

DALIBO

L'expertise PostgreSQL



pgBadger V.4

PostgreSQL Conference Europe 2013
Jean-Paul Argudo <jean-paul.argudo@dalibo.com>

Table des matières

1pgBadger V.4.....	3
1.1 About.....	3
1.2 License.....	3
1.3 Agenda.....	4
1.4 About Gilles.....	4
1.5 About me.....	4
1.6 About us.....	5
1.7 What is pgBadger.....	5
1.8 Why use pgBadger.....	6
1.9 pgBadger History : pgFouine.....	6
1.10 pgBadger History : total rewrite.....	6
1.11 pgBadger History : V.1 results.....	7
1.12 V.1 branch.....	7
1.13 V.2 branch (1/2).....	8
1.14 V.2 branch (2/2).....	9
1.15 V.3 branch (1/3).....	10
1.16 V.3 branch (2/3).....	10
1.17 V.3 branch (3/3).....	11
1.18 V.4 branch !!!.....	12
1.19 V.4 Design.....	12
1.20 V.4 top menu.....	13
1.21 V.4 Global section.....	14
1.22 V.4 Key values section (1/2).....	15
1.23 V.4 Key values section (2/2).....	16
1.24 V.4 Graphics.....	16
1.25 V.4 Pie charts.....	17
1.26 V.4 new informations shown.....	18
1.27 V.4 new reports.....	18
1.28 V.4 new exclude-time option.....	18
1.29 V.4 more awesomeness.....	19
1.30 Prepare postgresql.conf (1/3).....	19
1.31 Prepare postgresql.conf (2/3).....	20
1.32 Prepare postgresql.conf (3/3).....	20
1.33 How to use pgBadger (1/3).....	21
1.34 How to use pgBadger (2/3).....	21
1.35 How to use pgBadger (3/3).....	21
1.36 Demo Time !.....	22

1 pgBadger V.4

1.1 About



- » Author : Jean-Paul Argudo jpa@dalibo.com
- » Company : [DALIBO](#)
- » Date : October 31st 2013
- » URL : [pgBadger V.4 @Dalibo Knowledge Base](#)

1.2 License



- Creative Common BY-NC-SA
- You are free
 - to Share
 - to Remix
- Under the following conditions
 - Attribution
 - Noncommercial
 - Share Alike

1.3 Agenda

pgBadger V.4



- About Gilles, me and us
- Whats pgBadger and how to use it
- pgBadger history
- What's in V.4
- Demo, if we have time

1.4 About Gilles



- Long-term Perl hacker (take a look at [his website](#))
- Creator of multiple tools for PostgreSQL, including [Ora2Pg](#)
- Dalibo Senior PostgreSQL Consultant since 2010

1.5 About me



- PostgreSQL.fr founder (2004)
- PostgreSQLFr non-profit co-founder (2005→)
- PostgreSQLEu non-profit co-founder (2008→2013)
- Dalibo's co-founder (2005→2011)
- Dalibo's elected CEO (2012→)

1.6 About us



- Dalibo is [PostgreSQL Platinum Sponsor](#)
 - Helping businesses and governmentals since 2005
 - Involved in PostgreSQL projects (not only code)
 - 20% of the working time for R & D
 - Funding events like [pgConf.eu](#), [pgday.fr](#)

1.7 What is pgBadger



- a PostgreSQL log parser that:
 - extracts valuable information
 - regroups queries by type
 - does stats, graphs and tables
 - different reports and points of view

1.8 Why use pgBadger



- if you want to audit your whole server
- or your application
- shall it be production or not
 - helpful to benchmark applications
 - and new configurations too

1.9 pgBadger History : pgFouine



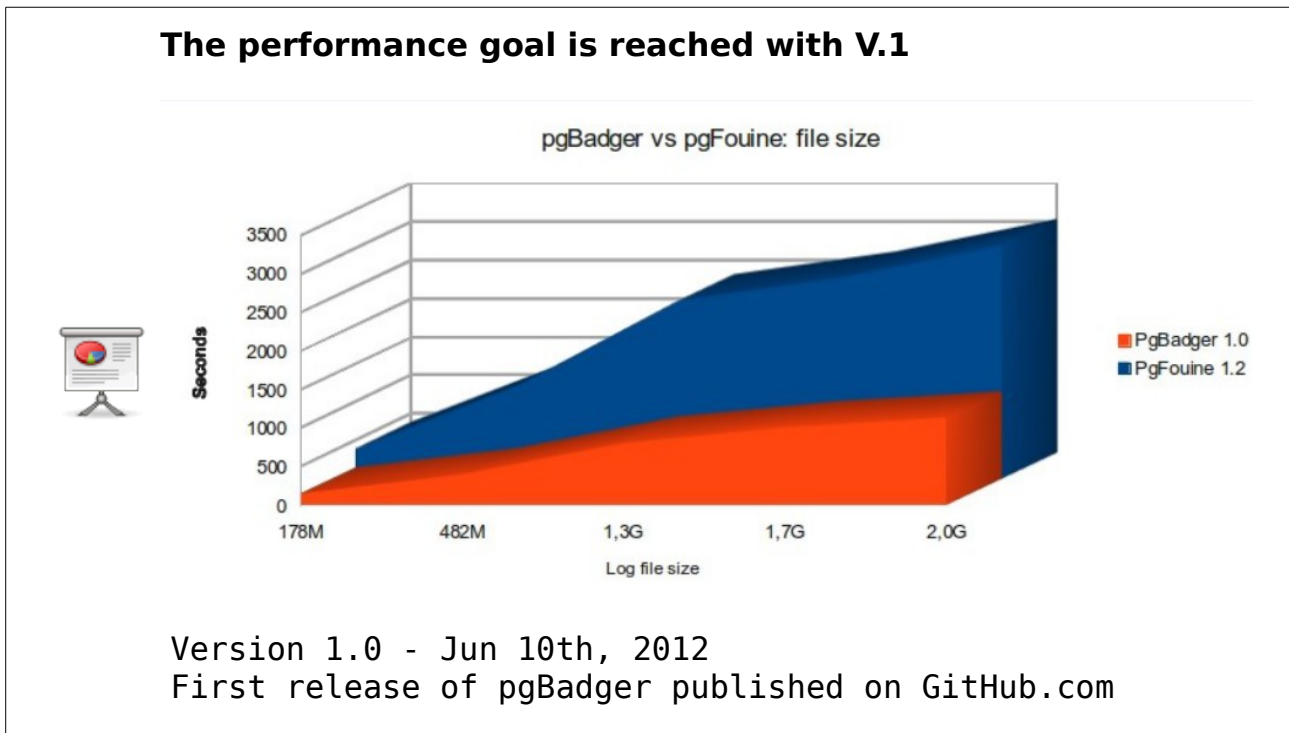
- We all waited long while pgFouine was running
- php-cli : slow, lots of memory used
- discontinued ?
 - wasn't reporting last PG's logs entries (checkpoints, temp files, locks, etc)

1.10 pgBadger History : total rewrite



- Gilles recoded everything in Perl
- Same output as pgFouine
- Additional reports
- Up to 3 times faster than pgFouine

1.11 pgBadger History : V.1 results



1.12 V.1 branch



- **Version 1.1 - Jul 04, 2012** :bugfixes and several features added
- **Version 1.2 - Jul 15, 2012** :(big)bugfix and new reports added

1.13 V.2 branch (1/2)

- **Version 2.0 - Aug 21, 2012**

- Support for incremental runs to avoid computing statistics on old data
- SQL format beautifier added on each reported query
- The CSV log parser has been entirely rewritten.



- **Version 2.1 - Oct 12, 2012**

- Major feature added: allows any custom log_line_prefix to be used with pgBadger
- Support to log_duration instead of log_min_duration_statement to allow reports simply based on duration and queries count without queries detail.
- Support compressed files with gzip, bzip2 and zip utilities.
- Bugfixes

1.14 V.2 branch (2/2)

- **Version 2.2 - Nov 13, 2012**

- Major features like tsung output, speed improvement with csvlog,
 - report of shut down events, new command line options to generate report excluding some user(s),
 - to build report based on select queries only, to specify regex of the queries that must only be included in the report and to remove comments from queries.
- Lots of bug fixes



- **Version 2.3 - Jan 17, 2013**

- Fixes several major issues especially with csvlog and a memory leak
- New reports of number of queries by database and application
- Mouse over reported queries show time, database, user, remote client and application name where they are executed
- New binary input/output format to allow saving or reading precomputed statistics. This allow incremental reports based on periodical runs of pgbader.
- Several SQL code beautifier improvement from pgFormatter have also been merged.

1.15 V.3 branch (1/3)



- **Version 3.0 - Feb 20, 2013**

- Adds parallel log processing by using as many cores as wanted, the performances gain is directly related to the number of cores specified.

- New reports about autovacuum and autoanalyze
- Bugfixes

- **Version 3.1 - Feb 21, 2013**

- Quick release to fix missing reports in V.3.0

1.16 V.3 branch (2/3)



- **Version 3.2 - Apr 07, 2013**

- Mostly a bug fix release

- **Version 3.3 - May 01, 2013**

- Adds four more useful reports about queries that generate locks and temporary files.
- Another new report about restartpoint on slaves
- Several bugfixes and cosmetic change

1.17 V.3 branch (3/3)

- **Version 3.4 - June 18, 2013**

- Lots of graphics improvements and a better rendering with logs over few hours
- Bugfixes on the report of queries that generate the most temporary files.

- **Version 3.5 - July 7th, 2013**



- Bugfixes release
- Adds some pretty print of Y axis number on graphs
- New graph that groups queries duration series that was shown as second Y axis on graphs
- New graph with number of temporary file that was also used as second Y axis

- **Version 3.6 - September 17th, 2013**

- Final bugfixes release
- **Last release of the 3.x branch**

1.18 V.4 branch !!!

- **“Say goodbye to the fouine” release**
- **Full rewrite** of the report design
- Intuitive user experience
- Professional look: [bootstrap](#) based



1.19 V.4 Design


art is code

design, développement
& contenu pour le web




- We know how to code
- But don't know anything about design
- We then contracted with [Art is code](#), a small desing an web agency in Paris

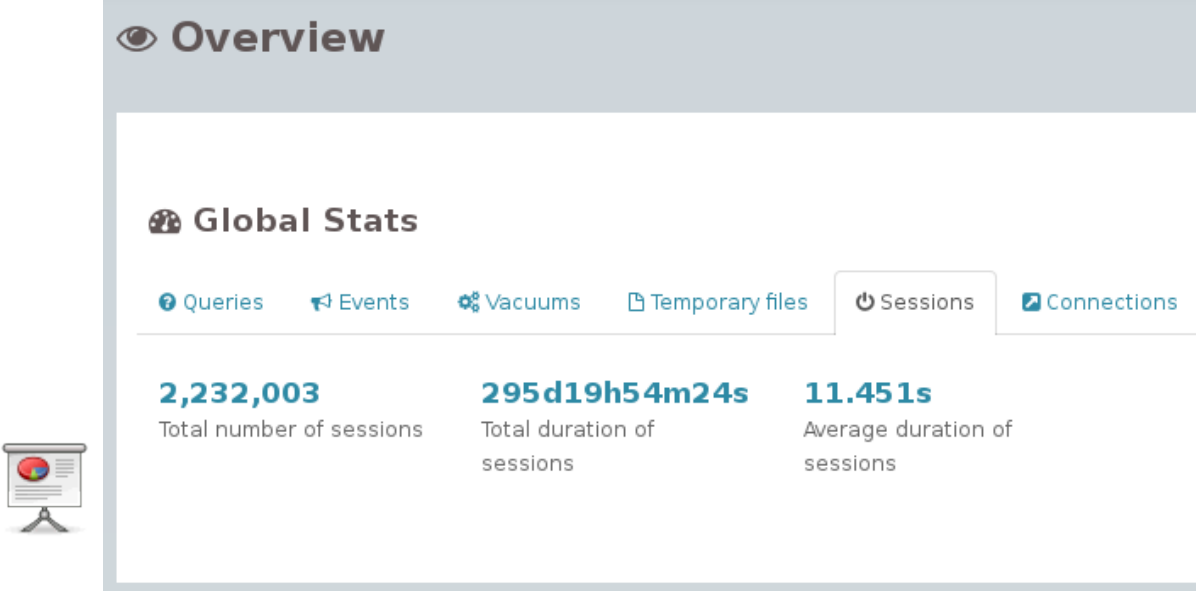
1.20 V.4 top menu

 pgBadger Overview ▾ Connections ▾ Sessions ▾ Checkpoints ▾ Temp Files ▾ Vacuums ▾ Locks ▾ Queries ▾ Top ▾ Events ▾

- Dynamic menu with the help of the embedded [bootstrap](#) library
- Menues and submenues based
- Embedded FontAwesome webfont to beautify the report
- Embedded logos and icons: a single “all-in-one” HTML



1.21 V.4 Global section



The screenshot displays the 'Overview' section of the pgBadger V.4 interface. It features a 'Global Stats' section with a navigation bar containing 'Queries', 'Events', 'Vacuums', 'Temporary files', 'Sessions', and 'Connections'. The 'Sessions' tab is selected, showing three key metrics: 2,232,003 total sessions, a total duration of 295d19h54m24s, and an average duration of 11.451s per session. A small icon of a presentation board is visible on the left side of the screenshot.

- Shows immediately most relevant values about
 - Queries
 - Events
 - Vacuums
 - Temporary files
 - Sessions
 - Connections

1.22 V.4 Key values section (1/2)

Statistical reports now include a “key values” section reports immediately some of the most relevant informations

A SQL Traffic

KEY VALUES

744 queries/s

Query Peak

2012-12-07 02:02:02

Date



↗ Size of temporary files

KEY VALUES

50.45 MiB

Temp Files Peak

2012-03-21 14:31:49

Date

1.23 V.4 Key values section (2/2)

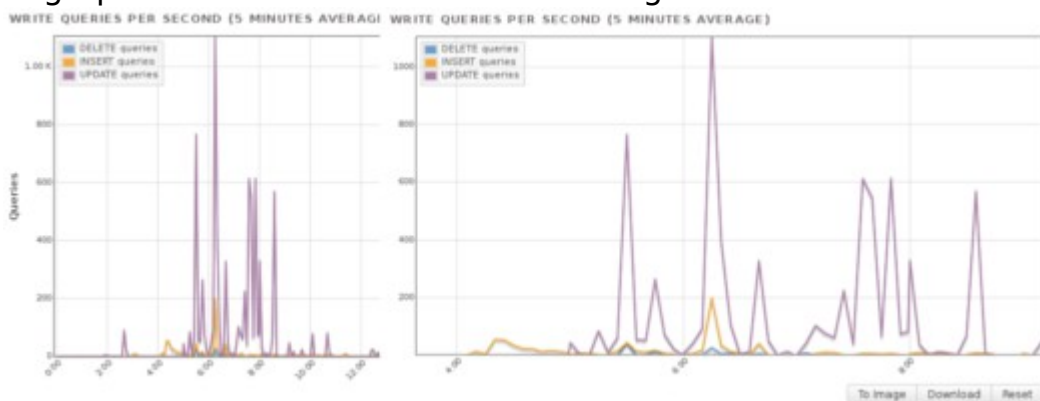
In the “key values” sections you can find information on those Peaks:



- Queries, read queries and write queries peaks
- Connections peak
- Checkpoints and WAL files usage peaks
- Checkpoints warnings peak
- Temporary file size and number peaks

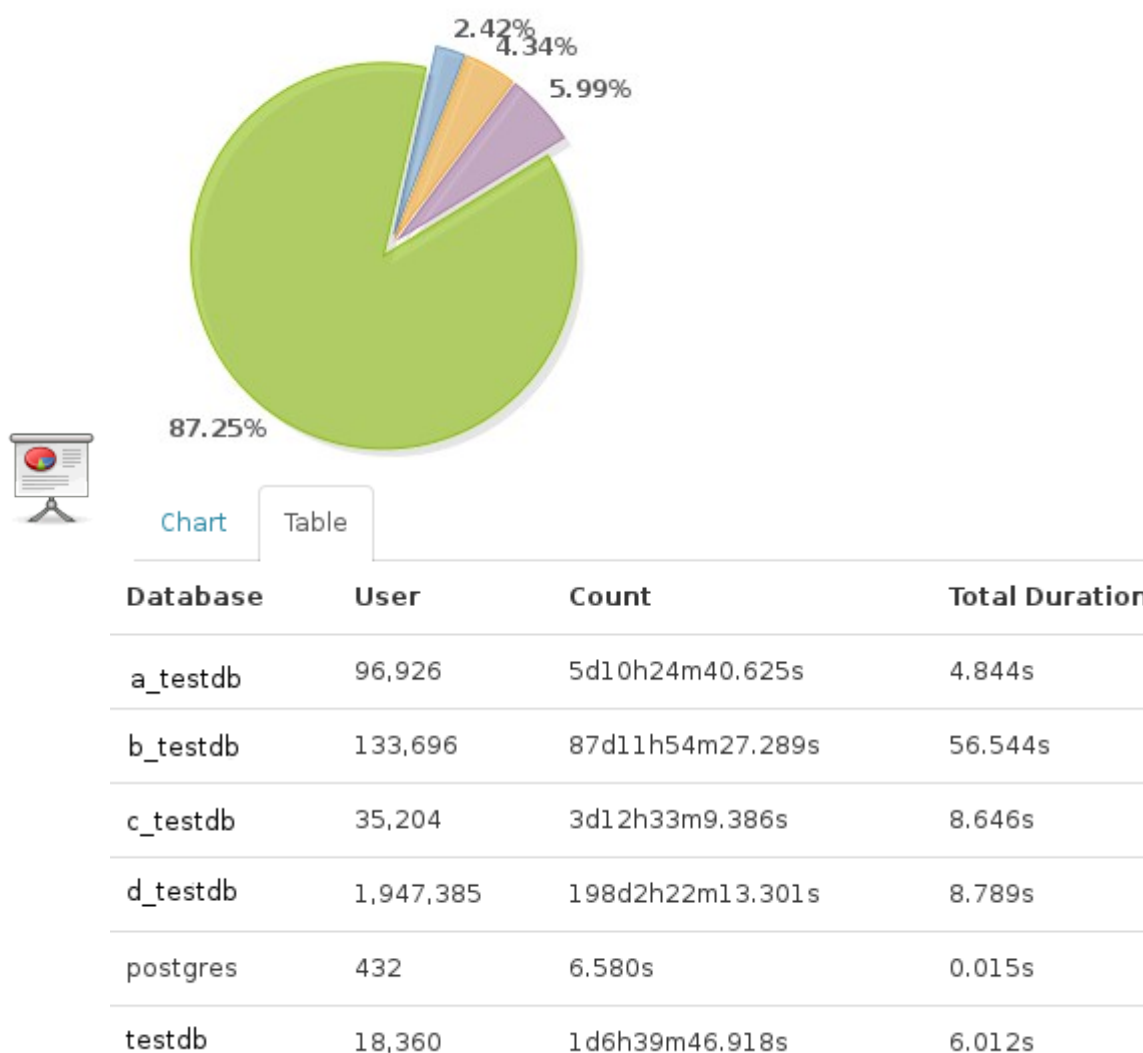
1.24 V.4 Graphics

All graphs are downloadable as PNG images and zoomable



1.25 V.4 Pie charts

Pie charts have also been separated from their data tables using two tabs, one for the chart and the other one for the data.



1.26 V.4 new informations shown



pgBadger keeps following PG's innovation, so the Checkpoint Activity report now shows

- additional information about checkpoints
- number of synced files
- the longest sync and the average of sync time per file

1.27 V.4 new reports



- Prepared queries ratio (execute vs prepare)
- Prepared over normal queries
- Queries (select, insert, update, delete) per user/host/application
- Pie charts for tables with the more tuples and pages removed during vacuum

1.28 V.4 new exclude-time option



- We can now exclude portions of times from the analysis
- `exclude-time` accepts a regexp
- example: `exclude 1:00pm to 1:39pm and 10:00pm to 10:39pm`
- `pgbadger --exclude-time '2013-09-\d+ (13|22):[0-3]' postgresql.log`
- use it when needed, this option has an high cost on the parser performances

1.29 V.4 more awesomeness

Can generate [Tsung](#) XML scenario files since V.2.2: based on your logged queries, pgBadger can produce a realistic benchmark scenario.



```
-x | --extension: output format. Values: text, html or  
tsung. Default: html
```

pgBadger has become parallel, did you notice ?

```
-j | --jobs number: number of jobs to run at same time.  
Default is 1, run as single process.
```

1.30 Prepare postgresql.conf (1/3)



- postgresql.conf must be adapted for pgBadger
- 1st parameter to adapt is `log_min_duration_statement >= 0`
 - what's the minimum duration time of queries you want pgBadger to handle?

1.31 Prepare postgresql.conf (2/3)

log_line_prefix should be like

- with stderr output



- '%t [%p]: [%l-1] '
- '%t [%p]: [%l-1] user=%u,db=%d '
- '%t [%p]: [%l-1] db=%d,user=%u '

- with syslog output

- 'db=%d,user=%u '

1.32 Prepare postgresql.conf (3/3)

Some additional informations that could be collected and reported



- log_checkpoints = on
- log_connections = on
- log_disconnections = on
- log_lock_waits = on
- log_temp_files = 0
- log_autovacuum_min_duration = 0

1.33 How to use pgBadger (1/3)

Trivial example



```
$ pgbadger mylogfile.log
```

will produce a single HTML file with defaults values and options

1.34 How to use pgBadger (2/3)

Consider only 5min of the activity with a custom prefix in the logs



```
$ perl pgbadger -b "2013-10-25 10:55:00" \  
-e "2013-10-25 11:00:00" \  
-prefix '%m %u@d %p %r %a : ' \  
/pglog/postgresql.log
```

1.35 How to use pgBadger (3/3)

Generate report every week using incremental behavior

In the crontab



```
0 4 * * 1 /usr/bin/pgbadger \  
-q `find /var/log/ -mtime -7 -name "postgresql.log*"` \  
-o /var/reports/pg_errors-`date +%F`.html \  
-l /var/reports/pgbadger_incremental_file.dat
```

This supposes that your log file and HTML report are also rotated every week.

1.36 Demo Time !



- **Send us feedback** and why not **your anonymised logs**
- V.4 Release date: CURRENT_TIMESTAMP !!!
- Get it at <http://dalibo.github.io/pgbadger>
- Follow the project on twitter: @pgbadger

Enough slides! **Let's demo now**