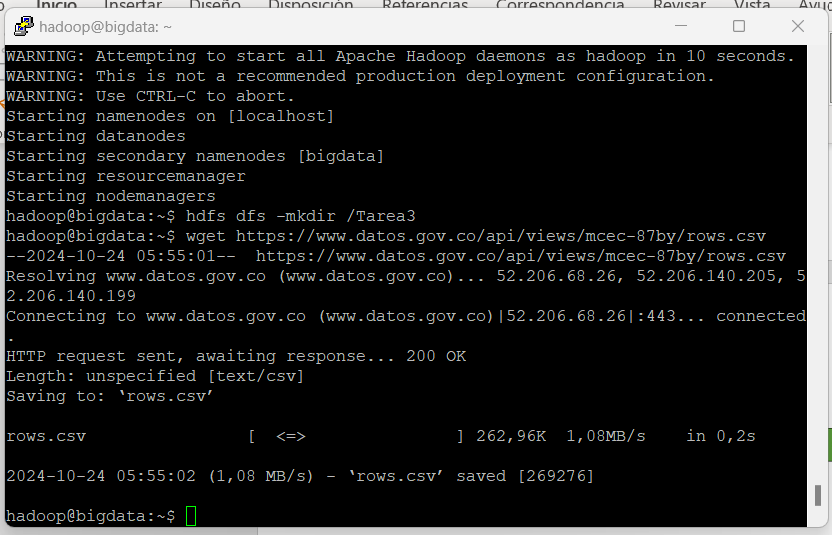
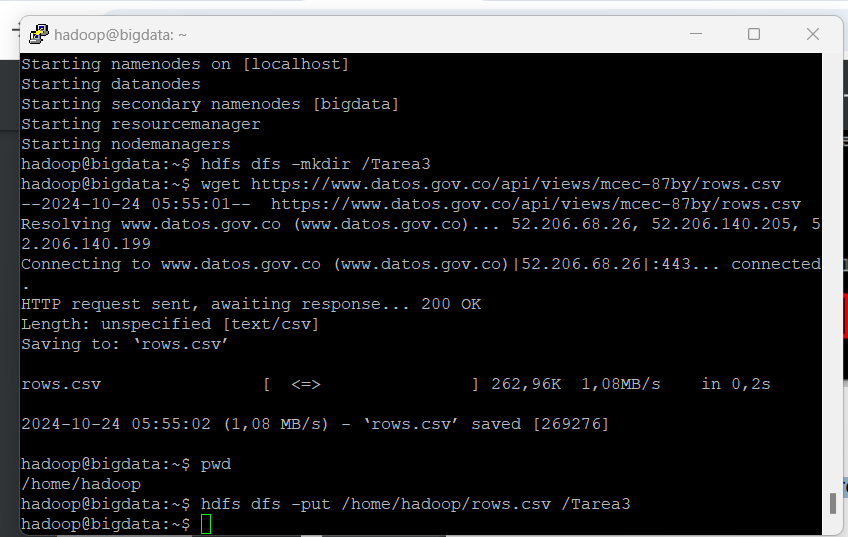


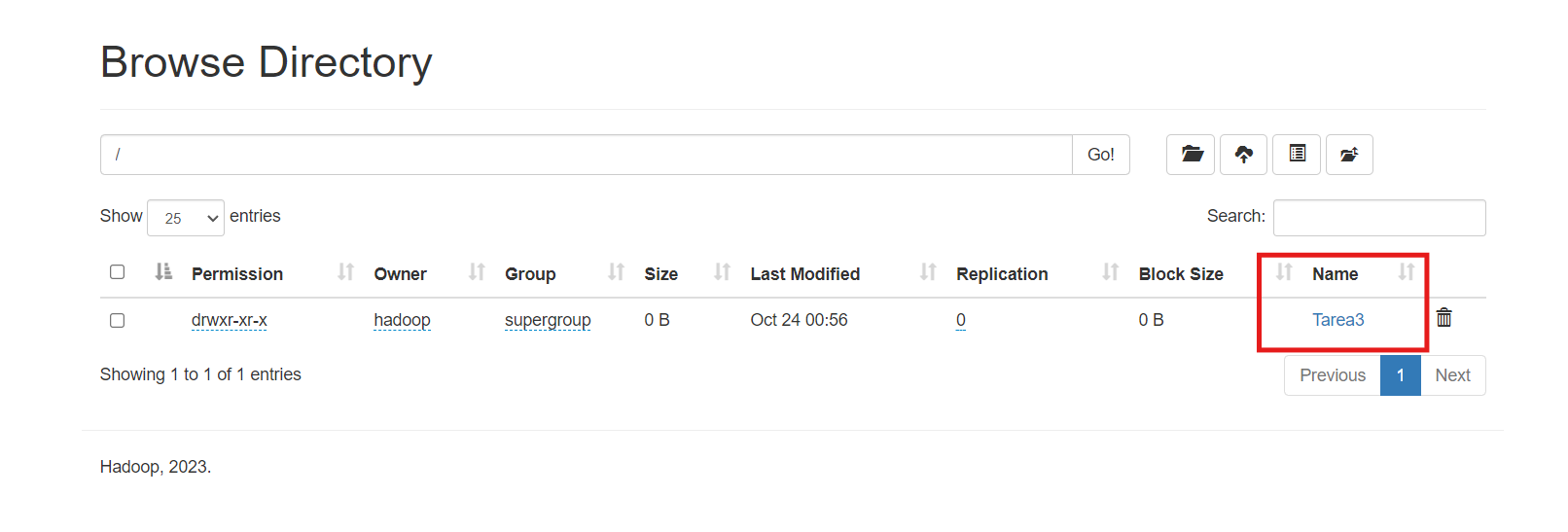
<https://www.datos.gov.co/api/views/mcec-87by/rows.csv>

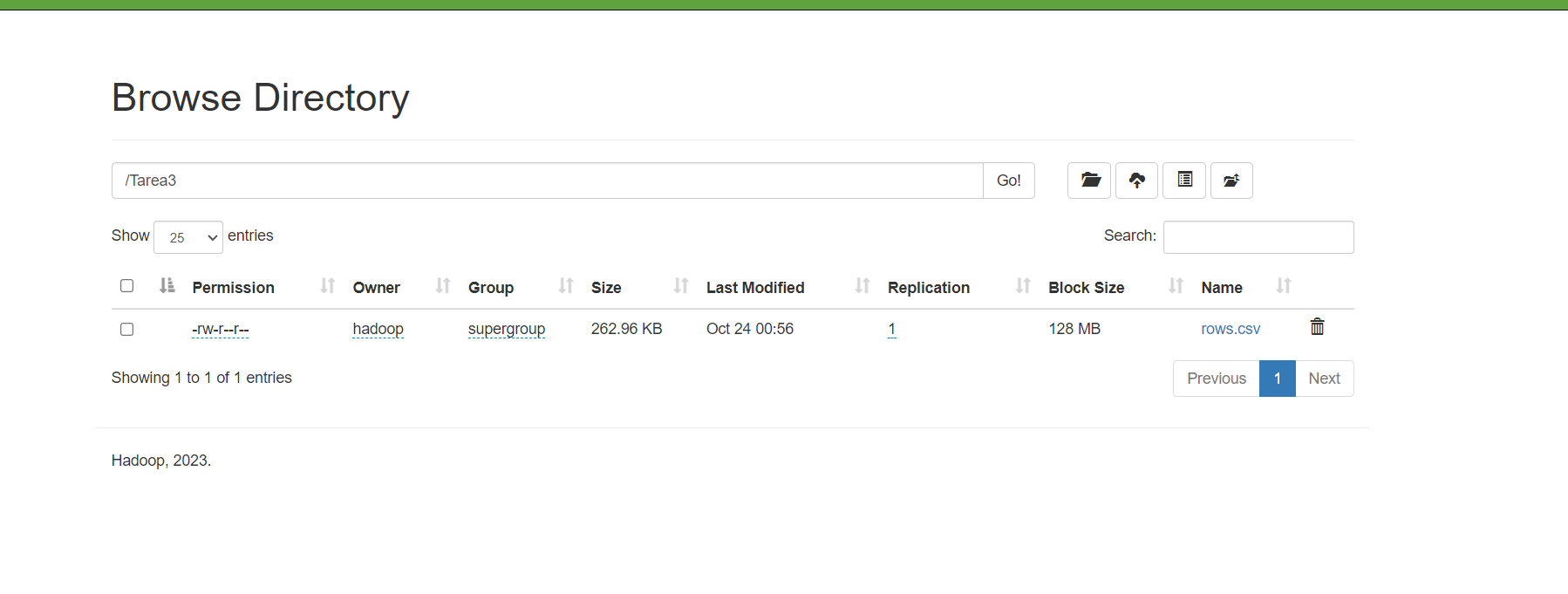
Se descarga el data set en la VM



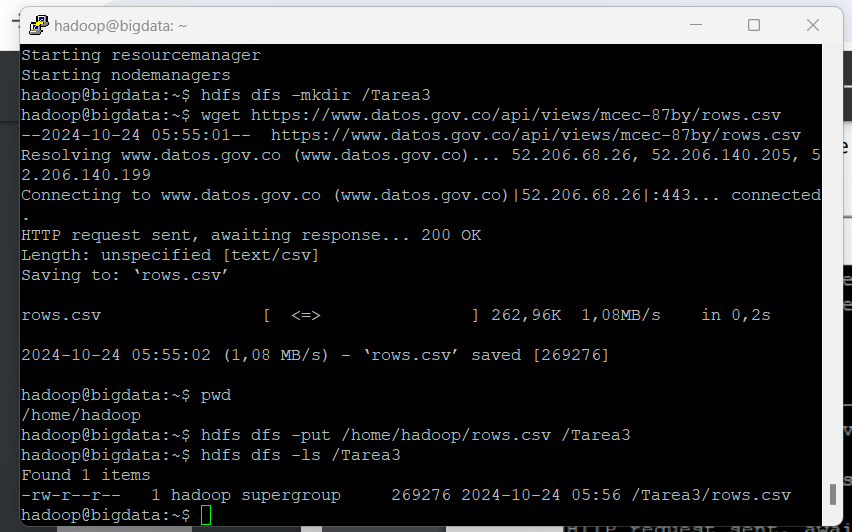
Copiamos el dataset a la carpeta creada:

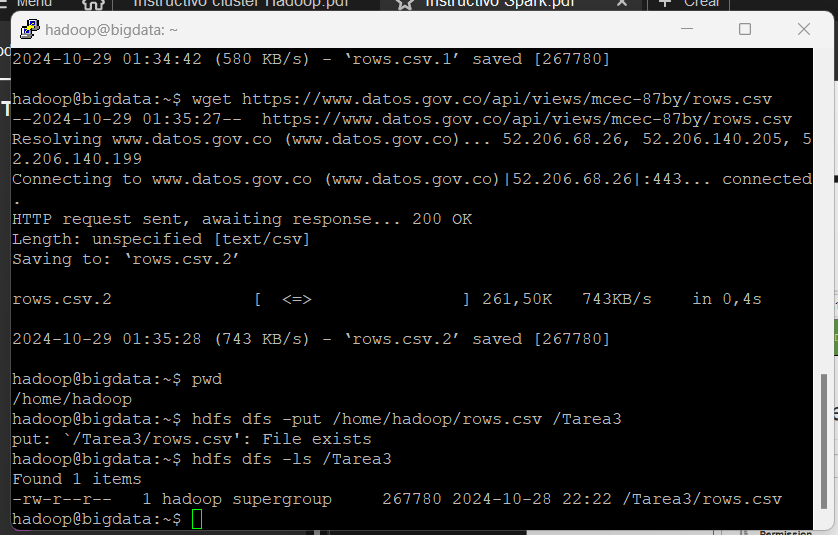


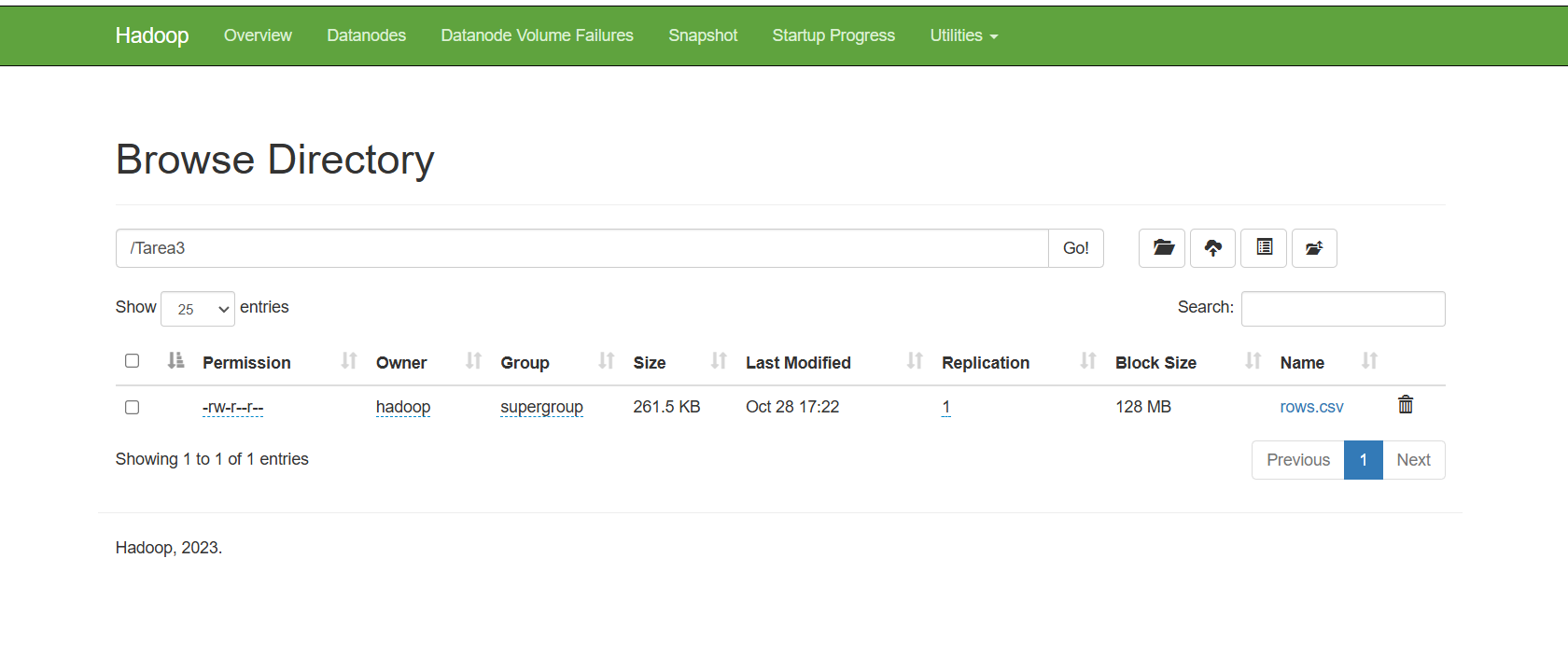




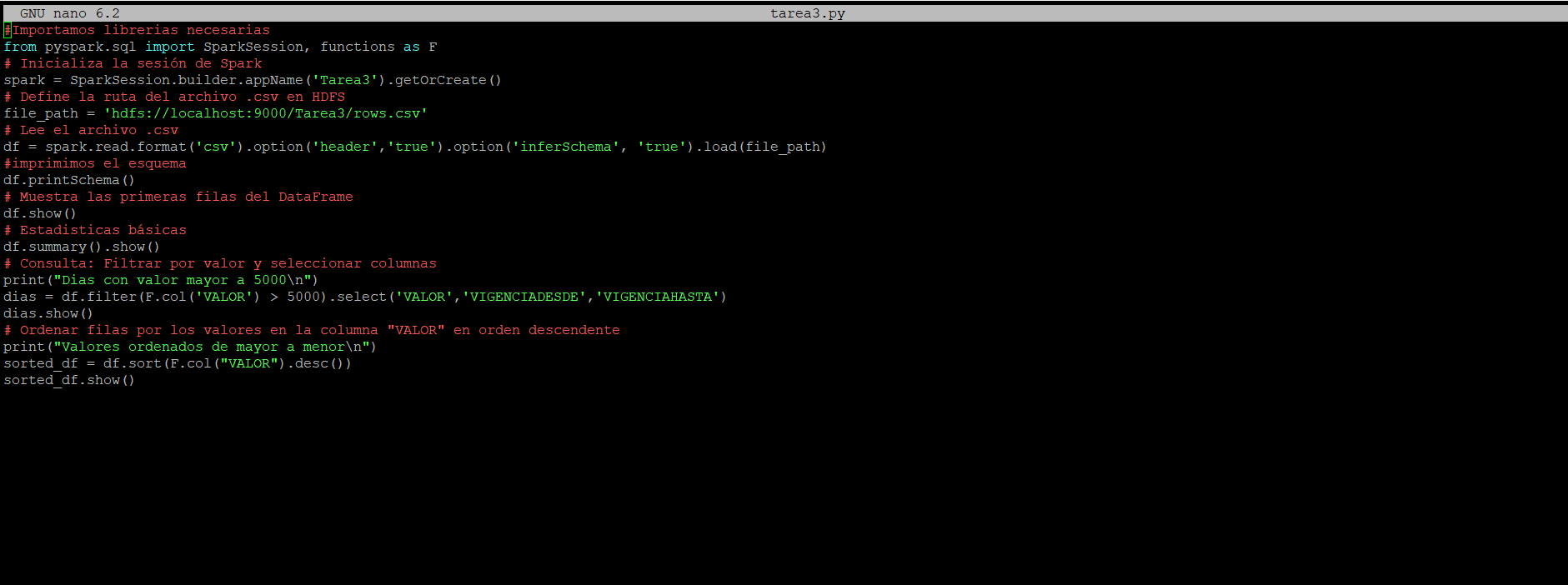
Se puede ver el archivo del dataset en la carpeta HDFS llamada Tarea 3

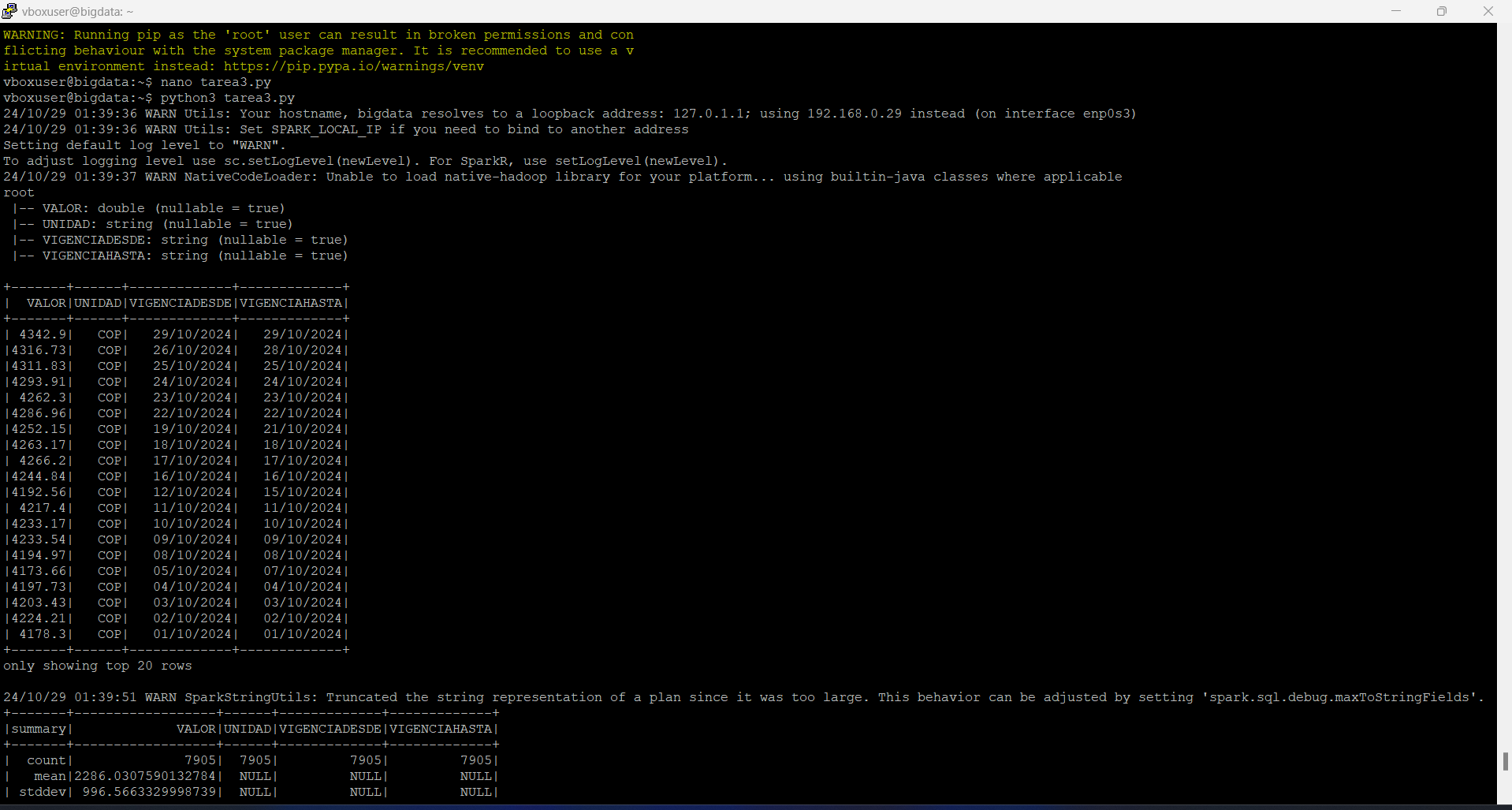


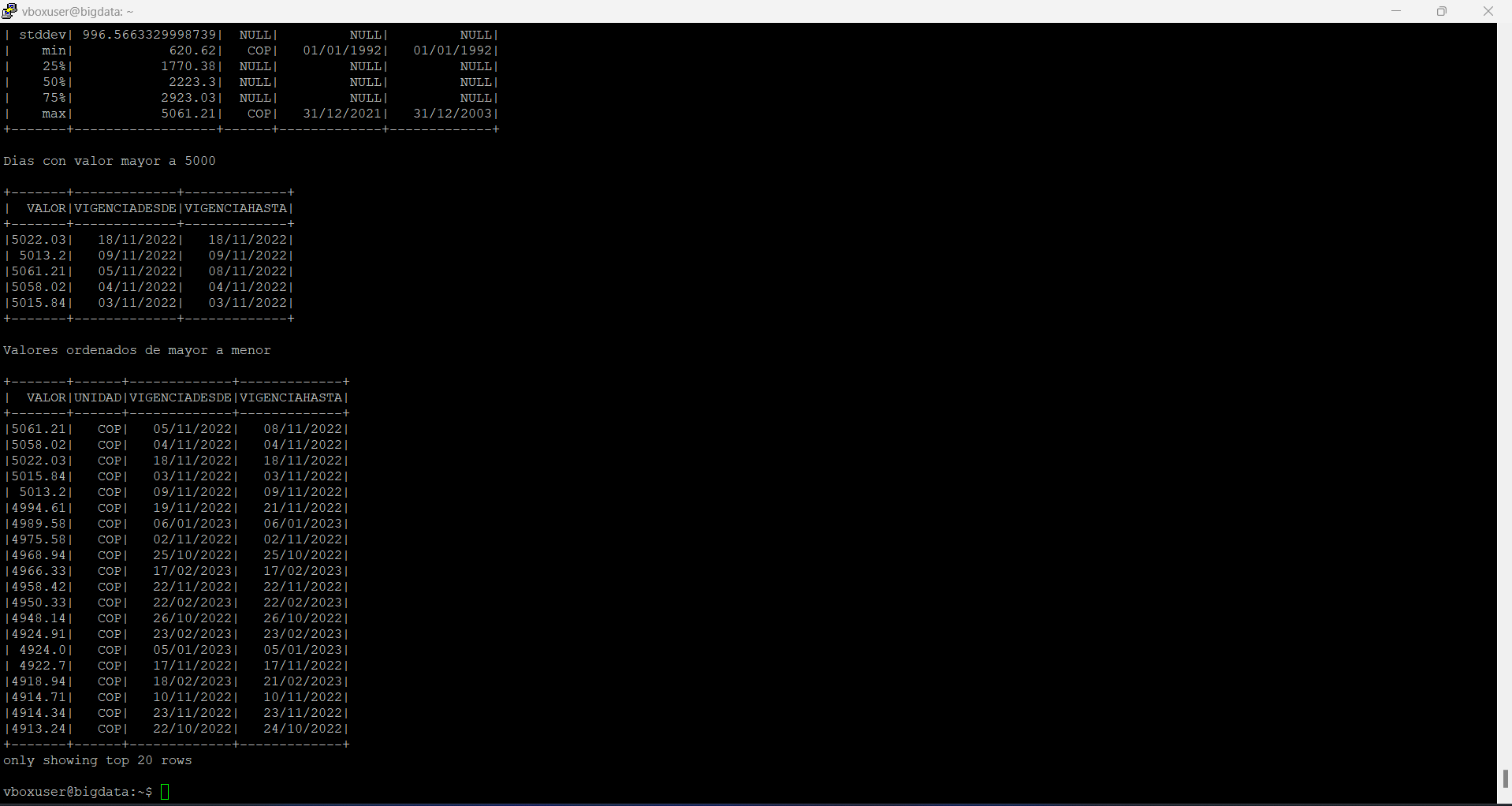


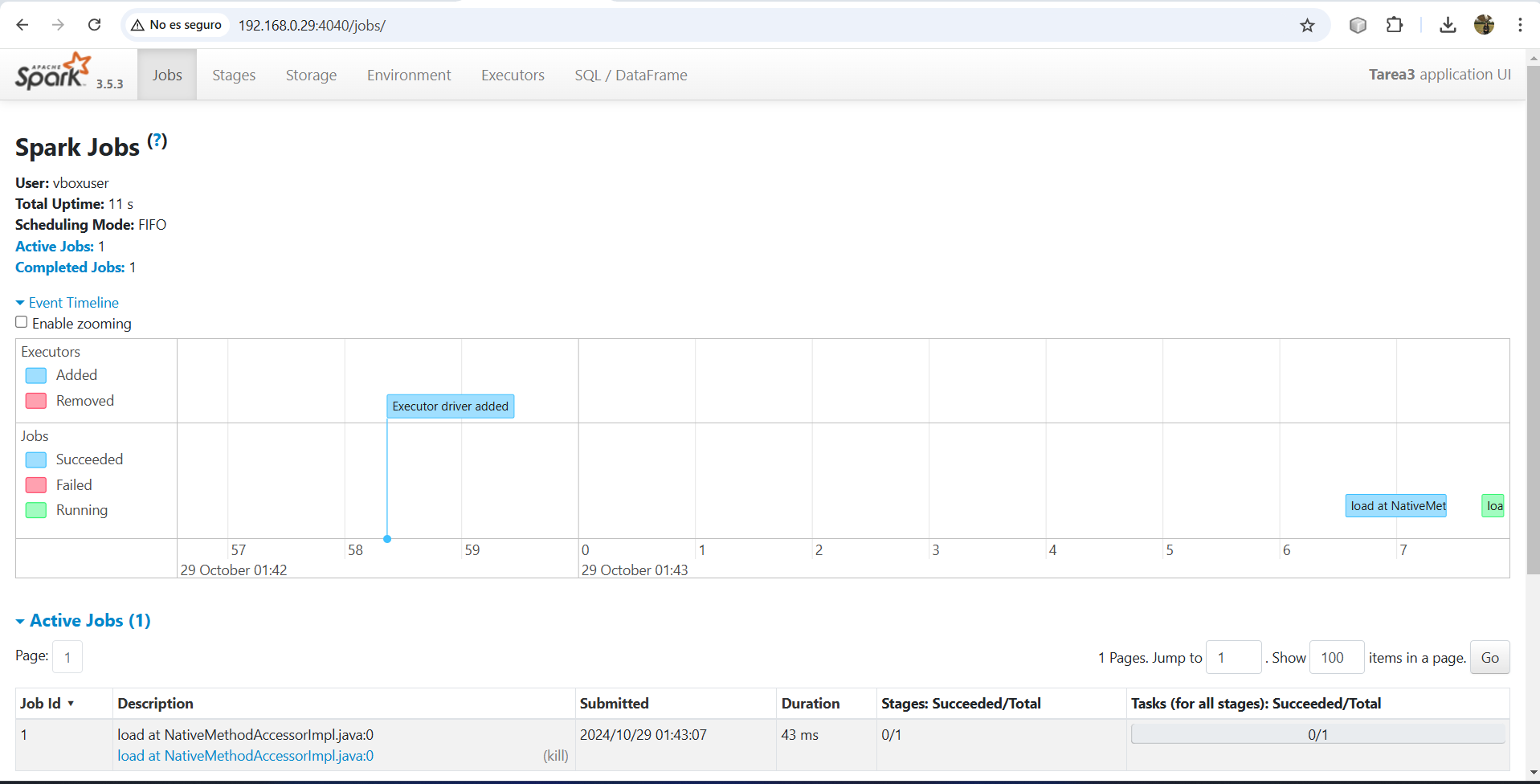


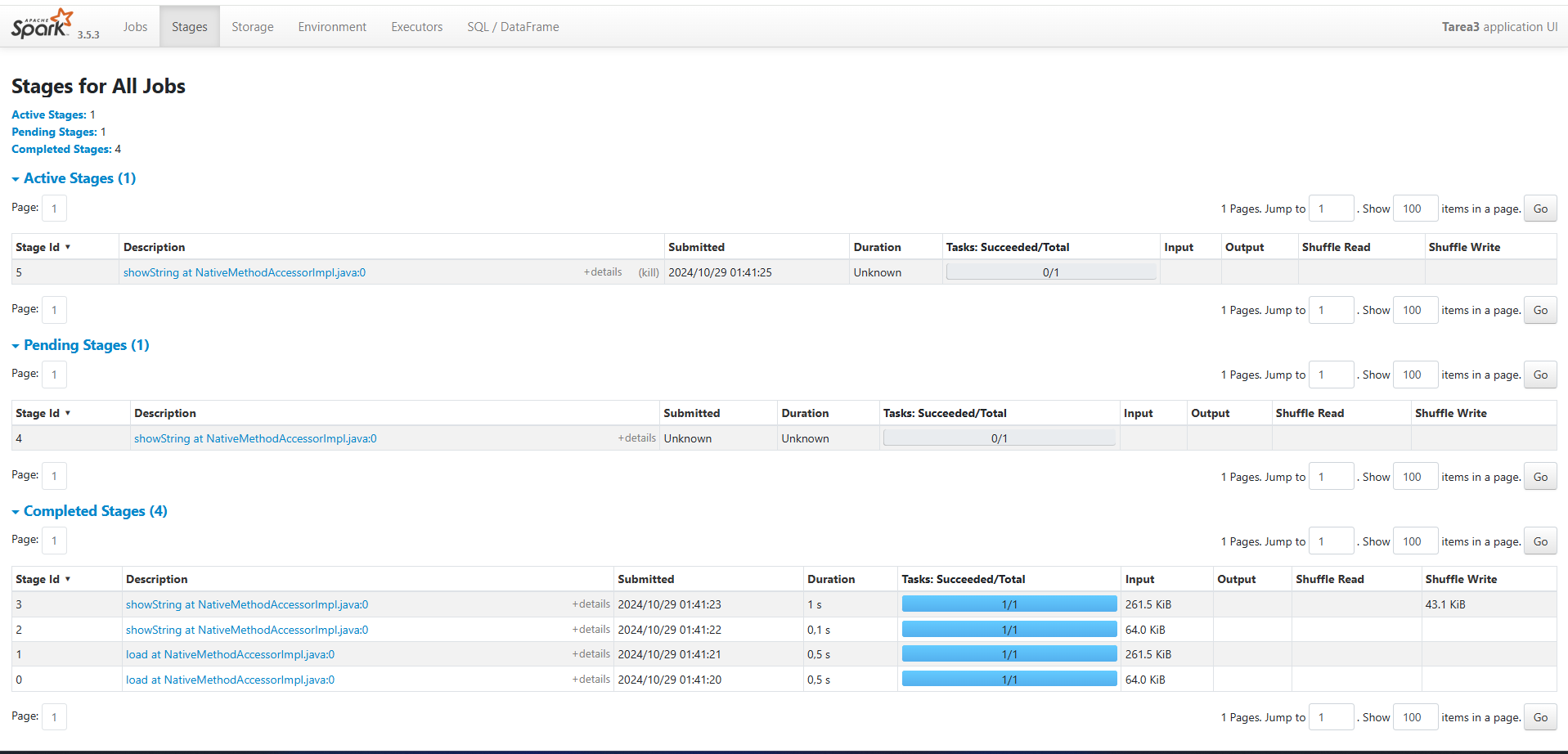
<https://www.datos.gov.co/api/views/mcec-87by/rows.csv>

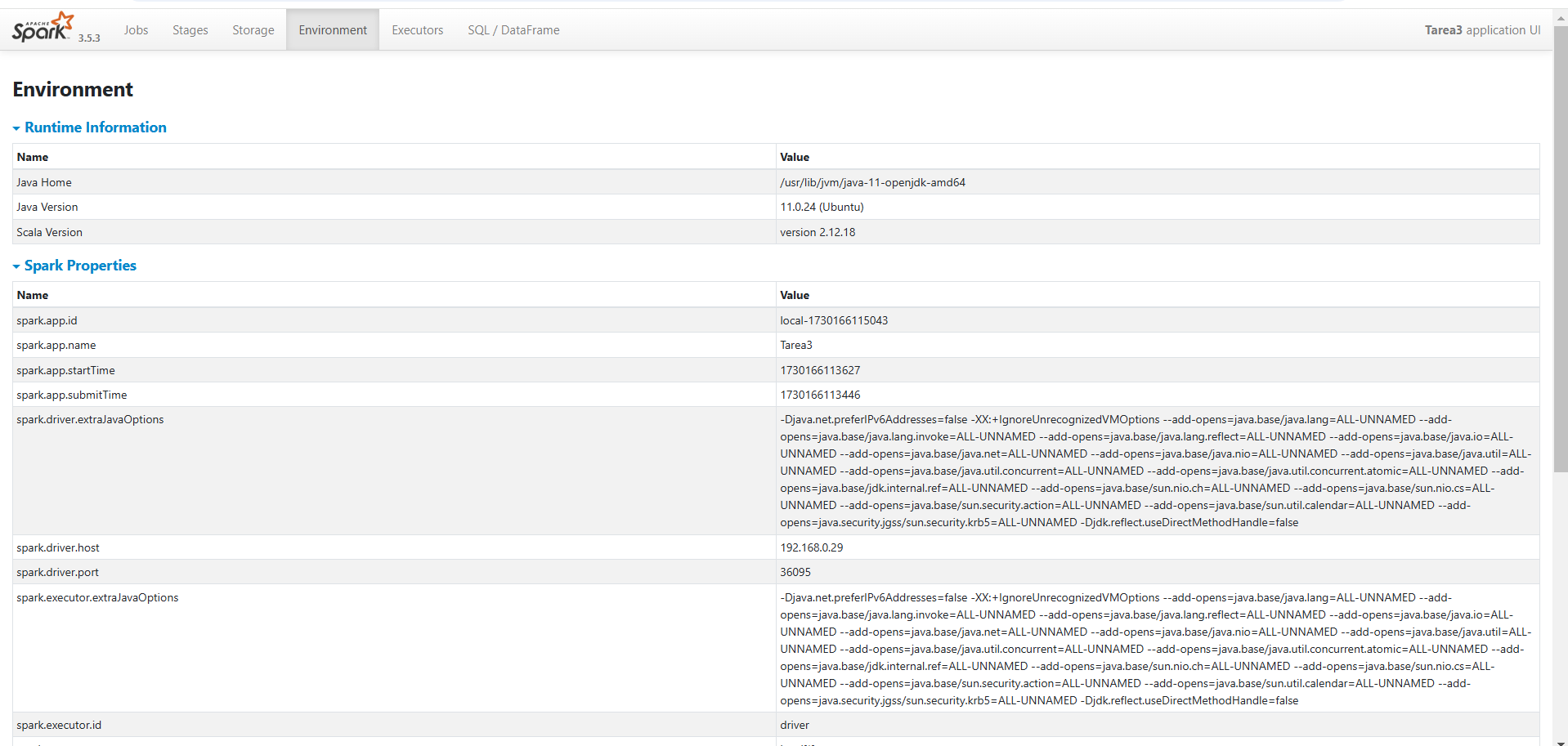


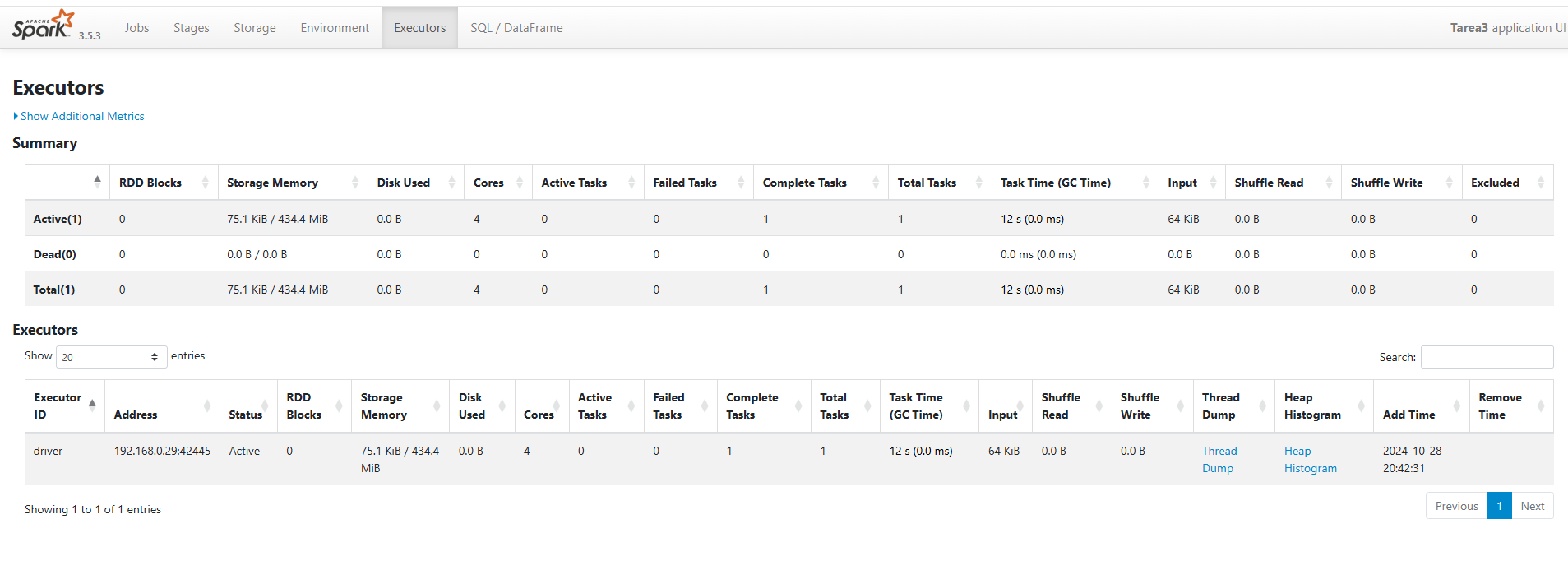




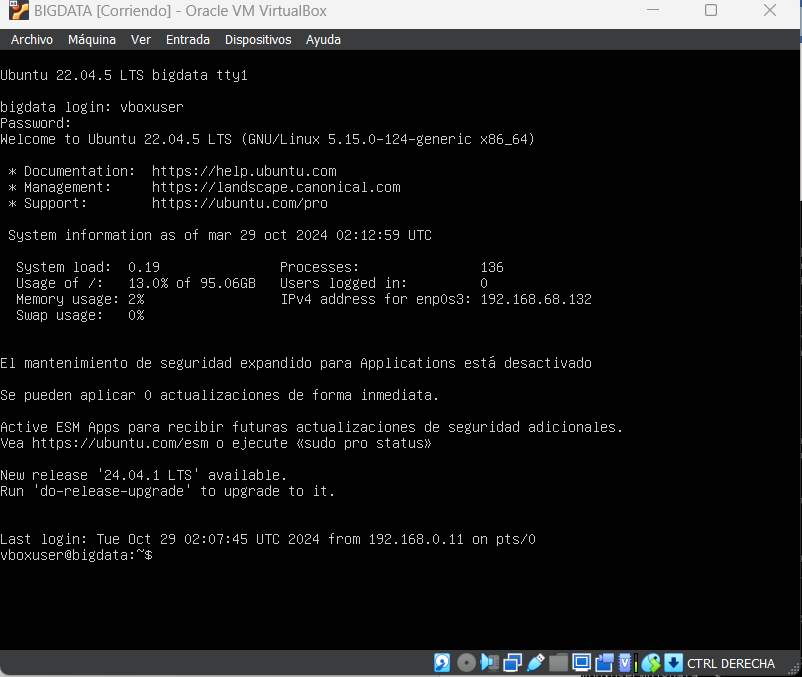


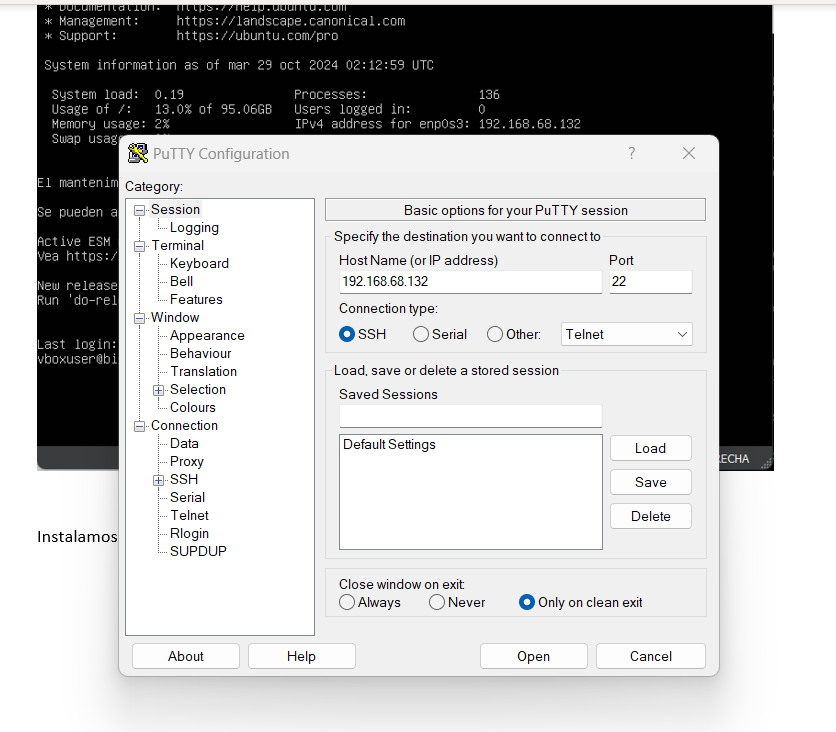




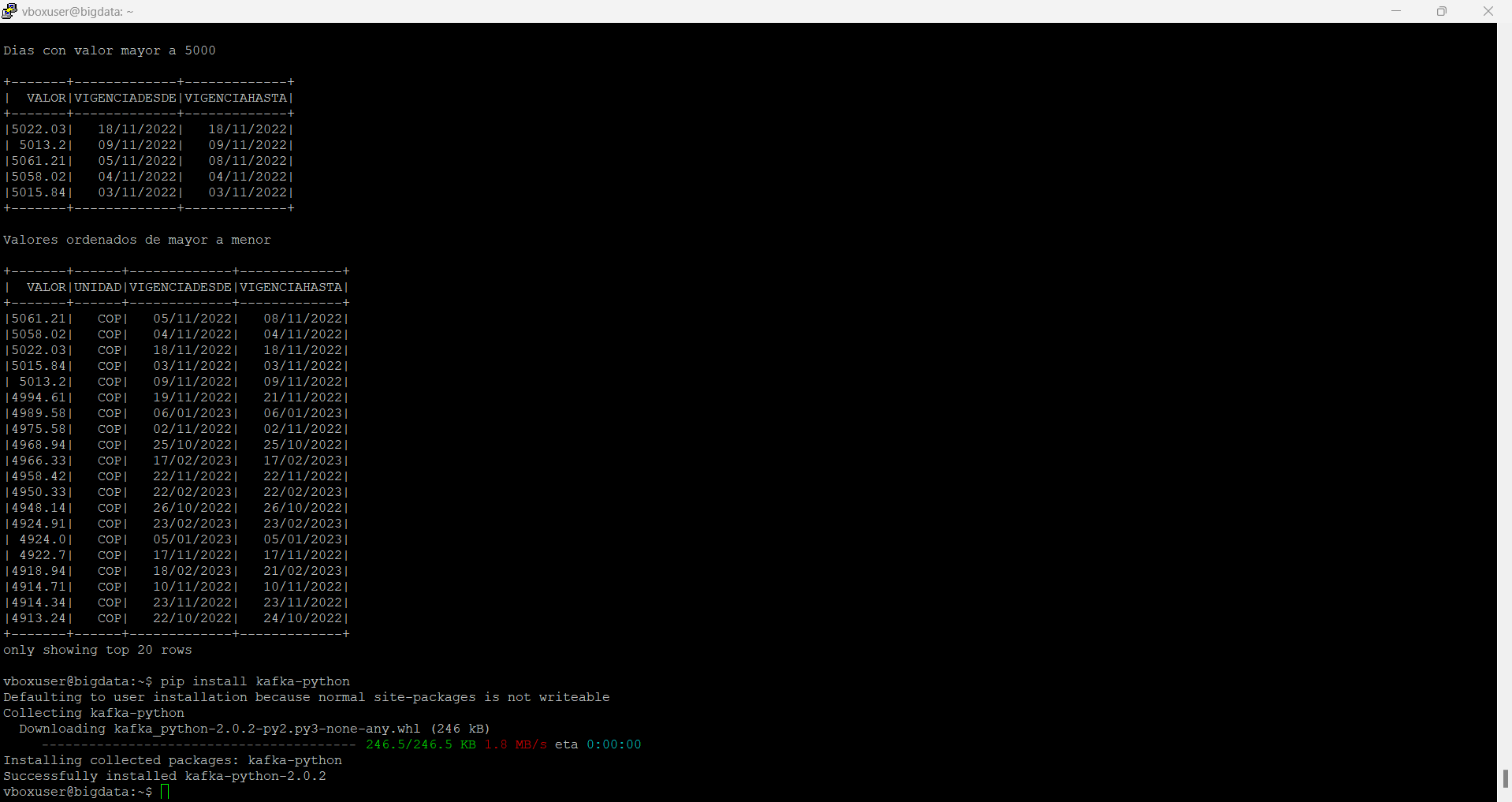


**Análisis en tiempo real con Spark Streaming y Kafka**

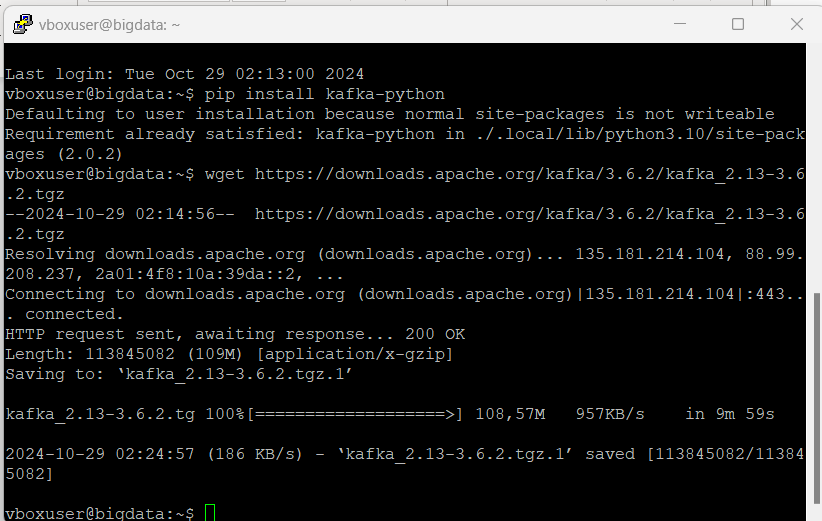
****

****

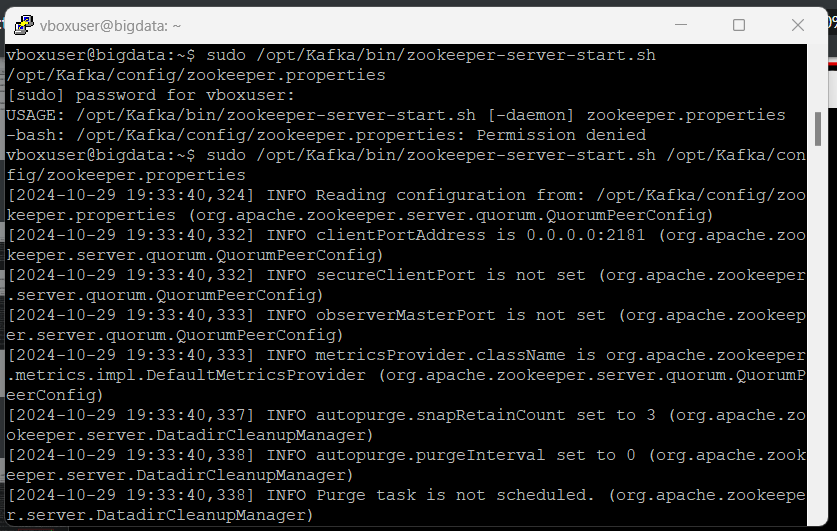
Instalamos mediante PIP la librería de Python Kafka



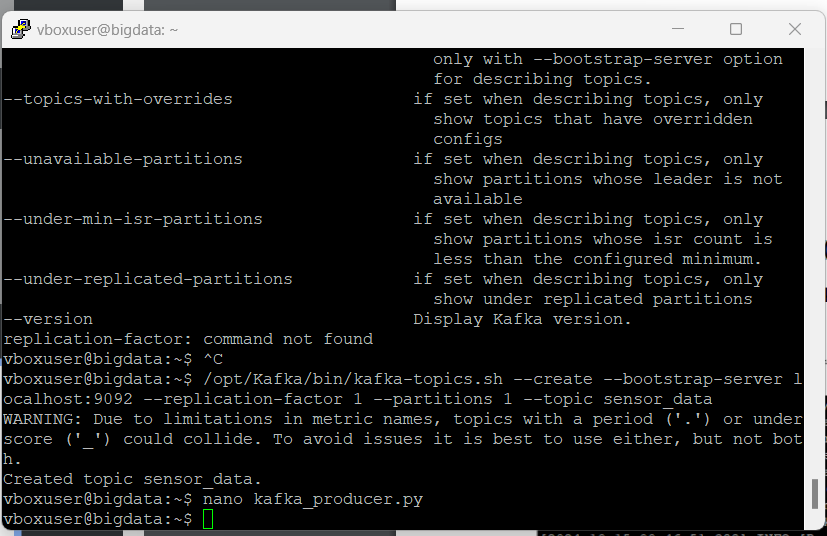
Se descarga, descomprime y mueve de carpeta Apache Kafka



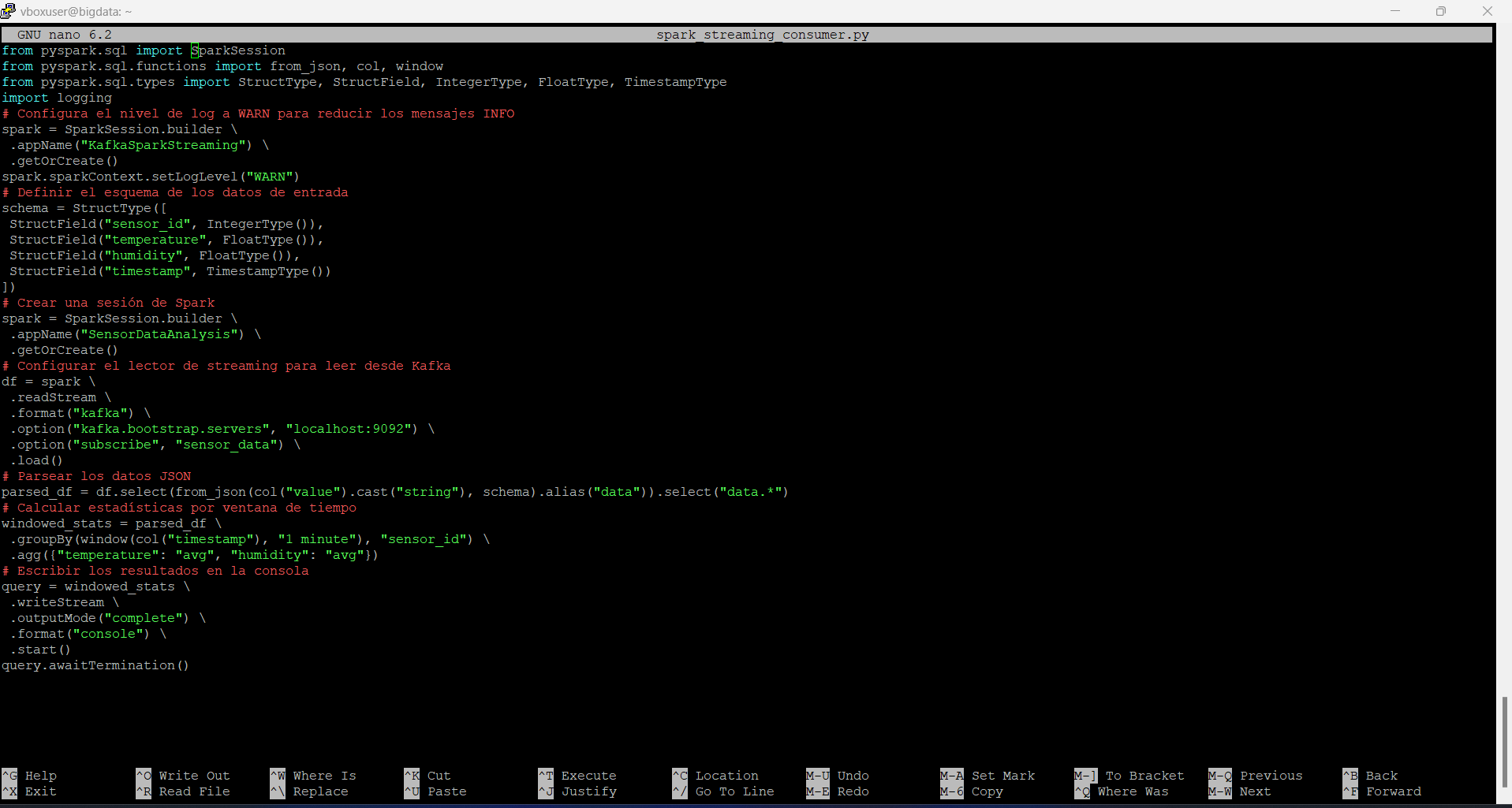
Se inicia ZooKeeper



Implementación del productor(producer) de Kafka

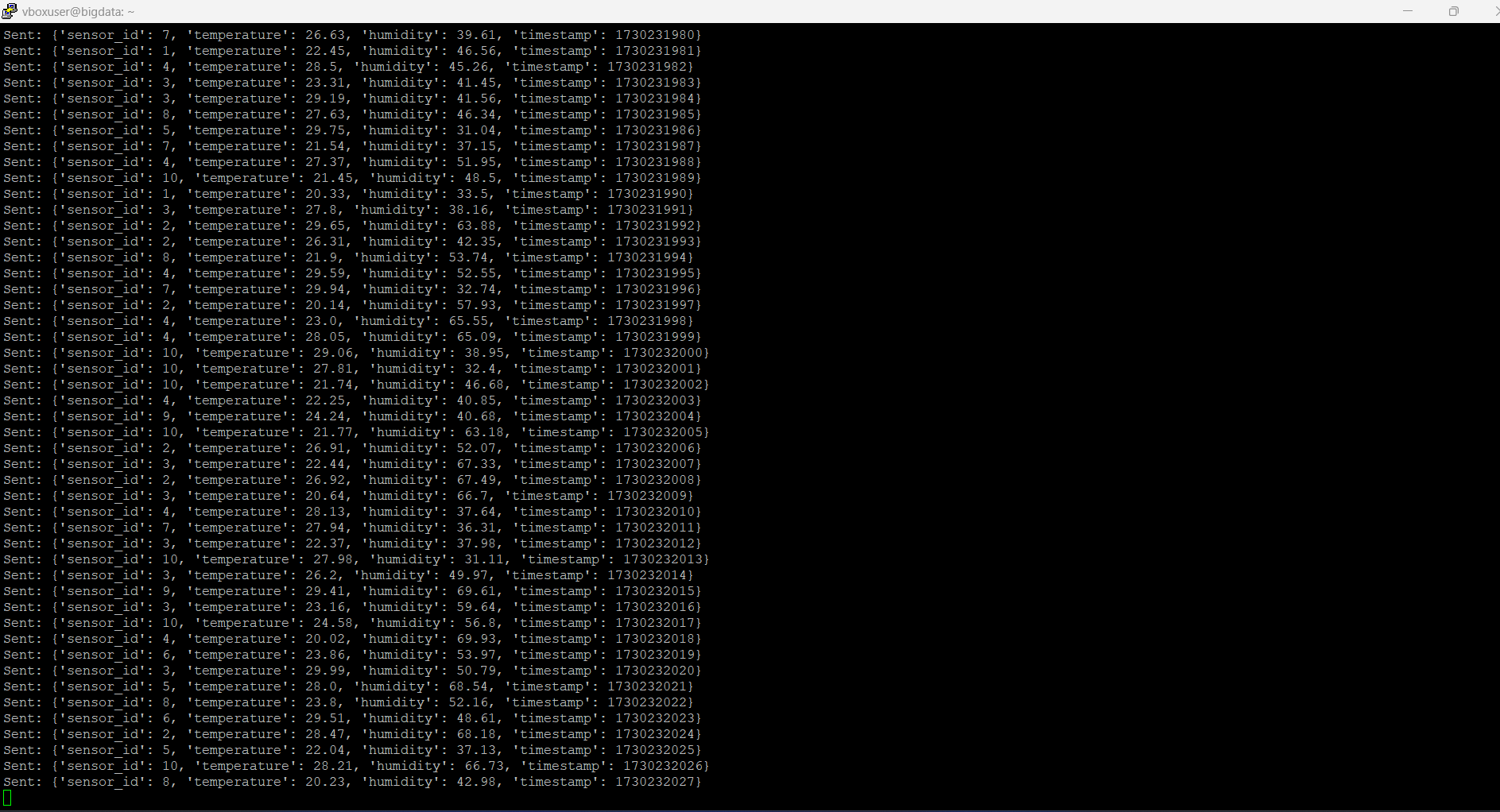


Implementación del consumidor con Spark Streaming



Ejecución y análisis

Ejecución del productor



Ejecución del Streaming

