


Declaración de uso del estudiante

Todo estudiante que utilice IA generativa en una actividad debe declarar su uso de manera transparente. Para ello, se debe utilizar este para declarar el uso de IA.

Nombre del estudiante	Adrián Fernández Galán, César López Mantecón, Manuel Gómez Plana
Asignatura	Arquitectura de Datos
Título de la tarea	Práctica 2
Fecha de entrega	08/12/2024
1. Uso de IA generativa	
- Herramienta utilizada	ChatGPT
- Fecha de uso	18/11/2024 – 06/11/2024
- Descripción del uso	Utilizamos ChatGPT para recibir explicaciones sobre cassandra y pyspark, tales como el funcionamiento de algunas operaciones.
2. Prompts utilizados	
- Prompt 1	¿Se puede optimizar de alguna manera el siguiente comando? spark-submit --packages com.datastax.spark:spark-cassandra-connector_2.12:3.3.0 PySpark.py
- Prompt 2	¿Cómo puedo hacer un group by en pyspark?
- Prompt 3	Tengo esta tabla original llamada <i>jsondf</i> y necesito renombrar todas las columnas que sean varias palabras para que estén separadas por
¿Cómo has usado o cambiado el resultado generado por IA?	Dado que los prompts buscan recibir una explicación de un concepto en concreto no ha sido necesario tratar los resultados obtenidos por IA
Resultados generados por la IA	<ul style="list-style-type: none">- Respuesta al Prompt 1: Resumen de varias formas en las que se puede optimizar el comando como, por ejemplo: evitar descargar paquetes repetidamente, configurar recursos como la memoria y el número de ejecutores, aprovechar configuraciones predefinidas...- Respuesta al Prompt 2: Explicación de cómo se hace un group by en pyspark seguido del siguiente código:

Edición: 2024/2025 Arquitectura de Datos Grado en Ingeniería Informática ASUNTO: Declaración del estudiante en herramientas de IA generativa	 Universidad Carlos III de Madrid
--	--

	<pre>df.groupBy("Name", "Subject").agg(avg("Score").alias("Average Score")).show()</pre> <p>- Respuesta al Prompt 3:</p> <p>Para renombrar todas las columnas con múltiples palabras separadas por espacios a un formato que use guiones bajos (_), necesitas recorrer el esquema de la tabla, identificar las columnas anidadas y aplicar los cambios recursivamente. Aquí está un ejemplo de cómo hacerlo en PySpark:</p> <p>Código para renombrar columnas</p> <pre>from pyspark.sql import DataFrame def rename_columns(df: DataFrame) -> DataFrame: def rename_struct_fields(schema): """Renombra campos dentro de estructuras anidadas.""" from pyspark.sql.types import StructType, StructField new_fields = [] for field in schema.fields: new_name = field.name.replace(" ", "_") # Reemplazar espacios por "_" if isinstance(field.dataType, StructType): # Si el campo es una estructura anidada new_field = StructField(new_name, rename_struct_fields(field.dataType), field.nullable) else: new_field = StructField(new_name, field.dataType, field.nullable) new_fields.append(new_field) return StructType(new_fields) new_schema = rename_struct_fields(df.schema) # Aplicar renombrado al esquema # Aplicar los cambios al DataFrame return spark.createDataFrame(df.rdd, new_schema)</pre> <p>1. Aplicar la función al DataFrame original</p> <pre>renamed_df = rename_columns(json_df)</pre> <p>2. Mostrar el nuevo esquema</p> <pre>renamed_df.printSchema()</pre>
Confirmación de cumplimiento con las directrices de derechos de autor	No se han ingresado materiales con derechos de autor, sin ser referenciados correctamente
Firma del estudiante	No será necesaria si la firma si se entrega digitalmente por Aula Global. En caso de entregarse en papel, sí será necesaria.