Performance Evaluation of Spark program on Latin Docs

For 2-word co-occurrence

No of Documents vs Runtime
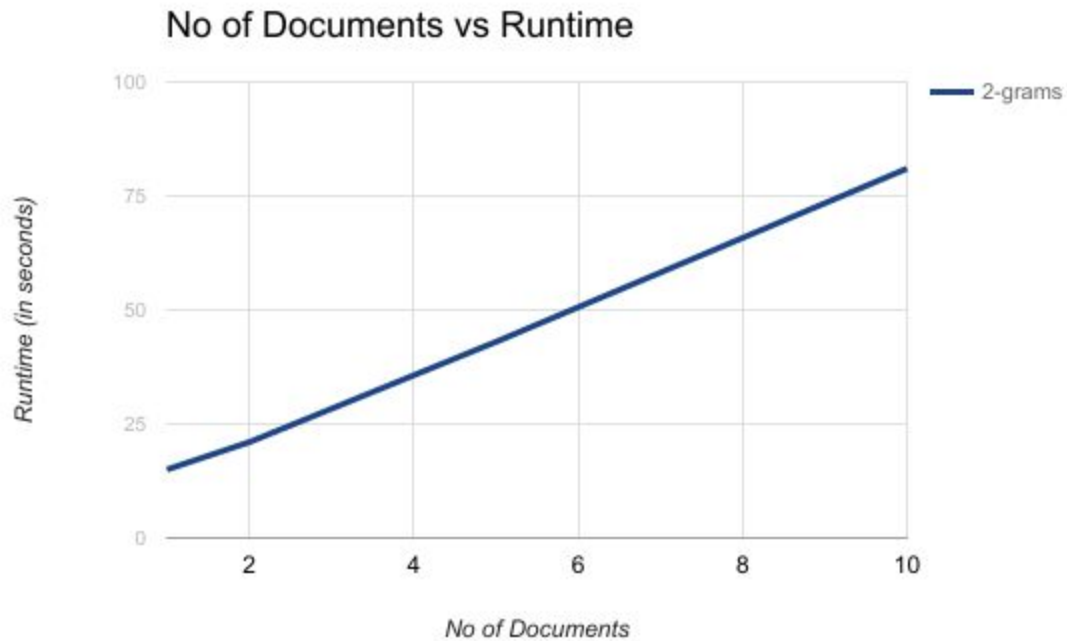
For 2-word co-occurrence there is an almost linear relationship between the number of documents and the running time. Even for 10 documents the execution time is only 81 seconds. From this we can say that our spark programs scalability on our system is quite good for 2-word co-occurrence.

For 3-word Co-occurrence

No of Documents vs Runtime
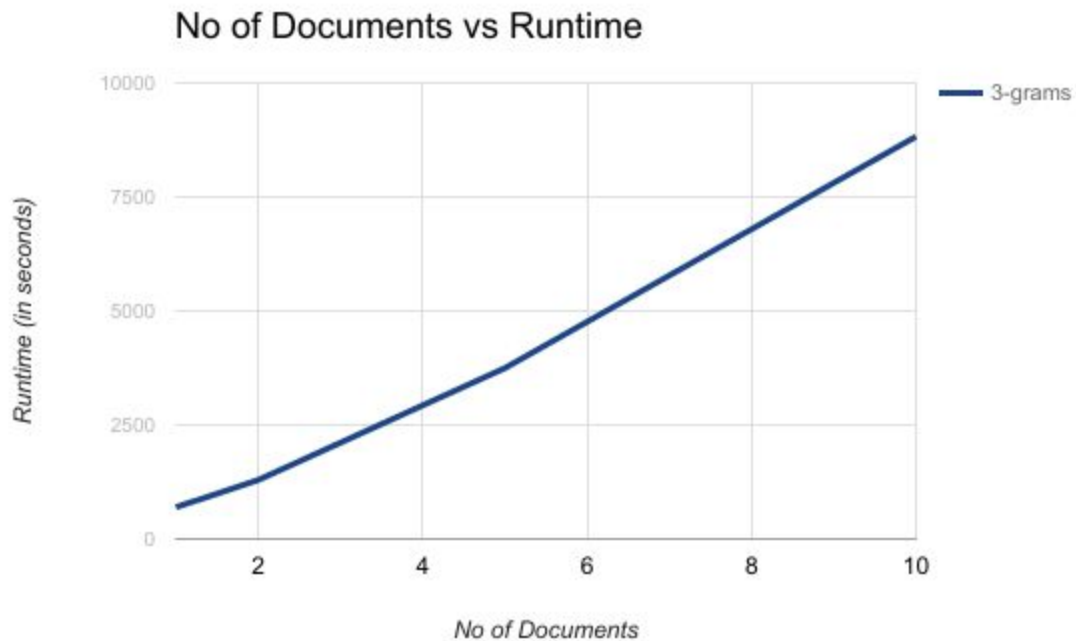
For 3-word co-occurrence the running time for the same number of documents as 2-word co-occurrence is exponentially greater. The increase in the running time with an increase in no of documents does not follow a linear relationship either. Due to limitations in computational resources we were unable to obtain a result for more than 6 documents. The results displayed for 7-10 documents have been extrapolated.