SAUCE LABS CONTINUOUS TESTING BENCHMARK REPORT

Q2 2019

A NOTE FROM THE CEO

If you think about it, continuous testing is really about continuous improvement. If your business relies on web and mobile applications to connect with and deliver products and services to customers --- and what business today doesn't? --- so much of your success depends on your ability to continually improve the quality of your offering. Faster releases. New features. Enhanced functionality. Better performance. Greater visual appeal. You name it, we're all on the hook to deliver it each and every day.

That's why, when I started my journey at Sauce Labs, I was surprised to find that there was really no standard industry benchmark that organizations could use to see how their continuous testing efforts stacked up not only against key best practices, but against those of other organizations endeavoring to do the same. An industry that's all about continuous improvement had no reliable benchmark against which to track improvement. It just didn't add up.

I'm biased, of course, but one of the things that I believe sets Sauce Labs apart is that when we see a gap in the market, we don't wait around for someone else to address it. So when I started asking around about the lack of an industry standard continuous testing benchmark, I kept getting the same response: let's create one!

Many hours of hard work and detailed data analysis later, we're both proud and excited to share with you the first iteration of the Sauce Labs Continuous Testing Benchmark. We may have created it, but your collective commitment to continuous testing excellence is what makes this report possible. Wherever you are on your continuous testing journey, we hope it serves you well.

CHARLES RAMSEY
CEO, SAUCE LABS

VISA splunk> GoDaddy Walmart

TABLE OF CONTENTS

EXECUTIVE SUMMARY

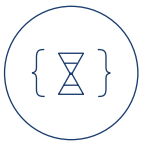
What is the Sauce Labs Continuous Testing Benchmark?

The Sauce Labs Continuous Testing Benchmark defines the best practices necessary to achieve continuous testing excellence, and tracks organizations' collective performance against those metrics based on real-world data. It is derived by observing patterns embedded in the more than 2 billion tests that have been performed on the company's continuous testing platform.

COMPONENTS OF THE SAUCE LABS CONTINUOUS TESTING BENCHMARK

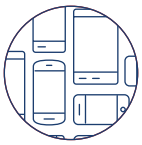
Test Quality

Benchmark for Excellence: Pass at least 90% of all tests run



Test Run Time

Benchmark for Excellence: Test run times averaging 2 minutes or less



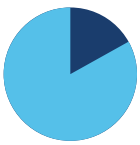
Test Platform Coverage

Benchmark for Excellence: Test against at least 5 platforms on average

Test Concurrency

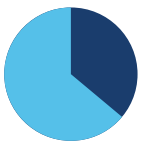
Benchmark for Excellence: Leverage at least 75% of available test capacity during peak testing periods

HOW ORGANIZATIONS PERFORMED



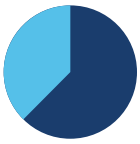
Test Quality

18.8% of organizations pass at least 90% of the tests they run



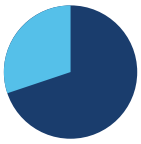
Test Run Time

35.9% of organizations complete their tests in an average of 2 minutes or less



Test Platform Coverage

62.5% of organizations test across 5 or more platforms on average



Test Concurrency

70.9% of organizations utilize at least 75% of their available testing capacity during peak testing

JUST 6.2% OF ORGANIZATIONS ACHIEVED THE BENCHMARK FOR EXCELLENCE IN ALL FOUR CATEGORIES

DATA AND METHODOLOGY

The data used in this report comes from the Sauce Labs Continuous Testing Cloud, the largest and most comprehensive testing platform in the world. The Sauce Labs platform is used by hundreds of the world's largest companies to test their web and mobile applications. It includes cross-browser testing, mobile emulators and simulators, live desktop and mobile testing for the web, live app testing, and the Sauce Labs Real Device Cloud, which allows users to test applications on real mobile devices. In total, the Sauce Labs platform enables users to test across more than 900 browser/OS combinations, more than 200 mobile emulators and simulators, and more than 2,000 real devices.

The report itself is divided into two sections. The first section defines the four core components of the newly established Sauce Labs Continuous Testing Benchmark, and then examines organizations' collective performance against those metrics. For this section, we anonymized and analyzed Sauce Labs test data from June 2018 through December 2018. Of note, though some organizations are running a significantly larger number of tests on the Sauce Labs platform than others, for the purposes of this study, each organization's impact on the overall performance against a particular metric is weighted equally.

The second section takes a closer look at current testing implementations among Sauce Labs users in that same timeframe to better understand, among other things, the browsers and devices against which organizations are most commonly testing. A small portion of the data in this section is collected from external sources, and these instances are clearly credited.

INTRODUCTION

FROM "THE STATE OF TESTING" TO "THE CONTINUOUS TESTING BENCHMARK"

Welcome to the inaugural Sauce Labs Continuous Testing Benchmark Report. In previous years, we have conducted "State of Testing" reports based on survey results from developers and quality assurance teams. Those reports consistently show that while the majority of organizations have shifted to DevOps-driven, agile development methodologies, and the majority of development teams are keenly aware of the important role testing plays in those processes, they are still behind in actually implementing effective continuous testing throughout the software development lifecycle.

We realize, however, that there is no data like direct data that measures actual user behaviors and choices in the real world. This is especially true with testing. Therefore, for 2019, Sauce Labs is transitioning from an industry survey to a benchmark report driven by the anonymized and aggregated testing activities of actual users of the Sauce Labs Continuous Testing Cloud.

Our aim is two-fold. We want to provide organizations with a clearly defined set of metrics against which they can measure their own performance and toward which they can strive in their pursuit of continuous testing excellence. Equally important, we want to provide a snapshot of how current users of the Sauce Labs platform perform against those benchmark metrics, so as to give development teams --- whether they use the Sauce Labs platform or not --- a window into how their efforts stack up against those of their peers in the testing world.

What qualifies Sauce Labs to create such a benchmark based solely on observations and data derived from its own platform? For starters, more than 2 million tests are processed every day on the Sauce Labs Continuous Testing Cloud, and more than 2 billion have been processed since its inception, making it the most widely used continuous testing platform in the world. By observing the patterns embedded in these billions of tests and aggregating the real-world performance of actual end-users, we can provide the kind of benchmark report that has never before existed in the continuous testing world, one that combines our point of view as the industry's continuous testing leader with hard data gleaned from millions of anonymized user tests.

Ultimately, this report is for you --- the testing community. We hope you enjoy reading it as much as we enjoyed creating it.

PART 1: INTRODUCING THE SAUCE LABS CONTINUOUS TESTING BENCHMARK

Organizations that understand the value of a fast quality feedback cycle are rapidly adopting automated testing technology to accelerate application development and shorten the release cycle. The Sauce Labs Continuous Testing Benchmark (CTB) is intended to help organizations and their development teams understand and measure how their automated testing stacks up against key benchmark metrics, as well as against the performance of other forward-thinking organizations that are also adopting automated testing.

The Sauce Labs CTB is comprised of the following four equally weighted performance metrics, each of which is described in greater detail later in this report:

- **Test Quality**, for which excellence is defined by passing at least 90% of all tests run
- **Test Run Time**, for which excellence is defined by running tests in an average of two minutes or less
- **Test Platform Coverage**, for which excellence is defined by testing against at least 5 platforms on average
- **Test Concurrency**, for which excellence is defined by leveraging at least 75% percent of available test capacity during peak testing periods to drive maximum efficiency and parallelization

These metrics are averaged together to produce an overall CTB score, which ranges from 0-100. An organization that earns a score of 100 is achieving excellence across all four of these performance metrics, and is thus fully leveraging the value of continuous testing.

BENCHMARK METRIC #1: TEST QUALITY

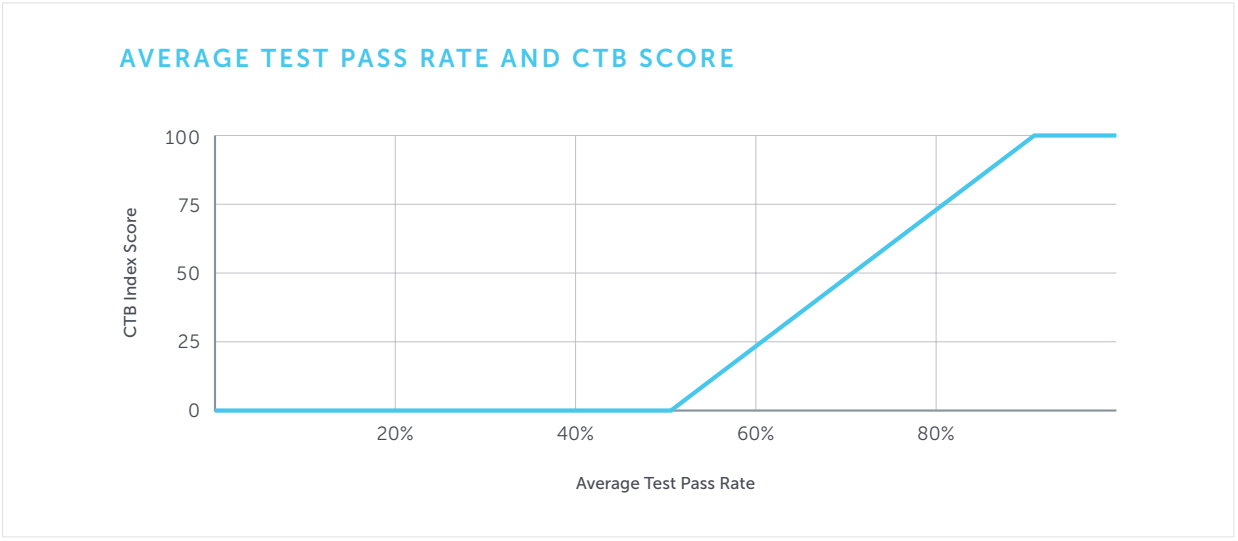
When a test fails, a developer should be able to safely assume that something broke in their application. As simple as that sounds, many organizations struggle with "flakiness" in their test suites, which causes them to treat a failed test as inconclusive. When this happens, manual follow-up is required to determine if the failure truly was caused by a code change, or if there was a problem with the test suite itself.

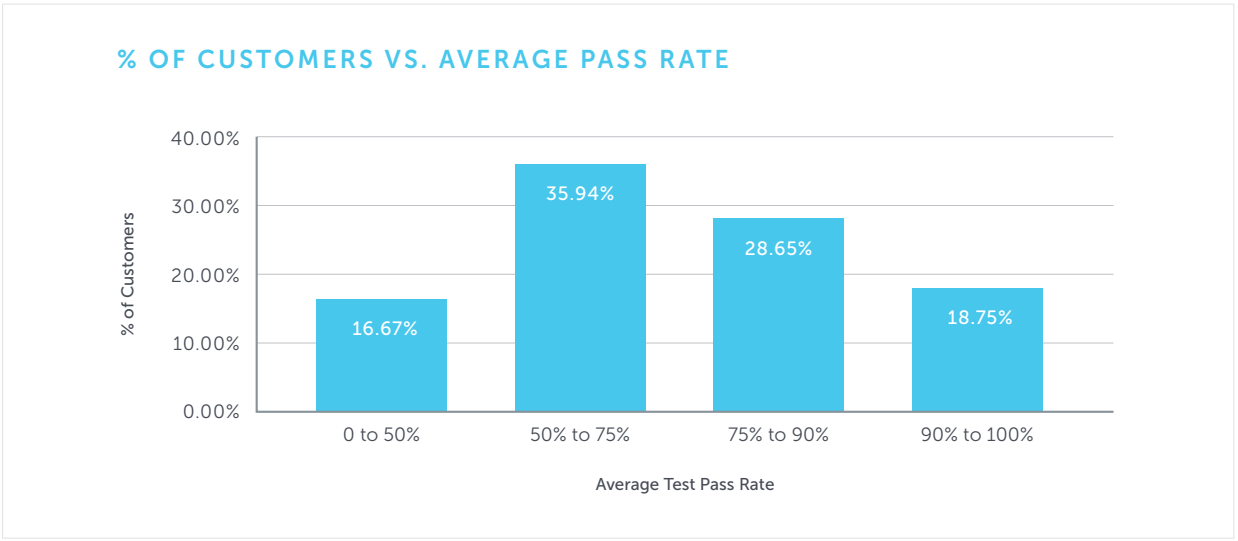
Why it matters: Testers should strive to have as low a failure rate as possible, so that a failed test carries a high signal-to-noise ratio, and developers can feel confident that failed test truly detected a breaking change as opposed to the "flaky" noise of an unreliable test suite.

The Benchmark for Excellence: The CTB Test Quality metric looks at the typical pass rate of an organization's tests. Having the highest possible pass rate is, of course, important. As is outlined above, however, the primary pain point with respect to a failed test is the manual follow up it necessitates. What's more important than the

specific pass rate then is the extent to which an organization has the bandwidth and resources to manually follow up on its failed tests. Test quality becomes a major concern when an organization reaches the point at which the number of failed tests it incurs exceeds its bandwidth to manually follow up on those failures. For most organizations, that breaking point tends to occur when they're passing less than 90% of tests.

Therefore, to achieve a test quality score of 100 signifying excellence, the average pass rate of all tests should be at or above 90%. However, for smaller organizations with greater bandwidth constraints, pass rates may need to be even higher to truly achieve excellence. Pass rates below 50% get a score of 0, signifying that a failed test happens just as often as a passing test.





How Organizations Performed: Only 18.75% of organizations pass at least 90% of their tests, while 16.67% of users pass fewer than 50% of their tests.

The Takeaway: As detailed earlier, not all failed tests are created equal, and in fact, not all failed tests are necessarily a bad thing. After all, the reason you run tests in the first place is to discover bugs and fix them before

an application reaches production and the user experience is compromised. So when a test fails because new code caused something in the application to break, that test has served its purpose well. On the other hand, when a test fails because of a failure to maintain the underlying test script, it's a sign that an organization's test suite is not keeping pace with the application changes it's pushing.

For the purposes of this assessment, however, no distinction is made between tests that failed on account of code changes and tests that failed on account of test suite maintenance. Given that organizations generally fix faulty code immediately upon identifying it, it's likely the high volume of failed tests evident in this data is reflective of organizations' ongoing struggle to maintain their test suites.

Whatever the case, the user performance data demonstrates a clear need for organizations to place greater emphasis on designing and maintaining test suites in a manner that will lead to better pass/fail ratios, and allow them to avoid scenarios where the number of failed tests exceeds their bandwidth to implement the appropriate manual follow up.

BENCHMARK METRIC #2: TEST RUN TIME

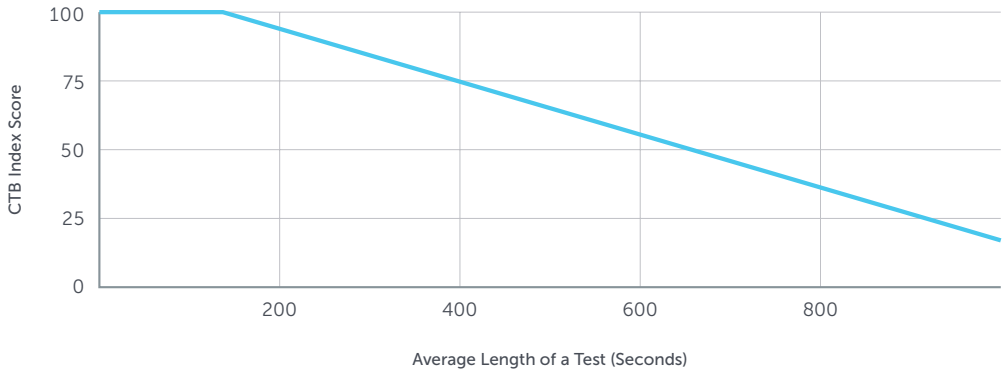
One of the primary reasons to invest in automated testing is the ability to quickly know that an application is working as expected as new code is rapidly pushed. In the world of agile development and fast release cycles, speed matters. The longer your tests take to run, the longer it takes to get feedback to your developers, and the longer it takes to push new code into production. That's why a major goal of any automated test suite should be short, targeted, fast-running tests that quickly evaluate a specific unit of application functionality.

Why it Matters: Enforcing test run time standards ensures that you don't end up with non-specific tests that assess many different pieces of functionality while traveling a long, meandering path through an application. Long tests make it easy to lump multiple functional tests together, which in turn makes it harder to quickly understand precisely what's gone wrong when a test fails. It is far better to have 10 short tests, each responsible for signaling when a specific application function is broken, rather than one long test that accounts for 10 completely different things.

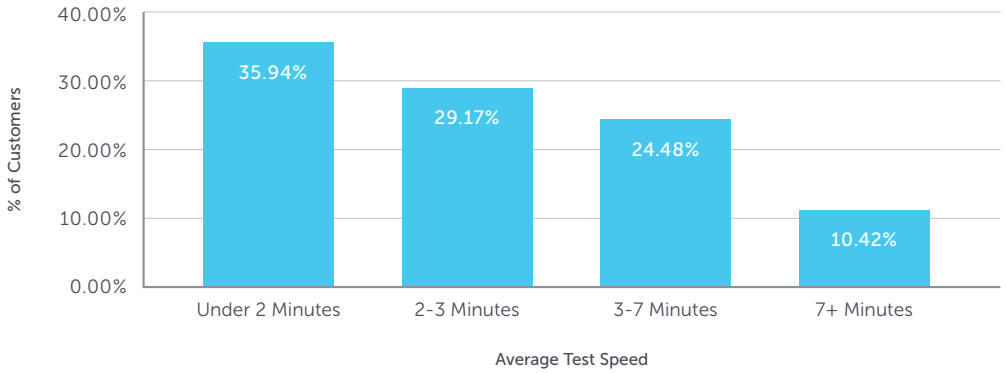
Long tests are also more likely to lead to flakiness, as the longer a test is, the more difficult it is to maintain. That's why Sauce Labs strongly encourages tests that are both atomic (meaning they focus on just a single application function) and autonomous (meaning they run completely independent of other tests).

The Benchmark for Excellence: The CTB Test Run Time metric looks at the average run time of the tests organizations are running on the Sauce Labs cloud. To achieve a score of 100 signifying excellence, the average run time of an organization's tests should be 2 minutes or less. Based on the entirety of test data assessed in this benchmark, tests that complete in 2 minutes or less are twice as likely to pass as tests lasting longer than two minutes. In other words, the longer a test takes to run, the more likely it is to fail.

AVERAGE TEST EXECUTION SECONDS AND CTB SCORE



% OF CUSTOMERS VS. AVERAGE TEST SPEED



How Organizations Performed: Only 35.94% of organizations complete their tests in an average of 2 minutes or less.

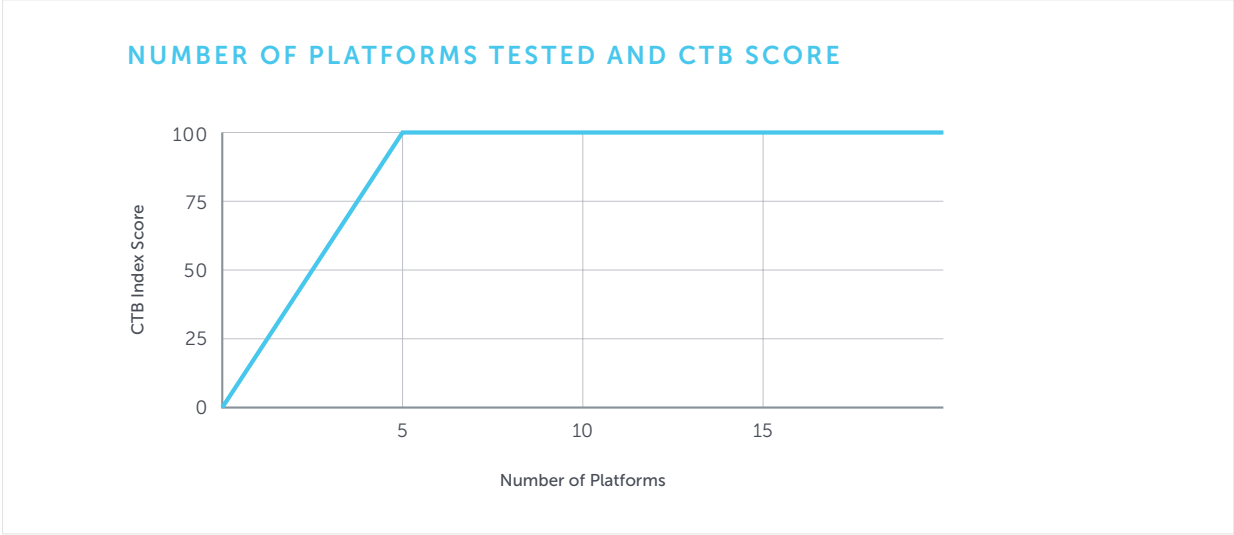
The Takeaway: With nearly two-thirds of organizations failing to achieve the desired benchmark for excellence, test run time stands out as a key area of potential improvement. This is especially true given the aforementioned correlation between test quality and test run time. Running atomic and autonomous tests increases the speed with which developers receive feedback while decreasing the rate at which tests fail. That's a win-win for agile development.

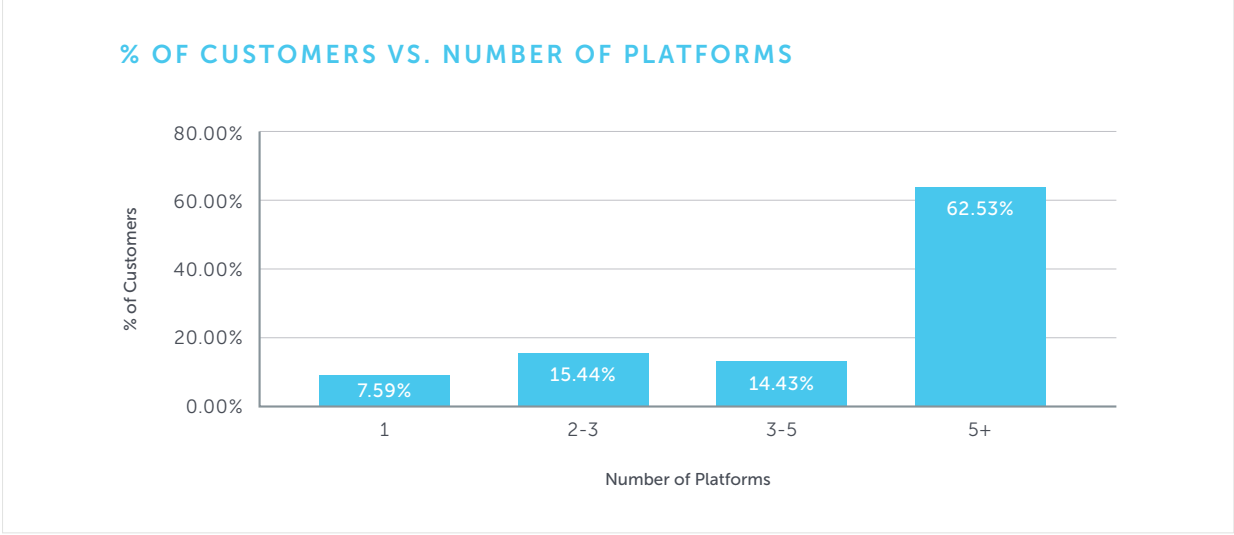
BENCHMARK METRIC #3: TEST PLATFORM COVERAGE

The ability to quickly determine if an application functions correctly across a wide variety of platforms is a critical component of effective continuous testing. For the purposes of this report, a platform is defined as any combination of an operating system (desktop or mobile) and a browser. The definition does not, however, distinguish between different versions of the same browser.

Why it matters: Customers in today's increasingly digital world consume information and services across an ever-growing range of platforms and devices. Delivering a flawless digital experience means rapidly delivering apps that work as intended whenever, wherever and however customers wish to access them.

The Benchmark for Excellence: The CTB Test Platform Coverage metric looks at the number of platforms against which an organization tests, and gives full marks if at least 5 platforms are included in their typical testing activity.





How Organizations Performed: Nearly two-thirds (62.53%) of organizations test across 5 or more platforms on average, and more than three-quarters of users (76.96%) test across at least 3 or more platforms on average.

The Takeaway: Though there is room for improvement --- 7.59% of users are still testing against just one platform on average --- the performance data indicates that the majority of organizations have a strong understanding of the importance of testing against multiple platforms, and that robust test platform coverage is a priority for most development teams.

BENCHMARK METRIC #4: TEST CONCURRENCY

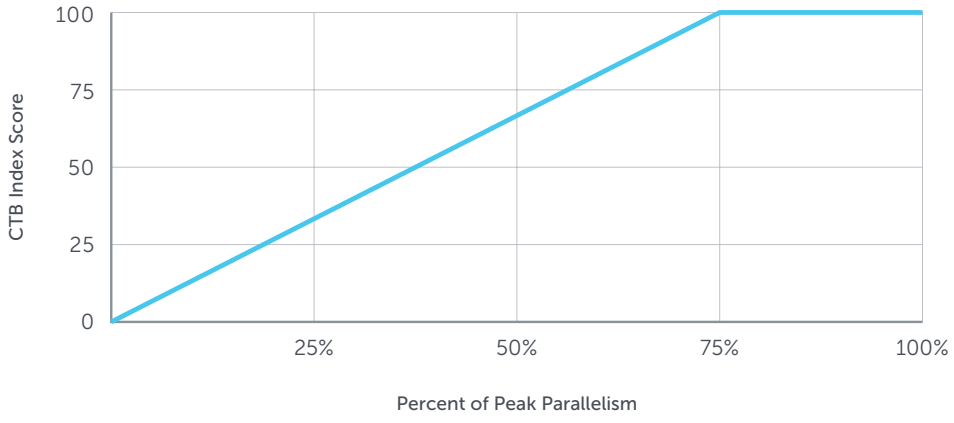
Though run time, quality and coverage are all critical components in their own right (and thus make up three-quarters of the overall benchmark), it can be argued that the secret sauce behind effective continuous testing is the ability to run multiple tests concurrently.

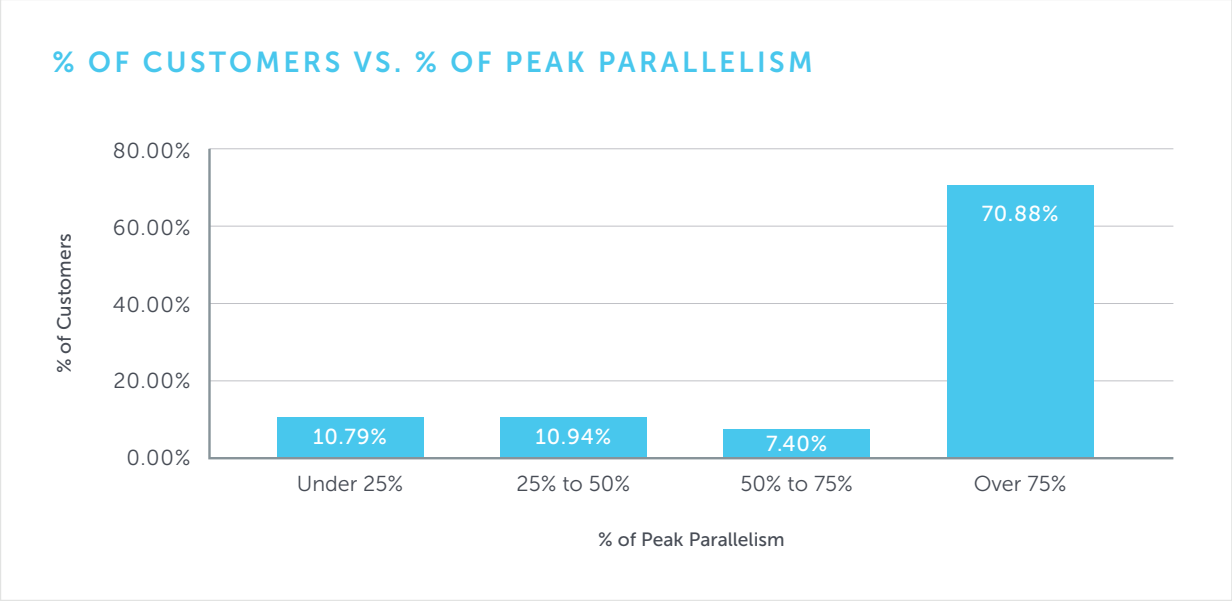
Why it matters: Take the simple example of a test suite where a typical test takes 2 minutes to run, and the suite includes 100 functional tests. If all 100 tests are run concurrently, the entire process will take just 2 minutes. If an organization is capacity constrained with regards to concurrency, however, tests have to wait in line, and the same suite of 100 tests can take more than 3 hours to complete. This dramatically slows down the evaluation of application quality, which in turn slows down the entire development cycle.

The hypothetical example above presumes that an organization needs additional test capacity, but simply does not have it, and therefore cannot leverage concurrency to run their tests efficiently. An equally common and arguably more concerning scenario is the opposite one, in which organizations have the necessary test capacity, but simply do not use it. Akin to paying for a gym membership and never going, failure to maximize available test capacity is inefficient from both an operational and cost standpoint.

The Benchmark for Excellence: The CTB Concurrency metric measures the percentage of provisioned testing capacity an organization utilizes when a test suite is running. The greater the percentage of test capacity leveraged, the more efficiently an organization is utilizing concurrency. To achieve a score of 100 signifying excellence, an organization should utilize at least 75% of their available test capacity during peak testing periods.

PARALLELISM AND CTB SCORE





How Organizations Performed: More than 70% of organizations currently utilize at least 75% of their available testing capacity during peak testing periods.

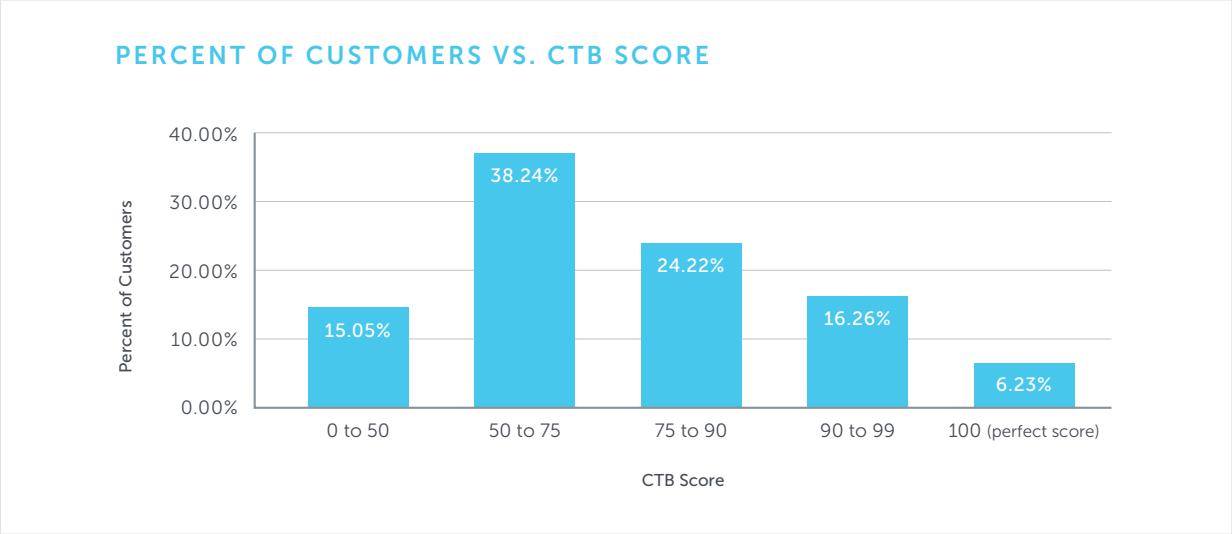
The Takeaway: Though there is once again room for improvement --- roughly 20% of organizations are utilizing less than half of their available capacity when running a test suite --- the large majority of organizations are indeed operating efficiently and making strong use of their available test capacity to run tests concurrently. Those that aren't should pay added attention to designing their test environment to fully saturate their available capacity when running a test suite.

PUTTING IT ALL TOGETHER: THE CONTINUOUS TESTING BENCHMARK

Combining the aforementioned metrics --- test quality, test run time, test platform coverage, and test concurrency --- into a single, equally weighted metric creates the overall Sauce Labs Continuous Testing Benchmark.

Why it Matters: A first-of-its-kind for the continuous testing space, the benchmark provides organizations with a single viewpoint into the efficacy of their overall continuous testing efforts.

Benchmark for Excellence: To achieve a perfect score (100), one signifying complete and total continuous testing excellence, organizations must meet or exceed the excellence benchmark in each of the four component categories as previously outlined.



How Organizations Performed: Only 6.23% of organizations achieved a perfect score of 100 signifying excellence across all four of the benchmark's component metrics.

The Takeaway: Perhaps not surprisingly given the relatively nascent stage of continuous testing as a strategic imperative in the DevOps era, room for improvement exists across the entire spectrum of continuous testing excellence, most notably in the areas of test quality and test run time. However, most organizations are achieving excellence with respect to test platform coverage and test concurrency, putting to smart use the breadth and scalability afforded to them by the Sauce Labs Continuous Testing Cloud. As organizations continue to prioritize continuous testing as the foundation of their agile development efforts, there is ample reason to expect their performance against these metrics to consistently improve over time.

PART 2: USER IMPLEMENTATIONS AT A GLANCE

In the first section of this report, we created the industry's first true continuous testing benchmark, not only providing organizations with clear metrics they can use to measure and assess the proficiency of their own testing efforts, but also with real-world data showing how other organizations currently perform against those metrics. This next section looks behind the curtain of those tests to offer a clear understanding of the platforms, browsers and devices against which Sauce Labs users most often test. Additionally, we break down testing trends across key industry verticals, and take a closer look at the proliferation of tests across desktop and mobile apps.

Though there are a number of existing industry studies that report on this type of information using third-party surveys, this is the first one to do so based on user data taken from actual tests run on the world's largest continuous testing platform.

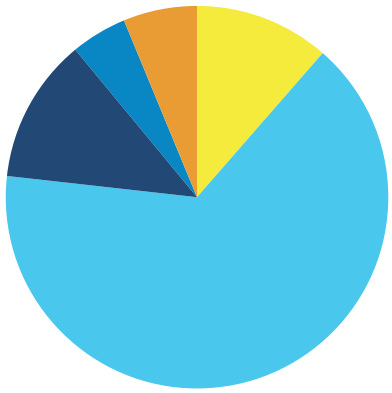
1. TESTING BY DESKTOP BROWSER

This section breaks down the percentage of tests running on the Sauce Labs Continuous Testing Cloud by desktop browser.

Key Findings:

- Nearly two-thirds (65.5%) of all desktop browser tests in the Sauce Labs cloud are run on Google Chrome. This is relatively consistent with the overall market, as Google Chrome has 70.88% of global desktop market share as of January 2019, according to StatCounter.
- Internet Explorer was a distant second at 12.4%, which is nonetheless more than double its global market share of 5.74% according to StatCounter.
- Firefox placed third at 10.6%, slightly above its global market share of 9.5% according to StatCounter.

PROPORTION OF TESTING - DESKTOP BROWSER



Firefox	10.6%
Safari	6.2%
Microsoft Edge	4.7%
iExplore	12.4%
Google Chrome	65.5%

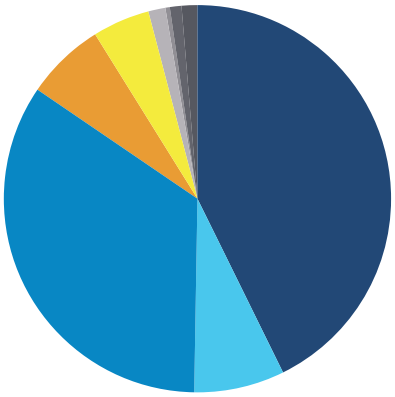
2. DESKTOP BROWSER TESTS BY INDUSTRY VERTICAL

This section looks at the breakdown all of desktop browser tests across specific industry verticals.

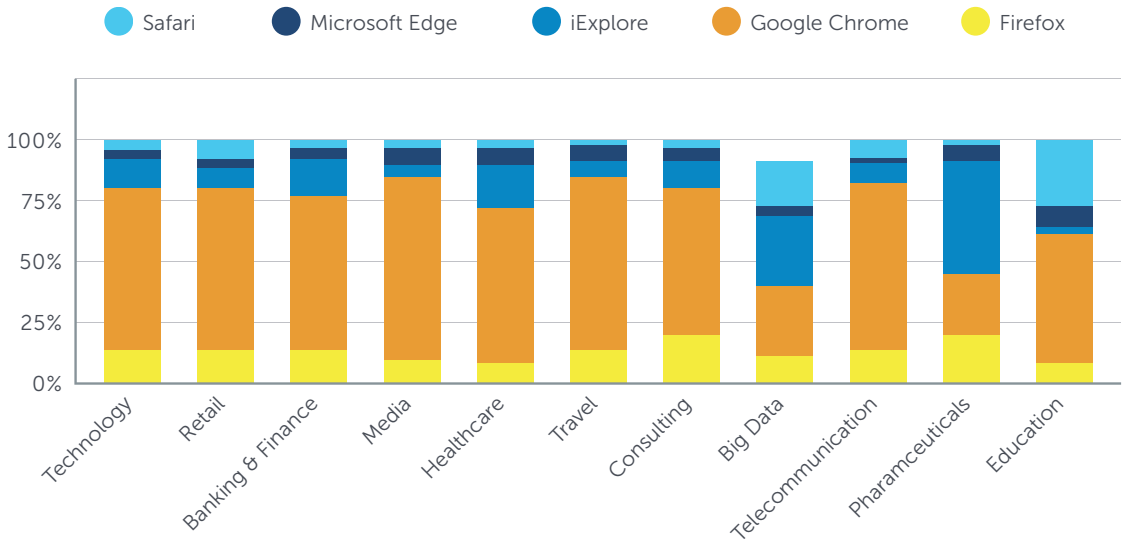
Key Findings:

- The technology and banking sectors lead the way by a significant margin, accounting for 43.1% and 33.9% of all tests, respectively.
- Retail, media and healthcare round out the top five, in that order, but significantly lag the technology and retail sectors.
- Across those top five verticals, Google Chrome is by far the most commonly tested browser, accounting for between 67% and 77% of all desktop browser testing activity.

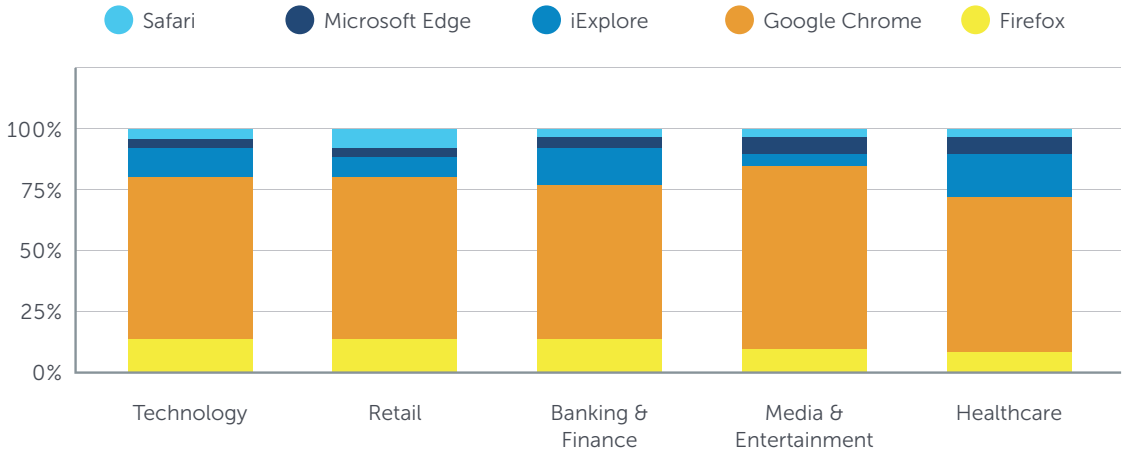
PERCENTAGE OF TESTS RUN



Technology	43.1%
Healthcare	4.7%
Media	6.6%
Banking	33.9%
Retail	7.5%



TOP 5 INDUSTRIES - DESKTOP BROWSERS



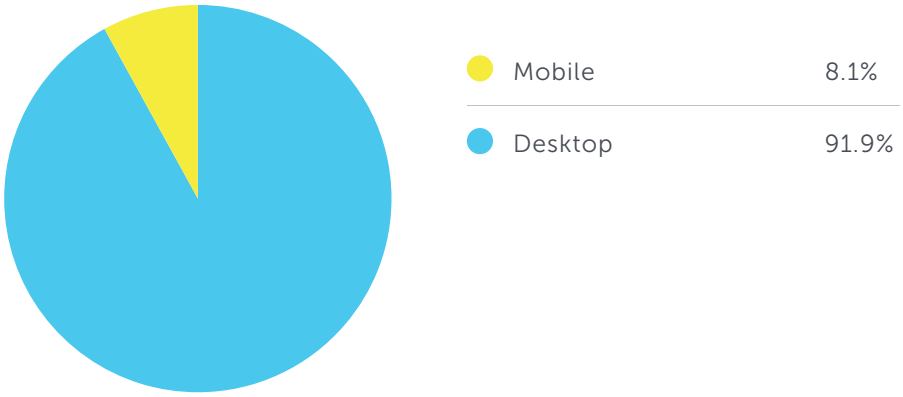
3. TESTING BY PLATFORM: DESKTOP VS. MOBILE

This section looks at the breakdown of tests between desktop and mobile platforms, as well as between mobile web and mobile native apps.

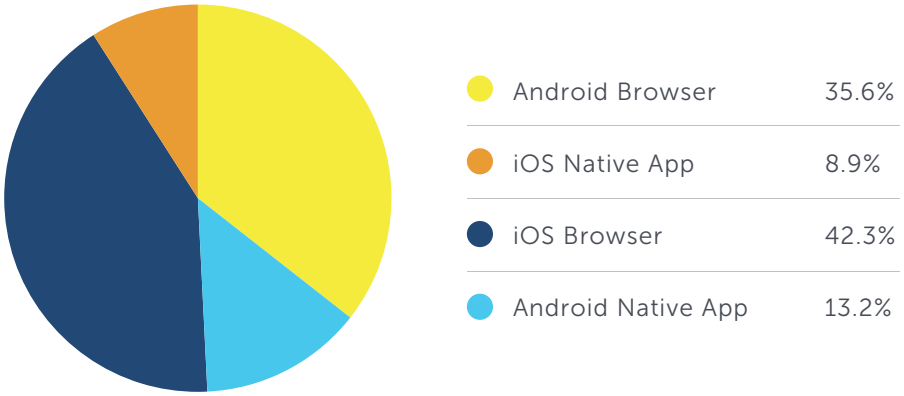
Key Findings:

- Nearly 92% of all tests run on the Sauce Labs Continuous Testing Cloud are conducted on desktop browsers, compared to just 8% on mobile platforms, which includes both mobile browsers and mobile native apps.
- Specific to mobile testing, the large majority of tests (77.9%) are run on mobile web browsers, compared to just 22.1% of tests run on native mobile applications.

PROPORTION OF TESTS - MOBILE VS. DESKTOP



MOBILE APP AND MOBILE WEB TESTING



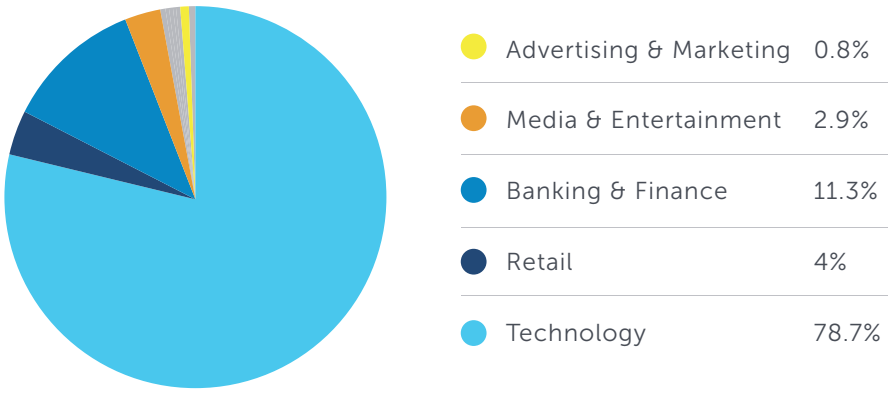
4. MOBILE TESTS BY INDUSTRY VERTICAL

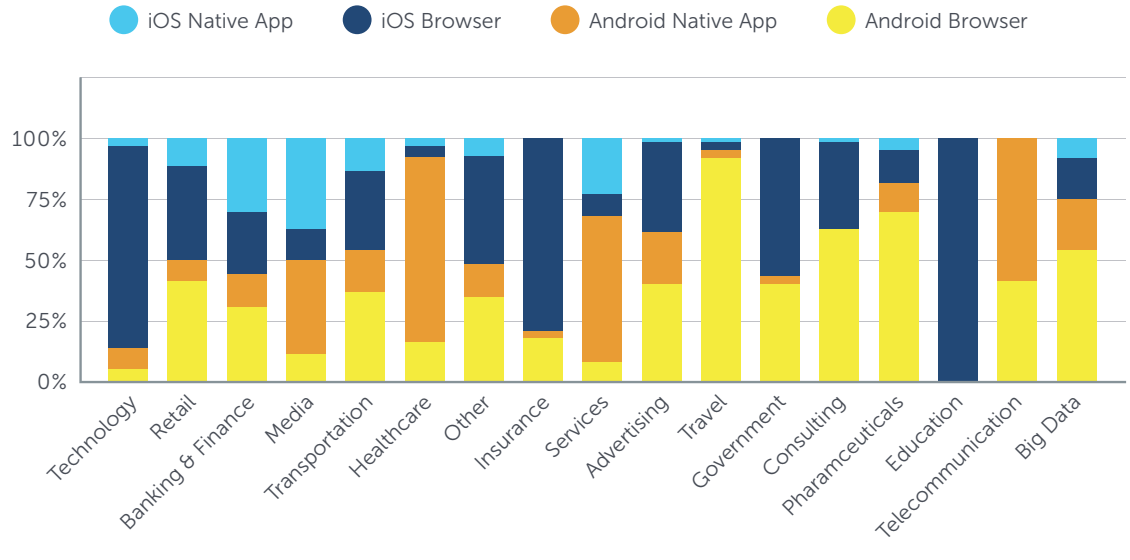
This section looks at the breakdown of all mobile tests (web and native) across specific industry verticals. The data is specifically based on mobile emulator/simulator tests run using the Sauce Labs Continuous Testing Cloud.

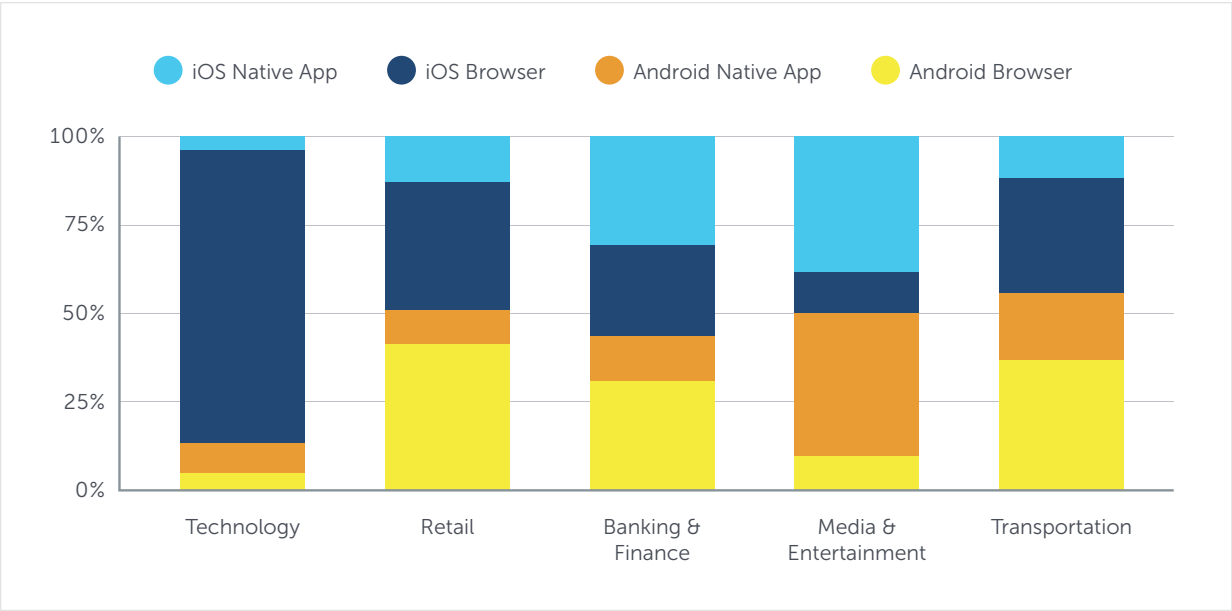
Key Findings:

- As was the case with desktop testing, the technology sector leads the way, accounting for 78.7% of all mobile (web or native) tests run on the Sauce Labs Continuous Testing Cloud.
- The banking and finance sector is a distant second, accounting for 11.3% of all mobile tests.
- Within the technology sector, the majority of tests were conducted on iOS browsers and/or native apps. Across all non-technology sectors, tests were relatively evenly split across iOS and Android browsers and/or native apps.

TESTS RUN BY INDUSTRY







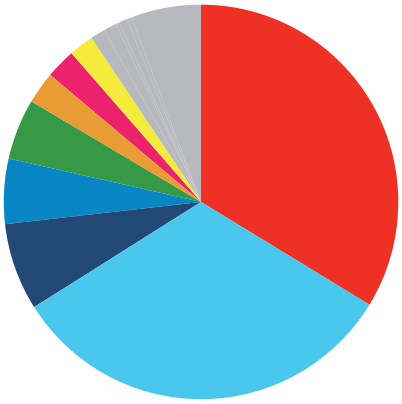
5. REAL DEVICE TESTING

This section specifically examines mobile testing trends on the Sauce Labs Real Device Cloud (RDC), in which tests are conducted using actual physical devices, as opposed to mobile emulators and simulators.

Key Findings:

- Apple and Samsung represent nearly two-thirds (66%) of all testing activity in the Sauce Labs RDC, respectively accounting for 33.6% and 32.4% of tests run. This is not surprising given that Apple and Samsung are the dominant manufacturers and thus the most attractive and most widely used for testing.
- Google, LG and Motorola round out the top five, in that order, with Google achieving 7.2% of the testing share.

RDC: TESTING BY MANUFACTURER



HTC	2.6%
OnePlus	2.0%
Apple	33.6%
Huawei	2.6%
Motorola	5.1%
LG	5.3%
Google	7.2%
Samsung	32.4%

RDC: ACTIVITY BY OS

iOS	35.4%
Android	64.6%

CONCLUSION

Most organizations are still just beginning their continuous testing journeys. The Sauce Labs Continuous Testing Benchmark offers testing teams a set of metrics that can guide them in an effective continuous testing strategy. It also provides real-world data they can use to benchmark the success of their efforts relative to those of other organizations.

The four pillars of the Sauce Labs Continuous Testing benchmark - test quality, test run time, test platform coverage and test concurrency - apply to all organizations that want to succeed with continuous testing. As most continuous testing programs are still in their infancy, the relatively uneven performance of most organizations against these four metrics is not all that surprising. Generally speaking, organizations are indeed achieving the level of test platform coverage and test concurrency needed to succeed with continuous testing. Test quality and test run time, on the other hand, stand as clear areas of improvement for most organizations.

However, the most important metric in this report is not run time or quality, nor is it coverage or concurrency. The most important metric is the overall continuous testing benchmark that combines all four metrics to assess the entirety of an organization's continuous testing efforts. Improving in one area at the expensive of another is counterproductive. Success requires a commitment to excellence and constant improvement across all four pillars of the benchmark. Over time, as organizations become more familiar with the nuances of continuous testing, and development teams increasingly adopt the mindset that quality is everyone's responsibility, that's exactly what we expect to see.

How is my organization doing?

Want to know how the tests your team is running perform against the Sauce Labs Continuous Testing Benchmarks? If you are a current Sauce Labs customer and would like a custom report showing your Continuous Testing Benchmark scores and recommendations, please contact your Customer Success Manager. If you're not a Sauce Labs customer but would like to learn more about improving your overall continuous testing performance, please contact hello@saucelabs.com.

ABOUT SAUCE LABS

Sauce Labs ensures the world's leading apps and websites work flawlessly on every browser, OS and device. Its award-winning Continuous Testing Cloud provides development and quality teams with instant access to the test coverage, scalability, and analytics they need to rapidly deliver a flawless digital experience. Sauce Labs is a privately held company funded by Toba Capital, Salesforce Ventures, Centerview Capital Technology, IVP, Adams Street Partners and Riverwood Capital. For more information, please visit saucelabs.com.

SAUCELABS

SAUCE LABS INC. - HQ
116 NEW MONTGOMERY STREET, 3RD FL
SAN FRANCISCO, CA 94105 USA

SAUCE LABS EUROPE GMBH
C/O WEWORK STRALAUER
ALLEE 6 10245 BERLIN DE

SAUCE LABS INC. - CANADA
134 ABBOTT ST #501
VANCOUVER, BC V6B 2K4 CANADA