**LAPORAN**

**RENCANA TUGAS MANDIRI (RTM) Ke-3**

**MATA KULIAH BIG DATA**

**“MAPREDUCE PROGRAM WORD COUNT PADA HADOOP DAN PYTHON”**



**DISUSUN OLEH:**

Chelsea Ayu Adhigiadany ( 21083010028 )

**DOSEN PENGAMPU:**

Tresna Maulana Fahrudin S.ST., M.T. (NIP. 199305012022031007)

**PROGRAM STUDI SAINS DATA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR**

**2022**

# **Daftar Isi**

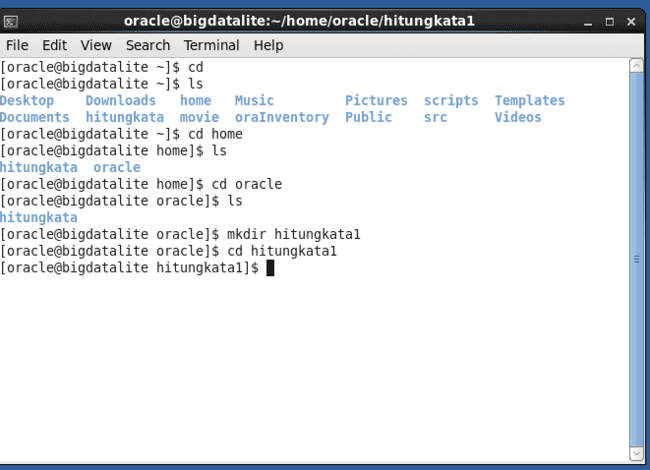
[**Daftar Isi** 2](#_Toc129551636)

[**MAPREDUCE PADA HADOOP** 3](#_Toc129551637)

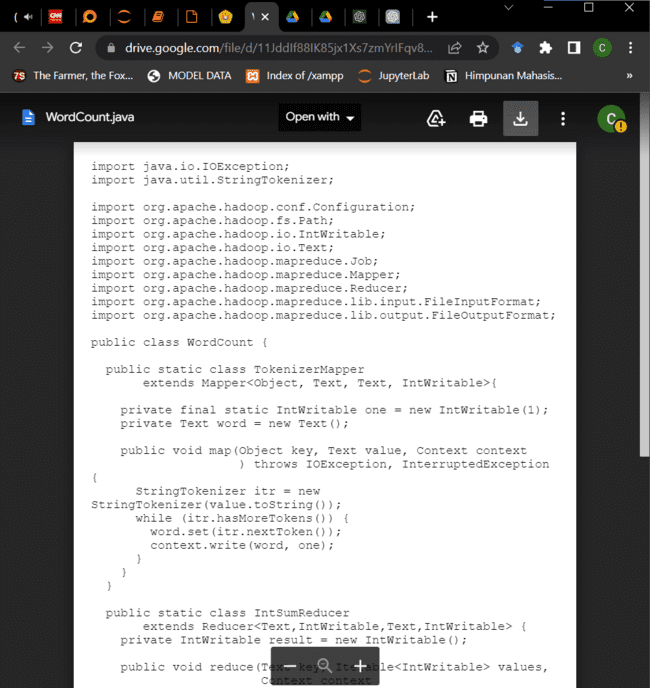
[**MAPREDUCE PADA PYTHON** 12](#_Toc129551638)

# **MAPREDUCE WORD COUNT PADA HADOOP**

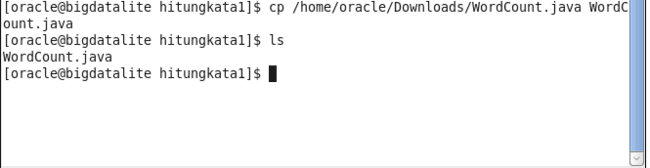
1. Masuk ke direktori home, masuk lagi ke dalam direktori oracle. Buat direktori baru bernama hitungkata1.



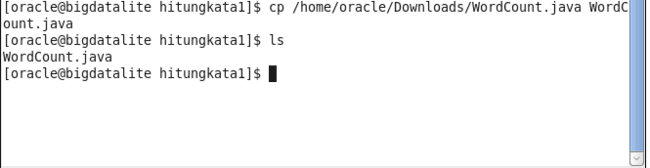
1. Download program WordCount.java



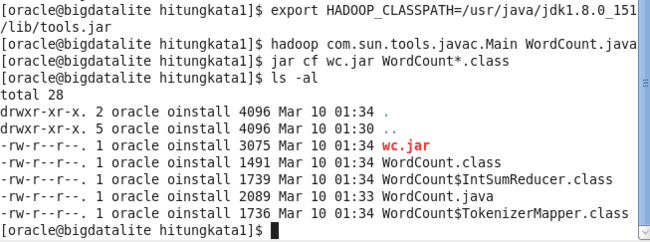
1. Copy WordCount.java pada direktori hitungkata1



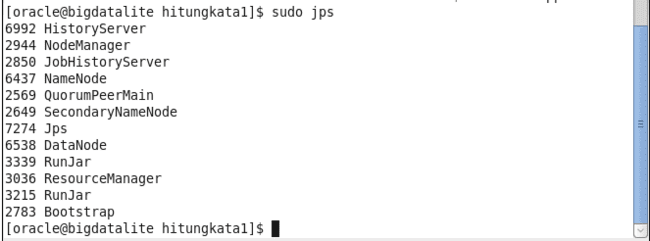
1. Cek apakah sudah ada dalam direktori hitungkata1



1. Compile program java dan membuat JAR file



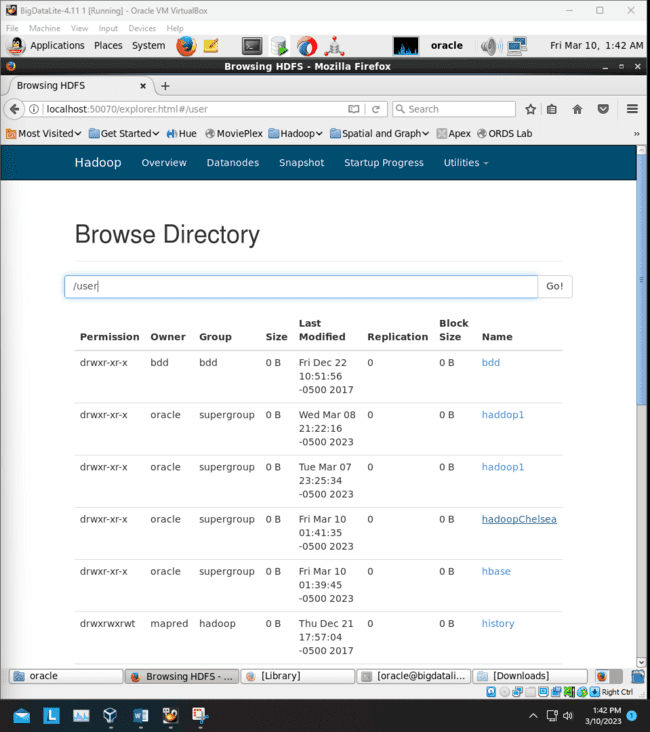
1. Cek melalui sudo jps untuk memastikan apakah Hadoop sudah berjalan dengan baik



1. Membuat direktori di HDFS



1. Periksa apakah direktori sudah ada di HDFS melalui Browsing HDFS pada link localhost:50070/explorer.html#/user

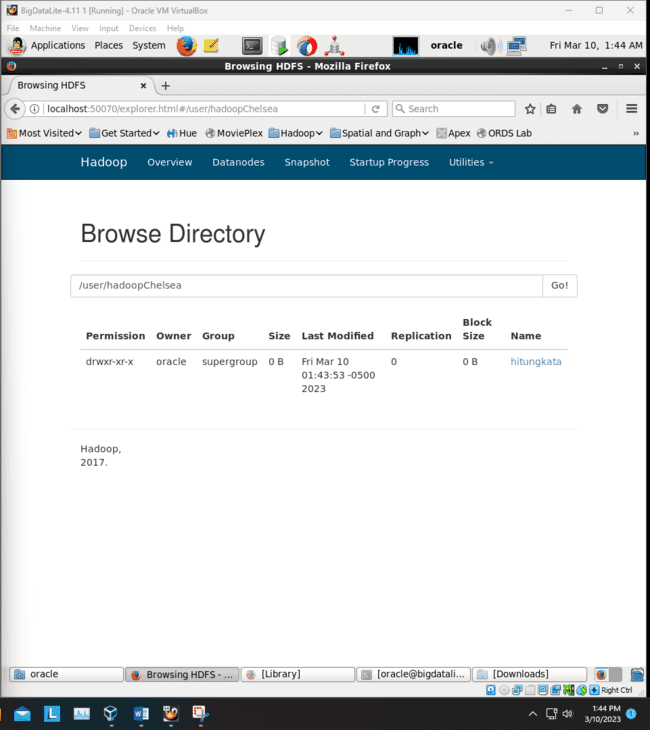


1. Buat direktori baru di Hadoop untuk menyimpan program WordCount beserta input kata yang akan diproses

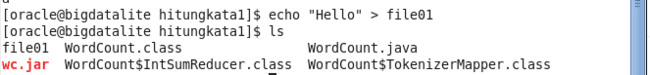


1. Buat direktori hitungkata di Hadoop



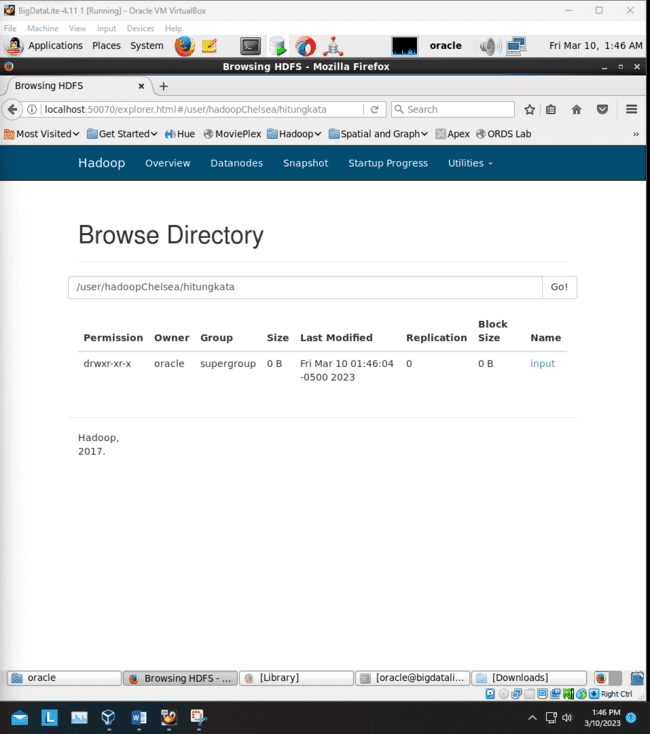


1. Buat file input yang nantinya akan dilakukan proses WordCount



1. Buat direktori input di Hadoop

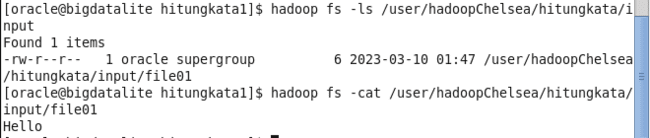


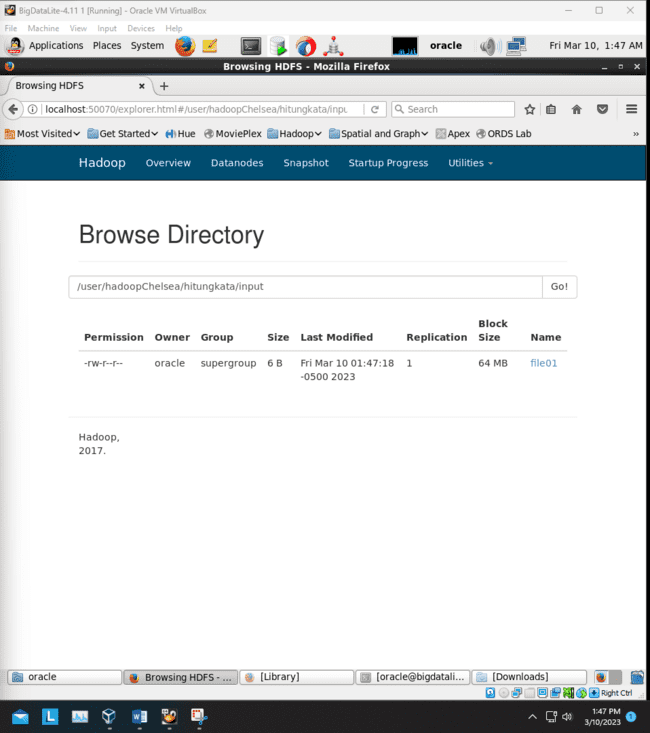


1. Memindahkan file input ke direktori input Hadoop

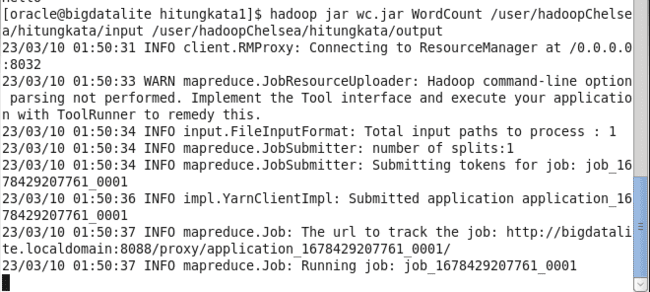


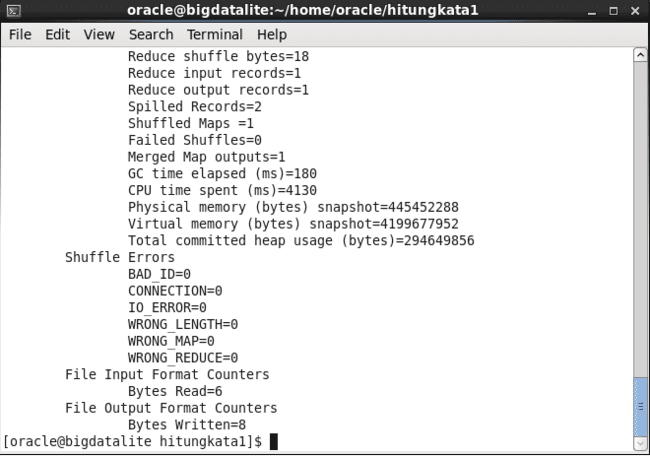
1. Cek apakah file sudah ada di direktori input Hadoop melalui terminal dan juga browsing HDFS





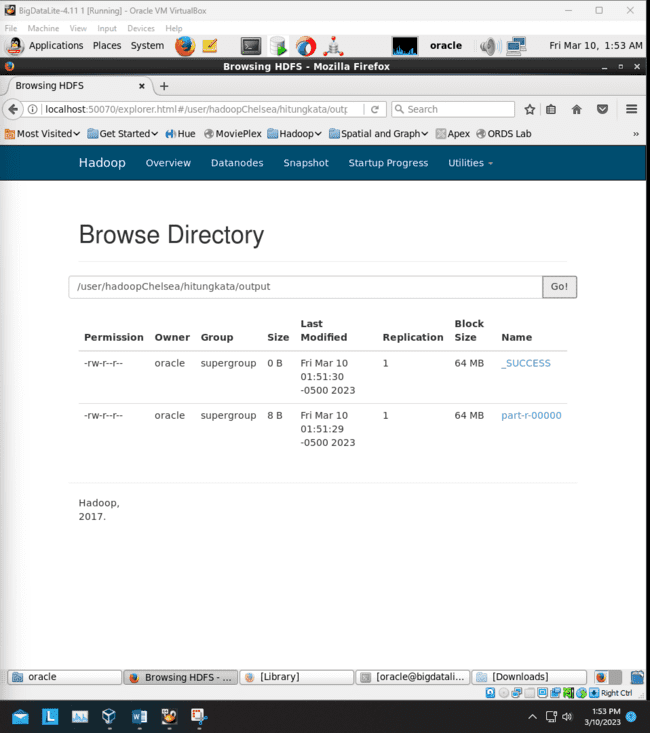
1. Menjalankan JAR dan membuat direktori output sebagai tempat hasil MapReduce





Output jika proses dinyatakan sukses:





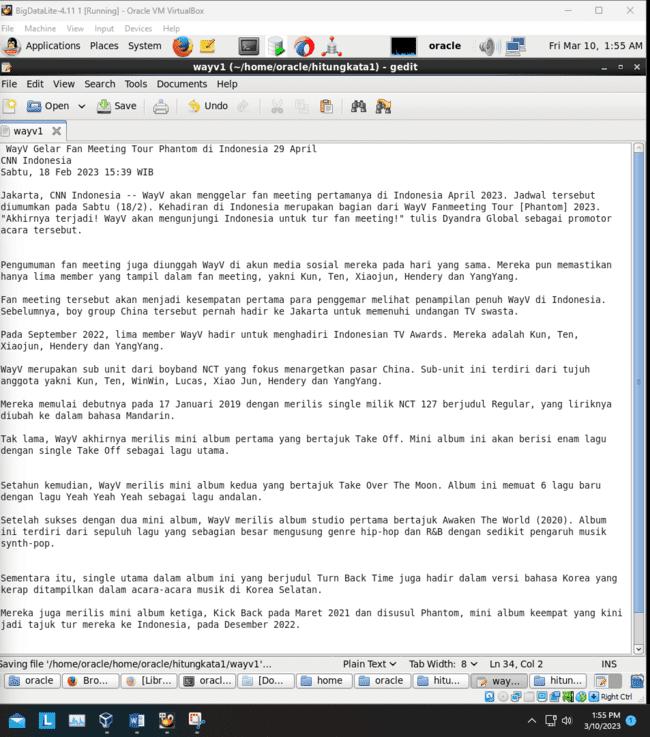
1. Mencari dua berita dengan topik WayV

Berita 1 = <https://yoursay.suara.com/entertainment/2023/03/07/175723/unjuk-bakat-baru-xiaojun-wayv-akan-jadi-mc-pogram-musik-sbs-mtv-the-show>

Berita 2 = <https://www.cnnindonesia.com/hiburan/20230218114512-227-914736/wayv-gelar-fan-meeting-tour-phantom-di-indonesia-29-april>

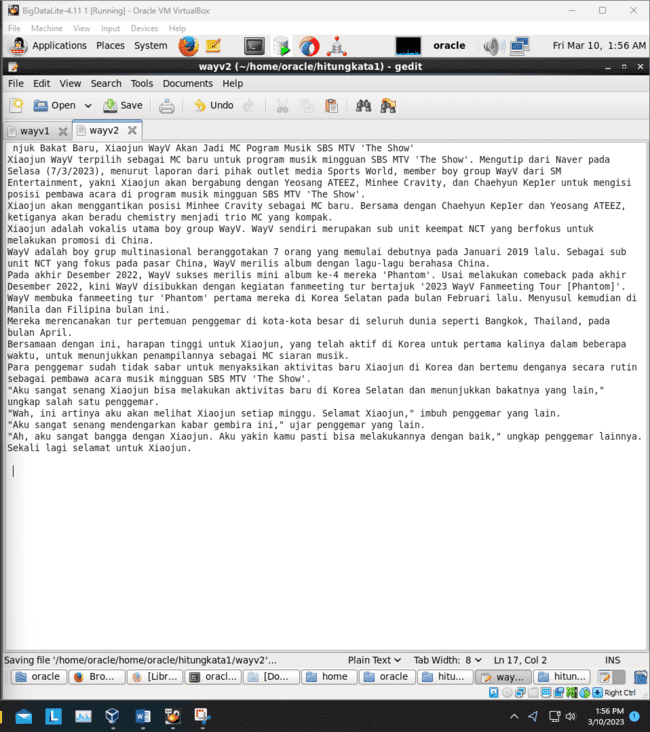
1. Membuat file input kosong wayv1 yang berisi berita 1, jangan lupa save setelah mengisi berita



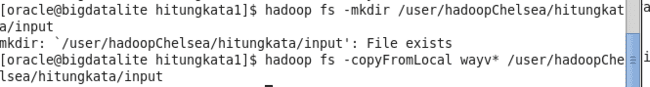


1. Membuat file input kosong wayv2 yang berisi berita 2, jangan lupa save setelah mengisi berita

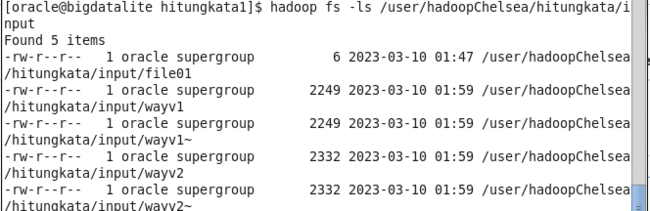


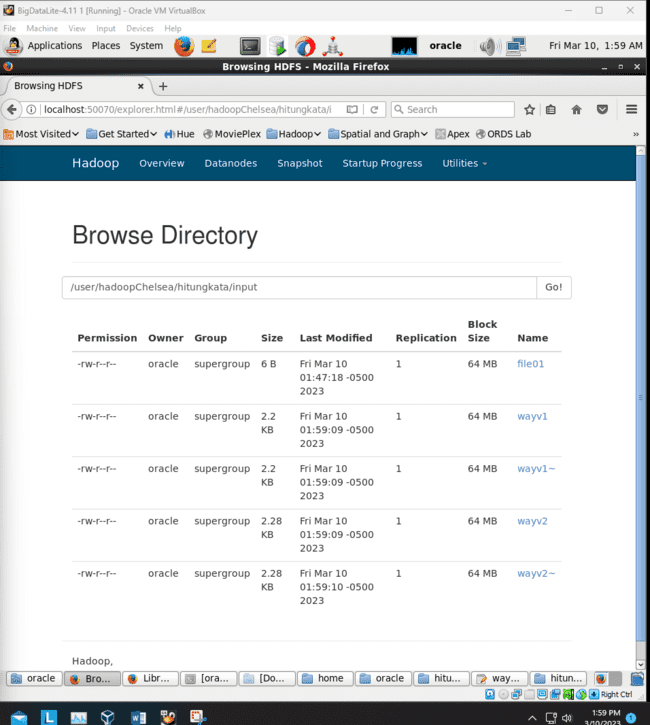


1. Memindahkan file input baru ke direktori input di Hadoop

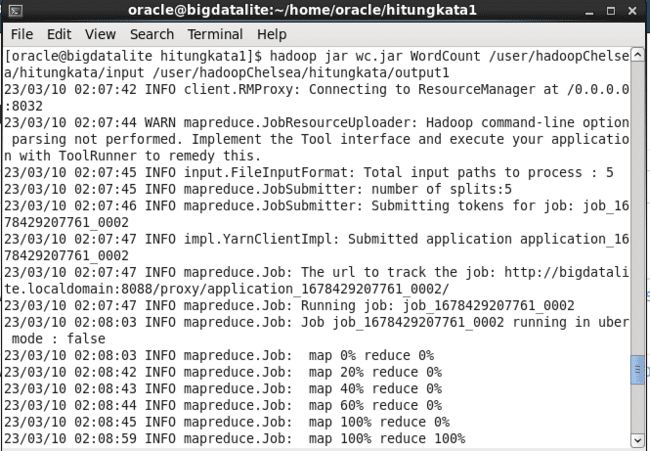


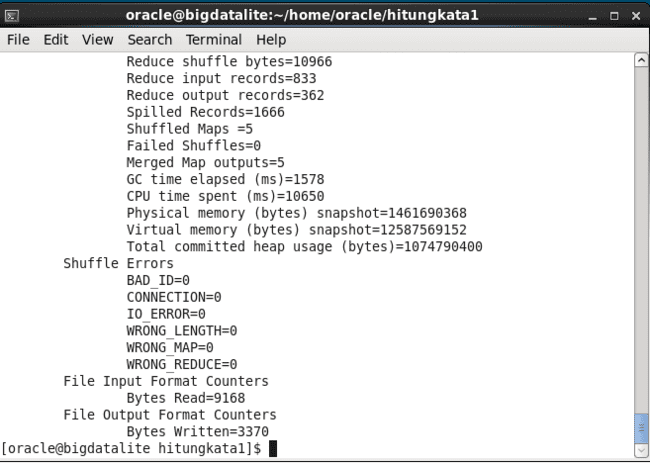
1. Cek apakah file sudah ada di direktori input Hadoop melalui terminal dan juga browsing HDFS



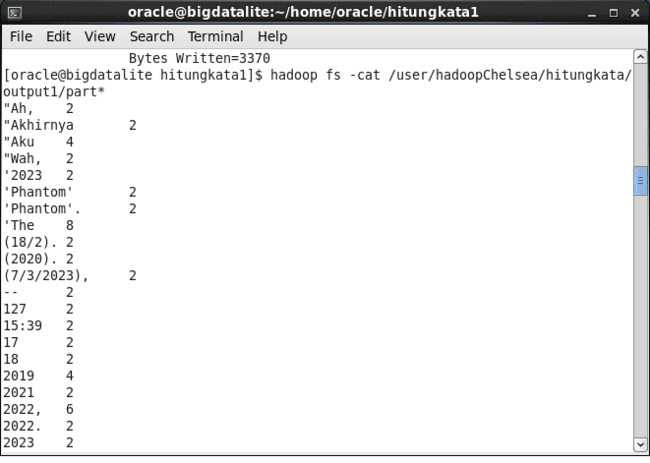


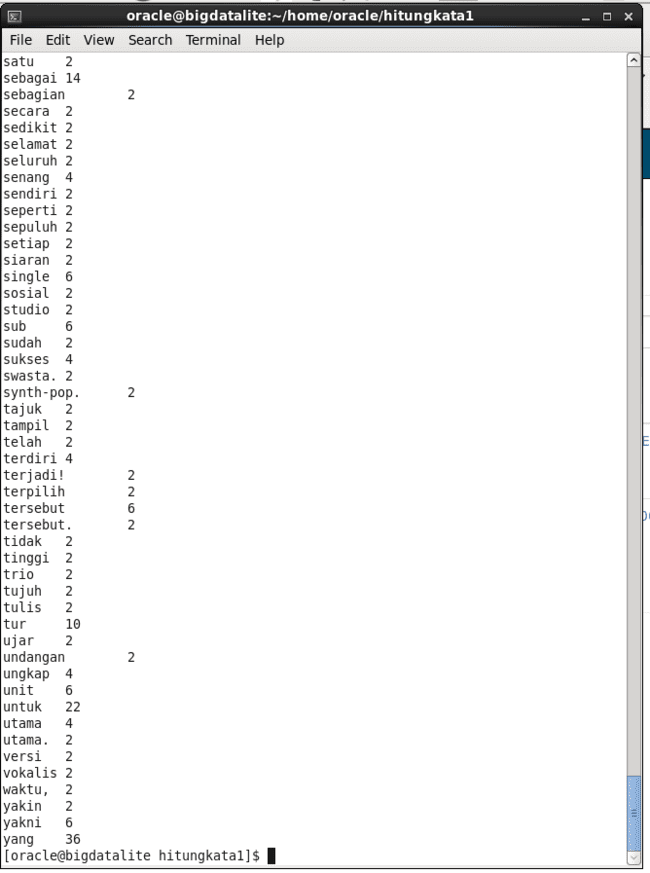
1. Menjalankan JAR dan membuat direktori output1 sebagai tempat hasil MapReduce untuk berita WayV

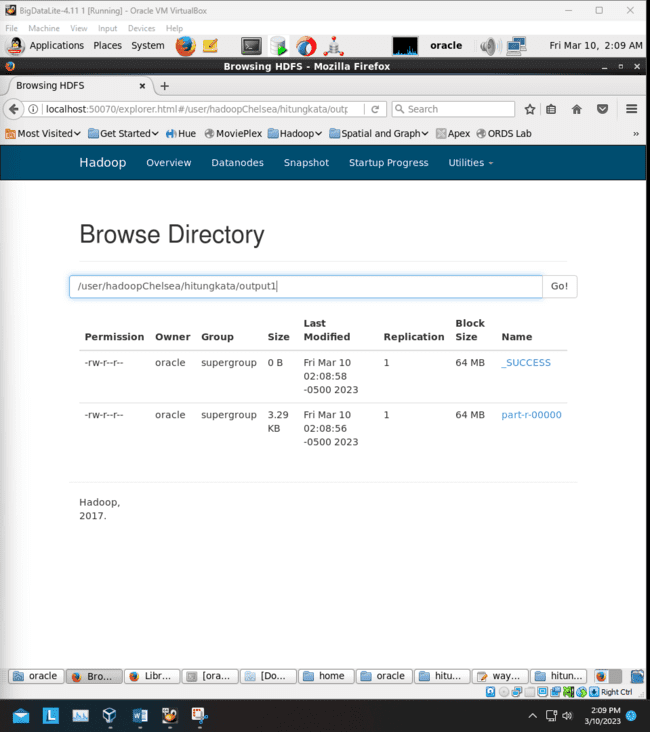




Output ketika proses MapReduce berhasil :



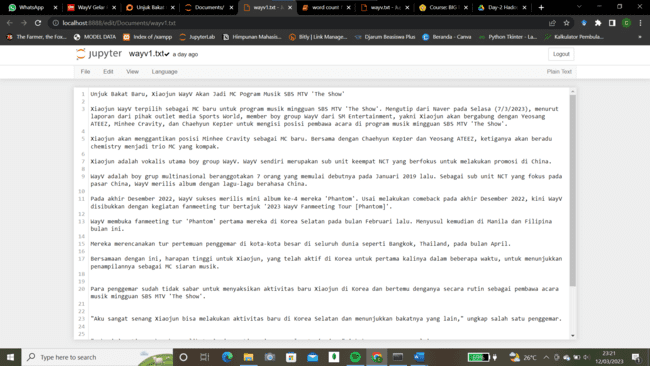


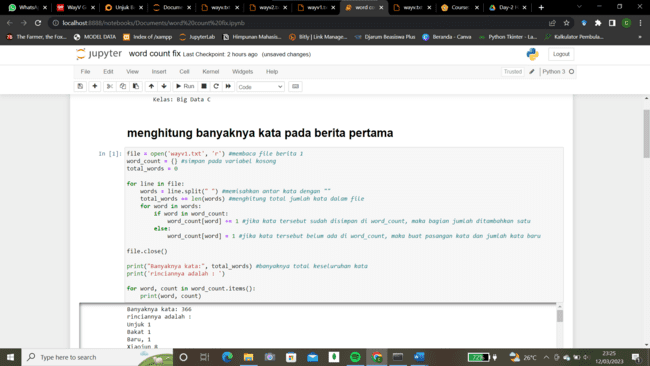


# **MAPREDUCE WORD COUNT PADA PYTHON**

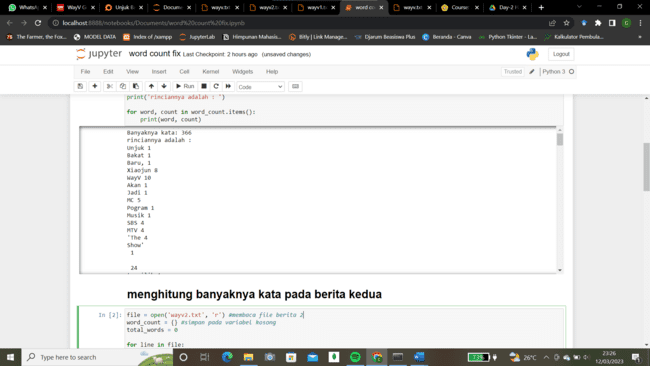
Tahapan MapReduce Word Count pada python yaitu melakukan pengecekan pada berita 1, berita 2, dan Ketika 2 berita digabung.

1. Berita 1





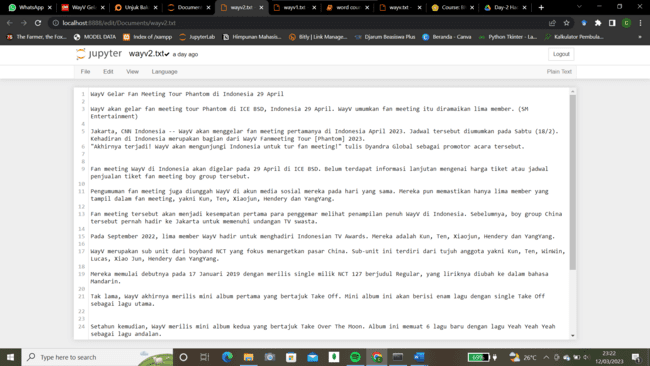
Output :

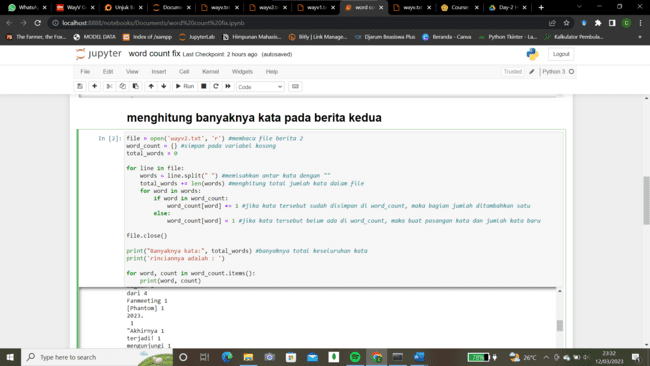


Penjelasan :

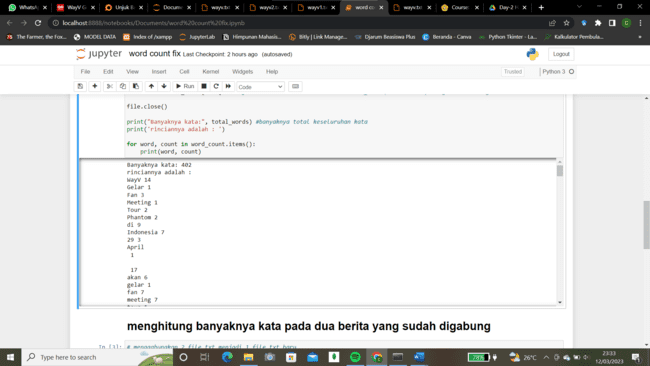
Langkah pertama tentunya membaca file txt berisi berita yang akan dikenalan proses MapReduce word count. Karena nantinya perhitungan dilakukan perkata, maka dibutuhkan tempat untuk menyimpan hasil kata yang sudah dihitung sebelumnya. Pemisahan per-kata dilakukan dengan memberikan tanda “ “ pada masing-masing kata kemudian dihitung berapa banyak kata tersebut ada pada berita 1. Karena MapReduce juga bekerja menjadikan satu kata-kata yang sama, maka dilakukan perulangan for, jika kata sudah ada pada word\_count maka banyaknya kata tersebut ditambah 1. Jika kata tersebut belum ada pada word\_count, maka membuat pasangan kata dan banyaknya kata baru. Pemangggilan output dilakukan dengan cara memanggil kata dan juga banyaknya kata dari word\_count, karena output yang diinginkan per kata maka ditambahkan items().

1. Berita 2





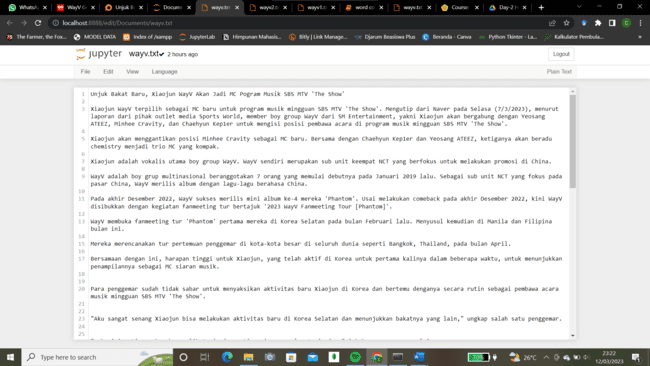
Output :

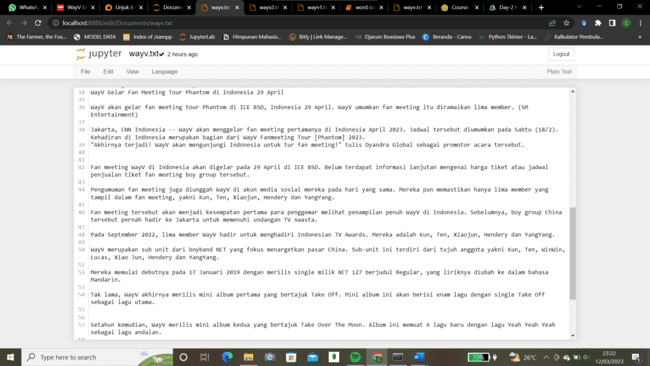


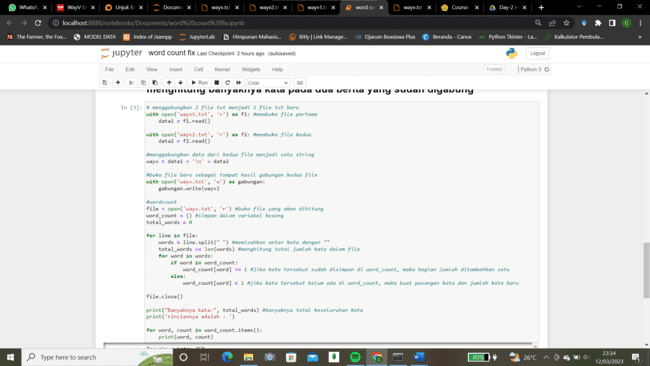
Penjelasan :

Langkah pertama tentunya membaca file txt berisi berita yang akan dikenalan proses MapReduce word count. Karena nantinya perhitungan dilakukan perkata, maka dibutuhkan tempat untuk menyimpan hasil kata yang sudah dihitung sebelumnya. Pemisahan per-kata dilakukan dengan memberikan tanda “ “ pada masing-masing kata kemudian dihitung berapa banyak kata tersebut ada pada berita 2. Karena MapReduce juga bekerja menjadikan satu kata-kata yang sama, maka dilakukan perulangan for, jika kata sudah ada pada word\_count maka banyaknya kata tersebut ditambah 1. Jika kata tersebut belum ada pada word\_count, maka membuat pasangan kata dan banyaknya kata baru. Pemangggilan output dilakukan dengan cara memanggil kata dan juga banyaknya kata dari word\_count, karena output yang diinginkan per kata maka ditambahkan items().

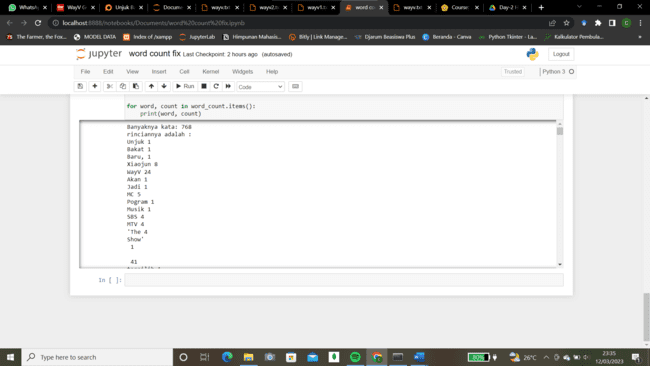
1. Kedua berita digabung







Output :



Penjelasan :

Karena 2 berita yang digabung menjadi 1, maka Langkah pertamanya tentu saja menggabungkan kedua file berita tersebut menjadi 1 text yang nantinya akan dipanggil saat melakukan proses MapReduce word count. Data1 mewakili proses baca berita 1, data2 mewakili proses baca berita 3. Wayv mewakili proses penggabungan dua berita yang dipisahkan oleh enter. Membuat file baru Bernama wayv.txt yang digunakan sebagai tempat penggabungan dua berita tersebut. Setelah berita sudah tergabung menjadi satu, langkah selanjutnya yaitu membaca file txt berisi berita yang akan dikenalan proses MapReduce word count. Karena nantinya perhitungan dilakukan perkata, maka dibutuhkan tempat untuk menyimpan hasil kata yang sudah dihitung sebelumnya. Pemisahan per-kata dilakukan dengan memberikan tanda “ “ pada masing-masing kata kemudian dihitung berapa banyak kata tersebut ada pada berita gabungan. Karena MapReduce juga bekerja menjadikan satu kata-kata yang sama, maka dilakukan perulangan for, jika kata sudah ada pada word\_count maka banyaknya kata tersebut ditambah 1. Jika kata tersebut belum ada pada word\_count, maka membuat pasangan kata dan banyaknya kata baru. Pemangggilan output dilakukan dengan cara memanggil kata dan juga banyaknya kata dari word\_count, karena output yang diinginkan per kata maka ditambahkan items().