

## CDAP - Práctica 8

### Introducción a Hadoop y Map/reduce

#### Ejecución ejemplo *wordcount*

La instalación de hadoop proporciona una implementación del ejemplo sencillo *wordcount*. A continuación se proporciona un ejemplo de uso.

Abrimos un terminal y comprobamos que todo funciona correctamente:

```
$ cd /usr/local/hadoop/share/hadoop/mapreduce
$ hadoop jar hadoop-mapreduce-examples-3.1.0.jar wordcount
Usage: wordcount <in> [<in>...] <out>
```

Vemos que la ejecución de *wordcount* requiere un parámetro de entrada y otro de salida. Primero creamos dos ficheros en el directorio home del usuario:

```
$ echo "Hello world in HDFS" > /home/hduser/testfile1
$ echo "Hadoop word count example in HDFS" > /home/hduser/testfile2
```

Y ahora creamos el directorio para los datos en el sistema de ficheros hdfs:

```
$ hdfs dfs -mkdir /simpletest
$ hdfs dfs -mkdir /simpletest/input
```

Podemos comprobar la existencia del directorio:

```
$ hdfs dfs -ls /simpletest
```

Copiamos los dos ficheros de prueba al sistema de ficheros hdfs:

```
$ hdfs dfs -put /home/hduser/testfile1 /simpletest/input
$ hdfs dfs -put /home/hduser/testfile2 /simpletest/input
```

Podemos examinar parámetros de los ficheros en el sistema hdfs:

```
$ hdfs fsck /simpletest/input -files -blocks -locations -racks
```

Ejecutamos el ejemplo:

```
$ hadoop jar hadoop-mapreduce-examples-2.8.0.jar wordcount
/simpletest/input /simpletest/output
```

Obsérvese que el directorio de salida es creado por el ejemplo. Si existe previamente obtendremos un error.

Podemos revisar el directorio de salida:

```
$ hdfs dfs -ls /simpletest
```

Observar el resultado:

```
$ hdfs dfs -cat /simpletest/output/part-r-00000
```

O recuperarlo del sistema hdfs:

```
$ hdfs dfs -get /simpletest/output/part-r-00000
/home/hduser/wordcount.txt
```

## Ejercicio

Se propone generar tres ficheros de texto con los tres primeros capítulos de “El Quijote” de Miguel de Cervantes (disponible en la biblioteca Proyecto Gutenberg). Ejecutando el programa *wordcount* de forma similar al ejemplo anterior, hallar el número de veces que aparece la palabra “Quijote”.