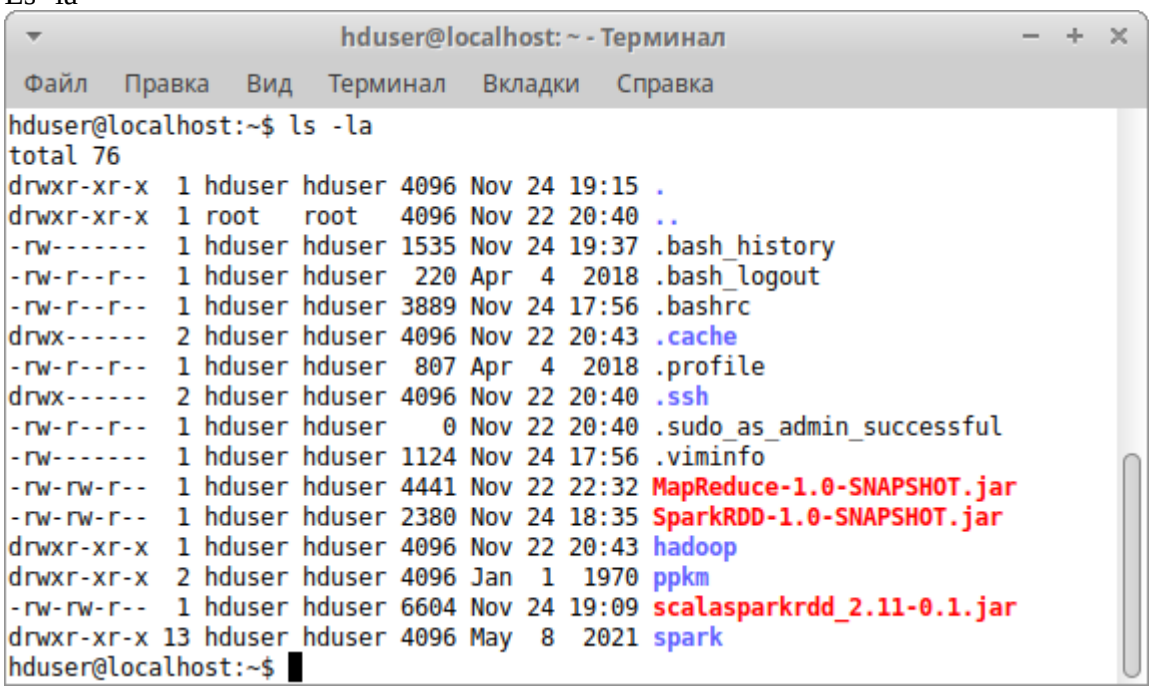
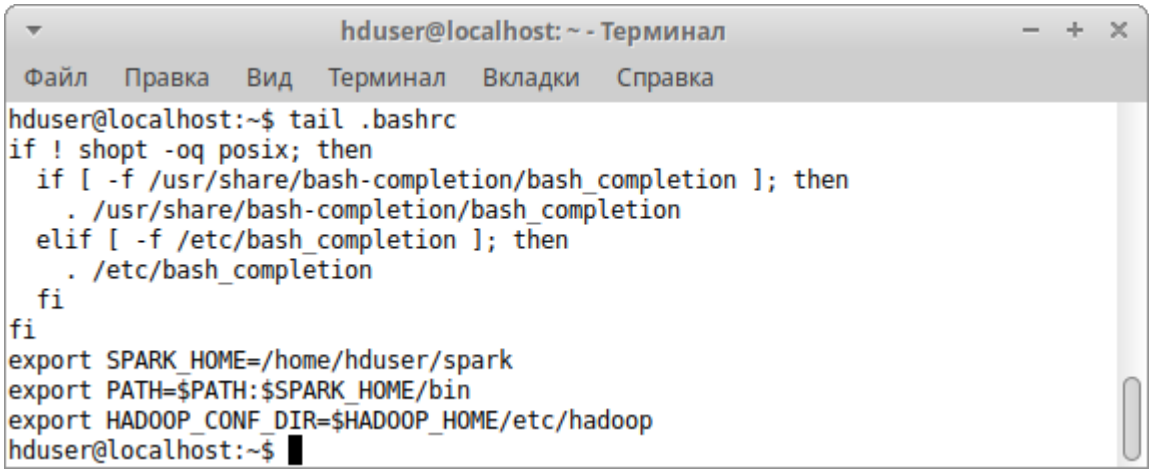
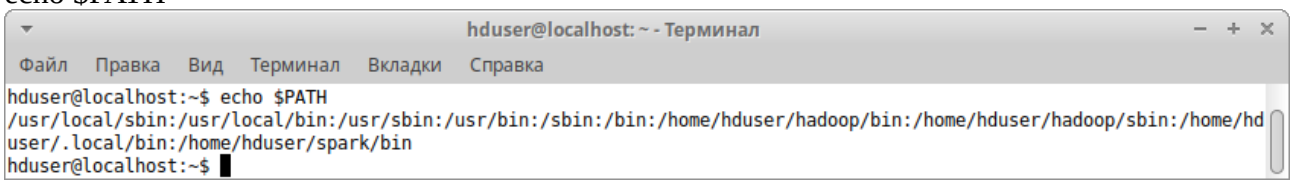
Ls -la



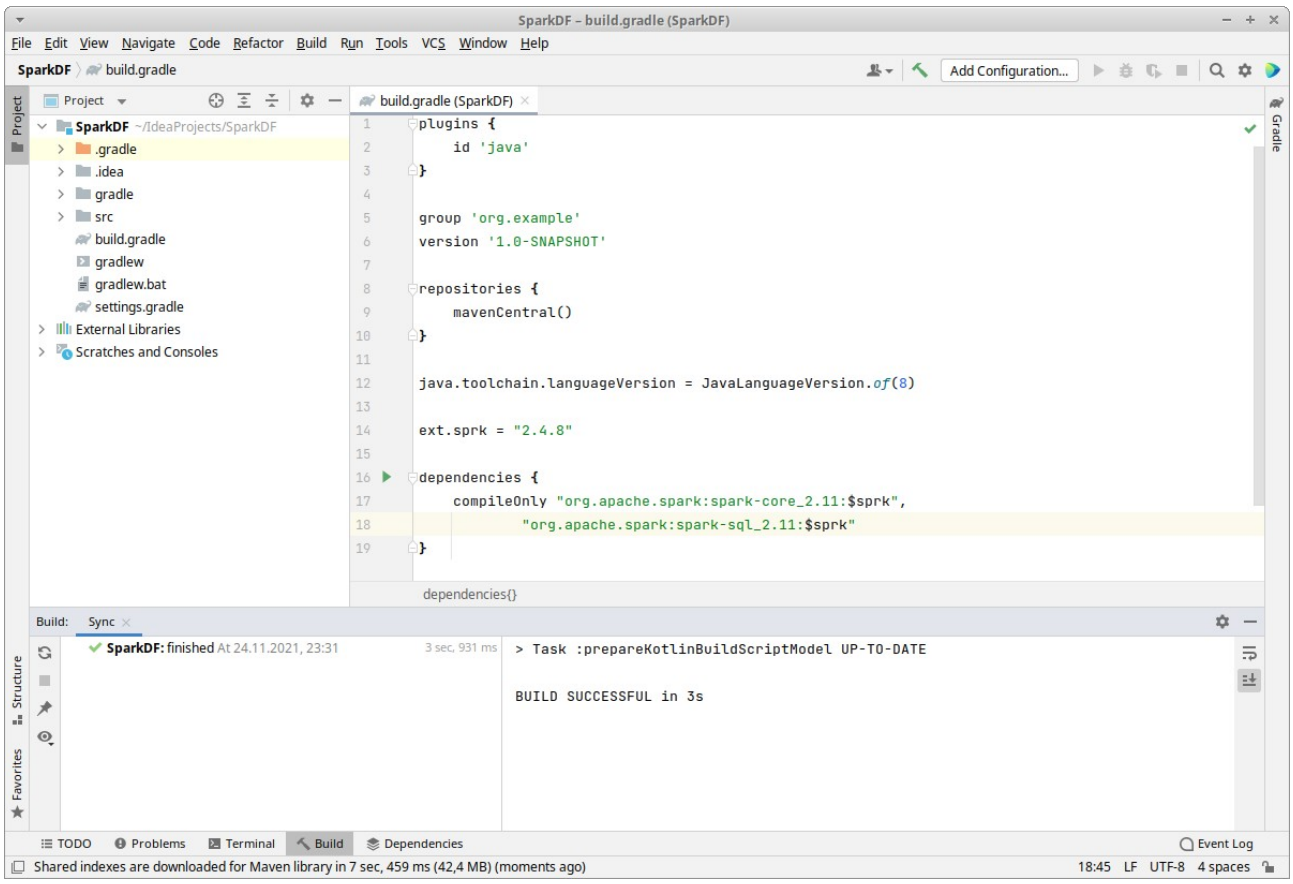
tail .bashrc



echo $PATH

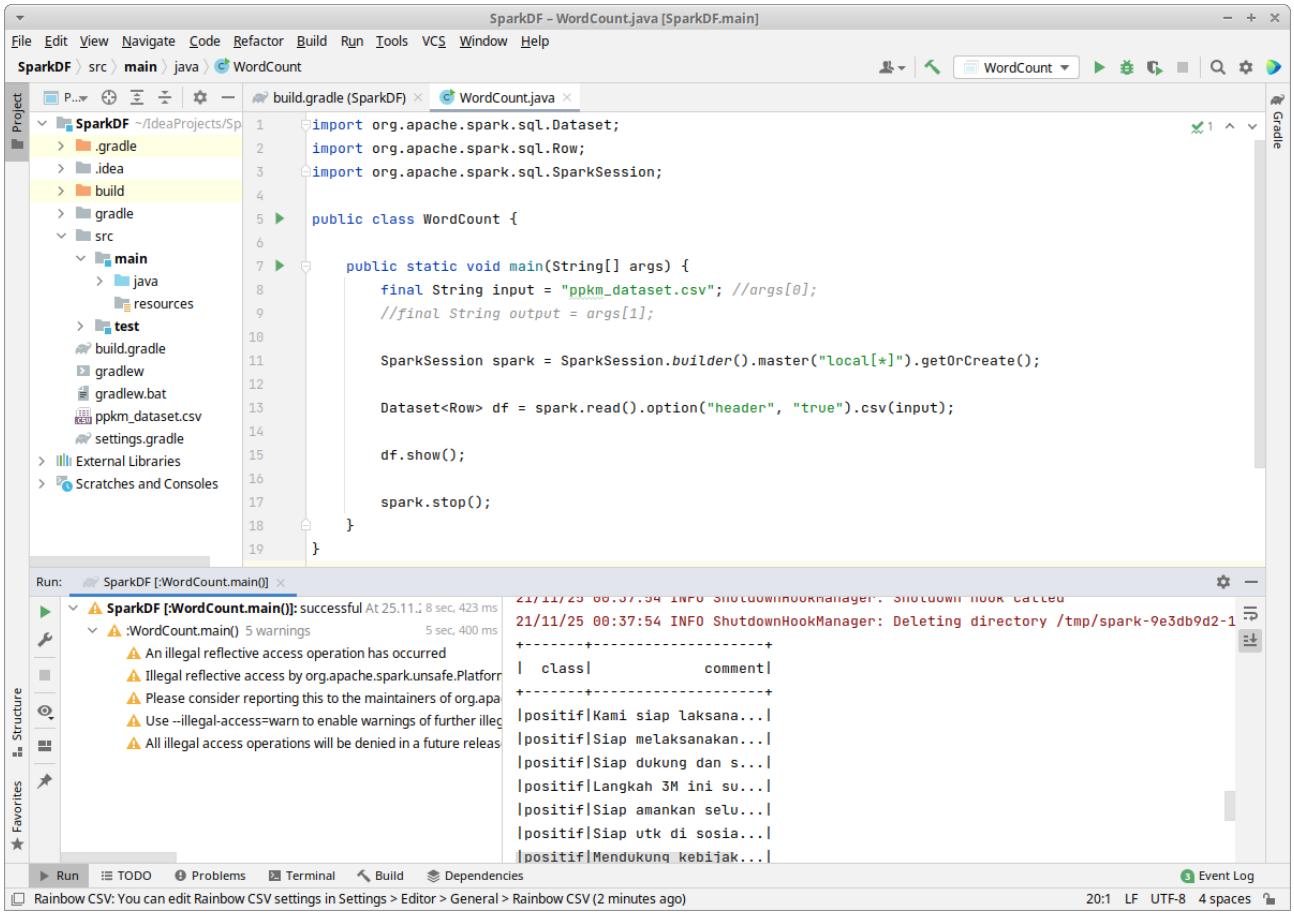


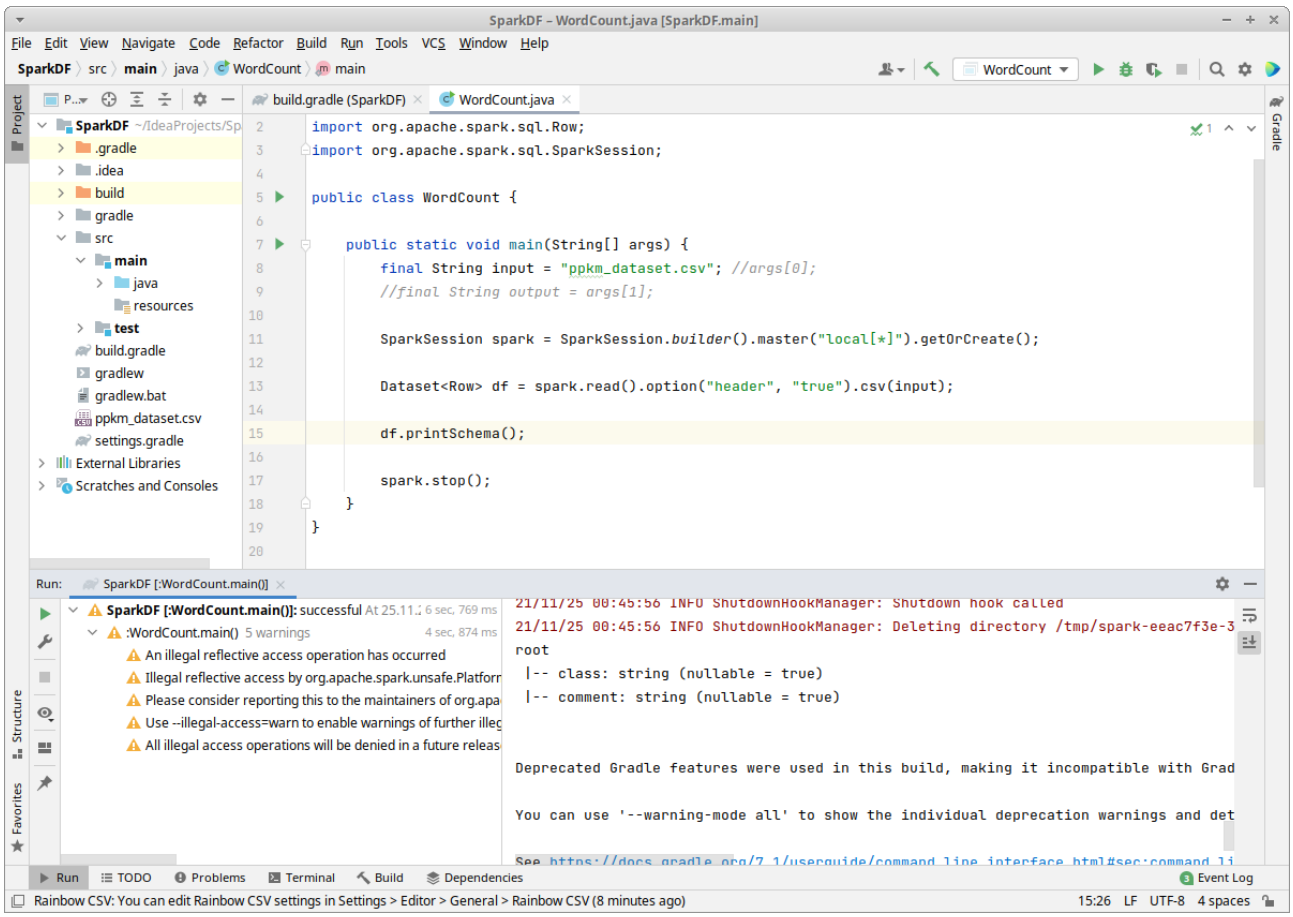
JAVA WORD COUNT: ПРОЕКТ В IDEA
JAVA WORD COUNT: ЗАВИСИМОСТИ
Добавим в build.gradle зависимость на Apache Spark:

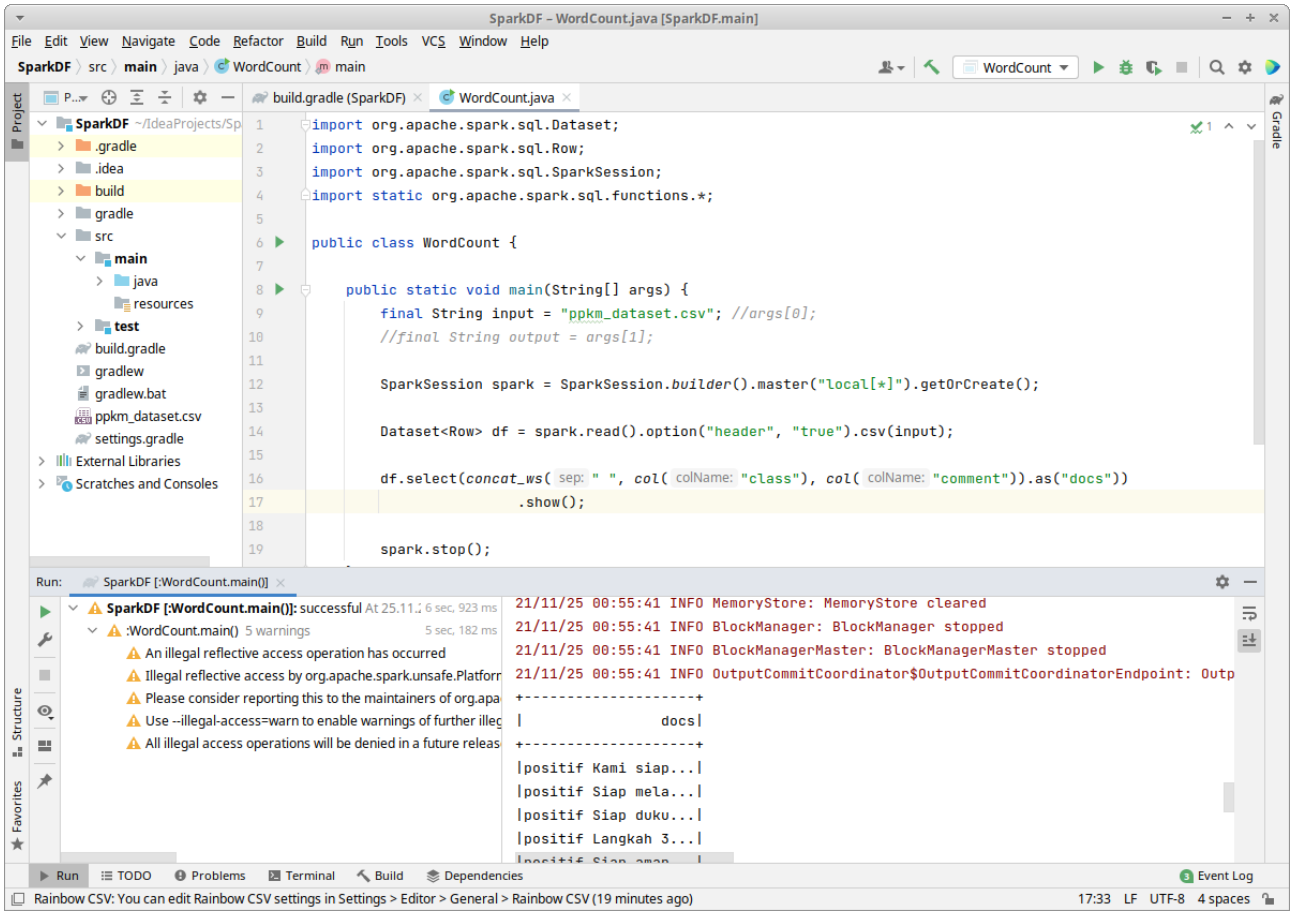


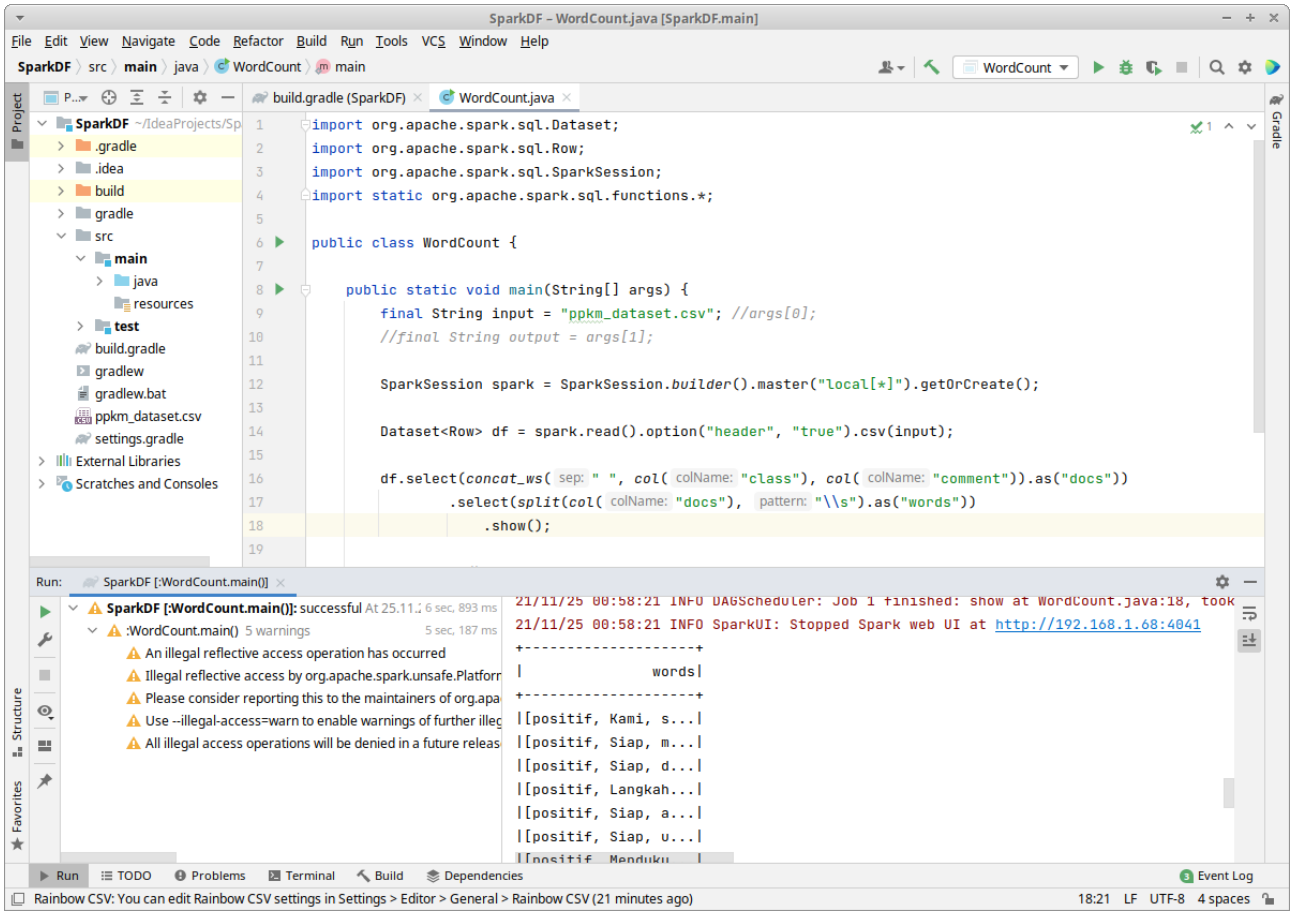
JAVA WORD COUNT: КОД

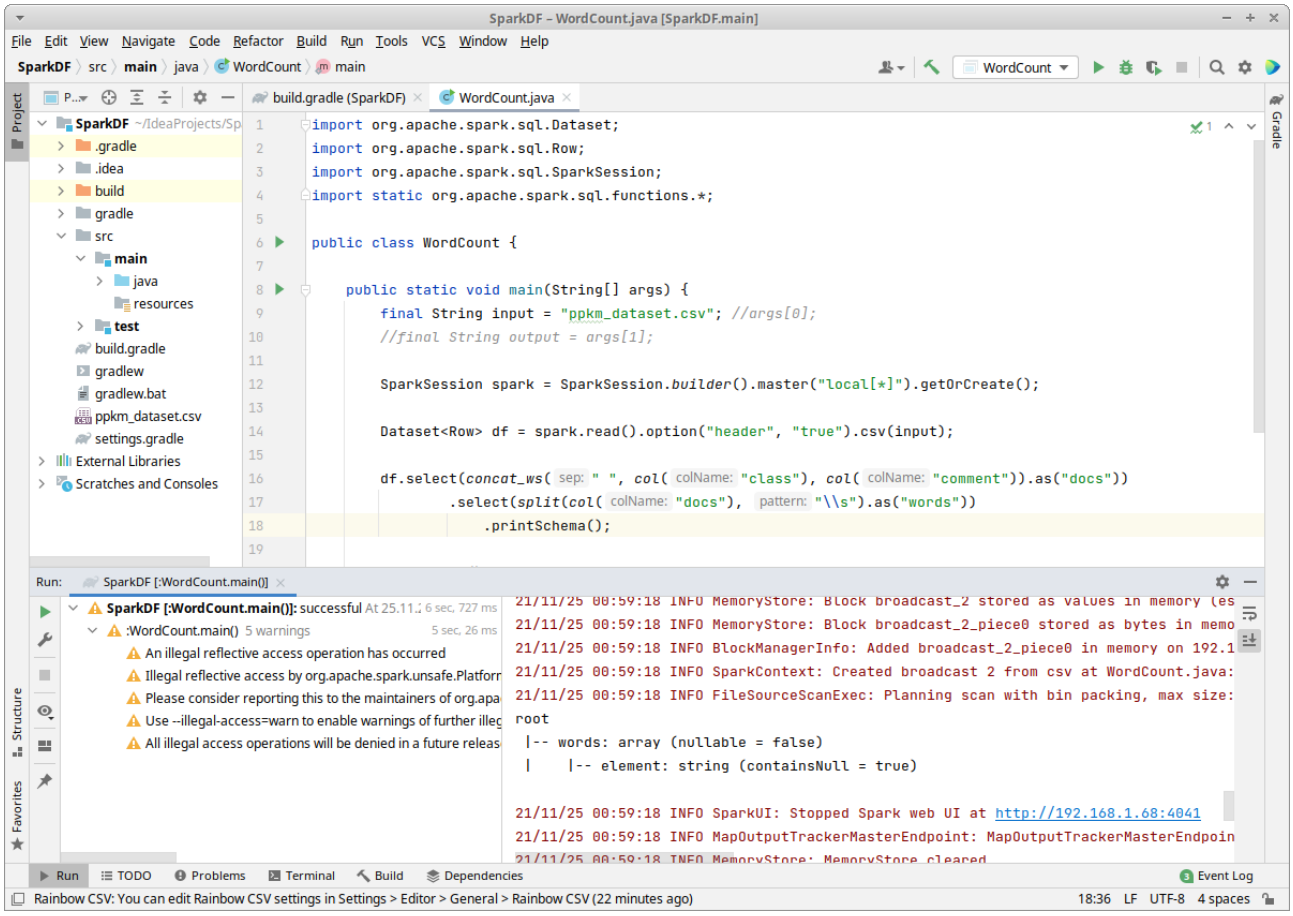
Создадим WordCount.java:

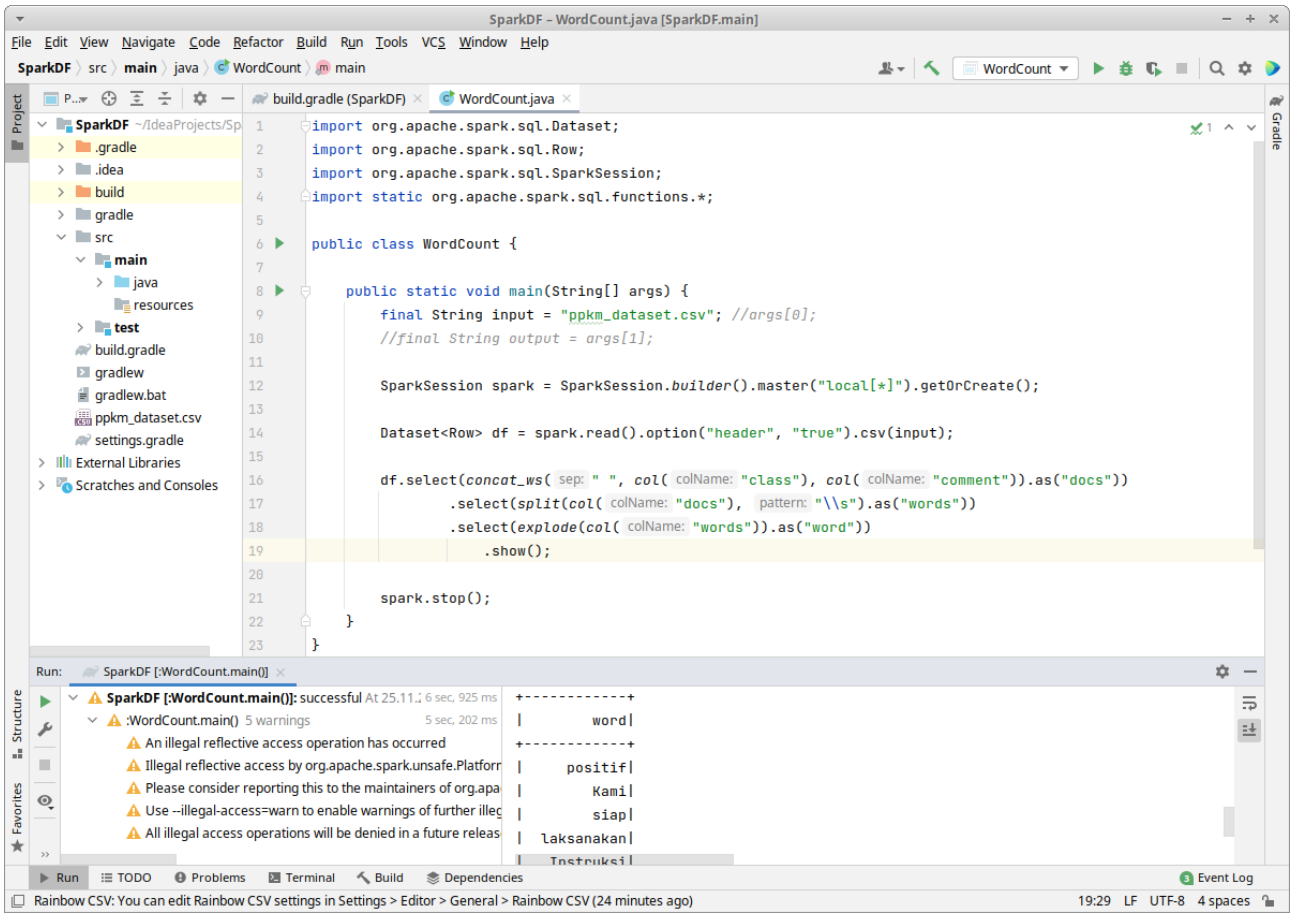



SparkDF – WordCount.java [SparkDF.main]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDF › src › main › java › WordCount › main
WordCount
build.gradle (SparkDF)
WordCount.java
import org.apache.spark.sql.Row;
import org.apache.spark.sql.SparkSession;
public class WordCount {
public static void main(String[] args) {
final String input = "ppkm_dataset.csv"; //args[0];
//final String output = args[1];
SparkSession spark = SparkSession.builder().master("local[*]").getOrCreate();
Dataset<Row> df = spark.read().option("header", "true").csv(input);
df.printSchema();
spark.stop();
}
}
Run: SparkDF [:WordCount.main()]
SparkDF [:WordCount.main()]: successful At 25.11.2 6 sec, 769 ms
:WordCount.main() 5 warnings 4 sec, 874 ms
An illegal reflective access operation has occurred
Illegal reflective access by org.apache.spark.unsafe.Platform
Please consider reporting this to the maintainers of org.apa
Use --illegal-access=warn to enable warnings of further illeg
All illegal access operations will be denied in a future releas
21/11/25 00:45:56 INFO ShutdownHookManager: Shutdown hook called
21/11/25 00:45:56 INFO ShutdownHookManager: Deleting directory /tmp/spark-eeac7f3e-3
root
|-- class: string (nullable = true)
|-- comment: string (nullable = true)
Deprecated Gradle features were used in this build, making it incompatible with Grad
You can use '--warning-mode all' to show the individual deprecation warnings and det
Run TODO Problems Terminal Build Dependencies Event Log
Rainbow CSV: You can edit Rainbow CSV settings in Settings > Editor > General > Rainbow CSV (8 minutes ago)
15:26 LF UTF-8 4 spaces

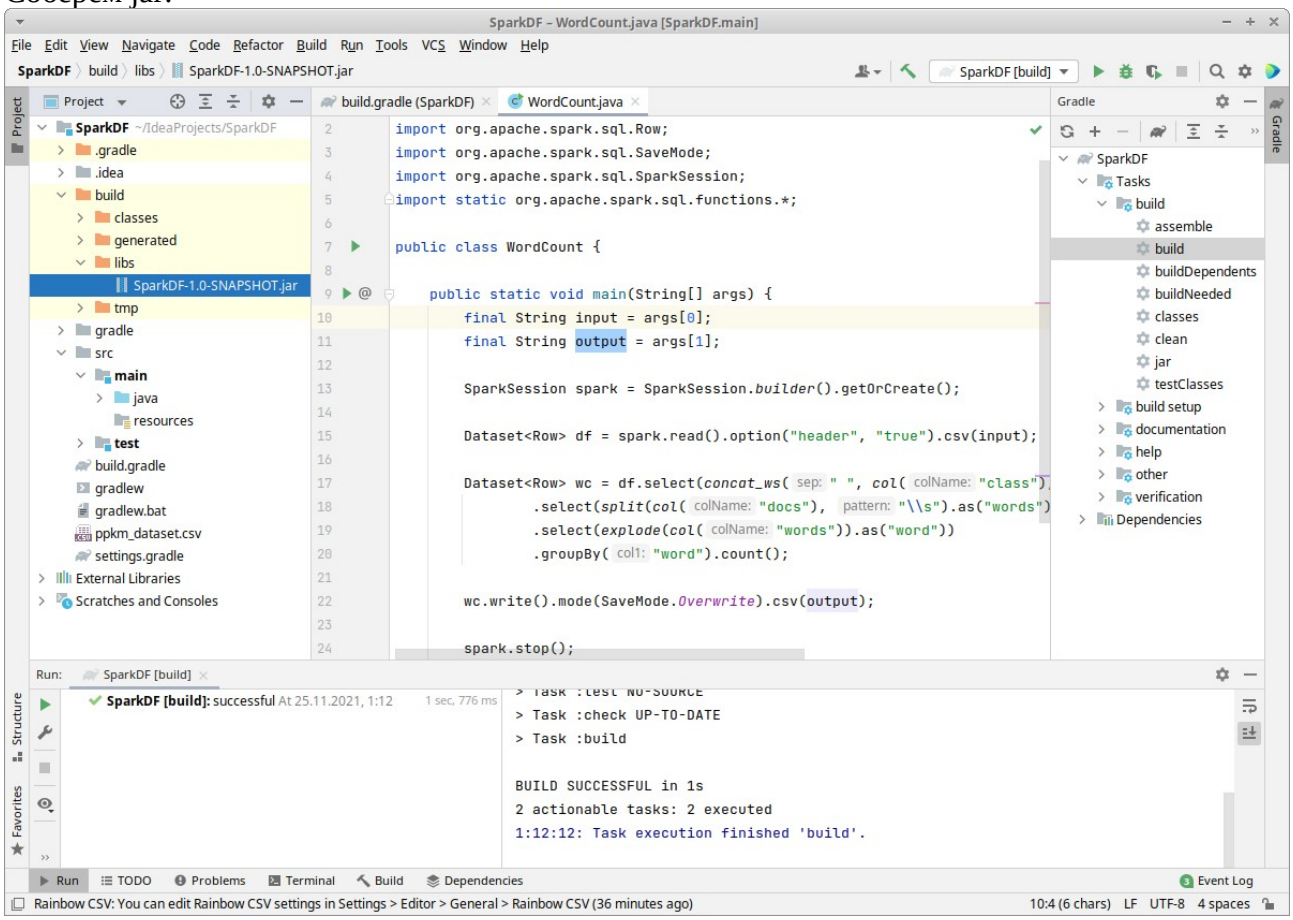

SparkDF – WordCount.java [SparkDF.main]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDF › src › main › java › WordCount › main
WordCount
build.gradle (SparkDF)
WordCount.java
import org.apache.spark.sql.Dataset;
import org.apache.spark.sql.Row;
import org.apache.spark.sql.SparkSession;
import static org.apache.spark.sql.functions.*;
public class WordCount {
public static void main(String[] args) {
final String input = "ppkm_dataset.csv"; //args[0];
//final String output = args[1];
SparkSession spark = SparkSession.builder().master("local[*]").getOrCreate();
Dataset<Row> df = spark.read().option("header", "true").csv(input);
df.select(concat_ws(sep: " ", col(colName: "class"), col(colName: "comment")).as("docs"))
.show();
spark.stop();
Run: SparkDF [:WordCount.main()]
SparkDF [:WordCount.main()]: successful At 25.11.2 6 sec, 923 ms
:WordCount.main() 5 warnings 5 sec, 182 ms
An illegal reflective access operation has occurred
Illegal reflective access by org.apache.spark.unsafe.Platform
Please consider reporting this to the maintainers of org.apa
Use --illegal-access=warn to enable warnings of further illeg
All illegal access operations will be denied in a future releas
21/11/25 00:55:41 INFO MemoryStore: MemoryStore cleared
21/11/25 00:55:41 INFO BlockManager: BlockManager stopped
21/11/25 00:55:41 INFO BlockManagerMaster: BlockManagerMaster stopped
21/11/25 00:55:41 INFO OutputCommitCoordinator$OutputCommitCoordinatorEndpoint: Outp
+--------------------+
| docs|
+--------------------+
|positif Kami siap...|
|positif Siap mela...|
|positif Siap duku...|
|positif Langkah 3...|
Run TODO Problems Terminal Build Dependencies Event Log
Rainbow CSV: You can edit Rainbow CSV settings in Settings > Editor > General > Rainbow CSV (19 minutes ago)
17:33 LF UTF-8 4 spaces


SparkDF – WordCount.java [SparkDF.main]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDF › src › main › java › WordCount › main
WordCount
build.gradle (SparkDF)
WordCount.java
import org.apache.spark.sql.Dataset;
import org.apache.spark.sql.Row;
import org.apache.spark.sql.SparkSession;
import static org.apache.spark.sql.functions.*;
public class WordCount {
public static void main(String[] args) {
final String input = "ppkm_dataset.csv"; //args[0];
//final String output = args[1];
SparkSession spark = SparkSession.builder().master("local[*]").getOrCreate();
Dataset<Row> df = spark.read().option("header", "true").csv(input);
df.select(concat_ws(sep: " ", col(colName: "class"), col(colName: "comment")).as("docs"))
.select(split(col(colName: "docs"), pattern: "\\s").as("words"))
.show();
Run: SparkDF [:WordCount.main()]
SparkDF [:WordCount.main()]: successful At 25.11.2 6 sec, 893 ms
:WordCount.main() 5 warnings 5 sec, 187 ms
An illegal reflective access operation has occurred
Illegal reflective access by org.apache.spark.unsafe.Platform
Please consider reporting this to the maintainers of org.apa
Use --illegal-access=warn to enable warnings of further illeg
All illegal access operations will be denied in a future releas
21/11/25 00:58:21 INFO DAGScheduler: Job 1 finished: show at WordCount.java:18, took
21/11/25 00:58:21 INFO SparkUI: Stopped Spark web UI at http://192.168.1.68:4041
+--------------------+
| words|
+--------------------+
|[positif, Kami, s...|
|[positif, Siap, m...|
|[positif, Siap, d...|
|[positif, Langkah...|
|[positif, Siap, a...|
|[positif, Siap, u...|
Run TODO Problems Terminal Build Dependencies Event Log
Rainbow CSV: You can edit Rainbow CSV settings in Settings > Editor > General > Rainbow CSV (21 minutes ago)
18:21 LF UTF-8 4 spaces

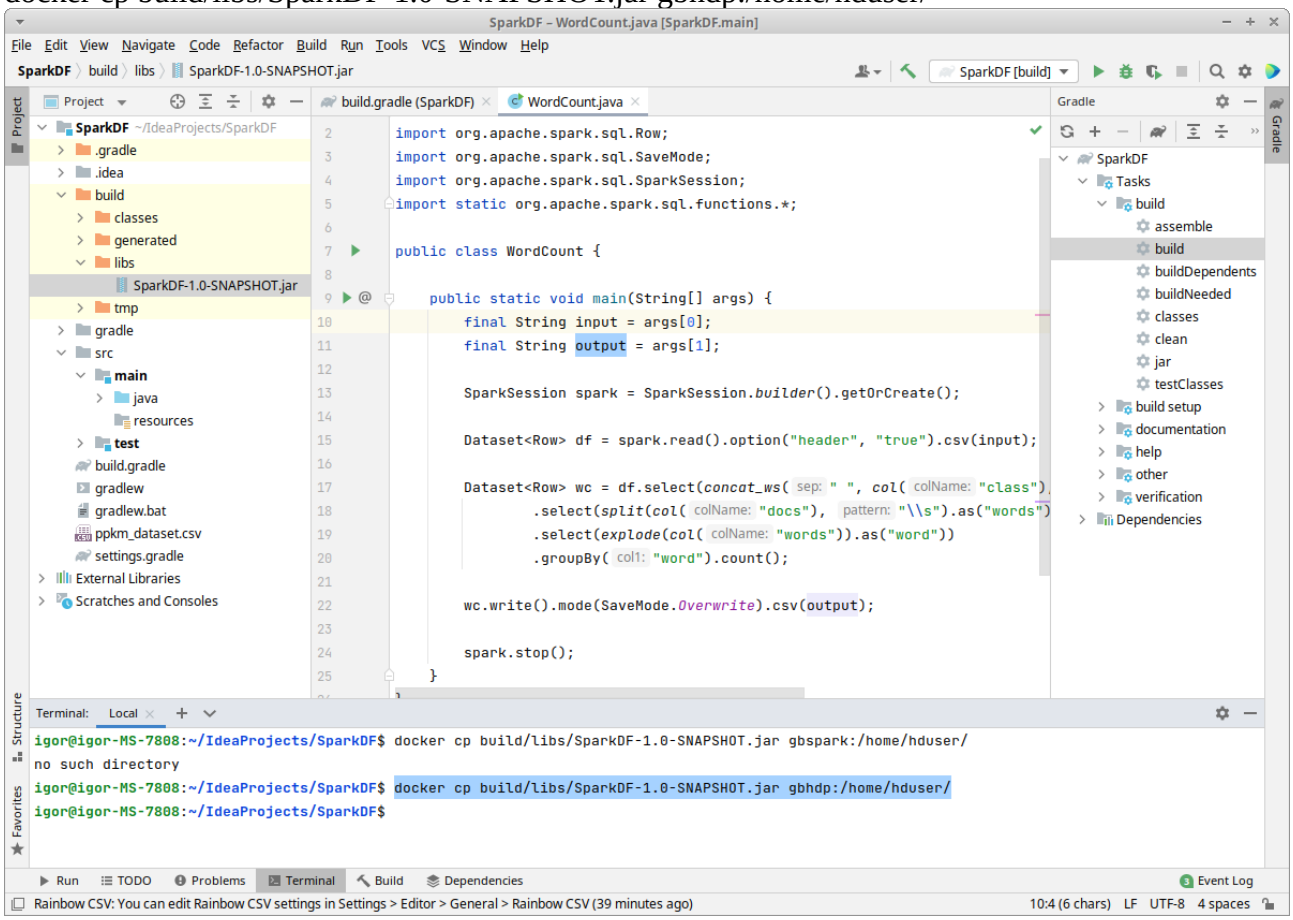

SparkDF – WordCount.java [SparkDF.main]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDF › src › main › java › WordCount › main
WordCount
build.gradle (SparkDF)
WordCount.java
import org.apache.spark.sql.Dataset;
import org.apache.spark.sql.Row;
import org.apache.spark.sql.SparkSession;
import static org.apache.spark.sql.functions.*;
public class WordCount {
public static void main(String[] args) {
final String input = "ppkm_dataset.csv"; //args[0];
//final String output = args[1];
SparkSession spark = SparkSession.builder().master("local[*]").getOrCreate();
Dataset<Row> df = spark.read().option("header", "true").csv(input);
df.select(concat_ws(sep: " ", col(colName: "class"), col(colName: "comment")).as("docs"))
.select(split(col(colName: "docs"), pattern: "\\s").as("words"))
.printSchema();
Run: SparkDF [:WordCount.main()]
SparkDF [:WordCount.main()]: successful At 25.11.2 6 sec, 727 ms
:WordCount.main() 5 warnings 5 sec, 26 ms
An illegal reflective access operation has occurred
Illegal reflective access by org.apache.spark.unsafe.Platform
Please consider reporting this to the maintainers of org.apa
Use --illegal-access=warn to enable warnings of further illeg
All illegal access operations will be denied in a future releas
21/11/25 00:59:18 INFO MemoryStore: Block broadcast_2 stored as values in memory (es
21/11/25 00:59:18 INFO MemoryStore: Block broadcast_2_piece0 stored as bytes in memo
21/11/25 00:59:18 INFO BlockManagerInfo: Added broadcast_2_piece0 in memory on 192.1
21/11/25 00:59:18 INFO SparkContext: Created broadcast 2 from csv at WordCount.java:
21/11/25 00:59:18 INFO FileSourceScanExec: Planning scan with bin packing, max size:
root
|-- words: array (nullable = false)
| |-- element: string (containsNull = true)
21/11/25 00:59:18 INFO SparkUI: Stopped Spark web UI at http://192.168.1.68:4041
21/11/25 00:59:18 INFO MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEndpoin
Run TODO Problems Terminal Build Dependencies Event Log
Rainbow CSV: You can edit Rainbow CSV settings in Settings > Editor > General > Rainbow CSV (22 minutes ago)
18:36 LF UTF-8 4 spaces

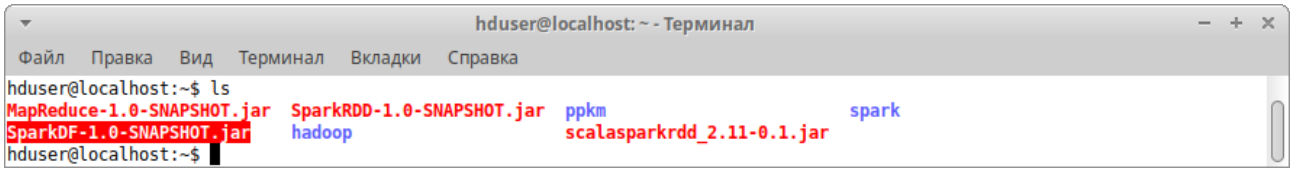

SparkDF – WordCount.java [SparkDF.main]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDF › src › main › java › WordCount › main
WordCount
build.gradle (SparkDF)
WordCount.java
import org.apache.spark.sql.Dataset;
import org.apache.spark.sql.Row;
import org.apache.spark.sql.SparkSession;
import static org.apache.spark.sql.functions.*;
public class WordCount {
public static void main(String[] args) {
final String input = "ppkm_dataset.csv"; //args[0];
//final String output = args[1];
SparkSession spark = SparkSession.builder().master("local[*]").getOrCreate();
Dataset<Row> df = spark.read().option("header", "true").csv(input);
df.select(concat_ws(sep: " ", col(colName: "class"), col(colName: "comment")).as("docs"))
.select(split(col(colName: "docs"), pattern: "\\s").as("words"))
.select(explode(col(colName: "words")).as("word"))
.show();
spark.stop();
}
}
Run: SparkDF [:WordCount.main()]
SparkDF [:WordCount.main()]: successful At 25.11.
:WordCount.main() 5 warnings 5 sec, 202 ms
An illegal reflective access operation has occurred
Illegal reflective access by org.apache.spark.unsafe.Platform
Please consider reporting this to the maintainers of org.apa
Use --illegal-access=warn to enable warnings of further illeg
All illegal access operations will be denied in a future releas
+------------+
| word|
+------------+
| positif|
| Kami|
| siap|
| laksanakan|
| Instruksi|
Run TODO Problems Terminal Build Dependencies Event Log
Rainbow CSV: You can edit Rainbow CSV settings in Settings > Editor > General > Rainbow CSV (24 minutes ago)
19:29 LF UTF-8 4 spaces

Соберем jar:

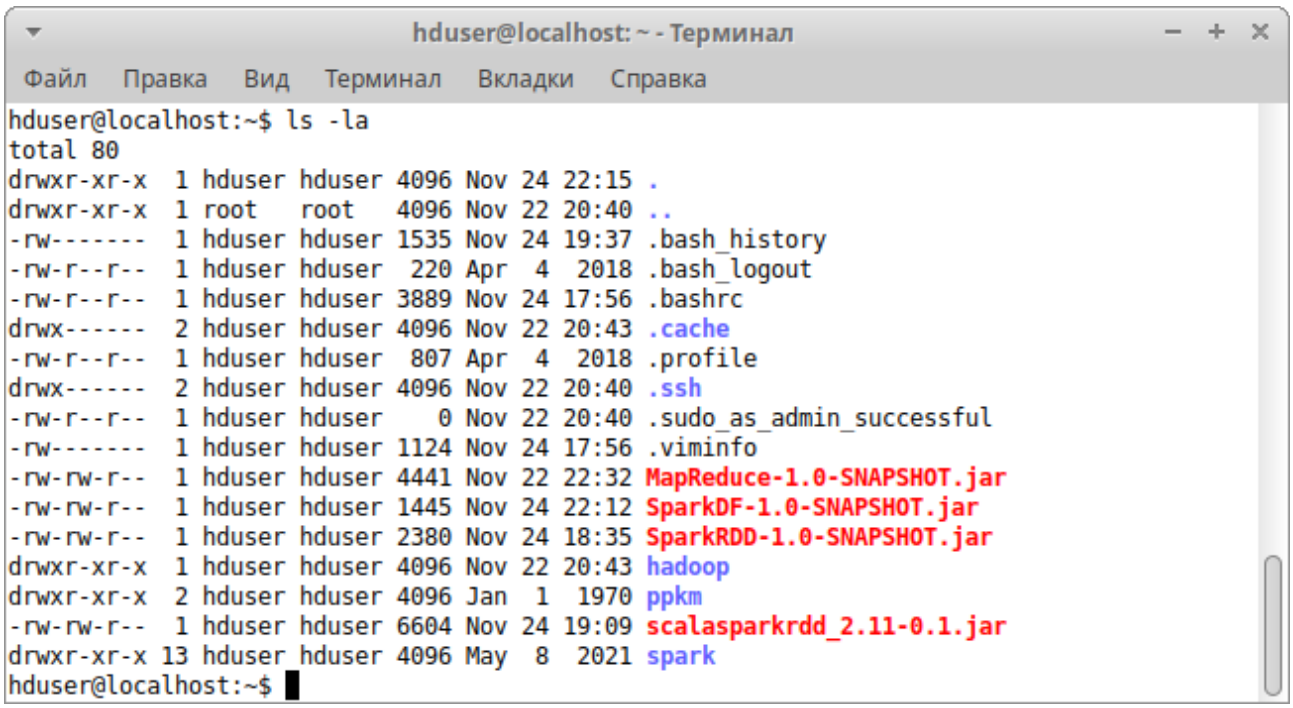

SparkDF – WordCount.java [SparkDF.main]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDF › build › libs › SparkDF-1.0-SNAPSHOT.jar
SparkDF [build]
build.gradle (SparkDF)
WordCount.java
import org.apache.spark.sql.Row;
import org.apache.spark.sql.SaveMode;
import org.apache.spark.sql.SparkSession;
import static org.apache.spark.sql.functions.*;
public class WordCount {
public static void main(String[] args) {
final String input = args[0];
final String output = args[1];
SparkSession spark = SparkSession.builder().getOrCreate();
Dataset<Row> df = spark.read().option("header", "true").csv(input);
Dataset<Row> wc = df.select(concat_ws(sep: " ", col(colName: "class"),
.select(split(col(colName: "docs"), pattern: "\\s").as("words"))
.select(explode(col(colName: "words")).as("word"))
.groupBy(col1: "word").count();
wc.write().mode(SaveMode.Overwrite).csv(output);
spark.stop();
Gradle
SparkDF
Tasks
build
assemble
build
buildDependents
buildNeeded
classes
clean
jar
testClasses
build setup
documentation
help
other
verification
Dependencies
Run: SparkDF [build]
SparkDF [build]: successful At 25.11.2021, 1:12 1 sec, 776 ms
> Task :check UP-TO-DATE
> Task :build
BUILD SUCCESSFUL in 1s
2 actionable tasks: 2 executed
1:12:12: Task execution finished 'build'.
Run TODO Problems Terminal Build Dependencies Event Log
Rainbow CSV: You can edit Rainbow CSV settings in Settings > Editor > General > Rainbow CSV (36 minutes ago)
10:4 (6 chars) LF UTF-8 4 spaces

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



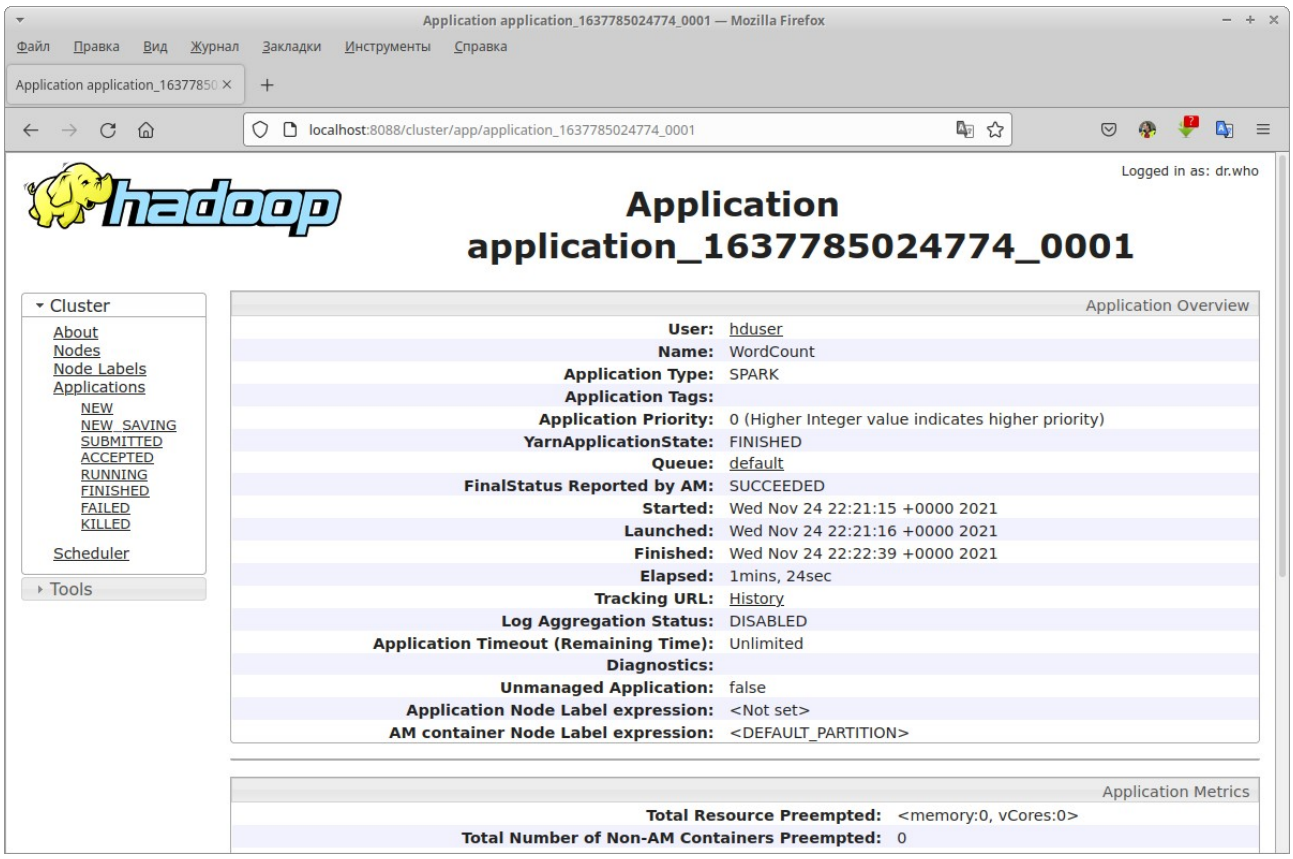


ls -la

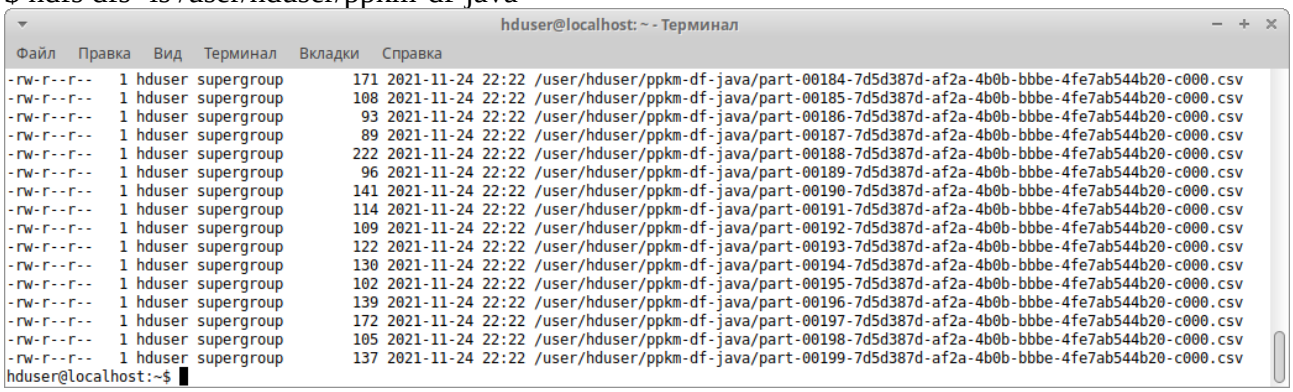


Запускаем приложение:
$ spark-submit \
--class WordCount \
--master yarn \
--deploy-mode cluster \
SparkDF-1.0-SNAPSHOT.jar \
/user/hduser/ppkm/ppkm_dataset.csv /user/hduser/ppkm-df-java

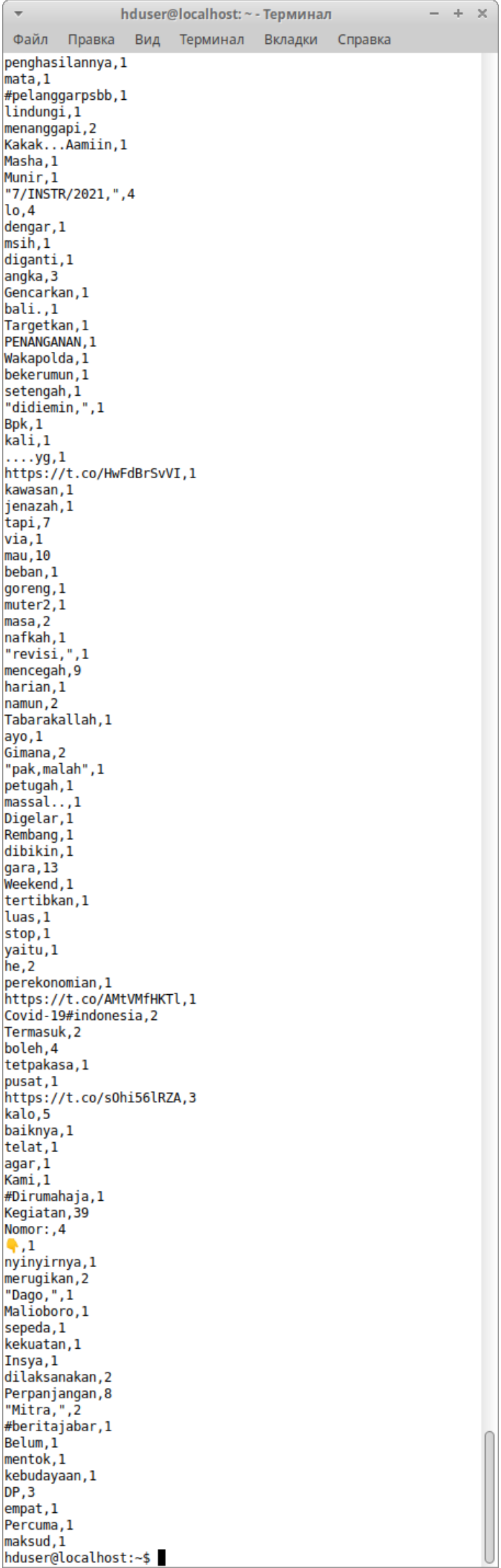
```
21/11/24 22:21:15 INFO yarn.Client: Submitting application application_1637785024774_0001 to ResourceManager
21/11/24 22:21:15 INFO impl.YarnClientImpl: Submitted application application_1637785024774_0001
21/11/24 22:21:16 INFO yarn.Client: Application report for application_1637785024774_0001 (state: ACCEPTED)
21/11/24 22:21:16 INFO yarn.Client:
         client token: N/A
         diagnostics: AM container is launched, waiting for AM container to Register with RM
         ApplicationMaster host: N/A
         ApplicationMaster RPC port: -1
         queue: default
         start time: 1637792475463
         final status: UNDEFINED
         tracking URL: http://localhost:8088/proxy/application_1637785024774_0001/
         user: hduser
21/11/24 22:21:17 INFO yarn.Client: Application report for application_1637785024774_0001 (state: ACCEPTED)
21/11/24 22:21:18 INFO yarn.Client: Application report for application_1637785024774_0001 (state: ACCEPTED)
21/11/24 22:21:19 INFO yarn.Client: Application report for application_1637785024774_0001 (state: ACCEPTED)
21/11/24 22:21:20 INFO yarn.Client: Application report for application_1637785024774_0001 (state: ACCEPTED)
21/11/24 22:21:21 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:21 INFO yarn.Client:
         client token: N/A
         diagnostics: N/A
         ApplicationMaster host: localhost
         ApplicationMaster RPC port: 34997
         queue: default
         start time: 1637792475463
         final status: UNDEFINED
         tracking URL: http://localhost:8088/proxy/application_1637785024774_0001/
         user: hduser
21/11/24 22:21:22 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:23 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:24 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:25 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:26 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:27 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:28 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:29 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:30 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:31 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:32 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:33 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:34 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:35 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:36 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:37 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:38 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:39 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:40 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:41 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:42 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:43 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:44 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:45 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:46 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:47 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:48 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:49 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:50 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:51 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:52 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:53 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:54 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:55 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:56 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:57 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:58 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:21:59 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:00 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:01 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:02 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:03 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:04 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:05 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:06 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:07 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:08 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:09 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:10 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:11 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:12 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:13 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:14 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:15 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:16 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:17 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:18 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:19 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:20 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:21 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:22 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:23 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:24 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:25 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:26 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:27 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:28 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:29 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:30 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:31 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:32 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:33 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:34 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:35 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:36 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:38 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:39 INFO yarn.Client: Application report for application_1637785024774_0001 (state: RUNNING)
21/11/24 22:22:40 INFO yarn.Client: Application report for application_1637785024774_0001 (state: FINISHED)
21/11/24 22:22:40 INFO yarn.Client:
         client token: N/A
         diagnostics: N/A
         ApplicationMaster host: localhost
         ApplicationMaster RPC port: 34997
         queue: default
         start time: 1637792475463
         final status: SUCCEEDED
         tracking URL: http://localhost:8088/proxy/application_1637785024774_0001/
         user: hduser
21/11/24 22:22:40 INFO util.ShutdownHookManager: Shutdown hook called
21/11/24 22:22:40 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-1d116ebe-707a-405b-85ba-82db26ee9496
21/11/24 22:22:40 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-0850b4e8-1dc1-45e8-a412-d66a2e7bf9b0
hduser@localhost:~$
```



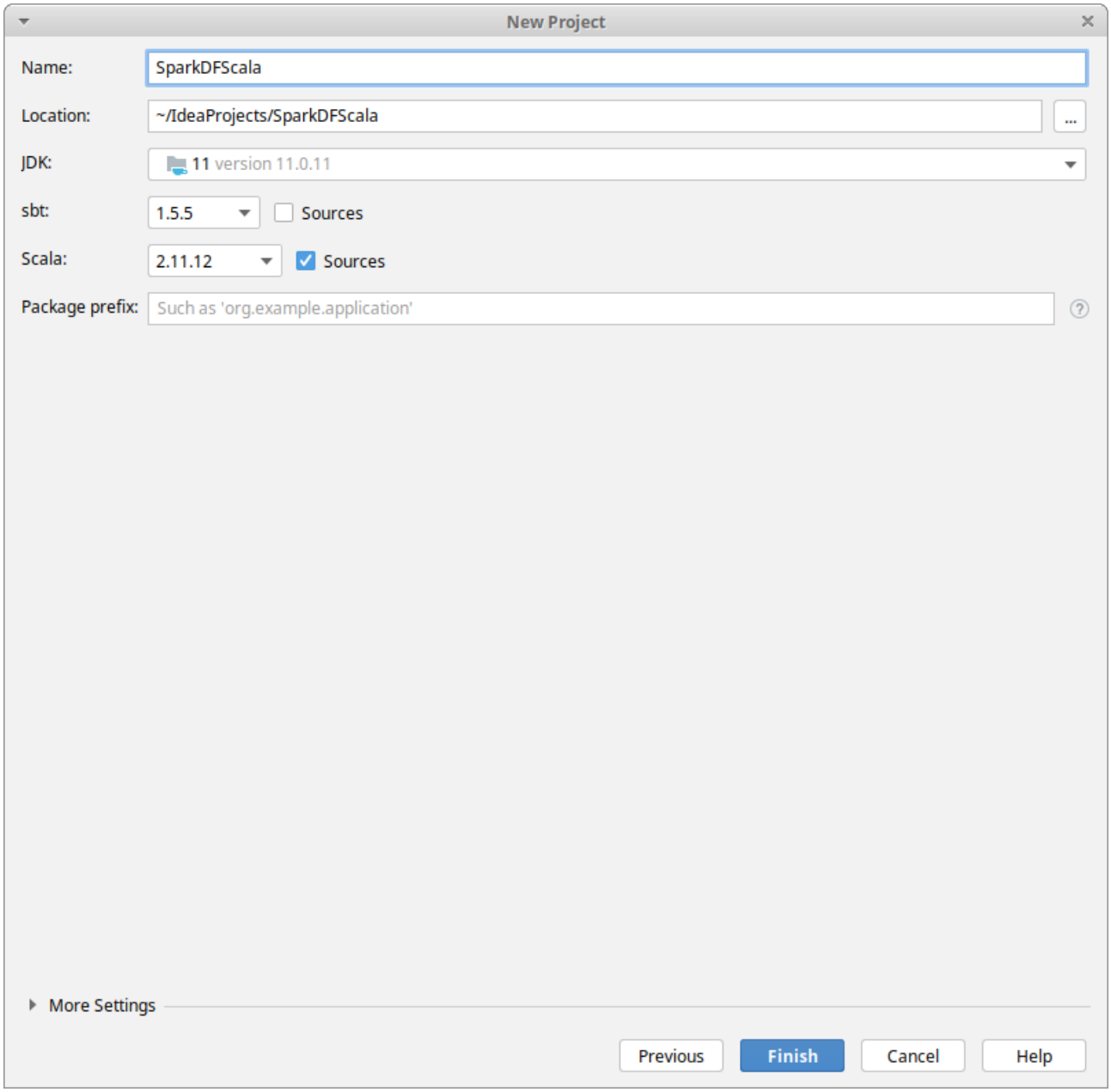
Смотрим результат:
$ hdfs dfs -ls /user/hduser/ppkm-df-java



$ hdfs dfs -cat /user/hduser/ppkm-df-java/*

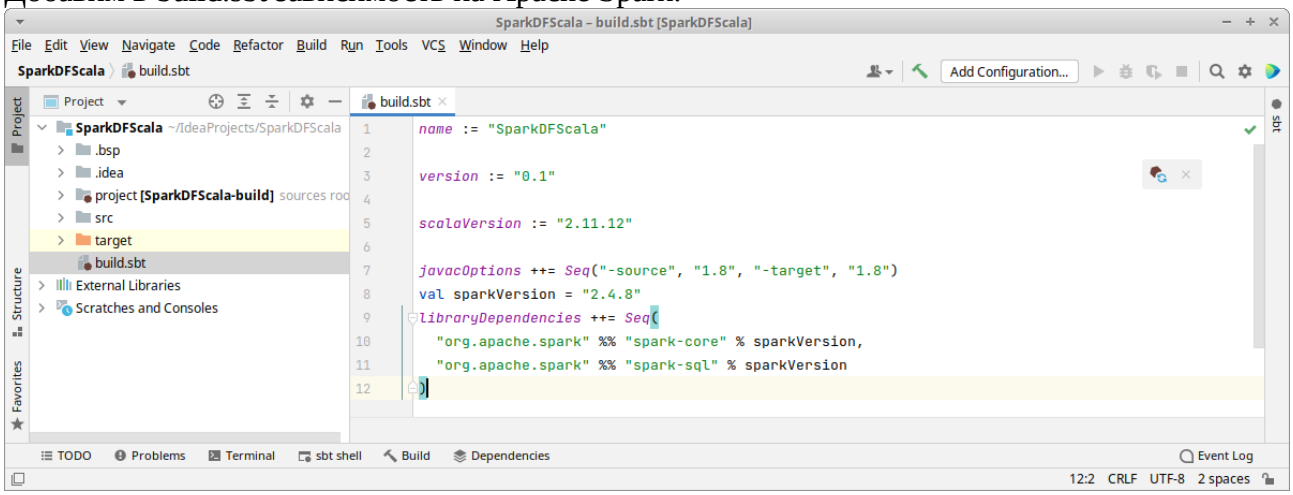

hduser@localhost: ~ - Терминал
Файл Правка Вид Терминал Вкладки Справка
penghasilannya,1
mata,1
#pelanggarpsbb,1
lindungi,1
menanggapi,2
Kakak...Aamiin,1
Masha,1
Munir,1
"7/INSTR/2021,",4
lo,4
dengar,1
msih,1
diganti,1
angka,3
Gencarkan,1
bali.,1
Targetkan,1
PENANGANAN,1
Wakapolda,1
bekerumun,1
setengah,1
"didiemin,",1
Bpk,1
kali,1
....yg,1
https://t.co/HwFdBrSvVI,1
kawasan,1
jenazah,1
tapi,7
via,1
mau,10
beban,1
goreng,1
muter2,1
masa,2
nafkah,1
"revisi,",1
mencegah,9
harian,1
namun,2
Tabarakallah,1
ayo,1
Gimana,2
"pak,malah",1
petugah,1
massal..,1
Digelar,1
Rembang,1
dibikin,1
gara,13
Weekend,1
tertibkan,1
luas,1
stop,1
yaitu,1
he,2
perekonomian,1
https://t.co/AMtVMfHKTl,1
Covid-19#indonesia,2
Termasuk,2
boleh,4
tetpakasa,1
pusat,1
https://t.co/sOhi56lRZA,3
kalo,5
baiknya,1
telat,1
agar,1
Kami,1
#Dirumahaja,1
Kegiatan,39
Nomor:,4
👇,1
nyinyirnya,1
merugikan,2
"Dago,",1
Malioboro,1
sepeda,1
kekuatan,1
Insya,1
dilaksanakan,2
Perpanjangan,8
"Mitra,",2
#beritajabar,1
Belum,1
mentok,1
kebudayaan,1
DP,3
empat,1
Percuma,1
maksud,1
hduser@localhost:~$

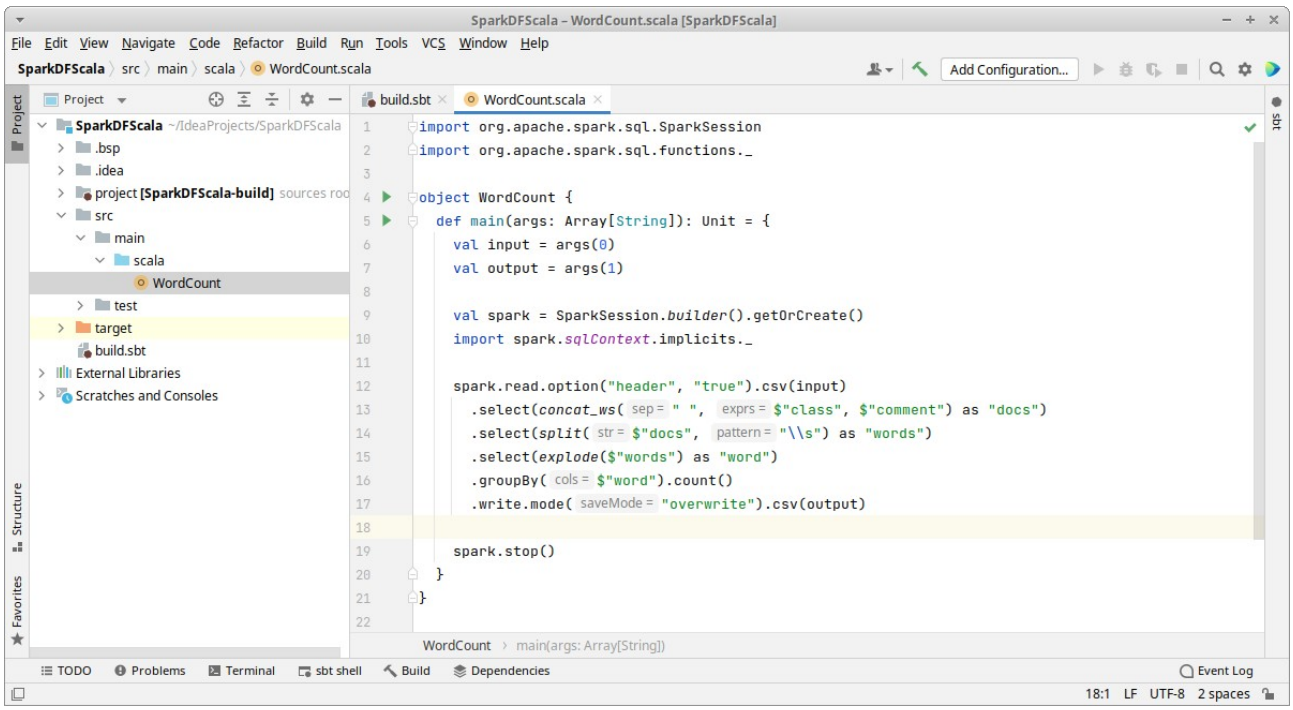
SCALA WORD COUNT: ПРОЕКТ В IDEA

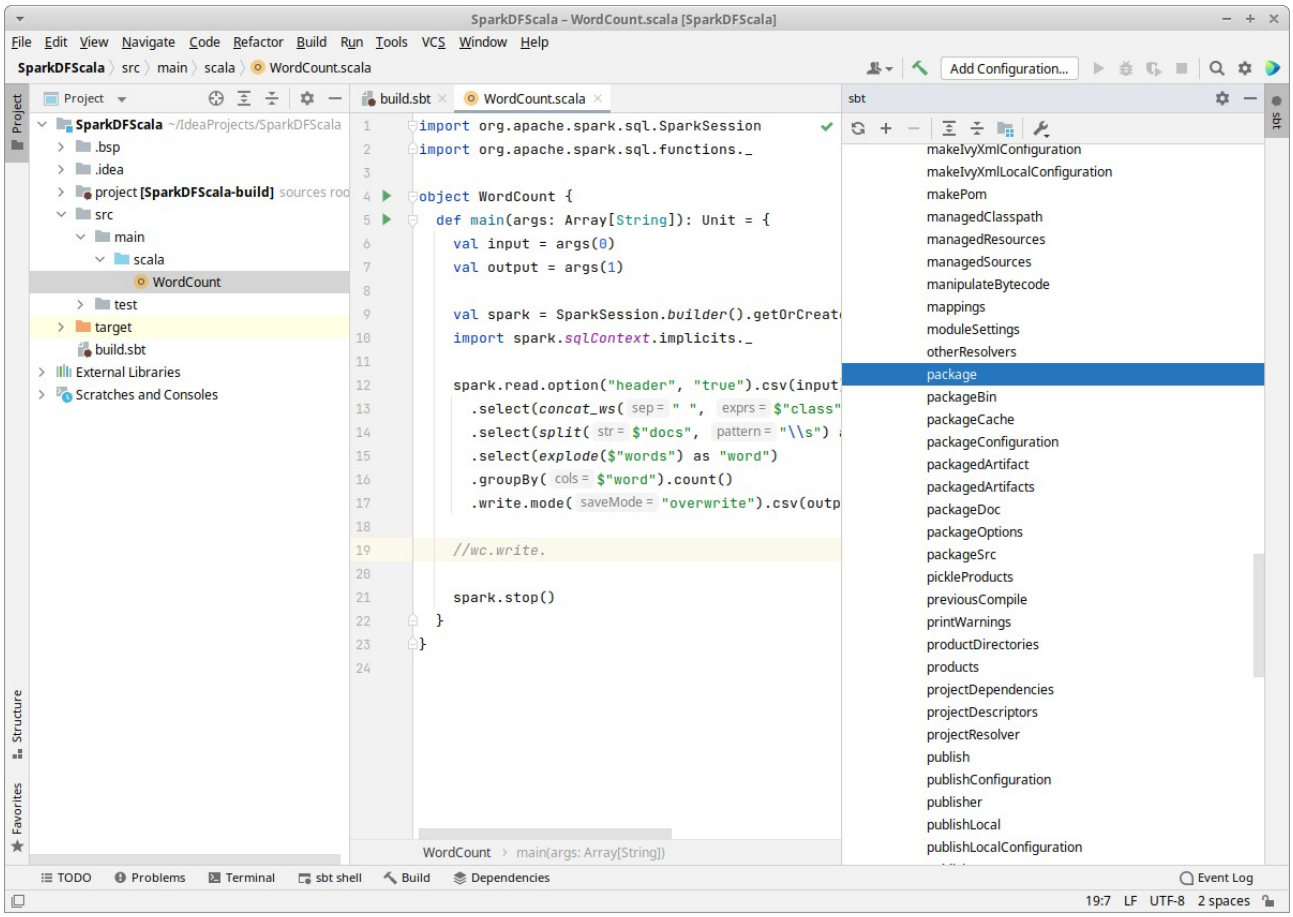


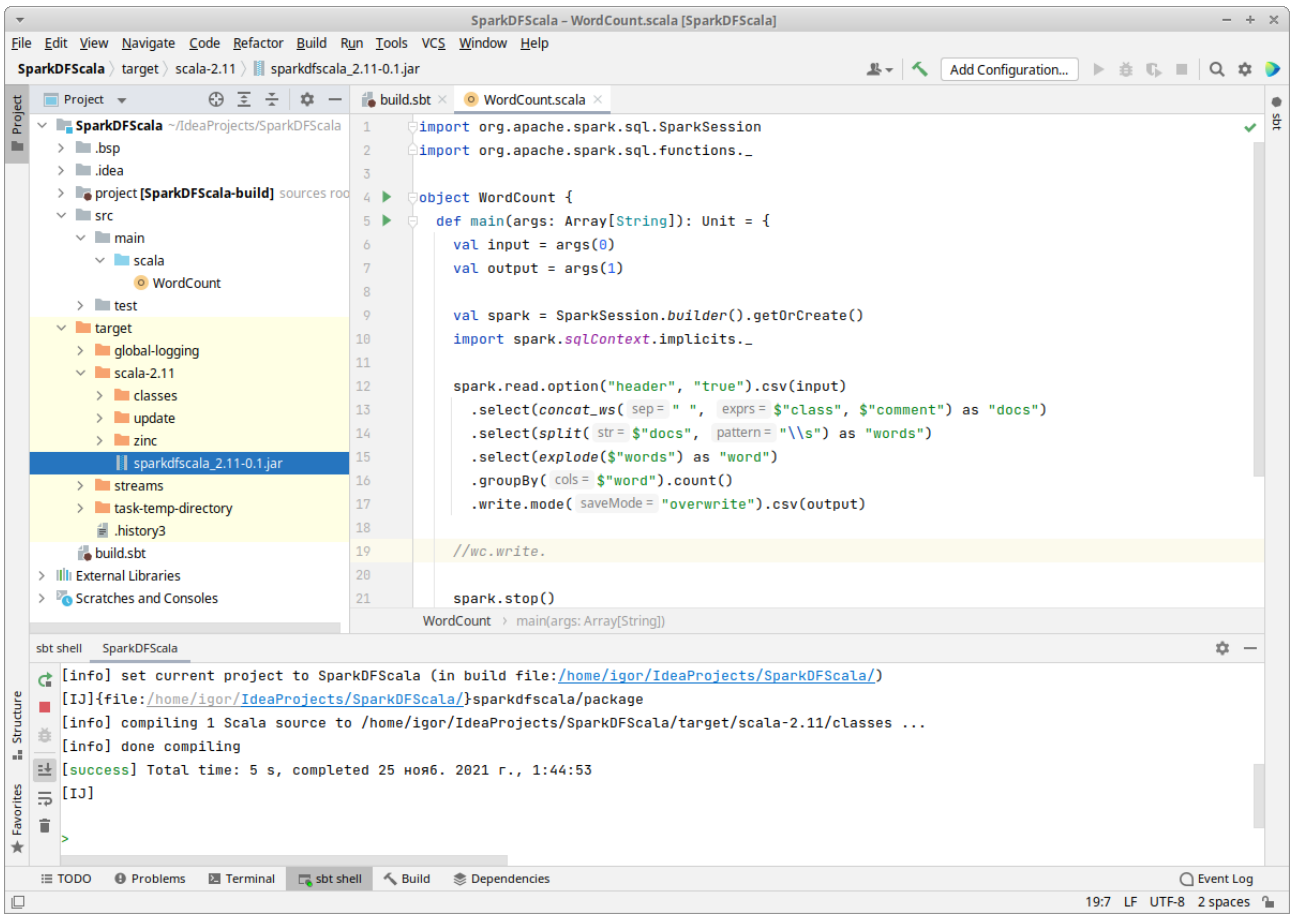
SCALA WORD COUNT: ЗАВИСИМОСТИ

Добавим в build.sbt зависимость на Apache Spark:

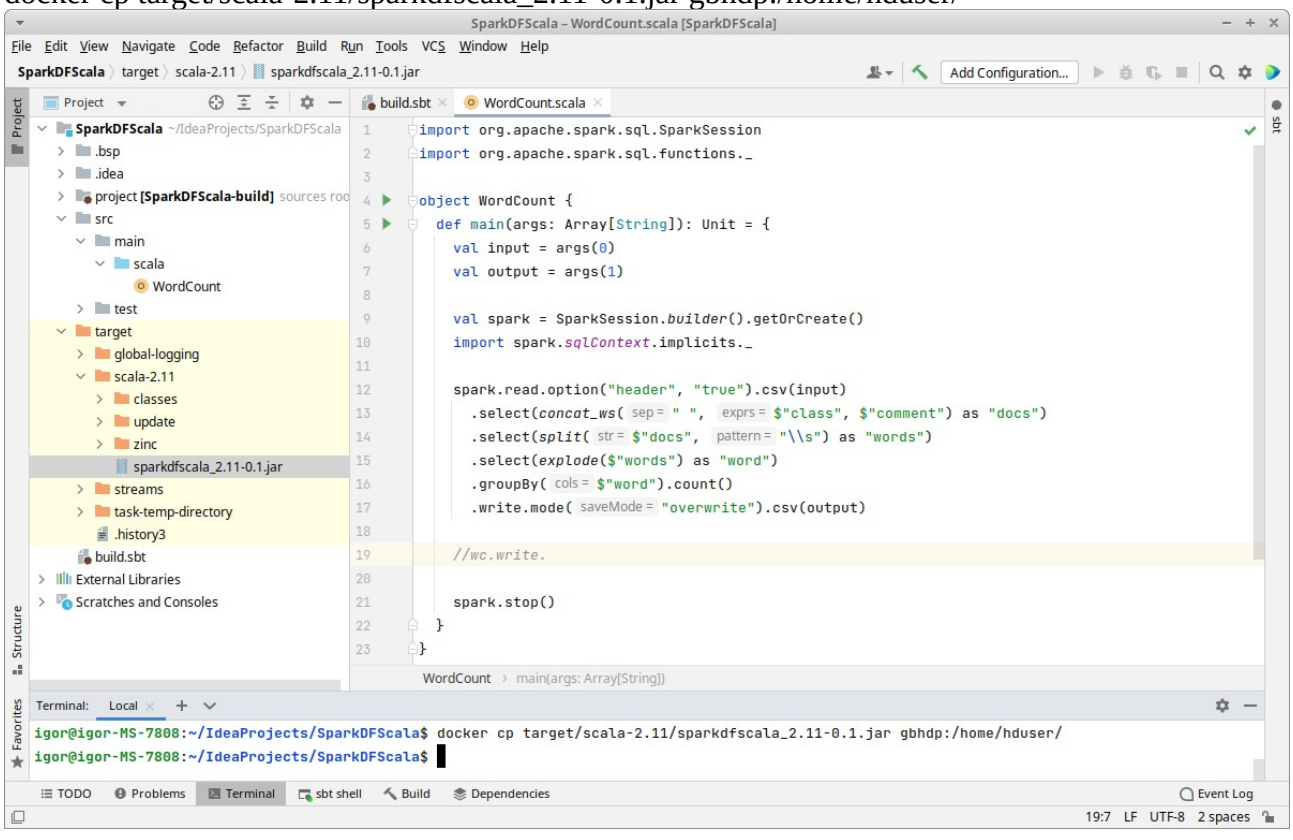


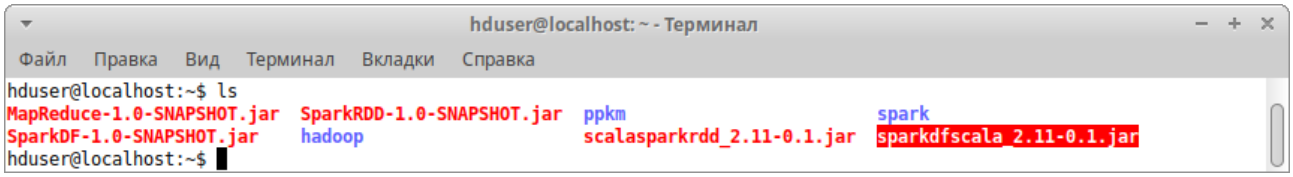

SparkDFScala – WordCount.scala [SparkDFScala]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDFScala src main scala WordCount.scala
Add Configuration...
Project
SparkDFScala ~/IdeaProjects/SparkDFScala
.bsp
.idea
project [SparkDFScala-build] sources roo
src
main
scala
WordCount
test
target
build.sbt
External Libraries
Scratches and Consoles
build.sbt
WordCount.scala
import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions._
object WordCount {
def main(args: Array[String]): Unit = {
val input = args(0)
val output = args(1)
val spark = SparkSession.builder().getOrCreate()
import spark.sqlContext.implicits._
spark.read.option("header", "true").csv(input)
.select(concat_ws(sep = " ", exprs = $"class", $"comment") as "docs")
.select(split(str = $"docs", pattern = "\\s") as "words")
.select(explode($"words") as "word")
.groupBy(cols = $"word").count()
.write.mode(saveMode = "overwrite").csv(output)
spark.stop()
}
}
WordCount main(args: Array[String])
TODO Problems Terminal sbt shell Build Dependencies Event Log
18:1 LF UTF-8 2 spaces
Structure
Favorites


SparkDFScala – WordCount.scala [SparkDFScala]
File Edit View Navigate Code Refactor Build Run Tools VCS Window Help
SparkDFScala src main scala WordCount.scala
Add Configuration...
Project
SparkDFScala ~/IdeaProjects/SparkDFScala
.bsp
.idea
project [SparkDFScala-build] sources roo
src
main
scala
WordCount
test
target
build.sbt
External Libraries
Scratches and Consoles
build.sbt
WordCount.scala
import org.apache.spark.sql.SparkSession
import org.apache.spark.sql.functions._
object WordCount {
def main(args: Array[String]): Unit = {
val input = args(0)
val output = args(1)
val spark = SparkSession.builder().getOrCreat
import spark.sqlContext.implicits._
spark.read.option("header", "true").csv(input
.select(concat_ws(sep = " ", exprs = $"class"
.select(split(str = $"docs", pattern = "\\s")
.select(explode($"words") as "word")
.groupBy(cols = $"word").count()
.write.mode(saveMode = "overwrite").csv(outp
//wc.write.
spark.stop()
}
}
WordCount main(args: Array[String])
sbt
makeIvyXmlConfiguration
makeIvyXmlLocalConfiguration
makePom
managedClasspath
managedResources
managedSources
manipulateBytecode
mappings
moduleSettings
otherResolvers
package
packageBin
packageCache
packageConfiguration
packagedArtifact
packagedArtifacts
packageDoc
packageOptions
packageSrc
pickleProducts
previousCompile
printWarnings
productDirectories
products
projectDependencies
projectDescriptors
projectResolver
publish
publishConfiguration
publisher
publishLocal
publishLocalConfiguration
TODO Problems Terminal sbt shell Build Dependencies Event Log
19:7 LF UTF-8 2 spaces
Structure
Favorites



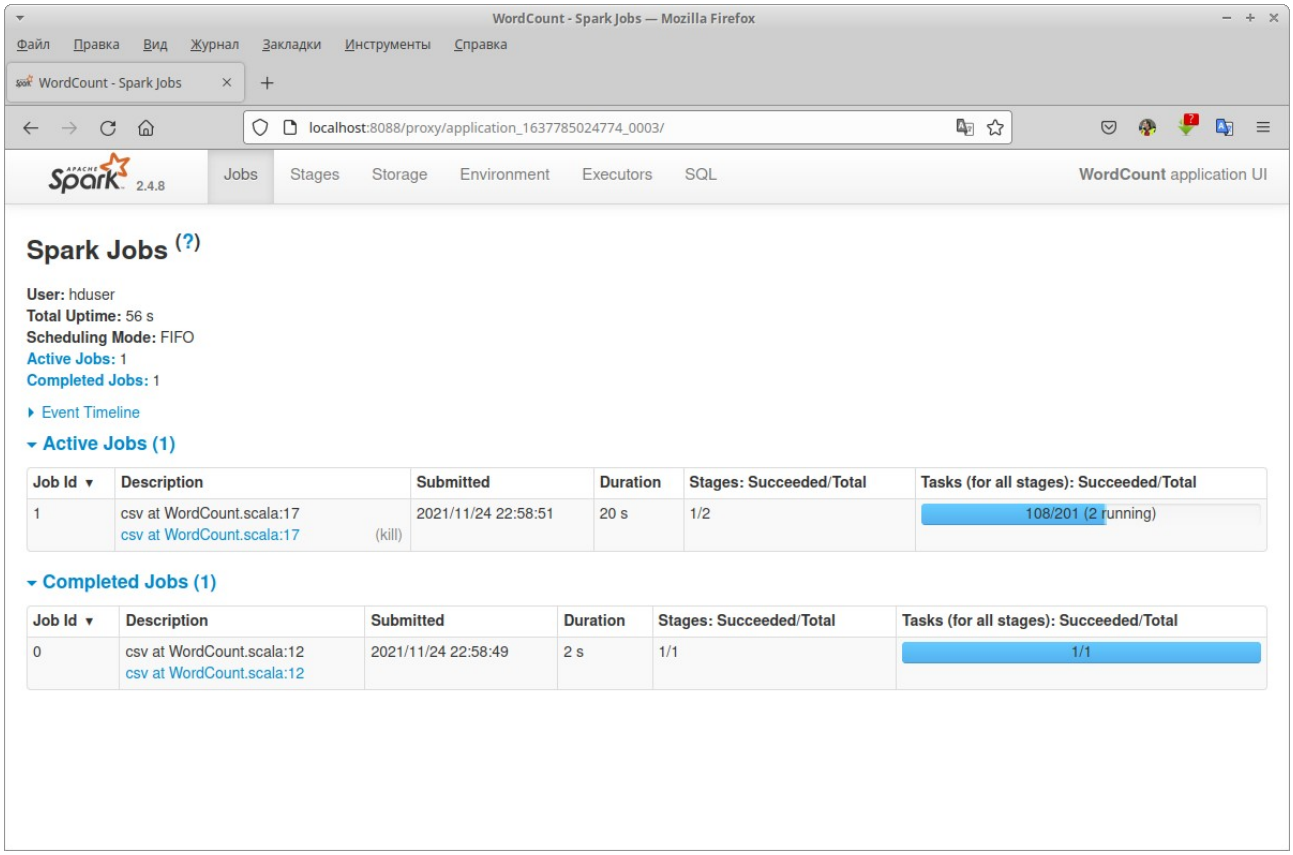
docker cp target/scala-2.11/sparkdfscala_2.11-0.1.jar gbhdp:/home/hduser/





SCALA WORD COUNT: ЗАПУСК
Запускаем приложение:

spark-submit \
--class WordCount \
--master yarn \
--deploy-mode cluster \
sparkdfscala_2.11-0.1.jar \
/user/hduser/ppkm/ppkm**_dataset.csv** /user/hduser/ppkm-df-scala

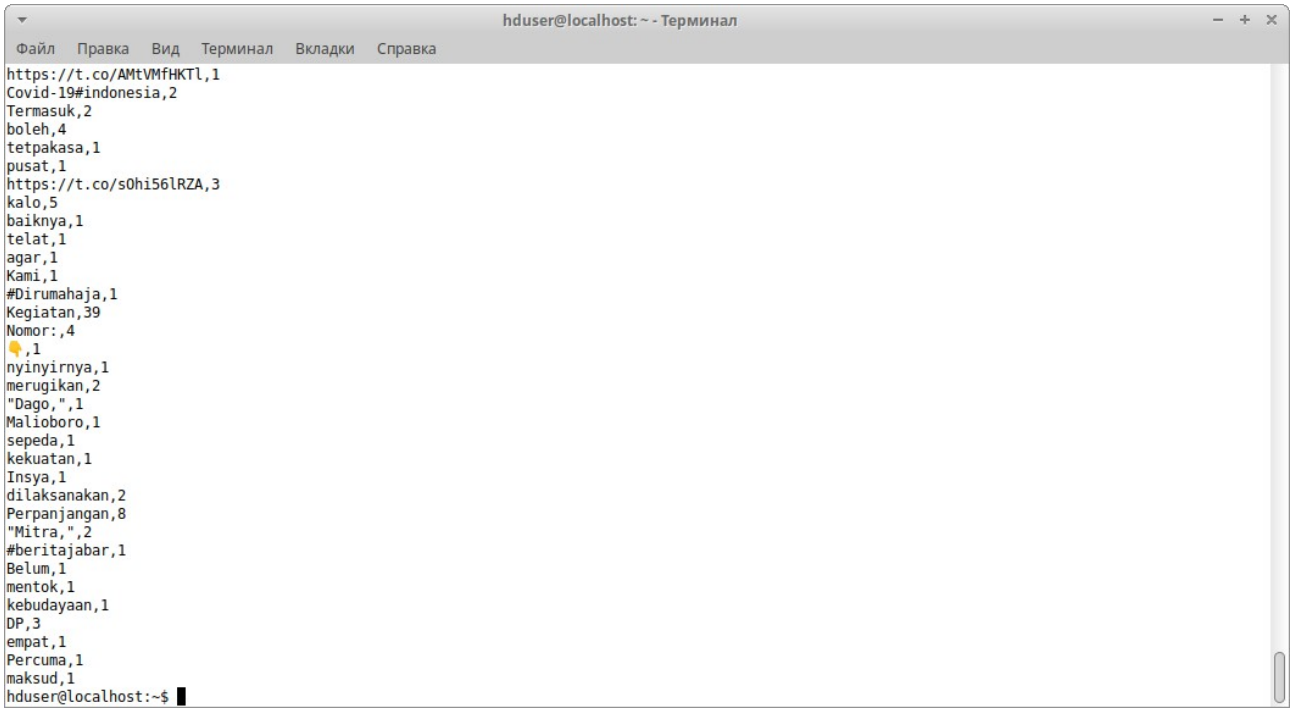


```
	client token: N/A
	diagnostics: AM container is launched, waiting for AM container to Register with RM
	ApplicationMaster host: N/A
	ApplicationMaster RPC port: -1
	queue: default
	start time: 1637794692338
	final status: UNDEFINED
	tracking URL: http://localhost:8088/proxy/application_1637785024774_0003/
	user: hduser
21/11/24 22:58:14 INFO yarn.Client: Application report for application_1637785024774_0003 (state: ACCEPTED)
21/11/24 22:58:15 INFO yarn.Client: Application report for application_1637785024774_0003 (state: ACCEPTED)
21/11/24 22:58:16 INFO yarn.Client: Application report for application_1637785024774_0003 (state: ACCEPTED)
21/11/24 22:58:17 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:17 INFO yarn.Client:
	client token: N/A
	diagnostics: N/A
	ApplicationMaster host: localhost
	ApplicationMaster RPC port: 44617
	queue: default
	start time: 1637794692338
	final status: UNDEFINED
	tracking URL: http://localhost:8088/proxy/application_1637785024774_0003/
	user: hduser
21/11/24 22:58:18 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:19 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:20 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:21 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:22 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:23 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:24 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:25 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:26 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:27 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:28 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:29 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:30 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:31 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:32 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:33 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:34 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:35 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:36 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:37 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:38 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:39 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:40 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:41 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:42 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:43 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:44 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:45 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:46 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:47 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:48 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:49 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:50 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:51 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:52 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:53 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:54 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:55 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:56 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:57 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:58 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:58:59 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:00 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:01 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:02 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:03 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:04 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:05 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:06 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:07 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:08 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:09 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:10 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:11 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:12 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:13 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:14 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:15 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:16 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:17 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:18 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:19 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:20 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:21 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:22 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:23 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:24 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:25 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:26 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:27 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:28 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:29 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:30 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:31 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:32 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:33 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:34 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:35 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:36 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:37 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:38 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:39 INFO yarn.Client: Application report for application_1637785024774_0003 (state: RUNNING)
21/11/24 22:59:40 INFO yarn.Client: Application report for application_1637785024774_0003 (state: FINISHED)
21/11/24 22:59:40 INFO yarn.Client:
	client token: N/A
	diagnostics: N/A
	ApplicationMaster host: localhost
	ApplicationMaster RPC port: 44617
	queue: default
	start time: 1637794692338
	final status: SUCCEEDED
	tracking URL: http://localhost:8088/proxy/application_1637785024774_0003/
	user: hduser
21/11/24 22:59:40 INFO util.ShutdownHookManager: Shutdown hook called
21/11/24 22:59:40 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-3ca436eb-a002-4ede-a9d8-aca5c4921ec2
21/11/24 22:59:40 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-3ce6fb59-fda9-4b01-96ee-6fc7cacf3f25
hduser@localhost:~$
```

hduser@localhost: ~ - Терминал

Файл Правка Вид Терминал Вкладки Справка

```
-rw-r--r--   1 hduser supergroup         82 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00081-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         94 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00082-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        135 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00083-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        145 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00084-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        133 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00085-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        132 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00086-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        162 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00087-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         87 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00088-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        133 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00089-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        102 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00090-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         59 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00091-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         49 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00092-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         95 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00093-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        142 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00094-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        107 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00095-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         64 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00096-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         85 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00097-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        135 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00098-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        135 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00099-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         79 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00100-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        131 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00101-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        180 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00102-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        133 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00103-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        116 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00104-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         67 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00105-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        114 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00106-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        207 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00107-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        169 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00108-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         91 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00109-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         95 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00110-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        160 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00111-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         89 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00112-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        178 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00113-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         80 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00114-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        111 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00115-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         60 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00116-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         60 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00117-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        102 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00118-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         80 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00119-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         83 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00120-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        107 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00121-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        115 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00122-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         50 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00123-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        125 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00124-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        103 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00125-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        175 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00126-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        137 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00127-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        108 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00128-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         99 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00129-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        115 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00130-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        144 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00131-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        148 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00132-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        153 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00133-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        162 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00134-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        186 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00135-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        104 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00136-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        100 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00137-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         43 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00138-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         65 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00139-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        209 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00140-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        155 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00141-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         75 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00142-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        150 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00143-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        103 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00144-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         63 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00145-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        126 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00146-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        129 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00147-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         79 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00148-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        146 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00149-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        177 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00150-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         56 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00151-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         86 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00152-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         61 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00153-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        130 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00154-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        131 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00155-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        150 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00156-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         81 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00157-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        306 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00158-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         56 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00159-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        150 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00160-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        187 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00161-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         46 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00162-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        227 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00163-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        102 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00164-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        139 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00165-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         76 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00166-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         38 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00167-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         83 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00168-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        115 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00169-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        102 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00170-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        226 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00171-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        163 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00172-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         93 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00173-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        148 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00174-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         87 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00175-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        167 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00176-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        149 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00177-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        104 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00178-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        209 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00179-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         96 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00180-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         94 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00181-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        107 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00182-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        216 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00183-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        171 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00184-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        108 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00185-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         93 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00186-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         89 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00187-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        222 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00188-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup         96 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00189-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        141 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00190-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        114 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00191-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        109 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00192-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        122 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00193-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        130 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00194-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        102 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00195-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        139 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00196-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        172 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00197-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        105 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00198-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
-rw-r--r--   1 hduser supergroup        137 2021-11-24 22:59 /user/hduser/ppkm-df-scala/part-00199-ccccb2c9-73a8-4768-a78f-638a20e7be7e-c000.csv
hduser@localhost:~$
```


hduser@localhost: ~ - Терминал
Файл Правка Вид Терминал Вкладки Справка
https://t.co/AMtVMfHKTl,1
Covid-19#indonesia,2
Termasuk,2
boleh,4
tetpakasa,1
pusat,1
https://t.co/sOhi56lRZA,3
kalo,5
baiknya,1
telat,1
agar,1
Kami,1
#Dirumahaja,1
Kegiatan,39
Nomor:,4
👇,1
nyinyirnya,1
merugikan,2
"Dago,",1
Malioboro,1
sepeda,1
kekuatan,1
Insya,1
dilaksanakan,2
Perpanjangan,8
"Mitra,",2
#beritajabar,1
Belum,1
mentok,1
kebudayaan,1
DP,3
empat,1
Percuma,1
maksud,1
hduser@localhost:~$