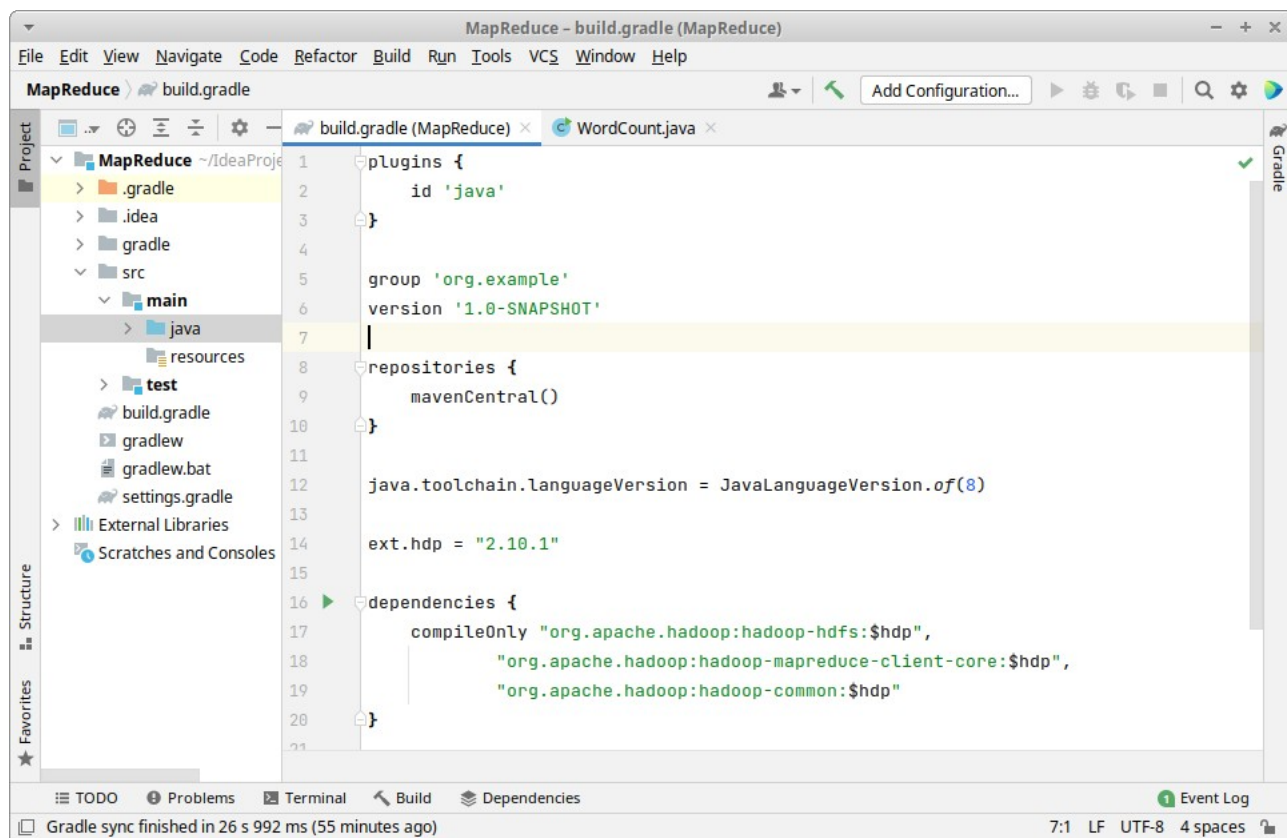


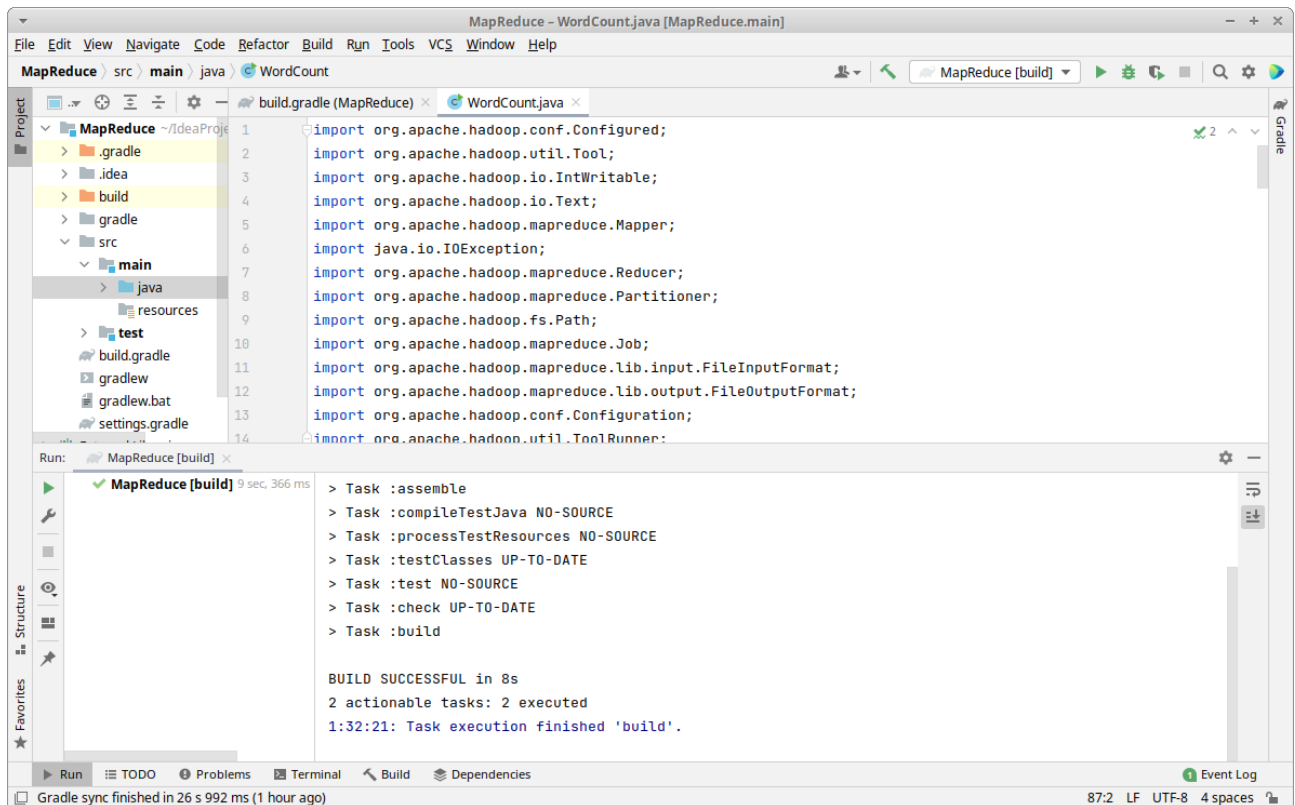
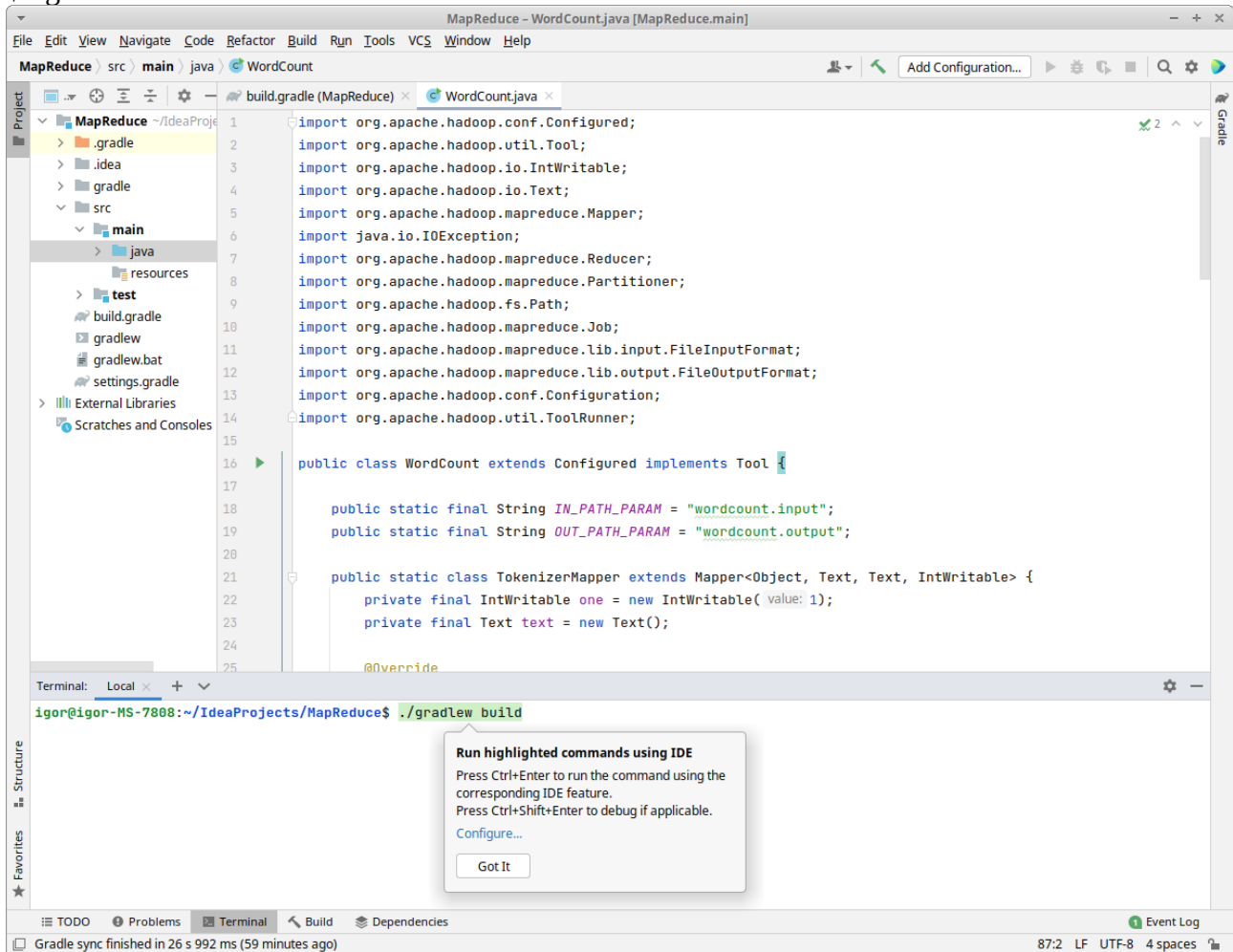
MAPREDUCE ON JAVA
ОТКРЫВАЕМ INTELLIJ IDEA
СОЗДАЕМ JAVA ПРОЕКТ НА GRADLE
WORD COUNT:
В build.gradle добавим зависимости хадупа:

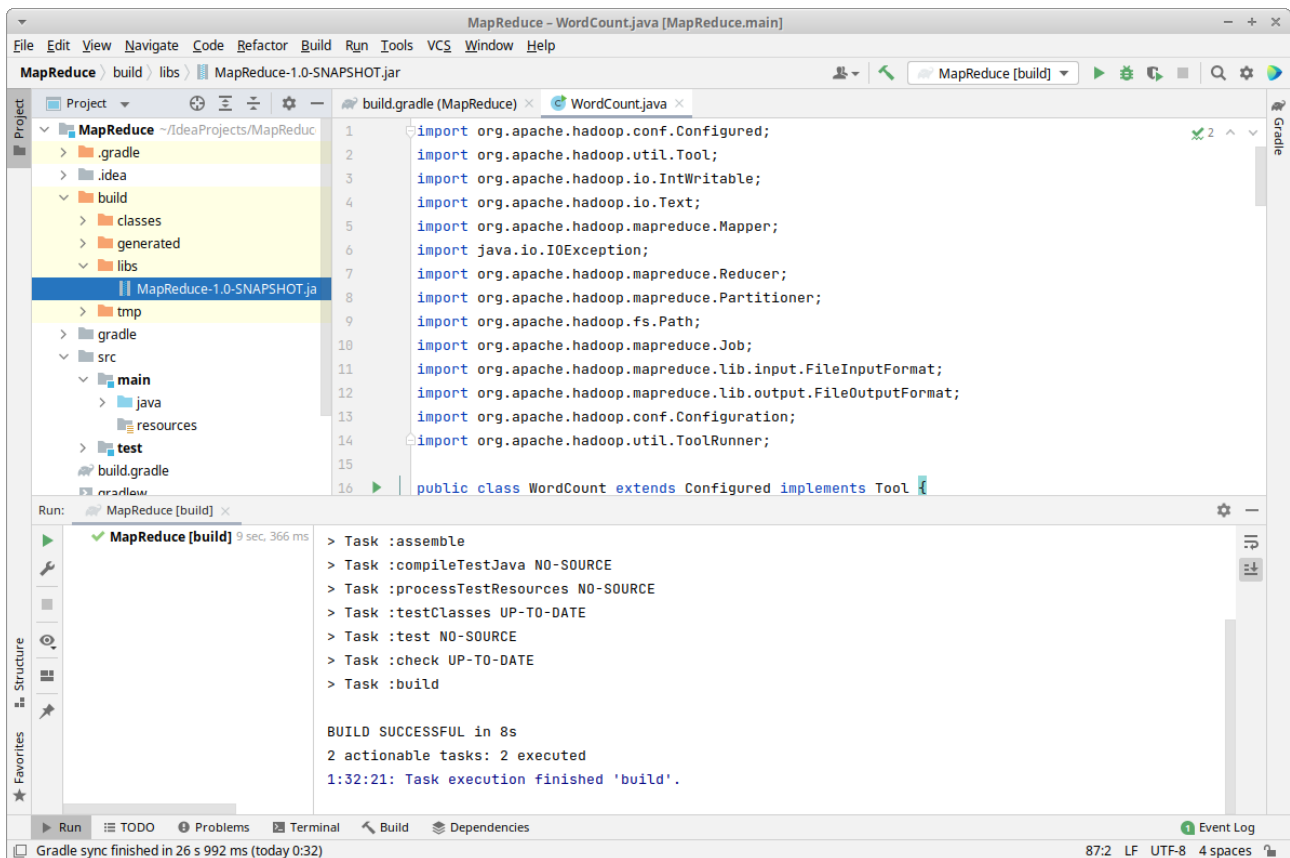


В src/main/java создадим WordCount.java:



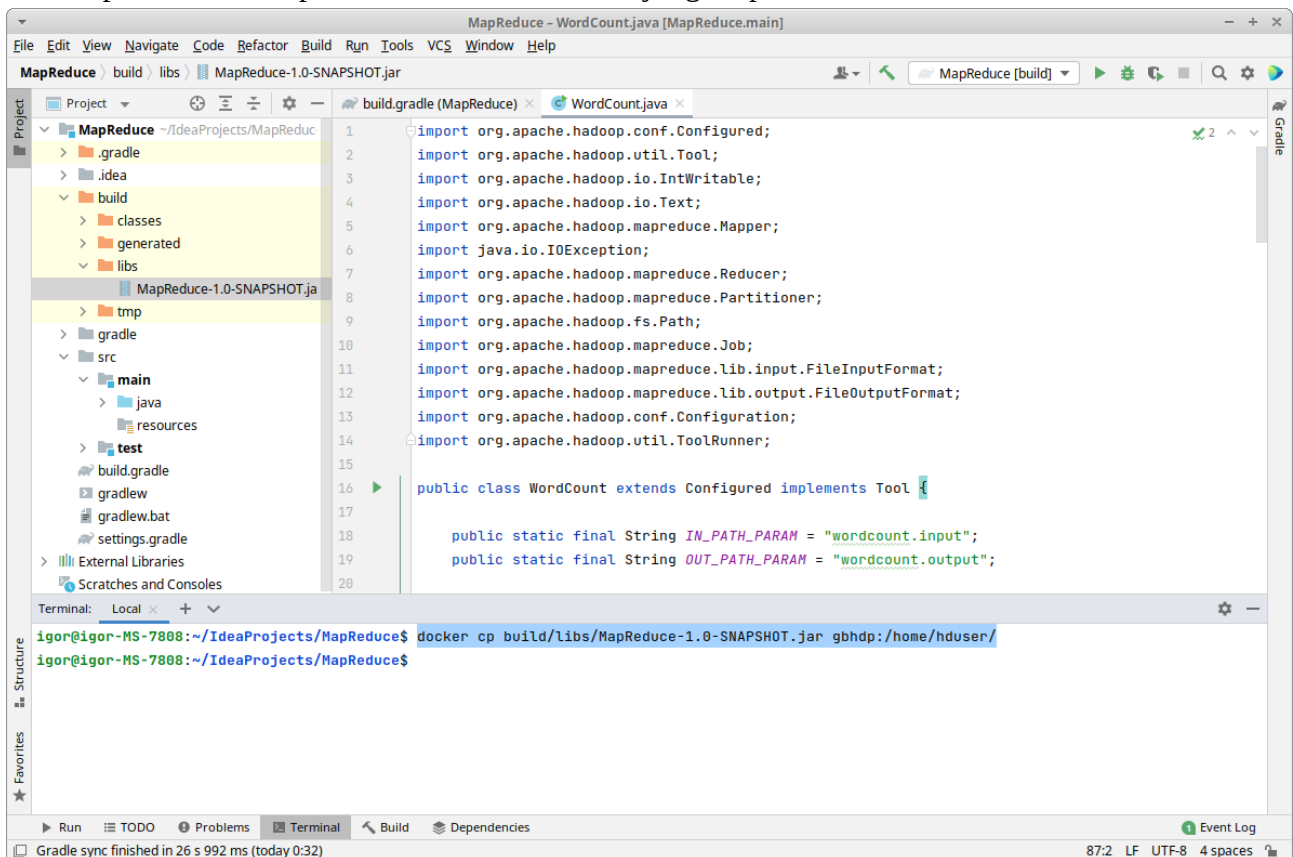
Соберем jar:
\$./gradlew build

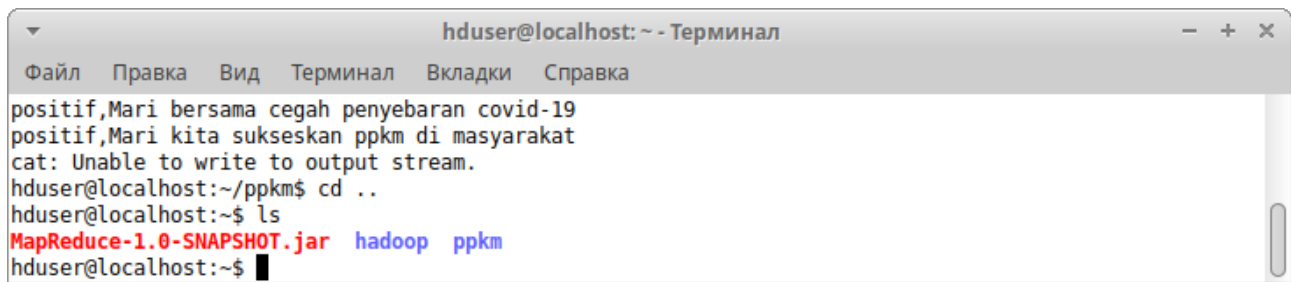




Перенесем его на кластер:

docker cp build/libs/MapReduce-1.0-SNAPSHOT.jar gbhdhp:/home/hduser/





```
hduser@localhost: ~ - Терминал
Файл  Правка  Вид  Терминал  Вкладки  Справка
positif,Mari bersama cegah penyebaran covid-19
positif,Mari kita sukseskan ppkm di masyarakat
cat: Unable to write to output stream.
hduser@localhost:~/ppkm$ cd ..
hduser@localhost:~$ ls
MapReduce-1.0-SNAPSHOT.jar  hadoop  ppkm
hduser@localhost:~$
```

Запускаем приложение:

```
$ hadoop jar MapReduce-1.0-SNAPSHOT.jar \
WordCount \
-Dwordcount.input=/user/hduser/ppkm \
-Dwordcount.output=/user/hduser/ppkm_out
```

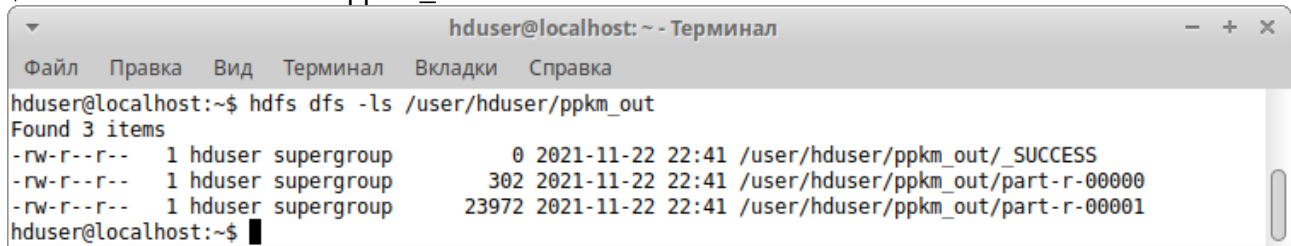


```
hduser@localhost: ~ - Терминал
Файл  Правка  Вид  Терминал  Вкладки  Справка

FILE: Number of bytes read=152790
FILE: Number of bytes 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=5
Map-Reduce Framework
  Combine input records=0
  Combine output records=0
  Reduce input groups=1962
  Reduce shuffle bytes=79482
  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
  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 mapred.LocalJobRunner: reduce task executor complete.
21/11/22 22:41:07 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
  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=9
Map-Reduce Framework
  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: 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=9
Map-Reduce Framework
  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
hduser@localhost:~$
```

Смотрим результат:

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



A terminal window titled "hduser@localhost: ~ - Терминал" with a menu bar containing "Файл", "Правка", "Вид", "Терминал", "Вкладки", and "Справка". The terminal shows the command "hdfs dfs -ls /user/hduser/ppkm_out" and its output. The output indicates that 3 items were found. The items are listed in a table-like format with permissions, owner, group, size, timestamp, and path.

Permissions	Owner	Group	Size	Timestamp	Path
-rw-r--r--	1 hduser	supergroup	0	2021-11-22 22:41	/user/hduser/ppkm_out/_SUCCESS
-rw-r--r--	1 hduser	supergroup	302	2021-11-22 22:41	/user/hduser/ppkm_out/part-r-00000
-rw-r--r--	1 hduser	supergroup	23972	2021-11-22 22:41	/user/hduser/ppkm_out/part-r-00001

The terminal prompt "hduser@localhost:~\$" is visible at the bottom.

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

```
hduser@localhost: ~ - Терминал
Файл  Правка  Вид  Терминал  Вкладки  Справка

mencegah      10
menerapkan    10
saja          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  18
mulai         19
atau          20
#covid-19     21
#jogjaelinglanwaspada 21
https://t.co/roomoqobgc 21
netral,rt     21
pemberlakuan  21
pengumuman    21
#jogjaistimewa 25
@humas_jogja: 25
covid         25
mikro.        25
covid-19      27
maret         27
untuk         27
yang          29
ada           30
berbasis      31
perpanjangan  38
pembatasan    42
kegiatan      44
yg            56
masyarakat    58
#ppkmmikro    60
di            77
mikro         78
dan           88
ppkm          100
hduser@localhost:~$
```



```
$ 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-000000
41
!      1
"      2
&      1
(      1
,      4
-      3
.      6
..     3
0      1
0%     3
1      5
10     1
14     1
19     5
2      4
22     11
3      3
35     1
3m     10
5      1
5m     8
7      1
75     1
8      2
9      8
:      2
:)     1
?      2
ah     1
bs     1
bu     2
d      5
dg     3
di     77
dk     1
dp     3
dr     5
eh     2
g      1
ga     9
gk     6
he     2
hp     1
ii     1
jd     3
ke     10
lo     4
n      1
ni     1
om     1
pb     1
rm     1
rs     1
sm     1
to     1
tp     3
tu     1
w      1
wi     1
y      1
ya     6
yg     56
|      1
hduser@localhost:~$
```

SCALA MAPREDUCE
ОТКРЫВАЕМ INTELLIJ IDEA
УСТАНАВЛИВАЕМ SCALA PLUGIN
СОЗДАЕМ SBT ПРОЕКТ
SCALA MAPREDUCE:
В build.sbt пропишем зависимости хадупа:

