

Científico de Datos

Nivel Básico

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

Tema:

Bases de datos y Analítica Moderna

Aliados:



Microsoft

Vigilada Mineducación



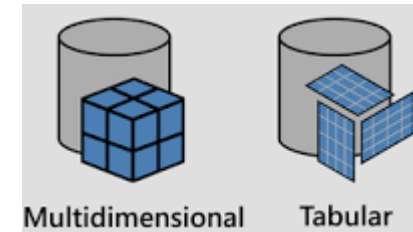
Bases de datos relacionales y no relacionales

student_id	age	score
1	12	77
2	12	68
3	11	75



```
[  
  {  
    "student_id":1,  
    "age":12,  
    "score":77  
  },  
  {  
    "student_id":2,  
    "age":12,  
    "score":68  
  },  
  {  
    "student_id":3,  
    "age":11,  
    "score":75  
  }  
]
```

- Modelos tabulares (2D)
- Cubos



Aliados:



Microsoft

Vigilada Mineducación



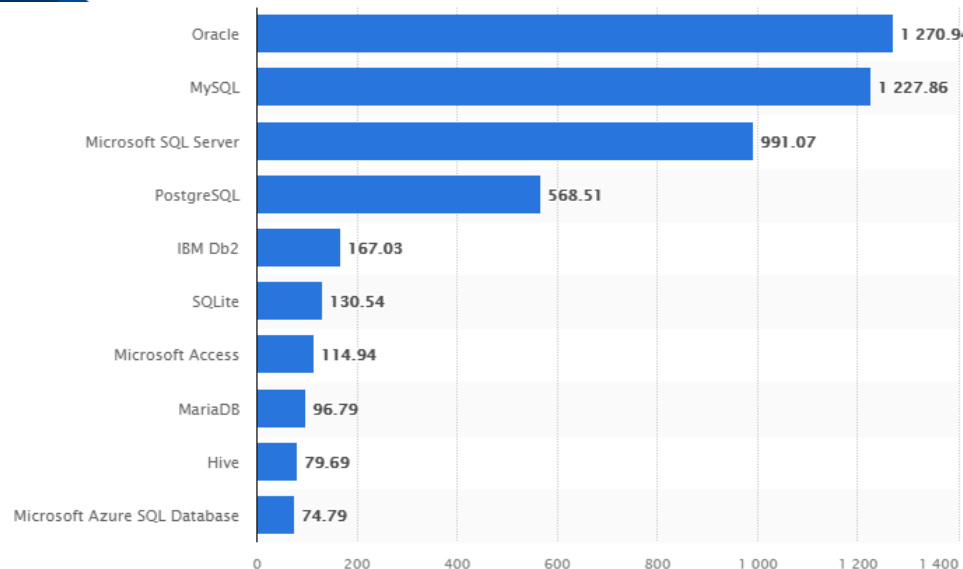
Advanced analytics for business

Bases de datos SQL y MySQL

Comparativa

Criterio	SQL	MySQL
Cost	Licenciado	Open source
Supported programming languages	Java, PHP, C++, Python, Ruby, Delphi, Go, R	PHP + Apache, Java, C++, Python, Ruby, Delphi, Go, R
Suported platforms	Windows, Linux	Linux, Windows, Solaris, Mac OS
Data file Manipulation	Not allowed	Allowed
Operational Storage Space	Large	Less

RDBMS



Servidores y estructura

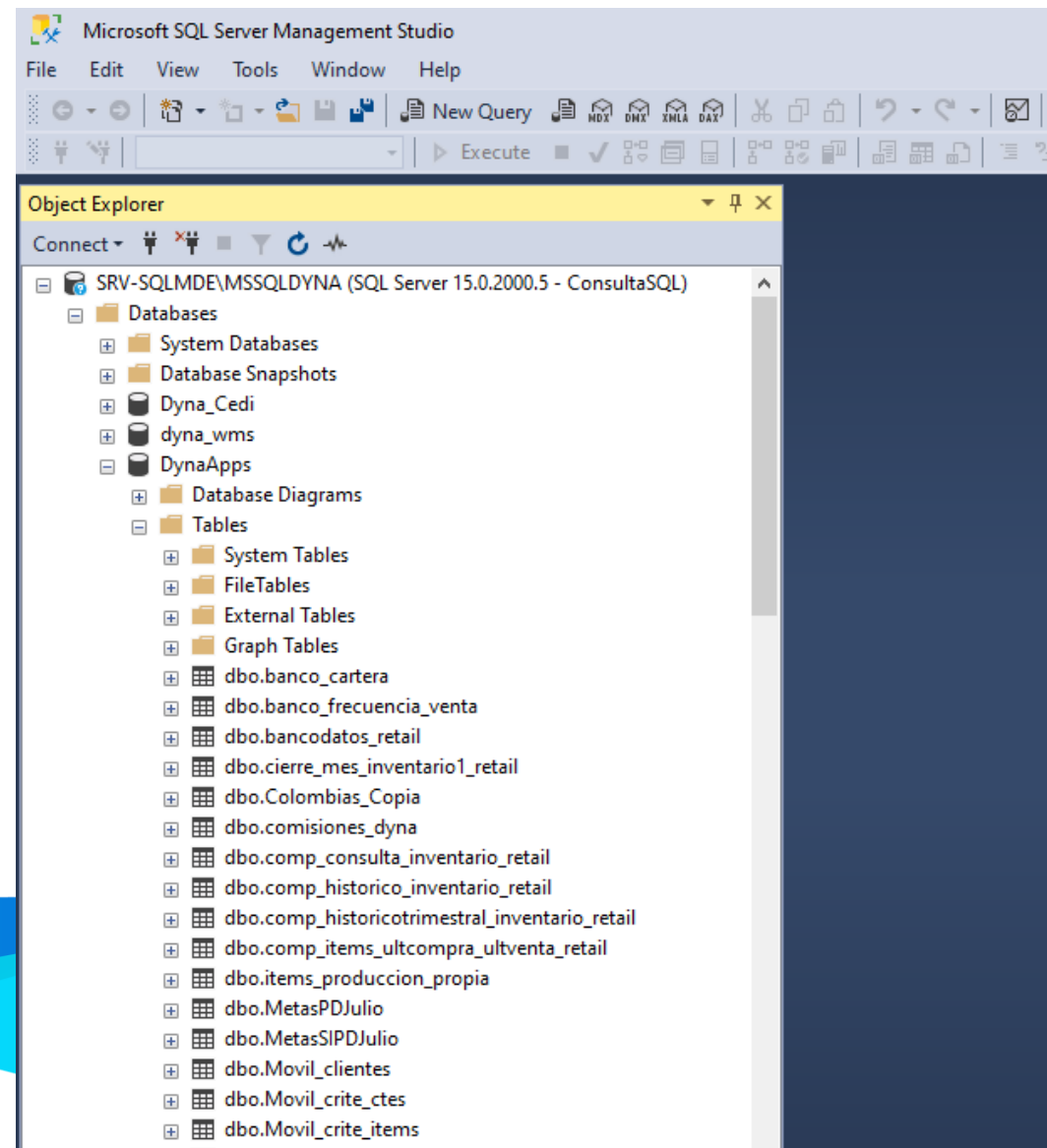
Servidores (cloud – on-premise)

Instancias

Bases de datos

Tablas (tipos de tablas)

Otros elementos (vistas, procedimientos almacenados)



Aliados:



Vigilada Mineducación



Advanced analytics for business

SQL

Para ir a las tablas hay que saber hacer las consultas en SQL

Lógica:

1. Crear / Eliminar tabla
2. Selección de campos
3. De cuál tabla con los campos (si hay varias tablas, tipos de cruce)
4. Condiciones a la consulta (filtros de partición)
5. Group (sum, avg, count)
6. Ordenar

Aliados:



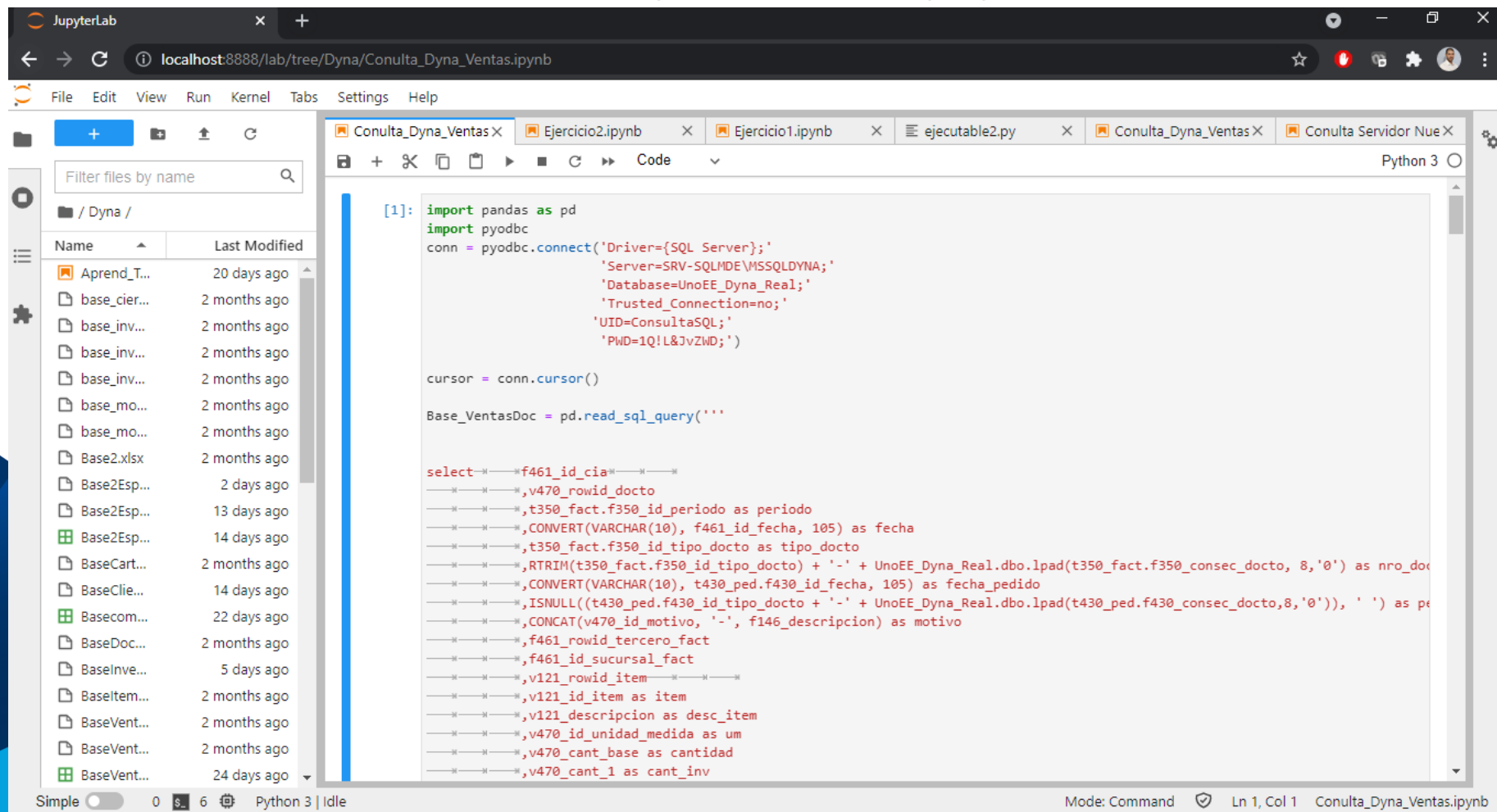
Microsoft

Vigilada Mineducación



Advanced analytics for business

Bonus 1: conexión Python (pyodbc)



The screenshot shows a JupyterLab environment with a browser window at `localhost:8888/lab/tree/Dyna/Consulta_Dyna_Ventas.ipynb`. The left sidebar displays a file explorer for the `/ Dyna /` directory, listing various files like `Aprend_T...`, `base_cier...`, `base_inv...`, `Base2.xlsx`, `Base2Esp...`, `BaseCart...`, `BaseClie...`, `Basecom...`, `BaseDoc...`, `BaseInve...`, `BaseItem...`, `BaseVent...`, and `BaseVent...`.

The main area shows a code cell with the following Python code:

```
[1]: import pandas as pd
import pyodbc
conn = pyodbc.connect('Driver={SQL Server};'
                      'Server=SRV-SQLMDEVMSQLDYNA;'
                      'Database=UnoEE_Dyna_Real;'
                      'Trusted_Connection=no;'
                      'UID=ConsultasSQL;'
                      'PWD=1Q!L&JvZWD;')

cursor = conn.cursor()

Base_VentasDoc = pd.read_sql_query(''

select * f461_id_cia *
* f470_rowid_docto
* t350_fact.f350_id_periodo as periodo
* CONVERT(VARCHAR(10), f461_id_fecha, 105) as fecha
* t350_fact.f350_id_tipo_docto as tipo_docto
* RTRIM(t350_fact.f350_id_tipo_docto) + '-' + UnoEE_Dyna_Real.dbo.lpad(t350_fact.f350_consec_docto, 8, '0') as nro_doc
* CONVERT(VARCHAR(10), t430_ped.f430_id_fecha, 105) as fecha_pedido
* ISNULL((t430_ped.f430_id_tipo_docto + '-' + UnoEE_Dyna_Real.dbo.lpad(t430_ped.f430_consec_docto, 8, '0')), ' ') as pe
* CONCAT(v470_id_motivo, '-', f146_descripcion) as motivo
* f461_rowid_tercero_fact
* f461_id_sucursal_fact
* v121_rowid_item
* v121_id_item as item
* v121_descripcion as desc_item
* v470_id_unidad_medida as um
* v470_cant_base as cantidad
* v470_cant_1 as cant_inv
```

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

Bonus 2: Buenas prácticas al hacer consultas SQL

- Peso de las consultas
- Formato para guardar las tables (parquet, ORC)
- Identificación de las tablas
- Cant. Máx. Case when
- Usar Compute Stats y ver Plan de ejecución
- Cant. Máx de joins (en una misma consulta)

Aliados:



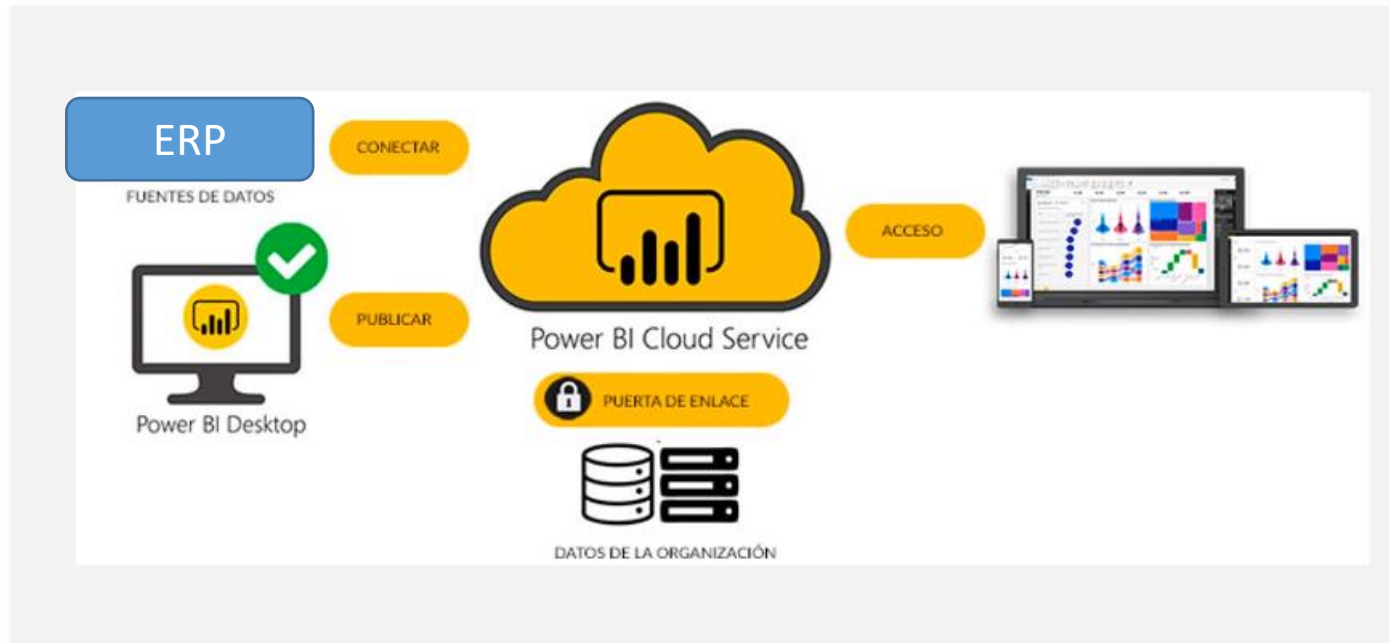
Microsoft

Vigilada Mineducación



Advanced analytics for business

Flujo de datos tradicional y básico



Aliados:



Vigilada Mineducación

Procesos de Analítica moderna

Definiciones

- Inteligencia de Negocios
- Analítica de datos
- Conceptos de Data Mart y Data Lake

Aliados:



Vigilada Mineducación



Procesos de Analítica moderna

Herramienta	On-Premise	Cloud
ETL	Integration Services*	Data Factory
Análisis en memoria	Analysis Services	Azure Analysis Services**
Dashboards y reportes	Power BI	Power BI
Python		Azure Functions***
Opciones de almacenamiento		Azure Blob Storage***

Paas Azure: Power BI con cloud implica integración y uso de servicios como Azure Synapse Analytics, Azure Data Lake Storage, Azure Cosmos DB y los servicios de Azure Analysis Services.

IaaS: permite administrar y orquestar estos recursos de acuerdo a las necesidades.

Recomendación (buenas practicas de computación en la nube): aislamiento de recursos para garantizar la disponibilidad y un punto de falla menos centralizado.

*Viene con el motor de base de datos.

**Esto para modelos tabulares y metadata (hay versiones Developer, Basic, Standard)

*** Servicio SaaS en donde se puede ejecutar el código Python

****Optimizado para el almacenamiento de cantidades masivas de datos no estructurados (como texto o datos binarios)

Aliados:



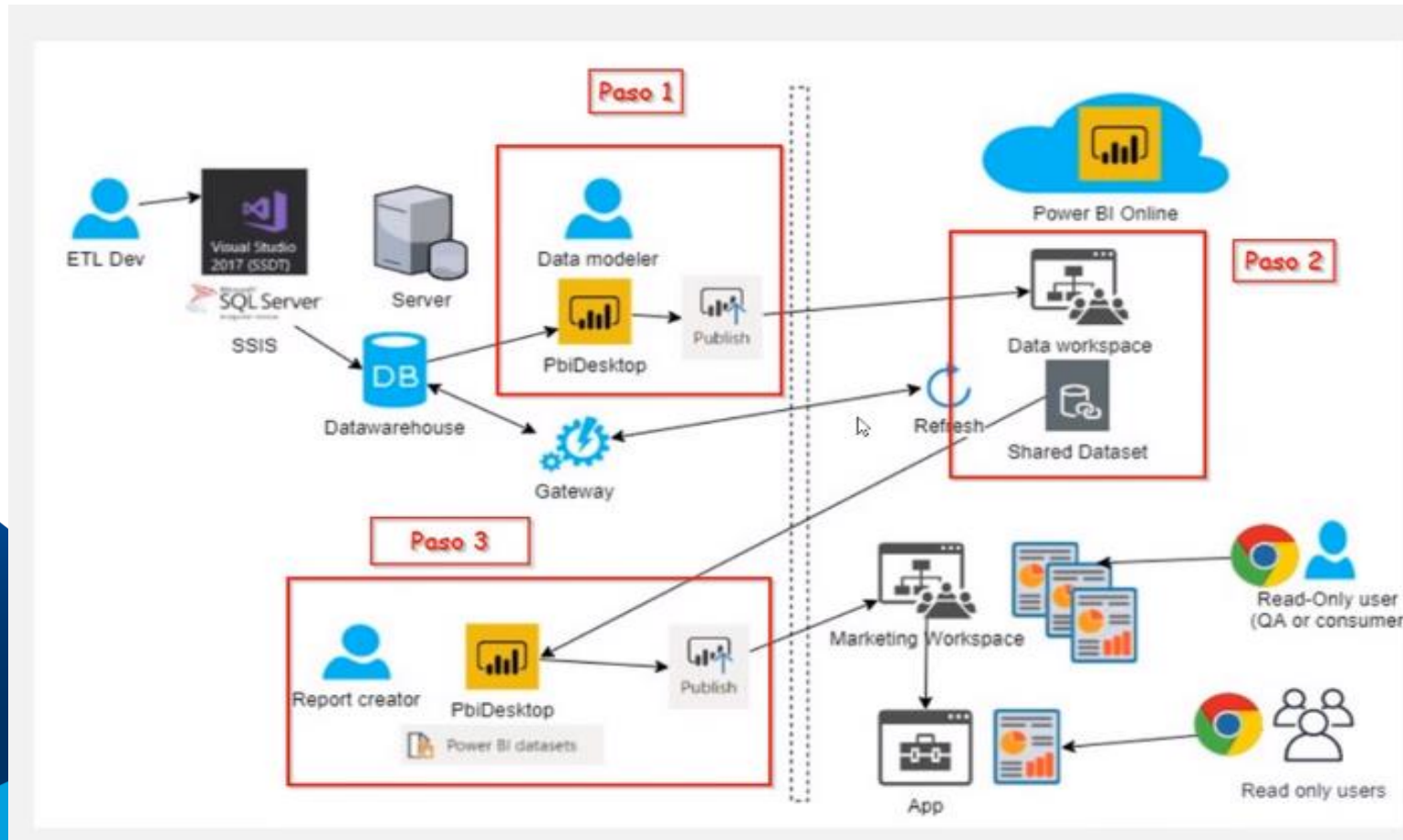
Microsoft

Vigilada Mineducación



Advanced analytics for business

Nuevos ciclos de datos



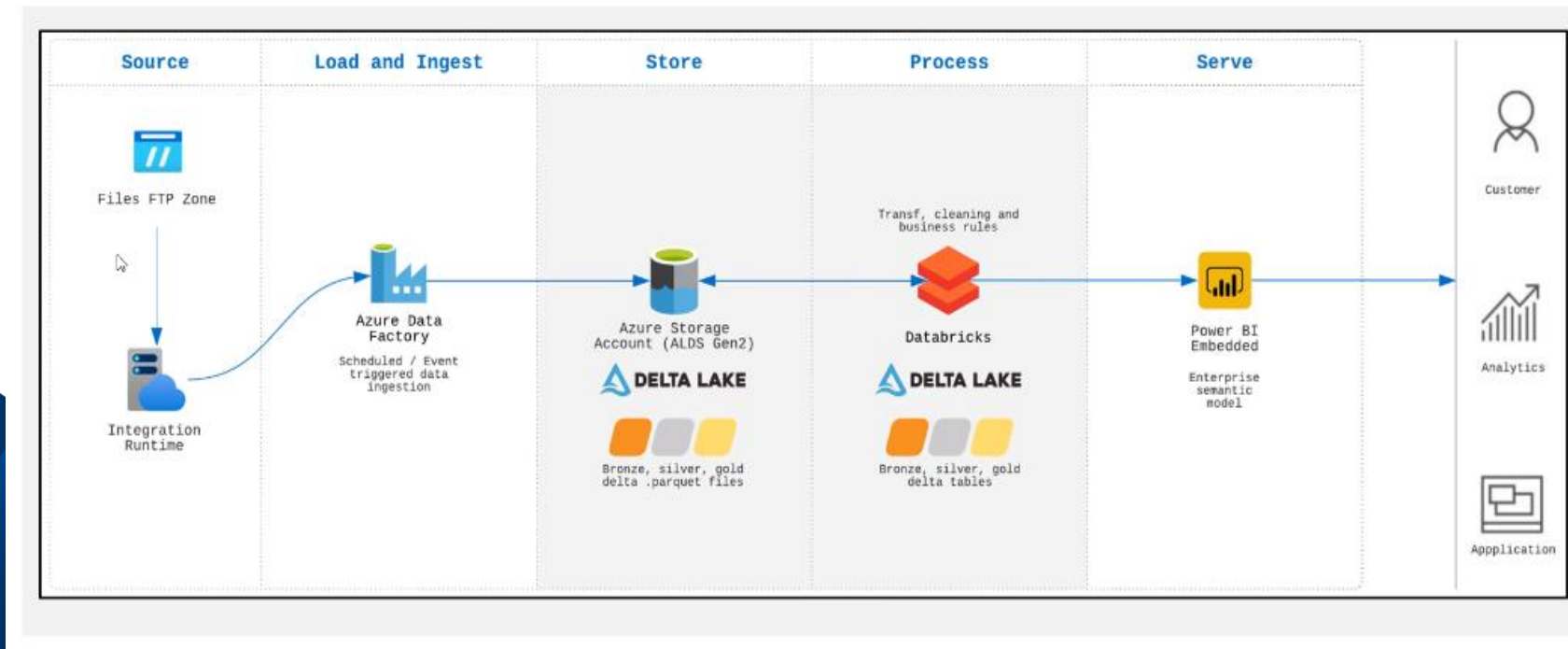
Aliados:



Vigilada Mineducación



Nuevos ciclos de datos



Aliados:

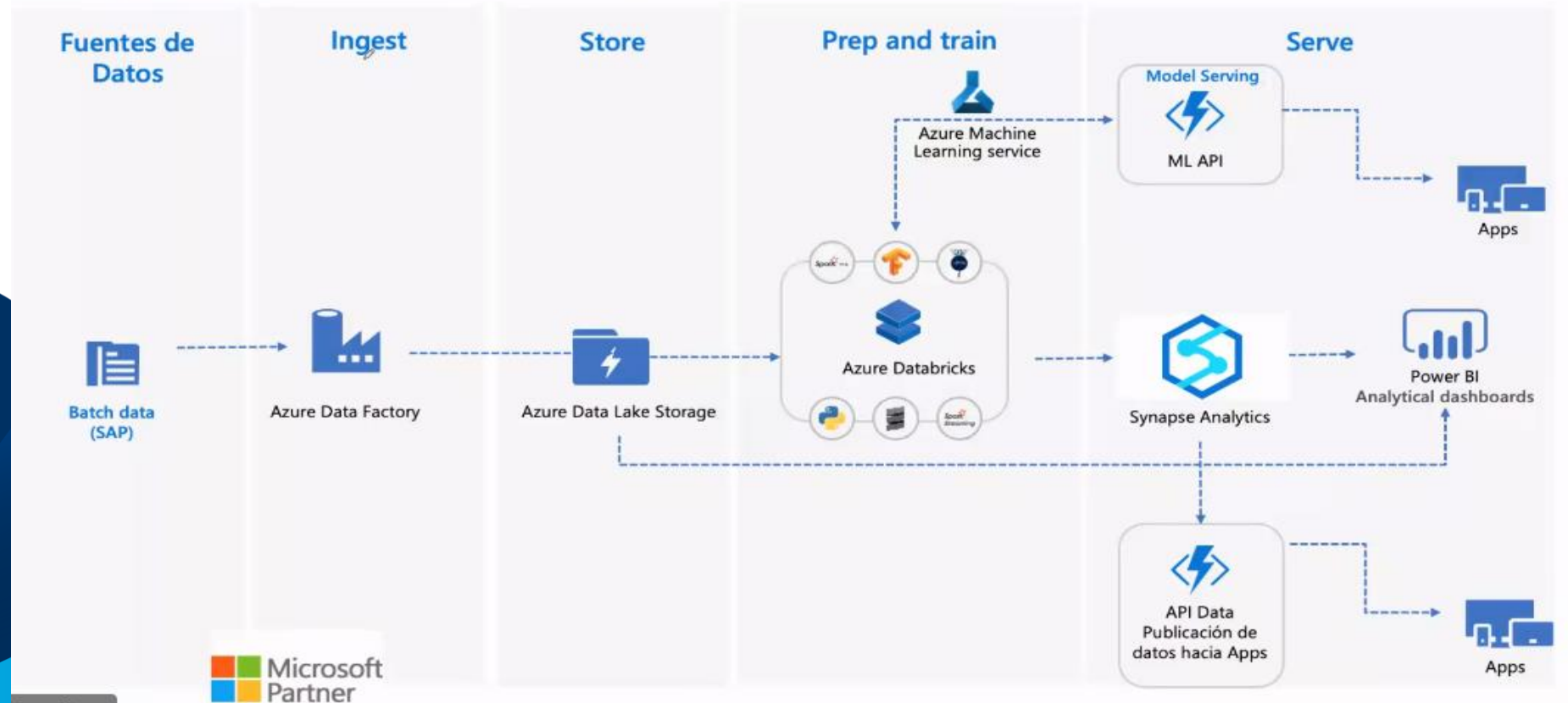


Vigilada Mineducación

Advanced analytics for business

Nuevos ciclos de datos

Arquitectura DWH Moderno

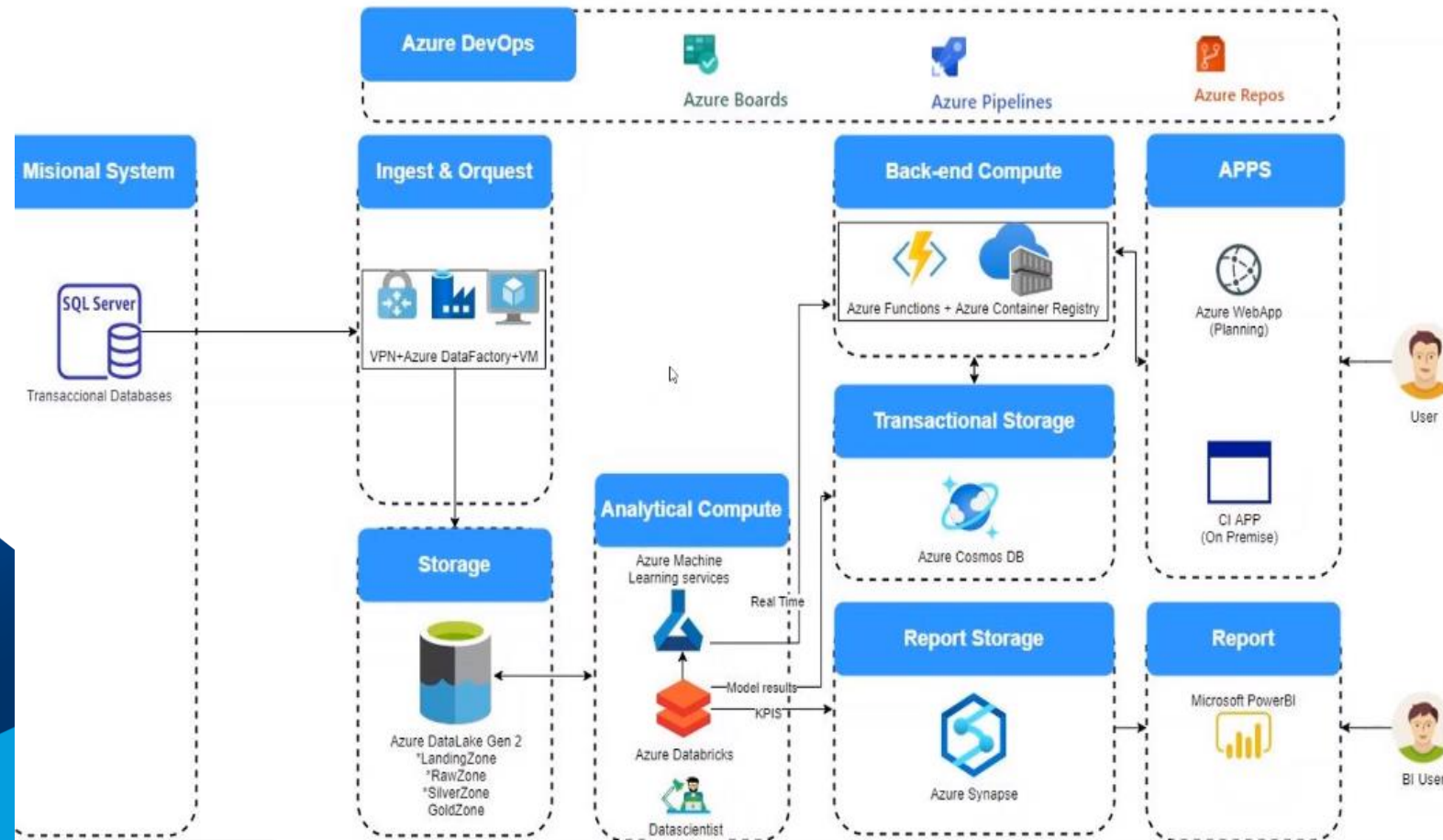


Aliados:



Vigilada Mineducación

Nuevos ciclos de datos



Aliados:



Vigilada Mineducación



Contenido asincrónico

- Resto de actividades en plataforma Interactiva Virtual

Aliados:



Microsoft

Vigilada Mineducación



¡Gracias!

Aliados:



Microsoft

Vigilada Mineducación

