

Title

Big data

Problem Description

Download the file “bigData.txt”. It will be our input file for this problem.

Warning: bigData.txt is too large to be opened by Notepad.

Computers are used for data analysis because the sheer amount of data that is generated is impossible to work with by hand. In this problem we are going to put our computer to work on a large dataset that we’ve created for you.

The data is organized in two columns separated by a space. You do not know how many rows the data has (hint: it’s a lot!). Open the file bigData.txt and compute the means of each column. Output the result to the screen.

Warning part 2: this actually takes quite a while to process. If you find your program doesn’t seem to be outputting anything for ten or twenty seconds, that’s OK. There’s a lot of data to work with in this file.

The first three lines of the file are shown below:

```
0.81472 5.4498
```

```
0.90579 14.083
```

```
0.12699 4.4615
```

Restrictions: This problem is solvable without the use of arrays. You are free to use arrays if you feel that’s the only way to solve it, **but you must declare the array to have a fixed size when you write the program.** You are not allowed to search the file, see how long it is, and then make a new array while the program is running.

Testing

The mean of the first column is approximately 0.5 and the mean of the second column is approximately 0. Your answer should be close to these values, with a margin of error of less than 0.01.

Time Target

- *** less than 10 minutes
- ** 10-15 minutes
- * greater than 15 minutes

Section

File IO