



Julio 16 de 2024

# **Estrategia laboratorio de datos SNUIRA AGROSAVIA**

Moderador: Victor Manuel Mondragon M.

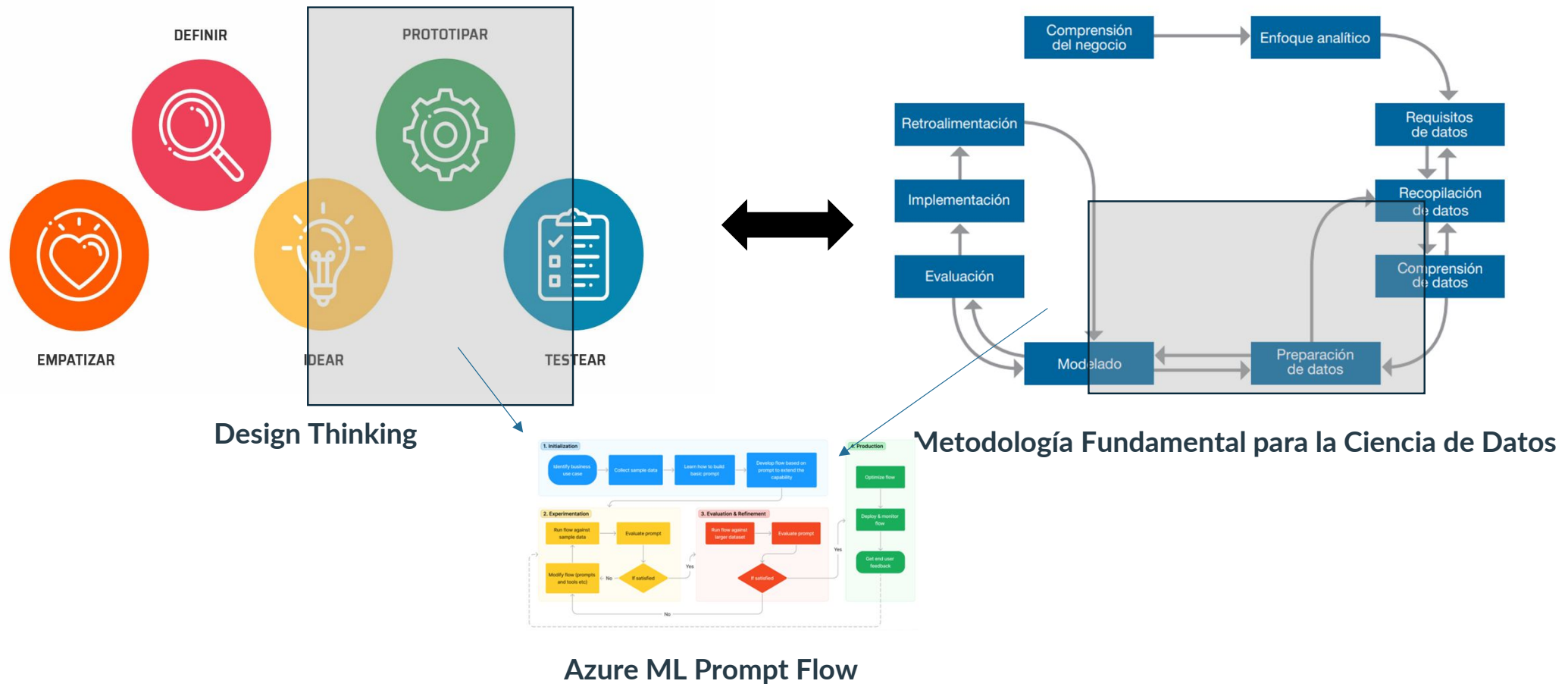
[www.upra.gov.co](http://www.upra.gov.co)

## Agenda - objetivos

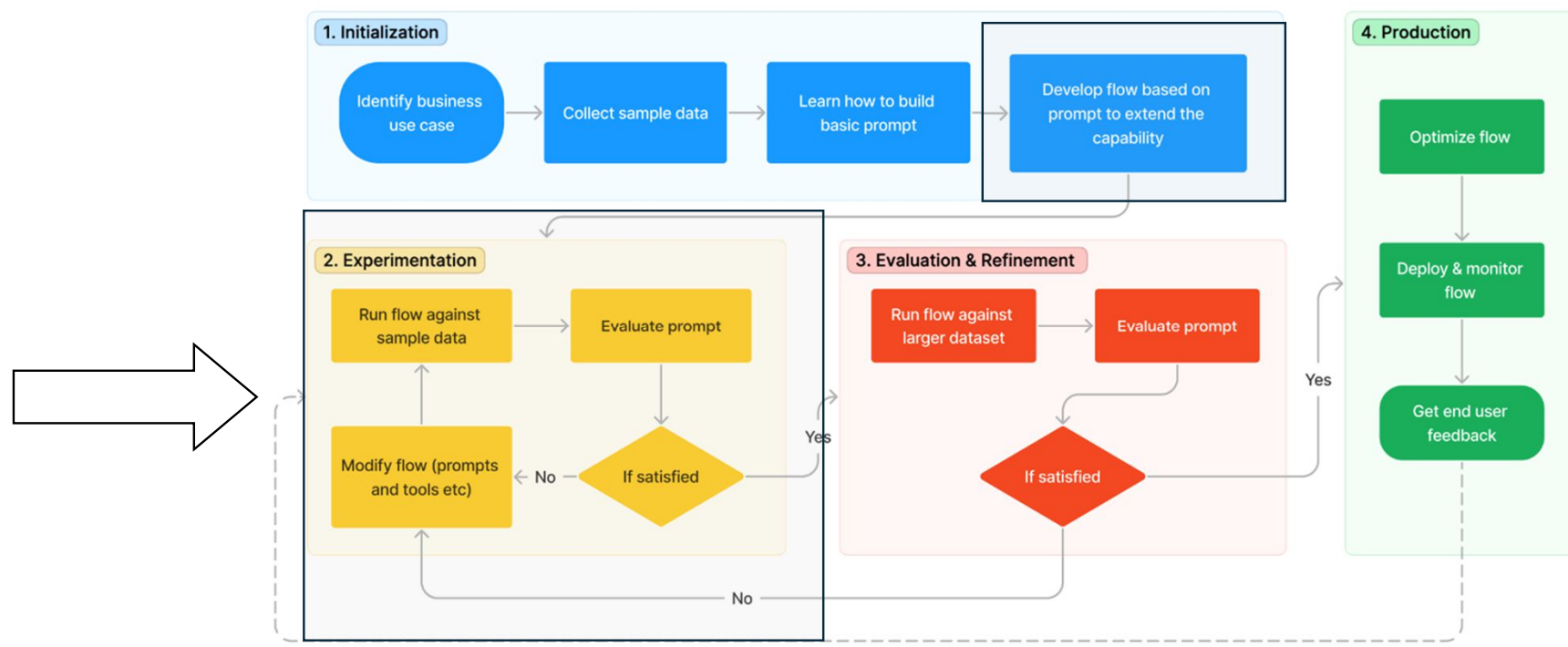
1. DataFactory con inter-operabilidad de BAC a través de la API PRIMO.
2. Almacenamiento de archivos PDF, mpg4 y otros en Azure Blob Storage.
3. Pruebas de carga utilizando índice vectorial para la recuperación de generación aumentada (RAG).
4. Ajuste de cronograma para cierre y entrega del prototipo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

# 0. Contexto

# Metodología AI Lab SNUIRA



# Actividad feedback con extensionistas

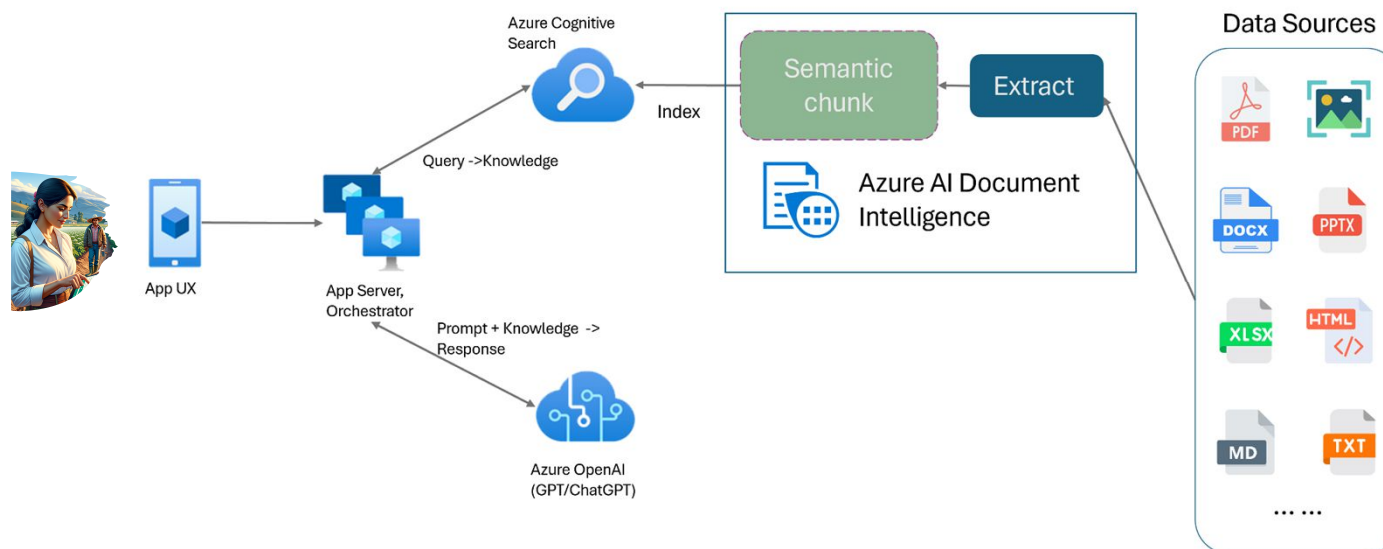


GPT Prompt Engineering

Solution Architecture

Prompt engineering techniques

# Generación aumentada de recuperación (RAG)

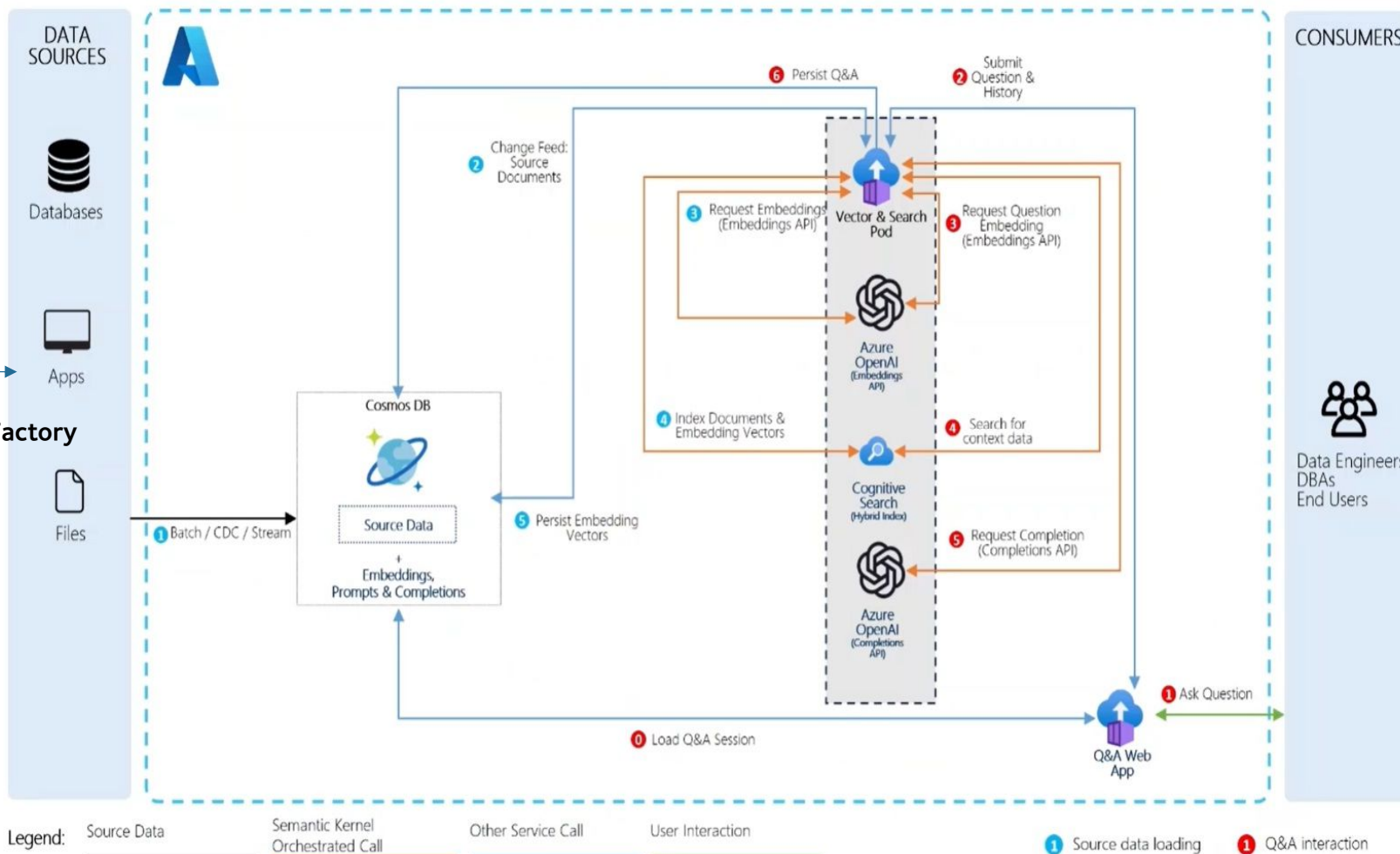


Fuente: Adaptada de Microsoft

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



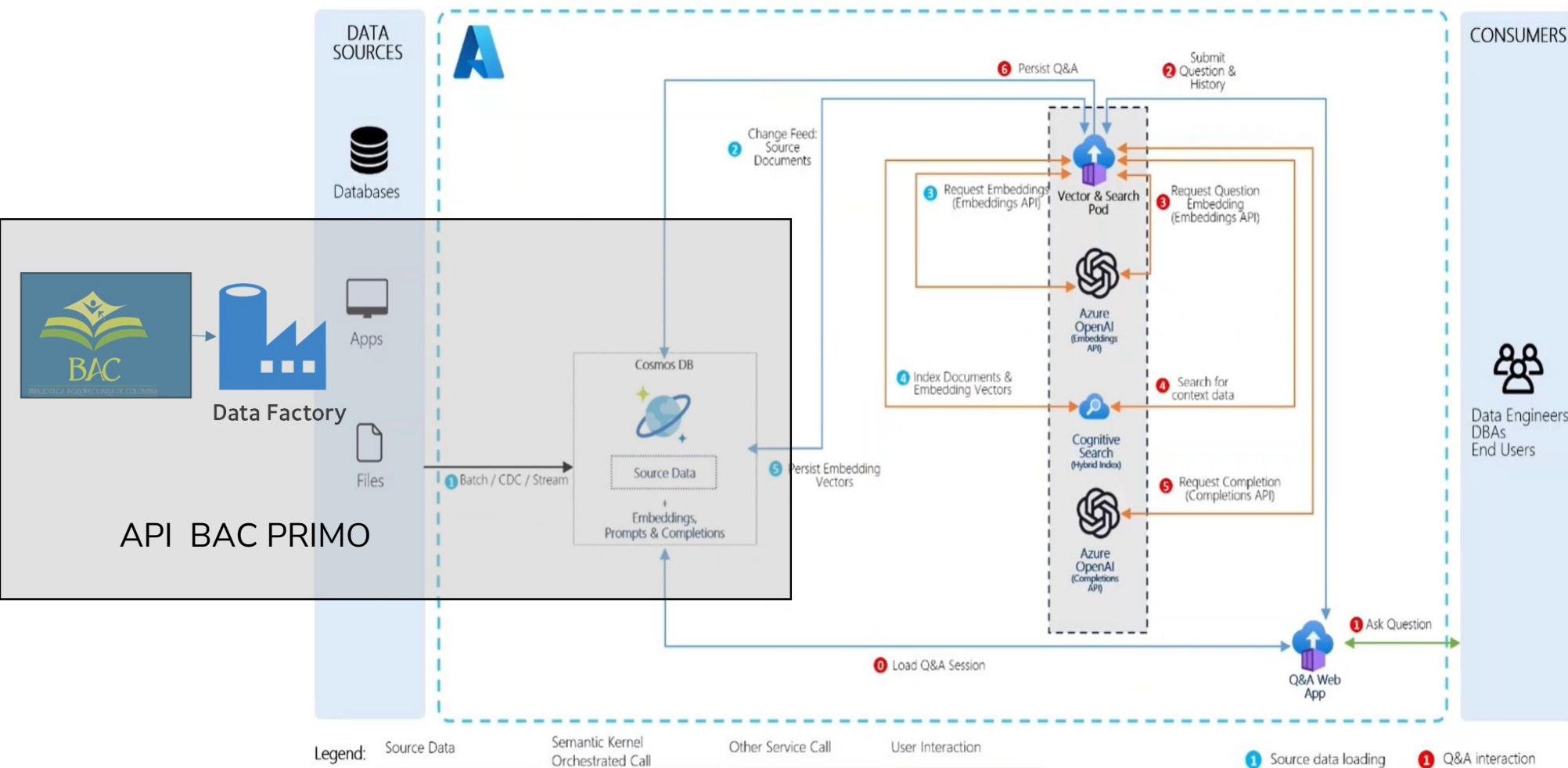
**Data Factory**





# 1. API inter-operabilidad de BAC PRIMO



Fuentes de Datos- YoCampo



# Funcionalidades API



## Producto obtenido

Nombre	Estado	Fecha de modificación	Tipo	Tamaño
OAI_FINAL_Storage_Account.ipynb		15/07/2024 3:13 p. m.	Archivo de origen ...	653 KB
oai_final_storage_account.py		15/07/2024 3:14 p. m.	#code	13 KB

- 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.

**Nota actualizar credenciales del Container**



carga\_cacao20240714183705.csv

## 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

# Analisis

## Tipo\_PDF\_Imagen\_texto.ipynb

	Tipo de Archivo	Cantidad
1	pdf	956
2	other	71
3	video	53
4	excel	2

```
df = pd.DataFrame(files_data)

# Guardar la tabla en un archivo CSV temporalmente
output_csv_path = '/tmp/resultado_detectar_type_pdf.csv'
df.to_csv(output_csv_path, index=False)

# Subir el archivo CSV al Blob Storage
output_blob_name = "resultado_detectar_type_pdf.csv"
blob_client = container_client.get_blob_client(output_blob_name)

with open(output_csv_path, "rb") as data:
    blob_client.upload_blob(data, overwrite=True)

print(f"\nArchivo {output_blob_name} subido a Azure Blob Storage en la ruta / del contenedor {container_name}")
```

```
Archivo: pdf/Ver_documento_36615.pdf, Peso: 3646.79 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36679.pdf, Peso: 1070.40 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36699.pdf, Peso: 42828.11 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36710.pdf, Peso: 7539.63 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36730.pdf, Peso: 6282.16 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36839.pdf, Peso: 2696.69 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36840.pdf, Peso: 683.67 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36841.pdf, Peso: 1899.43 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36842.pdf, Peso: 1825.15 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36843.pdf, Peso: 1600.15 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36844.pdf, Peso: 925.42 KB, Tipo: Texto
Archivo: pdf/Ver_documento_36845.pdf, Peso: 1653.82 KB, Tipo: Texto
```

Cantidad de PDFs de texto: 848  
Cantidad de PDFs de imagen: 108

# Producto

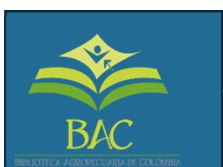
- 02-Blob Detección tipo PDF
  - tipo\_pdf\_imagen\_texto.py
  - Tipo\_PDF\_Imagen\_texto.ipynb



resultado\_detectar\_type\_pdf.csv

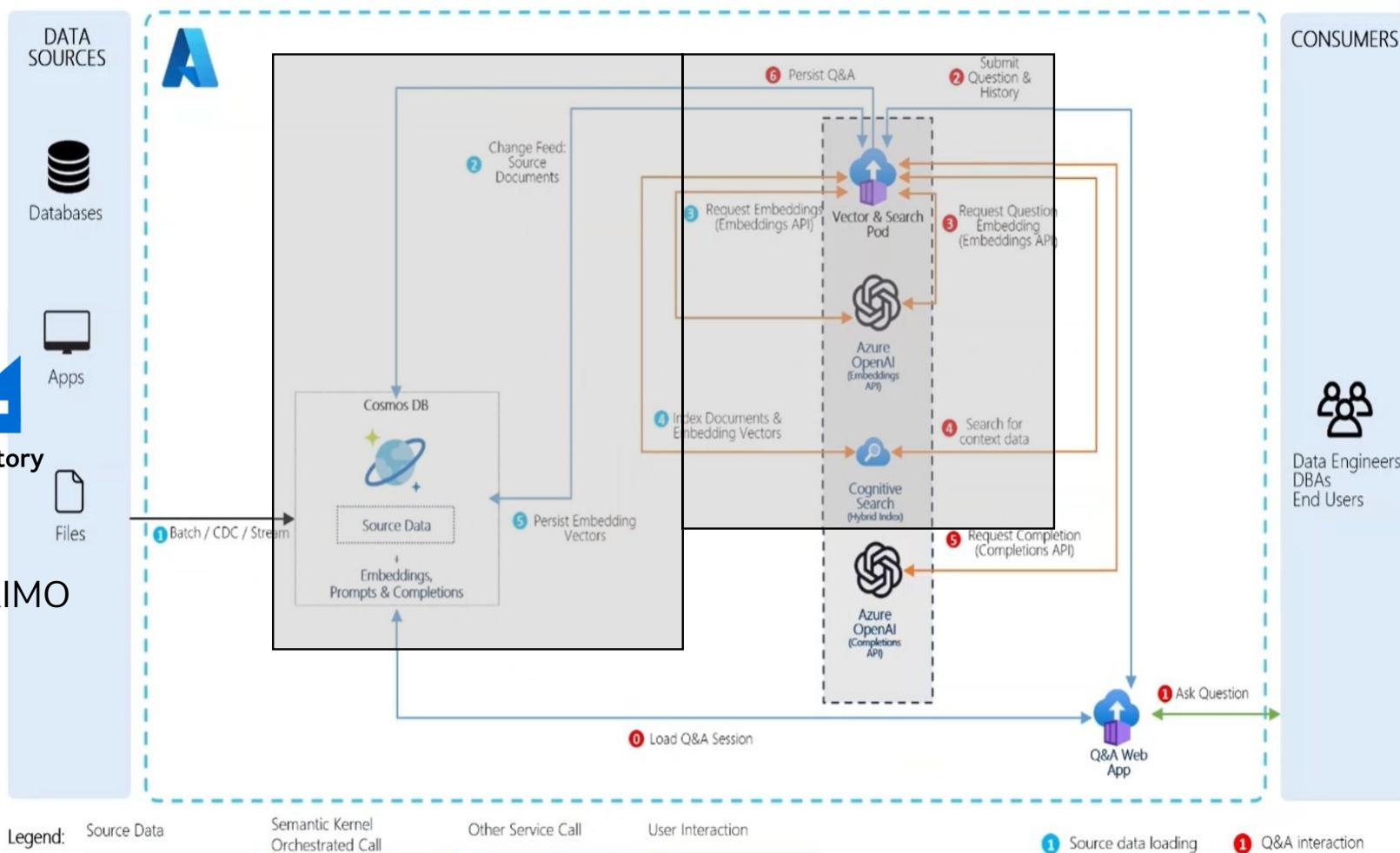


### 3. índice vectorial para la recuperación de generación aumentada (RAG).



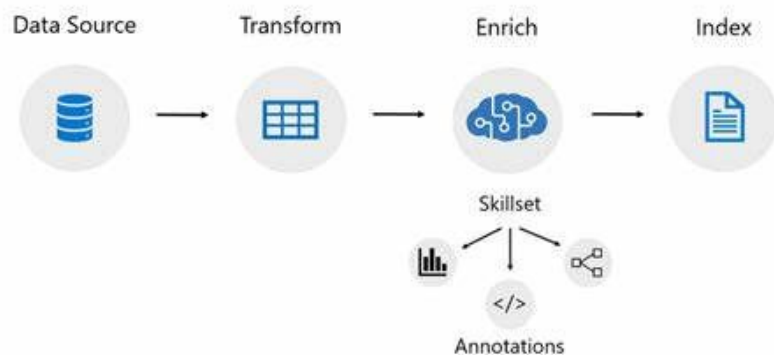
**Data Factory**

API BAC PRIMO

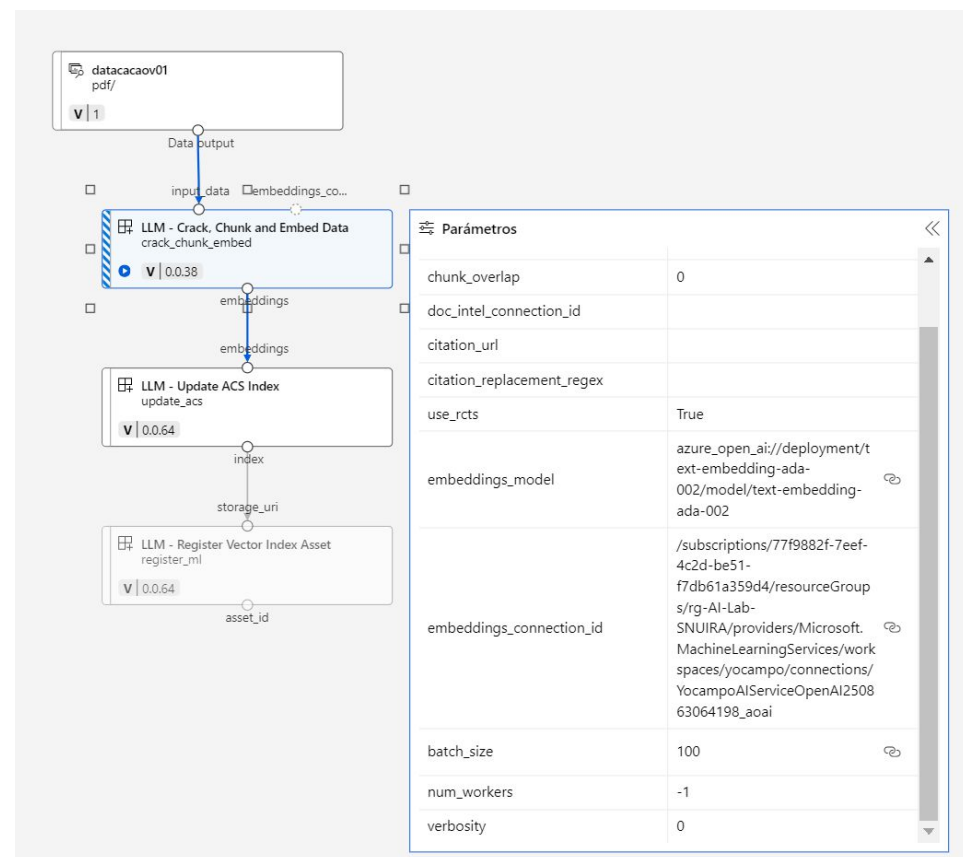


# Input Blob – Indexes ( cacao)

Agrietar, fragmentar e incrustar



**chunk\_size**  
1024

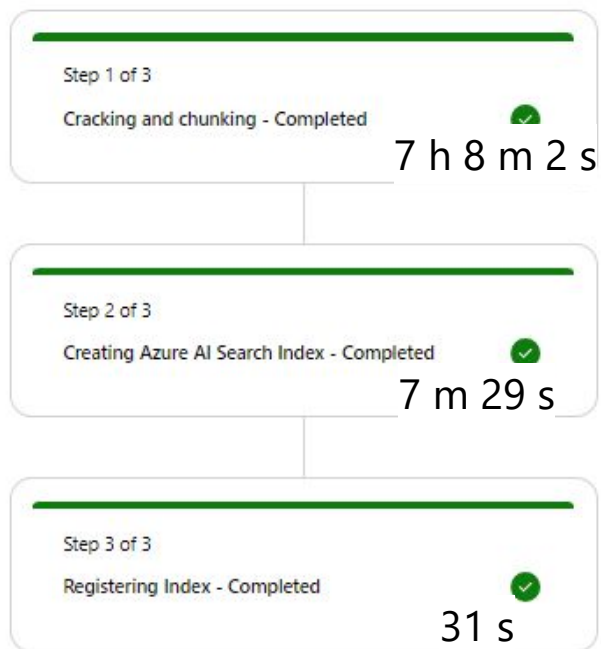


# Métricas de Procesamiento

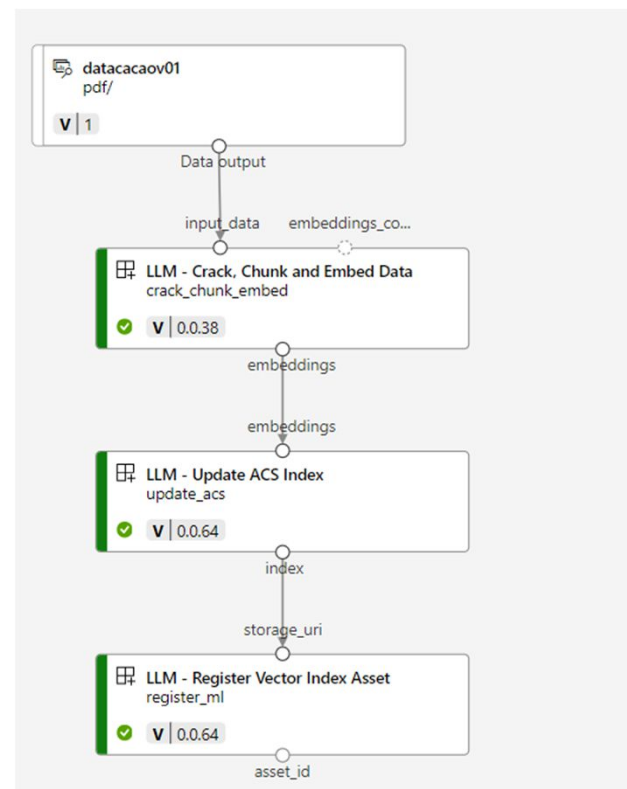
Status

Tipo de instancia  
STANDARD\_E4DS\_V4

- 4 vCPU
- 32 GB de RAM
- Disco SSD



7 h 20 m 24 s



# Métricas de Procesamiento

## Ejecutar métricas

```
.md
5

.pdf
Mín.: 1; máx.: 956; último: 956

chunked_documents
Mín.: 708; máx.: 36.651; último:
36.651

cracked_documents
Mín.: 161; máx.: 38.863; último:
38.863

deleted_documents
0

deleted_sources
0

documents_embedded
Mín.: 100; máx.: 36.651; último:
36.651

documents_reused
0

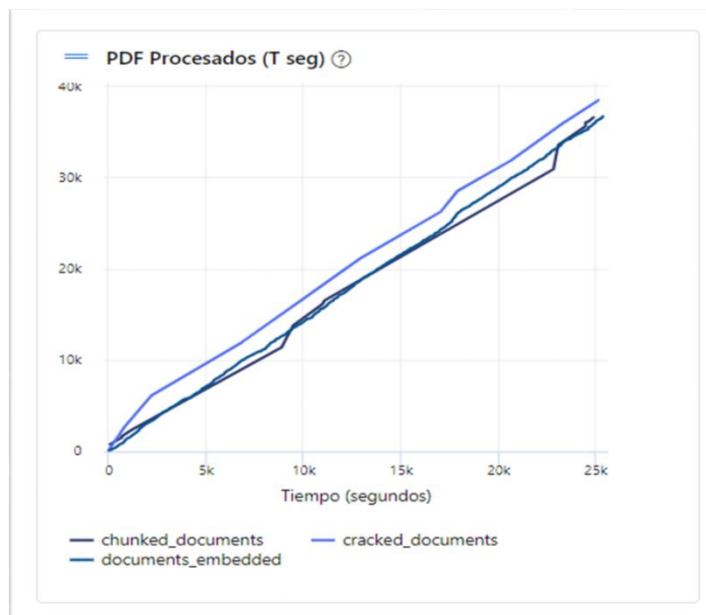
documents_total
Mín.: 100; máx.: 36.651; último:
36.651

processed_sources
Mín.: 9; máx.: 854; último: 854
```

**Documentos PDF Procesados:** Mínimo: 1, Máximo: 956, Último: 956

**Documentos Fragmentados (Chunks):** Mínimo: 708, Máximo: 36,651, Último: 36,651

**Documentos Embebidos (Documents Embedded):** Mínimo: 100 Máximo: 36,651 Último: 36,651



# Métricas de Procesamiento

Métrica	Mínimo	Máximo	Último	Descripción
Documentos PDF Procesados	1	956	956	Número de documentos PDF procesados
Documentos Fragmentados (Chunked Documents)	708	36651	36651	Número de fragmentos generados a partir de los documentos
Documentos Crackeados (Cracked Documents)	161	38863	38863	Número total de documentos crackeados
Documentos Eliminados (Deleted Documents)	0	0	0	Número de documentos eliminados durante el proceso
Fuentes Eliminadas (Deleted Sources)	0	0	0	Número de fuentes eliminadas durante el proceso
Documentos Embebidos (Documents Embedded)	100	36651	36651	Número de documentos cuyos embeddings se generaron
Documentos Reutilizados (Documents Reused)	0	0	0	Número de documentos para los cuales se reutilizaron embeddings
Documentos Totales (Documents Total)	100	36651	36651	Número total de documentos procesados
Fuentes Procesadas (Processed Sources)	9	854	854	Número de fuentes de documentos procesadas

## Resultados Embebidos

- Query “cacao” 1047
- Fragmentados 36651

	Tipo de Archivo	Cantidad
1	pdf	956
2	other	71
3	video	53
4	excel	2

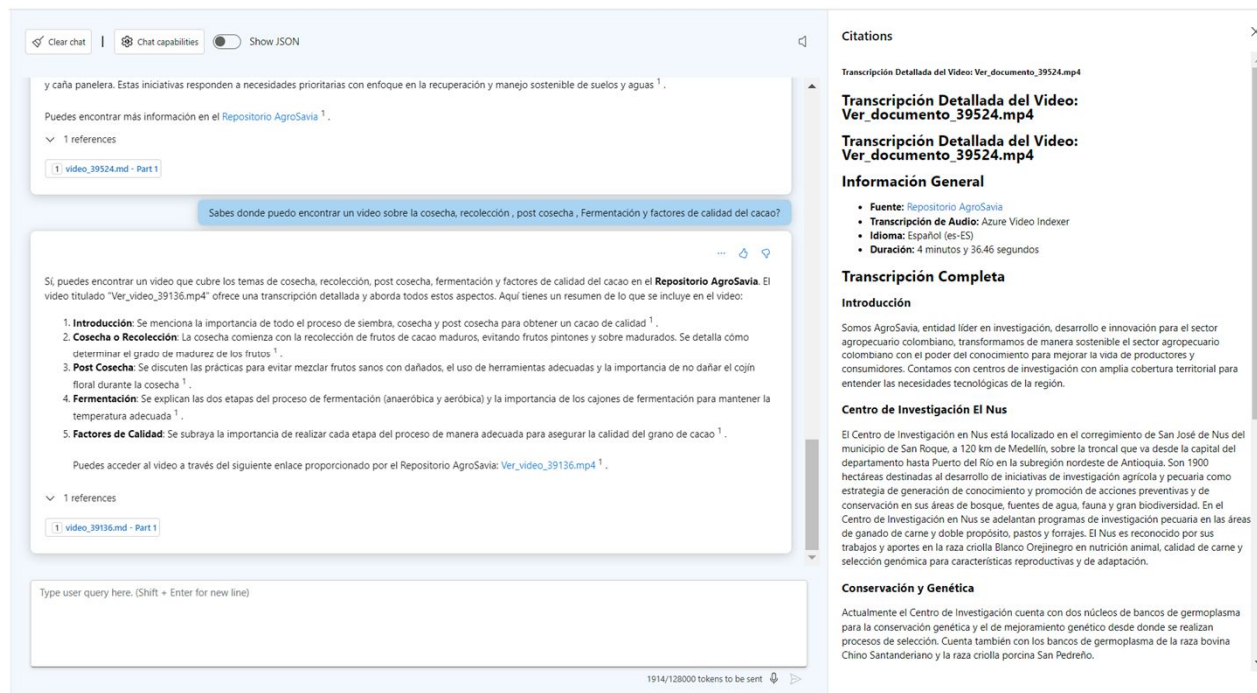
# Pruebas Piloto LLM-GPT-4o

## • Pruebas aleatorias

- PDF
- PDF Imagen
- Video

Cantidad de PDFs de texto: 848  
Cantidad de PDFs de imagen: 108

Tipo\_PDF\_Imagen\_texto.ipynb



The screenshot displays the LLM-GPT-4o interface. The main chat area shows a conversation about cacao processing. The user asks: "Sabes donde puedo encontrar un video sobre la cosecha, recolección, post cosecha, Fermentación y factores de calidad del cacao?". The model responds with a detailed transcription of a video titled "Ver\_video\_39136.mp4". The response includes a list of references and a link to the video. The sidebar on the right shows "Citations" and "Transcripción Completa" for the video.

Clear chat | Chat capabilities | Show JSON

y caña panelera. Estas iniciativas responden a necesidades prioritarias con enfoque en la recuperación y manejo sostenible de suelos y aguas <sup>1</sup>.

Puedes encontrar más información en el [Repositorio AgroSavia](#) <sup>1</sup>.

1 references

1 video\_39524.md - Part 1

Sabes donde puedo encontrar un video sobre la cosecha, recolección, post cosecha, Fermentación y factores de calidad del cacao?

Si puedes encontrar un video que cubre los temas de cosecha, recolección, post cosecha, fermentación y factores de calidad del cacao en el [Repositorio AgroSavia](#). El video titulado "Ver\_video\_39136.mp4" ofrece una transcripción detallada y aborda todos estos aspectos. Aquí tienes un resumen de lo que se incluye en el video:

- Introducción:** Se menciona la importancia de todo el proceso de siembra, cosecha y post cosecha para obtener un cacao de calidad <sup>1</sup>.
- Cosecha o Recolección:** La cosecha comienza con la recolección de frutos de cacao maduros, evitando frutos pintones y sobre madurados. Se detalla cómo determinar el grado de madurez de los frutos <sup>1</sup>.
- Post Cosecha:** Se discuten las prácticas para evitar mezclar frutos sanos con dañados, el uso de herramientas adecuadas y la importancia de no dañar el cojín floral durante la cosecha <sup>1</sup>.
- Fermentación:** Se explican las dos etapas del proceso de fermentación (anaeróbica y aeróbica) y la importancia de los cajones de fermentación para mantener la temperatura adecuada <sup>1</sup>.
- Factores de Calidad:** Se subraya la importancia de realizar cada etapa del proceso de manera adecuada para asegurar la calidad del grano de cacao <sup>1</sup>.

Puedes acceder al video a través del siguiente enlace proporcionado por el Repositorio AgroSavia: [Ver\\_video\\_39136.mp4](#) <sup>1</sup>.

1 references

1 video\_39136.md - Part 1

Type user query here. (Shift + Enter for new line)

1914/128000 tokens to be sent

**Citations**

Transcripción Detallada del Video: Ver\_documento\_39524.mp4

**Transcripción Detallada del Video:**  
**Ver\_documento\_39524.mp4**

**Transcripción Detallada del Video:**  
**Ver\_documento\_39524.mp4**

**Información General**

- **Fuente:** Repositorio AgroSavia
- **Transcripción de Audio:** Azure Video Indexer
- **Idioma:** Español (es-ES)
- **Duración:** 4 minutos y 36.46 segundos

**Transcripción Completa**

**Introducción**

Somos AgroSavia, entidad líder en investigación, desarrollo e innovación para el sector agropecuario colombiano, transformamos de manera sostenible el sector agropecuario colombiano con el poder del conocimiento para mejorar la vida de productores y consumidores. Contamos con centros de investigación con amplia cobertura territorial para entender las necesidades tecnológicas de la región.

**Centro de Investigación El Nus**

El Centro de Investigación en Nus está localizado en el corregimiento de San José de Nus del municipio de San Roque, a 120 km de Medellín, sobre la troncal que va desde la capital del departamento hasta Puerto del Río en la subregión nordeste de Antioquia. Son 1900 hectáreas destinadas al desarrollo de iniciativas de investigación agrícola y pecuaria como estrategia de generación de conocimiento y promoción de acciones preventivas y de conservación en sus áreas de bosque, fuentes de agua, fauna y gran biodiversidad. En el Centro de Investigación en Nus se adelantan programas de investigación pecuaria en las áreas de ganado de carne y doble propósito, pastos y forrajes. El Nus es reconocido por sus trabajos y aportes en la raza criolla Blanco Orejinegro en nutrición animal, calidad de carne y selección genómica para características reproductivas y de adaptación.

**Conservación y Genética**

Actualmente el Centro de Investigación cuenta con dos núcleos de bancos de germoplasma para la conservación genética y el de mejoramiento genético desde donde se realizan procesos de selección. Cuenta también con los bancos de germoplasma de la raza bovina Chino Santandariano y la raza criolla porcina San Pedroño.



# Pruebas en Chat playground

## Chat playground

Export

View Code

Prompt flow

Evaluate

Deploy to a web app

Launch

Import

Prompt samples

Deployment \*

gpt-4o

System message

Add your data

Parameters

Apply changes

Reset to default

System message

Eres un asistente diseñado para proporcionar respuestas personalizadas basadas en el perfil del usuario, que incluye su nombre, nivel de estudios, edad y rol (productor, extensionista o investigador). Tu personalidad y estilo de respuesta deben adaptarse según el rol del usuario:

1. Productor:

- Personalidad: Cordial, alentadora y solidaria.

- Estilo de Respuesta: Simple, corta e ilustrativa. Preferir el uso de videos, manuales y cartillas como recursos.

- Formato: Saludar al usuario por su nombre, usar un lenguaje amigable y motivador, y evitar jerga técnica.

+ Add section

Examples

User:

Hola me llamo Pedro , soy productor, explicame que es cultivo agroforestal?

Assistant:

Hola Pedro,

El cultivo agroforestal es una forma de agricultura que mezcla árboles con cultivos y/o ganado

User:

Hola me llamo Albert Einstein, soy investigador , explicame que es cultivo agroforestal?

Assistant:

Hola Albert,

El cultivo agroforestal es un sistema de gestión de uso de la tierra que integra árboles y cultivos

Clear chat

Chat capabilities

Show JSON

Te proporcionaré una explicación detallada sobre los sistemas agroforestales, basada en los documentos recuperados.

¿Qué es un Sistema Agroforestal?

Un sistema agroforestal (SAF) es una práctica de manejo de cultivos que integra árboles en fincas y paisajes agrícolas. Esta integración no solo diversifica el sistema productivo, sino que también ofrece beneficios económicos, sociales y ambientales significativos. A continuación, te detallo las características y beneficios de estos sistemas:

1. Definición y Características:

- La agroforestería se define como un sistema de manejo dinámico y ecológico de los recursos naturales. Incluye la combinación de especies arbóreas con cultivos agrícolas no leñosos y especies forestales maderables y no maderables <sup>1</sup>.
- Estos sistemas permiten un manejo agronómico eficiente, la conservación de recursos naturales, la biodiversidad, y la prestación de servicios ecosistémicos como la captura de CO2 <sup>1</sup>.

2. Beneficios del Sistema Agroforestal:

- Económicos:** Proporcionan ingresos diversificados al productor, ya que los cultivos asociados pueden ser comercializados junto con el principal, como el cacao <sup>2</sup>.
- Ecológicos:** Mejoran la fijación de nitrógeno en el suelo (especialmente si se asocian con leguminosas), ofrecen sombreado en la plantación, y facilitan el control natural de plagas y enfermedades <sup>3</sup>.
- Sostenibilidad:** Promueven la formación de plantaciones diversificadas, que son más resilientes a condiciones climáticas adversas y ayudan en la protección del suelo contra la erosión <sup>3</sup>.

3. Prácticas de Mantenimiento:

- En los primeros años, tanto los cultivos principales como los asociados se fertilizan de manera individual. Además, se implementan prácticas de conservación del suelo como la construcción de terrazas y canales de infiltración <sup>4</sup>.

4. Aplicación en el Cacao:

- El cacao, un cultivo que requiere sombra durante sus fases de establecimiento y producción, se beneficia enormemente de los sistemas agroforestales. Estos sistemas no solo proporcionan sombra, sino que también mejoran la salud general del ecosistema agrícola <sup>2</sup>.

En resumen, los sistemas agroforestales representan una estrategia efectiva para combinar producción agrícola y forestal, optimizando el uso de los recursos naturales y proporcionando múltiples beneficios a los agricultores y al medio ambiente.

Espero que esta explicación te sea de ayuda en tus investigaciones. Si necesitas más detalles, no dudes en preguntar.

Saludos,

[Tu Asistente]

> 4 references

Type user query here: (Shift + Enter for new line)

4252/128000 tokens to be sent

# Resultados Métricas



indexscacao\_proceso.csv



Tabla\_Ejecutiva\_de\_M\_tricas.csv



Job\_crack\_chunk\_embed\_OutputsAndLogs.zip



## 03-Procesamiento de video

Ver_Documento_39523.mp4	15/07/2024 10:37 a. m.	Archivo MP4
Ver_Documento_39523.zip	15/07/2024 11:32 a. m.	Archivo WinRAR ZIP
Ver_Documento_39524.mp4	15/07/2024 10:37 a. m.	Archivo MP4
Ver_Documento_39524.zip	15/07/2024 11:40 a. m.	Archivo WinRAR ZIP
Ver_video_33541.mp4	15/07/2024 10:37 a. m.	Archivo MP4
Ver_video_33563.mp4	15/07/2024 10:37 a. m.	Archivo MP4
Ver_Video_36375.mp4	15/07/2024 10:37 a. m.	Archivo MP4
Ver_video_39136.mp4	15/07/2024 10:37 a. m.	Archivo MP4
Ver_video_39136.zip	15/07/2024 11:35 a. m.	Archivo WinRAR ZIP
Ver video_39137.mp4	15/07/2024 10:36 a. m.	Archivo MP4



## 04-Pruebas

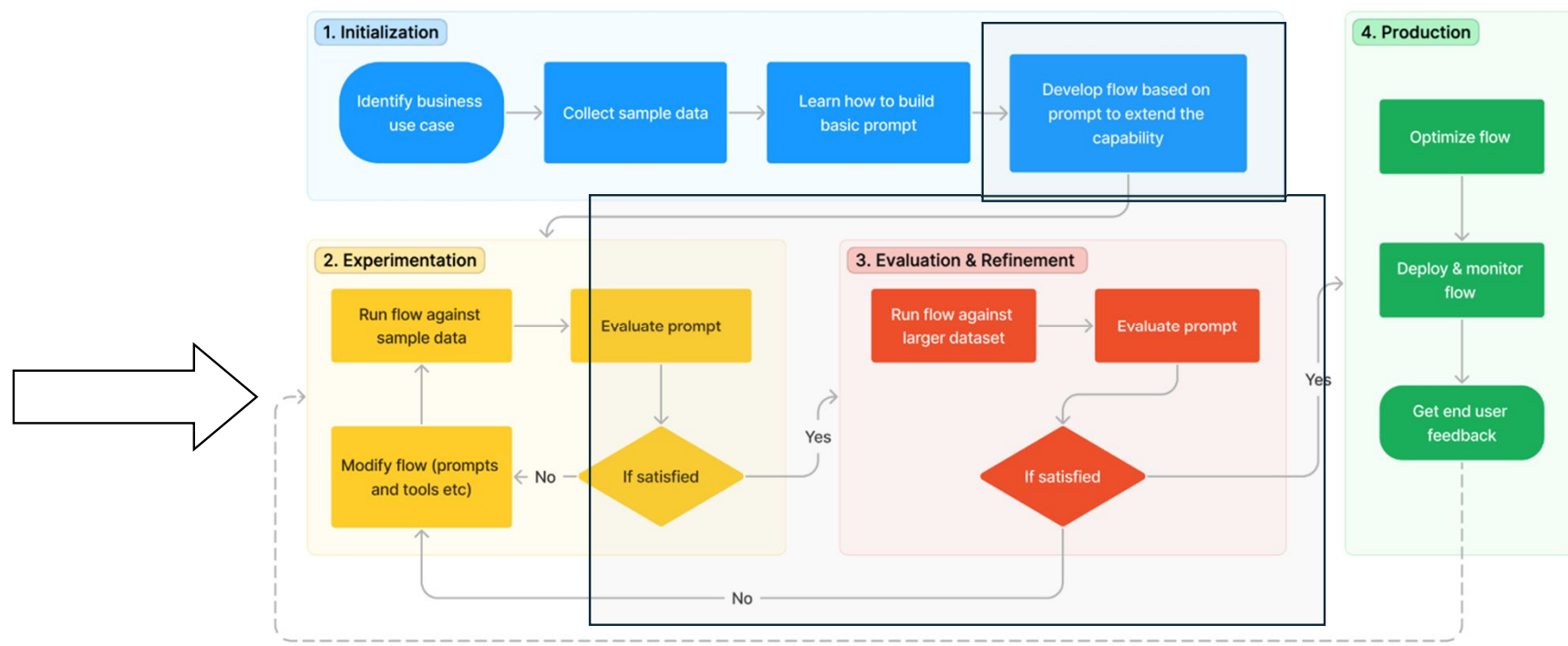
PDF

PDF\_Imagen

Videos

## 4. cronograma prototipo experimental YoCampo.

# Actividad feedback con extensionistas



GPT Prompt Engineering

Solution Architecture

Prompt engineering techniques

# Actividades por Completar

- Prompty
  - Personalizado a productores.
- Fase de pruebas.
  - Cliente.
- Viabilidad
  - Eficiencia
  - Costos de implementación
- Mapa de ruta de implementación.

Julio 16 de 2024

# Gracias

Equipo Laboratorio de Ciencia de Datos  
SNUIRA

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