



# Arquitectura en Azure para Python

***Adrián J. Fernández Zenteno***

*Sr. Cloud Solution Architect - Azure Advanced Analytics*

Email: [adrian.fernandez@microsoft.com](mailto:adrian.fernandez@microsoft.com)

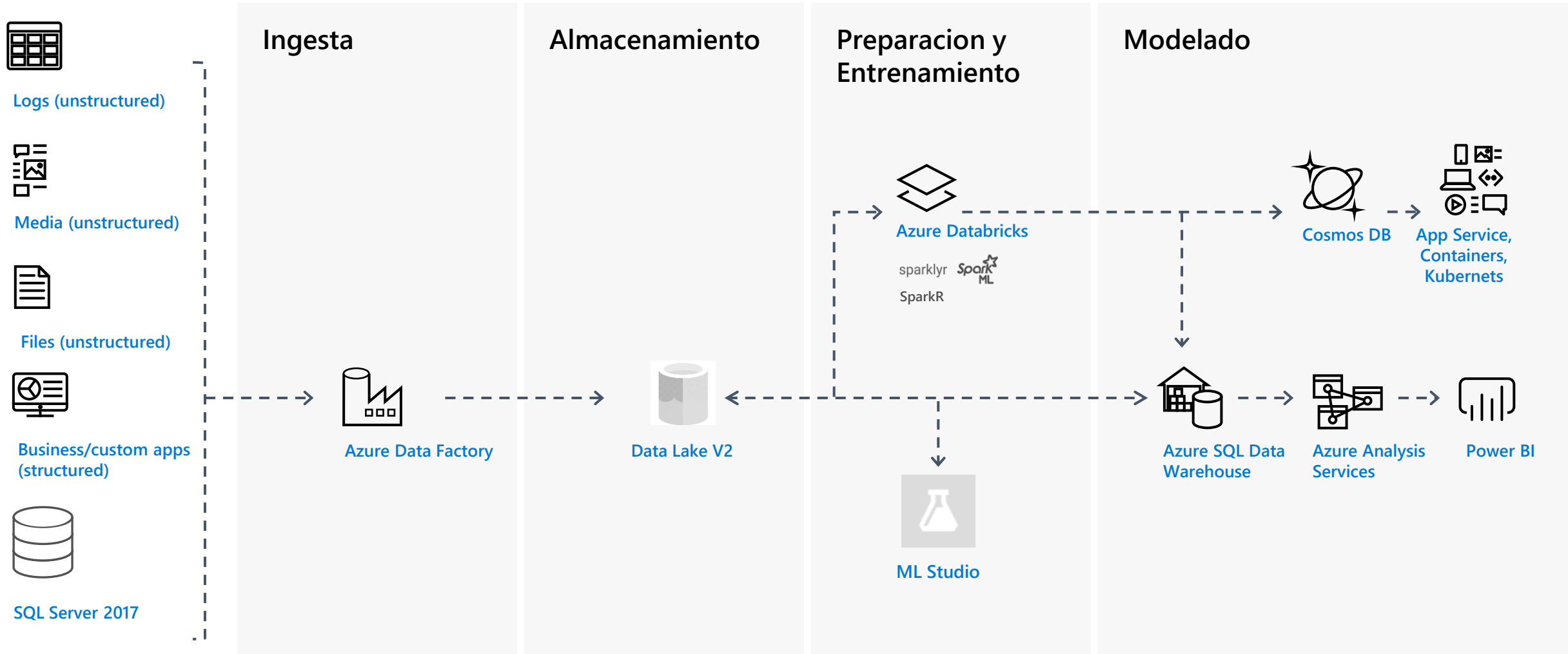
Twitter: @AdrianFZ10



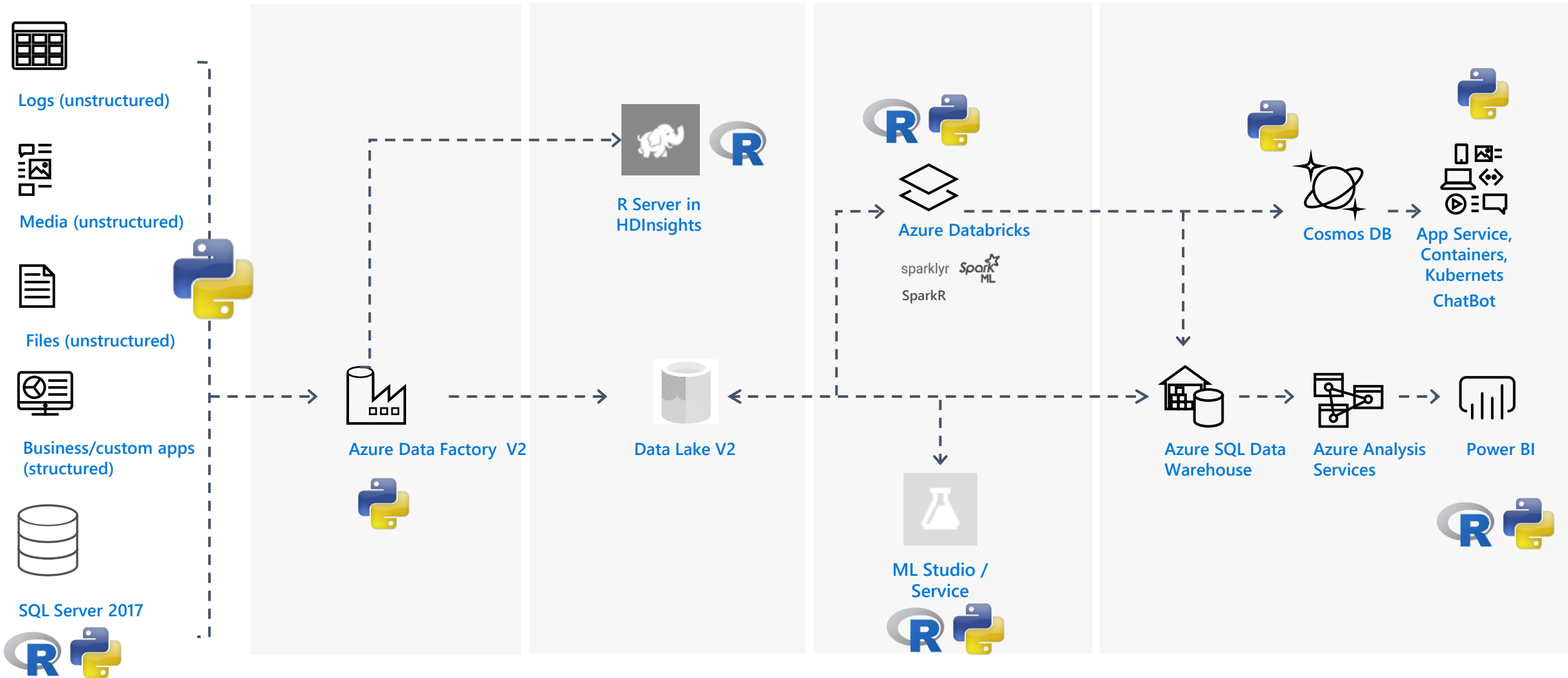
# Agenda

Architecture reference, ml studio, datasets,  
azure notebooks, databricks, notebook vm,  
automl

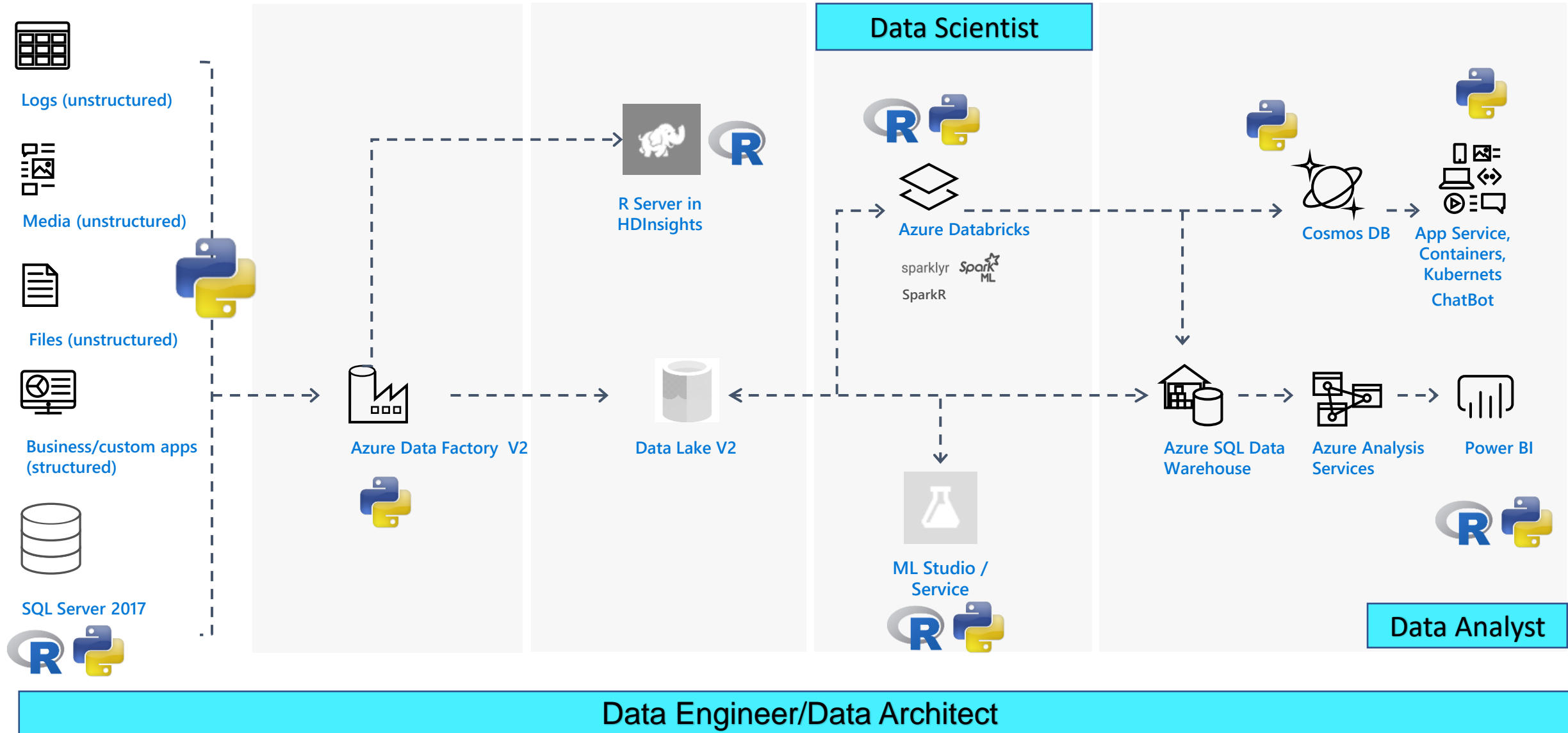
# Arquitectura de Referencia



# Arquitectura de Referencia

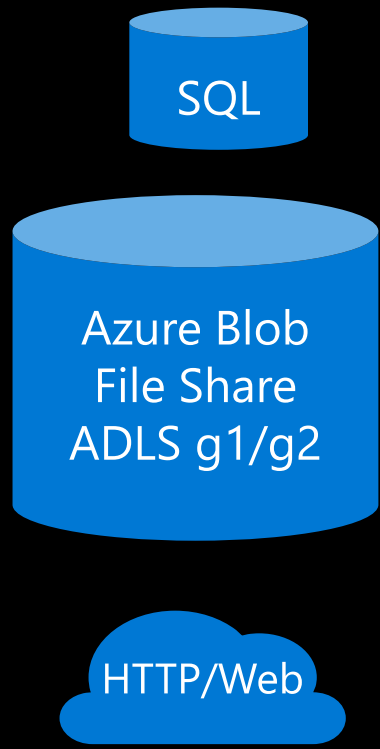


# Arquitectura de Referencia

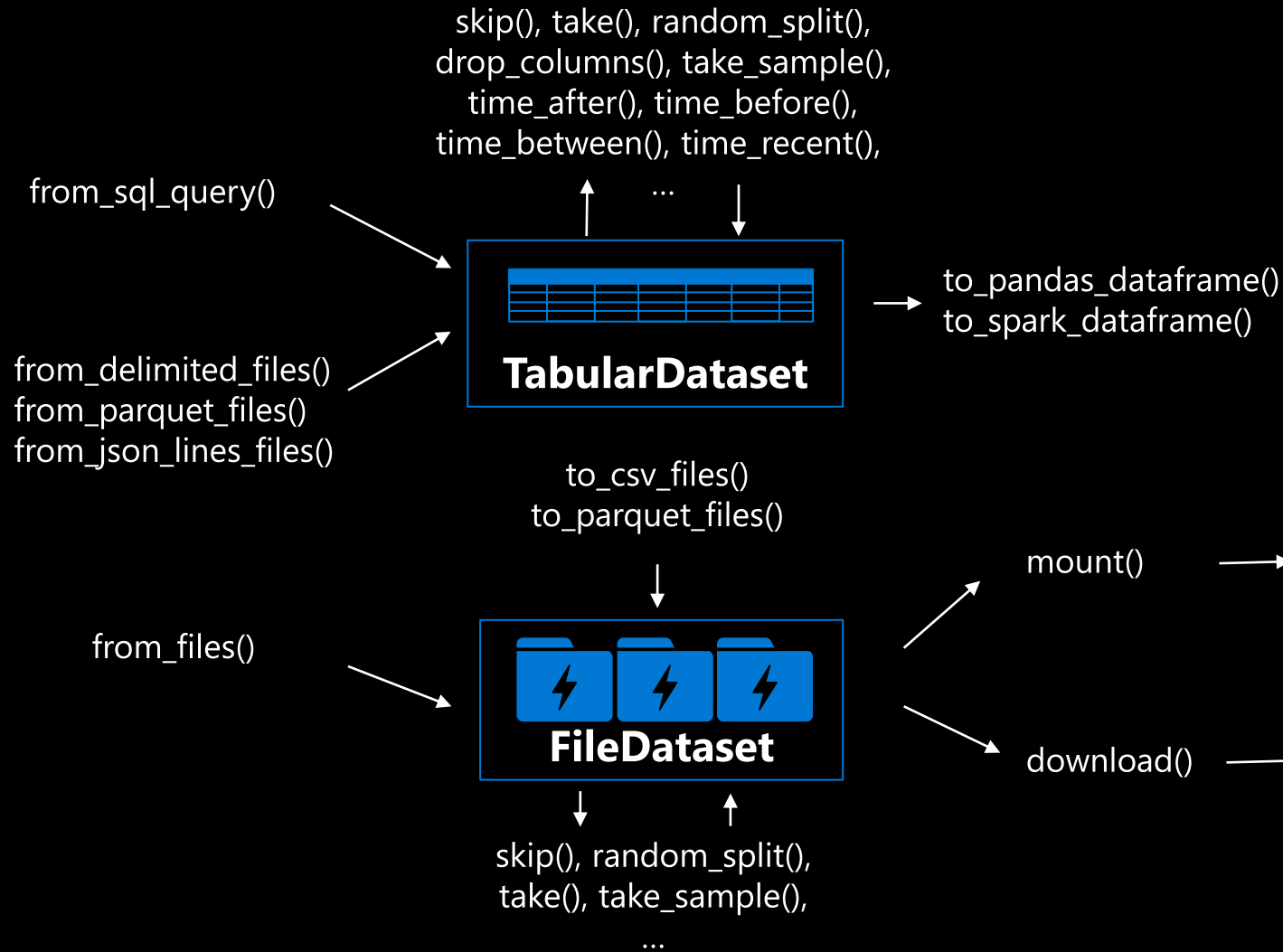


# Datasets cheat sheet

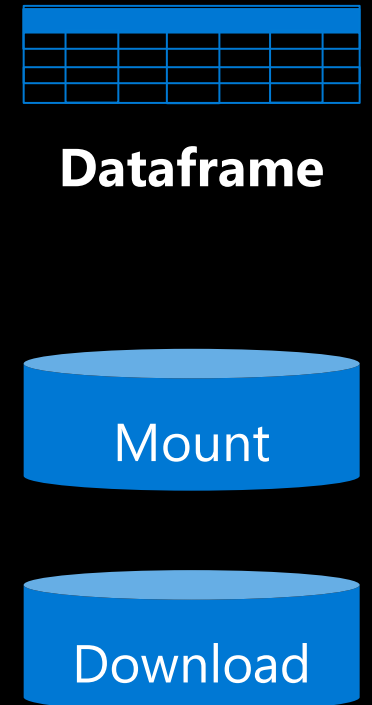
## Datastores



## Datasets



## Consumption



# Azure Machine Learning Tools

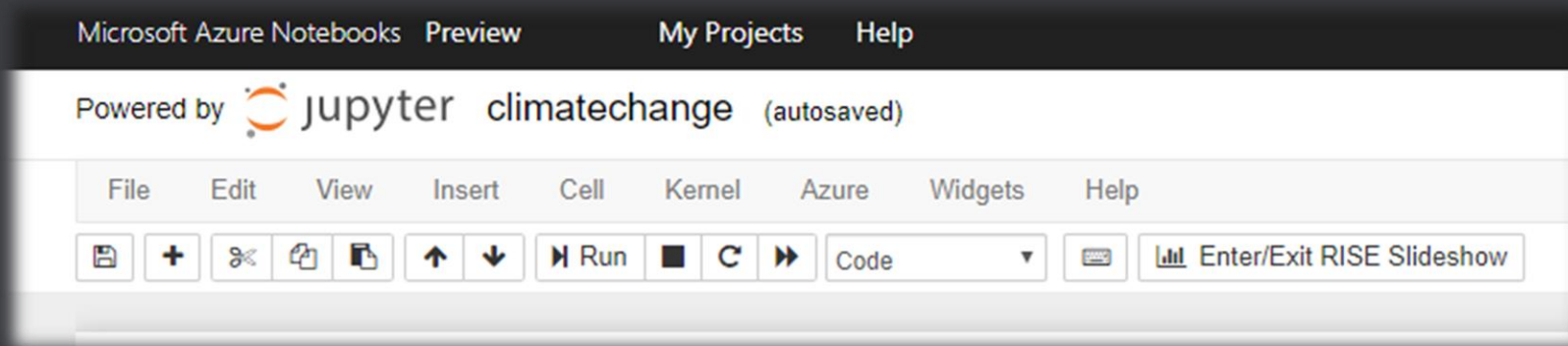
- Azure Notebooks
- Azure Databricks
- Data Science Virtual Machine
- Azure Machine Learning Studio
  - Notebooks
  - AutoML
  - Visual Interface
- Azure Machine Learning Compute Instance
- Visual Studio & VS Code with python

# Azure Notebooks



Azure Notebooks es un servicio manejado, y gratuito para desarrollar y ejecutar Jupyter Notebooks en nube y sin necesidad de instalación previa.

<https://notebooks.azure.com/>





# Azure Databricks

- Plataforma de analítica optimizada para Azure basada en Apache Spark
- Aspectos Relevantes
  - Rapidez ( Procesamiento en Memoria )
  - Fácil de usar ( Pocos Clics, Ambiente Interactivo )
  - Colaborativo (Integración con Servicios de Azure )
  - Procesamiento de datos a gran escala



# Cluster Creation

- You can create two types of clusters – *Standard* and *Serverless Pool* (see next slide)
- While creating a cluster you can specify:
  - Number of nodes
  - Autoscaling and Auto Termination policy
  - Auto Termination policy
  - Spark Configuration details
  - The Azure VM instance types for the Driver and Worker Nodes

| General Purpose               |                                   |
|-------------------------------|-----------------------------------|
| ✓ Standard_DS3_v2             | 14.0 GB Memory, 4 Cores, 0.75 DBU |
| Standard_DS4_v2               | 28.0 GB Memory, 8 Cores, 1.5 DBU  |
| Standard_DS5_v2               | 56.0 GB Memory, 16 Cores, 3 DBU   |
| Standard_D8s_v3               | 32.0 GB Memory, 8 Cores, 1.5 DBU  |
| Standard_D16s_v3              | 64.0 GB Memory, 16 Cores, 3 DBU   |
| Standard_D32s_v3              | 128.0 GB Memory, 32 Cores, 6 DBU  |
| Standard_D64s_v3              | 256.0 GB Memory, 64 Cores, 12 DBU |
| General Purpose (HDD)         |                                   |
| Standard_D3_v2                | 14.0 GB Memory, 4 Cores, 0.75 DBU |
| Standard_D8_v3                | 32.0 GB Memory, 8 Cores, 1.5 DBU  |
| Standard_D16_v3               | 64.0 GB Memory, 16 Cores, 3 DBU   |
| Standard_D32_v3               | 128.0 GB Memory, 32 Cores, 6 DBU  |
| Standard_D64_v3               | 256.0 GB Memory, 64 Cores, 12 DBU |
| Memory Optimized (Remote HDD) |                                   |
| Standard_D12_v2               | 28.0 GB Memory, 4 Cores, 1 DBU    |

Microsoft Azure

## Create Cluster

### New Cluster

Cancel Create Cluster

2-8 Workers: 28.0-112.0 GB Memory, 8-32 Cores, 1.5-6 DBU  
1 Driver: 14.0 GB Memory, 4 Cores, 0.75 DBU Cost \$0.55

Cluster Name  
MyDemoCluster

Cluster Mode ?  
Standard

Databricks Runtime Version ?  
Runtime: 5.2 (Scala 2.11, Spark 2.4.0)

Python Version ?  
3

New The default Python version for clusters v

Autopilot Options  
☒ Enable autoscaling ?  
☒ Terminate after 120 minutes of inactivity ?

Worker Type  
Standard\_DS3\_v2 14.0 GB Memory, 4 Cores, 0.75 DBU

Min Workers 2 Max Workers 8

Driver Type  
Same as worker 14.0 GB Memory, 4 Cores, 0.75 DBU

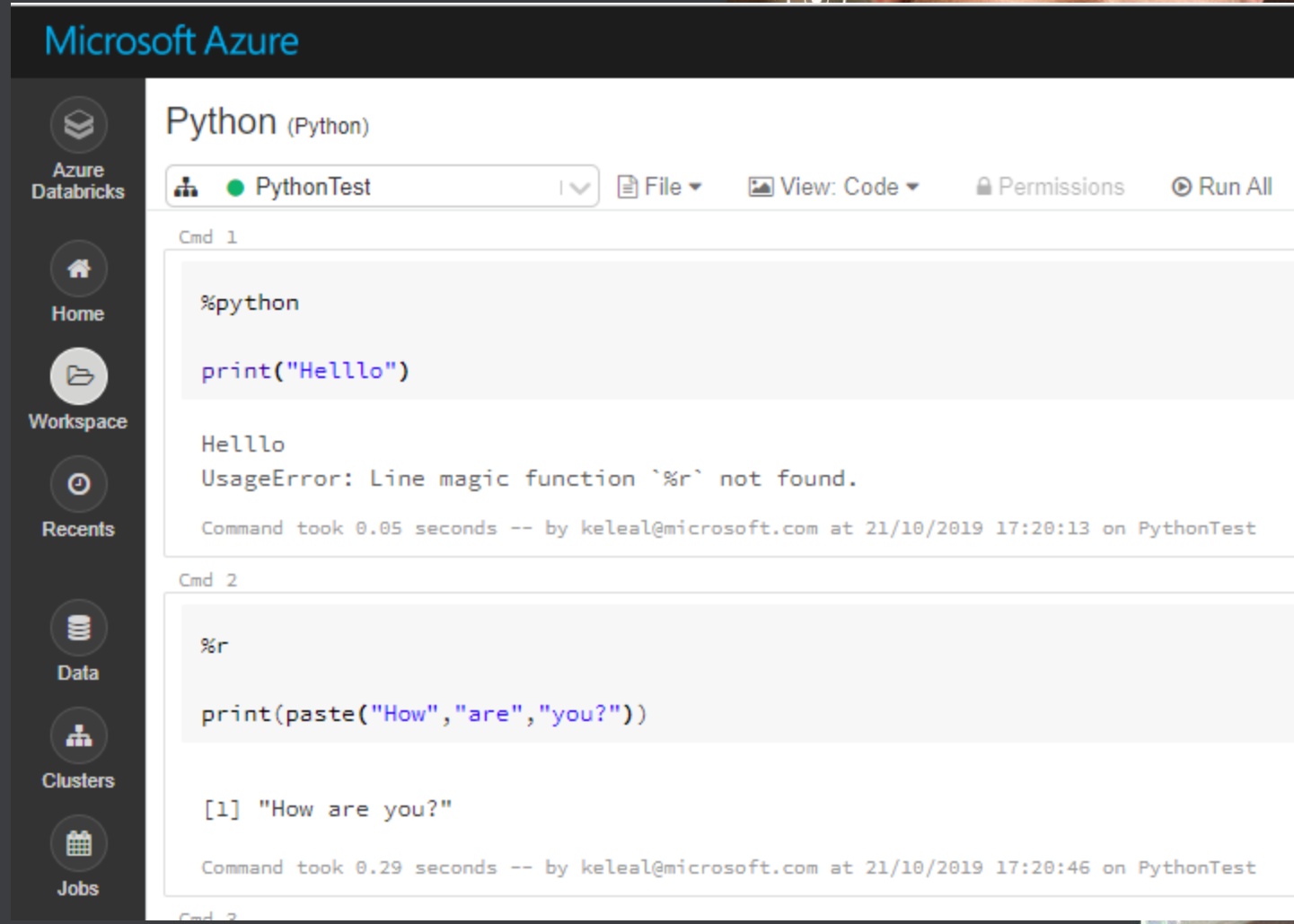
Advanced Options

Graphical wizard in the Azure Databricks portal to create a Standard Cluster

# Azure Databricks Mezclar lenguajes Microsoft

Normalmente un cuaderno está asociado a un lenguaje específico. Sin embargo, con los cuadernos de Azure Databricks, usted puede mezclar varios lenguajes en el mismo cuaderno.

- %python
- %sql
- %r
- %scala
- %sh



Microsoft Azure

Python (Python)

PythonTest | File | View: Code | Permissions | Run All

Cmd 1

```
%python
print("Hello")
```

Hello  
UsageError: Line magic function `%r` not found.

Command took 0.05 seconds -- by keleal@microsoft.com at 21/10/2019 17:20:13 on PythonTest

Cmd 2

```
%r
print(paste("How","are","you?"))
```

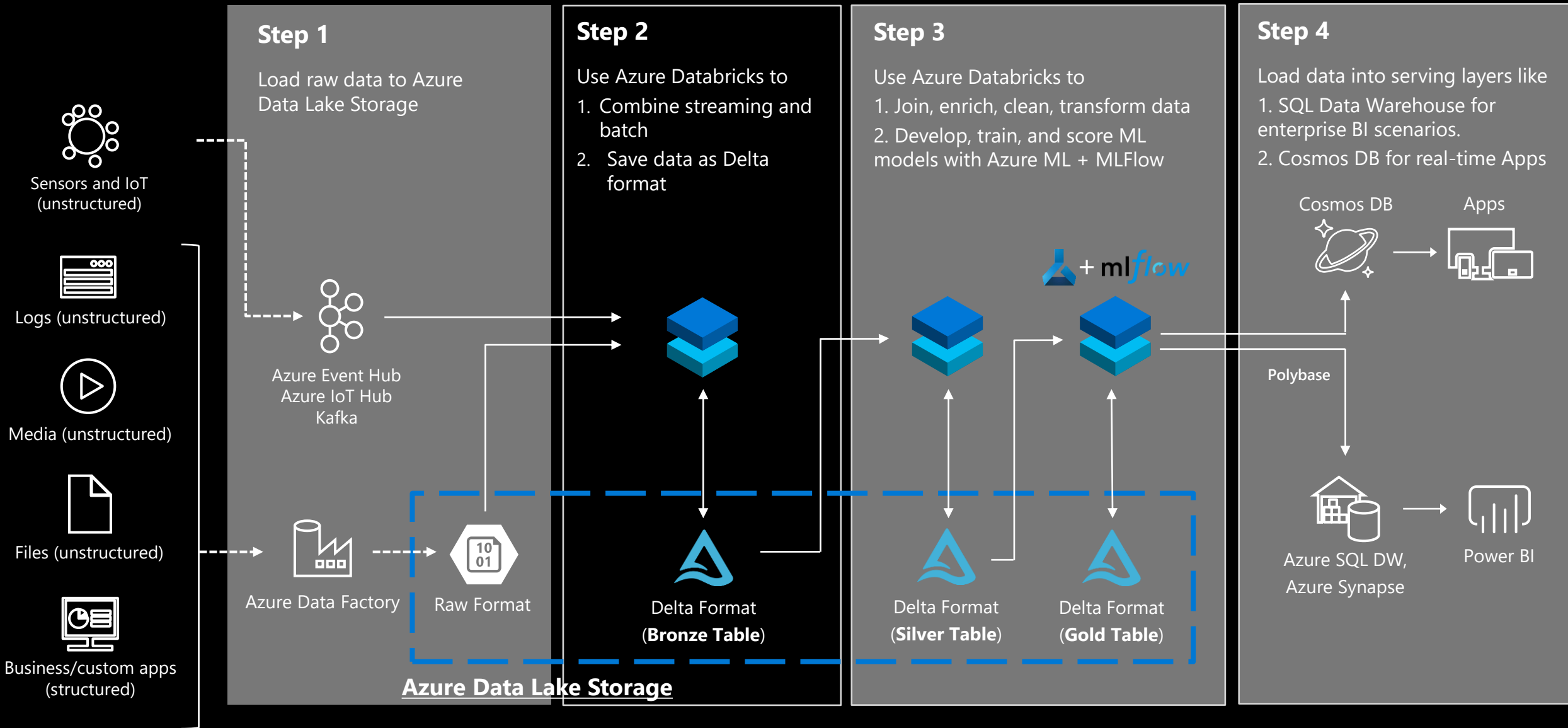
[1] "How are you?"

Command took 0.29 seconds -- by keleal@microsoft.com at 21/10/2019 17:20:46 on PythonTest

Cmd 3



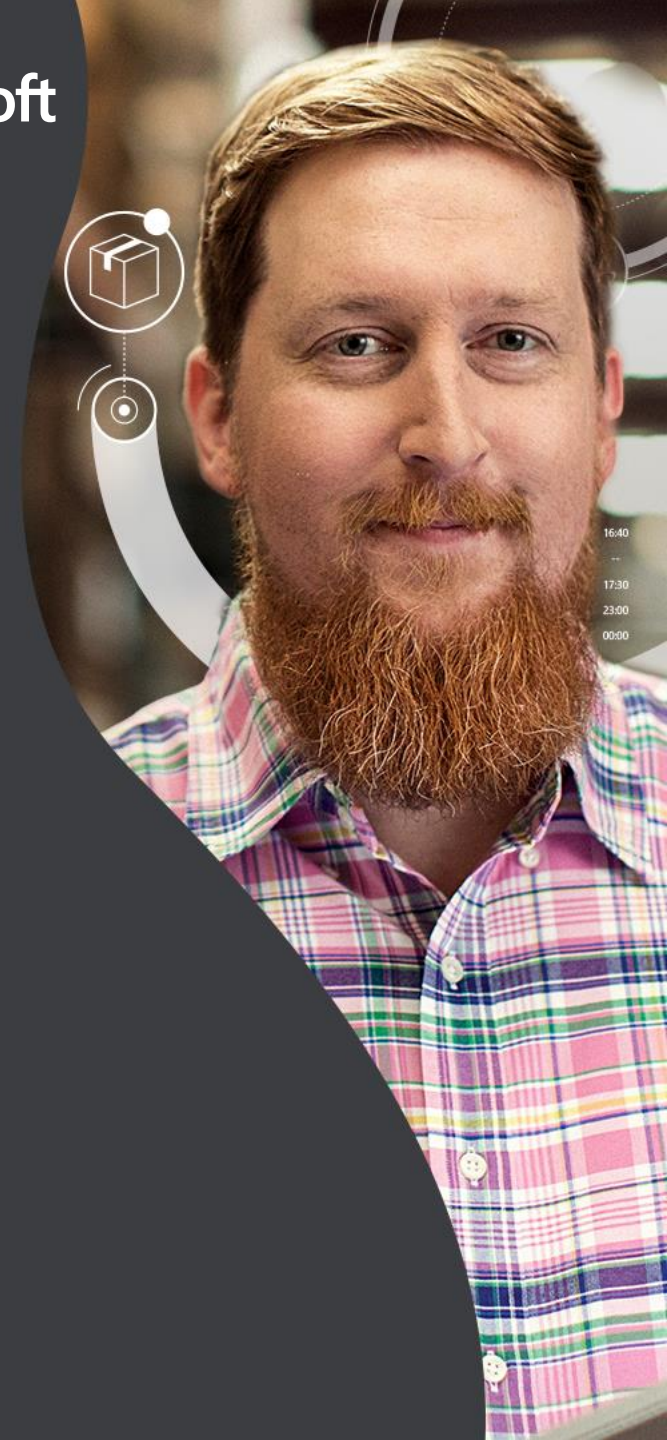
# Azure Databricks – Delta Lake at Scale on Azure



# Data Science Virtual Machine DSVM



- [Azure Machine Learning SDK](#)
- [Microsoft ML Server - Dev Edition](#) (Scalable R & Python)
- Anaconda Python
- SQL Server 2017 Dev. Edition - With In-Database R and Python analytics
- Microsoft Office 365 ProPlus BYOL - Shared Computer Activation
- Julia Pro + Juno Editor
- Jupyter notebooks
- Visual Studio Community Ed. + Python, R & node.js tools
- Power BI Desktop
- Deep learning tools e.g. Microsoft Cognitive Toolkit (CNTK, TensorFlow, Chainer, & mxnet)
- ML algorithm libraries e.g. xgboost, Vowpal Wabbit
- Azure SDKs + libraries for various Azure Cloud offerings. Integration tools are included for:
  1. Azure Machine Learning
  2. Azure Data Factory
  3. Stream Analytics
  4. SQL Data Warehouse
  5. Hadoop + Apache Spark (HDICluster)
  6. Data Lake
  7. Blob storage
  8. ML & Data Science tutorials as Jupyter notebooks



# Data Science Virtual Machine DSVM

[Dashboard](#) > [All resources](#) > [New](#) > [Marketplace](#)

## Marketplace

My Saved List

Recently created

Service Providers

### Categories

Get Started

AI + Machine Learning

Analytics

Blockchain

Compute

Containers

Databases

**Private** You have private offers available. [Click here to see.](#)

Pricing : All

Operating System : All

Publisher : All

Showing All Results



Free trial

### Data Science Virtual Machine for Linux (CentOS)

Microsoft

Virtual machine with tools for data science and machine learning



### Data Science Virtual Machine for Linux (Ubuntu)

Microsoft

Virtual machine image with deep learning frameworks and tools for machine learning and data science.



### Data Science Virtual Machine - Windows 2016

Microsoft

Development and modeling tools for AI, data science and analytics



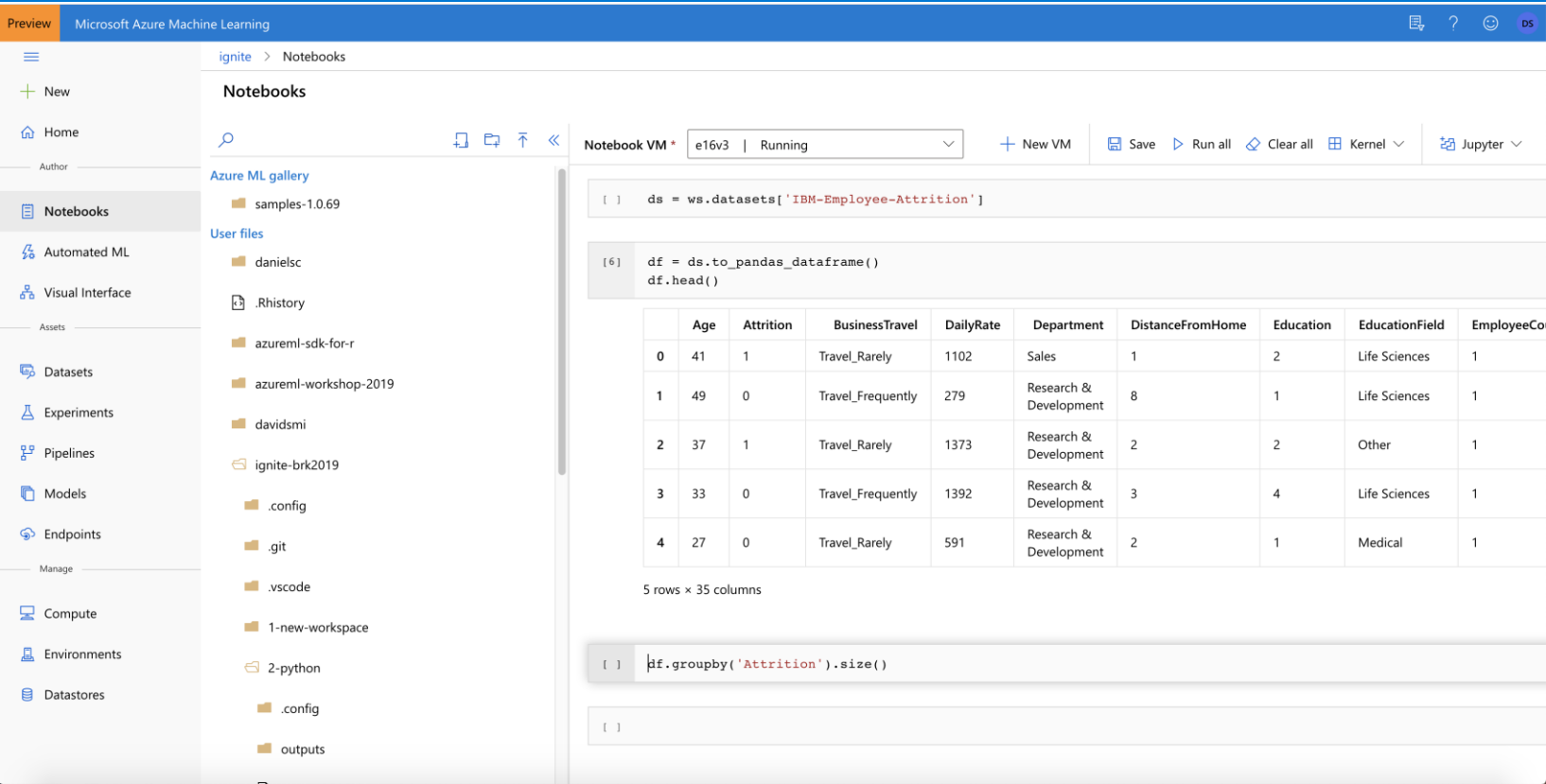
### Deep Learning Virtual Machine

Microsoft

A pre-configured environment for deep learning using GPU instances



# Demo: Azure Machine Learning



The screenshot displays the Microsoft Azure Machine Learning web interface. On the left, a sidebar contains navigation options: New, Home, Author, Notebooks, Automated ML, Visual Interface, Assets, Datasets, Experiments, Pipelines, Models, Endpoints, Manage, Compute, Environments, and Datastores. The main area is titled 'ignite > Notebooks' and shows a file explorer with 'Azure ML gallery' (samples-1.0.69) and 'User files' (danielsc, .Rhistory, azureml-sdk-for-r, azureml-workshop-2019, davidsmi, ignite-brk2019, .config, .git, .vscode, 1-new-workspace, 2-python, .config, outputs). The central pane shows a Jupyter Notebook with a 'Notebook VM' dropdown set to 'e16v3 | Running'. The notebook code includes:

```
[ ] ds = ws.datasets['IBM-Employee-Attrition']
```

```
[6] df = ds.to_pandas_dataframe()
df.head()
```

The output of the second cell is a table with 10 columns: Age, Attrition, BusinessTravel, DailyRate, Department, DistanceFromHome, Education, EducationField, and EmployeeCount. The table displays the first 5 rows of data.

|   | Age | Attrition | BusinessTravel    | DailyRate | Department             | DistanceFromHome | Education | EducationField | EmployeeCount |
|---|-----|-----------|-------------------|-----------|------------------------|------------------|-----------|----------------|---------------|
| 0 | 41  | 1         | Travel_Rarely     | 1102      | Sales                  | 1                | 2         | Life Sciences  | 1             |
| 1 | 49  | 0         | Travel_Frequently | 279       | Research & Development | 8                | 1         | Life Sciences  | 1             |
| 2 | 37  | 1         | Travel_Rarely     | 1373      | Research & Development | 2                | 2         | Other          | 1             |
| 3 | 33  | 0         | Travel_Frequently | 1392      | Research & Development | 3                | 4         | Life Sciences  | 1             |
| 4 | 27  | 0         | Travel_Rarely     | 591       | Research & Development | 2                | 1         | Medical        | 1             |

Below the table, it indicates '5 rows x 35 columns'. The notebook also shows the next code cell:

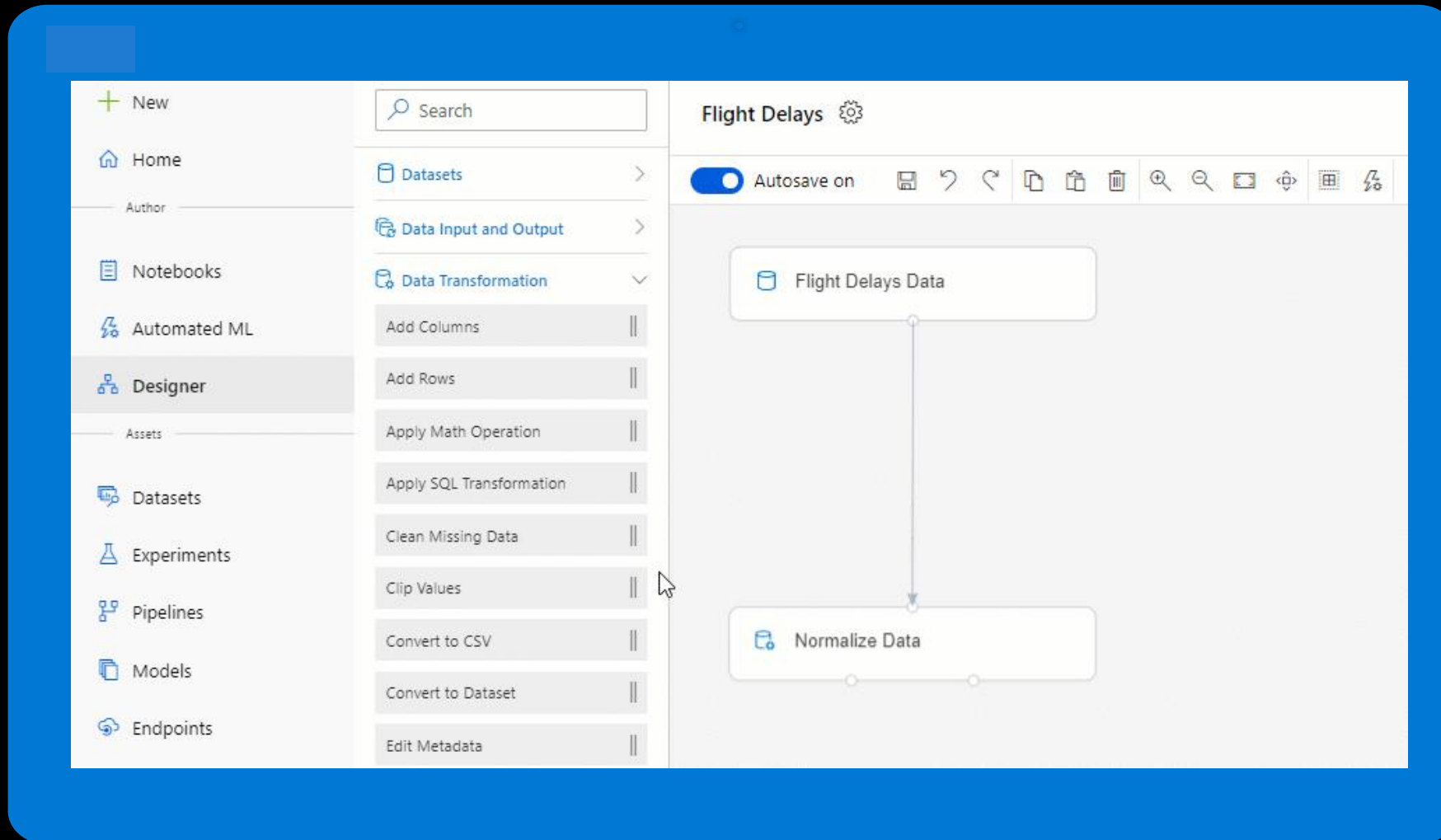
```
[ ] df.groupby('Attrition').size()
```

```
[ ]
```

Azure Machine Learning provides Python and R SDKs  
The "drag-and-drop" designer to build and deploy machine learning models.

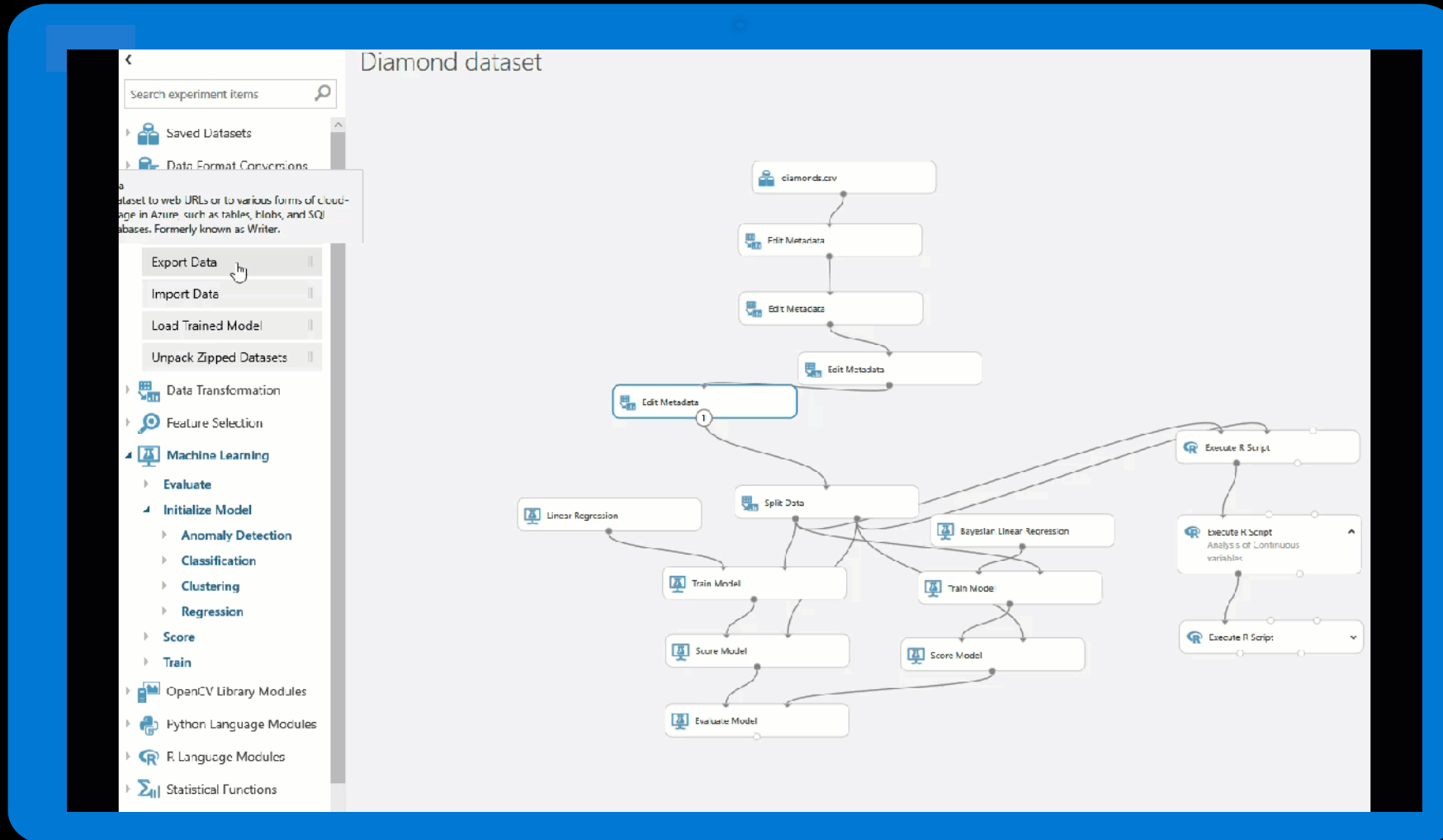


# Demo: Azure ML Designer





# Demo: Azure ML Designer



# Demo: Train and Deploy Model in Python and R

The screenshot displays a JupyterLab environment with a Python notebook titled "1-ExploreR" and a terminal window. The notebook contains five code cells:

- Load the workspace**:

```
In [2]: ws = load_workspace_from_config()
```
- Retrieve the Attrition Dataset from the workspace**:

```
In [3]: attrition = ws.datasets['IBM-Employee Attrition']
```
- Get a FileDataset for the Tabular Data**:

```
In [4]: mc = attrition.to_csv_files().mount(mc.start())
csv_file = paste(mc.mount_point, 'attrition.csv')
df = fread(file=csv_file)
mc.stop()
```
- Do some nice plots with ggplot**:

```
In [5]: gg <- ggplot(df, aes(x=Age, y=MonthlyIncome))
gg = gg + geom_point(size=0.5, color='red')
gg = gg + geom_smooth(aes())
gg = gg + facet_grid(Department ~ .)
gg`geom_smooth()` using method = 'lm'
```

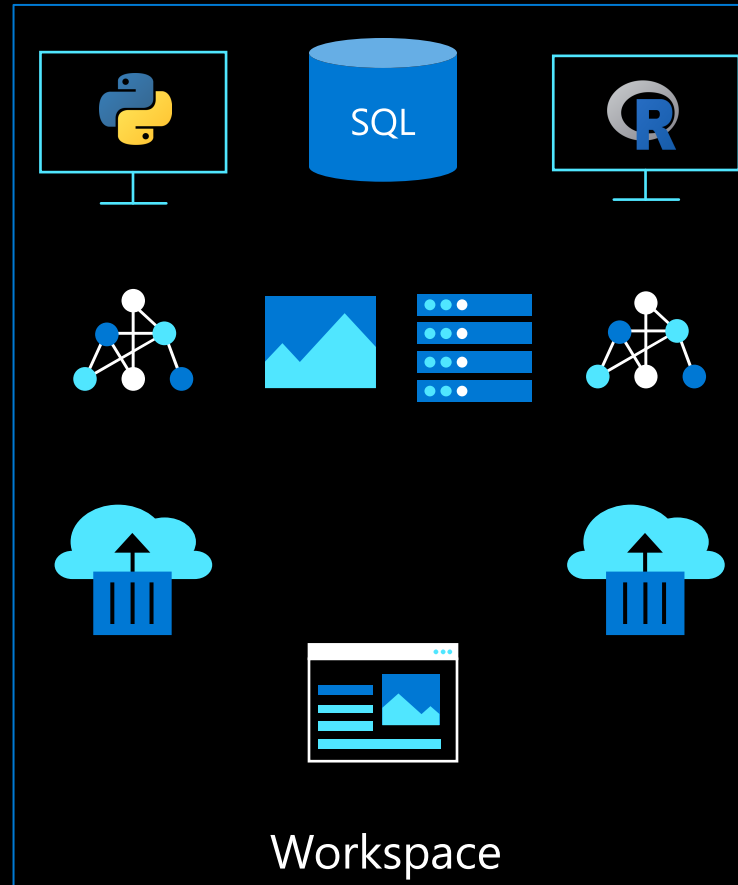
The terminal window shows the output of the training process:

```
load dataset
Age Attrition BusinessTravel ... YearsInCurrentRole YearsSinceLastPromo
0 41 1 Travel_Rarely ... 4
1 49 0 Travel_Frequently ... 7
2 37 1 Travel_Rarely ... 0
3 33 0 Travel_Frequently ... 7
4 27 0 Travel_Rarely ... 2

[5 rows x 35 columns]
train model
```

# What we did

- Created Dataset pointing to SQL data in the AzureML studio, tested it in Python
- Created a Python model using the Dataset in VSCode remote running on our Notebook VM, deploy the model as a webservice

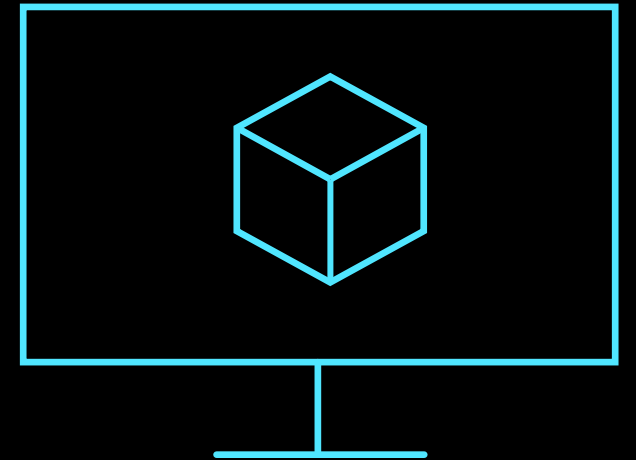


- Used the Dataset in RStudio server running on our Notebook VM to plot some charts
- Created an R model using the Dataset in RStudio server using a compute cluster and deployed it
- Created a Shiny app to score both models in Rstudio server on the Compute Instance



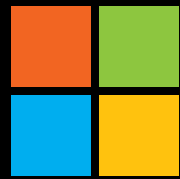
# Azure Machine Learning Compute Instance

- Collaborative
  - Share notebooks, compute instances\*, files
- Managed and Secure
  - Automated provisioning\*, RBAC\*, VNet\*, SSH policy\*, AAD-gated access
- Preconfigured for ML
  - Python, R, GPU drivers, DL Frameworks
- Fully customizable
  - Broad set of VM sizes (incl. GPU),
- Productive
  - Code & Debug with Jupyter, JupyterLab, Rstudio server, VS Code



An Azure Machine Learning compute instance (preview) is a fully-managed cloud-based workstation for data scientists.

\*new compared to Notebook VM



# Microsoft Azure

---

Invent with purpose.