MARKET NOTE

Object Vendor Storj Surfaces a New, Crowdsourced Model for Cloud-Based Storage

Eric Burgener

EXECUTIVE SNAPSHOT

FIGURE 1

Executive Snapshot: The Rise of Economically Compelling Cloud-Based Object Storage

Storage start-up Storj sells a cloud-based object storage service that costs up to 80% less than service from public cloud storage vendors like Amazon Web Services, Microsoft, and Google. The secret is primarily in how it builds, maintains, and grows its supporting infrastructure, using a unique crowdsourced model that underpins a highly available, highly durable, geo-distributed service. As data continues to grow at a rapid rate, 75–80% of the data created in the next five years will be unstructured, so there will be an increasing need for cost-effective unstructured storage solutions like the one Storj is offering.

Key Takeaways

- Put everything you think you know about "crowdsourced" infrastructure aside — Storj sells a cloud-based object storage service that delivers "eleven nines" of durability, default multiregion high availability, and 256-bit security for data both in flight and at rest and guarantees 99.95% availability for its close to 25,000 customers.
- What makes this such a compelling offering is the price point — it is undercutting the object storage services pricing of hyperscaler competitors by as much as 80% while providing reliable storage with extreme durability for less demanding enterprise data used by development, streaming services, content distribution, and data protection workloads.
- The Storj Decentralized Cloud Storage software offering takes tens of thousands of geographically distributed, internet-connected computers with internal storage that are not necessarily highly available and turns them into an easy-to-use object-based storage service.
- Storj does not yet promote the service for certain enterprise workloads, including those subject to compliance regimes or those considered mission critical. The vendor's unique model should be viewed as a "shot across the bow" of existing solutions handling cloud-based enterprise data, and this analyst believes that it is an approach that has the potential to significantly disrupt that market within the next two to three years.

Source: IDC, 2021

June 2021, IDC #US47959821

IN THIS MARKET NOTE

Storj is a vendor of object-based cloud storage services with a very unique model for building and maintaining its web infrastructure: its software product (Storj Decentralized Cloud Storage [DCS]) distributes data across a highly available, highly durable, geo-distributed public cloud-based infrastructure by combining resources from tens of thousands of storage nodes. This approach, discussed in this Market Note, allows Storj to undercut the costs to store data in popular hyperscalers like Amazon Web Services (AWS), Microsoft, and Google by as much as 80%.

IDC'S POINT OF VIEW

As most enterprises undergo digital transformation — the evolution to much more data-centric business models and decision making — data growth is exploding. In 2020 alone, over 44 zettabytes (ZB) of data was created, and data is expected to grow at 30–35% per year for the foreseeable future. By the end of 2020, almost 50% of all data resided in the cloud. With this rapid increase in capacity, the cost to store data is a perennial concern. The advent of the public cloud over a decade ago provided an interesting alternative to purchasing, deploying, and managing wholly owned on-premises infrastructure, and since their humble beginnings, cloud-based storage services have grown into a major component of overall storage infrastructure spend. In 2022, the overall spend worldwide on cloud-based storage infrastructure will be greater than that for on-premises storage infrastructure spend for the first time.

Cloud-based storage brought with it a number of benefits that initially differentiated it from on-premises infrastructure. It significantly increased the agility with which storage could be brought online (and retired), introduced an operational expenditure consumption model that tied costs much more closely with actual usage, enabled nondisruptive multigenerational technology refresh, and offloaded infrastructure management to a third party. The agility appealed in particular to constituencies like developers who were outside the traditional IT operations purview, and a strong argument can be made that cloud-based services in general helped foment the rise of "shadow IT." As cloud offerings matured, they have in many cases become much easier to bring under the umbrella of IT governance while still meeting individual constituency agility requirements.

Public cloud did, however, have its challenges. The market has become dominated by just a few large players — AWS, Microsoft, and Google alone hold an almost 70% market share in cloud storage. This dominance, combined with a pricing model that charges users not only to store their data in the cloud but also to get their data out of the cloud, has led to unpredictable costs and created somewhat of a lock-in that limits data mobility in practice. The actual cost to maintain a single terabyte (TB) of object data in the public cloud per year is over $3,300, a number that might surprise many users who are using that storage but not actually paying the bills for it. While this number may compare favorably in many cases to the costs of traditional on-premises storage infrastructure, there is no doubt that a lower-cost model would be a welcome addition for many users. More predictably, lower-cost offerings for business use are really only of interest when certain other requirements for access, performance, data integrity, availability, and security can be met as well.

Crowdsourcing Storage Infrastructure: A Newly Viable Model

Storj is a storage start-up that is successfully leveraging a very interesting new model to provide cloud-based storage services for commercial use. The vendor has developed an S3-compatible cloud object storage service that is geographically distributed across tens of thousands of storage nodes, each of which contributes both performance (in terms of bandwidth) and storage capacity that provides the

infrastructure foundation for its offering. While storage as a service is not new, making a service available at a price point as low as $4 per terabyte per month with a 99.95% availability guarantee, "eleven nines" of durability, default multiregion high availability, and 256-bit encryption absolutely is. This pricing is up to 80% less than traditional cloud storage providers, providing a compelling economic value proposition.

The way Storj builds out the infrastructure to support its object-based storage service takes advantage of a key data point about computer resource utilization along with a confluence of trends. Research has shown that, across all types of computers (servers, clients, etc.), 70% of hard disk drives are less than 25% full, composing a huge pool of untapped storage capacity that today is effectively wasted. The fact that most of these computers are connected to the internet makes that capacity potentially accessible by the right business model. Storj effectively "rents" this capacity from tens of thousands of internet-connected systems across the world. Owners of that capacity are incented to rent it to Storj because this model lets them monetize storage that would otherwise remain unused. These "suppliers" are called "node operators." In evaluating potential node operators, Storj looks for a minimum of bandwidth, met through an internet connection, and capacity, met through local disk. The rate Storj pays for this capacity is low because the service is designed to take advantage of existing infrastructure versus acquiring new hardware. This existing capacity is frequently fully depreciated and otherwise generates no revenue, and node operators can choose to bow out of the distributed infrastructure at any time (if and when those resources become needed locally).

While the crowdsourcing approach is interesting (not only because of the extremely low cost and the fact that Storj does not have to "maintain" the infrastructure that supports its service), it potentially raises all sorts of issues around availability, durability, and security. This is where its open source, decentralized cloud object storage service software platform comes into play. Storj DCS has three software components:

- The **Storage Nodes** make up the underlying worldwide internet-connected infrastructure to store data.
- The **Satellites** handle all the data management, including data placement, data protection, security, and recovering from failures, and act as the "metadata" servers.
- The **Uplink** is a small software component, loaded on client computers, that allows clients to store and retrieve data using Storj DCS; the Uplink is also implemented as a globally distributed set of hosted S3-compatible gateways.

The data and management capabilities on the Storage Nodes and Satellites are highly distributed, while the Uplink component is specific to a particular client computer. Storing data is a four-stage process: upload, encrypt, split, and distribute. A file to be uploaded is first encrypted (using AES 256-bit encryption) on the client, split into segments, and then distributed (using a modified erasure coding technique) across at least 80 different storage nodes. The erasure coding algorithm uses 29 data slices and 51 parity slices, with no two slices stored on the same physical node. When data must be retrieved, only 29 of the storage nodes need to respond to provide the data. This means that 52 storage nodes would have to be lost before data integrity would be compromised, providing far higher levels of resiliency than that available in most enterprise storage systems that are used for mission-critical workloads. When data needs to be retrieved, the storage nodes with relevant data that can respond fastest are identified by the satellites, the segments are reconstituted and decrypted, and the file is assembled. Satellites monitor storage node availability and recover lost slices on failed nodes on other nodes within the infrastructure in real time, effectively providing a highly resilient data store that is multiregion by default and self-healing with no single point of failure.

This highly distributed erasure coding approach was developed to enable Storj to use potentially inexpensive storage nodes that are not "highly available" for a storage service that is highly available. Storj's current infrastructure is composed of more than 13,000 node operators, spread across more than 90 countries, and many of those node operators (although certainly not all of them) are vendors themselves. Storj is adding new node operators all the time.

Storj DCS is a highly secure storage platform. Encryption keys are maintained on each client, and data is encrypted both in flight and at rest. Because of the highly distributed nature of the erasure coded storage process, the data stored on any single storage node is completely unusable. The storage nodes whose data comprises a file are randomly distributed as data is stored. The platform uses a "zero trust security" model that verifies identities each time data is accessed, providing a much more secure storage system that was engineered with the "everything is connected" world in mind. The satellites (which are also geographically redundant and distributed) manage access controls.

The highly distributed nature of Storj DCS, favorable economics, and generous data exit policies remove any sort of "data gravity" lock-in that public cloud storage may have. Customers storing large data sets in public cloud providers like AWS, Microsoft, and Google may be hesitant to move to other service providers that may offer better economics because of the time and cost (egress costs) associated with data migration. That is clearly not the case with Storj, whose data is already extremely widely distributed across tens of thousands of node operators.

The design of Storj's software-defined storage platform clearly addresses availability, durability, and security requirements, but what about performance? Because of its ability to service data requests from the "closest" nodes (as determined in real time by the satellites), the performance of this environment is surprisingly good. Storj currently has close to 25,000 customers and is storing almost 260 million objects. Common use cases for the platform include video storage and streaming (for those customers not large enough to have their own private content delivery networks), cloud-native applications, software and large file distribution, backup, and log retention.

While this all might look enticing to enterprise IT personnel for less critical workloads, Storj is actually not selling this service to that constituency. Today, the service is primarily being sold to developers and for the non-latency and availability-sensitive workloads mentioned previously. But enterprise IT managers should take note: it is common in the technology industry for new products and services that were initially used for the most forgiving application environments to evolve and over time successfully challenge legacy technologies that sought to differentiate themselves by their "enterprise class" nature. The "crowdsourcing" model has many attractions both for users and for vendors of cloud-based services. If interest in this approach begins to grow, vendors that got in early like Storj will bring valuable experience that may help them make the leap to other types of workloads more easily.

The existing use cases will put this approach to the test and prove (or disprove) its viability. There are enterprise workloads that seem to already be a good fit for the performance, availability, durability, and security profile met by Storj DCS. Developers are likely not paying $3,300 a year for a terabyte of object storage, but they are probably also not paying as little as Storj charges. Enterprises, however, may easily be paying that much if they are doing more than just storing data for long-term retention. Storj very intelligently has chosen to prove the mettle of its service in less demanding segments first. The economics of the crowdsourcing approach are undeniably attractive, and if it proves out, this model has the potential to be a significant disruptor of cloud-based enterprise storage within the next two to three years.

Storj DCS Go-To-Market Model

While Storj sells its storage services in conjunction with channel partners like Fastly, Plesk, and various managed service providers, systems integrators, and cloud service brokers, it is responsible for "maintaining" the underlying distributed infrastructure (most of which is "outsourced" to the node operators themselves). Storj has automated the payment process for node operators, tracking capacity and bandwidth allocation and usage and paying them in cryptocurrency utility tokens (STORJ tokens) in a blockchain-based system. Customers can choose to pay with traditional mechanisms such as credit cards as well as STORJ tokens.

Pricing for public cloud storage can be complex and confusing. The combination of ingress and egress charges, transaction fees, and a variety of hidden charges, driven by actual data usage patterns that may not have been forecast correctly by users, can lead to unpredictably high costs for data stored in the public cloud. Multiregion availability add-ons for those workloads that require them can also add to costs. Research performed by IDC in 2020 indicates that 84% of enterprises have repatriated at least one workload from the public cloud back into on-premises infrastructure, and a top driver of that repatriation for those workloads has been unpredictably high costs.

Given the open source background of some of the executive team at Storj, the vendor has implemented a simple pricing model that leads to easily predictable costs. Users can buy 1TB of storage per month for $4. Data usage patterns are measured by bandwidth, for which customers pay $7 per terabyte of bandwidth per month. All data is multiregion by default and guaranteed for 99.95% availability. The vendor notes that since it first began selling service in 2019, it has delivered 100% durability to all its customers. There is also a "freemium" tier to get customers started that provides 150GB of storage per month and 150GB of bandwidth per month.

LEARN MORE

Related Research

- *Worldwide Service Provider Infrastructure Forecast, 2021-2025* (IDC #US46385521, May 2021)

Synopsis

This IDC Market Note examines Storj, a vendor of object-based cloud storage services with a very unique model for building and maintaining its web infrastructure: its software product (Storj Decentralized Cloud Storage [DCS]) distributes data across a highly available, highly durable, geo-distributed public cloud-based infrastructure by combining resources from tens of thousands of geographically distributed storage nodes. This approach allows Storj to undercut the costs to store data in popular hyperscalers like AWS, Microsoft, and Google by as much as 80%. While this service is not currently being targeted for mainstream enterprise use, if the model proves itself out over time, it could be a significant market disruptor for cloud-based enterprise storage within the next two to three years.

About IDC

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

Global Headquarters

140 Kendrick Street
Building B
Needham, MA 02494
USA
508.872.8200
Twitter: @IDC
blogs.idc.com
www.idc.com