ex60

August 20, 2022

[ ]:

```python
from pyspark import SparkContext, SparkConf
from pyspark.streaming import StreamingContext

batch_size = 2

conf = SparkConf().setAppName("ex60")
sc = SparkContext(conf=conf)
ssc = StreamingContext(sc, batch_size)
```

[2]:

```python
inputPath = "data/Ex60/data/"
```

[3]:

```python
lines = ssc.textFileStream(inputPath)
```

[5]:

```python
fullStations = lines\
    .filter(lambda station : int(station.split(",")[1])==0)\
        .map(lambda station : station.split(",")[0])\
            .transform(lambda batchRDD : batchRDD.distinct())\
                .pprint()
```

[ ]:

```python
ssc.start()
```

[ ]:

```python
ssc.awaitTerminationOrTimeout(20)
ssc.stop(stopSparkContext=False)
```