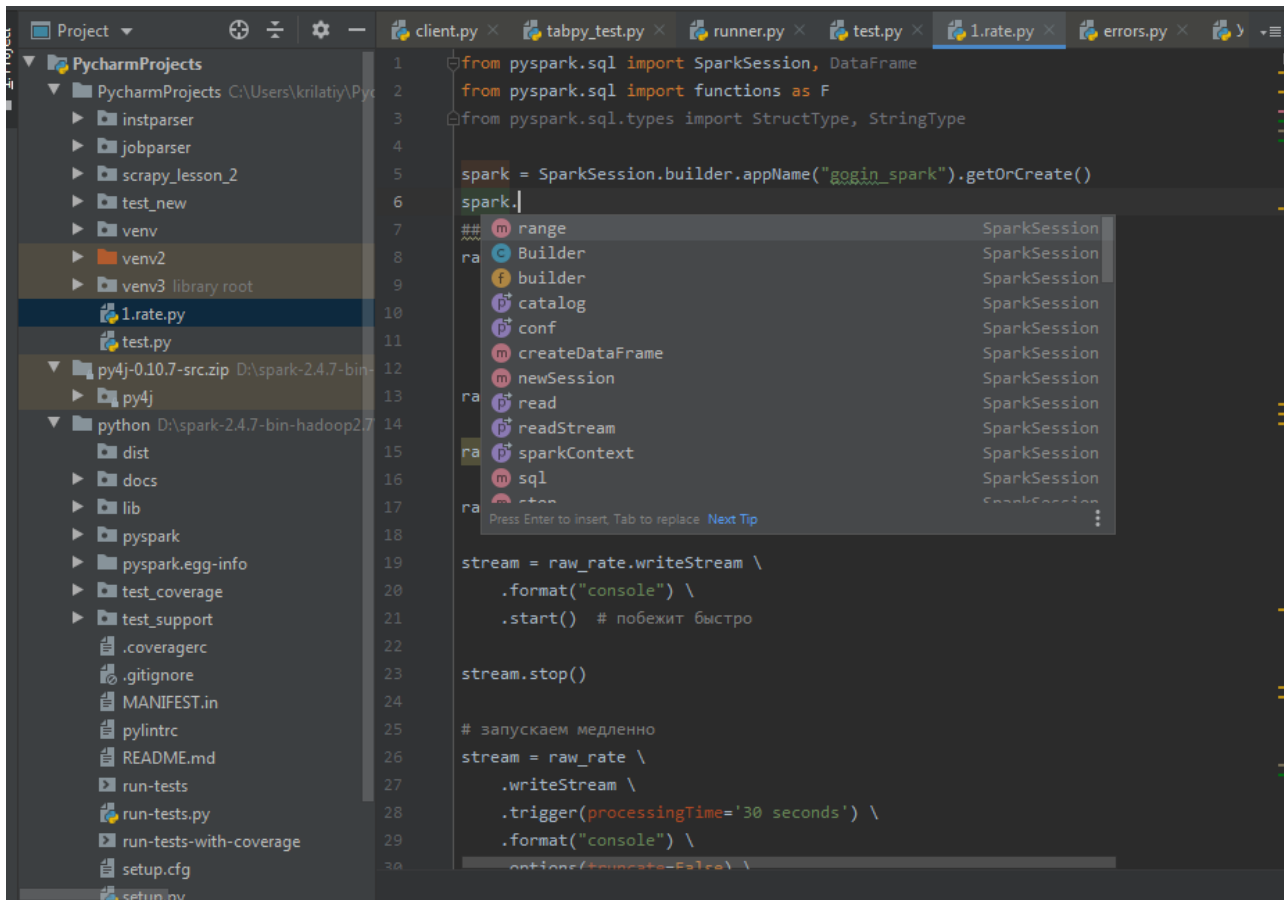
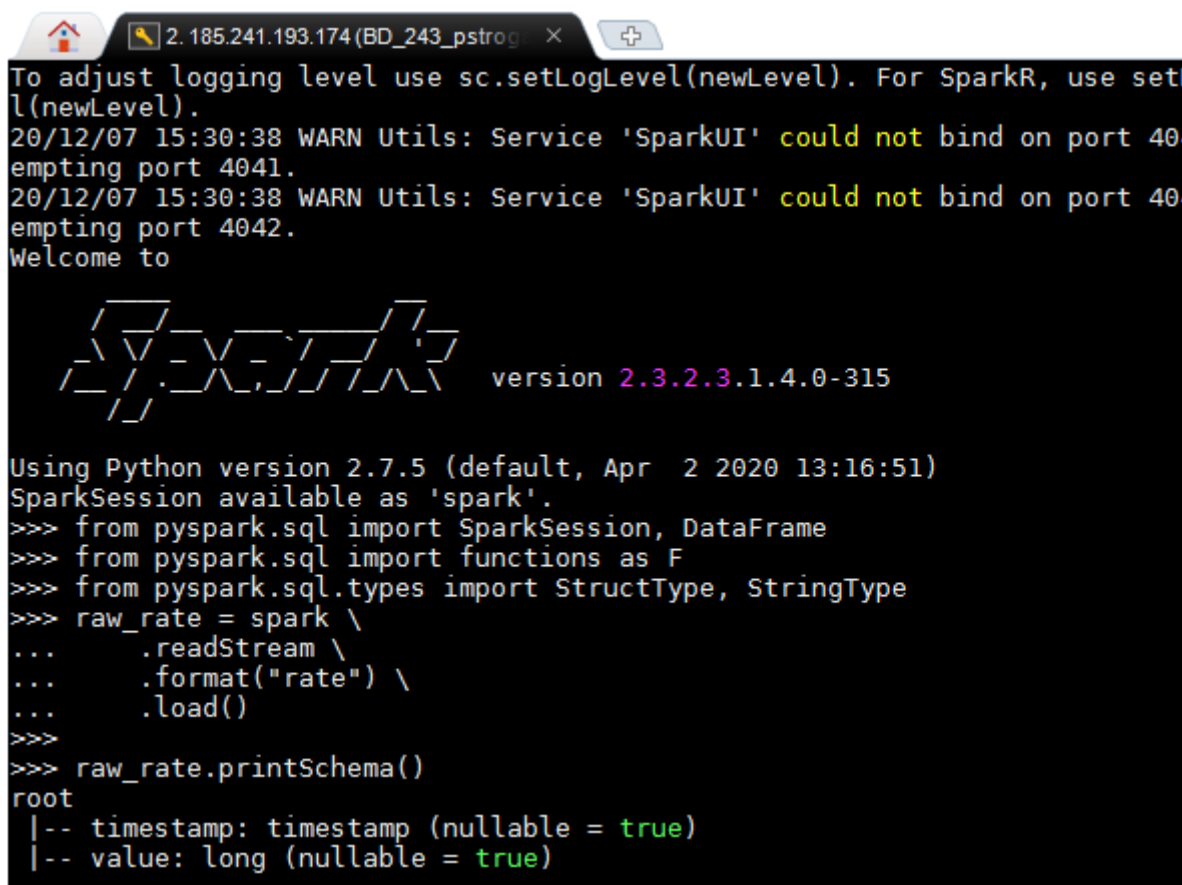


1. Скачал интерпретатор python 2.7 и установил для него pyspark с помощью видео инструкции.
2. В IDE открыл файл с занятия, интеллектуальный ввод подхватился.



3. Зашел под своим логином на сервер и как на занятии последовательно все ввел и получил такой же результат.



```

2. 185.241.193.174 (BD_243_pstrog X
>>> raw_rate.isStreaming
True
>>> raw_rate.show()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/usr/hdp/current/spark2-client/python/pyspark/sql/dataframe.py", line 350, in show
    print(self._jdf.showString(n, 20, vertical))
  File "/usr/hdp/current/spark2-client/python/lib/py4j-0.10.7-src.zip/py4j/java_gateway.py", line 1257, in __call__
  File "/usr/hdp/current/spark2-client/python/pyspark/sql/utils.py", line 69, in deco
    raise AnalysisException(s.split(':', 1)[1], stackTrace)
pyspark.sql.utils.AnalysisException: u'Queries with streaming sources must be executed with writeStream.start();;\\nrate'
>>> stream = raw_rate.writeStream \
...     .format("console") \
...     .start()
>>> -----
Batch: 0
-----
+-----+-----+
|timestamp|value|
+-----+-----+
-----

Batch: 1
-----
+-----+-----+
|          timestamp|value|
+-----+-----+
|2020-12-07 15:35:...|    0|
+-----+-----+
-----

Batch: 2
-----
+-----+-----+
|          timestamp|value|
+-----+-----+
|2020-12-07 15:35:...|    1|
|2020-12-07 15:35:...|    2|
|2020-12-07 15:35:...|    3|
|2020-12-07 15:35:...|    4|
|2020-12-07 15:35:...|    5|
+-----+-----+
-----

Batch: 3
-----
+-----+-----+
|          timestamp|value|
+-----+-----+
|2020-12-07 15:35:...|    6|
+-----+-----+
-----

Batch: 4

```

```

+-----+-----+
|      timestamp|value|
+-----+-----+
|2020-12-07 15:35:...|    7|
+-----+-----+

```

```

stream.stop()-----
Batch: 5

```

```

+-----+-----+
|      timestamp|value|
+-----+-----+
|2020-12-07 15:35:...|    8|
+-----+-----+

```

```

Batch: 6

```

```

+-----+-----+
|      timestamp|value|
+-----+-----+
|2020-12-07 15:35:...|    9|
+-----+-----+

```

```

20/12/07 15:35:16 WARN Client: interrupted waiting to send rpc request to server
java.lang.InterruptedException
    at java.util.concurrent.FutureTask.awaitDone(FutureTask.java:404)
    at java.util.concurrent.FutureTask.get(FutureTask.java:191)
    at org.apache.hadoop.ipc.Client$Connection.sendRpcRequest(Client.java:11
43)
    at org.apache.hadoop.ipc.Client.call(Client.java:1396)
    at org.apache.hadoop.ipc.Client.call(Client.java:1354)
    at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEng
ine.java:228)
    at org.apache.hadoop.ipc.ProtobufRpcEngine$Invoker.invoke(ProtobufRpcEng
ine.java:116)
    at com.sun.proxy.$Proxy11.addBlock(Unknown Source)
    at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolTranslatorPB.
addBlock(ClientNamenodeProtocolTranslatorPB.java:510)
    at sun.reflect.GeneratedMethodAccessor32.invoke(Unknown Source)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAcces
sorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invokeMethod(RetryI
nvocationHandler.java:422)
    at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeMethod(R
etryInvocationHandler.java:165)
    at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invoke(RetryIn
vocationHandler.java:157)
    at org.apache.hadoop.io.retry.RetryInvocationHandler$Call.invokeOnce(Ret
ryInvocationHandler.java:95)
    at org.apache.hadoop.io.retry.RetryInvocationHandler.invoke(RetryInvocat
ionHandler.java:359)
    at com.sun.proxy.$Proxy12.addBlock(Unknown Source)
    at org.apache.hadoop.hdfs.DFSOutputStream.addBlock(DFSOutputStream.java:
1078)
    at org.apache.hadoop.hdfs.DataStreamer.locateFollowingBlock(DataStreamer
.java:1865)

```

```

2.185.241.193.174 (BD_243_pstrog) x
>>> stream = raw_rate \
...     .writeStream \
...     .trigger(processingTime='30 seconds') \
...     .format("console") \
...     .options(truncate=False) \
...     .start()
>>> -----
Batch: 0
-----
+-----+-----+
|timestamp|value|
+-----+-----+
-----
Batch: 1
-----
+-----+-----+
|timestamp|value|
+-----+-----+
|2020-12-07 15:35:35.434|0|
|2020-12-07 15:35:36.434|1|
|2020-12-07 15:35:37.434|2|
|2020-12-07 15:35:38.434|3|
|2020-12-07 15:35:39.434|4|
|2020-12-07 15:35:40.434|5|
|2020-12-07 15:35:41.434|6|
|2020-12-07 15:35:42.434|7|
|2020-12-07 15:35:43.434|8|
|2020-12-07 15:35:44.434|9|
|2020-12-07 15:35:45.434|10|
|2020-12-07 15:35:46.434|11|
|2020-12-07 15:35:47.434|12|
|2020-12-07 15:35:48.434|13|
|2020-12-07 15:35:49.434|14|
|2020-12-07 15:35:50.434|15|
|2020-12-07 15:35:51.434|16|
|2020-12-07 15:35:52.434|17|
|2020-12-07 15:35:53.434|18|
|2020-12-07 15:35:54.434|19|
+-----+-----+
only showing top 20 rows

stream.explain()
== Physical Plan ==
WriteToDataSourceV2 org.apache.spark.sql.execution.streaming.sources.MicroBatchW
riter@70a76825
+- Scan ExistingRDD[timestamp#153,value#154L]
>>> stream.isActive
True
>>> stream.lastProgress
{u'stateOperators': [], u'name': None, u'timestamp': u'2020-12-07T15:36:00.000Z'
, u'processedRowsPerSecond': 90.22556390977444, u'inputRowsPerSecond': 0.9799918
334013884, u'numInputRows': 24, u'batchId': 1, u'sources': [{u'description': u'R
ateSource[rowsPerSecond=1, rampUpTimeSeconds=0, numPartitions=2]', u'endOffset':
24, u'processedRowsPerSecond': 90.22556390977444, u'inputRowsPerSecond': 0.9799
918334013884, u'numInputRows': 24, u'startOffset': 0}], u'durationMs': {u'queryP
lanning': 6, u'getOffset': 0, u'addBatch': 166, u'getBatch': 8, u'walCommit': 86
, u'triggerExecution': 266}, u'runId': u'effb5e2f-4d5f-434a-acf3-b35f2bfce2c6',
u'id': u'7b88227b-bcc3-402d-b5fc-ad77d0d94c78', u'sink': {u'description': u'org.

```

```

2.185.241.193.174 (BD_243_pstrog) x
>>> stream.status
{'message': u'Waiting for next trigger', u'isTriggerActive': False, u'isDataAvailable': True}
>>> stream.stop()
>>> def console_output(df, freq):
...     return df.writeStream \
...         .format("console") \
...         .trigger(processingTime='%s seconds' % freq) \
...         .options(truncate=False) \
...         .start()
...
>>> out = console_output(raw_rate, 5)
>>> -----
Batch: 0
-----
+-----+-----+
|timestamp|value|
+-----+-----+
+-----+-----+

-----
Batch: 1
-----
+-----+-----+
|timestamp|value|
+-----+-----+
|2020-12-07 15:36:51.202|0|
|2020-12-07 15:36:52.202|1|
|2020-12-07 15:36:53.202|2|
+-----+-----+

out.stop()
>>> filtered_rate = raw_rate \
...     .filter(F.col("value") % F.lit("2") == 0)
>>> out = console_output(filtered_rate, 5)
>>> -----
Batch: 0
-----
+-----+-----+
|timestamp|value|
+-----+-----+
+-----+-----+

-----
Batch: 1
-----
+-----+-----+
|timestamp|value|
+-----+-----+
|2020-12-07 15:37:29.543|0|
|2020-12-07 15:37:31.543|2|
|2020-12-07 15:37:33.543|4|
+-----+-----+

-----
Batch: 2
-----
+-----+-----+
|timestamp|value|
+-----+-----+

```

```

+-----+-----+
+-----+-----+

Batch: 1
+-----+-----+
|timestamp|value|
+-----+-----+
|2020-12-07 15:37:29.543|0|
|2020-12-07 15:37:31.543|2|
|2020-12-07 15:37:33.543|4|
+-----+-----+

Batch: 2
+-----+-----+
|timestamp|value|
+-----+-----+
|2020-12-07 15:37:35.543|6|
|2020-12-07 15:37:37.543|8|
+-----+-----+

out.stop()
>>> extra_rate = filtered_rate \
...     .withColumn("my_value",
...     F.when((F.col("value") % F.lit(10) == 0), F.lit("jubilee"))
...     .otherwise(F.lit("not yet")))
>>> out = console_output(extra_rate, 5)
>>>

Batch: 0
+-----+-----+-----+
|timestamp|value|my_value|
+-----+-----+-----+

Batch: 1
+-----+-----+-----+
|timestamp|value|my_value|
+-----+-----+-----+
|2020-12-07 15:38:16.968|0|jubilee|
|2020-12-07 15:38:18.968|2|not yet|
+-----+-----+-----+

pstroganov/
Batch: 2
+-----+-----+-----+
|timestamp|value|my_value|
+-----+-----+-----+
|2020-12-07 15:38:20.968|4|not yet|
|2020-12-07 15:38:22.968|6|not yet|
+-----+-----+-----+

out.stop()
>>>

```