

Tipos de instalación de Apache Spark

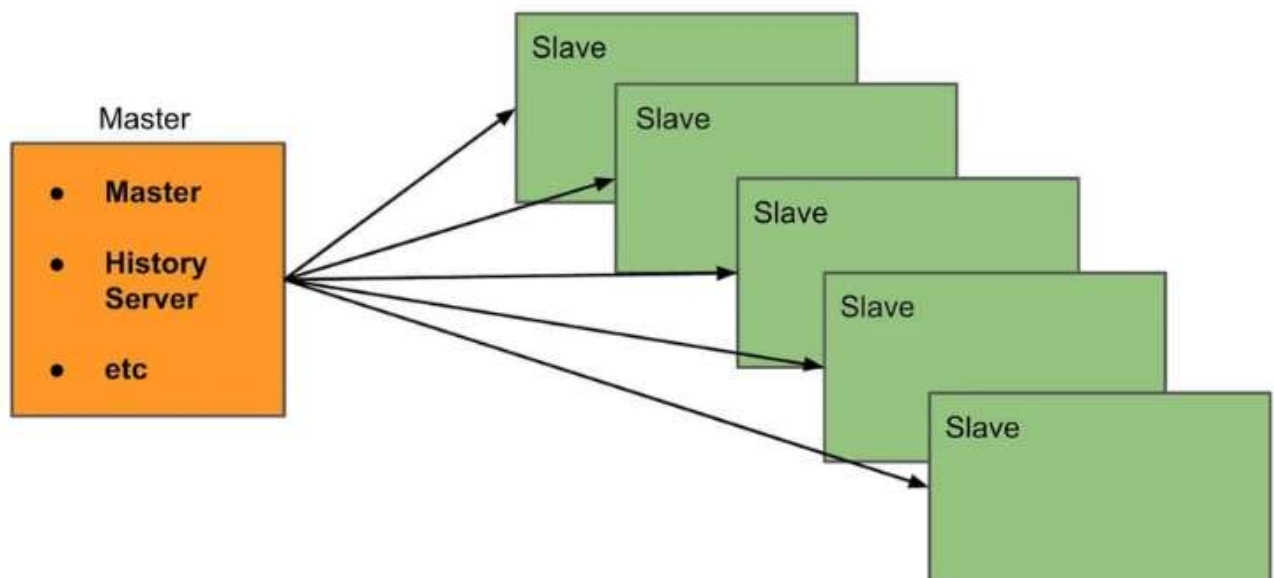
Existen varios modos de instalar Apache Spark:

- **Modo Standalone** : Tenemos el HDFS (Hadoop Distributed File System), es decir, el sistema de archivos distribuido de Hadoop, y por encima estaría Apache Spark.
- **Hadoop V1 (SIMR)** : Tenemos el HDFS y por encima Map Reduce, y por encima del mismo estaría Apache Spark.
- **Hadoop V2 (YARN)** : Tenemos nuestro HDFS y por encima un gestor de recursos, como puede ser YARN o Mesos, y por encima del mismo estaría Apache Spark. Ese gestor de recursos se va a encargar de gestionar todos los recursos de nuestro clúster.

Cómo realizar la instalación de Apache Spark

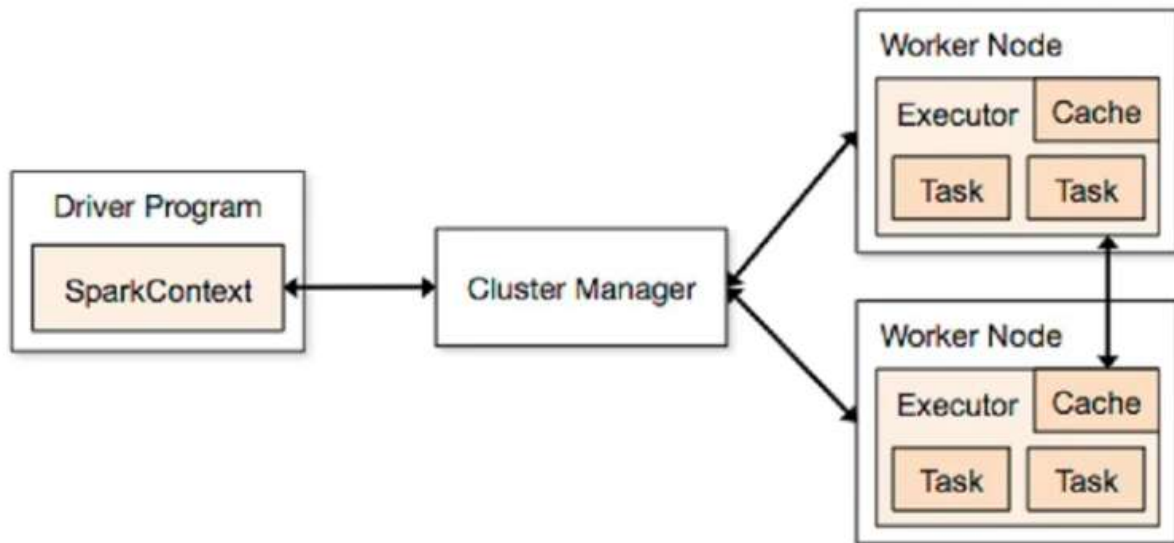
Modo Standalone

Es la manera más simple de instalar Apache Spark, porque únicamente requiere descargar una instancia o una versión compilada del mismo y añadirla a cada uno de los nodos del clúster



Spark sobre Mesos

Mesos es un gestor de recursos que se va a encargar del reparto de tareas. En esta instalación tendríamos nuestro driver o main, que va a levantar nuestro SparkContext, y el clúster manager o Mesos, que se va a encargar de repartir las tareas en los distintos Worker Node, que son las distintas máquinas de nuestro clúster.



Spark sobre YARN

Al igual que el anterior modo, tendríamos a YARN como gestor de recursos, que también se encuentra dentro de Apache Hadoop y es el que se encarga de los procesos mapreduce y demás. Es la instalación más habitual.

