

# **SQL performance tutorial**

Roan Kattouw  
Berlin Hackathon 2012

# Indexes

- P

r

e

p -

a s

	<i>page_id</i>	<i>page_len</i>
--	----------------	-----------------

g	0	
e1	1	10775

_67	t	1305
-----	---	------

i	69	e	548
---	----	---	-----

d	70	d	34
---	----	---	----

R	74		50085
---	----	--	-------

I	77	i	10292
---	----	---	-------

M	...	s	...
---	-----	---	-----

A	t		
---	---	--	--

I

N

	<i>page_id</i>	<i>page_len</i>
--	----------------	-----------------

D		
E	1320629	15

X	9609464	15
---	---------	----

f	340226	16
---	--------	----

o	725948	16
---	--------	----

(	940255	16
---	--------	----

P	1065064	16
---	---------	----

a	...	...
---	-----	-----

g		
---	--	--

S  
E  
L  
E  
C  
T

T

<i>page_id</i>	<i>page_touched</i>
1	2012052218512
*	4
67	2012051512253
	1
F <sub>69</sub>	2012011507260
R	1
O <sub>70</sub>	2012012312473
	0
M <sub>74</sub>	2012052017485
	4



<i>page_id</i>	<i>page_touched</i>
10212	2005062604445
	2
13394	2005062604450
	3
13395	2005062604450
	3
28307	2005062604452
	2
31045	2005062604452
	3

S  
E  
L  
E  
C

T

<u>page_id</u>	<u>page_len</u>
<u>1320629</u>	<u>15</u>
* <u>9609464</u>	<u>15</u>
<u>340226</u>	<u>16</u>
F <u>725948</u>	<u>16</u>
R <u>940255</u>	<u>16</u>
O1065064	16
M..	...



<u>page_id</u>	<u>page_len</u>
132062	15
9	
960946	15
4	
340226	16
725948	16
940255	16

S  
E  
L

<i>page_id</i>	<i>page_is_new</i>
C1	0
T67	0
69	0
*70	0
74	0
F77	0
R78	<u>1</u>
O83	0
M	...



<i>page_id</i>	<i>page_is_new</i>
78	1
110	1
221	1
262	1
263	1

S

E

L

*page\_id*

C1

T67

69

\*...

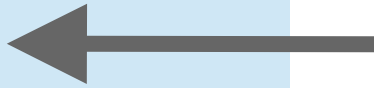
94106

F94109

R94112

O...

M



**b  
i  
n  
a  
r**

S  
E  
L  
E  
C  
T  
\*  
F  
R  
O  
M

*page\_id*

1  
67

...

94106

94109

94112

94115

94123

...

← b  
i  
n  
a  
r  
y

S  
E  
L  
E  
C  
T  
\*  
F  
R  
O  
M

<i>page_id</i>	<i>page_len</i>
C <sub>1</sub>	10775
T <sub>67</sub>	1305
...	...
*94106	465
<u>94109</u>	<u>688</u>
F <u>94112</u>	<u>533</u>
R <u>94115</u>	<u>31</u>
O <u>94123</u>	<u>287</u>
M..	...



<i>page_id</i>	<i>page_len</i>
1320629	15
9609464	15
340226	16



S

E

L

***img\_name***

C...

Flag\_of\_Břežany.jpeg

Flag\_of\_CANU.svg

\*Flag\_of\_CARICOM.svg

Flag\_of\_CEFTA.svg

Flag\_of\_CELAC.png

R...

O

M



***img\_name***

Flag\_of\_CANU.svg

Flag\_of\_CARICOM.svg

Flag\_of\_CEFTA.svg

S

E

L

E

C

T...

<i>img_name</i>	<i>img_size</i>
Flag_of_Břežany. jpeg	9264

<u>Flag_of_CANU.svg</u>	<u>11071</u>
-------------------------	--------------

<u>Flag_of_CARICO</u> <u>M.svg</u>	<u>3651</u>
---------------------------------------	-------------

...	...
-----	-----

<u>Flag_of_Cúcuta_(</u> <u>Norte_de_Santan</u> <u>der).svg</u>	<u>35869</u> <u>4</u>
--	--------------------------



<i>img_name</i>	<i>img_size</i>
-----------------	-----------------

Flag_of_Chh atarpur.svg	181
----------------------------	-----

Flag_of_Ceuta without_co a.svg	192
--------------------------------------	-----

Flag_of_Coss onay.svg	192
--------------------------	-----

S  
E  
L  
E  
C  
T  
\*  
F  
R  
O  
M

***img\_name***  
*FI-90-0020-Hitchhiking.jpg*  
FI-map-natfor.gif  
FI-resultaat.pdf  
\* FI-waldeck-Grab.JPG  
  
*Fl.jpg*  
FI.png  
...  
O



***img\_name***  
FI-90-0020-Hitchhiking.jpg  
  
FI.jpg  
  
FI10\_001.jpg

# Index on two fields

Works just like a phone book

<b>lastname</b>	<b>firstname</b>	<b>phone</b>
Stevens	Daniel	526-3401
Stevens on	Amy	945-8547
Stevens on	John	324-0625
Stevin	Aaron	777-6004

<b>page_name</b>	<b>page_title</b>	<b>page_id</b>
0	Main_ Page	3401
0	Nanobots	526
1	Main_ Page	8547
2	Catrop e	9

S

E

L

E

C

T...

Rostocker

\*

Rotatebot

R

O

Rotatebot

*img\_timesta  
mp*

2012030520  
0249

20080831180  
021

2009052800  
0041

2009090818  
0012

2009100518  
0217



**img\_user\_t  
ext**

**img\_timesta  
mp**

Rotatebot

2008083118  
0021

Rotatebot

2009052800  
0041

Rotatebot

2009090818  
0012

S  
E  
L  
E  
C  
T  
T  
R  
O  
M

<i>img_user_t ext</i>	<i>img_timesta mp</i>
Rnt20	2005050221 0737
<i>RoRo</i> *	2006121519 0509
<i>RoRo</i>	2007030316 5159
<i>RoRo</i>	2007030316 5610
<i>RoadyTom</i>	2006090609 4405
...	...



<i>img_user_t ext</i>	<i>img_timesta mp</i>
Roger Zenner	2005031717 3844
Romanm	2005073022 3825
Rosenzweig	2005081315 1322

S  
E  
L  
E

<i>Cimg_user_t</i>	<i>img_timesta</i>
<i>ext</i>	<i>mp</i>
T Rnt20	2005050221 0737
* <u>RoRo</u>	<u>2006121519</u> <u>0509</u>
<u>RoRo</u>	<u>2007030316</u> <u>5159</u>
F <u>RoRo</u>	<u>2007030316</u> <u>5610</u>
R <u>RoRo</u>	<u>2007030316</u> <u>5610</u>
O <u>RoadyTom</u>	<u>2006090609</u> <u>4405</u>
M	



<i>img_user_t</i>	<i>img_timesta</i>
<i>ext</i>	<i>mp</i>
RoRo	2006121519 0509
RoRo	2007030316 5159
RoRo	2007030316 5610

...

...

# LIMIT yourself

- Limit number of rows returned (LIMIT)
- Limit number of rows *scanned* (harder)



# DON'T

- Run unindexed queries
  - WHERE on rarely false conditions usually OK
  - Unindexed ORDER BY (*filesort*) **never OK**
- ORDER BY expression --> filesort == bad
- Page with OFFSET (or LIMIT 50,50)
  - LIMIT 10 OFFSET 5000 scans 5010 rows
  - Use WHERE foo\_id >= 12345 instead
- Use COUNT(), SUM(), GROUP BY, etc
  - No limit on rows scanned
  - MAX()/MIN() of indexed field on entire table is OK

# DO

- Think about how the DB will run your query
- Add indexes where needed
- Batch queries (when it makes sense)
- In some cases, denormalize for performance
  - Add information duplicated from other tables
  - Summary tables, counter tables, cache tables, etc.

# ContributionScores query

```
SELECT user_id, user_name, user_real_name, page_count, rev_count,  
page_count+SQRT(rev_count-page_count)*2 AS wiki_rank FROM `bw_user` u JOIN (  
(  
  SELECT rev_user, COUNT(DISTINCT rev_page) AS page_count, COUNT(rev_id) AS rev_count  
  FROM `bw_revision`  
  WHERE rev_user NOT IN (SELECT ug_user FROM `bw_user_groups` WHERE ug_group='bot')  
  GROUP BY rev_user ORDER BY page_count DESC LIMIT 50  
) UNION (  
  SELECT rev_user, COUNT(DISTINCT rev_page) AS page_count, COUNT(rev_id) AS rev_count  
  FROM `bw_revision`  
  WHERE rev_user NOT IN (SELECT ug_user FROM `bw_user_groups` WHERE ug_group='bot')  
  GROUP BY rev_user ORDER BY rev_count DESC LIMIT 50  
) s ON (user_id=rev_user) ORDER BY wiki_rank DESC LIMIT 50;
```