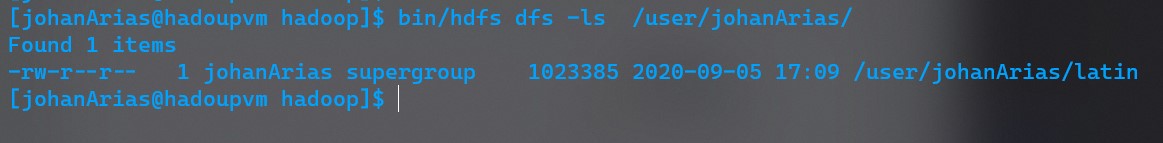
**Laboratorio 5 – Hadoop Quick Start**

**Screenshots**

**Parte a)**



**FIG 1: Archivo cargado al directorio de hdfs**

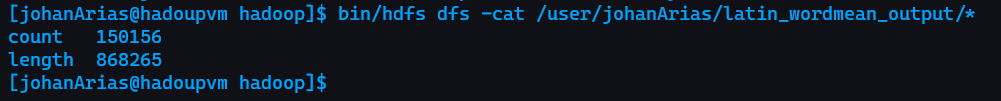
**(bin/hdfs dfs -put latin.txt latin)**

**Parte b)**



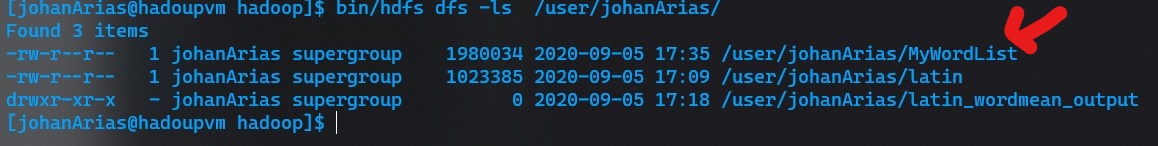
**FIG 2 : Al ejecutar el jar**

**(bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.9.2.jar wordmean <inputFile> <outputFile>)**

****

**FIG 3 : Conteo de palabras y longitud calculadas exitosamente**

**Parte c)**



**FIG 4: Cargando mi propio archivo de palabras a hdfs**

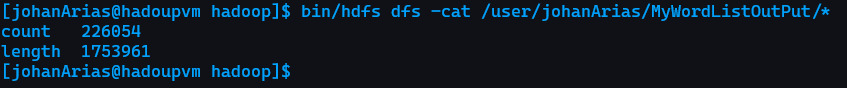
**(bin/hdfs dfs -put**

**directory-list-2.3-medium.txt MyWordlist)**

****

**FIG 5: Calculando la media con mi propia lista de palabras**

**(bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.9.2.jar wordmean MyWordList MyWordListOutPut)**

****

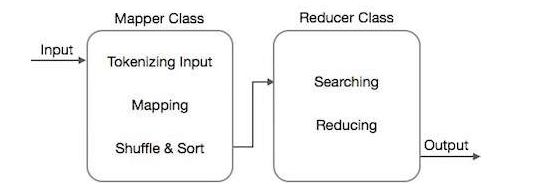
**FIG 6: Verificando que el conteo y la longitud se hayan calculado correctamente**

**(bin/hdfs dfs -cat /user/johanArias/MyWordListOutPut/\*)**

**Parte d) Investigación**

El algoritmo MapReduce contiene dos tareas importantes, Mapear y Reducir. La tarea de mapear es realizada por medio de la clase **“Mapper”** y la tarea de reducir la realiza la clase **“Reducer”**

La clase Mapper toma la información, la muestra, la mapea y la clasifica. La salida de la clase Mapper se utiliza como entrada por la clase Reducer, que a su vez busca los pares coincidentes y los reduce.



MapReduce implementa varios algoritmos matemáticos para dividir una tarea en pequeñas partes y asignarlas a múltiples sistemas. En términos técnicos, el algoritmo MapReduce ayuda a enviar las tareas de Map & Reduce a los servidores apropiados en un clúster.

El algoritmo de MapReduce tiene 3 fases:

1. Map Function.
2. Suffle Function
3. Reduce Function

**1. Map Function**

Este es el primer paso del Algoritmo MapReduce. Toma los conjuntos de datos y los distribuye en subtareas más pequeñas. Esto se hace más adelante en dos pasos, la división y el mapeo. La división toma el conjunto de datos de entrada y divide el conjunto de datos mientras que el mapeo toma esos subconjuntos de datos y realiza la acción requerida. La salida de esta función es un par clave-valor.

2**. Shuffle Function**

Esto también se conoce como función de combinación e incluye la fusión y la clasificación. La fusión combina todos los pares clave-valor. Todos ellos tendrán las mismas claves. La clasificación toma la información del paso de fusión y ordena todos los pares clave-valor haciendo uso de las claves. Este paso también regresará a los pares clave-valor. La salida será ordenada.

3**.Reduce Function**

Este es el último paso de este algoritmo. Toma los pares clave-valor del shuffle y reduce la operación.

