1. Запустил свой скрипт. Все получилось.

```
[BD_243_pstroganov@bigdataanalytics-worker-2 ~]$ /spark2.4/bin/spark-submit --driver-m
emory 512m --driver-cores 1 --master local[1] my_script.py
Warning: Ignoring non-Spark config property: hive.metastore.uris
20/12/29 10:05:10 WARN util.NativeCodeLoader: Unable to load native-hadoop library for
 your platform... using builtin-java classes where applicable
20/12/29 10:05:11 INFO spark.SparkContext: Running Spark version 2.4.7
20/12/29 10:05:11 INFO spark.SparkContext: Submitted application: my_spark
20/12/29 10:05:11 INFO spark.SecurityManager: Changing view acls to: BD_243_pstroganov
20/12/29 10:05:11 INFO spark.SecurityManager: Changing modify acls to: BD_243_pstrogan
ov
20/12/29 10:05:11 INFO spark.SecurityManager: Changing view acls groups to:
20/12/29 10:05:11 INFO spark.SecurityManager: Changing modify acls groups to:
20/12/29 10:05:11 INFO spark.SecurityManager: SecurityManager: authentication disabled
; ui acls disabled; users  with view permissions: Set(BD_243_pstroganov); groups with
view permissions: Set(); users  with modify permissions: Set(BD_243_pstroganov); group
s with modify permissions: Set()
20/12/29 10:05:11 INFO util.Utils: Successfully started service 'sparkDriver' on port
45121.
20/12/29 10:05:11 INFO spark.SparkEnv: Registering MapOutputTracker
20/12/29 10:05:11 INFO spark.SparkEnv: Registering BlockManagerMaster
20/12/29 10:05:11 INFO storage.BlockManagerMasterEndpoint: Using org.apache.spark.stor
age.DefaultTopologyMapper for getting topology information
20/12/29 10:05:11 INFO storage.BlockManagerMasterEndpoint: BlockManagerMasterEndpoint
up
20/12/29 10:05:11 INFO storage.DiskBlockManager: Created local directory at /tmp/block
mgr-cbc844d2-45ff-46c4-8b43-390fc7e8cc36
20/12/29 10:05:11 INFO memory.MemoryStore: MemoryStore started with capacity 93.3 MB
20/12/29 10:05:11 INFO spark.SparkEnv: Registering OutputCommitCoordinator
20/12/29 10:05:11 INFO util.log: Logging initialized @2497ms
20/12/29 10:05:11 INFO server.Server: jetty-9.3.z-SNAPSHOT, build timestamp: unknown,
git hash: unknown
20/12/29 10:05:11 INFO server.Server: Started @2588ms
20/12/29 10:05:11 INFO server.AbstractConnector: Started ServerConnector@4d304898{HTTP
/1.1,[http/1.1]}{0.0.0.0:4040}
20/12/29 10:05:11 INFO util.Utils: Successfully started service 'SparkUI' on port 4040
.
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1
702668f{/jobs,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4
b08994b{/jobs/json,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@a
d41531{/jobs/job,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6
ce2b1f6{/jobs/job/json,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@6
3afbfe0{/stages,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4
9784339{/stages/json,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@2
afb93f1{/stages/stage,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4
6cbfc6d{/stages/stage/json,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@3
0d99b5e{/stages/pool,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@1
cc8123e{/stages/pool/json,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@5
9570c1b{/storage,null,AVAILABLE,@Spark}
20/12/29 10:05:11 INFO handler.ContextHandler: Started o.s.j.s.ServletContextHandler@4
```

```
20/12/29 10:05:19 INFO parquet.ParquetWriteSupport: Initialized Parquet WriteSupport w
ith Catalyst schema:
{
  "type" : "struct",
  "fields" : [ {
    "name" : "Name",
    "type" : "string",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "Author",
    "type" : "string",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "User",
    "type" : "string",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "Rating",
    "type" : "float",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "Reviews",
    "type" : "integer",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "Price",
    "type" : "integer",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "Year",
    "type" : "integer",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "Genre",
    "type" : "string",
    "nullable" : true,
    "metadata" : { }
  }, {
    "name" : "p_date",
    "type" : "string",
    "nullable" : false,
    "metadata" : { }
  } ]
}
and corresponding Parquet message type:
message spark_schema {
  optional binary Name (UTF8);
  optional binary Author (UTF8);
  optional binary User (UTF8);
  optional float Rating;
  optional int32 Reviews;
```

```
  optional float Rating;
  optional int32 Reviews;
  optional int32 Price;
  optional int32 Year;
  optional binary Genre (UTF8);
  required binary p_date (UTF8);
}


20/12/29 10:05:19 INFO compress.CodecPool: Got brand-new compressor [.snappy]
20/12/29 10:05:20 INFO datasources.FileScanRDD: Reading File path: hdfs://bigdataanaly
tics-head-0.novalocal:8020/user/BD_243_pstroganov/for_stream, range: 0-3145, partition
 values: [empty row]
20/12/29 10:05:20 INFO codegen.CodeGenerator: Code generated in 43.055564 ms
20/12/29 10:05:20 WARN csv.CSVDataSource: Number of column in CSV header is not equal
to number of fields in the schema:
 Header length: 7, schema size: 8
CSV file: hdfs://bigdataanalytics-head-0.novalocal:8020/user/BD_243_pstroganov/for_str
eam
20/12/29 10:05:20 INFO hadoop.InternalParquetRecordWriter: Flushing mem columnStore to
 file. allocated memory: 149
20/12/29 10:05:20 INFO output.FileOutputCommitter: Saved output of task 'attempt_20201
229100519_0000_m_000000_0' to hdfs://bigdataanalytics-head-0.novalocal:8020/user/BD_24
3_pstroganov/my_submit_parquet_files/p_date=20201229100517/_temporary/0/task_202012291
00519_0000_m_000000
20/12/29 10:05:20 INFO mapred.SparkHadoopMapRedUtil: attempt_20201229100519_0000_m_000
000_0: Committed
20/12/29 10:05:20 INFO executor.Executor: Finished task 0.0 in stage 0.0 (TID 0). 2245
 bytes result sent to driver
20/12/29 10:05:20 INFO scheduler.TaskSetManager: Finished task 0.0 in stage 0.0 (TID 0
) in 1463 ms on localhost (executor driver) (1/1)
20/12/29 10:05:20 INFO scheduler.DAGScheduler: ResultStage 0 (parquet at NativeMethodA
ccessorImpl.java:0) finished in 1.640 s
20/12/29 10:05:20 INFO scheduler.TaskSchedulerImpl: Removed TaskSet 0.0, whose tasks h
ave all completed, from pool
20/12/29 10:05:20 INFO scheduler.DAGScheduler: Job 0 finished: parquet at NativeMethod
AccessorImpl.java:0, took 1.731869 s
20/12/29 10:05:21 INFO datasources.FileFormatWriter: Write Job c97bebb5-e7a6-49d0-8eb6
-0dfdd521d039 committed.
20/12/29 10:05:21 INFO datasources.FileFormatWriter: Finished processing stats for wri
FINISHED BATCH LOADING. TIME = 20201229100517
: Stopped Spark@4d304898{HTTP/1.1,[http
/1.1]}{0.0.0.0:4040}
20/12/29 10:05:21 INFO ui.SparkUI: Stopped Spark web UI at http://bigdataanalytics-wor
ker-2.novalocal:4040
20/12/29 10:05:21 INFO spark.MapOutputTrackerMasterEndpoint: MapOutputTrackerMasterEnd
point stopped!
20/12/29 10:05:21 INFO memory.MemoryStore: MemoryStore cleared
20/12/29 10:05:21 INFO storage.BlockManager: BlockManager stopped
20/12/29 10:05:21 INFO storage.BlockManagerMaster: BlockManagerMaster stopped
20/12/29 10:05:21 INFO scheduler.OutputCommitCoordinator$OutputCommitCoordinatorEndpoi
nt: OutputCommitCoordinator stopped!
20/12/29 10:05:21 INFO spark.SparkContext: Successfully stopped SparkContext
20/12/29 10:05:21 INFO util.ShutdownHookManager: Shutdown hook called
20/12/29 10:05:21 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-cf9d2ec
4-81c1-4f6f-9ba7-a62384000c59
20/12/29 10:05:21 INFO util.ShutdownHookManager: Deleting directory /tmp/spark-cf9d2ec
4-81c1-4f6f-9ba7-a62384000c59/pyspark-7772f030-77d7-479e-b1c3-0c243143139d
```