

```
JAVA:                jdk 1.8.0_65
PYTHON:              python 3.7.4
SPARK:               spark 2.4.3
JUPYTER NOTEBOOK    6.0.0
```

FOR ERROR SOLUTION:

**** Restart kernel for each run**

**** Make necessary updates:**

```
!pip install psutil
!python -m pip install --upgrade pip
!pip install findspark
```

****Add these into import and create sections**

```
%config IPCompleter.greedy=True
import findspark
findspark.init()
from pyspark import SparkConf
from pyspark import SparkContext
```

```
conf=SparkConf().setMaster("local[*]").setAppName("MyApp")
sc = SparkContext.getOrCreate(conf=conf)
```

**** Create just one SparkContext (sc) for each run.**

**** Do not print big sized files using collect() method. This cause memory exceed. Instead, just print small amount using take(). Ex: take(10)**

**** Do not use same variable for reassigning. Change the name of variable.**

**** After each transformation function, use action function Ex: use take(10) after reduceByKey**