

ex61

August 20, 2022

```
[ ]: from pyspark import SparkConf, SparkContext
    from pyspark.streaming import StreamingContext

    batch_size = 2

    conf = SparkConf().setAppName("ex61")
    sc = SparkContext(conf=conf)
    ssc = StreamingContext(sc, batch_size)
```

```
[2]: inputPath = "data/Ex61/data/"
```

```
[3]: lines = ssc.textFileStream(inputPath)
```

```
[4]: stationBatch = lines.map(lambda station : station.split(",")[1])\
    .reduce(lambda v1, v2 : max(v1, v2))
```

```
[ ]: stationBatch.pprint()
```

```
[ ]: ssc.start()
```

```
[ ]: ssc.awaitTerminationOrTimeout(20)
    ssc.stop(stopSparkContext=False)
```