

MAPPER

```
package bin;

import java.io.IOException;
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 AverageMapper extends Mapper<LongWritable, Text, Text, IntWritable> {
    public static final int MISSING = 9999;

    public void map(LongWritable key, Text value, Mapper<LongWritable, Text, Text,
IntWritable>.Context context) throws IOException, InterruptedException {
        int sumsquare;
        String line = value.toString();
        String year = line.substring(15, 19);
        if (line.charAt(87) == '*') {
            sumsquare= Integer.parseInt(line.substring(1,2,3,4,5));
        } else {
            sumsquare= Integer.parseInt(line.substring(1,4,7,8,9));
        }
        String quality = line.substring(92, 93);
        if (sumsquare!= 9999 && quality.matches("[01459]"))
            context.write(new Text(year), new IntWritable(sumsquare));
    }
}
```

REDUCER

```
package bin;

import java.io.IOException;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Reducer;

public class AverageReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
    public void reduce(Text key, Iterable<IntWritable> values, Reducer<Text, IntWritable, Text,
IntWritable>.Context context) throws IOException, InterruptedException {
        int sum_square= 0;
```

```

int count = 0;
for (IntWritable value : values) {
    sum_square+= value.get();

    count++;
}
context.write(key, new IntWritable(count*count));
}
}

```

DRIVER

```

package bin;

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 AverageDriver {
    public static void main(String[] args) throws Exception {
        if (args.length != 2) {
            System.err.println("Please Enter the input and output parameters");
            System.exit(-1);
        }
        @SuppressWarnings("deprecation")
        Job job = new Job();
        job.setJarByClass(AverageDriver.class);
        job.setJobName("sum of square");
        FileInputFormat.addInputPath(job, new Path(args[0]));
        FileOutputFormat.setOutputPath(job, new Path(args[1]));
        job.setMapperClass(SumMapper.class);
        job.setReducerClass(SumReducer.class);
        job.setOutputKeyClass(Text.class);
        job.setOutputValueClass(IntWritable.class);
        System.exit(job.waitForCompletion(true) ? 0 : 1);
    }
}

```

```

hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
hduser@localhost's password:
localhost: starting namenode, logging to /usr/local/hadoop/logs/hadoop-hduser-namenode-bmsce-Precision-T1700.out
hduser@localhost's password:
localhost: starting datanode, logging to /usr/local/hadoop/logs/hadoop-hduser-datanode-bmsce-Precision-T1700.out
Starting secondary namenodes [0.0.0.0]
hduser@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /usr/local/hadoop/logs/hadoop-hduser-secondarynamenode-bmsce-Precision-T1700.out
starting yarn daemons
starting resourcemanager, logging to /usr/local/hadoop/logs/yarn-hduser-resourcemanager-bmsce-Precision-T1700.out
hduser@localhost's password:
localhost: starting nodemanager, logging to /usr/local/hadoop/logs/yarn-hduser-nodemanager-bmsce-Precision-T1700.out
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ jps
11872 org.eclipse.equinox.launcher_1.5.600.v20191014-2022.jar
15424 NameNode
16421 Jps
15594 DataNode
16299 NodeManager
15967 ResourceManager
15807 SecondaryNameNode
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hdfs dfs -ls /
Found 17 items
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:26 /FFF
drwxr-xr-x - hduser supergroup 0 2022-06-04 10:28 /LLL
drwxr-xr-x - hduser supergroup 0 2022-06-06 12:14 /abc
drwxr-xr-x - hduser supergroup 0 2022-06-15 15:24 /cs228
drwxr-xr-x - hduser supergroup 0 2022-06-06 12:23 /hadoop_lab
drwxr-xr-x - hduser supergroup 0 2022-06-06 14:57 /harshita
drwxr-xr-x - hduser supergroup 0 2022-05-31 09:42 /hello
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:14 /nnn
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:02 /output
drwxr-xr-x - hduser supergroup 0 2019-10-24 12:16 /output1
drwxr-xr-x - hduser supergroup 0 2022-06-20 15:44 /rgs
drwxr-xr-x - hduser supergroup 0 2022-06-03 11:56 /sam
drwxr-xr-x - hduser supergroup 0 2022-06-03 12:39 /sampreeth
drwxr-xr-x - hduser supergroup 0 2019-10-24 11:01 /sid
drwxrwxr-x - hduser supergroup 0 2019-08-01 16:19 /tmp

```

```

hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hdfs dfs -ls /cs228
Found 6 items
-rw-r--r-- 1 hduser supergroup      0 2022-06-01 14:35 /cs228/copied_foo.txt
-rw-r--r-- 1 hduser supergroup      0 2022-06-01 14:23 /cs228/foo.txt
drwxr-xr-x - hduser supergroup      0 2022-06-15 15:24 /cs228/output_wc
-rw-r--r-- 1 hduser supergroup    26 2022-06-01 14:27 /cs228/renamed_cfl.txt
-rw-r--r-- 1 hduser supergroup    26 2022-06-01 14:26 /cs228/test_cs228.txt
drwxr-xr-x - hduser supergroup      0 2022-06-15 15:04 /cs228/wordcount
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hdfs dfs -mkdir /cs228/avgtemp
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hdfs dfs -put ~/Downloads/1901.txt
/cs228/avgtemp/1901.txt
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hadoop jar AvgTemp.jar /cs228/avgtemp/1901.txt
/cs228/avgtemp/output
Exception in thread "main" java.lang.ClassNotFoundException: /cs228/avgtemp/1901/txt
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:348)
    at org.apache.hadoop.util.RunJar.run(RunJar.java:214)
    at org.apache.hadoop.util.RunJar.main(RunJar.java:136)
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hadoop jar bin.AvgTemp.jar
/cs228/avgtemp/1901.txt /cs228/avgtemp/output
Not a valid JAR: /home/hduser/eclipse-workspace/cs228/bin.AvgTemp.jar
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hadoop jar AvgTemp.jar bin.AverageDriver
/cs228/avgtemp/1901.txt /cs228/avgtemp/output
22/06/22 15:05:56 INFO Configuration.deprecation: session.id is deprecated. Instead, use
dfs.metrics.session-id
22/06/22 15:05:56 INFO jvm.JvmMetrics: Initializing JVM Metrics with processName=JobTracker, sessionId=
22/06/22 15:05:56 WARN mapreduce.JobSubmitter: Hadoop command-line option parsing not performed.
Implement the Tool interface and execute your application with ToolRunner to remedy this.
22/06/22 15:05:56 INFO input.FileInputFormat: Total input paths to process : 1
22/06/22 15:05:56 INFO mapreduce.JobSubmitter: number of splits:1
22/06/22 15:05:56 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_local255818709_0001
22/06/22 15:05:56 INFO mapreduce.Job: The url to track the job: http://localhost:8080/
22/06/22 15:05:56 INFO mapreduce.Job: Running job: job_local255818709_0001
22/06/22 15:05:56 INFO mapred.LocalJobRunner: OutputCommitter set in config null
22/06/22 15:05:56 INFO mapred.LocalJobRunner: OutputCommitter is
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter
22/06/22 15:05:56 INFO mapred.LocalJobRunner: Waiting for map tasks
22/06/22 15:05:56 INFO mapred.LocalJobRunner: Starting task: attempt_local255818709_0001_m_000000_0
22/06/22 15:05:56 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
22/06/22 15:05:56 INFO mapred.MapTask: Processing split:
hdfs://localhost:54310/cs228/avgtemp/1901.txt:0+888190
22/06/22 15:05:56 INFO mapred.MapTask: (EQUATOR) 0 kvi 26214396(104857584)
22/06/22 15:05:56 INFO mapred.MapTask: mapreduce.task.io.sort.mb: 100
22/06/22 15:05:56 INFO mapred.MapTask: soft limit at 83886080

```

```
22/06/22 15:05:56 INFO mapred.MapTask: Map output collector class =
org.apache.hadoop.mapred.MapTask$MapOutputBuffer
22/06/22 15:05:56 INFO mapred.LocalJobRunner:
22/06/22 15:05:56 INFO mapred.MapTask: Starting flush of map output
22/06/22 15:05:56 INFO mapred.MapTask: Spilling map output
22/06/22 15:05:56 INFO mapred.MapTask: bufstart = 0; bufend = 59076; buvoid = 104857600
22/06/22 15:05:56 INFO mapred.MapTask: kvstart = 26214396(104857584); kvend = 26188144(104752576);
length = 26253/6553600
22/06/22 15:05:56 INFO mapred.MapTask: Finished spill 0
22/06/22 15:05:56 INFO mapred.Task: Task:attempt_local255818709_0001_m_000000_0 is done. And is in the
process of committing
22/06/22 15:05:56 INFO mapred.LocalJobRunner: map
22/06/22 15:05:56 INFO mapred.Task: Task 'attempt_local255818709_0001_m_000000_0' done.
22/06/22 15:05:56 INFO mapred.LocalJobRunner: Finishing task: attempt_local255818709_0001_m_000000_0
22/06/22 15:05:56 INFO mapred.LocalJobRunner: map task executor complete.
22/06/22 15:05:56 INFO mapred.LocalJobRunner: Waiting for reduce tasks
22/06/22 15:05:56 INFO mapred.LocalJobRunner: Starting task: attempt_local255818709_0001_r_000000_0
22/06/22 15:05:56 INFO mapred.Task: Using ResourceCalculatorProcessTree : [ ]
22/06/22 15:05:56 INFO mapred.ReduceTask: Using ShuffleConsumerPlugin:
org.apache.hadoop.mapreduce.task.reduce.Shuffle@e512f99
22/06/22 15:05:56 INFO reduce.MergeManagerImpl: MergerManager: memoryLimit=349752512,
maxSingleShuffleLimit=87438128, mergeThreshold=230836672, ioSortFactor=10,
memToMemMergeOutputsThreshold=10
22/06/22 15:05:56 INFO reduce.EventFetcher: attempt_local255818709_0001_r_000000_0 Thread started:
EventFetcher for fetching Map Completion Events
22/06/22 15:05:56 INFO reduce.LocalFetcher: localfetcher#1 about to shuffle output of map
attempt_local255818709_0001_m_000000_0 decomp: 72206 len: 72210 to MEMORY
22/06/22 15:05:56 INFO reduce.InMemoryMapOutput: Read 72206 bytes from map-output for
attempt_local255818709_0001_m_000000_0
22/06/22 15:05:56 INFO reduce.MergeManagerImpl: closeInMemoryFile -> map-output of size: 72206,
inMemoryMapOutputs.size() -> 1, commitMemory -> 0, usedMemory ->72206
22/06/22 15:05:56 INFO reduce.EventFetcher: EventFetcher is interrupted.. Returning
22/06/22 15:05:56 INFO mapred.LocalJobRunner: 1 / 1 copied.
22/06/22 15:05:56 INFO reduce.MergeManagerImpl: finalMerge called with 1 in-memory map-outputs and 0
on-disk map-outputs
22/06/22 15:05:56 INFO mapred.Merger: Merging 1 sorted segments
22/06/22 15:05:56 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size:
72199 bytes
22/06/22 15:05:56 INFO reduce.MergeManagerImpl: Merged 1 segments, 72206 bytes to disk to satisfy
reduce memory limit
22/06/22 15:05:56 INFO reduce.MergeManagerImpl: Merging 1 files, 72210 bytes from disk
22/06/22 15:05:56 INFO reduce.MergeManagerImpl: Merging 0 segments, 0 bytes from memory into reduce
22/06/22 15:05:56 INFO mapred.Merger: Merging 1 sorted segments
22/06/22 15:05:56 INFO mapred.Merger: Down to the last merge-pass, with 1 segments left of total size:
```

```
mapreduce.job.skiprecords
22/06/22 15:05:56 INFO mapred.Task: Task:attempt_local255818709_0001_r_000000_0 is done. And is in the
process of committing
22/06/22 15:05:56 INFO mapred.LocalJobRunner: 1 / 1 copied.
22/06/22 15:05:56 INFO mapred.Task: Task attempt_local255818709_0001_r_000000_0 is allowed to commit
now
22/06/22 15:05:57 INFO output.FileOutputCommitter: Saved output of task
'attempt_local255818709_0001_r_000000_0' to
hdfs://localhost:54310/cs228/avgtemp/output/_temporary/0/task_local255818709_0001_r_000000
22/06/22 15:05:57 INFO mapred.LocalJobRunner: reduce > reduce
22/06/22 15:05:57 INFO mapred.Task: Task 'attempt_local255818709_0001_r_000000_0' done.
22/06/22 15:05:57 INFO mapred.LocalJobRunner: Finishing task: attempt_local255818709_0001_r_000000_0
22/06/22 15:05:57 INFO mapred.LocalJobRunner: reduce task executor complete.
22/06/22 15:05:57 INFO mapreduce.Job: Job job_local255818709_0001 running in uber mode : false
22/06/22 15:05:57 INFO mapreduce.Job: map 100% reduce 100%
22/06/22 15:05:57 INFO mapreduce.Job: Job job_local255818709_0001 completed successfully
22/06/22 15:05:57 INFO mapreduce.Job: Counters: 38
  File System Counters
    FILE: Number of bytes read=153156
    FILE: Number of bytes written=722972
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=1776380
    HDFS: Number of bytes written=8
    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=6565
    Map output records=6564
    Map output bytes=59076
    Map output materialized bytes=72210
    Input split bytes=110
    Combine input records=0
    Combine output records=0
    Reduce input groups=1
    Reduce shuffle bytes=72210
    Reduce input records=6564
    Reduce output records=1
    Spilled Records=13128
    Shuffled Maps =1
    Failed Shuffles=0
    Merged Map outputs=1
```

```

HDFS: Number of large read operations=0
HDFS: Number of write operations=4
Map-Reduce Framework
  Map input records=6565
  Map output records=6564
  Map output bytes=59076
  Map output materialized bytes=72210
  Input split bytes=110
  Combine input records=0
  Combine output records=0
  Reduce input groups=1
  Reduce shuffle bytes=72210
  Reduce input records=6564
  Reduce output records=1
  Spilled Records=13128
  Shuffled Maps =1
  Failed Shuffles=0
  Merged Map outputs=1
  GC time elapsed (ms)=60
  CPU time spent (ms)=0
  Physical memory (bytes) snapshot=0
  Virtual memory (bytes) snapshot=0
  Total committed heap usage (bytes)=999292928
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=888190
File Output Format Counters
  Bytes Written=8
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hdfs dfs -ls /cs228/avgtemp/output
Found 2 items
-rw-r--r-- 1 hduser supergroup 0 2022-06-22 15:05 /cs228/avgtemp/output/_SUCCESS
-rw-r--r-- 1 hduser supergroup 8 2022-06-22 15:05 /cs228/avgtemp/output/part-r-000000
hduser@bmsce-Precision-T1700:~/eclipse-workspace/cs228$ hdfs dfs -cat /cs228/avgtemp/output/part-r-000000
1901 46

```