

Instalación de Apache Spark

Para instalar spark en local basta con descargarse un fichero con la versión precompilada. La descarga se puede hacer desde

<http://spark.apache.org/downloads.html>

Spark está escrito en Scala y se ejecuta bajo una máquina virtual de Java.

Por ello para ejecutar Apache Spark, es necesario tener instalado Java 6 o superior.

Descargamos Apache Spark:



Download Apache Spark™

1. Choose a Spark release:
2. Choose a package type:
3. Choose a download type:
4. Download Spark: [spark-2.1.0-bin-hadoop2.7.tgz](#)
5. Verify this release using the [2.1.0 signatures and checksums](#) and [project release KEYS](#).

Note: Starting version 2.0, Spark is built with Scala 2.11 by default. Scala 2.10 users should download the Spark source package and build with [Scala 2.10 support](#).

Dentro de la página de descarga seleccionamos la última release. Aquí se nos presentan varias opciones:

- Código fuente listo para compilar por nosotros mismos que permite el uso de la versión de Hadoop que deseemos.
- Pre-compilado a falta de la dependencia de Hadoop.
- Pre-compilado con diferentes versiones de Hadoop integradas.

Cuando se haya completado la descarga, lo descomprimos y ubicamos en el directorio donde alojemos nuestros programas:

```
tar -xvf spark-2.1.0-bin-hadoop2.7.tgz
sudo mv spark-2.1.0-bin-hadoop2.7 /usr/local/spark
```

Estructura de directorios

La estructura de directorios de Spark se centra en **3 directorios**:

- **bin**: Contiene los entornos de ejecución en los que se puede arrancar, además de los comandos para lanzar aplicaciones y ejemplos. Cuando hablo de entornos de ejecución me refiero a arrancarlo usando la API de Scala (spark-shell), Python (pyspark), R (sparkR) o SQL (spark-sql).
- **conf**: En este directorio se encuentran los archivos de configuración del clúster.
- **sbin**: Scripts para levantar y parar el clúster y sus componentes.

Para probar que todo funciona correctamente, dentro de la carpeta **bin** ejecutamos el comando `./spark-shell` y tendremos como resultado una consola de Spark.