



Julio 08 de 2024

# **Estrategia laboratorio de datos SNUIRA AGROSAVIA**

Moderador: Victor Manuel Mondragon M.

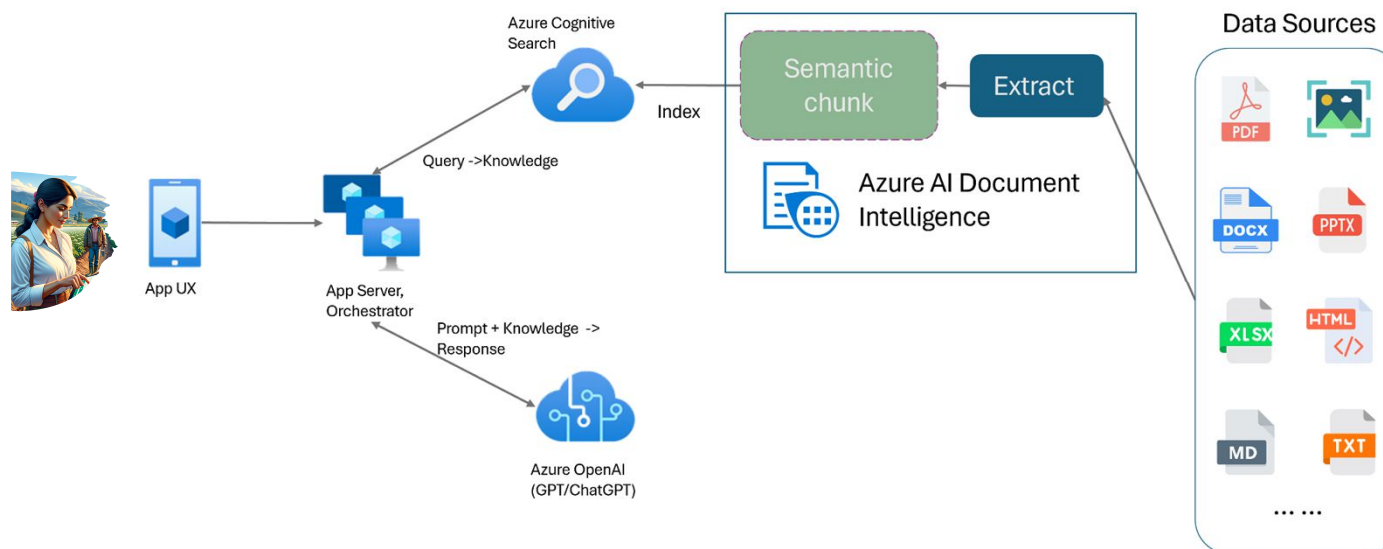
[www.upra.gov.co](http://www.upra.gov.co)

## Agenda - objetivos

1. DataFactory con inter-operabilidad de BAC a través de la API PRIMO.
2. Almacenamiento de archivos PDF, mpg4 y otros en Azure Blob Storage.
3. Pruebas de carga utilizando índice vectorial para la recuperación de generación aumentada (RAG).
4. Ajuste de cronograma para cierre y entrega del prototipo experimental YoCampo.

# 0. Contexto

# Generación aumentada de recuperación (RAG)

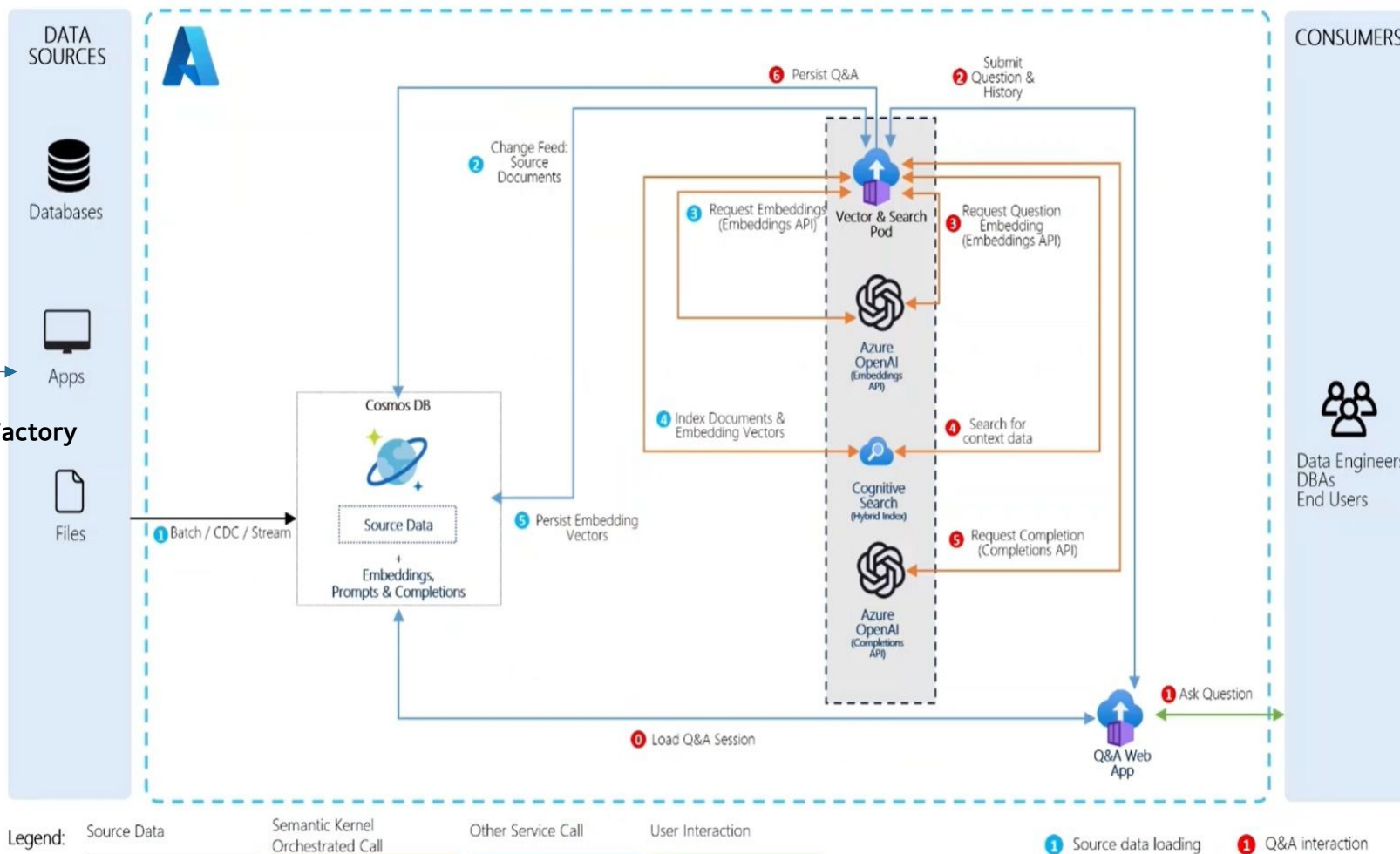


Fuente: Adaptada de Microsoft

- LLM se entrena previamente mediante datos públicos, pero genera **respuestas aumentadas por información del recuperador** ( GPT)

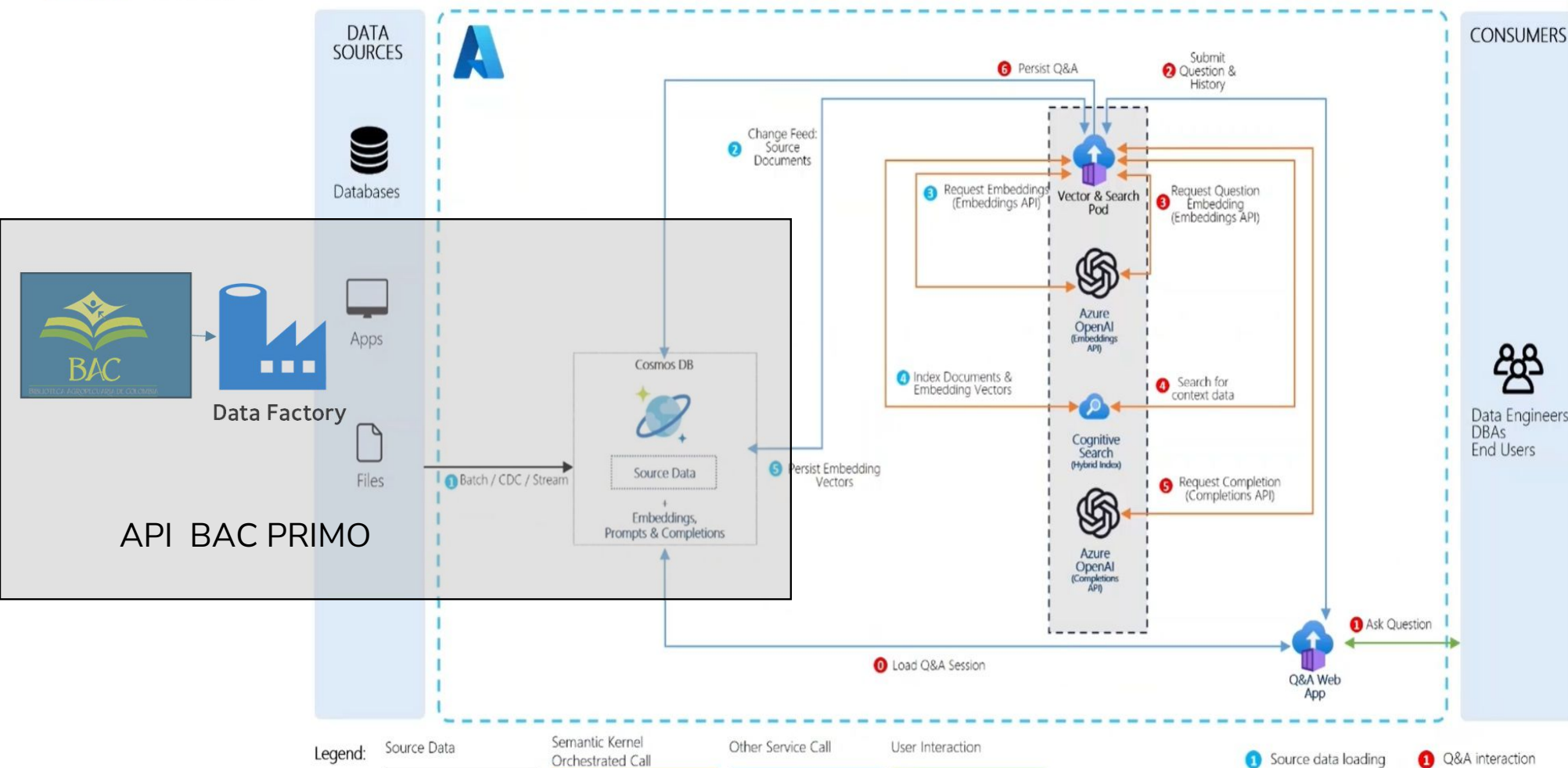


**Data Factory**



# 1. API inter-operabilidad de BAC PRIMO

Fuentes de Datos- YoCampo





# Funcionalidades API



## Producto obtenido

- Archivo OAI\_FINAL\_Storage\_Account.ipynb
  - Para ser ejecutado desde un ambiente Notebook ( Google Colab, databricks, Anaconda entre otros)
- Archivo oai\_final\_storage\_account.py
  - Para ser ejecutado desde un ambiente pyhton.

## 2. Almacenamiento de Datos

### Storage account

# Blob – YoCampo ( cacao)

QUERY = 'Cacao'

Home > storageyocampo

storageyocampo | Storage browser

Storage account

Search

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Storage Mover

Partner solutions

Data storage

Containers

storageyocampo

Favorites

Recently viewed

Blob containers

\$logs

cacao20240714183705

View all

Authentication method: Access key (Switch to Microsoft Entra user account)

Add filter

Search blobs by prefix (case-sensitive)

Showing all 7 items

Name	Last modified
excel	
md	
other	
pdf	
video	
word	

Total items found: 1047

QUERY = 'agroforestal'

storageyocampo

Favorites

Recently viewed

Blob containers

\$logs

agroforestal20240715203848

View all

Authentication method: Access key (Switch to Microsoft Entra user account)

Add filter

Search blobs by prefix (case-sensitive)

Only show active blobs

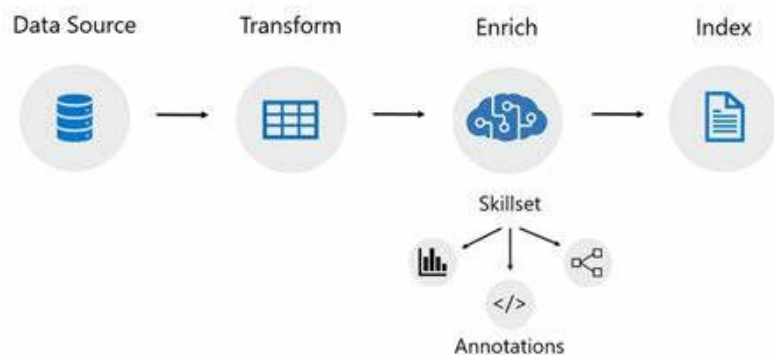
Showing all 5 items

Name	Last modified	Access tier
excel		
other		
pdf		
video		
word		

Total items found: 702

# Input Blob – Indexes ( cacao)

Agrietar, fragmentar e incrustar



**chunk\_size**  
1024

