

# CS 4530 Software Engineering

## Module 14: Continuous Development Processes

---

Adeel Bhutta, Jan Vitek and Mitch Wand

Khoury College of Computer Sciences

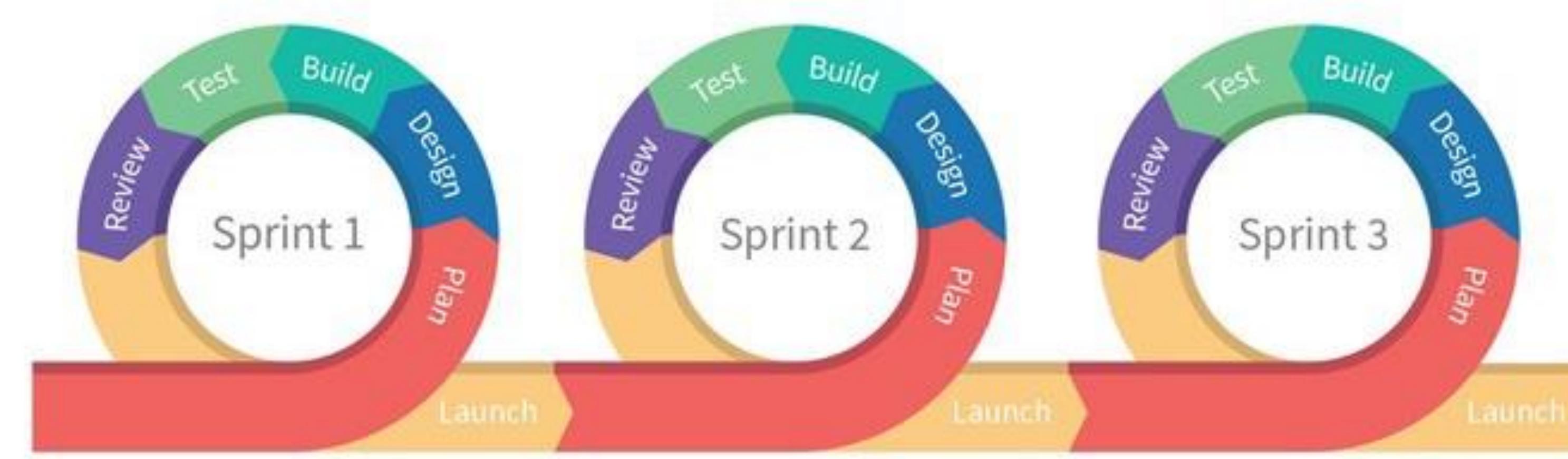
# Learning objectives for this lesson

By the end of this lesson, you should be able to...

- Describe the benefits of a culture of code review
- Describe how code review and continuous integration help to catch errors sooner in the software lifecycle
- Describe strategies for performing quality-assurance on software as and after it is delivered

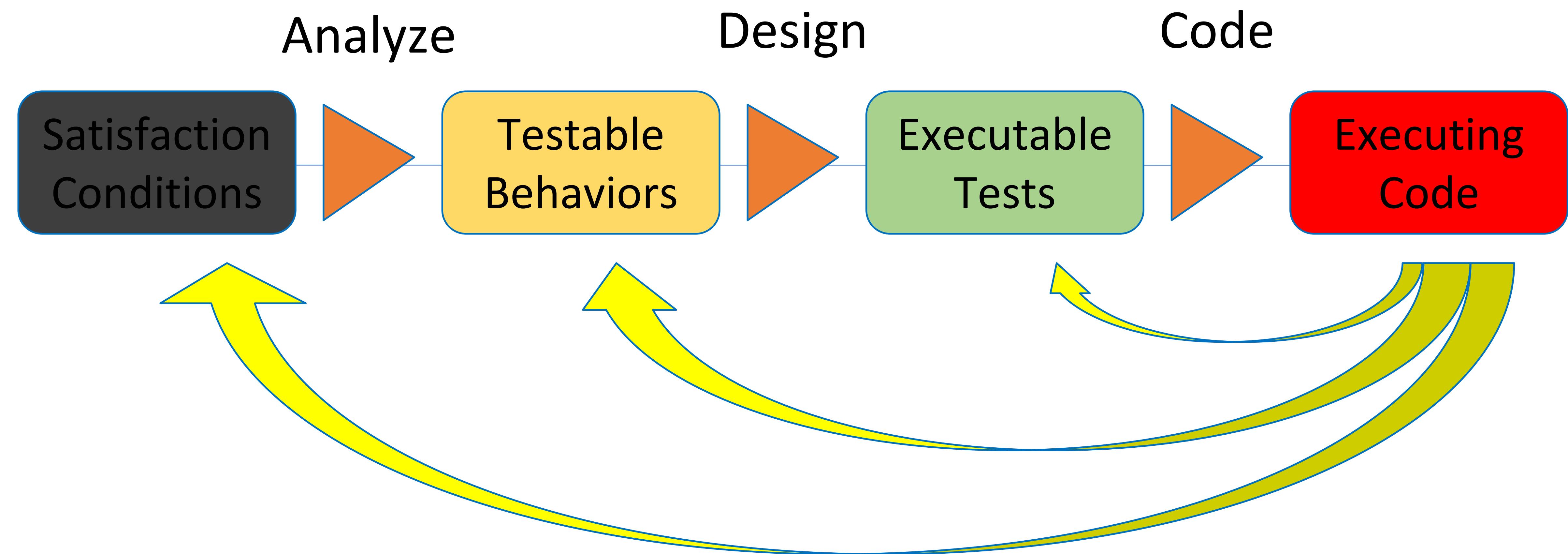
# The Agile Cycle (from Module 06)

---

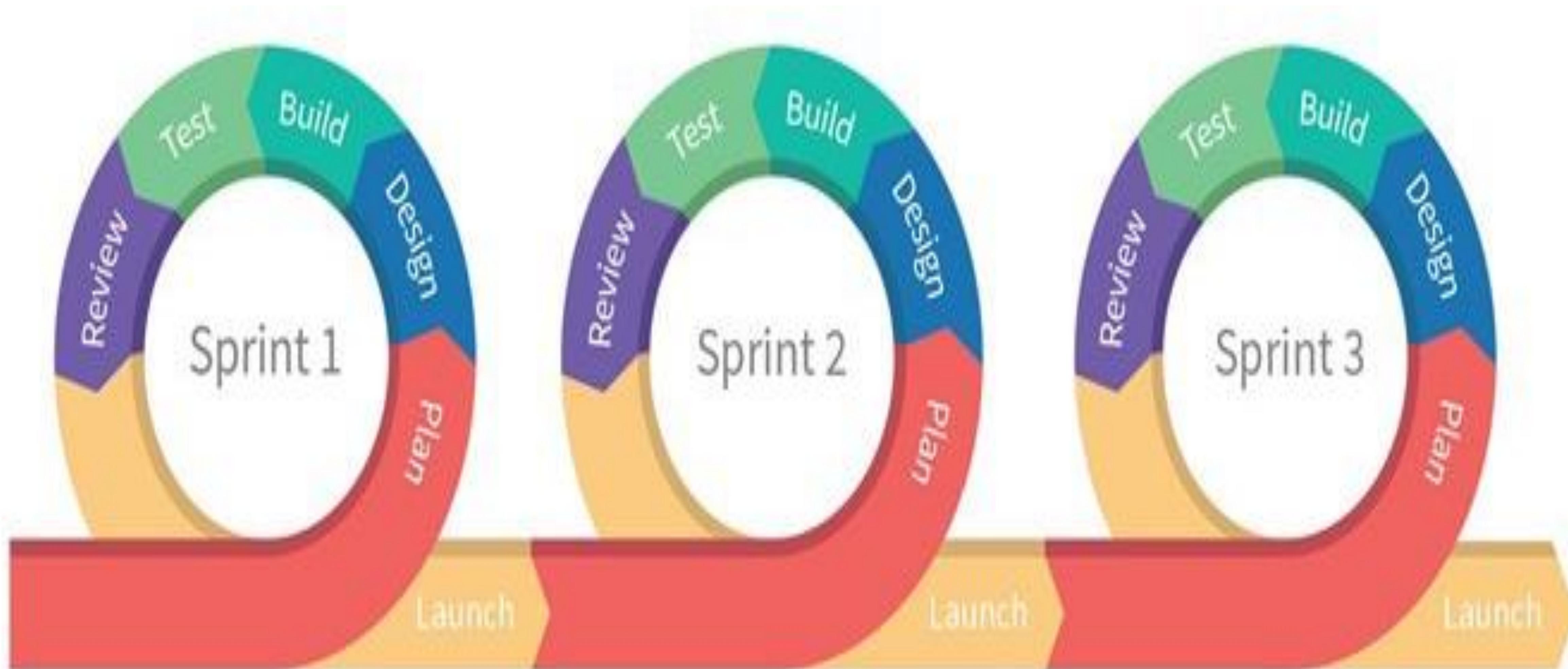


**Key Idea:** Small Continuous Releases

# The TDD Cycle (from Module 02)



# Let's look at the Agile Cycle a little more closely



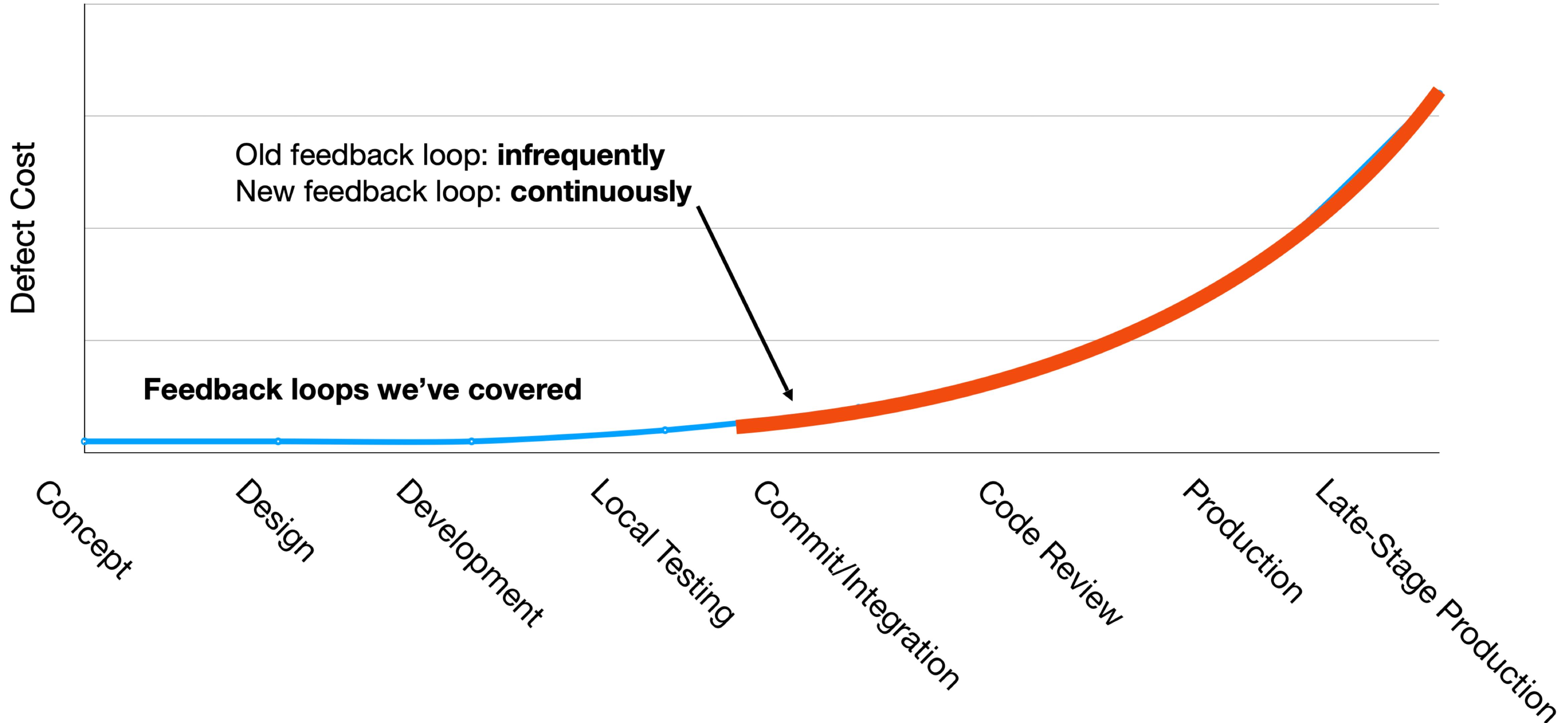
# How does the TDD cycle fit?

---

- TDD Covers:
  - Plan
  - Design
  - Build
  - Test (though not in the order in the picture....)
- Today we'll talk about the last two steps in this picture
  - Review
  - Release

# Agile values fast quality feedback loops

Faster feedback = lower cost to fix bugs



# What is a code review?

---

A code review is a formal process in which the code's author explains it to peers:

- When should it be done?
- What should it do?
- How does it do it?
- How confident are we in it?
- What are results of running tests?

A code review often concerns a code change (“diff”)

# Code review: example on pull request

...re-api/src/main/java/org/apache/maven/surefire/booter/CommandReader.java Hide resolved

```
        case BYE_ACK:  
            //After SHUTDOWN no more commands can come. Hence, do NOT go back to blocking in IO  
            callListeners( command );  
            return;  
        default:  
            callListeners( command );
```

 Tibor17 on Nov 12, 2019 Contributor ...  
The listeners are called here. But we can put IF condition:  
`IF BYE_ACK -> return` at the end of the default case.

 Tibor17 on Nov 12, 2019 Contributor ...  
Instead of calling the `return` we can make softer exit with `CommandReader.this.state.set(TERMINATED)`.

 eolivelli on Dec 17, 2019 Contributor ...  
Yes, I came to this same conclusion, change the state to TERMINATED.

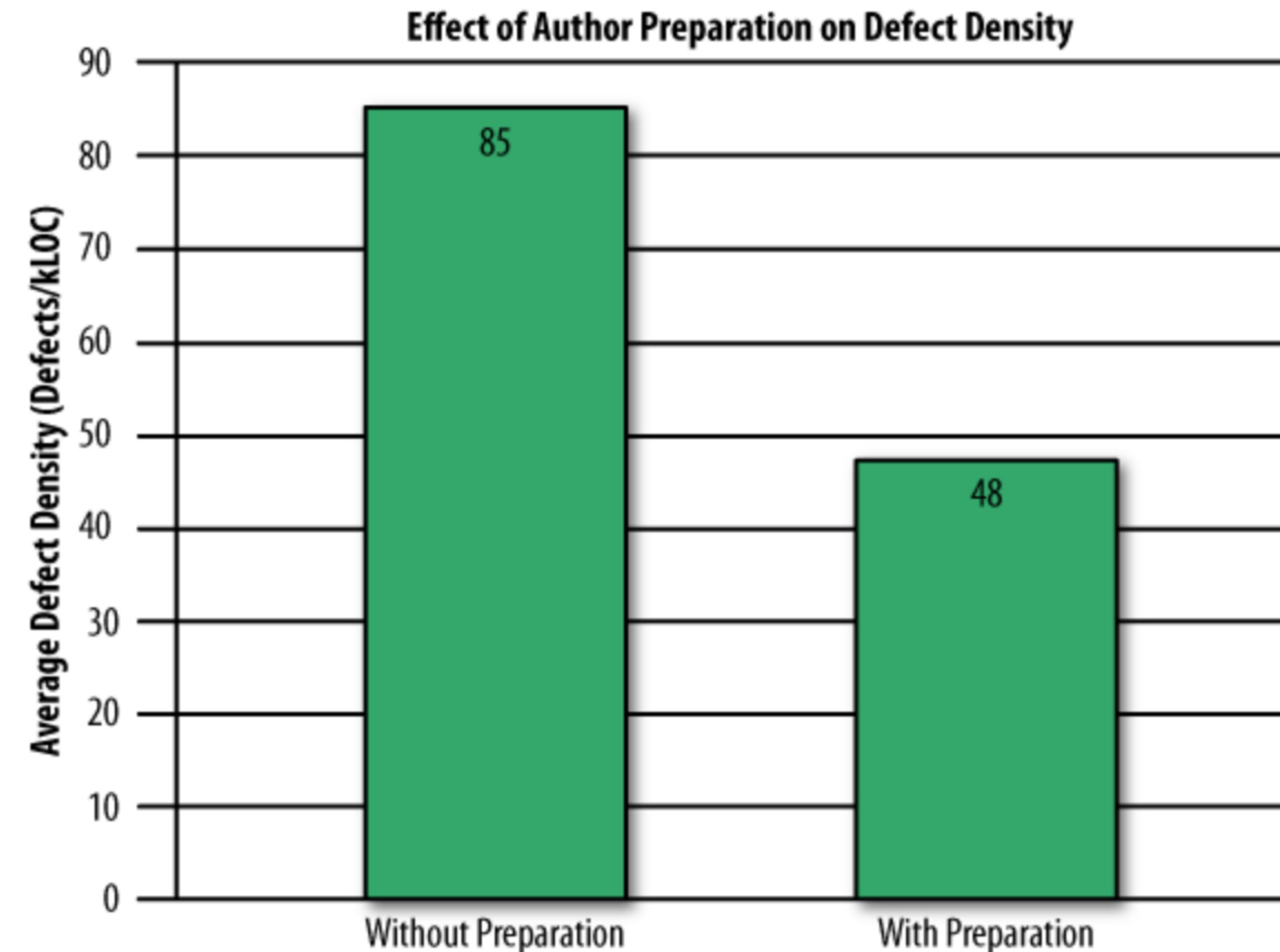
 jon-bell on Dec 19, 2019 Author Contributor ...  
Changed.

 Reply...

Unresolve conversation jon-bell marked this conversation as resolved.

# Self-review is no substitute for peer review

Study of 300 reviews at Cisco in 2006



Even if developers pre-review their code, many defects still found in peer review

# Code Review vs Code Inspection

---

Formal process of reading through code as a group;

- Applied to all project documents;
- A 3-5 person team reads the code aloud and explains what is being done;
- Each person has a specific role (moderator, reviewer, reader, scribe, observer, author)
- Usually a 60 minute meeting;
- Less efficient (defects/cost) than modern review processes.
- Very waterfall.
- Traceable, measurable

# Why should we perform code review?

Code review increases breadth of knowledge of code:

- Other people "know" the code
- Easier to handle someone cycling off project

Verbalizing decisions improves their quality:

- The process of writing an explanation encourages critical thinking

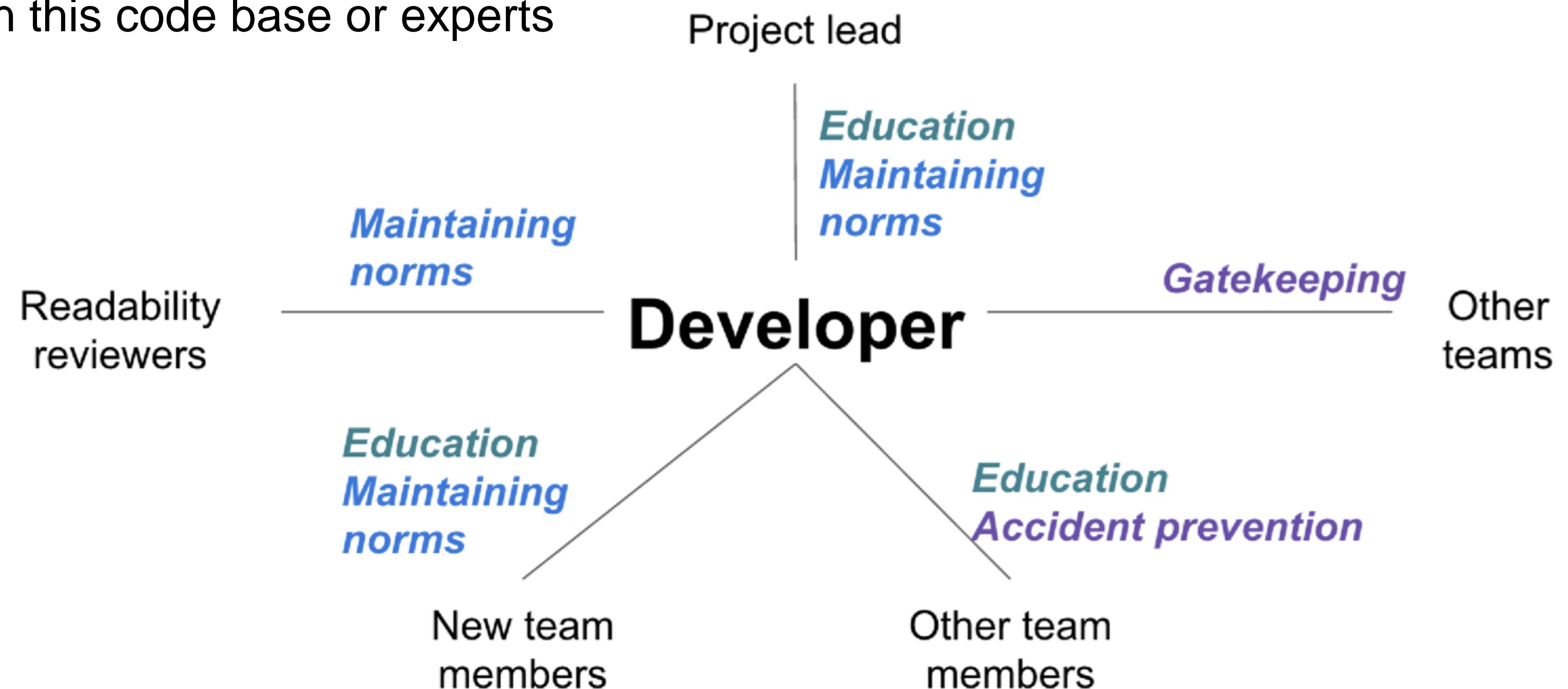
Code reviews improve quality of code base:

- Knowing code is reviewed pushes devs to make it more presentable and understandable

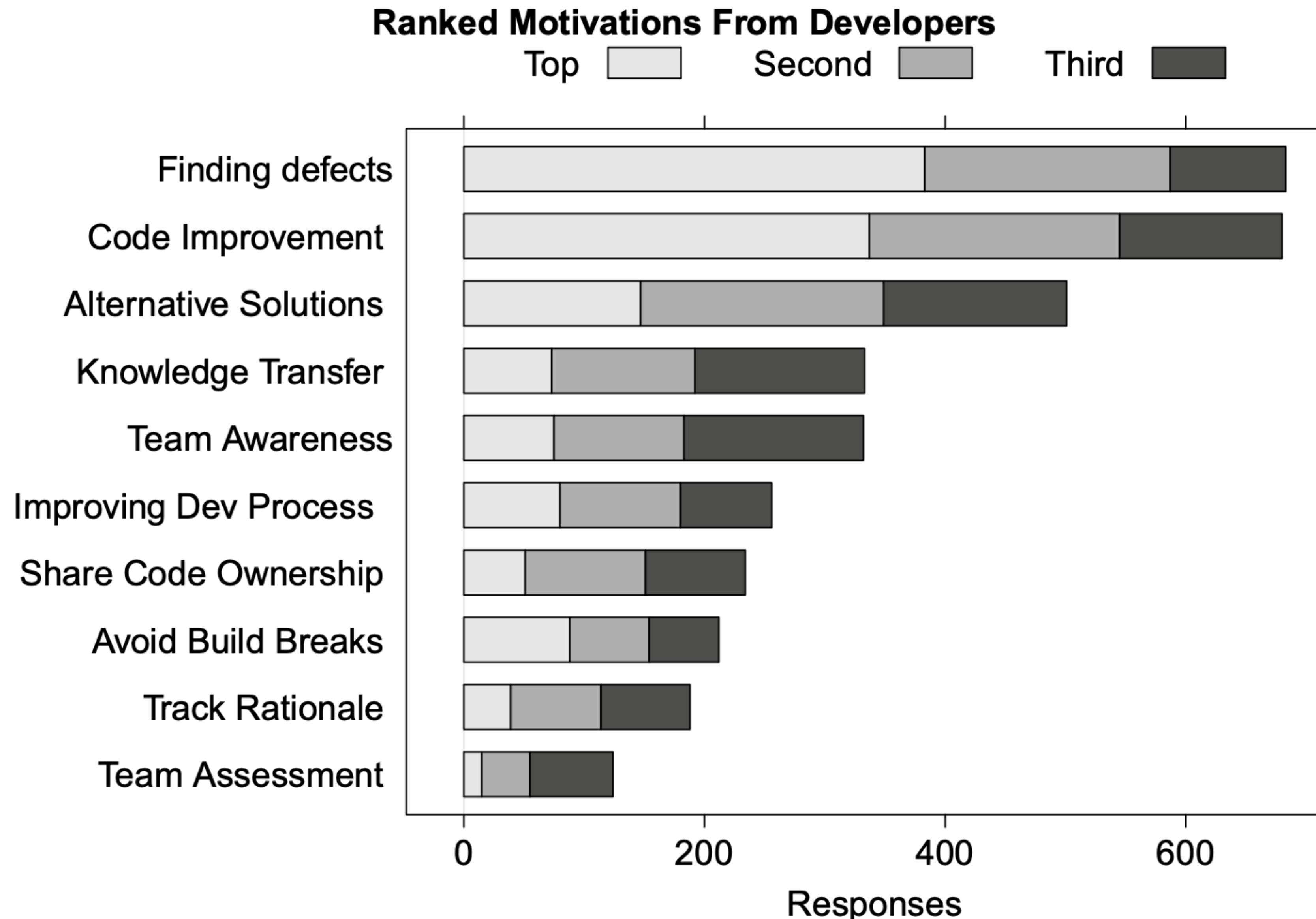
# Many stakeholders can benefit from code review

Reviewers might be...

- An owner of the code being changed or added to
- Someone to verify that the code meets standards.
- Someone to ensure documentation is consistent.
- Other people interested in this code base or experts



# Code reviews have many benefits



# Code review: How they do it at Google

At Google, reviewers get changes, explanation and all test results: review is asynchronous.

Elsewhere reviews can be in person:

- More heavyweight, cannot be as common.

Review must be professional and impersonal:

- No one is being “attacked” (or, no one should be).

Don’t rehash design arguments (defer to author).

All suggestions and criticisms must be addressed:

- At least in the negative.

# Code review checklist

---

Consider:

- Am I able to understand the code easily?
- Does the code follow our style guidelines?
- Is the same code duplicated more than once?
- Is this file (or change) too big?
- Does this code meet our non-functional requirements?
- Is this code maintainable?
- Does this code have unintended side-effects?

# Code Reviews and Programmer's Ego

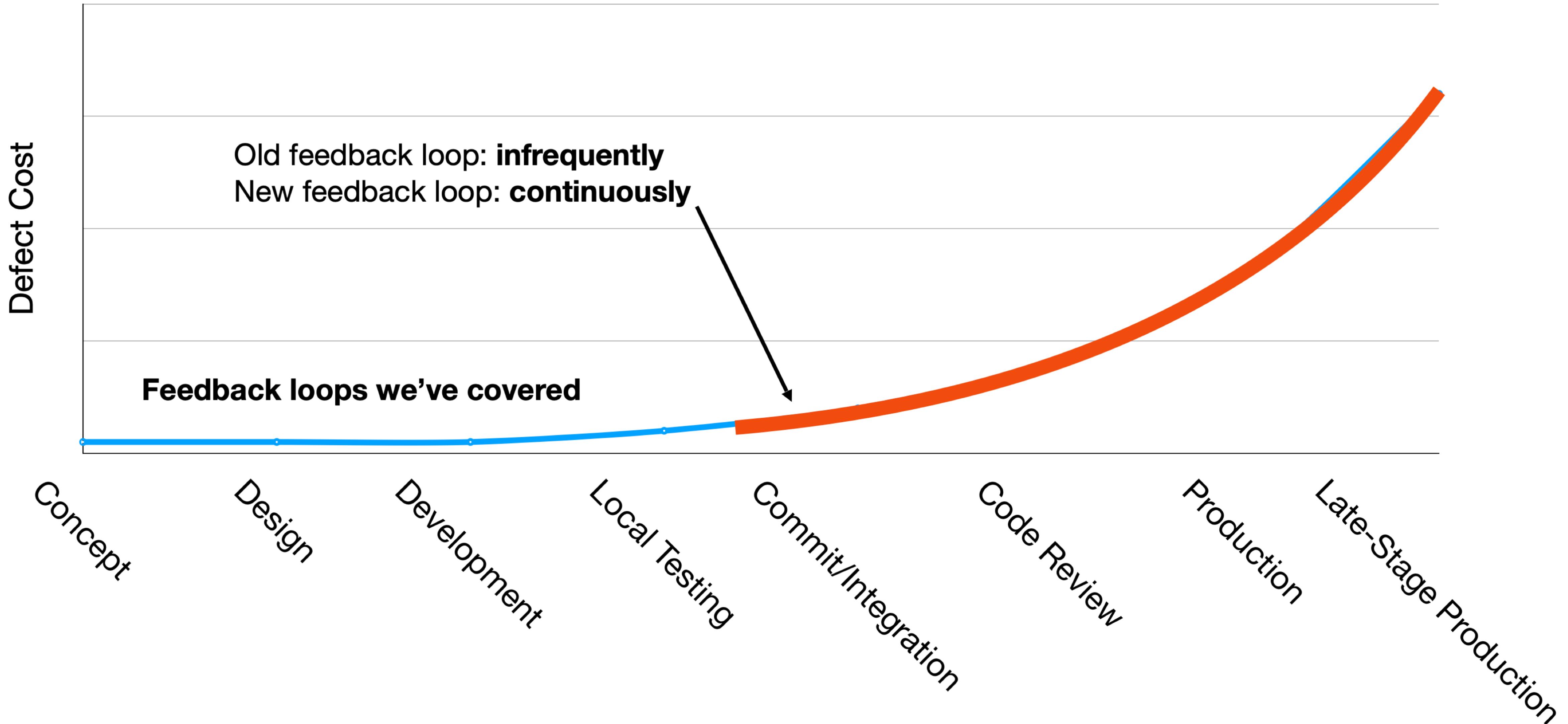
---

Remember:

- Code review means someone's looking over your work
- You might have some attachment to it
- Criticisms: sometimes hard not to take personally
- Acknowledge a criticism and move on
- Acknowledgment doesn't imply that the author agrees with the content of the criticism
- The review is not about you, the goal is to improve code

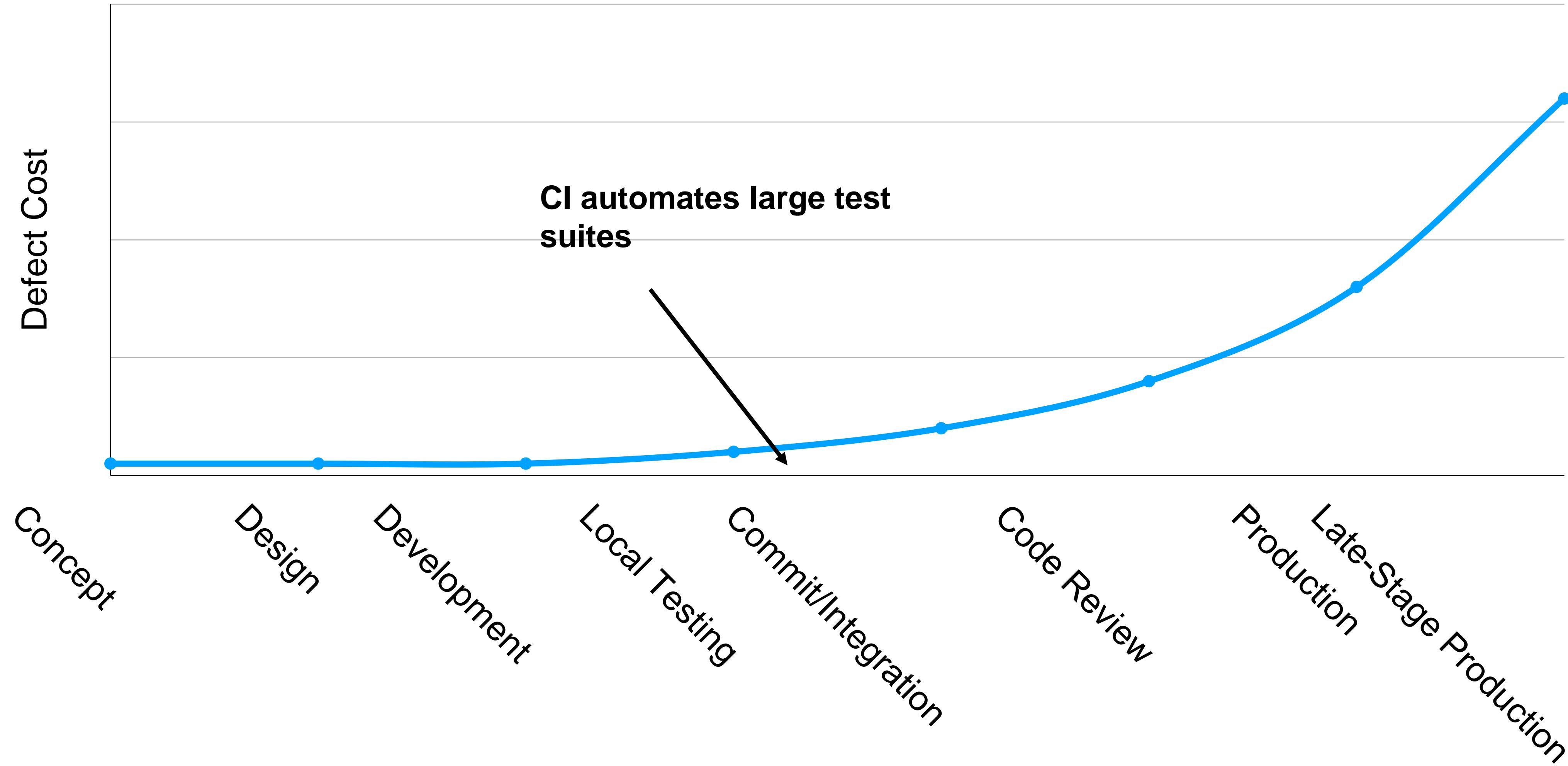
# Agile values fast quality feedback loops

Faster feedback = lower cost to fix bugs



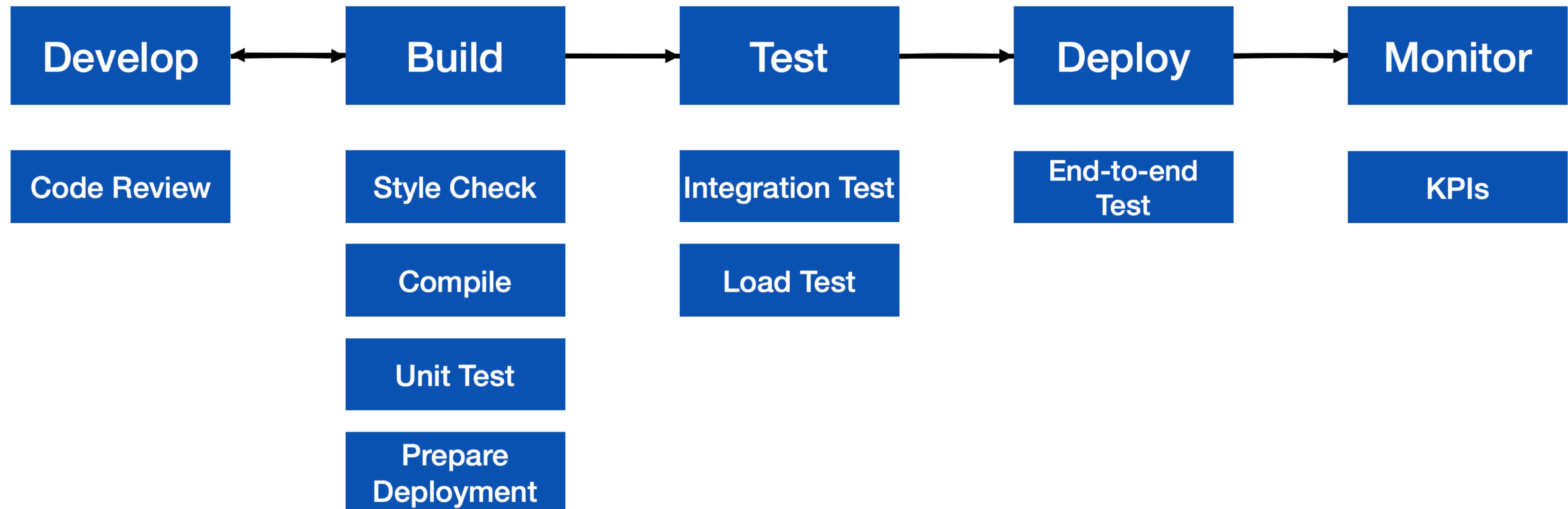
# Continuous Integration

Fast feedback on integration errors



# Continuous Development

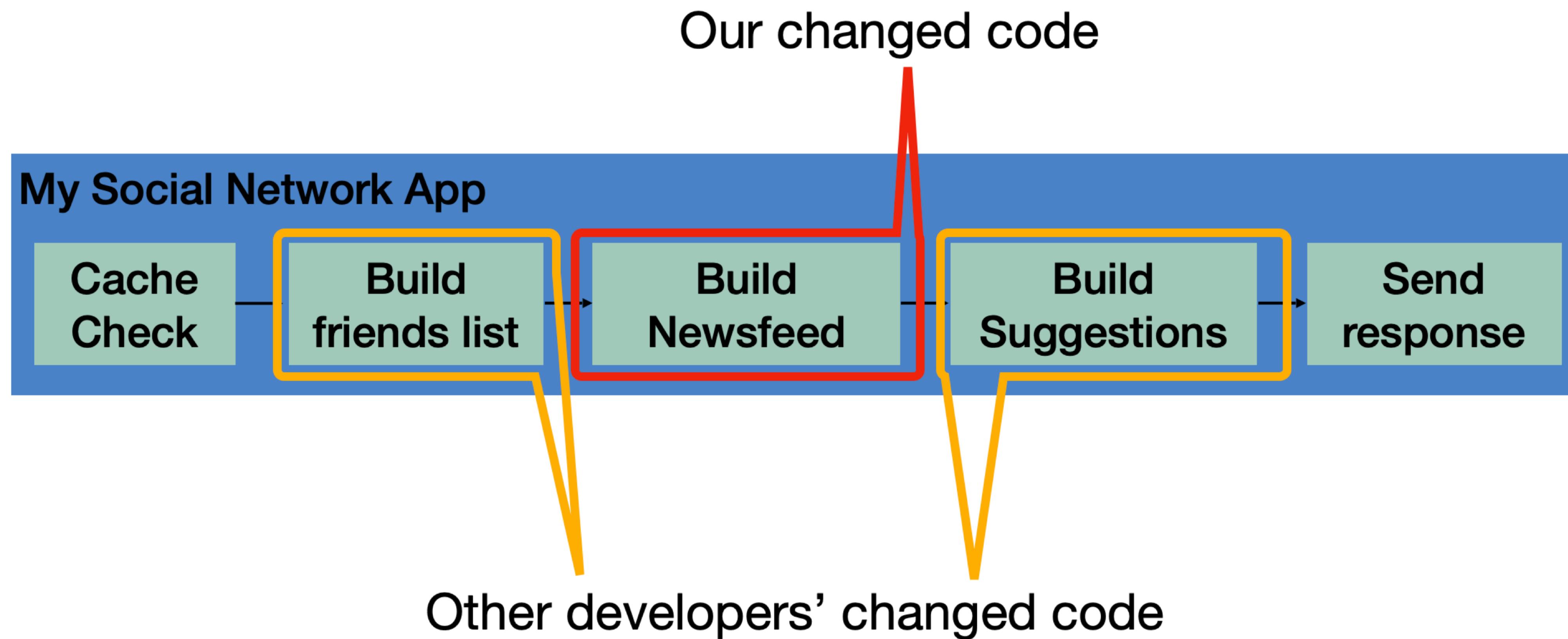
Improving quality & velocity with frequent, fast feedback loops



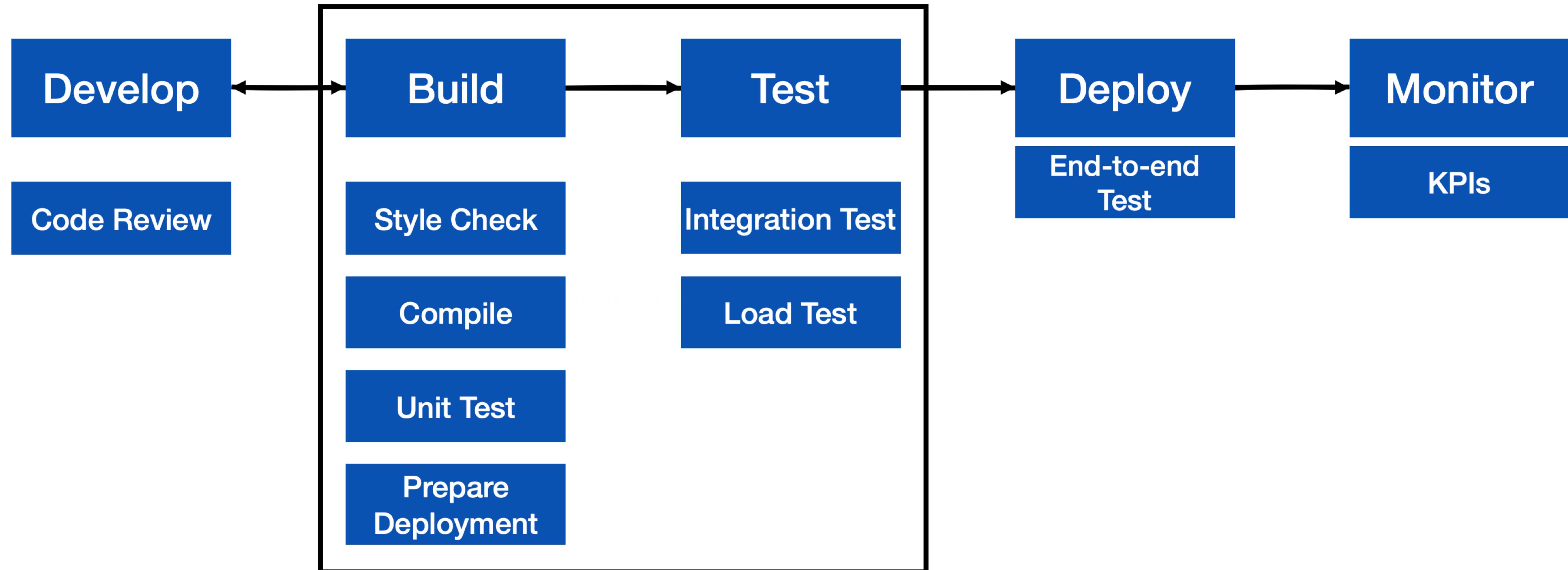
# Continuous Integration

## Motivation

- Our systems involve many components, some of which might even be in different version control repositories
- How does a developer get feedback on their (local) change?



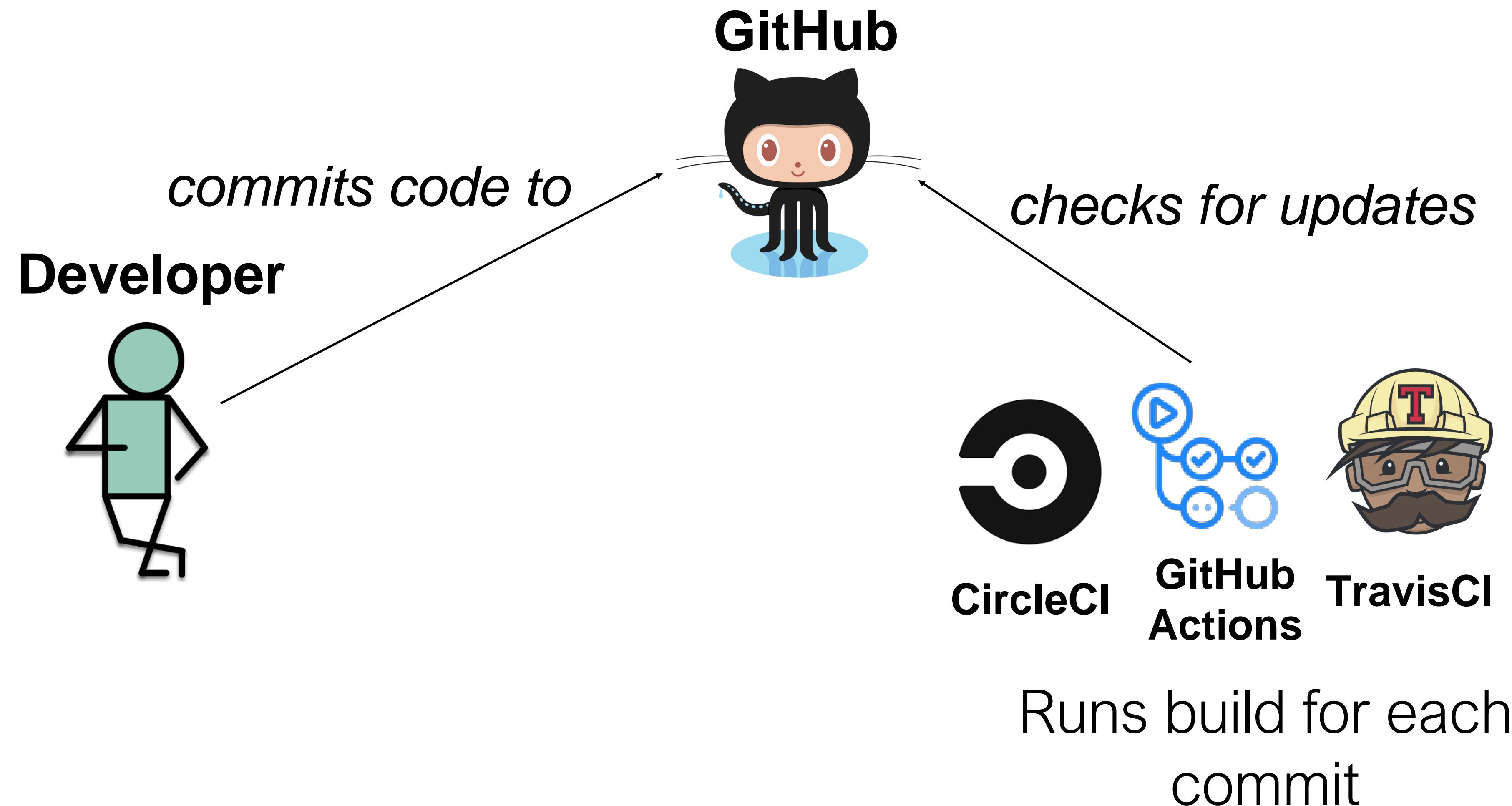
# CI is a software pipeline



**Automate this centrally, provide a central record of results**

# CI in practice

Small scale, with a service like CircleCI, GitHub Actions or TravisCI



# Example CI Pipeline

## Open source project: PrestoDB

prestodb / presto  build passing

Current Branches Build History Pull Requests

### Pull Request #15372 Fix extracting logic in dynamic filtering when

When integrating with filter pushdown, we extract dynamic filter

– Commit cde9e65 ↗

↳ #15372: Fix extracting logic in dynamic filtering when integrated with

↳ Branch master ↗

 Ke

 #52304 failed

 Ran for 17 min 40 sec

 Total time 10 hrs 26 min 10 sec

 10 hours ago

Build jobs				View config	
 # 52304.1	 AMD64	 Trusty	 Java		MAVEN_CHECKS=true  10 min 5 s
 # 52304.2	 AMD64	 Trusty	 Java		WEBUI_CHECKS=true  58 sec
 # 52304.3	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  6 min 7 s
 # 52304.4	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  24 min 5 s
 # 52304.5	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  7 min 45 s
 # 52304.6	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  8 min 4 s
 # 52304.7	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  7 min 11 s
 # 52304.8	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  8 min 23 s
 # 52304.9	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  8 min 28 s
 # 52304.10	 AMD64	 Trusty	 Java		TEST_SPECIFIC_MODULES=presto-tests  8 min 47 s

<https://travis-ci.com/github/prestodb/presto>

# Example CI Pipeline - TravisCI

At a glance. see history of build



Current Branches **Build History** Pull Requests

More options

✓ master	This patch bumps Alluxio dependency to 2.3.0-2	→ #52300 passed	⌚ 10 hrs 49 min 31 sec
⌚ James Sun	36392a2 ↗	⌚ 2 days ago	
! master	Handle query level timeouts in Presto on Spark	→ #52287 errored	⌚ 11 hrs 6 min 44 sec
⌚ Andrii Rosa	aa55ea7 ↗	⌚ 2 days ago	
! master	Fix flaky test for TestTempStorageSingleStreamSp	→ #52284 errored	⌚ 11 hrs 50 min 37 sec
⌚ Wenlei Xie	193a4cd ↗	⌚ 2 days ago	
✓ master	Check requirements under try-catch	→ #52283 passed	⌚ 11 hrs 3 min 20 sec
⌚ Andrii Rosa	fff331f ↗	⌚ 2 days ago	
✓ master	Update TestHiveExternalWorkersQueries to create	→ #52282 passed	⌚ 10 hrs 55 min 37 sec
⌚ Maria Basmanova	746d7b5 ↗	⌚ 2 days ago	
✓ master	Introduce large dictionary mode in SliceDictionary	→ #52277 passed	⌚ 10 hrs 43 min 30 sec
⌚ Maria Basmanova	a90d97a ↗	⌚ 2 days ago	

<https://travis-ci.com/github/prestodb/presto>

# CI In Practice: autograder

test.yml (CI workflow file)

```
name: 'Build and Test the Grader'
on: # rebuild any PRs and main branch changes
  pull_request:
  push:
    branches:
      - main
      - 'releases/*'
jobs:
  build:
    runs-on: self-hosted
    steps:
      - uses: actions/checkout@v2
      - uses: actions/setup-node@v2
        with:
          node-version: '16'
      - run: |
        npm install
  test:
    runs-on: self-hosted
    strategy:
      matrix:
        submission: [a, b, c, ts-ignore, linting-error, non-green-tests, empty]
    steps:
      - uses: actions/checkout@v2
      - uses: actions/setup-node@v2
        with:
          node-version: '16'
      - uses: ./
        with:
          submission-directory: solutions/${{ matrix.submission }}
```

GitHub Actions Results

test.yml

on: push

build

30s

Matrix: test

test (a)

3m 6s

test (b)

3m 3s

test (c)

2m 58s

test (ts-ignore)

5s

test (linting-error)

31s

test (non-green-tests)

35s

test (empty)

4s

# Attributes of effective CI processes

- Do not allow builds to remain broken for a long time
- CI should run for every change
- CI should be fast, providing feedback within minutes or hours
- CI should not completely replace pre-commit testing

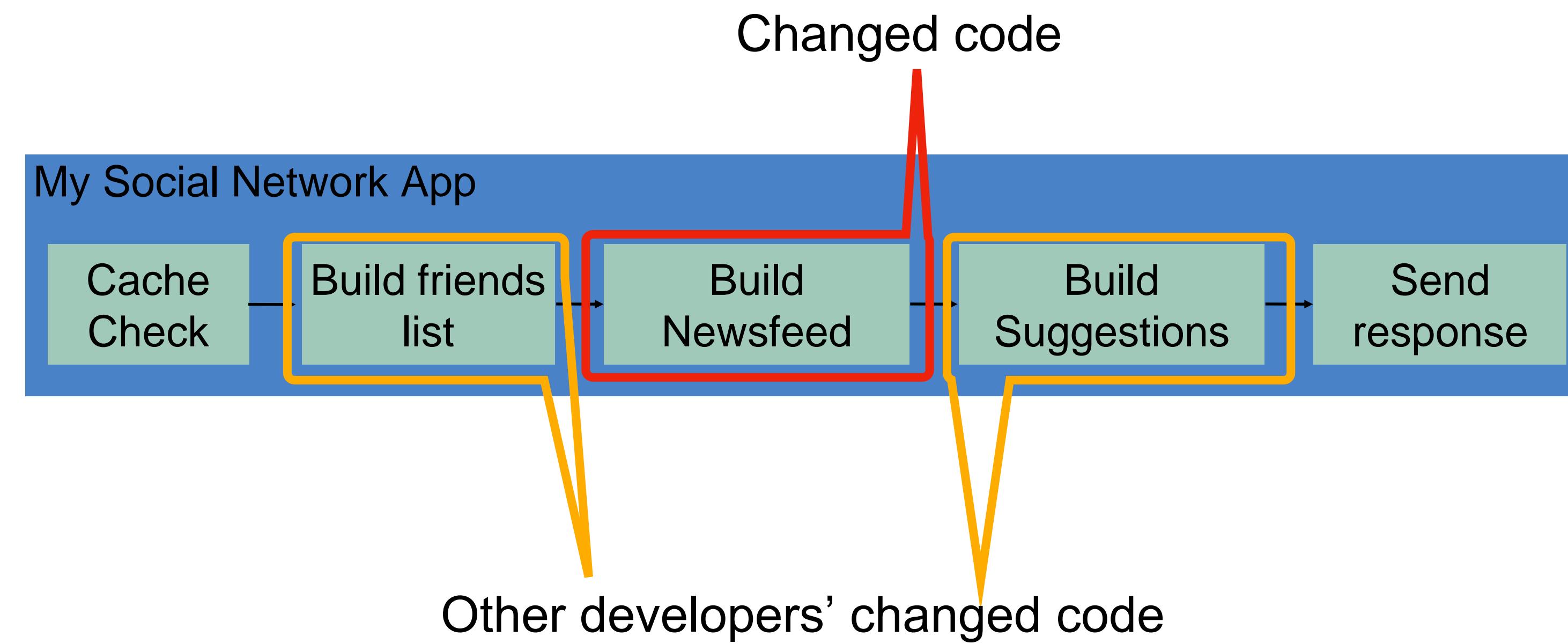
A screenshot of a CI build status page. At the top, a green checkmark icon and the text "Output the full test name" are displayed. Below this, a summary box states "All checks have passed" and "9 successful checks". A list of five successful checks is shown, each with a green checkmark icon, a GitHub icon, a descriptive name, a status message, and a "Details" link. The checks are: "Build and Test the Grader / build (push) Successfu...", "Check dist/ / check-dist (push) Successful in 30s", "Build and Test the Grader / test (reference) (push) ...", "Build and Test the Grader / test (b) (push) Succes...", and "Build and Test the Grader / test (ts-ignore) (push)".

A screenshot of a GitHub pull request interface. The title of the pull request is "Tools: extract\_features.py: correct define name for AP\_RPM\_ENABLED". Below the title, it shows a commit by "peterbarker" from 5 days ago. The next section is "AP\_Mission: prevent use of uninitialised stack data" with a commit by "peterbarker" from 5 days ago. The third section is "AP\_HAL\_ChibiOS: disable DMA on I2C on bdshot boards to free up DMA ch..." with a commit by "andyp1per" and "tridge" from 6 days ago. The fourth section is "SITL: Fixed rounding lat/Ing issue when running JSBSim SITL" with a commit by "ShivKhanna" and "tridge" from 6 days ago. The fifth section is "AP\_HAL\_ChibiOS: define skyviper short board names" with a commit by "yuri-rage" and "tridge" from 6 days ago.

# How do we apply continuous integration?

Testing the right things at the right time

- Do we integrate changes immediately, or do a pre-commit test?
- Which tests do we run when we integrate?
- How do we compose the system under test at each point?



# Use scalable cloud resources for CI

---

## Example: Developing a Fuzzer

- *Fuzzers* automatically generate inputs to programs that cover code and reveal bugs
- Fuzzers are non-deterministic: to evaluate with confidence, need repeated, long-running trials
- Evaluating fuzzers is time consuming, determining which changes impact performance is confusing
- How to run experiments in the cloud?

# CI Pipelines automate performance testing

## eval-10m-5x.yml

on: push

✓ evaluate / build-matrix

5s

Matrix: evaluate / run-fuzzer

✓ evaluate / run-fuzzer (... 12m 21s)

✓ evaluate / run-fuzzer ... 12m 25s

✓ evaluate / run-fuzzer ... 12m 23s

✓ evaluate / run-fuzzer (... 12m 27s)

✓ evaluate / run-fuzzer (... 12m 13s)

✓ evaluate / run-fuzzer ... 12m 24s

✓ evaluate / run-fuzzer (... 12m 21s)

✓ evaluate / run-fuzzer ... 12m 23s

✓ evaluate / run-fuzzer (... 12m 27s)

✓ evaluate / run-fuzzer (... 12m 13s)

✓ evaluate / run-fuzzer ... 12m 24s

✓ evaluate / run-fuzzer ... 12m 25s

✓ evaluate / run-fuzzer ... 12m 26s

✓ evaluate / run-fuzzer ... 12m 26s

Every commit: Run 10 minute performance test on 5 benchmarks, repeating each test 5 times (25 concurrent jobs)

## eval-24h-20x.yml

on: workflow\_dispatch

✓ evaluate / build-matrix

2s

Matrix: evaluate / run-fuzzer

✓ evaluate / run-fuzzer (an... 1d 0h)

✓ evaluate / run-fuzzer (bc... 1d 0h)

✓ evaluate / run-fuzzer (cl... 1d 0h)

✓ evaluate / run-fuzzer (m... 1d 0h)

✓ evaluate / run-fuzzer (rh... 1d 0h)

✓ evaluate / run-fuzzer (an... 1d 0h)

✓ evaluate / run-fuzzer (bc... 1d 0h)

✓ evaluate / run-fuzzer (cl... 1d 0h)

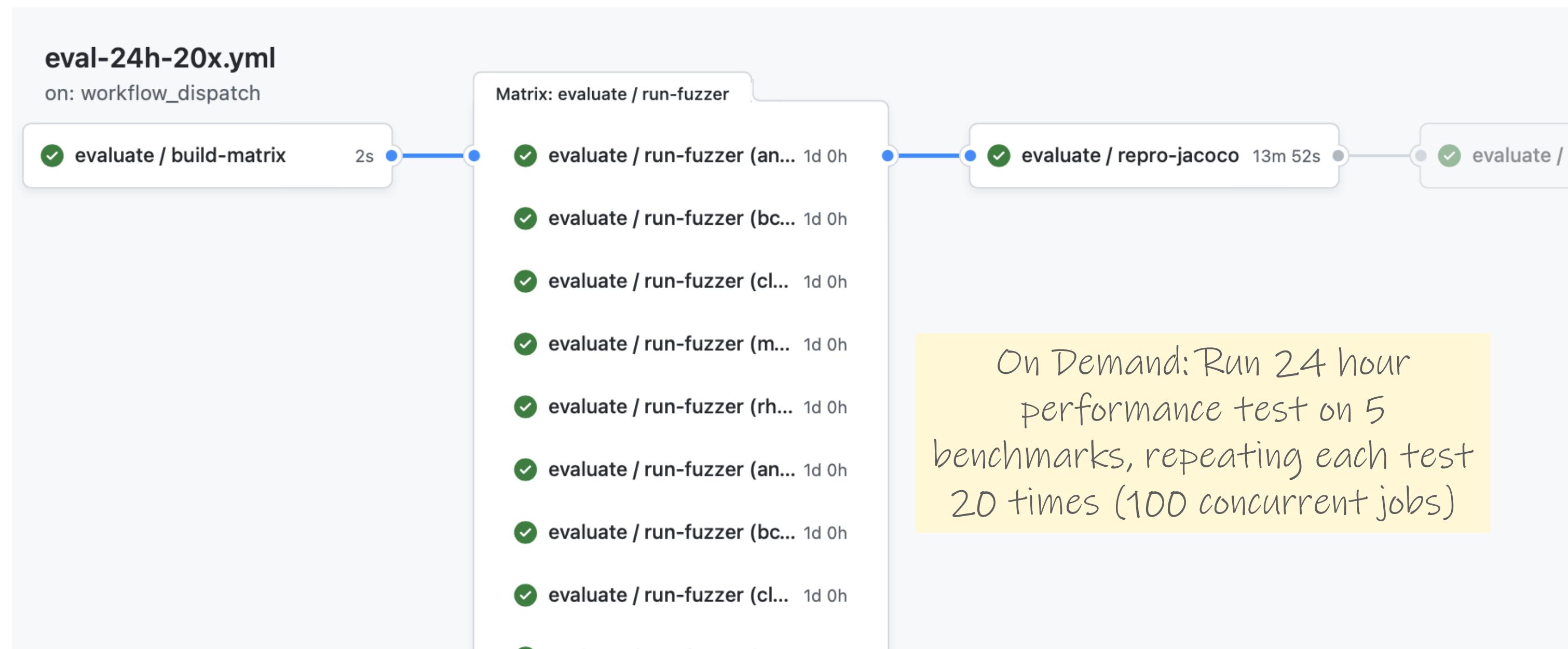
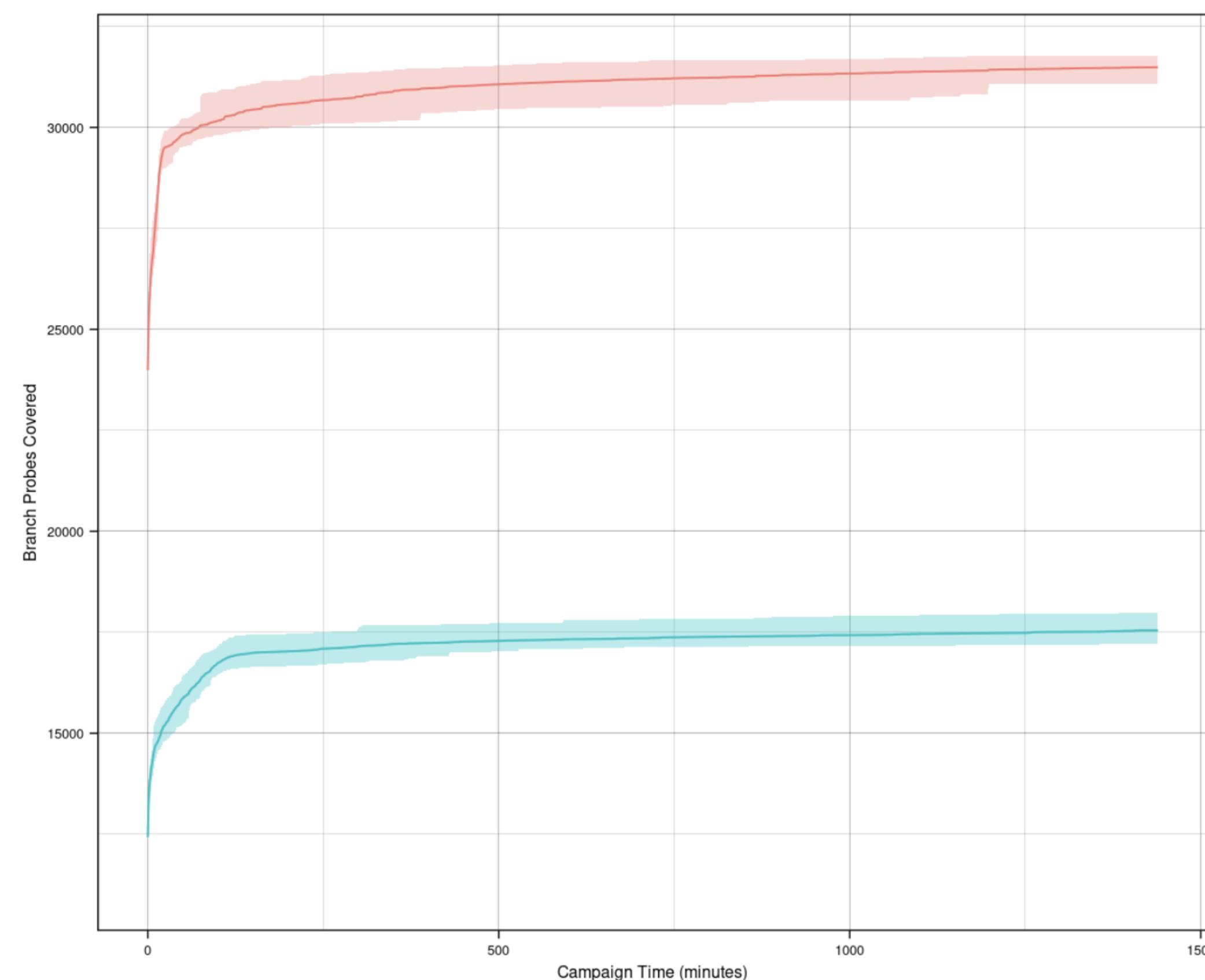
On Demand: Run 24 hour performance test on 5 benchmarks, repeating each test 20 times (100 concurrent jobs)

<https://github.com/neuse/CONFETTI/actions>

# CI Pipelines automate benchmarking

closure

## Branch Probes Over Time



# CI in practice

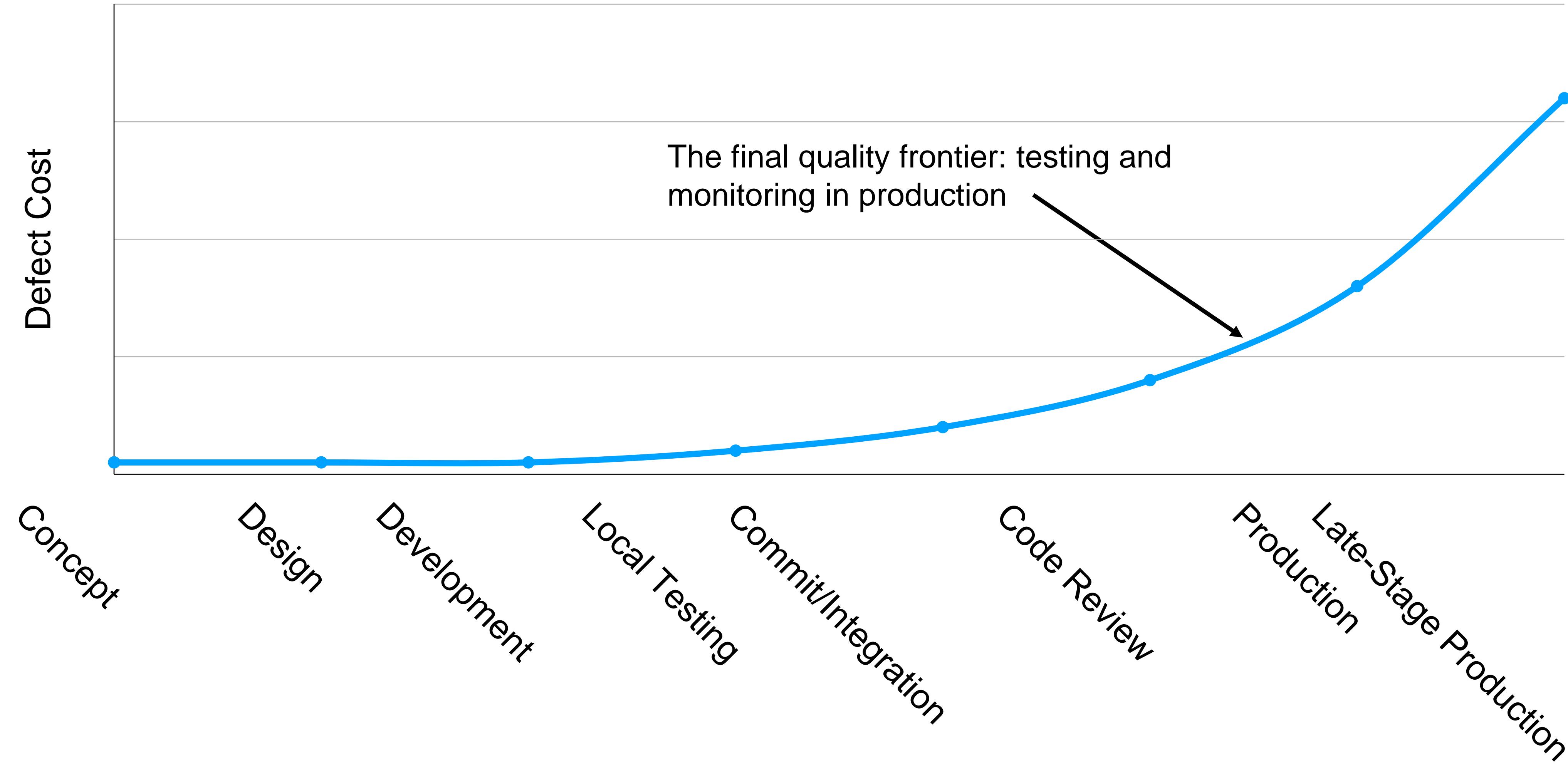
---

## Large scale example: Google TAP

- 50,000 unique changes per-day, 4 billion test cases per-day
- Pre-submit optimization: run fast tests for each individual change (before code review).  
Block merge if they fail.
- Then: run all affected tests; “build cop” monitors and acts immediately to roll-back or fix
- Build cop monitors integration test runs
- Average wait time to submit a change: 11 minutes

# Cost to fix a defect over time

Rough estimate



# Deployment Philosophy: Instagram

“Faster is safer”



“If stuff blows up it affects a very small percentage of people”



**Instagram cofounder and CTO Mike Krieger**

# Continuous Delivery

“Faster is safer”: Key values of continuous delivery

- Release frequently, in small batches
- Maintain key performance indicators to evaluate the impact of updates
- Phase roll-outs
- Evaluate business impact of new features

# Staging environments

---

## Enabling Continuous Delivery

As software gets more complex with more dependencies, it's impossible to simulate the whole when testing

Idea: Deploy to a complete production-like environment, but don't have all use it

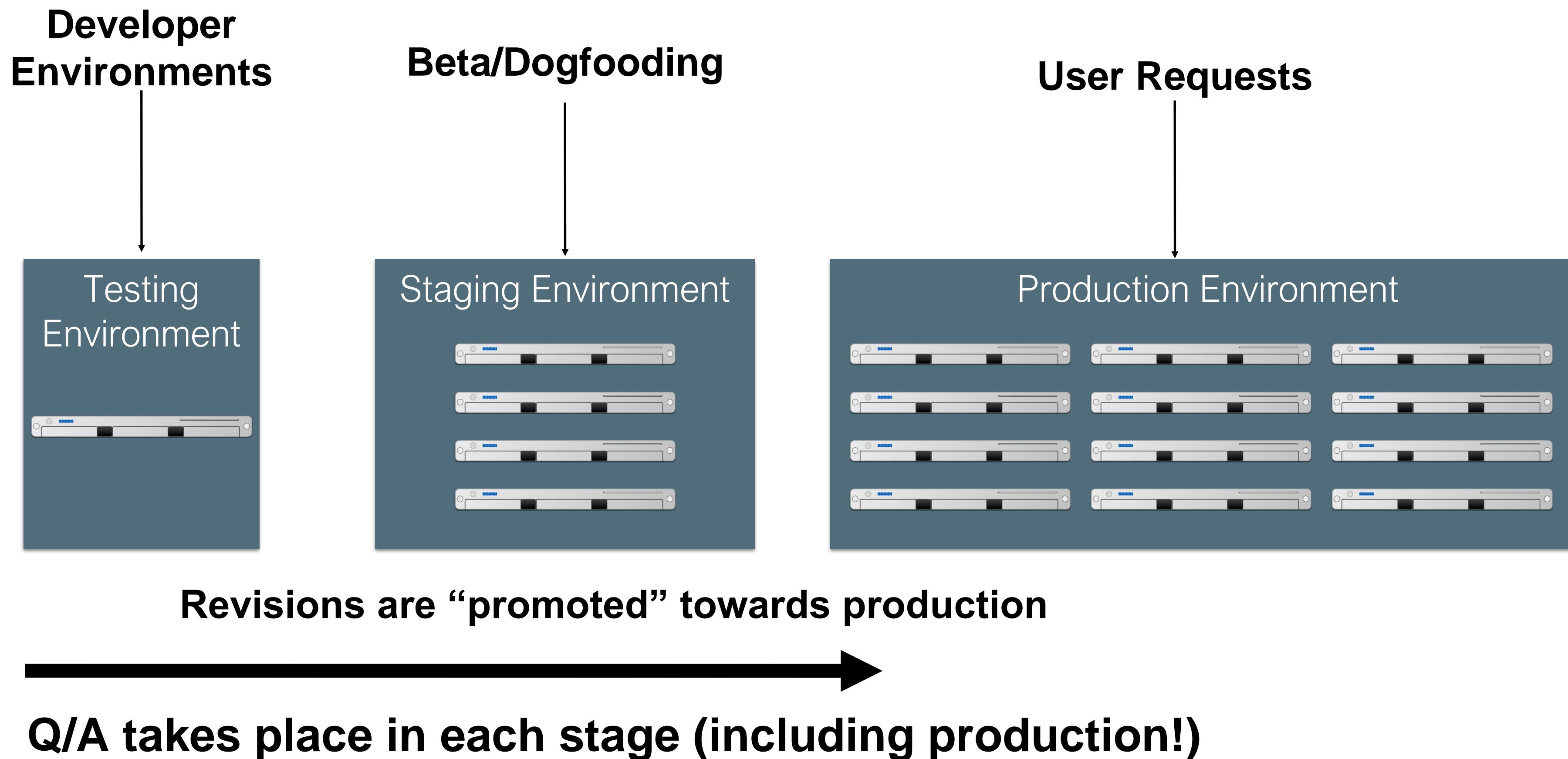
### Examples:

- ◎ “Eat your own dogfood”
- ◎ Beta/Alpha testers

Lower risk if a problem occurs in staging than in production

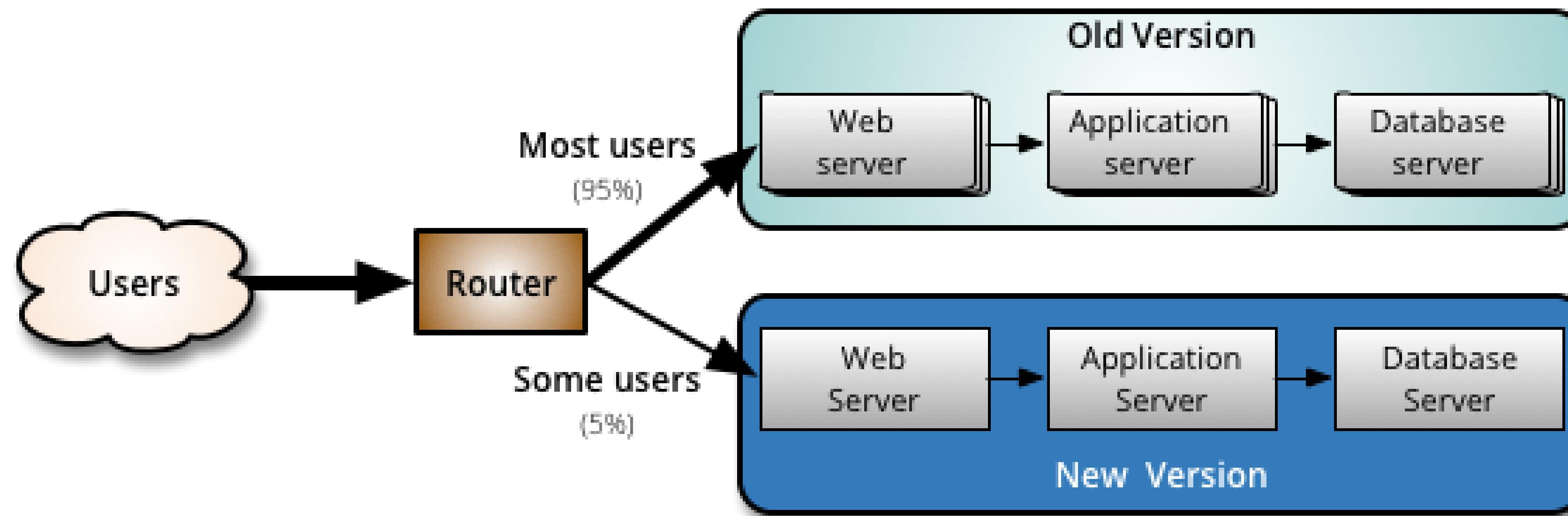
# Test-Stage-Production

## Continuous Delivery in Action



# A/B Deployments with Canaries

Mitigating risk in continuous delivery



Monitor both:  
But minimize impact of problems in new version

# Operations Responsibility

DevOps in a slide

Once we deploy, someone has to monitor, make sure it's running OK, no bugs, etc

Assume 3 environments: Test, Staging, Production

Whose job is it?

	Developers			Operators		
Waterfall				Test	Staging	Production
Agile	Test				Staging	Production
DevOps	Test	Staging	Production		Production	

# Release Pipelines

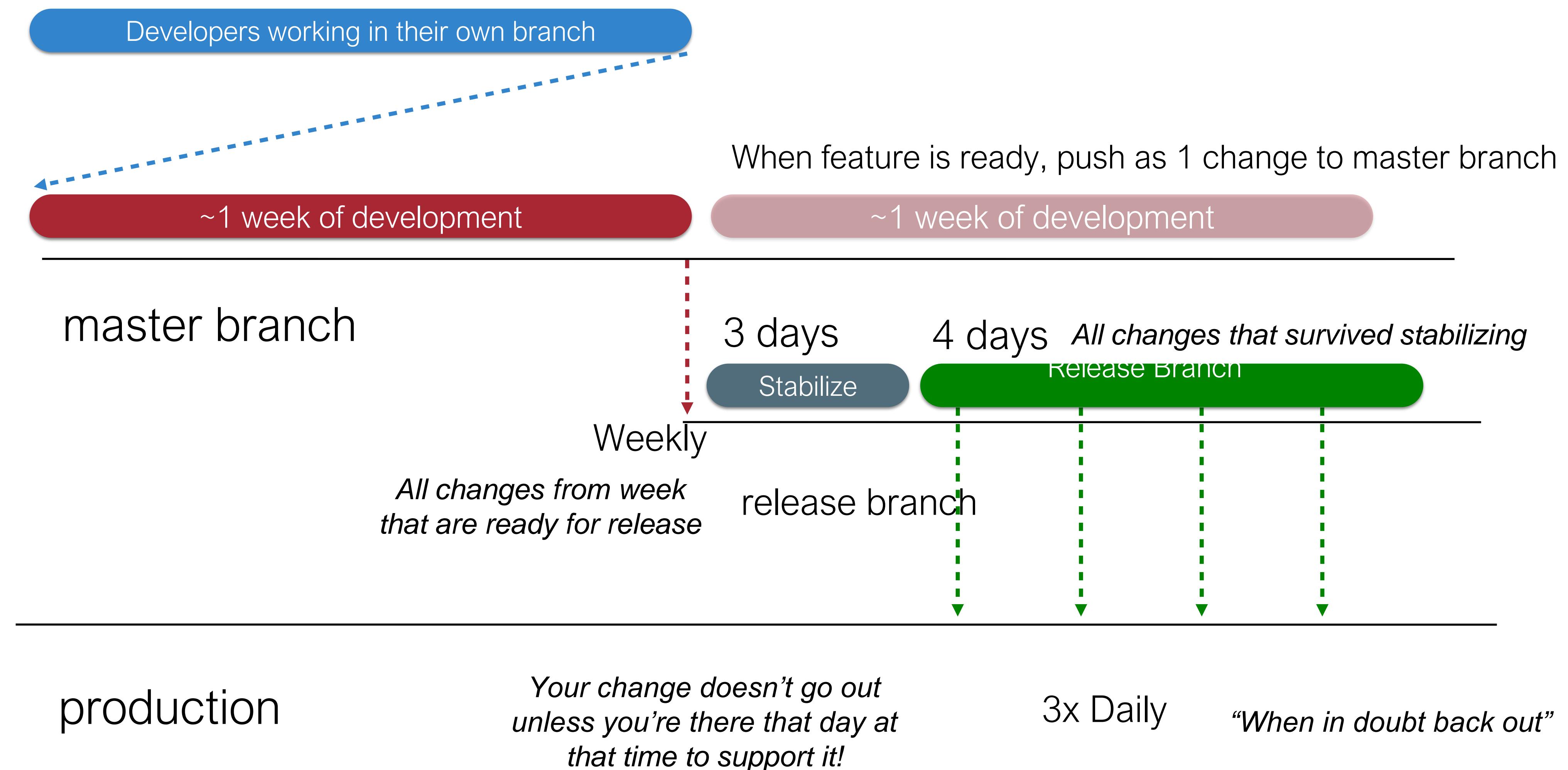
---

How quickly is my change deployed?

- Even if you are deploying every day, you still have some latency
- A new feature I develop today won't be released today
- But, a new feature I develop today can begin the release pipeline today (minimizes risk)
- Release Engineer: gatekeeper who decides when code ready to go out, oversees deployment process

# Deployment Example: Facebook.com

Pre-2016



# Deployment Example

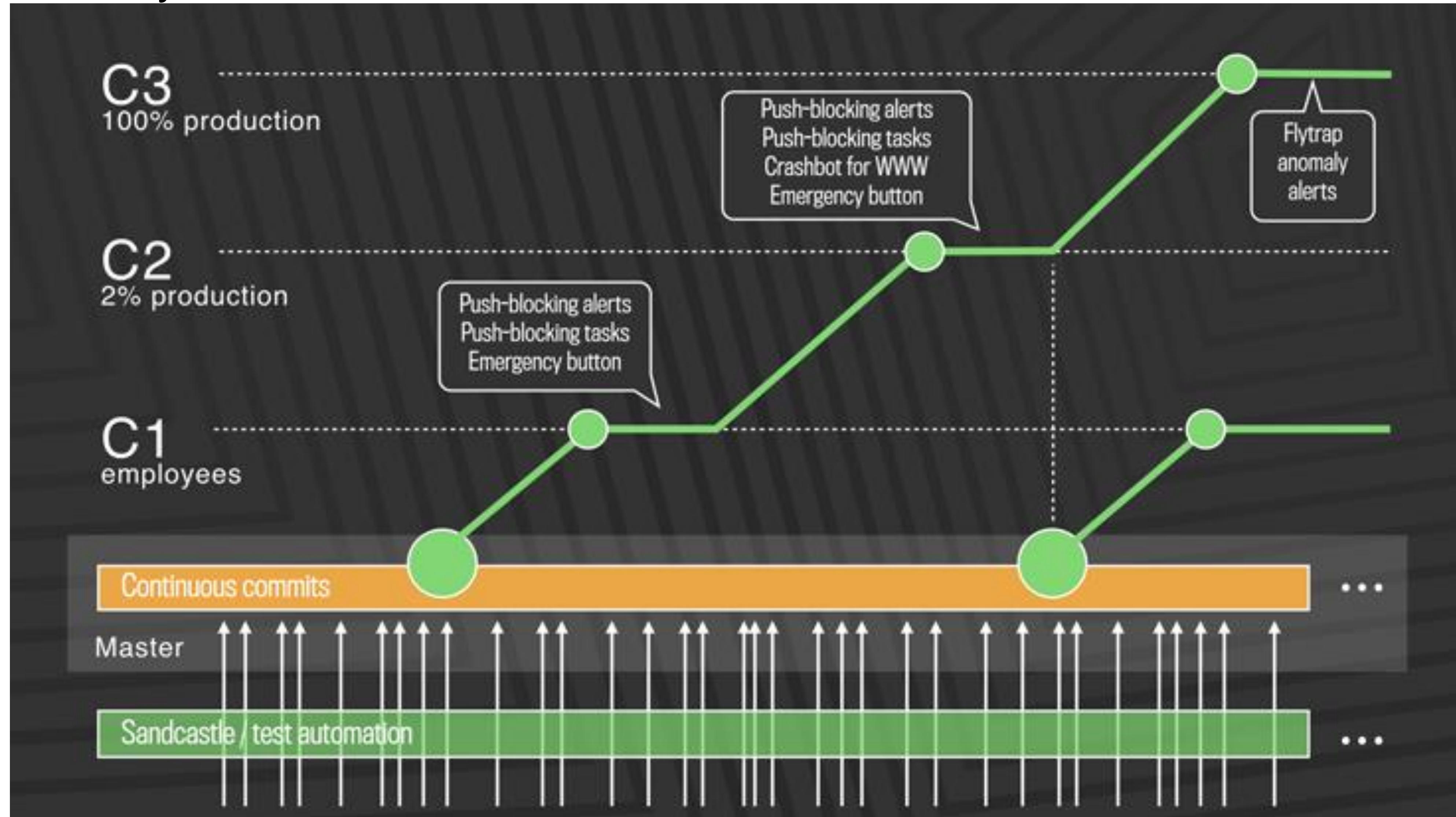
Chuck Rossi, Director Software Infrastructure & Release Engineering @ Facebook



“Our main goal was to make sure that the new system made people’s experience better — or at least, didn’t make it worse. After a year of planning and development, over the course of three days **we enabled 100% of our production web servers to run code deployed directly from master”**

# Deployment Example

Post-2016: Truly continuous releases from master branch



# Monitoring

---

The last step in continuous deployment: track metrics

## Hardware

- Voltages, temperatures, fan speeds, component health

## OS

- Memory usage, swap usage, disk space, CPU load

## Middleware

- Memory, thread/db connection pools, connections, response time

## Applications

- Business transactions, conversion rate, status of 3rd party components

# Monitoring services aggregate system status

# Monitoring dashboards help gather insights



# Monitoring services take automated actions



## Icinga

Search ...

- Dashboard
- Problems
- Overview
- History
- Event Grid
- Event Overview
- Notifications**
- Timeline
- Documentation
- System
- Configuration
- jon

Notifications		
<a href="#">View</a> <a href="#">Edit</a> <a href="#">Delete</a>		
Page	1	2
#	25	Sort by
Notification Start	Z	A
OK	2022-02-18 08:49:05	Slurm Nodes on nagios
OK	2022-02-18 08:49:05	Slurm Nodes on nagios
WARNING	2022-02-18 08:45:05	Slurm Nodes on nagios
WARNING	2022-02-18 08:45:05	Slurm Nodes on nagios
CRITICAL	2022-02-18 08:42:05	Slurm Nodes on nagios
CRITICAL	2022-02-18 08:42:05	Slurm Nodes on nagios
WARNING	2022-02-18 08:40:05	Slurm Nodes on nagios
WARNING	2022-02-18 08:40:05	Slurm Nodes on nagios
CRITICAL	2022-02-18 08:34:07	Slurm Nodes on nagios

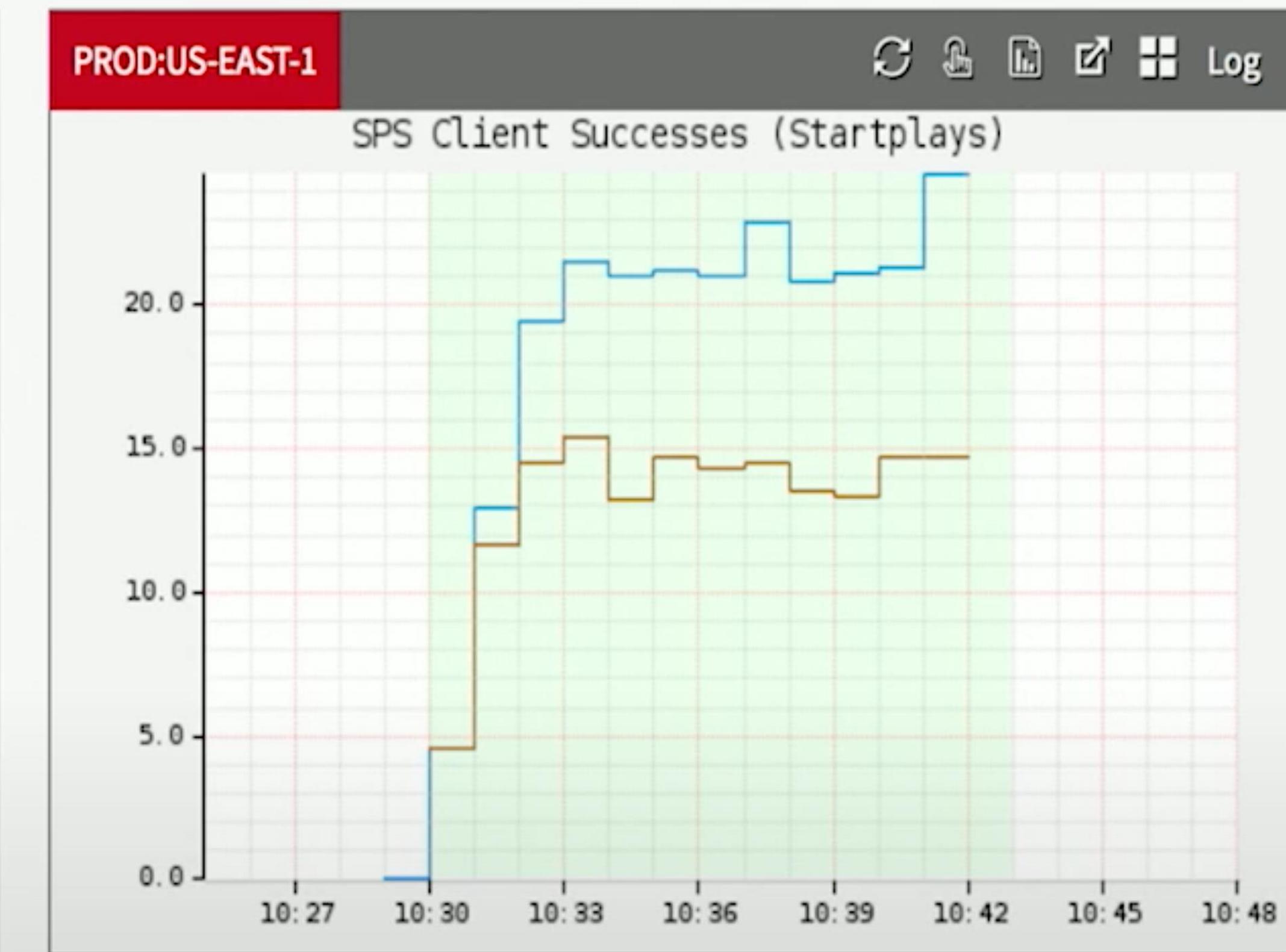
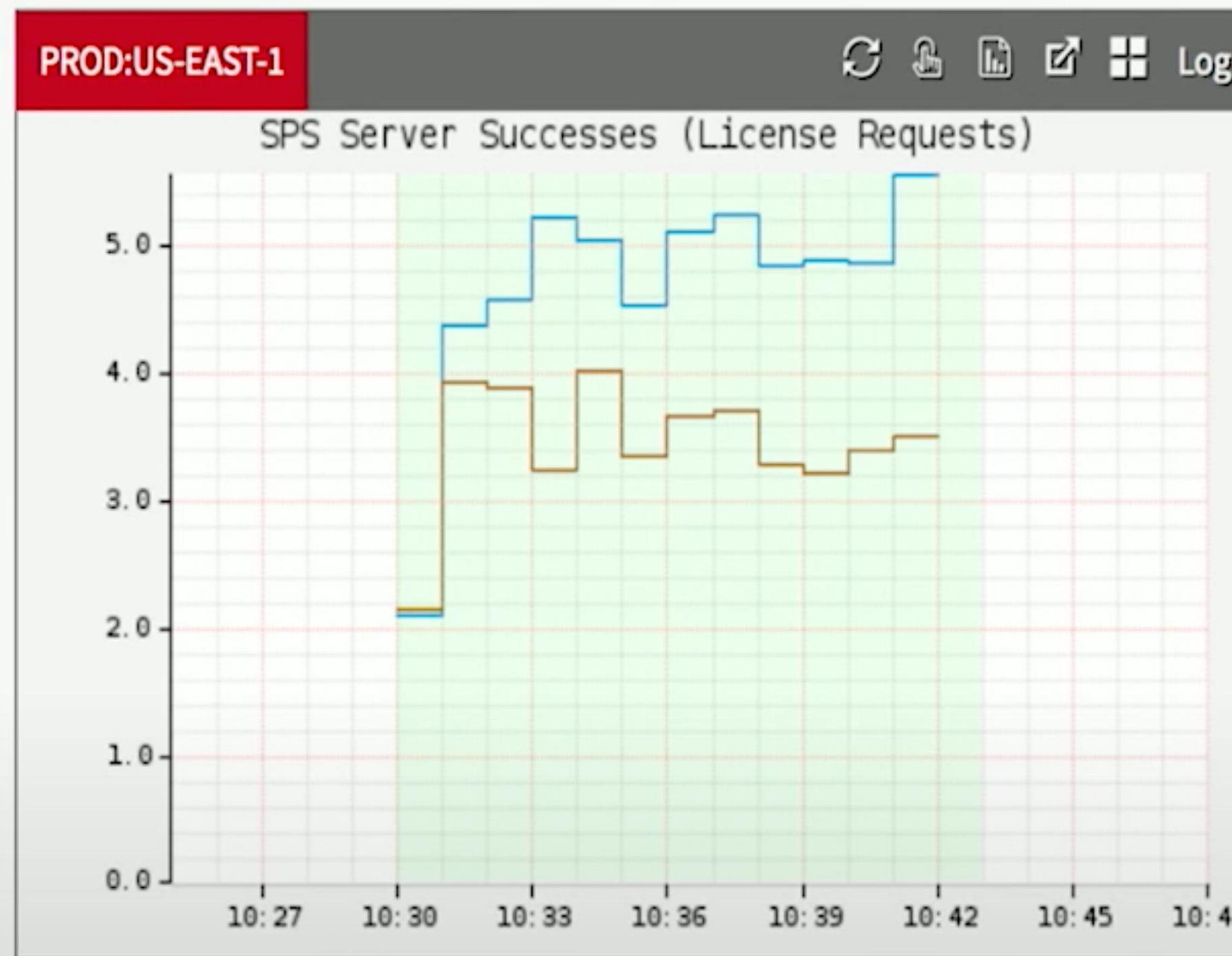
Notification		
<a href="#">View</a> <a href="#">Edit</a> <a href="#">Delete</a>		
<b>Current Service State</b>		
UP	nagios	since 2021-11-12 127.0.0.1
OK	Service: Slurm Nodes	for 1m 52s
<b>Event Details</b>		
Type	Notification	
Start time	2022-02-18 08:42:05	
End time	2022-02-18 08:42:05	
Reason	Normal notification	
State	CRITICAL	
Escalated	No	
Contacts notified	2	
Output	CRITICAL - 65 nodes unreachable, 161 reachable	

# Monitoring services take automated actions

Automatically detecting irregular behavior at Netflix

SPS

Legend: Experiment Control



MONITORING!

<https://www.youtube.com/watch?v=qyzymLlj9ag>

# Case study of a failed deployment

## Knightmare: A DevOps Cautionary Tale

 D7  DevOps  April 17, 2014  6 Minutes

I was speaking at a conference last year on the topics of DevOps, Configuration as Code, and Continuous Delivery and used the following story to demonstrate the importance making deployments fully automated and repeatable as part of a DevOps/Continuous Delivery initiative. Since that conference I have been asked by several people to share the story through my blog. This story is true – this really happened. This is my telling of the story based on what I have read (I was not involved in this).

This is the story of how a company with nearly \$400 million in assets went bankrupt in 45-minutes because of a failed deployment.



“In the week before go-live, a Knight engineer manually deployed the new RLP code in SMARS to its 8 servers. However, he made a mistake and did not copy the new code to one of the servers. Knight did not have a second engineer review the deployment, and neither was there an automated system to alert anyone to the discrepancy. “

# What could Knight capital have done better?

---

Use capture/replay testing instead of driving market conditions in a test

Avoid including “test” code in production deployments

Automate deployments

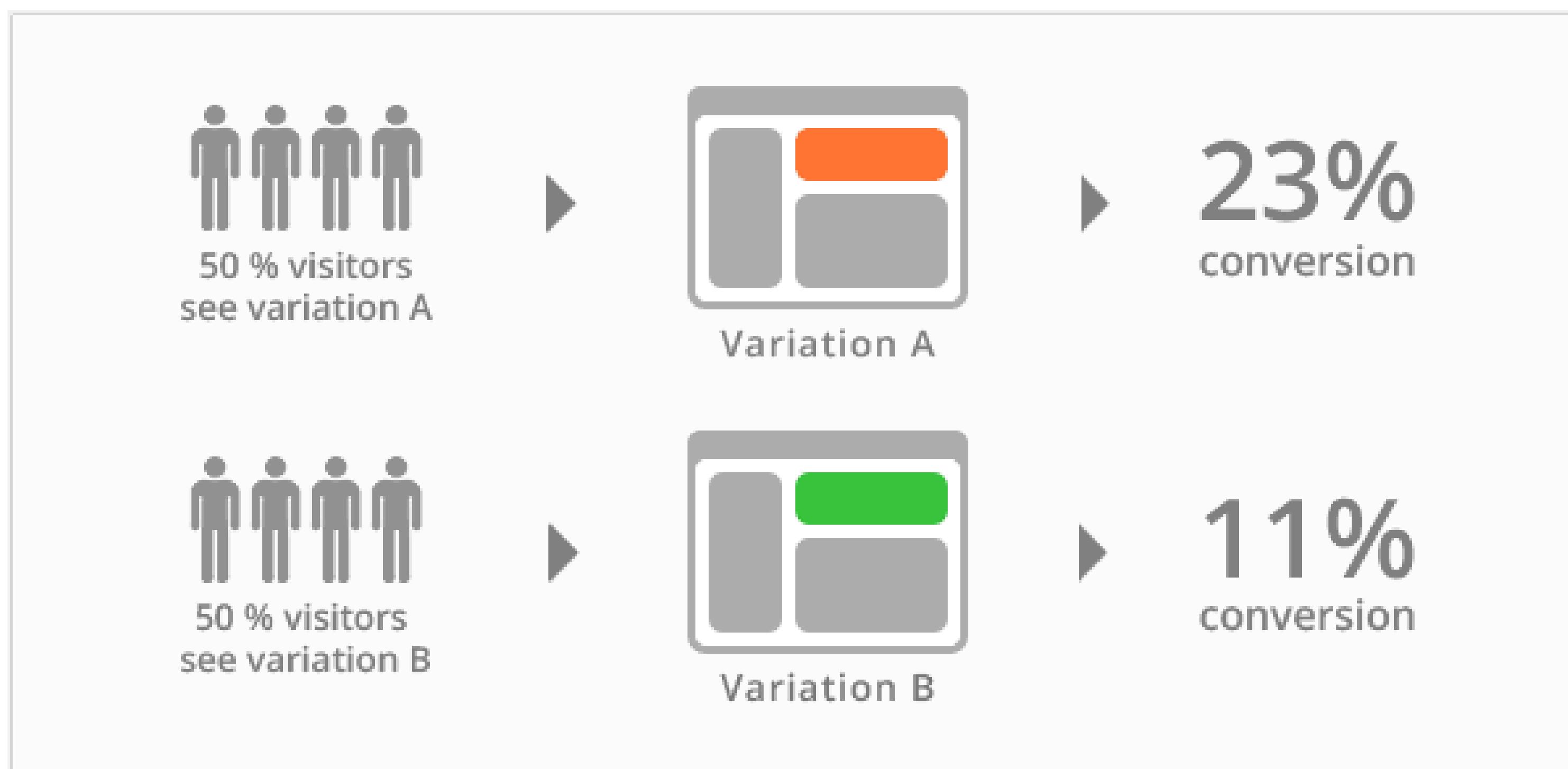
Define and monitor risk-based KPIs

Create checklists for responding to incidents

# Usability testing in continuous development

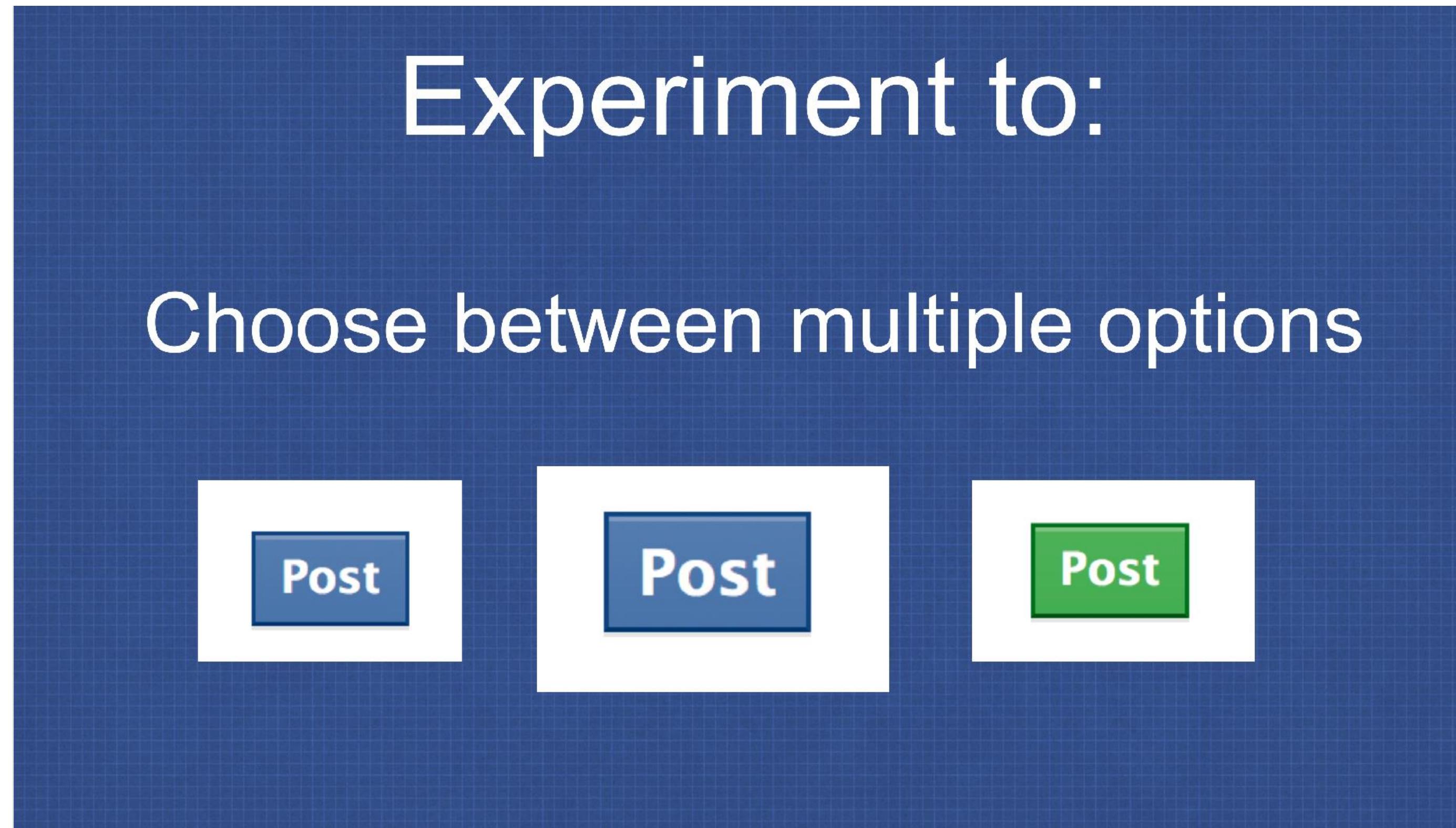
## A/B Testing

- Ways to test new features for usability, popularity, performance without a focus group
- Show 50% of your site visitors version A, 50% version B, collect metrics on each, decide which is better



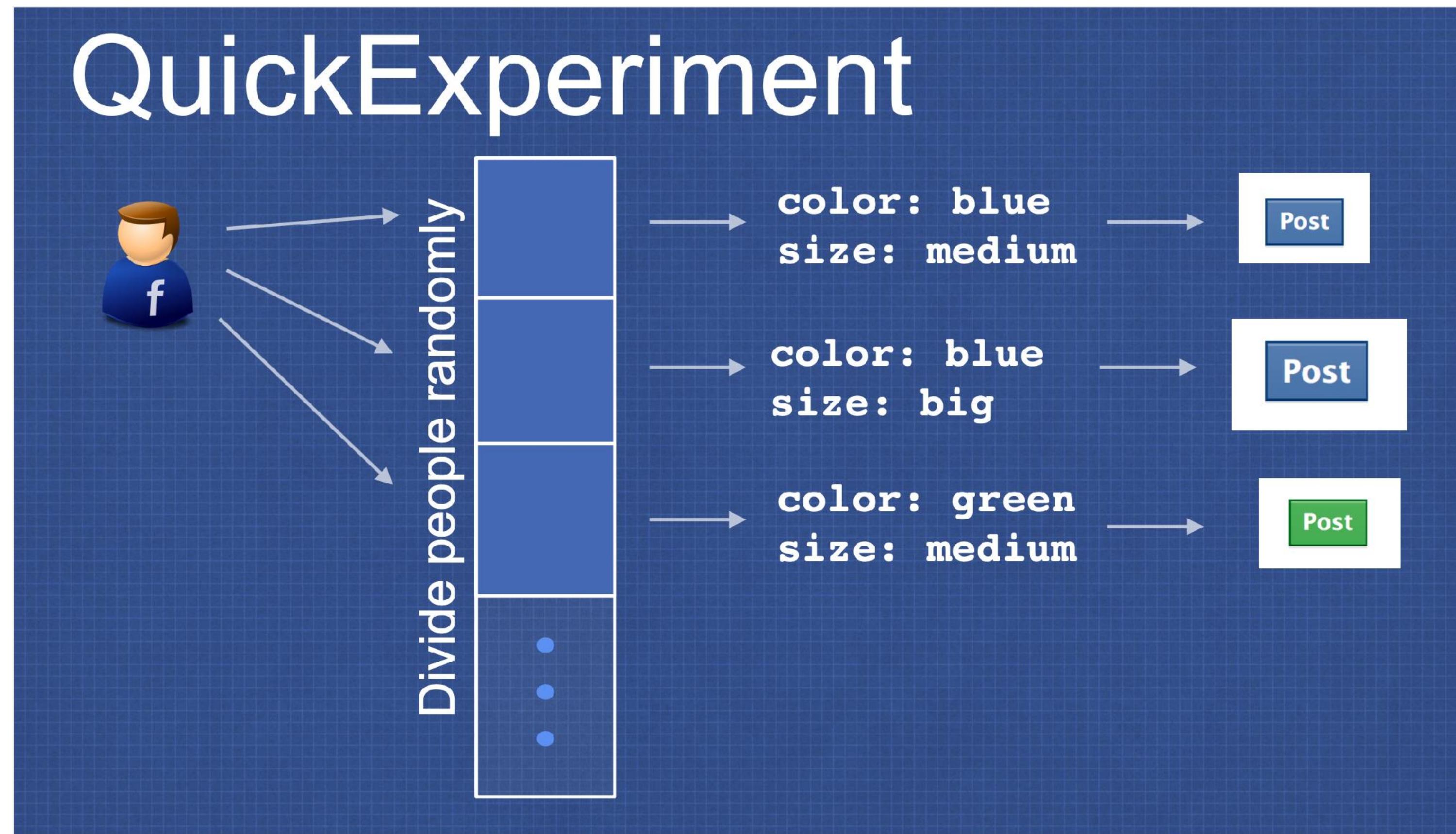
# Usability testing in continuous development

A/B Testing: PlanOut from Facebook (“N=10<sup>9</sup> user study”)



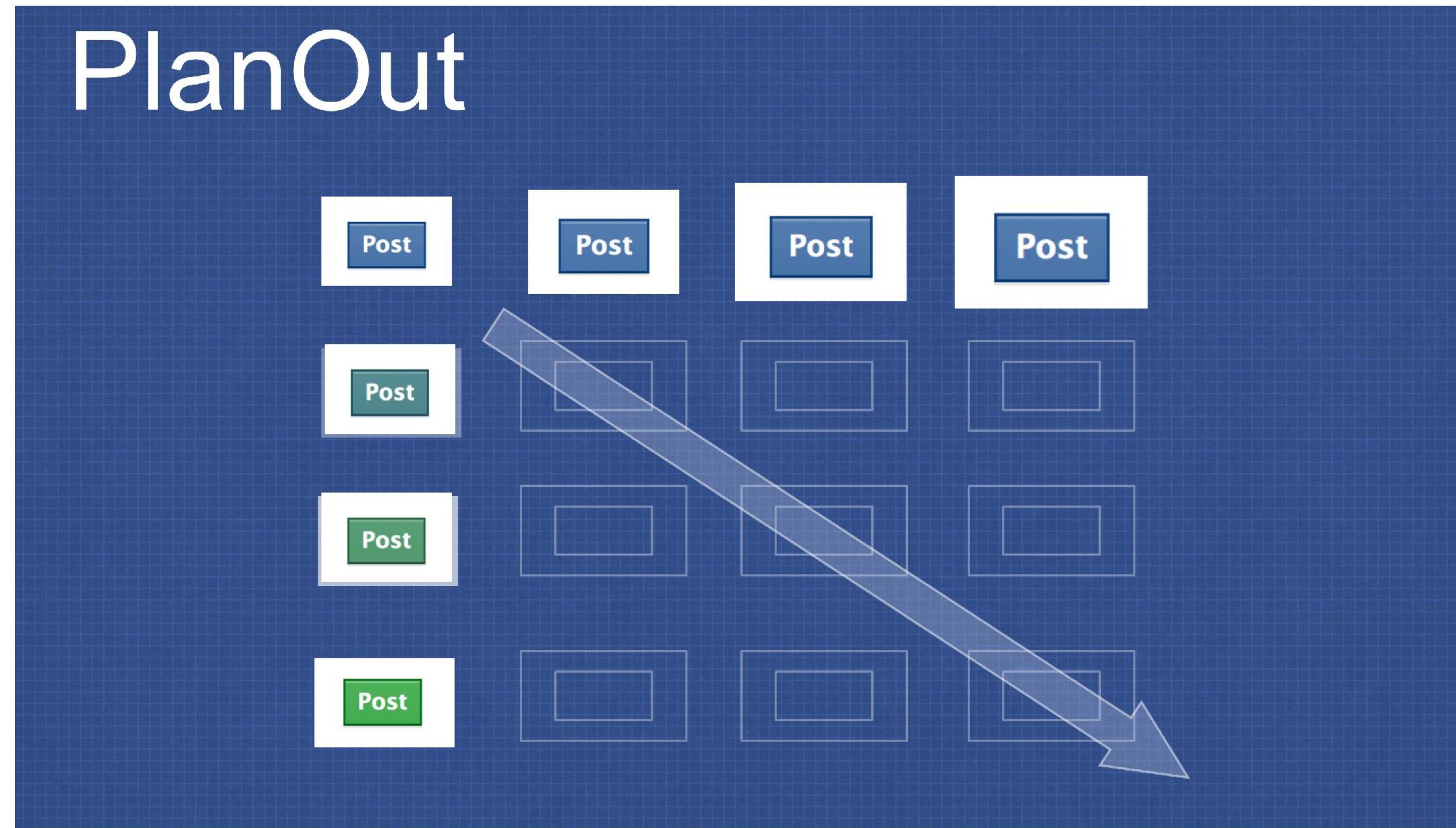
# Usability testing in continuous development

A/B Testing: PlanOut from Facebook (“N=10<sup>9</sup> user study”)



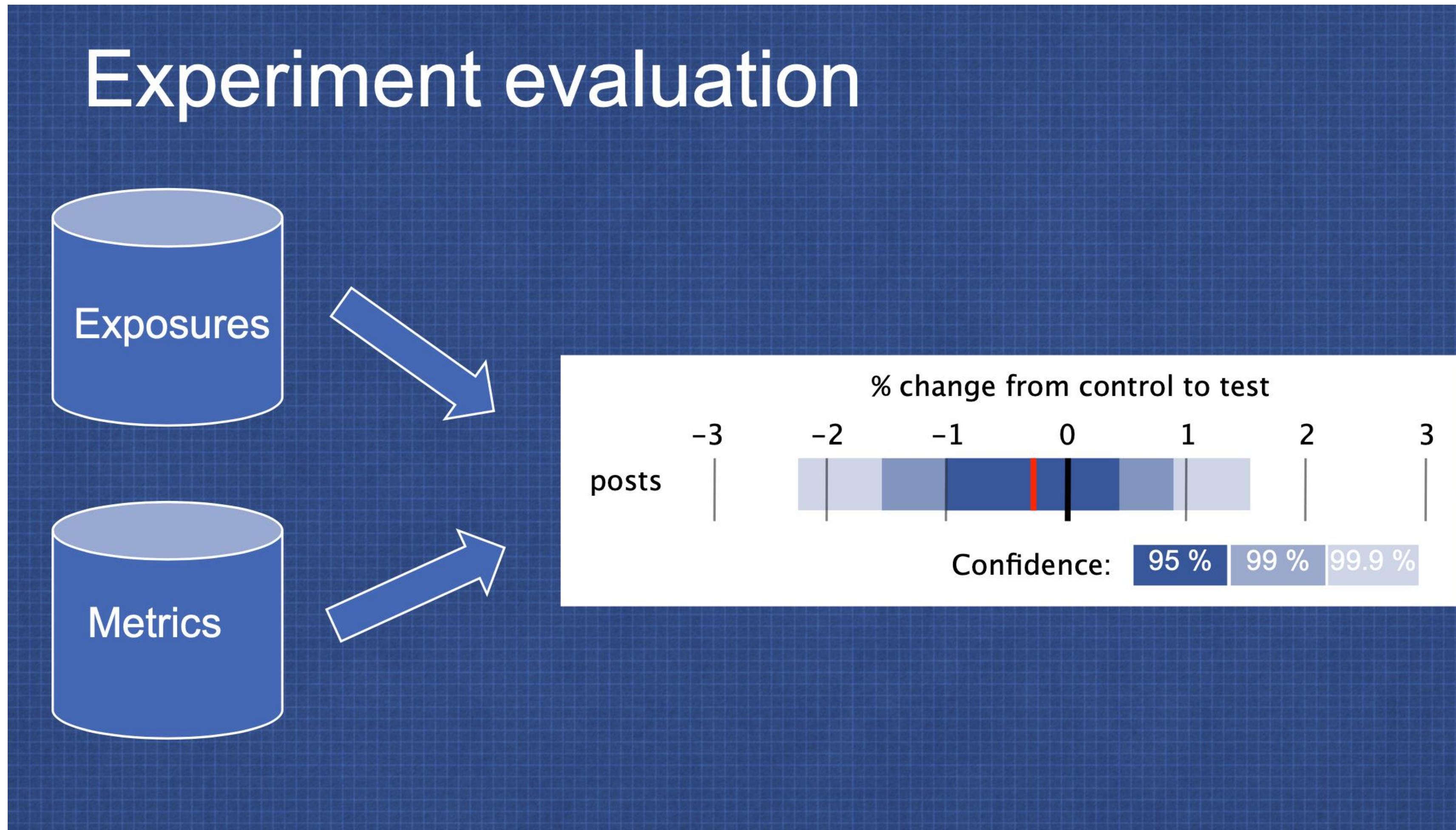
# Usability testing in continuous development

A/B Testing: PlanOut from Facebook (“N=10<sup>9</sup> user study”)



# Usability testing in continuous development

A/B Testing: PlanOut from Facebook (“N=10<sup>9</sup> user study”)



# Beware of Metrics

## McNamara Fallacy

- Measure whatever can be easily measured
- Disregard that which cannot be measured easily
- Presume that which cannot be measured easily is not important
- Presume that which cannot be measured easily does not exist



# Review

---

By now, you should be able to...

- Describe how continuous integration helps to catch errors sooner in the software lifecycle
- Describe the benefits of a culture of code review
- Describe strategies for performing quality-assurance on software as and after it is delivered