Continuous integration in a social-coding world

Bogdan Vasilescu, Stef van Schuylenburg, Jules Wulms, Alexander Serebrenik, Mark G. J. van den Brand
Integration

“[used to be] a long and unpredictable process” (Martin Fowler, 2000)
Continuous Integration

speeds up collaborative software development by reducing integration problems

integration with every commit
Continuous Integration

speeds up collaborative software development by reducing integration problems

integration with every commit

Report

Initiate CI process

Tests

Test

Build
“There are 7.1M people collaborating right now across 15.9M repositories on GitHub. Developers from all around the world are building amazing things together. Their story is our story.”

https://github.com/about/press

How are GitHub developers using CI?
How are GitHub developers using CI?

Most projects are configured to use CI, but less than half actually do.

(1) Most projects are configured to use CI, but less than half actually do.
GitHub workflow

Contributing to IronCat’s repository

Octocat

Jean-Luc Picat

Spocktocat

<created>

Repository

IronCat
GitHub workflow

Octocat

Collaborators

Jean-Luc Picat

Spocktocat

IronCat

Repository

<created>

<push> (commit)
GitHub workflow

I want to add a new feature / fix an issue but I don't have commit rights
GitHub workflow
I can implement the change in my own fork. But how to signal IronCat?
GitHub workflow

Can you please pull my change into your repository?

<pull request>
<push> (commit)
<created>
<forked>
<created>
GitHub workflow

Nice work! I’ll merge your change
GitHub workflow
GitHub workflow

Direct contributions (Collaborators)

Indirect contributions (Pull requests)
unprecedented low barrier to entry for potential contributors


scalability challenges when integrating (many) outside contributions

R. Pham, L. Singer, O. Liskin, K. Schneider et al., “Creating a shared understanding of testing culture on a social coding site,” in ICSE 2013
How are GitHub developers using CI?

(1) Most projects are configured to use CI, but less than half actually do.

(2) Pull requests are much more likely to result in successful builds than direct commits.
GHTorrent sample

G. Gousios, “The GHTorrent dataset and tool suite,” in MSR 2013

G. Gousios, B. Vasilescu, A. Serebrenik, A. Zaidman, “Lean GHTorrent: GitHub data on demand,” in MSR 2014

223 large and active repositories

* not forks
* ≥ 1 year
* both commits and pull requests
* Java, Python, or Ruby
* ≥ 10 changes last month
* ≥ 10 contributors
GitHub + Travis-CI

Octocat → Repository

IronCat → Repository

Jean-Luc Picat → Repository

Spocktocat → Repository

Direct contributions (Collaborators)

Indirect contributions (Pull requests)

@b_vasilescu @aserebrenik #icsme14
GitHub + Travis-CI

Octocat <push> Jean-Luc Picat <push>
<pull request> Spocktocat

GitHub + Travis-CI

Octocat
IronCat

Spocktocat

master - MySQL: set collation_connection to be the same as collation_database

Samuel Lebeau authored and committed
Commit 1538763

#13043: MySQL: set collation_connection to be the same as collation_database
ran for 31 min 56 sec
about 15 hours ago

master - Correctly generate default url for singular resource

Gert Goet authored and committed
Commit dd03371

#17066: Let's fix form_for for singular resources
ran for 7 hrs 15 min 45 sec
18 minutes ago
Methods and Data
Methods and Data
Methods and Data

- Travis CI configuration
- Methods and data analysis
- Programming languages: Java, Python, Ruby
- Travis.yml configuration
- Passes and failures
- Commits and pull requests
- Python and R project analysis

@b_vasilescu  @aserebrenik  #icsme14
Methods and Data

Distribution of p-values from the Chi^2 tests

Does type matter for passed/failed?

\( \chi^2 + \text{Stouffer} \)
Methods and Data

Which type succeeds more often?

Odds ratios
How are GitHub developers using CI?

(1) Most projects are configured to use CI, but less than half actually do.

(2) Pull requests are much more likely to result in successful builds than direct commits.