

Mapp Digital DMP coding exercise

At Mapp Digital DMP we analyze and process a lot of data on a daily basis. We ingest multiple gigabytes of data per minute. For demo and testing setups we can however not use data from our clients. As of such we need to generate valid, consistent and comparable data.

The Exercise

Write an application that generates test data based on your own configuration file. The configuration file should specify columns, types of data in those columns and anything else that is needed in order to produce data in a flexible way. The output format should be TSV with column headers, preferably compressed with an efficient algorithm.

It should also support taxonomy (relation in data between columns) in the data. For example:

Car Brand	Car Model
BMW	M3
BMW	1 series Coupe
Audi	A3

The M3 value could never exist under the brand Audi. This is a specific example involving cars. We, however, are no car dealers, so try to implement abstraction regarding the data.

Working vs Good

You can write a working piece of code, or you can write very good code. We value:

- modularity: easily replacing components of the code (e.g. we used to have TSV output, but can now also support Protocol Buffers)
- testability: make sure we can verify the behavior of the application
- balance: the right balance between performance, resource usage, lines of code, etc.

Programming language

We would like you to use Go. We think it is very suited for this process due to the way you can easily write concurrent tasks with very few lines of code. Helpful materials:

<https://golang.org/>

<https://gobyexample.com/>

Good luck and we look forward to your awesome piece of code!