

Práctica 10

Análisis personal obligatorio

Jordi Blasco Lozano
Computación de alto rendimiento
Grado en Inteligencia Artificial

¿Qué pasos seguiste para construir tu solución y por qué?

Lo primero que hice al leer la practica fue intentar entender spark y la Shell de scala, para esto me puse a mirar todos los ejemplos propuestos y las diapositivas de clase sobre el middleware y sobre cómo funcionaba spark. Probé la ejecución mediante del contenedor scala que se proporcionan en los ejemplos cuando ya comprendía que comandos se usaban cree un Dockerfile para que me fuera más fácil volverlo a ejecutar. Y finalmente cuando ya conseguí entender cómo funcionaba tanto el contenedor spark como la Shell de scala me puse a crear el archivo.scala, y con ayuda de chatgpt fui integrando las funciones que necesitaba para realizar la práctica.

¿En qué parte tuviste dificultades y cómo las resolviste?

Lo más difícil fue entender el funcionamiento de Spark, pese a que realmente no es tan complicado, como es una herramienta totalmente nueva que nunca había visto, entender cómo funciona fue lo más complicado. La shell de scala al ser tan diferente a lo que estaba acostumbrado también supuso un reto tedioso. Pero entre ejemplos, diapositivas, chatGPT y tutoriales supe entender la herramienta correctamente.

¿Qué crees que está ocurriendo “por dentro” cuando Spark ejecuta tu código?

Lo primero que ocurre es la generación del trazado más optimo para resolver todos los pasos que se planteen. Después se cargan los datos el texto en diferentes bloques, de esta manera se paraleliza la tarea, cada bloque se procesa por un executor que aplica una función Map para convertir el texto en tuplas ('texto', 1). Finalmente se reagrupan las tuplas iguales y se suman todas las ocurrencias de las palabras iguales. Al acabar el resultado de este conteo se guarda en HDFS.

¿Qué aprendiste sobre el middleware distribuido que no conocías antes?

Sobre el middleware distribuido solo sabíamos un poco de teoría general vista en sistemas operativos y distribuidos el cuatrimestre anterior. Poner en practica un ejemplo de middleware clasificador es algo completamente nuevo que he aprendido gracias a la práctica y aprender a manejar grandes volúmenes de datos de manera sencilla para poder clasificar la información es algo muy útil en los entornos de inteligencia artificial.