ex2

June 13, 2021

```
[2]: import sys
     from sys import argv
     from random import random
     from operator import add

     from pyspark.sql import SparkSession
```

```
[3]: if __name__ == "__main__":
         """
             Usage: pi [partitions]
         """
         spark = SparkSession\
             .builder\
             .appName("PythonPi")\
             .getOrCreate()

         partitions = abs(int(sys.argv[1], 16)) if len(sys.argv) > 1 else 2
     #     partitions = 2
         n = 100000 * partitions

         def f():
             x = random() * 2 - 1
             y = random() * 2 - 1
     #         return 1 if x ** 2 + y ** 2 <= 1 else 0
             if (x ** 2 + y ** 2) <= 1:
                 return 1
             else:
                 return 0

         count = spark.sparkContext.parallelize(range(1, n + 1), partitions).map(f).
     ↪reduce(add)
         print("Pi is roughly %f" % (4.0 * count / n))

         spark.stop()
```

```
    ---------------------------------------------------------------------------
    Py4JJavaError                             Traceback (most recent call last)
```

```
<ipython-input-3-51f4d32ae6dc> in <module>
     21             return 0
     22
---> 23     count = spark.sparkContext.parallelize(range(1, n + 1), partitions)
 ↪map(f).reduce(add)
     24     print("Pi is roughly %f" % (4.0 * count / n))
     25


/opt/conda/lib/python3.8/site-packages/pyspark/rdd.py in reduce(self, f)
    995             yield reduce(f, iterator, initial)
    996
--> 997         vals = self.mapPartitions(func).collect()
    998         if vals:
    999             return reduce(f, vals)


/opt/conda/lib/python3.8/site-packages/pyspark/rdd.py in collect(self)
    947         """
    948         with SCCallSiteSync(self.context) as css:
--> 949             sock_info = self.ctx._jvm.PythonRDD.collectAndServe(self.
 ↪_jrdd.rdd())
    950         return list(_load_from_socket(sock_info, self.
 ↪_jrdd_deserializer))
    951


/opt/conda/lib/python3.8/site-packages/py4j/java_gateway.py in __call__(self,
 ↪*args)
   1302
   1303         answer = self.gateway_client.send_command(command)
-> 1304         return_value = get_return_value(
   1305             answer, self.gateway_client, self.target_id, self.name)
   1306


/opt/conda/lib/python3.8/site-packages/pyspark/sql/utils.py in deco(*a, **kw)
    109     def deco(*a, **kw):
    110         try:
--> 111             return f(*a, **kw)
    112         except py4j.protocol.Py4JJavaError as e:
    113             converted = convert_exception(e.java_exception)


/opt/conda/lib/python3.8/site-packages/py4j/protocol.py in
 ↪get_return_value(answer, gateway_client, target_id, name)
    324             value = OUTPUT_CONVERTER[type](answer[2:], gateway_client)
    325             if answer[1] == REFERENCE_TYPE:
--> 326                 raise Py4JJavaError(
    327                     "An error occurred while calling {0}{1}{2}.\n".
    328                     format(target_id, ".", name), value)
```

```
Py4JJavaError: An error occurred while calling z:org.apache.spark.api.python.
↪PythonRDD.collectAndServe.
: org.apache.spark.SparkException: Job aborted due to stage failure: Task 0 in␣
↪stage 0.0 failed 1 times, most recent failure: Lost task 0.0 in stage 0.0 (TID␣
↪0) (jupyter-ntiana55 executor driver): org.apache.spark.api.python.
↪PythonException: Traceback (most recent call last):
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/worker.py", line 604, in␣
↪main
    process()
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/worker.py", line 596, in␣
↪process
    serializer.dump_stream(out_iter, outfile)
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/serializers.py", line␣
↪259, in dump_stream
    vs = list(itertools.islice(iterator, batch))
 File "/opt/conda/lib/python3.8/site-packages/pyspark/rdd.py", line 992, in func
    initial = next(iterator)
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/util.py", line 73, in␣
↪wrapper
    return f(*args, **kwargs)
TypeError: f() takes 0 positional arguments but 1 was given


        at org.apache.spark.api.python.BasePythonRunner$ReaderIterator.
↪handlePythonException(PythonRunner.scala:517)
        at org.apache.spark.api.python.PythonRunner$$anon$3.read(PythonRunner.
↪scala:652)
        at org.apache.spark.api.python.PythonRunner$$anon$3.read(PythonRunner.
↪scala:635)
        at org.apache.spark.api.python.BasePythonRunner$ReaderIterator.
↪hasNext(PythonRunner.scala:470)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterator.
↪scala:37)
        at scala.collection.Iterator.foreach(Iterator.scala:941)
        at scala.collection.Iterator.foreach$(Iterator.scala:941)
        at org.apache.spark.InterruptibleIterator.foreach(InterruptibleIterator.
↪scala:28)
        at scala.collection.generic.Growable.$plus$plus$eq(Growable.scala:62)
        at scala.collection.generic.Growable.$plus$plus$eq$(Growable.scala:53)
        at scala.collection.mutable.ArrayBuffer.$plus$plus$eq(ArrayBuffer.scala:
↪105)
        at scala.collection.mutable.ArrayBuffer.$plus$plus$eq(ArrayBuffer.scala:
↪49)
        at scala.collection.TraversableOnce.to(TraversableOnce.scala:315)
        at scala.collection.TraversableOnce.to$(TraversableOnce.scala:313)
        at org.apache.spark.InterruptibleIterator.to(InterruptibleIterator.scala:
↪28)
        at scala.collection.TraversableOnce.toBuffer(TraversableOnce.scala:307)
        at scala.collection.TraversableOnce.toBuffer$(TraversableOnce.scala:307
```

```
        at org.apache.spark.InterruptibleIterator.toBuffer(InterruptibleIterator.
↪scala:28)
        at scala.collection.TraversableOnce.toArray(TraversableOnce.scala:294)
        at scala.collection.TraversableOnce.toArray$(TraversableOnce.scala:288)
        at org.apache.spark.InterruptibleIterator.toArray(InterruptibleIterator.
↪scala:28)
        at org.apache.spark.rdd.RDD.$anonfun$collect$2(RDD.scala:1030)
        at org.apache.spark.SparkContext.$anonfun$runJob$5(SparkContext.scala:
↪2242)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Executor.
↪scala:497)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:500)
        at java.base/java.util.concurrent.ThreadPoolExecutor.
↪runWorker(ThreadPoolExecutor.java:1128)
        at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.
↪run(ThreadPoolExecutor.java:628)
        at java.base/java.lang.Thread.run(Thread.java:834)

Driver stacktrace:
        at org.apache.spark.scheduler.DAGScheduler.
↪failJobAndIndependentStages(DAGScheduler.scala:2253)
        at org.apache.spark.scheduler.DAGScheduler.
↪$anonfun$abortStage$2(DAGScheduler.scala:2202)
        at org.apache.spark.scheduler.DAGScheduler.
↪$anonfun$abortStage$2$adapted(DAGScheduler.scala:2201)
        at scala.collection.mutable.ResizableArray.foreach(ResizableArray.scala:
↪62)
        at scala.collection.mutable.ResizableArray.foreach$(ResizableArray.scala:
↪55)
        at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:49)
        at org.apache.spark.scheduler.DAGScheduler.abortStage(DAGScheduler.scala:
↪2201)
        at org.apache.spark.scheduler.DAGScheduler.
↪$anonfun$handleTaskSetFailed$1(DAGScheduler.scala:1078)
        at org.apache.spark.scheduler.DAGScheduler.
↪$anonfun$handleTaskSetFailed$1$adapted(DAGScheduler.scala:1078)
        at scala.Option.foreach(Option.scala:407)
        at org.apache.spark.scheduler.DAGScheduler.
↪handleTaskSetFailed(DAGScheduler.scala:1078)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.
↪doOnReceive(DAGScheduler.scala:2440)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.
↪onReceive(DAGScheduler.scala:2382)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.
↪onReceive(DAGScheduler.scala:2371)
```

```
        at org.apache.spark.util.EventLoop$$anon$1.run(EventLoop.scala:49)
        at org.apache.spark.scheduler.DAGScheduler.runJob(DAGScheduler.scala:86 )
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2202)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2223)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2242)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2267)
        at org.apache.spark.rdd.RDD.$anonfun$collect$1(RDD.scala:1030)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.
↪scala:151)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope.
↪scala:112)
        at org.apache.spark.rdd.RDD.withScope(RDD.scala:414)
        at org.apache.spark.rdd.RDD.collect(RDD.scala:1029)
        at org.apache.spark.api.python.PythonRDD$.collectAndServe(PythonRDD.
↪scala:180)
        at org.apache.spark.api.python.PythonRDD.collectAndServe(PythonRDD.scal )
        at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.
↪invoke0(Native Method)
        at java.base/jdk.internal.reflect.NativeMethodAccessorImpl.
↪invoke(NativeMethodAccessorImpl.java:62)
        at java.base/jdk.internal.reflect.DelegatingMethodAccessorImpl.
↪invoke(DelegatingMethodAccessorImpl.java:43)
        at java.base/java.lang.reflect.Method.invoke(Method.java:566)
        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.base/java.lang.Thread.run(Thread.java:834)
Caused by: org.apache.spark.api.python.PythonException: Traceback (most recent
↪call last):
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/worker.py", line 604, i
↪main
    process()
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/worker.py", line 596, i
↪process
    serializer.dump_stream(out_iter, outfile)
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/serializers.py", line
↪259, in dump_stream
    vs = list(itertools.islice(iterator, batch))
 File "/opt/conda/lib/python3.8/site-packages/pyspark/rdd.py", line 992, in fu c
    initial = next(iterator)
 File "/usr/local/spark/python/lib/pyspark.zip/pyspark/util.py", line 73, in
↪wrapper
    return f(*args, **kwargs)
TypeError: f() takes 0 positional arguments but 1 was given
```

```
        at org.apache.spark.api.python.BasePythonRunner$ReaderIterator.
↪handlePythonException(PythonRunner.scala:517)
        at org.apache.spark.api.python.PythonRunner$$anon$3.read(PythonRunner.
↪scala:652)
        at org.apache.spark.api.python.PythonRunner$$anon$3.read(PythonRunner.
↪scala:635)
        at org.apache.spark.api.python.BasePythonRunner$ReaderIterator.
↪hasNext(PythonRunner.scala:470)
        at org.apache.spark.InterruptibleIterator.hasNext(InterruptibleIterator
↪scala:37)
        at scala.collection.Iterator.foreach(Iterator.scala:941)
        at scala.collection.Iterator.foreach$(Iterator.scala:941)
        at org.apache.spark.InterruptibleIterator.foreach(InterruptibleIterator
↪scala:28)
        at scala.collection.generic.Growable.$plus$plus$eq(Growable.scala:62)
        at scala.collection.generic.Growable.$plus$plus$eq$(Growable.scala:53)
        at scala.collection.mutable.ArrayBuffer.$plus$plus$eq(ArrayBuffer.scala
↪105)
        at scala.collection.mutable.ArrayBuffer.$plus$plus$eq(ArrayBuffer.scala
↪49)
        at scala.collection.TraversableOnce.to(TraversableOnce.scala:315)
        at scala.collection.TraversableOnce.to$(TraversableOnce.scala:313)
        at org.apache.spark.InterruptibleIterator.to(InterruptibleIterator.scala
↪28)
        at scala.collection.TraversableOnce.toBuffer(TraversableOnce.scala:307)
        at scala.collection.TraversableOnce.toBuffer$(TraversableOnce.scala:307
        at org.apache.spark.InterruptibleIterator.toBuffer(InterruptibleIterator.
↪scala:28)
        at scala.collection.TraversableOnce.toArray(TraversableOnce.scala:294)
        at scala.collection.TraversableOnce.toArray$(TraversableOnce.scala:288)
        at org.apache.spark.InterruptibleIterator.toArray(InterruptibleIterator
↪scala:28)
        at org.apache.spark.rdd.RDD.$anonfun$collect$2(RDD.scala:1030)
        at org.apache.spark.SparkContext.$anonfun$runJob$5(SparkContext.scala:
↪2242)
        at org.apache.spark.scheduler.ResultTask.runTask(ResultTask.scala:90)
        at org.apache.spark.scheduler.Task.run(Task.scala:131)
        at org.apache.spark.executor.Executor$TaskRunner.$anonfun$run$3(Executor.
↪scala:497)
        at org.apache.spark.util.Utils$.tryWithSafeFinally(Utils.scala:1439)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:500
        at java.base/java.util.concurrent.ThreadPoolExecutor.
↪runWorker(ThreadPoolExecutor.java:1128)
        at java.base/java.util.concurrent.ThreadPoolExecutor$Worker.
↪run(ThreadPoolExecutor.java:628)
        … 1 more
```

[ ]: