# Preparación de tablas PostgreSQL

El primer paso para realizar la práctica es insertar la información de los usuarios. Para ello es necesario crear una base de datos PostgreSQL en Google Cloud, de modo que mediante la Serving Layer se cargue la tabla user\_metadata. También será necesario crear las tablas que van a ser usadas en los pasos posteriores.

# Simulación de antenas

Para la recogida de información “simulada” de las antenas se ha creado una VM de Google Cloud en la cual se ejecutará zookeeper y Kafka. Para enviar datos a Kafka se ha tenido que ejecutar varias veces el data simulator debido a que si no ningún cliente llegaba a la cuota límite asignada en la user\_metadata y por lo tanto no se ingestaba la tabla user\_quota\_limit.

# Ejecución de la Speed Layer

Mediante la Speed Layer, gracias a spark streaming se ejecuta un proceso que empleando microbatching va leyendo de la cola Kafka la información y calcula métricas agregadas que guarda en las diferentes tablas creadas en el paso 1 mencionado anteriormente. Por último guarda la información que lee de Kafka en formato Parquet en Google Cloud Storage.

# Ejecución de la Batch Layer

Finalmente, en la Batch Layer, mediante el uso de spark leemos la información guardada en formato Parquet en la capa anterior y calculamos los bytes por antena, usuario y app. Para acabar cruzamos la información de la tabla user\_metadata con los bytes por usuario para calcular que cliente se ha pasado del limite de bytes y guardarlo en la tabla user\_quota\_limit.