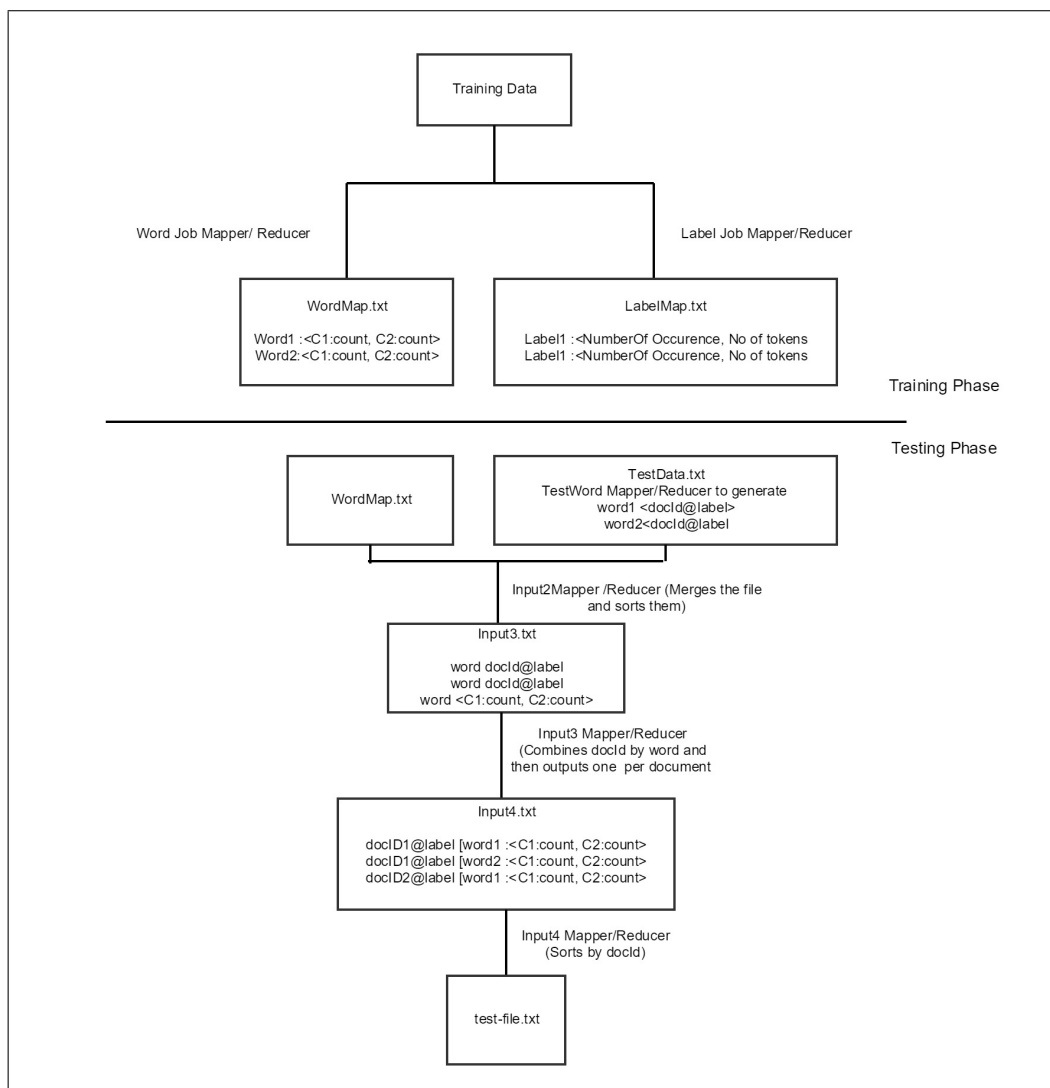


DS 222 Assignment-1

Abhinav Anand, SR - 13954

1 Algorithm Overview



2 Number of parameters(Large Dataset)

Number Of Classes - 50

Number of Distinct Words in Training - 423121

Total Parameters is calculated using the size of wordMap dictionary which contains word as key and classes in which it occurred in training data as value

Thus, Total Parameters = $1325407(\text{Size of wordMap dictionary}) + 50 = 1325457$

Local Implementation

2.1 Very Small

- Training Time(Train Data) : **7.48 s**
- Testing Time(Test Data) : **10.155 s**

	Training	Development	Test
Accuracy	0.8790	0.8952	0.8884

2.2 Small

- Training Time(Train Data) : **12.61 s**
- Testing Time(Test Data) : **16.06 s**

	Training	Development	Test
Accuracy	0.8470	0.8622	0.8610

2.3 Full

- Training Time : **78.21 s**
- Testing Time : **561.51 s**

	Training	Development	Test
Accuracy	0.8426	0.7991	0.7890

Hadoop Implementation

All observation are stated for full dataset

	Training	Development	Test
Accuracy	0.8417	0.7093	0.7837

2.4 Number Of Reducers = 2

- Training Time(2 Jobs) : **30 s**
- Testing Time(4 jobs) : **77 secs**

2.5 Number Of Reducers = 4

- Training Time(2 Jobs) : **21 s**
- Testing Time(4 Jobs) : **86 s**

2.6 Number Of Reducers = 6

- Training Time(2 Jobs) : **15 s**
- Testing Time(Test Data) : **85 s**

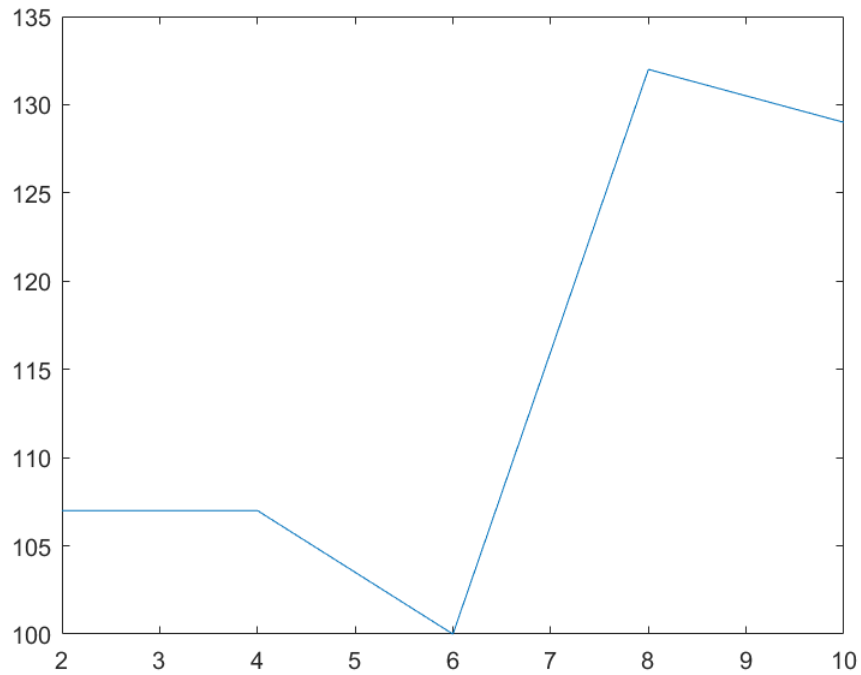
2.7 Number Of Reducers = 8

- Training Time(2 Jobs) : **24 s**
- Testing Time(Test Data) : **108 s**

2.8 Number Of Reducers = 10

- Training Time(2 Jobs) : **47 s**
- Testing Time(4 jobs) : **82 s**

3 Plot of Reducers vs Execution Time



Slight dip in the execution time is observed for 6 clusters

4 Job Log Files

4.1 10 Reducers

- Job 1
- Job 2
- Job 3
- Job 4
- Job 5
- Job6