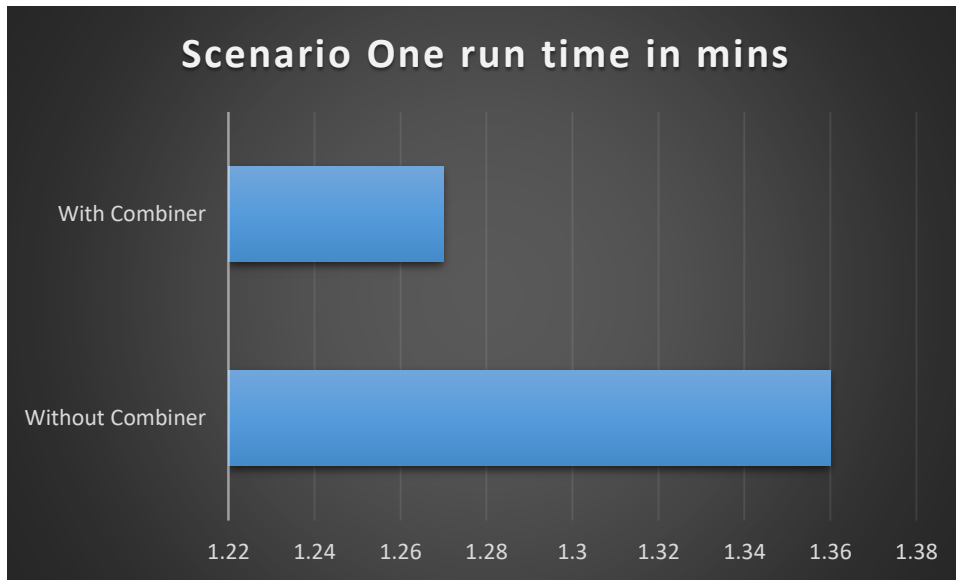


Pseudodistributed mode setup was done following the instructions in “Hadoop: the definitive guide”, Appendix-A. The vagrant box RAM size was set to 3072MB of RAM.

Scenario One Result(1990 dataset):

Scenario One	Run Time
Without Combiner	1.36
With Combiner	1.27

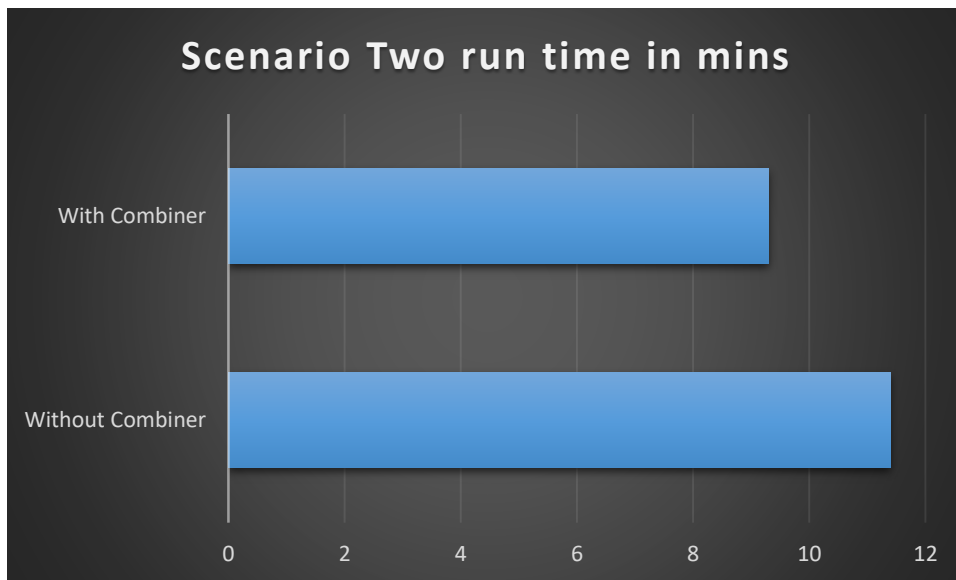


Hadoop with combiner ran for 1.27 minutes and Hadoop without combiner ran for 1.36 minutes.

It was more or less the same.

Scenario Two Result(1990 and 1992 datasets):

Scenario Two	Run Time
Without Combiner	11.4
With Combiner	9.3

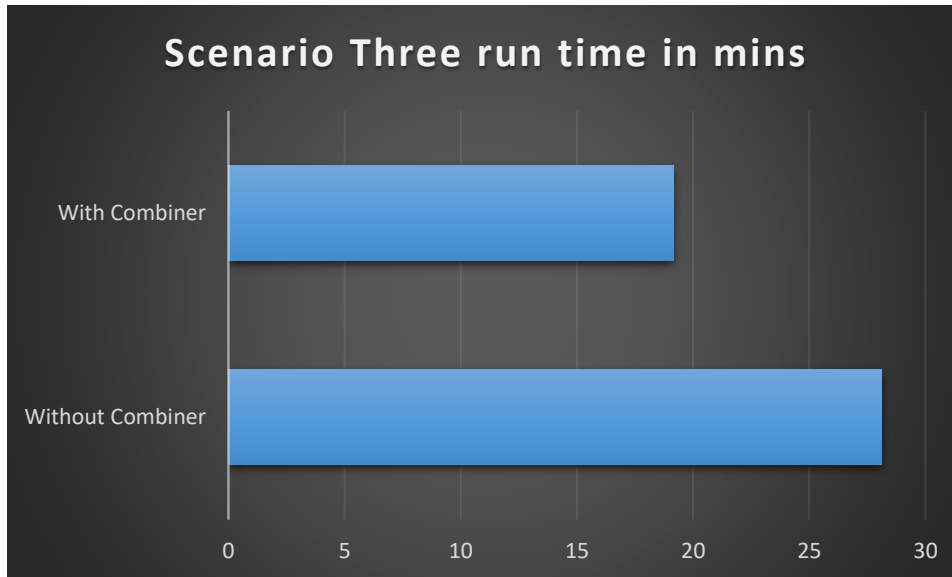


Hadoop without combiner for datasets 1990 and 1992 ran for 11.4 minutes and with combiner ran for 9.3 minutes.

It was also more or less the same.

Scenario Three Result(1990,1991,1992,1993 dataset):

Scenario Three	Run Time
Without Combiner	28.12
With Combiner	19.15

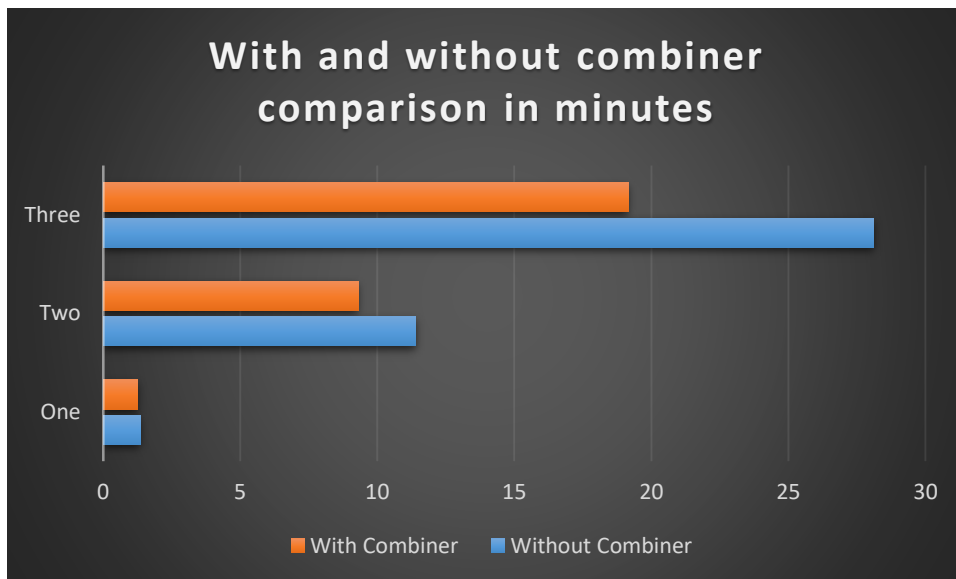


Hadoop with combiner for datasets 1990,1991,1992,1993 ran for 19.15 minutes and without combiner ran for 28.12 minutes.

The running time had reduced by 27%.

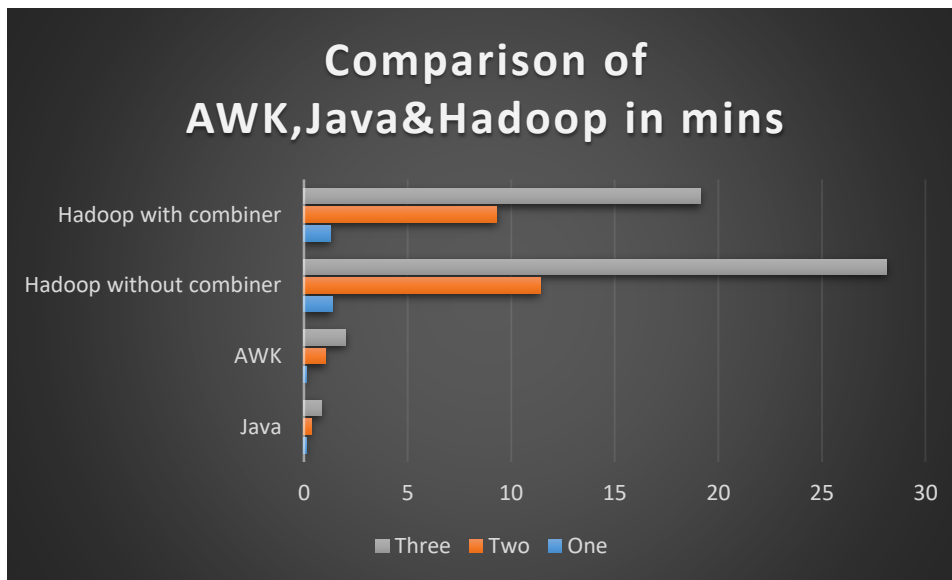
Comparing all the three scenarios:

Scenario	Without Combiner	With Combiner
One	1.36	1.27
Two	11.4	9.3
Three	28.12	19.15



Comparing the Hadoop results with AWK and Java results:

Scenario	Java	AWK	Hadoop without combiner	Hadoop with combiner
One	0.1102	0.114	1.36	1.27
Two	0.35	1.03	11.4	9.3
Three	0.83	2.008	28.12	19.15



It can be clearly observed that as data increases, the time consumed also increases in AWK, Java and Hadoop.

AWK takes much lesser time than Hadoop and Java.