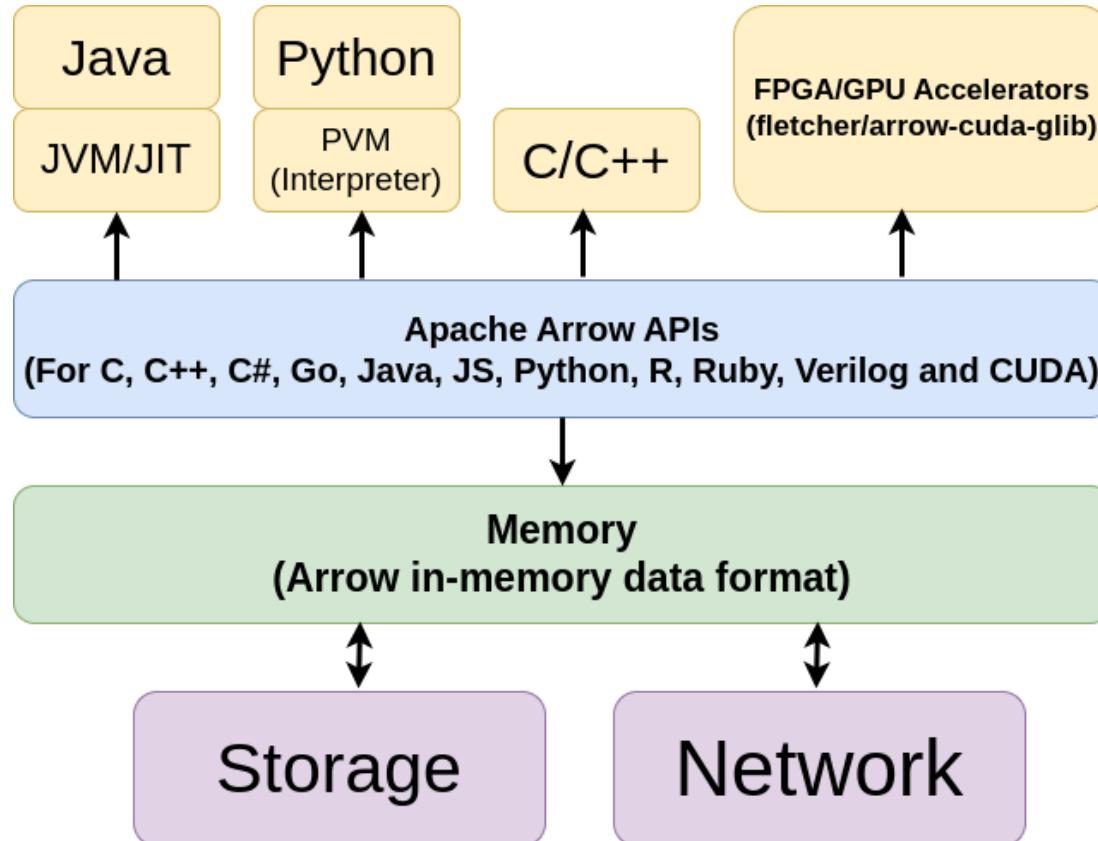


# **Apache Arrow**

Hermilo

4 de Septiembre 2025

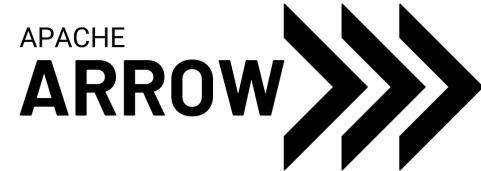
# ¿Qué es Apache Arrow?



- Estandar de intercambio de datos.
- Un formato in-memory.
- Un formato de Red.
- Un formato de Almacenamiento.
- Biblioteca de I/O
- Biblioteca de cómputo vectorizado.
- Biblioteca de manejo de dataframes.
- Un query engine.
- Administrador de datasets particionados.

# PyArrow Arrays

```
1 import pyarrow as pa
2
3 ## Arrays can be made of numbers
4 >>> pa.array([1, 2, 3, 4, 5])
5
6 ## Or strings
7 >>> pa.array(["A", "B", "C", "D", "E"])
8
9 ## Or even complex objects
10 >>> pa.array([{"a" : 5}, {"b" : 10}])
```



- Estandar de intercambio de datos.
- Un formato in-memory.
- Un formato de Red.
- Un formato de Almacenamiento.
- Biblioteca de I/O
- Biblioteca de cómputo vectorizado.
- Biblioteca de manejo de dataframes.
- Un query engine.
- Administrador de datasets particionados.