

README.md

Задание

MemcLoad

Задание: нужно переделать однопоточную версию memc_load.py в более производительный вариант. Сам скрипт парсит и заливает в мемкеш поминутную выгрузку логов трекера установленных приложений. Ключом является тип и идентификатор устройства через двоеточие, значением являет protobuf сообщение.

```
$ ls -lh /Users/s.stupnikov/Coding/otus/*.tsv.gz
-rw-r--r--  1 s.stupnikov  staff   506M 29 сен 12:09
/Users/s.stupnikov/Coding/otus/20170929000000.tsv.gz
-rw-r--r--  1 s.stupnikov  staff   506M 29 сен 12:17
/Users/s.stupnikov/Coding/otus/20170929000100.tsv.gz
-rw-r--r--  1 s.stupnikov  staff   506M 29 сен 12:25
/Users/s.stupnikov/Coding/otus/20170929000200.tsv.gz

$ gunzip -c /Users/s.stupnikov/Coding/otus/20170929000000.tsv.gz | head -3
idfa      e7e1a50c0ec2747ca56cd9e1558c0d7c      67.7835424444      -22.8044005471
7942,8519,4232,3032,4766,9283,5682,155,5779,2260,3624,1358,2432,1212,528,8182,9061,9628,2055,4821,3550,4964

idfa      f5ae5fe6122bb20d08ff2c2ec43fb4c4      -104.68583244      -51.24448376
4877,7862,7181,6071,2107,2826,2293,3103,9433,2794,4303,7500,5637,8935,6772,2481,1614,3946,7013,690,9474,165

gaid      3261cf44cbe6a00839c574336fdf49f6      137.790839567      56.8403675248
7462,1115,5205,6700,865,5317,4967,2104,7993,6357,2385,8639,2306,5712,5326,9929,7781,1402,8830,1978,6443,337
```

Ссылки на tsv.gz файлы:

- <https://cloud.mail.ru/public/2hZL/Ko9s8R9TA>
- <https://cloud.mail.ru/public/DzSX/oj8RxGX1A>
- <https://cloud.mail.ru/public/LoDo/SfsPEzoGc>

Пример запуска

```
$ python memc_load.py --pattern=/Users/s.stupnikov/Coding/otus/*.tsv.gz --dry
[2017.09.29 12:39:52] I Memc loader started with options: {'dry': True, 'log': None, 'pattern':
'/Users/s.stupnikov/Coding/otus/*.tsv.gz', 'idfa': '127.0.0.1:33013', 'dvid': '127.0.0.1:33016', 'test':
False, 'adid': '127.0.0.1:33015', 'gaid': '127.0.0.1:33014'}
[2017.09.29 12:39:52] I Processing /Users/s.stupnikov/Coding/otus/20170929000000.tsv.gz
[2017.09.29 12:39:52] D 127.0.0.1:33013 - idfa:e7e1a50c0ec2747ca56cd9e1558c0d7c -> apps: 7942 apps: 8519
apps: 4232 apps: 3032 apps: 4766 apps: 9283 apps: 5682 apps: 155 apps: 5779 apps: 2260 apps: 3624 apps:
1358 apps: 2432 apps: 1212 apps: 528 apps: 8182 apps: 9061 apps: 9628 apps: 2055 apps: 4821 apps: 3550
apps: 4964 apps: 6924 apps: 6737 apps: 3784 apps: 5428 apps: 6980 apps: 8137 apps: 2129 apps: 8751 apps:
3000 apps: 5495 apps: 5674 apps: 3023 apps: 818 apps: 2864 apps: 8250 apps: 768 apps: 6931 apps: 3493
apps: 3749 apps: 8053 apps: 8815 apps: 8448 apps: 8757 apps: 272 apps: 5951 apps: 2831 apps: 7186 apps:
157 apps: 1629 apps: 2021 apps: 3338 apps: 9020 apps: 6679 apps: 8679 apps: 1477 apps: 7488 apps: 3751
apps: 7399 apps: 8556 apps: 5500 apps: 5333 apps: 3873 apps: 7070 apps: 3018 apps: 2734 apps: 4273 apps:
3723 apps: 4528 apps: 4657 apps: 4014 lat: 67.7835424444 lon: -22.8044005471
[2017.09.29 12:39:52] D 127.0.0.1:33013 - idfa:f5ae5fe6122bb20d08ff2c2ec43fb4c4 -> apps: 4877 apps: 7862
apps: 7181 apps: 6071 apps: 2107 apps: 2826 apps: 2293 apps: 3103 apps: 9433 apps: 2794 apps: 4303 apps:
7500 apps: 5637 apps: 8935 apps: 6772 apps: 2481 apps: 1614 apps: 3946 apps: 7013 apps: 690 apps: 9474
apps: 1655 apps: 9718 apps: 4862 apps: 3367 apps: 3869 apps: 4255 apps: 9431 apps: 7333 apps: 5471 apps:
3267 apps: 7439 apps: 7202 apps: 7310 apps: 7875 apps: 1468 apps: 8146 apps: 9617 apps: 4336 apps: 8747
apps: 7815 lat: -104.68583244 lon: -51.24448376
[2017.09.29 12:39:52] D 127.0.0.1:33014 - gaid:3261cf44cbe6a00839c574336fdf49f6 -> apps: 7462 apps: 1115
apps: 5205 apps: 6700 apps: 865 apps: 5317 apps: 4967 apps: 2104 apps: 7993 apps: 6357 apps: 2385 apps:
```

```
8639 apps: 2306 apps: 5712 apps: 5326 apps: 9929 apps: 7781 apps: 1402 apps: 8830 apps: 1978 apps: 6443
apps: 3372 apps: 6379 apps: 5426 apps: 7847 apps: 8485 apps: 8983 apps: 1938 apps: 4809 apps: 2095 apps:
6887 apps: 2720 apps: 1074 apps: 1499 apps: 7165 apps: 4922 apps: 5969 apps: 6655 apps: 241 apps: 8738
apps: 1336 apps: 7334 apps: 465 apps: 4866 apps: 3402 apps: 634 apps: 9813 apps: 6343 apps: 2686 apps:
2214 apps: 7140 apps: 4818 apps: 92 apps: 1072 apps: 4784 apps: 2376 apps: 2086 apps: 5639 apps: 6314
apps: 9411 apps: 7888 apps: 8187 apps: 8766 apps: 7524 apps: 9002 apps: 3567 apps: 3563 apps: 360 apps:
8949 apps: 2780 apps: 8299 apps: 2460 apps: 1274 apps: 1474 apps: 3030 apps: 2221 apps: 3467 apps: 9094
apps: 9014 apps: 4475 apps: 4180 apps: 2205 apps: 481 apps: 7125 apps: 1237 lat: 137.790839567 lon:
56.8403675248
```

Важно обрабатывать файлы в хронологическом порядке. В данном случае под этим имеется в виду, что после обработки нужно переименовывать файл, префиксировав имя точкой, последовательно и хронологически. Заливать же можно параллельно.

Цель задания: реализовать типичную задачу конкурентной обработки данных, поэкспериментировать с различными подходами, получить опыт взаимодействия со стандартной библиотекой.

Критерии успеха: задание **обязательно**, критерием успеха является работающий согласно заданию код, для которого написаны минимальные тесты, проверено соответствие pep8, написана документация с примерами запуска (боевого и тестов), в README, например. Далее успешность определяется code review.

Deadline

Задание нужно сдать через неделю. То есть ДЗ, выданное в понедельник, нужно сдать до следующего занятия в понедельник. Код, отправленный на ревью в это время, рассматривается в первом приоритете. Нарушение дедлайна (пока) не карается, но может повлиять на ранжирование при выборе топа студентов при окончании курса, пытаться сдать ДЗ можно до конца курса. Но код, отправленный с опозданием, когда по плану предполагается работа над более актуальным ДЗ, будет рассматриваться в более низком приоритете без гарантий по высокой скорости проверки

Обратная связь

Студент коммитит все необходимое в свой github/gitlab репозиторий. Далее необходимо зайти в ЛК, найти занятие, ДЗ по которому выполнялось, нажать "Чат с преподавателем" и отправить ссылку. После этого ревью и общение на тему ДЗ будет происходить в рамках этого чата.