

Práctica de recuperación

El objetivo de esta práctica de recuperación es trabajar con el *Partitioner* y el *ToolRunner* de Hadoop. Para ello, responde a las siguientes preguntas:

1. Explica qué es y para qué sirve tanto el *Partitioner* como el *ToolRunner* de Hadoop.
2. Busca en Internet un *dataset* interesante que te permita demostrar el funcionamiento del *Partitioner* y el *ToolRunner*. Describe el *dataset* y justifica por qué es apropiado para esta práctica.
3. Propón un ejercicio no trivial en el que usar un *Partitioner* y el *ToolRunner* (para pasar algún parámetro por línea de comandos) sobre el *dataset* que has elegido en la pregunta anterior. Asegúrate de que es un ejercicio novedoso y no está resuelto en Internet.
4. Escribe el código java que resuelva el ejercicio propuesto. Explica el funcionamiento de cada clase así como las precondiciones y postcondiciones de todos los métodos.
5. Muestra tantos ejemplos de ejecución para demostrar que el código funciona correctamente (incluyendo *edge cases*).