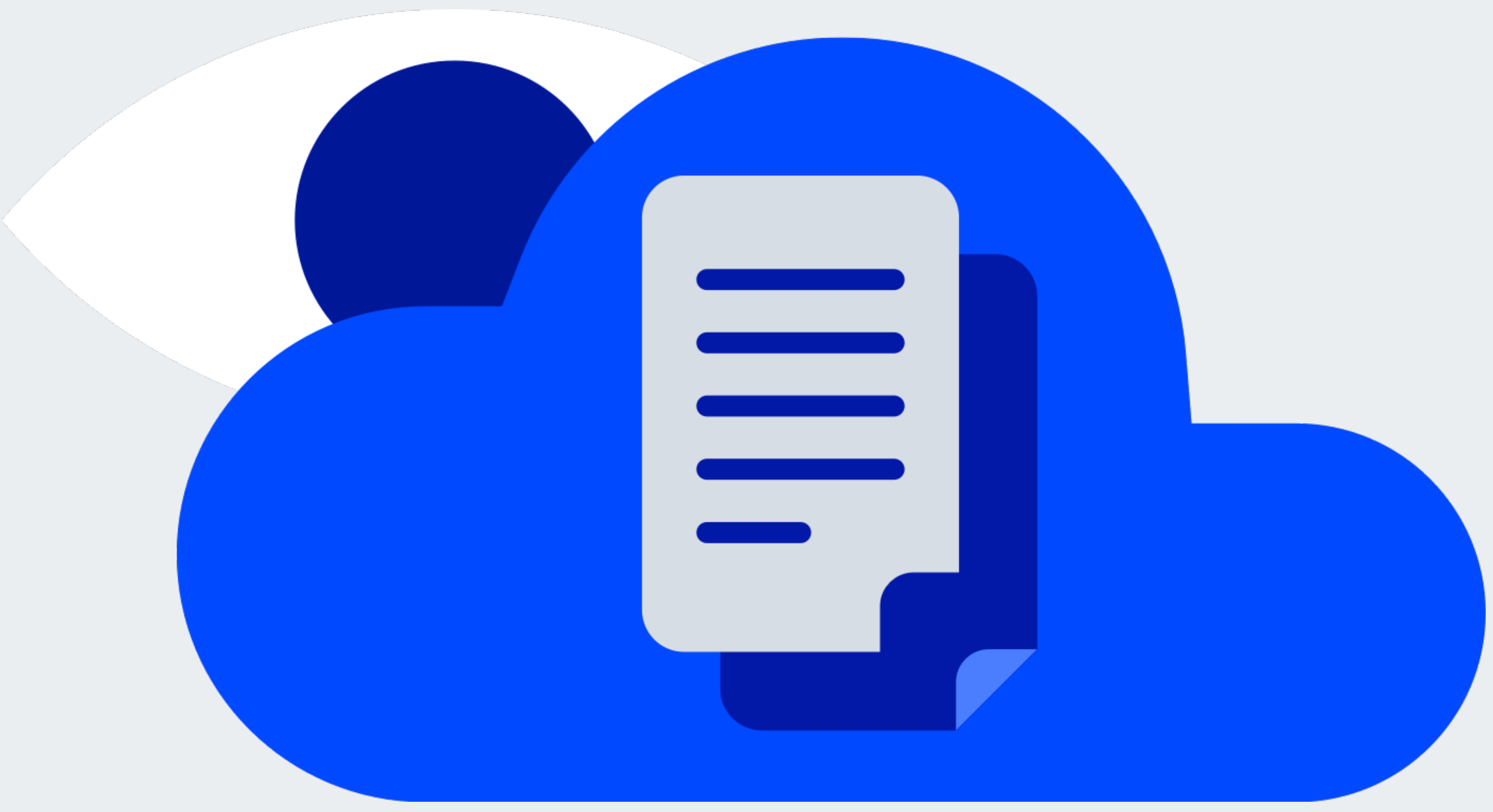
Cloud Storage and Privacy



Data Privacy–Topping the List of Concerns with Cloud Storage

Centralized cloud storage, like that offered by hyperscalers AWS S3, Google Cloud, and Microsoft Azure, were once promising alternatives to on-premise storage. But recently, companies are finding that centralized storage can be problematic due to issues like rising costs and complexities as data grows—plus increasing concerns around security and privacy.

These are the hard facts creating mounting operational challenges for companies who need and rely on accessibility, performance, privacy and scale from their cloud storage solutions.

While rising costs and security risks are motivating companies to look for alternatives, the most significant driver is a perceived lack of data privacy. Users are more aware of privacy in general and concerned about what personal data is collected and shared when they utilize a service or application. The largest centralized cloud storage providers have built their overall businesses on products that use personal data and activity to increase their revenue. At a time when large technology companies have been fined hundreds of millions of dollars for violating data privacy laws, organizations, developers and product teams are sometimes suspicious of the motivations of those same companies in storing their data.

In addition, insider threats and negligence often rank as the #1 offender in terms of privacy and security breaches. For example, according to research from Verizon Business, 85% of breaches involve the human element. So, the more people involved in touching and managing your data, the more likely a potential breach. And the complexities and costs around implementing a Zero Trust implementation to avoid these risks are significant.

Many companies have recognized these as valid concerns and are trying to show their users they're doing all they can to protect their data privacy. A part of this effort is finding a viable alternative for centralized cloud storage to ensure data privacy. Decentralized cloud storage has evolved to a point where it can now be compared to the centralized storage model as not only a viable alternative, but one that delivers better privacy and security - with less cost and complexity. Keep reading to learn how centralized and decentralized storage services handle data privacy and the risks associated with both.

Concerns Companies Have with Cloud Storage

1. Ensuring user data will be kept private

2. Rising costs of centralized cloud storage

3. Keeping data secure from hackers/leaks

4. Complexity of managing data globally

> “What you've got is a few hyperscale companies dominating the space. It's becoming an oligarchy where a few companies have all of the data. And many of those companies are mining that data as part of that service and business model. And frankly, consumers are rejecting that proposal in greater numbers.”

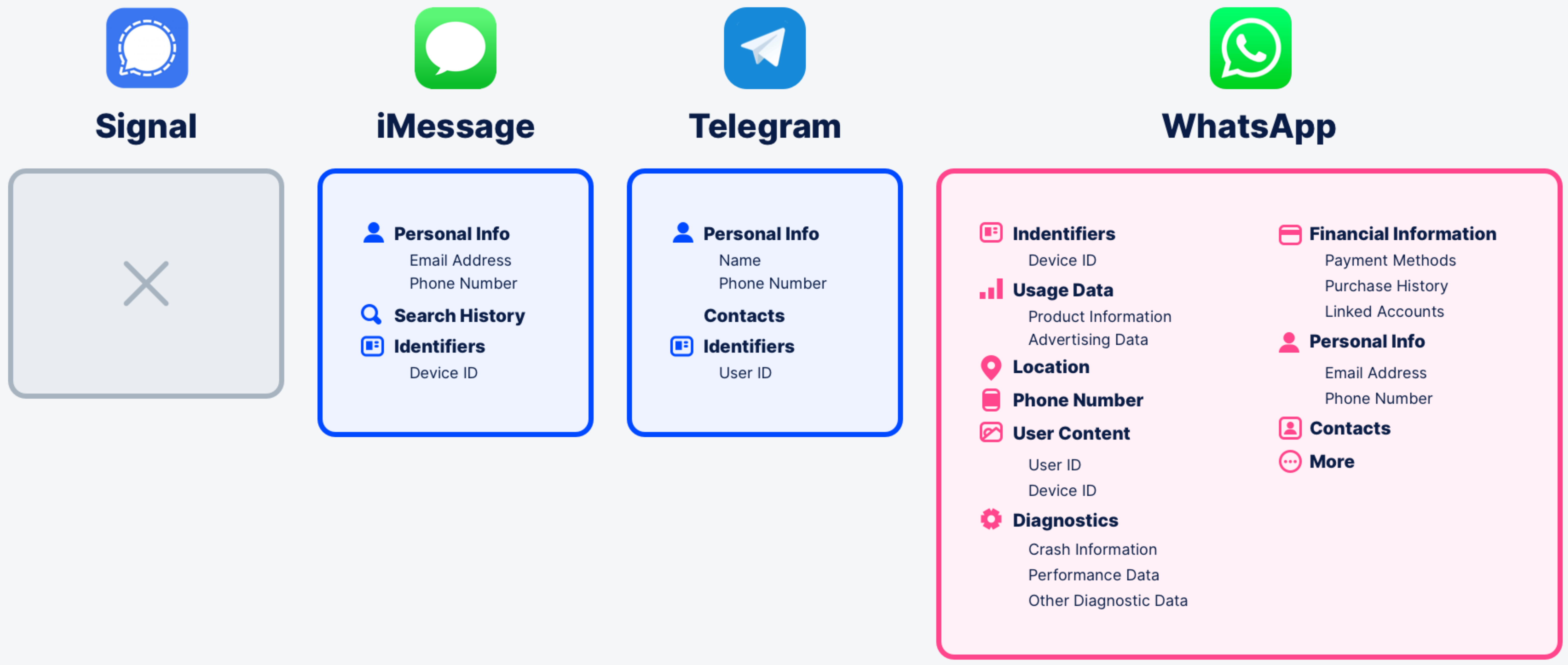
JT Olio
CTO at Storj

WhatsApp and Facebook—Why Data Privacy is in Question for Cloud Storage

Data security is about making sure any data provided is protected from malicious hackers. Data privacy, on the other hand, is focused on metadata, the data about your data, and ensuring partners involved in an application—and that includes cloud storage providers—aren’t harvesting metadata.

WhatsApp came under fire earlier this year as users of the secure messaging app on iPhones were shocked by a new iMessage privacy update. They found that WhatsApp was collecting much more metadata than competitors, and they were then harvesting the metadata and sharing it with Facebook. WhatsApp tried to make a privacy policy update back in January 2021, which resulted in millions of users abandoning the app for competitors like Signal.

Data Linked to You



Signal

×

iMessage

Personal Info
Email Address
Phone Number
Search History
Identifiers
Device ID

Telegram

Personal Info
Name
Phone Number
Contacts
Identifiers
User ID

WhatsApp

Indentifiers
Device ID
Usage Data
Product Information
Advertising Data
Location
Phone Number
User Content
User ID
Device ID
Diagnostics
Crash Information
Performance Data
Other Diagnostic Data

Financial Information
Payment Methods
Purchase History
Linked Accounts
Personal Info
Email Address
Phone Number
Contacts
More

The reality is that WhatsApp data sharing has been happening with Facebook since 2016. Users generally don't have a choice and are forced to accept current privacy policy changes or potentially lose access to the app entirely. Some may leave for competitors, some may stay, but the drama with WhatsApp and Facebook has raised two important issues.

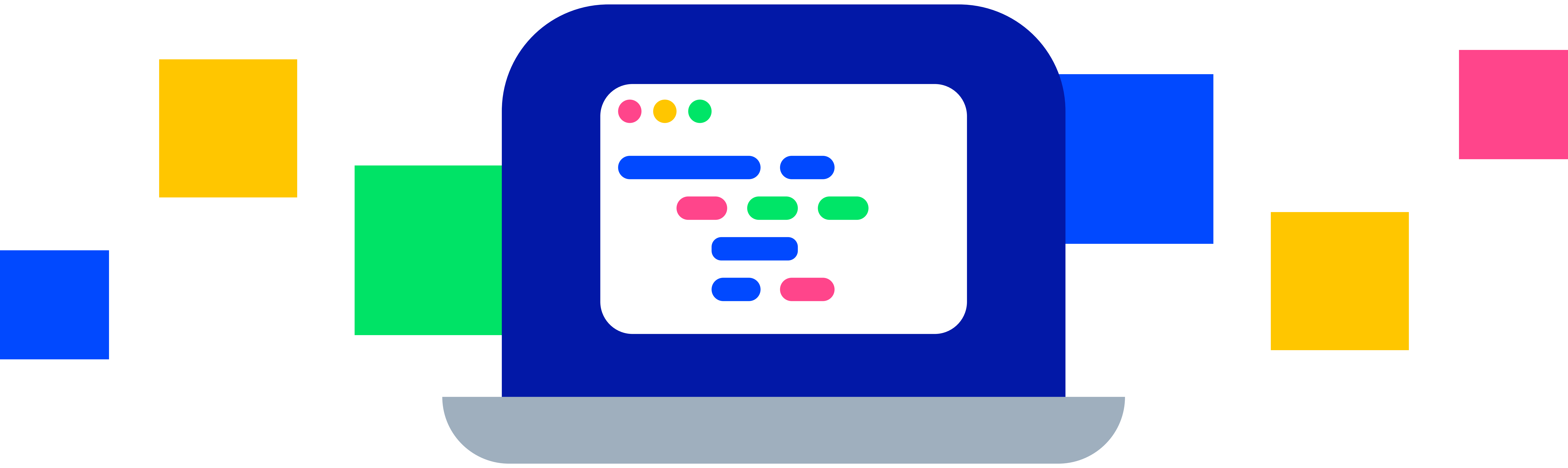
Two Important Issues Raised with WhatsApp and Facebook

1. End-users are more aware than ever before about data privacy

2. Data privacy is important enough that significant revenue is won or lost over this issue

> "WhatsApp's privacy label is awful. It's the only leading secure messenger that harvests data linked to you, including your device ID, for developer's advertising and marketing. It also collects your contact info, user ID, and device ID for ominously vague other purposes. Other messengers collect your data to tailor functionality. WhatsApp is harvesting it for other reasons."

Zak Doffman
Cybersecurity CEO & Forbes Contributor



Developers Are Making Changes

Developers have always been wary about data privacy, but even more so today. They're more careful about what metadata they're collecting—that said, metadata can be an important part of business value to understand user behavior. Understanding trends in how users utilize the application can help developers improve their products, and so some collection may be necessary. If a business decision is made to collect metadata, developers need to ensure it isn't shared with partners who use it for their gain—especially if this sharing isn't in line with your user's wishes.

Connecting the dots makes utilizing AWS, Google, and Microsoft cloud storage a perceived concern by end-users relative to their data privacy. Each of these companies has other business products relying on personal information, giving them the motivation to access the metadata stored on their service to enhance their product offerings. If you've got a great application, end-users are willing to provide personal data in order to use it. But they aren't so keen on that information being used for other purposes—particularly for revenue gain.

Despite what privacy policies might say, developers are now worried that users may perceive a lack of data privacy when using these storage providers. And as seen in the recent WhatsApp example, this poses a significant risk to their revenue model.

Developer Checklist for Data Privacy

☑ Minimize metadata being collected

☑ Ensure integration partners don't have access to user metadata

☑ Select cloud storage provider that has no possible use case to collect metadata

"With centralized cloud storage, there are large privacy concerns. If your data is stored in a non-encrypted way, all in one piece at any institution, it can be mined. If you have your data being stored by organizations who are in the business of selling ads and selling merchandise, it's really hard to believe that it won't be used in ways that perhaps you didn't intend."

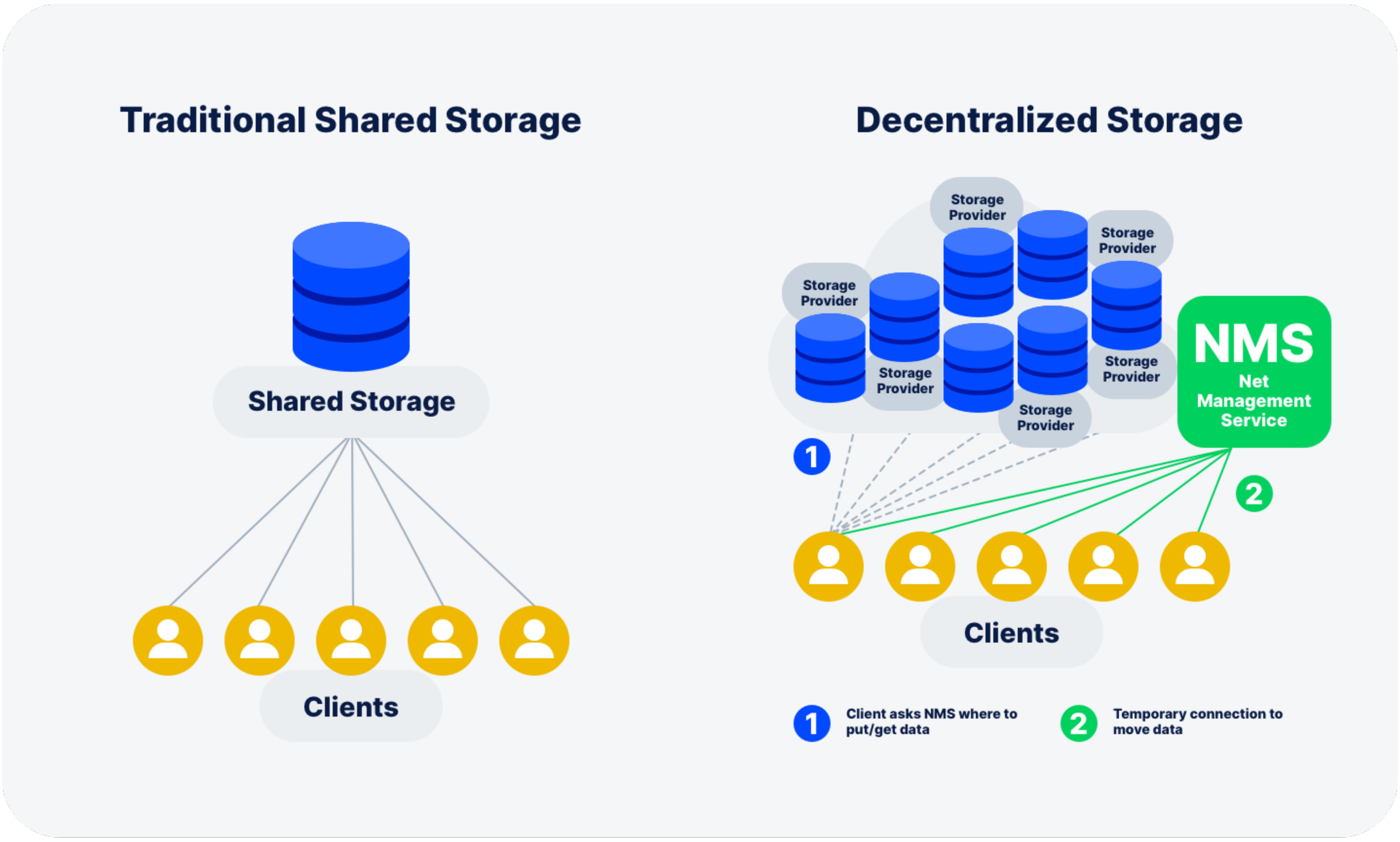
Ben Golub
CEO at Storj

"Developers want to make private software. Not just because their customers want it, but because it's the right thing to do. They want to make sure that what they're building is contributing to the betterment of the people who use what they're building."

JT Olio
CTO at Storj

Comparing Cloud Storage Architecture and Design for Data Privacy

Cloud storage applies to any service in which data is transmitted and stored on remote storage systems. Centralized and decentralized storage have significant differences in their architecture and design, and these architectural differences have an enormous impact on data privacy.



Source: GigaOm Sonar 2021

In a recent GigaOm Sonar Report, GigaOm compared and contrasted centralized and decentralized cloud storage as shown in this basic illustration below. This illustration represents how data is transferred and stored in centralized vs. decentralized storage systems, and it's a stark difference.

How Centralized Cloud Storage Works

The basic model of centralized cloud storage is built on an infrastructure of virtualized storage, where cloud-based data gets stored in logical pools across commodity storage servers. These centralized servers are in data centers located in various regions throughout the world to provide availability. Efficiencies are gained by metering these resources within the data centers, thereby making it cost-effective for data storage.

Most centralized cloud storage providers replicate storage across multiple data centers—they do this to ensure data is available should a region experience a natural disaster, power outage, or malicious attack, which would mean the data being stored in that location wouldn't be available. Take, for instance, AWS—they have three zones of replication, and in theory, if one zone should go down, the other two should be able to keep data available for their customers. Cloud storage data centers do their best to employ security measures to keep data protected. Still, ultimately, the model involves "honeypots" of data that are centralized within physical buildings—WikiLeaks has even made these locations public knowledge.

Centralized cloud storage providers also offer data encryption, but this is always at an additional cost and is not included in the base offer. The real problem with this is that unless you're willing to pay extra for client-side encryption, these providers can access your data whenever they want.

The defining issues with data privacy in centralized cloud storage:

1. The data is held in centralized physical locations with no encryption or minimal encryption
2. The storage provider has access to that data, particularly the metadata

How Decentralized Cloud Storage Works

Decentralized cloud storage is a collection of services and responsibilities brought together under a zero-trust architecture to ensure data privacy and security. Instead of data centers, decentralized storage utilizes shared hard drives worldwide, referred to as Storage Nodes.

Storage Nodes are computers, servers, and other storage devices run by individuals or companies loaning out unused space who get paid for their storage capacity and bandwidth. All of the data stored on Storage Nodes are default encrypted and erasure-coded. This means that in the decentralized model, the people hosting the data can't ever access it.

Decentralized cloud storage also includes network management services to monitor the health of the overall network and ensure data integrity. These services can be centralized or distributed depending upon the model. One of the most secure decentralized cloud storage providers uses Satellite uplinks to store and retrieve data across Storage Nodes. The uplink client automatically encrypts object data and metadata.

Data encryption is a standard part of decentralized cloud data storage as it is necessary to maintain the zero-trust security posture, where the system is designed assuming no entity that interacts with the data can be trusted.

> “Decentralized systems, just by their design, make it impossible for people to compromise the metadata. Which means that it's just far more difficult for anybody to compromise privacy.”

Ben Golub
CEO at Storj

Here’s How Encryption Works on Decentralized Cloud Storage

One of the easiest ways to illustrate how the decentralized network functions is to walk through the lifecycle of an object stored on the network:

- Uploaded objects are encrypted, then split into anywhere from 5 to 80 or more pieces depending upon the provider
- Pieces get distributed across tens of thousands of Storage Nodes and ISPs worldwide
- For downloads, the pieces needed to reconstitute an object (a fraction of the total) are located on the closest and fastest Storage Nodes, which are then retrieved, decrypted, and re-assembled
- This is an automatic process with every upload and download and adds significant security, availability, and performance benefits

Developers Have Complete Control Over Data Access

Within decentralized cloud storage, it's straightforward for developers to build applications that enable file sharing where no component in the network other than the developer’s application has access to the data. All access or sharing of any data is completely controlled on the client-side—this means any server that is serving up a piece of data only gets an access request. The server doesn’t know anything about users or who has access to what. The server simply recognizes whether the request is valid or not valid. If the request is valid, it returns encrypted data—after this, it is decrypted client-side.

The decentralized architecture makes it simple for developers to automate the management of privacy and access management in entirely separate ways. There is no additional work for developers to do, unlike centralized services—they simply share a file. The whole process is automated and the ‘behind the scenes’ technology separates access management from encryption to keep that Zero Trust service in place.

Fundamentally, security and privacy are built into a decentralized cloud storage model. Are you interested in a more detailed synopsis? Learn more details on how it works here. Are you interested in an even deeper dive? Learn about the technical infrastructure relative to protecting data privacy here.

The defining advantages for data privacy in decentralized cloud storage:

1. Data is automatically split, encrypted, and held in thousands of geographically diverse locations
2. Storage Nodes don’t have access to data or metadata, nor do they know what files are being stored

> “When you're interacting with Google or Microsoft or Amazon, these large companies have to some degree lost a lot of trust by being motivated by things that you're not paying them for. End users recognize this and are concerned with the perceived incentive model for centralized cloud storage providers to access their data.”

JT Olio
CTO at Storj

Comparing Privacy Policies in Cloud Storage

Application users don’t generally understand long and complex privacy policies, and most don’t care to try and understand them. Part of the complexity comes from adapting to data regulations, and part comes from the many partnerships and entities that interact with data from a single application. A recent survey of technology lawyers revealed their biggest challenge is “data and privacy legal and regulatory concerns,” followed closely by “the need for technical understanding of complex subject matter.” If lawyers are struggling with privacy policies, then it comes as no surprise that end users are frustrated too.

The largest centralized cloud storage providers have done a great job of trying to simplify the language of their policies and organize them with bullets and dropdown menus to make them easier to read, but the real issue users have is at the start of all of these policies—Data We Collect. Users want this statement to be followed by the answer “none.” The reality is much different. Beyond collecting data, these policies go on to talk about “How We Share Data” and “How We Use Data,” as well as a host of other notions that are concerning to users. The first page of the big three cloud providers’ policies are shown below with links to view their complete statements.

aws
Products Solutions Pricing Documentation Learn Partner Network AWS Marketplace
aws

Privacy Notice

Last Updated: February 9, 2021

This Privacy Notice describes how we collect and use your personal information in relation to ... Privacy Notice (together, “AWS Offerings”).

This Privacy Notice does not apply to the “content” processed, stored, or hosted by our ...

AWS Data Privacy Policy ›

Microsoft | Privacy Privacy dashboard Privacy report Privacy resources
Microsoft Azure

Microsoft Privacy Statement

Microsoft offers a wide range of products, including server products used to help operate enterprises ... software that students use at school, and services developers use to create and host what's next. ... include Microsoft services, websites, apps, software, servers, and devices.

Please read the product-specific details in this privacy statement, which provide additional relevan...

Azure Data Privacy Policy ›

Google Cloud Privacy Noti...

Effective Date: July 1, 2021

GCS Data Privacy Policy ›

"Decentralized storage gives developers a cost-effective option while providing better control over data with embedded security, geo-distribution, and other features that are very expensive and more difficult to manage in traditional environments."

Enrico Signoretti
GigaOm

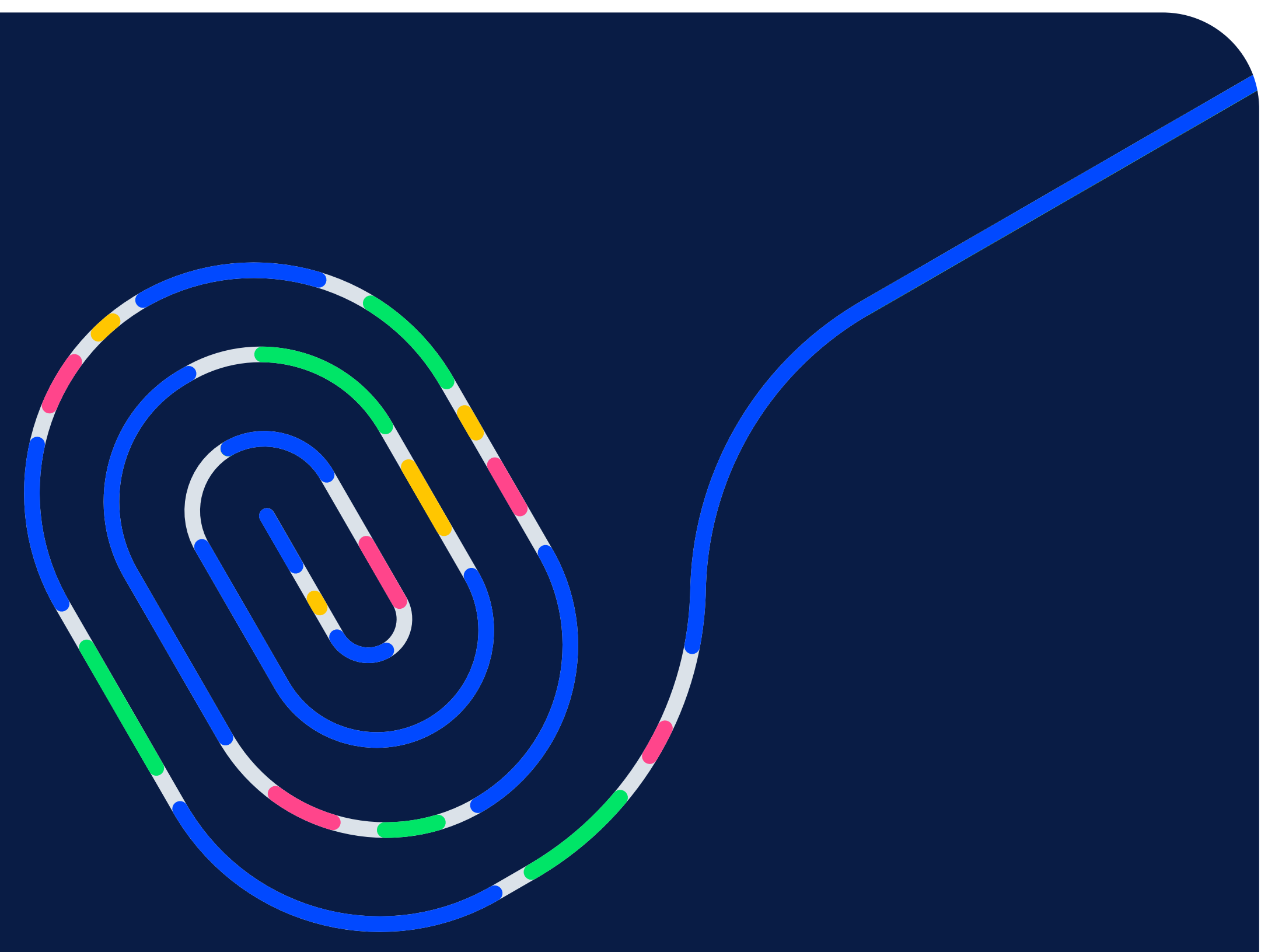
Beware of "Free" Storage

Many early-stage projects get started using reduced price or free tiers of storage. But just as free apps make you face advertisements, free storage comes with the cost of data sharing. These services often have privacy policies that allow cloud providers to mine that data, and while no individual person may be mining the data, AI may be scanning it for marketing and ad targeting purposes.

How Do End Users Feel About Their Online Data Privacy?

The Conference Board, in collaboration with Nielsen recently surveyed more than 30,000 consumers across 63 global markets about their attitudes on online data practices. Over 20 percent of respondents report having reduced or abandoned their use of a brand or company due to data privacy concerns. Additionally, 19% report having switched to or selected a competitor company for its better data policies.

The main problems users have when it comes to privacy comes down to how much they trust the provider they are giving their data to. Centralized cloud storage providers have fairly protective terms of service for themselves, and they have permissive default settings for themselves to access or use your data. While most providers offer additional features to have better encryption, most companies find this cost prohibitive and end up on their basic offerings, which is storage, where the providers have access to all of that data.



"Decentralized storage changes the dynamic with privacy in a way that is much more favorable to the end-user. At the end of the day, no matter what you're doing on the internet, you're interacting with someone else's server. So the question is who and why, and what are they motivated by?"

JT Olio
CTO at Storj

How Decentralized Cloud Storage Helps Eliminate Data Privacy Concerns

The beauty of decentralized cloud storage is that it's impossible for the provider to access stored data. All objects and associated metadata are encrypted using default AES 256 CTR Encryption. Segments are encrypted using a salted, randomized encryption key that is then encrypted with the user's encryption passphrase and stored in the object metadata. This truly is end-to-end encryption—which is the most secure and private encryption available. No entity involved in decentralized storage ever has access to encryption information, nor is there any way for a malicious actor to access the data.

This means that data privacy policies can be greatly simplified when using decentralized cloud storage. Companies can reassure users that their data can't be shared or accessed by third parties, which means a lot to today's potential users.

Comparing Data Privacy Risk in Cloud Storage

There are two main risks when it comes to data privacy:

1. Losing business because users no longer trust in the way you handle their data
2. Getting fined by regulatory bodies for non-compliant data handling

Both of these risks come with costs that can bankrupt large enterprises, and would decimate small startups. The WhatsApp example provided earlier demonstrates the genuine concern about how easy it can be to lose trust because of a real or perceived wrongdoing with data sharing.

Regarding regulatory compliance, laws are being rapidly passed to protect consumer data better. There's no single law that governs data privacy for each country. Instead, laws are being made at the state level in the United States, with a few federal laws for special use cases. Internationally, the General Data Protection Regulation is the most well known, but many countries have put specific regulations into place.

While many of these laws are still new, many companies have already been fined significant amounts of money for breaking them. None have been more prominent than the GDPR fine to Amazon in 2021 of $877 million. The biggest GDPR fines also include Google, which has been hit twice with GDPR fines totaling $65 million. The GDPR has fined an estimated 1,000 companies in the past two years with cumulative penalties of $1.25 billion.

> "If a cloud service provider doesn't ever look at your data and you're the only one who ever looks at it, then great, your privacy hasn't been violated. But can we trust these companies to do that—now, and indefinitely into the future—when their clearly stated business goals are so anti-user sometimes?"

JT Olio
CTO at Storj

Major Data Privacy Regulations

Gartner predicts that by 2023, 65% of people worldwide will have their personal data protected by privacy regulations. The most notable regulations today include the European Union's General Data Protection Regulation, the first major comprehensive data privacy law affecting multiple countries, and California's Consumer Data Privacy Act, which was the first broad U.S. data privacy law. They are notable because they are the first comprehensive privacy regulations, and these continue to grow in countries like China.

A list of data privacy regulations throughout the world can be found here. Some of the more active data privacy regulations are shown below.

United States

Federal
The Federal Trade Commission Act (FTC Act)

Federal
Children's Online Privacy Protection Act (COPPA)

Federal
Health Insurance Portability and Accounting Act (HIPAA)

Federal
Gramm Leach Bliley Act (GLBA)

Federal
Fair Credit Reporting Act (FCRA)

California
California Consumer Privacy Act (CCPA)

Virginia
Consumer Data Protection Act (CDPA)

Colorado
The Colorado Privacy Act

New York
Stop Hacks and Improve Electronic Data Security (SHIELD) Act

European Union

All EU Members except UK
General Data Protection Regulation (GDPR)

United Kingdom
UK GDPR

Brazil

Federal
General Data Protection Law

Federal
Lei Geral de Proteção de Dados (LGPD)

South Africa

Federal
Protectionof Personal Information Act (POPIA)

Bahrain

Federal
Data Protection Law

Philippines

Federal
Data Privacy Act of 2012

Canada

Federal
Personal Information Protection and Electronic Documents Act (PIPEDA)

What Can Companies Do to Lower Risk?

As a company trying to operate a business online, this is a daunting amount of risk to undertake. All the more reason for companies to be looking closely at data governance and how data is being protected and kept private.

A great way to reduce the risk of non-compliance is to utilize services that are architected for zero trust. Decentralized cloud storage is a perfect example as it is designed to eliminate the chance of non-compliance because of malicious actors. Decentralized cloud storage gives developers the control over their data that you just can't fully achieve with centralized cloud storage.

Summative Comparison of Data Privacy in Cloud Storage Models

Data Privacy Variables	Centralized Cloud Storage	Decentralized Cloud Storage
Public perception of trust	Very low	High
Access to metadata	High	None
Incentive to access metadata	High	None
Possible access by malicious actors	Low-Med	None
Standard end-to-end data encryption	None	High
Developer control over data access	Low-Med	High

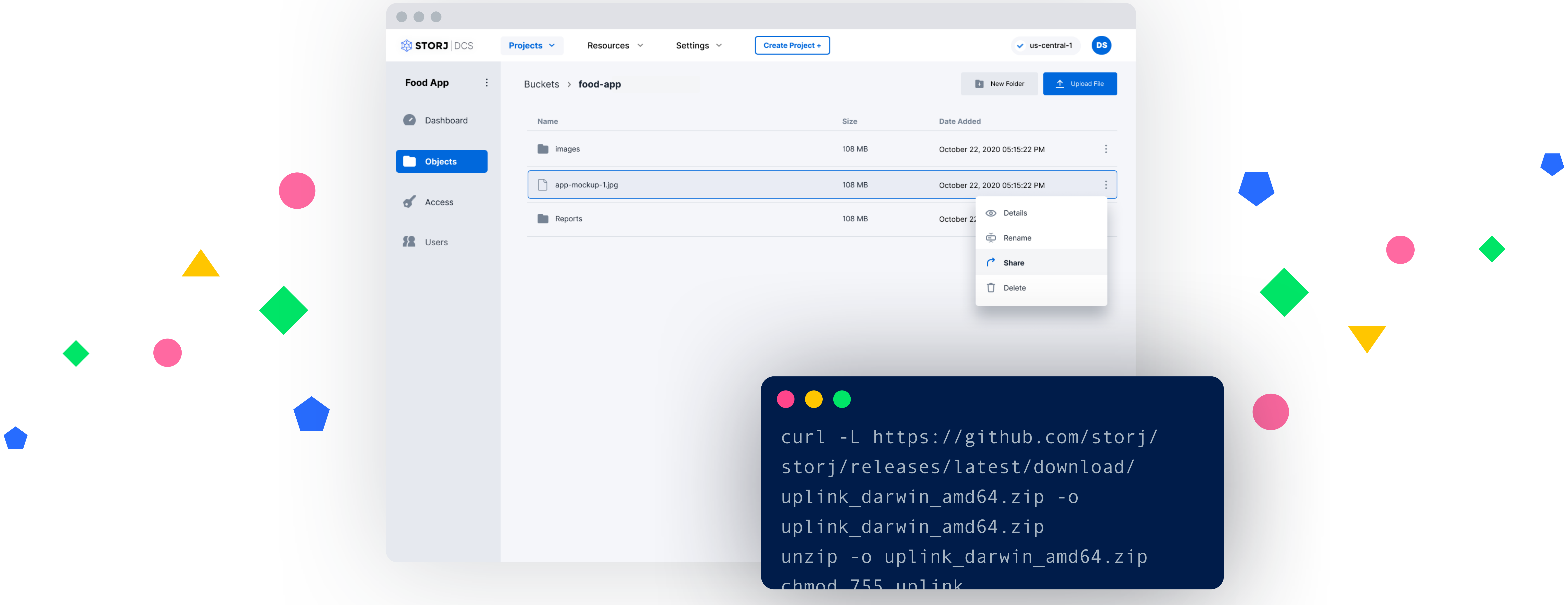
When looking for cloud object storage that ensures data privacy for your business, decentralized cloud storage is a solid new alternative worth exploring. The GigaOm Sonar report for Decentralized Storage is a great resource to evaluate vendors.

Here are some other helpful resources on privacy gains through decentralized cloud storage:

- Practical Application of Storj DCS Edge-based Privacy in Your Application
- Security and Privacy Benefits of Decentralized Cloud Object Storage | A TechTarget Review
- Cloud Storage Priorities | Privacy & Security
- Decentralization: A New Standard in Data Security
- On-Demand Webinar: Privacy & Security without Complexity

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. For more information on how Storj DCS can help your development team and organization secure your data, minimize storage costs, reduce complexity and increase performance of your backups, visit www.storj.io.



Name	Size	Date Added
images	108 MB	October 22, 2020 05:15:22 PM
app-mockup-1.jpg	108 MB	October 22, 2020 05:15:22 PM
Reports	108 MB	October 22

```
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