At First,

```
! pip3 install mrjob
```

```
WARNING: Running pip install with root privileges is generally not a good idea. Try `pip3 install --user` in
stead.
Collecting mrjob
  Downloading http://mirrors.aliyun.com/pypi/packages/75/1e/421eb3ad0a359f7b2d0619699d5d908c10b52dcd51ab4cac
fc8e8ab048b9/mrjob-0.6.11-py2.py3-none-any.whl (435kB)
    100% |████████████████████████████████| 440kB 73.2MB/s
Requirement already satisfied: PyYAML>=3.10 in /usr/lib64/python3.6/site-packages (from mrjob)
Collecting google-cloud-logging>=1.9.0 (from mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/b9/06/3a89173f80d02d275ab4fdca5283a7fdb84845e214f3383b
5d52c29f4b5f/google_cloud_logging-1.14.0-py2.py3-none-any.whl (134kB)
    100% |████████████████████████████████| 143kB 4.9MB/s
Collecting botocore>=1.6.0 (from mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/8e/7b/88f10115b4748f86be6b7b1d8761ba5023fccf6e6cbe762e
368f63eddcf9/botocore-1.12.253-py2.py3-none-any.whl (5.7MB)
    100% |████████████████████████████████| 5.8MB 82.5MB/s
Collecting google-cloud-dataproc>=0.3.0 (from mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/2d/7e/78e34a5f68252c3fd853cf11bc334d301cd9d098649bf241
931327b32d13/google_cloud_dataproc-0.5.0-py2.py3-none-any.whl (239kB)
    100% |████████████████████████████████| 245kB 481kB/s
Collecting google-cloud-storage>=1.13.1 (from mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/38/83/40049ccfb9e4504edf4988117a8b4652111708b150028dd6
63acfa061300/google_cloud_storage-1.20.0-py2.py3-none-any.whl (70kB)
    100% |████████████████████████████████| 71kB 578kB/s
Collecting boto3>=1.4.6 (from mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/f6/fa/6397049020b312f71c397fff8d10247c2e49da760e2807af
7d21e3c23695/boto3-1.9.253-py2.py3-none-any.whl (128kB)
    100% |████████████████████████████████| 133kB 80.2MB/s
Collecting google-api-core[grpc]<2.0.0dev,>=1.14.0 (from google-cloud-logging>=1.9.0->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/29/3a/c528ef37f48d6ffba16f0f3c0426456ba21e0dd32be9c61a
2ade93e07faa/google_api_core-1.14.3-py2.py3-none-any.whl (68kB)
    100% |████████████████████████████████| 71kB 4.2MB/s
Collecting google-cloud-core<2.0dev,>=1.0.3 (from google-cloud-logging>=1.9.0->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/ee/f0/084f598629db8e6ec3627688723875cdb03637acb6d86999
bb105a71df64/google_cloud_core-1.0.3-py2.py3-none-any.whl
Collecting python-dateutil<3.0.0,>=2.1; python_version >= "2.7" (from botocore>=1.6.0->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/41/17/c62faccbfbd163c7f57f3844689e3a78bae1f403648a6afb
1d0866d87fbb/python_dateutil-2.8.0-py2.py3-none-any.whl (226kB)
    100% |████████████████████████████████| 235kB 77.9MB/s
Collecting jmespath<1.0.0,>=0.7.1 (from botocore>=1.6.0->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/83/94/7179c3832a6d45b266ddb2aac329e101367fbdb11f425f13
771d27f225bb/jmespath-0.9.4-py2.py3-none-any.whl
Collecting docutils<0.16,>=0.10 (from botocore>=1.6.0->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/22/cd/a6aa959dca619918ccb55023b4cb151949c64d4d5d55b3f4
ffd7eee0c6e8/docutils-0.15.2-py3-none-any.whl (547kB)
    100% |████████████████████████████████| 552kB 9.4MB/s
Collecting urllib3<1.26,>=1.20; python_version >= "3.4" (from botocore>=1.6.0->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/e0/da/55f51ea951e1b7c63a579c09dd7db825bb730ec1fe9c0180
fc77bfb31448/urllib3-1.25.6-py2.py3-none-any.whl (125kB)
    100% |████████████████████████████████| 133kB 80.5MB/s
Collecting google-auth>=1.2.0 (from google-cloud-storage>=1.13.1->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/c5/9b/ed0516cc1f7609fb0217e3057ff4f0f9f3e3ce79a369c6af
4a6c5ca25664/google_auth-1.6.3-py2.py3-none-any.whl (73kB)
    100% |████████████████████████████████| 81kB 66.8MB/s
Collecting google-resumable-media!=0.4.0,<0.5dev,>=0.3.1 (from google-cloud-storage>=1.13.1->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/96/d7/b29a41b01b854480891dfc408211ffb0cc7a2a3d5f15a3b6
740ec18c845b/google_resumable_media-0.4.1-py2.py3-none-any.whl
Collecting s3transfer<0.3.0,>=0.2.0 (from boto3>=1.4.6->mrjob)
  Downloading http://mirrors.aliyun.com/pypi/packages/16/8a/1fc3dba0c4923c2a76e1ff0d52b305c44606da63f718d14d
3231e21c51b0/s3transfer-0.2.1-py2.py3-none-any.whl (70kB)
    100% |████████████████████████████████| 71kB 61.3MB/s
Collecting protobuf>=3.4.0 (from google-api-core[grpc]<2.0.0dev,>=1.14.0->google-cloud-logging>=1.9.0->mrjob
)
  Downloading http://mirrors.aliyun.com/pypi/packages/a8/52/d8d2dbff74b8bf517c42db8d44c3f9ef6555e6f5d6caddfa
3f207b9143df/protobuf-3.10.0-cp36-cp36m-manylinux1_x86_64.whl (1.3MB)
```

```
             100% |█████████████████████████████| 1.3MB 79.7MB/s
Collecting googleapis-common-protos<2.0dev,>=1.6.0 (from google-api-core[grpc]<2.0.0dev,>=1.14.0->google-clo
ud-logging>=1.9.0->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/eb/ee/e59e74ecac678a14d6abefb9054f0bbcb318a6452a30df37
76f133886d7d/googleapis-common-protos-1.6.0.tar.gz
Collecting requests<3.0.0dev,>=2.18.0 (from google-api-core[grpc]<2.0.0dev,>=1.14.0->google-cloud-logging>=1
.9.0->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/51/bd/23c926cd341ea6b7dd0b2a00aba99ae0f828be89d72b2190
f27c11d4b7fb/requests-2.22.0-py2.py3-none-any.whl (57kB)
             100% |█████████████████████████████| 61kB 59.9MB/s
Requirement already satisfied: setuptools>=34.0.0 in /usr/lib/python3.6/site-packages (from google-api-core[
grpc]<2.0.0dev,>=1.14.0->google-cloud-logging>=1.9.0->mrjob)
Collecting pytz (from google-api-core[grpc]<2.0.0dev,>=1.14.0->google-cloud-logging>=1.9.0->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/e7/f9/f0b53f88060247251bf481fa6ea62cd0d25bf1b11a87888e
53ce5b7c8ad2/pytz-2019.3-py2.py3-none-any.whl (509kB)
             100% |█████████████████████████████| 512kB 75.4MB/s
Requirement already satisfied: six>=1.10.0 in /usr/lib/python3.6/site-packages (from google-api-core[grpc]<2
.0.0dev,>=1.14.0->google-cloud-logging>=1.9.0->mrjob)
Collecting grpcio<2.0dev,>=1.8.2; extra == "grpc" (from google-api-core[grpc]<2.0.0dev,>=1.14.0->google-clou
d-logging>=1.9.0->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/e8/cb/ebf7b54c5d4ad521d88ee7826dfa0fc3ac84502361ad7e5c
b739ea5057a4/grpcio-1.24.1-cp36-cp36m-manylinux1_x86_64.whl (2.3MB)
             100% |█████████████████████████████| 2.3MB 79.6MB/s
Collecting cachetools>=2.0.0 (from google-auth>=1.2.0->google-cloud-storage>=1.13.1->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/2f/a6/30b0a0bef12283e83e58c1d6e7b5aabc7acfc4110df81a44
71655d33e704/cachetools-3.1.1-py2.py3-none-any.whl
Collecting pyasn1-modules>=0.2.1 (from google-auth>=1.2.0->google-cloud-storage>=1.13.1->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/52/50/bb4cefca37da63a0c52218ba2cb1b1c36110d84dcbae8aa4
8cd67c5e95c2/pyasn1_modules-0.2.7-py2.py3-none-any.whl (131kB)
             100% |█████████████████████████████| 133kB 77.6MB/s
Collecting rsa>=3.1.4 (from google-auth>=1.2.0->google-cloud-storage>=1.13.1->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/02/e5/38518af393f7c214357079ce67a317307936896e961e3545
0b70fad2a9cf/rsa-4.0-py2.py3-none-any.whl
Collecting certifi>=2017.4.17 (from requests<3.0.0dev,>=2.18.0->google-api-core[grpc]<2.0.0dev,>=1.14.0->goo
gle-cloud-logging>=1.9.0->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/18/b0/8146a4f8dd402f60744fa380bc73ca47303cccf8b9190fd1
6a827281eac2/certifi-2019.9.11-py2.py3-none-any.whl (154kB)
             100% |█████████████████████████████| 163kB 80.6MB/s
Requirement already satisfied: idna<2.9,>=2.5 in /usr/lib/python3.6/site-packages (from requests<3.0.0dev,>=
2.18.0->google-api-core[grpc]<2.0.0dev,>=1.14.0->google-cloud-logging>=1.9.0->mrjob)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in /usr/lib/python3.6/site-packages (from requests<3.0.
0dev,>=2.18.0->google-api-core[grpc]<2.0.0dev,>=1.14.0->google-cloud-logging>=1.9.0->mrjob)
Collecting pyasn1<0.5.0,>=0.4.6 (from pyasn1-modules>=0.2.1->google-auth>=1.2.0->google-cloud-storage>=1.13.
1->mrjob)
    Downloading http://mirrors.aliyun.com/pypi/packages/a1/71/8f0d444e3a74e5640a3d5d967c1c6b015da9c655f35b2d30
8a55d907a517/pyasn1-0.4.7-py2.py3-none-any.whl (76kB)
             100% |█████████████████████████████| 81kB 69.5MB/s
Installing collected packages: protobuf, googleapis-common-protos, certifi, urllib3, requests, pytz, cacheto
ols, pyasn1, pyasn1-modules, rsa, google-auth, grpcio, google-api-core, google-cloud-core, google-cloud-logg
ing, python-dateutil, jmespath, docutils, botocore, google-cloud-dataproc, google-resumable-media, google-cl
oud-storage, s3transfer, boto3, mrjob
    Running setup.py install for googleapis-common-protos ... done
Successfully installed boto3-1.9.253 botocore-1.12.253 cachetools-3.1.1 certifi-2019.9.11 docutils-0.15.2 go
ogle-api-core-1.14.3 google-auth-1.6.3 google-cloud-core-1.0.3 google-cloud-dataproc-0.5.0 google-cloud-logg
ing-1.14.0 google-cloud-storage-1.20.0 google-resumable-media-0.4.1 googleapis-common-protos-1.6.0 grpcio-1.
24.1 jmespath-0.9.4 mrjob-0.6.11 protobuf-3.10.0 pyasn1-0.4.7 pyasn1-modules-0.2.7 python-dateutil-2.8.0 pyt
z-2019.3 requests-2.22.0 rsa-4.0 s3transfer-0.2.1 urllib3-1.25.6
```

Check it,

```
! py3 -c "import mrjob"
```

Now, Be a Code,

```
! cat 032.Test_Mrjob_MapReduce.py


#!/usr/bin/env python
# -*- coding:utf-8 -*-


from mrjob.job import MRJob


class MRWordCounter(MRJob):
    def mapper(self,key,line):
        for word in line.split():
            yield word,1


    def reducer(self,word,occurrences):
        yield word,sum(occurrences)


if __name__ == '__main__':
    MRWordCounter.run()
```

Now, Be a word list,

```
! cat datalist


foo foo quux labs foo bar quux abc bar see you by test welcome test
abc labs foo me python hadoop ab ac bc bec python
```

Run! Amazing

```
! py3 run.py datalist -o out


No configs found; falling back on auto-configuration
No configs specified for inline runner
Running step 1 of 1...
Creating temp directory /tmp/run.root.20191022.010813.221449
job output is in out
Removing temp directory /tmp/run.root.20191022.010813.221449...
```

See Result...

```
! ll


total 8
-rw-r--r-- 1 root root 118 Oct 22 01:06 datalist
drwxr-xr-x 6 root root 192 Oct 22 01:08 out
-rw-r--r-- 1 root root 324 Oct 22 00:52 run.py


! cat out/part-0000*


"ab"    1
"abc"   2
"ac"    1
"bar"   2
"bc"    1
"bec"   1
"by"    1
"foo"   4
"hadoop"    1
"labs"  2
"me"    1
"python"    2
"quux"  2
"see"   1
"test"  2
"welcome"   1
"you"   1
```

Test Big Data!

032.Test_Mrjob_MapReduce/trdata.ipynb

```python
import random
p = {
    0: "XXX ",
    1: "YYY ",
    2: "ZZZ \n",
    3: "PAN ",
    4: "ZIP ",
    5: "QQQ ",
}
n = 1000000
with open("r1_list", "w") as f:
    for i in range(n):
        f.write(p[random.randint(0, 5)])
```

Take a look.

```
! tail -n 10 r1_list
```

```
PAN ZZZ
QQQ YYY XXX QQQ QQQ ZZZ
YYY XXX ZIP PAN PAN XXX XXX YYY ZZZ
YYY QQQ PAN XXX ZZZ
YYY ZZZ
XXX QQQ QQQ ZZZ
XXX PAN YYY QQQ ZZZ
PAN PAN YYY ZZZ
ZIP XXX PAN ZZZ
ZIP YYY QQQ YYY ZIP PAN QQQ ZIP ZIP QQQ PAN XXX PAN PAN YYY XXX PAN XXX QQQ PAN QQQ
```

Result

```
! py3 run.py r1_list -o out_r1
```

```
No configs found; falling back on auto-configuration
No configs specified for inline runner
Running step 1 of 1...
Creating temp directory /tmp/run.root.20191022.035100.093909
job output is in out_r1
Removing temp directory /tmp/run.root.20191022.035100.093909...
```

```
! cat out_r1/part-0000*
```

```
"PAN"    166990
"QQQ"    166897
"XXX"    166374
"YYY"    166436
"ZIP"    166476
"ZZZ"    166827
```