

Taller: Usando sparklyr

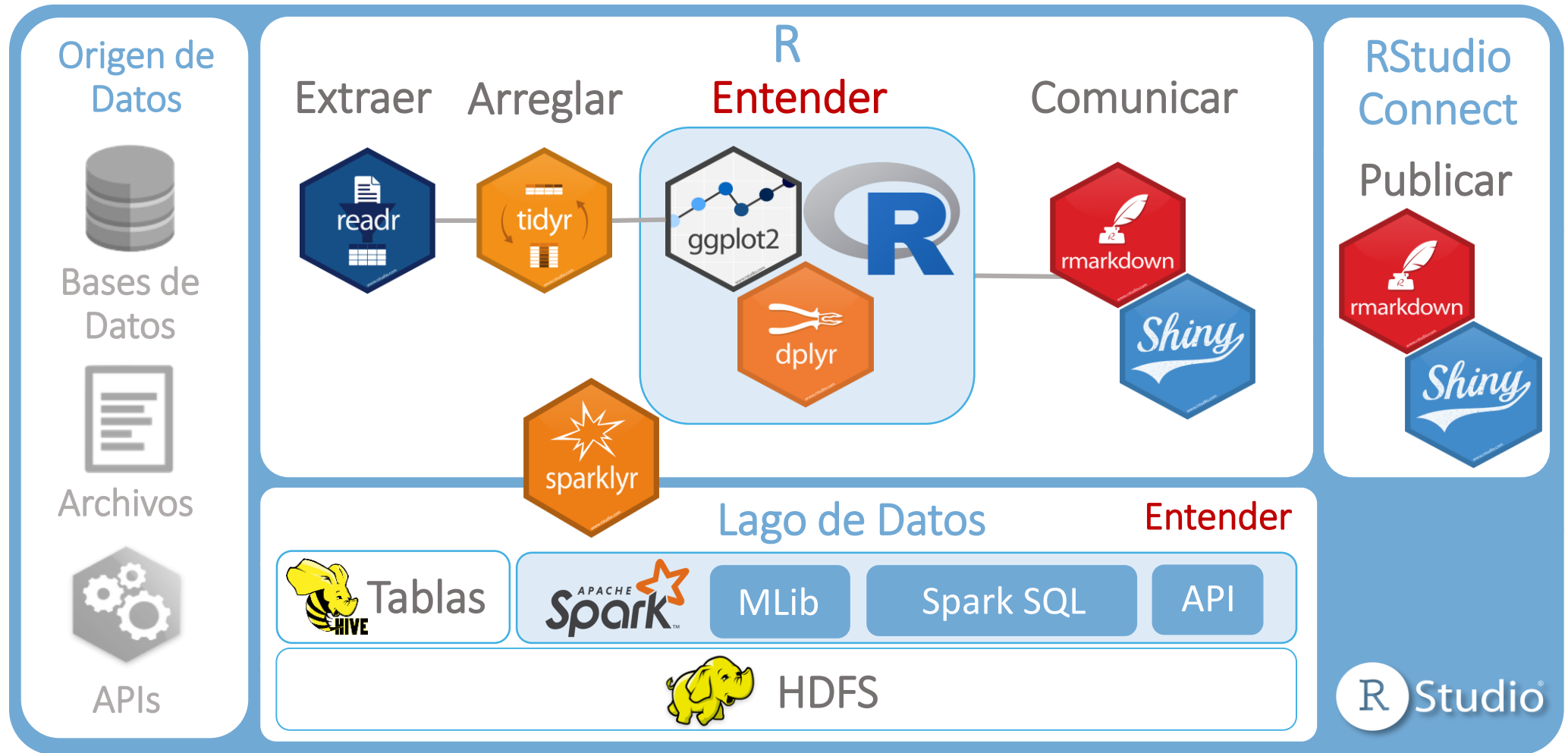
Edgar Ruiz

@theotheredgar

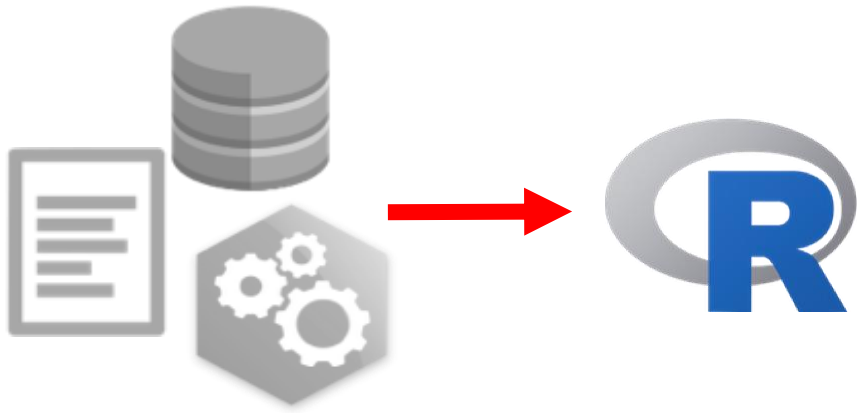
linkedin.com/in/edgararuiz

15 de marzo del 2018

Ciencia de Datos con R y Spark



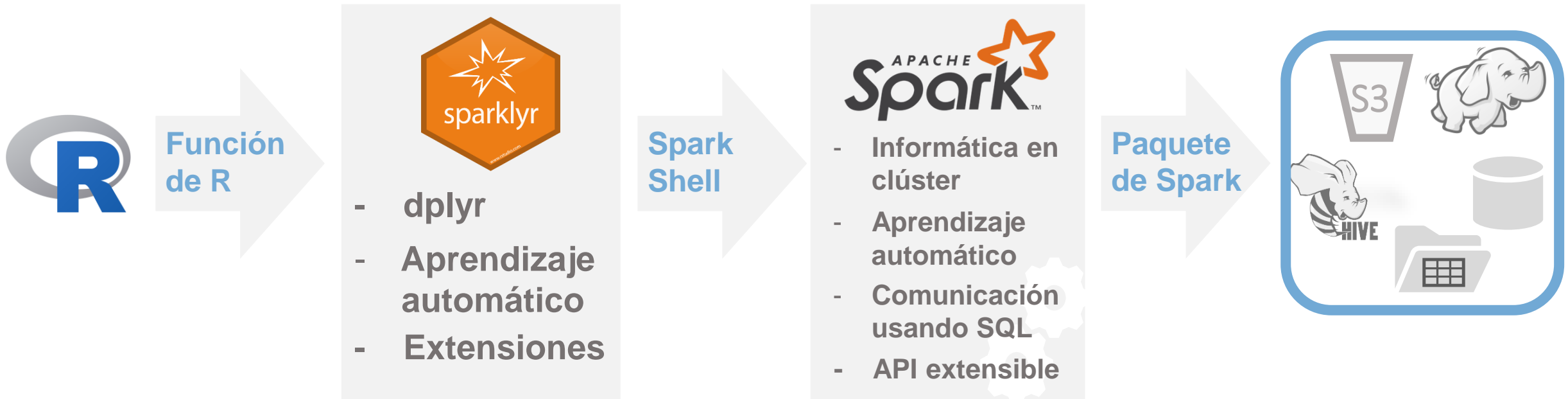
La idea principal



Extraer
datos



Como funciona sparklyr



Opciones para manejar memoria

Opción 1

Los datos de cada resultado vienen desde el origen.



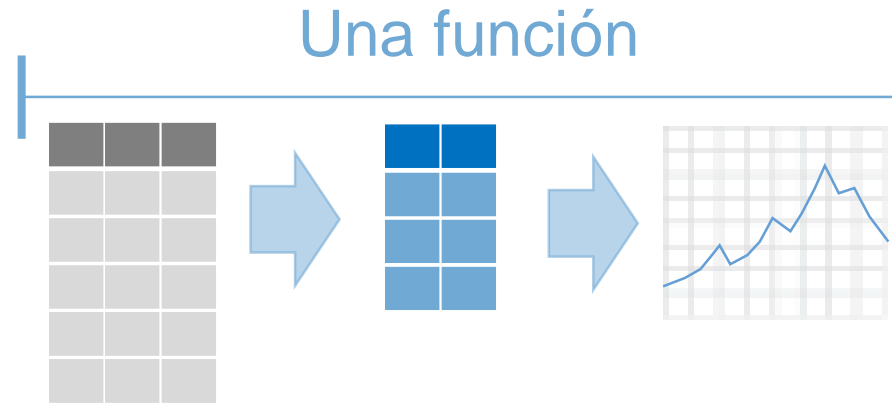
Opción 2

Copiar los datos en Spark. Los resultados están en la memoria.

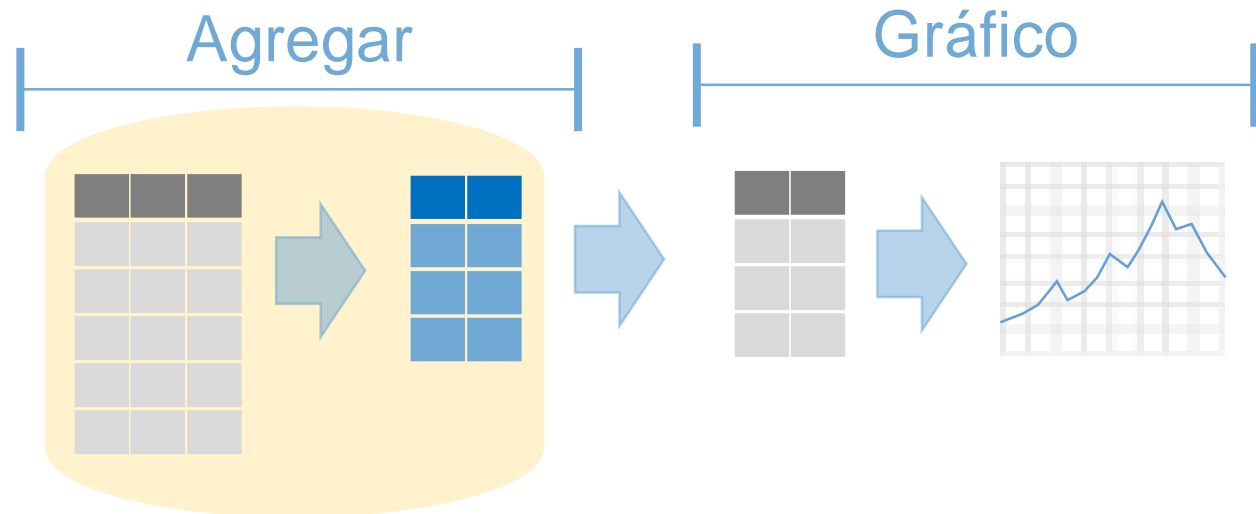


Visualizar datos

Datos locales

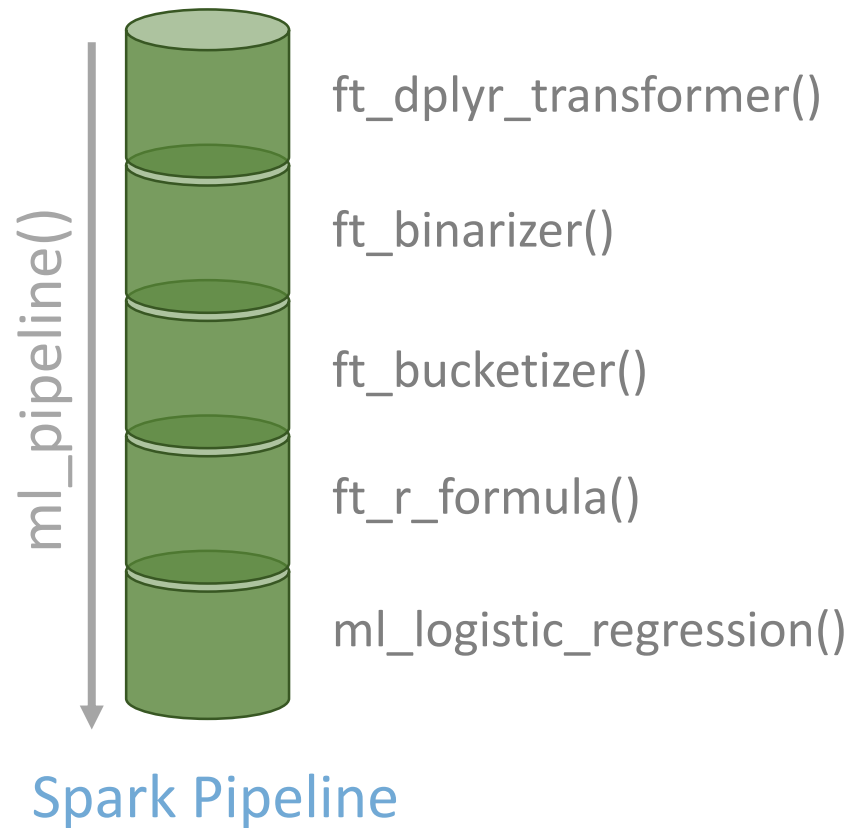


Datos lejanos

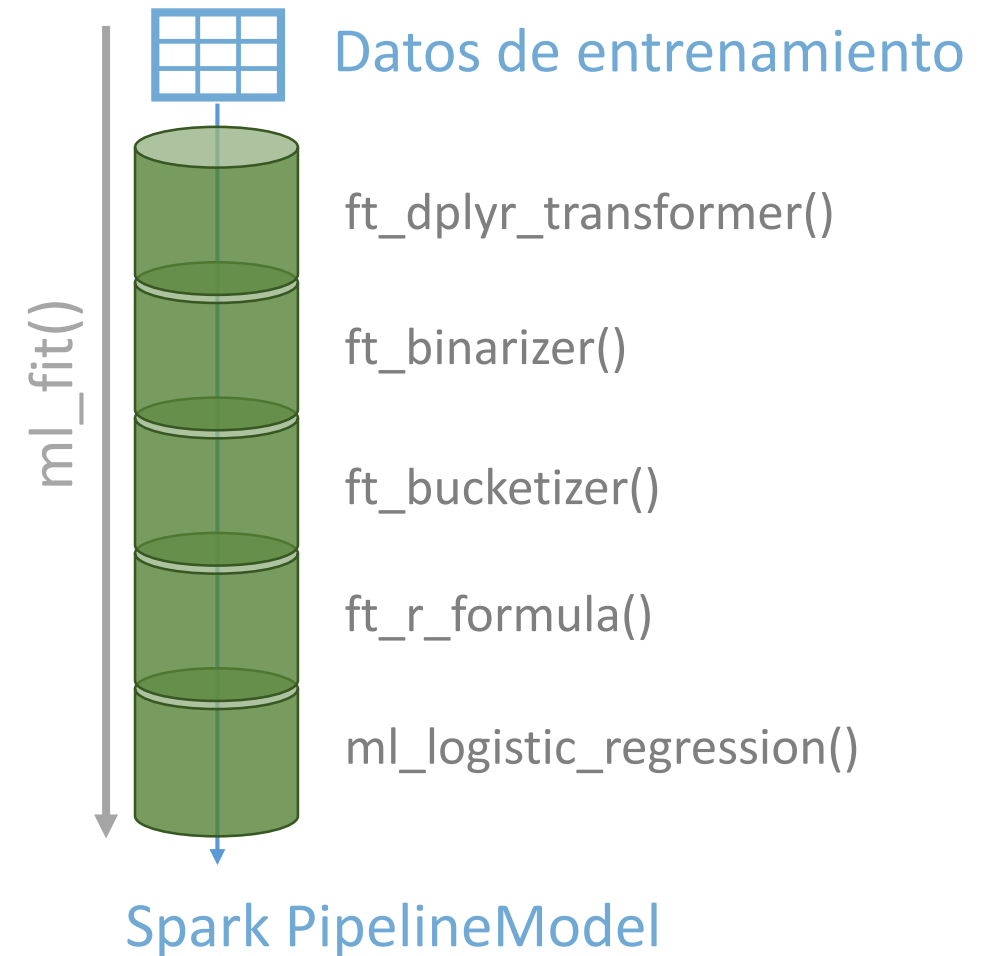


Spark Pipelines (Tubería)

Estimator



Transformer



Links

Sitio official de sparklyr: <http://spark.rstudio.com/>

Sitio official de Spark: <http://spark.apache.org/>