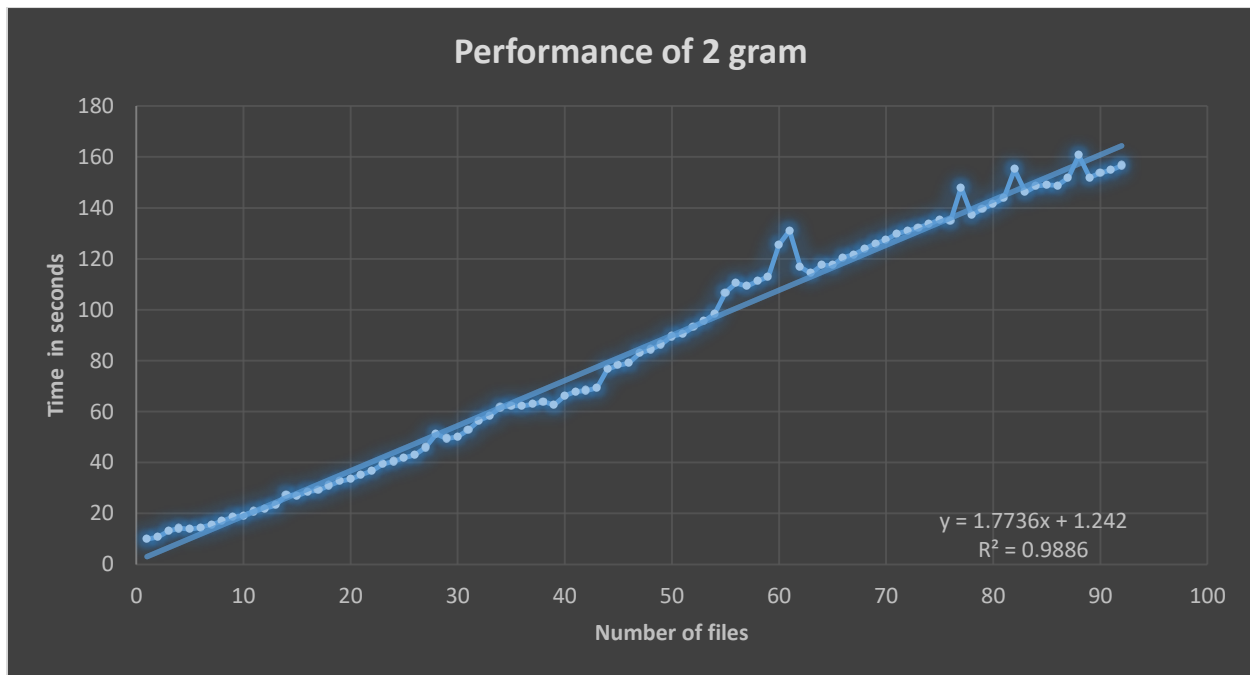


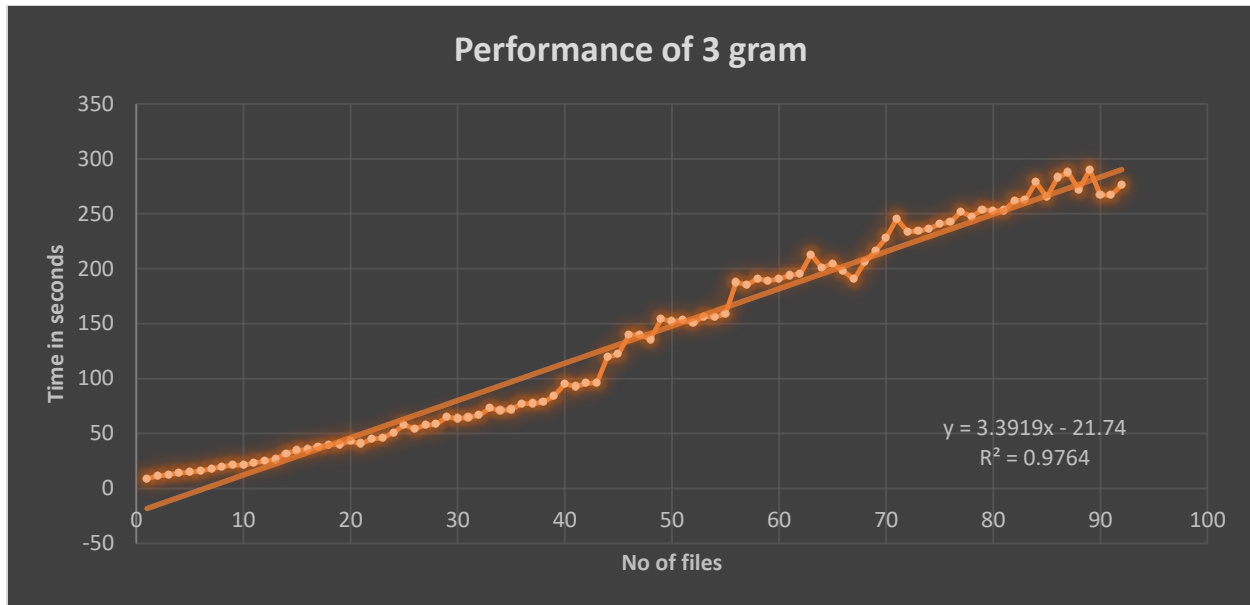
# Performance analysis of n-gram in Apache Spark

Participants: Ashwini Sarode(50205928) & Vivek Singh(50208473)

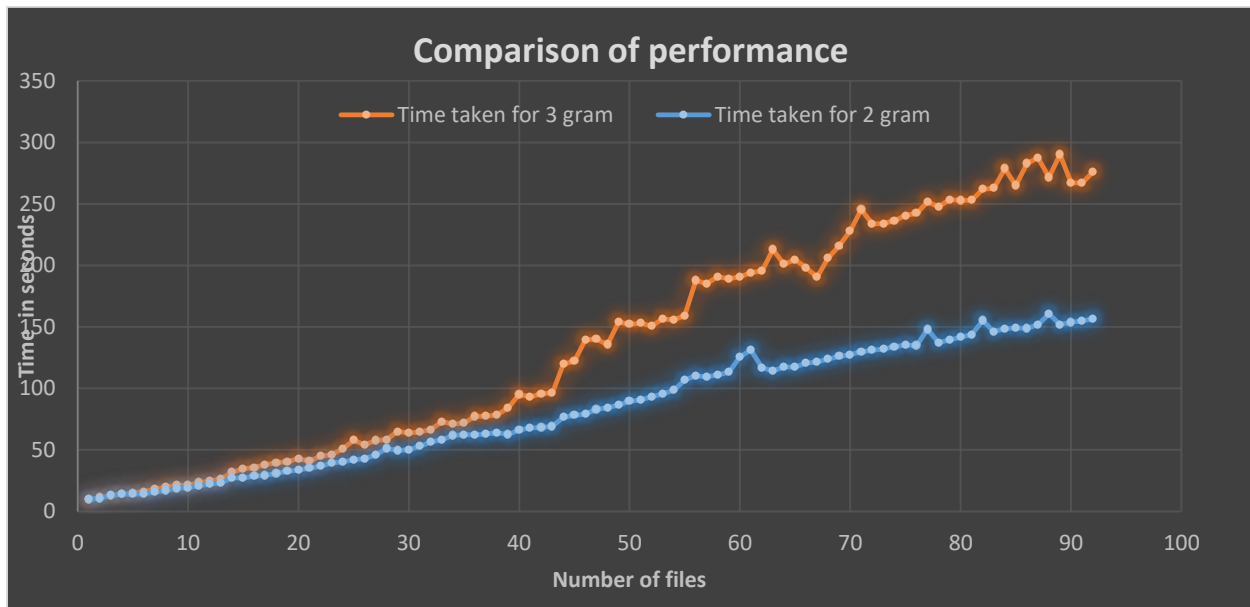
The following describes the performance of Spark while executing 2gram word co-occurrence.



The following describes the performance of Spark while executing 2gram word co-occurrence.



The following compares the performance of Spark while executing 2gram word co-occurrence with the performance of Spark while executing 3gram word co-occurrence.



Observations: From the above graph we can observe that:

- Time taken for n gram (both n=2 & n=3) increases linearly with increase in number of files.
- Time taken for computing 3 gram is more as compared to 2 gram.