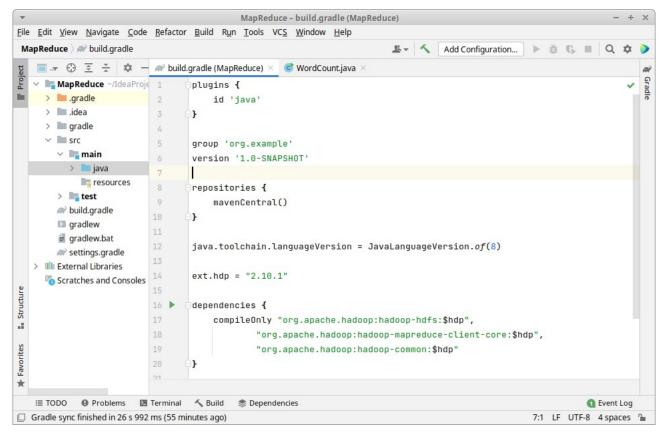
MAPREDUCE ON JAVA OTKPЫBAEM INTELLIJ IDEA CO3ДAEM JAVA ПРОЕКТ НА GRADLE WORD COUNT:

B build.gradle добавим зависимости хадупа:



В src/main/java создадим WordCount.java:

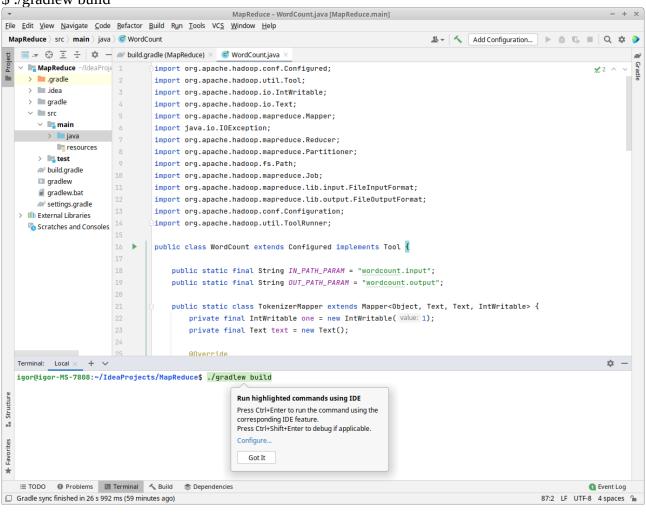
```
MapReduce - WordCount.java [MapReduce.main]
<u>F</u>ile <u>E</u>dit <u>V</u>iew <u>N</u>avigate <u>C</u>ode <u>R</u>efactor <u>B</u>uild <u>Run <u>T</u>ools VC<u>S <u>W</u>indow <u>H</u>elp</u></u>
MapReduce > src > main > java > © WordCount
                                                                                                           ♣ - 🔨 Add Configuration... 🕨 🇯 🖏 🗏 Q 🌣 🕨
   i → ⊕ ∑ 🛣 🌣 — 🔊 build.gradle (MapReduce) × 💣 WordCount.java
   ✓ MapReduce ~/IdeaProje 9
                                    import org.apache.hadoop.fs.Path;
                                                                                                                                                   *2 ^ v
    > ____ .gradle
                                   import org.apache.hadoop.mapreduce.Job;
     > 🗎 .idea
                                   import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
     > I gradle
     ∨ 🗎 src
                                   import org.apache.hadoop.conf.Configuration;
        v 📭 main
                                   import org.apache.hadoop.util.ToolRunner;
       > igava 15
            resources
                           16
                                   public class WordCount extends Configured implements Tool {
        > lightest
        w build.gradle
                                         public static final String IN_PATH_PARAM = "wordcount.input";
        ■ gradlew
                                         public static final String OUT_PATH_PARAM = "wordcount.output";
        gradlew.bat
        e settings.gradle
                                        public static class TokenizerMapper extends Mapper<Object, Text, Text, IntWritable> {
   > III External Libraries
                                            private final IntWritable one = new IntWritable( value: 1);
     Scratches and Consoles
                                             private final Text text = new Text();
                                             @Override
                           26 0 @
                                             public void map(Object key, Text value, Context context)
                                                     throws IOException, InterruptedException {
                           28
                                                 String line = value.toString().toLowerCase();
                                                 String[] words = line.split( regex: "\\s");
                                                 for (String word: words) {
                                                    text.set(word):
                                                     context.write(text, one);
                                             1
                                        }
                                         public static class IntSumReducer extends Reducer<Text, IntWritable, Text, IntWritable> {
                                             private final IntWritable result = new IntWritable();
                           40 01 @
                                             public void reduce(Text key, Iterable<IntWritable> values, Context context)
                                                     throws IOException, InterruptedException {
                                                  int sum = 0;
                                                 for (IntWritable val : values) {
                                                   sum += val.get();
                                                 result.set(sum);
                                                 context.write(key, result);
                                         public static class AlphabetPartitioner extends Partitioner<Text, IntWritable> {
                           54 🐠
                                             public int getPartition(Text key, IntWritable value, int numPartitions) {
                                                if (numPartitions == 1) {
                                                    return 0; // all in one
                                                 return (key.getLength() > 2) ? 1 : 0;
                                             }
                                        1
                                         public int run(String[] args) throws Exception {
                           64 1
                                             Job job = Job.getInstance(getConf(), jobName: "Word Count");
                                             job.setJarByClass(WordCount.class);
                                             job.setMapperClass(TokenizerMapper.class);
                                             job.setReducerClass(IntSumReducer.class);
                                             job.setOutputKeyClass(Text.class);
                           70
                                             job.setOutputValueClass(IntWritable.class);
                                             iob.setNumReduceTasks(2):
                                             job.setPartitionerClass(AlphabetPartitioner.class);
                                             FileInputFormat.addInputPath(job, new Path(getConf().get(IN_PATH_PARAM)));
                                             FileOutputFormat.setOutputPath(job, new Path(getConf().get(OUT_PATH_PARAM)));
                                             int exitCode = iob.waitForCompletion( verbose: true) ? 0 : 1:
                                             System.out.println(iob.getCounters().toString());
                                             return exitCode;
                           81
... Structure
                           83
                                         public static void main(String[] args) throws Exception {
                                             int res = ToolRunner.run(new Configuration(), new WordCount(), args);
                                             System.exit(res);
                                         1
                                    }

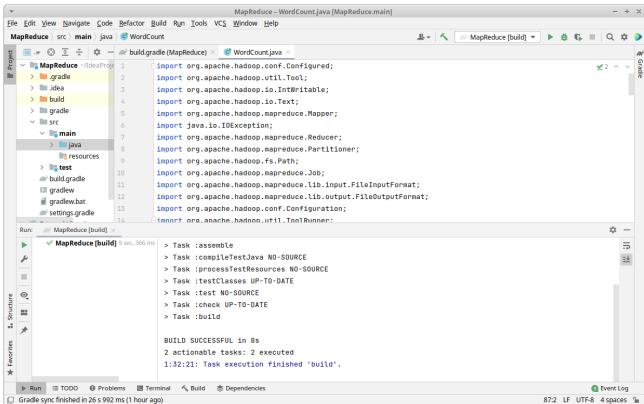
   Event Log

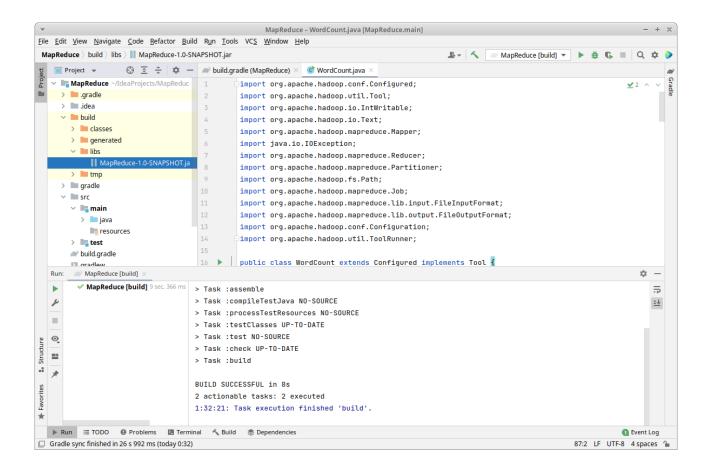
   III TODO ● Problems III Terminal < Build ● Dependencies
☐ Gradle sync finished in 26 s 992 ms (57 minutes ago)
                                                                                                                                      87:2 LF UTF-8 4 spaces 🚡
```

Соберем јаг:

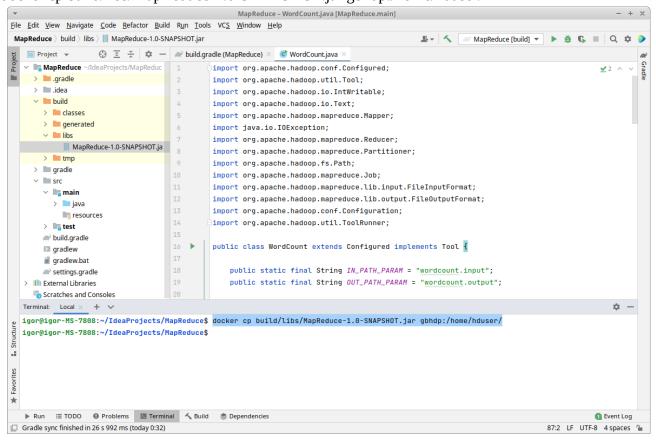
\$./gradlew build

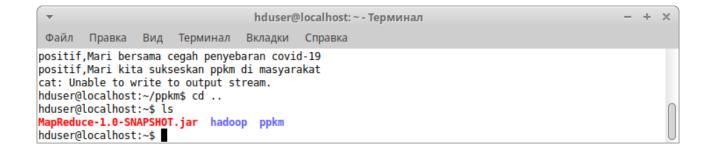






Перенесем его на кластер: docker cp build/libs/MapReduce-1.0-SNAPSHOT.jar gbhdp:/home/hduser/





Запускаем приложение:

\$ hadoop jar MapReduce-1.0-SNAPSHOT.jar \ WordCount \

- -Dwordcount.input=/user/hduser/ppkm $\$
- -Dwordcount.output=/user/hduser/ppkm_out

```
hduser@localhost: ~ - Терминал
   Файл Правка Вид Терминал Вкладки Справка
                                         FILE: Number of bytes read=152790
                                        FILE: Number of bytes read=152790
FILE: Number of bytes written=644541
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=43320
HDFS: Number of bytes written=24274
HDFS: Number of read operations=11
HDFS: Number of large read operations=0
HDFS: Number of write operations=0
HDFS: Number of write operations=5
Unce Framework
                    Map-Reduce Framework
                                         Combine input records=0
                                         Combine output records=0
                                         Reduce input groups=1962
Reduce shuffle bytes=70482
                                         Reduce input records=4678
Reduce output records=1962
                                         Spilled Records=4678
                                         Shuffled Maps =1
Failed Shuffles=0
                                         Merged Map outputs=1
GC time elapsed (ms)=0
                                         Total committed heap usage (bytes)=347602944
                      Shuffle Errors
                                         BAD ID=0
                                          CONNECTION=0
                                         IO ERROR=0
                                         WRONG_LENGTH=0
WRONG_MAP=0
                    WRONG_REDUCE=0
File Output Format Counters
File Output Format Counters
Bytes Written=23972
21/11/22 22:41:06 INFO mapred.LocalJobRunner: Finishing task: attempt_local1141479519_0001_r_000001_0
21/11/22 22:41:06 INFO mapreduce.Job: Job job_local1141479519_0001 running in uber mode : false
21/11/22 22:41:07 INFO mapreduce.Job: map 100% reduce 100%
21/11/22 22:41:07 INFO mapreduce.Job: Job job_local1141479519_0001 completed successfully
21/11/22 22:41:07 INFO mapreduce.Job: Counters: 35
File System Counters
                  22 22:41:07 IMFO mapreduce.Job: Counters: 35
File System Counters
   FILE: Number of bytes read=169177
   FILE: Number of bytes written=1789658
   FILE: Number of read operations=0
   FILE: Number of large read operations=0
   FILE: Number of write operations=0
   HDFS: Number of bytes read=129960
   HDFS: Number of bytes written=24576
   HDFS: Number of read operations=24
   HDFS: Number of large read operations=0
   HDFS: Number of write operations=0
   HDFS: Number of write operations=9
Map-Reduce Framework
Map input records=301
                                        Map input records=301
Map output records=5026
                                        Map output bytes=63419
Map output materialized bytes=73483
Input split bytes=120
Combine input records=0
Combine output records=0
                                         Reduce input groups=2026
Reduce shuffle bytes=73483
                                         Reduce input records=5026
Reduce output records=2026
                                         Spilled Records=10052
                                         Shuffled Maps =2
Failed Shuffles=0
                                         Merged Map outputs=2
GC time elapsed (ms)=0
                                         Total committed heap usage (bytes)=1042808832
                    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=43320
                    File Output Format Counters
                                        Bytes Written=24274
 Counters: 35
                    File System Counters
                   File System Counters

FILE: Number of bytes read=169177

FILE: Number of bytes written=1789658

FILE: Number of read operations=0

FILE: Number of large read operations=0

HDFS: Number of bytes read=129960

HDFS: Number of bytes written=24576

HDFS: Number of read operations=24

HDFS: Number of large read operations=0

HDFS: Number of write operations=9

Map-Reduce Framework

Map input records=301
                                        Map input records=301
                                        Map output records=5026
Map output bytes=63419
Map output materialized bytes=73483
Input split bytes=120
                                         Combine input records=0
Combine output records=0
                                         Reduce input groups=2026
Reduce shuffle bytes=73483
Reduce input records=5026
                                         Reduce output records=2026
Spilled Records=10052
                                         Shuffled Maps =2
                                          Failed Shuffles=0
                                         Merged Map outputs=2
GC time elapsed (ms)=0
Total committed heap usage (bytes)=1042808832
                    Shuffle Errors
BAD_ID=0
                                         CONNECTTON=0
                                        IO_ERROR=0
WRONG_LENGTH=0
                                        WRONG_MAP=0
WRONG_REDUCE=0
                    File Input Format Counters
Bytes Read=43320
File Output Format Counters
Bytes Written=24274
hduser@localhost:~$
```

Смотрим результат:

\$ hdfs dfs -ls /user/hduser/ppkm_out

\$ hdfs dfs -cat /user/hduser/ppkm_out/* | sort -nk2

· h	duser@l	localhost: ~ -	Терминал		-	+	×
Файл Правка	Вид	Терминал	Вкладки	Справка			
mencegah	10						
menerapkan saja 10	10						
seperti 10							
warga 10							
#covid19	11						
22 11 dalam 11							
gara 11							
masih 11							
mau 11 menjaga 11							
pemberlakuan	11						
tetap 11							
aja 12							
banyak 12 dukung 12							
guna 12							
kembali 12							
lagi 12 tentang 12							
hingga 13							
kesehatan	13						
kita 13							
prokes 13 rakyat 13							
bisa 14							
kasus 14							
masker 14							
tidak 14 ini 15							
itu 15							
protokol	15						
sampai 15 2021. 16							
dari 16							
desa 16							
kebijakan	16						
kota 16 orang 16							
pak 16							
penyebaran	17						
diperpanjang mulai 19	18						
atau 20							
#covid-19	21						
#jogjaelinglanwaspada 21							
https://t.co/roomoqobgc 21 netral,rt 21							
pemberlakukan	21						
pengumuman	21						
<pre>#jogjaistimewa @humas_jogja:</pre>	25 25						
covid 25	23						
mikro. 25							
covid-19	27						
maret 27 untuk 27							
yang 29							
ada 30							
berbasis	31 38						
perpanjangan pembatasan	38 42						
kegiatan	44						
yg 56							
masyarakat #ppkmmikro	58 60						
#ppkillilikro	00						
mikro 78							
dan 88							
ppkm 100 hduser@localhos	st:~\$ ■						
ווממפר שנוטכמנווט	ос.~р						0

\$ hdfs dfs -cat /user/hduser/ppkm_out/part-r-00000

```
hduser@localhost: ~ - Терминал
 Файл
        Правка Вид Терминал Вкладки Справка
hduser@localhost:~$ hdfs dfs -cat /user/hduser/ppkm_out/part-r-00000
        41
        1
        2
        1
        1
(
        4
        3
        6
        3
        1
0
        3
0%
        5
1
        1
10
        1
14
        5
19
        4
2
22
        11
3
        3
35
        1
3m
        10
        1
        8
5m
        1
75
        1
        2
8
        8
9
        2
        1 2
:)
        1
ah
        1
bs
        2
bu
        5
        3
dg
        77
di
dk
        1
        5 2
dp
dr
eh
        1
g
        9
ga
        6
gk
he
        2
hp
        1
ii
        1
        3
jd
ke
        10
lo
        4
n
        1
ni
        1
om
        1
pb
        1
rm
        1
rs
        1
sm
        1
to
        1
        3
tp
tu
        1
        1
W
wi
        1
        1
y
        6
ya
        56
yg
        1
hduser@localhost:~$
```

SCALA MAPREDUCE OTKPЫBAEM INTELLIJ IDEA УСТАНАВЛИВАЕМ SCALA PLUGIN CO3ДAEM SBT ПРОЕКТ SCALA MAPREDUCE:

В build.sbt пропишем зависимости хадупа:

