


Data / Analytics Engineer Homework Exercise

 Thank you for applying to Payoneer Germany!

As part of your application, we would like to see you write some code.

We're conscious of your time - this exercise should be relatively straightforward. We're not looking for you to add embellishments or additional features, just to cover what is requested.

Things we want:

1. We want to check that you are comfortable writing code
2. We want to provide some indication about the role (we're looking for someone to work with data pipelines more than we are looking for data scientists or people to do AI/ML)
3. We want to have something to discuss and work with in later interviews

Things we don't want:

1. We don't want to impose too much on your time - this should be clearly scoped and straightforward, not hours of open-ended work
2. We don't want to assess you on things we haven't asked for - we value documentation and tests, but the documentation and tests you write for a simple proof of concept are different than the documentation and tests you write for critical production infrastructure

We run **NGINX** in front of our web services and it logs request data in a relatively structured format - "the NGINX combined format":

```
$remote_addr - $remote_user [$time_local] "$request" $status  
$body_bytes_sent "$http_referer" "$http_user_agent"
```

An example of this would be:

```
93.180.71.3 - - [17/May/2015:08:05:32 +0000] "GET /downloads  
/product_1 HTTP/1.1" 304 0 "-" "Debian APT-HTTP/1.3 (0.8.16  
~expl2ubuntu10.21)"
```

More examples are available at https://github.com/elastic/examples/blob/fc8edfb876bb92251d7b53e7a5fac68192fb9b9e/Common%20Data%20Formats/nginx_logs/nginx_logs

Could you please:

- Design a suitable data schema for storing log lines (e.g. in SQLite or PostgreSQL)
- Write a small service/script that, when run, will open and continually read log lines from a constantly-growing log file, and parse and store them in a database with the schema you designed
- Write a small script / tool that, when run, will query the database and print a count of all requests grouped by status code
- Provide some brief notes on how we can run and test your solution

When you've done that, please zip it all up and send it to us. Thank you!