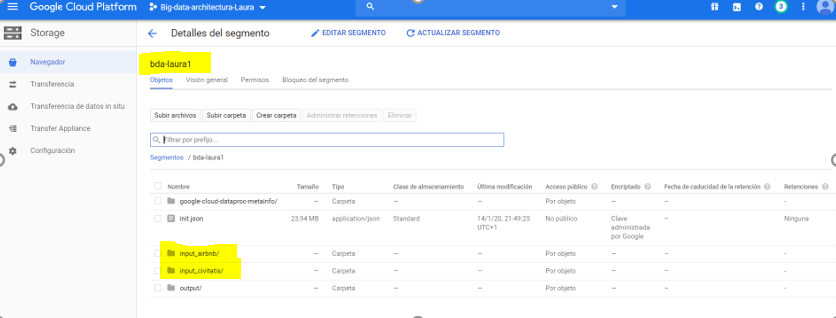
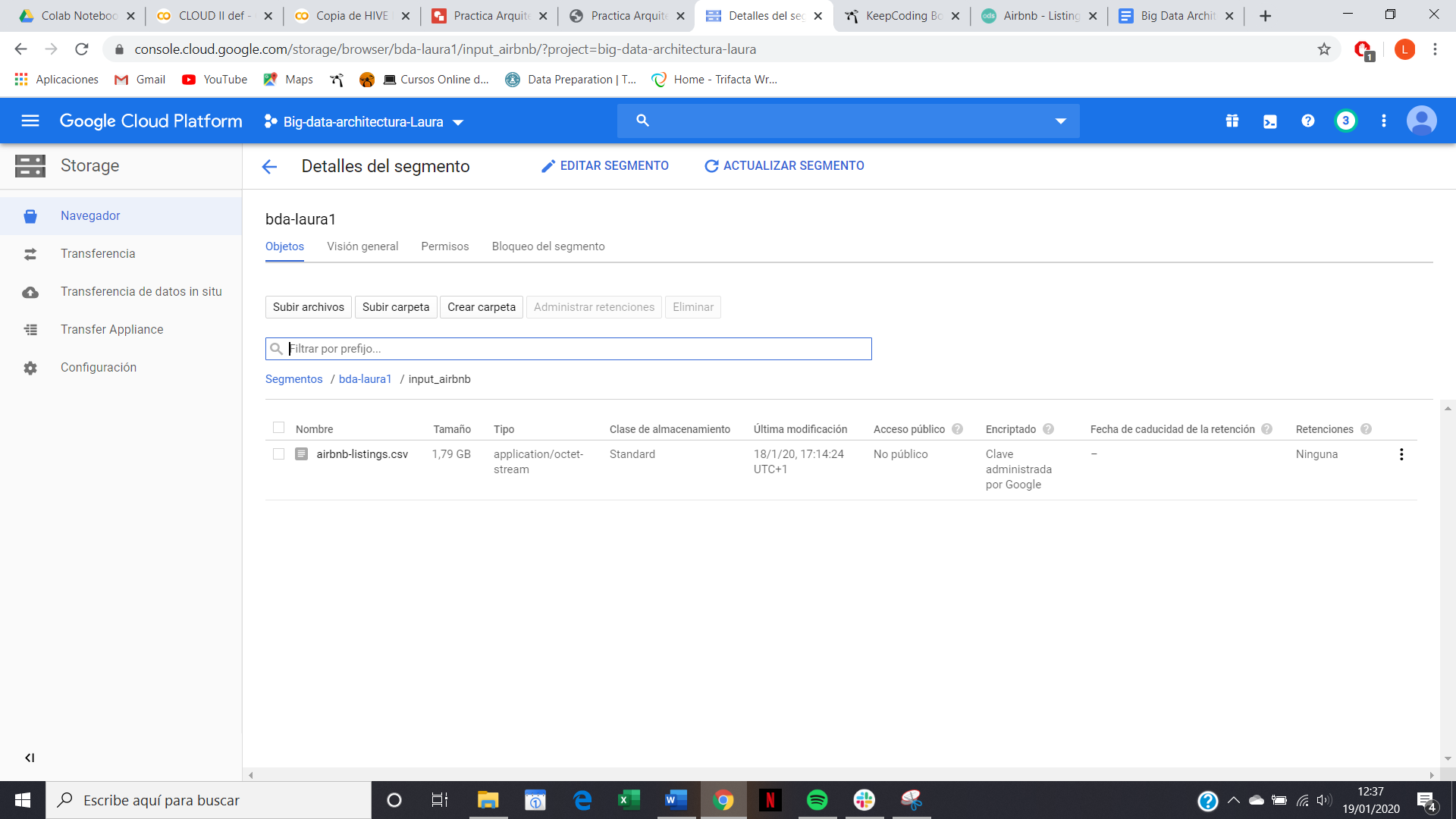
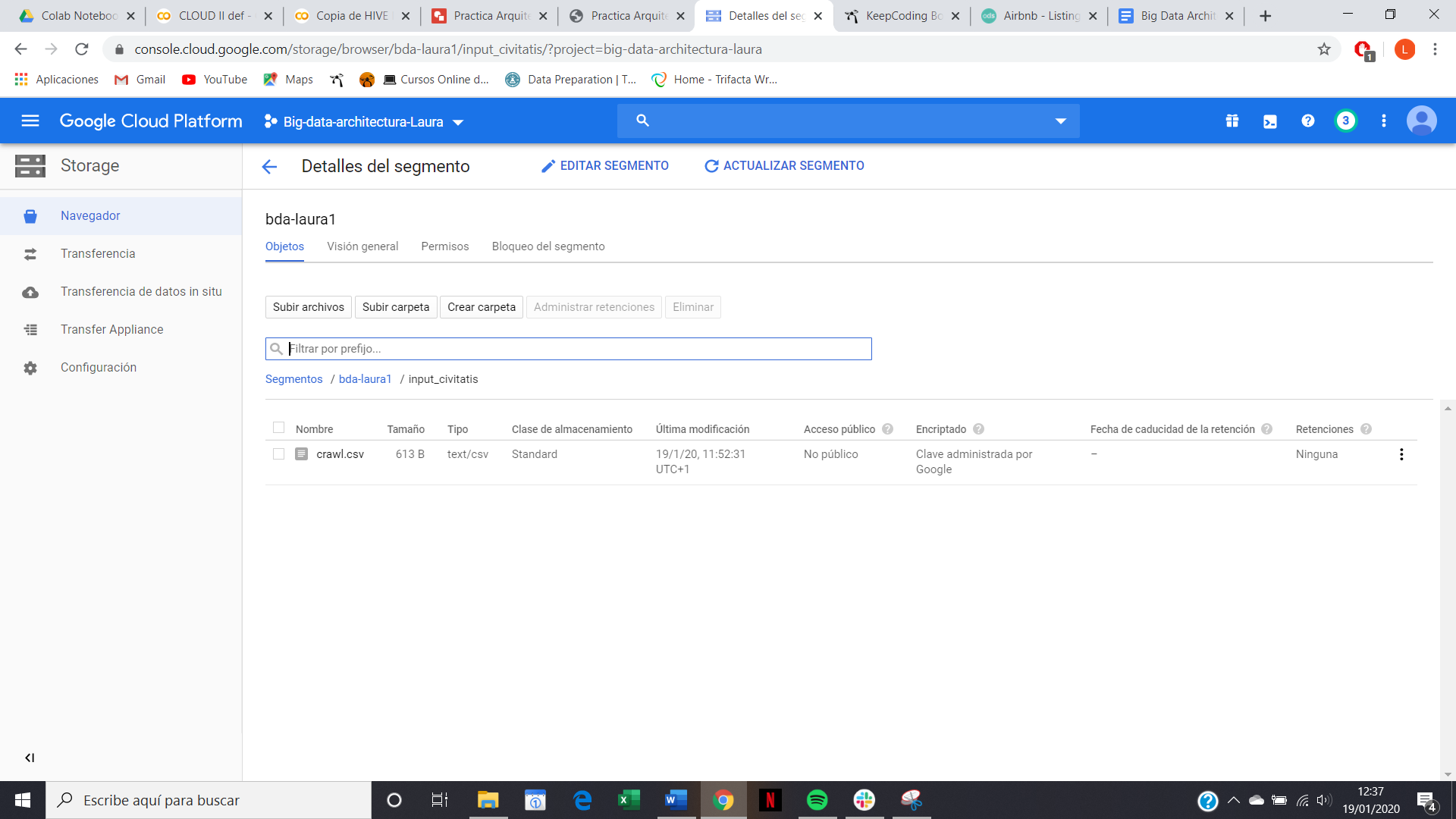
**Desarrollo de la práctica**

1. Creación del Segmento
   * Subida del fichero a mano

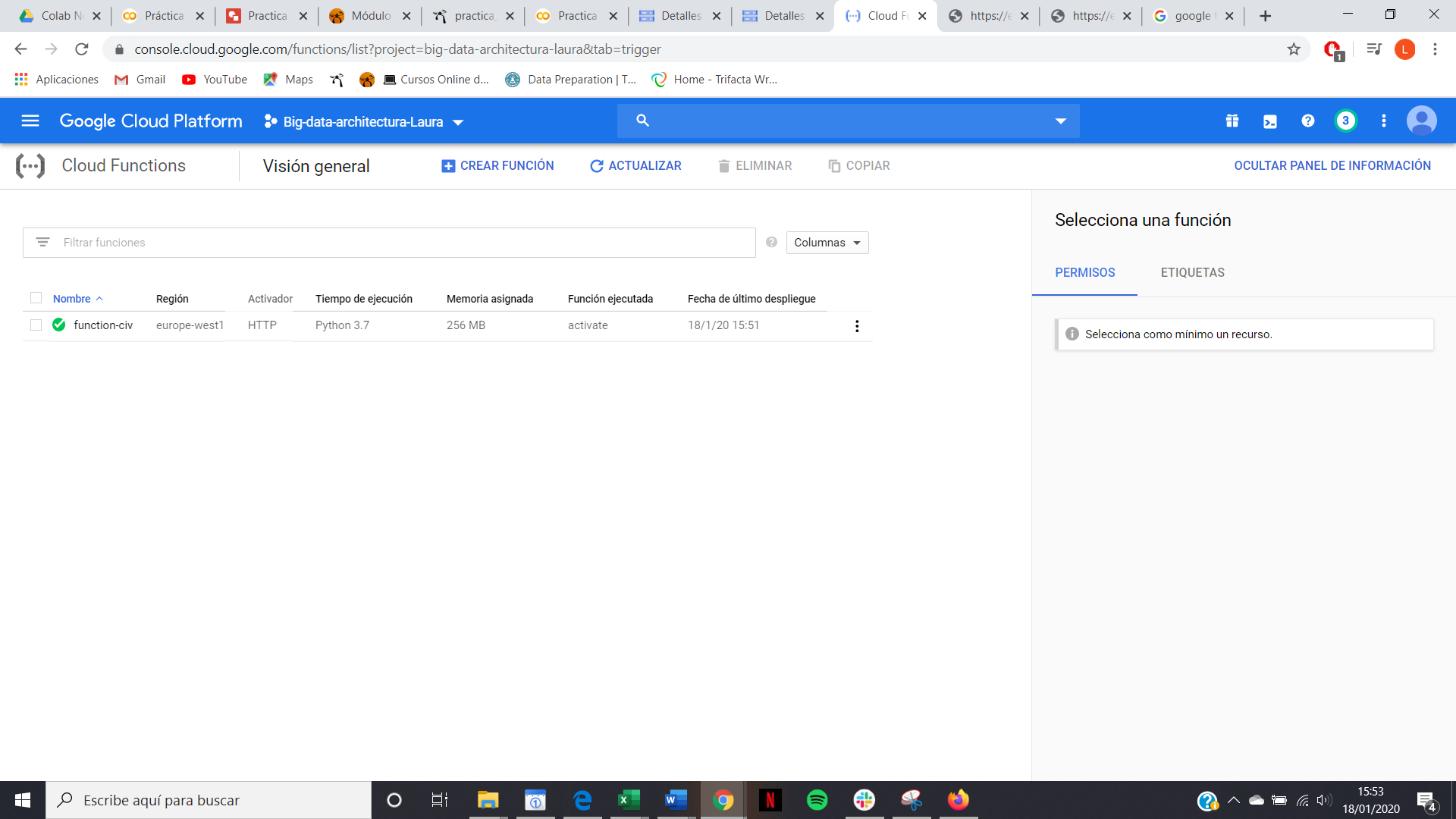


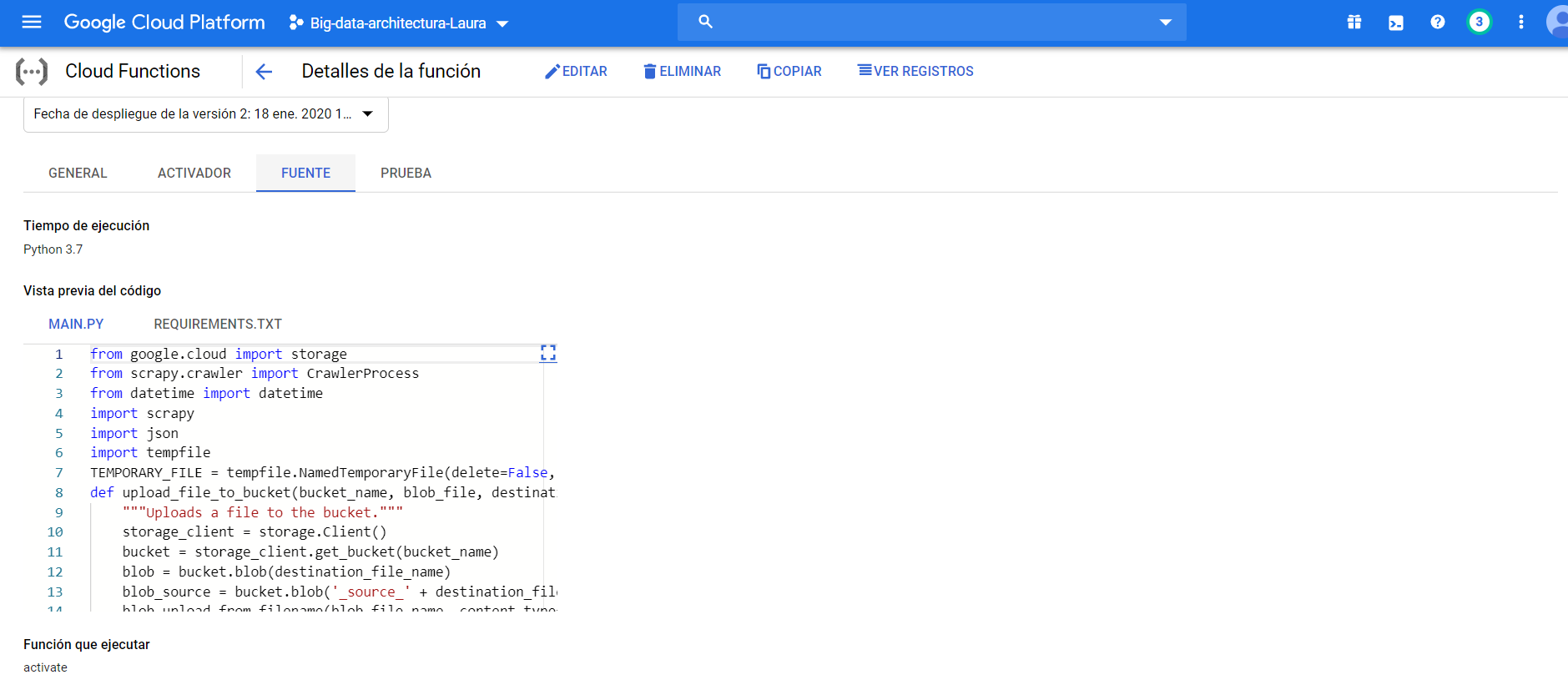
* El crawler entra de forma automática tras activar la función



1. Despliegue la Cloud Function

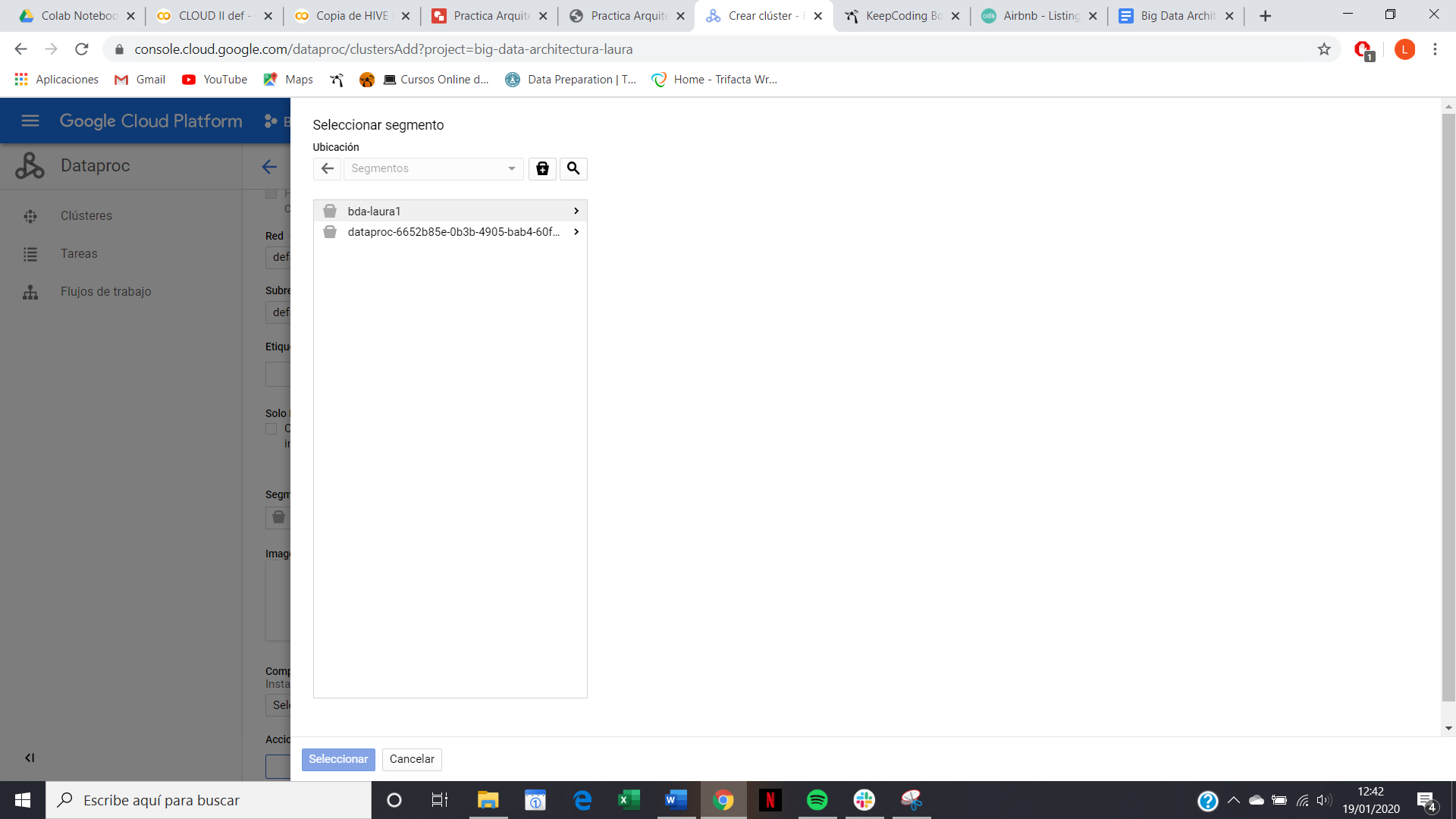
<https://europe-west1-big-data-architectura-laura.cloudfunctions.net/function-civ>

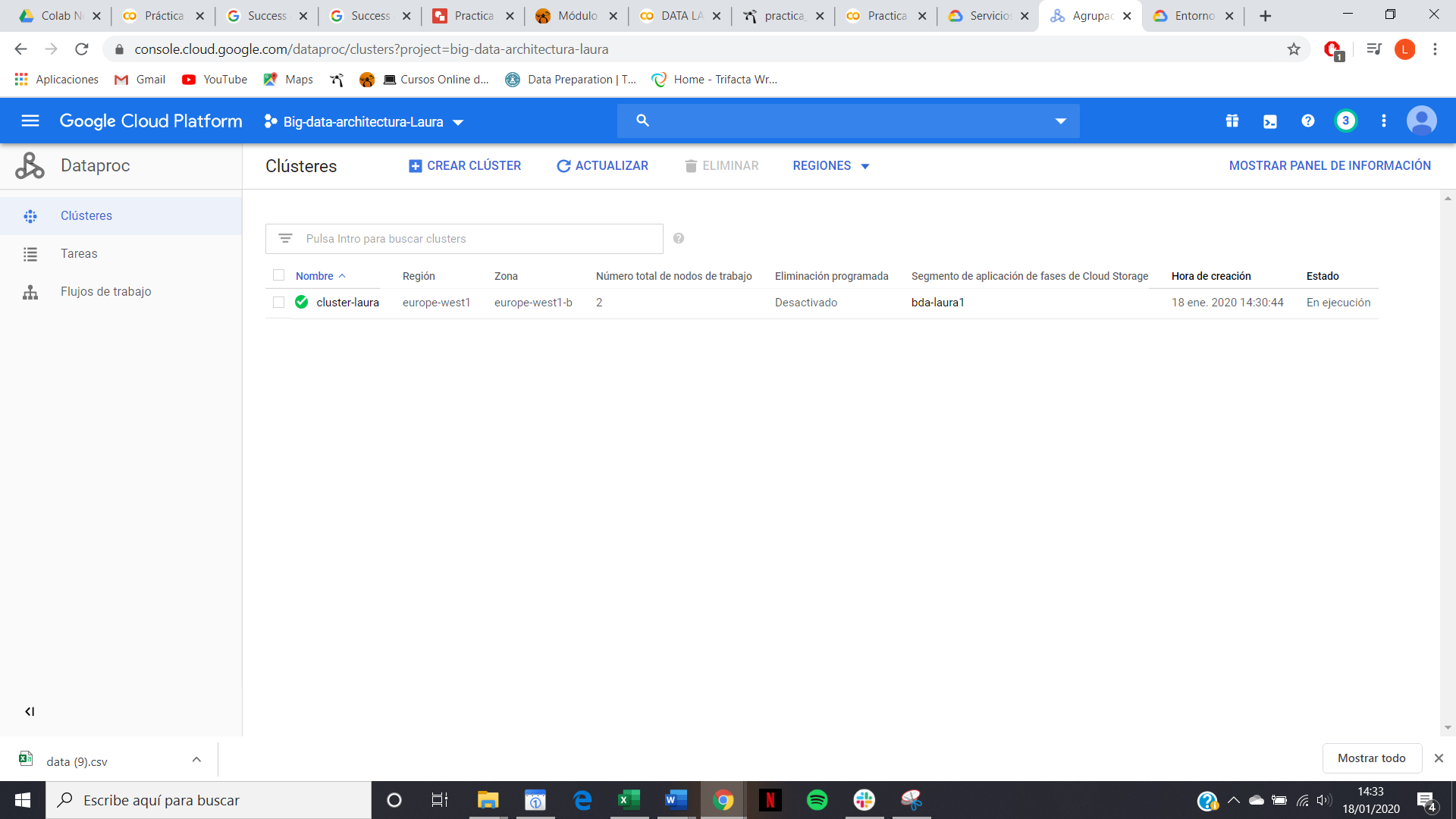




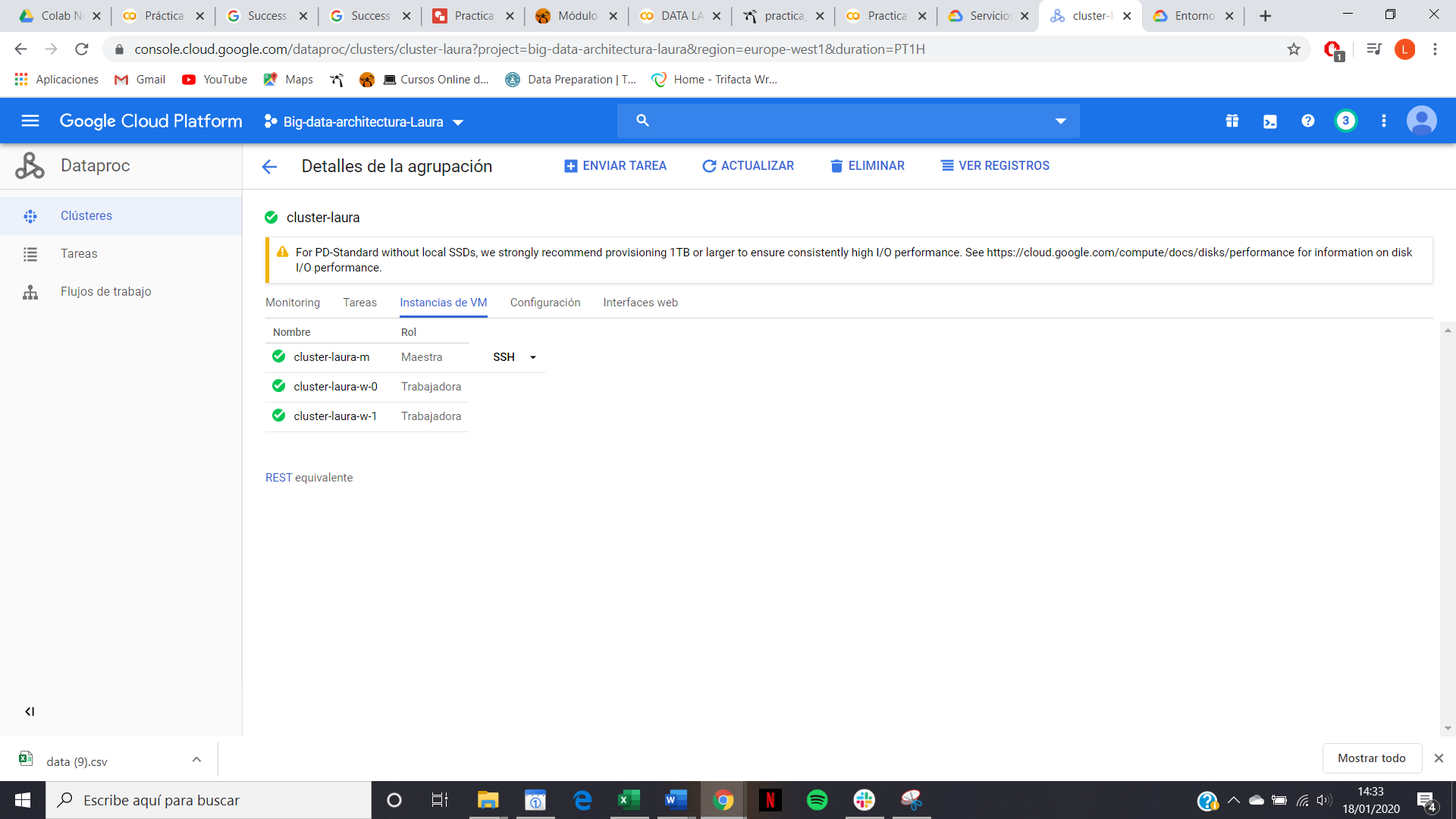
1. Levantar el cluster

* Escogemos el segmento ya creado con los datos.

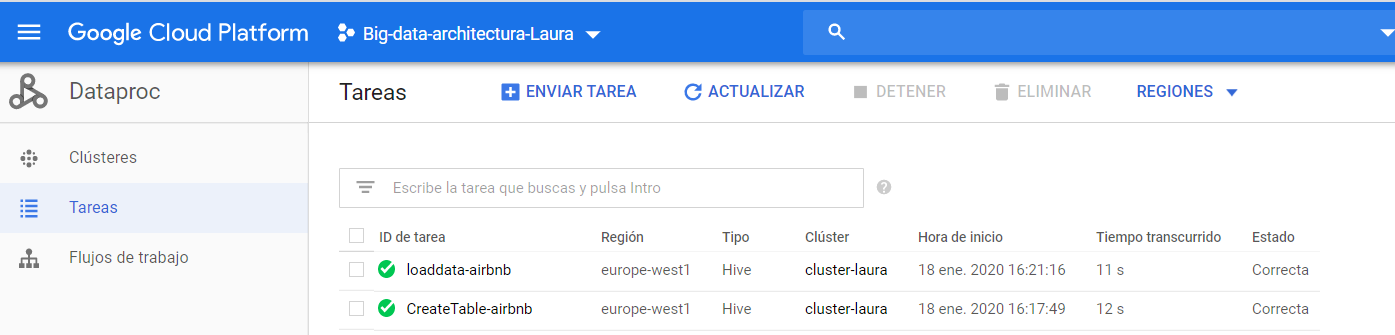




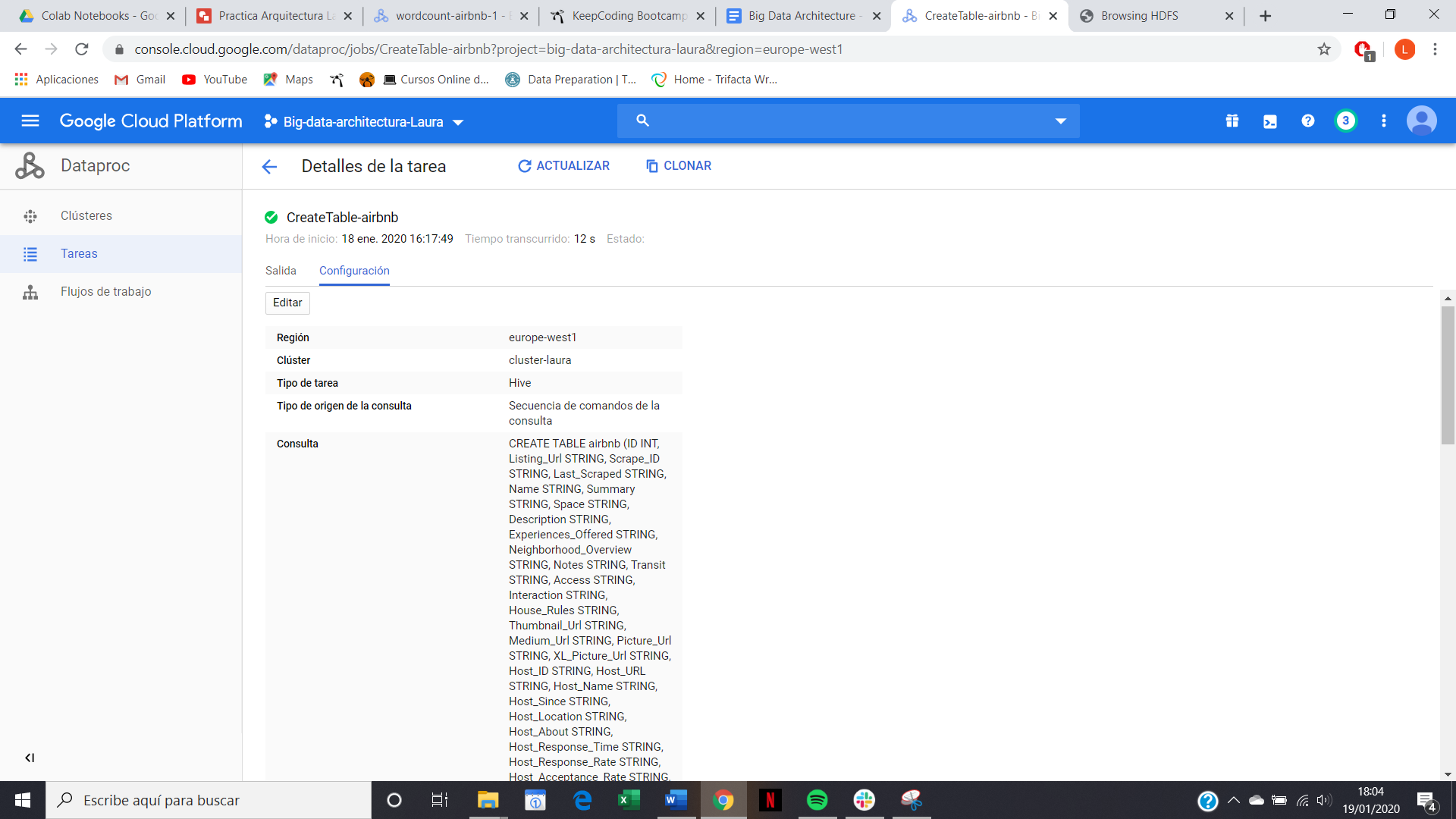
* Resultado cluster configurado



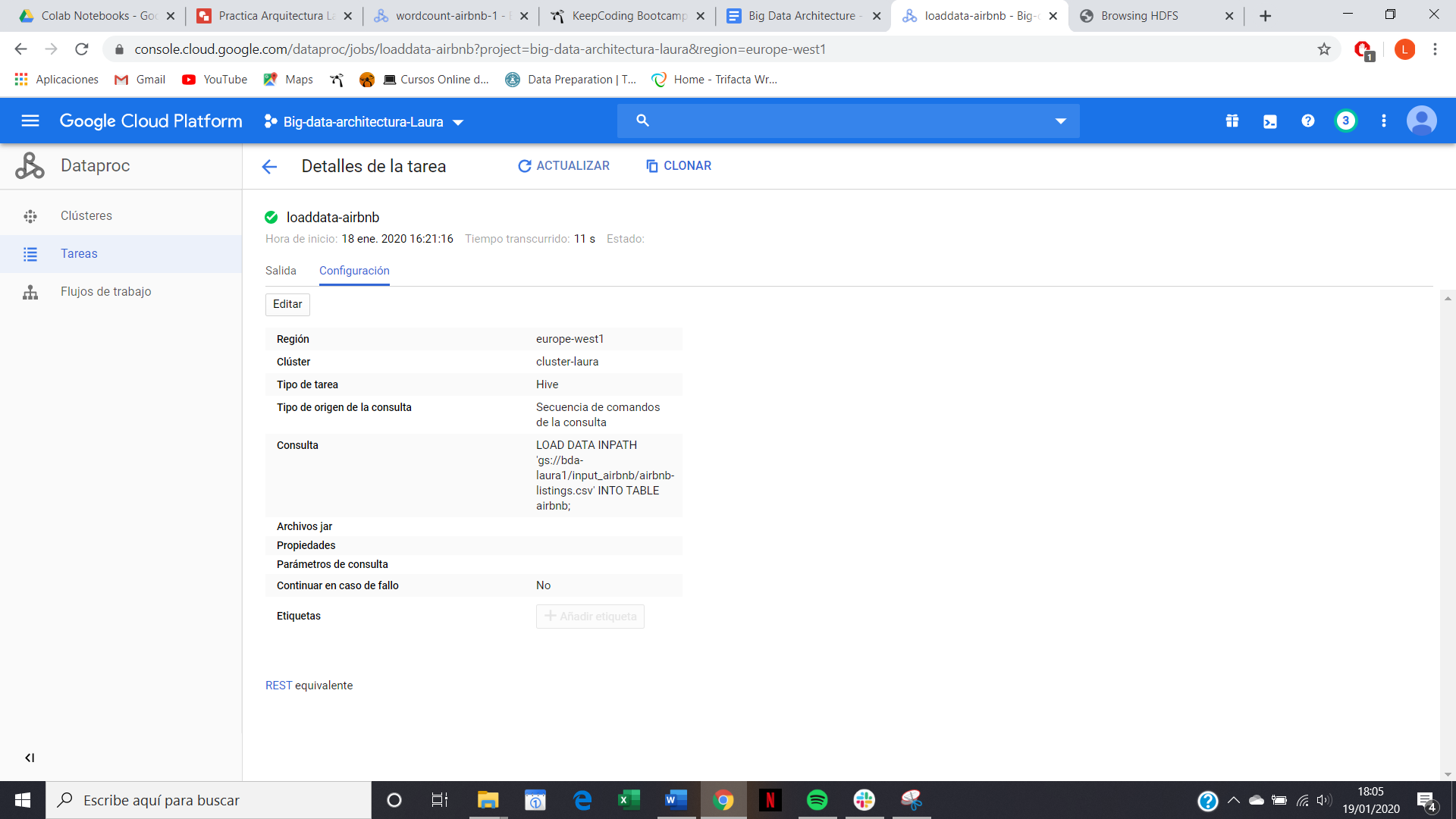
1. Tabla Airbnb



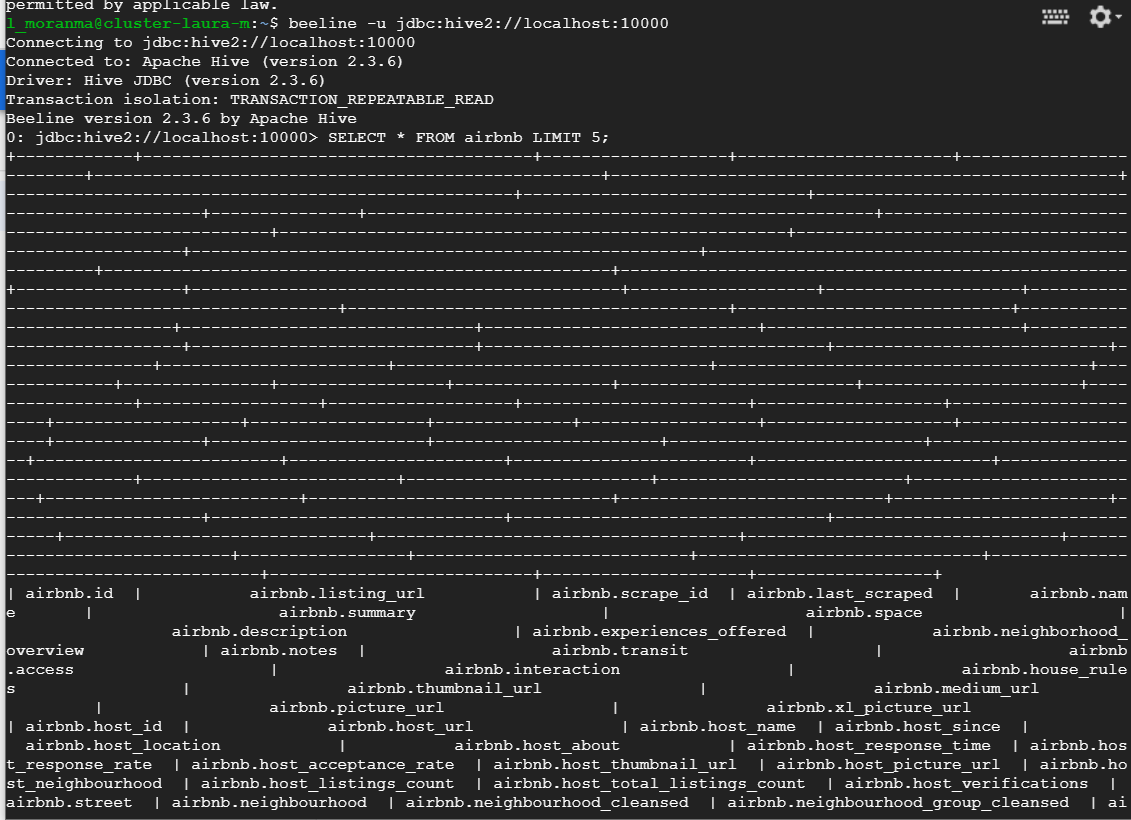
* Create table:

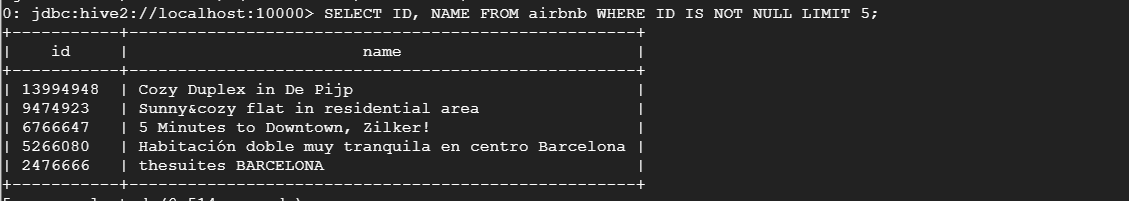


* Load data del Google storage:

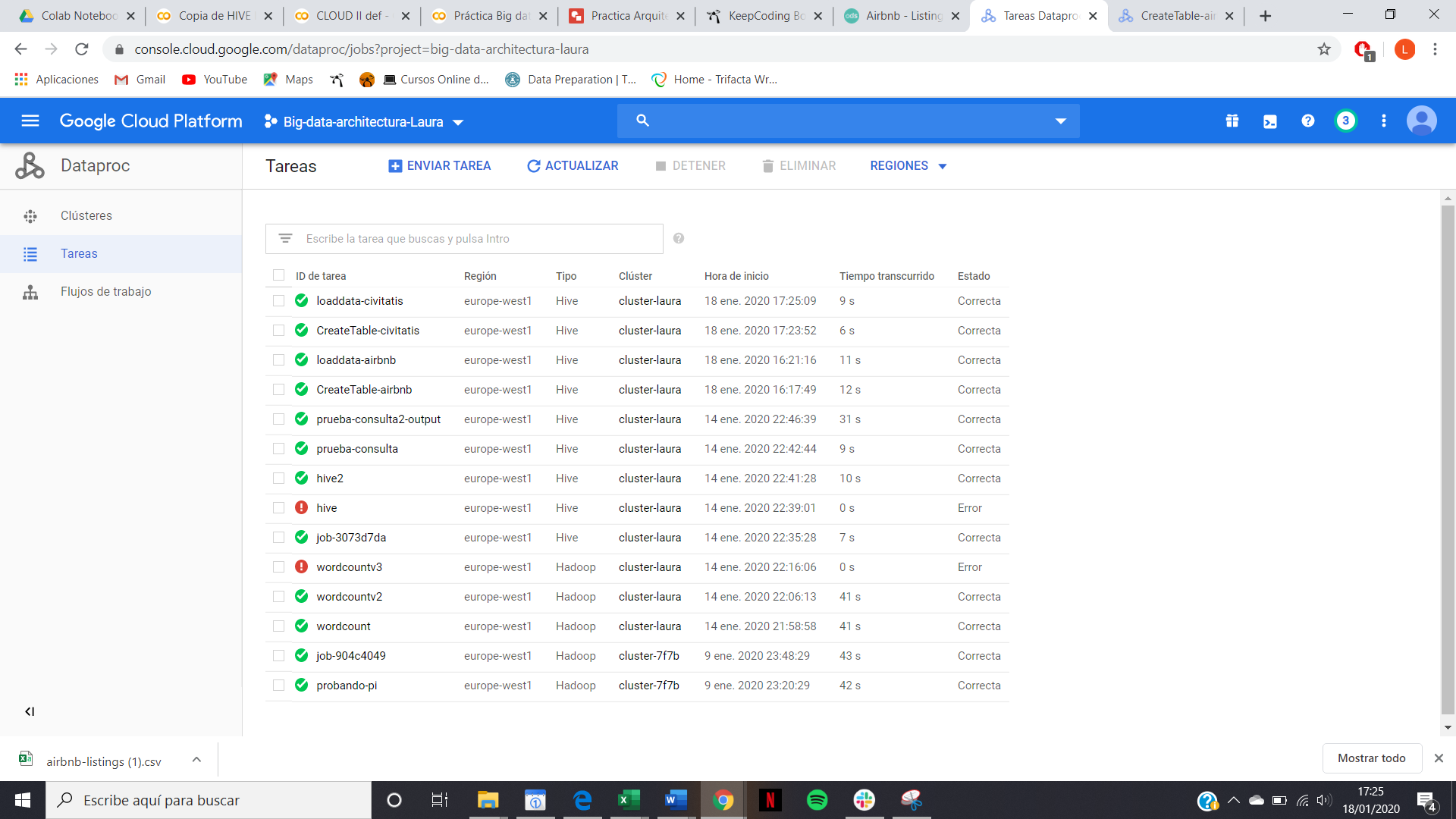


* Dentro de la rama master nos conectamos via SSH a Beeline para realizar consultas:

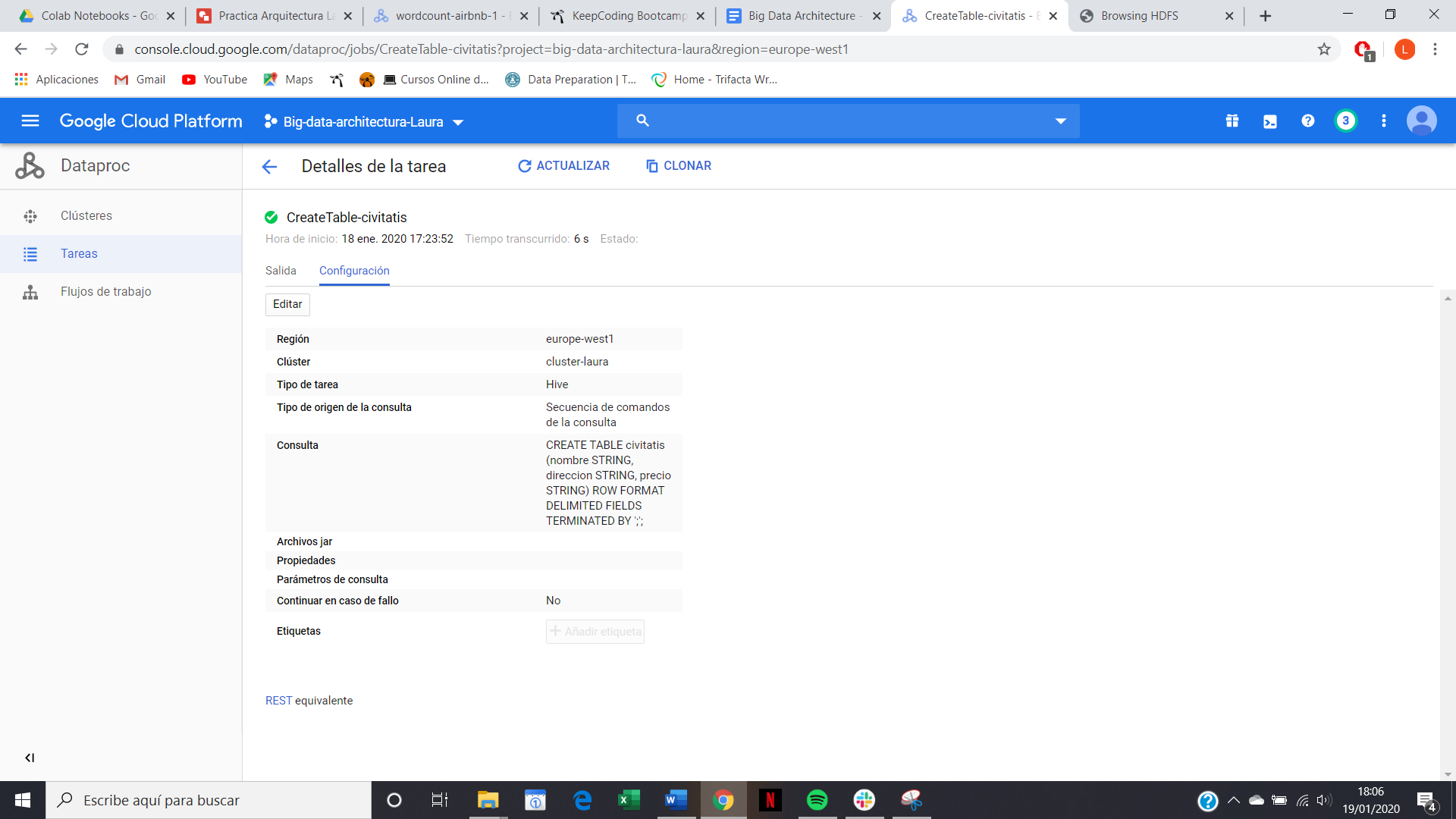




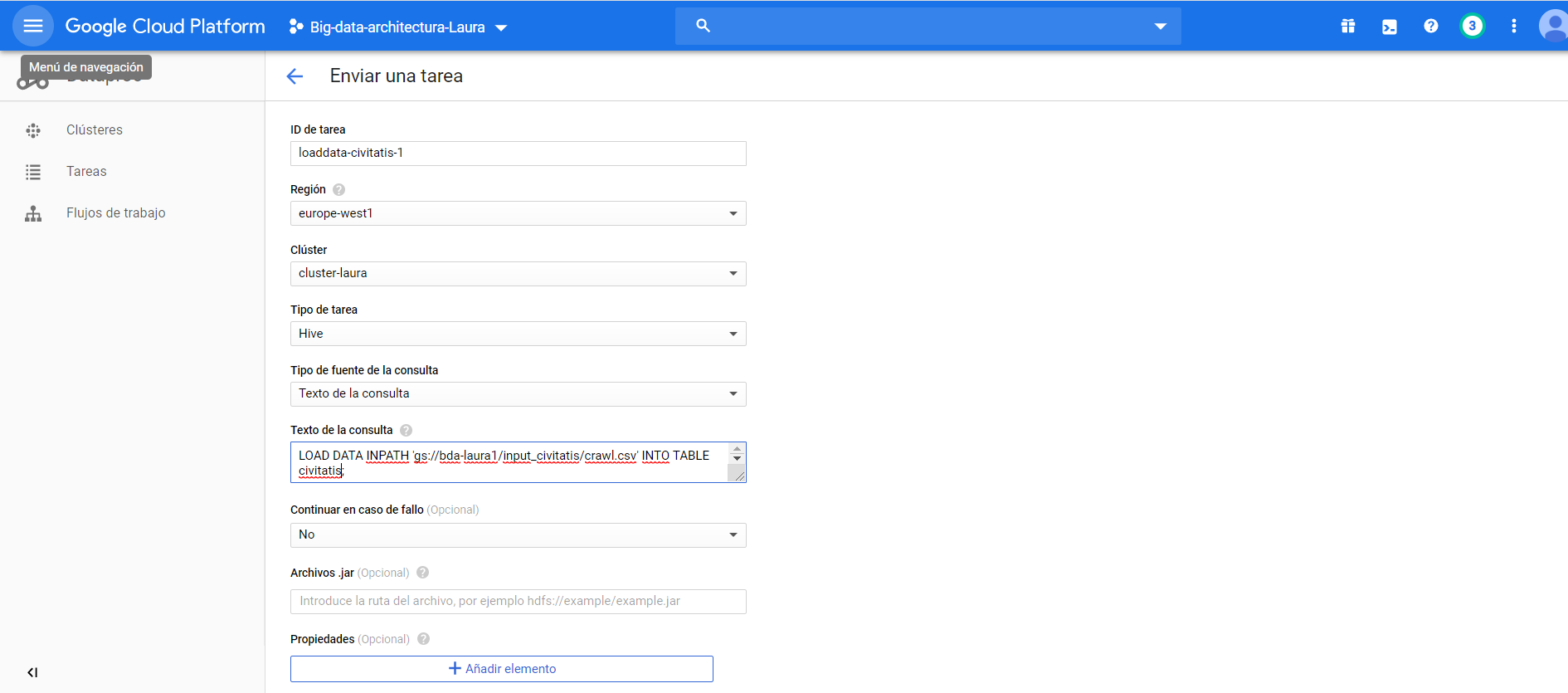
1. Tabla civitatis



* Create table:



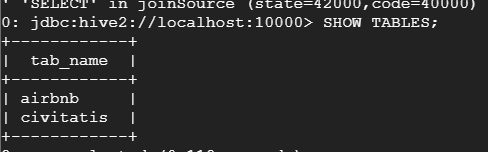
* Load data del Google storage:



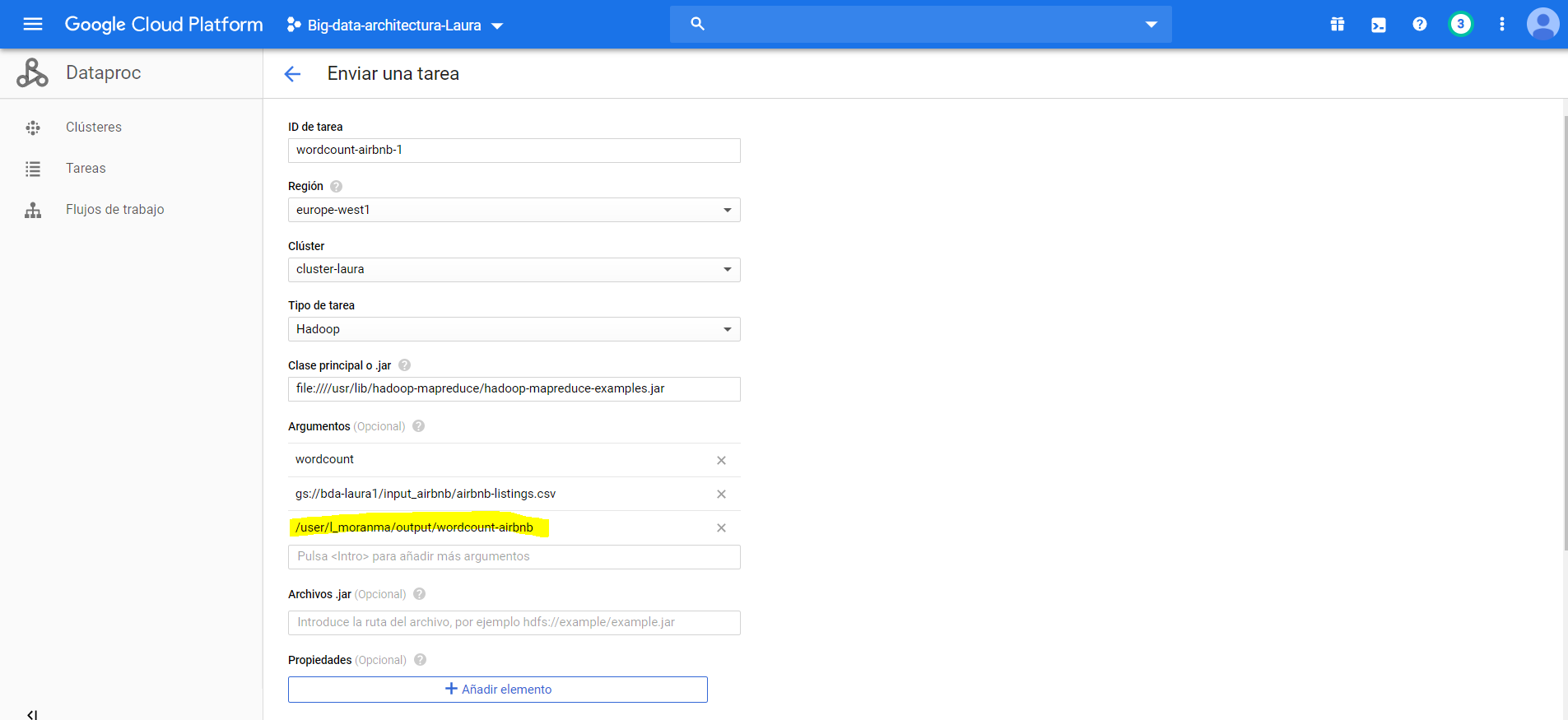
* Contenido de la tabla

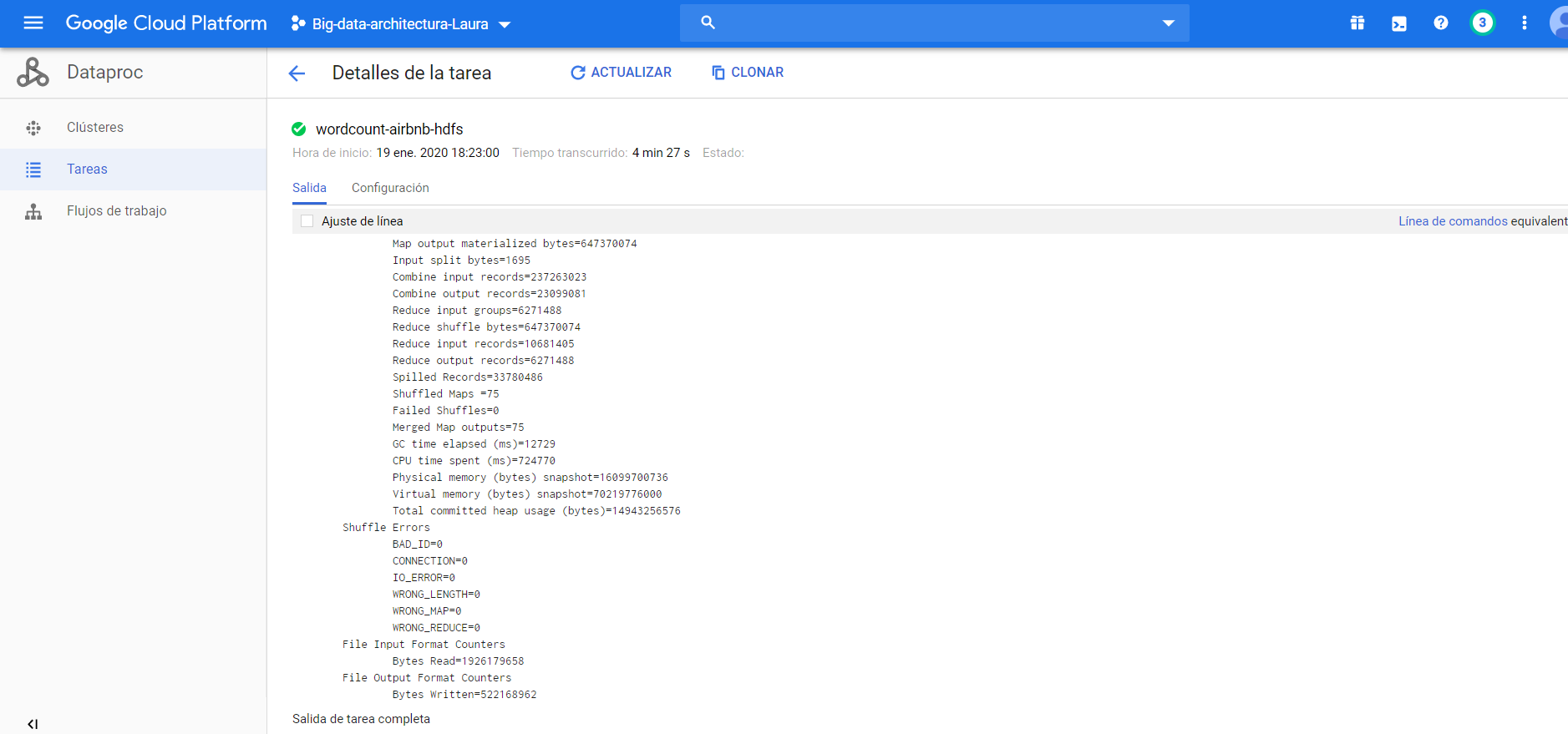


* Ambas tablas en Hive

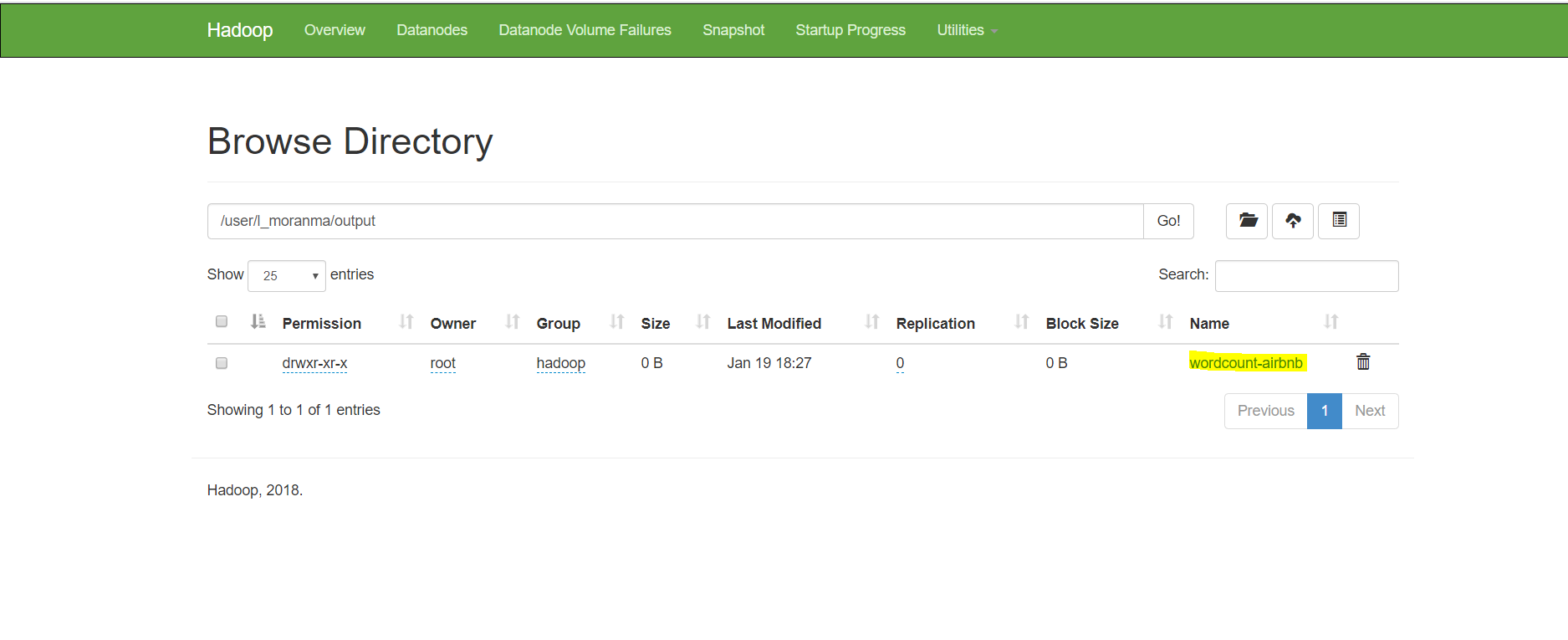


1. (Extra) Procesamiento de datos con Wordcount

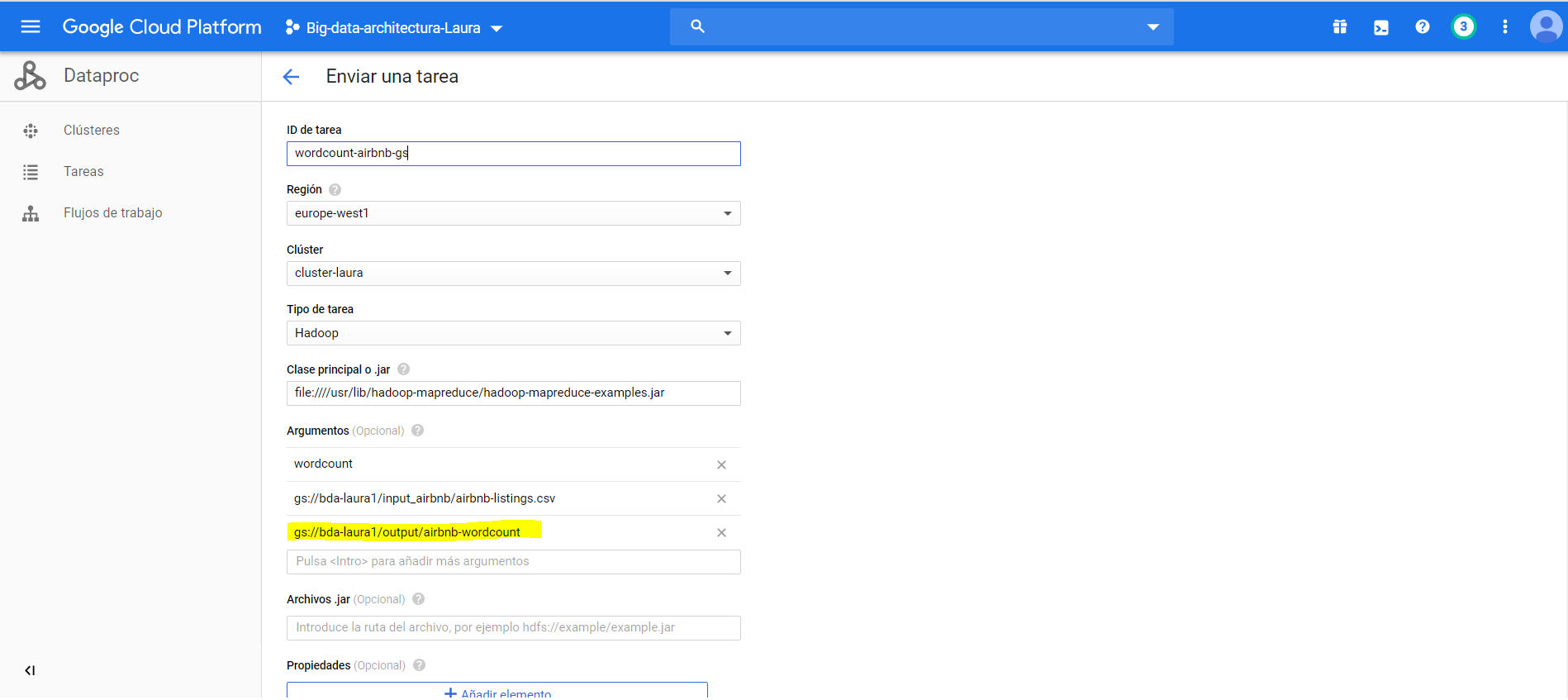


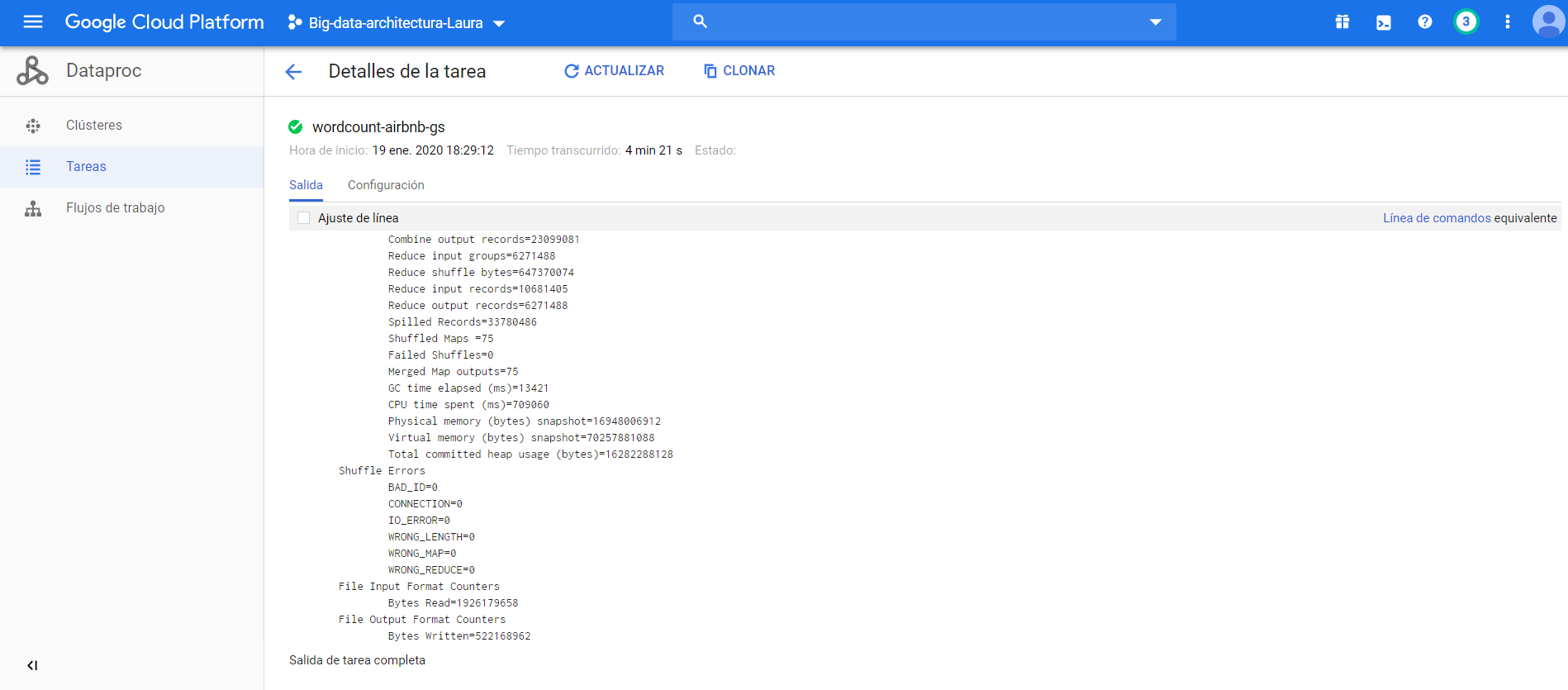


* De esta forma se guardaría el resultado en el hdfs y no en el segmento

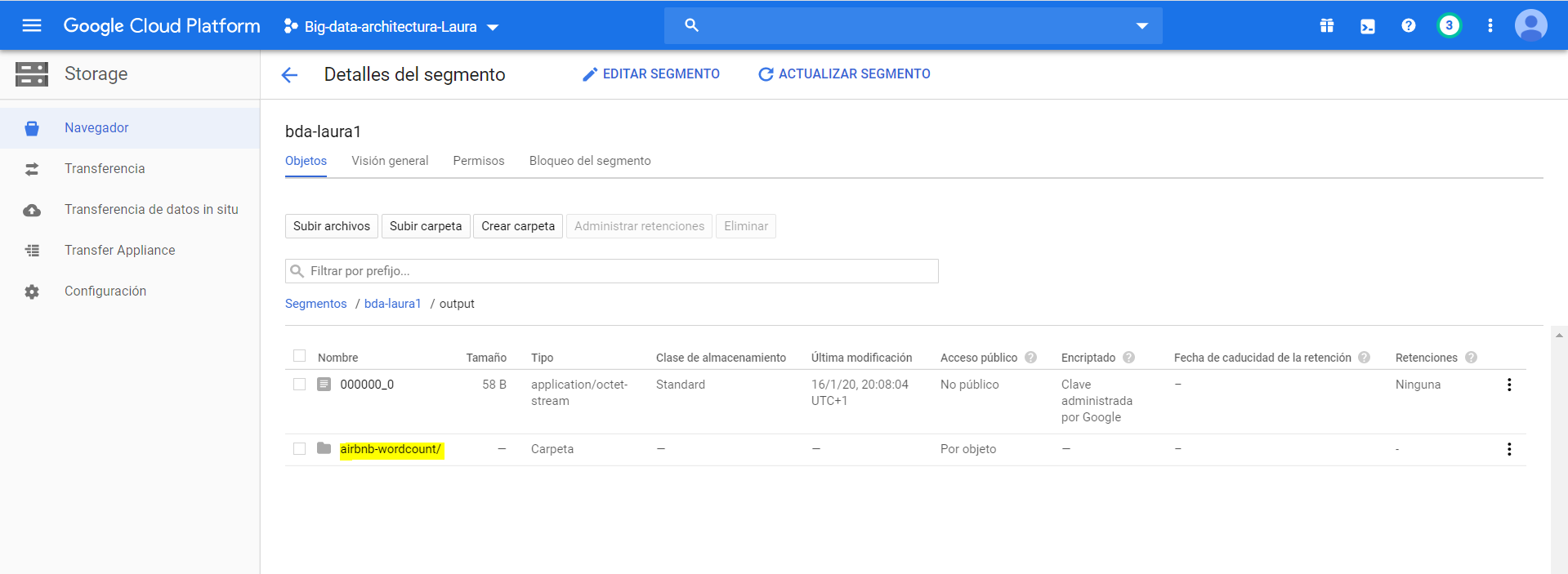


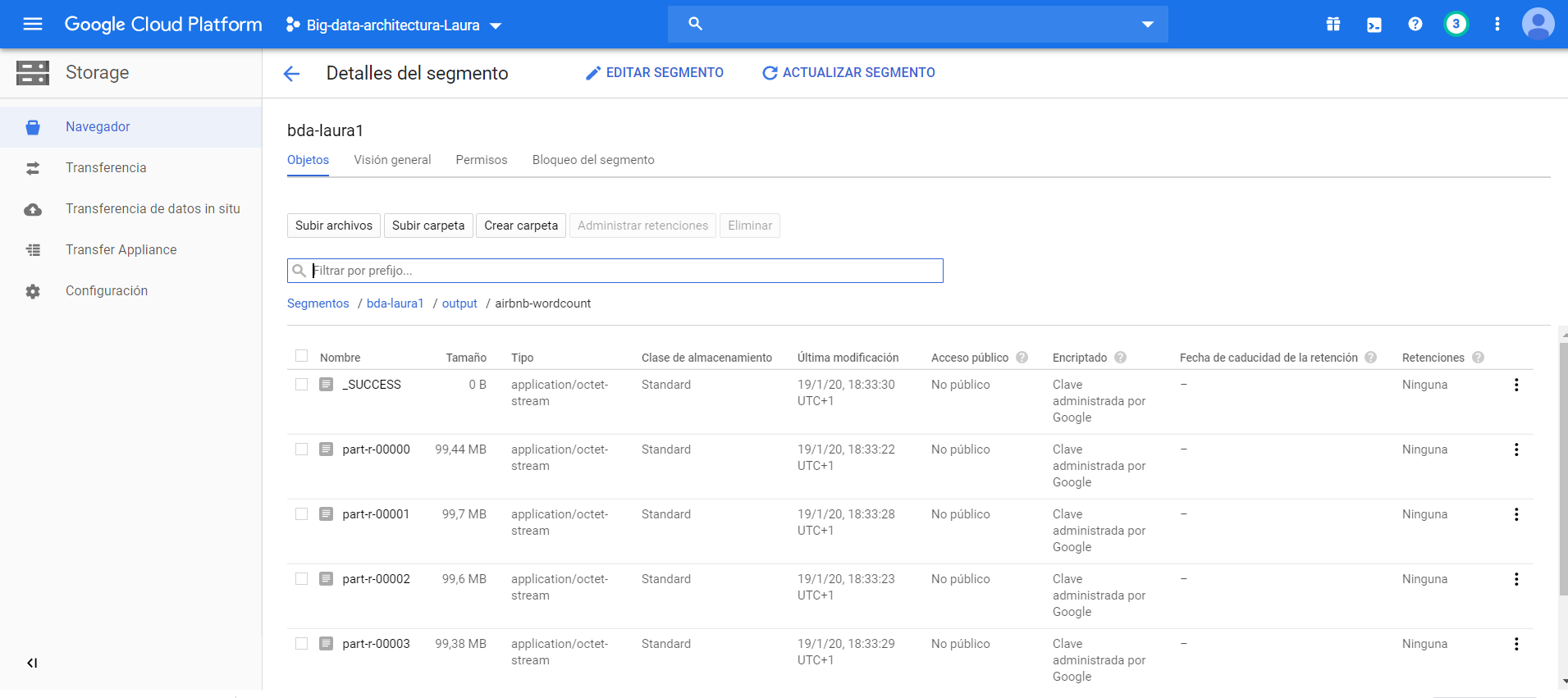
* También lo podemos guardar en el mismo segmento si lo preferimos





* Fichero output del Google storage





* Output: fichero final procesado

