Universidad de San Carlos de Guatemala

Facultad de Ingeniería

Escuela de Ciencias y Sistemas

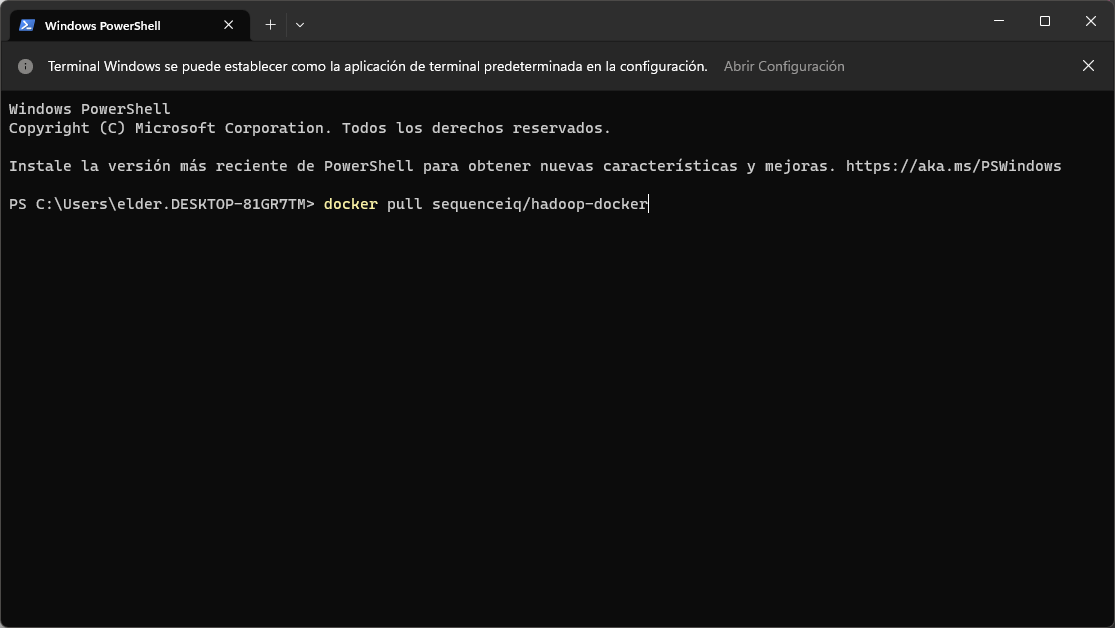
Seminario de Sistemas 2

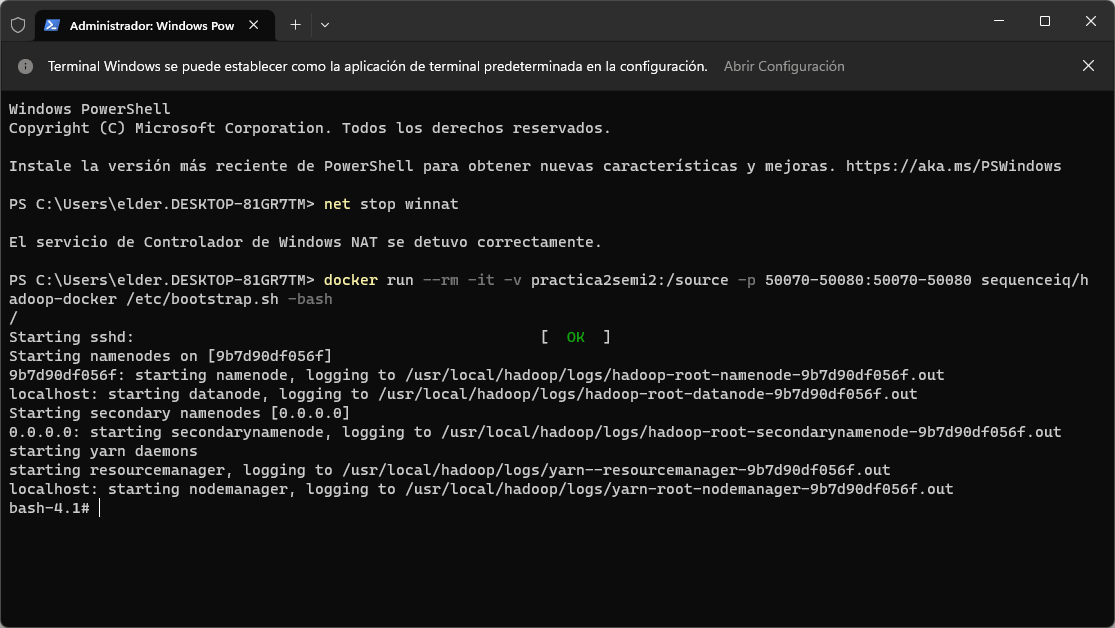
**DOCUMENTACIÓN PRÁCTICA #2**

Elder Anibal Pum Rojas 201700761

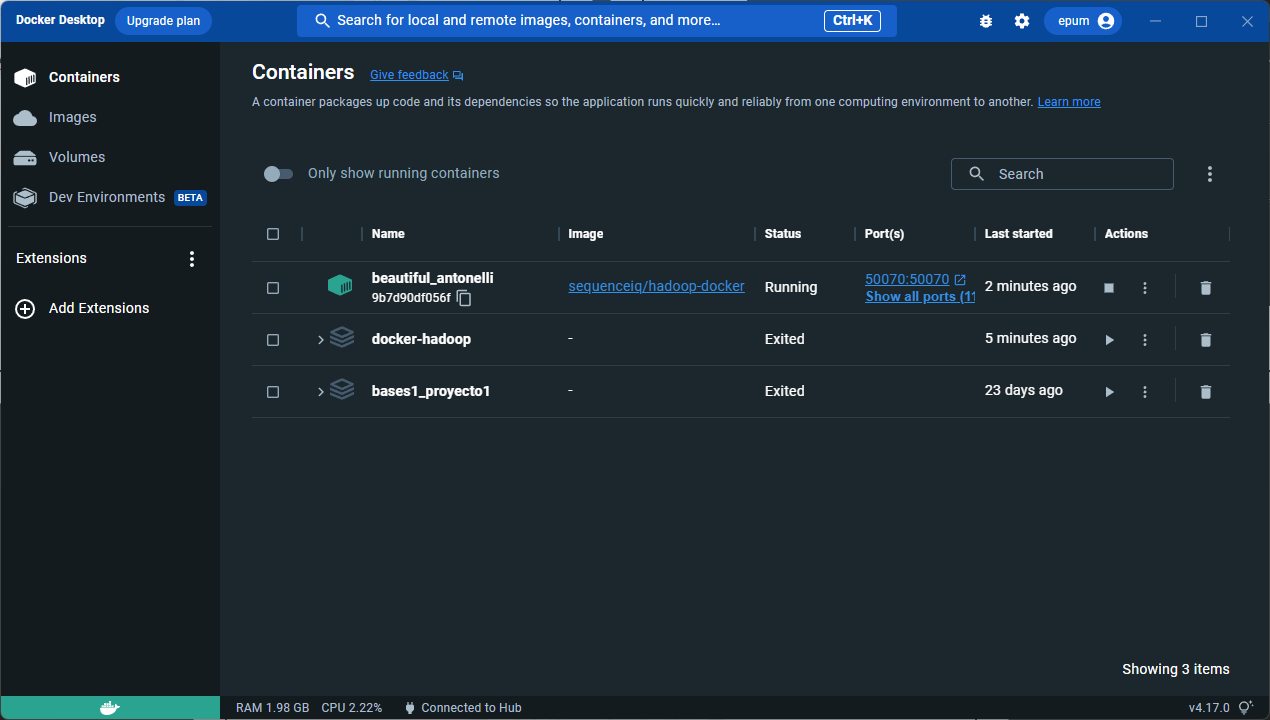
Guatemala, 2 de Abril del 2023

**Pasos a seguir para completar la práctica**

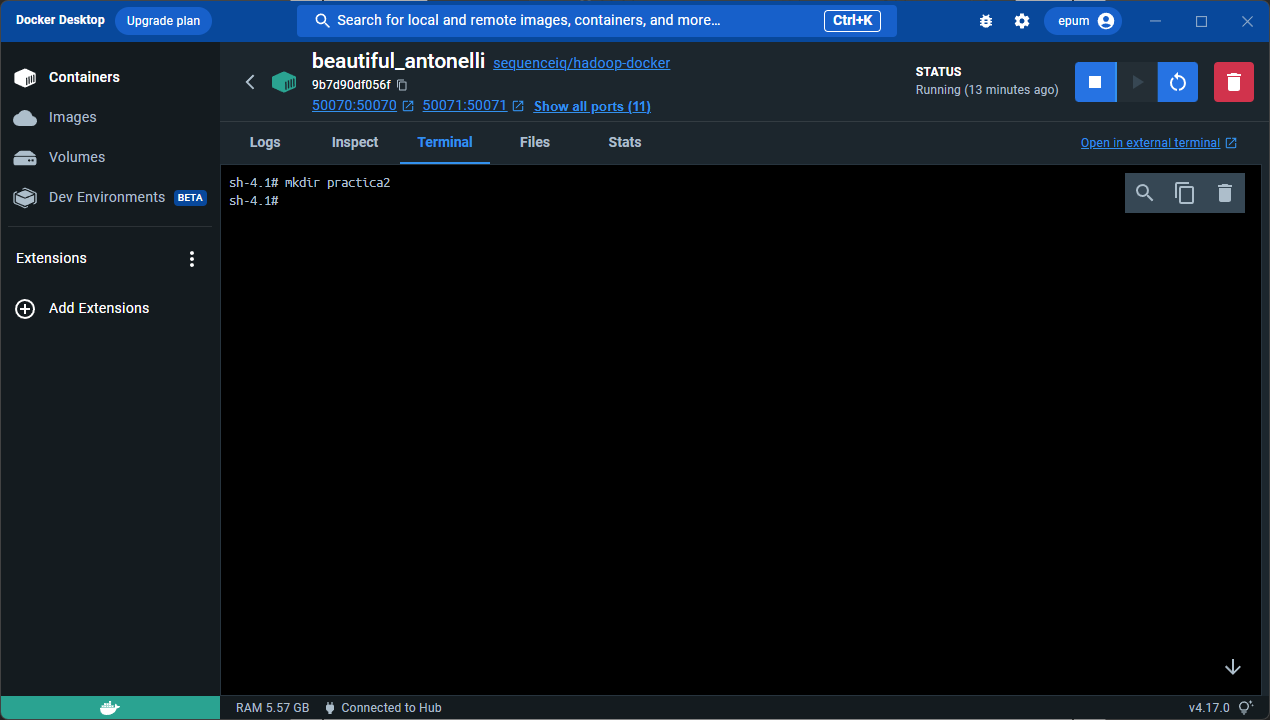
1. Primeramente, necesitamos descargar la imagen de Hadoop, para esto utilizaremos Docker (ya sea en Windows o Linux) y correremos el siguiente comando en una terminal:
2. Después de que se descargue la imagen de Hadoop, vamos a correrla utilizando el siguiente comando (tener en cuenta que cada vez que se apaga la imagen, esta se elimina y por ende, hay que volver a correr el comando. Además, hay que matar el proceso “winnat” porque usa los puertos que utilizaremos para Hadoop):



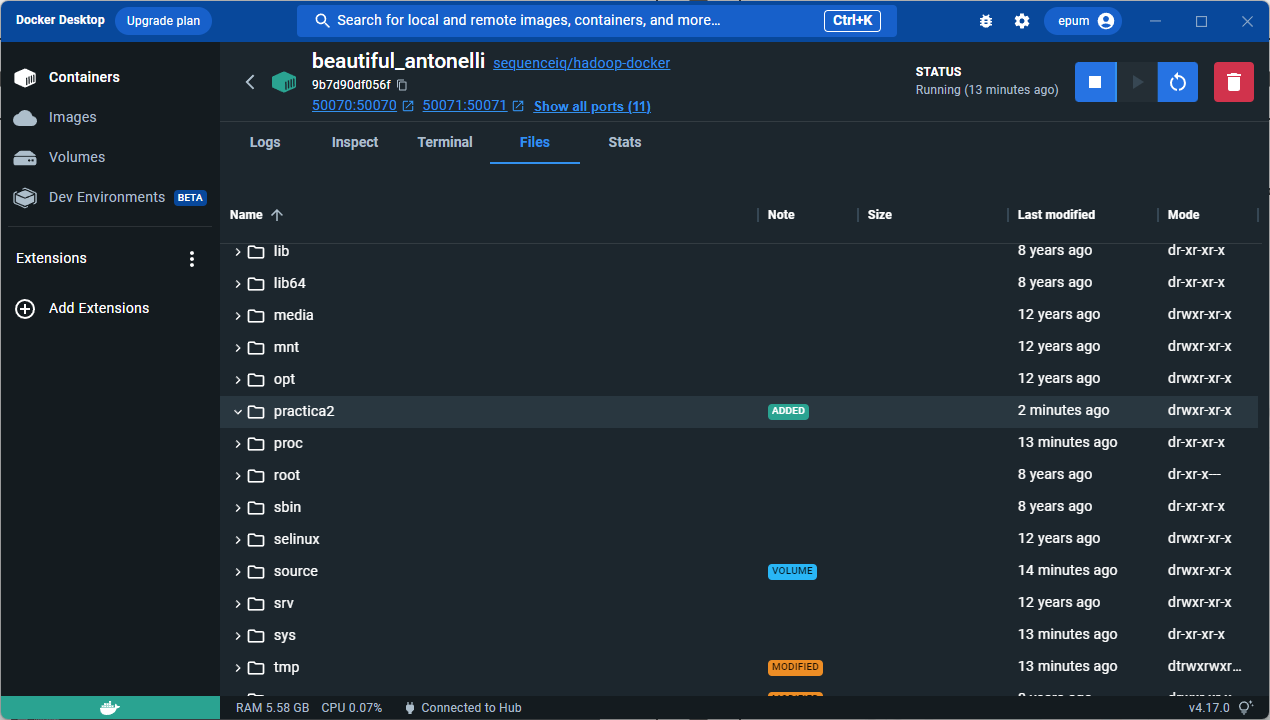
1. Revisando en Docker Desktop tenemos que el contenedor efectivamente está corriendo:



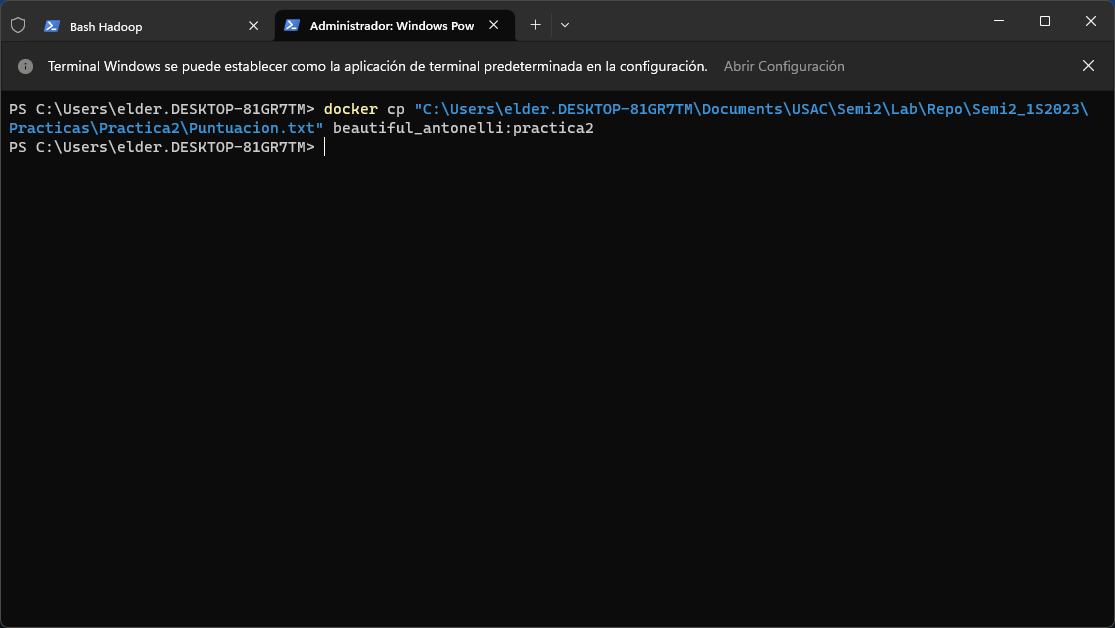
Ahora lo que vamos a hacer es crear una carpeta llamada “practica2” dentro del contenedor, para lo cual vamos a utilizar la terminal del contenedor en Docker Desktop y meter el siguiente comando:



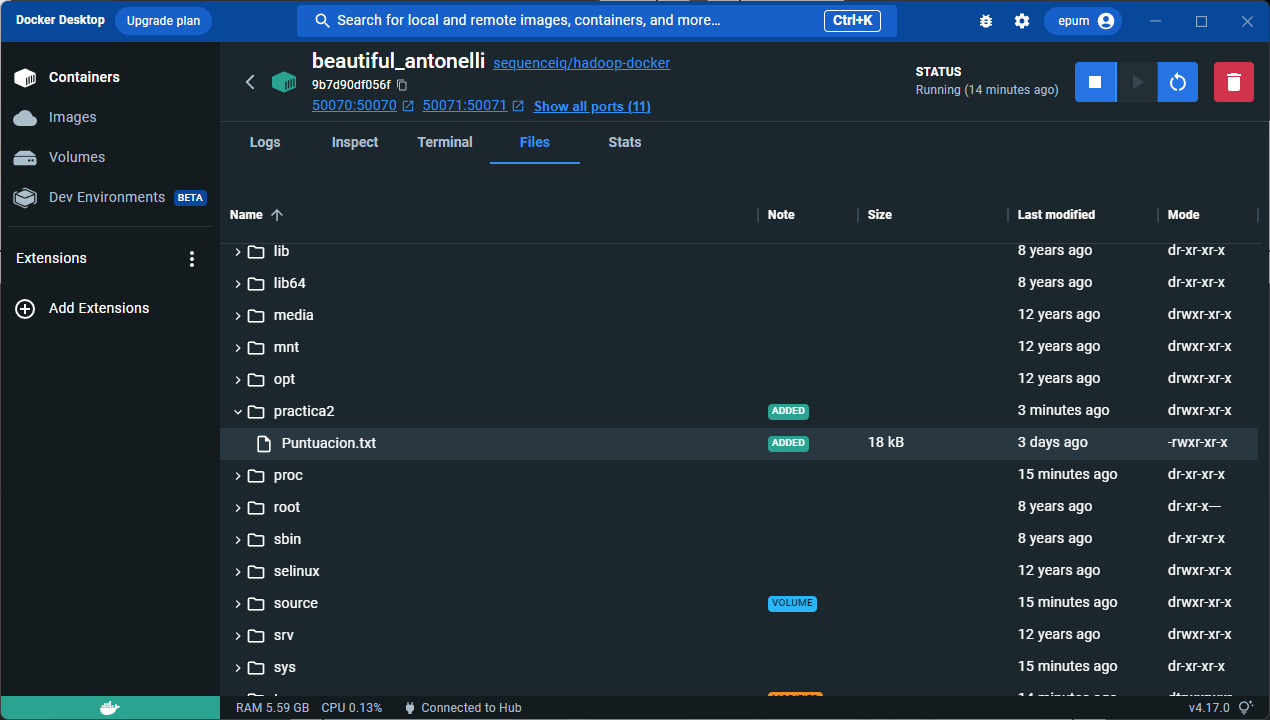
Y revisando en Docker Desktop en la parte de Files veremos que se ha creado la carpeta correctamente:



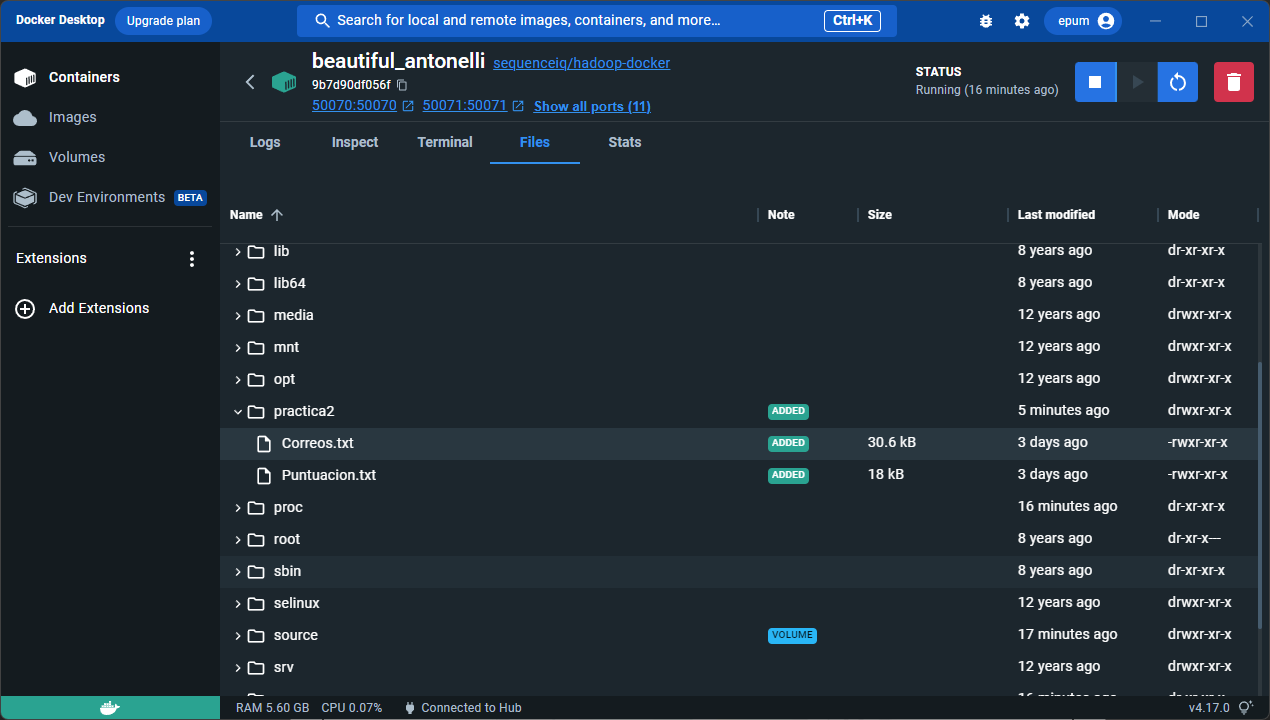
Ahora lo que vamos a hacer es copiar el archivo “Puntuacion.txt” dentro de la carpeta “practica” que acabamos de crear, para ello, vamos a utilizar una nueva terminal de Windows y usar el siguiente comando:



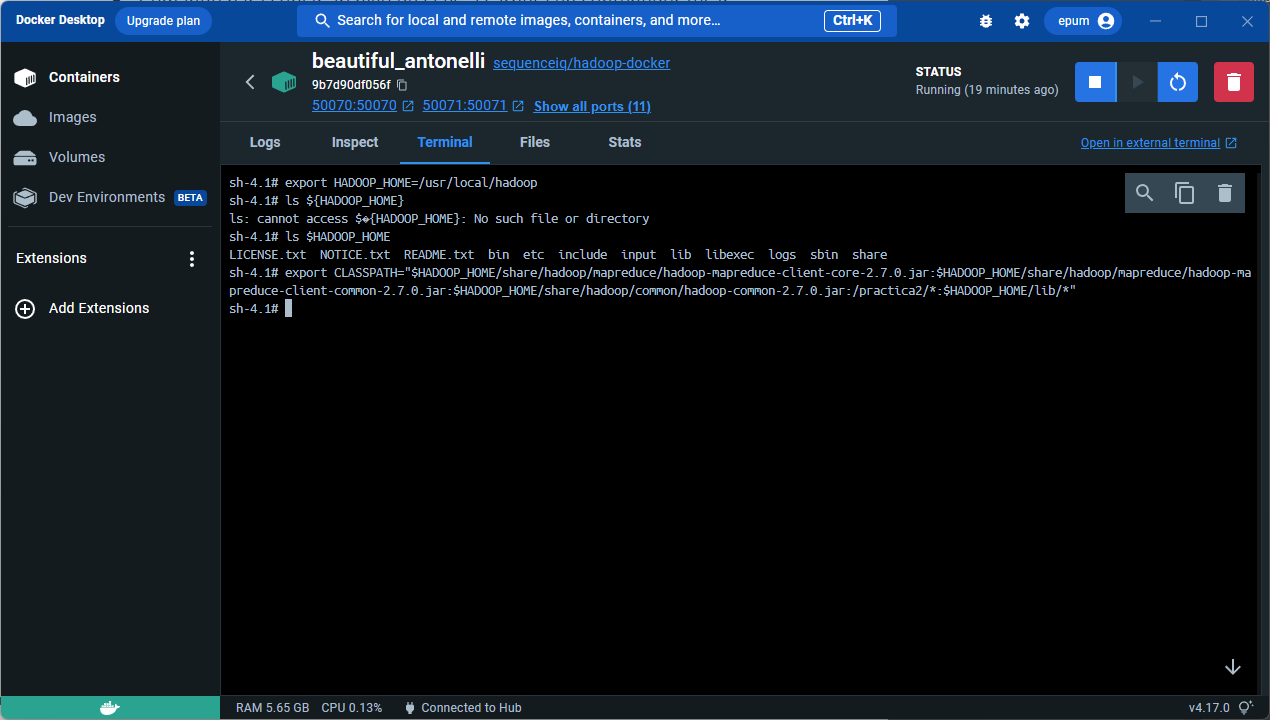
Y nuevamente verificamos que el archivo se ha agregado:



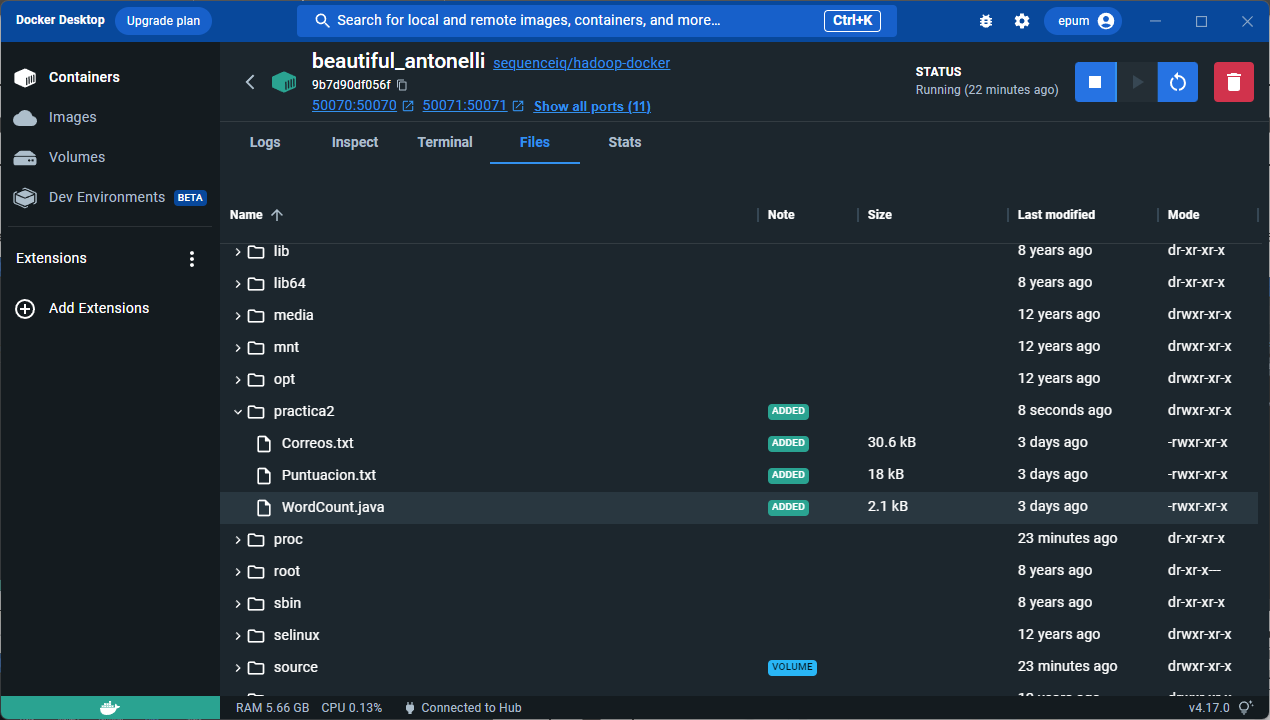
1. Vamos a hacer lo mismo para el archivo “Correos.txt” y verificaremos que el archivo esté dentro de la misma carpeta:



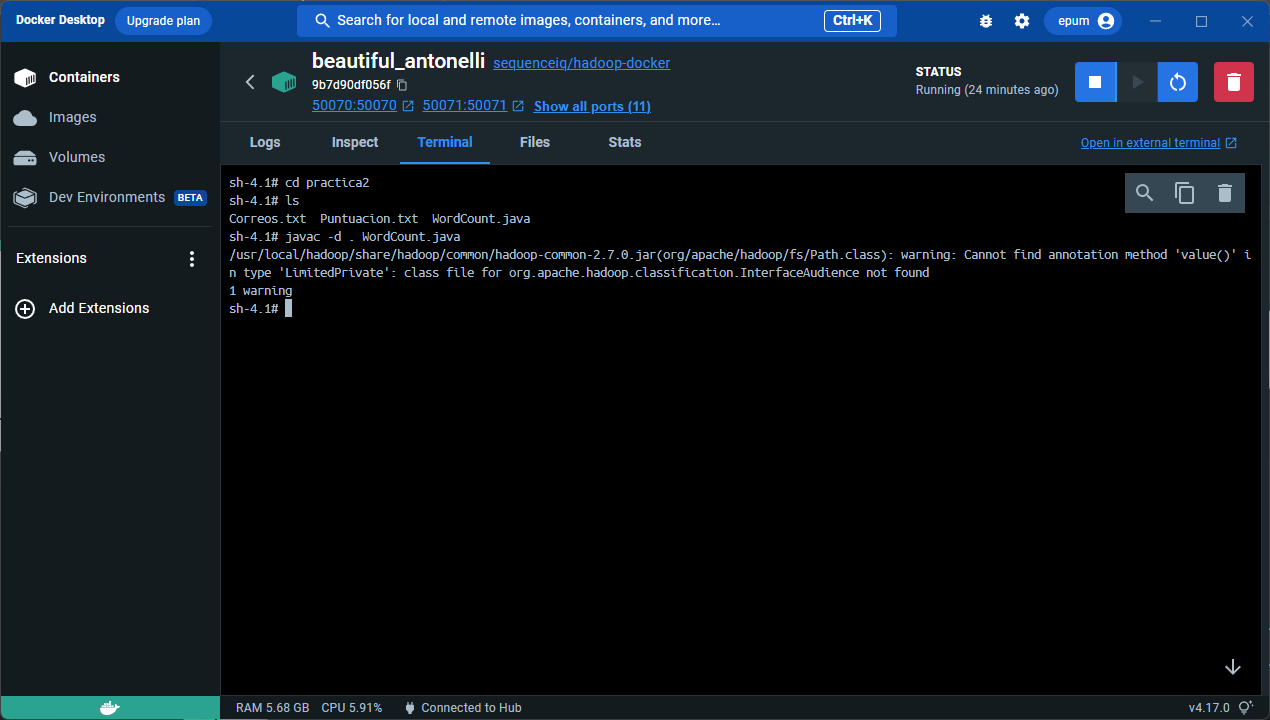
1. Ahora dentro de la terminal del contenedor vamos a iniciar las variables HADOOP\_HOME y CLASSPATH:



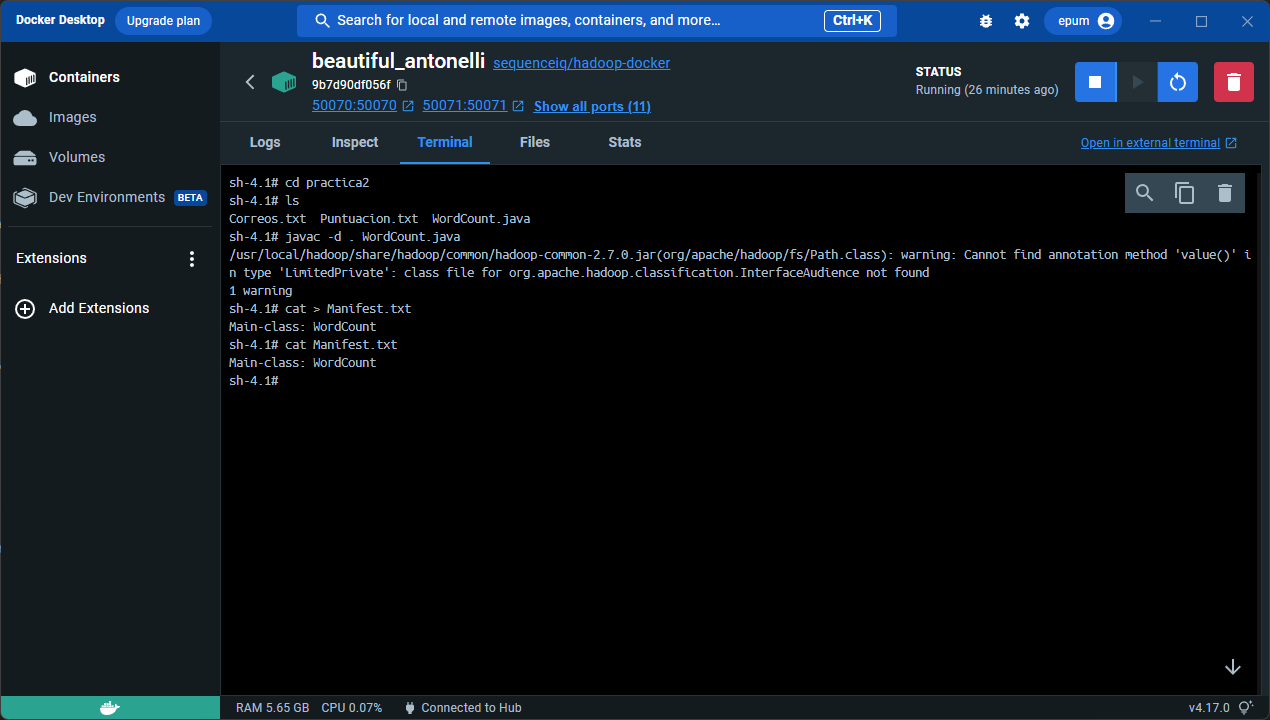
1. Antes de poder compilar, necesitamos subir igualmente al contenedor el archivo “WordCount.java” en la carpeta “practica2” utilizando el mismo método que utilizamos para subir los 2 archivos de texto anteriormente, luego de eso verificaremos que se encuentre el archivo dentro del contenedor:



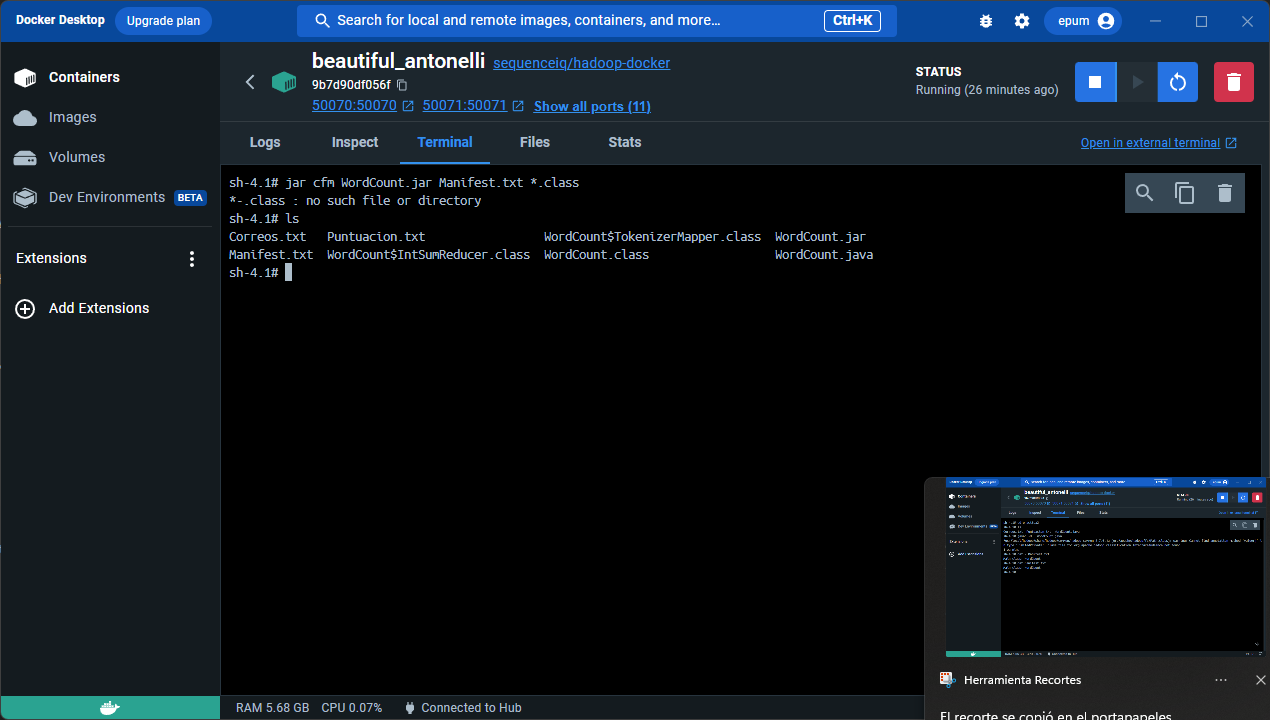
1. Vamos a compilar el archivo java con el siguiente comando:



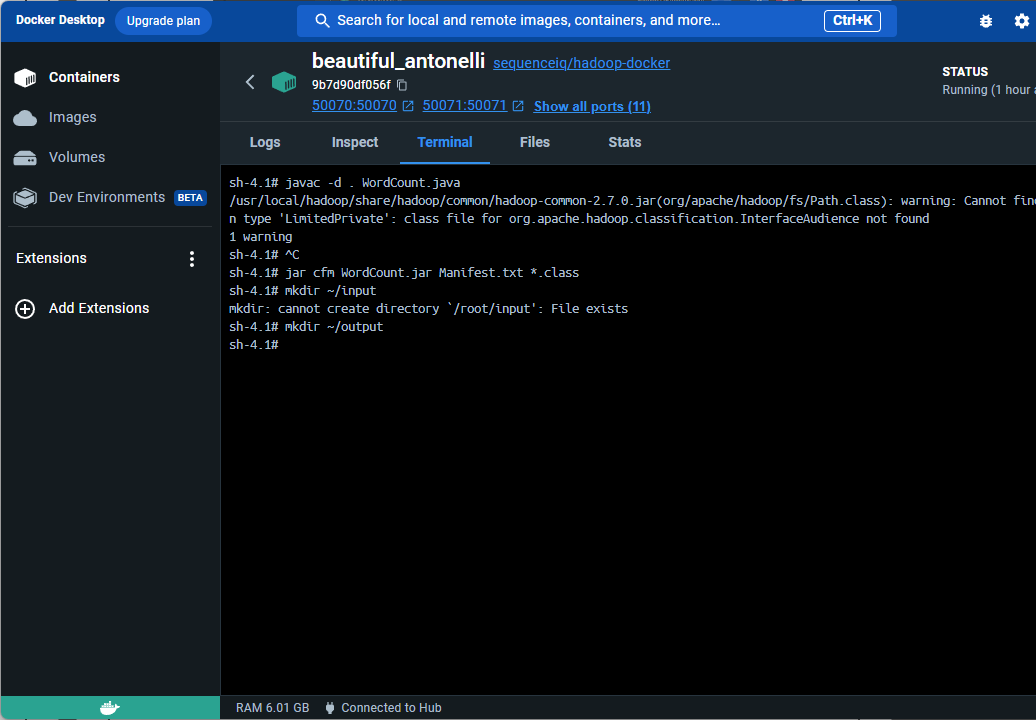
1. Ahora creamos un archivo Manifest.txt para indicar cual es el archivo que contiene el método Main del archivo java, para esto, utilizamos los siguientes comandos:

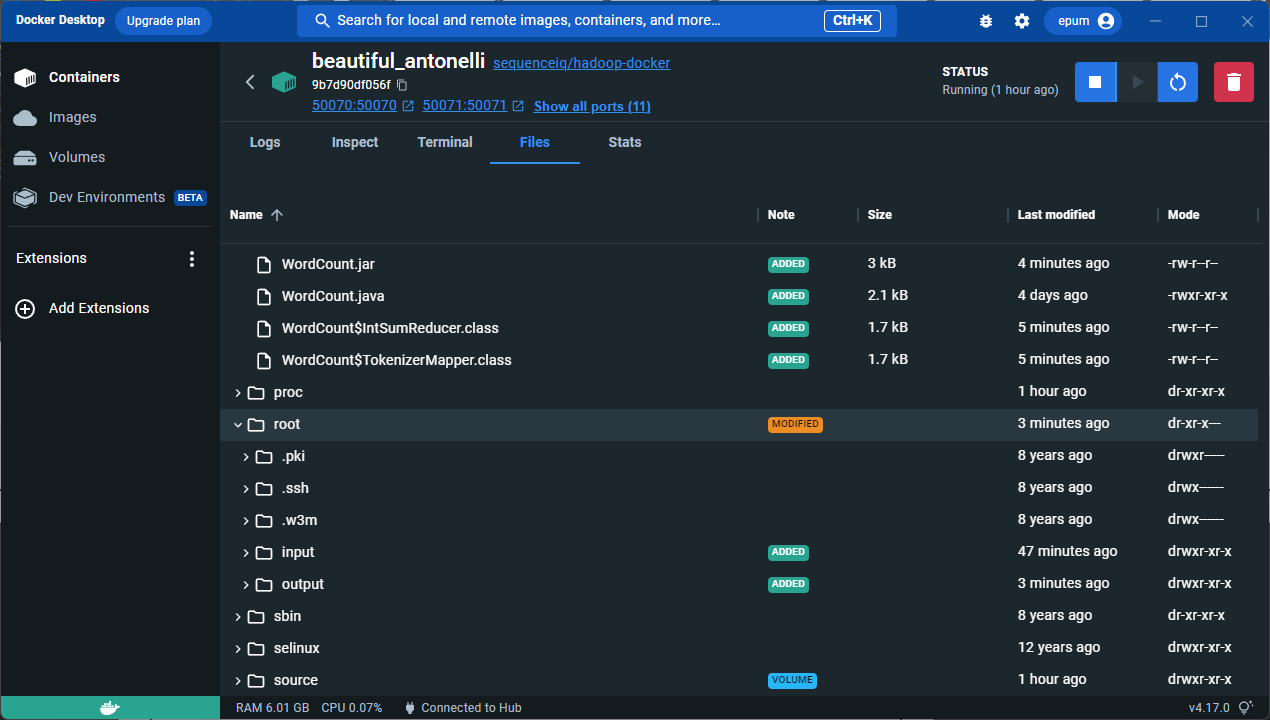


1. Ahora creamos un archivo .jar utilizando los siguientes comandos:

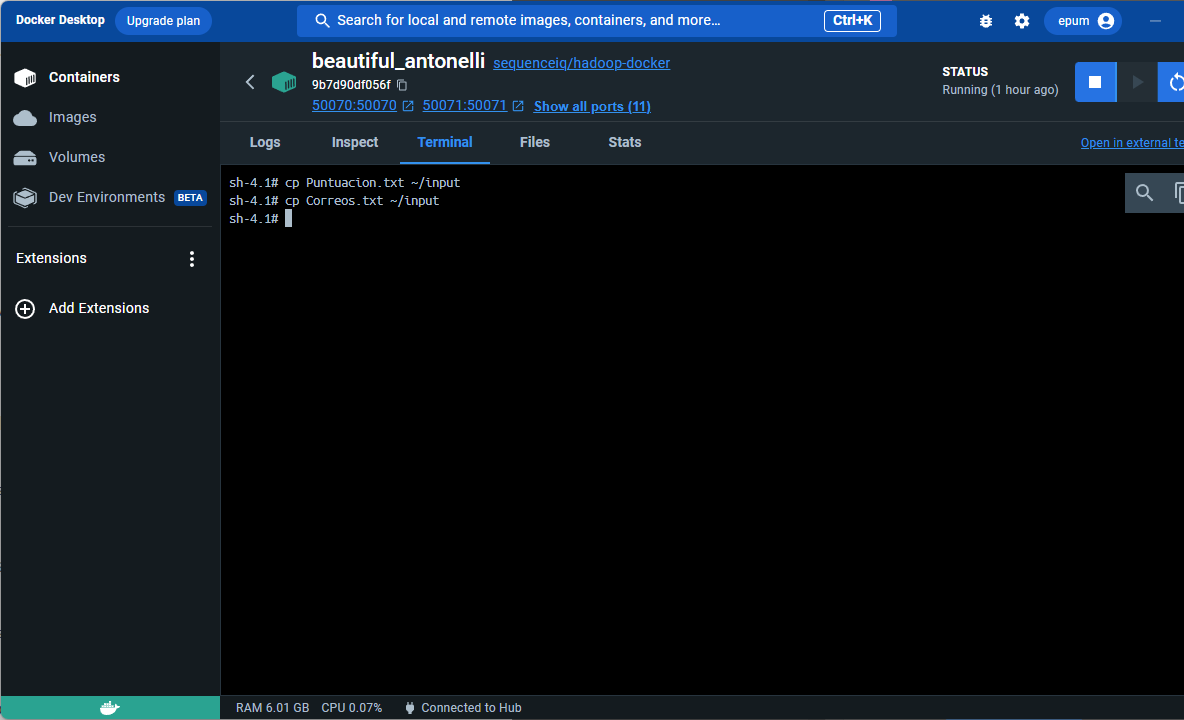


1. Ahora vamos a crear dentro de la carpeta “root” del contenedor las carpetas “input” y “output” utilizando los siguientes comandos:

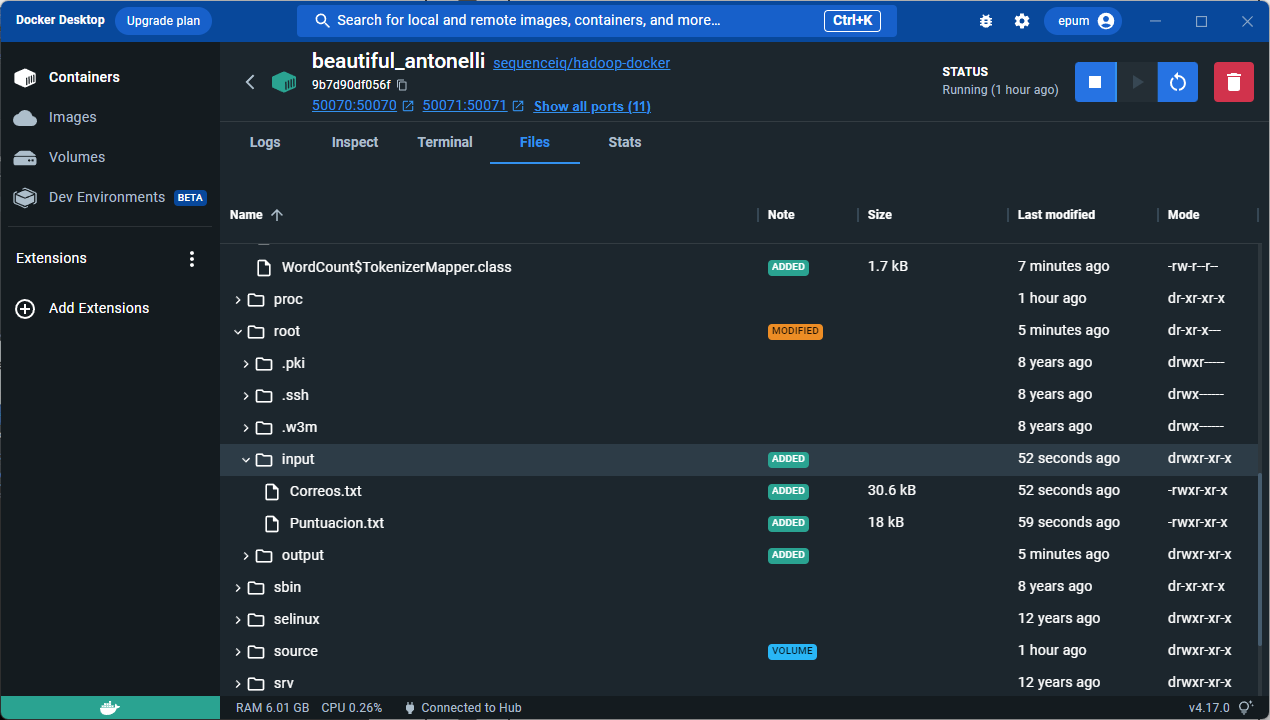




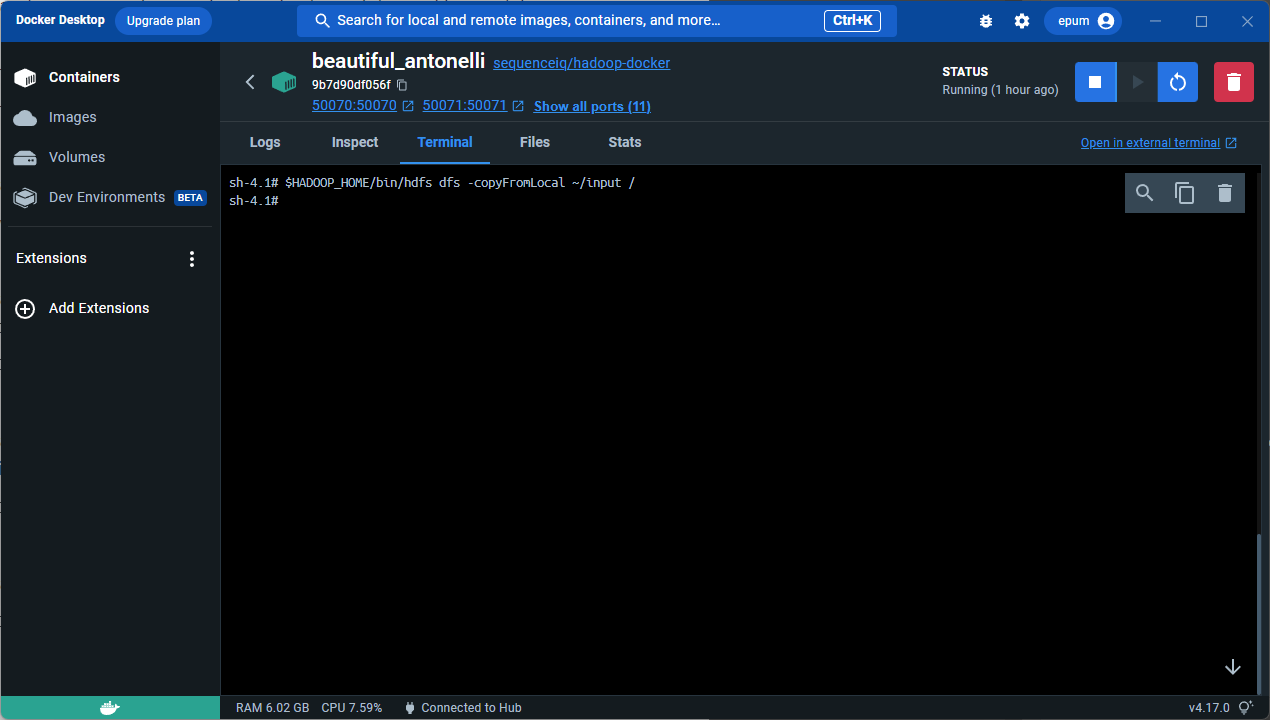
1. Copiamos los archivos de entrada dentro de la carpeta “input” utilizando los siguientes comandos:



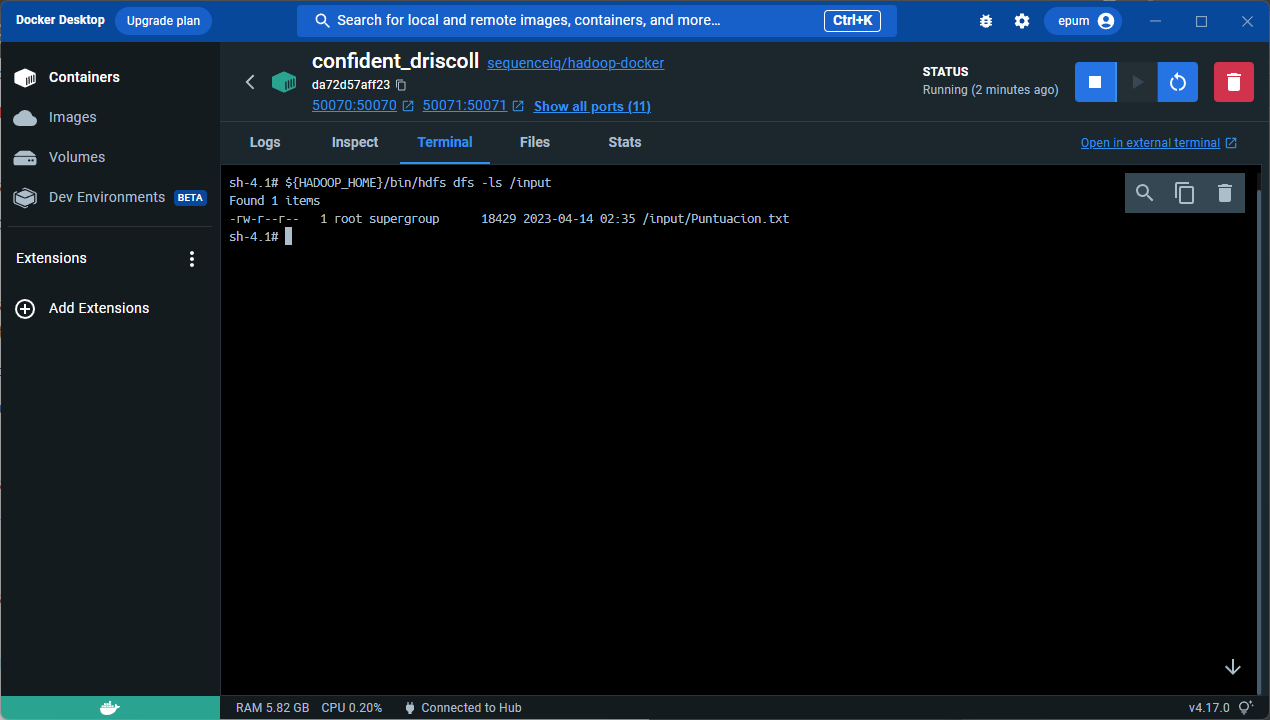
Verificamos que correctamente se hayan copiado los archivos:



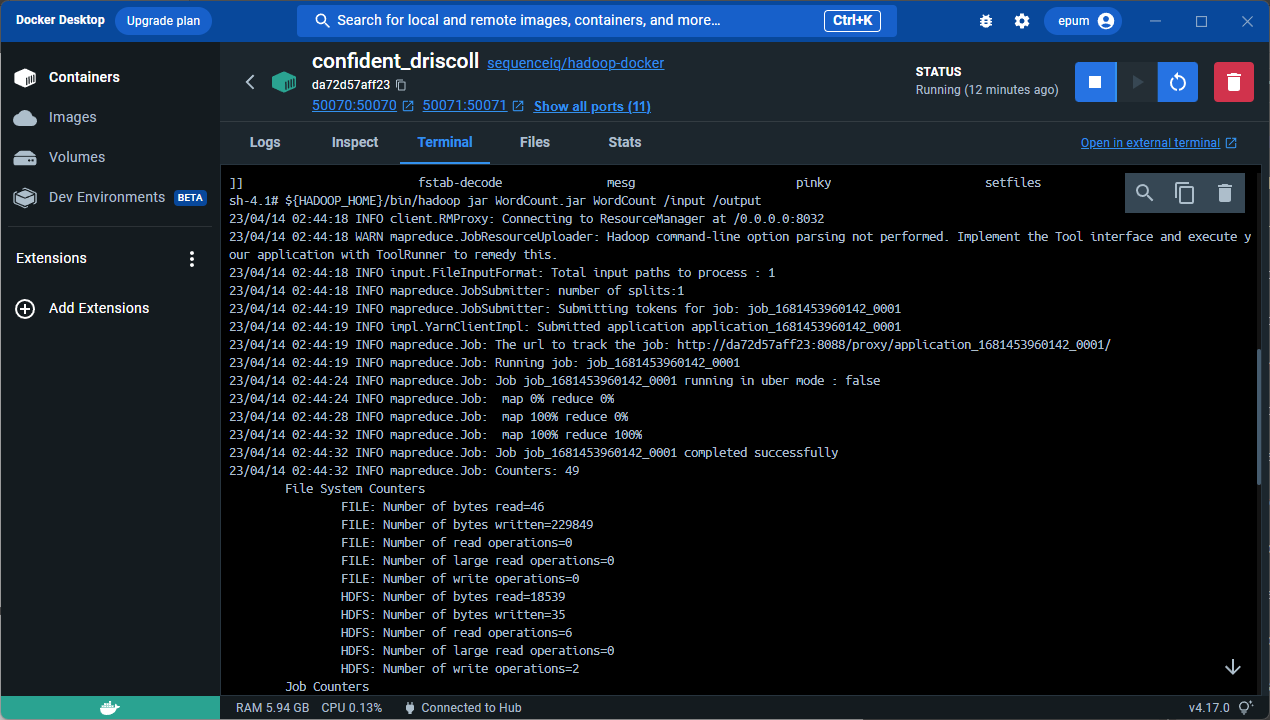
1. Ahora vamos a copiar los archivos de entrada en el sistema de archivos de Hadoop utilizando el siguiente comando:



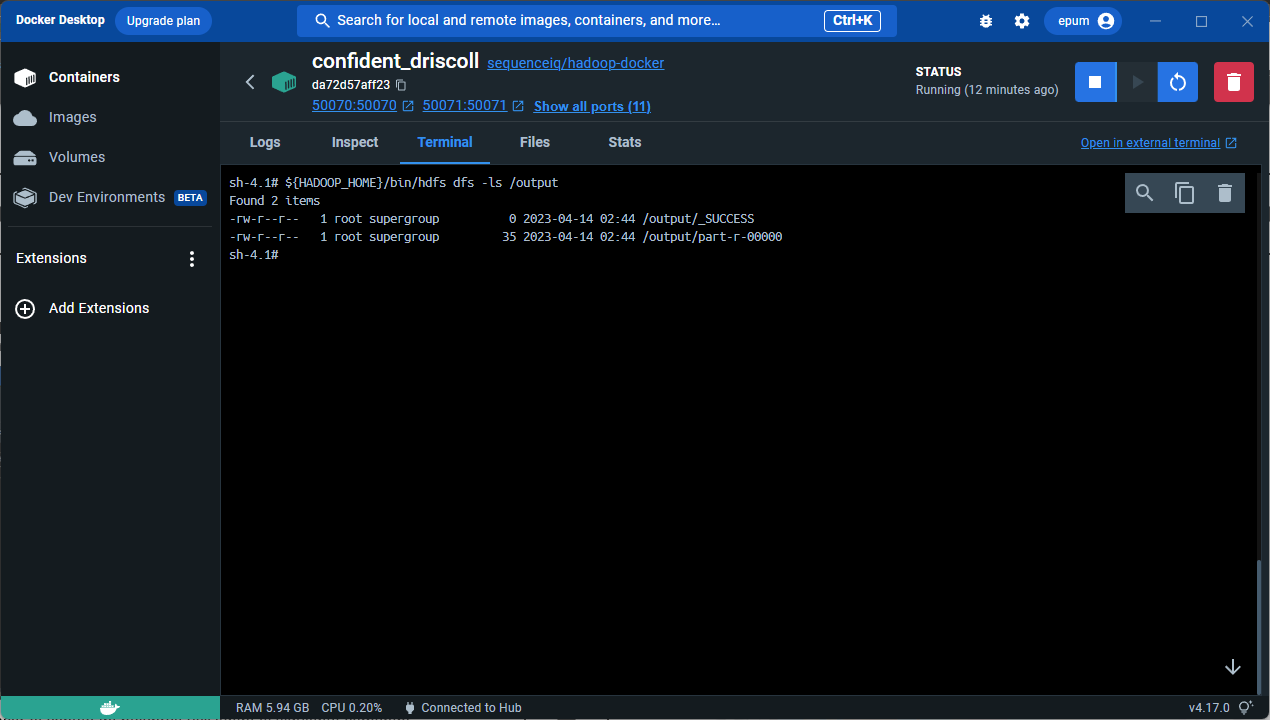
1. Verificamos que el comando anterior funcionó correctamente utilizando el siguiente comando en la terminal del contenedor:



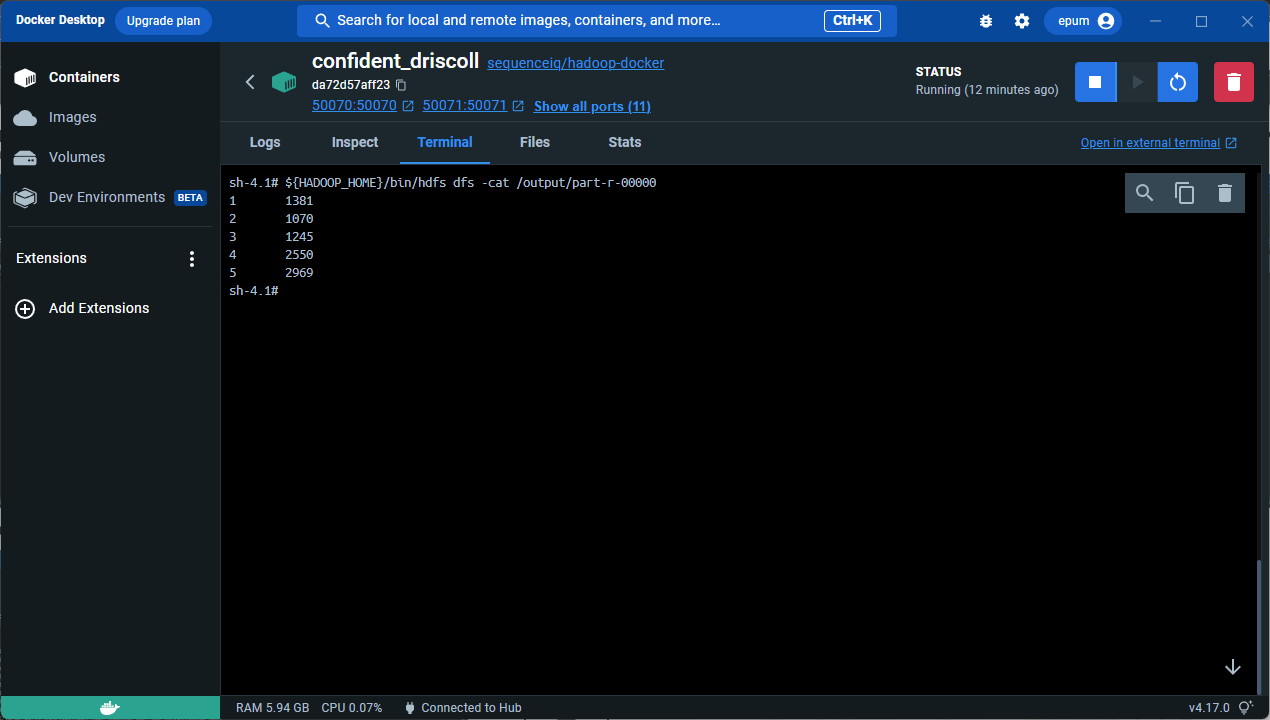
1. Ahora toca hacer el conteo de palabras utilizando el siguiente comando:



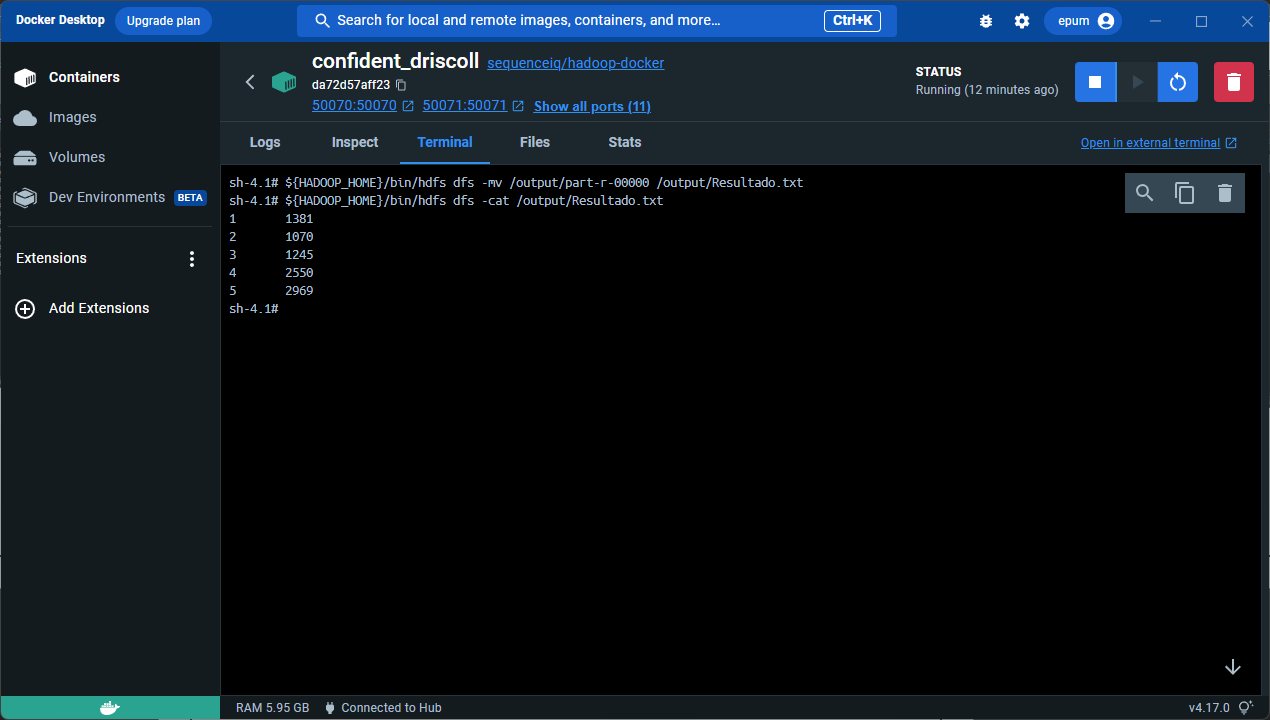
1. Para verificar que el comando anterior funcionó, ingresamos el siguiente comando:



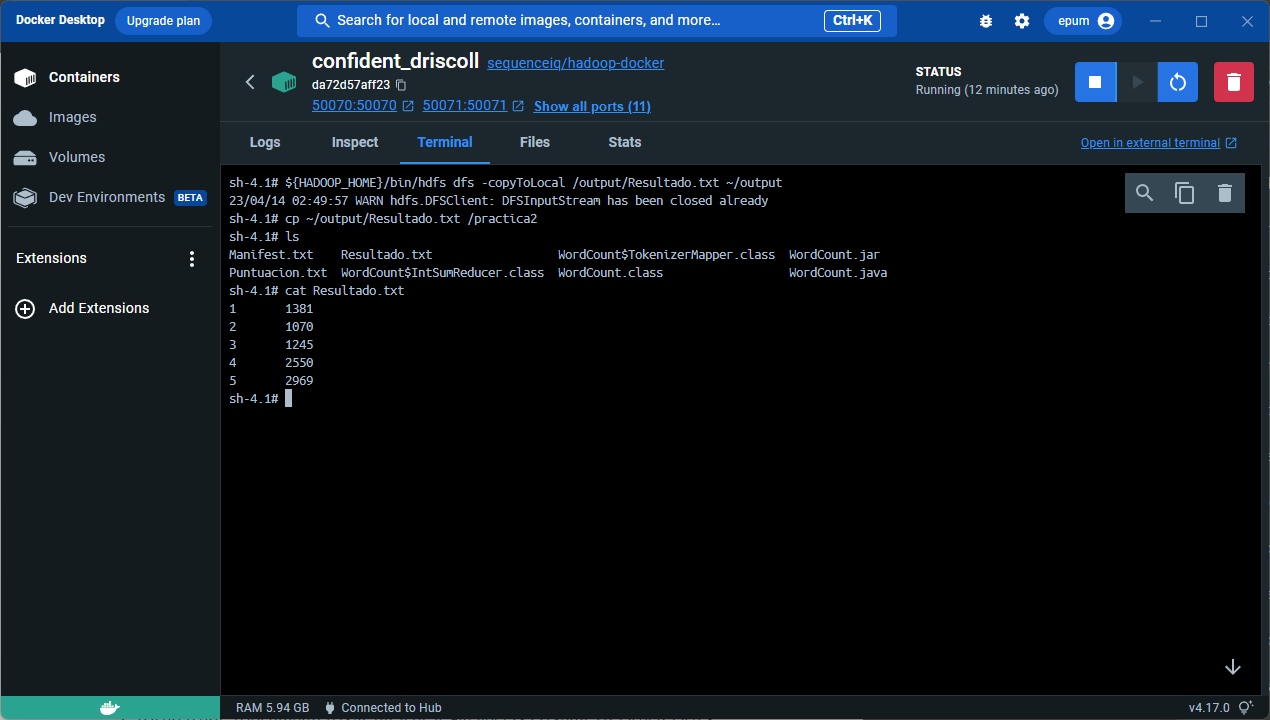
1. Para ver el resultado del conteo del archivo “Puntuacion.txt” que es con el que se está probando de primero, ingresamos el siguiente comando:



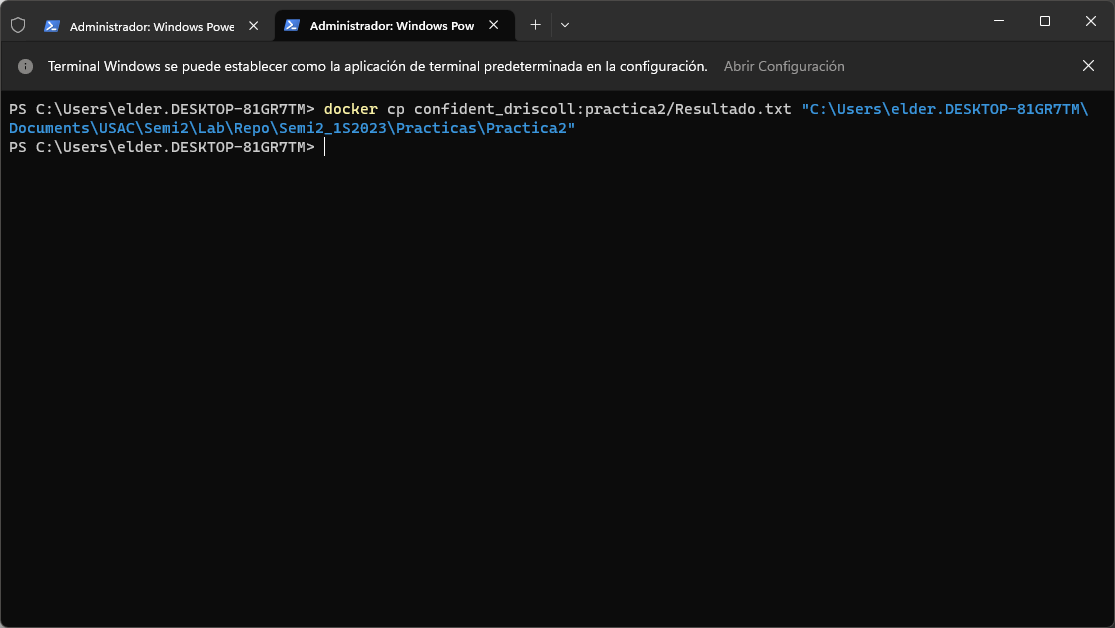
1. Ahora lo que vamos a hacer es renombrar el archivo de salida a “Resultado.txt” y además, vamos a ver que la información sea la correcta, para eso, ingresamos los siguientes comandos:



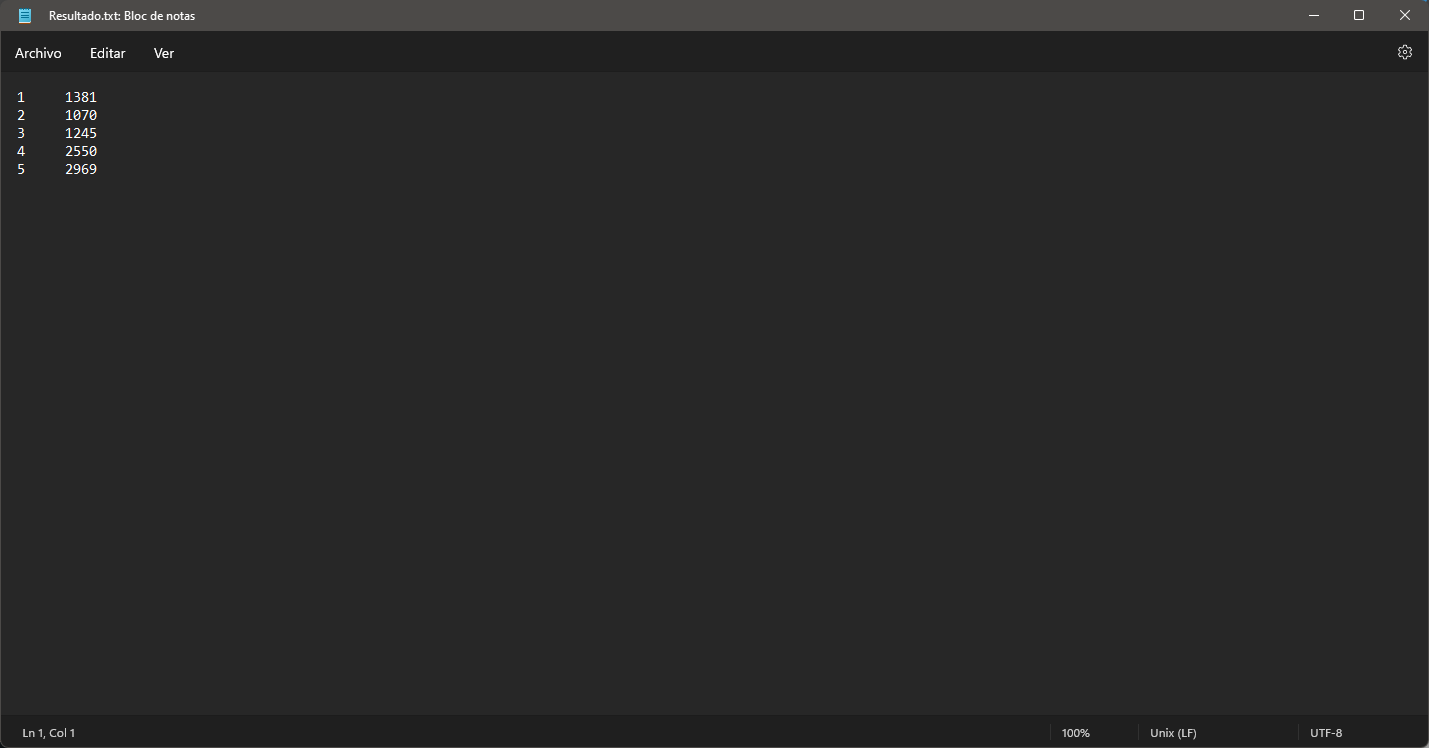
1. Ahora, vamos a copiar el archivo “Resultado.txt” del sistema de archivos de Hadoop a la carpeta que creamos al inicio llamada “practica2” utilizando los siguientes comandos:



1. Vamos a extraer el archivo “Resultado.txt” del contenedor y lo vamos a copiar a nuestra computadora utilizando la terminal de Windows/Linux, para esto, vamos a ingresar el siguiente comando:

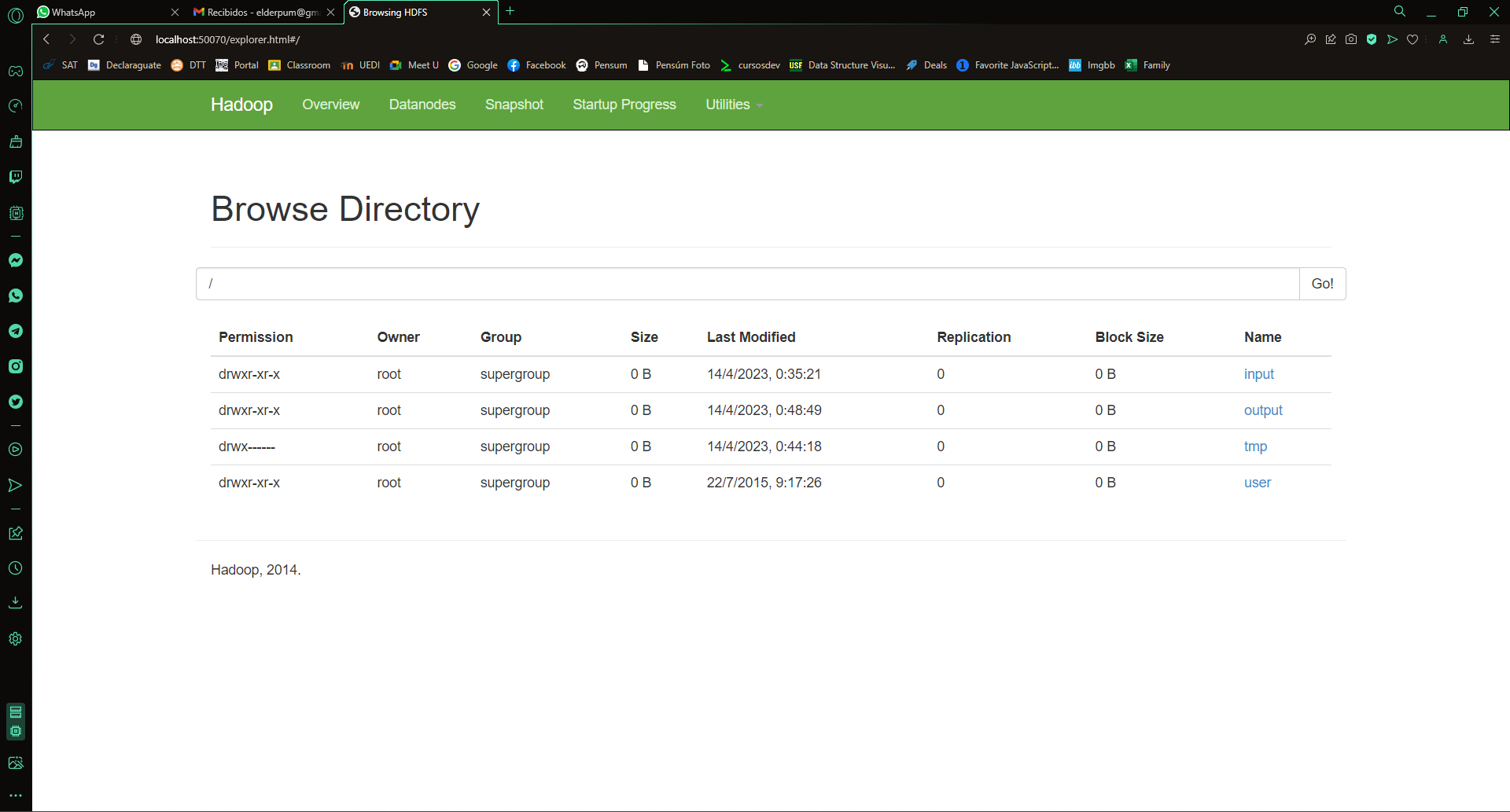


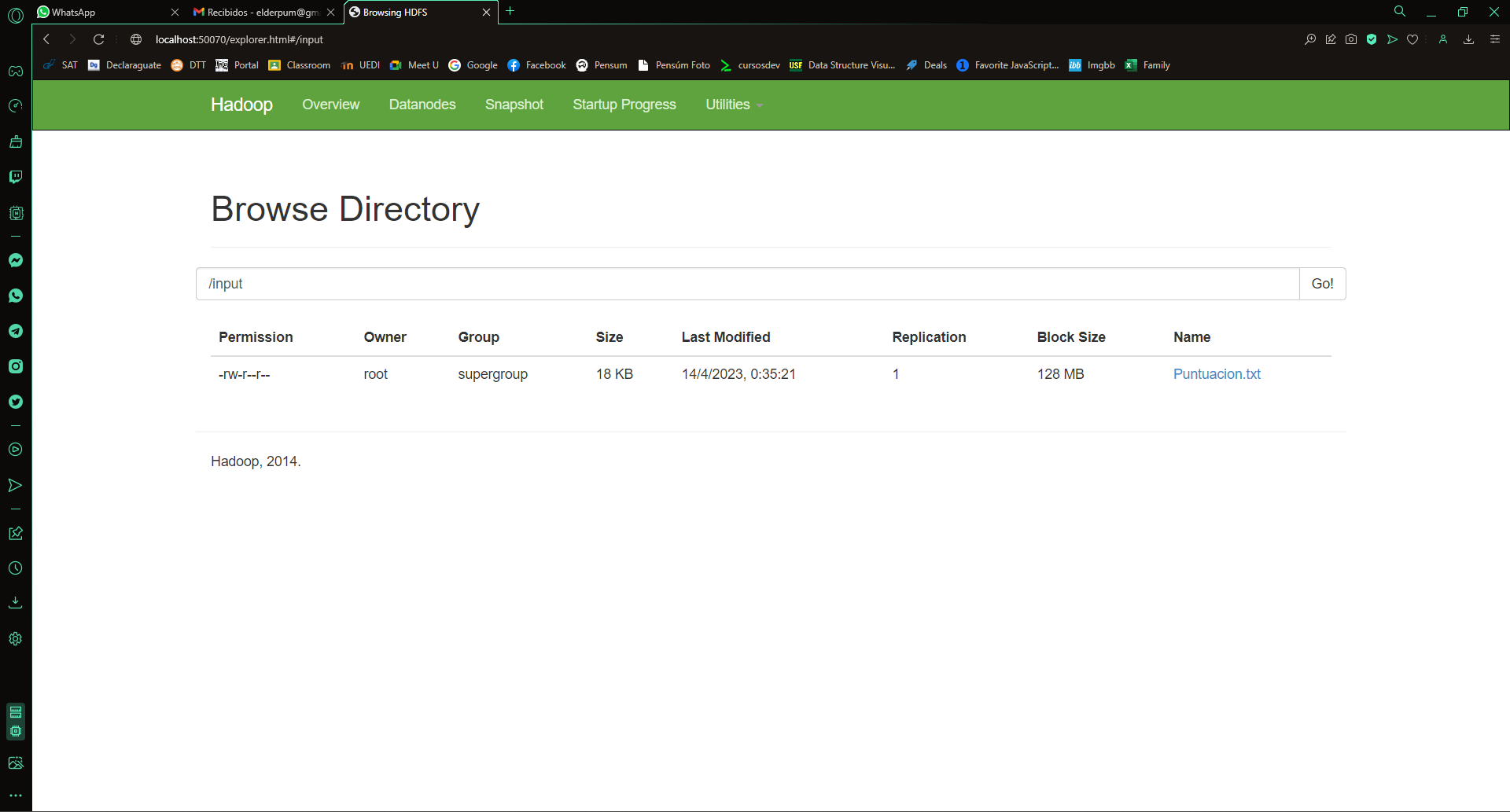
1. Verificamos en la ruta donde guardamos el archivo del Resultado:

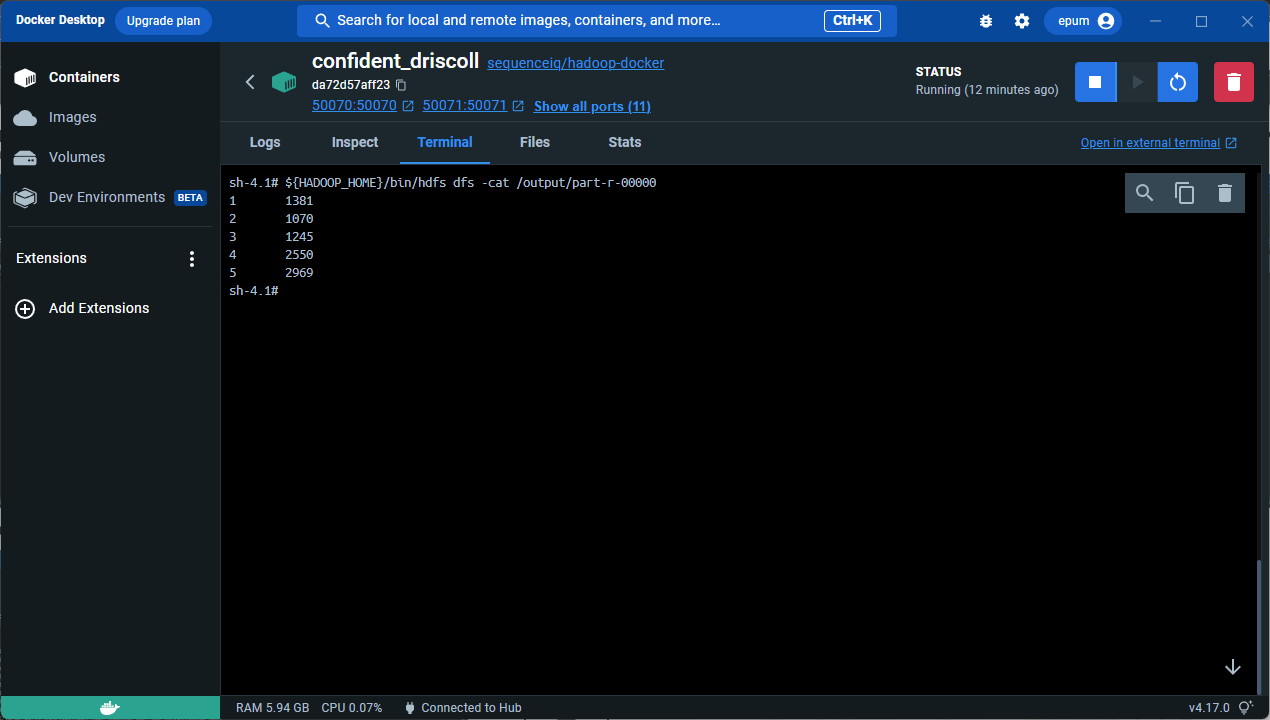


Cabe resaltar que para que todo esto funcione, se tiene que hacer 1 vez por archivo para que no se mezclen resultados y el WordCount.jar pueda funcionar correctamente y sin complicaciones, además de que se tiene que crear un nuevo contenedor por archivo a revisar.

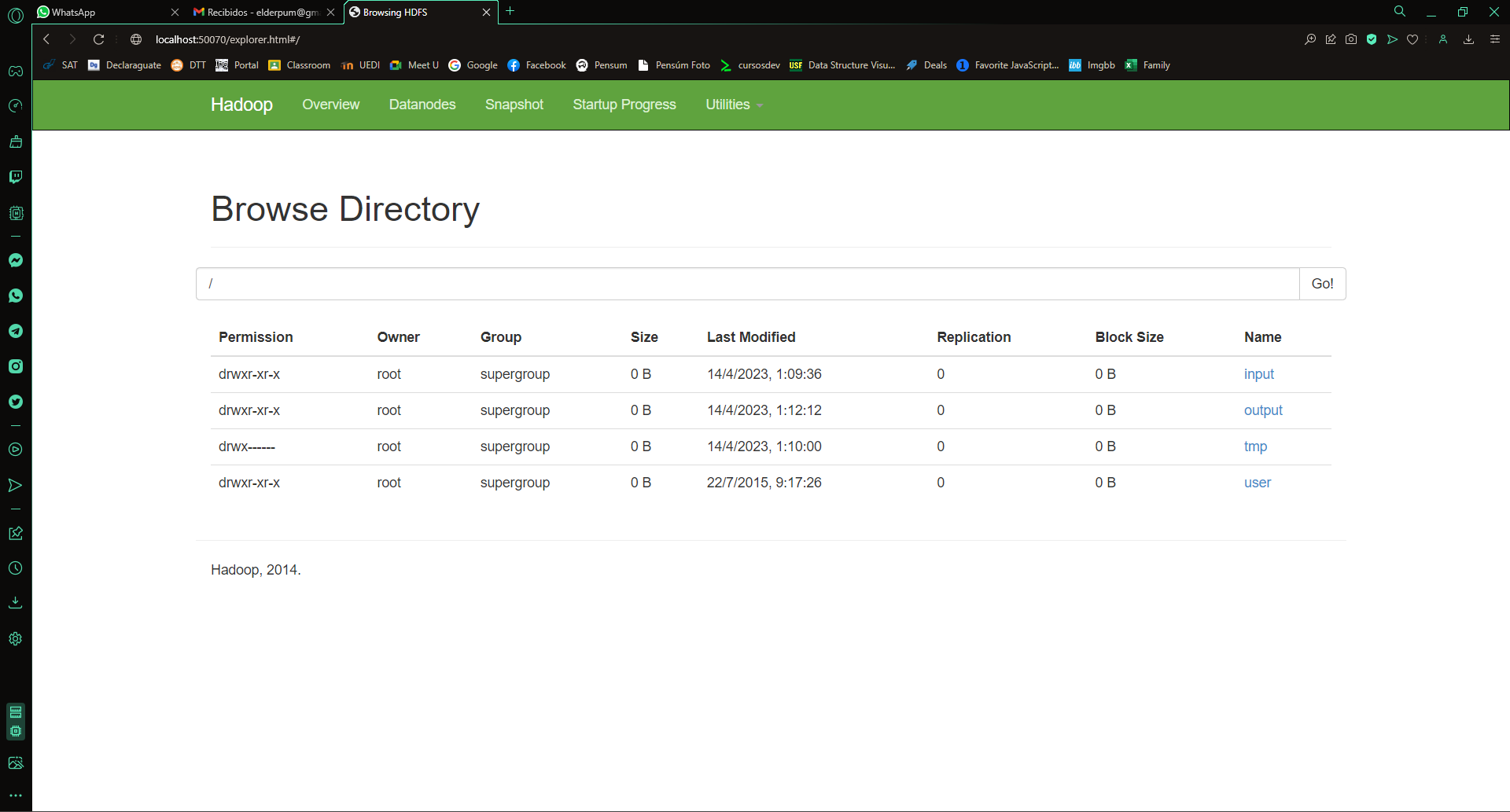
**Resultados de Puntuacion.txt**

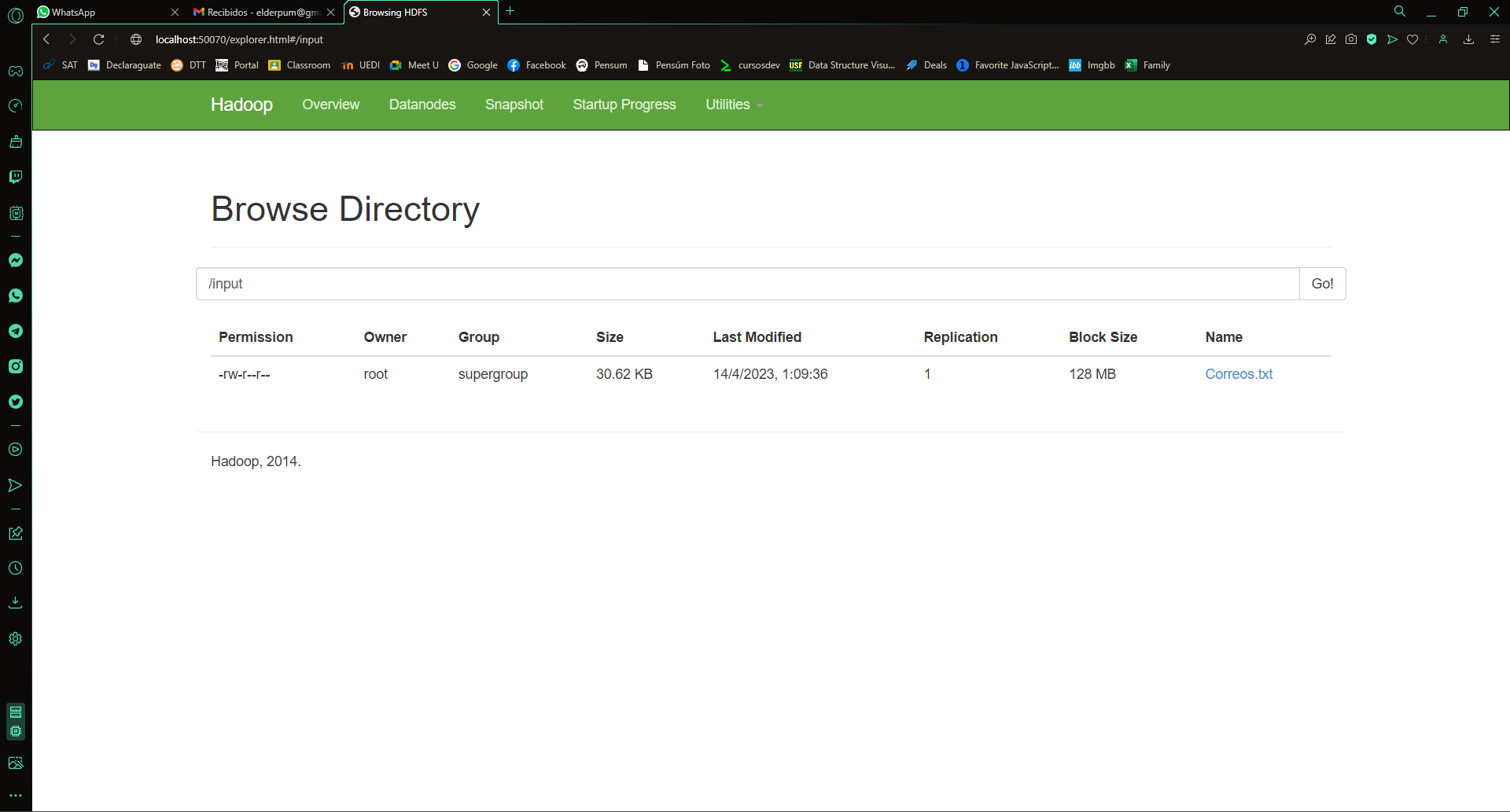


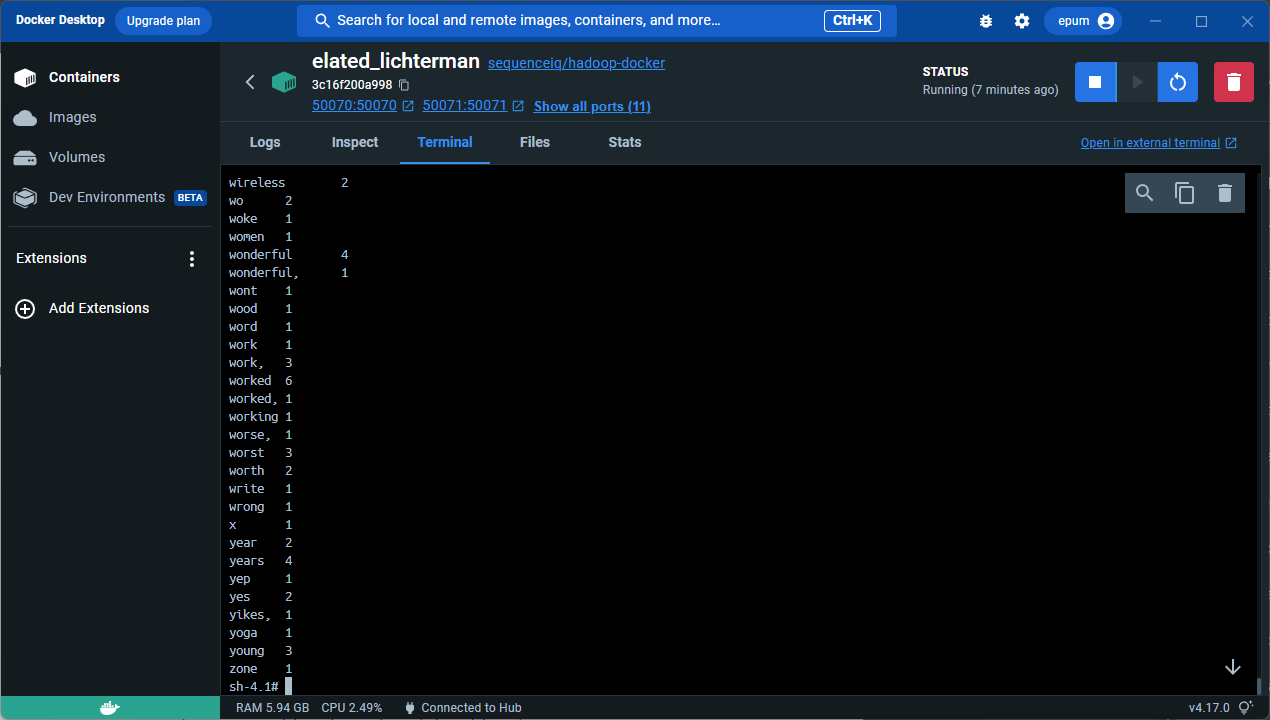




**Resultados de Correos.txt**

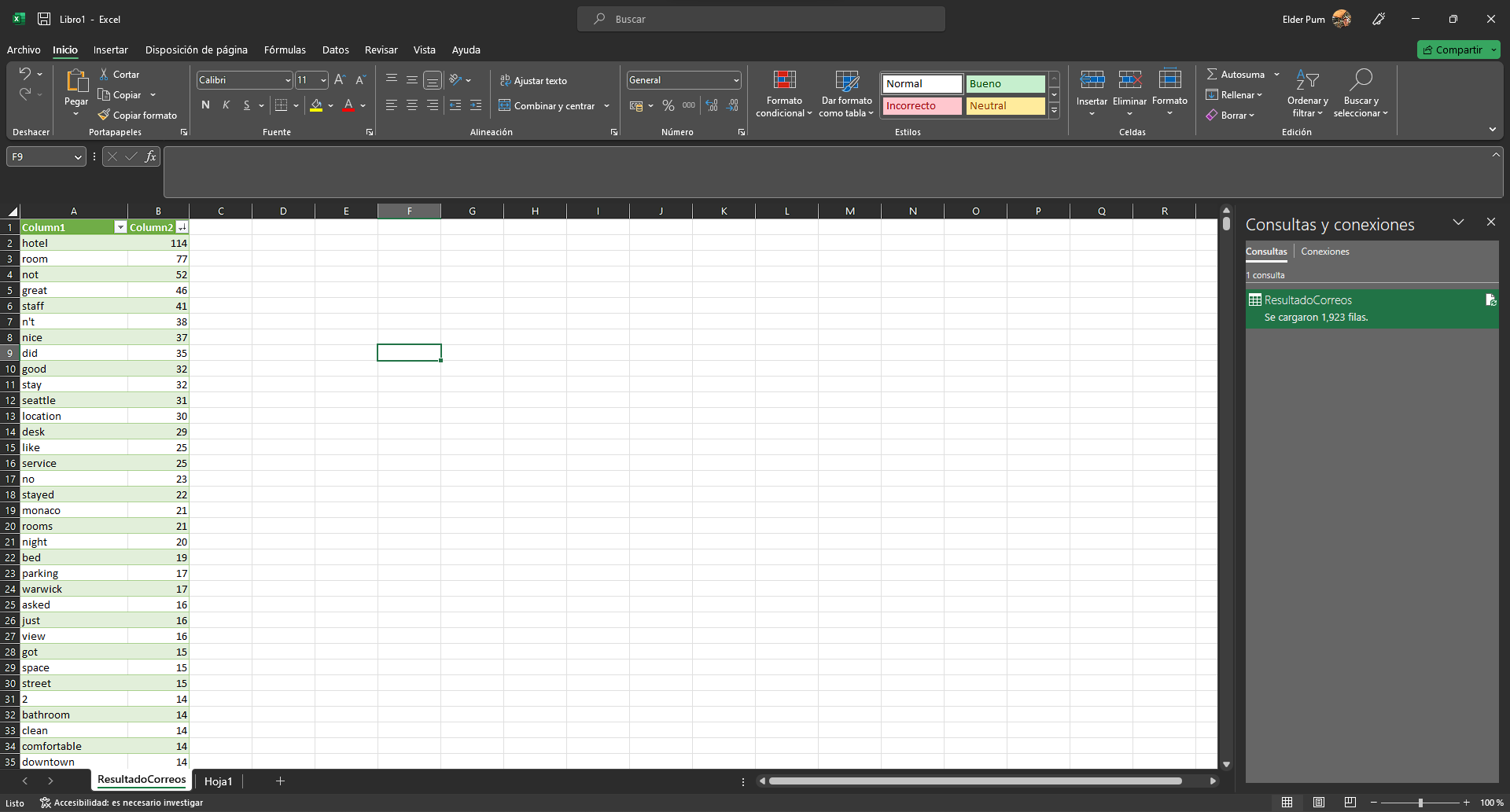






**Resultados de los datos obtenidos del archivo Correos.txt**

Después de obtener el archivo de texto en nuestra computadora, pasamos los datos a una tabla de datos en Excel y aplicándole un filtro de mayor a menor, nos queda los datos de la siguiente forma:



Básicamente las palabras claves que se repiten son:

|  |  |
| --- | --- |
| Hotel | 114 |
| Room | 77 |
| Not | 52 |
| Great | 46 |
| Staff | 41 |
| N’t | 38 |
| Nice | 37 |
| Did | 35 |
| Good | 32 |
| Stay | 32 |
| Seattle | 31 |
| Location | 30 |
| Desk | 29 |
| Like | 25 |
| Service | 25 |
| Rooms | 21 |

Si analizamos las palabras claves obtenidas del filtro de información, podemos deducir lo siguiente:

1. Principalmente que lo que más buscan las personas son hoteles. Se pueden deducir varias cosas, pero por lo general la gente lo que busca son hoteles fiables, con buen servicio en general y habitaciones cómodas para descansar después de un viaje agotador. Además, que el personal esté bien capacitado y sepa brindar un buen servicio.
2. Después podemos deducir que buscan buenas localizaciones para hospedarse, quizá buenas vistas desde el hotel.
3. Por último, las búsquedas indican que la gente busca servicios sanitarios dentro de los hoteles de óptimas condiciones, higiénicos y con una muy buena presentación.

Si ampliamos nuestras palabras claves usando Excel podemos saber que en este archivo toda la información trata sobre referencias de hoteles que cuentan con buena ubicación, parqueos óptimos, que el precio en la noche sea bueno, así como también habitaciones cómodas, aunque de precios elevados estos sean de excelente calidad, así como también el personal sea capacitado y por último, que la atención del lugar sea del agrado del cliente.

**Resultados de los datos obtenidos del archivo Puntuacion.txt**

El archivo Puntuacion.txt contiene calificaciones de distintos usuarios que varían entre las puntuaciones 1-5, donde 1 es la puntuación más baja y 5 la más alta.

Teniendo eso en cuenta y habiendo analizado el archivo de Correos, podremos darnos cuenta de que algunos factores principales que los usuarios toman en cuenta para emitir su calificación son los siguientes:

* Precio
* Disponibilidad
* Servicio
* Ubicación
* Presentación

Obviamente hay usuarios que tienen otros factores a tomar en cuenta, pero sumando esos factores junto a los principales descritos anteriormente, tenemos los siguientes resultados obtenidos del análisis del archivo de Puntuación:

|  |  |
| --- | --- |
| 1 | 1381 |
| 2 | 1070 |
| 3 | 1245 |
| 4 | 2550 |
| 5 | 2969 |

Es cierto que el puntaje más alto es también el más otorgado por los clientes, pero hay que darse cuenta que entre el punteo 1-3 juntos forman más que el punteo 4 o 5 individualmente. ¿Eso qué significa? Que hay bastantes clientes que consideran que el servicio es malo y no llenan sus expectativas, cosa alarmante para el hotel.

¿Cómo se arregla esto? Invirtiendo más en los factores principales que se detallaron anteriormente. Por ejemplo, regular el precio para lo que ofrecen, ofrecer disponibilidad 24/7 para abarcar mayor clientela, así como también capacitar al personal del hotel para ofrecer un excelente servicio de manera general e invertir en la presentación tanto externa como interna del hotel.