

Unit Testing Performance in Java Projects: Are We There Yet?

Petr Stefan, *Vojtěch Horký*, Lubomír Bulej and Petr Tůma

Charles University

Department of
Distributed and
Dependable
Systems



Unit Testing Performance in Java Projects: Are We There Yet?

Unit Testing Performance in Java Projects: Are We There Yet?

Short, isolated performance tests of individual components.
Suitable to be run during nightly builds.

Unit Testing Performance in **Java Projects**: Are We There Yet?

Where else :-)

Unit Testing Performance in Java Projects: Are We There Yet?

How many projects have such tests?

...

“Are We There Yet?” Expanded

- How many projects have performance unit tests?
- Which framework is used for these tests?
- What kind of projects care about performance?
- How often are performance unit tests changed (maintained)?
- Are the tests short enough for “after every commit” execution?
- What (de)motivates developers towards performance tests?

Setup

Getting the Data

Semi-automated exploration of open source projects.

Survey for developers that actually added/updated the performance tests.

Getting the Projects

GitHub as one of the prominent hosting platforms.

Almost 2.5 million of Java projects.

Getting the Projects

GitHub as one of the prominent hosting platforms.

Almost 2.5 million of Java projects.

We limit ourselves to projects with a fork to filter out abandoned repositories, toy examples and school assignments.

Getting the Projects

GitHub as one of the prominent hosting platforms.

Almost 2.5 million of Java projects.

We limit ourselves to projects with a fork to filter out abandoned repositories, toy examples and school assignments.

That totals to almost 100 thousand projects and over 3TB of data.

Developer's Survey

Targeted committers that modified the performance tests.

Developer's Survey

Targeted committers that modified the performance tests.

483 invitations, 111 responses.

Questions steered by results of source code analysis.

Results

Which Frameworks Do the Projects Use?

Methodology: parse the Java code, look for typical annotations or base classes . . .

Which Frameworks Do the Projects Use?

Methodology: parse the Java code, look for typical annotations or base classes ...

Example: this marks JMH (or Caliper) test (we use full-fledged parser to resolve ambiguities through imports):

```
@Benchmark public void x() { ... }
```


Which Frameworks Do the Projects Use?

Methodology: parse the Java code, look for typical annotations or base classes ...

Example: this marks JMH (or Caliper) test (we use full-fledged parser to resolve ambiguities through imports):

```
@Benchmark public void x() { ... }
```

We look for two functional testing frameworks and five performance testing frameworks.

Testing Frameworks

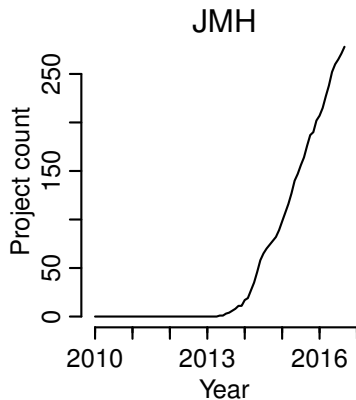
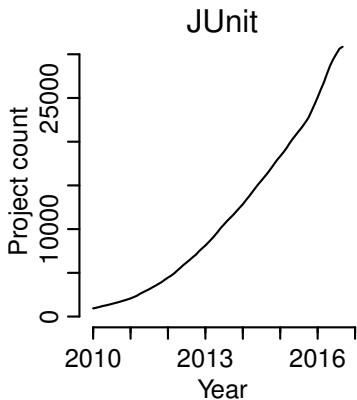
Framework	Repositories	Relative usage
JUnit 4	30871	31.177 %
TestNG	2053	2.073 %
Caliper	12	0.012 %
ContiPerf	17	0.017 %
Japex	52	0.053 %
JMH	278	0.281 %
JUnitPerf	11	0.011 %

Testing Frameworks

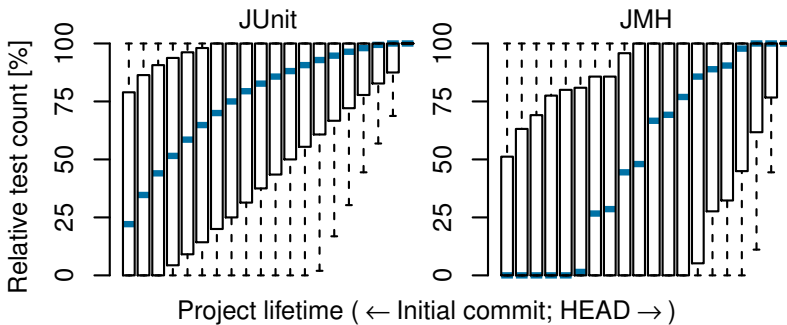
Framework	Repositories	Relative usage
JUnit 4	30871	31.177 %
TestNG	2053	2.073 %
Caliper	12	0.012 %
ContiPerf	17	0.017 %
Japex	52	0.053 %
JMH	278	0.281 %
JUnitPerf	11	0.011 %

From the survey: JMH is the most popular because developers trusts the results, like the documentation and active maintenance.

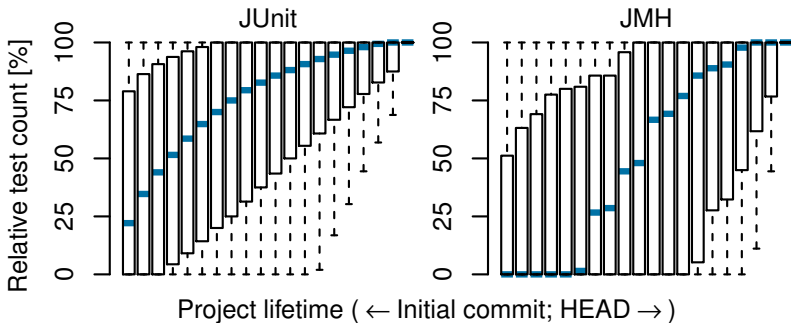
Project Count Over Time



When Tests Are Introduced



When Tests Are Introduced



From the survey: less than one third of developers maintains performance code regularly. Two thirds updates the tests only when addressing performance issues.

Home-baked Performance Testing

Not everybody is using testing frameworks . . .

Home-baked Performance Testing

Not everybody is using testing frameworks ...

1. Took 1 000 projects randomly.

Home-baked Performance Testing

Not everybody is using testing frameworks . . .

1. Took 1 000 projects randomly.
2. Look for code that queries system clock.

Home-baked Performance Testing

Not everybody is using testing frameworks ...

1. Took 1 000 projects randomly.
2. Look for code that queries system clock.
3. Determine whether it is used for benchmarking.

Home-baked Performance Testing

Not everybody is using testing frameworks . . .

1. Took 1 000 projects randomly.
2. Look for code that queries system clock.
3. Determine whether it is used for benchmarking.

About 3% of the projects use it for benchmarking
(we err on the optimistic side).

Exploring Projects with JMH Tests

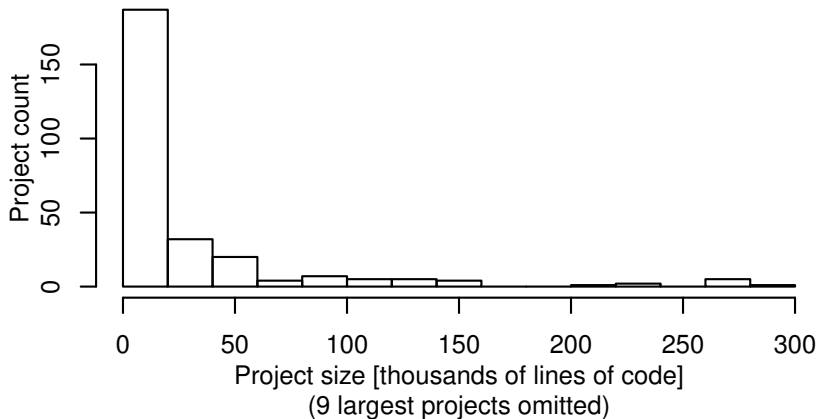
JMH-based projects only.

- Only framework with at least permille representation.
- Standardized execution.
- Apart from Caliper the only framework maintained till today.

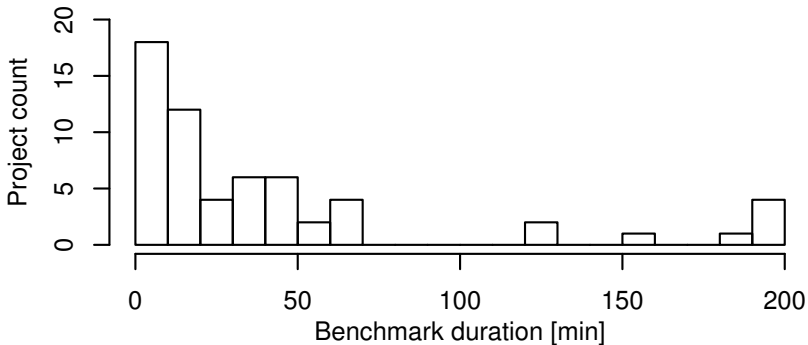
JMH Projects Classification

Category	Count
Database (ORM, SQL ...)	33
Tutorials and examples	30
Networking and distributed systems	29
Algorithms	27
Data structures	22
Object serialization, parsers (XML, JSON, ...)	22
Web frameworks or plugins	18

Size of Projects Using JMH

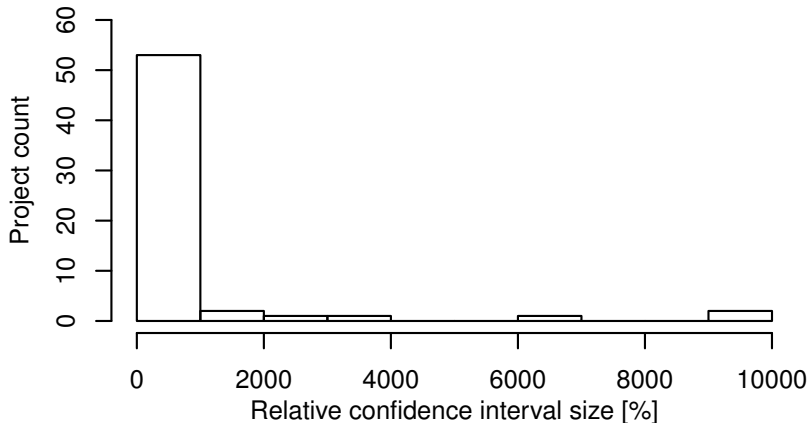


Can We Run the Tests After Each Commit?

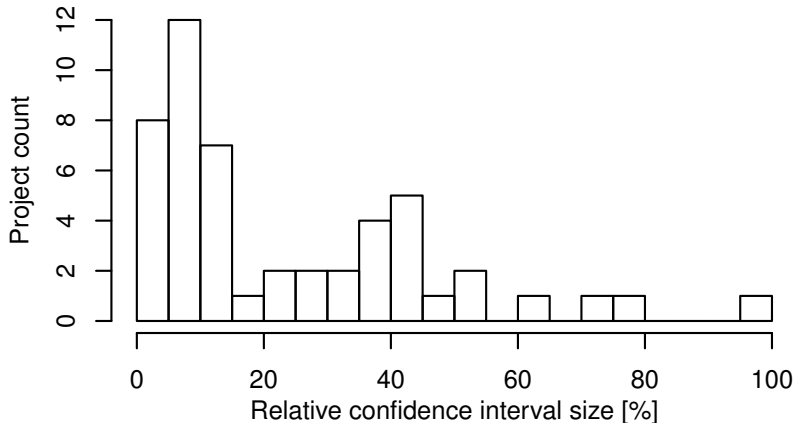


From the survey: less than one half of developers run the tests regularly.

One Hour Budget: Accuracy of Performance Data



One Hour Budget: Accuracy of Performance Data



Greater Adoption of Performance Tests

In the survey, we have learned that

- not enough time/money to maintain the tests (27%)

Greater Adoption of Performance Tests

In the survey, we have learned that

- not enough time/money to maintain the tests (27%)
- regular performance testing is rare (42%)

Greater Adoption of Performance Tests

In the survey, we have learned that

- not enough time/money to maintain the tests (27%)
- regular performance testing is rare (42%)
- tool integration with build infrastructure is too complex (50%)

Greater Adoption of Performance Tests

In the survey, we have learned that

- not enough time/money to maintain the tests (27%)
- regular performance testing is rare (42%)
- tool integration with build infrastructure is too complex (50%)
- developers miss (simple) automated evaluation (60%)

Greater Adoption of Performance Tests

In the survey, we have learned that

- not enough time/money to maintain the tests (27%)
- regular performance testing is rare (42%)
- tool integration with build infrastructure is too complex (50%)
- developers miss (simple) automated evaluation (60%)

Towards Simple Evaluation

From the survey: for one third of developers JMH is complex.

Towards Simple Evaluation

From the survey: for one third of developers JMH is complex.

Our extension to JMH

- To JMH itself: dump raw data (more of them).
- New Maven plugin: compares two runs of a JMH benchmark to simplify regression benchmarking.

Unit Testing Performance in Java Projects: Are We There Yet?

Unit Testing Performance in Java Projects: Are We There Yet?

Not really ;-)

Unit Testing Performance in Java Projects: Are We There Yet?

Performance tests are in less than 3% of Java projects
(functional tests are present in about one third).

Unit Testing Performance in Java Projects: Are We There Yet?

Performance tests are in less than 3% of Java projects
(functional tests are present in about one third).

The existing tests are not suitable for build-time testing
(complex setup, limited budget, little pressure to do so).

Unit Testing Performance in Java Projects: Are We There Yet?

Performance tests are in less than 3% of Java projects
(functional tests are present in about one third).

The existing tests are not suitable for build-time testing
(complex setup, limited budget, little pressure to do so).

Complexity of the benchmarking tools and a limited budget
seems to be the biggest obstacles.

Unit Testing Performance in Java Projects: Are We There Yet?

Performance tests are in less than 3% of Java projects
(functional tests are present in about one third).

The existing tests are not suitable for build-time testing
(complex setup, limited budget, little pressure to do so).

Complexity of the benchmarking tools and a limited budget
seems to be the biggest obstacles.

<http://d3s.mff.cuni.cz/resources/icpe2017/>

Thank you!