

## 1. Hadoop Word Count

### Python (using Hadoop Streaming with Python Mapper & Reducer)

#### mapper.py

```
#!/usr/bin/env python3
import sys

for line in sys.stdin:
    words = line.strip().split()
    for word in words:
        print(f"{word}\t1")
```

#### reducer.py

```
#!/usr/bin/env python3
import sys

current_word = None
current_count = 0

for line in sys.stdin:
    word, count = line.strip().split('\t')
    count = int(count)

    if word == current_word:
        current_count += count
    else:
        if current_word:
            print(f"{current_word}\t{current_count}")
        current_word = word
        current_count = count

if current_word:
    print(f"{current_word}\t{current_count}")
```

#### To run in Hadoop:

```
hadoop jar /path/to/hadoop-streaming.jar \
-input /input/dir \
-output /output/dir \
-mapper mapper.py \
-reducer reducer.py
```