LAPORAN PRAKTIKUM INFRASTRUKTUR BIG DATA
PERTEMUAN 3

MAP REDUCE

Oleh :

Nama : Jauhari Ahmad
No. Mhs : 205411167
Jurusan : Teknik Informatika / S1

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
AKAKOM
YOGYAKARTA
2020

PENDAHULUAN

A. TUJUAN

Mahasiswa menggunakan map reduce pada hadoop.

B. DASAR TEORI

MapReduce adalah sebuah model pemograman yang didesain untuk dapat melakukan pemrosesan data dengan jumlah yang sangat besar dengan cara membagi pemrosesan tersebut ke beberapa tugas yang indipenden satu sama lain.

Pengembang aplikasi MapReduce akan membutuhkan beberapa hal berikut dalam melakukan analisis data.

1. Berkas masukan. Berkas masukan ini dapat berupa berkas-berkas teks yang tersimpan di dalam sebuah media penyimpanan terdistribusi seperti Google File System (GFS), Hadoop File System(HDFS), AWS S3, Google Cloud Storage, dan lain-lain.
2. Fungsi Map & Reduce. Untuk membuat sebuah aplikasi MapReduce yang dapat dieksekusi secara paralel (misalkan dengan Hadoop MapReduce), pengembang aplikasi menyediakan fungsi khusus yang digunakan untuk melakukan pemrosesan pada fase map dan reduce. Seluruh hal yang berkaitan dengan penjadwalan, mekanisme penanganan eror, dll. akan dilakukan oleh MapReduce framework yang digunakan.

Berikut beberapa contoh fungsi MapReduce yang dapat dibuat oleh pengembang piranti lunak:

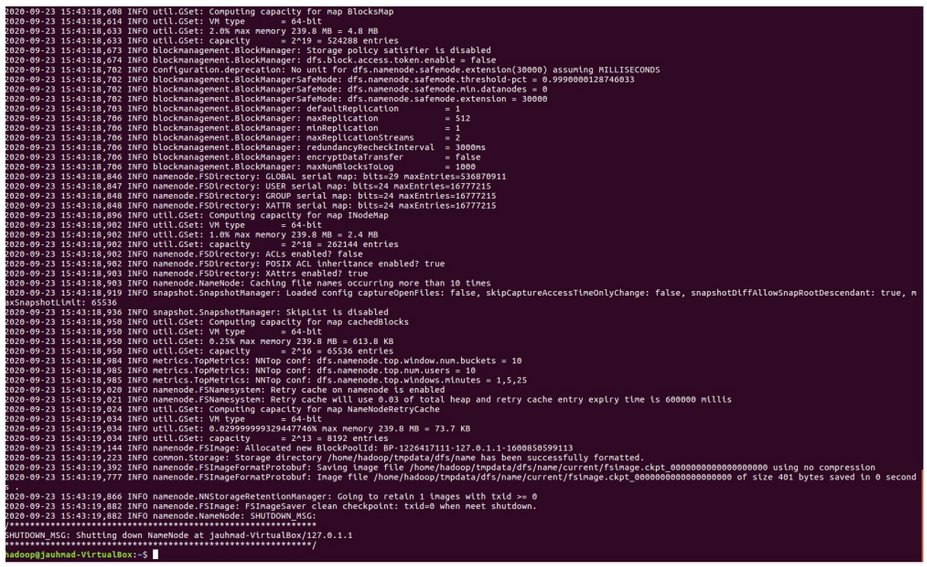
- Word Count
- Inverted Index, dll.

PEMBAHASAN

1. Menjalankan sistem operasi Linuk Ubuntu yang sudah terinstal hadoop dan menjalankan Hadoop

   a. Perintah:

```
hadoop@jauhmad-VirtualBox:~$ hdfs namenode -format
```



```
2020-09-23 15:43:18,608 INFO util.GSet: Computing capacity for map BlocksMap
2020-09-23 15:43:18,614 INFO util.GSet: VM type       = 64-bit
2020-09-23 15:43:18,633 INFO util.GSet: 2.0% max memory 239.8 MB = 4.8 MB
2020-09-23 15:43:18,633 INFO util.GSet: capacity      = 2^19 = 524288 entries
2020-09-23 15:43:18,673 INFO blockmanagement.BlockManager: Storage policy satisfier is disabled
2020-09-23 15:43:18,674 INFO blockmanagement.BlockManager: dfs.block.access.token.enable = false
2020-09-23 15:43:18,702 INFO Configuration.deprecation: No unit for dfs.namenode.safemode.extension(30000) assuming MILLISECONDS
2020-09-23 15:43:18,702 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.threshold-pct = 0.9990000128746033
2020-09-23 15:43:18,702 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.min.datanodes = 0
2020-09-23 15:43:18,702 INFO blockmanagement.BlockManagerSafeMode: dfs.namenode.safemode.extension = 30000
2020-09-23 15:43:18,703 INFO blockmanagement.BlockManager: defaultReplication         = 1
2020-09-23 15:43:18,706 INFO blockmanagement.BlockManager: maxReplication             = 512
2020-09-23 15:43:18,706 INFO blockmanagement.BlockManager: minReplication             = 1
2020-09-23 15:43:18,706 INFO blockmanagement.BlockManager: maxReplicationStreams      = 2
2020-09-23 15:43:18,706 INFO blockmanagement.BlockManager: redundancyRecheckInterval  = 3000ms
2020-09-23 15:43:18,706 INFO blockmanagement.BlockManager: encryptDataTransfer        = false
2020-09-23 15:43:18,706 INFO blockmanagement.BlockManager: maxNumBlocksToLog          = 1000
2020-09-23 15:43:18,846 INFO namenode.FSDirectory: GLOBAL serial map: bits=29 maxEntries=536870911
2020-09-23 15:43:18,847 INFO namenode.FSDirectory: USER serial map: bits=24 maxEntries=16777215
2020-09-23 15:43:18,848 INFO namenode.FSDirectory: GROUP serial map: bits=24 maxEntries=16777215
2020-09-23 15:43:18,848 INFO namenode.FSDirectory: XATTR serial map: bits=24 maxEntries=16777215
2020-09-23 15:43:18,896 INFO util.GSet: Computing capacity for map INodeMap
2020-09-23 15:43:18,902 INFO util.GSet: VM type       = 64-bit
2020-09-23 15:43:18,902 INFO util.GSet: 1.0% max memory 239.8 MB = 2.4 MB
2020-09-23 15:43:18,902 INFO util.GSet: capacity      = 2^18 = 262144 entries
2020-09-23 15:43:18,902 INFO namenode.FSDirectory: ACLs enabled? false
2020-09-23 15:43:18,902 INFO namenode.FSDirectory: POSIX ACL inheritance enabled? true
2020-09-23 15:43:18,903 INFO namenode.FSDirectory: Xattrs enabled? true
2020-09-23 15:43:18,903 INFO namenode.NameNode: Caching file names occurring more than 10 times
2020-09-23 15:43:18,919 INFO snapshot.SnapshotManager: Loaded config captureOpenFiles: false, skipCaptureAccessTimeOnlyChange: false, snapshotDiffAllowSnapRootDescendant: true, m
axSnapshotLimit: 65536
2020-09-23 15:43:18,936 INFO snapshot.SnapshotManager: SkipList is disabled
2020-09-23 15:43:18,950 INFO util.GSet: Computing capacity for map cachedBlocks
2020-09-23 15:43:18,950 INFO util.GSet: VM type       = 64-bit
2020-09-23 15:43:18,950 INFO util.GSet: 0.25% max memory 239.8 MB = 613.8 KB
2020-09-23 15:43:18,950 INFO util.GSet: capacity      = 2^16 = 65536 entries
2020-09-23 15:43:18,984 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.window.num.buckets = 10
2020-09-23 15:43:18,985 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.num.users = 10
2020-09-23 15:43:18,985 INFO metrics.TopMetrics: NNTop conf: dfs.namenode.top.windows.minutes = 1,5,25
2020-09-23 15:43:19,020 INFO namenode.FSNamesystem: Retry cache on namenode is enabled
2020-09-23 15:43:19,021 INFO namenode.FSNamesystem: Retry cache will use 0.03 of total heap and retry cache entry expiry time is 600000 millis
2020-09-23 15:43:19,024 INFO util.GSet: Computing capacity for map NameNodeRetryCache
2020-09-23 15:43:19,034 INFO util.GSet: VM type       = 64-bit
2020-09-23 15:43:19,034 INFO util.GSet: 0.029999999329447746% max memory 239.8 MB = 73.7 KB
2020-09-23 15:43:19,034 INFO util.GSet: capacity      = 2^13 = 8192 entries
2020-09-23 15:43:19,144 INFO namenode.FSImage: Allocated new BlockPoolId: BP-1226417111-127.0.1.1-1600850599113
2020-09-23 15:43:19,223 INFO common.Storage: Storage directory /home/hadoop/tmpdata/dfs/name has been successfully formatted.
2020-09-23 15:43:19,392 INFO namenode.FSImageFormatProtobuf: Saving image file /home/hadoop/tmpdata/dfs/name/current/fsimage.ckpt_0000000000000000000 using no compression
2020-09-23 15:43:19,777 INFO namenode.FSImageFormatProtobuf: Image file /home/hadoop/tmpdata/dfs/name/current/fsimage.ckpt_0000000000000000000 of size 401 bytes saved in 0 second
s .
2020-09-23 15:43:19,866 INFO namenode.NNStorageRetentionManager: Going to retain 1 images with txid >= 0
2020-09-23 15:43:19,882 INFO namenode.FSImage: FSImageSaver clean checkpoint: txid=0 when meet shutdown.
2020-09-23 15:43:19,882 INFO namenode.NameNode: SHUTDOWN_MSG:
/************************************************************
SHUTDOWN_MSG: Shutting down NameNode at jauhmad-VirtualBox/127.0.1.1
************************************************************/
hadoop@jauhmad-VirtualBox:~$
```

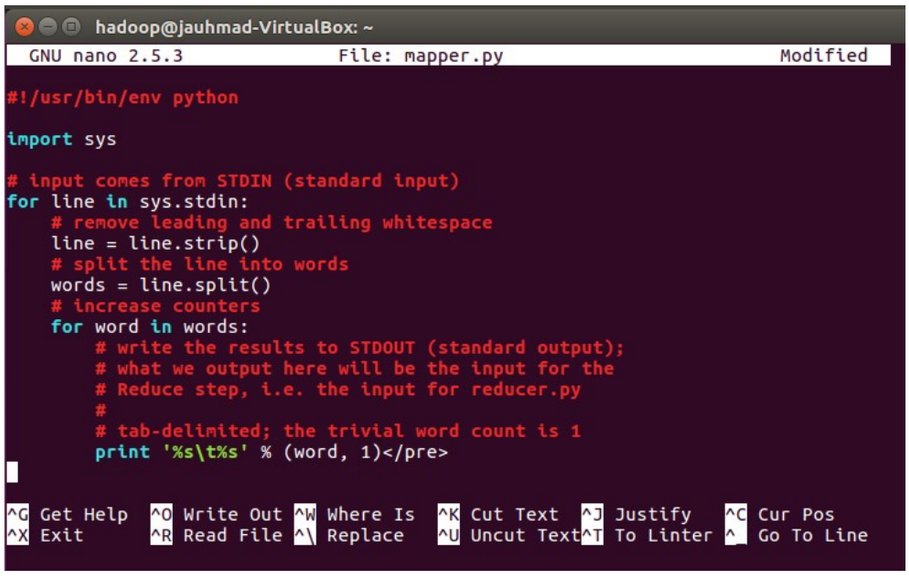
   b. Perintah

```
hadoop@jauhmad-VirtualBox:~$ start-dfs.sh
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [jauhmad-VirtualBox]
```

   c. Perintah :

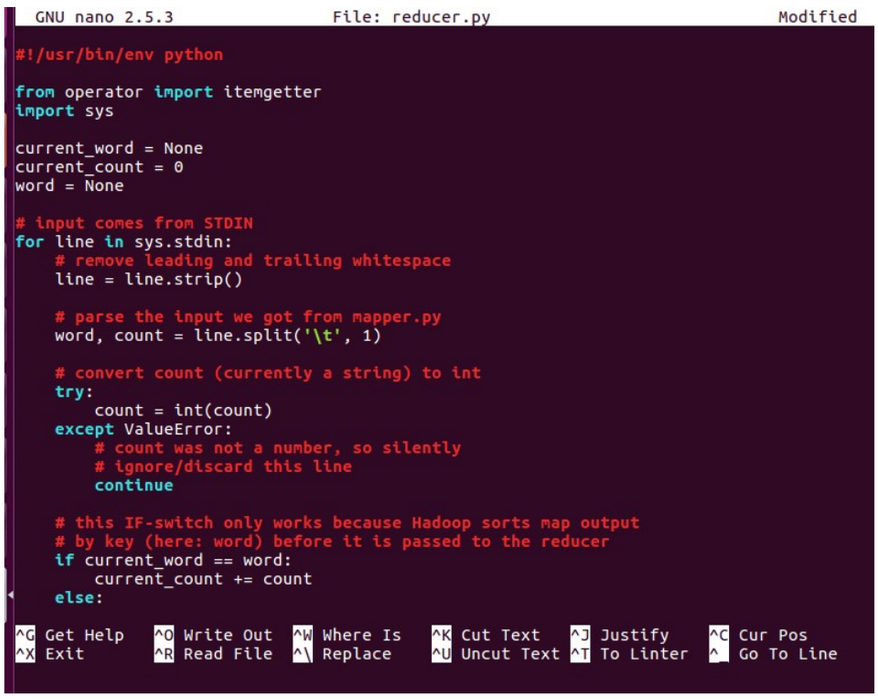
```
hadoop@jauhmad-VirtualBox:~$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
```

d. Membuat mapper.py pada /home/hadoop



Ubah mapper.py menjadi executable

e. Membuat file reducer.py pada /home/hadoop



Ubah menjadi file executable

```
hadoop@jauhmad-VirtualBox:~$ chmod +x /home/hadoop/reducer.py
```

f. Unduh file profile.txt, simpan pada /home/hadoop

```
  GNU nano 2.5.3                File: profile.txt

Profil STMIK AKAKOM

Dalam abad ke-20 ini dunia banyak diwarnai dengan berbagai kemajuan secara menakjubkan$

Terhitung mulai 1 Maret 1983, Akademi Aplikasi Komputer (AKAKOM), diubah menjadi Akade$

Agar lembaga tersebut mampu menghasilkan tenaga-tenaga profesional maupun akademik yan$

Saat ini STMIK AKAKOM Yogyakarta memunyai 2 program sarjana (Prodi. Teknik Informatika$
```

2. Mengamati perintah:

```
hadoop@jauhmad-VirtualBox:~$ cat /home/hadoop/profile.txt | /home/hadoop/mapper.py | so
rt -k1,1 | /home/hadoop/reducer.py
```

Hasil:

```
hadoop@jauhmad-VirtualBox:~$ cat /home/hadoop/profile.txt | /home/hadoop/mapper.py | sort -k1,1 | /home/hadoop/reducer.py
1       1
1979    1
1983,   1
1985,   1
1992,   1
2       2
262/DIKTI/Kep/1992,     1
3       1
30      1
8       1
abad    1
Agar    1
akademi 1
Akademi 5
akademik        1
AKAKOM  5
(AKAKOM),       1
AKAKOM. 3
akreditasi      1
Akuntansi)      1
AMIK    1
(AMIK)  1
Aplikasi        2
atas,   1
B.      1
baik    1
Bakti,  1
banyak  1
bentuknya       1
berbagai        1
berbobot        1
berdasarkan     1
bernama 1
bertujuan       1
bidang  1
dalam   1
Dalam   1
dan     14
dengan  3
Departemen      1
di      2
dibakukan       1
dibidang        1
diciptakan      1
didirikan       1
```

Penjelasan:

Perintah berfungsi untuk menghitung jumlah suatu kata yang terdapat di dalam suau kalimat.

mapper.py

Befungsi untuk membaca data dari STDIN, membaginya menjadi kata-kata dan mengeluarkan daftar baris yang memetakan kata-kata ke jumlah STDOUT. Skrip Map tidak akan menghitung jumlah dari kemunculan sebuah kata. Dalam kasus ini, membiarkan langkah “mengurangi” berikutnya melakukan penghitungan jumlah akhir. Kurangi langkah:

reducer.py

Berfungsi akan membaca hasil mapper.py dari STDIN (jadi format keluaran mapper.py dan format masukan yang diharapkan reducer.py

harus cocok) dan menjumlahkan kemunculan setiap kata ke hitungan akhir, dan kemudian menampilkan hasilnya ke STDOUT.

KESMIPULAN

1. mapper.py befungsi untuk membaca data dari STDIN, membaginya menjadi kata-kata dan mengeluarkan daftar baris yang memetakan kata-kata ke jumlah STDOUT.
2. reducer.py berfungsi akan membaca hasil mapper.py dari STDIN (jadi format keluaran mapper.py dan format masukan yang diharapkan reducer.py harus cocok) dan menjumlahkan kemunculan setiap kata ke hitungan akhir, dan kemudian menampilkan hasilnya ke STDOUT
