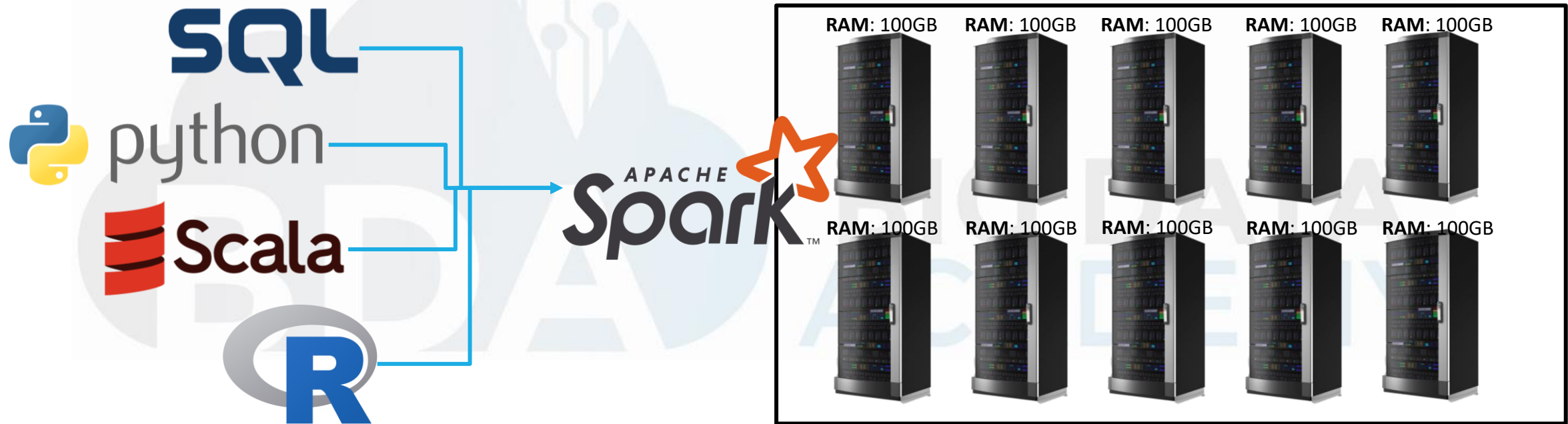

Librerías de Kafka para PySpark

BIG DATA ACADEMY

Lenguajes de Programación sobre SPARK



Los lenguajes de programación no ejecutan los comandos, sólo envían los comandos a SPARK, el cual los recibe y paraleliza su ejecución sobre el clúster

Instalación de Librerías sobre SPARK

Cuando instalamos librerías sobre SPARK, **se instalan a dos niveles, sobre el lenguaje que se usa y sobre SPARK**. Por ejemplo, si queremos procesar datos desde Kafka con Python y SPARK debemos de instalar las siguientes librerías:

SOBRE PYTHON:

- **Kafka-Python**: Librería utilitaria para que Python pueda comunicarse con Kafka

SOBRE SPARK:

- **SPARK STREAMING KAFKA**: Permite a Spark leer desde un tópico Kafka
- **SPARK SQL KAFKA**: Permite procesar con Spark los registros leídos desde Kafka

