

Развитие модуля анализа исторической нагрузки `pg_profile/pgpro_pwr`

Андрей Зубков
Руководитель группы систем мониторинга
Postgres Professional

План презентации

Top SQL by elapsed time

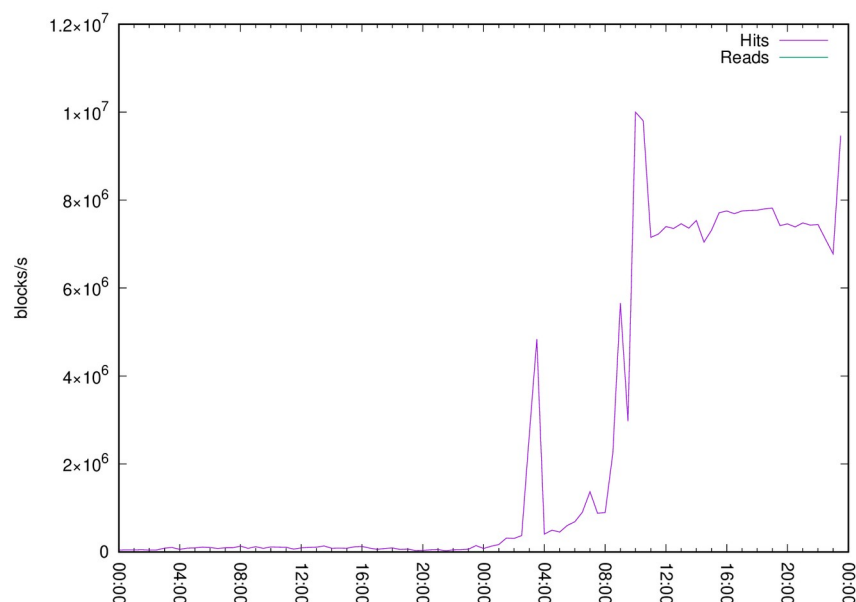
Query ID	Database	%Total
19e8277abe [43597a8eddd81e73]	bench	79.49
addb42436a [4e4ba95462874464]	postgres	7.47
52ab3d87cc [be42bf4c76d88a2f]	bench	4.09
adf5be70bc [24f4367c34e788bd]	bench	2.60
e7bcf9c5ff [e27327334db76a96]	bench	0.74

- Общие сведения о расширении
- Демонстрация на примере инцидента
- Развитие расширения
- Сложности в мониторинге Postgres
- Наши слоны в pgpro_pwr

pg_profile/pgpro_pwr

- Расширение на pl/pgsql
- Собирает статистическую информацию в редких снимках
- С нескольких серверов СУБД
- Строит отчеты между любыми двумя снимками
- Возможна выгрузка и загрузка наблюдений

Представления производительности



- `pg_stat_database`
- `pg_stat_tablespace`
- `pg_stat_bgwriter`
- `pg_stat_archiver`
- `pg_stat_wal`

- `pg_stat_all_tables`
- `pg_stat_all_indexes`
- `pg_statio_all_tables`
- `pg_statio_all_indexes`

- `pg_stat_statements`
- `pg_stat_kcache`
- *pgpro_stats*

- `pg_stat_user_functions`
- `pg_settings`

Процедура установки

- **Копирование файлов**
- Подготовка к установке
- Установка расширения
- Первый снимок
- Настройка планировщика

- Скачать последний релиз с гитхаба
- Распаковать файлы расширения:

```
tar xzf pg_profile-0.3.4.tar.gz \  
    --directory $(pg_config --sharedir)/extension
```

Процедура установки

- Копирование файлов
- **Подготовка к установке**
- Установка расширения
- Первый снимок
- Настройка планировщика

- Понадобится *dblink*:
`CREATE EXTENSION dblink;`
- Опционально *pg_stat_statements*, *pg_stat_kcache*:
`CREATE EXTENSION pg_stat_statements;`
`CREATE EXTENSION pg_stat_kcache;`

Процедура установки

- Копирование файлов
- Подготовка к установке
- **Установка расширения**
- Первый снимок
- Настройка планировщика

- Создадим схему

```
CREATE SCHEMA profile;
```

- И расширение:

```
CREATE EXTENSION pg_profile SCHEMA profile;
```

Процедура установки

- Копирование файлов
- Подготовка к установке
- Установка расширения
- **Первый снимок**
- Настройка планировщика

- Выполняем снимок

```
postgres=# SELECT * FROM profile.take_sample();
```

server	result	elapsed
--------	--------	---------

-----+	-----+	-----
--------	--------	-------

local	OK	00:00:00.6
-------	----	------------

(1 row)

- Исправляем ошибки

Процедура установки

- Копирование файлов
- Подготовка к установке
- Установка расширения
- Первый снимок
- **Настройка планировщика**

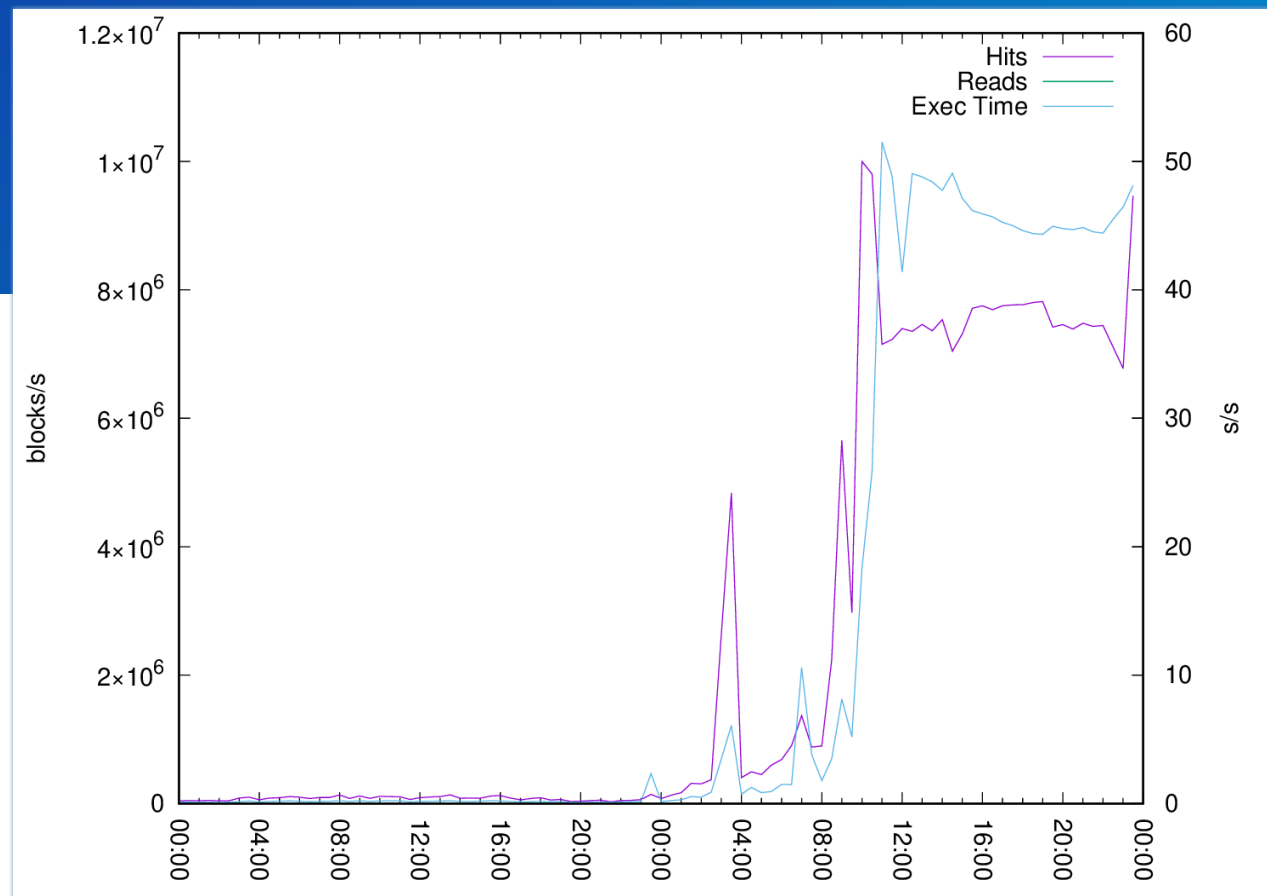
- Например, *cron*:

```
*/30 * * * * psql -c 'SELECT profile.take_sample()'
```

(Два снимка в час)

Пример инцидента

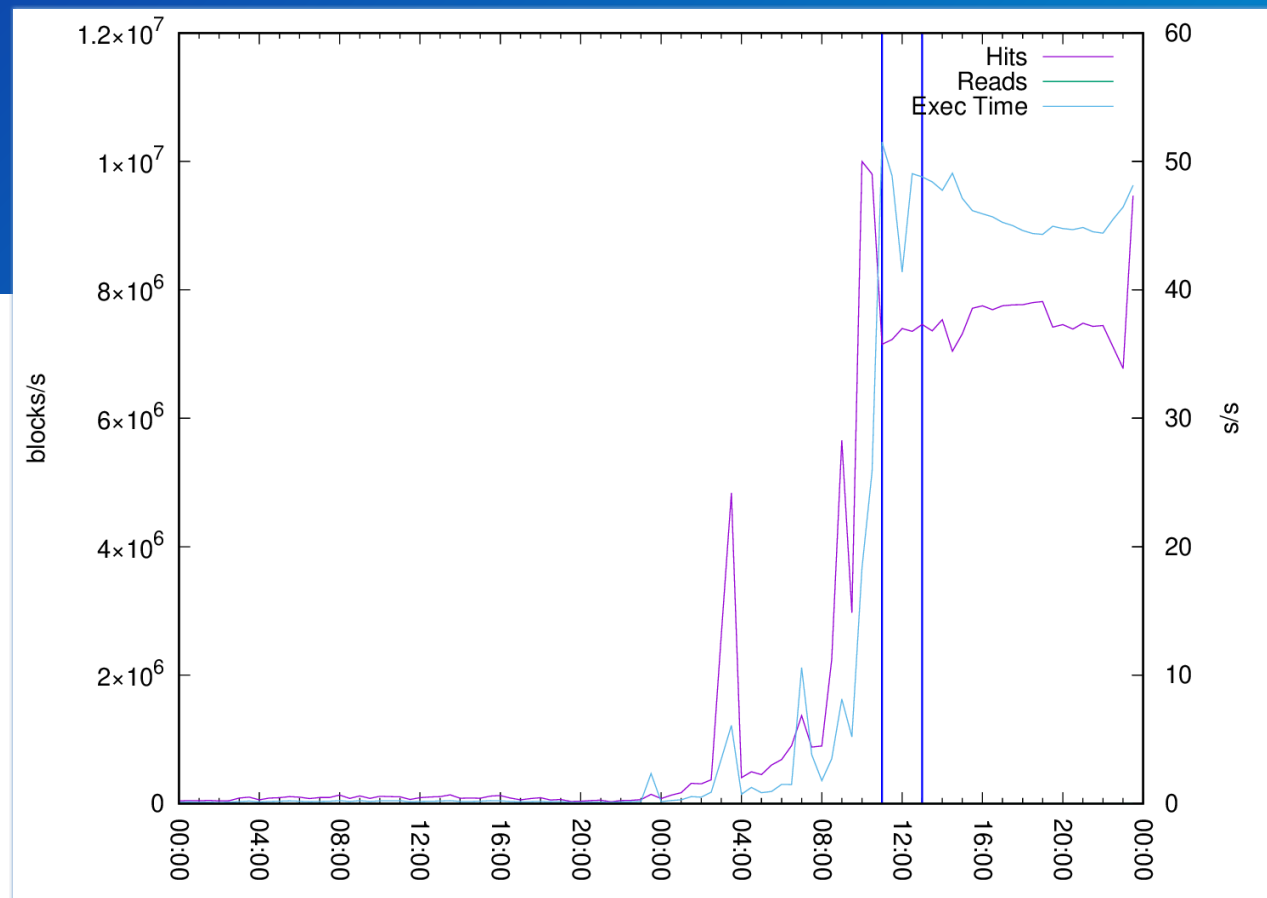
- Наблюдается существенная деградация производительности
- Имеется выраженное изменение нагрузки



Пример инцидента

Мы наблюдаем следствие, но хотим понять причину

Построим отчет по интервалу инцидента



Top SQL by execution time

Query ID	Database	Exec (s)	%Total	I/O time (s)		Rows	Execution times (ms)				Executions
				Read	Write		Mean	Min	Max	StdErr	
825ec9dfe2 [ba551e58a1220972]	demodb	17677.69	26.10			4976275	436.777	131.332	1268.142	87.207	40473
fc2b6ac0db [7058521854c25be5]	demodb	13814.61	20.40			435	124.600	33.169	523.835	53.414	110872
32e15bfa2b [2c3252b0a9ecf099]	demodb	12796.92	18.90				224.436	62.513	725.970	85.991	57018
9972b38b9c [711d687fdb6583af]	demodb	6819.08	10.07				216.334	63.142	628.219	86.583	31521
581a0cb27e [42c019fd344ccda3]	demodb	6763.17	9.99				2318.536	0.110	4348.005	515.938	2917
476c08c031 [de37a7b16ab1d9ec]	demodb	6257.31	9.24			3100	1098.931	0.012	4284.943	1233.238	5694
bb9daa91f5 [19858c316e39b93a]	demodb	820.60	1.21			19459	42.600	12.135	261.705	25.354	19263

65%

75%

Top SQL by shared blocks fetched

Query ID	Database	blks fetched	%Total	Hits(%)	Elapsed(s)	Rows	Executions
581a0cb27e #5 [42c019fd344ccda3]	demodb	7294930558	41.42	100.00	6763.2		2917
Fc2b6ac0db #2 [7058521854c25be5]	demodb	6232122854	35.38	100.00	13814.6	435	110872
825ec9dfe2 #1 [ba551e58a1220972]	demodb	2583010863	14.66	100.00	17677.7	4976275	40473
32e15bfa2b #3 [2c3252b0a9ecf099]	demodb	774440330	4.40	100.00	12796.9		57018
9972b38b9c #4 [711d687fdb6583af]	demodb	427932206	2.43	100.00	6819.1		31521
615932b6c7 #11 [3865b6f15c706793]	demodb	33263701	0.19	100.00	225.8	77099	962
5fadf658f1 #9 [2e77692b5dff5c21]	demodb	33045697	0.19	100.00	304.9	1268	1381

91%

- ранг в *Top SQL by execution time*

Top tables by blocks fetched

DB	Tablespace	Schema	Table	Heap		Ix		TOAST		TOAST-Ix	
				Blks	%Total	Blks	%Total	Blks	%Total	Blks	%Total
demodb	pg_default	i6c	i6_n_m	9481392382	53.64	357341693	2.02				
demodb	pg_default	i6c	i6_d_t_d	1656550069	9.37	4975941917	28.15				
demodb	pg_default	i6c	i6_d_t_m	696556409	3.94	31248149	0.18	4	0.00	66	0.00
demodb	pg_default	i6c	i6_n_as	91667256	0.52	4863695	0.03				

Top indexes by blocks fetched

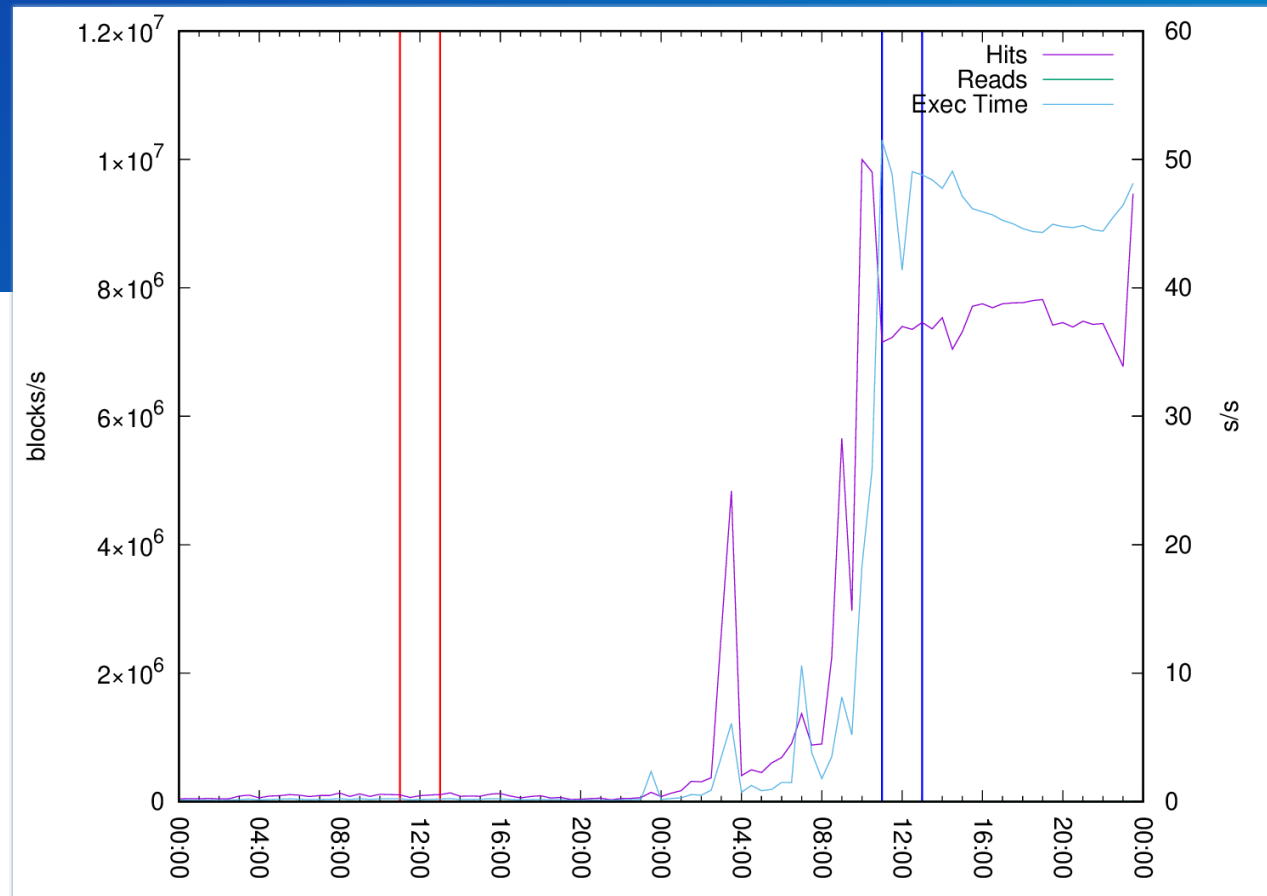
DB	Tablespace	Schema	Table	Index	Scans	Blks	%Total
demodb	pg_default	i6c	i6_d_t_d	i6_d_t_d_pk	1654718992	4975805807	28.15
demodb	pg_default	i6c	i6_n_m	i6_n_m_ix1	91626609	322193350	1.82
demodb	pg_default	i6c	i6_n_m	i6_n_m_ix2	88555	34801154	0.20
demodb	pg_default	i6c	i6_d_u	i6_d_u_uk	8529766	25684839	0.15

Что изменилось?

«Вчера всё работало!»

Дифференциальный отчет

- Строится по двум интервалам
- Значения метрик располагаются рядом



Database statistics

Database	I	Transactions			Block statistics			Tuples					Size	Growth
		Commits	Rollbacks	Deadlocks	Hit(%)	Read	Hit	Ret	Fet	Ins	Upd	Del		
demodb	1	1554854	1804		99.79	461906	218000570	701175479	90047284	27420	88226	9603	29 GB	7512 kB
	2	3439169	7307		99.99	1615331	17662792017	61934924526	25927080482	225683	357037	48921	29 GB	53 MB
aux	1	284			100.00		25876	162460	8712				7901 kB	
	2	284			100.00		24812	161980	8232				7901 kB	
postgres	1	2963			99.90	475	495350	1563478	89189	33269	736	32028	22 MB	296 kB
	2	2970			99.95	658	1459023	5186650	437424	33524	747	32063	27 MB	368 kB
Total	1	1558101	1804		99.79	462381	218521796	702901417	90145185	60689	88962	41631	29 GB	7808 kB
	2	3442423	7307		99.99	1615989	17664275852	61940273156	25927526138	259207	357784	80984	29 GB	53 MB

Top SQL by execution time

Query ID	Database	I	Exec (s)	%Total	Rows	Execution times (ms)				Executions
						Mean	Min	Max	StdErr	
825ec9dfe2 [ba551e58a1220972]	demodb	1	1.85	0.76	347	0.011	0.005	9.809	0.046	171201
		2	17677.69	26.10	4976275	436.777	131.332	1268.142	87.207	40473
fc2b6ac0db [7058521854c25be5]	demodb	1	1.74	0.71	135	0.010	0.003	11.883	0.067	172062
		2	13814.61	20.40	435	124.600	33.169	523.835	53.414	110872
32e15bfa2b [2c3252b0a9ecf099]	demodb	1	1.08	0.44		0.009	0.003	9.954	0.062	114724
		2	12796.92	18.90		224.436	62.513	725.970	85.991	57018
581a0cb27e [42c019fd344ccda3]	demodb	1	62.76	25.76		965.550	870.881	1085.988	58.479	65
		2	6763.17	9.99		2318.536	0.110	4348.005	515.938	2917
9972b38b9c [711d687fdb6583af]	demodb	1	0.61	0.25		0.010	0.003	7.502	0.057	58624
		2	6819.08	10.07		216.334	63.142	628.219	86.583	31521
476c08c031 [de37a7b16ab1d9ec]	demodb	1								
		2	6257.31	9.24	3100	1098.931	0.012	4284.943	1233.238	5694

Top tables by blocks fetched

DB	Tablespace	Schema	Table	I	Heap		Ix		TOAST		TOAST-Ix	
					Blks	%Total	Blks	%Total	Blks	%Total	Blks	%Total
demodb	pg_default	i6c	i6_n_m	1	536961	0.25	2418551	1.10				
				2	9481392382	53.64	357341693	2.02				
demodb	pg_default	i6c	i6_d_t_d	1	36214123	16.53	108835089	49.68				
				2	1656550069	9.37	4975941917	28.15				
demodb	pg_default	i6c	i6_d_t_m	1	17166613	7.84	1239130	0.57				
				2	696556409	3.94	31248149	0.18	4	0.00	66	0.00
demodb	pg_default	i6c	i6_n_a	1	574752	0.26	692931	0.32				
				2	91667256	0.52	4863695	0.03				

Развитие: Postgres 14

Представление `pg_stat_wal`

Содержит статистики журнала
упреждающей записи
(Masahiro Ikeda)

Metric	Value
WAL generated	7562 kB
WAL per second	222 kB
WAL records	12281
WAL FPI	679
WAL buffers full	
WAL writes	764
WAL writes per second	22.47
WAL sync	757
WAL syncs per second	22.26
WAL write time (s)	0.03
WAL write duty	0.08%
WAL sync time (s)	10.64
WAL sync duty	31.29%

Развитие: Postgres 14

Представление pg_stat_database

Содержит новые счетчики сессий (Laurenz Albe)

Session statistics by database

Database	Timings (s)			Sessions			
	Total	Active	Idle(T)	Established	Abondoned	Fatal	Killed
bench	27.30	26.49	0.39	5			
contrib_regression	0.25	0.20		1			
postgres	3.18	2.08		5			
Total	30.74	28.78	0.39	11			

Размеры отношений

Проблема:

- **pg_class.relpages** может быть неточным
- **pg_relation_size()** - долго и тяжело

Компромиссное решение — редкое измерение размеров

- **window_start** — время начала суточного окна измерения размеров
- **window_duration** — протяженность окна во времени
- **sample_interval** — минимальный интервал между снимками с размерами

Размеры отношений

Редкое измерение размеров

- **window_start** — время начала суточного окна измерения размеров
- **window_duration** — протяженность окна
- **sample_interval** — минимальный интервал между снимками с размерами

В 0.3.4 появился параметр **limited_sizes_collection**. Позволяет собирать размеры:

- Таблиц, если в них были Seq Scan или VACUUM
- Индексов, если в их таблицах был вакуум

Проблемы

- Размеры отношений
- Ограниченность `pg_stat_statements/pg_stat_kcache`
- Недостаток собираемых в ядре статистик
- Получение данных только после завершения выполнения выражения

Сброс pg_stat_statements

Используется для надёжного обнаружения вытеснений

- + Просто — не требует хранения предыдущих значений
- + Надёжно
- Влияет на мониторинг
- Может вызывать конкуренцию

Обсуждается патч

Track statement entry timestamp (<https://commitfest.postgresql.org/35/3048/>)

- Сохранение времени занесения выражения в хэш-таблицу
- Надо еще сбрасывать метрики min/max

Статистики вакуума

Нагрузка на вакуум

Таблица

Статистики вакуума

Нагрузка на вакуум

Таблица

Индекс

Статистики вакуума

Нагрузка на вакуум

Таблица

Индекс

Статистики вакуума

Нагрузка на вакуум

Top indexes by estimated vacuum I/O load

DB	Tablespace	Schema	Table	Index	~Vacuum bytes	Vacuum cnt	Autovacuum cnt
demo	pg_default	tiger	small_table	idx1	191 GB		621
demo	pg_default	tiger	small_table	pk_small_table	131 GB		621
demo	pg_default	tiger	small_table	idx2	130 GB		621
demo	pg_default	tiger	small_table	idx3	127 GB		621

- Размер таблицы примерно 700 КБ
- Размеры индексов 200 – 300 МБ

pgpro_pwr

Поддерживает pgpro_stats:

- **Статистики ожиданий**
- Планы запросов
- Распределение нагрузки

Wait statistics by database

Database	Wait event type	Waited (s)	%Total
bench	IO	0.24	15.79
bench	LWLock	0.79	51.97
bench	Timeout	0.49	32.24
bench	*	1.52	100.00
Total		1.52	

Top wait events

Database	Wait event type	Wait event	Waited (s)	%Total
bench	LWLock	WALWrite	0.79	51.97
bench	Timeout	PgSleep	0.49	32.24
bench	IO	DataFileImmediateSync	0.15	9.87
bench	IO	DataFileExtend	0.06	3.95
bench	IO	DataFileRead	0.02	1.32
bench	IO	WALWrite	0.01	0.66

pgpro_pwr

Поддерживает pgpro_stats:

- **Статистики ожиданий**
- Планы запросов
- Распределение нагрузки

SQL Query wait statistics

All wait event types					
Query ID	Plan ID	Database	Waited (s)	%Total	Details
45654e2243 [31180549f1e4ef4b1]	ad207fa1e04076c	bench	0.86	56.58	LWLock : 0.79 IO : 0.07
6222ba64f6 [178814b0d62dacff1]	8e5a27db1a887f1a	bench	0.49	32.24	Timeout : 0.49
f73d643fcb [e27327334db76a961]		bench	0.08	5.26	IO : 0.08
9a6808ba37 [9dce1ab20005a8c1]		bench	0.05	3.29	IO : 0.05
2b6412b658 [d82be2f606bc0bdf1]	846b518026564258	bench	0.02	1.32	IO : 0.02
b0fb839b71 [f6939cc6b6a1b5dc1]		bench	0.02	1.32	IO : 0.02

pgpro_pwr

Поддерживает pgpro_stats:

- Статистики ожиданий
- **Планы запросов**
- Распределение нагрузки

SQL Query wait statistics

All wait event types					
Query ID	Plan ID	Database	Waited (s)	%Total	Details
45654e2243 [31180549f1e4ef4b1]	ad207fa1e04076c	bench	0.86	56.58	LWLock : 0.79 IO : 0.07
6222ba64f6 [178814b0d62dacff1]	8e5a27db1a887f1a	bench	0.49	32.24	Timeout : 0.49
f73d643fcb [e27327334db76a961]		bench	0.08	5.26	IO : 0.08
9a6808ba37 [9dce1ab20005a8c1]		bench	0.05	3.29	IO : 0.05
2b6412b658 [d82be2f606bc0bdf1]	846b518026564258	bench	0.02	1.32	IO : 0.02
b0fb839b71 [f6939cc6b6a1b5dc1]		bench	0.02	1.32	IO : 0.02

pgpro_pwr

Поддерживает pgpro_stats:

- Статистики ожиданий
- Планы запросов
- **Распределение нагрузки**

View "pg_catalog.pg_stat_activity"				
Column	Type	Collation	Nullable	Default
datid	oid			
datname	name			
pid	integer			
leader_pid	integer			
usesysid	oid			
username	name			
application_name	text			
client_addr	inet			
client_hostname	text			
client_port	integer			
backend_start	timestamp with time zone			
xact_start	timestamp with time zone			
query_start	timestamp with time zone			
state_change	timestamp with time zone			
wait_event_type	text			
wait_event	text			
state	text			
backend_xid	xid			
backend_xmin	xid			
query	text			
backend_type	text			

pgpro_pwr

Поддерживает pgpro_stats:

- Статистики ожиданий
- Планы запросов
- **Распределение нагрузки**

View "pg_catalog.pg_stat_activity"				
Column	Type	Collation	Nullable	Default
datid	oid			
datname	name			
pid	integer			
leader_pid	integer			
usesysid	oid			
username	name			
application_name	text			
client_addr	inet			
client_hostname	text			
client_port	integer			
backend_start	timestamp with time zone			
xact_start	timestamp with time zone			
query_start	timestamp with time zone			
state_change	timestamp with time zone			
wait_event_type	text			
wait_event	text			
state	text			
backend_xid	xid			
backend_xmin	xid			
query	text			
backend_type	text			

pgpro_pwr

Поддерживает pgpro_stats:

- Статистики ожиданий
- Планы запросов
- **Распределение нагрузки**

Top Applications Load Distribution

Resource	Load distribution
Total time (sec.)	psql: 5.75 pgbench: 0.9
Executed count	pgbench: 7684
I/O time (sec.)	pgbench: 0.02 psql: 0.
Blocks fetched	psql: 327899 : 221 pgbench: 221
Shared blocks read	pgbench: 579 psql: 109
Shared blocks dirtied	psql: 10636 pgbench: 10636
Shared blocks written	psql: 10535
Temp and Local blocks written	
Temp and Local blocks read	

pgpro_pwr

Поддерживает pgpro_stats:

- Статистики ожиданий
- Планы запросов
- **Распределение нагрузки**

Top Applications Load Distribution

Resource	Load distribution	
	Interval 1	Interval 2
Total time (sec.)	psql: 5.75 pgbench: 0.02	load app: 6.88 psql: 0.02 pgbench: 0.02
Executed count	pgbench: 7684	benchapp2: 4632 benchapp1: 3967
I/O time (sec.)	pgbench: 0.02 psql: 0.02	load app: 0.02 benchapp2: 0.02 benchapp1: 0.02
Blocks fetched	psql: 327899 pgbench: 1	load app: 270330 psql: 8476 pgbench: 1
Shared blocks read	pgbench: 579 psql: 1	benchapp1: 169 benchapp2: 152
Shared blocks dirtied	psql: 10636 pgbench: 1	load app: 10365 psql: 10365
Shared blocks written	psql: 10535	load app: 15597
Temp and Local blocks written		
Temp and Local blocks read		

Спасибо за внимание



Андрей Зубков
Postgres Professional, 2021
a.zubkov@postgrespro.ru