Instalación y configuración de PySpark

1. Bajar del drive los archivos que están en la carpeta Software!

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. La carpeta hadoop insertarla en el disco C

C: \hadoop

1. Ir a Variables de entorno, Variables de sistema y hacer dos nuevas variables HADOOP\_HOME y hadoop.home.dir en valor a las dos variables se le añade el mismo valor C: \hadoop

Nota: Para hace una nueva variable presionar nueva en donde aparecerá

Nombre

Valor

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

1. En la mima variables de entorno, variables del sistema en la variable “path” Editar, añadir el valor de HADOOP\_HOME\bin

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Instalar Java

1. Para sabe que versión de java tiene instalado en un cmd : java -vesrion

Texto

Descripción generada automáticamente

1. Si tiene instalado otra versión desístale e instale el jdk-11 que se le proporciono en el drive

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Configure las variables de entorno de java

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Realice una copia de los de los notebook en su env

Instale pyspark y fidspark como se lo indica en los notebook

Proceda a practicar los los notebook proporcionados

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente