

[illegible]

The illustration features a central lightbulb with a yellow glow and red and blue scribbles inside. To the left, there's a section titled "BUSINESS Idea" with various icons like gears, bar charts, and a pie chart. Above it, words like "genius", "creativity", and "PLAN" are written. Below the lightbulb, there's a bar chart with numbers 1, 2, 3, 4. To the right, there's a section titled "CONCEPT" with icons like a magnifying glass labeled "search", a target, a clock labeled "Time", a briefcase, a coffee cup, and a group of people labeled "Team". Below this, there's a flowchart showing "POWER" leading to "success". At the bottom right, there's a network diagram with nodes labeled "TEAM", "WORK", "IDEA", and "WEB", connected by arrows. There are also icons of a computer monitor, a server tower, and a bar chart with numbers 10, 50, 65. The entire background is filled with various business-related sketches and symbols.

Tabla de contenidos

-
1. Descripción del caso
 2. Objetivos
 3. Descripción de los de los datos
 4. Storage Account
 5. Data Lake
 6. Servidor de DB
 7. DB de dbRetail
 8. Data Factory
 9. Databricks
 10. Proximos pasos.
-



Descripción del caso

- **Te contratan como Arquitecto Cloud para una startup de IT dedicada a la consultoría y asesoría en temáticas orientadas a analítica de datos. Marcela, project manager de tu equipo, te comparte el pedido de que un cliente de la organización (Tech Consulting Group) necesita realizar un proceso de ETL sobre una base de datos que se encuentra de forma “on - premise” dentro de su compañía.**

Objetivos

Desarrollar una canalización de datos (pipeline) utilizando diferentes servicios de Azure.

Desarrollar Transformaciones ETL en la DB

Almacenar, procesar y orquestar los datos. Como requerimiento excluyente del proyecto, es necesario que todo el flujo de datos se ejecute sobre la nube de Azure dado que es el proveedor que actualmente posee contratado la organización.



Descripción de los datos

- **La base de datos a utilizar se llama: dbRetail y se encuentra compuesta por diferentes tablas de negocio como ser: Resulta importante destacar que el formato de la base de datos es de tipo bacpac.**

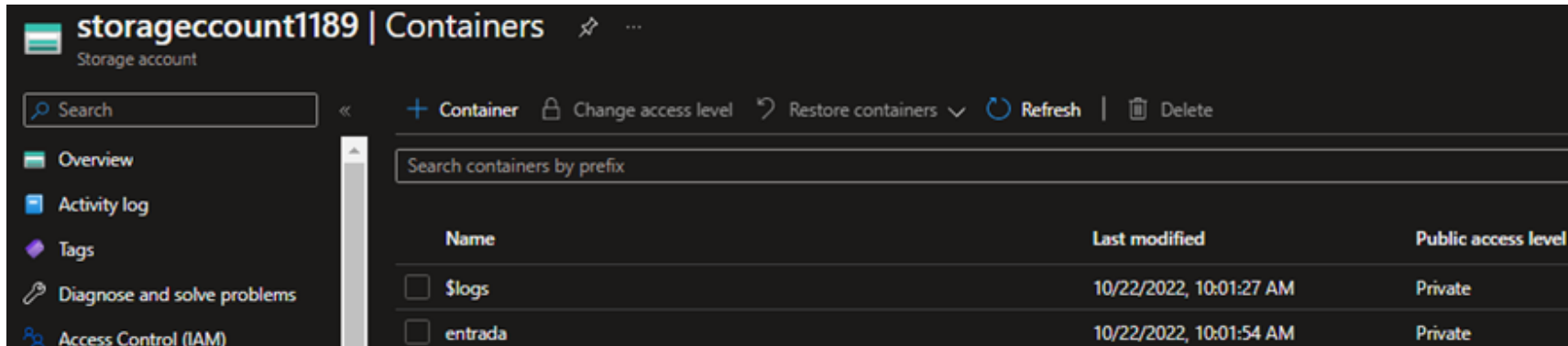


Storage Account

- La plataforma de Azure Storage **es la solución de almacenamiento en la nube de Microsoft para los escenarios modernos de almacenamiento de datos**. Azure Storage ofrece almacenamiento de alta disponibilidad, escalable de forma masiva, duradero y seguro para una gran variedad de objetos de datos en la nube.

Storage Account

- Para iniciar nuestro Proyecto, como punto inicial debemos crear nuestro servicio de Storage Account al cual le vamos a crear un contenedor llamado entrada que va a albergar nuestro archivo BACPAC con nuestras tablas para su analisis.



storageaccount1189 | Containers

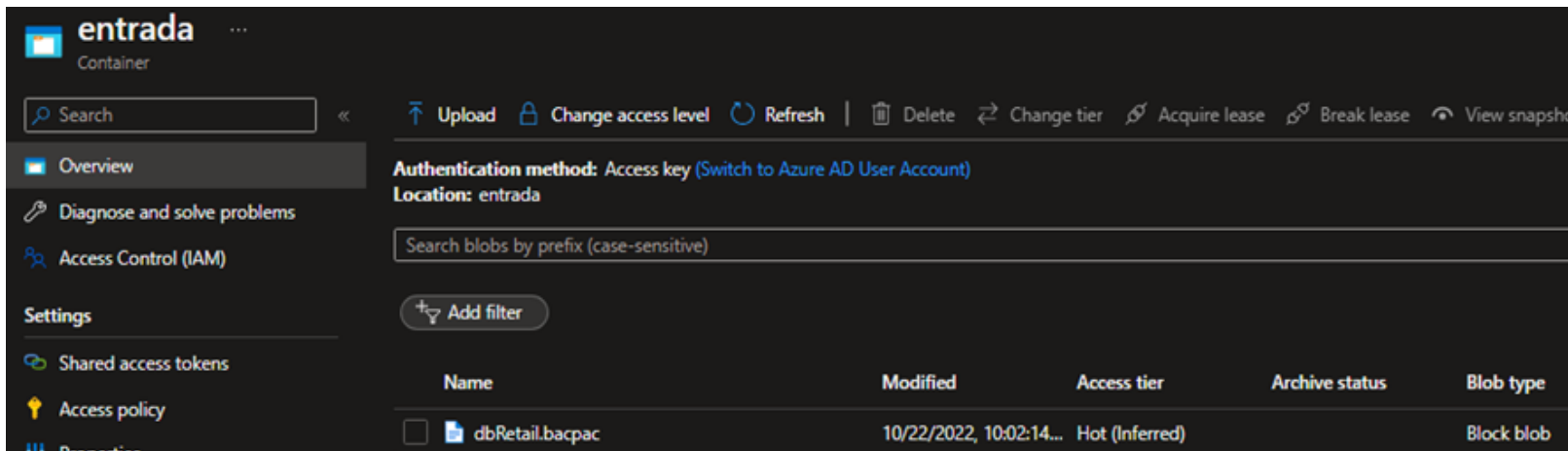
Storage account

Search

+ Container Change access level Restore containers Refresh Delete

Search containers by prefix

Name	Last modified	Public access level
\$logs	10/22/2022, 10:01:27 AM	Private
entrada	10/22/2022, 10:01:54 AM	Private



entrada

Container

Search

Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshot

Authentication method: Access key (Switch to Azure AD User Account)

Location: entrada

Search blobs by prefix (case-sensitive)

Add filter

Name	Modified	Access tier	Archive status	Blob type
dbRetail.bacpac	10/22/2022, 10:02:14...	Hot (Inferred)		Block blob

Data Lake Storage

- Azure Data Lake Storage Gen2 es un Servicio de almacenamiento en la nube dedicado al análisis de macrodatos e integrado en Azure Blob Storage.
- Data Lake Storage Gen2 combina las funcionalidades de Azure Blob Storage y Azure Data Lake Storage Gen1. El servicio resultante ofrece características de Azure Data Lake Storage Gen1, entre las que se incluyen la semántica del sistema de archivos, la seguridad de nivel de directorio y de archivo y la adaptabilidad. Junto con las funcionalidades de bajo costo, almacenamiento en capas, alta disponibilidad y recuperación ante desastres de Azure Blob Storage.

Data Lake Storage

- Para nuestra salida de datos, vamos a crear un Data Lake que va a recibir nuestro archivo luego de haberse echo las Transformaciones correspondientes.

The screenshot shows the Azure Data Lake Storage 'Containers' view for the storage account 'datalake1189'. The interface includes a left-hand navigation pane with options like Overview, Activity log, Tags, Diagnose and solve problems, and Access Control (IAM). The main area features a toolbar with actions such as '+ Container', 'Change access level', 'Restore containers', 'Refresh', and 'Delete'. Below the toolbar is a search bar labeled 'Search containers by prefix'. A table lists the existing containers:

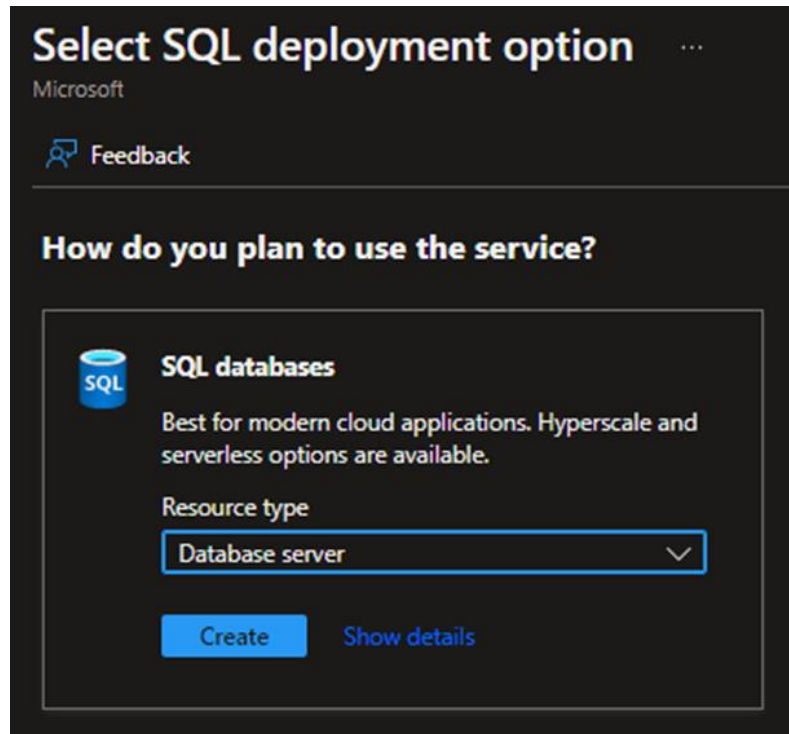
	Name	Last modified	Public access level
<input type="checkbox"/>	\$logs	10/22/2022, 10:17:22 AM	Private
<input type="checkbox"/>	salida	10/22/2022, 10:17:51 AM	Private

Servidor de DB en la nube

- Las bases de datos en la nube son una nueva modalidad de almacenamiento que difieren de las bases tradicionales. A diferencia de estas, **las bases de datos en cloud no se almacenan en un equipo o sistema local, sino que se ejecuta desde la infraestructura de un proveedor de servicios.**


Servidor de DB en la nube

- Para poder manipular nuestra base de datos, primero hay que crear el servicio de Azure SQL y dentro de este servicio, vamos a utilizar el recurso de servidor de base de datos al que le vamos a realizar un Backup y conectar nuestro archivo bacpac alojado en el contenedor del storage para crear una base de datos nueva.




Select SQL deployment option ...

Microsoft

 Feedback

How do you plan to use the service?

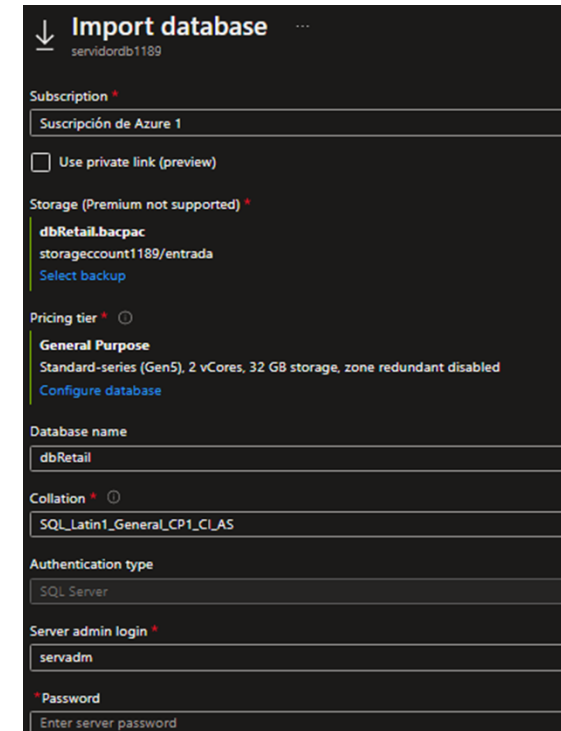
 **SQL databases**

Best for modern cloud applications. Hyperscale and serverless options are available.

Resource type

Database server

Create Show details



Import database ...

servidordb1189

Subscription *

Suscripción de Azure 1

☐ Use private link (preview)

Storage (Premium not supported) *

dbRetail.bacpac
storageaccount1189/entrada
[Select backup](#)

Pricing tier * ⓘ

General Purpose
Standard-series (Gen5), 2 vCores, 32 GB storage, zone redundant disabled
[Configure database](#)

Database name

dbRetail

Collation * ⓘ

SQL_Latin1_General_CP1_CI_AS

Authentication type

SQL Server

Server admin login *

servadm

* Password

Enter server password

Servidor de DB en la nube

- Para revisar que se esta realizando correctamente la importacion de la base de datos, podemos ingresar a la pestaña de import/export donde vamos a visualizar si efectivamente se esta realizando nuestra carga.

The screenshot displays the AWS Management Console interface for an SQL server instance named 'servidordb1189'. The page title is 'servidordb1189 | Import/Export history'. On the left, there is a navigation sidebar with the following menu items: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, and Quick start. The main content area shows the 'Import/Export history' section, which includes a 'Refresh' button and a table of operations.

Operation	Database
ImportToExistingDatabase	WideWorldImporters-Standard
ImportToExistingDatabase	dbRetail
ImportToExistingDatabase	dbRetail
ExportDatabase	dbRetail