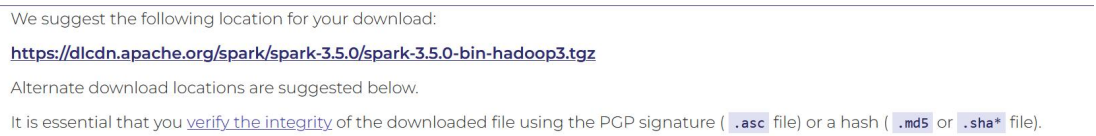


通过 Spark 实现词频统计。

通过官网进行下载: <https://spark.apache.org/downloads.html>



```
Welcome to
      _--_
     /  \_--_--_--_--_--_--_--_--_--_
    /  \_--_--_--_--_--_--_--_--_--_
   /  \_--_--_--_--_--_--_--_--_--_
  /  \_--_--_--_--_--_--_--_--_--_
 /  \_--_--_--_--_--_--_--_--_--_
/_  \_--_--_--_--_--_--_--_--_--_

version 3.5.0

Using Python version 3.11.5 (main, Sep 11 2023 08:31:25)
Spark context Web UI available at http://172.25.142.242:4040
Spark context available as 'sc' (master = local[*], app id = local-1701351226916
).
SparkSession available as 'spark'.
```

在 mac 上使用 `pyspark`，需要强行安装第三方包。将 `spark` 文件夹下的 `pyspark` 文件拷贝到 `python/site-packages` 文件夹目录下即可。

3. 程序运行

```
from operator import add
from pyspark import SparkContext
```

word.txt 内容如下:

```
文件  编辑  查看

good bad cool
hadoop spark mlib
good spark mlib
cool spark bad
```

结果如下:

```
good: 2
hadoop: 1
bad: 2
cool: 2
spark: 3
mlib: 2
```