

IMPLEMENTASI *HADOOP STANDALONE* PADA SISTEM OPERASI WINDOWS

Tasrif Hasanuddin

Program Studi : Sistem Informasi

STMIK Profesional Makassar

tasrif.h@gmail.com

Abstrack

Hadoop Mode Standalone dapat dijalankan pada PC berbasis Windows yang tentunya sudah terinstall Java Virtual Machine dan Java Eclipse IDE. Penggunaan Hadoop mode ini ditujukan untuk pengembangan aplikasi MapReduce (Map Reduce Application Development) dan testing program dengan data berukuran kecil. Dengan menggunakan Hadoop Standalone pada Eclipse, coding dan debug program dapat dilakukan lebih mudah dan cepat.

PENDAHULUAN

Big Data telah menjadi salah satu keunggulan strategis yang menjadikan Google berjaya sebagai raksasa internet serta berhasil menjadi bagian yang tak terpisahkan dari sebagian besar pengguna internet di dunia. Dalam hal ini, Google memiliki teknologi sendiri untuk memproses BIG Data, yaitu GFS, Google Map Reduce, dan Google Bigtable. Google juga telah mempublikasikan konsep dan spesifikasi dari ketiga teknologi Big Data tersebut. Sayangnya . Google tidak pernah mempublikasikan teknologi Big Data ciptaannya dalam bentuk framework ataupun software yang siap pakai.

Apache Software Foundation (ASF) yang juga melihat peluang besar pada fenomena Big Data, telah berhasil mengimplementasikan konsep dan spesifikasi

teknis dari GFS, MapReduce dan Bigtable ke dalam bentuk framework yang dikembangkan dengan bahasa Java. Kemudian pengembangan tersebut dikelola dalam bentuk proyek *Open Source* sehingga dapat diakses dan digunakan oleh semua orang .

Apache mengembangkan sistem terdistribusi yang diberik nama Hadoop, yang juga dikenala sebagai Apache Hadoop.

Hadoop adalah sistem terdistribusi *open source* yang menerpakan programming model sederhana yang ditujukan untuk memproses data berukuran raksasa dalam suatu cluster komputer.

TINJAUAN PUSTAKA

2.1 HDFS

HDFS (Hadoop Distributed File System) sesuai dengan namanya, adalah file

terdistribusi pada Hadoop. Yang dimaksud dengan system file terdistribusi adalah file system yang menyimpan data tidak dalam satu hard disk drive (HDD) atau media penyimpanan lainnya, tetapi data dipecah-pecah dan disimpan tersebar dalam suatu *cluster* yang terdiri atas sejumlah computer, bisa hanya puluhan, ratusan atau bahkan ribuan computer.

Jika disederhanakan, HDFS dapat dianggap sebagai media penyimpanan pada cluster Hadoop. HDFS didesain dan dikembangkan berdasarkan konsep dan spesifikasi teknis dari GFS, yang telah dipublikasikan oleh Google.

2.2 Hadoop Map Reduce

Hadoop Map Reduce dikembangkan berdasarkan konsep dan spesifikasi teknik Google Map Reduce yang diimplementasikan dengan menggunakan bahasa pemrograman Java. Olehnya itu Hadoop Map Reduce memiliki programming model yang identic dan Googl Map Reduce. Demikian juga dengan arsitekturnya yang ukuivalen dengan Google MapReduce .Perbedaannya adalah Google Map Reduce dijalankan dalam cluster GFS, sedangkan Hadoop Map Reduce dioperasikan dalam cluster Hadoop yang sudah tentunya sudah dalam format HDFS.

2.3 Model Pemrograman Hadoop Map Reduce

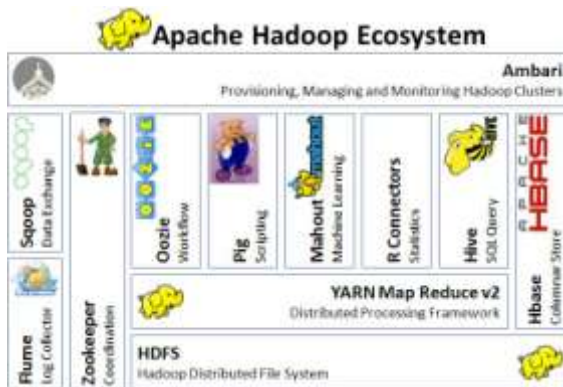
Hadoop Map Reduce memiliki programming model yang terdiri dari fase Map dan fase Reduce, terdapat juga fase Shuffle, namun demikian Shuffle tidak bersifat mutlak .

Pada fase Map, Mapper akan memproses blok-blok data dan menghasilkan intermediate key/value pair sebagai output. Kemudian pada fase shuffle, terdapat Comparator yang bertugas Menyortir berdasarkan key nya, dan partitioner yang bertugas untuk menentukan ke Reducer nomor berapa tiap intermediate key/value pair yang diterimanya mesti diserahkan. Selanjutnya, pada fase Reduce , value dari tiap-tiap key dari data yang telah di sortir tersebut akan dibaca secara berurutan untuk diproses sesuai dengan yang diprogramkan

2.4 Ekosistem Hadoop

Pada awalnya, Hadoop hanya terdiri atas dua framework, yaitu HDFS sebagai system file terdistribusi untuk menyimpan data dengan ukuran raksasa dan Map Reduce untuk memproses data raksasa tersebut secara terdistribusi dan parallel. Namun demikian saat ini Hadoop telah berkembang menjadi pondasi bagi frame work lain yang memiliki kemampuan untuk memproses data secara terdistribusi seperti halnya HBase, apalagi semenjak dirilisnya Hadoop 2, yaitu Hadoop yang telah dilengkapi dengan YARN yang

berfungsi sebagai operating System pada HDFS, Hadoop menjadi makin fleksibel dalam mewartahi aplikasi-aplikasi lain yang membutuhkan suatu system terdistribusi sebagai pondasinya. Olehnya itu Hadoop telah dianggap sebagai ekosistem bagi framework maupun aplikasi lain yang didesain untuk menjalankan fungsi tertentu.



Gambar 1. Hadoop Ecosystem

PEMBAHASAN

3.1 Instalasi Hadoop Mode Stand alone

Terlebih dahulu, akan dijelaskan langkah demi langkah tentang bagaimana menginstall dan menjalankan Hadoop Mode Standalone sekaligus menjalankan aplikasi *Wordcount* pada PC Windows. Sesungguhnya menggunakan PC Windows untuk menjalankan Hadoop adalah metoda yang lebih sedikit sulit disbanding megunakan platform Linux.

Tools yang digunakan yaitu :

1. Eclipse Java SE, bias di download di [Http://www.eclipse.org/downloads/](http://www.eclipse.org/downloads/)

2. Cygwin, dapatdidownload di <http://cygwin.com/install.html>
3. Hadoop-0.20.2.tar.gz file dapat di download dari <http://hadoop.apache.org/>
4. Ant library (ant.jar), dapat di download dari <http://ant.apache.org/bindownload.cgi> nama file apache-ant-1.8.4-bin.zip

Pertama silakan install Eclipse Java SE dan Cygwin sesuai prosedur. Setelah itu ikuti lahlangkah-langkah berikut untuk mengisntal/mengimpor Hadoopke Eclipse Java SE.

1. Buat New Java Project pada Eclipse dan beri nama hadoop-0.20.2
2. Import hadoop-0.20.2.tar.gz yang telah di download ke project hadoop-0.20.2 yang baru dibuat.
3. Copy ant.jar dari file apache dari file apache-ant-1.8.4-bin.zip yang telah diekstract dan paste pada folder lib dari project hadoop-0.20.2
4. Buka file hadoop-0.20.2/.classpath dan ganti isinya dengan code XML
Refresh project Hadoop-0.20.2. Caranya : klik kanan pada project hadoop-0,20,2 dan pilih “Refresh”.

Proses install Hadoop Standalone pada Eclipse Java SE telah selesai dan Hadoop telah siap untuk menjalankan aplikasi Map Reduce.

3.2 Uji Coba Sistem

Sekarang kita akan mencoba menjalankan satu contoh aplikasi Map Reduce yang dinamakan Word Count. Aplikasi Word Count berfungsi untuk menghitung jumlah atau frekuensi kemunculan tiap kata dalam satu atau sekumpulan text file yang akan kita simpan di dalam folder input. Langkah-langkahnya adalah sebagai berikut :

1. Buat New Java Project dengan nama WordCount. Klik Next pada New Java Project dan lakukan *setting* berikut :
 - a. Tambahkan project hadoop-0.20.2 pada tab Projects.
 - b. Tambahkan hadoop-0.20.2/hadoop-0.20.2/lib/ commons-cli-1.2.jar pada tab libraries.
 - c. Klik kanan pada project Word count > Properties.
 - d. Pada dialog Word Count klik java build path
2. Lakukan setting pada Run Configuration dengan langkah-langkah sebagai berikut:
 - a. Klik Run pada menu utama dan pilih Run Configuration.
 - b. Pada dialog Run Configuration klik tab environment>New
 - c. Pada dialog New Environment Variabel, set Name=Path, Value=c:\cygwin\bin pada kolom value

3. Buat File baru dengan nama log4j.properties dan letakkan pada folder src pada project Word Count. File ini berfungsi untuk menampilkan logs Hadoop pada console Eclipse.
4. Buat Class baru dengan nama Word Count.java dan simpan di folder src pada project Word Count.

Berikut ini source code yang kami coba dari web hadoop:

```
import java.io.IOException;
import java.util.StringTokenizer;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.
File Input Format;
import org.apache.hadoop.mapreduce.lib.output.
File Output Format;
public class Word Count {
public static class Tokenizer Mapper
extends Mapper<Object, Text, Text, Int
Writable>{
private final static Int Writable one = new Int
Writable(1);
private Text word = new Text();
```

```

public void map(Object key, Text value,
Context context) throws IO Exception,
InterruptedException {
String Tokenizeritr = new String
Tokenizer(value.toString());
while (itr.hasMoreTokens()) {
word.set(itr.nextToken());
context.write(word, one);
}
}
}

public static class Int Sum Reducer
extends Reducer<Text, Int Writable,Text, Int
Writable> {
private Int Writable result = new Int
Writable();
public void reduce(Text key,
Iterable<IntWritable> values,
Context context
) throws IO Exception,
InterruptedException {
int sum = 0;
for (Int Writableval : values) {
sum += val.get();
}
result.set(sum);
context.write(key, result);
}
}

```

```

public static void main(String[ ] args) throws
Exception {
Configuration conf = new Configuration();
Job job = Job.get Instance(conf, "word
count");
job.set Jar By Class(Word Count.class);
job.set Mapper Class(Tokenizer
Mapper.class);
job.set Combiner Class(IntSum
Reducer.class);
job.set Reducer Class(IntSum Reducer.class);
job.set Output Key Class(Text.class);
job.set Output Value Class(Int Writable.class);
File Input Format.add Input Path(job, new
Path(args[0]));
File Output Format.set Output Path(job, new
Path(args[1]));
System.exit(job.wait For Completion(true) ? 0
: 1);
}
}

```

5. Buat folder baru pada project WordCount dengan nama input (Word Count/input) dan tempatkan satu atau beberapa text file yang akan diproses dengan program Map Reduce Word Cunt.java
6. Klik menu Run pada main menu Eclipse dan pilih Run Configuration pada dialog Run Configuration klik tab Arguments dan tambahkan input

output pada program Arguments dan klik Apply. Kemudian klik Run. Demikian langkah-langkah yang harus dilakukan untuk menginstall dan menjalankan hadoop mode stand alone pada PC Windows menggunakan Eclipse Java SE.

KESIMPULAN

Implementasi *Hadoop Stand Alone* pada system operasi windows memang dapat dilakukan, dan sudah dapat menjalankan program word count, yang merupakan versi “hello word” untuk bahasa pemrograman yang lain. Walaupun lebih ribet karena diperlukan aplikasi lain seperti Cygwin.

DAFTAR PUSTAKA

Ant “Ant Library”

<http://ant.apache.org/bindownload.cgi>

Apache “Apache Hadoop for Windows Platform –CodeProject”

<http://www.codeproject.com>

Cygwin, “Cygwin for Windows”,

<http://cygwin.com/install.html>

Eclipse “Eclipse Java SE”,

<http://www.eclipse.org/downloads/>

Tariqmislam “Setting Up Hadoop 0.20.2 on Windows 7 With Cygwin

“<https://gist.github.com/tariqmislam/2159173>”

VladKorolev “Hadoop on Windows with Eclipse”<http://ebiquity.umbc.edu/Tutorials/Hadoop/00%20-%20Intro.html>

Wijaya. 2015. “Teknologi Big Data: SistemCanggihdibalik Google, Yahoo!, Facebook, IBM”. Google Play Book.