ATS Summit Spring 2024 Open Source Repo Analysis

Kit Chan / kichan@apache.org

TODO group

- Main site - https://github.com/todogroup
- Book - https://ospobook.todogroup.org/
- Landscape map - https://landscape.todogroup.org/

CHAOSS / TODO group

- Organization recommending Software for analyzing open source repositories & define metrics / metric models for open source projects
- https://chaoss.community
- Open Source Tools
  - GrimoireLab
    - Collect data from variety of channels - i.e. git, slacks, emails, etc
    - pre-made visualizations
    - From Bitergia
    - An application on top of ElasticSearch
  - Augur / 8Knot
    - Data from github/git labs
    - 8Knot is the frontend. Augur is the DB.
    - customizable queries using python
    - An application on top of relational DB
- Free & Hosted
  - GrimoireLab - cauldron.io - ATS - https://cauldron.io/project/8176
  - 8Knot - eightknot.osci.io - ATS - eightknot.osci.io / metrix.chaoss.io

Categorizing Repositories: Maturity Model

- With maturity Model -
  https://github.com/DSACMS/repo-scaffolder/blob/main/maturity-model-tiers.pdf

File	Ordinality of Documentation By Tier (M - Mandatory, R - Recommended, N - Not Recommended)				
	Tier 0: Private Repo	Tier 1: One-time Release	Tier 2: Close Collaboration	Tier 3: Working in Public	Tier 4: Community Governance
LICENSE	M	M	M	M	M
SECURITY.md	N	M	M	M	M
README.md	M	M	M	M	M
CONTRIBUTING.md	R	R	M	M	M
MAINTAINERS.md	N	N	R	M	M
GOVERNANCE.md	N	N	N	R	M
CODEOWNERS.md	N	N	R	M	M
COMMUNITY_GUIDELINES.md	N	N	M	M	M
CODE_OF_CONDUCT.md	N	N	M	M	M

Categorizing Repositories: Maturity Model

- Tool to identify - https://github.com/DSACMS/repo-scaffolder/tree/main
  - With repolinter config file
- Lint tool (from TODO group) - https://github.com/todogroup/repolinter
  - Can be used to define "cohorts"
  - And used to categorize repositories
- More information on categorizing repositories
  - https://static.sched.com/hosted_files/ossna2024/20/repository_cohorts_ossna24.pdf
  - https://static.sched.com/hosted_files/ossna2024/d4/CMS_OSPO_Baseline_Presentation_OSSNA24.pdf

Demo

Items
Authors | 1 |

Creation date: 2nd May 2024

Items
Authors | 273 | 193 | 160 |

Creation date: 16th Feb 2024

Items
Authors | 124 | 85 | 85 |

Creation date: 16th Feb 2024

Arkime

Data sources 5 5

	Commits	Issues	Reviews
Items	6267	1447	1417
Authors	137	427	157

Creation date: 16th Feb 2024

Yavin

Data sources 6 6

	Commits	Issues	Reviews
Items	2494	436	1404
Authors	77	18	38

Creation date: 16th Feb 2024

Bullet-db

Data sources 15 15

	Commits	Issues	Reviews
Items	1506	117	517
Authors	20	6	12

Creation date: 16th Feb 2024

Yahoo

Data sources 195 195

	Commits	Issues	Reviews
Items	37931	4204	13383
Authors	2045	2094	1163

Creation date: 14th Feb 2024

Athenz

Data sources 17 17

	Commits	Issues	Reviews
Items	4208	314	2680
Authors	162	69	109

Creation date: 30th Jan 2024

ATS

Data sources 1 1

	Commits	Issues	Reviews
Items	20854	2158	9082
Authors	569	339	304

Creation date: 29th Jan 2024

+ New report

shukitchan

ATS

Fork report

Report information

Repositories: 2. **Git:** 1. **GitHub:** 1. **Running:** 0

Data from some repositories have not been refreshed since April 28, 2024. Refresh all data sources

Summary | Activity | Community | Performance | Repos | CHAOSS | More details
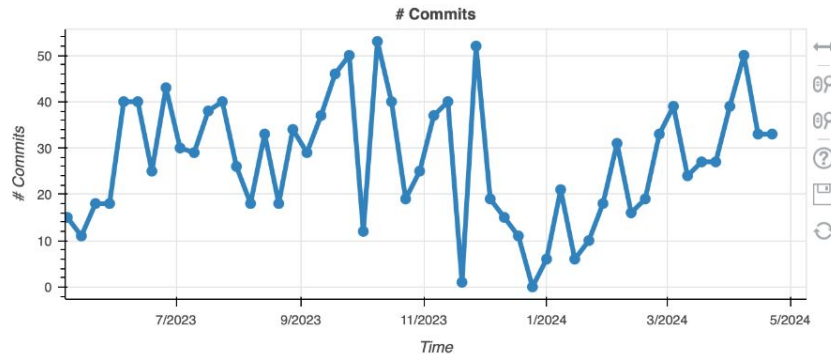
Select a repository... | Apply filter | 2023-05-08 - 2024-05-08

Activity

	Commits	Issues	Reviews
Total	20854	2158	9082
Total (Last year)	1394	335	1270
Year-over-year	64.58 %	101.81 %	84.33 %

Commits

Number of commits

Number of commits is the most common measure of activity in a git repository. To

Dashboard / Cauldron Public Dashboard

Full screen   Share   Clone   Reporting   Edit

Search

KQL   Last 6 months   Show dates   Refresh

NOT Empty Commits ×   + Add filter

Welcome

Public Overview Dashboard

This dashboard shows basic figures for git and Github data. There are more public dashboards available
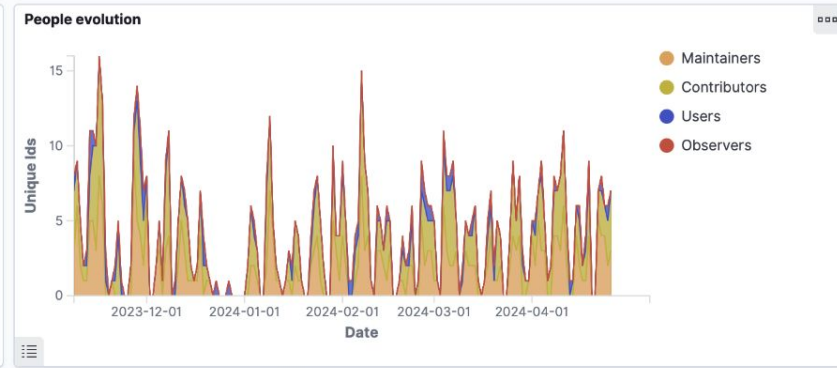
If you want to create your own dashboards, join Cauldron

If you need any help with this dashboard or Cauldron in general, join the Cauldron community and feel free to ask

People

30
Maintainers - Unique Ids

30
Contributors - Unique Ids
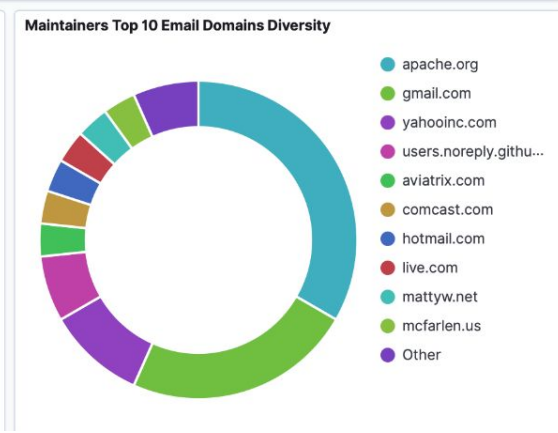
31
Users - Unique Ids

0
Observers - Unique Ids

People evolution

- Maintainers
- Contributors
- Users
- Observers

Code Activity

1
Repositories

588
Code commits

163,948
Added LoC

147,061
Removed LoC

Code commits

Maintainers Top 10 Email Domains Diversity

- apache.org
- gmail.com
- yahooinc.com
- users.noreply.githu...
- aviatrix.com
- comcast.com
- hotmail.com
- live.com
- mattyw.net
- mcfarlen.us
- Other
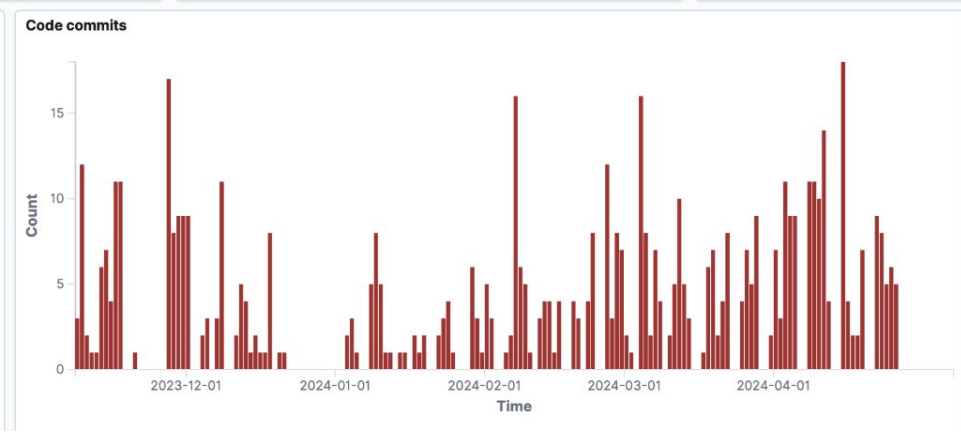
Select GitHub repos or orgs:

https://github.com/apache/trafficserver ✕

Search | Help | Repo List | ⬤ GitHub Bot Filter
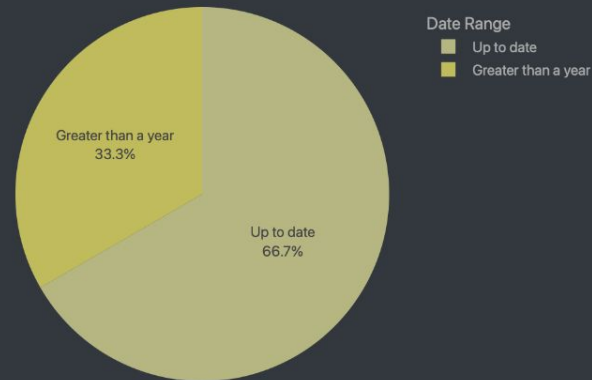
Data Ready

Search Bar Populated Analysis

File Language by File

C Header
23.1%

C++
28.2%

ReStructuredText
15.9%

Python
11%

YAML
5.49%

CMake
4.62%

Languages
- C++
- C Header
- ReStructuredText
- Python
- YAML
- CMake
- Shell
- Plain Text
- JSON
- Lua
- Markdown
- Autoconf
- Other
- C++ Header
- Perl
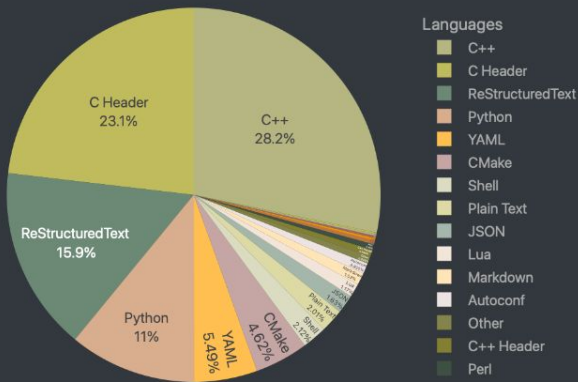
Graph View: ⬤ Files   ◯ Lines of Code

About Graph

Package Version Updates

Date Range
- Up to date
- Greater than a year

Greater than a year
33.3%

Up to date
66.7%

About Graph

OSSF Scorecard

Check Type	Score
Aggregate Score	7.2
Binary-Artifacts	9
Branch-Protection	8
CI-Tests	10
CII-Best-Practices	0
Code-Review	10
Contributors	10
Dangerous-Workflow	10
Dependency-Update-Tool	0
Fuzzing	10
License	10
Maintained	10
Packaging	-1
Pinned-Dependencies	0
SAST	7
Security-Policy	10
Signed-Releases	-1

Repo General Info

Section	Info
License	Apache License 2.0
Code of Conduct	File found
Contributor Guidelines	File found
Security Policy	File found
Number of Releases	0
Last Release Date	No Releases Found
Avg Time Between Releases	No Releases Found
Star Count	1727
Fork Count	777
Watcher Count	166
Issues Enabled	True