



Testing on Thin Ice

Chipping Away at Test
Unpredictability

LOTS OF RERUN TESTS



Brian Demers

Java Champion
Developer Advocate

 @bdemers



François Martin

Senior Full Stack
Software Engineer

 @martinfrancois



\$36 Billion/year (USD)

Loss Productivity

\$40 Billion/year (USD)

End World Hunger
by 2030

Topics

What is a Flaky Test?

How to find them

How to deal with them

How to fix them



Slides & eBook

What is a Flaky Test?

No Flake News!



Definition

Test that exhibits both a passing and failing test result with the same code if run multiple times



Timing related

Example:

1. Enter search term into textbox
2. Click on button "Search"
3. **Wait for one second**
4. **Click on first result**

Test data related

Examples:

- Input field for prename doesn't allow special characters
 - Some randomly generated prenames in tests contain special characters ("François")
- Tests reliant on predefined test data in the database
 - Dependencies between tests
 - Failing tests cause inconsistencies



Infrastructure related

Unstable infrastructure

Tests time out

Most difficult to reproduce

Examples:

- Build runners under heavy load
- Microservices with networking issues

Resource contention related

Mostly when running tests in parallel

Examples: Multiple tests trying to...

- ... write to the same file at the same time
 - File lock
 - Concurrent modification errors
- ... create a user with the same username at the same time
 - Unique constraint violations



Why Worry about Flaky Tests?

- Re-running tests wastes a lot of time
- Repetitive investigation efforts to determine if flaky or legitimate failure
- Team starts losing trust in test results
- Bad software quality
- Pilots frequently ignoring an alarm in cockpits of a Boeing 737 resulted in fatal plane crash in 2005

Unit Tests vs Integration Tests

UT vs IT



End-to-End Tests

Integration Tests

Unit Tests

Testing Pyramid

An iceberg floating in the ocean. The tip of the iceberg, which is above the water line, is relatively small and jagged. The much larger part of the iceberg is submerged below the water line, showing a complex, layered structure with many cracks and crevices. The water is a deep blue color. Three white rectangular boxes with black text are overlaid on the image, pointing to different parts of the iceberg.

Unit Tests

Integration Tests

Flaky Foundations

End-to-End Tests

Complexity

Unit Tests

Integration Tests

End-to-End Tests



Flakiness

How to find Flaky Tests

Flake Busters

Running tests in parallel

Can reveal many issues:

- Resource contention
- Dependency on test data
- Dependency on test execution order

Side benefit: Tests execute faster



- Playwright & WebdriverIO: parallel by default
- Cypress: only in cloud (paid)
- JUnit and Gradle: through configuration

Rerun Tests

Most test frameworks allow for executing a test multiple times.

Slides & eBook



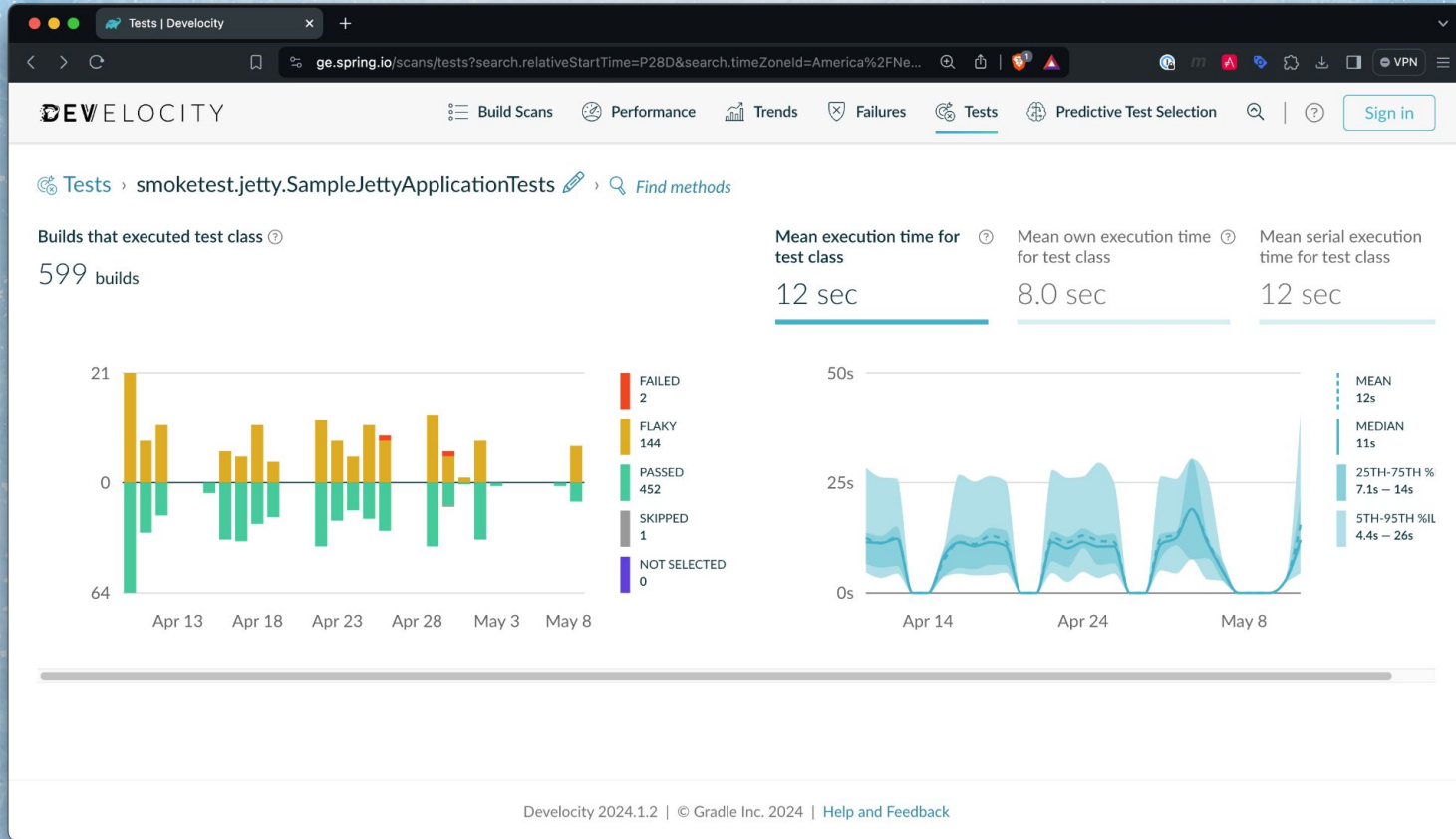
Maven Surefire Plugin

```
<plugin>
  <groupId>org.apache.maven.plugins</groupId>
  <artifactId>maven-surefire-plugin</artifactId>
  <configuration>
    <rerunFailingTestsCount>3</rerunFailingTestsCount>
    <failOnFlakeCount>0</failOnFlakeCount>
  </configuration>
</plugin>
```



Rerun Gotcha

Increased time to fail





Dealing with Flaky Tests



Prevention

In Pull / Merge Requests, run new or changed tests 3, 10, 50, or 1000 times (depending on resources and test type)

Alternatively run all tests multiple times nightly



Tell Your Team!

Don't let them waste time

Uncover deeper issues

Saves morale

New Sprint Checklist

- Fix a flaky test
- Pull a new issue



Flaky Test Days

Temporarily Quarantine Tests

```
// JUnit
@Disabled("quarantine: reason or link to issue here")
@Test
void flakyTest() {
}

// Jest
// quarantine: reason or link to issue here
it.skip('should throw an error', () => {
  expect(response).toThrowError(expected_error);
});
```



Quarantine

- Quarantine test after first failure
- Fix the quarantined tests timely
- Set a limit for tests in quarantine
- Define time limit for quarantined tests

Delete Flaky Tests

Does the test provide value?

Can it be fixed?

Keeping a flaky test provides negative value

SSH into CI Agent

```
$ ssh ci-agent  
  
cd /your-project  
  
# rerun build  
./mvnw clean install
```

How to fix Flaky Tests

Defeating Murphy's Law

Test Data Consistency

```
// JUnit 5
```

```
@BeforeEach  
void setup() {  
    // Create test data  
}
```

```
@AfterEach  
void teardown() {  
    // Delete test data  
}
```

```
// Jest
```

```
beforeEach(() => {  
    // Create test data  
});
```

```
afterEach(() => {  
    // Delete test data  
});
```


Test Data Consistency

// Using Testcontainers

```
// build.gradle
testImplementation "org.testcontainers:postgres:1.19.8"
testImplementation "org.testcontainers:junit-jupiter:1.19.8"

// JUnit Test
@Testcontainers
public class IntegrationTest {
    @Container
    PostgreSQLContainer<?> postgresContainer =
        new PostgreSQLContainer<>("postgres:16.3");
}
```



Isolate Tests

Don't Reuse System Resources

- Setting System Properties

```
System.setProperty("key", "value");
```

- Use unique files for each test

temp directories, output files, etc.

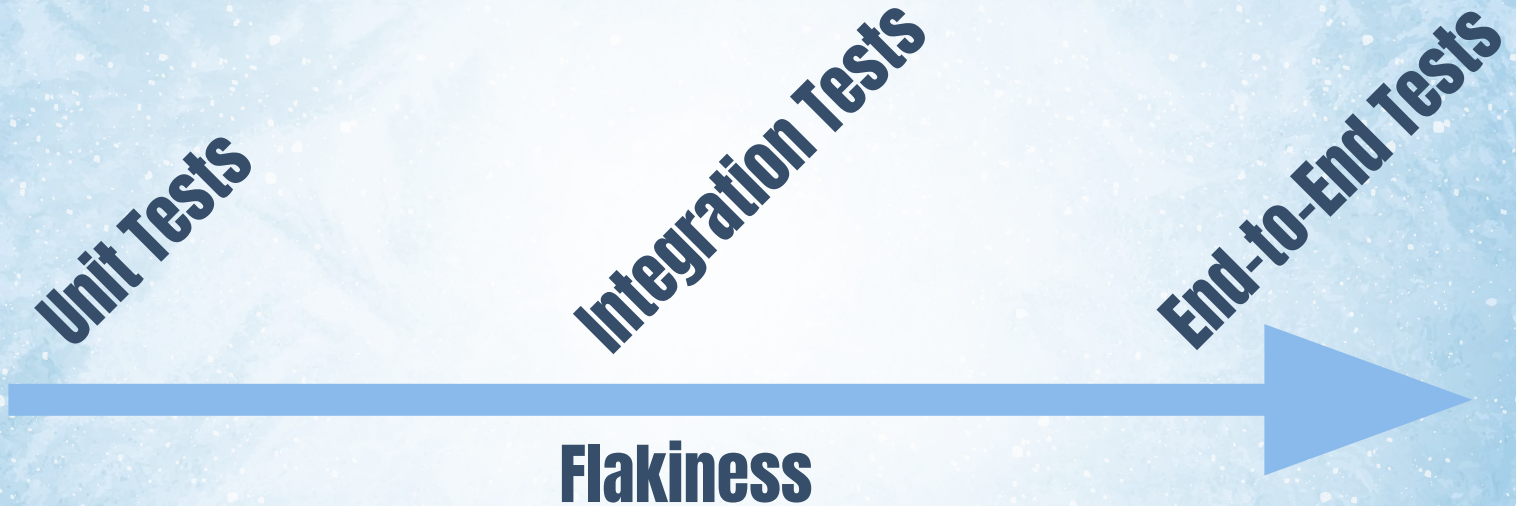
- Use a random **free** ephemeral port (49152 – 65535)

```
new ServerSocket(0).getLocalPort()
```

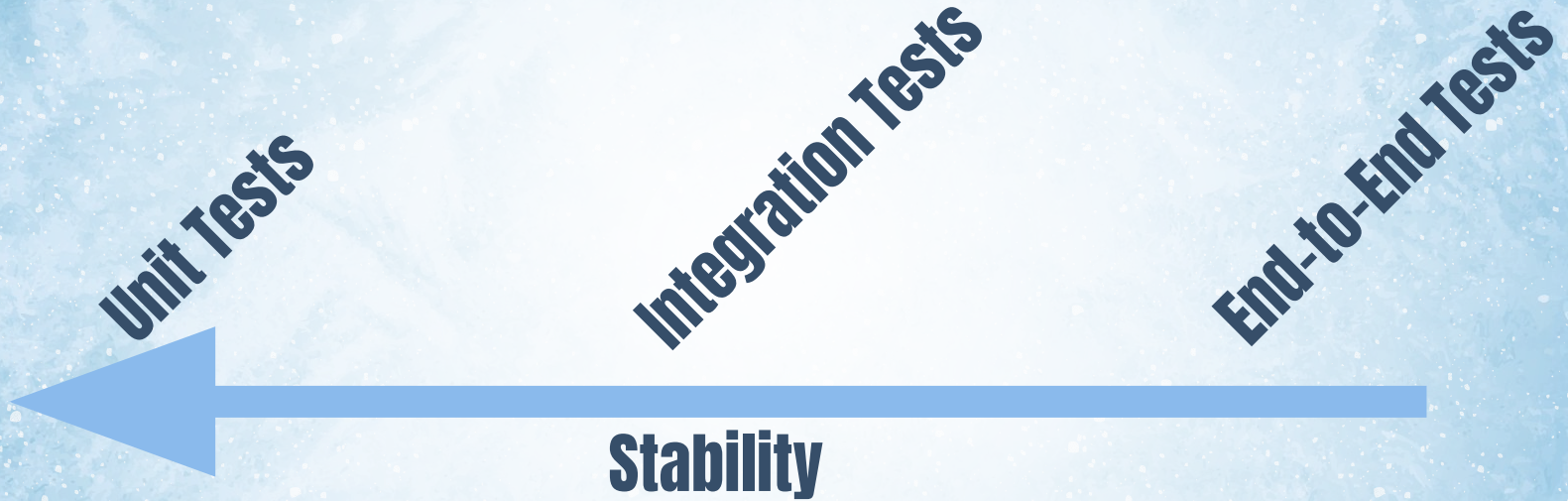
In E2E Tests

- Use deep links to navigate directly to the page to be tested
- Start the test with the test user already logged in
- Make it possible to run each test spec independently
- Do not wait in predefined time intervals, wait for a condition to be fulfilled with a timeout
- Use stable test ids in selectors
- Consider mocking external APIs


Convert ITs and E2E into UTs



Convert ITs and E2E into UTs



Fix the Culture of Flaky Tests

- Brown Bag / Lunch & Learn
- Make flaky tests visible
- Track them
- Use manager terms 

Slides & eBook



THANKS!

QUESTIONS?



bdemers.io
fmartin.ch



bdemers
FrançoisMartin

CREDITS: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#), and infographics & images by [Freepik](#) and images generated by OpenAI's DALL-E model via ChatGPT