ECOSISTEMA SPARK
Práctica 2: SparkSQL
TAREA 1

Vamos a seguir trabajando con el conjunto de datos del Heterogeneity Dataset for Human Activity Recognition (HHAR) que contiene información de los sensores de movimientos de teléfonos y relojes.

El objetivo será también agregar usando como clave primaria la terna usuario (User), modelo (Model) y movimiento ejecutado (gt).

En concreto, hay que crear un dataframe por cada fichero proporcionando el esquema.

A partir de los DataFrames iniciales, obtendremos un registro por cada usuario, modelo y clase con la media, desviación estándar y valor máximo y mínimo de la secuencia del movimiento ejecutado.

Una vez hecho esto, se deberá concatenar mediante join los registros de giróscopo y acelerómetro de los relojes por un lado y de los teléfonos por otro. Finalmente se creará un DataFrame único (mediante union) con los DataFrames de teléfonos y relojes.

Para manipular los DataFrames, podemos utilizar cualquiera de las dos opciones vistas en clase:

- Aplicar operaciones de la API
- Ejecutar consultas SQL

Para esta tarea se utilizará un único notebook que formará parte del archivo .zip correspondiente a la Práctica 2. No se deben incluir los ficheros de datos. Las funciones deben estar documentadas.

FECHA DE ENTREGA: 9 de enero