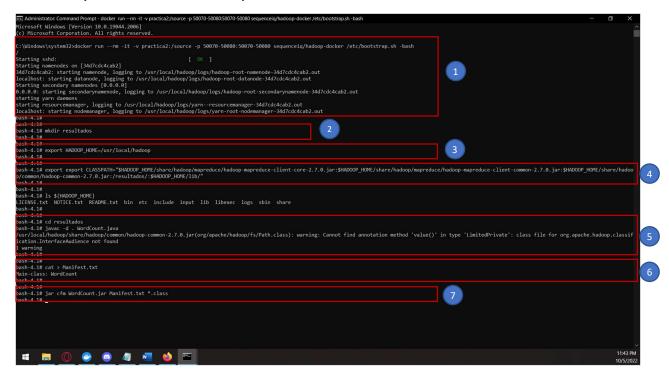
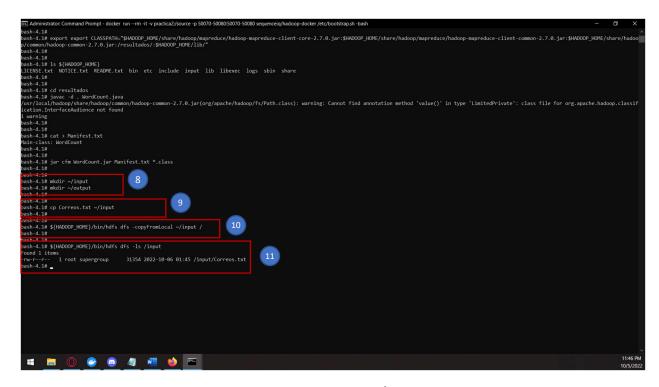
PRACTICA 2

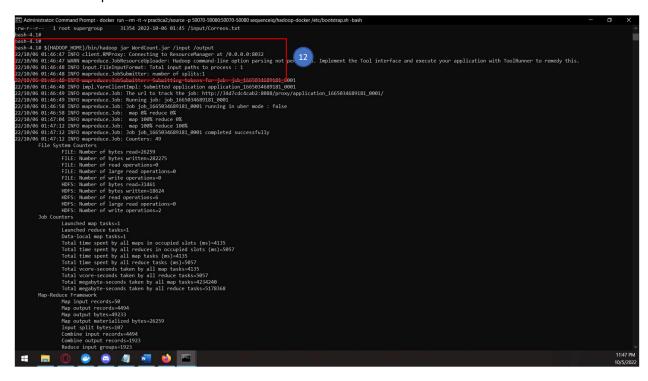
Comandos para realizar el conteo de palabras en el archivo Correos.txt



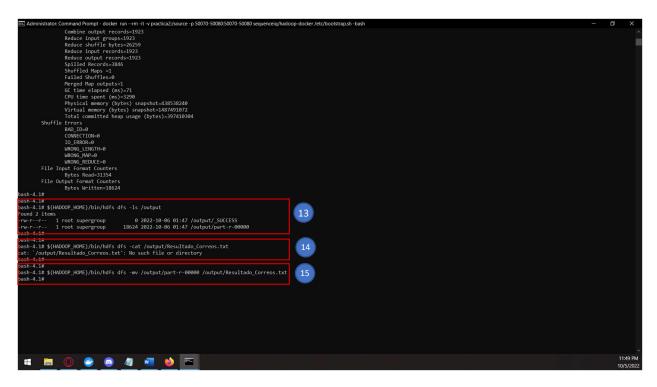
- 1. Comando utilizado para correr la imagen
- 2. Comando utilizado para crear la carpeta en la que se guardarán los archivos.
- 3. Comando utilizado para inicializar la variable HADOOP HOME
- 4. Comando utilizado para inicializar la variable CLASSPATH
- 5. Comando utilizado para compilar el archivo .java
- 6. Comando utilizado para crear el archivo Manifest.txt con el contenido "Main-class: WordCount"
- 7. Comando utilizado para crear el archivo .jar



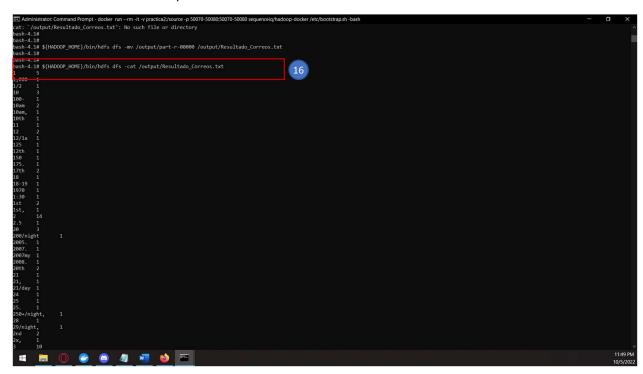
- 8. Comandos para crear las carpetas input y output en /root
- 9. Comando utilizado para copiar el archivo de entrada en la carpeta ~/input
- 10. Comando utilizado para copiar el archivo de entrada en el sistema de archivos de hadoop
- 11. Comando utilizado para verificar los archivos que fueron copiados en el sistema de archivos de hadoop



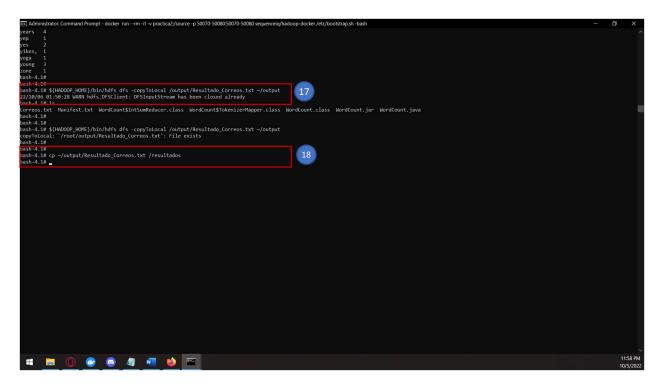
12. Comando utilizado para realizar el conteo de palabras



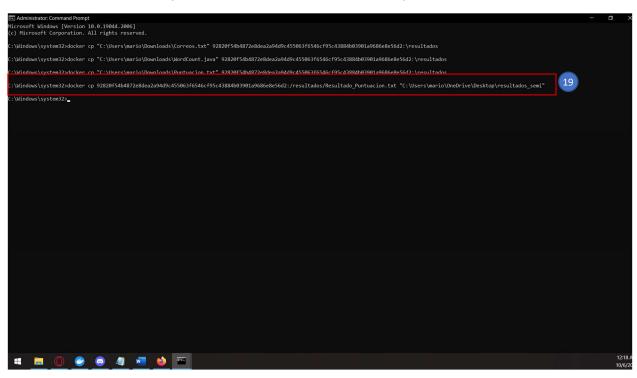
- 13. Comando utilizado para listar los archivos de salida
- 14. Comando utilizado para ver el contenido del archivo de salida
- 15. Comando utilizado para cambiar el nombre del archivo de salida



16. Comando utilizado para ver el contenido del archivo que fue renombrado

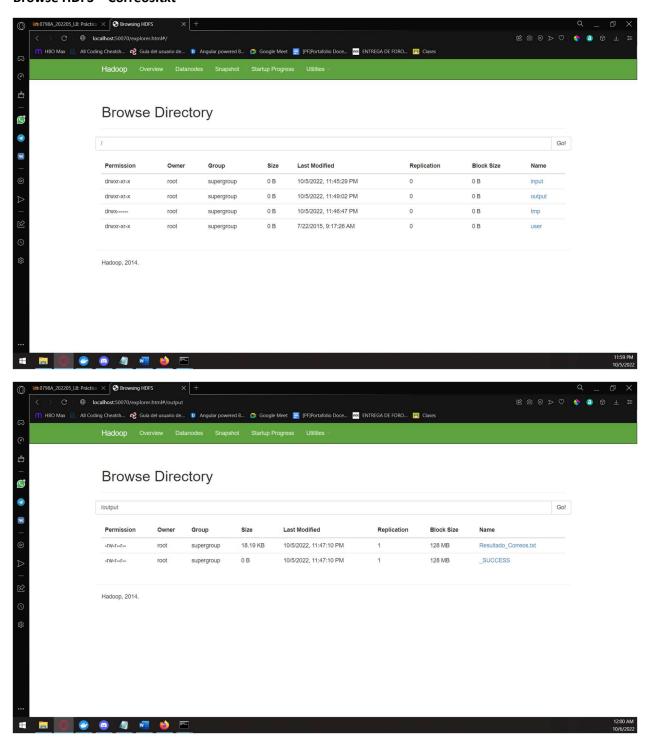


- 17. Comando utilizado para copiar el archivo de salida a la carpeta ~/output
- 18. Comando utilizado para mover el archivo de salida a la carpeta resultados del contenedor

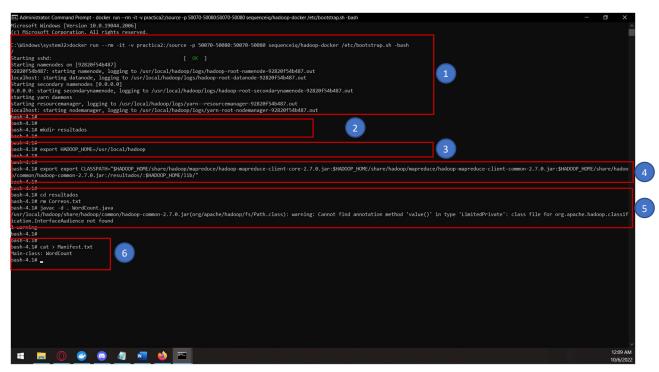


19. Comando utilizado para copiar el archivo de salida del contenedor a la PC

Browse HDFS - Correos.txt

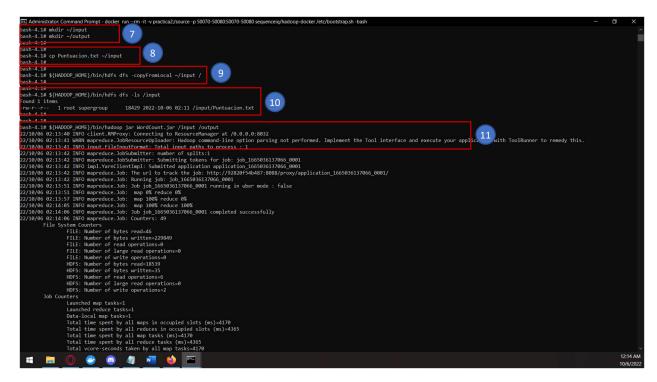


Comandos para realizar el conteo de palabras en el archivo Puntuaciones.txt

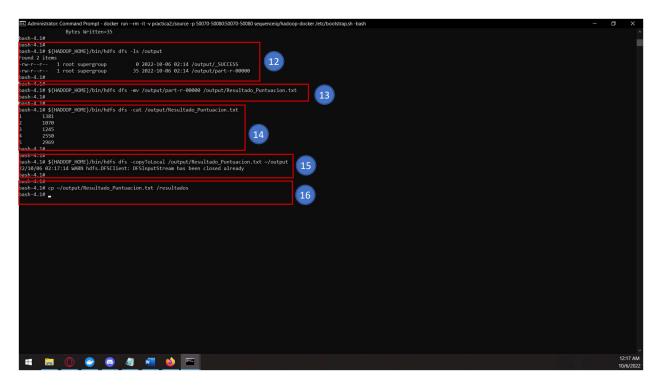


- 1. Comando utilizado para correr la imagen de secuenceiq/hadoop-docker
- 2. Comando utilizado para crear la carpeta en la que se guardarán los archivos.
- 3. Comando utilizado para inicializar la variable HADOOP HOME
- 4. Comando utilizado para inicializar la variable CLASSPATH
- 5. Comando utilizado para compilar el archivo .java
- 6. Comando utilizado para crear el archivo Manifest.txt con el contenido "Main-class: WordCount"

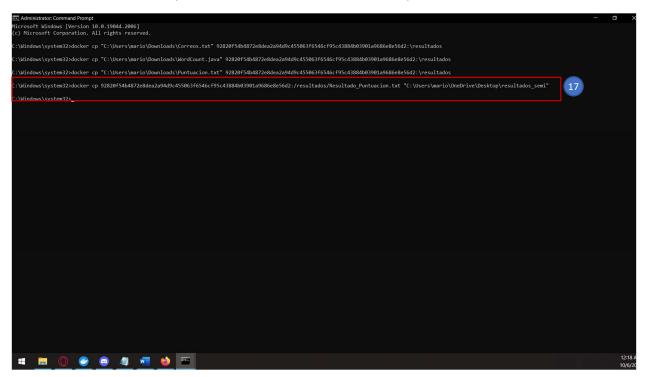
Posteriormente se utilizó el comando "jar cfm WordCount.jar Manifest.txt *.class" para crear el archivo .jar pero este no se mostró en pantalla.



- 7. Comandos para crear las carpetas input y output en /root
- 8. Comando utilizado para copiar el archivo de entrada en la carpeta ~/input
- 9. Comando utilizado para copiar el archivo de entrada en el sistema de archivos de hadoop
- 10. Comando utilizado para verificar los archivos que fueron copiados en el sistema de archivos de hadoop
- 11. Comando utilizado para realizar el conteo de palabras

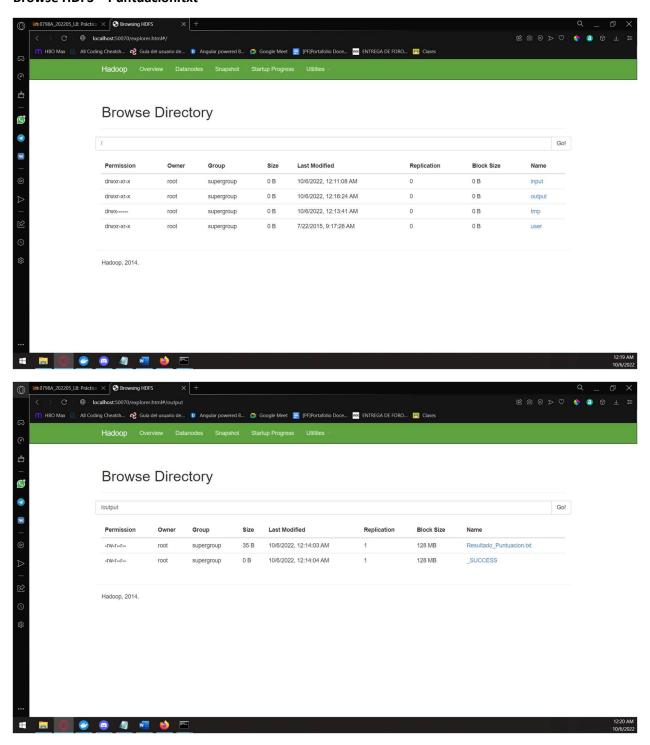


- 12. Comando utilizado para listar los archivos de salida
- 13. Comando utilizado para cambiar el nombre del archivo de salida
- 14. Comando utilizado para ver el contenido del archivo que fue renombrado
- 15. Comando utilizado para copiar el archivo de salida a la carpeta ~/output
- 16. Comando utilizado para mover el archivo de salida a la carpeta resultados del contenedor



17. Comando utilizado para copiar el archivo de salida del contenedor a la PC

Browse HDFS - Puntuacion.txt



Conclusiones de los resultados

- El documento de los resultados del archivo correos.txt muestra una gran cantidad de palabras y las veces que estas se encuentran repetidas, aunque en algunas ocasiones se puede notar que algunas palabras se repiten, pero con alguna variante como terminar en una coma, esto debido a que la herramienta realiza el conteo utilizando espacios como separador únicamente.
- En documento de resultados del archivo puntuacion.txt muestra una salida más pequeña siendo esto debido a que en el archivo solo se encuentran números del 1 al 5, que son las puntuaciones recolectadas de usuarios, también siendo un resultado más exacto que el anterior.

Uso de Hadoop en Big Data

 Hadoop resulta ser una herramienta de gran ayuda en el tema de Big Data, esto debido a las avanzadas capacidades que esta ofrece en el análisis de datos. Puede almacenar grandes cantidades de información para ser analizada y posteriormente procesada, es útil para mejorar la eficiencia de procesos y operaciones de las organizaciones, mediante esta se pueden descubrir patrones de comportamiento, y debido a su carácter flexible permite modificar el sistema de datos de las empresas u organismos.