Decentralized Cloud Computing

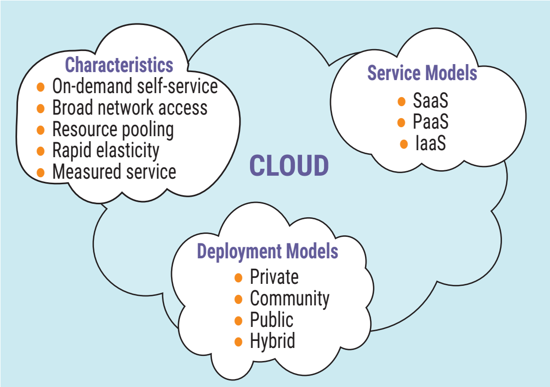
Computing has penetrated all aspects of everyday life from smartphones, cars, televisions and cameras to home appliances. As these smart devices are constantly enhanced with greater computing capabilities, they require new systems that can deliver advanced features and functions on the move.

Cloud computing provides a way to store and access data from anywhere by connecting applications using the Internet. Cloud usage has increased dramatically in the past few years and has seen a tremendous surge during the COVID-19 pandemic as enterprises have been forced to address issues related to remote or virtual working environments, along with unpredictable changes to IT requirements. As a result, enterprises have had to accelerate their journeys to the cloud and digitize quickly and effectively. Meanwhile, cloud computing also introduces risk factors that need to be managed. The on-demand self-service provisioning features of the cloud may enable an organization's personnel to provision additional services without engagement or even being supported by the organization's IT department. The use of such services decreases an organization's visibility and control of its network and data, thus presenting increased risk to the organization. There is a data security risk associated with loss, leakage, unauthorized access and unavailability of data. There is also a regulatory risk associated with noncompliance with various national or geographic legal or regulatory requirements.

Cloud Computing

Cloud computing has been one of the greatest success stories in recent years. The US National Institute of Standards and Technology (NIST) defines cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."[1] According to NIST, cloud computing is composed of the following (**figure 1**):

Figure 1—NIST Definition of Cloud Computing



- **Five essential characteristics**—On-demand self-service, broad network access, resource pooling, rapid elasticity and measured service
- **Three service models**—Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS)

Daniel D. Wu, Ph.D., CISA, CISM
Is an assistant professor of computer and information sciences at Cabrini University (Radnor, Pennsylvania, USA). He can be reached at daniel.d.wu@cabrini.edu.

Enjoying this article?

- Read *Managing Security Impacts in a Multicloud Environment. www.isaca.org/multicloud-security-impacts*
- Learn more about, discuss and collaborate on risk management in ISACA's Online Forums. *https://engage.isaca.org/onlineforums*

- **Four deployment models**—Private cloud, community cloud, public cloud and hybrid cloud

In a private cloud, the infrastructure is provided for the exclusive use of a single enterprise comprising multiple business units. It may be owned, managed and operated by the enterprise itself, by a third party or by some combination of the two, and it may exist on or off premises. In a private cloud, only internal users experience it as a cloud computing service. Therefore, a private cloud does not fully conform to the definition of cloud computing. When the term "cloud" is used, it normally refers to the public cloud.

The public cloud allows a user to run workloads remotely over the Internet in the cloud service provider's (CSP's) data center and gain new capabilities without investing in new hardware or software. The user either pays the CSP a regular subscription fee or pays for only the consumed resources.

The cloud industry is currently dominated by a few technology giants such as Amazon, Google, Microsoft, Dropbox and Salesforce. All these big players use the same basic architecture to sync and copy to a centralized cloud server cluster via the Internet. Thus, the cloud is the ultimate centralized processing network. Millions of users and their devices connect to these central cloud clusters to store and access their data.

Unfortunately, the centralized cloud has many shortcomings. First, it is susceptible to a single point of failure. On 14 December 2020, multiple Google Cloud services and websites, including YouTube, Gmail, Google Assistant and Google Docs, were down for approximately one hour due to a widespread outage.[2] Because cloud computing is Internet-based, service outages can happen at any time and for any reason, and users have very little control over these situations. In addition, centralized cloud servers are concentrated in one or several locations. In the event of an outage or other type of failure, a large number of related services are often paralyzed or even at risk of permanent data loss.

Second, the performance of a centralized cloud may be less than ideal. Centralized cloud servers are usually located in remote areas, resulting in delayed data transmission. Constant syncing of all users and their devices to the centralized cloud can also lead to increased demands for bandwidth, imposing undue burdens on network infrastructure.

Third, the cost of a centralized cloud is high. It requires an expensive data center that must be maintained and secured by skilled technical staff. It is also expensive to support data accessibility with duplication. The centralized cloud is very resource intensive, requiring multiple servers, load balancers and other facilities that must be meticulously managed and secured.

Fourth, the centralized cloud has serious security and privacy concerns. Cloud computing can be associated with data loss, data breaches and data unavailability. Data in the cloud may be accessed and exfiltrated, lawfully or illicitly, by the hosts, law enforcement officials or intruders. If a centralized server is compromised, user data may be compromised as well. As more legislation is passed related to data privacy and protection—such as the EU General Data Protection Regulation (GDPR) and the US Health Insurance Portability and Accountability Act (HIPAA)—complying with the various national, geographic, industry or service-specific legal and regulatory mandates is becoming more difficult and complex.

> "THE DECENTRALIZED CLOUD COMPUTING MODEL PROMISES TO SUPPORT SCALABLE APPLICATIONS WHILE RETAINING THE SAFEGUARDS OF A DECENTRALIZED, TRUST-MINIMIZED ECOSYSTEM."

A pragmatic solution to these shortcomings is to decentralize the cloud with artificial intelligence (AI) and blockchain. The decentralized cloud computing model promises to support scalable applications while retaining the safeguards of a decentralized, trust-minimized ecosystem.

Cloud Decentralization

Decentralization means that the system does not rely on centralized servers or data centers, so there can be no single point of failure. Decision-making is performed independently by all the participating nodes. The decentralized cloud also harnesses the power of edge computing by moving processes and storage to the device at the edge of the network. In an ideal world, the fully decentralized cloud is an architecture where every edge device can function as a cloud server. Edge devices have their own cloud functionality for remote access, sharing, streaming, collaboration and file management. They can process data locally, communicate with one another directly and share resources freely.

There are many advantages of a decentralized cloud. First, edge computing makes the decentralized cloud faster, more efficient and more scalable. Localization provides close physical proximity to the server, so the required computing resources are quickly accessible. For example, Ankr is a decentralized cloud solution that uses existing idle computing resources to drive the network.[3] Ankr's computing power is supplied by many different providers via their own enterprise-managed data centers. This provider decentralization is more efficient and more scalable.

Second, the decentralized cloud is more cost-efficient because it leverages unused computing resources. For example, Ankr is about half the cost of Amazon Web Services (AWS) for the same computing power.

Third, the decentralized cloud is more reliable. The redundancy across multiple nodes adds protection to ensure data integrity in case of an error in the storage or transmission of data and to prevent data loss. Again, using Ankr as an example, provider decentralization is capable of preventing a single point of failure.

Fourth, decentralization provides better security and privacy. Data are kept locally behind a firewall to protect their privacy and prevent exfiltration. Data are not duplicated to third-party servers or secondary locations. Thus, the risk of data compromise is reduced. In the decentralized cloud, end-to-end encryption is standard, making security much stronger. Each piece of data is only a small portion of the whole, so even if one piece of data is hacked, the whole is not compromised. In addition, each piece of data is encrypted locally before it is uploaded and spread out to uncorrelated nodes across the cloud. Only the user has access to the encryption keys, making it virtually impossible for data to be compromised or stolen.

> “IN THE DECENTRALIZED CLOUD, END-TO-END ENCRYPTION IS STANDARD, MAKING SECURITY MUCH STRONGER.”

There are many emerging decentralized cloud platforms, such as Ankr, Filecoin, MaidSafe, Siacoin, Storj, X Cloud, Dfinity and Sharder. Some of them are now available in the marketplace.

One of the underlying technologies behind the decentralized cloud is the InterPlanetary File System (IPFS).

The IPFS

The IPFS is an open-source project sponsored by Protocol Labs. On its official website, IPFS is defined as "a distributed system for storing and accessing files, websites, applications, and data."[4] IPFS aims to replace Hypertext Transfer Protocol (HTTP) to make a better and more efficient web. It is built on three fundamental principles:

1. Unique identification via content addressing
2. Content linking via directed acyclic graphs (DAGs)
3. Content discovery via distributed hash tables (DHTs)

Content addressing is used to identify content by what is in it rather than by where it is located. To link between content, IPFS uses Merkle DAGs, in which each node has a unique identifier that is a hash of the node's contents. A Merkle DAG in IPFS is optimized to represent directories and files. To find which peers are hosting the desired content, IPFS uses a DHT. A hash table is a database of keys to values, and a DHT is one in which the table is split across all the peers in a distributed network.

The various platforms differ in some areas but are, for the most part, similar. They are built on top of IPFS or use a variant of it.

How It Works: Storj

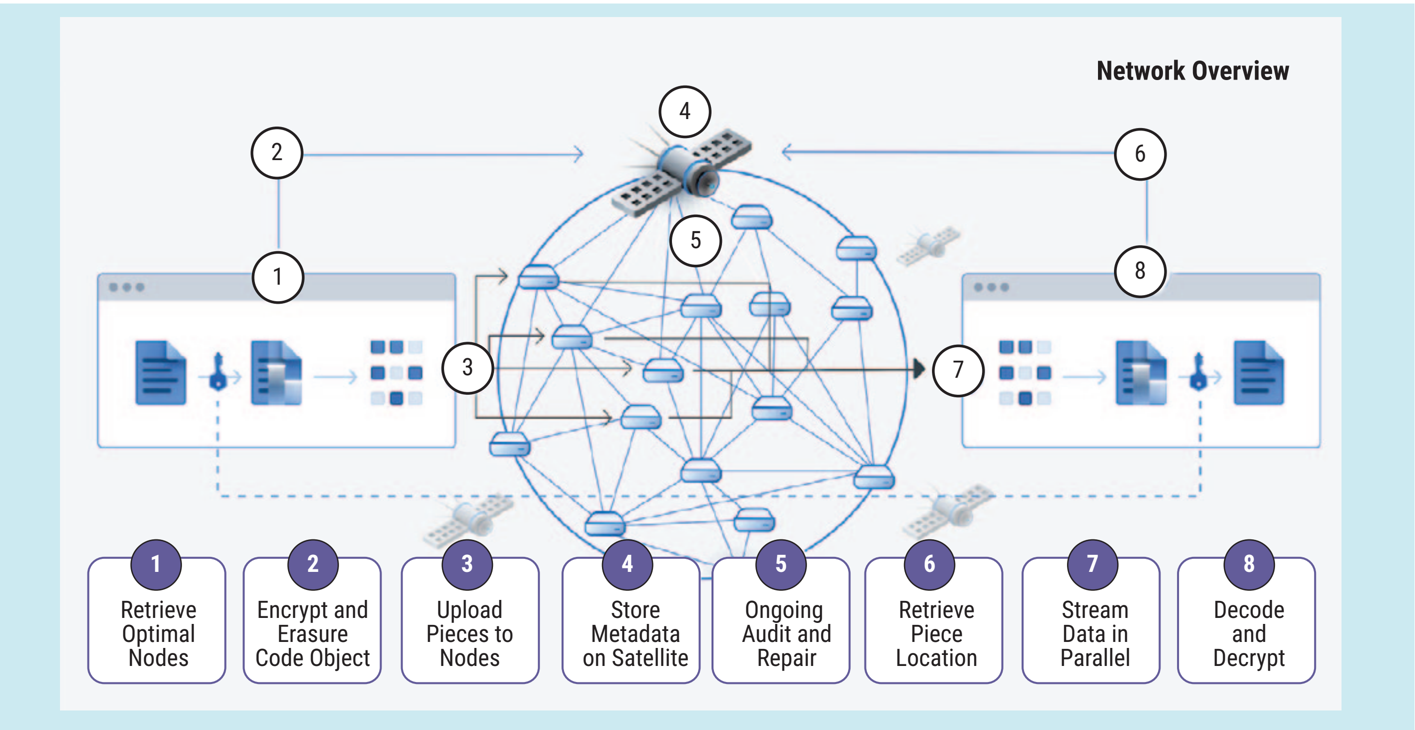
Sponsored by Storj Labs, Storj is an open-source Simple Storage Service (S3)–compatible platform and suite of decentralized applications that allow users to store data in a secure and decentralized manner.[5] According to Storj Labs, a separate brand called Tardigrade is used for the demand side of its business (primarily developers), whereas Storj focuses on the supply side of its business. Here, the term "Storj" refers to the decentralized cloud platform. As depicted in **figure 2**, the decentralized cloud is a large, distributed network composed of thousands of nodes across the globe that are independently owned and operated.[6] A node is simply a hard drive or a storage device that someone owns privately. The node owners store files on behalf of a provider (such as Storj) and are compensated for the usage.

The basic technical components of Storj are data storage, data retrieval and data management. When data are stored on the Storj network (**figure 3**), a user encrypts the data and breaks them up into multiple pieces. The pieces are distributed to peers across the network, and metadata are generated. When data are retrieved from the network, the distributed pieces are retrieved by referencing the metadata, and the original data are reconstructed on the user's local device. When the amount of redundancy drops below a certain threshold, the necessary data for the missing pieces are regenerated and replaced.[7]

There are four important aspects of the Storj network:

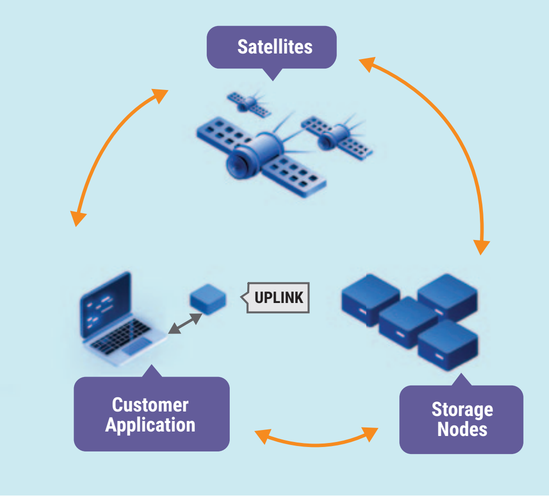
1. It is private and secure by default.
2. It is "trustless," meaning that users do not have to place their trust in any single enterprise, process or system to keep the network running.
3. Data are encrypted end to end. Advanced Encryption Standard (AES)-256 encryption occurs before data are uploaded to the network, ensuring that only the user can access data without authorization. Data on

Figure 2—Overview of Storj


Network Overview
1
2
3
4
5
6
7
8
1 Retrieve Optimal Nodes
2 Encrypt and Erasure Code Object
3 Upload Pieces to Nodes
4 Store Metadata on Satellite
5 Ongoing Audit and Repair
6 Retrieve Piece Location
7 Stream Data in Parallel
8 Decode and Decrypt

Source: Ford, P.; "What Is Decentralized Cloud Storage," Storj Labs, 16 October 2020, *https://storj.io/blog/2020/10/what-is-decentralized-cloud-storage*. Reprinted with permission.

Figure 3—How Storj Works



Source: Tardigrade, "Product Overview," *https://documentation.tardigrade.io/storage/considerations*. Reprinted with permission.

Storj are almost impossible to access without the proper keys or permissions. Because everything is encrypted locally, data are solely in the users' hands.

4. Data storage and retrieval are secure and efficient. After data are encrypted, they are split into smaller fragments that are completely indistinguishable from one another. Each of the pieces is at a different location with different operators, power supplies and networks. Splitting data yields tremendous security, performance and durability advantages. Because there are thousands of nodes around the world in the Storj network, it claims to be as fast as or even faster than centralized cloud services such as Amazon S3. A node operator does not know what data are stored and would have to track down all the other nodes to reconstitute the data and the keys to all the pieces.

Conclusion

The drawbacks of centralized cloud computing offer the rationale behind moving toward decentralization. Decentralized cloud computing is still considered an emerging technology, and some of the players (e.g., Filecoin) are still in the development phase. Even though all the systems in the decentralized cloud space claim better security and privacy, there are many obstacles to overcome before a fully decentralized cloud is achieved, where the dynamic ecosystems of "intelligent things" can operate in a fully distributed and decentralized manner. Nonetheless, both research and commercial endeavors to tackle the technological challenges of decentralized cloud computing continue.

> "EVEN THOUGH ALL THE SYSTEMS IN THE DECENTRALIZED CLOUD SPACE CLAIM BETTER SECURITY AND PRIVACY, THERE ARE MANY OBSTACLES TO OVERCOME BEFORE A FULLY DECENTRALIZED CLOUD IS ACHIEVED."

Endnotes

1 Mell, P.; T. Grance; *The NIST Definition of Cloud Computing*, Special Publication 800-145, National Institute of Standards and Technology (NIST), USA, September 2011, *https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf*

2 Google Cloud Status Dashboard, *https://status.cloud.google.com/incident/zall/20013*

3 Ankr, "Welcome to Ankr," *https://docs.ankr.com/*

4 InterPlanetary File System (IPFS), *https://ipfs.io*

5 Storj, "Introduction," *https://documentation.storj.io*

6 Ford, P.; "What Is Decentralized Cloud Storage," Storj Blog, 16 October 2020, *https://storj.io/blog/2020/10/what-is-decentralized-cloud-storage*

7 Tardigrade, "Product Overview," *https://documentation.tardigrade.io/storage/considerations*