

ex31

August 12, 2022

```
[ ]: from pyspark import SparkContext, SparkConf
conf = SparkConf().setAppName("ex31")
sc = SparkContext(conf=conf)
```

```
[2]: inputPath = "data/Ex31/data"
outputPath = "out31/"
```

```
[10]: def filterAndExtractIp(line):
      IPS = list()

      if line.lower().find("google")>=0:
          IP = line.split("-")[0]
          IPS.append(IP)
      return IPS
```

```
[11]: ipsRDD = sc.textFile(inputPath).flatMap(filterAndExtractIp).distinct()
```

```
[12]: ipsRDD.saveAsTextFile(outputPath)
```