

### 1. What is the question?

Get the line and line number having words like "loue", "course" and both the words in a same line from "Midsummer Night's Dream" text file and the line number should begin with the title of the play.

### 2. Why is it important?

It is important to analyze where and how words are used in Shakespeare's play "A Midsummer Night's Dream". To do so, it is necessary to get the line and line numbers of the play where specific words occur. Words used to analyze the Shakespeare's play are "loue", "course" and both the words in the same line.

### 3. What have others done to try and solve the question?

I referred a word count python example done by others in the following link <https://course.ie.cuhk.edu.hk/~engg4030/tutorial/tutorial3/> . Here they have used some options while running python code in hadoop. Particularly mapred.text.key.comparator.options helped me to sort the output according to the line number.

### 4. What will I do to solve the problem?

- a) Login to fast.cs.odu.edu
- b) Secure shell to the Hadoop server.
- c) Make sure these files are available mapper.py, reducer.py, mapredresult.py, makeFile.sh
- d) Run the makeFile.sh bash script. It will ask user to enter mapper file, reducer file. Further it will ask user to give directory name where output files need to be generated.
- e) Once the output file is generated. It will ask user to generate results.txt file.
- f) If user wants result.txt file then both output file (part-00000) and results.txt are generated in the given directory name or else only output file will be generated.

### 5. What will I do to prove that I have solved the problem?

I will generate the output file(part-00000) having line number and line with words "loue", "course" and both the words in the line and generate results.txt by segregating each of these condition.

### 6. What is the conclusion?

I was able to fetch the line number and line containing the words like "loue", "course" and both the words in the same line in the output file(part-00000) using mapper and reducer program in python. Further, I was able to generate results.txt in which these 3 conditions("loue", "course" and both the words in the same line) are segregated with their line and line number.