Dream Agility – Coding Challenge

Purpose of this coding challenge is to measure the candidates coding proficiency, attention to detail and ability to diagnose performance issues.

Problem Scenario

Client has requested that you perform calculations on his data set.

Each data entry is made up of a number columns/attributes and each entry has a unique ID.

The data is provided in 3 files, each file contains information about products but each file only contains a subset of columns/attributes for each entry. The rest can be found in the other files.

The files need to be merged into a single file containing all columns before the processing can take place.

Example

File A

id	attr1	attr2
a	1u	a1
b	2j	a2
c	3k	a3
d	4m	a4
e	5n	a5

File B

id	attr6	attr7
d	1a	a1q
e	2a	a2w
a	3v	a3e
c	4c	a4r
b	5d	a5t

File C

id	attr11	attr12
c	1y	a11
a	2u	a22
e	3i	a33
d	4l	a44
b	5m	a55

Merged File

id	attr1	attr2	attr6	attr7	attr11	attr12
a	1u	a1	3v	a3e	2u	a22
b	2j	a2	5d	a5t	5m	a55
c	3k	a3	4c	a4r	1y	a11
d	4m	a4	1a	a1q	5m	a55
e	5n	a5	2a	a2w	3i	a33

Task

Write a console program to merge the files provided in a **timely** and **efficient** manner. File paths should be supplied as arguments, so that the program can be evaluated on different data sets.

The merged file should be saved as CSV; use the id column as the unique key for merging; the program should do any necessary data cleaning and error checking. We will evaluate the submission based on code, extensibility, performance, readability, correctness and integrity of the merged files (Large Files will be used), validity of the output file schema etc. Notes and observations must be submitted along with the files along with instructions and any dependencies which will be required to run the software (version of Python as an example is very important.) Any potential errors should be handled in the code.

Submission

Send the source code with the output files via email with any notes/observations (if any) on the program/problem.

©Dream Agility Ltd 2018