Towards the Use of the Readily Available Tests from the Release Pipeline as Performance Tests. Are We There Yet?

Zishuo Ding

Jinfu Chen

Weiyi Shang

UNIVERSITÉ Concordia UNIVERSITY


PERFORMANCE

The short-battery-life issue of Surface Pro tablet was due to a software performance bug

Performance issues

Make changes

Too Expensive!

Analyze

Build and deploy

Testing environment

Produce at run-time

Testing is important to detect performance issues.

Functional tests are widely available and are executed often.

More than **4,000** functional tests available for *Hadoop*

These tests are executed on a **daily** basis

Can we use the readily available tests to evaluate performance?

Case study: Subject systems

指 Hadoop、HDFS、Yarn、MapReduce 四个软件

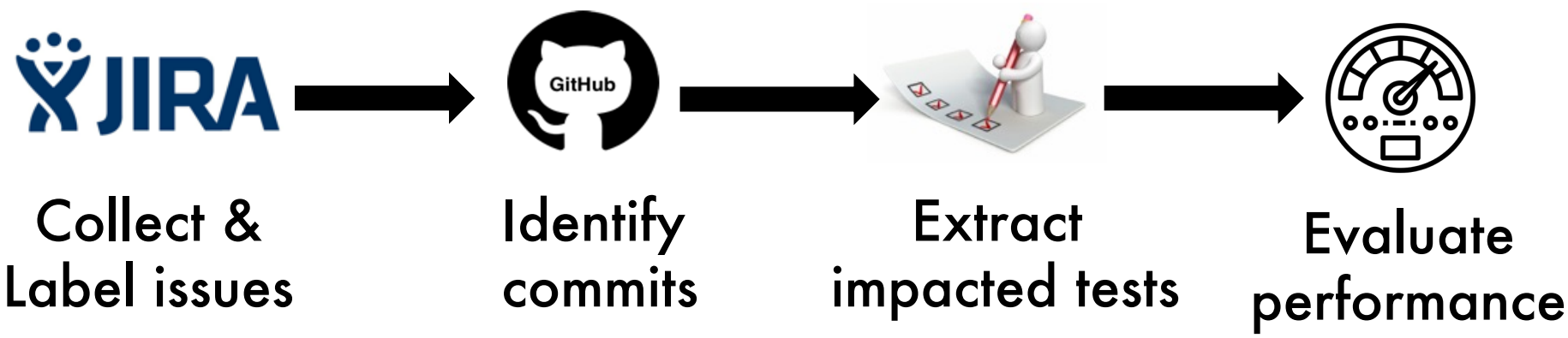
Hadoop

Distributed computing framework.
> 2.6M LOC

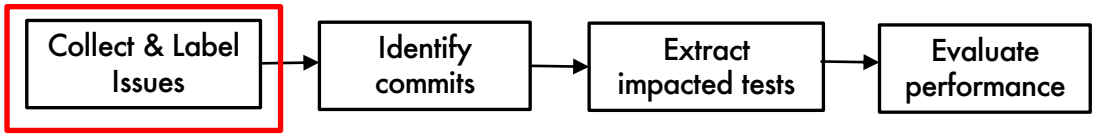
Cassandra

NoSQL database.
358K LOC

Our approach to collect performance data

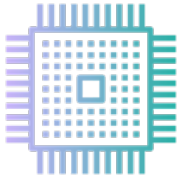


Step 1: Collect & label issues



Collect the performance issues and manually label each performance issue with performance metrics.


Response time


CPU


Memory

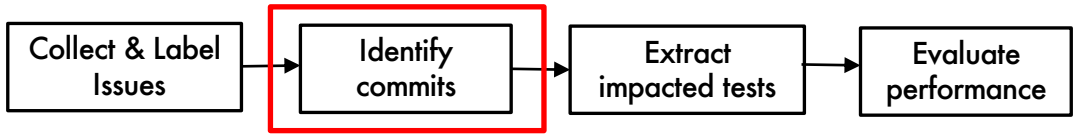

IO write


IO read

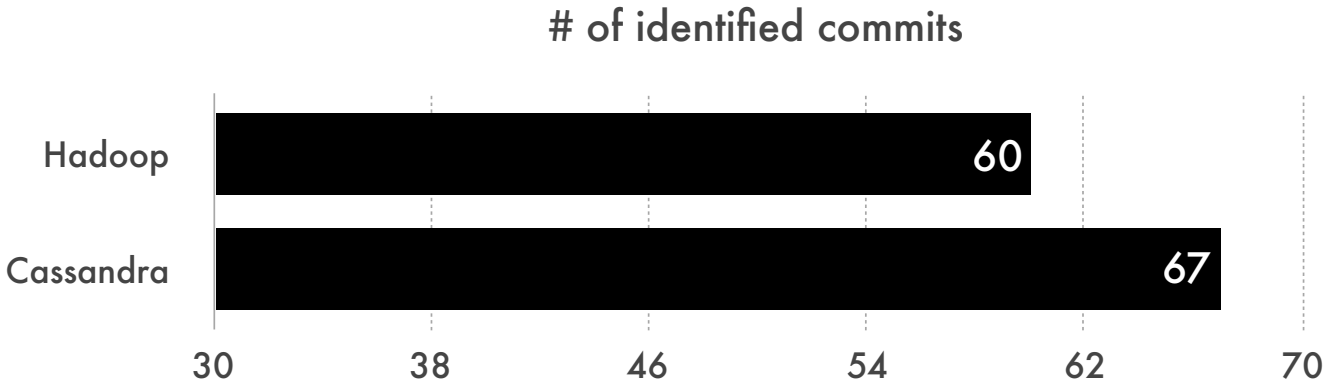
Cassandra / CASSANDRA-1246

Hadoop output SlicePredicate is slow and doesn't work as intended

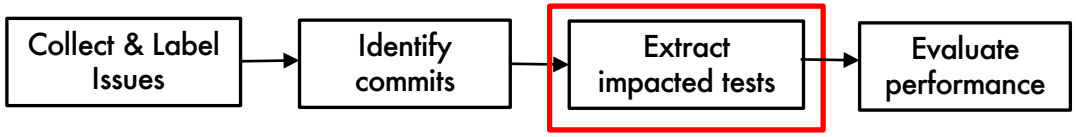
Step 2: Identify Commits



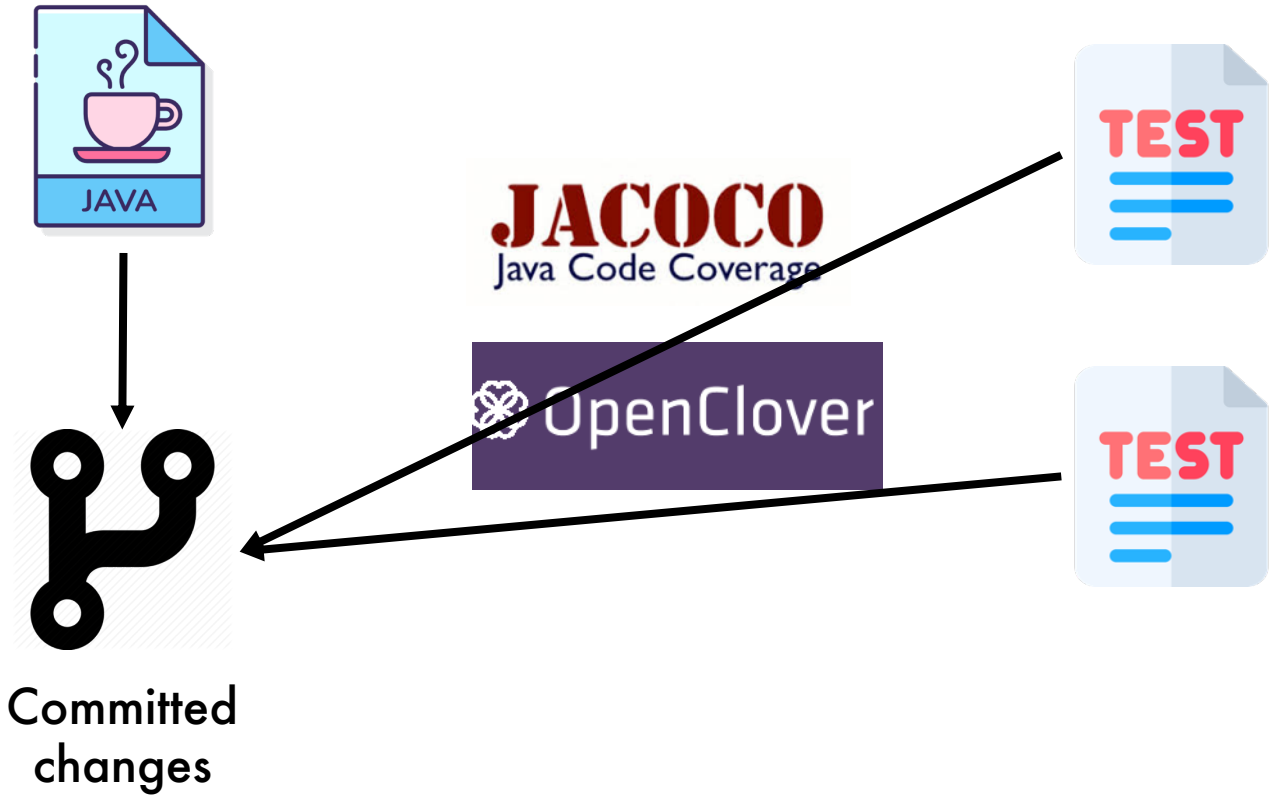
Identify the code commits that fix the collected performance issues.



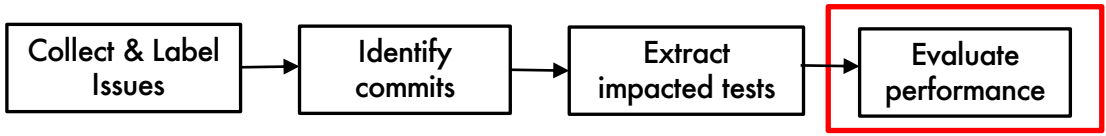
Step 3: Extract impacted tests



We first calculate the test coverage and the tests that cover changed lines are **impacted** tests in each commit.



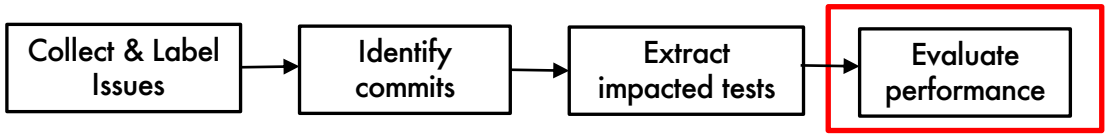
Step 4: Evaluate performance



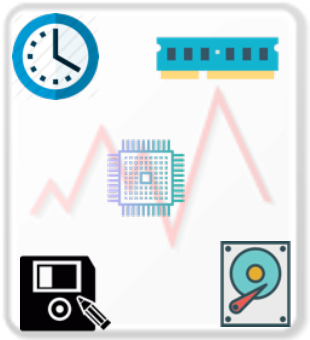
Execute the impacted tests and collect performance of each test.

psutil

Step 4: Evaluate performance



We perform statistical analyses to determine the existence and the magnitude of performance improvement.



psutil

Performance **before** fix

Performance **after** fix

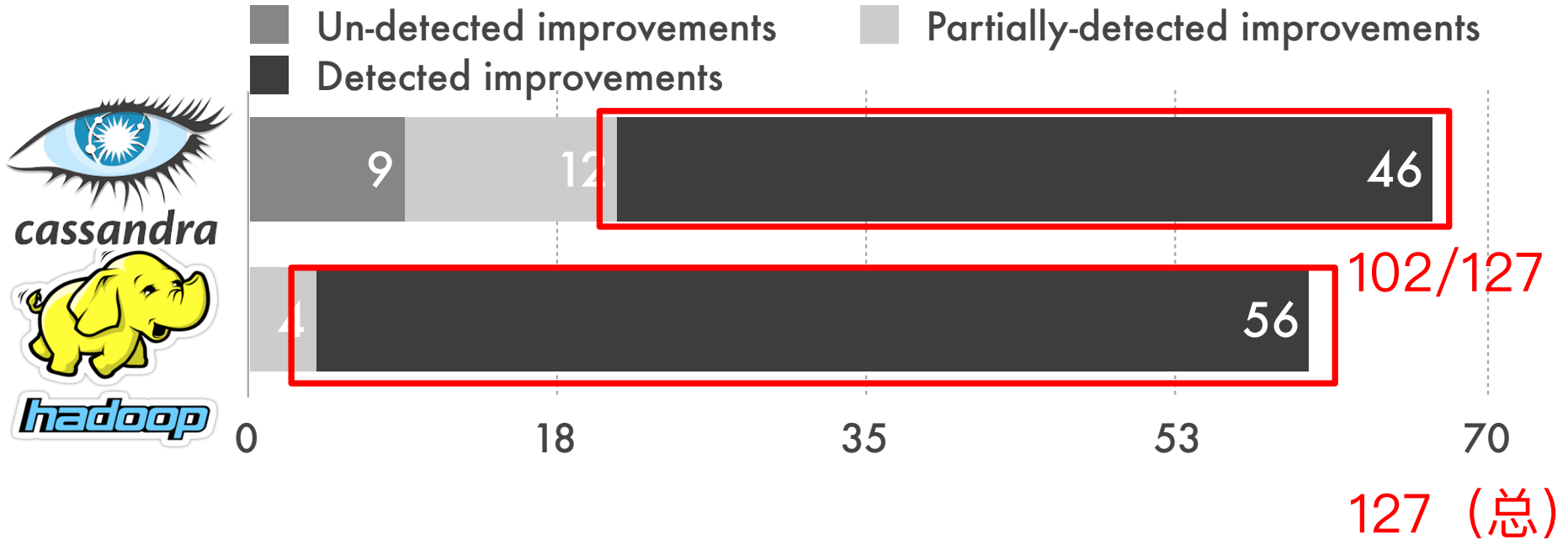
Mann–Whitney U test
Cliff's Delta

P-value
Effect sizes

RQ1: Can the readily available tests from the release pipeline demonstrate performance improvements from performance issues fixes?

Most of the performance improvements after an issue fix can be demonstrated by at least one test.

127 / 163 的issue至少有一个test case能够覆盖到fix commit改的代码

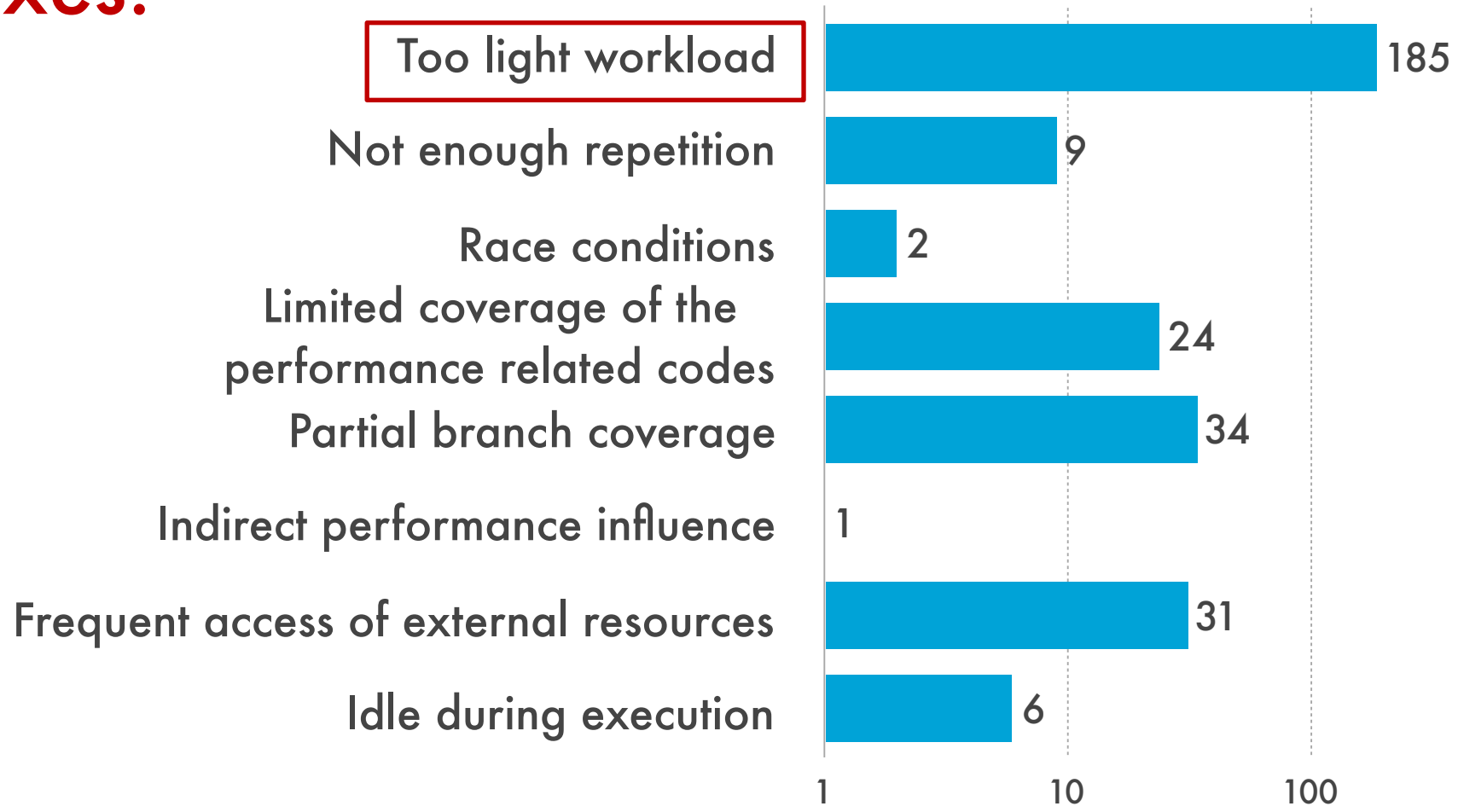


However, only a very small portion of the tests can demonstrate the performance improvements.

On the other hand, 76.9% and 74.3% of the tests in Cassandra and Hadoop, CANNOT demonstrate any performance improvement, even though these tests all execute the changed source code for the issue fixes.

RQ2: What are the reasons that some tests in the release pipeline cannot be used as performance tests?

We identify eight reasons that a test cannot demonstrate performance improvements from a performance issue fixes.



We identify eight reasons that a test cannot demonstrate performance

Cassandra / CASSANDRA-581

RandomPartitioner convertFromDiskFormat is slow

Description

convertFromDiskFormat in RandomPartitioner is slow. It uses split. We were testing with 1000+ keys using multi-get on a local node. We saw on average 200ms~, with the applied patch it went down to 76ms~.

RQ3: What are the important factors for a test to be useful as a performance test?

We build random forest classifiers to understand the importance of metrics.

Test code

Code committed for issues fix

JACOCO
Java Code Coverage

OpenClover

Code covered by the test

We build random forest classifiers to understand the importance of metrics.

Practitioners should focus on designing and selecting the tests.

Test code

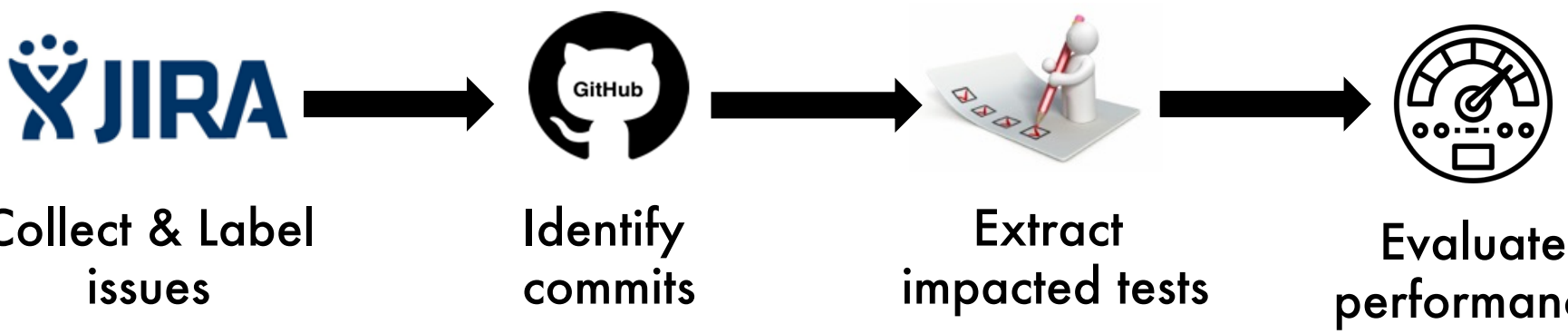
Code committed for issues fix

Code covered by the test

Can we use the readily available tests to evaluate performance?

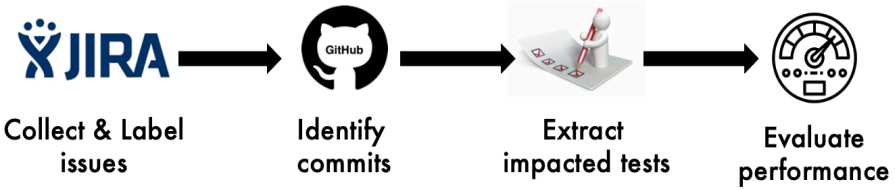
Can we use the readily available tests to evaluate performance?

Our approach to collect performance data



Can we use the readily available tests to evaluate performance?

Our approach to collect performance data



Research questions

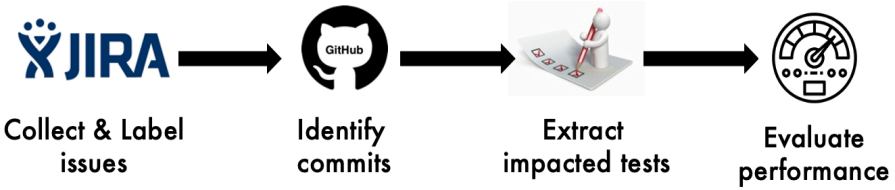
RQ1: Can the readily available tests from the release pipeline demonstrate performance improvements from performance issues fixes?

RQ2: What are the reasons that some tests in the release pipeline cannot be used as performance tests?

RQ3: What are the important factors for a test to be useful as a performance test?

Can we use the readily available tests to evaluate performance?

Our approach to collect performance data



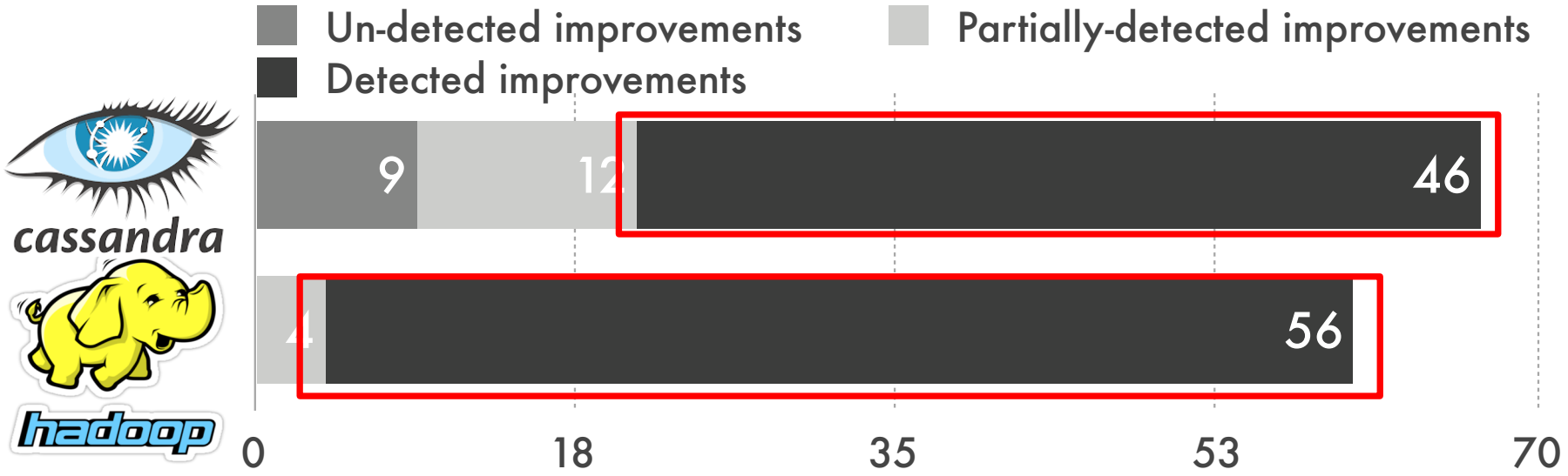
Research questions

RQ1: Can the readily available tests from the release pipeline demonstrate performance improvements from performance issues fixes?

RQ2: What are the reasons that some tests in the release pipeline cannot be used as performance tests?

RQ3: What are the important factors for a test to be useful as a performance test?

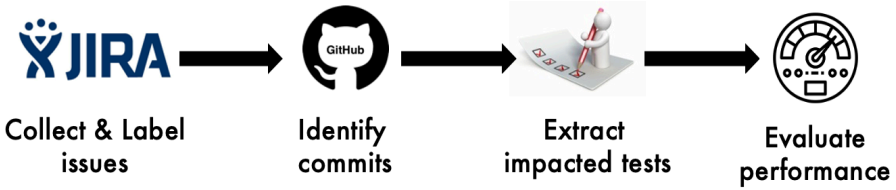
Most of the performance improvements after an issue fix can be demonstrated by at least one test.



However, only a very small portion of the tests can demonstrate the performance improvements.

Can we use the readily available tests to evaluate performance?

Our approach to collect performance data



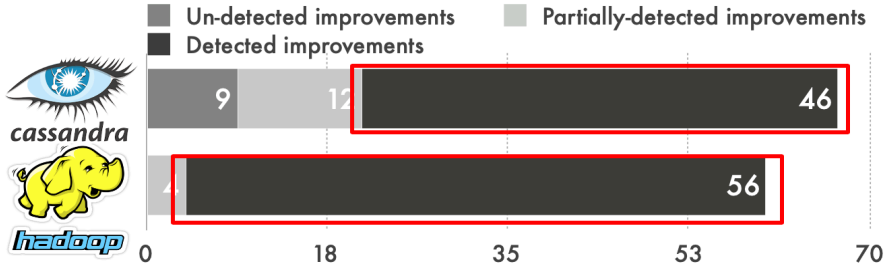
Research questions

RQ1: Can the readily available tests from the release pipeline demonstrate performance improvements from performance issues fixes?

RQ2: What are the reasons that some tests in the release pipeline cannot be used as performance tests?

RQ3: What are the important factors for a test to be useful as a performance test?

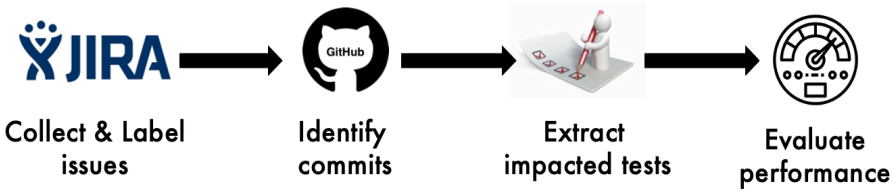
Most of the performance improvements after an issue fix can be demonstrated by at least one test.



However, only a very small portion of the tests can demonstrate the performance improvements.

Can we use the readily available tests to evaluate performance?

Our approach to collect performance data



Zishuo Ding
zi_ding@encs.concordia.ca

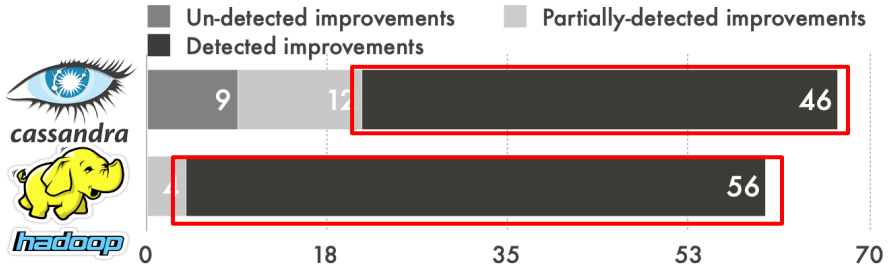
Rese

oy

release pipeline demonstrate performance improvements from performance issues fixes?

RQ2: What are the reasons that some tests in the release pipeline cannot be used as performance tests?

RQ3: What are the important factors for a test to be useful as a performance test?



However, only a very small portion of the tests can demonstrate the performance improvements.