DemoNotebook

January 29, 2019

In [4]:
```
from pyspark import SparkConf, SparkContext
sc = SparkContext(conf=SparkConf().setAppName("MyApp").setMaster("local"))

import re

def parse_article(line):
    try:
        article_id, text = unicode(line.rstrip()).split('\t', 1)
        text = re.sub("^\W+|\W+$", "", text, flags=re.UNICODE)
        words = re.split("\W*\s+\W*", text, flags=re.UNICODE)
        return words
    except ValueError as e:
        return []

wiki = sc.textFile("/data/wiki/en_articles_part/articles-part", 16).map(parse_article)
result = wiki.take(1)[0]
```

```
---------------------------------------------------------------------------

ValueError                                Traceback (most recent call last)

<ipython-input-4-bce5f1bc4ca2> in <module>()
  1 from pyspark import SparkConf, SparkContext
----> 2 sc = SparkContext(conf=SparkConf().setAppName("MyApp").setMaster("local"))
  3
  4 import re
  5


/usr/local/spark/python/pyspark/context.py in __init__(self, master, appName, sparkHome,
113         """
114         self._callsite = first_spark_call() or CallSite(None, None, None)
--> 115         SparkContext._ensure_initialized(self, gateway=gateway, conf=conf)
116         try:
117             self._do_init(master, appName, sparkHome, pyFiles, environment, batchSiz
```

```
        /usr/local/spark/python/pyspark/context.py in _ensure_initialized(cls, instance, gateway
        273                         " created by %s at %s:%s "
        274                         % (currentAppName, currentMaster,
    --> 275                            callsite.function, callsite.file, callsite.linenum))
        276                 else:
        277                     SparkContext._active_spark_context = instance


        ValueError: Cannot run multiple SparkContexts at once; existing SparkContext(app=MyApp,
```

```
In [2]: for word in result[:50]:
            print word
```

```
Anarchism
Anarchism
is
often
defined
as
a
political
philosophy
which
holds
the
state
to
be
undesirable
unnecessary
or
harmful
The
following
sources
cite
anarchism
as
a
political
philosophy
Slevin
Carl
Anarchism
The
Concise
Oxford
```

```
Dictionary
of
Politics
Ed
Iain
McLean
and
Alistair
McMillan
Oxford
University
Press
2003
However
others
argue
```

```
In [23]: a = sc.parallelize([1, 2, 3, 4, 5])
         a
         a.getNumPartitions()
         a.collect()
         b=a.map(lambda x:2*x)
         b
         b.collect()
         from __future__ import print_function
         b=a.map(lambda x:(print(x),2*x)[1])
         b.collect()

Out[23]: [2, 4, 6, 8, 10]
```