

Panduan Praktikum Pengembangan Aplikasi Terdistribusi – Hadoop 1

1. Untuk praktikum Hadoop sesi 1, tidak ada hasil yang harus dikumpulkan. Target anda hari ini adalah familiarisasi dengan lingkungan hadoop dan dapat melakukan kompilasi program sederhana hadoop. Tugas hadoop yang dikumpulkan adalah mencari pengguna dengan jumlah follower level 2 terbanyak dari data twitter (dijelaskan pada dokumen terpisah).
2. Cluster hadoop telah di-setup pada 9 komputer pada Labdas 5, dengan detail sbb:

- Ld5-01: 167.205.35.25
- Ld5-02: 167.205.35.26
- Ld5-04: 167.205.35.28
- Ld5-05: 167.205.35.29
- Ld5-06: 167.205.35.30
- Ld5-07: 167.205.35.31
- Ld5-08: 167.205.35.32
- Ld5-09: 167.205.35.33
- Ld5-10: 167.205.35.34

Ld5-01 sebagai NameNode dan ResourceManager, dan semua node lain (termasuk Ld5-01) sebagai DataNode dan NodeManager. Untuk login ke setiap komputer tersebut, anda dapat menggunakan user: hduser, password: hadoop.

3. Instalasi Hadoop terletak pada direktori /home/hduser/Application/hadoop-2.7.1
4. Untuk melihat web interface untuk NameNode, gunakan browser pada alamat <http://167.205.35.25:50070>. Gunakan menu Utilities pada page tersebut untuk melihat file system yang tersimpan pada HDFS.
5. Buatlah direktori tempat anda bekerja pada HDFS, pada subdirektori /user dengan perintah sbb:

```
$ cd /home/hduser/Application/hadoop-2.7.1  
$ bin/hdfs dfs -mkdir /user/<nama_dir_baru>
```

Periksa direktori yang baru dibuat tersebut melalui <http://167.205.35.25:50070>
6. Anda akan mencoba menjalankan program grep yang tersedia pada hadoop-examples. Source code untuk program ini dapat dilihat pada komputer 167.205.35.25 pada direktori /home/hduser/Application/hadoop-2.7.1-src/hadoop-mapreduce-project/hadoop-mapreduce-examples/src/main/java/org/apache/hadoop/examples/Grep.java
Untuk melihat web interface untuk ResourceManager, gunakan browser pada alamat <http://167.205.35.25:8088>. Job history server dapat diakses pada alamat <http://167.205.35.25:19888>

7. Siapkan data input yang akan anda proses pada hdfs. Misal jika pada langkah 3 di atas anda membuat direktori dengan nama `/user/myworkspace` dan data yang akan anda grep adalah file xml yang terdapat pada direktori `/home/hduser/Application/hadoop-2.7.1/etc/hadoop`, maka jalankan perintah:

```
$ bin/hdfs dfs -put /home/hduser/Application/hadoop-2.7.1/etc/hadoop/*.xml  
/user/myworkspace/input
```

8. Jalankan program grep pada hadoop dengan perintah (pada direktori `/home/hduser/Application/hadoop-2.7.1`):

```
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.7.1.jar  
grep <nama_direktori_input> <nama_direktori_output> <pattern_yang_akan_dicari>
```

Perhatikan eksekusi program melalui web interface untuk ResourceManager

9. Setelah berhasil, lihat hasil program dengan perintah:

```
$ bin/hdfs dfs -cat /user/myworkspace/output/part-r-00000
```

10. Cobalah lakukan kompilasi program hadoop sederhana sesuai dengan panduan yang terdapat pada <http://hadoop.apache.org/docs/current/hadoop-mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html>