PASS

Argentina Local Group PASS

DATA Summit 2019

Machine Learning en SQL Server

27 de mayo 2019

Maximiliano D. Accotto

Microsoft MVP Data Platform desde el 2005
Consultor especializado en Data Platform y Business Intelligence con mas de 15 años de experiencia sobre Plataforma Microsoft

Owner Triggerdb Consulting SRL.
www.triggerdb.com
www.triggerdb.com/blog
@maxiaccotto

Acerca de mi

Microsoft MVP Data Platform y speaker desde el año 2005.
Fundador de TriggerDB Consulting SRL.
Technical Solution Specialist Data Platform & BI con mas de 15 años de experiencia.

https://www.triggerdb.com

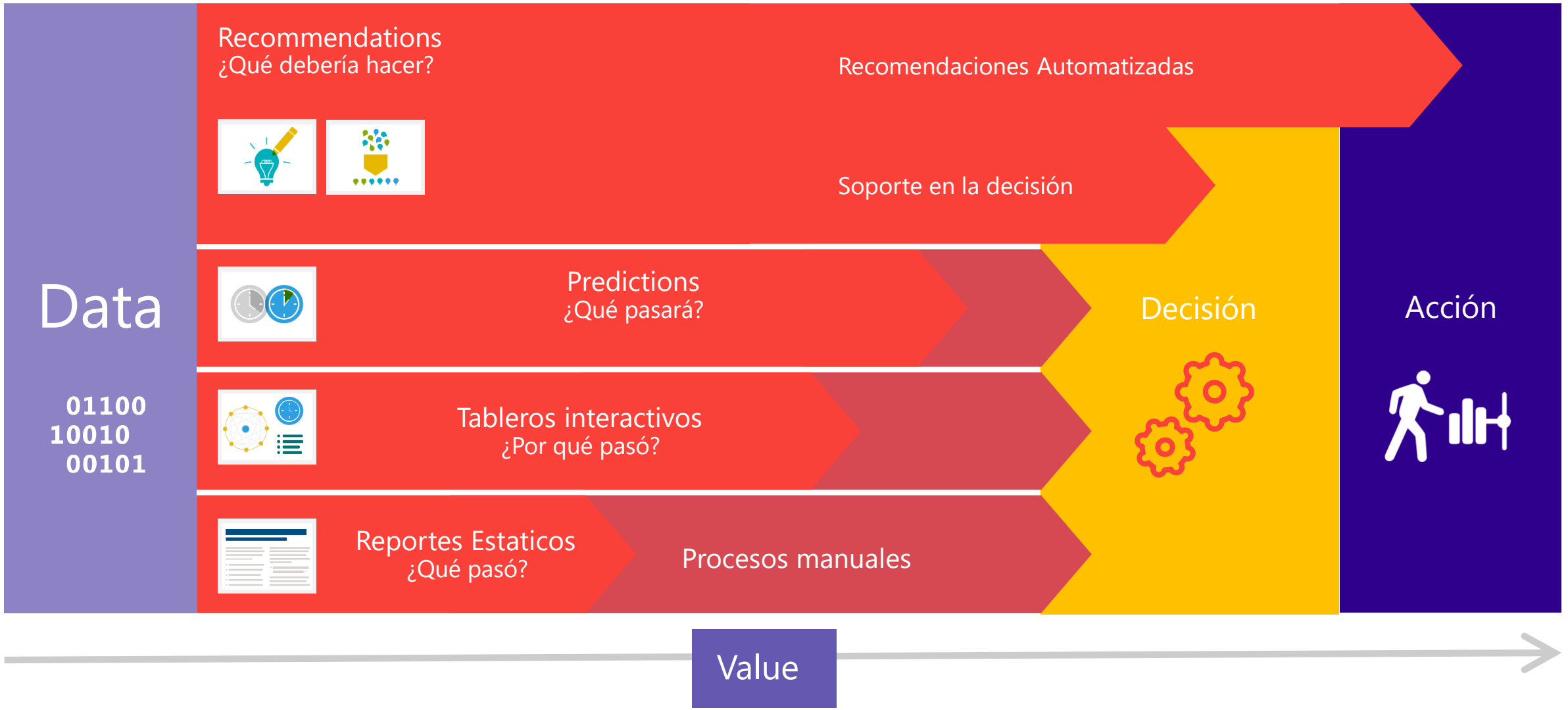
@maxiaccotto

https://www.linkedin.com/in/maxiaccotto

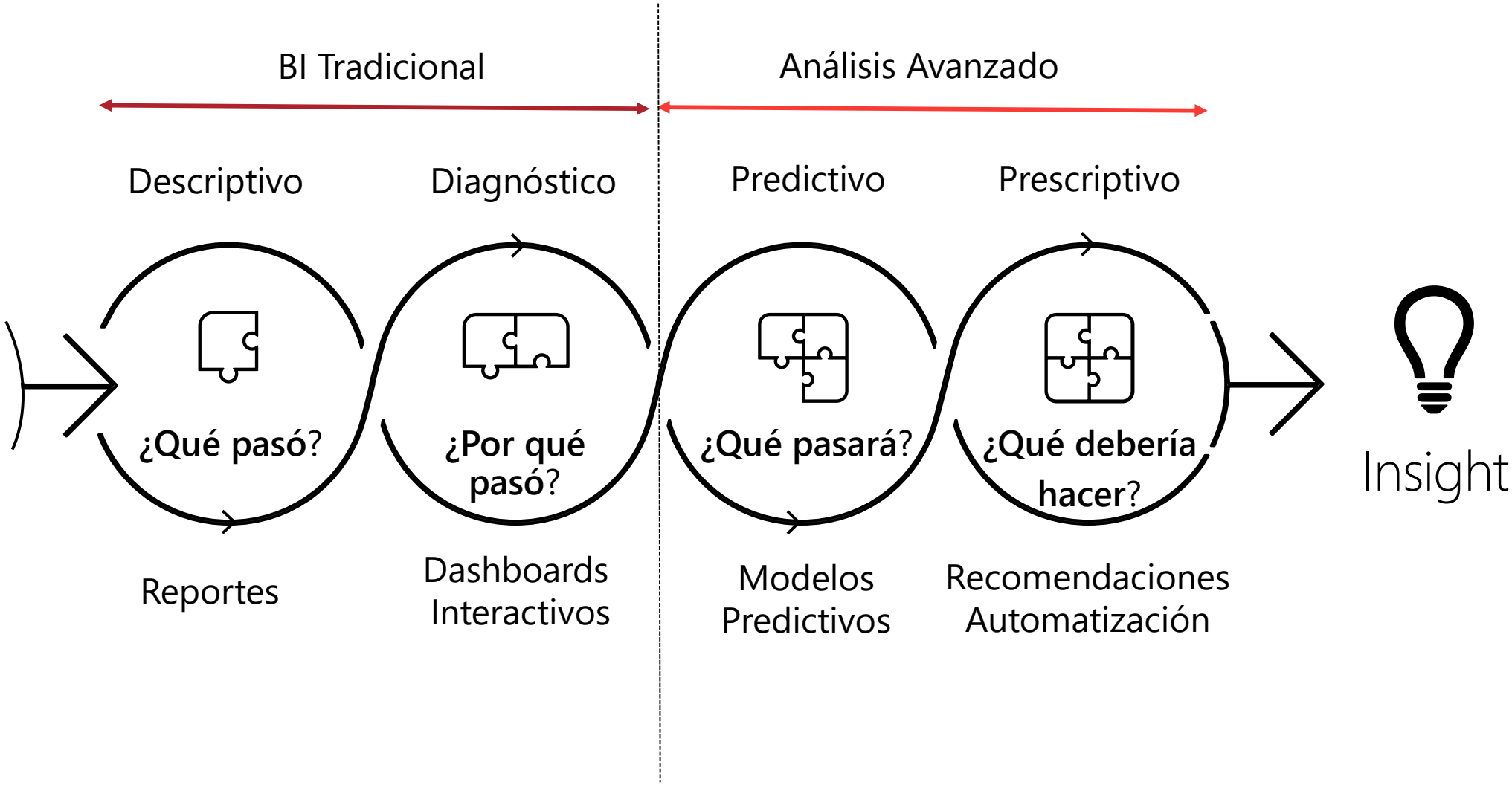
maxi@triggerdb.com

¿Que es Machine Learning?

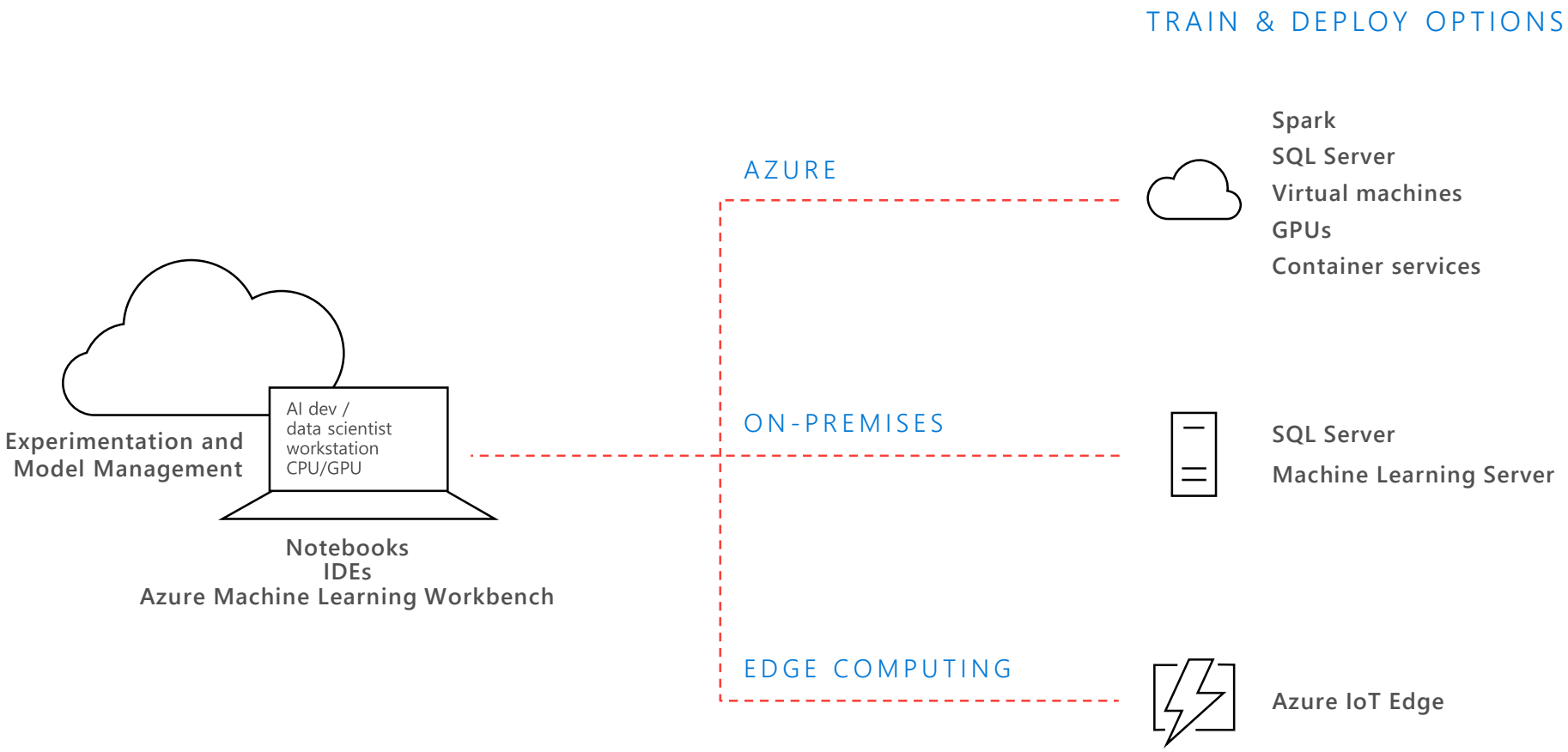
La creación de programas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos.


Recommendations
¿Qué debería hacer?
Recomendaciones Automatizadas
Soporte en la decisión
Data
01100
10010
00101
Predictions
¿Qué pasará?
Tableros interactivos
¿Por qué pasó?
Reportes Estaticos
¿Qué pasó?
Procesos manuales
Decisión
Acción
Value

Evolución del análisis

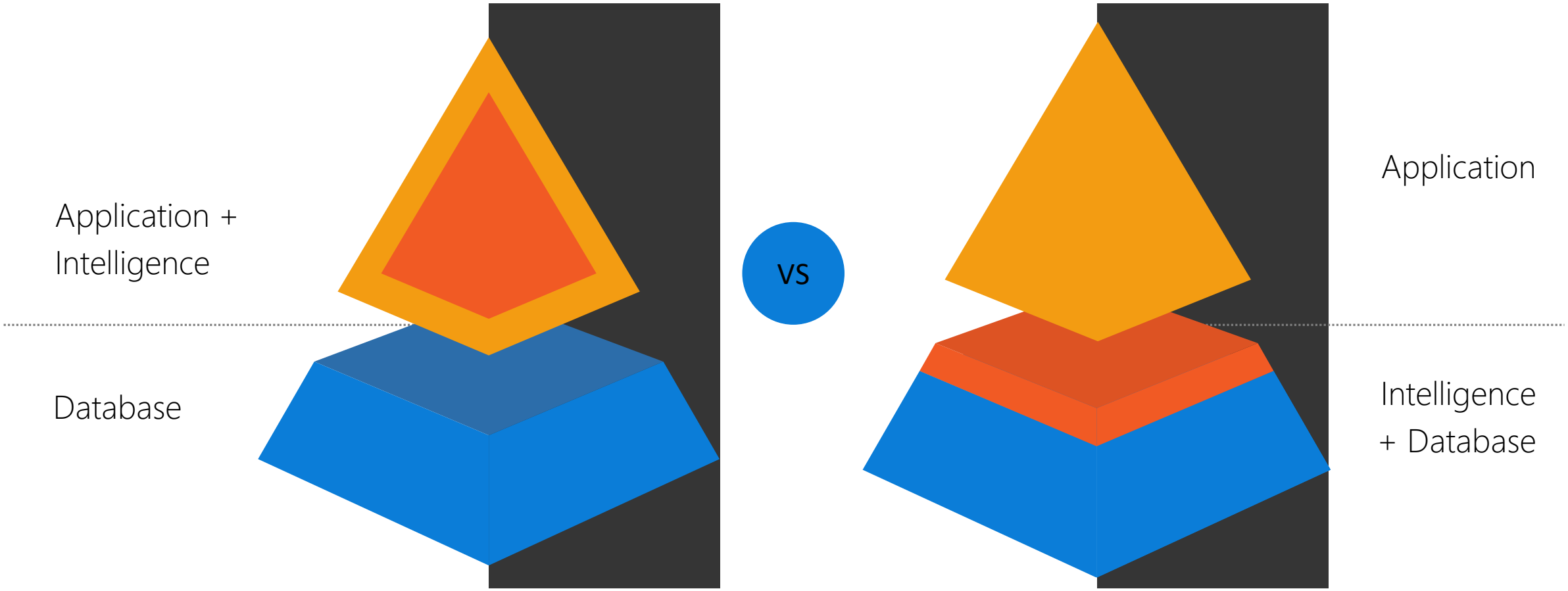


Microsoft Machine Learning



SQL Server Machine Learning

Llevar la inteligencia a donde están los datos



¿Porque ML en SQL Server?

Eliminar el movimiento de datos

Aprovechar la seguridad de la base de datos

Ejecutar los calculos de ML en la base de datos

Operación ML scripts y modelos

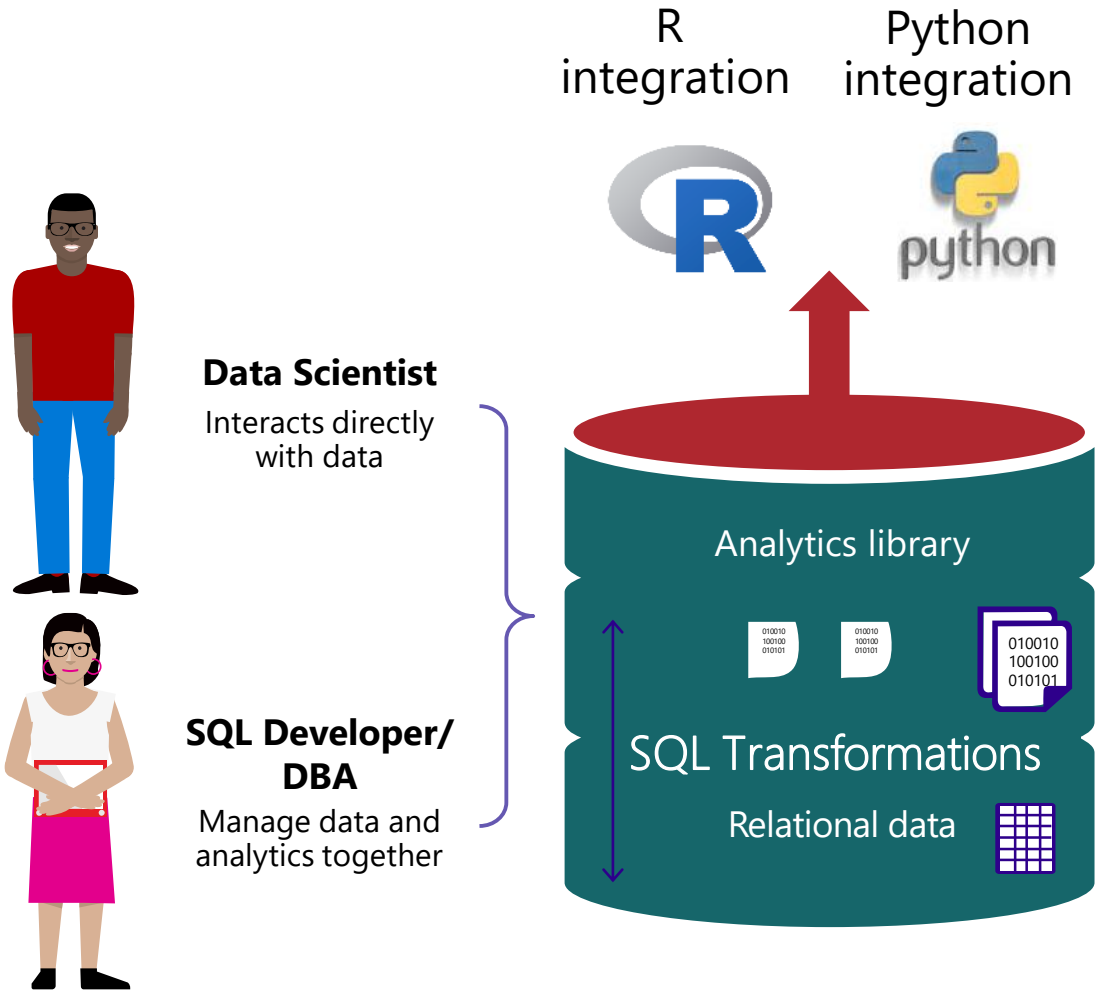
Usar store Procedures de T-SQL

Manejar los modelos en el SQL Server

Performance y escalabilidad Enterprise

Escalar con R y Python analytics usando multi-threading y parallel processing.

SQL Server security, compliance, resource governance, query performance, always on secondaries



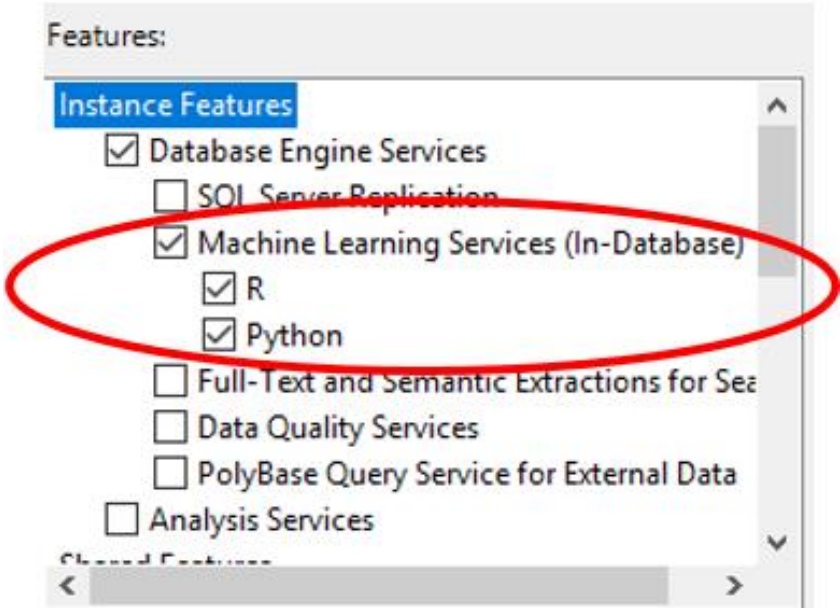
SQL Machine Learning Services

SQL Server 2016

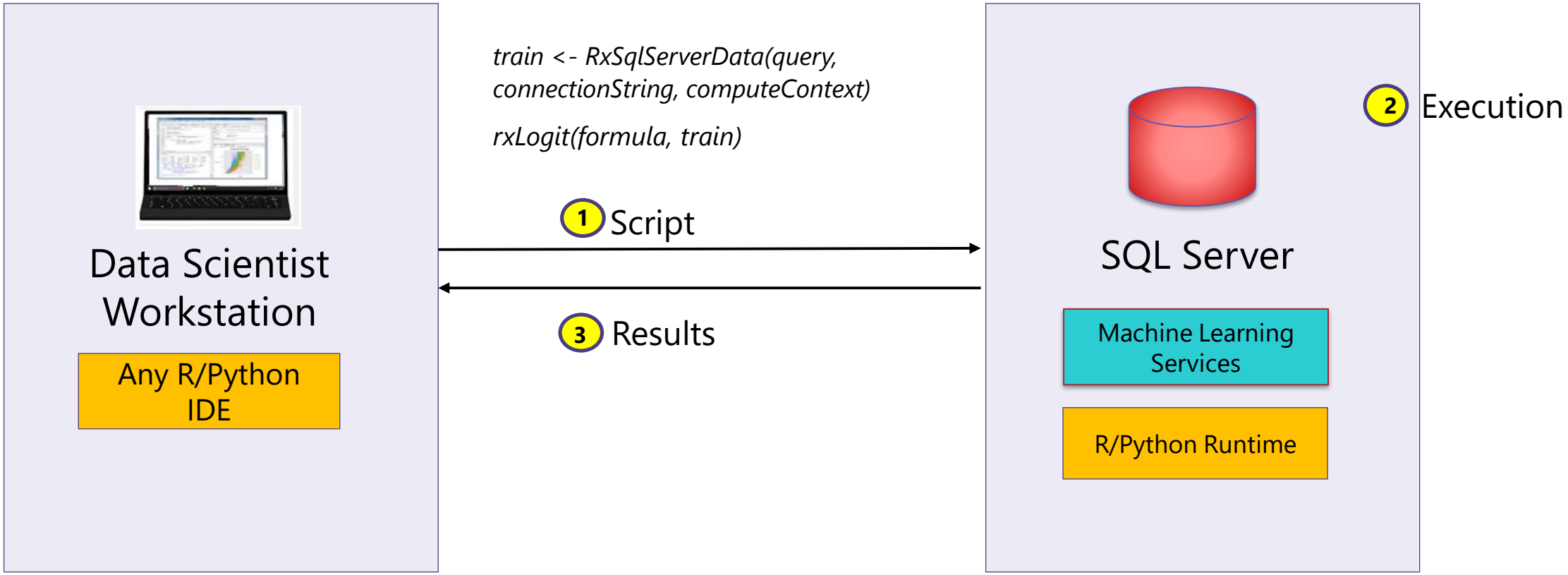
- R support (3.2.2 version)
- Microsoft R Server

SQL Server 2017

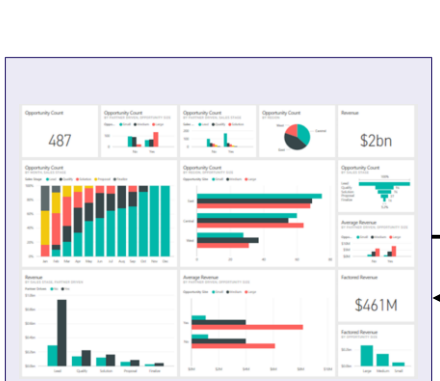
- Scoring native en TSQL usando PREDICT function
- EXTERNAL LIBRARY DDL para el manejo de paquetes R
- Ejecución en batch para la entrada de datos
- Soporte para R (3.3.3 version)
- Soporte para Python (Anaconda 3.5.2)



Data Scientists - Exploración de Datos y Desarrollo de Modelos



Application Developer - **Operacionalización de modelos**



Application

1 Stored Proc call

3 Results

SQL Server

Machine Learning Services

R/Python Runtime

2 Execution

The stored procedure contains R or Python code and executes in-database

exec `sp_execute_external_script`

```
    @language = 'Python'
,   @script =
-- Python code --
```

Trabajo del DBA: Habilitar ML en SQL Server ☺

Enable External scripts

- Exec sp_configure 'external scripts enabled', 1
- RCONFIGURE

sp_execute_external_script

```
EXEC sp_execute_external_script
    @language = N'R',
    @script = N'[Codigo]',
    @input_data_1 = N'[SQL input]'
    [ , @input_data_1_name = N'InputDataSet' ]
    [ , @output_data_1_name = N'OutputDataSet' ]
    [ , @params = N'parameter' ]
WITH RESULT SETS (([SQL output]));
```

Tipos de salida

1. **Dataset**
- Standard resultset of rows and columns
- Data types will vary
2. **Plot**
- Static images
- Binary
3. **Model**
- Trained models such as linear regression, naïve bayes, etc.
- Binary

Realtime Predictions usando Scoring nativo

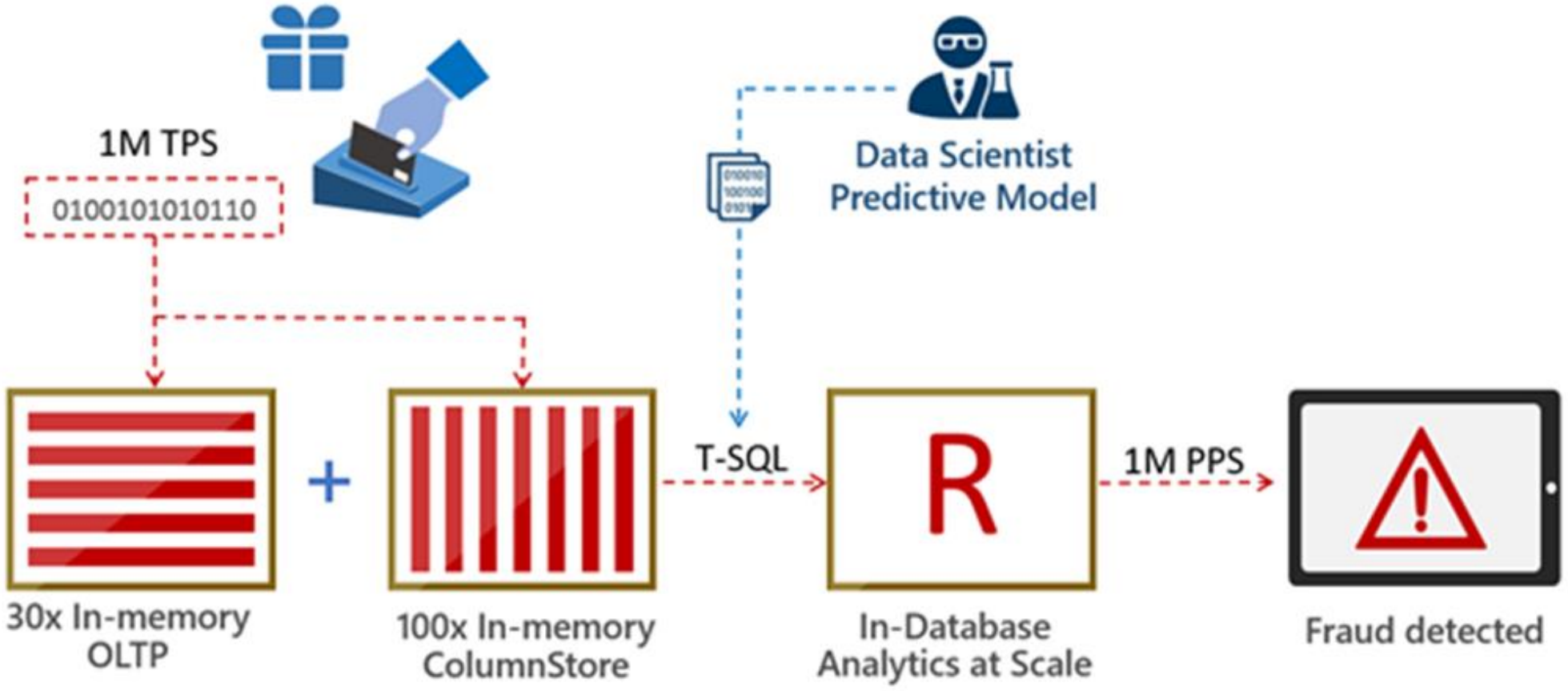
- PREDICT function
  - No depende de R o Python runtime
  - Habilitado on SQL Server tanto en Windows como Linux
- Uso
  - Single or small number of rows scoring
  - Highly concurrent scoring scenarios
  - Predict during INSERT, UPDATE, MERGE statements
- Requirements
  - Models built using RevoScaleR or revoscalepy
    - rxLinMod, rxLogit, rxBTrees, rxDTree, rxDForest
  - Serialized using rxSerializeModel (R) or rx_serialize_model (Python)

PREDICT syntax

```
PREDICT ( MODEL = @model | model_literal,
          DATA = object AS <table_alias> )
   WITH ({ {column_definition } [,...n ] } )


INSERT INTO loan_applications
    (c1, c2, c3, c4, risk_score)
SELECT d.c1, d.c2, d.c3, d.c4, p.score
  FROM PREDICT(MODEL = @model, DATA = @input as d)
  WITH(score float) as p;
```

1,000,000 predictions per second



https://blogs.technet.microsoft.com/dataplatforminsider/2016/10/11/1000000-predictions-per-second/

SQL Server Machine Server

SQL Machine Learning Server

Soporte Multi-plataforma

- Windows, Linux, Hadoop, SQL Server

Microsoft R Server

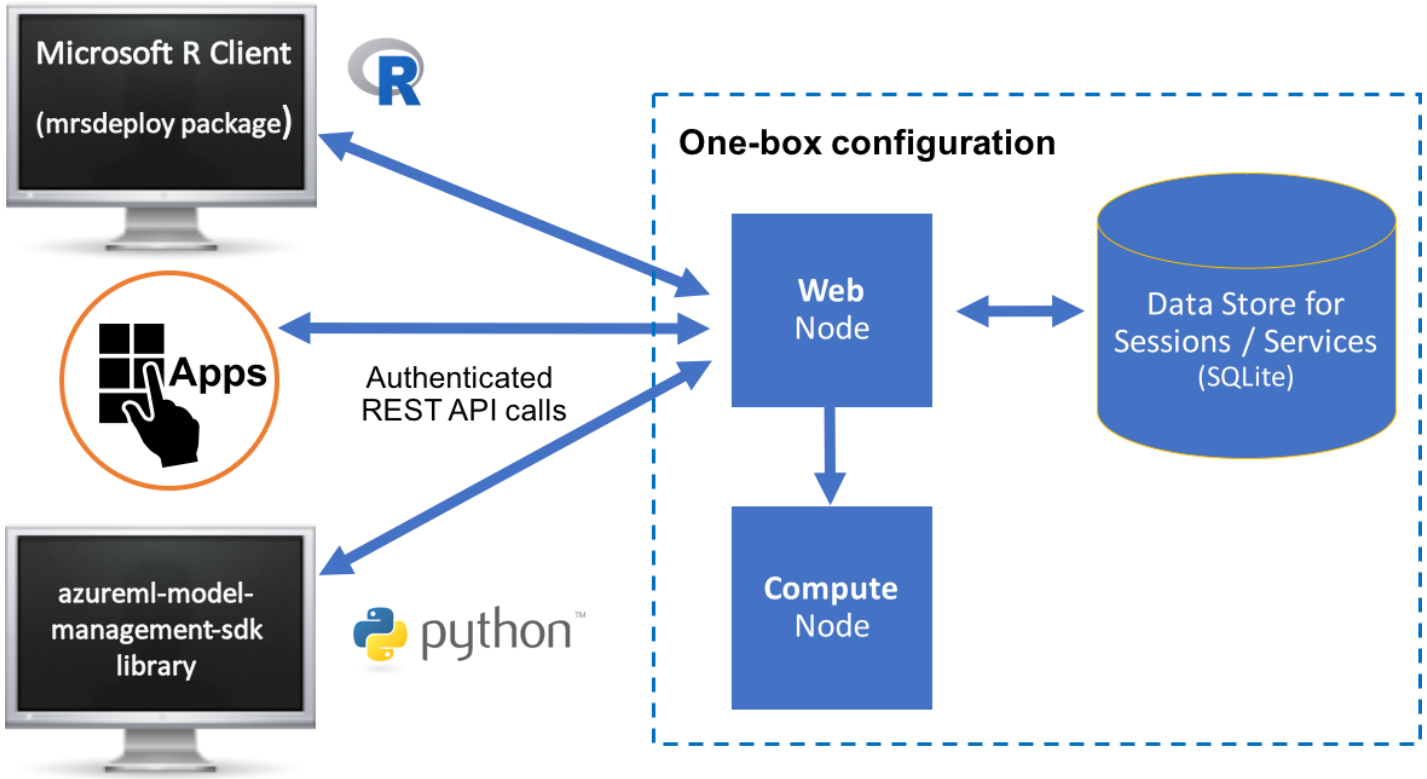
- RevoScaleR, MicrosoftML, olapR, sqlrutils packages
- **Uso de Web services para operar.**

Microsoft Machine Learning Server

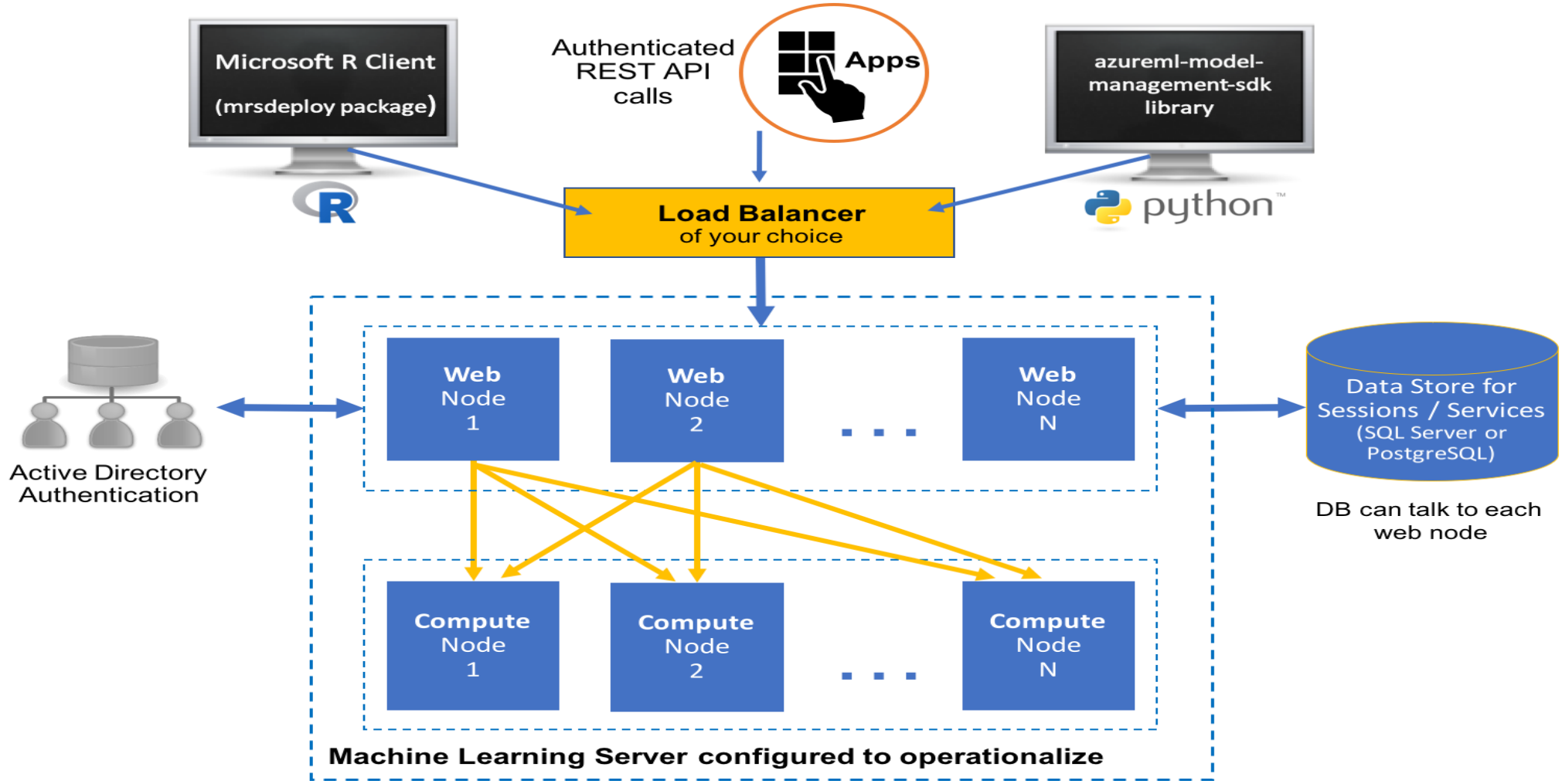
- Soporte de R & Python
- revoscalepy, microsoftml python libraries
- rxExecBy

SQL Machine Learning Server

- Setup one-box configuration
  `> az ml admin boostrap`
- Jupyter notebook deploy, and consume



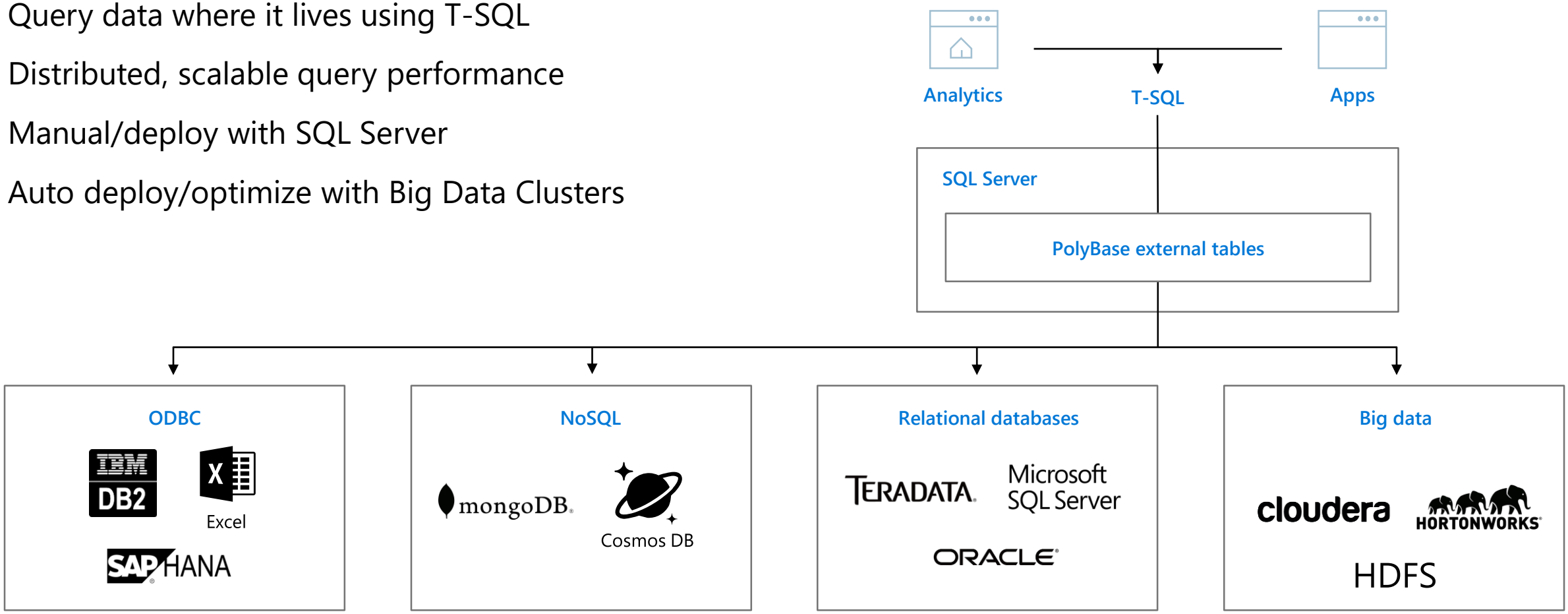
SQL Machine Learning Server – Multi Server Configuration



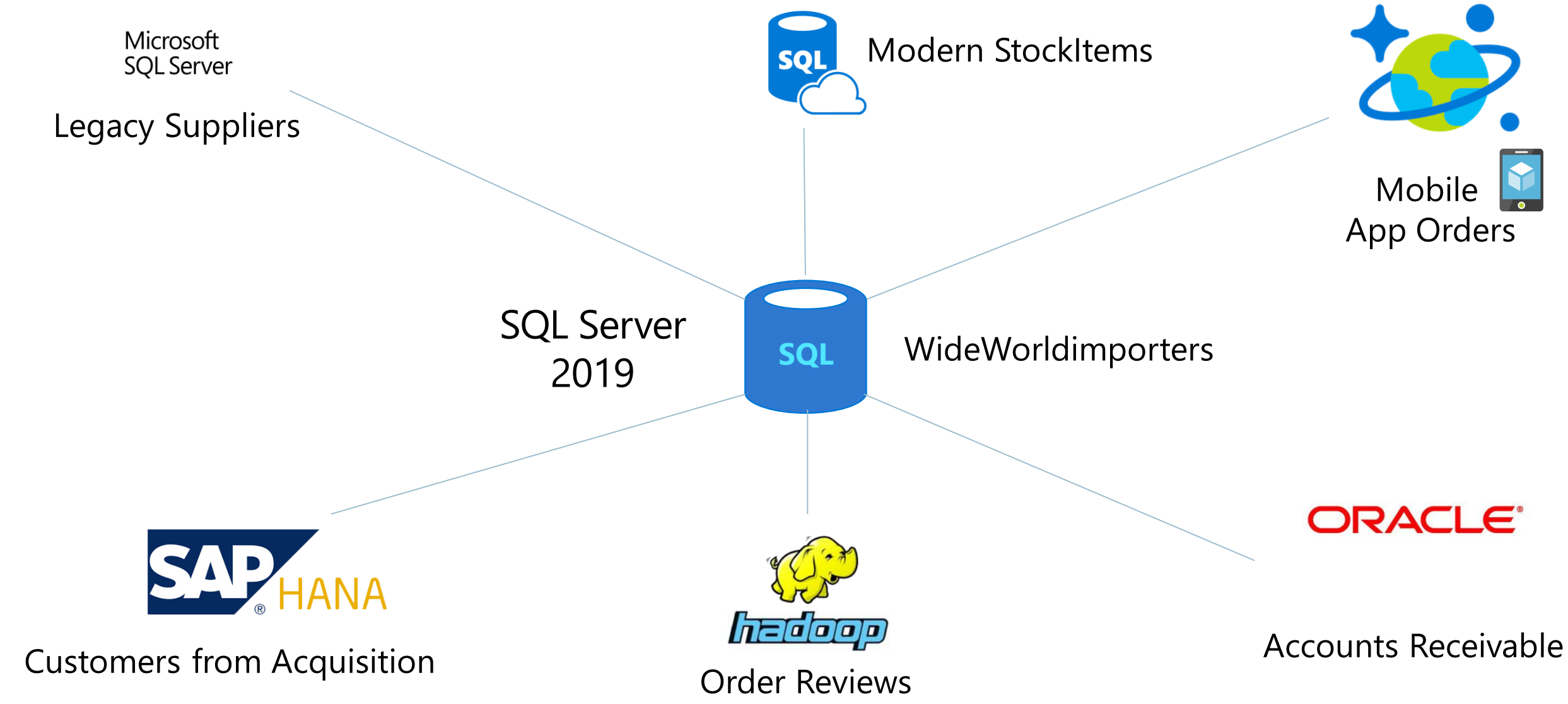
SQL 2019

What is SQL Server Polybase? "It's all about Data Virtualization"

- ✓ Distributed compute engine integrated with SQL Server
- ✓ Query data where it lives using T-SQL
- ✓ Distributed, scalable query performance
- ✓ Manual/deploy with SQL Server
- ✓ Auto deploy/optimize with Big Data Clusters

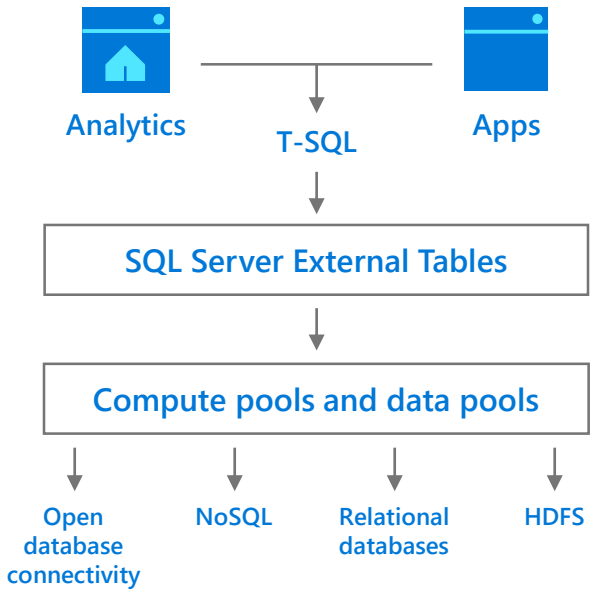


SQL Server 2019: Data Virtualization



SQL Server 2019 Big Data Clusters

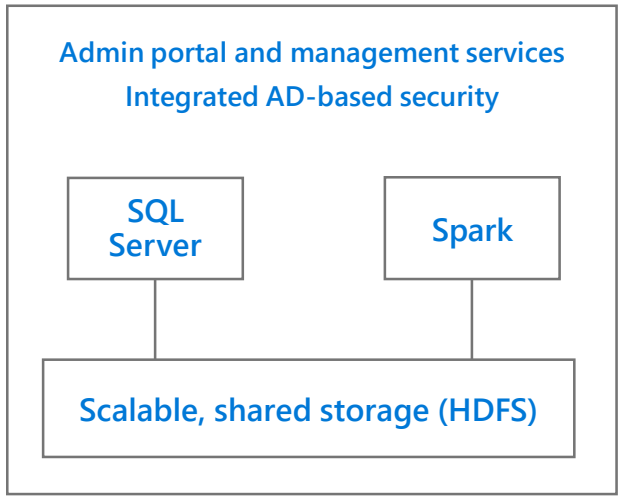
Data virtualization



Combine data from many sources without moving or replicating it

Scale out compute and caching to boost performance

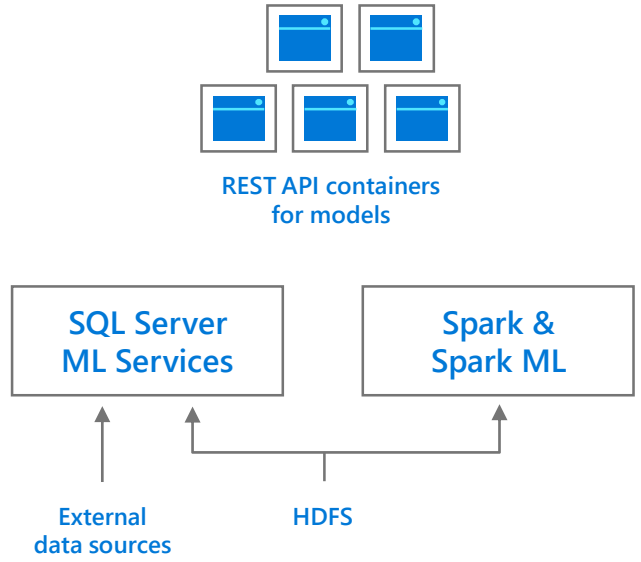
Managed SQL Server, Spark, and data lake



Store high volume data in a data lake and access it easily using either SQL or Spark

Management services, admin portal, and integrated security make it all easy to manage

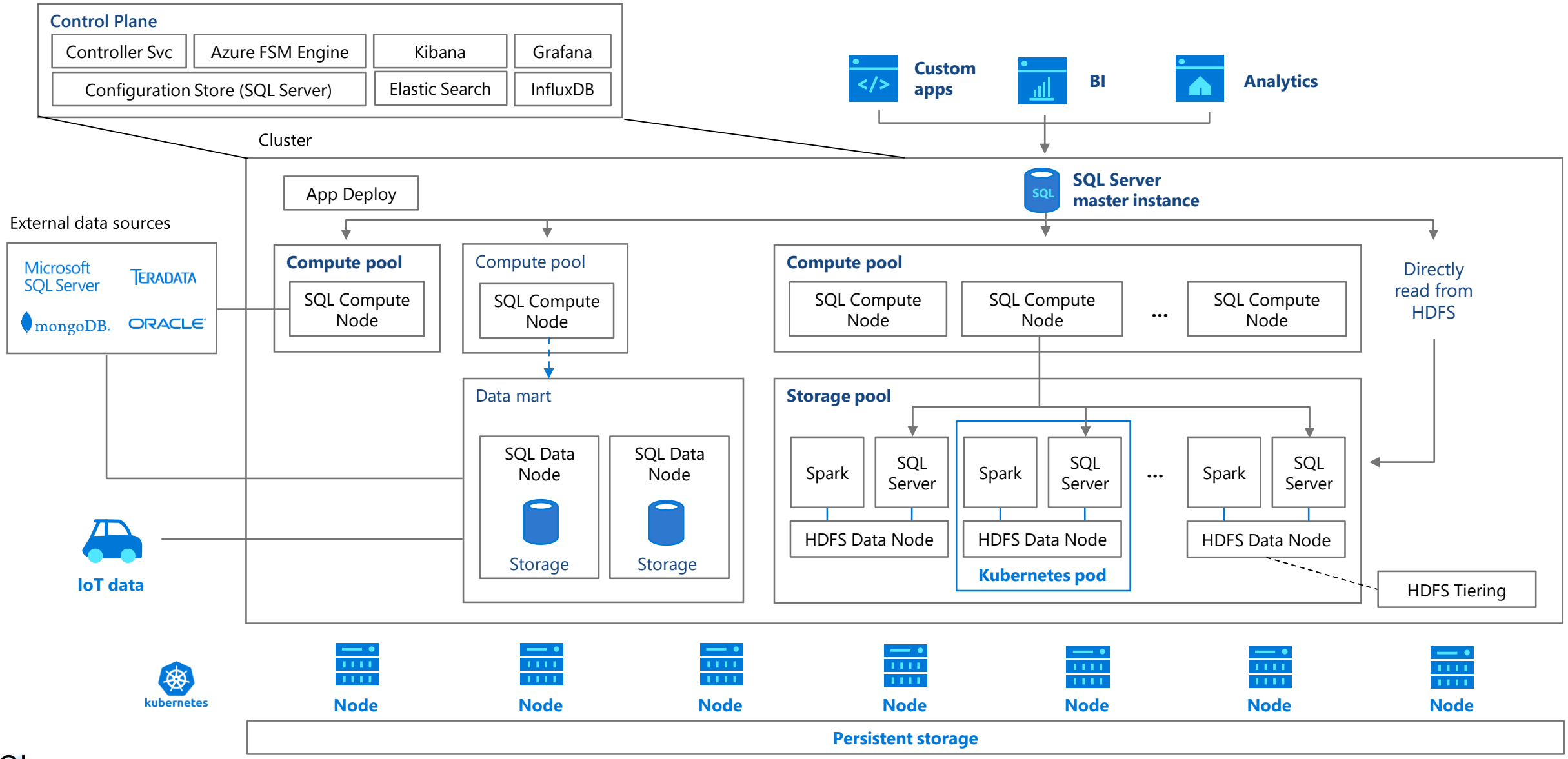
Complete AI platform



Easily feed integrated data from many sources to your model training

Ingest and prep data and then train, store, and operationalize your models all in one system

SQL Server Big Data Cluster Architecture



Microsoft Machine Learning

Maximiliano Accotto | MVP Data Platform


maxi@triggerdb.com


maxiaccotto


www.triggerdb.com

Para Conectarnos

Registrarse para una membresía gratuita y asociarse al grupo Local SQL Argentina:

pass.org
Sqlargentina.pass.org

https://www.facebook.com/groups/SQLArgentina/

@sqlargentina

https://www.linkedin.com/groups/3664741

https://www.youtube.com/channel/UCGa-ldNDwP8nAVfQwqoeYHw