Acceso a nodo Hadoop en CPD del CUMe.

Índice

1. [*Introducción 1*](#_TOC_250005)
2. [*Envío de datos de nuestro equipo al nodo Hadoop 2*](#_TOC_250004)
   1. [Acceso al nodo: 2](#_TOC_250003)
   2. [Instrucciones básicas de Hadoop 2](#_TOC_250002)
      1. Listar contenido de HDFS 2
      2. Lanzar una ejecución MapReduce 2
      3. Consulta de salida del procesamiento 3
      4. Obtener la salida del procesamiento a un fichero 3
   3. [Mover datos del nodo hadoop a nuestro equipo 3](#_TOC_250001)
3. [*Consideraciones a tener en cuenta de cómo crear el fichero.jar 3*](#_TOC_250000)

# Introducción.

Para la realización de la práctica 2 de la asignatura Sistemas de Información los alumnos deberán utilizar un nodo del CPD del Centro Universitario de Mérida donde se ha instalado previamente Hadoop.

El nodo está provisto con las librerías necesarias para poder ejecutar la práctica bajo el paradigma MapReduce.

# Envío de datos de nuestro equipo al nodo Hadoop.

#> scp fichero.jar hadoop@158.49.96.124:/home/hadoop

Se almacenará el fichero.jar en el directorio /home/hadoop del nodo Haddop.

## Acceso al nodo:

#> ssh [hadoop@158.49.96.124](mailto:hadoop@158.49.96.124) (pass: Hadoop\*CUMe23)

## Instrucciones básicas de Hadoop.

* + 1. Listar contenido de HDFS

#> hadoop fs -ls /

* + 1. Lanzar una ejecución MapReduce

#> hadoop jar fichero.jar paquete.aplicacion.claseprincipal /practica2/input

/practica2/salida

Como podemos ver se lanza el trabajo a través de fichero.jar que hemos subido al nodo, indicando el paquete que contiene la clase principal y la propia clase. Como argumentos del programa tenemos la ruta donde se encuentran los ficheros a procesar y la ruta donde dejaremos la salida del procesamiento.

* + 1. Consulta de salida del procesamiento

#> hadoop fs -cat /practica2/salida/par\*

* + 1. Obtener la salida del procesamiento a un fichero

#> hadoop fs -get /practica2/salida/par\* .

Con ello el fichero part\* se descargaría de HDFS al directorio /home/hadoop

## Mover datos del nodo hadoop a nuestro equipo

Desde una terminal de nuestro equipo ejecutar:

#> scp hadoop@158.49.96.124:/home/hadoop/par\* ruta\_windows/linux

# Consideraciones a tener en cuenta de cómo crear el fichero.jar

En el nodo Hadoop instalado en el CPD del CUMe no se encuentran las librerías Java necesarias para poder ejecutar ciertos programas, por ello es necesario construir el fichero.jar incluyendo las librerías necesarias. Para ello se debe incorporar la siguiente directiva en la etiqueta build. Una vez compilado el proyecto maven generará un fichero.jar mayor de lo habitual, pero incluirá todas las librerías necesarias para la correcta ejecución en el nodo Hadoop.

**<build>**

**<plugins>**

**<plugin>**

**</plugin>**

**<plugin>**

**<groupId>org.apache.maven.plugins</groupId>**

**<artifactId>maven-compiler-plugin</artifactId>**

**<configuration>**

**<encoding>${project.build.sourceEncoding}</encoding>**

**<source>1.7</source>**

**<target>1.7</target>**

**</configuration>**

**<artifactId>maven-assembly-plugin</artifactId>**

**<executions>**

**<execution>**

**<phase>package</phase>**

**<goals>**

**</goals>**

**</execution>**

**</executions>**

**<configuration>**

**<descriptorRefs>**

**<goal>single</goal>**

**</build>**

**</plugin>**

**</plugins>**

**<descriptorRef>jar-with-dependencies</descriptorRef>**

**</descriptorRefs>**

**</configuration>**