

超级简历-Python数据分析实习岗笔试题说明

附件中的日志来自于某项目。都是纯文本内容，格式为：

[IP] -- [日期] "[request方式] [url] [http版本]" [返回码] [response长度] [referer] [user agent] [缓存是否命中]

例如：

58.253.69.253 -- [26/Apr/2019:23:59:09 +0800] "GET /companies.json?term=淘宝dian&lang=cn HTTP/1.1" 304 20 "<https://www.wondercv.com/zh-CN/cvs/sDMwxn8/editor>" "Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.25 Safari/537.36 Core/1.70.3650.400 QQBrowser/10.4.3341.400" --

对于上面这一行日志，各个部分的说明分别是：

IP: 58.253.69.253

日期: [26/Apr/2019:23:59:09 +0800]

request方式: GET

url: /companies.json?term=淘宝dian&lang=cn

http版本: HTTP/1.1

返回码: 304

response长度 20

referer(浏览器访问的前一个url): <https://www.wondercv.com/zh-CN/cvs/sDMwxn8/editor>

user agent: "Mozilla/5.0 (Windows NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/70.0.3538.25 Safari/537.36 Core/1.70.3650.400

QQBrowser/10.4.3341.400"

缓存是否命中: 无缓存 (如果命中了, 就是HIT, 未命中: MISS)

请分析附件中：

1. 合计多少次访问
2. GET, POST, DELETE各是多少次
3. 多少个独立IP。
4. 每个IP出现了多少次。

附加

5. 如果不考虑URL的参数 (也就是URL中?和后面的部分都不考虑)的话, 都出现过哪些URL, 每个URL出现多少次? 并把结果从大到小排序, 然后输出出来。 例如：

51 /companies.json
40 /example_entries
32 /universities.json

要求：

1. 使用你最擅长的语言，不用非用python
2. 源代码需要可运行。可以附加一个说明文档，包括语言的版本，运行环境，运行的命令，你的操作系统。