

Задача. Парсинг событий лога

имя входного файла: input.txt

имя выходного файла: output.txt

Сервис состоит из фронтенда и бекендов. Бекенды бывают нескольких типов. Для увеличения отказоустойчивости бекенды одного типа реплицированы и составляют группу реплик (далее ГР). Бекенды из одной ГР одинаковы, и для оформления выдачи фронтенду достаточно получить ответ хотя бы с одного бекенда из каждой ГР. Фронтенд может сделать несколько попыток обращения к бекендам ГР, пока не получит результат.

У фронтенда есть три фазы обработки запроса:

- опрос бекендов
- мерджинг результатов, полученных с бекендов
- отправка результатов пользователю

В случае, когда фаза опроса бекендов занимает слишком много времени, фронтенд может принудительно завершить опрос и перейти к фазе мерджинга результатов с неполным набором данных.

В процессе работы фронтенд пишет в лог файл события, возникающие при обработке запроса.

Вам нужно по данному логу событий посчитать:

1. 95-й перцентиль времени обработки запросов фронтендом
- 2 .количество запросов к фронтенду на которые фронтенд не смог собрать данные со всех ГР.

Формат входных данных:

Входной файл состоит из строчек в формате:

<время события> <идентификатор запроса на фронтенд> <тип события>
[<дополнительные параметры>]

Разделителем полей выступает символ табуляции.

Строки отсортированы по времени события - UNIX timestamp в микросекундах.

Идентификатор запроса это уникальное целое число. Отфильтровав лог по записям с выбранным идентификатором запроса можно узнать все об обработке соответствующего пользовательского запроса.

Типы событий:

StartRequest

Начало обработки запроса

BackendConnect

Установка tcp соединения с бекендом.

Дополнительные параметры: <номер ГР> <URL запроса на бекенд>

BackendRequest

Отправка запроса на бекенд.

Дополнительные параметры: <номер ГР>

BackendOk

Отметка успешного получения ответа с бекенда. Работа с ГР завершается.

Дополнительные параметры: <номер ГР>

BackendError

Ошибка работы с бекендом. Следует после **BackendConnect** или **BackendRequest**.

Дополнительные параметры: <номер ГР> <строка с текстом возникшей ошибки>

StartMerge

Означает конец фазы опроса бекендов и начало фазы мерджа результатов.

StartSendResult

Означает конец фазы мерджа и начало отправки результата пользователю.

FinishRequest

Конец обработки запроса.

Формат выходных данных:

Выведите в свободной форме ответ на поставленные вопросы.

Посмотреть, что такое 95-й перцентиль можно тут:

ru.wikipedia.org/wiki/Квантиль

Время ответа пользователю можно вычислить как разницу времени между

StartSendResult и FinishRequest.

Бэкенды, с которыми производилась работа можно узнать разобрав URL в событии BackendConnect.

Запросы, на которые ответили не все ГР не имеют событий BackendOK хотя бы для одной из своих ГР.

Мы ожидаем, что решение будет оформлено на Python 3.5+

Можно пользоваться любыми библиотеками с `ruv`, но не забудьте описать зависимости в `pip-requirements.txt`.

В общем случае событийный лог может быть достаточно большим.