

Tutorial Hadoop

1. Introducción

En esta guía vamos a encontrar las instrucciones para ejecutar un programa de prueba de Hadoop. El programa que vamos a correr cuenta las ocurrencias de cada palabra en un archivo de texto. Es importante que necesitas tener instalado un entorno de desarrollo Java para poder seguir este tutorial.

2. Ejecutar un programa sencillo en Hadoop

Para descargar Hadoop debemos ir al siguiente link:

<https://hadoop.apache.org/releases.html>

Y bajar la última versión de Hadoop. Se debería descargar un archivo comprimido que en su interior tiene las carpetas **bin**, **src**, **lib**, entre otros. Este archivo tiene todo lo necesario para correr un programa en Hadoop de forma local (algo especialmente útil para *debuggear* programas que corren sobre Hadoop).

Luego debes tomar el programa **Ejemplo01.java** que está subido al repositorio del curso junto a esta guía. Luego, lo debes mover a la carpeta que descargaste en el punto anterior y correr los siguientes comandos:

```
bin/hadoop com.sun.tools.javac.Main Example01.java
```

```
jar cf ExampleHadoop.jar Example01*.class
```

Donde estos comandos te ayudan a compilar el programa en un **.jar** (el formato de los ejecutables de Java). Una vez compilado podemos ejecutar el programa con el siguiente comando.

```
bin/hadoop jar ExampleHadoop.jar Example01 <input_folder> <output_foldeer>
```

Donde **<input_folder>** es una carpeta que contiene un archivo de texto en su interior, al que le vamos a contar las palabras. **<output_folder>** es la carpeta que será creada donde guardaremos el *output* del programa.