

# Отчет по заданиям для стажировки

## Песоцкий Дмитрий Игоревич

### Задание 1.

#### Постановка задачи.

Выполнить разбор предлагаемого файла лога с заполнением таблиц БД: В таблицу message должны попасть только строки прибытия сообщения (с флагом <=). Поля таблицы должны содержать следующую информацию: created - timestamp строки лога id - значение поля id=xxxx из строки лога int\_id - внутренний id сообщения str - строка лога (без временной метки) В таблицу log записываются все остальные строки: created - timestamp строки лога int\_id - внутренний id сообщения str - строка лога (без временной метки) address - адрес получателя.

#### Выполнение.

Для выполнения разбора файла лога и внесения полученной информации в таблицы из базы данных был выбран язык программирования Python 3. База данных была создана при помощи Postgresql 12-й версии. В связи с недостатком опыта в разработке проектов данного типа проект содержит только файлы с кодом программы, которые для работы сперва необходимо скомпилировать в среде разработки, поддерживающей Python 3. Для изменений каких-либо параметров необходимо редактировать код в файлах.

Предполагается, что база данных уже создана в Postgresql 12 при помощи кода, предоставленного в условии задачи, поэтому в коде программы по фильтрации файла лога не предусмотрено повторное создание базы данных и таблиц для нее.

В Postgresql 12 мною была создана база данных под названием "forlogs" с паролем "12345". Для смены базы данных под необходимую достаточно изменить параметры по подключению (строчки 3-9 файла Tasl 1.py), предоставленные на следующем рисунке:

```
con = psycopg2.connect(
    user="postgres",
    dbname = "forlogs",
    password="12345",
    host="127.0.0.1",
    port="5432"
)
```

Перед заполнением происходит полная очистка таблиц, если они уже были ранее созданы и чем-то заполнены, Затем происходит фильтрация данных при помощи регулярных выражений и заполнение таблиц согласно условию задачи.

Таблица message представлена на следующем рисунке:

created	id	int_id	str	status
2012-02-13 14:39:22	1	1Rwt3a-0009R1-2d	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1289 id=128213143629.COM_FM_END.206359@whois.somehost.ru	
2012-02-13 14:39:22	2	1Rwt3a-000AF3-3B	<> R=1RookS-000Pg8-VO U=mailnull P=local S=3968	
2012-02-13 14:39:22	3	1Rwt3a-0009R1-5L	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1268 id=128213143629.COM_FM_END.126492@whois.somehost.ru	
2012-02-13 14:39:22	4	1Rwt3a-000AG5-7A	<> R=1Rm0kE-00027I-1Y U=mailnull P=local S=2551	
2012-02-13 14:39:22	5	1Rwt3a-0009R1-7W	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=2229 id=128213143628.DOMAIN_FEEDBACK_MAIL.903141@whois.somehost.ru	
2012-02-13 14:39:22	6	1Rwt3a-0009R1-9n	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1266 id=128213143628.COM_FM_END.613957@whois.somehost.ru	
2012-02-13 14:39:22	7	1Rwt3a-0009R1-D4	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1258 id=128213143628.COM_FM_END.883547@whois.somehost.ru	
2012-02-13 14:39:22	8	1Rwt3a-0009R1-F4	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1278 id=128213143628.COM_FM_END.788785@whois.somehost.ru	
2012-02-13 14:39:22	9	1Rwt3a-0009R1-Oy	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1228 id=128213143628.COM_FM_END.691493@whois.somehost.ru	
2012-02-13 14:39:22	10	1Rwt3a-0009R1-ic	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1266 id=128213143628.COM_FM_END.613957@whois.somehost.ru	
2012-02-13 14:39:22	11	1Rwt3a-0009R1-KL	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1716 id=128213143628.BLOCKED.453962@whois.somehost.ru	
2012-02-13 14:39:22	12	1Rwt3a-0009R1-MA	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1592 id=128213143618.NIC-LTR-00100.967467@whois.somehost.ru	
2012-02-13 14:39:22	13	1Rwt3a-0009R1-OW	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1746 id=128213143618.BLOCKED.894434@whois.somehost.ru	
2012-02-13 14:39:22	14	1Rwt3a-0009R1-Qb	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1657 id=128213143617.AUCTION_RENEW.965150@whois.somehost.ru	
2012-02-13 14:39:22	15	1Rwt3a-0009R1-TL	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1658 id=128213143617.AUCTION_RENEW.146144@whois.somehost.ru	
2012-02-13 14:39:23	16	1Rwt3a-0009R1-VU	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1783 id=128213143616.AUCTION_RENEW.981122@whois.somehost.ru	
2012-02-13 14:39:23	17	1Rwt3b-0009R1-1j	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1783 id=128213143616.AUCTION_RENEW.843481@whois.somehost.ru	
2012-02-13 14:39:23	18	1Rwt3b-0009R1-3I	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1673 id=128213143616.AUCTION_RENEW.787528@whois.somehost.ru	
2012-02-13 14:39:23	19	1Rwt3b-0009R1-5o	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1673 id=128213143616.AUCTION_RENEW.789898@whois.somehost.ru	
2012-02-13 14:39:23	20	1Rwt3b-0009R1-7J	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=4818 id=128213143615.OI_STAT_CHNG_HOSTING_UPDATED.956003@whois.somehost.ru	
2012-02-13 14:39:23	21	1Rwt3b-0009R1-9N	tpxmuwr@somehost.ru H-mail.somehost.com 84.154.134.45] Psesmt S=1655 id=128213143615.AUCTION_RENEW.609165@whois.somehost.ru	
2012-02-13 14:39:23	22	1Rwt3b-000Ab2-9u	<> R=1RwtEg-0002zL-Mn U=mailnull P=local S=2383	

Всего в таблице message существует 1920 записей:

```
2012-02-13 15:09:41 1913 | 1RwtmV-000PdP-QV | <> R=1Rq0Bz-000Hd1-Sd U=mailnull P=local S=2541 |
2012-02-13 15:09:41 1914 | 1RwtmV-000MsX-Q1 | tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=smtp S=1711 id=128213150341.ANKETA_UPD.645427@whois.somehost.ru |
2012-02-13 15:09:42 1915 | 1RwtmV-000MsX-UD | tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=smtp S=1832 id=128213150385.O1_STAT_CHNG_SERVICE_OBJECT_UP.350119@whois.somehost.ru |
2012-02-13 15:09:42 1916 | 1RwtmV-000MsX-Qk | tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=smtp S=5664 id=128213150201.O1_STAT_CHNG_HOSTING_UPDATED.406721@whois.somehost.ru |
2012-02-13 15:09:55 1917 | 1RwtmV-0000Yc-3R | <> R=1RwtmV-000MsX-Jb U=mailnull P=local S=3926 |
2012-02-13 15:10:04 1918 | 1RwtmV-00057Q-Cs | <> R=1RwtmV-000MsX-Qk U=mailnull P=local S=6624 |
2012-02-13 15:10:15 1919 | 1RwtmV-00087I-0D | <> R=1RwtmV-000MsX-E0 U=mailnull P=local S=2917 |
2012-02-13 15:11:05 1920 | 1RwtmV-0004td-5w | <> R=1RwtmV-000MsX-Bg U=mailnull P=local S=2498 |
(1920 rows)
```

Таблица log представлена на следующем рисунке:

```
created | int_id | str | address
-----|-----|-----|-----
2012-02-13 14:39:22 | 1RwtJA-000AFB-07 | => :blackhole: <tpxmuwr@somehost.ru> R=blackhole_router |
2012-02-13 14:39:22 | 1RwtJA-000AFB-07 | Completed | tpxmuwr@somehost.ru
2012-02-13 14:39:22 | 1RookS-000Pg8-VO | == udbswscdnbegmloghuf@london.com R=dnsllookup T=remote_smtp defer (--44): SMTP error from remote mail server after RCPT TO:<udbswscdnbegmloghuf@london.com>; host mx0.gmx.com 74.208.5.90]: 450 4.3.2 Too many mails (mail bomb), try again in 1 hour(s) 25 minute(s) and see ( http://portal.gmx.net/serverrules ) {mx-us011} | udbswscdnbegmloghuf@london.com
2012-02-13 14:39:22 | 1RookS-000Pg8-VO | ** fwvparobkymbyemevz@london.com: retry timeout exceeded | fwvparobkymbyemevz@london.com
2012-02-13 14:39:22 | 1Qc0oJ-000DxQ-75 | Completed |
2012-02-13 14:39:22 | 1RwtJA-000AFJ-3B | => :blackhole: <tpxmuwr@somehost.ru> R=blackhole_router |
2012-02-13 14:39:22 | 1RwtJA-000AFJ-3B | Completed | tpxmuwr@somehost.ru
2012-02-13 14:39:22 | 1Rm0KE-00027I-IY | == udbswscdnbegmloghuf@london.com R=dnsllookup T=remote_smtp defer (--44): SMTP error from remote mail server after RCPT TO:<udbswscdnbegmloghuf@london.com>; host mx0.gmx.com 74.208.5.90]: 450 4.3.2 Too many mails (mail bomb), try again in 1 hour(s) 24 minute(s) and see ( http://portal.gmx.net/serverrules ) {mx-us014} | udbswscdnbegmloghuf@london.com
2012-02-13 14:39:22 | 1Rm0KE-00027I-IY | ** fwvparobkymbyemevz@london.com: retry timeout exceeded | fwvparobkymbyemevz@london.com
2012-02-13 14:39:22 | 1RookS-000Pg8-VO | Completed |
2012-02-13 14:39:22 | 1RwtJA-000AGs-7A | => :blackhole: <tpxmuwr@somehost.ru> R=blackhole_router |
2012-02-13 14:39:22 | 1RwtJA-000AGs-7A | Completed | tpxmuwr@somehost.ru
```

Всего в таблице log существует 6971 запись:

```
2012-02-13 15:11:45 | 1RwtmW-000MsX-Aw | -> qefpvgewrxy@gmail.com R=dnsllookup T=remote_smtp H=gmail-smtp-in.l.google.com 209.85.137.26] X=TLSv1:RC4-SHA:128 C="250 2.0.0 OK 1329131505 f2si4017581bl.67" | fwjatugliro@gmail.com
2012-02-13 15:11:45 | 1RwtmW-000MsX-Aw | Completed | qefpvgewrxy@gmail.com
2012-02-13 15:11:46 | 1RwtmW-000MsX-99 | => ggivvlar@gmail.com R=dnsllookup T=remote_smtp H=gmail-smtp-in.l.google.com 209.85.137.26] X=TLSv1:RC4-SHA:128 C="250 2.0.0 OK 1329131506 p7si52985061bz.61" | ggivvlar@gmail.com
2012-02-13 15:11:46 | 1RwtmW-000MsX-99 | -> fwjatugliro@gmail.com R=dnsllookup T=remote_smtp H=gmail-smtp-in.l.google.com 209.85.137.26] X=TLSv1:RC4-SHA:128 C="250 2.0.0 OK 1329131506 p7si52985061bz.61" | fwjatugliro@gmail.com
2012-02-13 15:11:46 | 1RwtmW-000MsX-99 | -> qefpvgewrxy@gmail.com R=dnsllookup T=remote_smtp H=gmail-smtp-in.l.google.com 209.85.137.26] X=TLSv1:RC4-SHA:128 C="250 2.0.0 OK 1329131506 p7si52985061bz.61" | qefpvgewrxy@gmail.com
2012-02-13 15:11:46 | 1RwtmW-000MsX-99 | Completed | qefpvgewrxy@gmail.com
2012-02-13 15:11:47 | 1RwtmW-000MsX-Lj | => vcncwzgdqxn@gmail.com R=dnsllookup T=remote_smtp H=gmail-smtp-in.l.google.com 209.85.137.26] X=TLSv1:RC4-SHA:128 C="250 2.0.0 OK 1329131507 ic2si45267451ab.70" | vcncwzgdqxn@gmail.com
(6971 rows)
```

Таким образом, программа фильтрует данные из файла лога out.txt и помещает их в 2 таблицы согласно условию задачи, но для смены данных о базе данных необходимо менять сам код программы, а также для работы с программой необходимо каждый раз компилировать файл Task 1.py и иметь созданную базу данных с таблицами, соответствующими предоставленному коду из условия задачи: «CREATE TABLE message ( created TIMESTAMP(0) WITHOUT TIME ZONE NOT NULL, id VARCHAR NOT NULL, int\_id CHAR(16) NOT NULL, str VARCHAR NOT NULL, status BOOL, CONSTRAINT message\_id\_pk PRIMARY KEY(id) ); CREATE INDEX message\_created\_idx ON message (created); CREATE INDEX message\_int\_id\_idx ON message (int\_id); CREATE TABLE log ( created TIMESTAMP(0) WITHOUT TIME ZONE NOT NULL, int\_id CHAR(16) NOT NULL, str VARCHAR, address VARCHAR ); CREATE INDEX log\_address\_idx ON log USING hash (address);».

## Задание 2.

### Постановка задачи.

Создать html-страницу с поисковой формой, содержащей одно поле (type="text") для ввода адреса получателя.

Результатом отправки формы должен являться список найденных записей ' ' из двух таблиц, отсортированных по идентификаторам сообщений (int\_id) и времени их появления в логе.

Отображаемый результат необходимо ограничить сотней записей, если количество найденных строк превышает указанный лимит, должно выдаваться соответствующее сообщение.

## Выполнение.

Для выполнения задачи был выбран язык программирования PHP версии 7.4.12. Для вывода html-страницы, использующей элементы программирования на языке PHP необходимо наличие локального сервера, поэтому мною также была установлена программа MAMP для создания локального сервера.

Интерфейс программы MAMP:



Для запуска сервера достаточно нажать на кнопку Start в правой верхней части окна приложения и в любом браузере перейти по адресу localhost:8888.

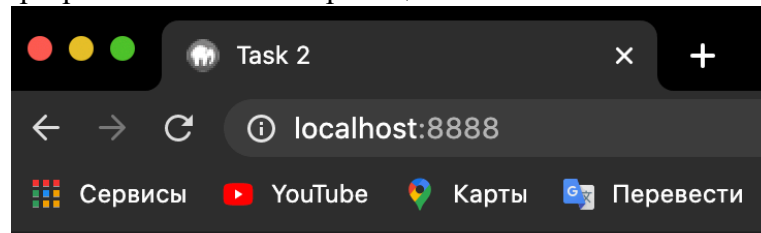
Для работы с PostgreSQL 12 в языке PHP предусмотрены специальные расширения, которые необходимо подключить в файле инициализации локального сервера. Файл инициализации php.ini прикреплен мною в директории Task 2 на удаленном репозитории с проектом. Подключенные расширения (строки 892-893 файла php.ini) представлены на следующем рисунке:

```
extension=pgsql.so
extension=pdo_pgsql.so
```

В связи с недостатком опыта в разработке проектов данного типа проект содержит только файлы с кодом (файл index.php в директории Task 2/htdocs) и файл инициализации (php.ini). Для корректной работы необходимо запустить локальный сервер, предварительно подключив к нему файл index.php из проекта, затем перейти в браузере по адресу этого локального сервера и пользоваться функциями, предоставленными на странице.

Для выполнения поиска данных из таблиц, заполненных в прошлом задании, пользователю необходимо ввести желаемый адрес электронной почты и нажать на кнопку “Показать результат”.

Интерфейс разработанной html-страницы:



**Введите e-mail:**

Показать результат

После ввода желаемого адреса электронной почты в соответствующее поле и нажатия кнопки “Показать результат” на странице будут отображены данные из таблиц log и message, относящиеся к введенному адресу электронной почты. При наличии более 100 упоминаний в одной из двух таблиц на экран также будет выведено сообщение о том, что количество записей превышает 100, из таблицы будут взяты данные только 100 строк. Также при помощи запроса "SELECT <имя столбцов> FROM <имя таблицы> WHERE <имя столбца, содержащего адрес электронной почты> = <введенный адрес электронной почты> ORDER BY <имя столбца с внутренним id сообщения>" (25-я и 54-я строки файла index.php) данные выводятся отсортированными по значениям из столбца с внутренним id сообщения.

Результат для адреса [tpxmuwr@somehost.ru](mailto:tpxmuwr@somehost.ru) из таблицы log:

**Введите e-mail:**

Показать результат

По адресу tpxmuwr@somehost.ru найдено следующее:  
Совпадения из таблицы log: 361 запрос превышает 100 строк  
2012-02-13 14:39:22 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:22 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:22 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:23 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:30 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:30 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:30 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:31 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:31 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:32 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:32 => :blackhole: R=blackhole\_router  
2012-02-13 14:39:32 => :blackhole: R=blackhole\_router

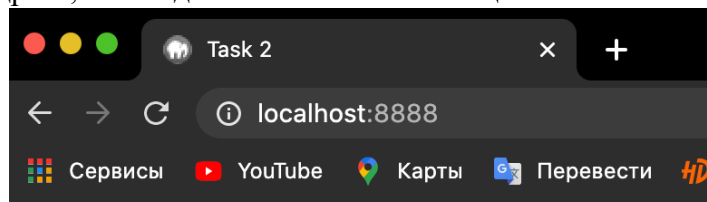
Данные выводятся из обеих таблиц последовательно, в отчете для экономии места привожу только части того, что видит пользователь в качестве результата.

## Результат для адреса [tpxmuwr@somehost.ru](mailto:tpxmuwr@somehost.ru) из таблицы message:

Совпадения из таблицы message: 1527 запрос превышает 100 строк

```
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1289
14:39:22 id=120213143629.COM_FM_END.205359@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1260
14:39:22 id=120213143629.COM_FM_END.125492@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=2229
14:39:22 id=120213143628.DOMAIN_FEEDBACK_MAIL.503141@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1268
14:39:22 id=120213143628.COM_FM_END.961254@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1250
14:39:22 id=120213143628.COM_FM_END.883547@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1270
14:39:22 id=120213143628.COM_FM_END.788705@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1220
14:39:22 id=120213143628.COM_FM_END.691493@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1266
14:39:22 id=120213143628.COM_FM_END.613957@whois.somehost.ru
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1716 id=120213143628.BLOCKED.453962@whois.somehost.ru
14:39:22
2012-02-13 tpxmuwr@somehost.ru H=mail.somehost.com 84.154.134.45] P=esmtп S=1592 id=120213143618.NIC-LTR-
14:39:22 00100.967467@whois.somehost.ru
```

## Результат для адреса, не найденного в обеих таблицах:



Введите e-mail:

Показать результат

По адресу [dima\\_pesotskiy@mail.ru](mailto:dima_pesotskiy@mail.ru) найдено следующее:

Совпадения из таблицы log: совпадений не найдено

Совпадения из таблицы message: совпадений не найдено

Таким образом, для работы проекта необходимы база данных и таблицы, созданные в задании 1. При желании можно изменить данные по подключению к базе данных. Для этого необходимо изменить строку 20 файла `index.php`, которая включает в себя название базы данных, пароль к ней, порт и тд аналогично пункту по подключению к базе данных из задания 1. Также для работы необходим локальный сервер, который в моем случае реализован при помощи программы MAMP. В случае с этой программой директория `htdocs` должна находиться там же, где и сама программа MAMP. Файл инициализации `php.ini` должен находиться по следующему пути: `"MAMP/bin/php/php7.4.12/conf"`. После запуска сервера следует перейти в браузере по адресу `localhost:8888` и начать пользоваться функциями появившейся html-страницы.