

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.
- ~~2. Gunakan komputer 167.205.33.58 atau 167.205.33.99. Kedua komputer ini telah dikonfigurasi untuk menjalankan hadoop dalam mode pseudo-distributed (masing-masing komputer menjalankan hadoop secara terpisah, tidak terhubung dalam sebuah cluster). Instalasi Hadoop terletak pada direktori /home/hadoop/hadoop-2.5.1~~
3. Gunakan komputer 167.205.32.201. user: hduser, pass: hadoop. Komputer ini telah dikonfigurasi untuk menjalankan cluster hadoop dalam mode cluster (terdiri atas 7 komputer). Instalasi Hadoop terletak pada direktori /home/hduser/Application/hadoop-2.2.0
4. Untuk melihat web interface untuk NameNode, gunakan browser pada alamat <http://167.205.32.201:50070>. 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/hadoop/hadoop-2.5.1  
$ bin/hdfs dfs -mkdir /user/<nama_dir_baru>
```


Periksa direktori yang baru dibuat tersebut melalui <http://167.205.32.201:50070>
6. Anda akan mencoba menjalankan program grep yang tersedia pada hadoop-examples. Source code untuk program ini dapat dilihat pada komputer 167.205.33.58 pada direktori /home/hadoop/hadoop-2.5.1-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://167.205.32.201:8088>
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/hadoop/hadoop-2.2.0/etc/hadoop, maka jalankan perintah:

```
$ bin/hdfs dfs -put /home/hduser/Application/hadoop-2.5.1/etc/hadoop/*.xml  
/user/myworkspace/input
```
8. Jalankan program grep pada hadoop dengan perintah (pada direktori /home/hadoop/hadoop-2.5.1):

```
$ bin/hadoop jar share/hadoop/mapreduce/hadoop-mapreduce-examples-2.5.1.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>