

Fundamentos de Apache Spark

1. Funcionalidades Básicas y RDDs

Ejercicio 1

Completa los ejercicios del notebook “Ejercicio_práctico_RDDs”. Has de sustituir los comentarios por el código que realice las funciones indicadas.

Solución: *Ejercicio_Solución_RDDs.ipynb*

2. Spark DataFrames y Apache Spark SQL

Ejercicio 1

Completa los ejercicios del notebook “Ejercicio_DataFrames_SQL”. Has de sustituir los comentarios por el código que realice las funciones indicadas.

Solución: *Ejercicio_Solución_DataFrames.ipynb*

3. Funciones avanzadas en Apache Spark

4. Analítica avanzada con Apache Spark

Ejercicio 1

En este ejercicio vas a tener que cargar el dataset de “data/stocks_price_final.csv”, con el esquema correcto de datos, y realizar los siguiente:

- Eliminar o renombrar la variable market.cap
- Filtrar los datos donde el valor de “open” es nulo y eliminar esas filas
- Calcular el mínimo y máximo de data, open, close y adjusted
- Convertir un DataFrame de Spark en un DataFrame de pandas
- Calcular la media de las variables opne, close y adjusted por industria
- Generar un gráfico de líneas donde se muestre la media de open por industria
- Generar un heatmap con seaborn donde se muestre la correlación entre las medias de open, close y adjusted
- Guardar en un archivo parquet una selección de datos

Solución: *Ejercicio_Solución_Analitica datos con PySpark.ipynb*