ORDERING LARGE AMOUNTS OF DATA

ASSIGNMENT OVERVIEW

This assignment is used to assess your skill level specific to the position you have applied for. The assignment is to order a file with a large amount of data.

The lines in the file use the following format:

id<tab>timestamp<tab>data\r\n

Where **id** is a number (in hexadecimal notation), **tab** is a regular tab (represented by **<tab>**), and **timestamp** is a Unix time stamp in milliseconds. **Data** is a string with different values for each entry. At the end of every line there is a carriage return followed by a new line (represented by **\r\n**).

The file to order has a size of around 2GB and can be generated using the included *CreateData.jar* file by executing the following command:

java -jar CreateData.jar

A file called *largefile.txt* is created in the directory from where you execute the command. In

case you receive problems generating this file, you can download it through the following link:

https://s3-eu-west-1.amazonaws.com/behaviosec-hiring-developer-tests/largefile.7z

ASSIGNMENT INSTRUCTIONS

We would like for you to create a small program that takes the *largefile.txt* file and generates a **new file** that is **ordered first** by **id** then by **date**.

For example:

Id1<tab>date1<tab>data\r\n
Id1<tab>date2<tab>data\r\n
Id2<tab>date1<tab>data\r\n
Id2<tab>date2<tab>data\r\n

… .

The name assigned to your new .txt file is your preference.

PROGRAM REQUIREMENTS

It is important that your program:

- sorts the lines in the new file by **id** and then by **date**.
- does not use more than 500 MB of RAM.
- does not use more than 5 GB of disk space (including input, output, and any temp file you might use).