



**Campus Monterrey**

**Materia**

Inteligencia artificial avanzada para la ciencia de datos II

**Tarea**

Actividad 1 Cloud Storage

**Estudiantes**

Cleber Gerardo Pérez Galicia - A01236390

Juan Pablo Bernal Lafarga - A01742342

Jacobo Hirsch Rodríguez - A00829679

Eryk Elizondo González - A01284899

**Profesor**

EdgFélix Ricardo Botello Urrutia

# Azure File Storage

Es un servicio de almacenamiento que permite a las aplicaciones y usuarios compartir archivos a través de la red usando el protocolo SMB (Server Message Block), el cual es ampliamente utilizado para compartir archivos en entornos de red. Además, también es compatible con el protocolo NFS (Network File System) para aplicaciones que requieren este tipo de sistema de archivos.

## Pricing

Los costos de este servicio dependen de múltiples factores, como es un sistema flexible realmente se cobra por el uso de los recursos tanto como de entrada como de salida. De igual manera estos servicios tienen a su vez parámetros que dependiendo de su configuración el precio cambia, aquí se muestra más a detalle como funciona:

### 1. Capacidad de almacenamiento:

Pagas por la cantidad de datos que almacenas en Azure Files. El costo se mide en **GiB** almacenados por mes. Existen diferentes **niveles de almacenamiento** que afectan el precio:

- **Nivel de transacciones frecuentes (Premium):** Para aplicaciones que requieren acceso rápido y frecuentes operaciones de lectura/escritura.
- **Nivel de uso estándar (Estándar - Hot, Cool, Archive):**
  - **Hot (caliente):** Para datos que se acceden con frecuencia.
  - **Cool (frío):** Para datos que se acceden de manera poco frecuente pero deben permanecer disponibles.
  - **Archive (archivo):** Para datos a los que raramente se accede, pero que deben mantenerse durante un largo periodo.

### 2. Operaciones de transacción (Transacciones IOPS):

Azure Files cobra por las operaciones de lectura, escritura y otros tipos de transacciones de archivo, como la creación, eliminación, enumeración de archivos, etc. La cantidad y tipo de transacciones influye en el costo:

- **Lecturas/escrituras de archivos.**
- **Operaciones de metadatos** (como listar o modificar atributos de archivos).

### 3. Costo de transferencia de datos:

- **Entrada de datos (ingreso):** Normalmente, no hay cargos por los datos que se cargan o ingresan a Azure desde fuera (desde tu red local hacia la nube de Azure).
- **Salida de datos (egreso):** Se cobran los datos que se descargan o transfieren desde Azure Files hacia un destino fuera de Azure, como tu red local. Esto se llama "egreso de datos" y tiene un costo asociado dependiendo de la cantidad de datos transferidos.

### 4. Replicación de datos:

Azure Files ofrece diferentes opciones de replicación para aumentar la durabilidad de los datos, lo que también impacta en el costo:

- **Locally-Redundant Storage (LRS):** Los datos se replican dentro de una sola región.
- **Zone-Redundant Storage (ZRS):** Los datos se replican en varias zonas dentro de una región.
- **Geo-Redundant Storage (GRS):** Los datos se replican entre dos regiones distantes geográficamente.
- **Read-Access Geo-Redundant Storage (RA-GRS):** Es similar a GRS pero ofrece acceso de solo lectura a los datos replicados en la segunda región.

### 5. Snapshots (instantáneas):

Si deseas tomar **instantáneas** de tus archivos o directorios para tener puntos de restauración o copias de seguridad, esto también genera costos adicionales basados en el tamaño de las instantáneas.

### 6. Almacenamiento reservado:

Azure también ofrece la opción de comprar **reservas de capacidad** para obtener descuentos si te comprometes a un cierto uso de almacenamiento a largo plazo (1 o 3 años). Esto es útil para organizaciones que tienen un uso predecible.

## 7. Acceso premium:

Para almacenamiento de archivos de alta velocidad, existe la opción de **Azure Files Premium**, que ofrece un rendimiento más alto y latencias menores, pero con un costo más elevado. Aquí, pagas también por la cantidad de aprovisionamiento de almacenamiento reservado (un número de GiB fijos).

Si se quiere saber más información sobre los precios dependiendo de las configuraciones, necesidades de consumo de datos, tráfico de red etc, se puede acceder al siguiente link para ver dichos costos:

<https://azure.microsoft.com/es-es/pricing/details/storage/files/>

## Storage

Las capacidades de almacenamiento se pueden dividir en dos grupos generales, que son los **Standard** y **Premium**, ambos tienen un file share máximo de 100 TiB, estos son los archivos que tienen uso compartido entre diferentes usuarios o aplicaciones. Cada file share puede tener un archivo de un tamaño máximo de 4 TiB, esto significa que ningún archivo file share puede superar este tamaño. La diferencia principal entre el Standard y el Premium es que el segundo ofrece una opción de redundancia llamada “almacenamiento localmente redundante” por sus siglas LRS, esto significa que si se borran por cualquier razón los datos se tiene un respaldo de los mismos, también aumenta la cantidad de operaciones de entrada y salida por segundo IOPS, pudiendo llegar en el paquete premium hasta 100,000 IOPS.

## Availability

Para los paquetes premium existen estas funciones que pueden ser de mucha utilidad, todos son elementos redundantes para la prevención de desastres o fallas en el sistema.

**Locally Redundant Storage (LRS):** Los datos se replican tres veces dentro de una sola instalación física en una región de Azure. Es la opción más económica, pero menos resistente a fallos de infraestructura a gran escala.

**Zone-Redundant Storage (ZRS):** Los datos se replican en varias zonas de disponibilidad dentro de la misma región, proporcionando una mayor disponibilidad en caso de fallos de zona.

**Geo-Redundant Storage (GRS):** Los datos se replican entre dos regiones geográficamente separadas, lo que asegura que en caso de que una región falle, los datos puedan recuperarse desde la otra región.

**Read-Access Geo-Redundant Storage (RA-GRS):** Similar a GRS, pero además permite acceder a los datos en modo de solo lectura desde la región secundaria en cualquier momento.

De acuerdo al acuerdo de nivel de servicio de azure files, microsoft ofrece una garantía de nivel de servicio para los paquetes Premium y Standard ZRS de 99.99% de disponibilidad. Los servicios de LRS y GRS: También proporcionan 99.9% de disponibilidad, lo que implica un tiempo máximo de inactividad mensual muy limitado.

## Security Policies

Las políticas de seguridad en Azure Files están diseñadas para proteger los datos almacenados y garantizar un acceso controlado y seguro. A continuación, un resumen de las principales políticas de seguridad:

1. Cifrado:
  - En reposo: Todos los datos se cifran con AES-256 automáticamente.
  - En tránsito: El cifrado se realiza utilizando el protocolo SMB 3.0 para proteger los datos mientras se transfieren.
2. Autenticación y control de acceso:
  - Soporte para Azure Active Directory (Azure AD) y Active Directory Domain Services (AD DS) para gestionar el acceso y la autenticación de usuarios de forma segura.
  - Control de acceso basado en roles (RBAC) para asignar permisos específicos a los usuarios según sus roles dentro de la organización.

3. Gestión de claves:
  - Opciones de Bring Your Own Key (BYOK) para que las organizaciones gestionen sus propias claves de cifrado a través de Azure Key Vault.
  - Microsoft también gestiona automáticamente las claves para los usuarios que no usan BYOK.
4. Supervisión y auditoría:
  - Azure Security Center permite monitorear actividades sospechosas, generar alertas de seguridad y proporcionar recomendaciones para mejorar la seguridad.
  - Las auditorías y diagnósticos permiten rastrear y registrar accesos a los archivos para garantizar el cumplimiento de normativas como GDPR y HIPAA.
5. Backups y recuperación:
  - Azure ofrece opciones de respaldo y recuperación ante desastres mediante Azure Backup, lo que asegura la protección de los datos.

## Scalability

Ofrece escalabilidad automática tanto en capacidad como en rendimiento, permitiendo que los file shares crezcan hasta 100 TiB sin preaprovisionamiento y ajustando dinámicamente el rendimiento (IOPS) según la demanda. Esto incluye la capacidad de "bursting" para manejar picos de actividad y opciones avanzadas de replicación geográfica con **Geo-Redundant Storage (GRS)**. Además, la integración con **Azure File Sync** facilita la gestión de grandes volúmenes de datos distribuidos, lo que garantiza tanto escalabilidad horizontal como resiliencia y disponibilidad a nivel global

## Amazon S3

Es un servicio de almacenamiento en la nube proporcionado por **Amazon Web Services (AWS)** que permite almacenar y recuperar grandes cantidades de datos a través de Internet. Amazon S3 es altamente escalable, duradero y seguro, diseñado para almacenar cualquier tipo de dato, desde pequeños archivos hasta grandes volúmenes de datos, como backups, archivos multimedia o bases de datos.

## Pricing

De la misma forma que con azure files, los precios dependen mucho de la cantidad de recursos que se utilicen, el tráfico de la red a cantidad de datos almacenados, el tipo de clase de almacenamiento, y las operaciones realizadas (lecturas, escrituras, etc.).

Existen 9 clases de almacenamiento, que se utilizan de acuerdo a las necesidades del clientes, dichas clases son las siguientes: S3 Standard, S3 Intelligent-Tiering, S3

Standard-Infrequent Access, S3 One Zone-Infrequent Access, S3 Express One Zone, S3 Glacier Instant Retrieval, S3 Glacier Flexible Retrieval (antes S3 Glacier) y S3 Glacier Deep Archive.

Amazon S3 no cobra generalmente por el input de datos sino por el output, adicionalmente tambien cobra por las operaciones CRUD correspondientes a la modificación, eliminación y lectura de los datos.

Si se quiere saber más información sobre los precios dependiendo de las configuraciones, necesidades de consumo de datos, trafico de red etc, se puede acceder al siguiente link para ver dichos costos:

<https://aws.amazon.com/es/s3/pricing/>

## Storage

Las opciones de almacenamiento se mencionaron en el punto anterior pero aqui hay una descripción mas detallada de cada una con sus usos y ventajas:

### 1. S3 Standard

- Diseñada para datos que se acceden frecuentemente, con baja latencia y alta disponibilidad.
- **Caso de uso:** Aplicaciones web, contenido dinámico y datos críticos.

### 2. S3 Intelligent-Tiering

- Optimiza automáticamente los costos moviendo datos entre capas de acceso frecuente e infrecuente.
- **Caso de uso:** Datos con patrones de acceso impredecibles.

### 3. S3 Standard-IA (Infrequent Access)

- Para datos que se acceden esporádicamente, pero necesitan estar disponibles rápidamente cuando se requieran.
- **Caso de uso:** Backups, recuperación ante desastres.

### 4. S3 One Zone-IA

- Similar a Standard-IA, pero almacena datos en una única zona de disponibilidad, lo que la hace más económica.
- **Caso de uso:** Datos recreables fácilmente, como backups secundarios.

### 5. S3 Glacier

- Almacenamiento de archivo a largo plazo, con tiempos de recuperación que van de minutos a horas.
- **Caso de uso:** Archivos históricos o de cumplimiento que se acceden rara vez.

## 6. S3 Glacier Deep Archive

- La opción más económica, diseñada para datos que se acceden muy raramente, con tiempos de recuperación de hasta 12 horas.
- **Caso de uso:** Almacenamiento a largo plazo para datos que requieren retención prolongada por razones legales o regulatorias.

## 7. S3 Outposts

- Almacena datos localmente en las instalaciones del cliente mediante AWS Outposts, manteniendo la misma experiencia que S3 en la nube.
- **Caso de uso:** Aplicaciones con baja latencia o con requisitos de residencia de datos específicos.

## 8. S3 Reduced Redundancy Storage (RRS)

- Diseñada para datos que se acceden con frecuencia y que no requieren la misma durabilidad que S3 Standard.
- **Caso de uso:** Datos temporales o replicados, donde la pérdida de algunos archivos es aceptable.

## 9. S3 Glacier Instant Retrieval

- Proporciona la capacidad de acceder instantáneamente a los datos almacenados en el nivel de archivo, con un costo de almacenamiento muy bajo.
- **Caso de uso:** Para datos de archivo que necesitan recuperarse con rapidez pero rara vez se acceden.

## Availability

Ofrece una alta disponibilidad y durabilidad. Los datos almacenados en S3 se replican automáticamente en múltiples dispositivos en al menos tres zonas de disponibilidad físicamente separadas dentro de una región de AWS. Amazon garantiza una disponibilidad del 99.99% y una durabilidad del 99.999999999% (11 nueves) para los datos de S3, lo que significa que los datos son extremadamente seguros contra pérdidas.



## Security Policies

1. **Control de acceso:** S3 permite controlar el acceso a los datos usando políticas basadas en AWS Identity and Access Management (IAM) que se pueden asignar a usuarios, grupos o roles específicos. También se pueden establecer políticas a nivel de bucket o a nivel de objeto individual.
2. **Cifrado de datos:** S3 ofrece opciones de cifrado tanto en tránsito como en reposo. El cifrado en tránsito se realiza utilizando SSL/TLS, mientras que para el cifrado en reposo puedes elegir entre el cifrado gestionado por S3 (SSE-S3), el cifrado con claves gestionadas por el cliente en AWS Key Management Service (SSE-KMS) o el cifrado con claves gestionadas por el cliente fuera de AWS (SSE-C).
3. **Bloqueo de objetos:** Amazon S3 ofrece características de bloqueo de objetos que protegen tus datos de ser borrados o sobrescritos durante un período de tiempo definido o indefinidamente.

## Scalability

Amazon S3 es altamente escalable, lo que permite almacenar y recuperar cualquier cantidad de datos, desde pequeños bytes hasta exabytes de datos. No hay límites en la cantidad de datos que se pueden almacenar en S3 y puedes acceder a tus datos desde cualquier lugar a través de Internet. El servicio está diseñado para soportar altas tasas de transferencia y el número de operaciones por segundo puede escalar automáticamente dependiendo de la demanda del sistema.

Además, con la reciente introducción de Amazon S3 Multi-Region Access Points, los datos en S3 pueden ser aún más accesibles y resilientes, proporcionando mejoras en la latencia y la carga del tráfico al acceder a datos almacenados en múltiples regiones geográficas.

# Diferencias y similitudes entre proveedores

## Similitudes

1. **Protocolos de Transferencia:** Ambos servicios permiten el almacenamiento y recuperación de archivos a través de protocolos estándares de red. Azure File Storage utiliza SMB y NFS, mientras que Amazon S3 opera a través de HTTP/S.
2. **Escalabilidad:** Ambos servicios ofrecen soluciones altamente escalables, permitiendo el almacenamiento desde pequeños hasta enormes volúmenes de datos.
3. **Seguridad:** Tanto Azure como Amazon proporcionan robustas políticas de seguridad, incluyendo cifrado de datos en tránsito y en reposo, así como opciones avanzadas de control de acceso y gestión de claves.
4. **Opciones de Replicación:** Ofrecen varias opciones de replicación para aumentar la durabilidad y disponibilidad de los datos, incluyendo replicación dentro de una región y entre múltiples regiones.

## Diferencias

1. **Modelos de Precios y Uso:**
  - **Azure File Storage:** Los costos varían según la capacidad de almacenamiento utilizada, las operaciones de transacción, la transferencia de datos y las opciones de replicación. Azure también ofrece acceso premium y almacenamiento reservado.
  - **Amazon S3:** El pricing se basa en la clase de almacenamiento seleccionada, operaciones CRUD, y la transferencia de datos, con una amplia variedad de clases de almacenamiento que se adaptan a diferentes necesidades de accesibilidad y costo.
2. **Clases de Almacenamiento:**
  - **Azure File Storage:** Diferentes niveles de almacenamiento como Premium (transacciones frecuentes) y Standard (Hot, Cool, Archive) dependiendo de la frecuencia de acceso a los datos.
  - **Amazon S3:** Ofrece una gama más amplia de clases de almacenamiento optimizadas para acceso frecuente, acceso infrecuente, y archivo a largo plazo, incluyendo opciones específicas como Glacier para archivo a largo plazo y Intelligent-Tiering para optimización automática de costos.
3. **Rendimiento y Latencia:**

- **Azure File Storage:** El acceso premium proporciona un rendimiento más alto y latencias menores para aplicaciones que requieren acceso rápido a los datos.
- **Amazon S3:** Diseñado para ofrecer alta disponibilidad y durabilidad, con soporte para altas tasas de transferencia y escalabilidad automática de operaciones por segundo.

#### 4. Capacidad de Almacenamiento Máxima:

- **Azure File Storage:** Limita el tamaño máximo de cada file share a 100 TiB.
- **Amazon S3:** No impone límites explícitos en la cantidad de datos que se pueden almacenar, adaptándose a necesidades desde bytes hasta exabytes.

#### 5. Integración y Servicios Adicionales:

- **Azure File Storage:** Integración con Azure Active Directory y otras herramientas de Microsoft para una gestión de datos y seguridad más integrada.
- **Amazon S3:** Ofrece integración con otros servicios de AWS y características adicionales como bloqueo de objetos para protección contra eliminación o modificación durante un período definido.

## Recomendación

recomendaría Amazon S3 debido a su robusta capacidad para almacenar y gestionar grandes volúmenes de datos estructurados y no estructurados. Amazon S3 se integra perfectamente con herramientas como Amazon SageMaker, que es un servicio completamente gestionado que proporciona a los desarrolladores y científicos de datos la capacidad de construir, entrenar y desplegar modelos de machine learning , facilitando el entrenamiento y despliegue de modelos, y AWS Lambda, para la automatización de tareas. Su escalabilidad y durabilidad son ideales para manejar extensos conjuntos de datos y garantizar su disponibilidad. Además, S3 ofrece opciones como S3 Standard-IA para datos de acceso infrecuente y S3 Intelligent-Tiering para optimizar costos, junto con políticas de ciclo de vida para manejar eficientemente los datos a lo largo del tiempo.

## Referencias:

Microsoft. (n.d.). *Azure files pricing*. Microsoft Azure. Retrieved October 3, 2024, from <https://azure.microsoft.com/en-us/pricing/details/storage/files/>

Microsoft. (n.d.). *Introduction to Azure files*. Microsoft Learn. Retrieved October 3, 2024, from <https://learn.microsoft.com/en-us/azure/storage/files/storage-files-introduction>

Amazon Web Services. (n.d.). *Amazon S3 – Cloud object storage*. Amazon Web Services. Retrieved October 3, 2024, from <https://aws.amazon.com/es/s3/>

Amazon Web Services. (n.d.). *Amazon S3 pricing*. Amazon Web Services. Retrieved October 3, 2024, from <https://aws.amazon.com/es/s3/pricing/>

Amazon Web Services. (n.d.). *Amazon S3 storage classes*. Amazon Web Services. Retrieved October 3, 2024, from <https://aws.amazon.com/es/s3/storage-classes/>