# EJERCICIOS

### EJERCICIO 1

Reproduce el ejemplo de DataFrames sobre el fichero de El Quijote en una máquina virtual con PySpark y Ubuntu24.

### EJERCICIO 2

Reproduce el ejemplo de DataFrames sobre el fichero de [El Quijote](https://aitor-medrano.github.io/iabd2223/spark/resources/el_quijote.txt) en Google Colab (el fichero el\_quijote.txt tiene que estar en drive y desde Google Colab montar tu drive para poder leer el fichero).

### EJERCICIO 3

Reproduce el ejemplo de DataFrames sobre el fichero de [El Quijote](https://aitor-medrano.github.io/iabd2223/spark/resources/el_quijote.txt) en tu equipo con Visual Studio Code.

### EJERCICIO 4

Reproduce el ejemplo de DataFrames sobre el fichero de [El Quijote](https://aitor-medrano.github.io/iabd2223/spark/resources/el_quijote.txt) en Databricks y visualiza el resultado en Spark UI dentro de Databricks.

### EJERCICIO 5

Crea manualmente un clúster con Spark y Hadoop y tal y como se indica en el documento "3\_2Cluster Spark con Hadoop" y reproduce los dos ejemplos que se indican.

Abre en una terminal una spark-shell (pyspark) y después abre Spark UI (http://master:4040).

### EJERCICIO 6

Crea un clúster con Spark y Hadoop utilizando contenedores docker y tal y como se indica en el documento "3\_2Cluster Spark con Hadoop" y reproduce los dos ejemplos que se indican.

### EJERCICIO 7

Crea un clúster con Spark y Hadoop utilizando AWS y tal y como se indica en el documento "3\_2Cluster Spark con Hadoop" y reproduce los dos ejemplos que se indican.