MAP-REDUCE ASSIGNMENT

Checking frequency of user login through particular IP address with input log file

**Mapper class :Mpr**

import java.io.IOException;

import java.util.regex.Matcher;

import java.util.regex.Pattern;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.LongWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Mapper;

public class Mpr extends Mapper<LongWritable, Text, Text, IntWritable> {

private static final IntWritable one=new IntWritable(1);

private Pattern patternIP=Pattern.compile("(?:([0-9]+[\\.][0-9]+[\\.][0-9]+[\\.][0-9]+))");

private Pattern patternURL=Pattern.compile("((https?|ftp|gopher|telnet|file):((//)|(\\\\))+[\\w\\d:#@%/;$()~\_?\\+-=\\\\\\.&]\*)");

private Text word=new Text();

public void map(LongWritable ikey, Text ivalue, Context context)

throws IOException, InterruptedException {

String url=null;

String iPAddress=null;

String[] entries=ivalue.toString().split("\r?\n");

for (String string : entries) {

Matcher matcherIP=patternIP.matcher(string);

Matcher matcherUrl=patternURL.matcher(string);

if(matcherIP.find())

{

iPAddress=matcherIP.group(1);

}

if(matcherUrl.find())

{

url=matcherUrl.group(1);

}

word.set(iPAddress+":"+url);

context.write(word,one);

}

}

}

**Reducer class: Redr**

import java.io.IOException;

import org.apache.hadoop.io.IntWritable;

import org.apache.hadoop.io.Text;

import org.apache.hadoop.mapreduce.Reducer;

public class Redr extends Reducer<Text, IntWritable, Text, IntWritable> {

public void reduce(Text \_key, Iterable<IntWritable> values, Context context)

throws IOException, InterruptedException {

// process values

int count=0;

for (IntWritable val : values) {

count += val.get();

}

context.write(\_key, new IntWritable(count));

}

}

**Driver class: Dri**

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.lib.input.FileInputFormat;

import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;

public class Dri {

public static void main(String[] args) throws Exception {

Configuration conf = new Configuration();

Job job = Job.getInstance(conf, "College");

job.setJarByClass(Dri.class);

job.setMapperClass(Mpr.class);

job.setReducerClass(Redr.class);

// TODO: specify output types

job.setOutputKeyClass(Text.class);

job.setOutputValueClass(IntWritable.class);

// TODO: specify input and output DIRECTORIES (not files)

FileInputFormat.setInputPaths(job, new Path(args[1]));

FileOutputFormat.setOutputPath(job, new Path(args[2]));

if (!job.waitForCompletion(true))

return;

}

}

**Input file : log.txt**

1 4 0 172.16.3.101 Sat Mar 14 13:09:20 IST 2015 Login

1 4 0 172.16.2.86 Sat Mar 14 15:07:14 IST 2015 Login

1 4 0 172.16.2.86 Sat Mar 14 15:10:48 IST 2015 Login

3 4 0 172.16.2.86 Sat Mar 14 15:12:07 IST 2015 Login

3 5 0 172.16.2.86 Sat Mar 14 15:13:08 IST 2015 USER-3/hadoop-2.6.0.tar.gz

3 4 0 172.16.2.86 Sat Mar 14 15:13:16 IST 2015 Login

4 4 0 172.16.2.86 Sat Mar 14 15:16:46 IST 2015 Login

4 5 0 172.16.2.86 Sat Mar 14 15:17:02 IST 2015 USER-4/tokeHarami

3 4 0 172.16.2.86 Sat Mar 14 15:17:25 IST 2015 Login

4 4 0 172.16.2.86 Sat Mar 14 15:19:02 IST 2015 Login

1 4 0 172.16.2.86 Sat Mar 14 15:20:42 IST 2015 Login

1 4 0 172.16.2.86 Sat Mar 14 15:25:24 IST 2015 Login

3 4 0 172.16.2.86 Sat Mar 14 15:25:44 IST 2015 Login

3 3 0 172.16.2.86 Sat Mar 14 15:26:24 IST 2015 ASD

3 4 0 172.16.2.86 Sat Mar 14 15:30:23 IST 2015 Login

3 4 0 172.16.2.86 Sat Mar 14 15:30:59 IST 2015 Login

3 4 0 172.16.2.86 Sat Mar 14 15:31:24 IST 2015 Login

5 4 0 172.16.2.110 Mon Mar 16 10:41:06 IST 2015 Login

4 4 0 172.16.2.110 Mon Mar 16 10:45:44 IST 2015 Login

4 5 0 172.16.2.110 Mon Mar 16 10:46:14 IST 2015 USER-4/StarToken-NG.msi

4 4 0 172.16.2.110 Mon Mar 16 10:46:38 IST 2015 Login

5 4 0 172.16.2.110 Mon Mar 16 10:53:57 IST 2015 Login

5 3 0 172.16.2.110 Mon Mar 16 10:54:03 IST 2015 CHETAN

5 3 0 172.16.2.110 Mon Mar 16 10:56:11 IST 2015 ABCD

5 5 0 172.16.2.110 Mon Mar 16 11:00:44 IST 2015 USER-5/New Microsoft Office Excel Worksheet (2).xlsx

5 5 0 172.16.2.110 Mon Mar 16 11:01:25 IST 2015 USER-5/New Microsoft Office Excel Worksheet (2).xlsx

5 5 0 172.16.2.110 Mon Mar 16 11:01:35 IST 2015 USER-5/Class room pc status ( RAJU SM).docx

5 5 0 172.16.2.110 Mon Mar 16 11:01:46 IST 2015 USER-5/Class room pc status ( RAJU SM).docx

5 5 0 172.16.2.110 Mon Mar 16 11:02:14 IST 2015 USER-5/Hadoop-Introduction.pdf

5 3 0 172.16.2.110 Mon Mar 16 11:09:26 IST 2015 ABCD

6 4 0 172.16.2.43 Mon Mar 16 12:38:04 IST 2015 Login

6 5 0 172.16.2.43 Mon Mar 16 12:39:41 IST 2015 USER-6/cal7.lisp

4 4 0 172.16.2.102 Tue Mar 17 11:13:51 IST 2015 Login

4 1 0 172.16.2.102 Tue Mar 17 11:14:39 IST 2015 USER-4/StarToken-NG.msi

4 4 0 172.16.2.102 Tue Mar 17 11:15:58 IST 2015 Login

4 4 0 172.16.2.102 Tue Mar 17 11:17:04 IST 2015 Login

3 4 0 172.16.2.33 Fri Jun 19 11:03:04 IST 2015 Login

3 4 0 172.16.2.33 Fri Jun 19 11:03:40 IST 2015 Login

3 4 0 172.16.2.33 Fri Jun 19 11:04:33 IST 2015 Login

3 4 0 172.16.2.33 Fri Jun 19 11:04:54 IST 2015 Login

3 4 0 172.16.2.33 Fri Jun 19 11:08:17 IST 2015 Login

3 1 0 172.16.2.33 Fri Jun 19 11:08:37 IST 2015 USER-3/SMART TEST.docx

3 3 0 172.16.2.33 Fri Jun 19 11:09:10 IST 2015 C

3 1 0 172.16.2.33 Fri Jun 19 11:09:16 IST 2015 /SharedData/4/C programming 1st\_edition.pdf

3 3 0 172.16.2.33 Fri Jun 19 11:13:10 IST 2015 CSYLLABUS

3 3 0 172.16.2.33 Fri Jun 19 11:13:14 IST 2015 YLLABUS

3 1 0 172.16.2.33 Fri Jun 19 11:13:17 IST 2015 /SharedData/2/Syllabus TE Computer Engineering 2012Course.pdf

3 3 0 172.16.2.33 Fri Jun 19 11:14:44 IST 2015 C

3 1 0 172.16.2.33 Fri Jun 19 11:14:58 IST 2015 /SharedData/4/C programming 1st\_edition.pdf

3 3 0 172.16.2.33 Fri Jun 19 11:15:43 IST 2015 YLLABUS

3 3 0 172.16.2.33 Fri Jun 19 11:15:49 IST 2015 SYLLABUS

3 3 0 172.16.2.33 Fri Jun 19 11:15:53 IST 2015 BUS

5 3 0 172.16.2.110 Wed Jul 08 15:15:51 IST 2015 DEMO

5 3 0 172.16.2.110 Wed Jul 08 15:15:51 IST 2015 DEMO

5 3 0 172.16.2.110 Wed Jul 08 15:15:51 IST 2015 DEMO

5 3 0 172.16.2.110 Wed Jul 08 15:15:51 IST 2015 DEMO

5 3 0 172.16.2.110 Wed Jul 08 15:15:51 IST 2015 DEMO

5 3 0 172.16.2.110 Wed Jul 08 15:15:52 IST 2015 DEMO

5 4 0 172.16.2.110 Wed Jul 08 15:21:45 IST 2015 Login

5 4 0 172.16.2.110 Wed Jul 08 15:24:44 IST 2015 Login

5 3 0 172.16.2.110 Wed Jul 08 15:27:15 IST 2015 A

3 4 0 172.16.3.100 Wed Jul 08 15:32:05 IST 2015 Login

**Output file: part-r-00000**

172.16.1.124:nul l 2

172.16.1.125:null 15

172.16.1.144:nul l 7

172.16.1.177:null 17

172.16.1.178:null 9

172.16.1.182:null 3

172.16.1.183:null 4

172.16.1.184:null 5

172.16.1.185:null 9

172.16.1.186:null 7

172.16.1.187:null 12

172.16.1.188:null 45

172.16.1.189:null 16

172.16.1.190:null 13

172.16.1.193:null 55

172.16.1.195:null 23

172.16.1.198:null 18

172.16.1.199:null 1

172.16.1.209:null 2

172.16.1.217:null 9

172.16.1.222:null 9

172.16.2.58:null 4

172.16.2.59:null 7

172.16.2.60:null 8

172.16.2.62:null 4

172.16.2.64:null 1

172.16.2.69:null 1

172.16.2.86:null 79

172.16.2.87:null 8

172.16.2.88:null 51

172.16.2.89:null 5

172.16.2.8:null 1

172.16.2.90:null 21

172.16.2.92:null 46

172.16.2.94:null 17

172.16.2.95:null 5

172.16.2.96:null 11

172.16.2.97:null 23

172.16.2.99:null 14

172.16.3.100:null 3

172.16.3.101:null 187

172.16.3.104:null 12

172.16.3.105:null 6

172.16.3.108:null 1

Terminal commands to start hadoop

***root@ccoew-desktop:/usr/local/hadoop/etc/hadoop# start-dfs.sh***

Starting namenodes on [localhost]

localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-root-namenode-ccoew-desktop.out

localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-root-datanode-ccoew-desktop.out

Starting secondary namenodes [0.0.0.0]

The authenticity of host '0.0.0.0 (0.0.0.0)' can't be established.

ECDSA key fingerprint is fc:7a:18:cf:cc:46:85:73:c0:1b:5e:fc:ac:bc:c8:01.

**Are you sure you want to continue connecting (yes/no)? yes**

0.0.0.0: Warning: Permanently added '0.0.0.0' (ECDSA) to the list of known hosts.

0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-root-secondarynamenode-ccoew-desktop.out

***root@ccoew-desktop:/usr/local/hadoop/etc/hadoop# start-yarn.sh***

starting yarn daemons

starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-root-resourcemanager-ccoew- desktop.out

***root@ccoew-desktop:/usr/local/hadoop/etc/hadoop# jps***

3559 SecondaryNameNode

4099 Jps

3388 DataNode

3243 NameNode

3699 ResourceManager

3816 NodeManager

***root@ccoew-desktop:/usr/local/hadoop/etc/hadoop#***

Terminal commands for mapreduce :

***root@ccoew-desktop:/home/ccoew# hdfs dfs -mkdir /ankita***

***root@ccoew-desktop:/home/ccoew# hdfs dfs -put /home/ccoew/log.txt /ankita***

***root@ccoew-desktop:/home/ccoew# hadoop jar /home/ccoew/p1.jar Dri /ankita/log.txt /ankita/op***

16/09/23 16:53:54 INFO Configuration.deprecation: session.id is deprecated. Instead, use dfs.metrics.session-id

16/09/23 16:53:54 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=

16/09/23 16:53:54 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed. Implement the Tool interface and execute your application with ToolRunner to remedy this.

16/09/23 16:53:54 INFO input.FileInputFormat: Total input paths to process : 1

16/09/23 16:53:54 INFO mapreduce.JobSubmitter: number of splits:1

16/09/23 16:53:54 INFO mapreduce.JobSubmitter: Submitting tokens for job: job\_local949013430\_0001

16/09/23 16:53:55 INFO mapreduce.Job: The url to track the job: http://localhost:8080/

16/09/23 16:53:55 INFO mapreduce.Job: Running job: job\_local949013430\_0001

16/09/23 16:53:55 INFO mapred.LocalJobRunner: OutputCommitter set in config null

16/09/23 16:53:55 INFO mapred.LocalJobRunner: OutputCommitter is org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter

16/09/23 16:53:55 INFO mapred.LocalJobRunner: Waiting for map tasks

16/09/23 16:53:55 INFO mapred.LocalJobRunner: Starting task: attempt\_local949013430\_0001\_m\_000000\_0

16/09/23 16:53:55 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]

16/09/23 16:53:55 INFO mapred.MapTask: Processing split: hdfs://localhost:9000/ankita/log.txt:0+124829

16/09/23 16:53:55 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)

16/09/23 16:53:55 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100

16/09/23 16:53:55 INFO mapred.MapTask: soft limit at 83886080

16/09/23 16:53:55 INFO mapred.MapTask: bufstart = 0; bufvoid = 104857600

16/09/23 16:53:55 INFO mapred.MapTask: kvstart = 26214396; length = 6553600

16/09/23 16:53:55 INFO mapred.MapTask: Map output collector class = org.apache.hadoop.mapred.MapTask$MapOutputBuffer

16/09/23 16:53:55 INFO mapred.LocalJobRunner:

16/09/23 16:53:55 INFO mapred.MapTask: Starting flush of map output

16/09/23 16:53:55 INFO mapred.MapTask: Spilling map output

16/09/23 16:53:55 INFO mapred.MapTask: bufstart = 0; bufend = 40241; bufvoid = 104857600

16/09/23 16:53:55 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26206908(104827632); length = 7489/6553600

16/09/23 16:53:55 INFO mapred.MapTask: Finished spill 0

16/09/23 16:53:55 INFO mapred.Task: Task:attempt\_local949013430\_0001\_m\_000000\_0 is done. And is in the process of committing

16/09/23 16:53:55 INFO mapred.LocalJobRunner: map

16/09/23 16:53:55 INFO mapred.Task: Task 'attempt\_local949013430\_0001\_m\_000000\_0' done.

16/09/23 16:53:55 INFO mapred.LocalJobRunner: Finishing task: attempt\_local949013430\_0001\_m\_000000\_0

16/09/23 16:53:55 INFO mapred.LocalJobRunner: map task executor complete.

16/09/23 16:53:55 INFO mapred.LocalJobRunner: Waiting for reduce tasks

16/09/23 16:53:55 INFO mapred.LocalJobRunner: Starting task: attempt\_local949013430\_0001\_r\_000000\_0

16/09/23 16:53:55 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]

16/09/23 16:53:55 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin: org.apache.hadoop.mapreduce.task.reduce.Shuffle@7bfbde18

16/09/23 16:53:55 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=333971456, maxSingleShuffleLimit=83492864, mergeThreshold=220421168, ioSortFactor=10, memToMemMergeOutputsThreshold=10

16/09/23 16:53:55 INFO reduce.EventFetcher: attempt\_local949013430\_0001\_r\_000000\_0 Thread started: EventFetcher for fetching Map Completion Events

16/09/23 16:53:55 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map attempt\_local949013430\_0001\_m\_000000\_0 decomp: 43989 len: 43993 to MEMORY

16/09/23 16:53:55 INFO reduce.InMemoryMapOutput: Read 43989 bytes from map-output for attempt\_local949013430\_0001\_m\_000000\_0

16/09/23 16:53:55 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 43989, inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->43989

16/09/23 16:53:55 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning

16/09/23 16:53:55 INFO mapred.LocalJobRunner: 1 / 1 copied.

16/09/23 16:53:55 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0 on-disk map-outputs

16/09/23 16:53:55 INFO mapred.Merger: Merging 1 sorted segments

16/09/23 16:53:55 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 43969 bytes

16/09/23 16:53:55 INFO reduce.MergeManagerImpl: Merged 1 segments, 43989 bytes to disk to satisfy reduce memory limit

16/09/23 16:53:55 INFO reduce.MergeManagerImpl: Merging 1 files, 43993 bytes from disk

16/09/23 16:53:55 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce

16/09/23 16:53:55 INFO mapred.Merger: Merging 1 sorted segments

16/09/23 16:53:55 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size: 43969 bytes

16/09/23 16:53:55 INFO mapred.LocalJobRunner: 1 / 1 copied.

16/09/23 16:53:55 INFO Configuration.deprecation: mapred.skip.on is deprecated. Instead, use mapreduce.job.skiprecords

16/09/23 16:53:56 INFO mapreduce.Job: Job job\_local949013430\_0001 running in uber mode : false

16/09/23 16:53:56 INFO mapreduce.Job: map 100% reduce 0%

16/09/23 16:53:56 INFO mapred.Task: Task:attempt\_local949013430\_0001\_r\_000000\_0 is done. And is in the process of committing

16/09/23 16:53:56 INFO mapred.LocalJobRunner: 1 / 1 copied.

16/09/23 16:53:56 INFO mapred.Task: Task attempt\_local949013430\_0001\_r\_000000\_0 is allowed to commit now

16/09/23 16:53:56 INFO output.FileOutputCommitter: Saved output of task 'attempt\_local949013430\_0001\_r\_000000\_0' to hdfs://localhost:9000/ankita/op/\_temporary/0/task\_local949013430\_0001\_r\_000000

16/09/23 16:53:56 INFO mapred.LocalJobRunner: reduce > reduce

16/09/23 16:53:56 INFO mapred.Task: Task 'attempt\_local949013430\_0001\_r\_000000\_0' done.

16/09/23 16:53:56 INFO mapred.LocalJobRunner: Finishing task: attempt\_local949013430\_0001\_r\_000000\_0

16/09/23 16:53:56 INFO mapred.LocalJobRunner: reduce task executor complete.

16/09/23 16:53:57 INFO mapreduce.Job: map 100% reduce 100%

16/09/23 16:53:57 INFO mapreduce.Job: Job job\_local949013430\_0001 completed successfully

16/09/23 16:53:57 INFO mapreduce.Job: Counters: 38

File System Counters

FILE: Number of bytes read=99538

FILE: Number of bytes written=642043

FILE: Number of read operations=0

FILE: Number of large read operations=0

FILE: Number of write operations=0

HDFS: Number of bytes read=249658

HDFS: Number of bytes written=1689

HDFS: Number of read operations=13

HDFS: Number of large read operations=0

HDFS: Number of write operations=4

Map-Reduce Framework

Map input records=1873

Map output records=1873

Map output bytes=40241

Map output materialized bytes=43993

Input split bytes=101

Combine input records=0

Combine output records=0

Reduce input groups=84

Reduce shuffle bytes=43993

Reduce input records=1873

Reduce output records=84

Spilled Records=3746

Shuffled Maps =1

Failed Shuffles=0

Merged Map outputs=1

GC time elapsed (ms)=28

CPU time spent (ms)=0

Physical memory (bytes) snapshot=0

Virtual memory (bytes) snapshot=0

Total committed heap usage (bytes)=496500736

Shuffle Errors

BAD\_ID=0

CONNECTION=0

IO\_ERROR=0

WRONG\_LENGTH=0

WRONG\_MAP=0

WRONG\_REDUCE=0

File Input Format Counters

Bytes Read=124829

File Output Format Counters

Bytes Written=1689

***root@ccoew-desktop:/home/ccoew#***