```
%%configure
{ "name":"Spark-to-Cosmos_DB_Change_Feed_Connector",
  "executorMemory": "8G",
  "executorCores": 2,
  "numExecutors": 2,
  "driverCores": 2,
  "jars": ["wasb:///example/jars/azure-cosmosdb-spark_2.4.0_2.11-1.4.1.jar",
"wasb:///example/jars/azure-documentdb-1.13.0.jar", "wasb:///example/jars/az
ure-documentdb-rx-0.9.0-rc2.jar", "wasb:///example/jars/json-20140107.jar",
"wasb:///example/jars/rxjava-1.3.0.jar", "wasb:///example/jars/rxnetty-0.4.2
0.jar"],
  "conf": {
    "spark.jars.excludes": "org.scala-lang:scala-reflect"
  }
}
```

Current session configs: {u'kind': 'pyspark', u'name': u'Spark-to-
Cosmos_DB_Change_Feed_Connector', u'numExecutors': 2, u'conf':
{u'spark.jars.excludes': u'org.scala-lang:scala-reflect'}, u'executorCores':
2, u'driverCores': 2, u'jars': [u'wasb:///example/jars/azure-cosmosdb-
spark_2.4.0_2.11-1.4.1.jar', u'wasb:///example/jars/azure-documentdb-
1.13.0.jar', u'wasb:///example/jars/azure-documentdb-rx-0.9.0-rc2.jar',
u'wasb:///example/jars/json-20140107.jar', u'wasb:///example/jars/rxjava-
1.3.0.jar', u'wasb:///example/jars/rxnetty-0.4.20.jar'], u'executorMemory':
u'8G'}
No active sessions.

```
database = "dc1-test-db"
collection = "states"

tweetsConfig = {
"Endpoint" : "https://dc1-cosmos-db.documents.azure.com:443/",
"Masterkey" : "48pZMByEhd5DNtylKgMK2n34AcAF0VTXsP0ewaTfJw1LMBBHaCG6AxcpZOgoe
ZT7VhNHU7Rs7dhjkvVzHbkFnQ==",
"Database" : database,
"Collection" : collection,
"preferredRegions" : "West US",
"SamplingRatio" : "1.0",
"schema_samplesize" : "200000",
"query_custom" : "select * FROM c"
}
```

Starting Spark application

YARN Application ID	Kind	State	SparkUI	Driver log	Current session?
application_1576190073047_0010	pyspark	idle	[Link](#)	[Link](#)	✔

SparkSession available as 'spark'.

```
tweets = spark.read.format("com.microsoft.azure.cosmosdb.spark").options(**tweetsConfig).load()
```

An error occurred while calling o97.load.
: org.apache.spark.SparkException: Job aborted due to stage failure: Task 0 in stage 0.0 failed 4 times, most recent failure: Lost task 0.3 in stage 0.0 (TID 3, wn1-dc1hdi.0t04vm2df5ve1a0cf2pxuar1sd.dx.internal.cloudapp.net, executor 2): java.lang.NoSuchMethodError: com.microsoft.azure.documentdb.FeedOptions.setResponseContinuationTokenLimitInKb(I)V
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.queryDocuments$1(CosmosDBRDDIterator.scala:195)

```
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader$l
zycompute(CosmosDBRDDIterator.scala:423)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader(C
osmosDBRDDIterator.scala:146)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.hasNext(
CosmosDBRDDIterator.scala:446)
        at scala.collection.Iterator$$anon$12.hasNext(Iterator.scala:439)
        at org.apache.spark.util.collection.ExternalSorter.insertAll(ExternalS
orter.scala:191)
        at org.apache.spark.shuffle.sort.SortShuffleWriter.write(SortShuffleWr
iter.scala:63)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:96)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:53)
        at org.apache.spark.scheduler.Task.run(Task.scala:109)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:34
5)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:748)

Driver stacktrace:
        at org.apache.spark.scheduler.DAGScheduler.org$apache$spark$scheduler$
DAGScheduler$$failJobAndIndependentStages(DAGScheduler.scala:1651)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply
(DAGScheduler.scala:1639)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply
(DAGScheduler.scala:1638)
        at scala.collection.mutable.ResizableArray$class.foreach(ResizableArra
y.scala:59)
        at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:48)
        at org.apache.spark.scheduler.DAGScheduler.abortStage(DAGScheduler.sca
la:1638)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFaile
d$1.apply(DAGScheduler.scala:831)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFaile
d$1.apply(DAGScheduler.scala:831)
        at scala.Option.foreach(Option.scala:257)
        at org.apache.spark.scheduler.DAGScheduler.handleTaskSetFailed(DAGSche
duler.scala:831)
```

```
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.doOnReceive
(DAGScheduler.scala:1872)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(D
AGScheduler.scala:1821)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(D
AGScheduler.scala:1810)
        at org.apache.spark.util.EventLoop$$anon$1.run(EventLoop.scala:48)
        at org.apache.spark.scheduler.DAGScheduler.runJob(DAGScheduler.scala:6
42)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2034)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2131)
        at org.apache.spark.rdd.RDD$$anonfun$aggregate$1.apply(RDD.scala:1124)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope
.scala:151)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope
.scala:112)
        at org.apache.spark.rdd.RDD.withScope(RDD.scala:363)
        at org.apache.spark.rdd.RDD.aggregate(RDD.scala:1117)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBSchema.schema(Cos
mosDBSchema.scala:71)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.com$micr
osoft$azure$cosmosdb$spark$schema$CosmosDBRelation$$lazySchema$lzycompute(Cos
mosDBRelation.scala:59)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.com$micr
osoft$azure$cosmosdb$spark$schema$CosmosDBRelation$$lazySchema(CosmosDBRelati
on.scala:44)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation$$anonfun
$schema$1.apply(CosmosDBRelation.scala:62)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation$$anonfun
$schema$1.apply(CosmosDBRelation.scala:62)
        at scala.Option.getOrElse(Option.scala:121)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.schema$l
zycompute(CosmosDBRelation.scala:62)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.schema(C
osmosDBRelation.scala:62)
        at org.apache.spark.sql.execution.datasources.DataSource.resolveRelati
on(DataSource.scala:432)
        at org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.s
cala:239)
        at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:227
)
        at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:164
)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
```

```
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImp
l.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAcc
essorImpl.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:748)
Caused by: java.lang.NoSuchMethodError: com.microsoft.azure.documentdb.FeedOp
tions.setResponseContinuationTokenLimitInKb(I)V
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.queryDoc
uments$1(CosmosDBRDDIterator.scala:195)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader$l
zycompute(CosmosDBRDDIterator.scala:423)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader(C
osmosDBRDDIterator.scala:146)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.hasNext(
CosmosDBRDDIterator.scala:446)
        at scala.collection.Iterator$$anon$12.hasNext(Iterator.scala:439)
        at org.apache.spark.util.collection.ExternalSorter.insertAll(ExternalS
orter.scala:191)
        at org.apache.spark.shuffle.sort.SortShuffleWriter.write(SortShuffleWr
iter.scala:63)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:96)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:53)
        at org.apache.spark.scheduler.Task.run(Task.scala:109)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:34
5)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        ... 1 more

Traceback (most recent call last):
  File "/usr/hdp/current/spark2-client/python/pyspark/sql/readwriter.py", lin
e 172, in load
```

```
    return self._df(self._jreader.load())
  File "/usr/hdp/current/spark2-client/python/lib/py4j-0.10.7-src.zip/py4j/ja
va_gateway.py", line 1257, in __call__
    answer, self.gateway_client, self.target_id, self.name)
  File "/usr/hdp/current/spark2-client/python/pyspark/sql/utils.py", line 63,
in deco
    return f(*a, **kw)
  File "/usr/hdp/current/spark2-client/python/lib/py4j-0.10.7-src.zip/py4j/pr
otocol.py", line 328, in get_return_value
    format(target_id, ".", name), value)
Py4JJavaError: An error occurred while calling o97.load.
: org.apache.spark.SparkException: Job aborted due to stage failure: Task 0 i
n stage 0.0 failed 4 times, most recent failure: Lost task 0.3 in stage 0.0 (
TID 3, wn1-dc1hdi.0t04vm2df5ve1a0cf2pxuar1sd.dx.internal.cloudapp.net, execut
or 2): java.lang.NoSuchMethodError: com.microsoft.azure.documentdb.FeedOption
s.setResponseContinuationTokenLimitInKb(I)V
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.queryDoc
uments$1(CosmosDBRDDIterator.scala:195)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader$l
zycompute(CosmosDBRDDIterator.scala:423)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader(C
osmosDBRDDIterator.scala:146)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.hasNext(
CosmosDBRDDIterator.scala:446)
        at scala.collection.Iterator$$anon$12.hasNext(Iterator.scala:439)
        at org.apache.spark.util.collection.ExternalSorter.insertAll(ExternalS
orter.scala:191)
        at org.apache.spark.shuffle.sort.SortShuffleWriter.write(SortShuffleWr
iter.scala:63)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:96)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:53)
        at org.apache.spark.scheduler.Task.run(Task.scala:109)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:34
5)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        at java.lang.Thread.run(Thread.java:748)

Driver stacktrace:
```

```
        at org.apache.spark.scheduler.DAGScheduler.org$apache$spark$scheduler$
DAGScheduler$$failJobAndIndependentStages(DAGScheduler.scala:1651)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply
(DAGScheduler.scala:1639)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$abortStage$1.apply
(DAGScheduler.scala:1638)
        at scala.collection.mutable.ResizableArray$class.foreach(ResizableArra
y.scala:59)
        at scala.collection.mutable.ArrayBuffer.foreach(ArrayBuffer.scala:48)
        at org.apache.spark.scheduler.DAGScheduler.abortStage(DAGScheduler.sca
la:1638)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFaile
d$1.apply(DAGScheduler.scala:831)
        at org.apache.spark.scheduler.DAGScheduler$$anonfun$handleTaskSetFaile
d$1.apply(DAGScheduler.scala:831)
        at scala.Option.foreach(Option.scala:257)
        at org.apache.spark.scheduler.DAGScheduler.handleTaskSetFailed(DAGSche
duler.scala:831)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.doOnReceive
(DAGScheduler.scala:1872)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(D
AGScheduler.scala:1821)
        at org.apache.spark.scheduler.DAGSchedulerEventProcessLoop.onReceive(D
AGScheduler.scala:1810)
        at org.apache.spark.util.EventLoop$$anon$1.run(EventLoop.scala:48)
        at org.apache.spark.scheduler.DAGScheduler.runJob(DAGScheduler.scala:6
42)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2034)
        at org.apache.spark.SparkContext.runJob(SparkContext.scala:2131)
        at org.apache.spark.rdd.RDD$$anonfun$aggregate$1.apply(RDD.scala:1124)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope
.scala:151)
        at org.apache.spark.rdd.RDDOperationScope$.withScope(RDDOperationScope
.scala:112)
        at org.apache.spark.rdd.RDD.withScope(RDD.scala:363)
        at org.apache.spark.rdd.RDD.aggregate(RDD.scala:1117)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBSchema.schema(Cos
mosDBSchema.scala:71)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.com$micr
osoft$azure$cosmosdb$spark$schema$CosmosDBRelation$$lazySchema$lzycompute(Cos
mosDBRelation.scala:59)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.com$micr
osoft$azure$cosmosdb$spark$schema$CosmosDBRelation$$lazySchema(CosmosDBRelati
on.scala:44)
```

```
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation$$anonfun
$schema$1.apply(CosmosDBRelation.scala:62)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation$$anonfun
$schema$1.apply(CosmosDBRelation.scala:62)
        at scala.Option.getOrElse(Option.scala:121)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.schema$l
zycompute(CosmosDBRelation.scala:62)
        at com.microsoft.azure.cosmosdb.spark.schema.CosmosDBRelation.schema(C
osmosDBRelation.scala:62)
        at org.apache.spark.sql.execution.datasources.DataSource.resolveRelati
on(DataSource.scala:432)
        at org.apache.spark.sql.DataFrameReader.loadV1Source(DataFrameReader.s
cala:239)
        at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:227
)
        at org.apache.spark.sql.DataFrameReader.load(DataFrameReader.scala:164
)
        at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
        at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImp
l.java:62)
        at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAcc
essorImpl.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:748)
Caused by: java.lang.NoSuchMethodError: com.microsoft.azure.documentdb.FeedOp
tions.setResponseContinuationTokenLimitInKb(I)V
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.queryDoc
uments$1(CosmosDBRDDIterator.scala:195)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader$l
zycompute(CosmosDBRDDIterator.scala:423)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.reader(C
osmosDBRDDIterator.scala:146)
        at com.microsoft.azure.cosmosdb.spark.rdd.CosmosDBRDDIterator.hasNext(
CosmosDBRDDIterator.scala:446)
        at scala.collection.Iterator$$anon$12.hasNext(Iterator.scala:439)
        at org.apache.spark.util.collection.ExternalSorter.insertAll(ExternalS
orter.scala:191)
```

```
        at org.apache.spark.shuffle.sort.SortShuffleWriter.write(SortShuffleWr
iter.scala:63)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:96)
        at org.apache.spark.scheduler.ShuffleMapTask.runTask(ShuffleMapTask.sc
ala:53)
        at org.apache.spark.scheduler.Task.run(Task.scala:109)
        at org.apache.spark.executor.Executor$TaskRunner.run(Executor.scala:34
5)
        at java.util.concurrent.ThreadPoolExecutor.runWorker(ThreadPoolExecuto
r.java:1149)
        at java.util.concurrent.ThreadPoolExecutor$Worker.run(ThreadPoolExecut
or.java:624)
        ... 1 more
```