



Reactions/Comments Streaming Counter

Ronald Ernesto Soto Nicoll

Descripción



Se envían dos notebooks :

- reactions-comments-youtube
- RCY5

El proyecto consistía en generar un contador de likes, dislikes y comentarios ordenados por código de vídeo y ventanas de tiempo de 5 min en una sola tabla. Los Likes y Dislikes provenían del tópico **reactions** y los comentarios del tópico **comments**, pero surgieron errores porque al parecer en Spark no se puede hacer join entre dataframes que ya han tenido agregaciones previas. Por lo tanto, se envía dos notebooks que son testimonio del avance hasta ese punto:

-El primero (reactions-comments-youtube), muestra que efectivamente se lograron realizar las agregaciones por separado y mostrar el total de reacciones y comentarios en dataframes diferentes. La idea era realizar un join entre estas tablas mediante un join pero al parecer no es factible. Por lo que se intentó otro camino. Solo se muestra el esquema de este en estas diapositivas.

- El segundo (RCY5) muestra un join entre ambas tablas pero esta vez sin agregaciones, el problema es que esta tabla puede generar resultados duplicados.

Descripción



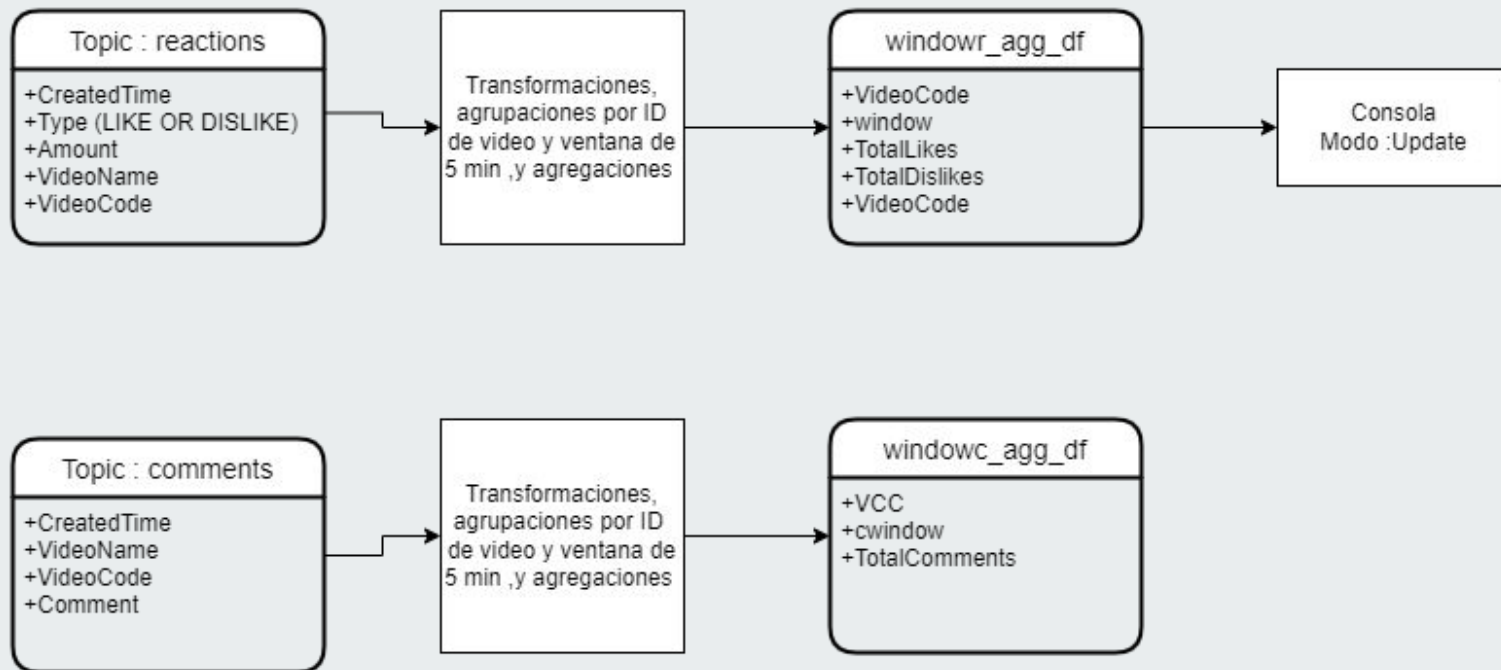
Además de los notebooks, se incluyen:

- reactions_data : Los datos que incluyen la información de los likes y dislikes, va en el tópico reactions.
- comments_data: Los datos que incluyen la información de los comentarios va en el tópico reactions.
- instrucciones: Las instrucciones para iniciar y crear los tópicos reactions y comments de kafka.
- Esquema Proyecto reactions-comments-youtube with Kafka and Apache Spark.drawio: Esquema del primer notebook. Solo se incluye este esquema pues es el que produce resultados veraces, el otro notebook puede producir resultados duplicados.

Esquema reactions-comments-youtube

Kafka

Apache Spark





¡GRACIAS!