

Using *Grimoire to analyze OpenStack

Jesus M. Gonzalez-Barahona

jgb@bitergia.com

<http://identi.ca/jgbarah> <http://twitter.com/jgbarah>

Bitergia

GSyC/LibreSoft (Universidad Rey Juan Carlos)

OpenStack Summit, Portland, April 15th, 2013





©2012, 2013 Bitergia

Some rights reserved. This presentation is distributed under the
“Attribution-ShareAlike 3.0” license, by Creative Commons, available at
<http://creativecommons.org/licenses/by-sa/3.0/>



We've done the Grizzly companies analysis

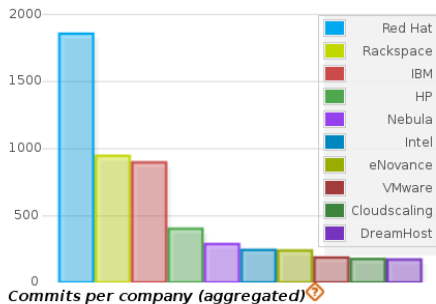
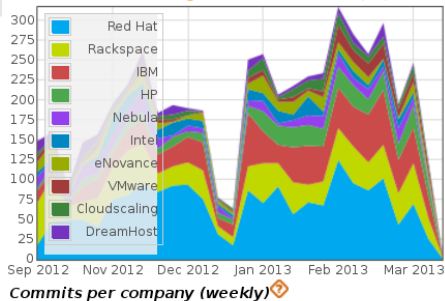
[Apr 4th 2013](#) ([methodology notes](#))

[s|Messages](#)): [Red Hat](#) | [Rackspace](#) | [IBM](#) | [HP](#) | [Nebula](#) | [Intel](#) | [eNovance](#) | [VMware](#) | [Cloudscaling](#) | [City of Melbourne](#) |



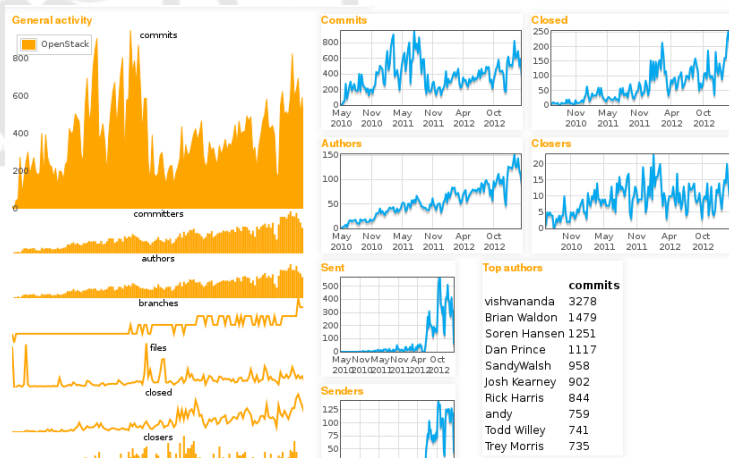
SUMMARY OF ACTIVITY

Overall View of the Changes to the Source Code (Git)



<http://blog.bitergia.com/2013/04/04/companies-contributing-to-openstack-grizzly-analysis/>

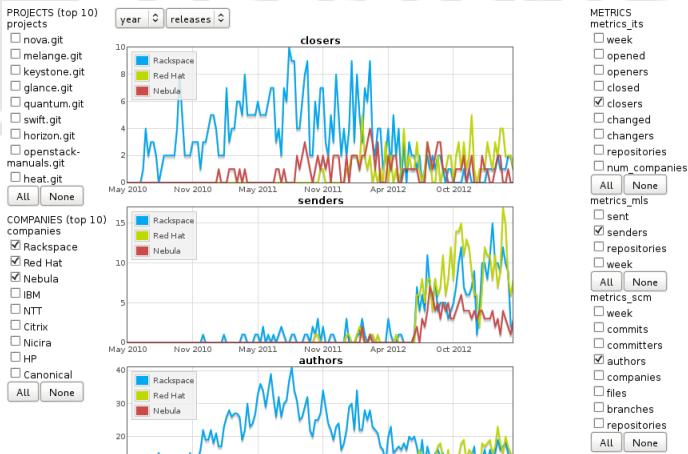
We're deploying the OpenStack activity dashboard




<http://activity.openstack.org/dash>

We're deploying the OpenStack activity dashboard (2)

This is a (preliminary) preview!!!



<http://activity.openstack.org/dash/dashboard> 

Information about code, community, development
for OpenStack
can be retrieved, organized, analyzed

Let's do it!



Why?

Bitergia

Open development: transparency

Complex information: knowledge extraction

Community-based decisions: objective information

Decision tracking: parameters to decide

...



Data has to be extracted, mined

Data lives in repositories
usually not designed to release it easily:

tools are needed to retrieve and extract

Data includes many complexities and details

tools are needed to assist in mining, analysis

Analyze free software with free software!




The MetricsGrimoire approach

Set of tools specialized in retrieving information from different kinds of repositories. Among them:

- CVSSAnALY: source code management (CVS, Subversion, git, etc.)
- Bicho: issue tracking systems (Bugzilla, Jira, SourceForge, Allura, Launchpad, Google Code, etc.) & code revision systems (Gerrit)
- MLStats: mailing lists (mbox files, Mailman archives, etc.)

Store all the information in SQL databases

<http://metricsgrimoire.github.com> 

Once information is ready for querying:

- it can be queried directly in the database
- it can be analyzed from R, Python, etc.
- it can be filtered, manually inspected, improved
- it can be combined, cross-analyzed
- it can be visualized

Set of tools to simplify & automate all of this

<https://vizgrimoire.github.com>



R package specialized in managing MetricsGrimoire information

Connects directly to the database and:

- gets the information from it
- filters & massages it
- does statistical analysis on it
- produces charts and WebGL 3D graphs
- produces JSON files to export to other tools

...and lets you unleash all the potential of R



JavaScript library producing visualizations

Retrieves JSON files and produces:

- live charts: evolution, pies, bars, etc.
- tables and text
- comparative charts
- actionable dashboards
- soon to support links to information in forge

Integration with HTML5 applications



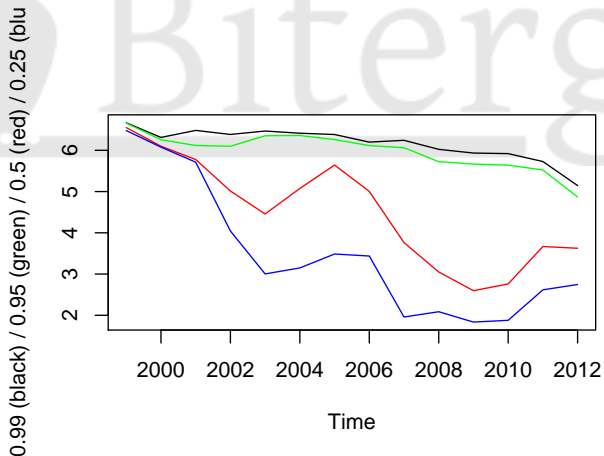
How the OpenStack dashboard is created

- Run MetricsGrimoire (CVSAnalY, Bicho, MLStats) on repositories
Example: determining which git repositories to mine
- Produce queries specific for OpenStack
Example: condition for deciding who closes a ticket
- Run customized Python & R scripts to produce JSON files
Example: produce results per OpenStack subproject
- Customize vizGrimoireJS to display charts
Example: remove bots from top tables
- Export the result via HTTP We still have performance issues...

Do all of this continuously



Future features: tracking other parameters



Time-to-close (quantiles over time): Time in minutes, log 10 scale



Future features: tracking other parameters (2)

99% of tickets open during the corresponding month were closed in less than x days



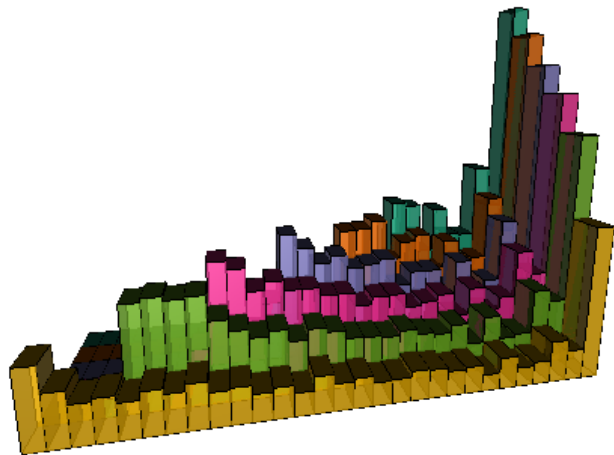
95% of tickets open during the corresponding month were closed in less than x days



Time-to-close (quantiles over time)



Future features: tracking other parameters (3)



Demographics (attraction rate, retention rate per generation)



In summary...

- OpenStack repositories have a wealth of information
- We all can do (and many do) our own analysis
- Free software to analyze free software development
- Let's define common formats to interface to different tools
- We can incrementally develop a powerful platform

What would you like to know about OpenStack?



Bitergia: an spin-off

Started operations in July 2012

Builds on the experience of LibreSoft R&D group

Offering professional products and services

Focused on:

- Metrics about software development (including community metrics)
- Specialized support for development forges (including metrics for projects)

<http://bitergia.com>



Thanks go to...

- Many LibreSoft developers who developed MetricsGrimoire
- The (small) community now maintaining MetricsGrimoire
- Some Bitergia developers producing vizGrimoire
- The (future) community maintaining vizGrimoire
- The many free software developers who produced all the software on which these tools rely
- You, who are producing OpenStack

<http://libresoft.es>

<http://bitergia.com>



This is the end, my friend

Please, provide ***any*** feedback

[I would love to know what interested you the most]
[...and the least]

Final note:

Let's find ways to understand OpenStack a bit better

