

# ESPECIALIZACIÓN

# Ingeniería de datos con Azure

Curso: Pipelines para Data No estructurada y  
Big Data

Docente: Richard Tadeo Zenteno

## REGLAS



Se requiere **puntualidad** para un mejor desarrollo del curso.



Para una mayor concentración **mantener silenciado el micrófono** durante la sesión.



Las preguntas se realizarán **a través del chat** y en caso de que lo requieran **podrán activar el micrófono**.



Realizar las actividades y/o tareas encomendadas en **los plazos determinados**.



**Identificarse** en la sala Zoom con el primer nombre y primer apellido.

## ITINERARIO

*07:00 PM – 07:30 PM      **Soporte técnico DMC***

*07:30 PM – 08:50 PM      **Agenda***

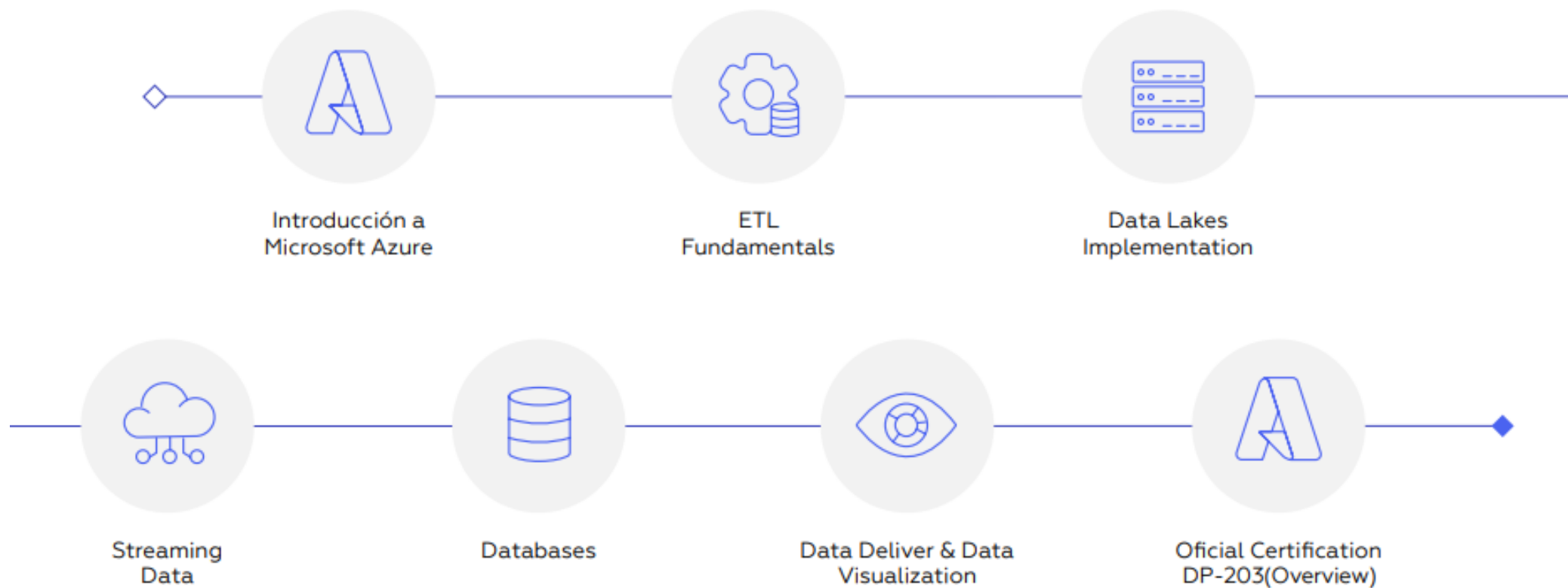
*08:50 PM – 09:00 PM      **Pausa Activa***

*09:00 PM – 10:30 PM      **Agenda***

*Horario de Atención Área Académica y Soporte*

*Lunes a Viernes 09:00 am a 10:30 pm / Sábado 09:00 am a 02:00pm*

# MALLA CURRICULAR



**CERIFICACIÓN FINAL**  
por **Aprobación** de la Especialización en **Ingeniería de Datos con Microsoft Azure** (48 horas académicas)

# CONTENIDO



## Introducción a Microsoft Azure

- Introducción a Cloud Computing. Proveedores de servicios Cloud, On-Premise vs. On-Cloud, principales servicios, descripción de los modelos de costos.
- Identity and Access Management (IAM). Overview de los roles principales, ejemplos de gestión de permisos.



## ETL Fundamentals

- Introducción a las soluciones ETL. Definición, descripción de sus etapas.
- Introducción a los servicios Azure Data Factory y Data Flow. Características generales, casos de uso.
- Taller: Implementación de un ETL Básico con Azure.



## Data Lakes Implementation

- Introducción a Data Lakes. Definición, arquitectura, capas (Raw, Stage, Analytics).
- Introducción a los servicios Azure Blob Storage y Storage Account.
- Taller: Implementación de un Datalake en Azure.

# CONTENIDO



## Streaming Data

- Introducción a procesamiento de datos Batch y Streaming. Diferencias Near-Real-Time y Real-Time.
- Introducción a IoT. Definición, uso de sensores, aplicaciones.
- Revisión de servicios: Azure EventHubs y IoT Hub. Características generales, ejemplos de implementación y uso.
- Taller: Manejo de Streaming al Data.



## Databases

- Introducción a las bases de datos Relacionales y No-Relacionales. Definición, características, casos de uso.
- Azure SQL Database for MariaDB. Descripción y características generales.
- Azure SQL Database for PostgreSQL. Descripción y características generales.
- Azure SQL Database for CosmosDB. Descripción y características generales.
- Taller: Diseño de una base de datos relacional y técnicas para poblarla.



## Data Deliver & Data Visualization

- Azure Synapse Analytics. Propósito del servicio, características generales.
- Fabric. Propósito del servicio, características generales.
- Taller: Conexión de Power BI a servicios de datos de Azure.

## AGENDA

01

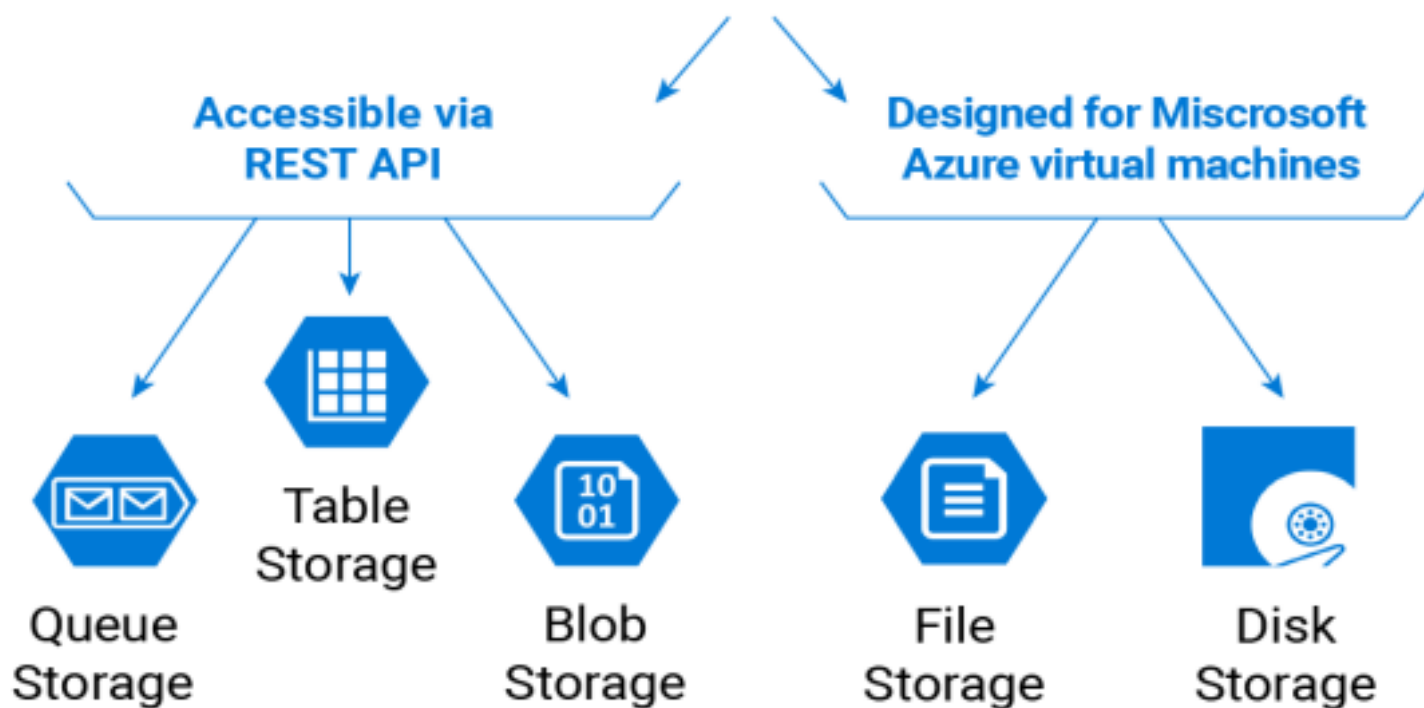
Introducción a los  
servicios Azure  
Storage Account

02

Introducción a los  
servicios Azure  
Blob Storage

## ¿Qué es Storage Account?

Un Storage Account es un contenedor que proporciona acceso a varios servicios de **almacenamiento** en Microsoft Azure. Permite almacenar y gestionar datos en la nube de forma segura y escalable.





## ¿Qué es REST API en Azure Storage?

En Microsoft Azure, los diferentes tipos de almacenamiento, como Blob Storage, Table Storage y Queue Storage, proporcionan acceso a través de una REST API.

1. **Interfaz de Comunicación: REST API** en Azure Storage es una interfaz que permite a los desarrolladores interactuar con los servicios de almacenamiento de Azure mediante solicitudes HTTP. Esta API define cómo se deben enviar las solicitudes al servidor y cómo se deben interpretar las respuestas.
2. **Operaciones Basadas en HTTP:** Utiliza los métodos estándar de HTTP (GET, POST, PUT, DELETE) para realizar operaciones sobre los recursos almacenados en Azure. Por ejemplo, puedes usar GET para recuperar datos, POST para agregar nuevos datos, PUT para actualizar datos existentes y DELETE para eliminar datos.
3. **Formato de Datos:** Los datos y metadatos se intercambian en formatos como JSON o XML. Por ejemplo, los blobs en Blob Storage pueden ser accedidos y manipulados en formato JSON.

## Componentes de REST API en Azure Storage

- 1. Recursos y Endpoints:** Cada tipo de almacenamiento en Azure tiene endpoints específicos que representan recursos. Por ejemplo:
  - **Blob Storage:** Los recursos son blobs y contenedores. Endpoints como <https://<account-name>.blob.core.windows.net/<container-name>/<blob-name>> permiten operaciones sobre blobs.
  - **Table Storage:** Los recursos son tablas y entidades. Endpoints como <https://<account-name>.table.core.windows.net/<table-name>> permiten operaciones sobre tablas.
  - **Queue Storage:** Los recursos son colas y mensajes. Endpoints como <https://<account-name>.queue.core.windows.net/<queue-name>> permiten operaciones sobre mensajes en las colas.
- 2. Autenticación y Autorización:** Las solicitudes a la REST API deben estar autenticadas. Azure Storage utiliza mecanismos como claves de acceso, Microsoft Entra ID para autenticar y autorizar el acceso a los recursos.
- 3. Operaciones y Métodos HTTP:**
  - GET:** Recuperar información sobre un recurso (por ejemplo, obtener el contenido de un blob).
  - POST:** Crear un nuevo recurso (por ejemplo, agregar un mensaje a una cola).
  - PUT:** Actualizar un recurso existente (por ejemplo, actualizar propiedades de un blob).
  - DELETE:** Eliminar un recurso (por ejemplo, eliminar un blob o mensaje de la cola).

## Ventajas de REST API en Azure Storage

- ✓ **Estándar Abierto:** REST API utiliza el protocolo HTTP, que es universalmente conocido y soportado por diversas plataformas y lenguajes de programación.
- ✓ **Interoperabilidad:** Permite la integración con diferentes servicios y aplicaciones que pueden realizar solicitudes HTTP.
- ✓ **Flexibilidad:** Facilita operaciones CRUD (Crear, Leer, Actualizar, Eliminar) sobre datos almacenados sin necesidad de interfaces específicas o clientes.

### Resumen

En **Microsoft Azure**, **REST API** es el mecanismo principal para interactuar con los servicios de almacenamiento en la nube. Permite realizar operaciones sobre recursos de almacenamiento mediante solicitudes HTTP estandarizadas. Esta interfaz es fundamental para acceder, manipular y gestionar datos en servicios como Blob Storage, Table Storage y Queue Storage, proporcionando una forma flexible y universal para interactuar con los servicios de almacenamiento de Azure.



# Diseño para Microsoft Azure Virtual Machine

Es un tipo específico de almacenamiento que está optimizado para ser utilizado en conjunto **con las máquinas virtuales (VMs) en Azure**. Este tipo de almacenamiento se ajusta a los requisitos específicos de las VMs para proporcionar un rendimiento óptimo, una alta disponibilidad y una integración fluida con los recursos de Azure.

## Componentes Clave del Almacenamiento Diseñado para Máquinas Virtuales

### 1. Azure Managed Disks

- I. **Descripción:** Los Managed Disks son discos virtuales administrados que se utilizan como almacenamiento para las máquinas virtuales de Azure. Microsoft se encarga de la administración del disco, incluida la replicación y el mantenimiento, lo que simplifica la gestión del almacenamiento.
- II. **Características:**
  - a. **Escalabilidad:** Permite la creación de discos de diferentes tamaños y tipos (por ejemplo, SSD, HDD) según las necesidades de rendimiento de las VMs.
  - b. **Alta Disponibilidad:** Los discos están replicados en múltiples ubicaciones dentro de una región para proporcionar alta disponibilidad y durabilidad.
  - c. **Tipos de Disco:** Incluye discos de alto rendimiento como Ultra Disk, Premium SSD, Standard SSD y Standard HDD, adecuados para diferentes niveles de carga de trabajo.

## Tipos de Almacenamiento



- **Blobs (Binary Large Objects):** Datos no estructurados como documentos, imágenes y videos.
- **Files:** Archivos en un sistema de archivos en la nube accesibles mediante SMB (Server Message Block).
- **Queues:** Mensajes para aplicaciones distribuidas.
- **Tables:** Datos estructurados en formato NoSQL.

## Niveles de Acceso

Los niveles de acceso en Azure Blob Storage definen el costo y el rendimiento del almacenamiento de blobs. Cada nivel está optimizado para diferentes escenarios de uso, desde acceso frecuente a datos hasta almacenamiento a largo plazo con bajo costo.

### Hot Tier

**Descripción:** El nivel **Hot** está diseñado para datos a los que se accede con frecuencia. Es ideal para aplicaciones y servicios que requieren acceso rápido y frecuente a los datos almacenados.

#### Características:

- **Costo:** Más alto en comparación con los niveles Cool y Archive para el almacenamiento de datos, pero ofrece el costo más bajo para las operaciones de lectura y escritura.
- **Rendimiento:** Proporciona el mejor rendimiento en términos de latencia y throughput para las operaciones de lectura y escritura.
- **Casos de Uso:** Datos de aplicaciones en uso activo, datos de análisis que se consultan frecuentemente, datos de medios en reproducción continua.

## Cool Tier

**Descripción:** El nivel **Cool** está diseñado para datos a los que se accede infrecuentemente, pero que deben estar disponibles de manera inmediata cuando se necesitan. Ofrece un costo de almacenamiento más bajo en comparación con el nivel Hot.

### Características:

- **Costo:** Menor costo de almacenamiento en comparación con el nivel Hot, pero mayores costos para las operaciones de lectura y escritura.
- **Rendimiento:** Menor rendimiento en comparación con el nivel Hot, pero adecuado para la mayoría de las aplicaciones que no requieren acceso frecuente a los datos.
- **Casos de Uso:** Datos que se acceden ocasionalmente, como archivos de respaldo que no se usan frecuentemente, datos históricos, datos de archivo de análisis que no se consultan regularmente.



## Archive Tier

**Descripción:** El nivel **Archive** está diseñado para el almacenamiento a largo plazo de datos que rara vez se accede y para los cuales el costo de almacenamiento es la principal preocupación. Este nivel ofrece el costo de almacenamiento más bajo de los tres niveles.

### Características:

- **Costo:** El costo de almacenamiento es significativamente más bajo en comparación con los niveles Hot y Cool, pero los costos de acceso a los datos (lectura y escritura) son más altos.
- **Rendimiento:** El tiempo de acceso a los datos en el nivel Archive es más largo, y la recuperación de datos puede llevar varias horas.
- **Casos de Uso:** Datos de archivo a largo plazo, datos de cumplimiento y regulaciones, copias de seguridad y datos históricos que se necesitan conservar pero rara vez se accede a ellos.



## Automatización de Transiciones

**Reglas de Ciclo de Vida:** Utiliza las políticas de ciclo de vida de Azure Blob Storage para automatizar la transición de datos entre niveles de acceso basándote en el tiempo de acceso y otros factores.

### Ejemplo de Política de Ciclo de Vida:

- ✓ **Regla de Transición:** Mover blobs al nivel Cool después de 30 días sin acceso.
- ✓ **Regla de Expiración:** Eliminar blobs del nivel Archive después de 5 años.

## Conclusión

Los niveles de acceso en **Azure Blob Storage** (Hot, Cool, Archive) están diseñados para optimizar el costo de almacenamiento en función de la frecuencia de acceso a los datos y las necesidades de recuperación. Elegir el nivel de acceso correcto puede ayudar a reducir costos y mejorar la eficiencia en la gestión de datos en la nube.

Microsoft Azure

Upgrade

Search resources, services, and docs (G+)

Copilot

docente1\_dmc@outloo...  
DMCPER01 (DOCENTE1DMCOUT...

Home > Storage accounts >

Storage accounts

DMCPer01 (docente1dmcoutlook.onmicrosoft.com)

Create

Restore

...

Filter for any field...

Name ↑

stracdmceucnshd001

strgacdmcd0001

stracdmceucnshd001

Storage account

Search

Microsoft Defender for Cloud

Data management

Storage tasks (preview)

Redundancy

Data protection

Object replication

Blob inventory

Static website

Lifecycle management

Azure AI Search

Settings

Configuration

Data Lake Gen2 upgrade

Upload

Open in Explorer

Delete

Move

Refresh

Open in mobile

Essentials

Resource group (move)

RSGRUCSHNDMCD0003

Location

eastus

Subscription (move)

Azure subscription DMC

Subscription ID

290ca738-ccf6-464f-a62a-68b7bda21ed1

Disk state

Available

Tags (edit)

Add tags

Performance

Standard

Replication

Locally-redundant storage (LRS)

Account kind

StorageV2 (general purpose v2)

Provisioning state

Succeeded

Created

12/8/2024, 20:25:04

Properties

Monitoring

Capabilities (7)

Recommendations (0)

Tutorials

Tools + SDKs

Blob service

Security

https://portal.azure.com/#@docente1dmcoutlook.onmicrosoft.com/resource/subscriptions/290ca738-ccf6-464f-a62a-68b7bda21ed1/resourceGroups/RSGRUCSHNDMCD0003/providers/Microsoft.Storage/storageAc...

Home > Storage accounts > stracdmceucnshd001 | Lifecycle management >

## Add a rule

Details **2 Base blobs**

Lifecycle management uses your rules to automatically move blobs to cooler tiers or to delete them. If you create multiple rules, the associated actions must be implemented in tier order (from hot to cool storage, then archive, then deletion).

If

Base blobs were \*

☒ Last modified

Move to cool storage

For infrequently accessed data that you want to keep on cool storage for at least 30 days.

Move to cold storage

For rarely accessed data that you want to keep for at least 90 days.

Move to archive storage

Use if you don't need online access and want to keep the object for 180 days or longer.

Delete the blob

Deletes the object per the specified conditions.

Delete the blob

Previous

Add

## Performance

El rendimiento de una **Storage Account** en Azure se refiere a la capacidad de la cuenta de almacenamiento para manejar operaciones de lectura y escritura. Azure ofrece diferentes opciones para ajustar el rendimiento de acuerdo con las necesidades de la aplicación o servicio.

### Tipos de Rendimiento:

#### 1. Standard Performance

Ofrece almacenamiento basado en discos duros (HDD) y es adecuado para la mayoría de las aplicaciones que no requieren un rendimiento extremadamente alto.

##### Características:

- **Costo:** Más económico en comparación con Premium.
- **Rendimiento:** Adecuado para aplicaciones con cargas de trabajo de acceso moderado o baja intensidad, como archivos de datos no críticos o aplicaciones de respaldo.
- **Tipos de Almacenamiento:** General Purpose v2 (GPv2), Blob Storage (con niveles de acceso Hot y Cool), Table Storage, Queue Storage.

## 2. Premium Performance

Ofrece almacenamiento basado en discos de estado sólido (SSD), proporcionando un alto rendimiento para aplicaciones que requieren baja latencia y alto throughput.

### Características:

- **Costo:** Más caro debido al alto rendimiento y la tecnología SSD.
- **Rendimiento:** Ideal para aplicaciones que requieren un acceso rápido y frecuente a los datos, como bases de datos de alto rendimiento, aplicaciones empresariales críticas y almacenamiento de datos en tiempo real.
- **Tipos de Almacenamiento:** Premium Blob Storage (con acceso a nivel Hot) y Premium File Storage.

#### Elección entre Standard y Premium:

**Standard:** Ideal para aplicaciones con requerimientos normales de rendimiento y donde el costo es una preocupación principal.

**Premium:** Perfecto para aplicaciones que necesitan un rendimiento de alto nivel y donde la velocidad y la latencia son críticos.



## Redundancy (Redundancia)

La **redundancia** en Azure Storage se refiere a las técnicas utilizadas para proteger los datos y asegurar su disponibilidad en caso de fallos o desastres. Azure ofrece varias opciones de redundancia para satisfacer diferentes necesidades de disponibilidad y durabilidad de datos.

### Tipos de Redundancia:

#### 1. Locally Redundant Storage (LRS):

Replica los datos dentro de una sola región, almacenando múltiples copias en discos físicos separados dentro del mismo centro de datos.

**Ventajas:** Menor costo y alta disponibilidad dentro de la región.

**Desventajas:** No protege contra fallos en el centro de datos o desastres regionales.

## 2. Geo-Redundant Storage (GRS):

Replica los datos en dos regiones geográficamente distintas. Primero almacena los datos en LRS en la región primaria, y luego replica esos datos a una región secundaria.

**Ventajas:** Alta durabilidad y disponibilidad en caso de desastres regionales.

**Desventajas:** Mayor costo y latencia para la replicación entre regiones.

## 3. Zone-Redundant Storage (ZRS):

Replica los datos en múltiples zonas dentro de una región. Las zonas son ubicaciones independientes dentro de la misma región, diseñadas para soportar fallos de zona.

**Ventajas:** Alta disponibilidad y durabilidad con protección contra fallos de zona dentro de una región.

**Desventajas:** No soporta Blob Storage en el nivel Archive, y puede ser más caro que LRS.

## 4. Geo-Zone-Redundant Storage (GZRS):

Combina la replicación geo-redundante con la replicación en zonas. Replica datos en múltiples zonas dentro de una región primaria y luego en una región secundaria.

**Ventajas:** Alta durabilidad y disponibilidad, con protección contra fallos de zona y regionales.

**Desventajas:** Costoso y puede tener mayor latencia debido a la replicación entre regiones y zonas.

# Consideraciones al elegir un tipo de cuenta de almacenamiento

**Requisitos de rendimiento:** Algunas cuentas están optimizadas para diferentes patrones de acceso y tipos de datos.

**Costos:** Los niveles de acceso y el tipo de almacenamiento (GPv2, Blob Storage, etc.) tienen estructuras de precios diferentes.

**Escenarios de uso:** Considera los servicios específicos que tu aplicación requiere (e.g., blobs, archivos, colas, tablas).

**Redundancia y durabilidad:** Algunas cuentas ofrecen opciones avanzadas de redundancia (LRS, GRS, RA-GRS, ZRS).



## Regulaciones

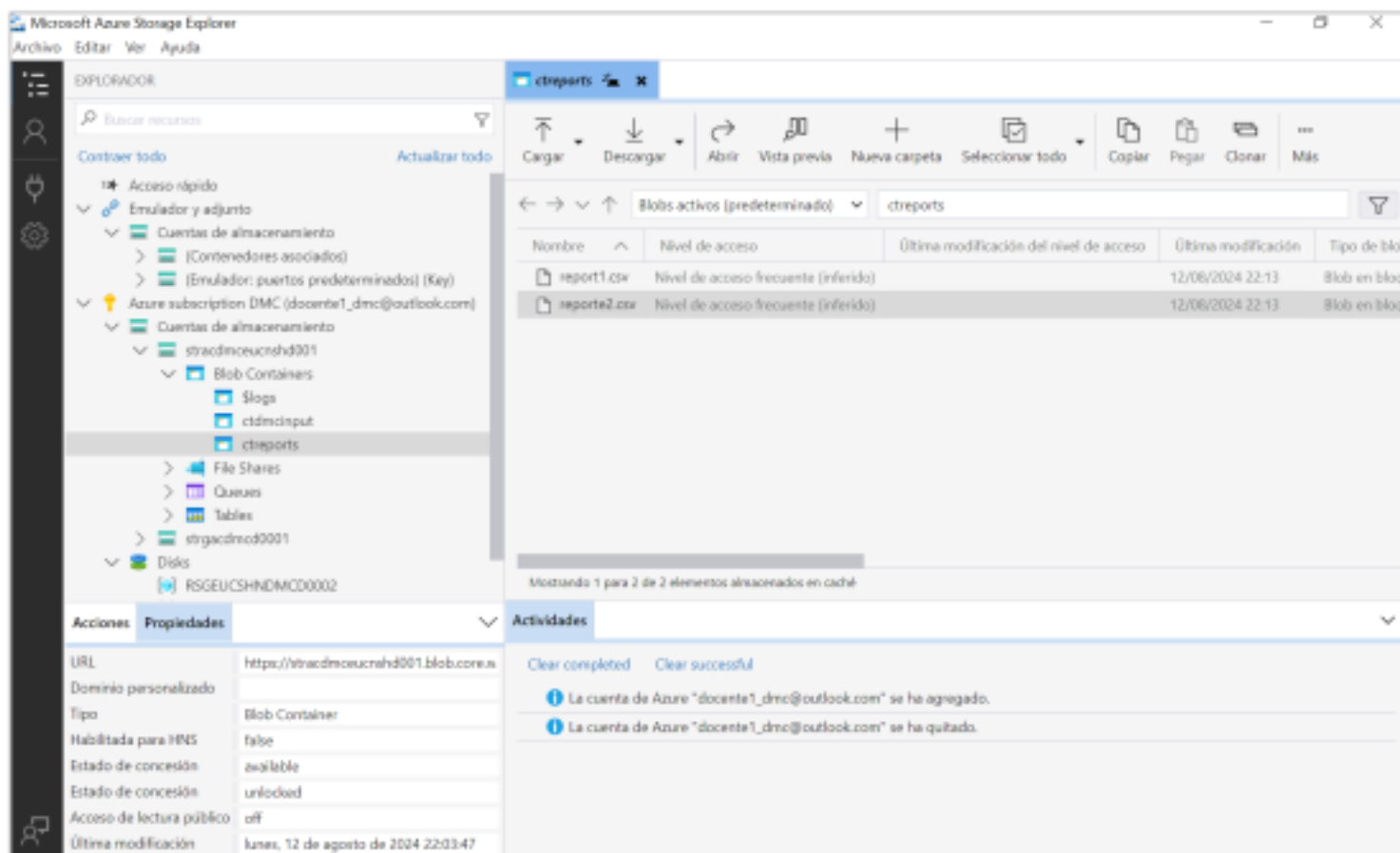
- Las organizaciones pueden cumplir con las regulaciones de privacidad de datos como GDPR, HIPAA o CCPA, y asegurar la protección de los datos personales y la confidencialidad de la información (USA - EU).
- En Perú, la regulación principal que rige la protección de datos personales es la **Ley de Protección de Datos Personales (Ley N° 29733)**, promulgada en julio de 2011, y su reglamento, aprobado por el Decreto Supremo N° 003-2013-JUS. Esta ley establece los principios, condiciones y obligaciones que deben cumplir las entidades públicas y privadas en el tratamiento de datos personales.



**GDPR**  
General Data  
Protection Regulation

# Microsoft Azure Storage Explorer

Es una aplicación de escritorio que ofrece una interfaz gráfica para gestionar y explorar datos en una cuenta de almacenamiento de Azure. Es útil para desarrolladores, administradores de sistemas y cualquier persona que necesite administrar datos en Azure Storage de manera visual y simplificada.



The screenshot displays the Microsoft Azure Storage Explorer application window. The interface is divided into several sections:

- Left Sidebar (EXPLORADOR):** Shows the navigation pane with 'Acceso rápido' (Quick access) and 'Emulador y adjunto' (Emulator and add-on). Under 'Emulador y adjunto', there are 'Cuentas de almacenamiento' (Storage accounts) and 'Cuentas de almacenamiento' (Storage accounts). The selected account is 'Azure subscription DMC (docente1\_dmc@outlook.com)'. Under this account, there are 'Cuentas de almacenamiento' (Storage accounts) and 'Cuentas de almacenamiento' (Storage accounts). The selected container is 'ctreports'.
- Top Bar:** Contains a search bar and a toolbar with icons for 'Cargar' (Upload), 'Descargar' (Download), 'Abrir' (Open), 'Vista previa' (Preview), 'Nueva carpeta' (New folder), 'Seleccionar todo' (Select all), 'Copiar' (Copy), 'Pegar' (Paste), 'Clonar' (Clone), and 'Más' (More).
- Main View:** Displays a table of blobs in the 'ctreports' container. The table has columns: 'Nombre' (Name), 'Nivel de acceso' (Access level), 'Última modificación del nivel de acceso' (Last modification of access level), 'Última modificación' (Last modification), and 'Tipo de blob' (Blob type).
 

Nombre	Nivel de acceso	Última modificación del nivel de acceso	Última modificación	Tipo de blob
report1.csv	Nivel de acceso frecuente (inferido)		12/06/2024 22:13	Blob en blob
report2.csv	Nivel de acceso frecuente (inferido)		12/06/2024 22:13	Blob en blob
- Bottom Panel:** Contains 'Acciones' (Actions), 'Propiedades' (Properties), and 'Actividades' (Activities).
  - Propiedades:** Shows details for the selected container, including 'URL' (https://vtracdmceucnhd001.blob.core.w...), 'Dominio personalizado' (Custom domain), 'Tipo' (Blob Container), 'Habilitada para HNS' (Enabled for HNS), 'Estado de concesión' (Grant status), 'Estado de concesión' (Grant status), 'Acceso de lectura público' (Public read access), and 'Última modificación' (Last modified).
  - Actividades:** Shows a list of activities, including 'Clear completed' and 'Clear successful'.

## Laboratorio: Manejo de Microsoft Azure Storage Explorer.

1. Recursos a utilizar Storage Account.
2. Ingresar al portal de Azure: <https://azure.microsoft.com/es-es/products/storage/storage-explorer>
3. Descargar e Instalar.
4. Asociar el contenedor asignado.
5. Interactúa con Storage Explorer

# RONDAS DE PREGUNTAS



**¡GRACIAS!**

