

Microsoft Fabric: Todo lo que necesitas para una Plataforma de Ciencia de Datos

Antonio Soto
@antoniosql



Sponsors:



Organizan:



#PowerBIDaysBilbao

¡Gracias sponsors!



#PowerBIDaysBilbao

Platinum Sponsor:



- Verne Tech, te ayudamos a **liderar** el camino en la era de los **Datos** y la **Inteligencia Artificial**
- **Ganadores** de 2 Datatones de IndesIA
- Tenemos como meta la **excelencia**
- Si tu objetivo es **aprender** tecnología, te **ayudamos** en **Verne Academy**, somos Microsoft Learning Partner
- Mas de +200 Profesionales, bien formados y motivados
- ¿Necesitas más información? 670 25 37 34 / marketing@verneit.com



Platinum Sponsor:



- Fours9s, la mejor herramienta de **observabilidad SQL Server** del mercado.
- Gracias a la información detallada **segundo a segundo** de la actividad de tu sistema, podrás **analizar y optimizar** tus procesos de base de datos con funcionalidades únicas.
- Somos la **solución** que dará respuesta a la **problemática** de tus bases de datos y te ayudaré a mantener tus **servidores Sql Server** en óptimo estado.
- ¿Quieres verlo **en acción**? Contacta con nosotros y te realizaremos una demo.
- info@four9s.es / 688 688 797



Gracias a los Sponsors y Colaboradores

Quiz Final + Regalos + Networking (18:15h)

Quédate al divertido Quiz
para participar en el
Sorteo
(sólo entre asistentes)

Y más ...!



Power BI Days

VOL. 2

- 📍 Alicante 17 Junio '23
 - 📍 Mallorca 23 Sept '23
 - 📍 Málaga 3 Feb '24
 - 📍 **Bilbao 15 Junio '24**
 - 📍 ANUNCIO HOY Oct '24
 - 📍 Marzo'25
 - 📍 FIN – Junio'25
-  **BAT**
B Accelerator Tower

@Pow
Tu canal Telex



Power BI Day Bilbao

VOL. 2

Antonio Soto

- Después de más de 25 años gestionando sistemas de información, principalmente en entornos de Microsoft, con especial atención a los sistemas de Business Intelligence y la gestión de datos, ahora estoy más centrado en la gestión de equipos y ventas. Siempre buscando nuevos retos para ayudar a nuestros clientes a optimizar sus inversiones en su plataforma de datos. En este momento se centra en Big data, Smart data y tecnologías IA con el fin de definir nuevas líneas de negocio



<https://www.linkedin.com/in/antoniosql/>



<https://twitter.com/antoniosql>



<https://medium.com/@antoniosql>

#PowerBIDaysBilbao



Objetivos

- Conocer las opciones de ML e IA que tenemos en Fabric
- Entender el proceso de Desarrollo de soluciones

Agenda

- Ciclo de Desarrollo y Componentes
- Servicios de IA incluidos
- AutoML

Ciclo de Desarrollo Data Science



Componentes

Data Source



Lakehouse &
Data Warehouse

Clean & prepare



Notebooks

Store



Lakehouse(s)

Expose



PBI



Lake Warehouse

Train test &
apply

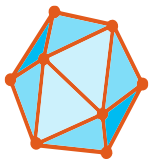


Models



Experiments

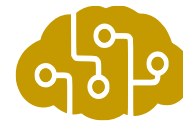
Componentes II: Synapse ML y AI



Entrenamiento
distribuido de
Modelos



Soporte MLflow



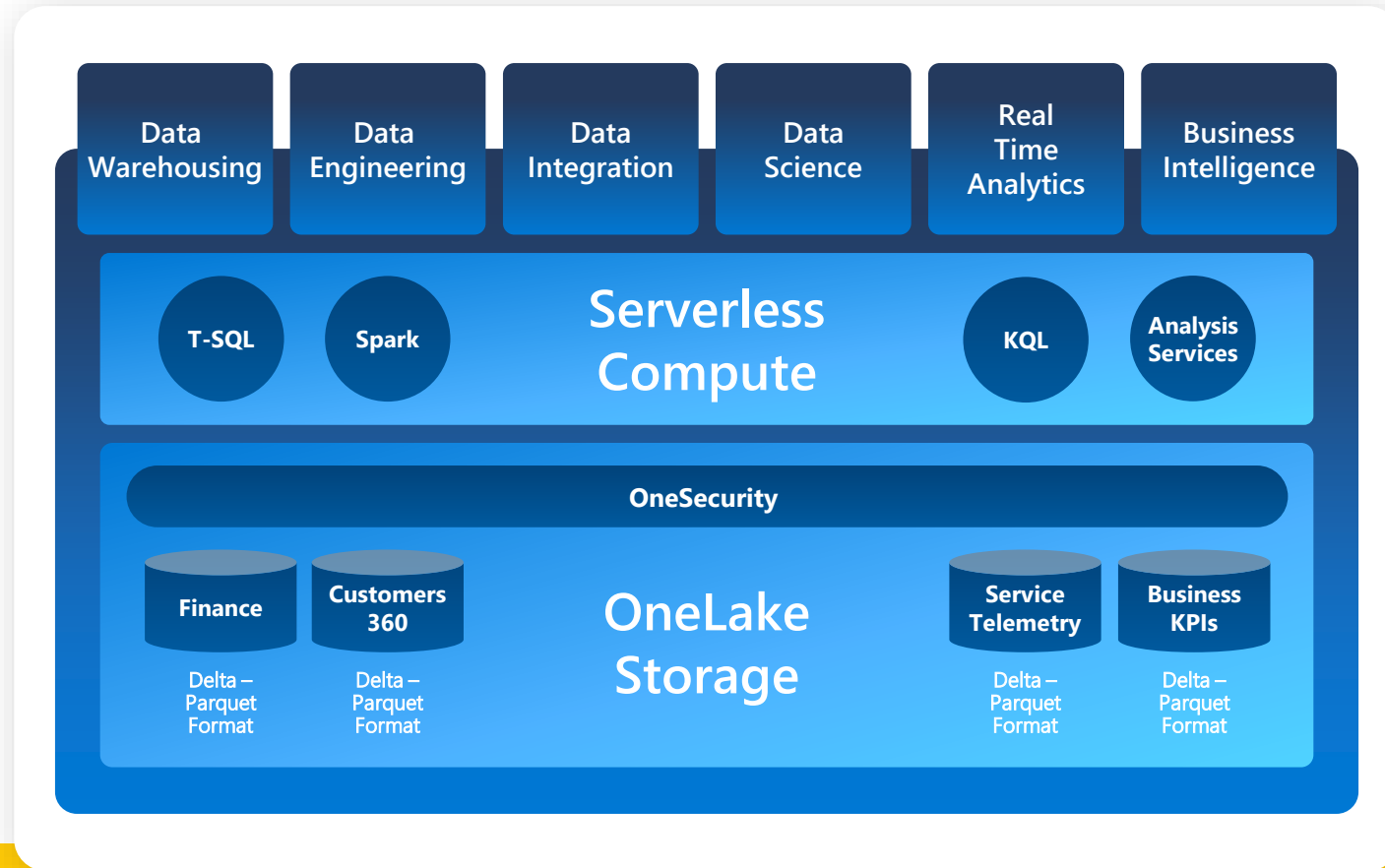
Cognitive
Services



OpenAI
LLMs



Descubrimiento de Datos



Preparación de Datos

Notebooks

Data Wrangler

The screenshot shows a Databricks notebook interface. The title bar indicates 'Demo Churn clientes' and 'Ninguna etiqueta'. The notebook title is 'Demo de Análisis de Abandono de Clientes'. The content includes a description in Spanish and three code cells executed in PySpark (Python). The first cell installs 'imblearn'. The second cell imports 'pandas', 'numpy', and 'pd'. The third cell loads a CSV file from AWS and creates a 'df_transactions' dataframe. The status bar at the bottom shows 'Session ready' and 'Save option: Automatic'.

Demo Churn clientes | Ninguna etiqueta | Guardado

Prueba de Fabric: Queda 59 días

Home Edit Run Data View

Save options Automatic

Search: Buscar

Editing Comments Share

Clear all output(s) Find and replace

Lakehouse explorer lakehousedemo

Tables Files aws-dunnhumby synthea

Demo de Análisis de Abandono de Clientes

En este notebook veremos primero como acceder a datos que están realmente alojados en AWS, en formato CSV, para cargarlos en un dataframe de pandas, limpiar y cruzar esos datos, y posteriormente entrenar un modelo de clasificación para detectar ese abandono. Veremos como utilizar Data Wrangler para analizar esos datos y poder limpiarlos, así como la integración con MIFlow para gestionar el ciclo de experimento --> modelo

```
1 #Instalamos las librerías necesarias
2
3 !pip install imblearn
```

[12] ✓ -Command executed in 8 sec 378 ms by Verne Dev Admin on 12:33:45 PM, 9/12/23

PySpark (Python)

Output is hidden

```
1 #Importamos las librerías necesarias
2 import pandas as pd
3 import numpy as np
```

[1] ✓ -Apache Spark session started in 11 sec 62 ms. Command executed in 2 sec 215 ms by Verne Dev Admin on 12:30:21 PM, 9/12/23

PySpark (Python)

```
1 #Cargamos con pandas los datos que están en AWS a través del enlace
2 df_transactions = pd.read_csv("/lakehouse/default/Files/aws-dunnhumby/transaction_data.csv")
3
```

[3] ✓ -Command executed in 8 sec 222 ms by Verne Dev Admin on 12:31:39 PM, 9/12/23

PySpark (Python)

```
1 #Calculamos el ticket medio de compra
2
3 df_transactions.groupby(by="BASKET_ID").agg({'SALES_VALUE': 'sum'}).sort_values(by="BASKET_ID", ascending=False)
```

Selected 1 of 17 cells

Session ready Save option: Automatic

The screenshot shows the Databricks Data Wrangler interface for a dataframe named 'df_transactions'. The top bar shows 'Demo Churn clientes' and 'Ninguna etiqueta'. The left sidebar lists operations: Find and replace (4), Format (7), Formulas (4), Numeric (4), Schema (5), Sort and filter (2), Custom operation, Group by and aggregate, and New column by example. The main area displays a table with columns: index, household_key, BASKET_ID, DAY, and PRODUCT_ID. Each column has a histogram and summary statistics. The right sidebar shows a 'Summary' panel with 'Data shape: 5,000 rows x 12 columns', 'Columns: 12', 'Rows (5,000)', 'with missing values: 0 (0.0%)', 'duplicate rows: 0 (0.0%)', and 'Missing Values (0)'. The bottom status bar shows 'Data frame loaded' and 'Showing the first 5000 rows'.

Demo Churn clientes | Ninguna etiqueta | Guardado

Prueba de Fabric: Queda 59 días

df_transactions (Data Wrangler)

+ Add code to notebook Copy code to clipboard Save as CSV

Search for operations...

Operations

- Find and replace (4)
- Format (7)
- Formulas (4)
- Numeric (4)
- Schema (5)
- Sort and filter (2)
- Custom operation
- Group by and aggregate
- New column by example

Cleaning steps

- Loaded variable 'df_transactions' from kernel sta...
- New operation

Preview code for all steps

2 New operation

1 Choose an operation to get started

Or, start typing code to see a live preview of the transformation on your data (e.g. df_transactions = df_transactions.drop(columns=['household_key']))

Choose an operation to generate preview

Apply Discard

Data shape 5,000 rows x 12 columns

Columns 12

Rows (5,000)

with missing values 0 (0.0%)

duplicate rows 0 (0.0%)

Missing Values (0)

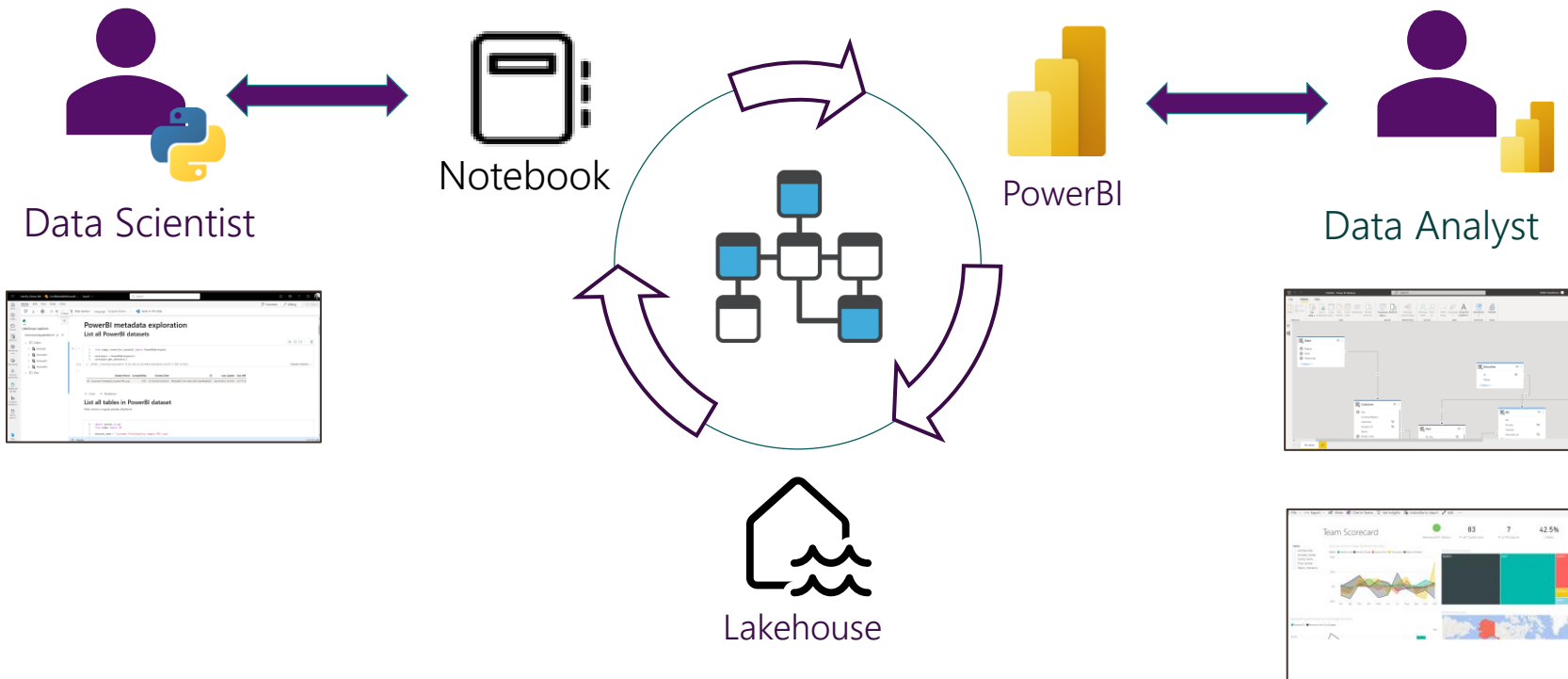
No missing values

Data frame loaded Showing the first 5000 rows

Kernel ready pandas 1.5.3

Preparación de Datos

Semantic Link



Experimentos y Modelos con MLFlow

The screenshot displays the MLFlow web interface for an experiment named 'churn-dunnhumby'. The interface includes a top navigation bar with a search bar, a 'Prueba de Fabrica' status, and user profile icons. Below the navigation bar, there are tabs for 'Inicio' and 'Ver'. A toolbar contains actions like 'Guardar como modelo', 'Descargar archivos de ejecución', and 'Eliminar ejecución'. The left sidebar shows a file explorer for the experiment, listing files like 'MLmodel', 'conda.yaml', 'model.pkl', 'python_env.yaml', and 'requirements.txt'. The main content area displays the 'Propiedades' (Properties) of the experiment, including a description, execution name, start date, duration, status, and execution ID. Below the properties, there is a table of metrics and a section for execution parameters and schemas.

Propiedades

Descripción

Nombre de ejecución	Fecha de inicio	Duración	Estado	# Id. de ejecución
shy_panda_7w...	9/12/2023 12:34 PM	10s	Completed	531660e...
Creado por	Origen	Nombre del experimento	Versiones del modelo	
Verne Dev Ad...	Demo Churn client...	churn-dunnhu...	demo-churn-dunnhumby: Versión 1	

Comparar ejecuciones

Compare los detalles de las ejecuciones en una lista. Seleccione ciertas ejecuciones para comparar las métricas de ejecución visualmente. Más información

Ver lista de ejecución

Nombre	Valor
Ejecución de métricas (2)	
score_test	0.9723618090452262
score_training	0.9765100671140939
Parámetros de ejecución (0)	
Esquema de entrada (0)	
Esquema de salida (0)	

Demo time!

Servicios de IA incluidos

Servicios Incluidos

- Azure AI Translate
- Azure AI Language
- Azure Open AI (F64 o superior y no en la trial)

Azure AI Translate

- Traducir Texto
- Transliterar.. Traduce texto entre scripts
- Detecta el Idioma
- Búsquedas y ejemplos de diccionario

Azure AI Language

- Análisis de Sentimiento
- Detección de Idioma
- Extracción de frases clave
- Reconocimiento de Identidades de Identificación Presonal (PII)
- Reconocimiento de Identidades con Nombre (NER)
- "Vinculación" de Entidad. Elimina ambigüedades de la identidad de entidades

Azure OpenAI

- GPT-35-turbo
- text-embedding-ada-002
- text-davinci-003
- code-cushman-002

Cómo podemos utilizarlo

- Utilizando API Rest de cada servicio
- Utilizando SynapseML
- Para Azure OpenAI también con el SDK de Python de OpenAI

Demo time!

AutoML

AutoML

FLAML

A Fast Library for Automated Machine Learning & Tuning

FLAML Getting Started - 5min ⌚



Find Quality Model at Your Fingertips

FLAML finds accurate models or configurations with low computational resources for common ML/AI tasks. It frees users from selecting models and hyperparameters for training or inference, with smooth customizability.



Adapt Large Language Models to Your Needs

By automatically adapting LLMs to applications, FLAML maximizes the benefits of expensive LLMs and reduce monetary cost. FLAML enables users to build and use intelligent adaptive AI agents with minimal effort.



Tune It Fast, Tune It As You Like

FLAML offers a fast auto tuning tool powered by a novel cost-effective tuning approach. It is capable of handling large search space with heterogeneous evaluation cost and complex constraints/guidance/early stopping.



Demo time!

¿Preguntas?

Microsoft Fabric: Todo lo que necesitas para una Plataforma de Ciencia de Datos

Antonio Soto
@antoniosql



Sponsors:



Organizan:



#PowerBIDaysBilbao