

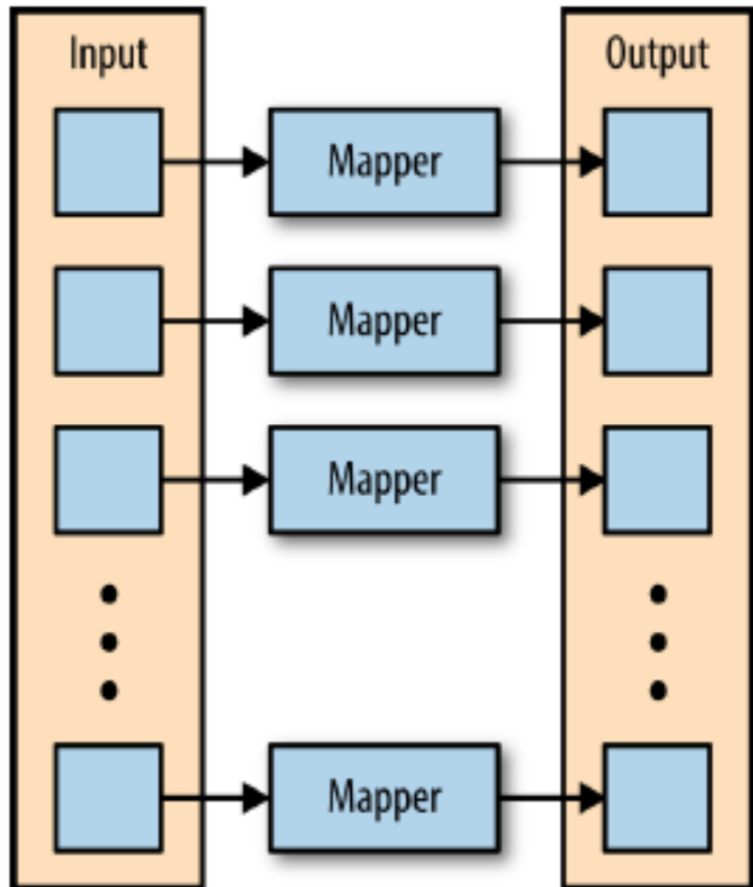
مثال‌های عملی Map-Reduce

خلاصه‌ای از Map-Reduce

► یک مدل برنامه‌نویسی است که پردازش موازی حجم زیادی از داده‌ها را برای ما راحت‌تر می‌سازد

► هر برنامه‌ی Map-Reduce لیستی از عناصر داده‌ی ورودی را به لیستی از عناصر خروجی تبدیل می‌کند

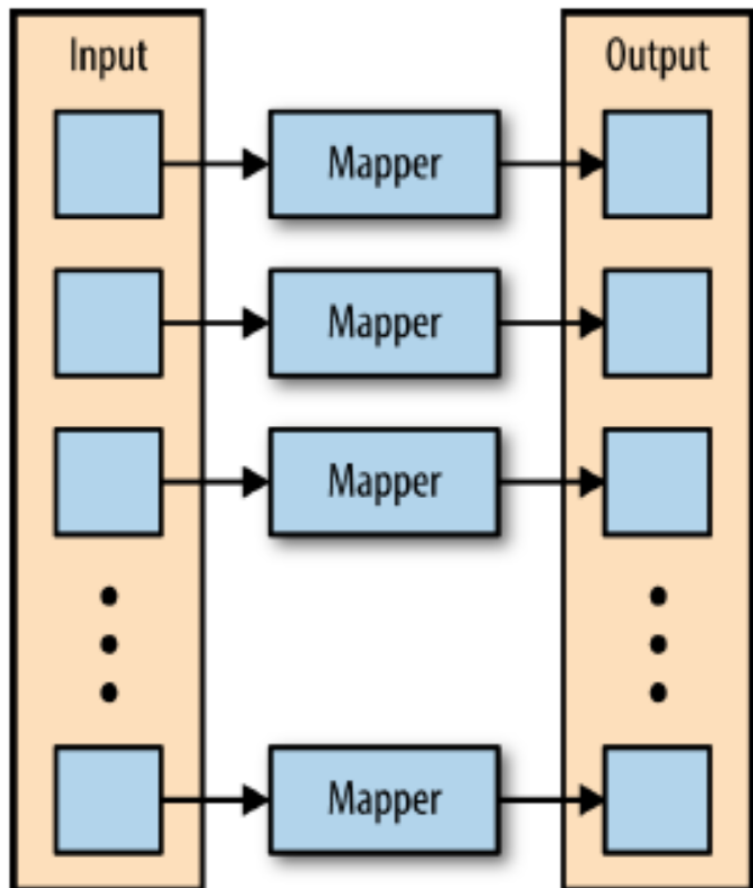
فاز Map



▶ اولین گام در برنامه‌نویسی با مدل Map-Reduce است

▶ تابع یا برنامه‌ای که در نقش Mapper هست مجموعه‌ای از داده را به عنوان ورودی می‌گیرد و معمولاً در خروجی مجموعه‌ای به صورت Key-Value تولید می‌کند

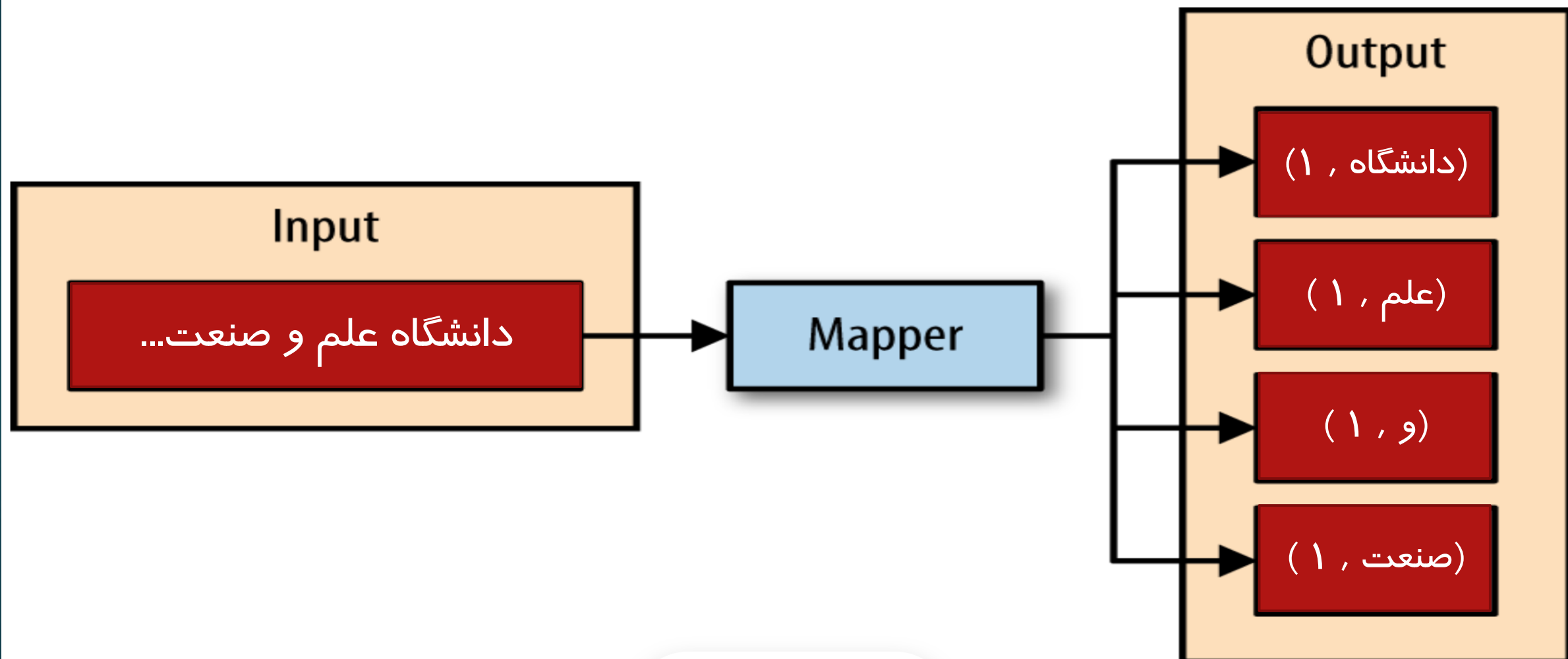
مثال اول - پیدا کردن تعداد لغات



► فرض کنید می خواهیم در
داکیومنتی که ۱ میلیون صفحه است
یا چند داکيومنت که مجموعاً ۱
میلیون صفحه هستند تعداد هر کلمه
را به دست بیاوریم (علم - صنعت -
کامپیوتر و ...)

► با توجه به تعداد Mapperها
تقسیم بندی می کنیم

فاز Map



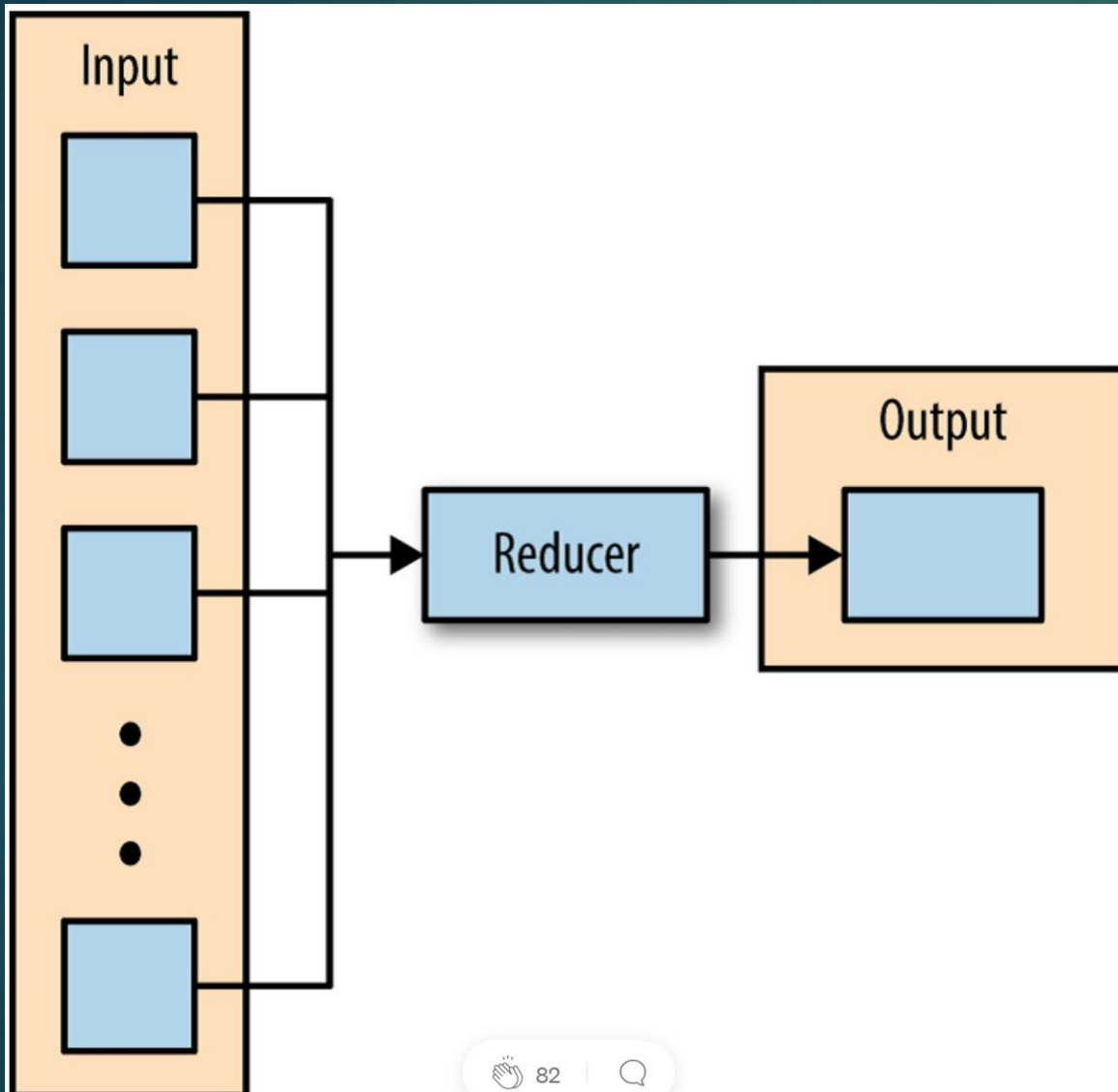
فاز Map مثال تعداد کلمات

```
1  #!/usr/bin/env python
2  import sys
3
4  # Read each line from stdin
5  for line in sys.stdin:
6
7      # Get the words in each line
8      words = line.split()
9
10     # Generate the count for each word
11     for word in words:
12         print '{0}\t{1}'.format(word, 1)
13
```

فاز Shuffle and sort

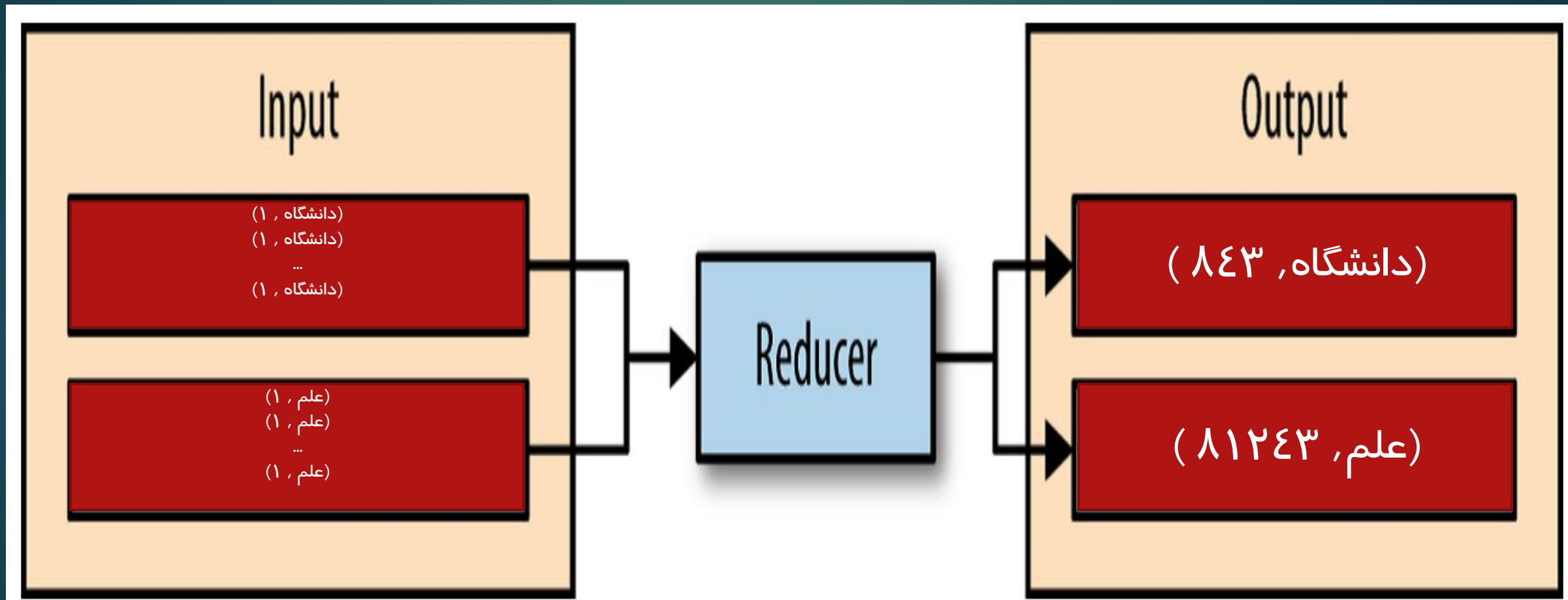
- ▶ دومین گام در برنامه‌نویسی با مدل Map-Reduce است
- ▶ خروجی‌های میانی از مرحله Map به Reducer ها منتقل می‌شوند (Shuffle)
- ▶ یک نوع مرتب‌سازی هم بر اساس کلیدها استفاده می‌شود

فاز Reduce



- ▶ آخرین گام در برنامه‌نویسی با مدل Map-Reduce است
- ▶ Reducer مقادیر هر کلید را جمع می‌کند و صفر یا چند جفت کلید-مقدار خروجی را به عنوان خروجی تولید می‌کند.

فاز Reduce



فاز Reduce مثال تعداد کلمات

```
#!/usr/bin/env python
import sys
curr_word = None
curr_count = 0

# Process each key-value pair from the mapper
for line in sys.stdin:

    # Get the key and value from the current line
    word, count = line.split('\t')

    # Convert the count to an int
    count = int(count)

    # If the current word is the same as the previous word,
    # increment its count, otherwise print the words count
    # to stdout

    if word == curr_word:
        curr_count += count
    else:
        # Write word and its number of occurrences as a key-value
        # pair to stdout
        if curr_word:
            print '{0}\t{1}'.format(curr_word, curr_count)
        curr_word = word
        curr_count = count

# Output the count for the last word
if curr_word == word:
    print '{0}\t{1}'.format(curr_word, curr_count)
```