Data storage services

Data storage is the retention of information using technology specifically developed to keep that data and have it as accessible as necessary. Data storage refers to the use of recording media to retain data using computers or other devices. The most prevalent forms of data storage are file storage, block storage, and object storage, with each being ideal for different purposes.

With data storage space, users can save data onto a device. And should the computer power down, the data is retained. And instead of manually entering data into a computer, users can instruct the computer to pull data from storage devices. Computers can read input data from various sources as needed, and it can then create and save the output to the same sources or other storage locations. Users can also share data storage with others.

Today, organizations and users require data storage to meet today's high-level computational needs like big data projects, artificial intelligence (AI), machine learning and the internet of things (IoT). And the other side of requiring huge data storage amounts is protecting against data loss due to disaster, failure or fraud. So, to avoid data loss, organizations can also employ data storage as backup solutions.

In simple terms, modern computers, or terminals, connect to storage devices either directly or through a network. Users instruct computers to access data from and store data to these storage devices. However, at a fundamental level, there are two foundations to data storage: the form in which data takes and the devices data is recorded and stored on.

Forms of data storage:

* File storage (Azure Files, AWS Elastic File System (EFS).)

File storage—also called file-level or file-based storage—is a hierarchical storage methodology used to organize and store data on a computer hard drive or on network-attached storage (NAS) device. In file storage, data is stored in files, the files are organized in folders, and the folders are organized under a hierarchy of directories and subdirectories. To locate a file, all you or your computer system need is the path—from directory to subdirectory to folder to file.

Hierarchical file storage works well with easily organized amounts of structured data. But, as the number of files grows, the file retrieval process can become cumbersome and time-consuming. Scaling requires adding more hardware devices or continually replacing these with higher-capacity devices, both of which can get expensive.

File storage benefits

If your organization requires a centralized, easily accessible, and affordable way to store files and folders, file-level storage is a good approach. The benefits of file storage include the following:

* **Simplicity:** File storage is the simplest, most familiar, and most straightforward approach to organizing files and folder on a computer’s hard drive or NAS device. You simply name files, tag them with metadata, and store them in folders under a hierarchy of directories and subdirectories. It is not necessary to write applications or code to access your data.
* **File sharing:** File storage is ideal for centralizing and sharing files on a Local Area Network (LAN). Files stored on a NAS device are easily accessible by any computer on the [network](https://www.ibm.com/topics/networking) that has the appropriate permission rights.
* **Common protocols:** File storage uses common file-level protocols such as Server Message Block (SMB), Common Internet File System (CIFS), or Network File System (NFS). If you utilize a Windows or Linux operating system (or both), standard protocols like SMB/CIFS and NFS will allow you to read and write files to a Windows-based or Linux-based server over your Local Area Network (LAN).
* **Data protection:** Storing files on a separate, LAN-connected storage device offers you a level of data protection should your network computer experience a failure. Cloud-based file storage services provide additional data protection and [disaster recovery](https://www.ibm.com/topics/backup-disaster-recovery) by replicating data files across multiple, geographically-dispersed data centers.
* **Affordability:** File storage using a NAS device allows you to move files off of expensive computing hardware and onto a more affordable LAN-connected storage device. Moreover, if you choose to subscribe to a cloud file-storage service, you eliminate the expense of on-site hardware upgrades and the associated ongoing maintenance and operation costs.

File storage use cases

File storage is a good solution for a wide variety of data needs, including the following:

* **Local file sharing:** If your data storage needs are generally consistent and straightforward, such as storing and sharing files with team members in the office, consider the simplicity of file-level storage.
* **Centralized file collaboration:**If you upload, store, and share files in a centralized library—located on-site, off-site, or in the cloud—you can easily collaborate on files with internal and external users or with invited guests outside of your network.
* **Archiving/storage:**You can cost-effectively archive files on NAS devices in a small data center environment or subscribe to a cloud-based file storage service to store and archive your data.
* **Backup/disaster recovery:**You can store backups securely on separate, LAN-connected storage devices. Or you can subscribe to a cloud-based file storage service to replicate your data files across multiple, geographically-dispersed data centers and gain the additional data protection of distance and redundancy.

Storing file data in the cloud also enables you to scale up capacity as needed and on demand. Cloud-based file storage services typically offer simple, pre-defined tiers with varying levels of storage capacity and workload performance requirements (total number of input/output operations per second, or IOPS), as well as data protection and replication to other data centers for business continuity—all for a predictable monthly fee. Or you can increase or decrease IOPS and expand data volumes dynamically, paying only for what you use.

* Block storage: (AWS Elastic Block Store (EBS), Azure Disk Storage, and Google Persistent Disk.)

Block storage, sometimes referred to as block-level storage, is a technology that is used to store data files on storage area networks (SANs) or cloud-based storage environments. Developers favor block storage for computing situations where they require fast, efficient, and reliable data transportation.

Block storage breaks up data into blocks and then stores those blocks as separate pieces, each with a unique identifier. The SAN places those blocks of data wherever it is most efficient. That means it can store those blocks across different systems and each block can be configured (or partitioned) to work with different operating systems.

Block storage also decouples data from user environments, allowing that data to be spread across multiple environments. This creates multiple paths to the data and allows the user to retrieve it quickly. When a user or application requests data from a block storage system, the underlying storage system reassembles the data blocks and presents the data to the user or application.

Block storage allows for the creation of raw storage volumes, which server-based operating systems can connect to. You can treat those raw volumes as individual hard drives. This lets you use block storage for almost any kind of application, including file storage, database storage, virtual machine file system (VMFS) volumes, and more.

Take, for example, the deployment of virtual machines across an enterprise. With block storage, you can easily create and format a block-based storage volume to store the VMFS. A physical server can then attach to that block, creating multiple virtual machines. What’s more, creating a block-based volume, installing an operating system, and attaching to that volume allows users to share files using that native operating system.

Private cloud deployments are another excellent use of block storage. For a deeper dive into private clouds and block storage, check out the IBM Garage’s explanation of virtualization for extending virtualized private cloud using block and file storage.

* Object data storage: (AWS S3, Azure Blob Storage, and Google Cloud Storage.)

Object storage, often referred to as object-based storage, is a data storage architecture for handling large amounts of unstructured data. This is data that does not conform to, or cannot be organized easily into, a traditional relational database with rows and columns. Today’s Internet communications data is largely unstructured. This includes email, videos, photos, web pages, audio files, sensor data, and other types of media and web content (textual or non-textual). This content streams continuously from social media, search engines, mobile, and “smart” devices.

Objects are discrete units of data that are stored in a structurally flat data environment. There are no folders, directories, or complex hierarchies as in a file-based system. Each object is a simple, self-contained repository that includes the data, metadata (descriptive information associated with an object), and a unique identifying ID number (instead of a file name and file path). This information enables an application to locate and access the object. You can aggregate object storage devices into larger storage pools and distribute these storage pools across locations. This allows for unlimited scale, as well as improved data resiliency and disaster recovery.

Objects (data) in an object-storage system are accessed via Application Programming Interfaces (APIs). The native API for object storage is an HTTP-based RESTful API (also known as a RESTful Web service). These APIs query an object’s metadata to locate the desired object (data) via the Internet from anywhere, on any device. RESTful APIs use HTTP commands like “PUT” or “POST” to upload an object, “GET” to retrieve an object, and “DELETE” to remove it. (HTTP stands for Hypertext Transfer Protocol and is the set of rules for transferring text, graphic images, sound, video, and other multimedia files on the Internet).

Beneﬁts

There are many reasons to consider an object-storage-based solution to store your data, particularly in this era of the Internet and digital communications that is producing large volumes of web-based, multimedia data at an increasing rate.

Storing/managing unstructured data

Object storage is seeing wide adoption in the era of [cloud computing](https://www.ibm.com/topics/cloud-computing) and for the management of unstructured data which analysts estimate will represent the vast majority of all data worldwide in the near future. The volume of web-generated content—emails, videos, social media, documents, sensor data produced by the Internet of Things (IoT) devices, and more—is massive and growing. Unstructured data is typically static (unchanging) but may be required at any time, anywhere (like images and video files, for example, or archived data backups).

Cloud-based object storage is ideal for long-term data retention. Use object storage to replace traditional archives, such as Network Attached Storage (NAS), reducing your IT infrastructure. Easily archive and store mandated, regulatory data that must be retained for extended periods of time. Cost-effectively preserve large amounts of rich media content (images, videos, etc.) that is not frequently accessed.

Scalability

Unlimited scale is perhaps the most significant advantage of object-based data storage. Objects, or discrete units of data (in any quantity), are stored in a structurally flat data environment, within a storage device such as a server. You simply add more devices/servers in parallel to an object storage cluster for additional processing and to support the higher throughputs required by large files such as videos or images.

Reduced complexity

Object storage removes the complexity that comes with a hierarchical file system with folders and directories. There is less potential for performance delay and you will realize efficiencies when retrieving data since there are no folders, directories or complex hierarchies to navigate. This improves performance, particularly when managing very large quantities of data.

Disaster recovery/availability

You can configure object storage systems so that they replicate content. If a disk within a cluster fails, a duplicate disk is available, ensuring that the system continues running with no interruption or performance degradation. Data can be replicated within nodes and clusters and among distributed data centers for additional back-up off-site and even across geographical regions.

Object storage is a more efficient alternative to tape backup solutions, which require tapes that need to be physically loaded into and removed from tape drives and moved off-site for geographic redundancy. You can use object storage to automatically back up on-premises databases to the cloud and/or to cost-effectively replicate data among distributed data centers. Add additional back-up off-site and even across geographical regions to ensure disaster recovery.

For a deeper dive on disaster recovery, check out "[Backup and Disaster Recovery: A Complete Guide](https://www.ibm.com/topics/backup-disaster-recovery)."

Customizable metadata

Remember that each object is a self-contained repository that includes metadata or descriptive information associated with it. Objects use this metadata for important functions such as policies for retention, deletion and routing, disaster recovery strategies (data protection), or validating content authenticity. You can also customize the metadata with additional context that can be later extracted and leveraged to perform business insights and analytics around customer service or market trends, for example.

Affordability

Object storage services use pay-as-you-go pricing that incurs no upfront costs or capital investment. You simply pay a monthly subscription fee for a specified amount of storage capacity, data retrieval, bandwidth usage, and API transactions. Pricing is usually tiered-based or volume-based, which means that you will pay less for very large volumes of data.

Additional cost savings come from the use of commodity server hardware since object storage solutions have limited hardware constraints and can be deployed on most properly configured commodity servers. This limits the need to purchase new hardware when deploying an object storage platform on-premises. You can even use hardware from multiple vendors.

Cloud compatibility

Object storage goes hand in hand with cloud or hosted environments that deliver multi-tenant storage as a service. This allows many companies or departments within a company to share the same storage repository, with each having access to a separate portion of the storage space. This shared storage approach inherently optimizes scale and costs. You will reduce your organization’s on-site IT infrastructure by using low-cost [cloud storage](https://www.ibm.com/topics/cloud-storage) while keeping your data accessible when needed. Your enterprise, for example, can use a cloud-based object storage solution to collect and store large amounts of unstructured IoT and mobile data for your smart device applications.

Cloud compliance

Cloud compliance is the art and science of complying with regulatory standards of cloud usage in accordance with industry guidelines and local, national, and international laws. Some common regulatory requirements include the Health Insurance Portability and Accountability Act (HIPAA), Payment Card Industry Data Security Standard (PCI DSS), and Gramm-Leach-Bliley Act (GLBA).

**Components of cloud compliance**

Cloud compliance requirements will vary depending on your industry and the regulations that guide your business.

We dig into the components that shape general cloud compliance below.

**Standards**

Certain industries outline specific instructions for properly handling data within the cloud. These are known as cloud security compliance standards.

For example, ISO includes cloud-specific security controls within ISO 27017 and ISO 27018. That means implementing specific security controls regarding the configuration of your cloud environment.

HIPAA also specifies that a covered entity and their cloud service provider (CSP) must enter into a business associate agreement where the CSP will be held liable for compliance with HIPAA Rules.

**Laws and regulations**

Laws and regulations — at the global, national, and state levels — also help shape cloud compliance requirements.

It’s important to understand your country’s laws and regulations for cloud compliance, data privacy, data protection and localization, and cybersecurity.

A few common regulations include HIPAA, PCI DSS, and SOX.

**Governance**

[Cloud governance controls](https://www.forbes.com/sites/forbestechcouncil/2020/01/17/cloud-governance-best-practices-how-to-create-a-framework-for-success/?sh=6c2d34448ce6) help manage a company’s data within the cloud and provide clear security policies on how to use (and how not to use) the cloud.

Companies should have guidelines on organizing, sharing, and tracking information on the cloud and expanding cloud usage. These should also cover ownership and responsibility of cloud strategy.

**Why is cloud compliance important?**

As of 2022,[over 60 percent of all corporate data](https://www.statista.com/statistics/1062879/worldwide-cloud-storage-of-corporate-data/) was stored in the cloud — double the amount that was stored in the cloud in 2015.

Organizations are continuing to move more data and workloads into the cloud, according to the[2023 Thales Global Cloud Security Study](https://www.techrepublic.com/article/thales-2023-cloud-security-report/). This year, 27% of organizations reported that 60% or more of their workloads are in the cloud, an increase from 23% of organizations in 2022.

With more data moving to the cloud, a business must understand its own role and responsibility for keeping that data safe, including achieving and maintaining compliance with cloud requirements. This is essential for not only building customer trust, but also for avoiding costly data breaches.

Cloud Availability

1. \*\*Geographic Redundancy\*\*: Imagine your cloud application is hosted across multiple data centers in different regions offered by your cloud provider, ensuring that even if one region experiences an outage, users can still access your application from another region.

2. \*\*Load Balancing as Traffic Manager\*\*: Think of load balancers as traffic managers in the cloud, distributing incoming requests across multiple instances or servers to prevent any single instance from becoming overwhelmed and ensuring consistent access to your application.

3. \*\*Auto-scaling for Elasticity\*\*: Picture your application automatically scaling up or down based on demand, adding more resources during peak times and reducing them during quieter periods, all without manual intervention, thus maintaining availability regardless of fluctuations in traffic.

4. \*\*Redundancy and Failover\*\*: Consider deploying redundant components of your application across multiple availability zones within a cloud region, so that if one zone experiences issues, traffic is automatically rerouted to healthy zones, minimizing downtime.

5. \*\*Real-time Monitoring and Alerts\*\*: Think of monitoring tools constantly watching over your cloud infrastructure, instantly notifying you via alerts if any performance issues arise, allowing you to promptly address them before they impact availability.

6. \*\*Data Redundancy and Backup\*\*: Imagine your data being replicated across multiple storage locations in the cloud, with regular backups ensuring that even in the event of a failure, your data remains intact and accessible.

7. \*\*High Availability Database Solutions\*\*: Think of database services offered by cloud providers with built-in redundancy and automatic failover mechanisms, ensuring continuous access to your data even if a database instance fails.

8. \*\*Network Redundancy for Seamless Connectivity\*\*: Consider redundant network configurations in the cloud, where if one network path fails, traffic automatically reroutes through alternate paths, ensuring uninterrupted connectivity to your application.

9. \*\*Disaster Recovery in the Cloud\*\*: Picture a comprehensive disaster recovery plan tailored for the cloud, with data replication to a secondary region and automated procedures for quickly restoring services in case of a major outage.

10. \*\*Continuous Optimization for Reliability\*\*: Think of regularly optimizing your cloud infrastructure and configurations to enhance reliability, ensuring that your application remains available and responsive to users' needs.

In summary, ensuring availability in the cloud involves leveraging various cloud services and features to build a resilient and highly available architecture that can withstand failures and provide uninterrupted service to users.

Cloud responsibilities

Cloud providers offer a range of security features and services to help customers protect their data, applications, and infrastructure in the cloud. Some of the key security offerings provided by cloud providers include:

1.\*\*Identity and Access Management (IAM)\*\*: Cloud providers offer IAM services to manage user identities, roles, and permissions, allowing customers to control access to their resources and data. IAM enables features like multi-factor authentication (MFA), role-based access control (RBAC), and identity federation.

### Sample IAM Scenarios:

1. \*\*Creating Users and Groups\*\*:

\*Scenario\*: A company wants to grant access to its AWS resources to its employees based on their departments.

\*Solution\*: The company creates IAM users for each employee and groups them based on their departments (e.g., "Engineering," "Marketing," "Finance"). Permissions are assigned to these groups to control access to relevant resources.

2. \*\*Assigning Roles for Temporary Access\*\*:

\*Scenario\*: An external contractor needs temporary access to manage resources within an organization's AWS account.

\*Solution\*: Instead of creating a permanent user account, the organization creates an IAM role with the necessary permissions. The contractor assumes this role when performing tasks, and access automatically expires after a specified period.

3. \*\*Implementing Least Privilege Access\*\*:

\*Scenario\*: A company wants to ensure that employees have access only to the resources they need to perform their jobs, following the principle of least privilege.

\*Solution\*: IAM policies are carefully crafted to grant only the minimum permissions required for each user's or role's tasks. For example, developers may have access to EC2 instances for development purposes but not to production databases.

4. \*\*Enforcing Multi-Factor Authentication (MFA)\*\*:

\*Scenario\*: An organization wants to enhance security by requiring users to authenticate using multiple factors, such as a password and a temporary authentication code.

\*Solution\*: IAM policies are configured to enforce MFA for users accessing sensitive resources or performing privileged actions. Users are required to provide a second form of authentication, such as a code from a mobile app or a hardware token, in addition to their password.

5. \*\*Implementing Identity Federation\*\*:

\*Scenario\*: An organization wants to allow its employees to access cloud resources using their existing corporate credentials without creating separate IAM accounts.

\*Solution\*: The organization sets up identity federation using services like AWS Single Sign-On (SSO) or AWS Identity and Access Management (IAM) roles for federated users. Employees authenticate using their corporate credentials, and temporary IAM credentials are generated to access AWS resources.

6. \*\*Monitoring and Auditing Access\*\*:

\*Scenario\*: An organization wants to track and review user activity to identify unauthorized access or suspicious behavior.

\*Solution\*: IAM access logs are enabled to record all IAM actions performed in the AWS account. The organization uses AWS CloudTrail to centrally collect and analyze these logs, allowing them to monitor user activity, investigate security incidents, and maintain compliance with regulatory requirements.

By effectively implementing IAM, organizations can manage access to their cloud resources securely, enforce security policies, and protect sensitive data from unauthorized access or misuse.

2. \*\*Encryption\*\*: Cloud providers support encryption at rest and in transit to protect data stored in their infrastructure and transmitted over networks. They offer services such as Key Management Systems (KMS) for managing encryption keys and Hardware Security Modules (HSMs) for enhanced security.

3. \*\*Network Security\*\*: Cloud providers offer features like virtual private clouds (VPCs), network access control lists (ACLs), and security groups to create isolated network environments and control traffic flow. They also provide Distributed Denial of Service (DDoS) protection to mitigate and prevent attacks targeting network infrastructure.

A Virtual Private Cloud (VPC) is a virtual network environment that closely resembles a traditional on-premises network. Cloud providers allow customers to create VPCs to logically isolate their resources within the cloud. Key features of VPCs include:

* Isolation: VPCs provide logical isolation for resources, allowing customers to create private networks within the cloud environment.
* Customization: Customers can customize network configurations, including IP address ranges, subnets, route tables, and network gateways, to meet their specific requirements.
* Security: VPCs include features such as network access control lists (ACLs) and security groups to control inbound and outbound traffic, restricting access to resources based on predefined rules.

4. \*\*Security Monitoring and Logging\*\*: Cloud providers offer monitoring and logging services to track and analyze activities, events, and security-related incidents within their environments. They provide tools for real-time monitoring, log aggregation, and analysis to help customers detect and respond to security threats.

5. \*\*Vulnerability Management\*\*: Cloud providers offer vulnerability scanning and assessment services to identify security vulnerabilities and weaknesses in customer deployments. They provide tools for automated scanning, patch management, and remediation to help customers address security issues proactively.

6. \*\*Compliance and Governance\*\*: Cloud providers offer compliance certifications and frameworks to demonstrate adherence to industry standards and regulations related to data privacy, security, and governance. They provide compliance tools, reports, and audit logs to help customers maintain regulatory compliance in the cloud.

7. \*\*Security Analytics and Threat Intelligence\*\*: Cloud providers offer security analytics and threat intelligence services to analyze security data, identify patterns, and detect anomalous behavior indicative of security threats. They provide threat detection, incident response, and forensics capabilities to help customers investigate and mitigate security incidents.

8. \*\*Application Security\*\*: Cloud providers offer application security services to help customers secure their applications throughout the software development lifecycle. They provide tools for code scanning, vulnerability assessment, and runtime protection to help customers identify and mitigate security risks in their applications.

9. \*\*Data Loss Prevention (DLP)\*\*: Cloud providers offer DLP services to help customers prevent the unauthorized disclosure of sensitive data. They provide features like data classification, encryption, and access controls to help customers protect their data from loss, leakage, or theft.

10. \*\*Security Consulting and Training\*\*: Cloud providers offer security consulting and training services to help customers design, implement, and maintain secure cloud deployments. They provide guidance, best practices, and training resources to help customers build and operate secure environments in the cloud.

Overall, cloud providers offer a comprehensive set of security features and services to help customers protect their assets and mitigate security risks in the cloud. However, it's important for customers to understand their shared responsibility model and take appropriate measures to secure their own applications and data deployed in the cloud.