**Implementación de Arquitectura de Datos y Gobierno con Azure`**

## 1. Preparación e Ingesta del Dataset

Para este proyecto, se utilizó el Brazilian E-Commerce Public Dataset de Olist, un conjunto de datos de Kaggle ampliamente reconocido que resulta ideal para probar este tipo de implementaciones.

### Infraestructura y Seguridad en Azure

Se establecieron los siguientes componentes para alojar los datos y garantizar su seguridad:

- **Creación de Resource Group:** Se configuró un Resource Group para aplicar límites de acceso a la infraestructura, asegurando que solo los usuarios autorizados tengan acceso a la infraestructura.
- **Base de Datos Azure SQL:** Se optó por una base de datos en modalidad **PaaS (Platform as a Service)**, eliminando la necesidad de gestionar una Máquina Virtual (VM) propia.
    - **Configuración:** Se eligió el modelo **Serverless** por sus ventajas en el ahorro de costes, a diferencia del modelo *Provisioned*.
- **Seguridad Implementada:**
    - Servidor SQL (`server-gobernance.database.windows.net`) configurado en modo **Serverless**.
    - **Firewall por IP** para restringir el acceso.
    - Uso de **Microsoft Entra ID** (autenticación moderna) para el acceso.
    - Los datos viajan y se almacenan de forma **cifrada** (Cifrado TLS).

### Carga de Datos (Ecosistema de Datos)

Se simuló la fase de ingeniería de datos, importando manualmente los archivos planos (`.csv`) a coste cero utilizando **SQL Server Management Studio (SSMS)**, resultando en la creación de **cuatro tablas principales** en la base de datos:

- **Clientes:** Información geográfica de los clientes.
- **Pedidos:** Ciclo de vida de las compras (fechas de compra, aprobación, etc.).
- **Pagos:** Montos y métodos financieros.
- **Productos:** Categorías y dimensiones físicas.

Durante la ingesta, se aplicaron reglas de gobernanza a nivel técnico:

- **Tipado de Datos:** Se ajustaron tamaños (`nvarchar(100)`) y tipos (`int`, `float`, `datetime2`) para asegurar coherencia.
- **Gestión de Nulos:** Se permitió el valor nulo en campos de fecha opcionales, una decisión técnica para evitar la interrupción de la carga de datos.

## 2. Despliegue de Azure Purview y Gobierno de Datos

Se procedió con el despliegue de **Purview Enterprise (con capacidades completas)** para implementar la solución de Data Governance.**Registro y Conexión de Recursos**

1. **Registro de Azure SQL como Source:** Se registró el recurso Azure SQL en el Data Map de Purview.
2. **Permisos para el Escaneo:** Se otorgaron los permisos necesarios a la identidad de Purview (`Gobernance`) sobre la base de datos SQL para permitir el escaneo de metadatos y linaje:
    - **Nivel Base de Datos:** `GRANT SELECT TO [Gobernance];`
    - **Visibilidad de Metadatos:** `GRANT VIEW DEFINITION TO [Gobernance];`
    - **Estado de Base de Datos:** `GRANT VIEW DATABASE STATE TO [Gobernance];`

**Gestión de Acceso Centralizado (Data Policy Enforcement)**

Se resalta la capacidad de **Data Policy Enforcement** de Purview, que permite:

- **Control Centralizado:** Gestionar permisos de lectura/escritura sin modificar código SQL, usando clics en Purview.
- **Políticas de Autoservicio:** Los usuarios pueden solicitar acceso a datos a través del catálogo, y Purview lo habilita automáticamente tras la aprobación del Data Governor.
- **Seguridad Avanzada:** Crear políticas para que ciertos usuarios solo vean datos bajo condiciones de seguridad específicas.

## 3. Activación del Catálogo de Datos y Gobernanza

Una vez registrado el recurso, el siguiente paso fue realizar el **Scan (Escaneo)** para generar los artefactos de gobernanza: Clasificaciones, Glosario y Capturas de Pantalla.

**Pasos como Analista de Gobernanza**

Para profesionalizar el proyecto en la pestaña **Overview** de la base de datos, los siguientes pasos son cruciales:

| Paso | Pestaña/Herramienta | Qué Buscar/Acción | Objetivo |
|---|---|---|---|

| | | | |
|---|---|---|---|
| **A: El Esquema** | **Schema** (Activo de Datos) | Revisar si Purview detectó **Classifications** automáticamente (Ej. PII). Si falta, añadir etiquetas manualmente (Ej. PII a nombres de clientes) mediante **Edit**. | Garantizar el cumplimiento legal y la identificación de datos sensibles. |
| **B: El Glosario** | **Business Glossary** (Menú Principal) | **Crear un término** (Ej. "Customer_ID") y definirlo (Ej. "Identificador único universal para clientes de la plataforma Olist"). Luego, **vincular** el término a la columna. | Crear un lenguaje de negocio unificado y evitar la ambigüedad de los campos. |
| **C: Contactos** | **Contacts** (Activo de Datos) | Añadirse como **Data Steward** (Administrador de datos). | Establecer la responsabilidad (Data Ownership) para saber a quién contactar ante fallos o dudas. |

**Colecciones (Organización por Negocio)**

Las **Colecciones** en Purview se utilizan para segmentar la visibilidad de los datos por departamento, permitiendo que:

- El equipo de Finanzas solo vea sus tablas.
- El equipo de Logística solo vea las suyas.

**Diferencia clave con Resource Group:** Los Resource Groups en Azure gestionan el acceso *total o denegado* a la infraestructura, mientras que las Colecciones en Purview gestionan la *visibilidad parcial* de los datos dentro del catálogo.

**4. El Valor de Azure Purview (Visión de Negocio)**

La diferencia fundamental entre el Portal de Azure y Azure Purview es el cambio de un enfoque técnico a uno de negocio:

| Aspecto | Portal de Azure (Técnico) | Azure Purview (Negocio) |
|---|---|---|
| **Rol** | El "Cómo" (Desarrollador) | El "Qué" y el "Quién" (Analista de Data Governance) |
| **Datos** | Datos "crudos" (Un nombre es un `nvarchar`) | **Activo de Negocio** (Un nombre es un "Dato Sensible PII" con riesgo legal - RGPD) |
| **Información** | Muestra el estado del servidor y permite consultas SQL. | Proporciona **Clasificación Automática**, **Glosario de Negocio** (contexto) y **Linaje de Datos** (trazabilidad). |
| **Utilidad** | Uso técnico y operativo (consultas, encendido/apagado). | **Buscador unificado** ("Google" para datos) entre todas las bases de datos de la empresa y gestión de políticas. |

Purview transforma una base de datos en un **Activo de Negocio** catalogado y comprensible para toda la organización.

**Próximos Pasos (Hoja de Ruta)**

Una vez que los datos están catalogados y gobernados, las siguientes acciones se centran en la extracción de valor y calidad:

1. **Glosario de Datos y Clasificación:** Definir y aplicar etiquetas de **PII** (Información de Identificación Personal) y **CDE** (Elementos Críticos de Datos) a las columnas del dataset de Olist.
2. **Scripts de Calidad (Data Quality):** Escribir consultas SQL avanzadas para detectar anomalías, como pedidos entregados antes de la compra o transacciones con valor cero.
3. **Análisis de Valor:** Crear una consulta SQL para unir las tablas y calcular, por ejemplo, el gasto total por ciudad (`JOIN`).

| Actual | Mejoras a Futuro |
|---|---|

| | |
|---|---|
| "Subí unos archivos a una base de datos SQL." | "Diseñar una **arquitectura de ingesta de datos** en Azure." |
| "Usé el asistente de importación de SSMS." | "Implementar **pipelines de ETL** con Azure Data Factory." |
| "Los datos están en la nube." | "Garantizar la **durabilidad y el linaje** usando un Data Lake." |