

```
In [1]: import findspark
        from pyspark.streaming import StreamingContext
        from pyspark.sql import SparkSession
        import math

        findspark.init()
```

```
In [11]: try: ssc.stop(True, True)
        except: pass
        try: spark.stop()
        except: pass
```

```
In [3]: spark=SparkSession.builder.appName("SparkStreaming-01").master('local[1]').getOrCreate()
        sc=spark.sparkContext
        # .config("spark.driver.allowMultipleContexts","true")
        spark
```

Out[3]: **SparkSession - in-memory**

SparkContext

[Spark UI](#)

Version	v2.4.8
Master	local[1]
AppName	SparkStreaming-01

```
In [4]: spark.conf.get("spark.sql.shuffle.partitions")
```

Out[4]: '200'

```
In [5]: # spark.conf.set("spark.sql.shuffle.partitions", 5)
        # spark.conf.get("spark.sql.shuffle.partitions")
```

```
In [6]: ssc=StreamingContext(sc, 1)
        ssc # 1=1 second
```

Out[6]: <pyspark.streaming.context.StreamingContext at 0x2151aeb0988>

```
In [7]: rdd1=ssc.textFileStream("hdfs://localhost:9000/data")
        # rdd1=ssc.textFileStream("data/")
```

```
In [8]: def add(x,y):
        return x+y

        rdd1=rdd1.flatMap(lambda x: x.split(' '))
        rdd1=rdd1.map(lambda x: (x,1)).reduceByKey(add)
        rdd1.pprint()
```

```
In [ ]: def printRDD(rdd):
        print(rdd.collect())
```

```
rdd1.foreachRDD(printRDD)
```

In [9]:

```
# close notebook session and restart notebook command again + reopen this file  
ssc.start()  
ssc.awaitTerminationOrTimeout(2) # 1=1 second
```

Out[9]:

False

```
-----  
Time: 2021-09-17 16:40:49  
-----  
  
-----  
Time: 2021-09-17 16:40:50  
-----  
  
-----  
Time: 2021-09-17 16:40:51  
-----  
  
-----  
Time: 2021-09-17 16:40:52  
-----  
  
-----  
Time: 2021-09-17 16:40:53  
-----  
  
-----  
Time: 2021-09-17 16:40:54  
-----  
  
-----  
Time: 2021-09-17 16:40:55  
-----  
  
-----  
Time: 2021-09-17 16:40:56  
-----  
  
-----  
Time: 2021-09-17 16:40:57  
-----  
  
-----  
Time: 2021-09-17 16:40:58  
-----  
  
-----  
Time: 2021-09-17 16:40:59  
-----  
  
-----  
Time: 2021-09-17 16:41:00  
-----  
  
-----  
Time: 2021-09-17 16:41:01  
-----
```

In []: