

PostgreSQL

Veritabanı Sunucusu

Statistics Collector

Devrim GÜNDÜZ

PostgreSQL Danışmanı

devrim@gunduz.org

devrim@PostgreSQL.org[.tr]

devrim@commandprompt.com

23.12.2008

Ankara

INET-TR 2008



Ajanda

1. Statistics Collector temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats_temp_directory
5. İstatistik viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

23.12.2008
Ankara

INET-TR 2008



Ajanda

1. Statistics Collector temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats_temp_directory
5. İstatistik viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

23.12.2008
Ankara

INET-TR 2008



PostgreSQL' de Statistics Collector - Temeller

- Statistics Collector, PostgreSQL'de veritabanı ile ilgili istatistikleri toplayan ve bunların sorgulanabilmesini sağlayan bir süreçtir.
- Postmaster sürecinin alt süreci olarak başlar.

PostgreSQL' de Statistics Collector - Temeller

- Ön tanımlı olarak açık gelir.
- 8.3 sürümü ile birlikte gelen gelişmeler

Ajanda

1. Statistics Collector Temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats_temp_directory
5. İstatistik viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

23.12.2008
Ankara

INET-TR 2008



PostgreSQL' de Statistics Collector - Yapılandırma

- postgresql.conf içindeki 3 parametre.
- Değişikliklerin geçerli olması için reload yeterlidir.
- SET ile oturum bazında ayarlanabilme

PostgreSQL' de Statistics Collector - Yapılandırma

- **track_counts** : Tabloların ve indexlerin istatistiklerinin toplanıp toplanmayacağını belirler.
- 8.3 öncesindeki *stats_start_collector*,
- *stats_block_level* ve *stats_row_level* parametrelerinin karşılığıdır.

PostgreSQL' de Statistics Collector - Yapılandırma

- **track_activities** : Sunucuda işletilen anlık komutun görünmesini sağlar.
- Eski sürümlerdeki *stats_command_string* parametresinin karşılığıdır.

PostgreSQL' de Statistics Collector - Yapılandırma

- **track_functions** : Kullanıcı tanımlı fonksiyonların istatistiklerini toplar.
- 8.3'de yeni gelen bir özelliktir.

Ajanda

1. Statistics Collector Temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats_temp_directory
5. İstatistik viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

23.12.2008
Ankara

INET-TR 2008



PostgreSQL' de Statistics Collector

– İstatistiklerin Gözlemlenmesi

- postgresql.conf dosyasındaki 4 parametre
- Reload yeterlidir.
- SET ile oturum başına ayarlanabilirler.
-

PostgreSQL' de Statistics Collector

– İstatistiklerin Gözlemlenmesi

- log_parser_stats
- log_planner_stats
- log_executor_stats
- log_statement_stats
 - Önemli: *log_statement_stats* değeri diğer değerlerle aynı anda etkinleştirilemez.

Ajanda

1. Statistics Collector Temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats_temp_directory
5. İstatistik viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

23.12.2008
Ankara

INET-TR 2008



PostgreSQL' de Statistics Collector – stats_temp_directory

- Bir diğer postgresql.conf parametresi
- 8.3 sürümü ile geldi
- pg_stat_tmp dizinlerinin başka bir disk/ramdiskte tutulması ve başarımın artması

Ajanda

1. Statistics Collector Temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats_temp_directory
5. İstatistik Viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

23.12.2008
Ankara

INET-TR 2008



PostgreSQL' de Statistics Collector – İstatistik Viewları

- pg_stat_activity
- pg_stat_database
- pg_stat_all_tables, pg_stat_user_tables,
pg_stat_sys_tables
- pg_stat_all_indexes,
pg_stat_user_indexes,
pg_stat_sys_indexes

PostgreSQL' de Statistics Collector – İstatistik Viewları

- pg_statio_all_tables ,
pg_statio_sys_tables ,
pg_statio_user_tables
- pg_statio_all_indexes ,
pg_statio_sys_indexes,
pg_statio_user_indexes :
- pg_statio_all_sequences,
pg_statio_sys_sequences,
pg_statio_user_sequences

PostgreSQL' de Statistics Collector

-

- pg_stat_bgwriter
- 8.3'deki favorilerimden birisi!
- Bgwriter istatistiklerinin sorgulanabilmesi ve başarımların ayarlanabilmesi.
- Örnek!

Ajanda

1. Statistics Collector Temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats_temp_directory
5. İstatistik viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

23.12.2008
Ankara

INET-TR 2008



PostgreSQL' de Statistics Collector – İstatistiklerin sorgulanması

- (Raporlama araçları)
- Asıl sorgulamayı sorgu planlayıcısı yapar
- pg_class tablosu
- Birazdan vereceğimiz örnekte pagila veritabanını kullandım:

<http://pgfoundry.org/projects/dbsamples>

PostgreSQL' de Statistics Collector

- İstatistiklerin sorgulanması

- ```
SELECT relname,
CASE
 WHEN relkind='r' THEN 'Table'
 WHEN relkind='i' THEN 'Index'
 WHEN relkind='S' THEN 'Sequence'
 WHEN relkind='t' THEN 'TOAST Table'
 WHEN relkind='c' THEN 'Composite Type'
END AS relkind,
reltuples, relpages FROM pg_class WHERE
relname LIKE

'%payment%';
```

# PostgreSQL' de Statistics Collector

## - İstatistiklerin sorgulanması

- VACUUM ANALYZE yapılmış halinde payments\_p2007\_02 tablosunda 17 sayfada 2312 kayıt görünüyor.
- *idx\_fk\_payment\_p2007\_02\_customer\_id* ve *idx\_fk\_payment\_p2007\_02\_staff\_id* indexlerinde de 2312 kayıt göreceksiniz – ve bunlar daha az sayfada olacaktır.
- Partial indexlerde genelde daha az kayıt vardır.

# PostgreSQL' de Statistics Collector - İstatistiklerin sorgulanması

- Kritik soru: Değerler ne zaman güncelleniyor?
- VACUUM, VACUUM ANALYZE, CREATE INDEX
- Reltuples ve relpages değerleri anlık olarak **güncellenmezler.**



# PostgreSQL' de Statistics Collector

## - İstatistiklerin sorgulanması

- Veriler neden güncel olmalıdır?

# PostgreSQL' de Statistics Collector

## - İstatistiklerin sorgulanması

- pg\_statistic - > pg\_stats
- pg\_statistic bir katalog tablosudur.
- pg\_stats ise pg\_statistic tablosunun bir view'udur.
- pg\_stats daha kullanıcı dostudur.

# PostgreSQL' de Statistics Collector

- İstatistiklerin sorgulanması

pg\_stats kolonları:

- **schemaname**
- **tablename**
- **attname**
- **null\_frac**
- **avg\_width**
- **n\_distinct**
- **most\_common\_vals**
- **most\_common\_freqs**
- **histogram\_bounds**
- **correlation**

# PostgreSQL' de Statistics Collector

## - İstatistiklerin sorgulanması

- histogram\_bounds ve most\_common\_vals kolonları ile default\_statistics\_target değerinin ilişkisi

# PostgreSQL' de Statistics Collector

## - İstatistiklerin sorgulanması

- `SELECT attname, n_distinct,  
most_common_vals FROM pg_stats  
WHERE tablename = 'payment_p2007_02';`

# PostgreSQL' de Statistics Collector - İstatistiklerin sorgulanması

- Sonucun yorumlanması  
(Bu kısım sadece seminere katılanlar içindir )

23.12.2008  
Ankara

INET-TR 2008



# Ajanda

---

1. Statistics Collector Temelleri
2. Yapılandırması
3. İstatistiklerin gözlemlenmesi
4. stats\_temp\_directory
5. İstatistik viewları
6. İstatistiklerin sorgulanması
7. İstatistiklerin kullanılması

**23.12.2008**  
**Ankara**

**INET-TR 2008**



# PostgreSQL' de Statistics Collector

## – İstatistiklerin kullanılması

- EXPLAIN!
- EXPLAIN SELECT \* from  
payment\_p2007\_02;

QUERY PLAN

-----  
Seq Scan on payment\_p2007\_02  
(cost=0.00..40.12 rows=2312 width=28)  
(1 row)



# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- PostgreSQL, bu aşamada tablodaki gerçek sayfa sayısını bulur. Eğer bu değer ile relpages değeri farklı ise, pg\_class'daki değer ile son bulunduğu değeri oranlar ve bu oranı reltuples ile çarpıp tahmin edilen satır sayısını bulur.

# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- EXPLAIN SELECT \* from payment\_p2007\_02 WHERE customer\_id < 100;  
QUERY PLAN

-----  
Bitmap Heap Scan on payment\_p2007\_02 (cost=11.37..33.41 rows=**403** width=28)

Recheck Cond: (customer\_id < 100)

-> Bitmap Index Scan on  
idx\_fk\_payment\_p2007\_02\_customer\_id (cost=0.00..11.27 rows=403 width=0)

Index Cond: (customer\_id < 100)  
(4 rows)

# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- default\_statistics\_target değerini değiştirdiğinde planda da değişiklikler olur:
- SET default\_statistics\_target = 100;  
VACUUM ANALYZE

...ve:

# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

```
•EXPLAIN SELECT * from payment_p2007_02 WHERE
customer_id < 100;
QUERY PLAN
```

---

```
Bitmap Heap Scan on payment_p2007_02
(cost=11.34..33.33 rows=399 width=28)
Recheck Cond: (customer_id < 100)
-> Bitmap Index Scan on
idx_fk_payment_p2007_02_customer_id
(cost=0.00..11.24 rows=399 width=0)
```

# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- Gerçek değer: 398 :)
- WHERE sorgusundaki operatör
  - pg\_operator tablosu ve oprrest kolonu
  - < için scalartltest

# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- Selectivity !
- `SELECT histogram_bounds FROM pg_stats WHERE tablename='payment_p2007_02' AND attname='customer_id';`
- $selectivity = (1 + (1000 - EnAltDeğer) / (EnÜstDeğer - EnAltDeğer)) / ToplamDeğerSayısı$
- `ToplamDeğerSayısı = default_statistics_target` (birçok durumda)

# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- = operatöründe ise selectivity için histogram\_bounds işe yaramayacaktır:
- *SELECT null\_frac, n\_distinct, most\_common\_vals, most\_common\_freqs FROM pg\_stats WHERE tablename = 'payment\_p2007\_02' AND attname = 'customer\_id';*

# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- Eğer sorgulanan değer mcv içinde değilse:
- $selectivity = (1 - (\text{most\_common\_freqs}$  tablosundaki değerlerin toplamı) / (n\_distinct değeri – most\_common\_values içinde kaç değer olduğu)



# PostgreSQL' de Statistics Collector

## - İstatistiklerin kullanılması

- Eğer sorgulanan değer mcv içindeyse:
- selectivity =mvc içindeki değerin mcf tarafındaki karşılığı

# PostgreSQL' de Statistics Collector - İstatistiklerin kullanılması

- Sorular?

23.12.2008  
Ankara

INET-TR 2008



# Kaynaklar

---

- <http://www.PostgreSQL.org/docs>
- <http://wiki.PostgreSQL.org>
- <http://archives.PostgreSQL.org>

**23.12.2008**  
**Ankara**

**INET-TR 2008**



# PostgreSQL

Veritabanı Sunucusu

## Statistics Collector

Devrim GÜNDÜZ

PostgreSQL Danışmanı

[devrim@gunduz.org](mailto:devrim@gunduz.org)

[devrim@PostgreSQL.org](mailto:devrim@PostgreSQL.org)[.tr]

[devrim@commandprompt.com](mailto:devrim@commandprompt.com)

23.12.2008

Ankara

INET-TR 2008

