HOMEWORK – 3

Data Intensive Computing

Nikhil Saji

50592329

# Task 1.1

Time taken for each set of sentences

* 10 Sentences:
  + Spark – 3.18 seconds
  + Multi-process – 0.432 seconds
  + For-loop – 0.066 seconds
* 50 Sentences:
  + Spark – 0.54 seconds
  + Multi-process – 1.466 seconds
  + For-loop – 3.402 seconds
* 100 Sentences:
  + Spark – 0.53 seconds
  + Multi-process – 3.438 seconds
  + For-loop – 14.735 seconds
* 250 Sentences:
  + Spark – 0.53 seconds
  + Multi-process – 18.900 seconds
  + For-loop – 88.130 seconds
* 500 Sentences:
  + Spark – 1.04 seconds
  + Multi-process – 77.052 seconds
  + For-loop – 344.748 seconds
* 1000 Sentences:
  + Spark – 3.58 seconds
  + Multi-process – 386.267 seconds
  + For-loop – 1564.122 seconds

# Task 1.2

Time taken for n\_inputs

* 1000 Inputs:
  + Spark – 0.49 seconds
  + Non-Spark – 0.18 seconds
* 5000 Inputs:
  + Spark – 0.35 seconds
  + Non-Spark – 0.96 seconds
* 10000 Inputs:
  + Spark – 0.31 seconds
  + Non-Spark – 2.8 seconds
* 50000 Inputs:
  + Spark – 0.38 seconds
  + Non-Spark – 350.29 seconds
* 100000 Inputs:
  + Spark – 0.4 seconds
  + Non-Spark – 847.21 seconds