Небольшое сопроводительное письмопо решению тестового задания **«Выполнить разбор файла почтового лога, залить данные в БД и организовать поиск по адресу получателя»**  
Кроме полученного задания с инструкцией, данными таблиц для хранения и примера файла почтового лога, я получил предложение выбрать самому варианты решения возникающих вопросов и обосновать их.  
  
На решение задачи я предположил затратить менее 7 дней, рассчитывая использовать для решения второй части задачи ( «фронтенд» - показ результата ) готовую наработку с не очень большими изменениями.

Самую большую проблему доставила загрузка лога в таблицы базы данных.  
В связи с тем, что мои знания языка Perl и MySQL за значительно время без употребления попортились, пришлось решать задачу исходя из оставшихся знаний и не всегда оптимальным вариантом.  
  
В таблицы базы данных были добавлены поля с целью обеспечения полноты выборки и, главное, закачки информации из лога в базу.  
судя по приложенной статистике многие записей не попали в базу по причине сходства основных полей и даже малых различий в дополнительной информации

|  |  |  |
| --- | --- | --- |
| в таблицу | всего | удачно |
| message | 1922 | 1562 |
| log | 8079 | 6795 |
| всего | 10000 | 8357 |

--

-- Структура таблицы `message`

--

CREATE TABLE `message` (

`created` timestamp NOT NULL,

`id` varchar(64) NOT NULL,

`int\_id` char(16) NOT NULL,

`flag` varchar(16) NOT NULL,

`address` varchar(64) CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci NOT NULL,

`str` varchar(512) CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci NOT NULL,

`message\_id\_pk` varchar(128) CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci NOT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

--

-- Индексы таблицы `message`

--

ALTER TABLE `message`

ADD PRIMARY KEY (`id`) USING BTREE,

ADD KEY `message\_created\_idx` (`created`) USING BTREE,

ADD KEY `message\_int\_id\_idx` (`int\_id`) USING BTREE;

--

-- Структура таблицы `log`

--

CREATE TABLE `log` (

`created` timestamp NOT NULL,

`int\_id` char(16) CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci NOT NULL,

`flag` varchar(16) CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci NOT NULL,

`str` varchar(256) CHARACTER SET utf8mb4 COLLATE utf8mb4\_0900\_ai\_ci NOT NULL,

`log\_id\_pk` varchar(128) NOT NULL,

`address` varchar(64) NOT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

--

-- Индексы таблицы `log`

--

ALTER TABLE `log`

ADD PRIMARY KEY (`log\_id\_pk`) USING BTREE,

ADD KEY `log\_address\_idx` (`address`);

Необходимо было решить организационные вопросы о том, как будет происходить загрузка данных из лога в таблицу.  
Это будет автоматически происходить при каждом запуске скрипта с предварительной очисткой ранее внесенных данных,   
по прямому указанию пользователя из определенного места или с «закачкой» лога с «фронтэнда»,

- автоматизированное определение изменения файла лога (его размера и временных меток) и загрузки данных только в случае их изменения. Для этого была создана таблица file\_attr c единственной строкой, в которую записываются размер файла и 3 временные метки.

--

-- Структура таблицы `file\_attr`

--

CREATE TABLE `file\_attr` (

`id` int NOT NULL,

`size` int NOT NULL,

`ctime` varchar(16) NOT NULL,

`mtime` varchar(16) NOT NULL,

`atime` varchar(16) NOT NULL

) ENGINE=InnoDB DEFAULT CHARSET=utf8mb4 COLLATE=utf8mb4\_0900\_ai\_ci;

--

-- Дамп данных таблицы `file\_attr`

--

INSERT INTO `file\_attr` (`id`, `size`, `ctime`, `mtime`, `atime`) VALUES

(0, 1, '2', '3', '4');

Для отладки скрипта использовались следующие запросы на очисmку mаблиц и "сброс" данных файла maillog

truncate table log;

truncate table message;

update file\_attr set size = 1, ctime = 2, mtime = 3, atime = 4 where id=0;

В целях уменьшения количества записей не попавших базу я пробовал собирать индекс из нескольких полей, но этот вариант не сработал ..

Результат работы скрипт возвращает для HTTP клиента в JSON формате.

В консоли можно только проверить ошибки и загрузить Базу Данных.

Для реализации выборки по почте не было задания.

Но результат загрузки БД и «не принятых» базой записей можно посмотреть, запустив скрипт из браузера  
/cgi-bin/test/mla/mla.pl

А также, сделать выборку по почтовому адресу или его части

/cgi-bin/test/mla/mla.pl?mail.ru

{

"name": "MLA BigData",

"work\_time": 0,

"it": {

"1": "есть запрос: mail.ru",

"SQL запрос": "( SELECT created, address, int\_id, flag, str FROM message WHERE address LIKE '%mail.ru%' ) UNION ( SELECT created, address, int\_id, flag, str FROM log WHERE address LIKE '%mail.ru%' ) ORDER BY int\_id, created ;",

"найдено всего: ": 652,

"length": 652

},

"result": {

"1": {

"flag": "==",

"int\_id": "1RvaMp-000FiI-S2",

"address": "hxyglqwidzclrii@bossmail.ru",

"N": 1,

"str": "hxyglqwidzclrii@bossmail.ru R=dnslookup T=remote\_smtp defer (60): Operation timed out",

"created": "2012-02-13 14:44:51"

},

"2": {

"created": "2012-02-13 15:01:56",

"str": "hxyglqwidzclrii@bossmail.ru R=dnslookup T=remote\_smtp defer (60): Operation timed out",

"N": 2,

"address": "hxyglqwidzclrii@bossmail.ru",

"int\_id": "1RvaMp-000FiI-S2",

"flag": "=="

},

………………………………………….

Скорее всего, я напрасно ограничил себя по времени выполнения задания, но некоторый негативный опыт с увеличение времени на работы уже приводил к полному отсутствию ответа.

Мне было бы очень интересно и полезно на будущее получить от Вас отзыв по реализации именно скрипта на Perl

Результат выложен по адресу

<https://github.com/fotobred/test_mla>

настройки подключения к базе данных и места размещения лога почты находятся в файле Mlam.pm  
настройки подключения к серверу в файле mla.js

С Уважением,

Трошин Алексей

PS попытка выложить результат для «живой» демонстрации к себе на сайт не удался по причине отсутствия грамотного технического сопровождения у провайдера и дешевизной тарифа не допускающего собственного участия. Вот еще одна польза от Вашего задания – стало понятно, что провайдера надо менять