

Instalación del entorno para Spark

Índice

- Spark
- Pyspark

SPARK

Apache Spark

- Descargar Apache Spark 2.4.8 Pre-built for Hadoop 2.7 de:
<https://spark.apache.org/downloads.html>
- Descomprimir archivo
- Crear variable de entorno SPARK_HOME, apuntando hacia la carpeta creada tras descomprimir
- Añadir a la variable de entorno PATH lo siguiente: “%SPARK_HOME%\bin”
- Crear una carpeta “RUTA\Hadoop\bin”
- Descargar a esa carpeta el archivo winutils.exe de AQUÍ
- Crear variable de entorno HADOOP_HOME, apuntando a la carpeta de Hadoop creada previamente

PYSPARK

Pyspark

- PIP:
 - Pip install pyspark
- Conda:
 - conda install pyspark