Introduction to GrimoireLab and more ...

Venu Vardhan Reddy Tekula

CHA OSS Beijing Meetup
08 July, 2021

chaoss.community
Links and slides

vchrombie.github.io/presentations
Agenda

• What is GrimoireLab
• A Brief History
• How does it work
• Demo of GrimoireLab
• How to Get Started
• Current Work
• How to Get Involved
About myself

Hyderabad, India
Graduate from Amrita Vishwa Vidyapeetham, Member @amfoss
GrimoireLab Community Manager
Previously Backend Developer at Bitergia
Google Summer of Code w/ CHAOSS and GrimoireLab

vchrombie.github.io
Community Projects Health

What is the data?

How do we collect the data?

What are the goals?

What are the tools?

Where do we store it?

How do we want to show it?

What about contributors information?
GrimoireLab

FREE, LIBRE OPEN SOURCE TOOLS FOR SOFTWARE DEVELOPMENT ANALYTICS

grimoirelab.github.io
A Brief History

2001-2015
Counting Potatoes: the Size of Debian 2.2

Jesús M. González-Barahona, Miguel A. Ortuño Pérez, Pedro de las Heras Quirós, José Centeno González and Vicente Matellán Olivera

Debian is the largest Free Software distribution, with well over 2,800 source packages in the latest stable release (Debian 2.2) and more than 4,000 source packages in the release currently in preparation. But, how large is “the largest”? In this paper, we use David Wheeler’s slccount system to determine the number of physical source lines of code (SLOC) of Debian 2.2 (aka Potato). We show that Debian 2.2 includes over 56,000,000 physical SLOC (almost twice than Red Hat 7.1, released about 8 months later), showing that the Debian development model (based on the work of a large group of voluntary developers spread around the world) is at least as capable as other development methods (like the more centralized one, based on the work of employees, used by Red Hat or Microsoft) to manage distributions of this size.

Keywords: Debian, Free Software, Libre Software, SLOC, Lines of Code, Linux

1 Introduction

On August 14th of 2000 the Debian Project announced Debian GNU/Linux 2.2, the “Joel ‘Espy’ Klecker” release [Debian22Ann] [Debian22Rel]. Code named “Potato”, it is the latest (to date) release of the Debian GNU/Linux Operating System. In this paper, we have counted this distribution, showing it is the largest this far.

2 Some Background about Debian

The Debian 2.2 GNU/Linux distribution is put together and maintained by the Debian project. In this section, we offer some background data about Debian as a project, and about the Debian 2.2 release.

The FLOSSMetrics project

FLOSSMetrics stands for Free/Libre Open Source Software Metrics.

The main objective of FLOSSMETRICS is to construct, publish and analyse a large scale database with information and metrics about libre software development coming from several thousands of software projects, using existing methodologies, and tools already developed.

FLOSSMetrics is providing access to dumps of this database (along with charts, tables and other quantitative information about FLOSS development projects) in the Melquiades website.

FLOSSMetrics is also providing the FLOSS Guide for SMEs (small and medium enterprises), which explains the benefits of FLOSS form SMEs, and how to take advantage of it.

More information about the project, including results and main achievements, can be found in the final report document.

FLOSSMetrics main targets

- Identify and evaluate sources of data and develop a comprehensive database structure, built upon the results of CALIBRE
- Integrate already available tools to extract and process such data into a complete platform
- Build and maintain an updated empirical database applying extraction tools to thousands of open source projects
- Develop visualisation methods and analytical studies, especially relating to benchmarking, identification of best practices, measuring and predicting success
MetricsGrimoire

[Update]: Currently, main development related with MetricsGrimoire has stopped. Main contributors are working on a new platform: GrimoireLab

MetricsGrimoire (pronounced /ˈmɛtəkrɪmɔɪr/) is a toolset to obtain data from repositories related to software development: source code management (aka version control), issue tracking (aka bug reporting) systems, mailing lists, etc. Data and metadata about the software development processes is retrieved from those repositories (information about commits, ticket management, communication in mailing lists, etc.), and then organized and stored into SQL databases that can later be mined for specific patterns or summaries of activity.

MetricsGrimoire tools support many kinds of repositories, including those provided by GitHub (git & GitHub issue tracking).

MetricsGrimoire has been already used to analyze many different projects and together with visualization tools like VizGrimoire it is possible to get dashboards or reports like the ones
The vizGrimoire project

vizGrimoire (usually pronounced /vizGrimoir/) is a toolset and framework to analyze and visualize data about software development. Currently, it is focused on data produced by the MetricsGrimoire tools (CVSAnalY, Bicho and MailingListStats).

vizGrimoire is promoted by Bitergia, the company providing software development analytics services, but it is a project open to the community: you are welcome to participate.

The following is a list of the current set of tools with a brief description. For more information, visit each of tool repositories, and/or read their install and readme files. You can also have a glimpse of the intended vizGrimoire architecture.

vizGrimoireR

Analyze data in databases produced by MetricsGrimoire tools from software development repositories (source code management systems, issue tracking systems, mailing lists, etc.).

Repository at GitHub

vizGrimoireJS

JavaScript framework to create dashboards and interactive reports and visualizations from data produced by vizGrimoireJS

Repository at GitHub
Extracting Knowledge From CHA OSS

CHA OSS will develop metrics, including:
- Project lifecycle
- Diversity and inclusion
- Risk and provenance
- Ecosystem impact

In addition...
- Develop a FLOSS reference implementation of defined metrics
- Integrate GrimoireLab, Prospector and Credible into a collaborative framework
- Develop a better understanding of how contributions happen

Backed by:
- Red Hat, Bitnami, Eclipse Foundation, Linaro, Mozilla, OpenStack, Polytechnique Montreal, Sauce Labs, Software Sustainability Institute, Symphony Software Foundation, University of Missouri, University of Mons, University of Nebraska at Omaha, and University of Victoria, Laval University, and Jono Bacon Consulting

THE LINUX FOUNDATION
How does it work
Where is the community? Where is the data?

Git
GitHub
GitLab
BitBucket
Jira
Gerrit
...

CHA OSS
Where is the community? Where is the data?

- Git
- GitHub
- GitLab
- BitBucket
- Jira
- Gerrit
- ...

Community doesn’t mean just code

Community means people, communications, and more
Where is the community? Where is the data?

Git
GitHub
GitLab
BitBucket
Jira
Gerrit
...

Wiki
Discourse
Mailing List
Slack
Meetup.com
StackOverflow
...

CHAOSs
Raw data:
- Get data from the data sources

Enriched data:
- Unify data, extract what is required
- Manage identities
- Calculate metrics

Useful data:
- Visualize, report
github.com/chaoss/grimoirelab#grimoirelab-components
Demo of GrimoireLab

sneak-peek
GrimoireLab powers Metrics aaS

There are some software that offer community metrics as a service, with GrimoireLab working under the hood.

- cauldrion.io

- lfx.linuxfoundation.org/tools/insights
Demo of GrimoireLab

There are some instances already deployed, you can visit any of the links

- chaoss.biterg.io

- dashboard.mautic.org
How to Get Started
How to Get Started

1. docker-compose (easiest of all)
2. docker run
3. pip packages
4. source code (mainly, for developers)
1. Install git, docker & docker-compose
2. Clone chaoss/grimoirelab repository
3. Go to docker-compose folder
4. Run docker-compose up -d
5. Open your browser to see the dashboard at localhost:5601
Simplifying how to install GrimoireLab

By Georg Link

When I talk to people who tried to install GrimoireLab, I get one consistent answer. It is difficult and our GrimoireLab tutorial is too complicated. I believe this status quo is hurting the adoption of GrimoireLab software and CHA OSS metrics. This blog post is about how to make it easier for anyone to start using GrimoireLab.
Current Work
SortingHat (muggle branch)
Bestiary (unicorn branch)
Grimoirelab is an open-source toolset for software development analytics. Grimoirelab provides a set of tools to collect, analyze and visualize software development metrics from a variety of sources like Git, Jira, Confluence, Slack, etc. In order to manage the identities of people across these different sources, Grimoirelab developed Sorting Hat. Sorting Hat manages the identities of people and related metadata.

As part of the metadata collected around identities, Sorting Hat stores organizational information such as the name and domains related to the organization. This project aims to add to this information by extending the existing Organization model to capture the internal structure of organizations such as departments, sub-organizations, and teams. This will help in annotating the identity information more meaningfully.
Expanding and restyling the GrimoireLab tutorial

Project ID: 210190824

Selected Students: Veerasamy Sevagen
Degree of Difficulty: Medium
Language: English
Numbers of Applicants: 1

GrimoireLab is a powerful open-source platform that provides support for monitoring and in-depth analysis of software projects. It produces a rich set of metrics with data extracted from more than 30 tools related to contributing to Open Source development such as version control systems, issue trackers and forums. These metrics are shown and exploited on Web dynamic dashboards, which can be easily inspected by decision-makers to help them understand the evolution and health of their projects. The main entry point to learn about GrimoireLab is the tutorial which provides a walkthrough of the platform and its components.

Mentor of Project: Venu Vardhan

Contact Information: <venuvardhanreddytekula8@gmail.com>
How to Get Involved
Check out the CONTRIBUTING Guidelines.

Feel free to contact the maintainers.

github.com/chaoss/grimoirelab/blob/master/CONTRIBUTING.md
Join the community

github.com/chaoss/grimoirelab

#grimoirelab on CHA OSS Slack

@grimoirelab

chaoss.github.io/grimoirelab
You can also help us with the different working groups

chaoss.community/participate
CHAOSS

Questions?

Twitter: @vchrombie
Mail to: venu@chaoss.community

Thanks to Georg Link, Yehui Wang for the support and providing the resources.