Laboratorio 5 - Hadoop Quick Start

Michael Sebastian Preciado Garzon

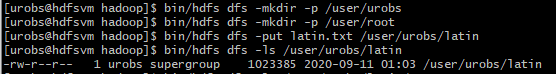
PENU 2020-2

Escuela Colombiana de ingeniería Julio Garavito

# Objetivo

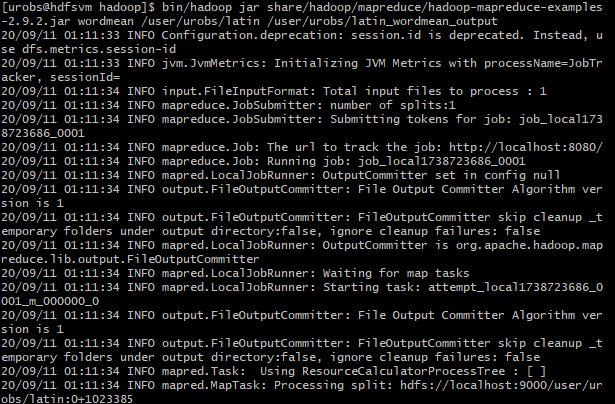
Conocer de manera introductoria el despliegue de un clúster Hadoop/HDFS para ejecutar un programa que realiza el conteo de palabras usando el algoritmo de Map Reduce.

## Subir el archivo

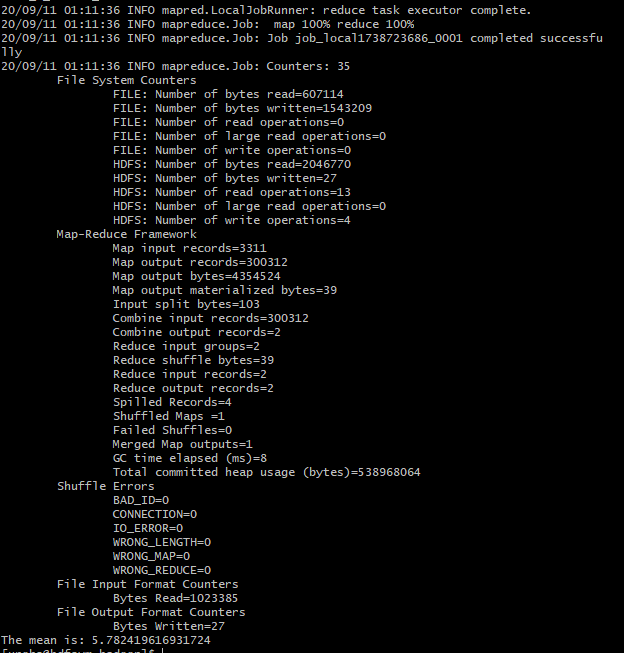


## Conteo de palabras

El inicio del comando



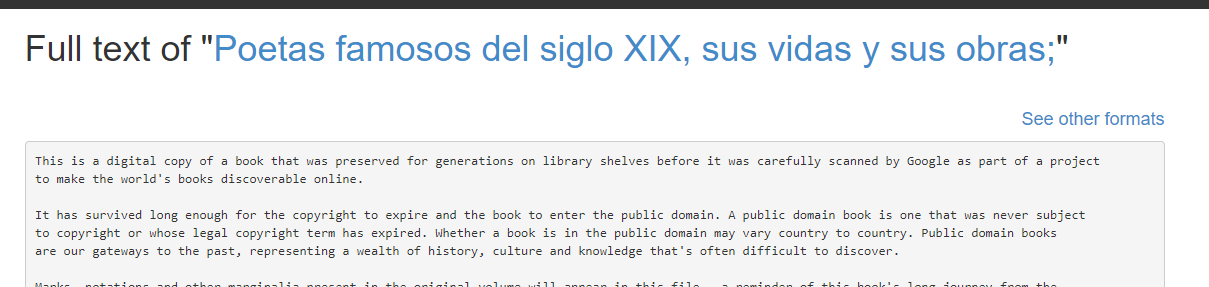
El fin del comando



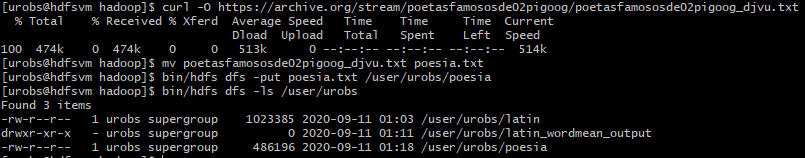
El resultado

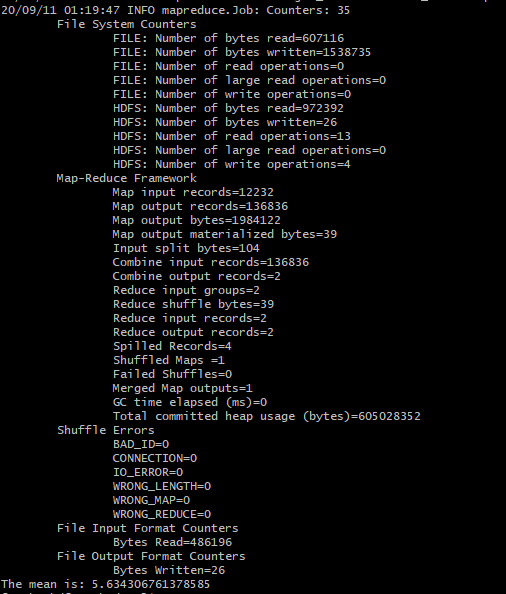
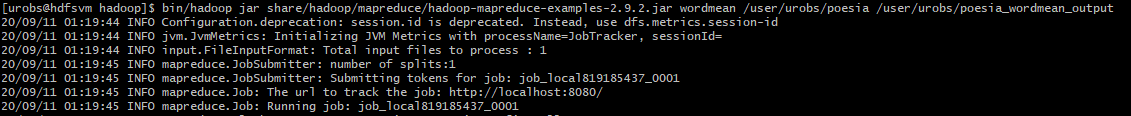


## El otro archivo



Subir el archivo

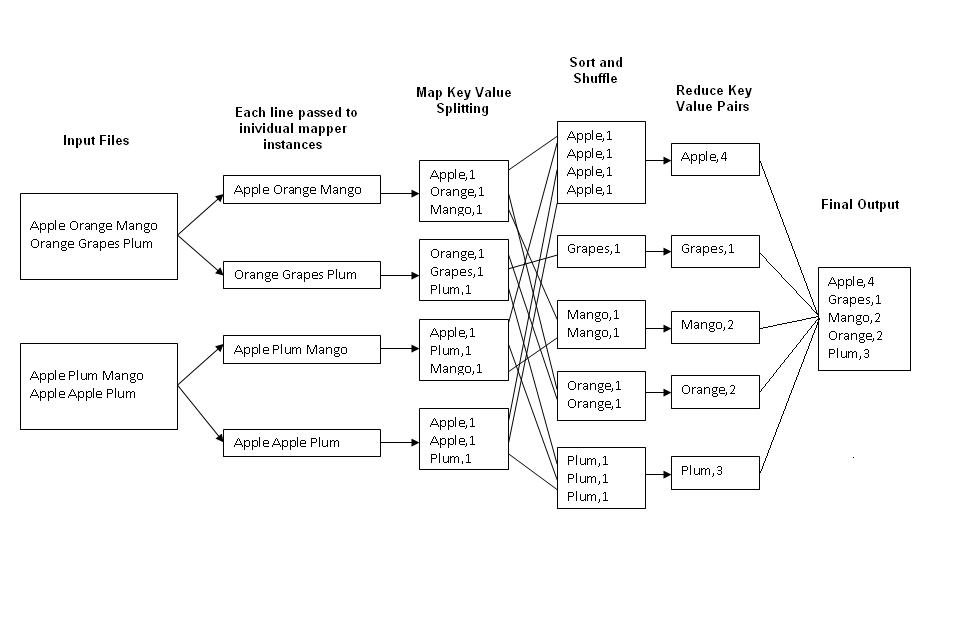


El conteo

El resultado



Explicación del mapreduce en el caso de WordCount



El algoritmo MapReduce usa diferentes funciones de mapeo de llave valor, se ejecutan en paralelo en la primera fase se mapea cada línea a un conjunto de llave valor de <palabra,1> luego se unen el proceso del resto de funcionen y se agrupan y ordenan, para luego ser fácilmente procesado por el reduce para contar los 1 para luego ser presentado en el output con la suma hecha