

1. Instruksi

Buatlah implementasi salah satu atau beberapa Teknologi Big Data (Hadoop, MapReduce, Spark, Flume, Sqoop, Hive, Pig, Oozie) dengan permasalahan sederhana yang anda tentukan sendiri dan tulis dalam bentuk laporan dengan format bebas. Implementasi dapat menggunakan komputer sendiri, virtual machine, Docker dan Kubernetes, atau cloud computing.

2. Permasalahan

Kelompok Tani Aglonema di sebuah desa yang terletak di Kabupaten Sleman memiliki anggota yang semakin bertambah. Begitu pula jumlah tanaman yang dimiliki setiap anggotanya. Padahal kelompok tani tersebut membutuhkan data terintegrasi setiap aset tanaman yang dimiliki oleh setiap anggota untuk perhitungan estimasi pencapaian target jumlah tanaman secara keseluruhan. Oleh karena itu, diperlukan teknologi big data karena dataset tersebut akan berukuran besar. Kelompok tani tersebut melakukan pendataan setiap tambahan aset tanaman yang terdiri dari kolom nama_anggota, jenis_tanaman, dan jumlah seperti yang ditunjukkan pada dataset Gambar 2.1. Kemudian untuk mempermudah implementasi Teknologi Big Data, dataset tersebut di-convert dalam format csv seperti pada Gambar 2.2:

nama_anggota	jenis_tanaman	jumlah
Zamroni	red_anjamani	2
Zamroni	legasi_merah	1
Zamroni	red_queen	3
Inggar Puriyono	red_anjamani	3
Inggar Puriyono	suksom	2
Inggar Puriyono	legasi_merah	6
Inggar Puriyono	red_queen	2
Anita Widhayanti	red_anjamani	4
Anita Widhayanti	suksom	5
Anita Widhayanti	legasi_merah	12
Anita Widhayanti	dud_white	9
Anita Widhayanti	red_exotic	1
Anita Widhayanti	red_queen	3
Dijan Ratri	red_anjamani	1
Kartikaningsih		

Gambar 2.1

Zamroni,red_anjamani,2	
Zamroni,legasi_merah,1	
Zamroni,red_queen,3	
Inggar Puriyono,red_anjamani,3	
Inggar Puriyono,suksom,2	
Inggar Puriyono,legasi_merah,6	
Inggar Puriyono,red_queen,2	
Anita Widhayanti,red_anjamani,4	
Anita Widhayanti,suksom,5	
Anita Widhayanti,legasi_merah,12	
Anita Widhayanti,dud_white,9	
Anita Widhayanti,red_exotic,1	
Anita Widhayanti,red_queen,3	
Dijan Ratri Kartikaningsih,red_anjamani,	

Gambar 2.2

3. Teknologi Big Data

Teknologi Big Data yang dimanfaatkan dalam implementasi kali ini adalah Apache Hadoop, Mapreduce, dan Apache Pig. Apache Hadoop adalah *software open-source framework* untuk penyimpanan terdistribusi dan pemrosesan big data. Hadoop berjalan dengan bahasa pemrograman Java menggunakan Mapreduce. Untuk mempermudah proses big data, terdapat *high-level programming language* yaitu Apache Pig dengan bahasa Pig Latin yang lebih mudah dipelajari dan bersifat *scripting language* artinya di-run setiap barisnya. Baris kode ini akan di-*compile* dalam pemrograman MapReduce oleh Apache Pig. Apache Pig ini berjalan di atas Hadoop, maka diperlukan instalasi Java, Hadoop, dan Pig.

4. Implementasi

- A. Implementasi dijalankan di mode lokal dengan eksekusi kode pig -x local pada Gambar 4.1

```
C:\windows\system32>pig -x local
2022-10-11 14:29:30,691 INFO org.apache.pig.ExecTypeProvider: Trying ExecType : LOCAL
2022-10-11 14:29:30,692 INFO org.apache.pig.ExecTypeProvider: Picked LOCAL as the ExecType
2022-10-11 14:29:31,080 [main] INFO org.apache.pig.Main - Apache Pig version 0.17.0 (r1797386) compiled Jun 02 2017, 15:41:58
2022-10-11 14:29:31,080 [main] INFO org.apache.pig.Main - logging error messages to: D:\BigData\hadoop-3.2.2\logs\pig_1665473371086.log
2022-10-11 14:29:31,122 [main] INFO org.apache.pig.impl.util.Utils - Default bootstrap file C:\Users\hp/.pigbootstrap not found
2022-10-11 14:29:31,237 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.job.tracker is deprecated. Instead, use mapreduce.jobtracker.address
2022-10-11 14:29:31,240 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
2022-10-11 14:29:31,657 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2022-10-11 14:29:31,684 [main] INFO org.apache.pig.PigServer - Pig Script ID for the session: PIG-default-b2ccf5ed-97ed-4f95-94c6-92bac78ba42e
2022-10-11 14:29:31,684 [main] WARN org.apache.pig.PigServer - ATS is disabled since yarn.timeline-service.enabled set to false
```

Gambar 4.1

- B. LOAD Dataset aglonema di mode lokal untuk menghubungkan dataset di memori lokal pada Gambar 4.2

```
grunt> aglonema = LOAD 'D:\BigData\Dataset\aglonema.csv' using PigStorage(',') as (nama_anggota:chararray, jenis_aglonema:chararray, jumlah:int);
2022-10-11 14:31:56,674 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
```

Gambar 4.2

- C. Untuk menjadikan dataset sebagai input gunakan operator DUMP, dengan command *dump aglonema*; maka dapat dilihat dataset aglonema.csv pada Gambar 4.3

```
Administrator: Command Prompt - pig -x local
2022-10-11 14:32:12,999 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized.
2022-10-11 14:32:12,987 [main] WARN org.apache.hadoop.io.nativeio.NativeIO - NativeIO.getStat error (3): the system cannot find the path specified.
-- file path: tmp/tmp22809488/tmp2102868577/par = 00000
2022-10-11 14:32:12,987 [main] WARN org.apache.hadoop.io.nativeio.NativeIO - NativeIO.getStat error (3): the system cannot find the path specified.
-- file path: tmp/tmp22809488/tmp2102868577/SUCCESS
2022-10-11 14:32:12,102 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-10-11 14:32:12,102 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(Zamroni,red_anjmani,2)
(Zamroni,legasi_merah,1)
(Zamroni,red_queen,3)
(Inggar_Puriyono,red_anjmani,3)
(Inggar_Puriyono,suksom,2)
(Inggar_Puriyono,legasi_merah,6)
(Inggar_Puriyono,red_queen,2)
(Anita_Kidhayanti,red_anjmani,4)
(Anita_Kidhayanti,suksom,5)
(Anita_Kidhayanti,legasi_merah,12)
(Anita_Kidhayanti,dud_white,9)
(Anita_Kidhayanti,red_exotic,1)
(Anita_Kidhayanti,red_queen,3)
(Oljan_Ratri_Kartikaningsih,red_anjmani,1)
(Oljan_Ratri_Kartikaningsih,suksom,1)
(Oljan_Ratri_Kartikaningsih,legasi_merah,1)
(Oljan_Ratri_Kartikaningsih,dud_white,8)
(Oljan_Ratri_Kartikaningsih,red_exotic,3)
(Oljan_Ratri_Kartikaningsih,red_queen,3)
(Oljan_Ratri_Kartikaningsih,red_cherry,1)
(Oljan_Ratri_Kartikaningsih,mahaseti,1)
(Lukman_Harun_M.,widuri,1)
(Lukman_Harun_M.,adila,1)
(Lukman_Harun_M.,red_anjmani,1)
(Lukman_Harun_M.,suksom,2)
(Lukman_Harun_M.,legasi_merah,6)
(Lukman_Harun_M.,red_exotic,1)
(Lukman_Harun_M.,red_queen,3)
(Lukman_Harun_M.,bidadadi,1)
(Zuli_Astanto,suksom,2)
(Zuli_Astanto,legasi_merah,1)
(Zuli_Astanto,dud_white,4)
(Zuli_Astanto,red_exotic,1)
(Zuli_Astanto,red_queen,3)
(Zuli_Astanto,bidadadi,1)
(Smauri,red_anjmani,2)
(Smauri,suksom,5)
(Smauri,legasi_merah,3)
(Smauri,dud_white,2)
(Smauri,red_exotic,1)
(Smauri,red_queen,3)
(Suradi,suksom,1)
(Suradi,legasi_merah,3)
```

Gambar 4.3

D. Untuk melihat jumlah dari setiap jenis tanaman maka dilakukan operasi pada Gambar 4.4

```
grunt> asetinfo = FOREACH(GROUP aglonema BY jenis_aglonema) GENERATE group AS nama_aglonema, SUM(aglonema.jumlah) as jumlah_per_jenis;  
grunt> describe asetinfo;  
asetinfo: {nama_aglonema: chararray,jumlah_per_jenis: long}
```

Gambar 4.4

Tabel asetinfo akan terdiri dari kolom nama_aglonema dan jumlah_per_jenis. Agar tabel tersusun dari aset tanaman dari paling banyak hingga paling sedikit dilanjutkan perintah berikut:

```
grunt> sorted = ORDER asetinfo BY jumlah_per_jenis DESC;
```

```
grunt> dump sorted;
```

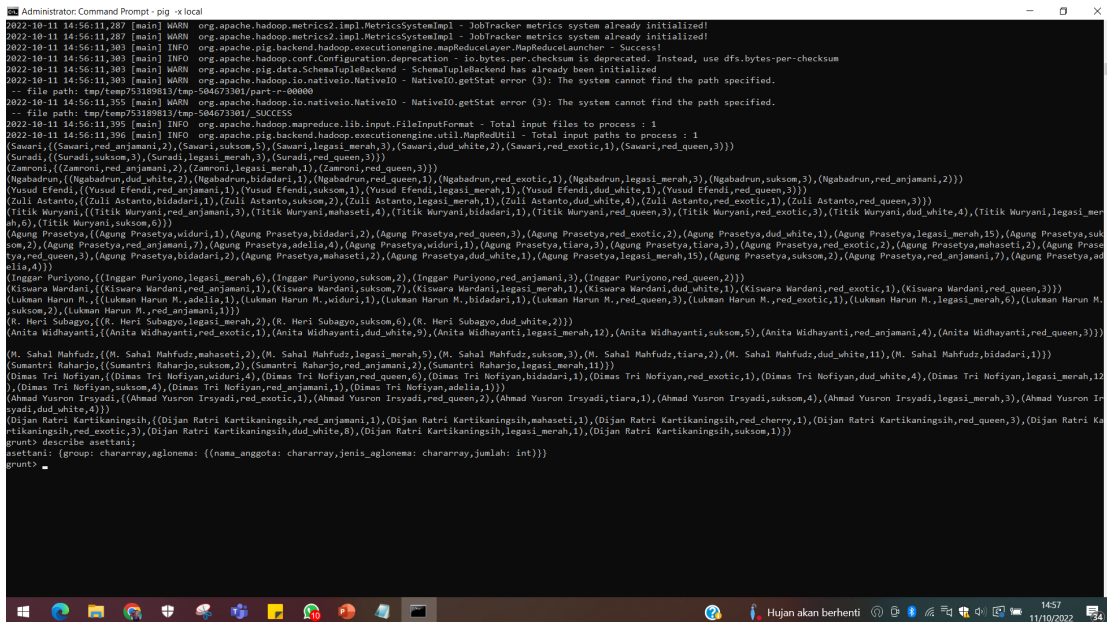
Hasil operator yaitu jumlah tanaman aglonema yang diklasifikasikan setiap jenisnya urut menurun ditunjukkan pada Gambar 4.5

```
(legasi_merah,107)  
(suksom,60)  
(dud_white,54)  
(red_queen,47)  
(red_anjamani,37)  
(red_exotic,18)  
(mahaseti,11)  
(bidadari,10)  
(adelia,10)  
(tiara,9)  
(widuri,7)  
(red_cherry,1)
```

Gambar 4.5

```
grunt> asettni = GROUP aglonema BY nama_anggota;
grunt> dump asettni;
2022-10-11 14:56:10 404 [main] INFO org.apache.pig tool
```

Operasi akan menghasilkan tuple yang berisi nama_anggota{(jenis_tanaman,jumlah)} yang menunjukkan setiap tanaman beserta jumlah yang dimiliki anggota tersebut seperti Gambar 4.7



Gambar 4.7

Berdasarkan implementasi Big Data yang telah diproses di bagian Implementasi D, dapat diketahui bahwa aset produksi tanaman aglonema **terbesar** di kelompok tani tersebut adalah aglonema jenis **legasi merah** dengan jumlah **107** dan aset produksi tanaman aglonema **terkecil** adalah jenis **red cherry** dengan jumlah **1**. Hal ini divalidasi oleh para petani karena jenis legasi merah mudah dirawat dan pasarnya bagus. Sedangkan red cherry merupakan jenis aglonema paling sensitif dan sulit dirawat sehingga rentan mati. Dari implementasi bagian E juga mempermudah para petani untuk mengakses data asetnya untuk setiap jenis tanaman.