

DATA SCIENCE HW4

I have used Cloudera image of Hadoop in VMware and executed the programs in Hadoop. I used python programming.

Task 1 Problem A

Word that appears the most is "**the**". It appears **31950** times.

Task 1 Problem B

Palindrome that appears the most is "**a**". It appears **10150** times.

TASK 2 Problem A

Party 3 won the election with **12071** votes

TASK 2 Problem B

The County is **277**

TASK 2 Problem C

The Counties are **107 178 201 220 241 244 274 332 334 359 390 424 474**.

TASK 2 Problem D

6297 voters changed their votes. The most common change is from party 1-3 and 1564 times