**INSTALACION DE PYSPARK EN W10**

**Paso 1: Instalar Java**

Apache Spark requiere Java para funcionar. Asegúrate de tener Java instalado en tu sistema. Puedes descargar e instalar Java desde el sitio web oficial de Oracle o utilizar una distribución de OpenJDK.

**Paso 2: Descargar Apache Spark**

Visita el sitio web oficial de Apache Spark (<https://spark.apache.org/downloads.html>) y descarga la última versión de Spark preconstruida para Hadoop. Elige la opción "Pre-built for Apache Hadoop" y descarga el archivo ZIP.

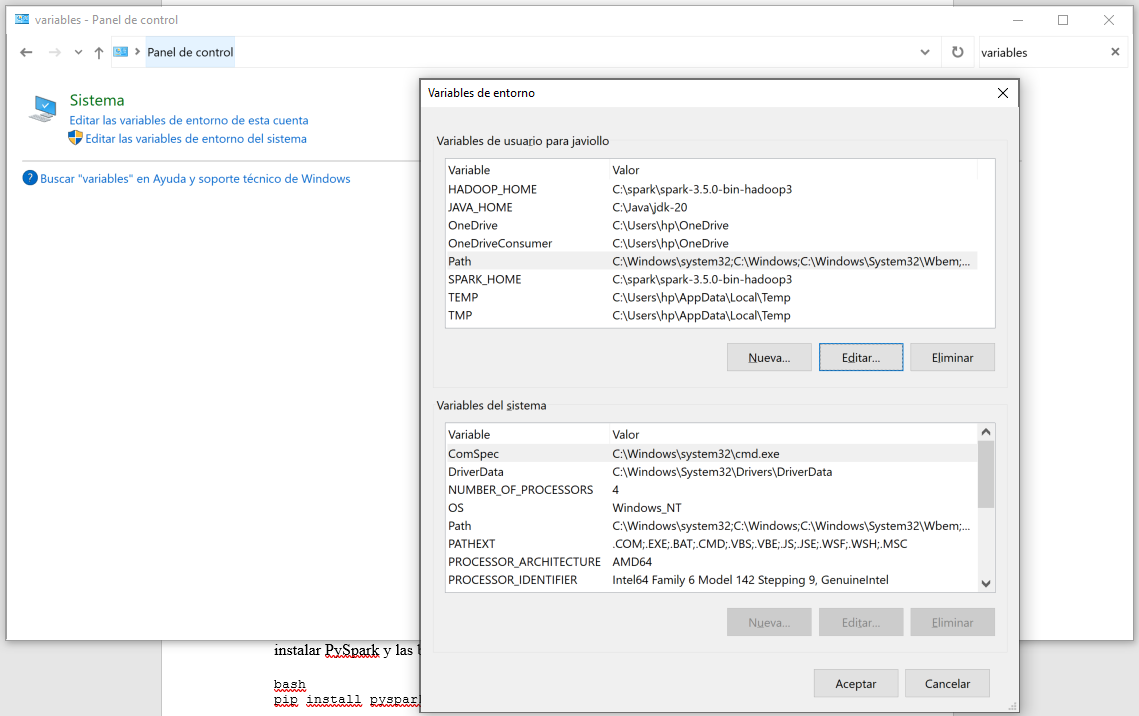
**Paso 3: Descomprimir Apache Spark**

Descomprime el archivo ZIP descargado en una ubicación de tu elección en tu sistema.

**Paso 4: Configurar las variables de entorno**

Para que Spark funcione correctamente, necesitas configurar algunas variables de entorno. Sigue estos pasos:

1. Abre el Panel de Control de Windows y busca "Variables de entorno".
2. Haz clic en "Editar las variables de entorno del sistema".
3. En la sección "Variables de sistema", busca la variable "JAVA\_HOME" y configúrala con la ruta a tu instalación de Java. Por ejemplo, C:\Program Files\Java\jdk1.8.0\_291.
4. En la misma sección, busca la variable "HADOOP\_HOME" (si no existe, créala) y configúrala con la ruta a la carpeta donde descomprimiste Apache Spark. Por ejemplo, C:\ruta\a\la\carpeta\de\spark.
5. Añade la carpeta "bin" de Apache Spark a tu variable "PATH". Por ejemplo, si descomprimiste Spark en C:\ruta\a\la\carpeta\de\spark, agrega C:\ruta\a\la\carpeta\de\spark\bin a tu variable "PATH".



**Paso 5: Instalar PySpark**

Es una buena práctica crear un entorno Conda dedicado para tu proyecto de PySpark para evitar conflictos con las dependencias de otros proyectos. Puedes crear un nuevo entorno con el siguiente comando:

conda create --name pyspark

Luego, activa el entorno recién creado:

conda activate pyspark

pip install pyspark

Esto instalará PySpark en tu sistema.

**Paso 6: Verificar la instalación**

Para verificar que PySpark se haya instalado correctamente, abre una ventana de comandos y ejecuta el siguiente comando:

pyspark

Esto iniciará una sesión interactiva de PySpark. Si todo está configurado correctamente, deberías ver el prompt de PySpark y poder ejecutar comandos de PySpark en esta sesión.

Con estos pasos, habrás instalado PySpark en Windows 10 y estarás listo para empezar a trabajar con Apache Spark en tu entorno local.

Y luego, si se quiere:

conda install -c anaconda jupyter