



CURSO: CMP 0575 - TÓPICOS 2 (DATA MINING)
COLEGIO: POLITÉCNICO
Semestre: Primer Semestre 2018/2019 NRC: 1068

Tarea 1: Ejercicio usando la técnica MapReduce

Problema:

1. Se desea implementar un programa que permita la lectura de un fichero (Word, texto) y devuelva un índice de palabras populares empleadas en el documento (la estructura del índice tiene que ser en el formato *<palabra, frecuencia>*).

Requisitos:

- Es obligatorio el uso de la filosofía MapReduce, en este caso sobre una arquitectura simple (no distribuida).
- Cargar al D2L los códigos implementados (fichero compactado) dentro del plazo de entrega de las distintas fases de evaluación.

Evaluación:

- Fase 1:
 - Aplicar correctamente el MapReduce sobre un fichero de entrada y obtener la salida que da solución a la problemática planteada. (50% de la nota final de la tarea)
- Fase 2:
 - Aplicar elementos de distribución a nivel de tareas Map y Reduce, y obtener la salida que da solución a la problemática planteada. (50% de la nota final de la tarea)

Nota: En cada fase de evaluación el profesor aplicará puntos de chequeo sobre el código implementado.