**TAREA 1 - Unidad 2: Contar palabras del Quijote usando Hadoop y Docker.**

**Objetivo de la práctica:**

En esta tarea, deberás configurar y utilizar un clúster de Hadoop montado en Docker para contar las ocurrencias de cada palabra en el texto de **El Quijote**. El texto se puede descargar desde el siguiente enlace: [El Quijote - Gist GitHub](https://gist.github.com/jsdario/6d6c69398cb0c73111e49f1218960f79).

**Instrucciones:**

**1. Preparación del entorno:**

* Asegúrate de que tu clúster de **Hadoop** esté configurado en **Docker**.
* Modifica la configuración de tu clúster de Hadoop para que el **NameNode** tenga una carpeta compartida con tu equipo anfitrión. Esto te permitirá interactuar con los archivos de entrada y salida de manera eficiente.

**2. Subir el archivo al sistema HDFS:**

Una vez que tu clúster de Hadoop esté en funcionamiento y YARN esté activo, sube el archivo **El Quijote** al sistema de archivos distribuido de Hadoop (HDFS) usando el siguiente comando:

hdfs dfs -put el\_quijote.txt /user/hdadmin

Este comando subirá el archivo **el\_quijote.txt** a la carpeta de usuario en HDFS.

**3. Ejecutar el programa WordCount:**

Hadoop incluye varios ejemplos prediseñados para demostrar el uso de **MapReduce**. Uno de estos ejemplos es el programa **WordCount**, que cuenta las ocurrencias de cada palabra en un archivo de texto.

Para ejecutar este programa en tu clúster, usa el siguiente comando de Hadoop:

hadoop jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar wordcount /user/hdadmin/el\_quijote.txt /user/hdadmin/salidaWC

También puedes ejecutar el mismo programa usando YARN:

yarn jar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar wordcount /user/iabd/el\_quijote.txt /user/hdadmin/salidaWC

Este comando ejecutará el programa **WordCount** en el archivo **el\_quijote.txt** y guardará los resultados en la carpeta **salidaWC** en HDFS.

**4. Configurar un Bind Mount en Docker (Windows):**

Si estás usando Docker en **Windows**, puedes usar un **Bind Mount** para compartir una carpeta entre tu sistema anfitrión y el contenedor Docker. Sigue estos pasos para configurar el Bind Mount:

1. **Seleccionar o crear un directorio en Windows:**

Elige o crea una carpeta en tu sistema, por ejemplo: C:\mi\_datos.

1. **Ejecutar el contenedor con el Bind Mount:**

Usa el siguiente comando para montar la carpeta en tu contenedor Docker:

docker run -v /c/mi\_datos:/home -itubuntubash

Aquí, /c/mi\_datos es la ruta de la carpeta en Windows, y /home es la ruta dentro del contenedor Docker.

1. **Trabajar con los archivos compartidos:**

Cualquier archivo colocado en **C:\mi\_datos** estará disponible dentro del contenedor en **/home**, y viceversa. Esto te permitirá trabajar con los archivos del clúster de manera más eficiente.

**Entrega del trabajo:**

El resultado final debe entregarse en formato PDF, y debe incluir un manual paso a paso detallado que explique cómo se ha realizado la configuración y ejecución de la práctica. Es importante que el manual esté estructurado de forma clara y contenga las capturas de pantalla, diagramas y otros materiales multimedia que consideres necesarios para mejorar la comprensión.

Asegúrate de que el manual cubra los siguientes aspectos:

1. Configuración de Hadoop en Docker.
2. Proceso para subir el archivo a HDFS.
3. Ejecución de MapReduce utilizando el ejemplo de WordCount.
4. Resultados obtenidos y cómo se han interpretado.
5. Posibles errores y cómo los has resuelto.

El uso de material multimedia como videos cortos, capturas de pantalla y diagramas es altamente recomendable para facilitar la comprensión de los pasos.

**Consideraciones importantes:**

* **Rutas en Docker para Windows**: En Docker para Windows, las rutas del sistema de archivos se mapean de una manera específica (por ejemplo, C:\ se mapea como /c/).
* **Permisos de archivos**: Los archivos creados dentro del contenedor pueden tener permisos y propietarios diferentes a los esperados en Windows, lo que puede afectar ciertos procesos.
* **Rendimiento**: El uso de BindMounts puede afectar el rendimiento en operaciones intensivas de entrada/salida (I/O).

**Resumen de comandos importantes:**

1. **Subir el archivo al HDFS:**

hdfsdfs -put el\_quijote.txt /user/hdadmin

1. **Ejecutar WordCount en Hadoop:**

hadoopjar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar wordcount /user/hdamin/el\_quijote.txt /user/hdamin/salidaWC

1. **Ejecutar WordCount en YARN:**

yarnjar $HADOOP\_HOME/share/hadoop/mapreduce/hadoop-mapreduce-examples-3.3.6.jar wordcount /user/hdamin/el\_quijote.txt /user/hdamin/salidaWC

1. **Configurar Bind Mount en Docker:**

docker run -v /c/mi\_datos:/home -itubuntubash