

```
!pip install mrjob
```

```
Collecting mrjob
  Downloading mrjob-0.7.4-py2.py3-none-any.whl (439 kB)
    439.6/439.6 kB 5.2 MB/s eta 0:00:00
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.10/dist-packages (from mrjob) (6.0.1)
Installing collected packages: mrjob
Successfully installed mrjob-0.7.4
```

```
%%writefile textfile.txt
Ali, 12-3-20, S1,S2,S3
Fahad, 14-04-22, S1,S2,S7,S3
Ali, 20-4-22, S1,S2
Sara, 20-4-22, S1,S7,S9
Amina, 20-4-22, S1
```

```
Overwriting textfile.txt
```

```
#1
%%writefile first.py
from mrjob.job import MRJob
class SongsPlayed(MRJob):

    def mapper(self, _, line):
        parts = line.strip().split(',')
        name = parts[0]
        date = parts[1]
        songs = parts[2:]

        if len(songs) > 5:
            yield name, 1
```

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

```
Overwriting first.py
```

```
!python first.py textfile.txt
```

```
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/first.root.20230821.174744.887423
Running step 1 of 1...
job output is in /tmp/first.root.20230821.174744.887423/output
Streaming final output from /tmp/first.root.20230821.174744.887423/output...
Removing temp directory /tmp/first.root.20230821.174744.887423...
```

```
#2
%%writefile second.py
from mrjob.job import MRJob
class SongsPlayedCount(MRJob):

    def mapper(self, _, line):
        parts = line.strip().split(',')
        name = parts[0]
        date = parts[1]
        songs = parts[2:]

        yield (name, date), len(songs)
```

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

```
Writing second.py
```

```
!python second.py textfile.txt
```

```
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/second.root.20230821.174829.864747
Running step 1 of 1...
job output is in /tmp/second.root.20230821.174829.864747/output
Streaming final output from /tmp/second.root.20230821.174829.864747/output...
["Ali", " 12-3-20"]      3
```

```
["Fahad", " 14-04-22"] 4
["Ali", " 20-4-22"] 2
["Sara", " 20-4-22"] 3
["Amina", " 20-4-22"] 1
Removing temp directory /tmp/second.root.20230821.174829.864747...
```

#3

```
%%writefile third.py
from mrjob.job import MRJob
```

```
class SongsPlayedTillNow(MRJob):
```

```
    def mapper(self, _, line):
        parts = line.strip().split(',')
        name = parts[0]
        songs = parts[2:]
```

```
        for song in songs:
            yield name, 1
```

```
    def combiner(self, name, counts):
        yield name, sum(counts)
```

```
    def reducer(self, name, counts):
        yield name, sum(counts)
```

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

```
Writing third.py
```

```
!python third.py textfile.txt
```

```
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/third.root.20230821.175234.985844
Running step 1 of 1...
job output is in /tmp/third.root.20230821.175234.985844/output
Streaming final output from /tmp/third.root.20230821.175234.985844/output...
"Ali" 5
"Amina" 1
"Fahad" 4
"Sara" 3
Removing temp directory /tmp/third.root.20230821.175234.985844...
```

#4

```
%%writefile forth.py
from mrjob.job import MRJob
class AverageSongsPerDay(MRJob):
```

```
    def mapper(self, _, line):
        parts = line.strip().split(',')
        songs = parts[2:]
```

```
        yield "total", len(songs)
        yield "count", 1
```

```
    def combiner(self, key, values):
        if key == "total":
            yield key, sum(values)
        elif key == "count":
            yield key, sum(values)
```

```
    def reducer(self, key, values):
        total_songs = 0
        total_days = 0
```

```
        for sub_key, value in zip(["total", "count"], values):
            if sub_key == "total":
                total_songs += value
            elif sub_key == "count":
                total_days += value
```

```
        if key == "total":
            yield "average", total_songs / total_days
```

```

if __name__ == '__main__':
    AverageSongsPerDay.run()

    Writing forth.py

!python forth.py textfile.txt

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/forth.root.20230821.175521.450445
Running step 1 of 1...
job output is in /tmp/forth.root.20230821.175521.450445/output
Streaming final output from /tmp/forth.root.20230821.175521.450445/output...
"average"      1.4
Removing temp directory /tmp/forth.root.20230821.175521.450445...

```

```

#5
%%writefile fifth.py
from mrjob.job import MRJob
class SongPlayedCount(MRJob):

    def mapper(self, _, line):
        parts = line.strip().split(',')
        songs = parts[2:]

        for song in songs:
            yield song, 1

    def combiner(self, song, counts):
        yield song, sum(counts)

    def reducer(self, song, counts):
        yield song, sum(counts)

if __name__ == '__main__':
    SongPlayedCount.run()

    Writing fifth.py

```

```

!python fifth.py textfile.txt

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/fifth.root.20230821.175702.806605
Running step 1 of 1...
job output is in /tmp/fifth.root.20230821.175702.806605/output
Streaming final output from /tmp/fifth.root.20230821.175702.806605/output...
"S1" 5
"S2" 3
"S3" 2
"S7" 2
"S9" 1
Removing temp directory /tmp/fifth.root.20230821.175702.806605...

```

```

#6
%%writefile sixth.py
from mrjob.job import MRJob
class MaxSongsUsers(MRJob):

    def mapper(self, _, line):
        parts = line.strip().split(',')
        name = parts[0]
        songs = parts[2:]

        yield name, len(songs)

    def combiner(self, name, song_counts):
        yield name, max(song_counts)

    def reducer_init(self):
        self.max_count = 0
        self.max_users = []

    def reducer(self, name, song_counts):
        count = max(song_counts)

```

```

        if count > self.max_count:
            self.max_count = count
            self.max_users = [name]
        elif count == self.max_count:
            self.max_users.append(name)

    def reducer_final(self):
        for user in self.max_users:
            yield user, self.max_count

if __name__ == '__main__':
    MaxSongsUsers.run()

    Overwriting sixth.py

!python sixth.py textfile.txt

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/sixth.root.20230821.181229.243381
Running step 1 of 1...
job output is in /tmp/sixth.root.20230821.181229.243381/output
Streaming final output from /tmp/sixth.root.20230821.181229.243381/output...
"Ali" 3
"Fahad" 4
"Sara" 3
Removing temp directory /tmp/sixth.root.20230821.181229.243381...

```

```

#7
%%writefile seventh.py
from mrjob.job import MRJob
class MaxSongsUpTillNow(MRJob):

    def mapper(self, _, line):
        parts = line.strip().split(',')
        name = parts[0]
        songs = parts[2:]

        yield name, len(songs)

    def combiner(self, name, song_counts):
        yield name, sum(song_counts)

    def reducer_init(self):
        self.max_count = 0
        self.max_users = []

    def reducer(self, name, song_counts):
        total_count = sum(song_counts)

        if total_count > self.max_count:
            self.max_count = total_count
            self.max_users = [name]
        elif total_count == self.max_count:
            self.max_users.append(name)

    def reducer_final(self):
        for user in self.max_users:
            yield user, self.max_count

if __name__ == '__main__':
    MaxSongsUpTillNow.run()

    Overwriting seventh.py

```

```

!python seventh.py textfile.txt

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/seventh.root.20230821.181355.462344
Running step 1 of 1...
job output is in /tmp/seventh.root.20230821.181355.462344/output
Streaming final output from /tmp/seventh.root.20230821.181355.462344/output...
"Ali" 5
"Fahad" 4
"Sara" 3
Removing temp directory /tmp/seventh.root.20230821.181355.462344...

```

#8 is same as task 6 which i have already implemented

✓

0s

completed at 11:13 PM

×