

POWER PLATFORM ESPAÑA - MALLORCA

Power Bi y Millones de registros
con Azure Databrick

Toni Ferrá



POWER PLATFORM

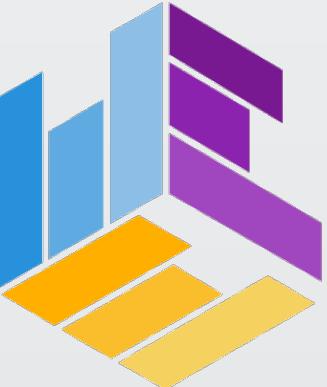


Palma de Mallorca, Spain



KingswaySoft

Toni Ferrá



MASTER DATA 360

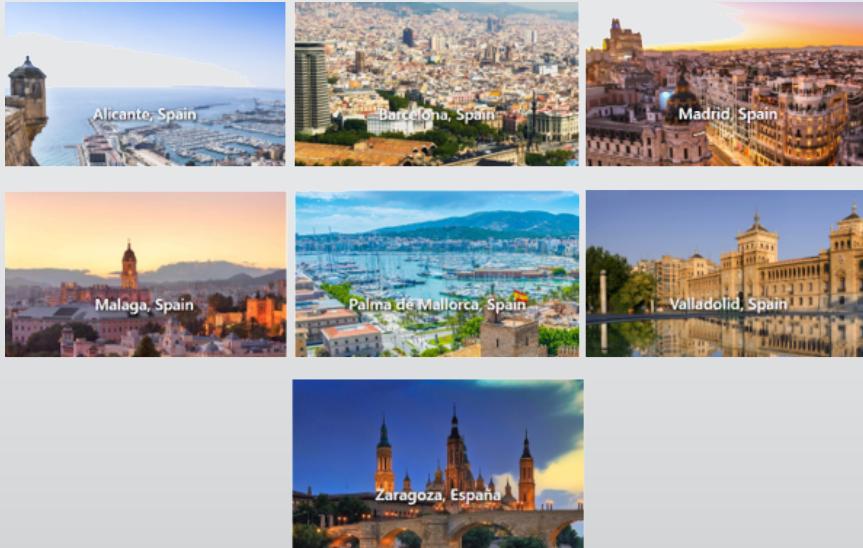
- CTO de Master Data 360:
- Certificado en Power BI en 2018
- Consultor en Azure Data Platform y Power Platform
- Profesor Colaborador de UOC en el Master de Big Data y en el Master de Data Science
- Estudiante eterno

<https://www.linkedin.com/in/antonioferra/>
aferra@mdata360.com



Power Platform España

- Comunidad



Global Power Platform Bootcamp España

Alicante

Barcelona

Madrid

Málaga

Palma de Mallorca

Valladolid

Zaragoza



Pablo Barrachina Pastor
SumamoOs



Ferran Chopo García
IT Consultant and Trainer



Ana María Bisbé York
Microsoft Data Platform
MVP



Pedro Sánchez Naranjo
Technology Team Leader
Bravent



Toni Granell
Dynamics Spain
Microsoft Data Platform
MVP



Yolanda Cuesta Altieri
Experta en Excel
MVP



Juan Ignacio Oller Aznar
IT Instructor, Consultor
MVP



[Return to Dynamic Communities](#)

Visit Our User Group Communities



Explore Conferences and Events



Power Platform Mallorca

@PowerMallorca



#PowerPlatformMallorca

Session Agenda

- Que es Spark
- Que es Azure Databricks
- Arquitecturas
- Conexión Power BI
- Composite Model
- Aggregations



POWER PLATFORM

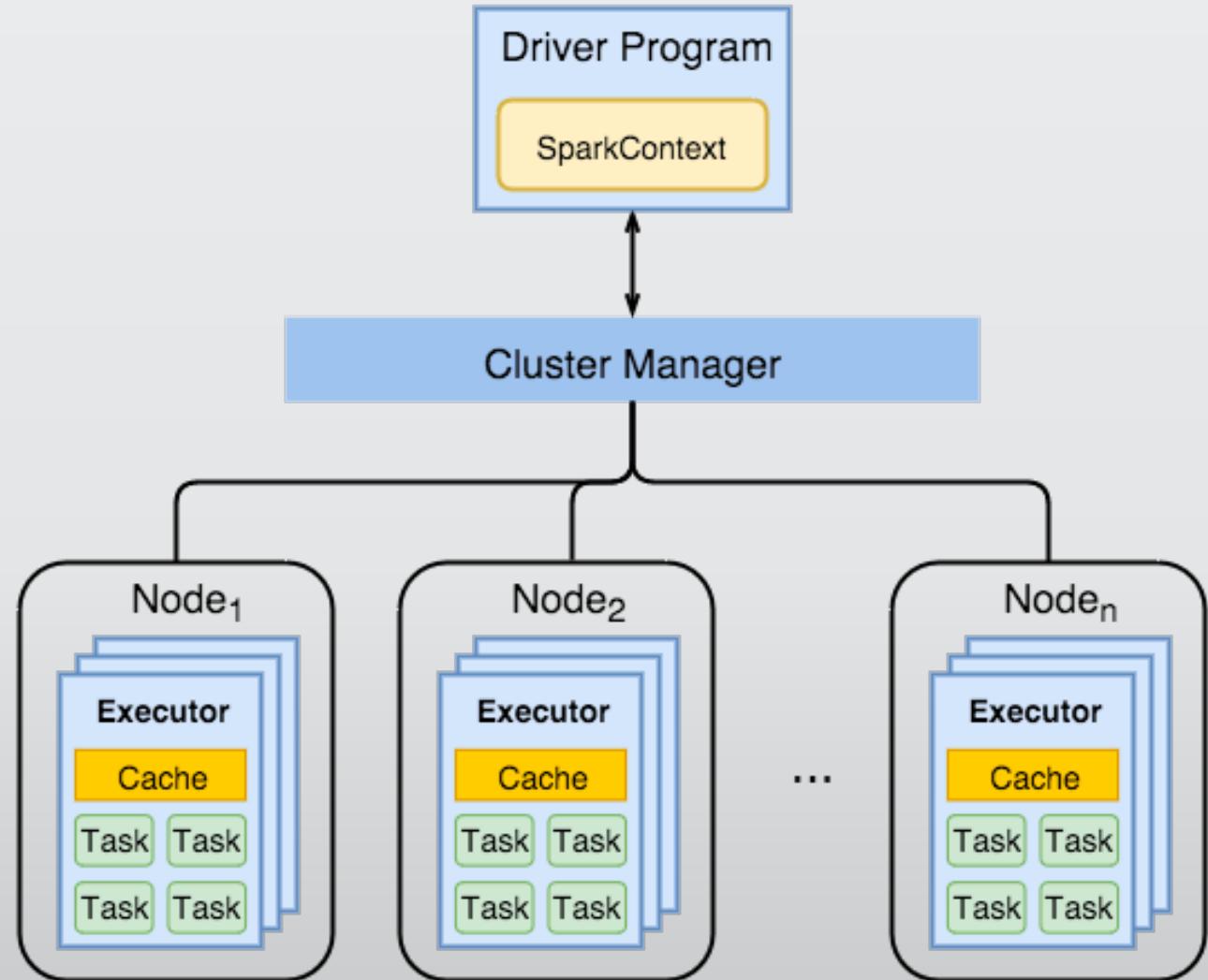
#PowerPlatformMallorca

Cluster Spark

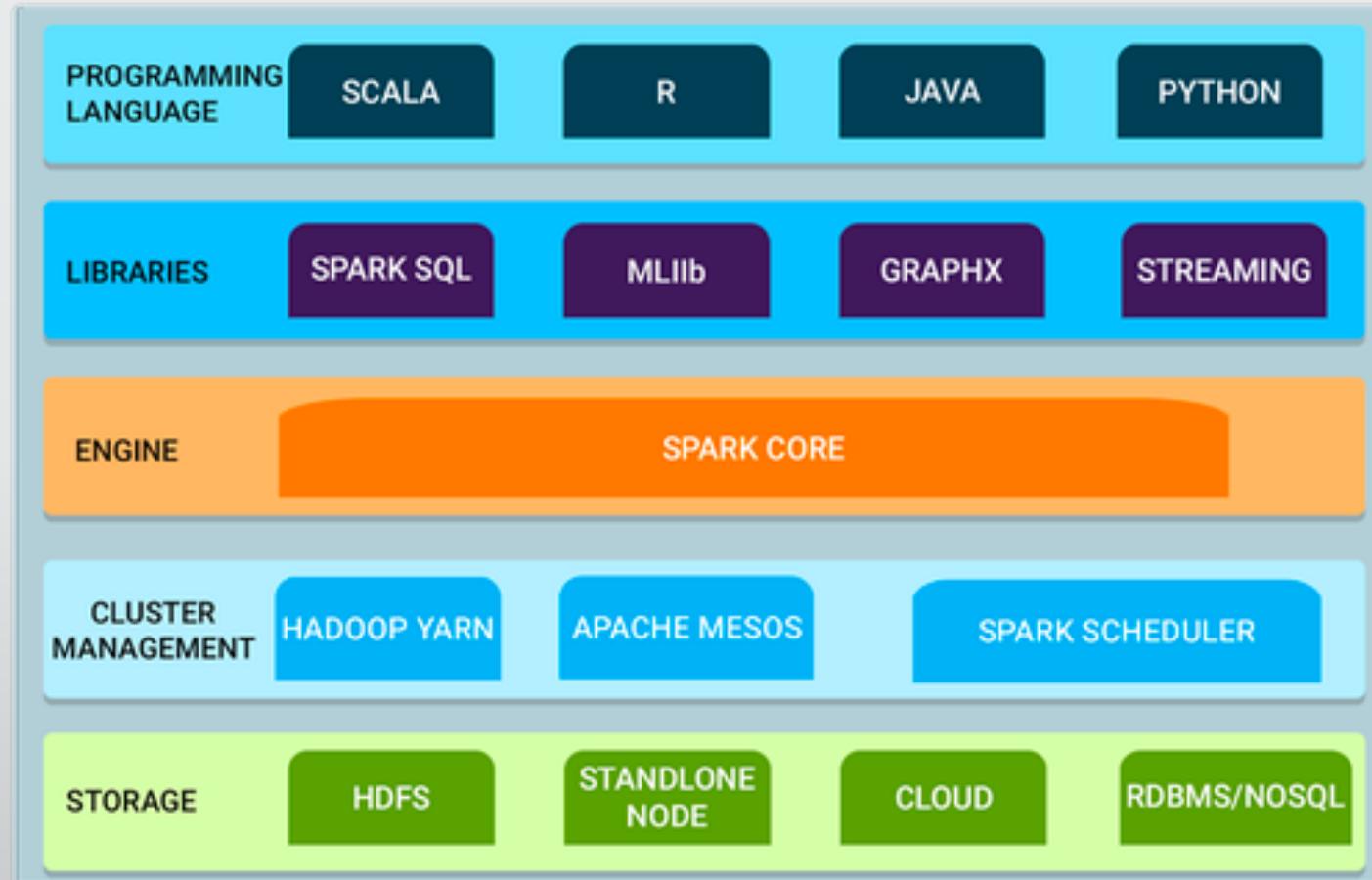
- Alta Disponibilidad
- Alta Concurrency
- Crecimiento horizontal

En el caso de Databricks

- Auto apagado / encendido
- Autoescalado



Ecosistema Spark



POWER PLATFORM

#PowerPlatformMallorca

A Z U R E D A T A B R I C K S

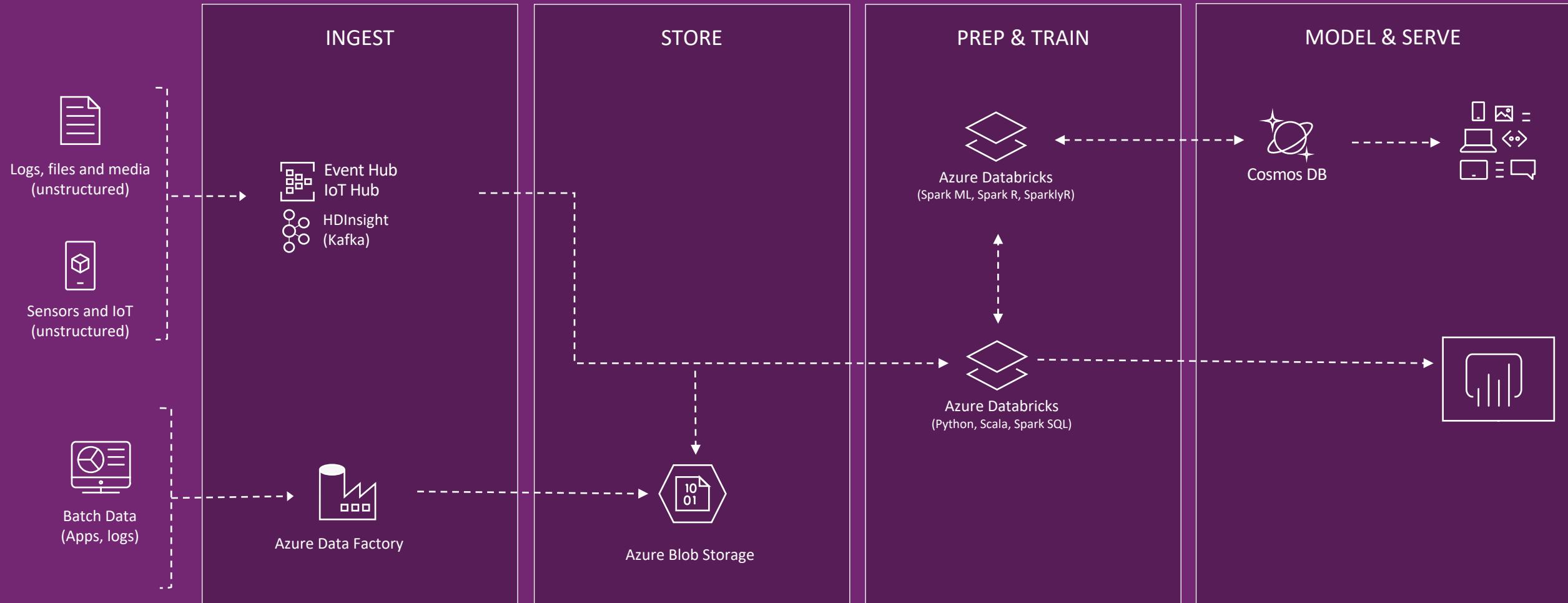
- Azure Databricks is a **first party** service on Azure.
 - Unlike with other clouds, it is **not** an Azure Marketplace or a 3rd party hosted service.
- Azure Databricks is integrated seamlessly with Azure services:
 - Azure Portal: Service can be launched directly from Azure Portal
 - Azure Storage Services: Directly access data in Azure Blob Storage and Azure Data Lake Store
 - Azure Active Directory: For user authentication, eliminating the need to maintain two separate sets of users in Databricks and Azure.
 - Azure SQL DW and Azure Cosmos DB: Enables you to combine structured and unstructured data for analytics
 - Apache Kafka for HDInsight: Enables you to use Kafka as a streaming data source or sink
 - Azure Billing: You get a single bill from Azure
 - Azure Power BI: For rich data visualization



Microsoft Azure

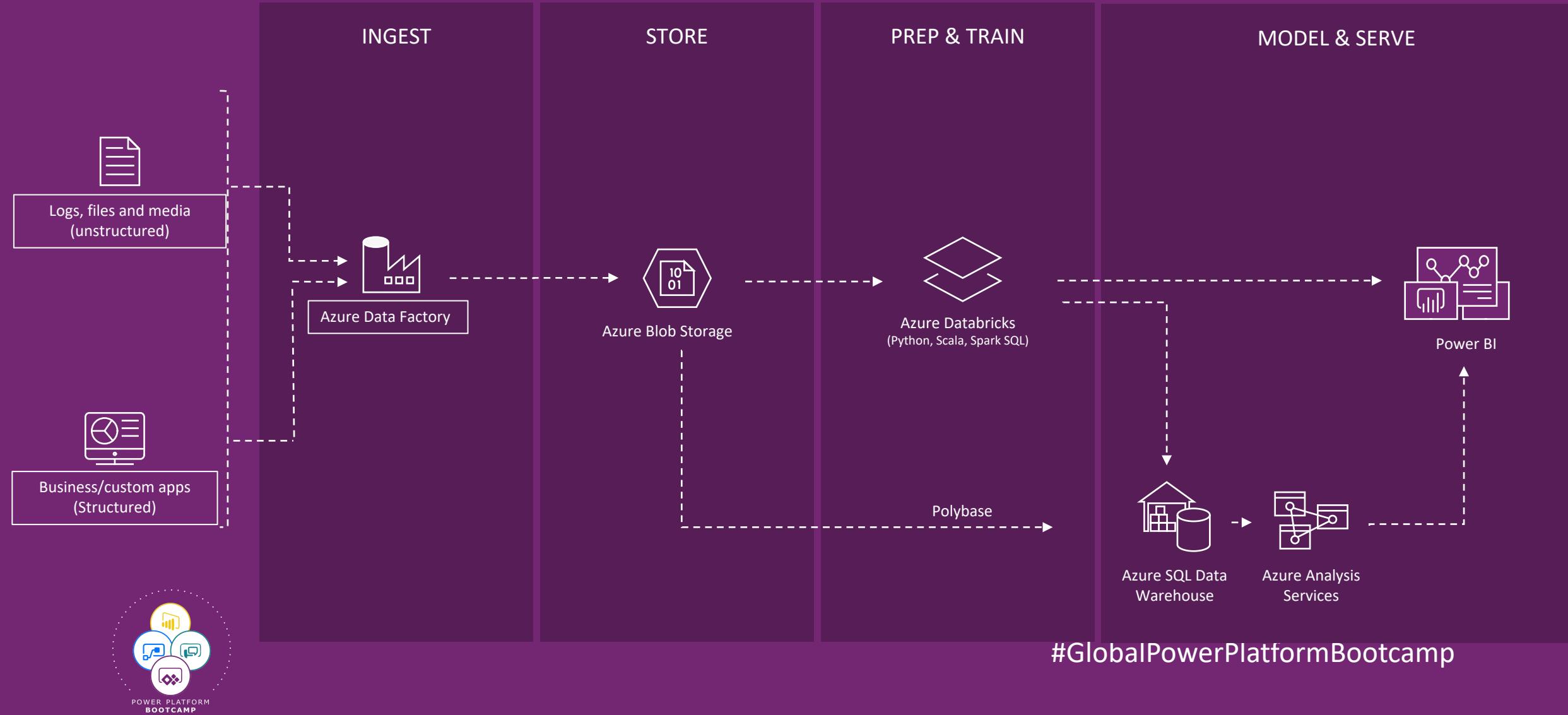
#GlobalPowerPlatformBootcamp

Real time analytics

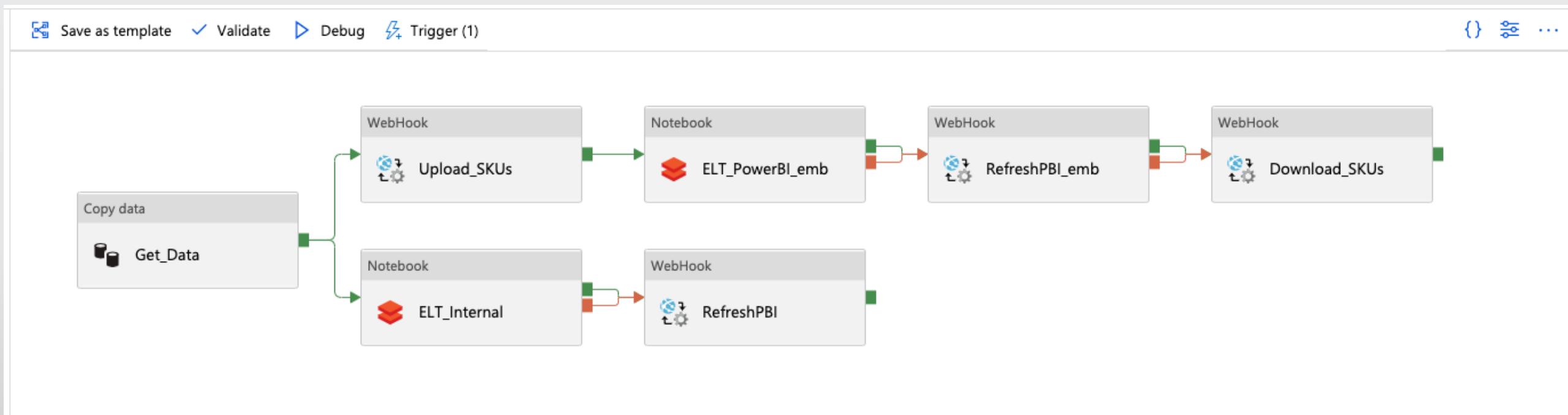


#GlobalPowerPlatformBootcamp

Modern Data Warehouse



Ejemplo ELT



POWER PLATFORM

#PowerPlatformMallorca

Conexión con Azure DataBricks

- Crear el Cluster
- Crear un Token de autentificación
- Cadena de conexión

The screenshot shows the Microsoft Azure Databricks Cluster configuration interface. On the left, there is a sidebar with icons for Azure Databricks, Home, Workspace, Recents, Data, Clusters (which has a red arrow pointing to it), Jobs, and Search. The main area displays a cluster named "Mdata" with status "Up". It includes tabs for Configuration, Notebooks (1), Libraries, Event Log, Spark UI, Driver Logs, Metrics, Apps, and Spark Cluster UI - Master. A note says "This Runtime version supports only Python 3." Under "Autopilot Options", two checkboxes are checked: "Enable autoscaling" and "Terminate after 45 minutes of inactivity". The "Worker Type" is set to "Standard_DS3_v2" with "1 Min Workers" and "6 Max Workers". The "Driver Type" is also "Standard_DS3_v2" with the same specifications. Below these, there is a section for "Advanced Options" which is expanded, showing options for "Azure Data Lake Storage Credential Passthrough" (disabled) and "Enable credential passthrough for user-level data access". There are tabs for Spark, Tags, Logging, Init Scripts, JDBC/ODBC (which is selected and has a red arrow pointing to it), and Permissions. The "JDBC/ODBC" tab contains fields for "Server Hostname" (northeurope.azuredatabricks.net), "Port" (443), "Protocol" (HTTPS), "HTTP Path" (sql/protocolv1/o/6036174880832815/1004-091837-kiwis897), and a large "JDBC URL" field which is highlighted with a red box and contains the value: `jdbc:spark://northeurope.azuredatabricks.net:443/default;transportMode=http;ssl=1;httpPath=sql/protocolv1/o/6036174880832815/1004-091837-kiwis897;AuthMech=3;UID=token;PWD=<personal-access-token>`. At the bottom of the JDBC URL field, there is a link: "Learn more about connecting your favorite BI tool to Databricks".



POWER PLATFORM

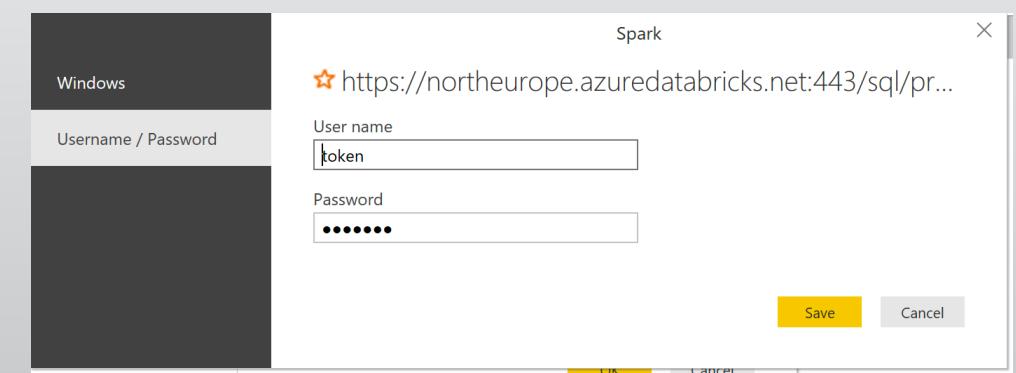
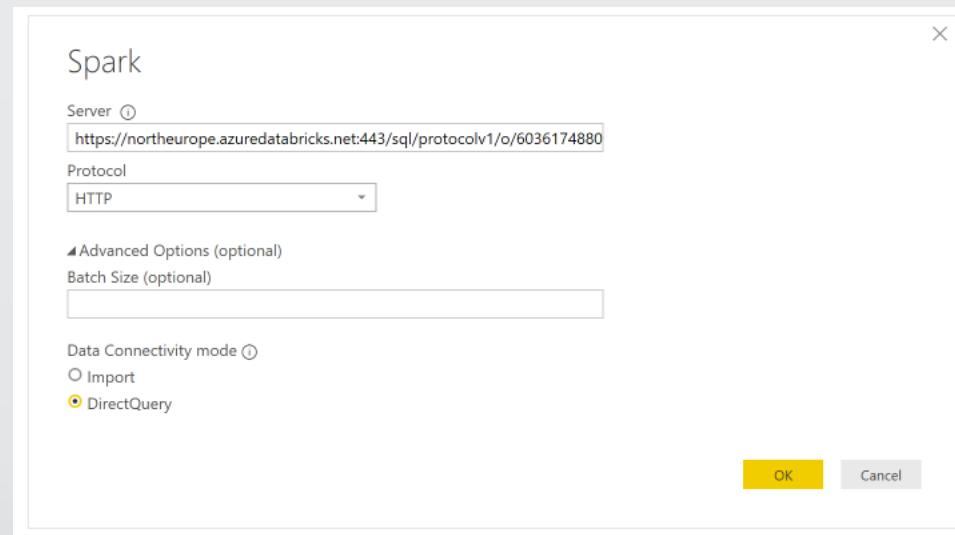
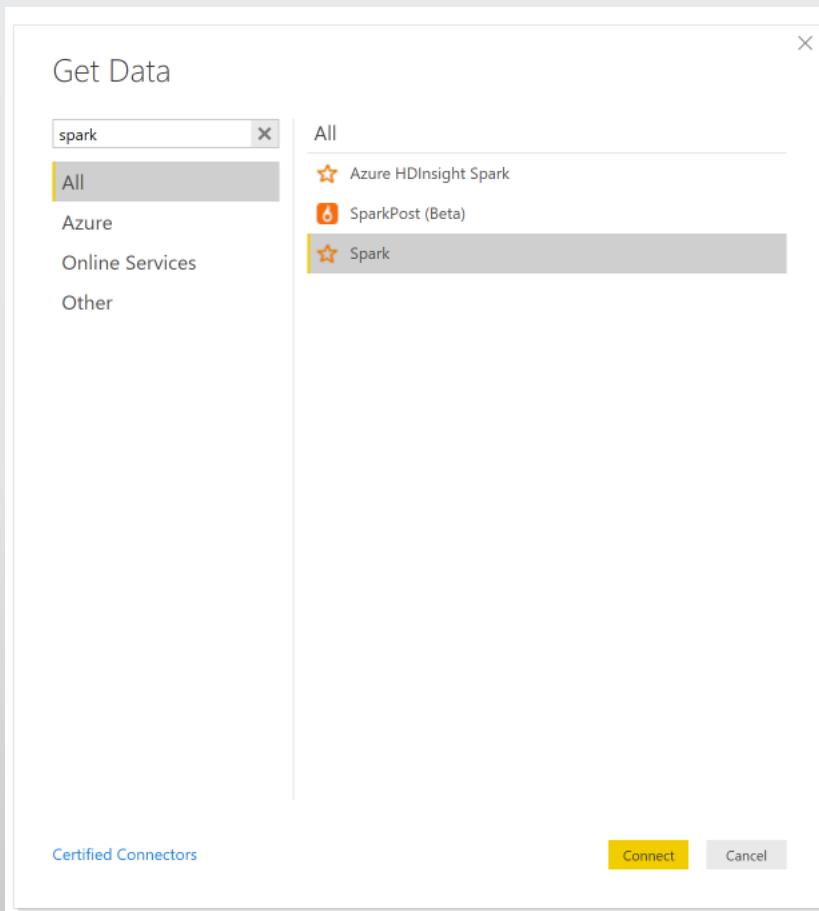
#PowerPlatformMallorca

“Corregir” la cadena de conexión

- ~~jdbc:spark~~
~~https://northeurope.azuredatabricks.net:443/default;trans~~
~~portMode=http;ssl=1;httpPath=sql/protocolv1/o/6036174880832815~~
~~/1004-091837-kiwis897;AuthMech=3;UID=token;PWD=<personal~~
~~access token>~~
- <https://northeurope.azuredatabricks.net:443/sql/protocolv1/o/6036174880832815/1004-091837-kiwis897>



“Corregir” la cadena de conexión



POWER PLATFORM

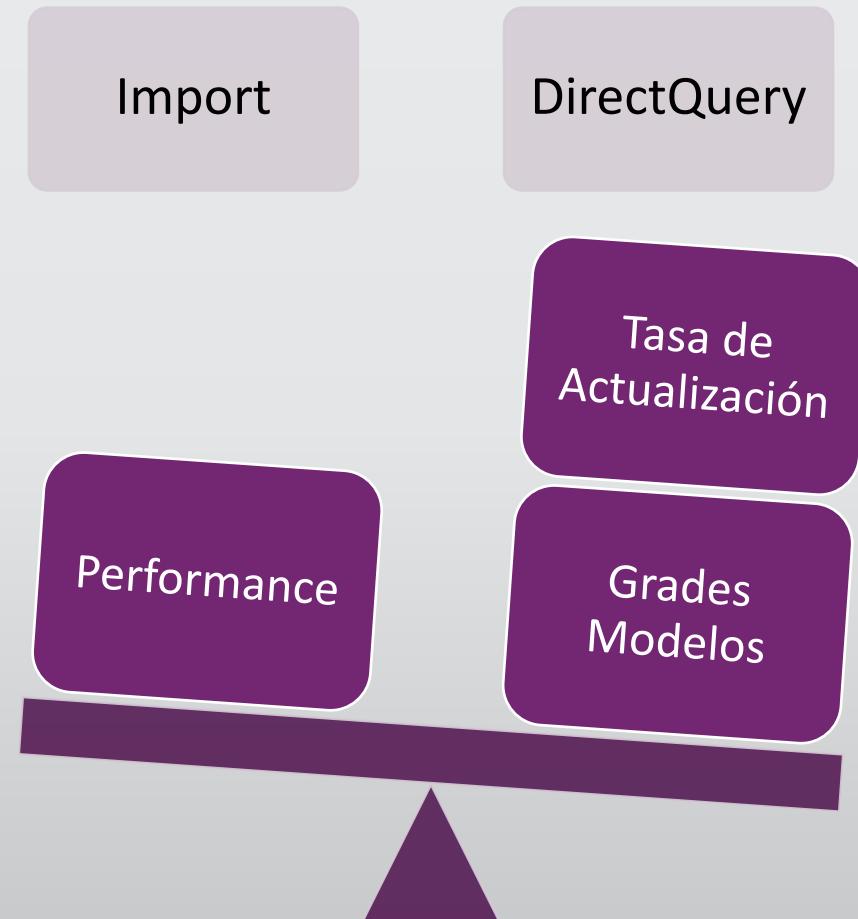
#PowerPlatformMallorca

Demo Conexión Spark

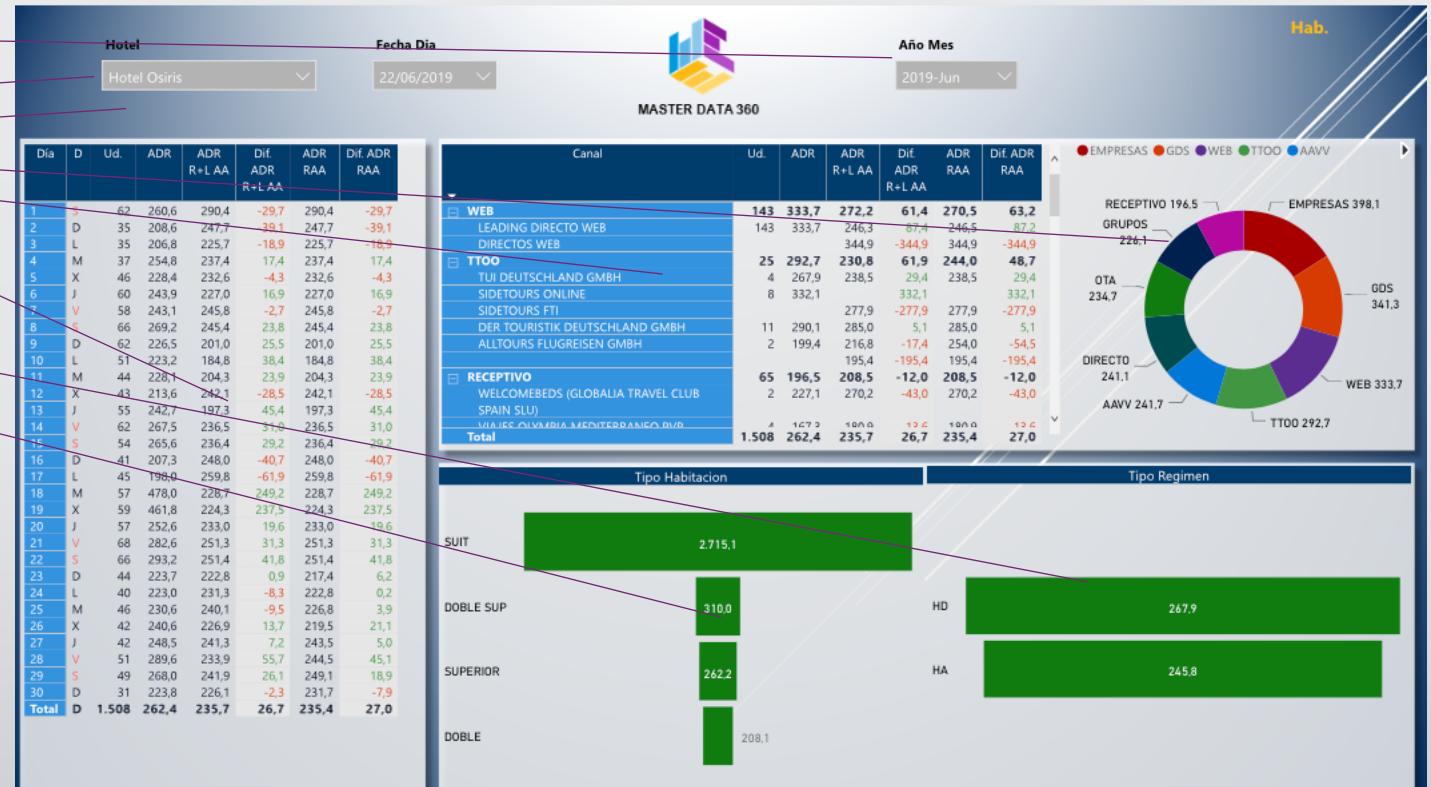


#GlobalPowerPlatformBootcamp

Import vs. DirectQuery



DirectQuery

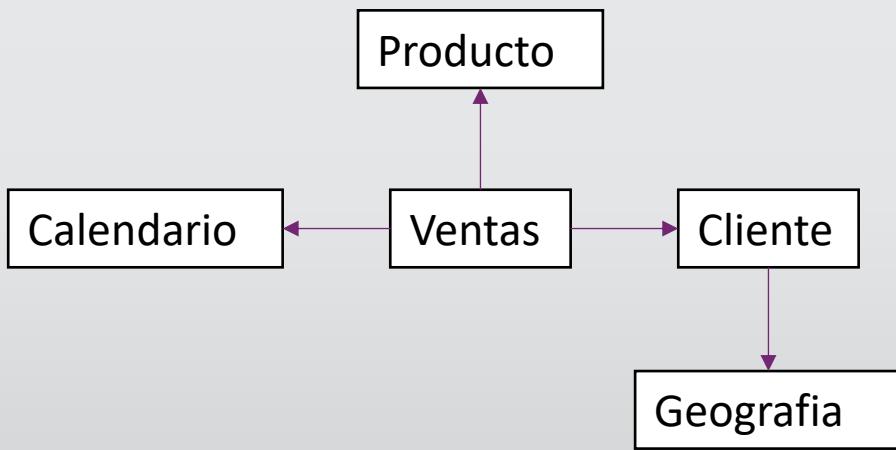


#PowerPlatformMallorca

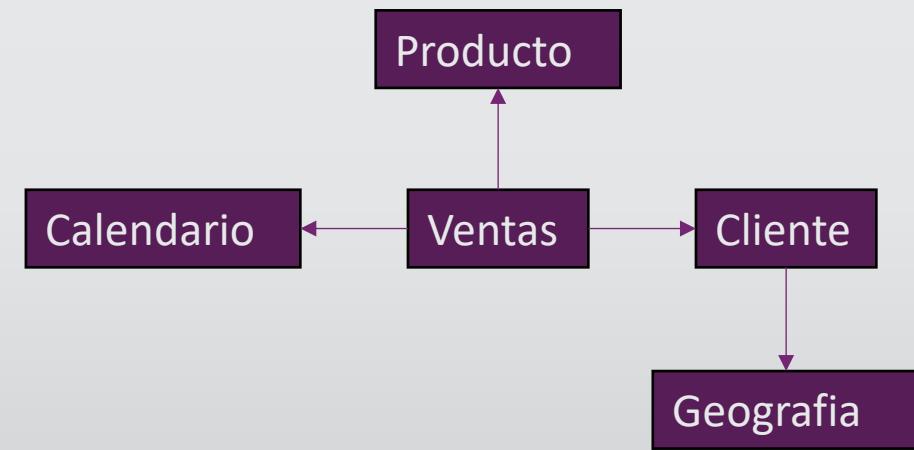


POWER PLATFORM

Import



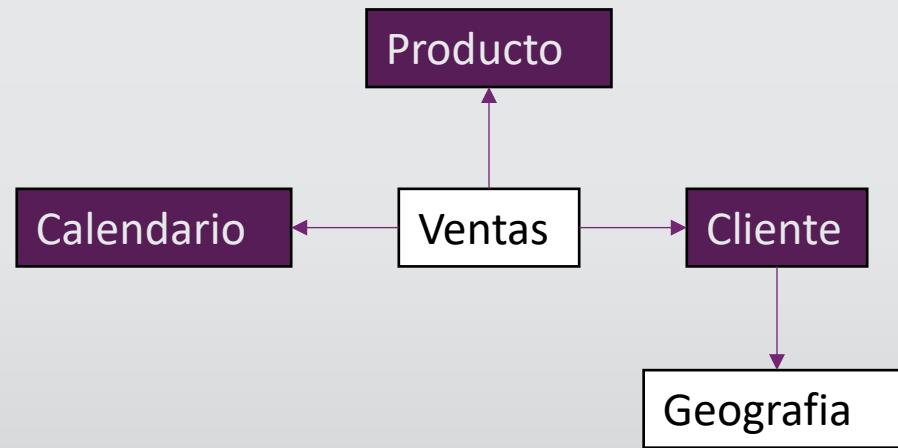
DirectQuery



POWER PLATFORM

#PowerPlatformMallorca

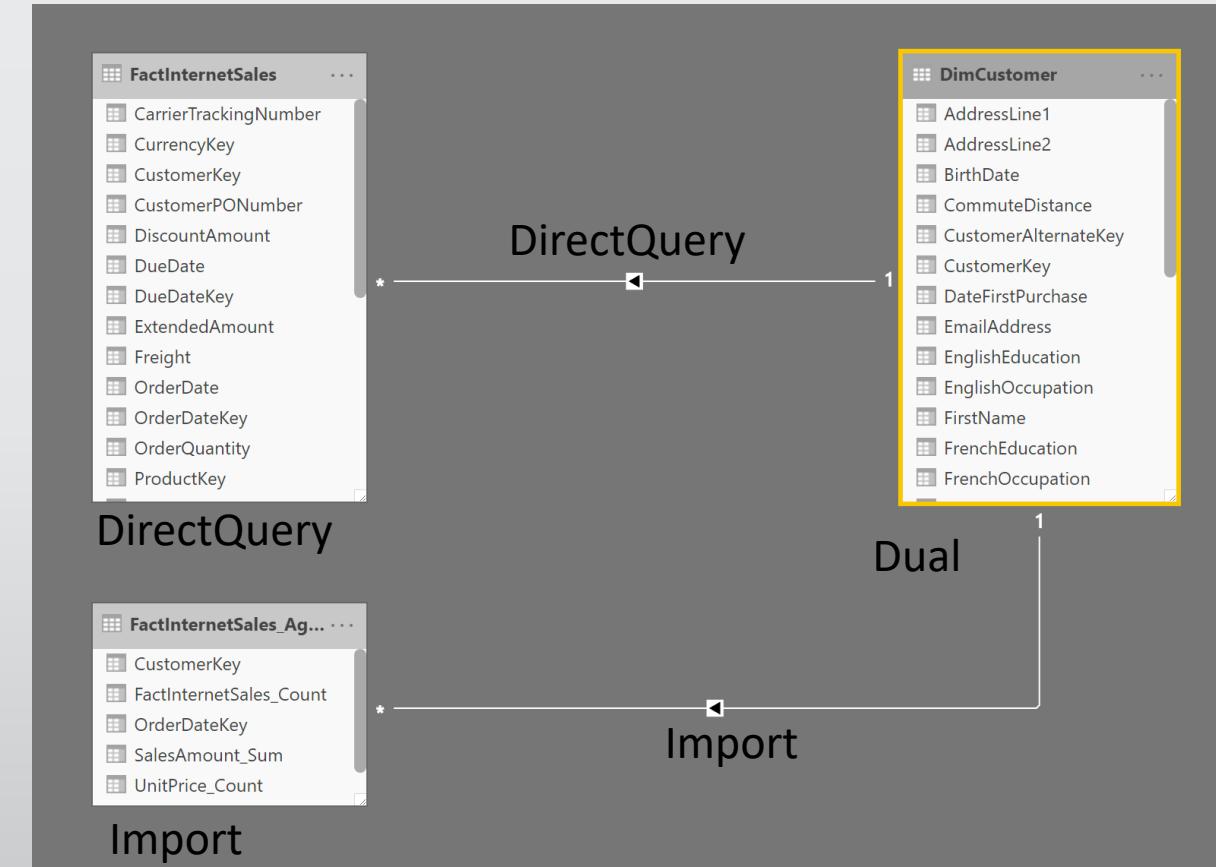
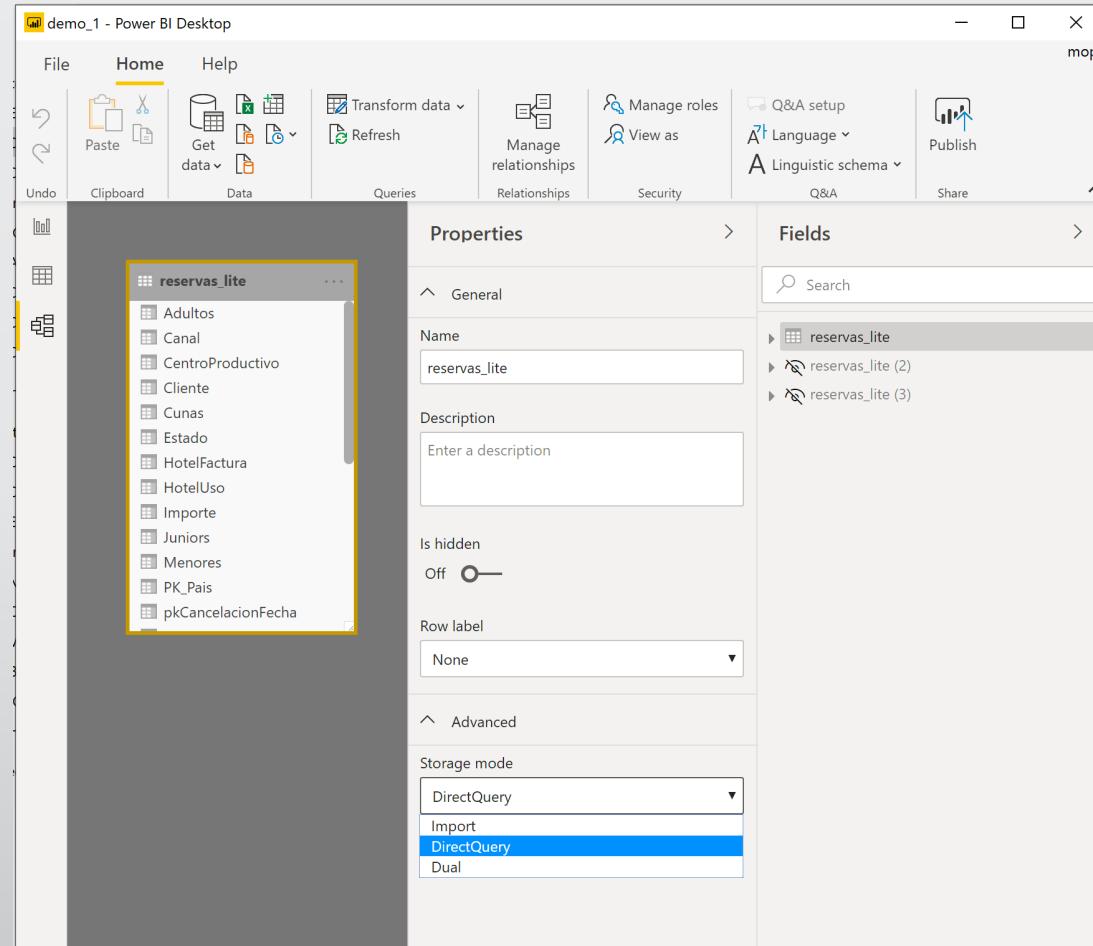
Composite Model



POWER PLATFORM

#PowerPlatformMallorca

Composite Model



POWER PLATFORM

#PowerPlatformMallorca

Aggregation

The screenshot shows the Power BI Data view interface. On the left, there's a sidebar with a 'Fields' section containing a search bar and a list of tables: 'reserva', 'Adultos', 'Canal', 'CentroProductivo', 'Cliente', 'Cunas', 'Estado', 'HotelFactura', 'HotelUso', 'Importe', 'Juniors', 'Menores', 'PK_Pais', 'pkCancelacionFecha', 'pkFecha', 'pkFechaEntrada', 'pkFechaSalida', 'pkVentaFecha', 'RegimenFactura', 'RegimenUso', 'Reserva', 'Segmento', 'TipoHabitacionFactura', 'TipoHabitacionUso', 'TipoHotelUso', 'Ventas', and 'Year'. A context menu is open over the 'reserva' table, with 'Manage aggregations' selected. The main area displays three tables: 'reservas_lite' (containing columns like Adultos, Canal, CentroProductivo, Cliente, Cunas, Estado, HotelFactura, HotelUso, Importe, Juniors, Menores, PK_Pais, pkCancelacionFecha, pkFecha, pkFechaEntrada, pkFechaSalida, pkVentaFecha, RegimenFactura, RegimenUso, Reserva, Segmento, TipoHabitacionFactura, TipoHabitacionUso, TipoHotelUso, Ventas, and Year), 'reservas_lite (2)' (containing columns like Adultos_Sum, Canal, Count, HotelFactura, HotelFactura_Count, Importe_Sum, and TipoHabitacionFactura), and 'reservas_lite (3)' (containing columns like Count, HotelFactura, and Importe_Sum). A 'Manage aggregations' dialog is open, titled 'Manage aggregations'. It includes a note: 'Aggregations accelerate query performance to unlock big-data sets. [Learn more](#)'. Below this, it shows an 'Aggregation table' dropdown set to 'reservas_lite (2)' and a 'Precedence' input field set to '0'. The main table area has four columns: AGGREGATION COLUMN, SUMMARIZATION, DETAIL TABLE, and DETAIL COLUMN. The rows define the following aggregations:

AGGREGATION COLUMN	SUMMARIZATION	DETAIL TABLE	DETAIL COLUMN
Adultos_Sum	Sum	reservas_lite	Adultos
Canal	GroupBy	reservas_lite	Canal
Count	Count table rows	reservas_lite	
HotelFactura	GroupBy	reservas_lite	HotelFactura
HotelFactura_Count	Count	reservas_lite	HotelFactura

At the bottom right of the dialog are 'Apply all' and 'Cancel' buttons.



POWER PLATFORM

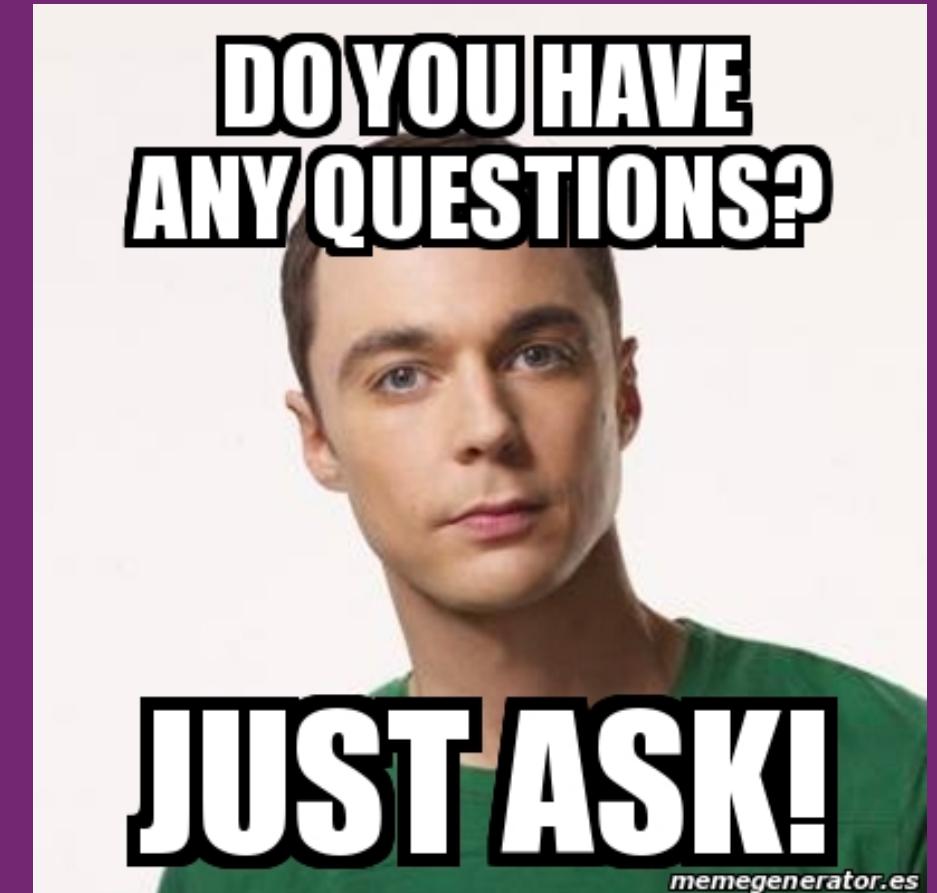
#PowerPlatformMallorca

Demo Aggregation y Composite Model



#GlobalPowerPlatformBootcamp

Alguna pregunta?

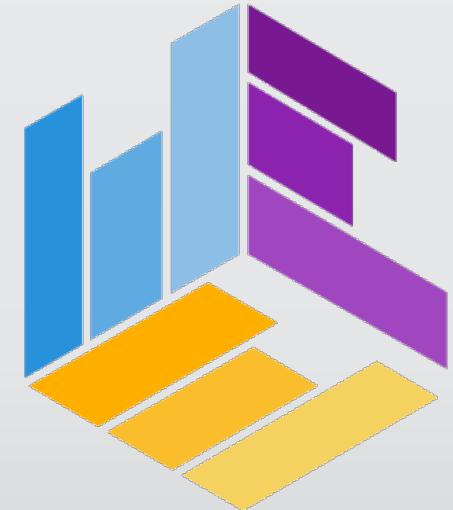


#GlobalPowerPlatformBootcamp

Toni Ferrá

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

aferra@mdata360.com



MASTER DATA 360