud development and the internet began to take shape. By the late 90's we were beginning to use the cloud term and applications started to be delivered in the cloud.

By 2011, we were seeing rapid adoption of cloud services and storage. Much of that has been via centralized cloud providers such as Amazon, Microsoft, IBM, Oracle and Google. Yet with adoption came complexity. The past 20 years have been fraught with data breaches and privacy concerns. Simultaneously, data creation has been exploding, causing many cost and scalability challenges.

Throughout the evolution of the cloud, one thing remains consistent—innovation. The internet and data storage started in a decentralized model, then overtime began to centralize to capitalize on efficiencies. As problems with cost, security and data privacy arose with centralized cloud storage, advancements in technology paved the way to begin to drive cloud storage and compute back to a decentralized model without sacrificing cost or performance.

In this paper, we'll walk through the journey of cloud storage evolution and explain the circumstances driving change. We'll also look to the future of cloud storage and how developers should prepare to take advantage of new options.

The 5 Phases of Cloud Storage Evolution

1 Ideation 1962-1974

Originating concepts and early experiments launched a host of projects to develop interconnected networking.

2 Development 1975-1999

Early concepts built upon each other to create the Internet and eventually to offer applications online.

3 Services and Storage 2000-2010

Cloud storage is a reality for both businesses and consumers and early adopters reap the benefits.

4 Adoption & Exploitation 2011-2015

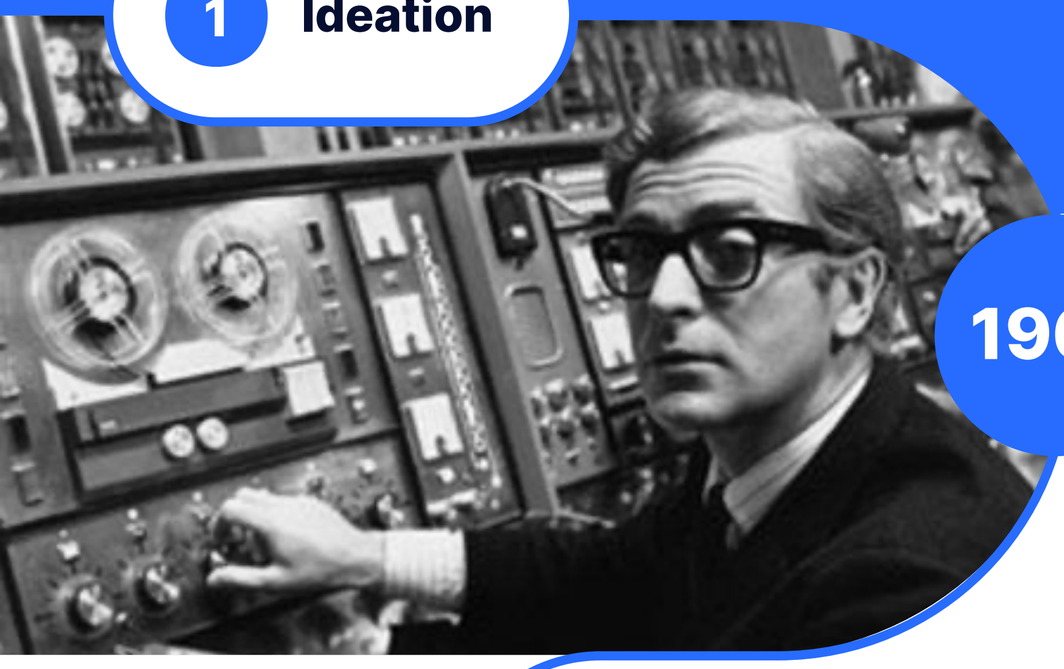
Cloud storage migration takes place for personal and business use. At the same time, security breaches increase in volume and impact.

5 Evolving to Decentralization Now

Cloud storage evolves to allow for decentralized storage that solves the security and privacy issues of traditional cloud storage.

The Cloud Evolution Timeline

1 Ideation



J.C.R. Licklider

1962 J.C.R. Licklider “Lick” proposed the idea for an Intergalactic Computer Network.

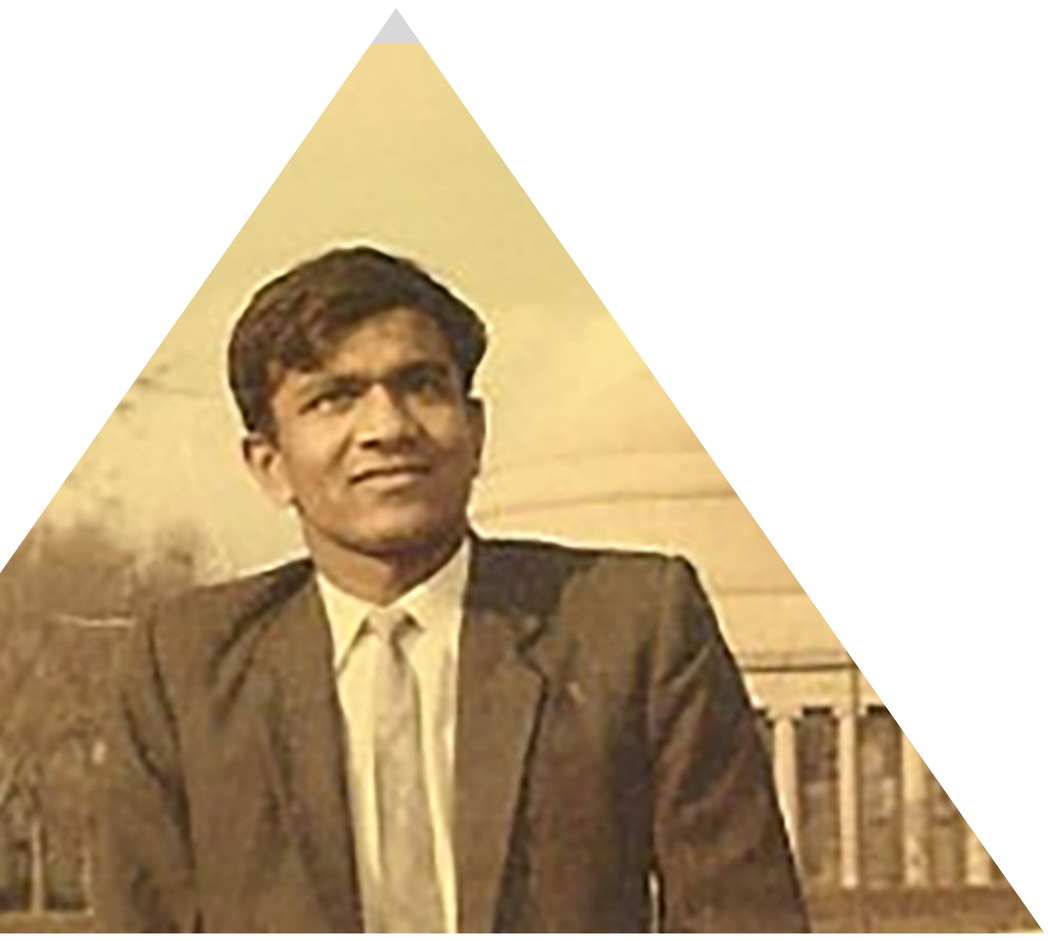
1966 ARPANET project initiated to enable access to remote computers.

1966 The Merit Network was formed to explore networking across universities.

ARPANET Project Computers, 1960’s

1969 ARPANET succeeded in connecting the first computers.

1971 File Transfer Protocol (FTP) was first written by Abhay Bhushan.



Abhay Bhushan, April 1971

1974 The word “Internet” was first coined.

2 Development

1975 ARPANET was declared operational and control was passed to the Defense Communications Agency.

1978 UUCP (Unix-to-Unix Copy Protocol) first in use at AT&T Bell Laboratories.

1980 Usenet discussion system was connected to ARPANET.

AOL 1.0 Floppy Disk, 1989

1983 The National Science Foundation started NSFNET.

1983 CompuServe offered consumers storage.

1989 America OnLine (AOL) was launched.

1989 The World Wide Web (Web 1.0) was invented.

1990 ARPANET was decommissioned.

1992 NetApp founded.

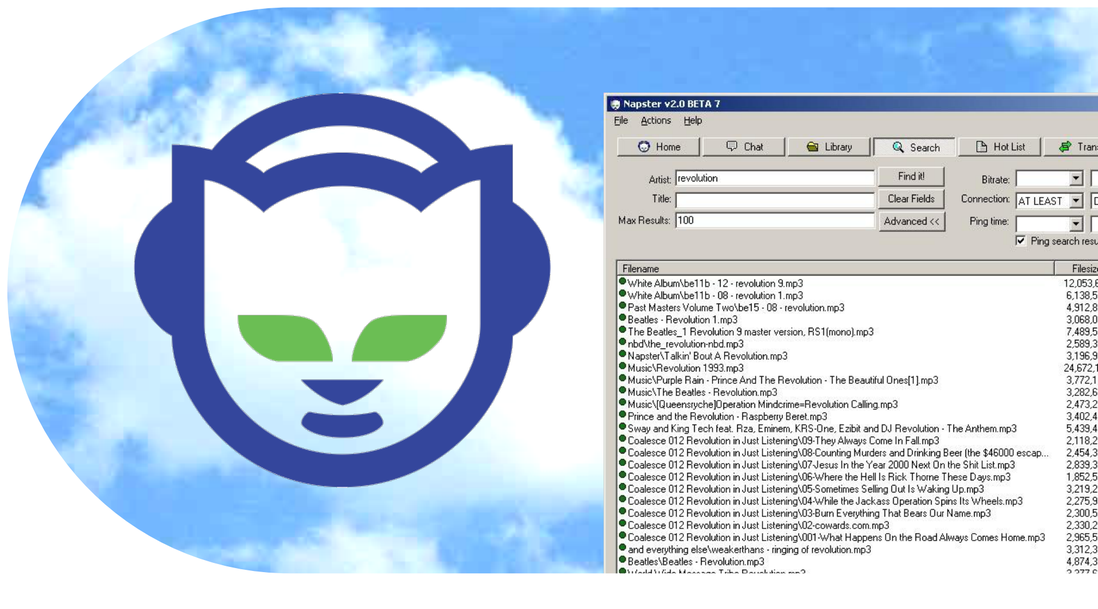
1994 AT&T launched PersonaLink Service.

1995 NSFNET was decommissioned and network access points were created.

1999 VMware launch its first Workstation.

1999 Salesforce.com launched.

1999 Napster founded.

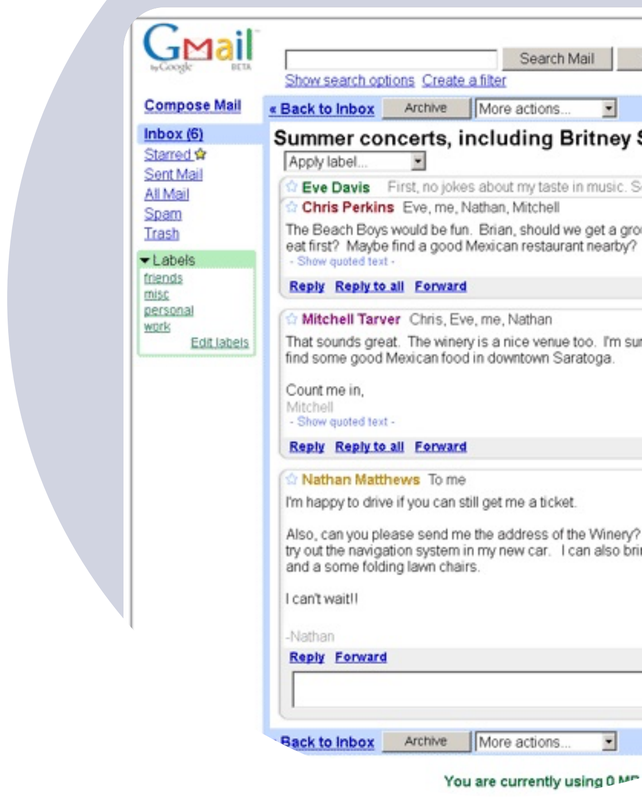


Napster, 1999

3 Services & Storage

2000 MojoNation launches encrypted decentralized file sharing.

2000 Freenet and Gnutella launched.



Gmail, 2004

2004 The second generation of the World Wide Web, Web 2.0, began to evolve.

2004 Gmail was launched offering 1GB storage.

2004 BitTorrent founded.

2005 Box, Mozy and Carbonite founded.

2006 Tahoe-LAFS (Least-Authority File Store) is launched at AllMyData.

2006 Amazon Web Services (AWS) launched their cloud storage service, AWS S3 (Simple Storage Service).

2007 DropBox founded.

2007 Netflix released its video streaming service.

Netflix, 2007

2009 Bitcoin, the first cryptocurrency, went into use.

2010 Microsoft launched Microsoft Azure.

2010 The OpenStack project began for open source cloud computing.

2010 The OpenStack project began for open source cloud computing.

4 Adoption & Exploitation

2011 Apple released their iCloud for backing up mobile device data.

2012 Oracle launched their Oracle Cloud.

2012 Dropbox and LinkedIn experienced the first major cloud data breaches.

2013 SpaceMonkey offers decentralized cloud storage via Kickstarter.

2013 Google launched Google Drive.

2013 Internet pioneer Yahoo suffers 1 billion user breach.

2014 First prototype developed for decentralized cloud storage using cryptocurrency.

2014 Initial offerings of decentralized cloud storage launched.

2015 Ethereum, the first cryptocurrency with programmable contracts, went live.

2015 InterPlanetary File System released.

Storj Founder Shawn Wilkinson, 2016

5 Evolving to Decentralization

2017 Fastly and Cloudflare launched serverless edge computing platforms.

2019 Facebook had 540 million user records exposed on AWS.

2019 Capital One hack exposing 100 million customers blamed on AWS.

2020 SolarWinds hack impacted 18,000 enterprise and government customers.

2020 Covid-19 pandemic hit forcing cloud storage and security adaptations for remote work.

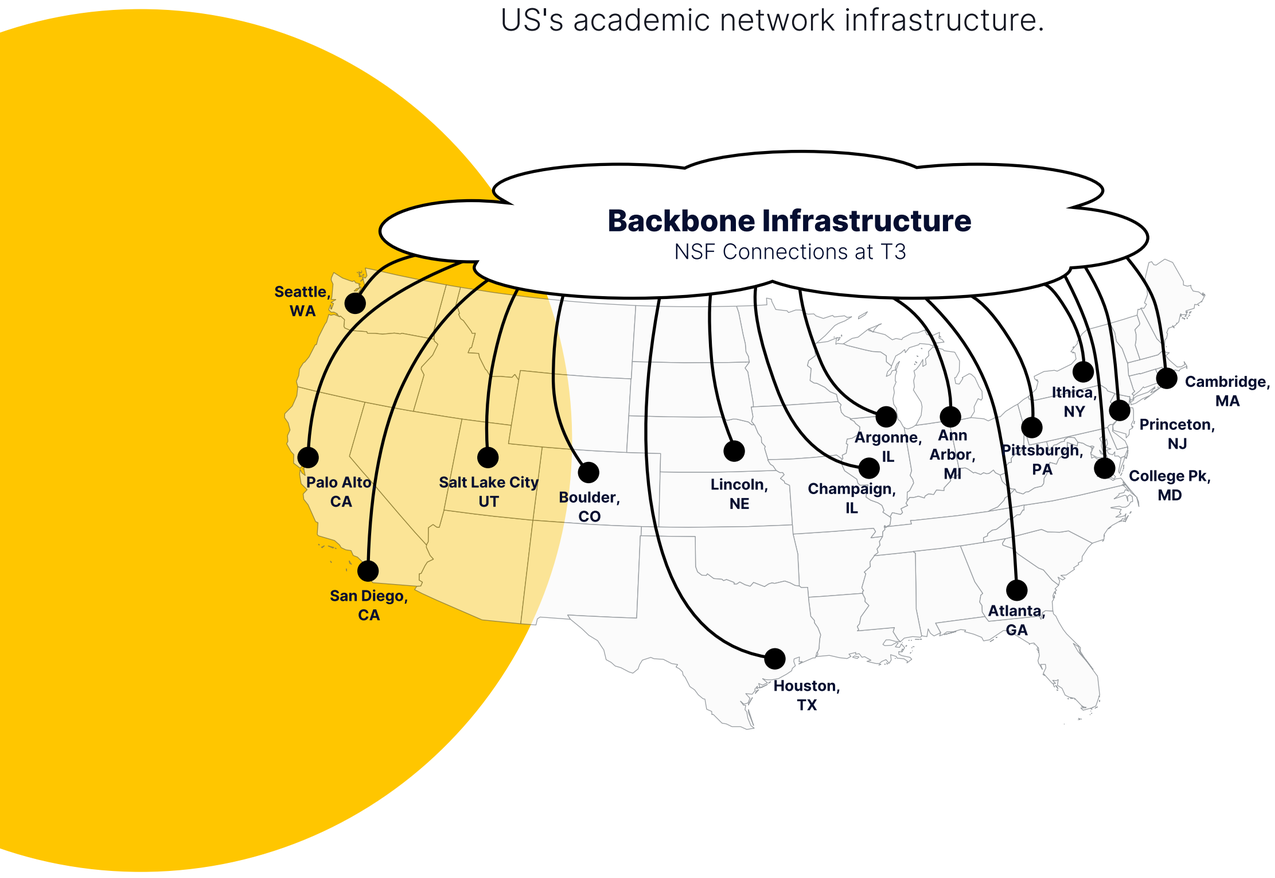
2021 Decentralized cloud object storage launches with proven benefits over centralized cloud storage alternatives.

How the Cloud Became the Cloud

So how did the Cloud really get its name? Sources have difficulty agreeing on a single person for coming up with the name “Cloud”. What we do know is that it came to fruition as both a spoken metaphor and graphical representation of a network too complex for most people to understand.

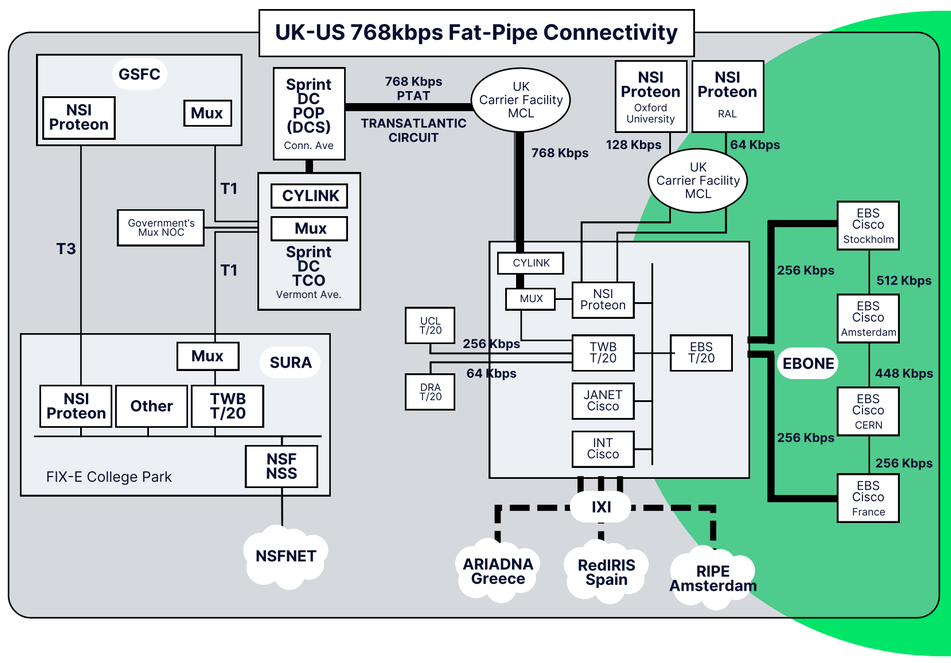
NSFNET 1991

US's academic network infrastructure.



JANET 1992

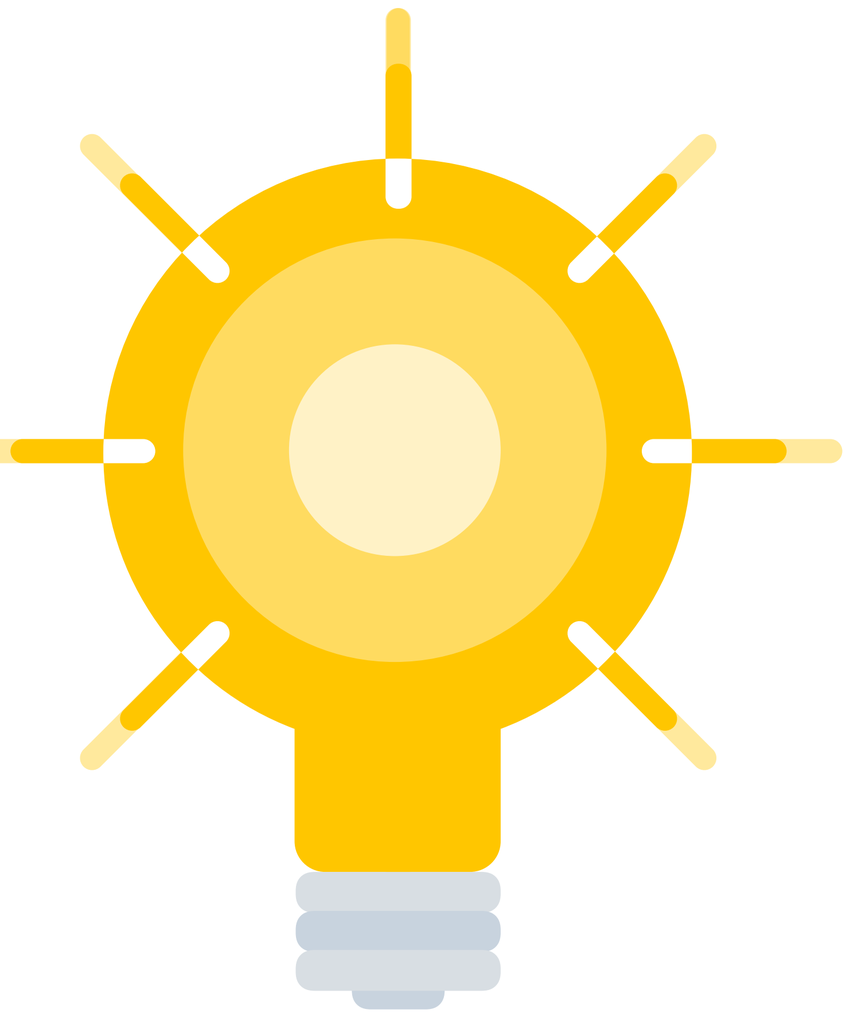
UK's academic network that tied into NSFNET.



The first use of the cloud image on schematics was likely by NSFNET. It was used to represent a network of computing equipment. It also shows up in schematics of UK’s version of NSFNET called JANET.

The first metaphorical usage of the term that matches how we use it today is often attributed to David Hoffman, an employee of General Magic. At the time, General Magic was a spin off from Apple collaborating with AT&T to pair their Telescript and PersonaLink technologies. At a conference in 1994, Hoffman was quoted as saying, “. . .instead of just having a device to program, we now have the entire Cloud out there, where a single program can go and travel to many different sources of information and create a sort of virtual service.”

Today, the Cloud is so complex and has so many aspects that it is truly a magical term for most people. Qualifiers are needed to specify compute or storage as well as where those things are happening—in a centralized location or “at the edge” in a location where the data is needed. Only two things are clear, 1) that the term “the Cloud” is here to stay, and 2) that it holds just as much potential and magic today as it did in the early 90’s.



The Cloud Ideation Phase (1962-1974)

The 60's saw the birth of the ideas and initial experiments that led to the internet as we know it today. Despite claims from Al Gore, most people attribute the invention of the internet to Joseph Carl Robnett Licklider (aka. J.C.R.,"Lick"). Licklider was a psychologist, computer scientist, and MIT professor who had a vision for an Intergalactic Computer Network. His idea was outlined in a memo to his colleagues at ARPA. His vision spurred the ARPANET project.

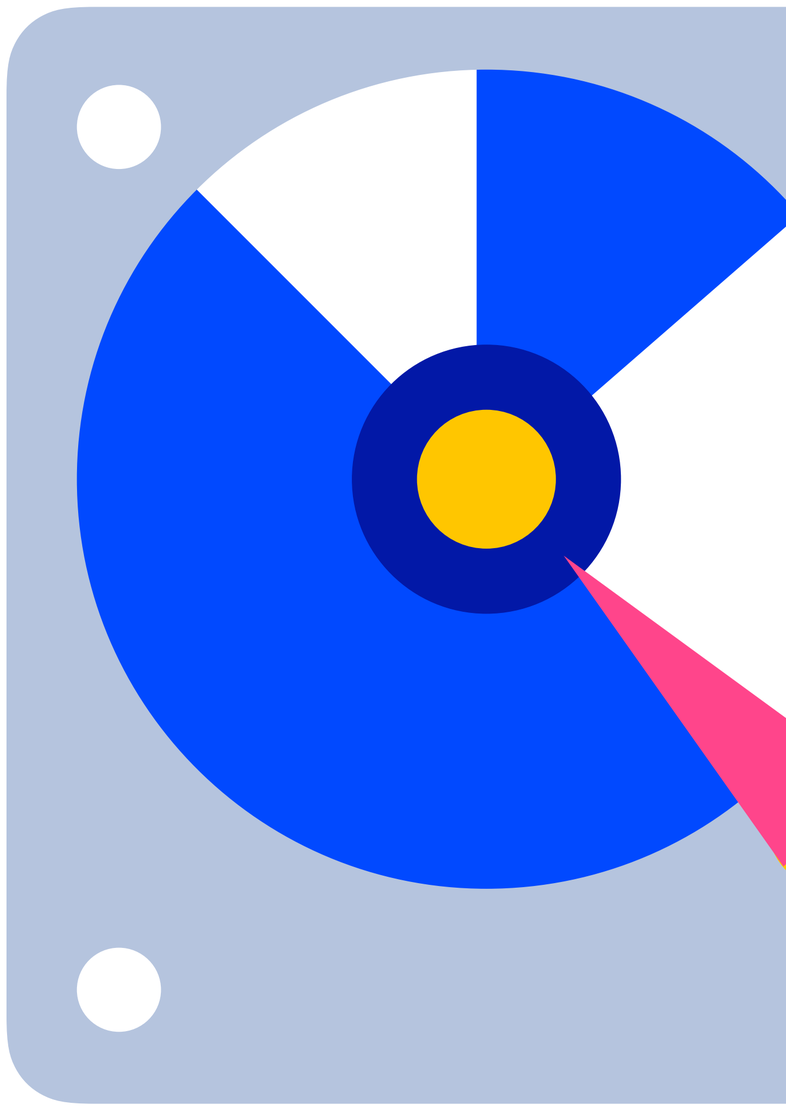
ARPANET was an endeavor to connect computers using the first wide-area packet-switching network. It succeeded in connecting the first computers in 1969. Meanwhile, three universities in Michigan also were working to create a network they called The Merit Network. They started their project in 1966 and successfully connected the three universities in 1972. These two initiatives later merged under Telenet.

The Cloud Development Phase (1975-1999)

In 1975, ARPANET was deemed a success and transferred over to the Defense Communications Agency to use for government and military computer networking. The 80's saw the rise of NSFNET, a National Science Foundation project which created a network across major research and university sites and quickly surpassed ARPANET. Consumers were also dabbling with the first online storage option from CompuServe who provided their customers a whopping 128K of file storage. The 80's finished strong with the invention of the World Wide Web (1.0) in 1989 and Internet Service Providers, like AOL, growing rapidly.

The 90's saw a decommissioning of the original networks (ARPANET and NSFNET) to pave the way for peering arrangements between commercial ISPs and the Internet as we know it today. Vmware launched in the late 90's as well as the birth of cloud-only applications with Salesforce.com's launch in 1999.

The Cloud Services and Storage Phase (2000-2010)

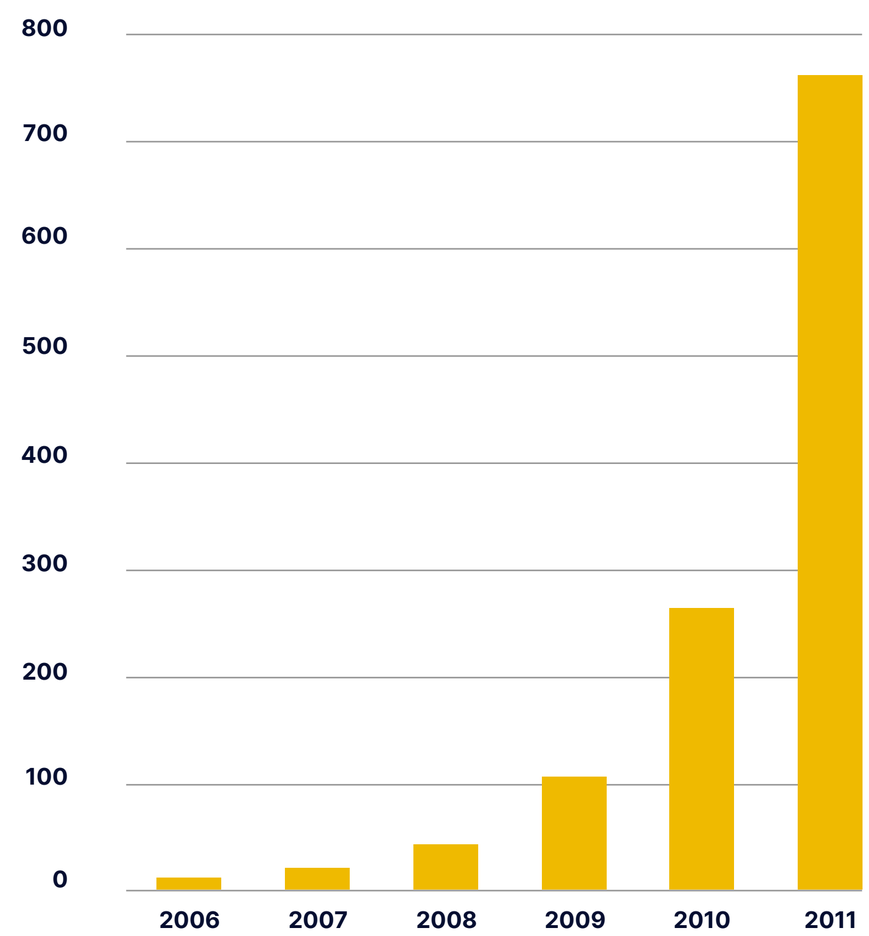


The early 2000's saw rapid advancement in peer-to-peer technologies with Napster, Freenet, Gnutella and BitTorrent. The World Wide Web advanced to 2.0—evolving to allow user-generated content. This paved the way for social media.

Cloud storage also took off in the 2000's. Email providers started to offer increasing amounts of disk space to customers, and online but off-premise backup providers began to gain more trust. Crucially, it became a possibility for smaller companies to store data without their own data centers with Amazon's launch of their AWS Simple Storage Service, or S3. Many followed suit with cloud storage services over the next few years. Additionally, more business models were operating in the cloud with Netflix releasing its video streaming service and Google releasing its cloud business applications.

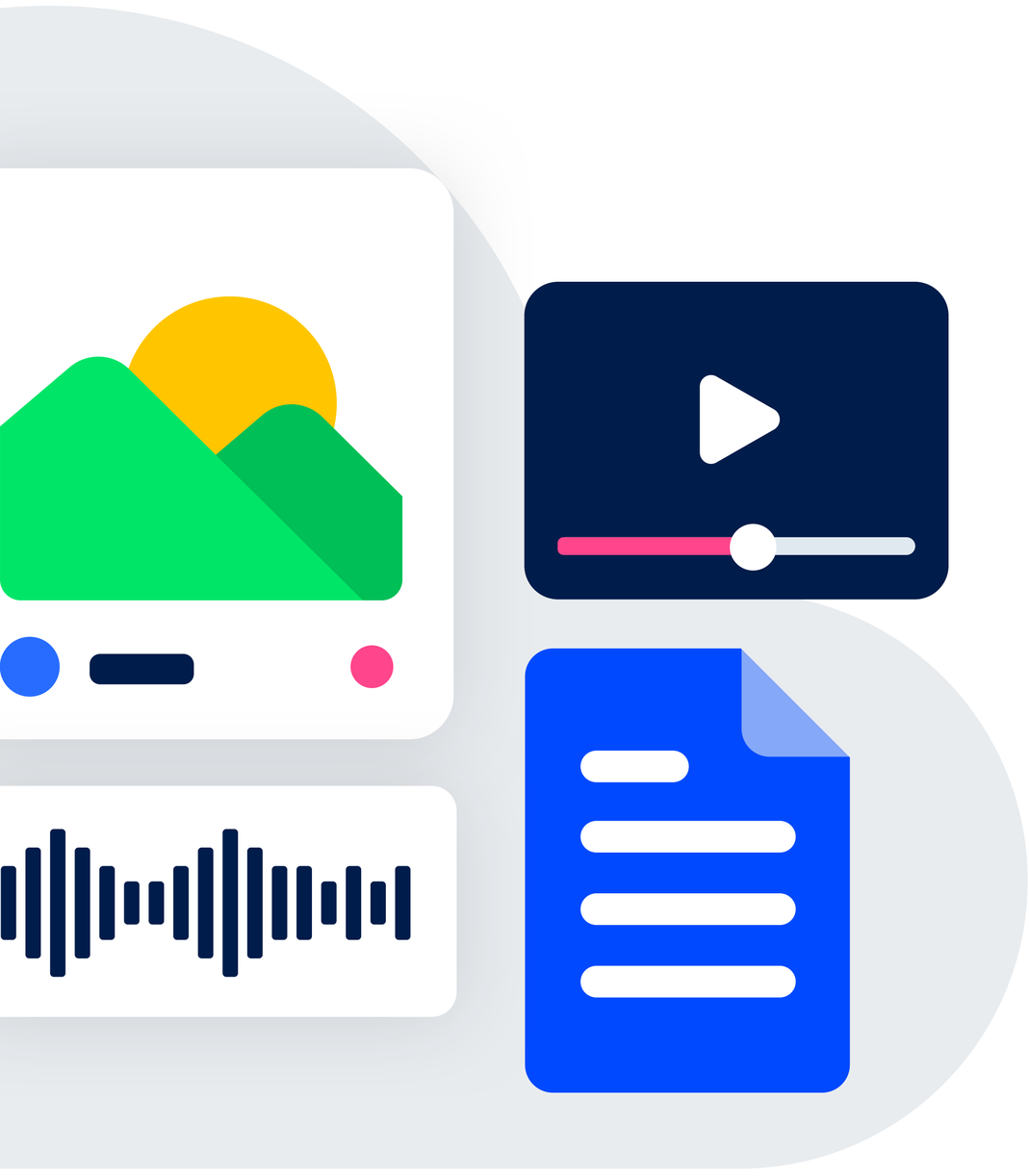
With the easy availability of centralized cloud storage followed by cloud compute, cloud-based applications were surfacing everywhere. Both business and consumer-facing applications were being adopted and became more accepted during this time period. Software as a Service (SaaS) became a sought after buzzword. This period finished out strong with the OpenStack project, which was the beginning of open source cloud computing initiatives.

Total Number of Objects Stored on Amazon S3 (Billions)



The Rapid Growth of Centralized Cloud Services and Storage

As delivering software in the cloud became mainstream, rapid adoption happened because the cloud was easily accessible and very affordable thanks to subscription pricing. Competition now had a way to break into existing markets and disrupt them thanks to the lower cost of entry. There was no longer a need to build a data center, acquire hardware and build infrastructure from scratch. Businesses could now go from zero to product release in a very short time. These were the years of limitless opportunity for companies. And with this rapid adoption came more and more cloud data to store. All of this made being a provider of cloud services and storage an incredibly profitable business to be in.



Centralized Cloud Storage Adoption & Exploitation Phase (2011-2015)

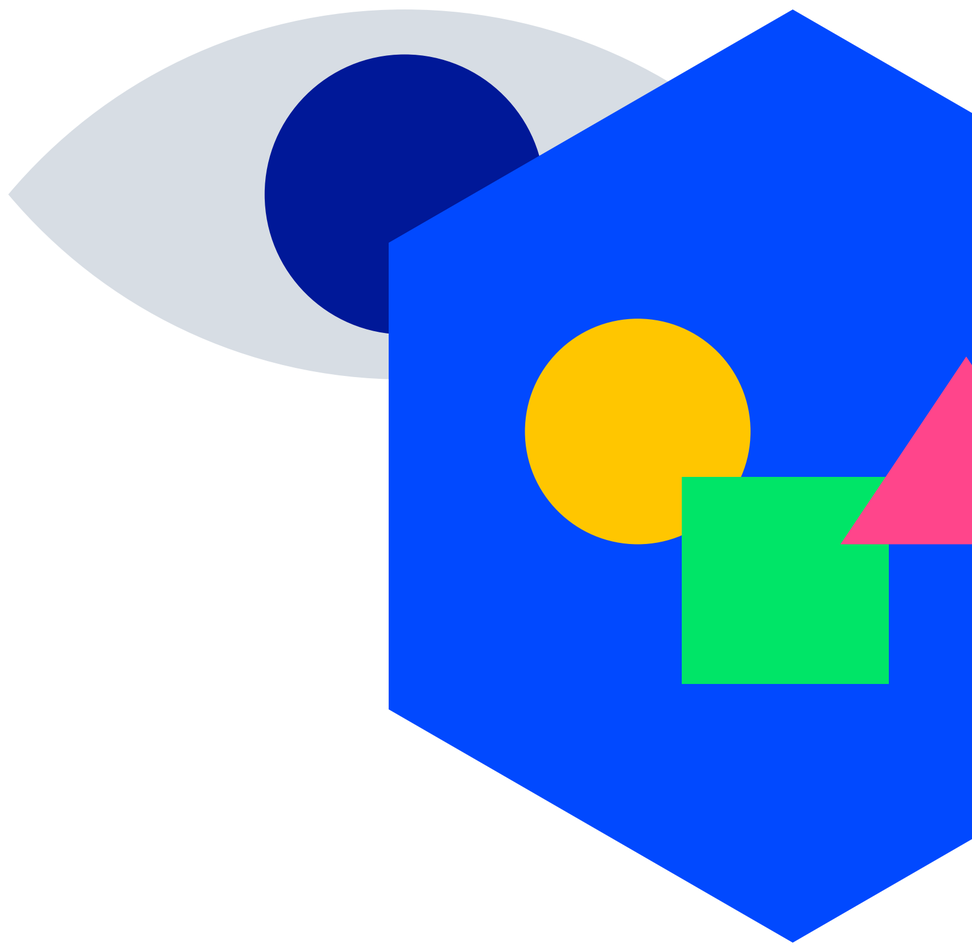
The boom of the 2000's for the cloud continued into the 2010's with more companies (namely Google and Microsoft) getting into the business of centralized cloud storage. It also brought a major migration to the cloud. Companies began migrating their existing data to the cloud to benefit from the perceived cost savings and better availability. By 2015, the cloud storage market was experiencing triple-digit annual growth.

With Popularity Come Malicious Players

The popularity of centralized cloud storage brought with it a new breed of cyber threats. Cloud data storage created a honeypot of data for cyber attackers to tap into at increasing levels of scale. Additionally, simple configuration errors could open the door for attackers to enter and steal data. 2012 saw major cloud data breaches for Dropbox and LinkedIn. Combined, the hacks resulted in over 235 million credentials being stolen and later sold on the dark web. Just a year later, cloud giant Yahoo suffered a data breach exposing 1 billion user credentials. This was just the start of security concerns surrounding centralized cloud storage as major data breaches became more frequent over the years.

Beyond Breaches—Centralized Cloud Suffers Privacy & Regulatory Issues

As data breaches continued, end users became more aware of and concerned with data privacy. A big part of this concern revolves around the fact that the largest centralized cloud storage providers (Amazon, Google and Microsoft) are also in the business of mining data to generate advertising or product revenue. Companies using these providers came out with lengthy privacy policies to attempt to clarify how they utilized data stored on their networks, but customers often weren't happy with the lack of protections for data privacy. Governments recognized the issue and began implementing regulations to try to protect their citizen's data privacy.



Centralized Cloud Problems Drive Innovative Alternatives

While the downsides of centralized cloud storage started appearing in the 2010's, entrepreneurs saw opportunity arise in decentralized cloud. As more capacity and bandwidth had become available in people's homes, the concept of decentralized cloud storage was becoming possible. Previous attempts had failed either due to lack of available bandwidth resources, or lack of fungible incentive systems (such as those provided by cryptocurrency). In 2014, the first decentralized cloud storage prototype with cryptocurrency payment was developed at the Texas Bitcoin Hackathon by Shawn Wilkinson. Ethereum went live soon after, which is a decentralized, open-source blockchain that later became a key platform for node operator payments on decentralized networks. At the tail end of this period the InterPlanetary File System (IPFS) was released. This is a peer-to-peer network similar to BitTorrent, but uses a decentralized system of user-operators.

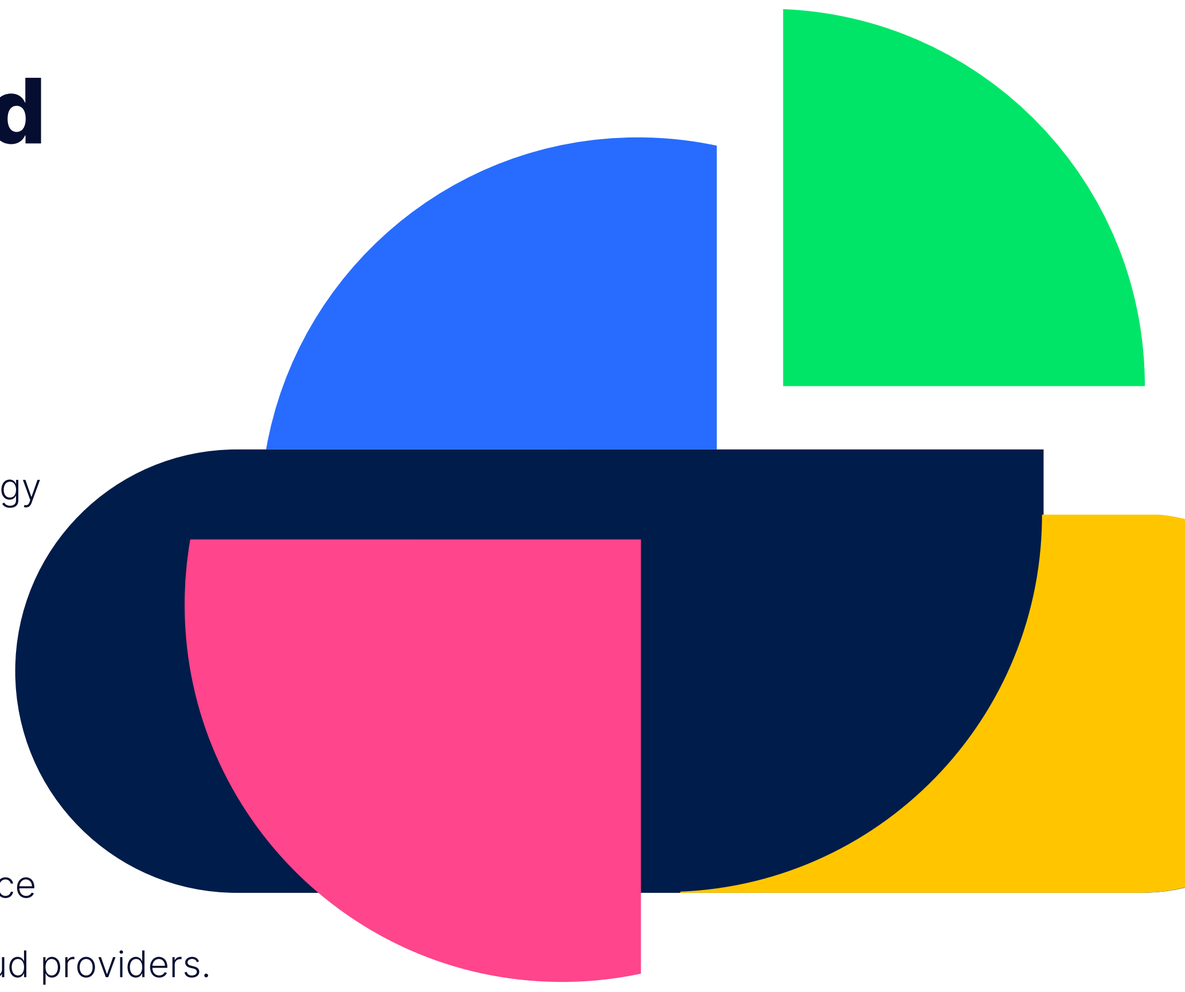
What is Decentralized Cloud Storage?

As opposed to storing data within large data centers run by a third party, decentralized cloud storage provides a new storage option for data objects by taking advantage of excess storage capacity on end user hard drives and the general availability of bandwidth.

Encrypt **Split** **Distribute**

Decentralized storage providers take fully encrypted data, break it up into lots of pieces and spread those pieces around to unique end users located throughout the world. This model is highly secure and private, highly performant and very cost effective. Altogether, it creates a new option for cloud storage that solves the predominant problems with centralized cloud storage.

Evolution to Decentralized Cloud Storage Phase (2016-present)



The past five years have been filled with a combination of technology advancements in edge computing and decentralized storage as well as increasing concerns over privacy and security in centralized cloud data storage. 2017 saw the launch of both Fastly and Cloudflare serverless edge computing platforms. Having compute at the edge increases accessibility and performance and provides a viable alternative to compute within centralized cloud providers. During this time decentralized cloud storage providers have begun adding compatibility with AWS, with Storj being the first major open source digital token-driven decentralized object storage provider with S3 compatibility. The ability to migrate away from centralized cloud storage to the more secure, private and performant decentralized model is now more viable than ever before.

Data breaches continued during this time, proving that even the highest levels of security from centralized cloud storage providers wasn't enough. The 2019 Facebook breach exposing 540 million user records was of data stored on Amazon Web Services. Likewise, AWS is being investigated for their role in not putting in security measures to prevent the Capital One hack exposing 100 million credentials. Then in 2020 IT Service Management provider SolarWinds had its network management software used by cyber attackers to infiltrate 18,000 enterprise and government customers.

The recent history of cloud storage would be incomplete without mention of the drastic changes caused by the Covid-19 pandemic. With many workforces distributed in home environments, this meant that data storage and security needed to adapt to models that better fit the decentralized workforce.

End Users Fight Back on Data Privacy

The late 2010's saw end users fatigued with data breaches. They were becoming numb to notifications of their data being compromised and frustrated with the many data policy updates that were lengthy with legal jargon which customers saw as an effort by large companies to limit their liability and responsibility to prevent such breaches. Eventually, users started speaking up about their frustrations, which only increased tensions between companies and their customers.

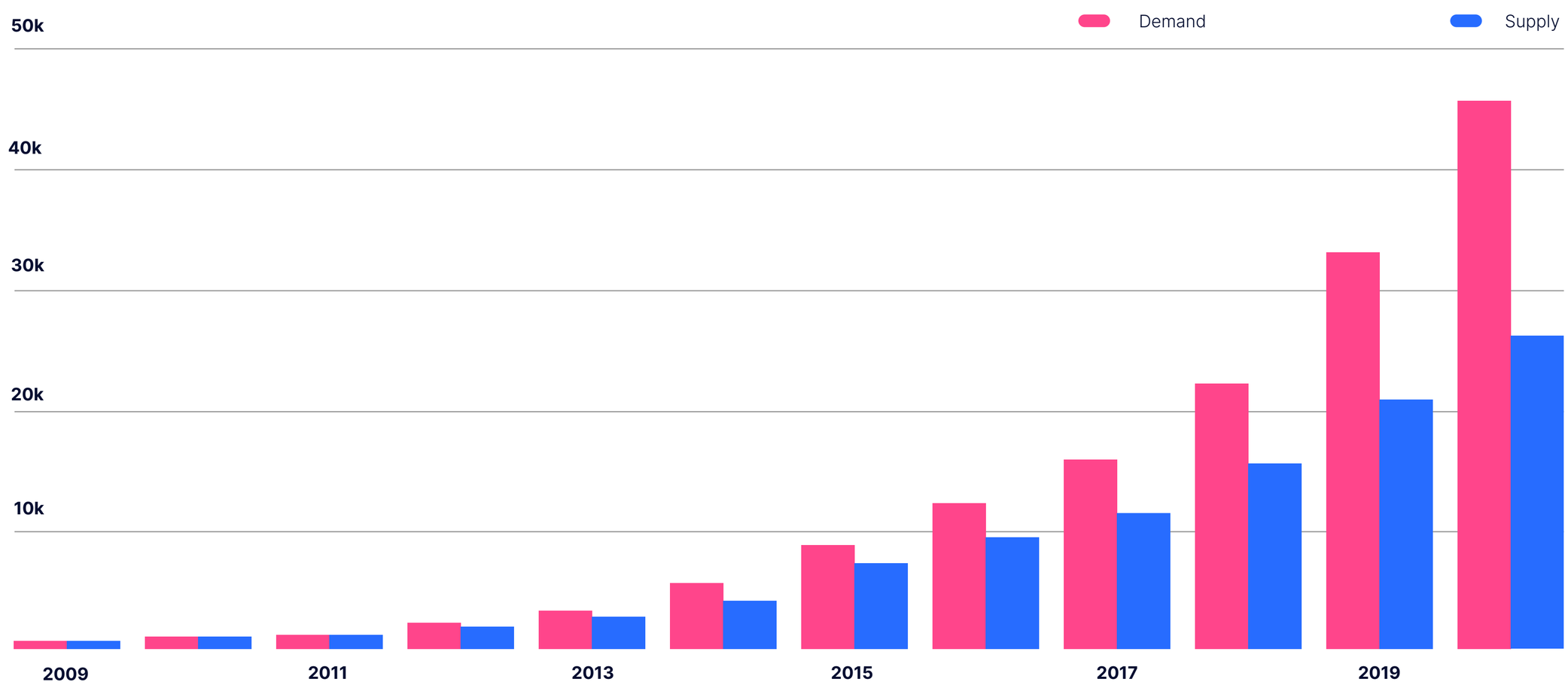
Recently, WhatsApp was the recipient of customer backlash over a privacy policy change about data sharing. Customers misunderstood the policy, believing that the company was changing to be able to read their conversations and personal data. While that was not the case, many users left the app and went to competing messaging services like Signal and Telegram that were perceived as having tighter privacy controls. WhatsApp did their best to reassure customers that it could not read their messages and that it used end-to-end encryption, but still lost an estimated 32.5 million users over the debacle.

Exponential Data Growth Creates Exponential Costs

With the incredible technology advancements made possible by centralized cloud computing has come an incredible increase in data. The International Data Corporation (IDC) is seeing a 61% compounded annual growth rate of data reaching 175 zettabytes by 2025*. With more data being produced, companies are facing increased costs for centralized cloud storage. There has been little to no reduction in the price of centralized cloud storage in the last 6 years, despite Moore's law. Instead, the need for multi-region and multi-geography redundancy has increased costs and fees have been added for data movement and egress. In fact, bills are getting so large and so complex that businesses (like Ducktools and Honeycomb) have arisen just to help companies make sense of their cloud storage bills. What once was an inexpensive way to start a business becomes a punishing expense when you become successful.

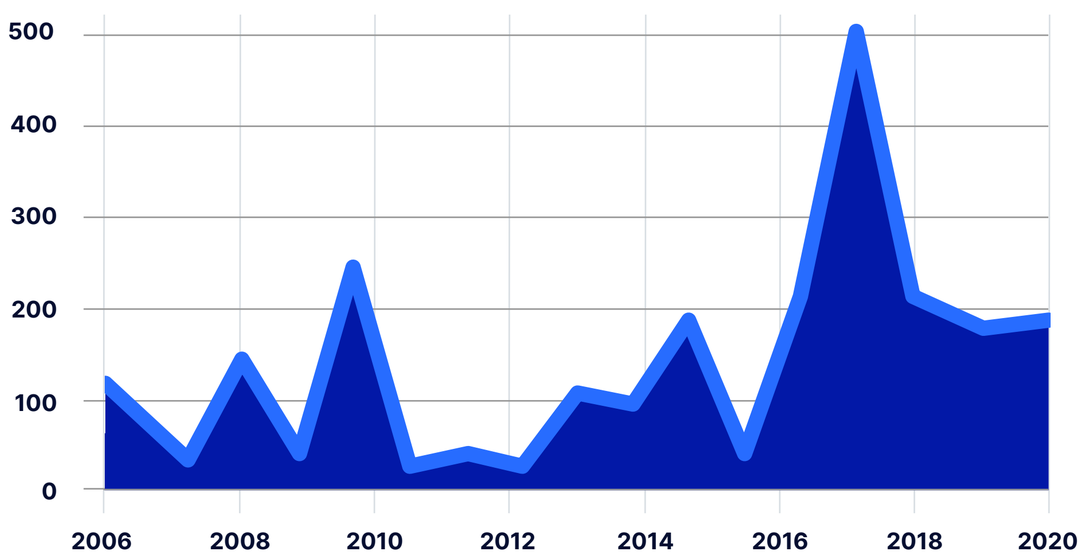
*IDC, Data Age 2025: The Digitization of the World, 2018.

Data Storage Supply & Demand (Exabytes) *Source: Statistica, 2017*



Cost, Privacy and Security Concerns Continue to Grow

Data Records Exposed (Millions) *Source: Statistica, 2021*



While centralized cloud storage has many benefits over on-premise data storage, the issues of rising costs, data privacy and security continued to grow over the past five years with no signs of stopping. This is creating the perfect storm of concern that is both driving further investment in decentralized and edge cloud services while also driving developers to look closely at these now viable alternatives.

How Decentralized Cloud Storage Solves Centralized Cloud Problems

Problem	Solution
Data Breaches	With decentralized cloud storage, the data is encrypted and split into many pieces and storage on different nodes. The data also requires an encryption key (which is held by the data owner) in order to make sense of it. Because node operators don't have the keys, no node operator or malicious actor can find, access, or decrypt the pieces to make sense of the data.
Data Outages	There is a realistic fear of data being stored in a centralized location where an act of terrorism or a major utility issue could cause the data to be inaccessible. With decentralized cloud storage, the data is spread out across geographies. Additionally, not all the pieces are needed in order to access the data so a power failure at a node operator does not impact data accessibility.
Scalability for Data Growth	Thanks to Moore's law and the continued improvements in bandwidth, the amount of storage capacity in people's homes continues to grow. With thousands and potentially millions of nodes hosting your data, the amount of available storage is significantly higher than with centralized operators.This provides the ability to scale with the rising generation of data.
Rising Costs	The availability of storage capacity at the edge and the lack of overhead in building and maintaining large data centers naturally leads to lower costs. Additionally, fair market prices can be achieved thanks to continuous competition amongst nodes where only the highest quality nodes will be utilized.
Data privacy	With decentralized cloud storage there is no business entity that can access your data in any way, unless you provide them the keys.

Where Does Cloud Storage Go From Here?

Over the next five years and beyond, all evidence points to the problems with centralized cloud storage only getting worse. As data increases exponentially, cost, privacy and security issues will be magnified. These issues will likely cause continued evolution in regulatory frameworks and continued awareness and activism from consumers.

We believe these pressures will drive the adoption of decentralized storage. To meet the demand, edge computing providers will likely work in partnership with decentralized storage providers to create a holistic alternative to centralized compute and storage services. In the near term, the obvious use cases for decentralized storage, such as archival and large file transfer, will begin shifting from centralized to decentralized storage models. During this time, awareness of node operation will proliferate, and the network of decentralized storage nodes will multiply.

Top 5 Predictions for the Next 5 Years of Cloud Storage

1. **Problems with centralized cloud storage will continue to worsen**
2. **Regulatory frameworks will evolve to better protect user's data privacy**
3. **Edge compute and decentralized storage will partner for a comprehensive alternative**
4. **Adoption of decentralized storage will grow for obvious use cases**
5. **Decentralized network of storage nodes will continue to grow**

Next Steps for Developers Considering Decentralized Cloud Storage

If you or your company is considering alternatives to centralized cloud storage, here are the key areas to evaluate for fit to ensure you are selecting a path that is affordable, fast and secure for your end users.

1. Compare security & privacy.

You'll want to understand the encryption method as well as how the files are split up, stored and assembled. The number of shards, handling of encryption keys, how sharing and access management work, and file backup are essential elements that should be compared.

2. Ensure long-term cost efficiency.

Ensure that the pricing will be beneficial now and into the future. Check that there are flat fees for capacity and bandwidth, no lock-in fees for egress or multi-region, and that the positive economics of scaling are clear and available to you.

3. Check SLAs, uptime and availability.

You'll want to understand the encryption method as well as how the files are split up, stored and assembled. The number of shards, handling of encryption keys, how sharing and access management work, and file backup are essential elements that should be compared.

4. Make sure your use case is a fit.

Make sure that your application is a fit for decentralized storage. The best use cases today are video storage and streaming, backup, point-to-port file transfer for large files, media serving and software distribution.

5. Evaluate compute compatibility.

If you need to mutate your data you'll want to make sure your decentralized storage provider has partnered with someone who can offer affordable egress bandwidth from your compute platform, such as Storj's hosted gateway's peering agreements.

6. Read up on model consistency.

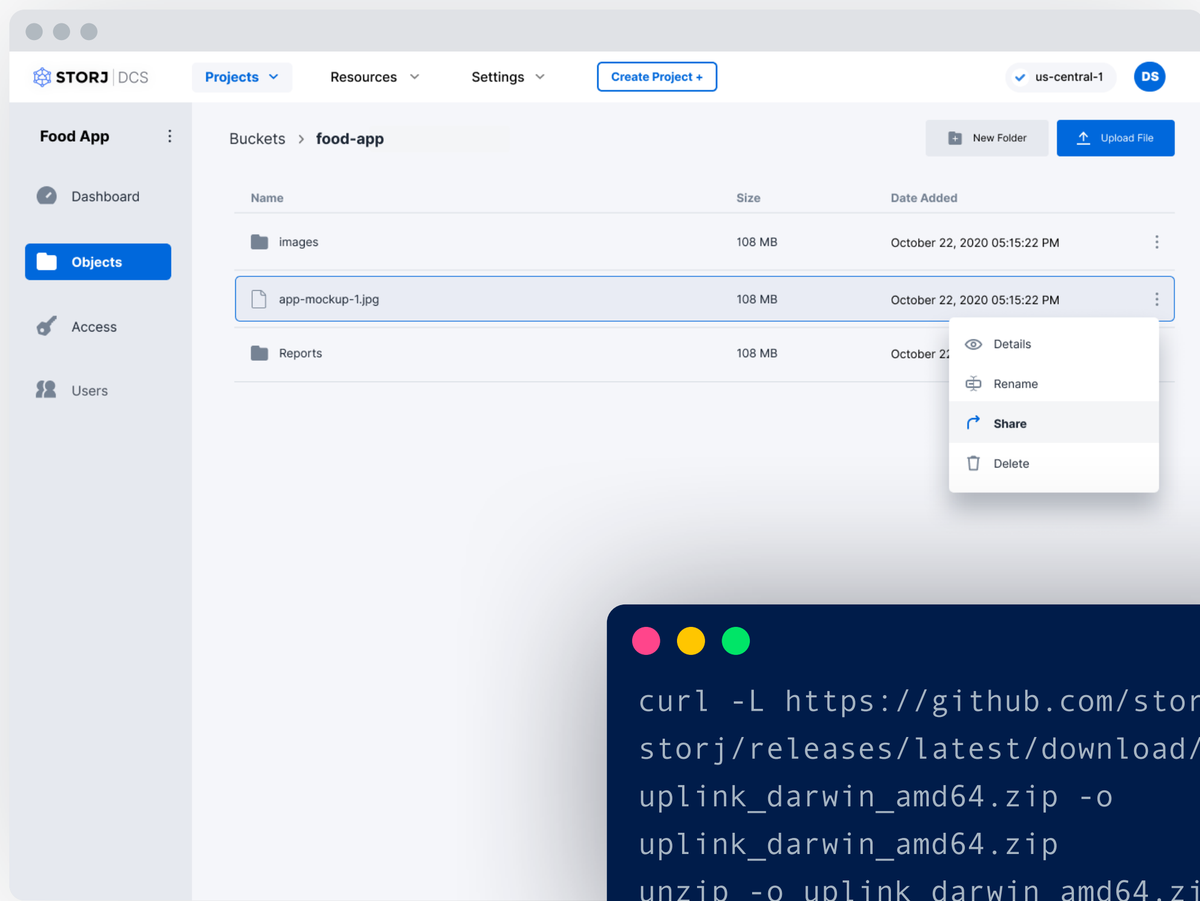
Decentralized storage providers have different behaviors when a data model changes. Which operations are atomic, and when do they show up to other clients? The consistency model of your storage provider will make a big difference to your application's semantics.

7. Pilot the tool set.

Decentralized storage providers have different behaviors when a data model changes. Which operations are atomic, and when do they show up to other clients? The consistency model of your storage provider will make a big difference to your application's semantics.

Experience Storj DCS today.

Decentralization is already here, and it's only going to get bigger, better and more mainstream as people discover the benefits of a decentralized model. Head over to www.storj.io and see how the unparalleled security, availability and management features of Storj DCS can start benefiting your project, organization and bottom line today.




```
curl -L https://github.com/storj/
storj/releases/latest/download/
uplink_darwin_amd64.zip -o
uplink_darwin_amd64.zip
unzip -o uplink_darwin_amd64.zip
chmod 755 uplink
```

STORJ | DCS

Start building on the decentralized cloud.

www.storj.io

@storj

github.com/Storj

storj.io/blog