

Хранение метрик потоков в БД Timescale

Timescale - БД временных рядов с открытым исходным кодом, основанная на PostgreSQL. В отличие от InfluxDB, не хранит все индексы в оперативной памяти, поэтому может быть использована для хранения больших объемов метрик. Кроме того, в этом случае используется одна и та же БД PostgreSQL для хранения метрик, статистики потоков и настроек сбора данных.

Установка PostgreSQL и TimescaleDB

В процессе настройки должна быть установлена определенная версия PostgreSQL (9.6, 10 или 11). Если на сервере установлена более ранняя версия PostgreSQL (например, из стандартных репозиториях Centos 7 устанавливается PostgreSQL 9.2), необходимо ее удалить, включая исполняемые файлы в `/usr/bin`.

Для того, чтобы настроить PostgreSQL 9.6+TimescaleDB на CentOS, необходимо сделать следующее:

1. Установить репозиторий PostgreSQL

```
yum install -y
https://download.postgresql.org/pub/repos/yum/11/redhat/rhel-7-
x86_64/pgdg-redhat-repo-latest.noarch.rpm
```

2. Создать файл `/etc/yum.repos.d/timescale_timescaledb.repo`

```
[timescale_timescaledb]
name=timescale_timescaledb
baseurl=https://packagecloud.io/timescale/timescaledb/e1/7/$basearch
repo_gpgcheck=1
gpgcheck=0
enabled=1
gpgkey=https://packagecloud.io/timescale/timescaledb/gpgkey
sslverify=1
sslcacert=/etc/pki/tls/certs/ca-bundle.crt
metadata_expire=300
```

3. Установить PostgreSQL

```
yum install -y postgresql96-server
```

4. Установить TimescaleDB

```
yum install -y timescaledb-postgresql-9.6
```

5. Инициализировать БД Postgresql

```
/usr/pgsql-9.6/bin/postgresql96-setup initdb
```

6. Настроить Postgresql для работы с TimescaleDB

```
timescaledb-tune --yes --pg-config=/usr/pgsql-9.6/bin/pg_config
```

7. Настроить доступ к таблицам Postgresql в

файле `/var/lib/pgsql/9.6/data/pg_hba.conf`

```
# IPv4 local connections:  
host    all             all             127.0.0.1/32      md5  
# IPv6 local connections:  
host    all             all             ::1/128           md5
```

8. Запустить Postgresql

```
su - postgres  
/usr/pgsql-9.6/bin/postgres &
```

9. Войти в консоль Postgresql

```
/usr/pgsql-9.6/bin/psql
```

10. Создать пользователя и БД

```
CREATE USER wcoam WITH PASSWORD 'wcoam';  
CREATE DATABASE wcsstat;
```

11. Дать пользователю права на доступ к БД

```
\c wcsstat  
GRANT ALL PRIVILEGES ON ALL TABLES IN SCHEMA public TO "wcoam";
```

12. Инициализировать расширение Timescale

```
CREATE EXTENSION IF NOT EXISTS timescaledb CASCADE;
```

13. Создать БД для хранения статистики потоков

```
CREATE DATABASE wcoam;  
\c wcoam  
GRANT ALL PRIVILEGES ON ALL TABLES IN SCHEMA public TO "wcoam";
```

14. Выйти из консоли Postgresql

```
\q  
exit
```

Настройка БД Timescale

Для настройки БД Timescale для хранения метрик необходимо:

1. Указать в файле настроек `wcsoam.properties` параметр

```
metric_store=timescale
```

2. Указать в файле настроек `init_tsdb.properties` размер фрагментов таблицы для хранения метрик. По умолчанию, размер одного фрагмента установлен в 2 дня

```
timescale_chunk_interval=2 days
```

3. Запустить скрипт настройки БД

```
./init_tsdb.sh
```

Настройка удаления устаревших данных при помощи cron

Для экономии дискового пространства устаревшие данные могут удаляться периодически при помощи cron. Например, проверка таблицы `conditions` каждые 5 минут и удаление данных старше 5 дней может быть настроена следующим образом:

```
*/5 * * * * PGPASSWORD="wcsoam" /usr/pgsql-9.6/bin/psql -h localhost -p 5432  
-U wcsoam -d wcsstat -w -c "SELECT drop_chunks(interval '5 days',  
'conditions');"
```

Структура БД

Postgresql должен содержать следующую таблицу для хранения метрик

Field	Type	Note
CONDITIONS		
time	timestampz	NOT NULL

Field	Type	Note
node_id	bigint	NOT NULL
media_id	character varying(255)	
video_height	bigint	
video_width	bigint	
video_rate	bigint	
video_sync	bigint	
video_fps	bigint	
video_nack	bigint	
video_pli	bigint	
video_codec	bigint	
audio_sync	bigint	
audio_rate	bigint	
audio_lost	bigint	
audio_codec	bigint	

Данные метрик фрагментируются и хранятся в подтаблицах, объем каждой из которых определяется заданной при настройке длительностью фрагмента. Для ограничения количества подтаблиц рекомендуется настройка [очистки устаревших данных при помощи cron](#) с указанием максимального времени хранения данных.