

# PySpark: java.lang.OutOfMemoryError: Java heap space

Asked 2015-09-01T16:45:02   Active 2 years ago   Viewed 83k times



63



16



I have been using PySpark with lpython lately on my server with 24 CPUs and 32GB RAM. Its running only on one machine. In my process, I want to collect huge amount of data as is give in below code:

```
train_dataRDD = (train.map(lambda x:getTagsAndText(x))
    .filter(lambda x:x[-1]!=[])
    .flatMap(lambda (x,text,tags): [(tag,(x,text)) for tag in tags])
    .groupByKey()
    .mapValues(list))
```

When I do

```
training_data = train_dataRDD.collectAsMap()
```

It gives me outOfMemory Error. Java heap space . Also, I can not perform any operations on Spark after this error as it looses connection with Java. It gives Py4JNetworkError: Cannot connect to the java server .

It looks like heap space is small. How can I set it to bigger limits?

## EDIT:

Things that I tried before running:

```
sc._conf.set('spark.executor.memory', '32g').set('spark.driver.memory', '32g').set('spark.driver.maxResultsSize', '0')
```

I changed the spark options as per the documentation here(if you do ctrl-f and search for spark.executor.extraJavaOptions) :

<http://spark.apache.org/docs/1.2.1/configuration.html>

It says that I can avoid OOMs by setting spark.executor.memory option. I did the same thing but it seem not be working.

java apache-spark out-of-memory heap-memory pyspark

Share Improve this question Follow

edited Sep 1, 2015 at 19:50

asked Sep 1, 2015 at 16:45



pg2455

4,483

13

45

73

Check this question [stackoverflow.com/questions/21138751/...](https://stackoverflow.com/questions/21138751/...) – Bruno Caceiro Sep 1, 2015 at 16:52

@bcaceiro: I see lot of spark options being set in the post. I dont use scala. I am using IPython. Do you know if I can set those options from within the shell?  
– pg2455 Sep 1, 2015 at 17:12

@bcaceiro : Updated the question with suggestion from the post that you directed me too. It seems like there is some problem with JVM. – pg2455 Sep 1, 2015 at 18:59

### 3 Answers

Active

Oldest

Votes



80



After trying out loads of configuration parameters, I found that there is only one need to be changed to enable more Heap space and i.e.  
`spark.driver.memory`.

```
sudo vim $SPARK_HOME/conf/spark-defaults.conf
#uncomment the spark.driver.memory and change it according to your use. I changed
it to below
spark.driver.memory 15g
# press : and then wq! to exit vim editor
```

Close your existing spark application and re run it. You will not encounter this error again. :)

Share Improve this answer Follow

answered Sep 3, 2015 at 15:42



pg2455

4,483

13

45

73

4 Can you change this conf value from the actual script (ie. `set('spark.driver.memory', '15g')`) ? – swdev Dec 21, 2015 at 7:31

1 I tried doing it but was not successful. I think it need to restart with new global parameters. – [pg2455](#) Dec 21, 2015 at 15:48

12 From docs: spark.driver.memory "Amount of memory to use for the driver process, i.e. where SparkContext is initialized. (e.g. 1g, 2g). Note: In client mode, this config must not be set through the SparkConf directly in your application, because the driver JVM has already started at that point. Instead, please set this through the --driver-memory command line option or in your default properties file." – [Răzvan Flavius Panda](#) May 27, 2016 at 15:15

I was running the Spark code using SBT run from IDEA SBT Console, the fix for me was to add `-Xmx4096M -d64` to the java VM parameters that get passed on the SBT Console launch. This is under `Other settings` -> `SBT` . – [Răzvan Flavius Panda](#) May 31, 2016 at 18:24

Spark keeps evolving. So you might have to look into its documentation and find out the configuration parameters that correlate to the memory allocation.  
– [pg2455](#) Nov 30, 2016 at 20:19



If you're looking for the way to set this from within the script or a jupyter notebook, you can do:

33

```
from pyspark.sql import SparkSession

spark = SparkSession.builder \
    .master('local[*]') \
    .config("spark.driver.memory", "15g") \
    .appName('my-cool-app') \
    .getOrCreate()
```



Share Improve this answer Follow

answered Feb 17, 2020 at 17:44



[louis\\_guitton](#)

**4,003**

1

26

30