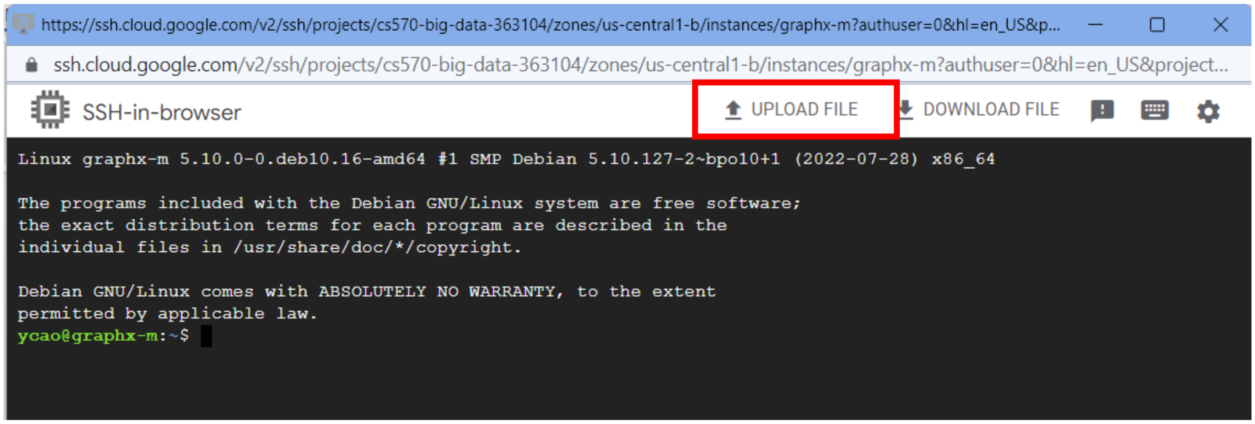
PySpark: DataFrames / SparkSQL + GraphFrames / GraphX

Step 1: Create Cluster on GCP

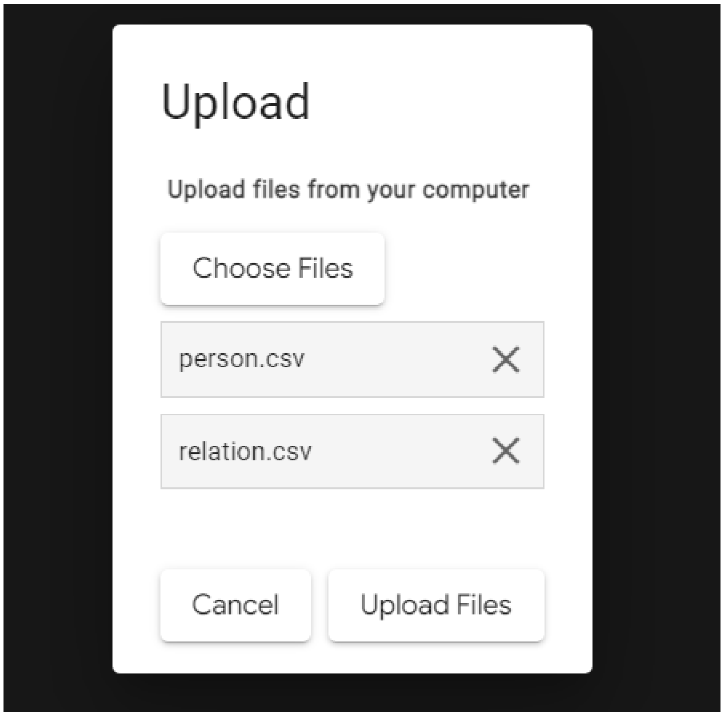
- ➔ Refer to previous HW to create
- ➔ Open a terminal through SSH


https://ssh.cloud.google.com/v2/ssh/projects/cs570-big-data-363104/zones/us-central1-b/instances/graphx-m?authuser=0&hl=en_US&p...
ssh.cloud.google.com/v2/ssh/projects/cs570-big-data-363104/zones/us-central1-b/instances/graphx-m?authuser=0&hl=en_US&project...
SSH-in-browser
UPLOAD FILE
DOWNLOAD FILE

```
Linux graphx-m 5.10.0-0.deb10.16-amd64 #1 SMP Debian 5.10.127-2~bpo10+1 (2022-07-28) x86_64

The programs included with the Debian GNU/Linux system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
ycao@graphx-m:~$
```

Step 2: Data Prepare

- ➔ Upload the csv data files from local to cluster



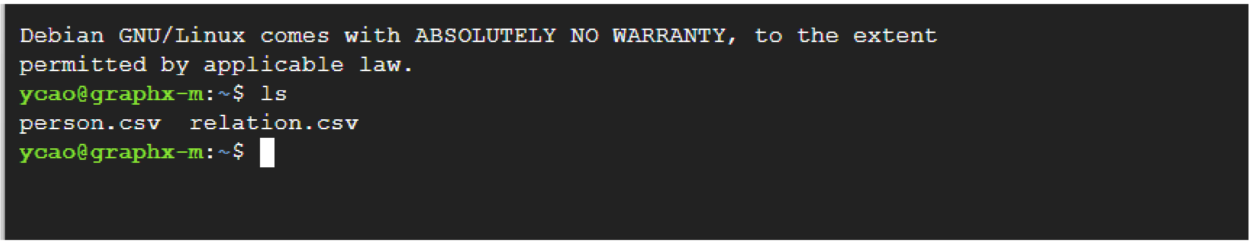
Upload

Upload files from your computer

person.csv

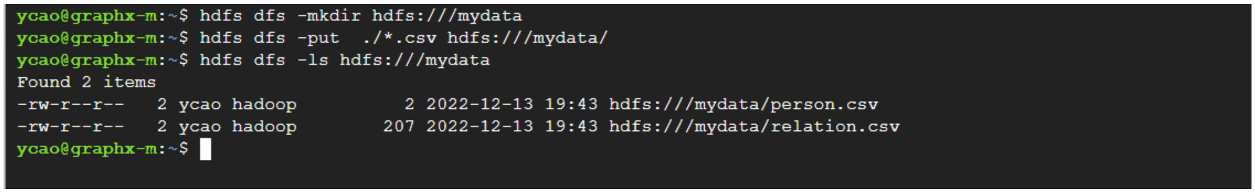
relation.csv

➔ Check Upload



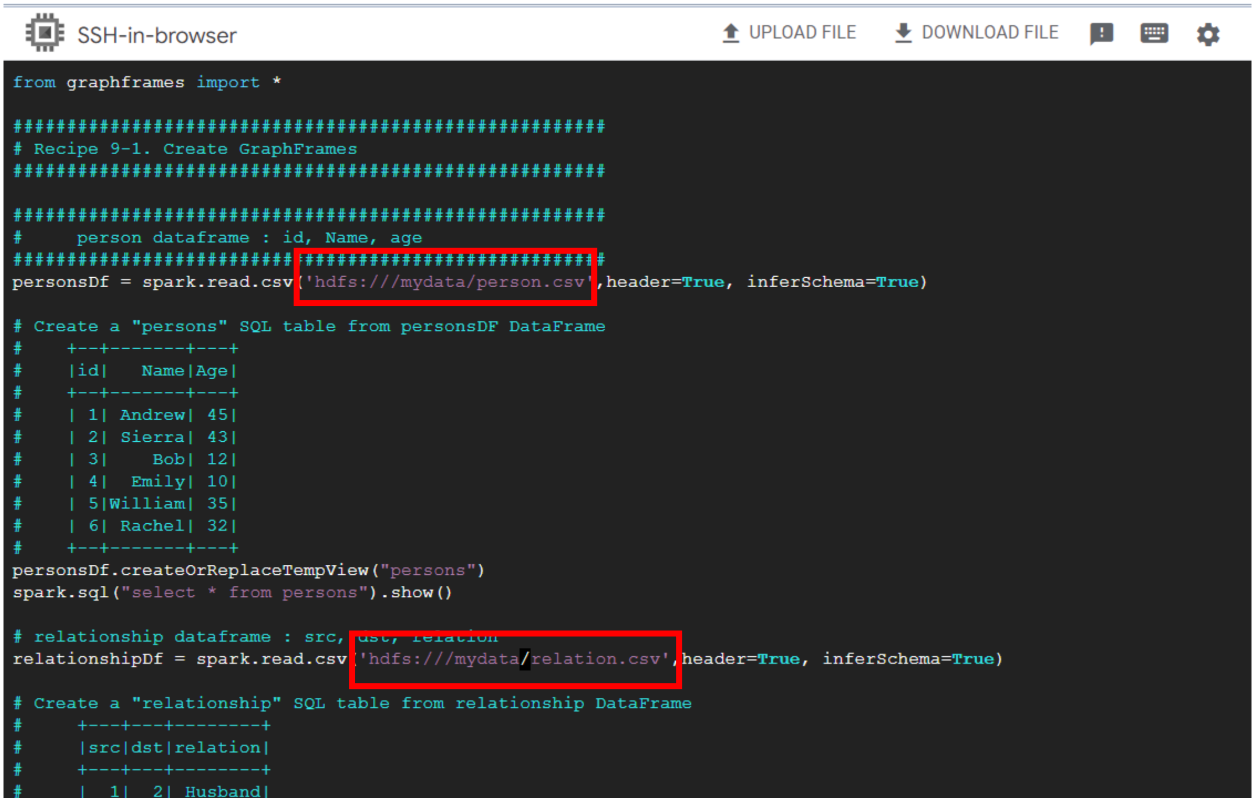
```
Debian GNU/Linux comes with ABSOLUTELY NO WARRANTY, to the extent
permitted by applicable law.
ycao@graphx-m:~$ ls
person.csv  relation.csv
ycao@graphx-m:~$ █
```

➔ Create HDFS file system and copy the data files to HDFS



```
ycao@graphx-m:~$ hdfs dfs -mkdir hdfs:///mydata
ycao@graphx-m:~$ hdfs dfs -put  ./*.csv hdfs:///mydata/
ycao@graphx-m:~$ hdfs dfs -ls hdfs:///mydata
Found 2 items
-rw-r--r--   2 ycao hadoop          2 2022-12-13 19:43 hdfs:///mydata/person.csv
-rw-r--r--   2 ycao hadoop        207 2022-12-13 19:43 hdfs:///mydata/relation.csv
ycao@graphx-m:~$ █
```

➔ Create the graphdemo.py file and change the path for data files



```
from graphframes import *

########################################################
# Recipe 9-1. Create GraphFrames
########################################################

########################################################
#     person dataframe : id, Name, age
########################################################
personsDf = spark.read.csv('hdfs:///mydata/person.csv',header=True, inferSchema=True)

# Create a "persons" SQL table from personsDF DataFrame
#     +--+-------+---+
#     |id|   Name|Age|
#     +--+-------+---+
#     | 1| Andrew| 45|
#     | 2| Sierra| 43|
#     | 3|    Bob| 12|
#     | 4|  Emily| 10|
#     | 5|William| 35|
#     | 6| Rachel| 32|
#     +--+-------+---+
personsDf.createOrReplaceTempView("persons")
spark.sql("select * from persons").show()

# relationship dataframe : src, dst, relation
relationshipDf = spark.read.csv('hdfs:///mydata/relation.csv',header=True, inferSchema=True)

# Create a "relationship" SQL table from relationship DataFrame
#      +---+---+--------+
#      |src|dst|relation|
#      +---+---+--------+
#      |  1|  2| Husband|
```

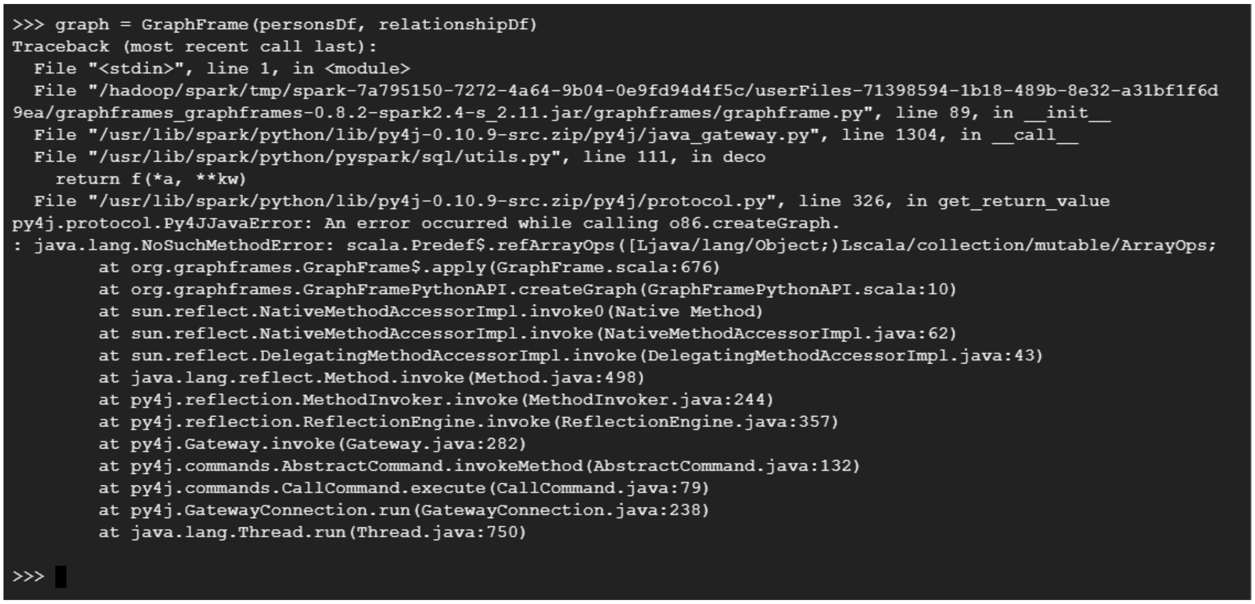
Step 3: Run the code in pyspark shell line by line

```
$ pyspark
```

- ➔ GraphFrame module not found, solve the problem by adding packages

```
$ pyspark --packages graphframes:graphframes:0.8.2-spark2.4-s_2.11
```

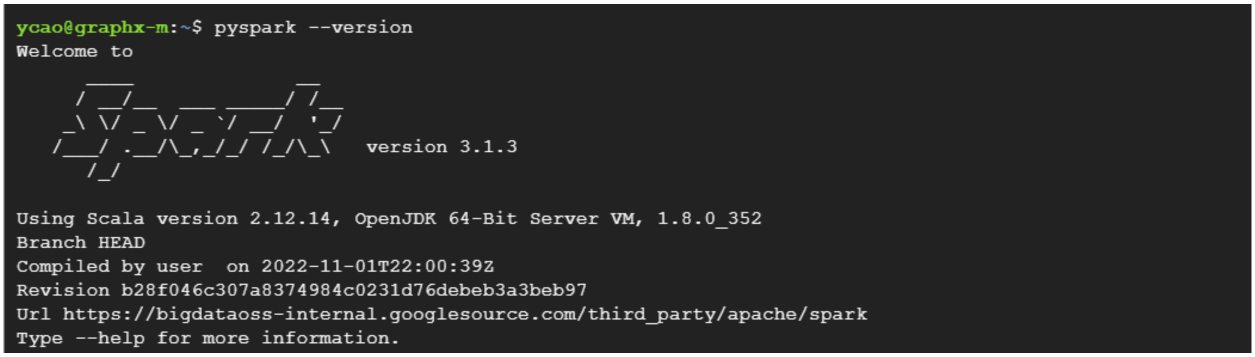
- ➔ New error



```
>>> graph = GraphFrame(personsDf, relationshipDf)
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/hadoop/spark/tmp/spark-7a795150-7272-4a64-9b04-0e9fd94d4f5c/userFiles-71398594-1b18-489b-8e32-a31bf1f6d
9ea/graphframes_graphframes-0.8.2-spark2.4-s_2.11.jar/graphframes/graphframe.py", line 89, in __init__
  File "/usr/lib/spark/python/lib/py4j-0.10.9-src.zip/py4j/java_gateway.py", line 1304, in __call__
  File "/usr/lib/spark/python/pyspark/sql/utils.py", line 111, in deco
    return f(*a, **kw)
  File "/usr/lib/spark/python/lib/py4j-0.10.9-src.zip/py4j/protocol.py", line 326, in get_return_value
py4j.protocol.Py4JJavaError: An error occurred while calling o86.createGraph.
: java.lang.NoSuchMethodError: scala.Predef$.refArrayOps([Ljava/lang/Object;)Lscala/collection/mutable/ArrayOps;
        at org.graphframes.GraphFrame$.apply(GraphFrame.scala:676)
        at org.graphframes.GraphFramePythonAPI.createGraph(GraphFramePythonAPI.scala:10)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
        at java.lang.reflect.Method.invoke(Method.java:498)
        at py4j.reflection.MethodInvoker.invoke(MethodInvoker.java:244)
        at py4j.reflection.ReflectionEngine.invoke(ReflectionEngine.java:357)
        at py4j.Gateway.invoke(Gateway.java:282)
        at py4j.commands.AbstractCommand.invokeMethod(AbstractCommand.java:132)
        at py4j.commands.CallCommand.execute(CallCommand.java:79)
        at py4j.GatewayConnection.run(GatewayConnection.java:238)
        at java.lang.Thread.run(Thread.java:750)

>>> 
```

- ➔ Check pyspark version and find correspondence jar package

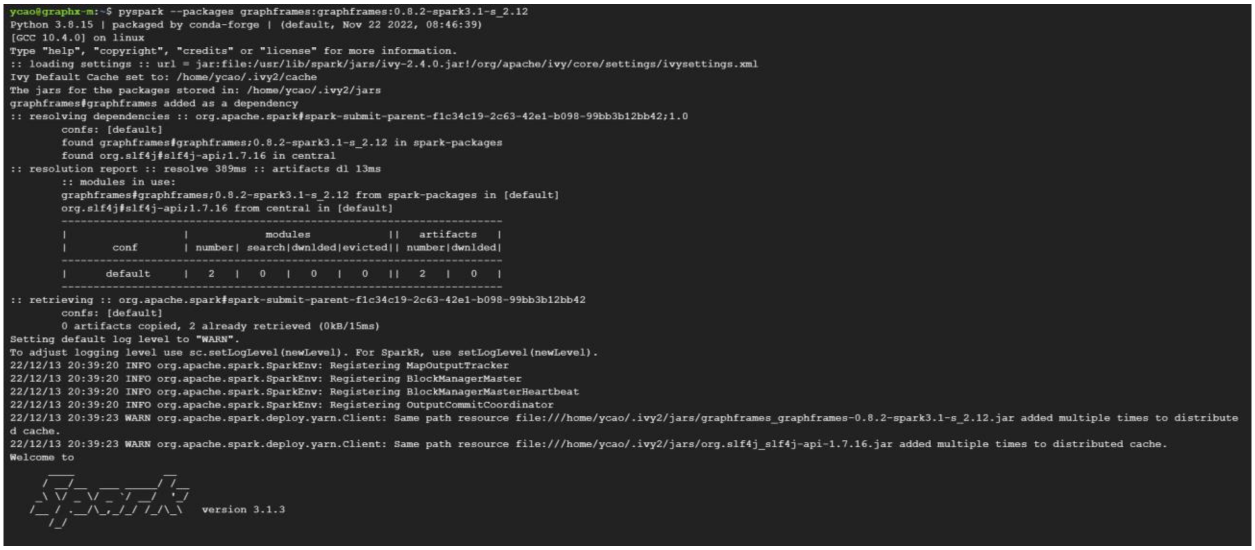
https://spark-packages.org/package/graphframes/graphframes



```
ycao@graphx-m:~$ pyspark --version
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /___/ .__/\_,_/_/ /_/\_\   version 3.1.3
      /_/

Using Scala version 2.12.14, OpenJDK 64-Bit Server VM, 1.8.0_352
Branch HEAD
Compiled by user  on 2022-11-01T22:00:39Z
Revision b28f046c307a8374984c0231d76debeb3a3beb97
Url https://bigdataoss-internal.googlesource.com/third_party/apache/spark
Type --help for more information.
```

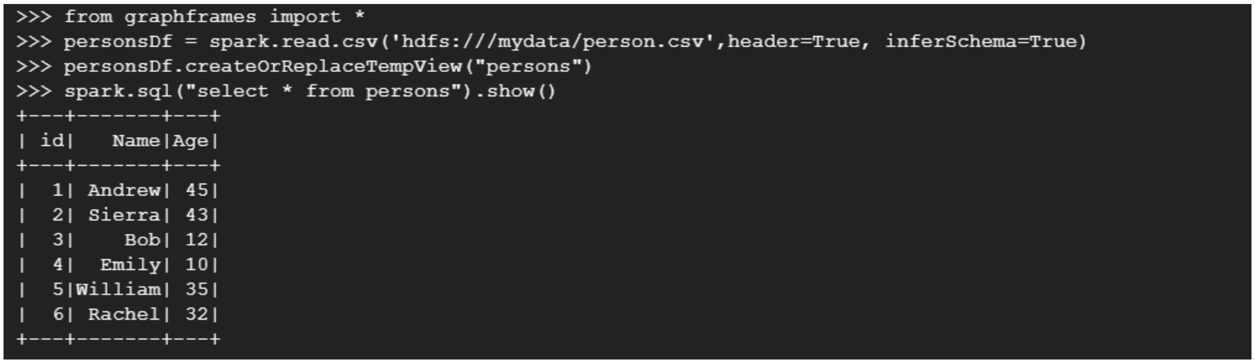
- ➔ Try again

```
$ pyspark --packages graphframes:graphframes:0.8.2-spark3.1-s_2.12
```

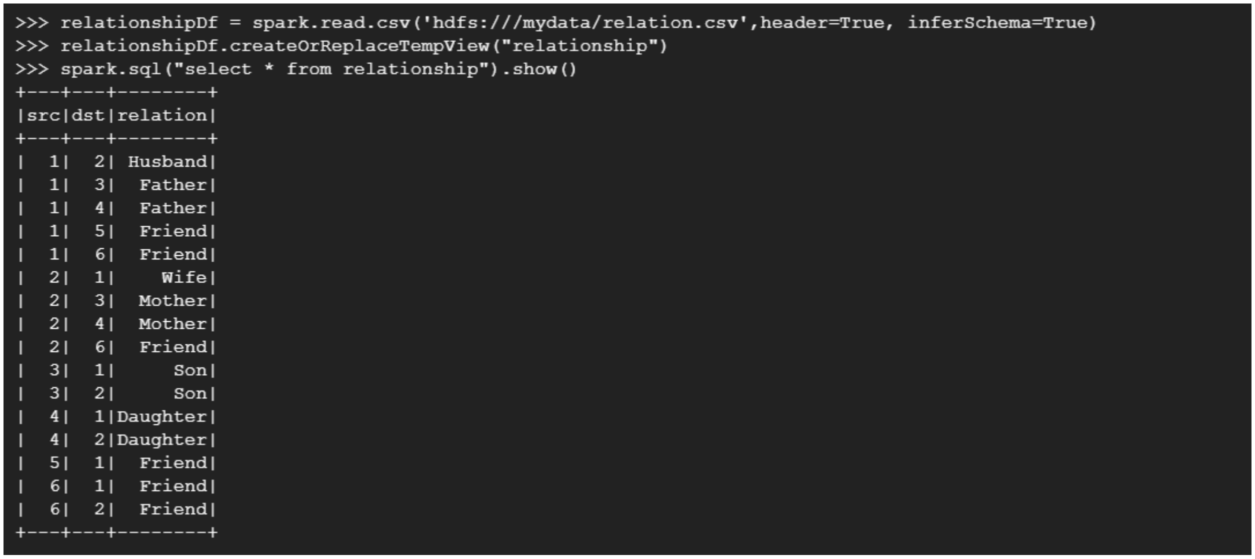


```
ycao@graphx-m:~$ pyspark --packages graphframes:graphframes:0.8.2-spark3.1-s_2.12
Python 3.8.15 | packaged by conda-forge | (default, Nov 22 2022, 08:46:39)
[GCC 10.4.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
:: loading settings :: url = jar:file:/usr/lib/spark/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: /home/ycao/.ivy2/cache
The jars for the packages stored in: /home/ycao/.ivy2/jars
graphframes#graphframes added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-f1c34c19-2c63-42e1-b098-99bb3b12bb42;1.0
        confs: [default]
        found graphframes#graphframes;0.8.2-spark3.1-s_2.12 in spark-packages
        found org.slf4j#slf4j-api;1.7.16 in central
:: resolution report :: resolve 389ms :: artifacts dl 13ms
        :: modules in use:
        graphframes#graphframes;0.8.2-spark3.1-s_2.12 from spark-packages in [default]
        org.slf4j#slf4j-api;1.7.16 from central in [default]
        ---------------------------------------------------------------------
        |                  |            modules            ||   artifacts   |
        |       conf       | number| search|dwnlded|evicted|| number|dwnlded|
        ---------------------------------------------------------------------
        |      default     |   2   |   0   |   0   |   0   ||   2   |   0   |
        ---------------------------------------------------------------------
:: retrieving :: org.apache.spark#spark-submit-parent-f1c34c19-2c63-42e1-b098-99bb3b12bb42
        confs: [default]
        0 artifacts copied, 2 already retrieved (0kB/15ms)
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
22/12/13 20:39:20 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
22/12/13 20:39:20 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
22/12/13 20:39:20 INFO org.apache.spark.SparkEnv: Registering BlockManagerMasterHeartbeat
22/12/13 20:39:20 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
22/12/13 20:39:23 WARN org.apache.spark.deploy.yarn.Client: Same path resource file:///home/ycao/.ivy2/jars/graphframes_graphframes-0.8.2-spark3.1-s_2.12.jar added multiple times to distribute
d cache.
22/12/13 20:39:23 WARN org.apache.spark.deploy.yarn.Client: Same path resource file:///home/ycao/.ivy2/jars/org.slf4j_slf4j-api-1.7.16.jar added multiple times to distributed cache.
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 3.1.3
      /_/
```

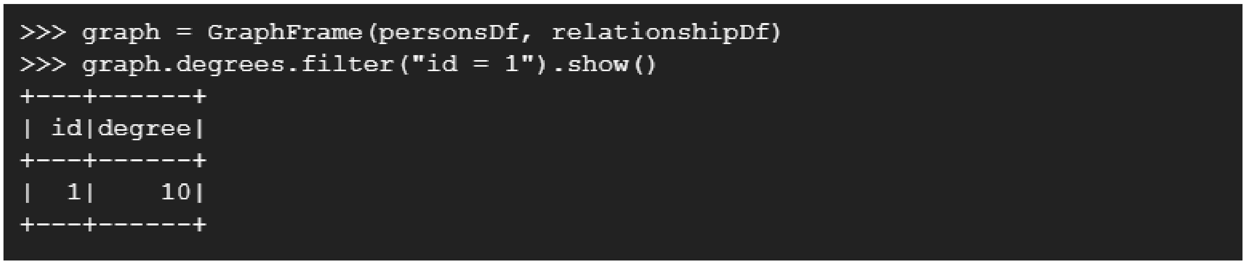
- ➔ Success in pyspark-shell, run remaining code



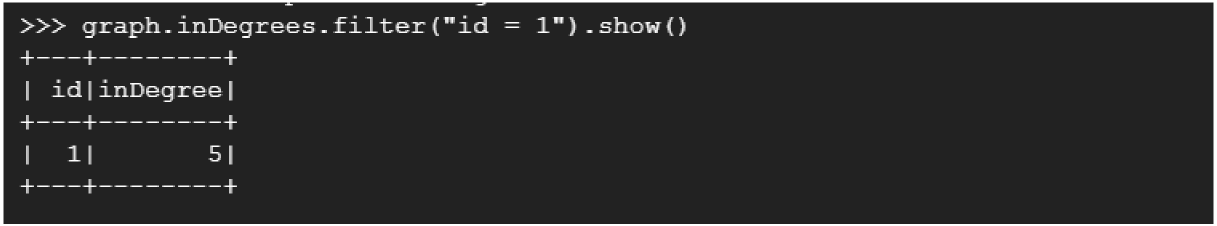
```
>>> from graphframes import *
>>> personsDf = spark.read.csv('hdfs:///mydata/person.csv',header=True, inferSchema=True)
>>> personsDf.createOrReplaceTempView("persons")
>>> spark.sql("select * from persons").show()
+---+-------+---+
| id|   Name|Age|
+---+-------+---+
|  1| Andrew| 45|
|  2| Sierra| 43|
|  3|    Bob| 12|
|  4|  Emily| 10|
|  5|William| 35|
|  6| Rachel| 32|
+---+-------+---+
```



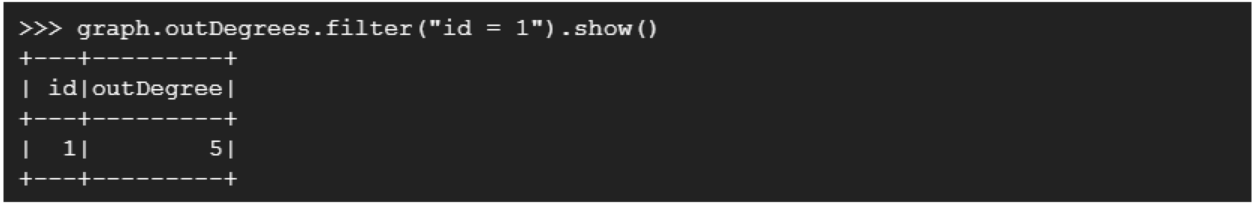
```
>>> relationshipDf = spark.read.csv('hdfs:///mydata/relation.csv',header=True, inferSchema=True)
>>> relationshipDf.createOrReplaceTempView("relationship")
>>> spark.sql("select * from relationship").show()
+---+---+--------+
|src|dst|relation|
+---+---+--------+
|  1|  2| Husband|
|  1|  3|  Father|
|  1|  4|  Father|
|  1|  5|  Friend|
|  1|  6|  Friend|
|  2|  1|    Wife|
|  2|  3|  Mother|
|  2|  4|  Mother|
|  2|  6|  Friend|
|  3|  1|     Son|
|  3|  2|     Son|
|  4|  1|Daughter|
|  4|  2|Daughter|
|  5|  1|  Friend|
|  6|  1|  Friend|
|  6|  2|  Friend|
+---+---+--------+
```



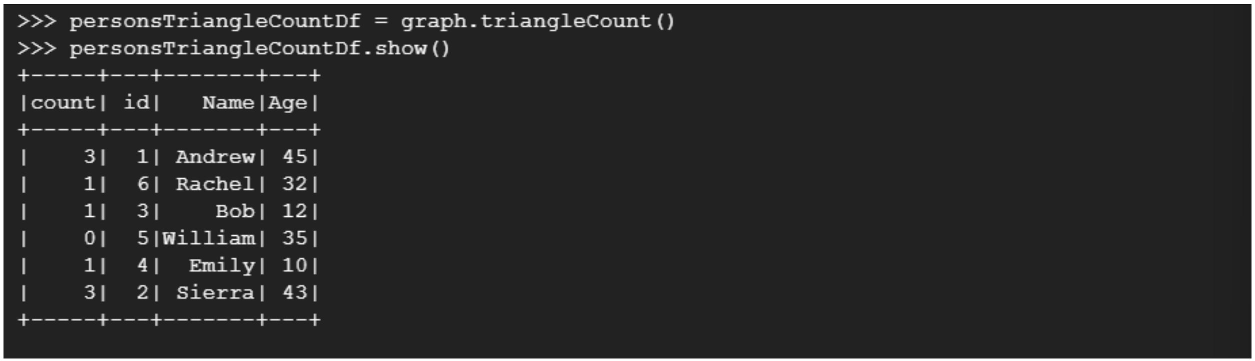
```
>>> graph = GraphFrame(personsDf, relationshipDf)
>>> graph.degrees.filter("id = 1").show()
+---+------+
| id|degree|
+---+------+
|  1|    10|
+---+------+
```



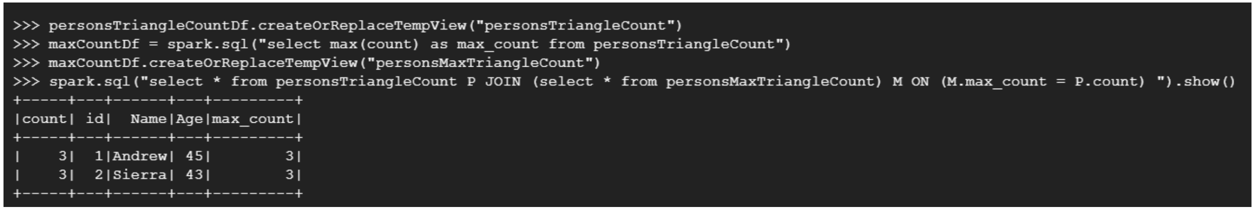
```
>>> graph.inDegrees.filter("id = 1").show()
+---+--------+
| id|inDegree|
+---+--------+
|  1|       5|
+---+--------+
```



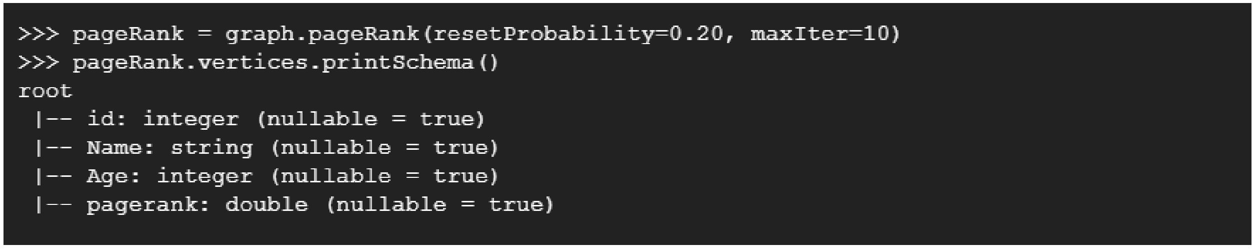
```
>>> graph.outDegrees.filter("id = 1").show()
+---+---------+
| id|outDegree|
+---+---------+
|  1|        5|
+---+---------+
```



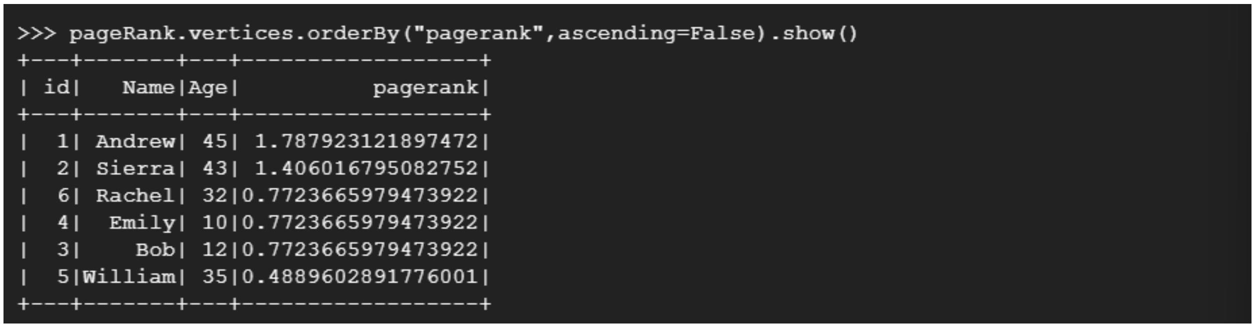
```
>>> personsTriangleCountDf = graph.triangleCount()
>>> personsTriangleCountDf.show()
+-----+---+-------+---+
|count| id|   Name|Age|
+-----+---+-------+---+
|    3|  1| Andrew| 45|
|    1|  6| Rachel| 32|
|    1|  3|    Bob| 12|
|    0|  5|William| 35|
|    1|  4|  Emily| 10|
|    3|  2| Sierra| 43|
+-----+---+-------+---+
```



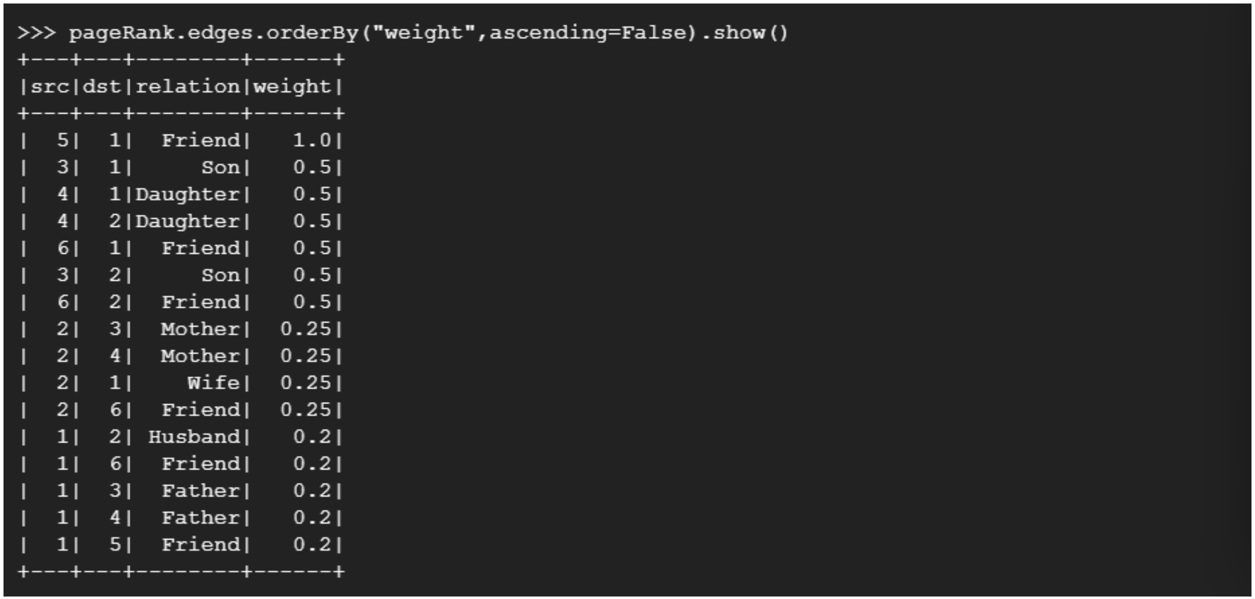
```
>>> personsTriangleCountDf.createOrReplaceTempView("personsTriangleCount")
>>> maxCountDf = spark.sql("select max(count) as max_count from personsTriangleCount")
>>> maxCountDf.createOrReplaceTempView("personsMaxTriangleCount")
>>> spark.sql("select * from personsTriangleCount P JOIN (select * from personsMaxTriangleCount) M ON (M.max_count = P.count) ").show()
+-----+---+------+---+---------+
|count| id|  Name|Age|max_count|
+-----+---+------+---+---------+
|    3|  1|Andrew| 45|        3|
|    3|  2|Sierra| 43|        3|
+-----+---+------+---+---------+
```



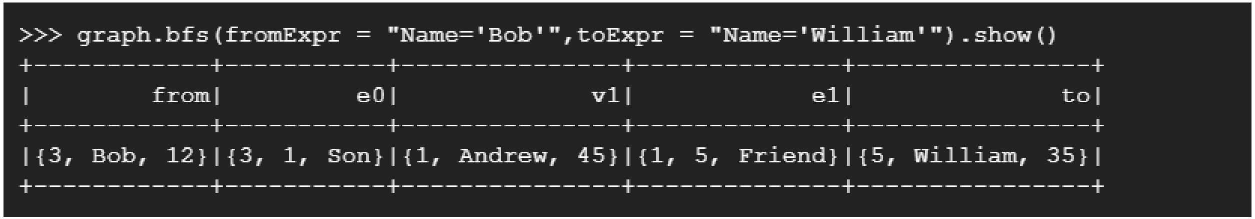
```
>>> pageRank = graph.pageRank(resetProbability=0.20, maxIter=10)
>>> pageRank.vertices.printSchema()
root
 |-- id: integer (nullable = true)
 |-- Name: string (nullable = true)
 |-- Age: integer (nullable = true)
 |-- pagerank: double (nullable = true)
```



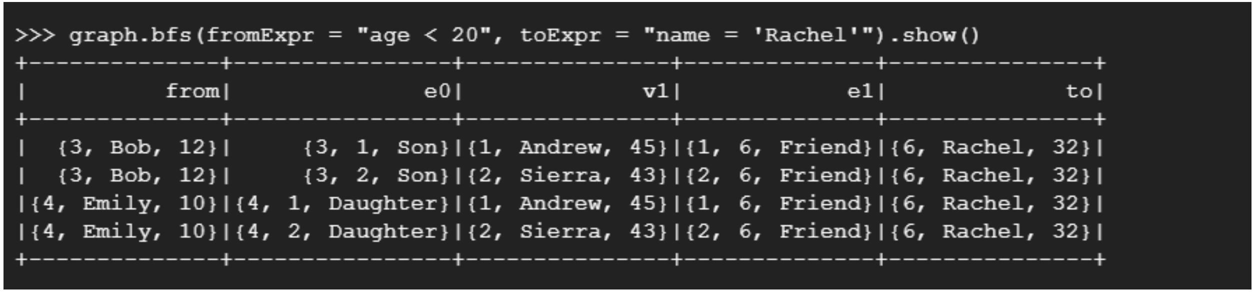
```
>>> pageRank.vertices.orderBy("pagerank",ascending=False).show()
+---+-------+---+------------------+
| id|   Name|Age|          pagerank|
+---+-------+---+------------------+
|  1| Andrew| 45| 1.787923121897472|
|  2| Sierra| 43| 1.406016795082752|
|  6| Rachel| 32|0.7723665979473922|
|  4|  Emily| 10|0.7723665979473922|
|  3|    Bob| 12|0.7723665979473922|
|  5|William| 35|0.4889602891776001|
+---+-------+---+------------------+
```



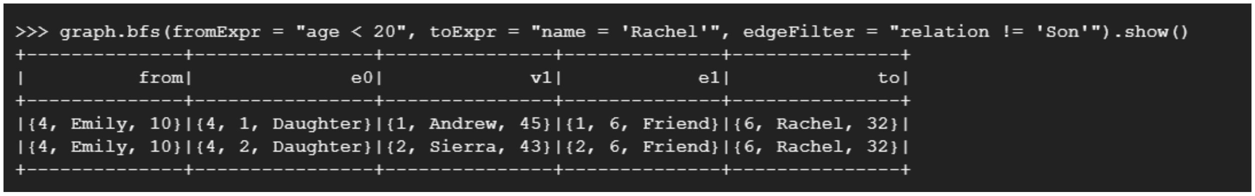
```
>>> pageRank.edges.orderBy("weight",ascending=False).show()
+---+---+--------+------+
|src|dst|relation|weight|
+---+---+--------+------+
|  5|  1|  Friend|   1.0|
|  3|  1|     Son|   0.5|
|  4|  1|Daughter|   0.5|
|  4|  2|Daughter|   0.5|
|  6|  1|  Friend|   0.5|
|  3|  2|     Son|   0.5|
|  6|  2|  Friend|   0.5|
|  2|  3|  Mother|  0.25|
|  2|  4|  Mother|  0.25|
|  2|  1|    Wife|  0.25|
|  2|  6|  Friend|  0.25|
|  1|  2| Husband|   0.2|
|  1|  6|  Friend|   0.2|
|  1|  3|  Father|   0.2|
|  1|  4|  Father|   0.2|
|  1|  5|  Friend|   0.2|
+---+---+--------+------+
```



```
>>> graph.bfs(fromExpr = "Name='Bob'",toExpr = "Name='William'").show()
+------------+-----------+---------------+--------------+----------------+
|        from|         e0|             v1|            e1|              to|
+------------+-----------+---------------+--------------+----------------+
|{3, Bob, 12}|{3, 1, Son}|{1, Andrew, 45}|{1, 5, Friend}|{5, William, 35}|
+------------+-----------+---------------+--------------+----------------+
```



```
>>> graph.bfs(fromExpr = "age < 20", toExpr = "name = 'Rachel'").show()
+--------------+----------------+---------------+--------------+---------------+
|          from|              e0|             v1|            e1|             to|
+--------------+----------------+---------------+--------------+---------------+
|  {3, Bob, 12}|     {3, 1, Son}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|  {3, Bob, 12}|     {3, 2, Son}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
|{4, Emily, 10}|{4, 1, Daughter}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|{4, Emily, 10}|{4, 2, Daughter}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
+--------------+----------------+---------------+--------------+---------------+
```



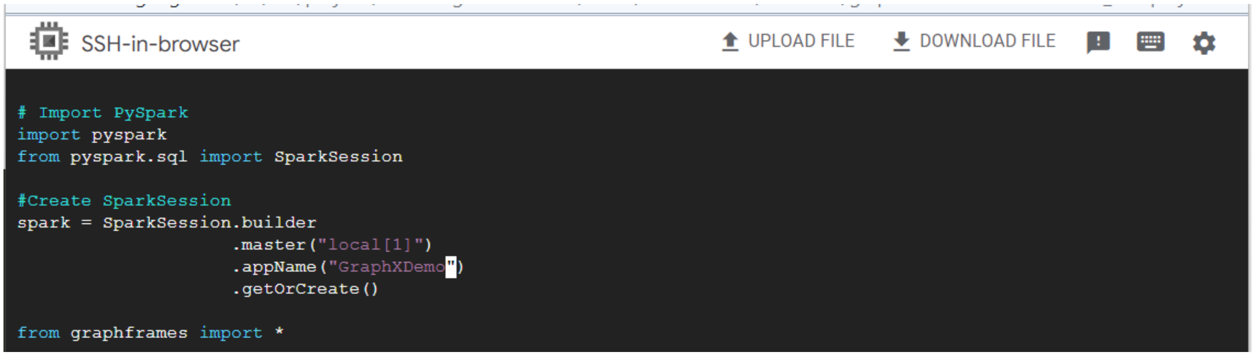
```
>>> graph.bfs(fromExpr = "age < 20", toExpr = "name = 'Rachel'", edgeFilter = "relation != 'Son'").show()
+--------------+----------------+---------------+--------------+---------------+
|          from|              e0|             v1|            e1|             to|
+--------------+----------------+---------------+--------------+---------------+
|{4, Emily, 10}|{4, 1, Daughter}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|{4, Emily, 10}|{4, 2, Daughter}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
+--------------+----------------+---------------+--------------+---------------+
```

Step 4: Run the code with spark-submit

- With experience of fixed errors earlier
- Run code

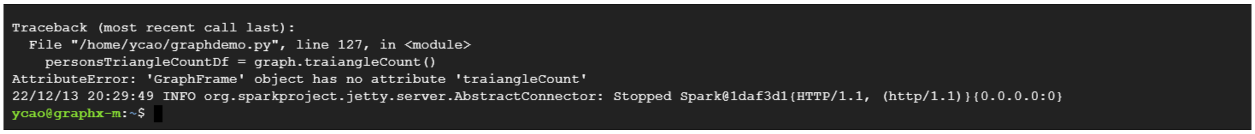
```
$ spark-submit --packages graphframes:graphframes:0.8.2-spark3.1-s_2.12 graphdemo.py
```

- No spark found, add initial code for spark session

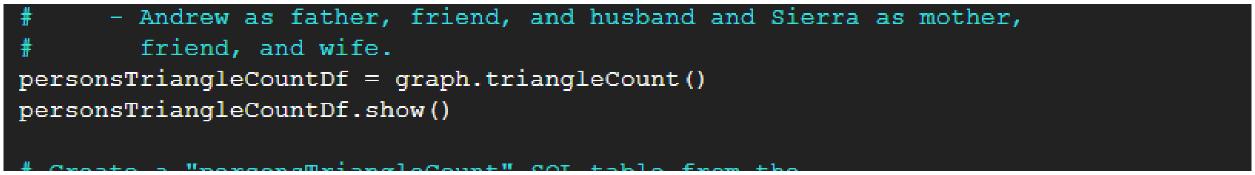


```
# Import PySpark
import pyspark
from pyspark.sql import SparkSession

#Create SparkSession
spark = SparkSession.builder
                    .master("local[1]")
                    .appName("GraphXDemo")
                    .getOrCreate()

from graphframes import *
```

- Run again
- Typo found

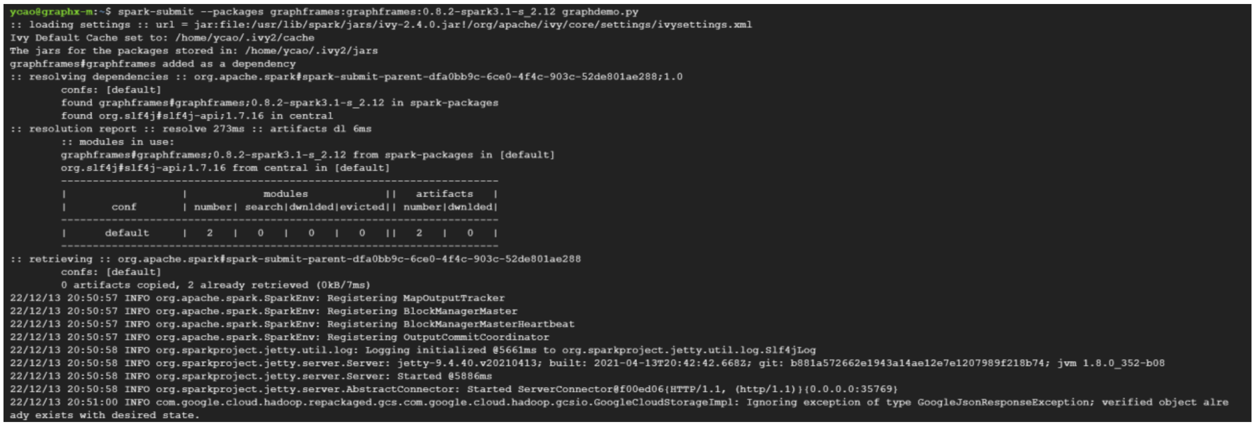


```
Traceback (most recent call last):
  File "/home/ycao/graphdemo.py", line 127, in <module>
    personsTriangleCountDf = graph.traiangleCount()
AttributeError: 'GraphFrame' object has no attribute 'traiangleCount'
22/12/13 20:29:49 INFO org.sparkproject.jetty.server.AbstractConnector: Stopped Spark@1daf3d1{HTTP/1.1, (http/1.1)}{0.0.0.0:0}
ycao@graphx-m:~$
```

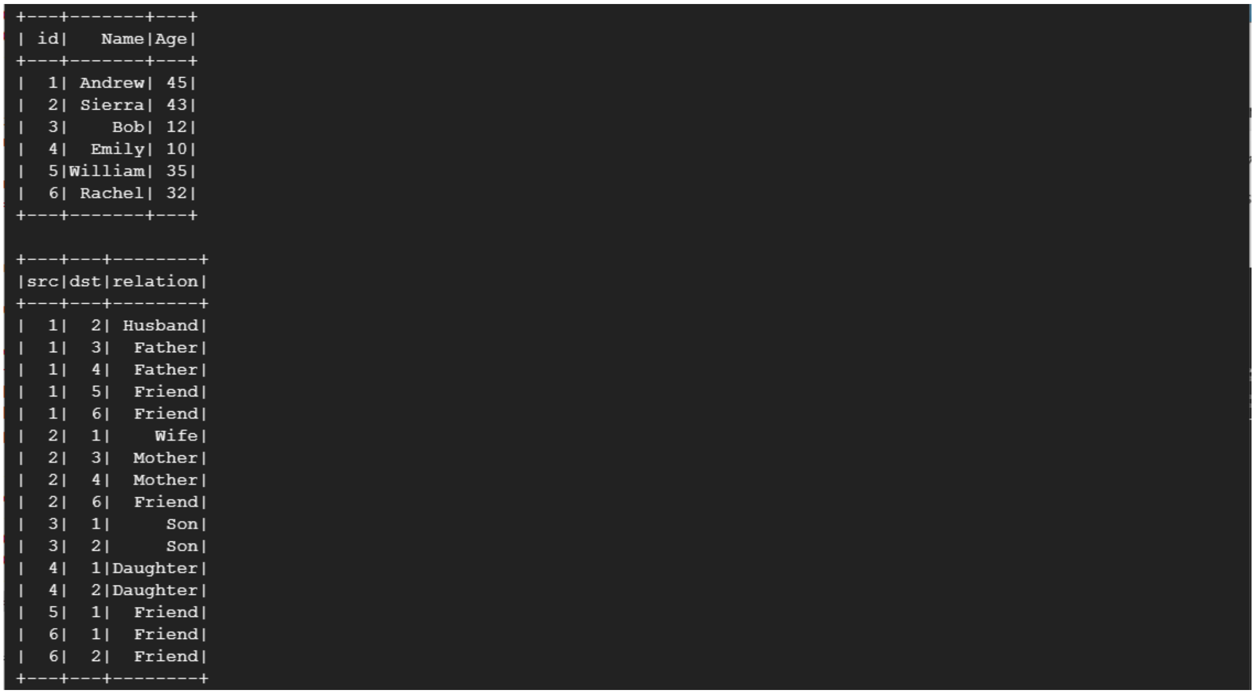


```
#       - Andrew as father, friend, and husband and Sierra as mother,
#         friend, and wife.
personsTriangleCountDf = graph.triangleCount()
personsTriangleCountDf.show()

# Create a "personsTriangleCount" SQL table from the
```

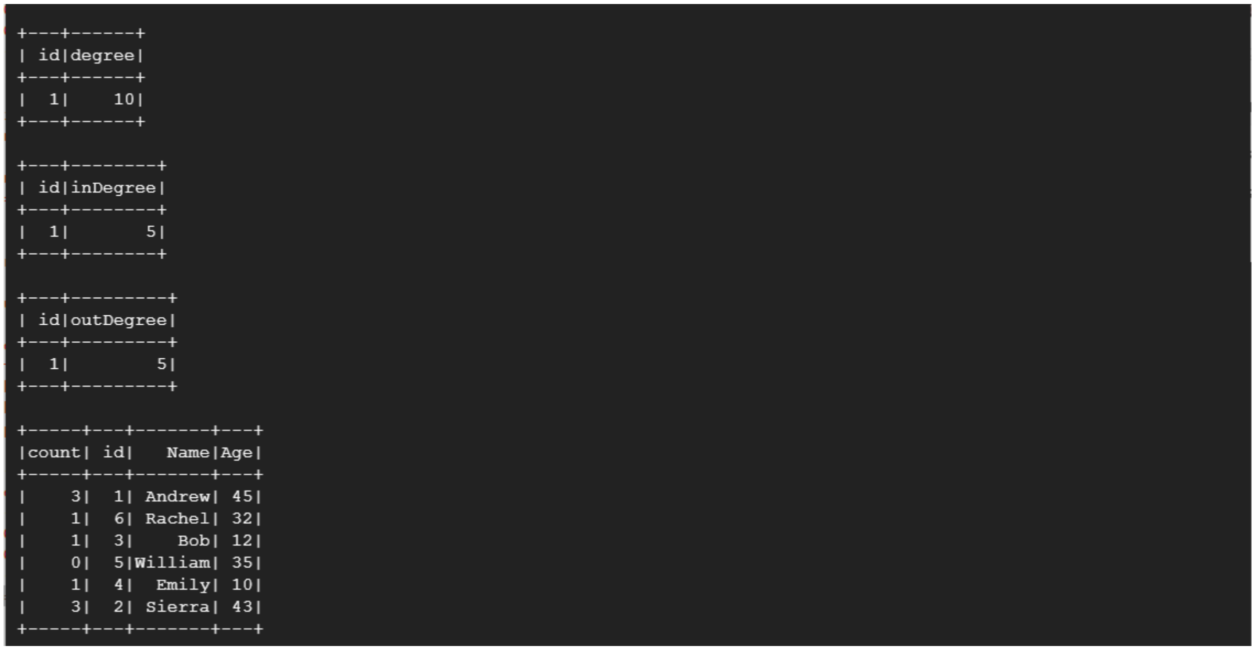
- Fixed and run, successfully execute



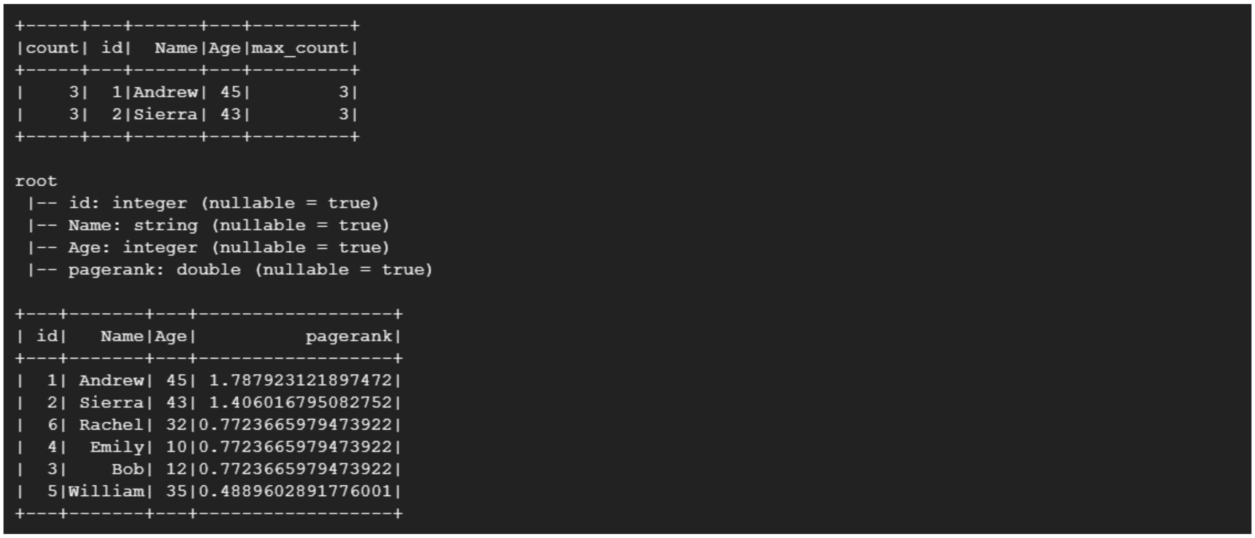
```
ycao@graphx-m:~$ spark-submit --packages graphframes:graphframes:0.8.2-spark3.1-s_2.12 graphdemo.py
:: loading settings :: url = jar:file:/usr/lib/spark/jars/ivy-2.4.0.jar!/org/apache/ivy/core/settings/ivysettings.xml
Ivy Default Cache set to: /home/ycao/.ivy2/cache
The jars for the packages stored in: /home/ycao/.ivy2/jars
graphframes#graphframes added as a dependency
:: resolving dependencies :: org.apache.spark#spark-submit-parent-dfa0bb9c-6ce0-4f4c-903c-52de801ae288;1.0
        confs: [default]
        found graphframes#graphframes;0.8.2-spark3.1-s_2.12 in spark-packages
        found org.slf4j#slf4j-api;1.7.16 in central
:: resolution report :: resolve 273ms :: artifacts dl 6ms
        :: modules in use:
        graphframes#graphframes;0.8.2-spark3.1-s_2.12 from spark-packages in [default]
        org.slf4j#slf4j-api;1.7.16 from central in [default]
        ---------------------------------------------------------------------
        |                  |            modules            ||   artifacts   |
        |       conf       | number| search|dwnlded|evicted|| number|dwnlded|
        ---------------------------------------------------------------------
        |      default     |   2   |   0   |   0   |   0   ||   2   |   0   |
        ---------------------------------------------------------------------
:: retrieving :: org.apache.spark#spark-submit-parent-dfa0bb9c-6ce0-4f4c-903c-52de801ae288
        confs: [default]
        0 artifacts copied, 2 already retrieved (0kB/7ms)
22/12/13 20:50:57 INFO org.apache.spark.SparkEnv: Registering MapOutputTracker
22/12/13 20:50:57 INFO org.apache.spark.SparkEnv: Registering BlockManagerMaster
22/12/13 20:50:57 INFO org.apache.spark.SparkEnv: Registering BlockManagerMasterHeartbeat
22/12/13 20:50:57 INFO org.apache.spark.SparkEnv: Registering OutputCommitCoordinator
22/12/13 20:50:58 INFO org.sparkproject.jetty.util.log: Logging initialized @5661ms to org.sparkproject.jetty.util.log.Slf4jLog
22/12/13 20:50:58 INFO org.sparkproject.jetty.server.Server: jetty-9.4.40.v20210413; built: 2021-04-13T20:42:42.668Z; git: b881a572662e1943a14ae12e7e1207989f218b74; jvm 1.8.0_352-b08
22/12/13 20:50:58 INFO org.sparkproject.jetty.server.Server: Started @5886ms
22/12/13 20:50:58 INFO org.sparkproject.jetty.server.AbstractConnector: Started ServerConnector@f00ed06{HTTP/1.1, (http/1.1)}{0.0.0.0:35769}
22/12/13 20:51:00 INFO com.google.cloud.hadoop.repackaged.gcs.com.google.cloud.hadoop.gcsio.GoogleCloudStorageImpl: Ignoring exception of type GoogleJsonResponseException; verified object alre
ady exists with desired state.
```



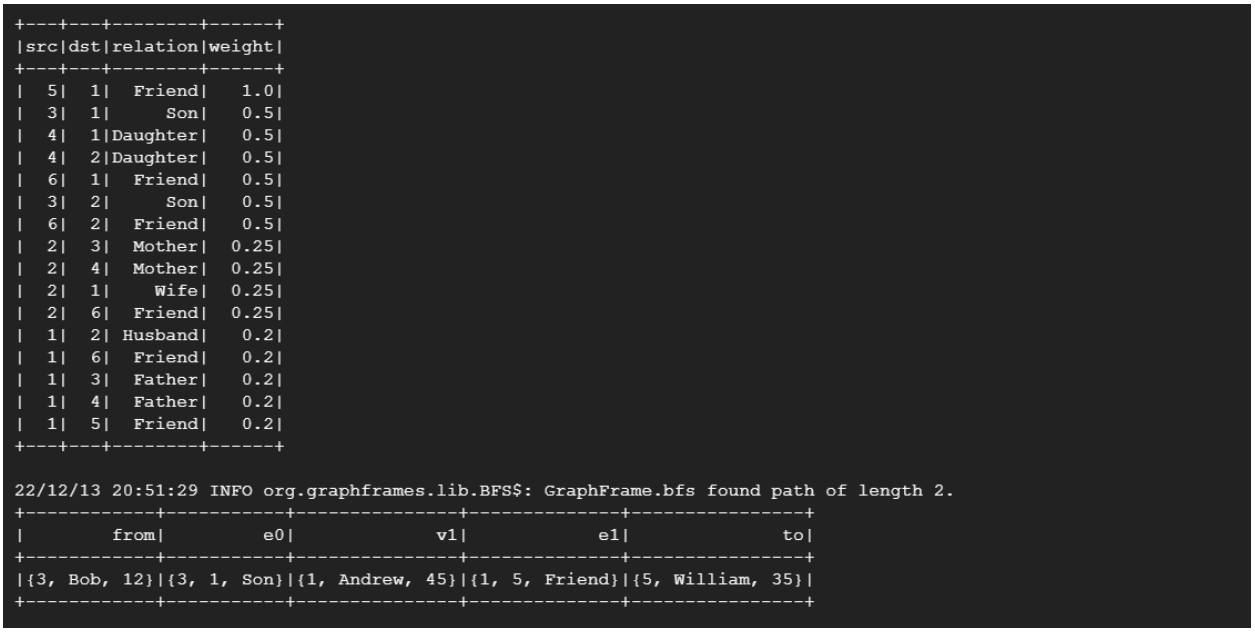
```
+---+-------+---+
| id|   Name|Age|
+---+-------+---+
|  1| Andrew| 45|
|  2| Sierra| 43|
|  3|    Bob| 12|
|  4|  Emily| 10|
|  5|William| 35|
|  6| Rachel| 32|
+---+-------+---+

+---+---+--------+
|src|dst|relation|
+---+---+--------+
|  1|  2| Husband|
|  1|  3|  Father|
|  1|  4|  Father|
|  1|  5|  Friend|
|  1|  6|  Friend|
|  2|  1|    Wife|
|  2|  3|  Mother|
|  2|  4|  Mother|
|  2|  6|  Friend|
|  3|  1|     Son|
|  3|  2|     Son|
|  4|  1|Daughter|
|  4|  2|Daughter|
|  5|  1|  Friend|
|  6|  1|  Friend|
|  6|  2|  Friend|
+---+---+--------+
```



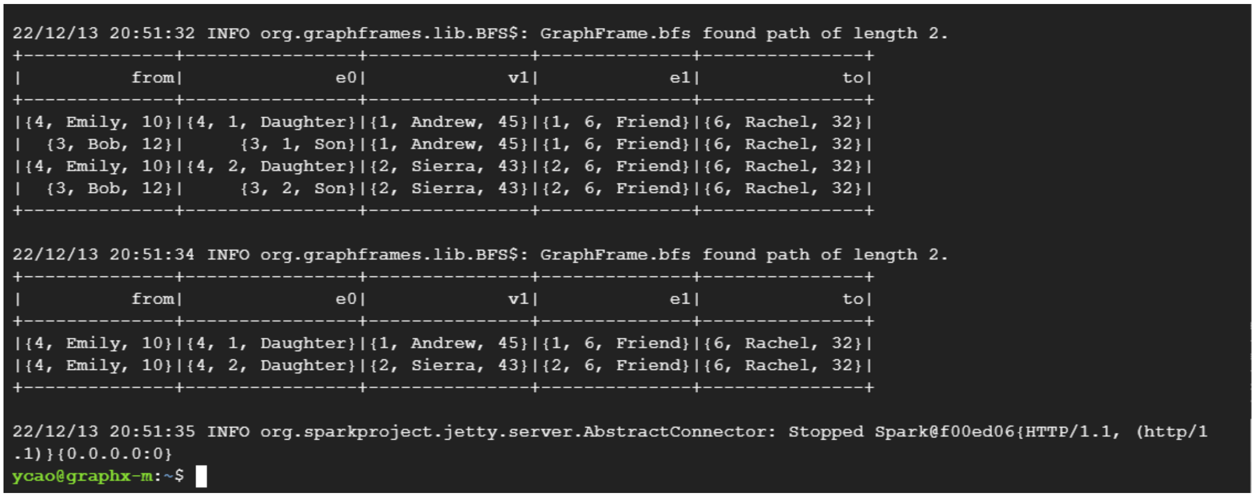
```
+---+------+
| id|degree|
+---+------+
|  1|    10|
+---+------+

+---+--------+
| id|inDegree|
+---+--------+
|  1|       5|
+---+--------+

+---+---------+
| id|outDegree|
+---+---------+
|  1|        5|
+---+---------+

+-----+---+-------+---+
|count| id|   Name|Age|
+-----+---+-------+---+
|    3|  1| Andrew| 45|
|    1|  6| Rachel| 32|
|    1|  3|    Bob| 12|
|    0|  5|William| 35|
|    1|  4|  Emily| 10|
|    3|  2| Sierra| 43|
+-----+---+-------+---+
```



```
+-----+---+------+---+---------+
|count| id|  Name|Age|max_count|
+-----+---+------+---+---------+
|    3|  1|Andrew| 45|        3|
|    3|  2|Sierra| 43|        3|
+-----+---+------+---+---------+

root
 |-- id: integer (nullable = true)
 |-- Name: string (nullable = true)
 |-- Age: integer (nullable = true)
 |-- pagerank: double (nullable = true)

+---+-------+---+------------------+
| id|   Name|Age|          pagerank|
+---+-------+---+------------------+
|  1| Andrew| 45| 1.787923121897472|
|  2| Sierra| 43| 1.406016795082752|
|  6| Rachel| 32|0.7723665979473922|
|  4|  Emily| 10|0.7723665979473922|
|  3|    Bob| 12|0.7723665979473922|
|  5|William| 35|0.4889602891776001|
+---+-------+---+------------------+
```



```
+---+---+--------+------+
|src|dst|relation|weight|
+---+---+--------+------+
|  5|  1|  Friend|   1.0|
|  3|  1|     Son|   0.5|
|  4|  1|Daughter|   0.5|
|  4|  2|Daughter|   0.5|
|  6|  1|  Friend|   0.5|
|  3|  2|     Son|   0.5|
|  6|  2|  Friend|   0.5|
|  2|  3|  Mother|  0.25|
|  2|  4|  Mother|  0.25|
|  2|  1|    Wife|  0.25|
|  2|  6|  Friend|  0.25|
|  1|  2| Husband|   0.2|
|  1|  6|  Friend|   0.2|
|  1|  3|  Father|   0.2|
|  1|  4|  Father|   0.2|
|  1|  5|  Friend|   0.2|
+---+---+--------+------+

22/12/13 20:51:29 INFO org.graphframes.lib.BFS$: GraphFrame.bfs found path of length 2.
+------------+-----------+---------------+--------------+----------------+
|        from|         e0|             v1|            e1|              to|
+------------+-----------+---------------+--------------+----------------+
|{3, Bob, 12}|{3, 1, Son}|{1, Andrew, 45}|{1, 5, Friend}|{5, William, 35}|
+------------+-----------+---------------+--------------+----------------+
```



```
22/12/13 20:51:32 INFO org.graphframes.lib.BFS$: GraphFrame.bfs found path of length 2.
+--------------+----------------+---------------+--------------+---------------+
|          from|              e0|             v1|            e1|             to|
+--------------+----------------+---------------+--------------+---------------+
|{4, Emily, 10}|{4, 1, Daughter}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|  {3, Bob, 12}|     {3, 1, Son}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|{4, Emily, 10}|{4, 2, Daughter}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
|  {3, Bob, 12}|     {3, 2, Son}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
+--------------+----------------+---------------+--------------+---------------+

22/12/13 20:51:34 INFO org.graphframes.lib.BFS$: GraphFrame.bfs found path of length 2.
+--------------+----------------+---------------+--------------+---------------+
|          from|              e0|             v1|            e1|             to|
+--------------+----------------+---------------+--------------+---------------+
|{4, Emily, 10}|{4, 1, Daughter}|{1, Andrew, 45}|{1, 6, Friend}|{6, Rachel, 32}|
|{4, Emily, 10}|{4, 2, Daughter}|{2, Sierra, 43}|{2, 6, Friend}|{6, Rachel, 32}|
+--------------+----------------+---------------+--------------+---------------+

22/12/13 20:51:35 INFO org.sparkproject.jetty.server.AbstractConnector: Stopped Spark@f00ed06{HTTP/1.1, (http/1
.1)}{0.0.0.0:0}
ycao@graphx-m:~$
```

Done!