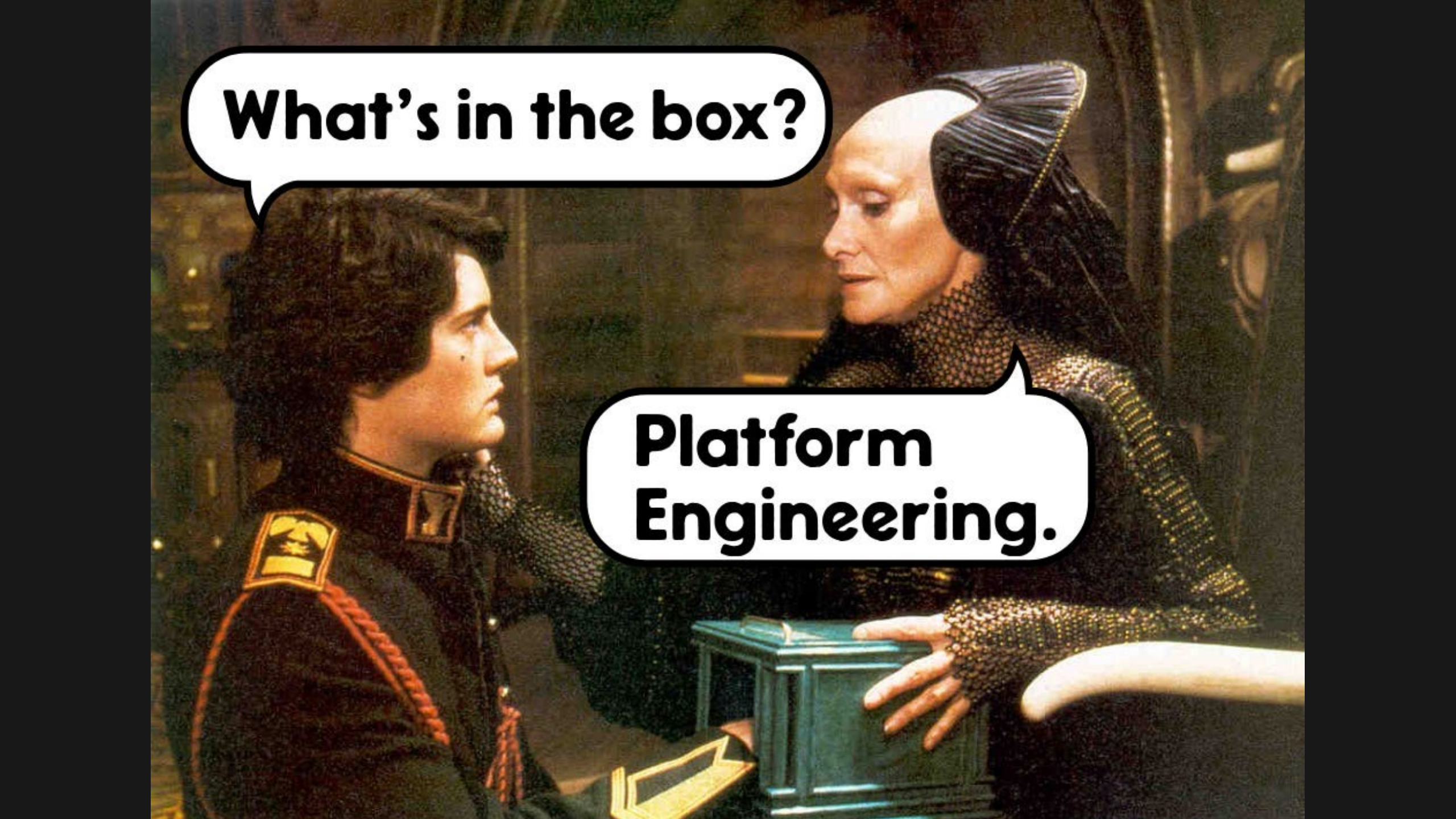


Platform Engineering for Private Cloud

Coté – SRE Day London- March 27th, 2025





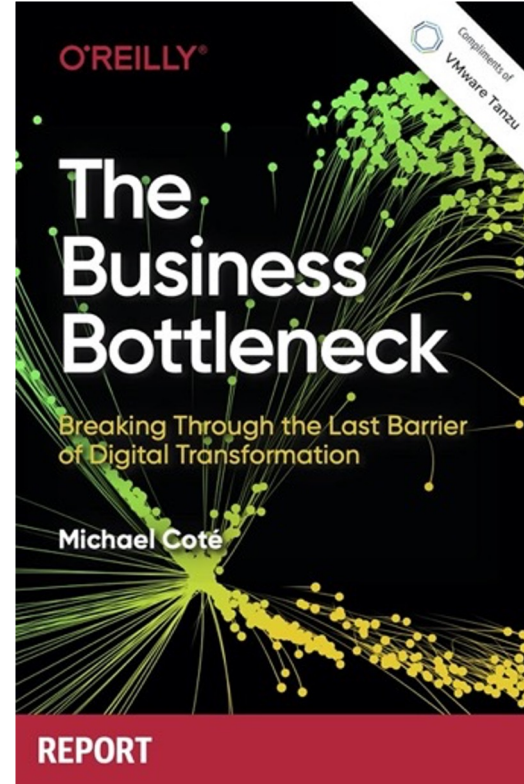
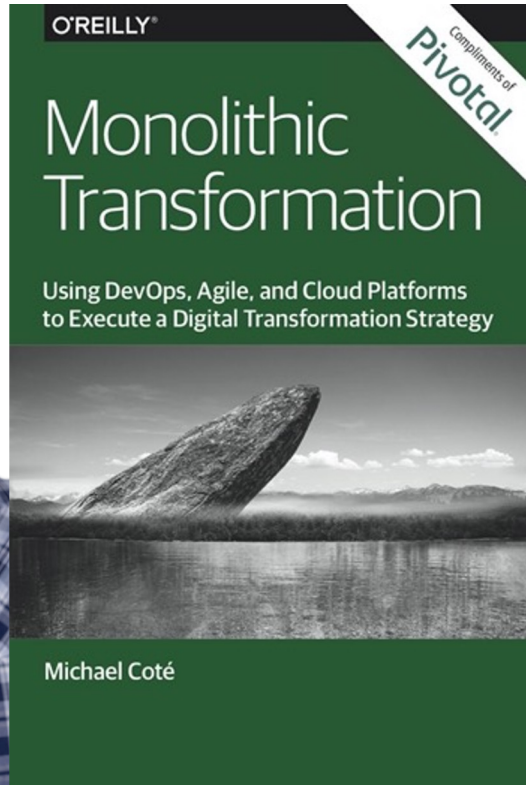
What's in the box?

**Platform
Engineering.**

50%
of enterprise apps
run on
private cloud

Coté

<https://newsletter.cote.io/> | cote@broadcom.com



Where are the apps?



“[W]e do have a lot of large customers that are running in AWS in the cloud today, and a huge number of them still have massive amounts of their estate on-premise. And so there’s a huge amount of growth available there. You can even take

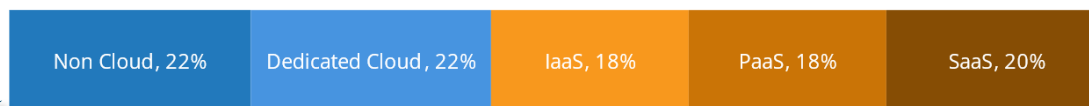
our largest customers, many of them only have 10, 20, 30, or 40 percent of their workloads in the cloud.”

Matt Garman, AWS CEO, January, 2025

Cloud buyers increased their level of cross-platform interoperability in 2024. They also relied more upon providers for the management of their IT environments

Where are all your organizations applications deployed?

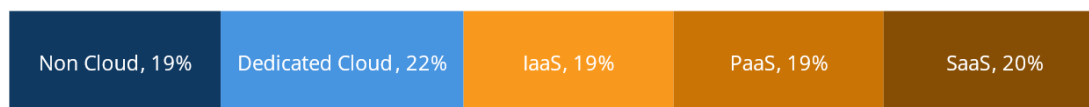
Q3 2024



44%

dedicated environment

Q3 2026



Q3 2028



Level of Interoperability Between Clouds

Public Cloud	41%	Hosted Dedicated Cloud
On Prem Dedicated Cloud	39%	Public Cloud
Hosted Dedicated Cloud	38%	On Prem Dedicated Cloud
Public Cloud	38%	Public Cloud

HOW THE EDGE IS MANAGED

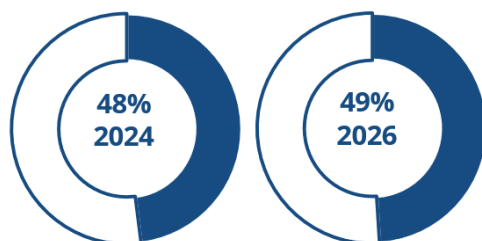
IAAS EDGE

- 49% of IaaS deployments are considered as remote of edge deployments

DEDICATED EDGE

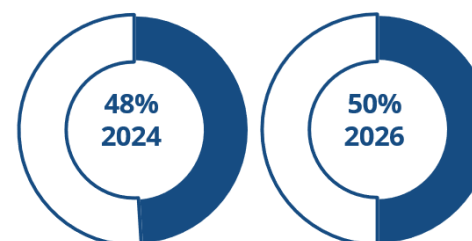
- 48% of dedicated cloud environments are classified as remote or edge deployments.

NON CLOUD



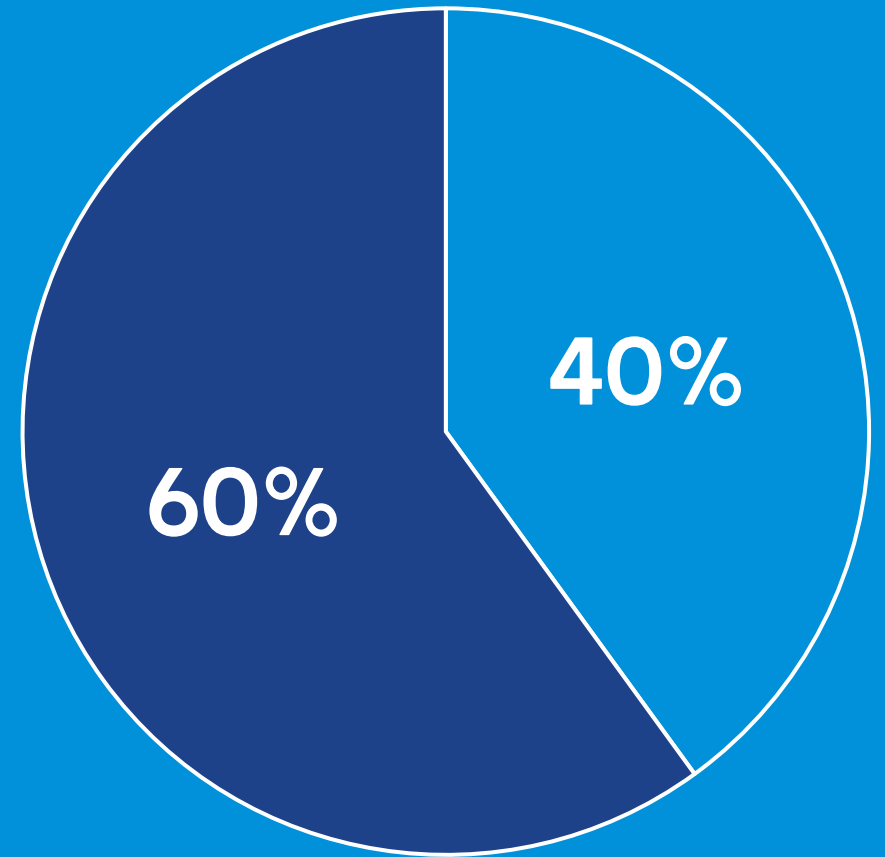
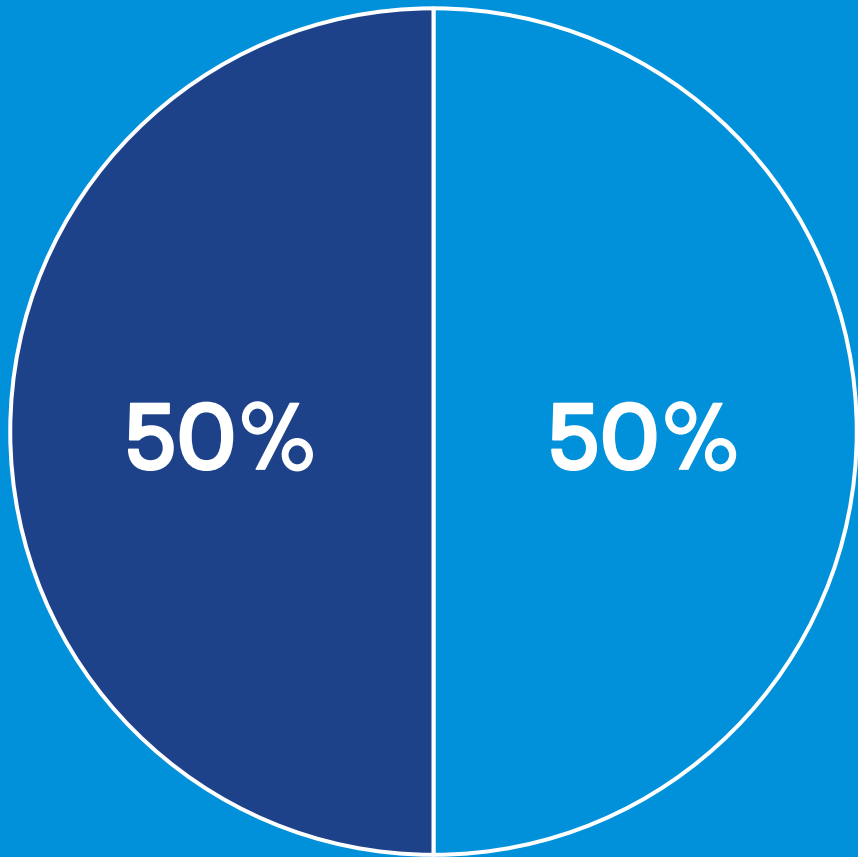
PROVIDER MANAGED ENVIRONMENTS 2024 versus 2026

DEDICATED CLOUD

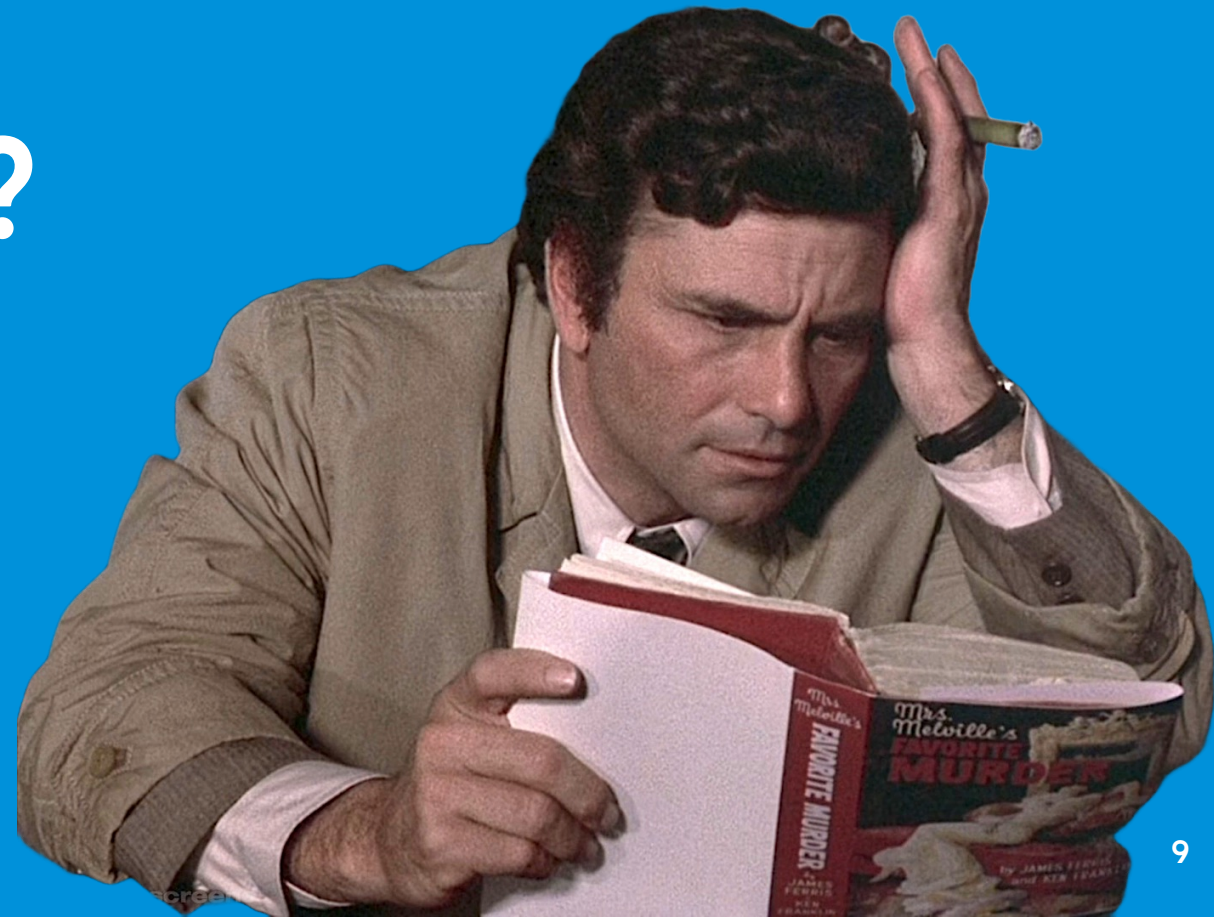


N = 1,724, QC6 Source: QDV1 3Q24 Cloud Pulse Survey, October 2024, IDC

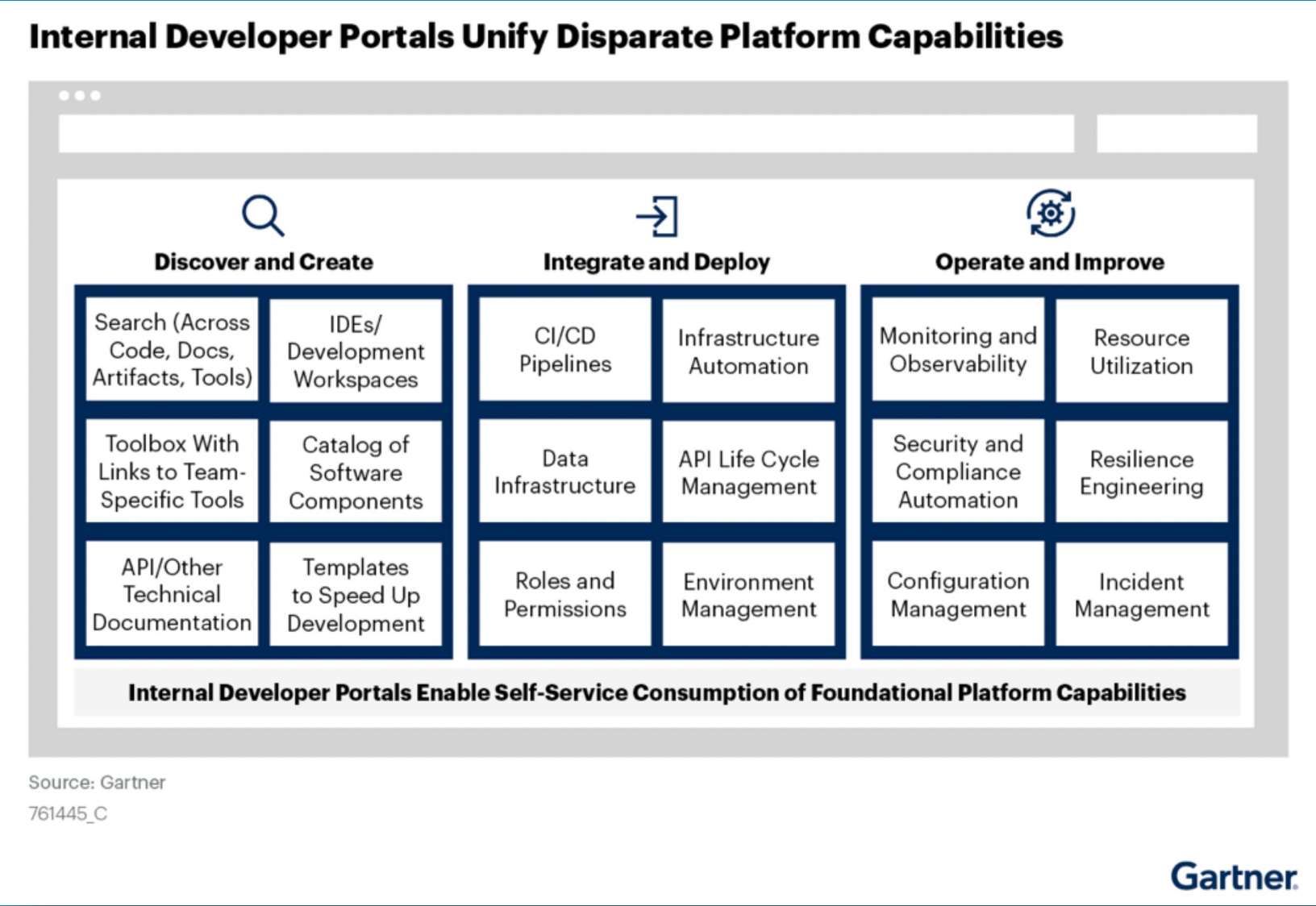
Where the workloads live, rough estimates



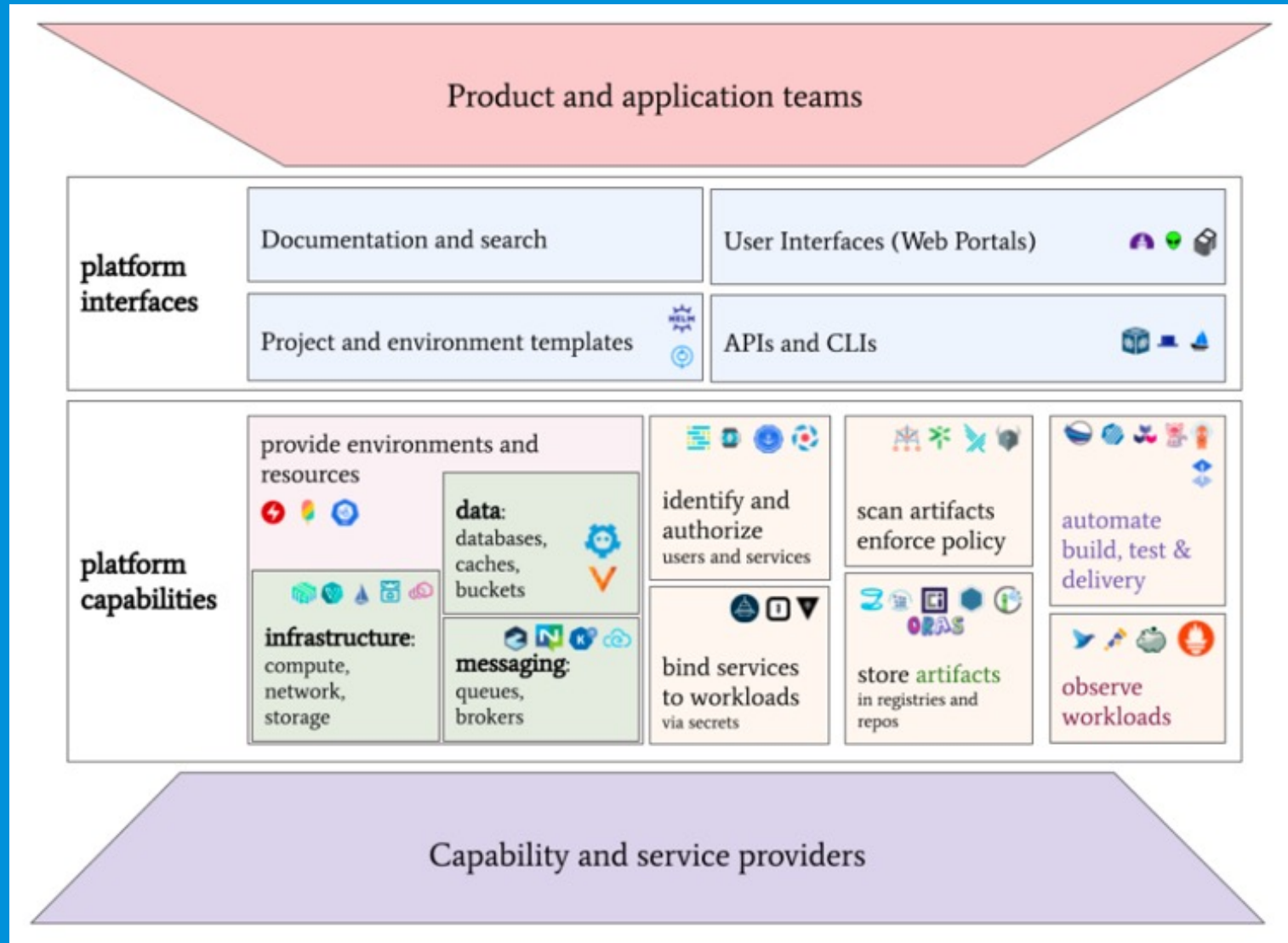
What is a platform?



February, 2022 – Internal Developer PORTAL (IDP)



March, 2023 – Internal Developer PLATFORM



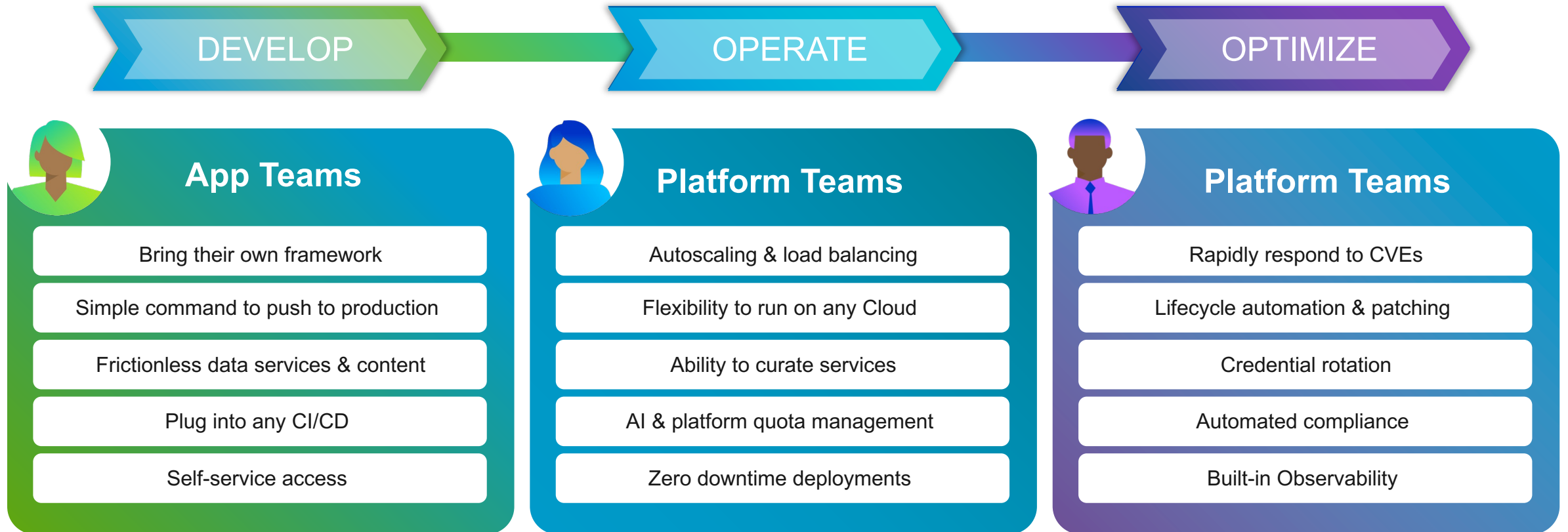


A digital platform is a foundation of self-service APIs, tools, services, knowledge and support which are arranged as a compelling internal product.

[SO THAT] Autonomous delivery teams can make use of the platform to deliver product features at a higher pace, with reduced co-ordination.

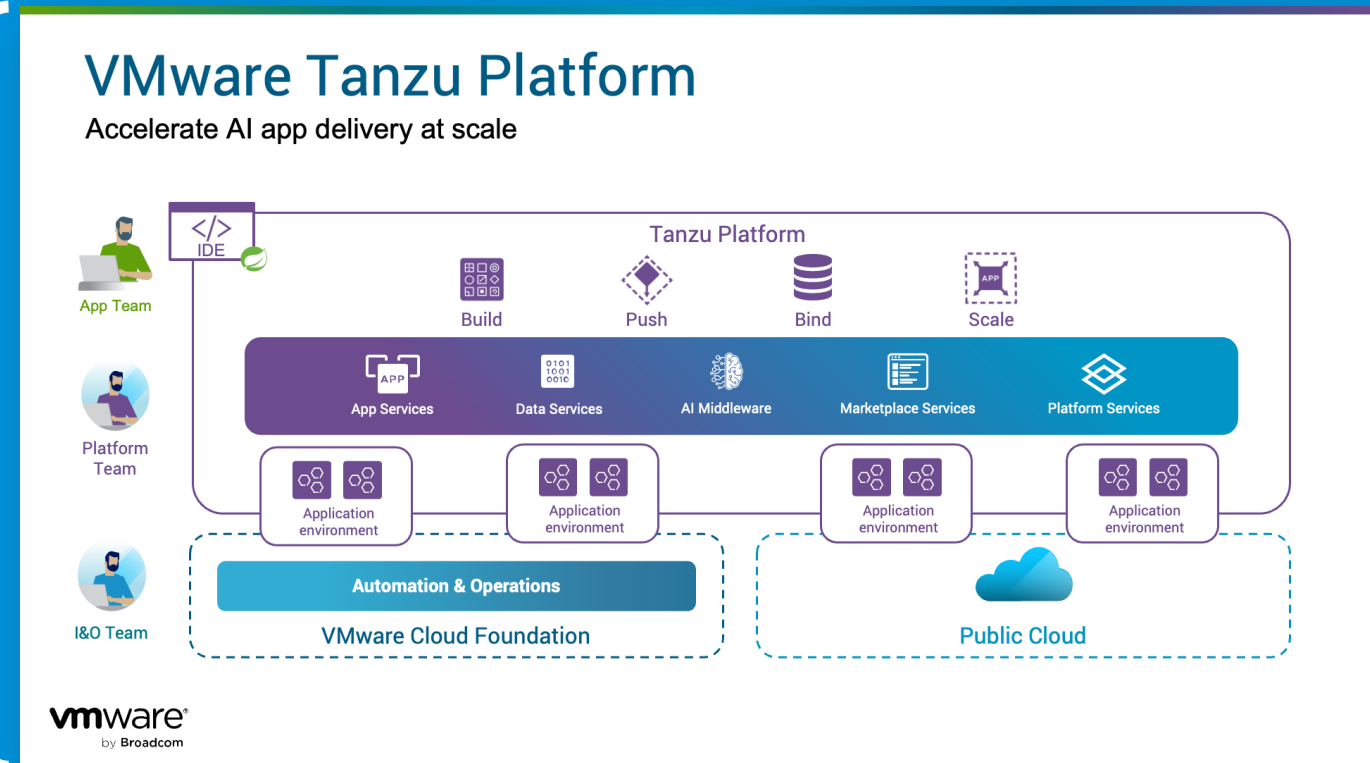
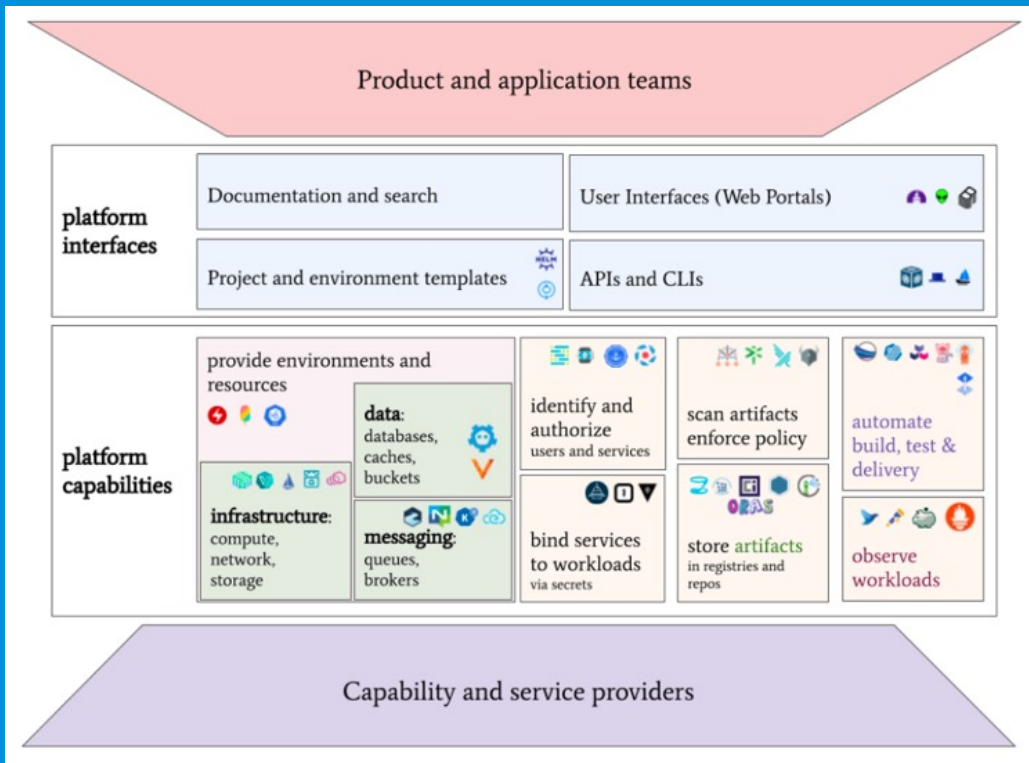
Evan Bottcher, March, 2018

Developer productivity & all the ops -illities



What is a *platform*?

Centralized, standardized stack for building, running, and managing in-house apps.



Is Kubernetes a
platform?

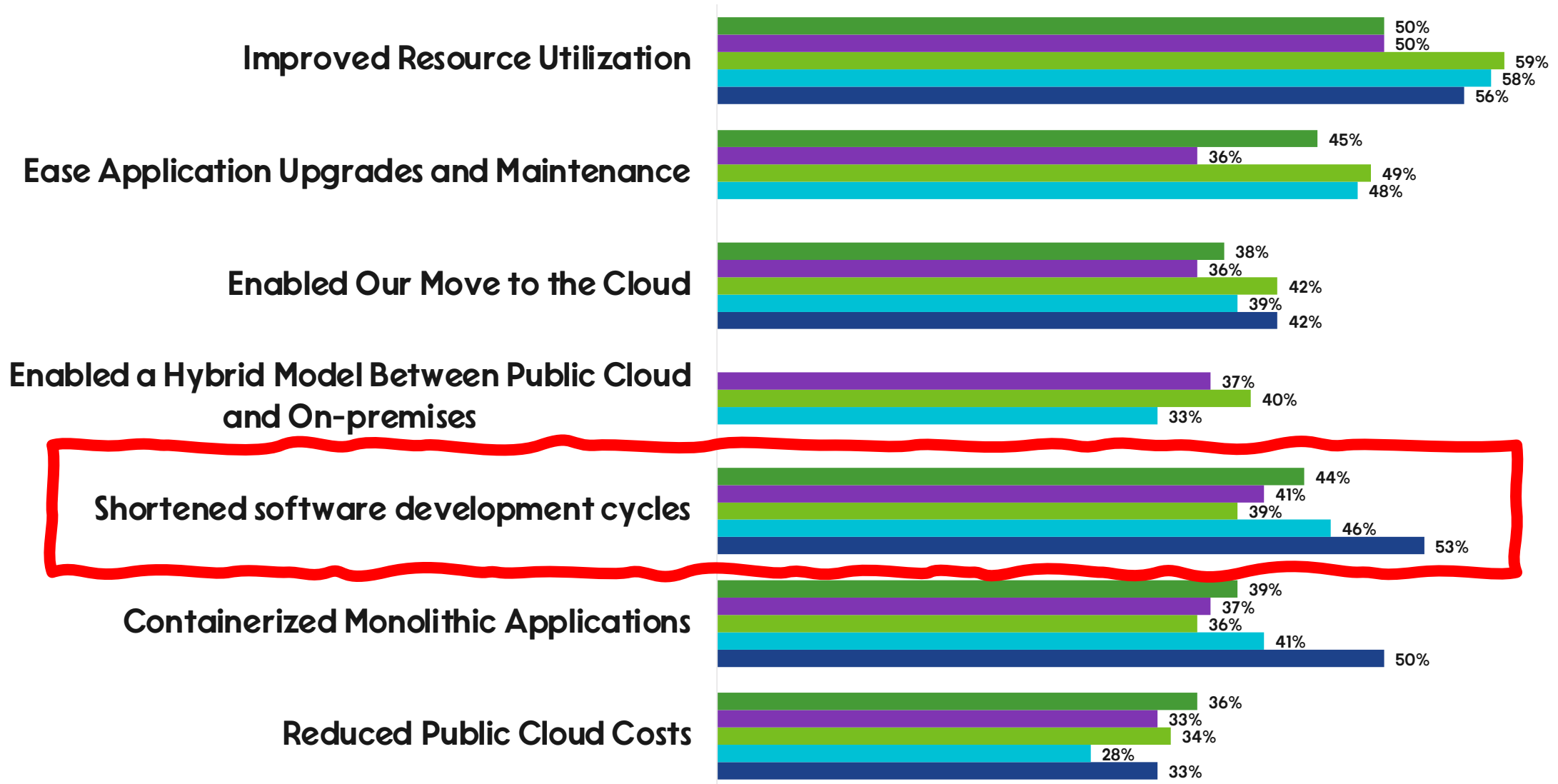


No.

(Not enough of it.)

What benefits has your organization realized from operating Kubernetes?

■ 2024 ■ 2023 ■ 2022 ■ 2021 ■ 2020



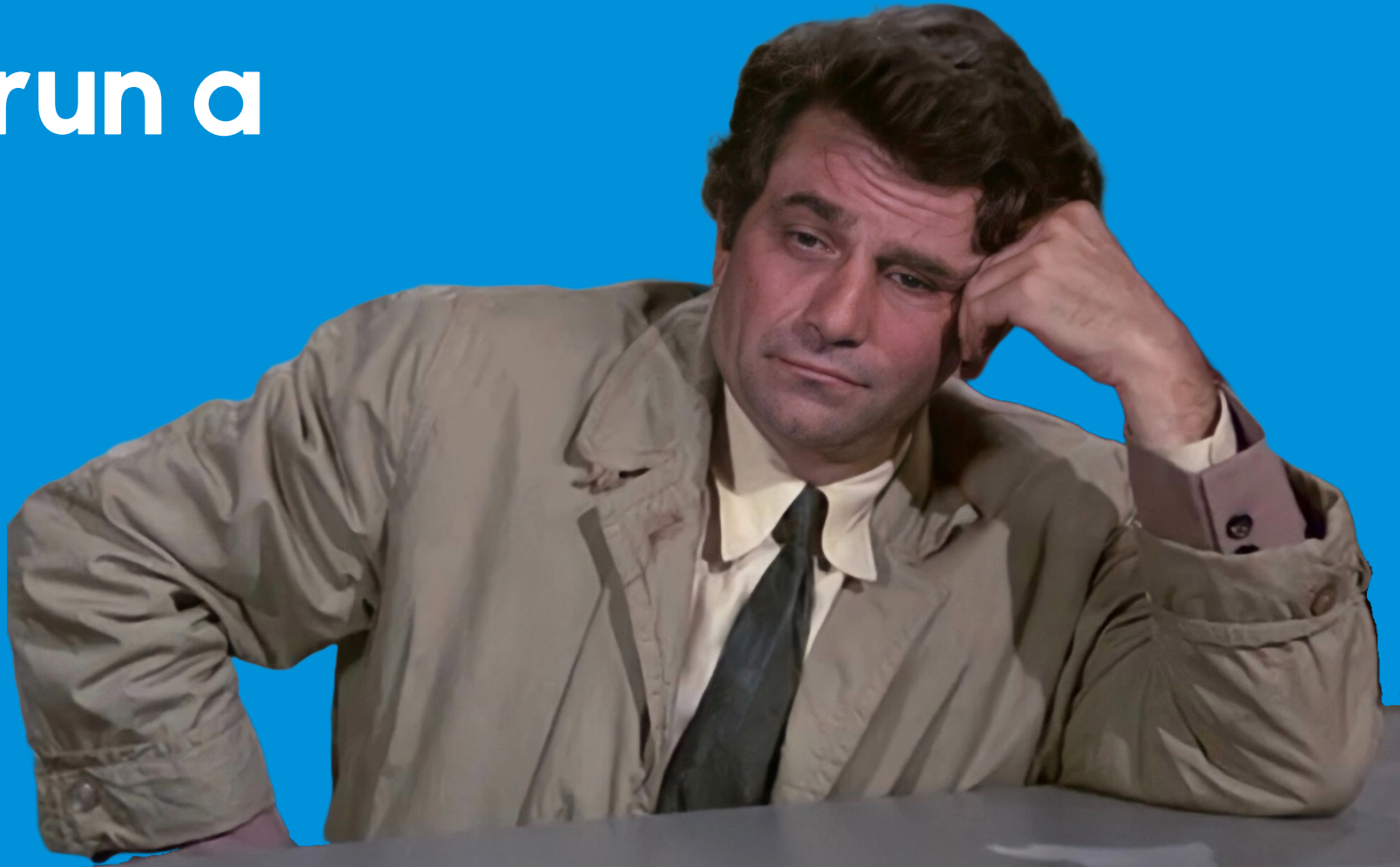
A man with a beard and glasses, wearing a blue button-down shirt, is speaking. He is gesturing with his right hand, which has a ring on the ring finger. The background is a bright, out-of-focus indoor space with large windows.

“The initial experience, that ‘wall of yaml,’ as we like to say, when you configure your first application can be a little bit daunting. And, I’m sorry about that. **We never really intended folks to interact directly with that subsystem.** It’s, more or less, developed a life of its own over time.”

Craig McLuckie, SpringOne 2021

How do you run a platform?

(in private cloud)



Developers



Operators

Stills from "The Mint Brothers," Bill Norton, sometime in the 2000s.

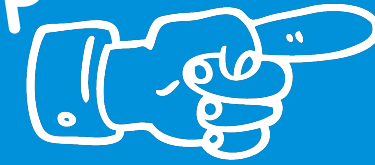
“We are building this platform not for us, we are building it for Mercedes-Benz developers.”

Thomas Müller, Mercedes-Benz



Platform marketing

Free Paper



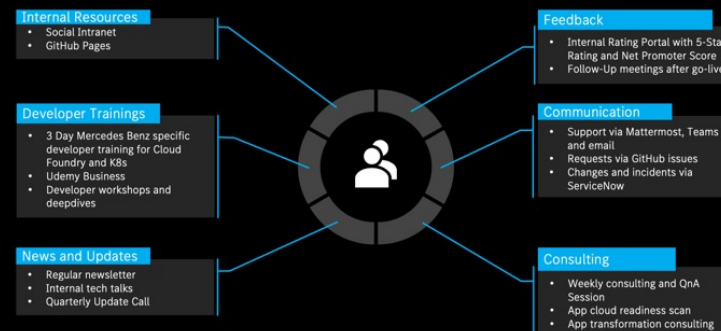
Organizational Learning



Focus on ways of working.....



Developer centric platform



Mercedes-Benz

Improving Developer Productivity with Platform as a Product | Mercedes-Benz

17



Sources: ING, 2023; BT Canvas team; MB.io; Duke Energy; Allstate; "Take DevOps to 11 and Sprinkle Cloud on it with Rainbows and Unicorns," Matt Curry, s1p 2017. "Improve Developer Productivity with Platform as a Product," VMware Explore, Nov. 2022; Kessel Run Wikipedia page (circa Feb 2025);

“What have you done for me lately?”

Tales of ROI, or,
Metrics == Money



Metrics for the LINE OF BUSINESS

Speed

Velocity is a vector comprised of speed and direction.

We bring a raw speed of advantage to the LOBs and also enable them to rapidly and reliably respond to changes in direction in the service of the business based on user feedback loops.

MEASUREMENTS

- ❑ Time to value (cycle time)
- ❑ Frequency of customer feedback
- ❑ Time between bug identification and fix
- ❑ Time from feedback to deployment of change
- ❑ Customer satisfaction (NPS)
- ❑ Business satisfaction

Stability

Reality is a complex landscape of changing priorities, emergent bugs, evolving architectures, and staffing changes.

We help the LOB achieve resiliency and low volatility as they deliver customer value in the face of this complex reality.

MEASUREMENTS

- ❑ Volatility (std dev in velocity / mean velocity)
- ❑ # of defects generated per developer - year
- ❑ % of software launches / upgrades delayed due to defects
- ❑ Employee satisfaction (ENPS)

Scalability

LOBs need to scale across two dimensions:

People - LOBs strive to attract developers and ramp productivity linearly with personnel.

Apps - LOBs need to rapidly scale their applications and their complexity to handle demand.

MEASUREMENTS

- ❑ # of products in development
- ❑ # of products measuring business success
- ❑ Investment ratios: spend developing software vs operating and systems
- ❑ Disruption caused by doubling workload
- ❑ Ability to attract and retain talent (# of internal referrals)

Security

To move rapidly the team needs to feel secure in making code changes aggressively. Automated test coverage provides this safety net.

To rapidly search for customer value LOBs must adopt a learning culture that fosters psychological safety necessary to fail and learn from failure.

MEASUREMENTS

- ❑ % teams using CI
- ❑ % teams doing TDD
- ❑ Time from commit to deployment

Savings

Teams must reduce risk and waste through small batch delivery and fast consumer feedback.

This drives significant savings as use of the product grows and is key to maintaining their trust and enabling them to go fast, forever.

MEASUREMENTS

- ❑ Fraction of developer time spend writing code and delivering value
- ❑ Product:dev ratio
- ❑ Business satisfaction
- ❑ # of go/no-go decisions based on business success

Metrics for the IT

Speed

IT can efficiently upgrade, patch, and manage the platform.

They rapidly onboard new application teams and provide the necessary services to quickly unblock teams and enable them to deliver consumer value.

MEASUREMENTS

- ❑ # prod/dev deploys per month
- ❑ # platform upgrades per month
- ❑ Platform upgrade speed
- ❑ # of new apps onboarded/month
- ❑ Team distribution of skills

Stability

Our customers entrust us with their production workloads and their developer productivity.

We must provide adequate SLOs to meet their needs and earn their trust by ensuring compatibility and uptime across platform upgrades.

MEASUREMENTS

- ❑ Minutes of prod outage per year
- ❑ Minutes of dev outage per year
- ❑ Mean time to recovery
- ❑ Mean time between failures
- ❑ # of upgrade-related failures

Scalability

IT needs to provide an “at-scale” service on-demand at the whim of the business.

They need to explore all options with minimal friction as they grapple with the mix of workloads on-premise and in the cloud.

MEASUREMENTS

- ❑ Queries per second
- ❑ # of AIs per foundation
- ❑ # of SIs per foundation
- ❑ # of foundations
- ❑ # of teams using the platform
- ❑ Does increasing workload on existing

Security

Security is a paramount concern for our customers. We earn their trust by providing a platform that is secure by default.

We solve for security and reduce security-related friction and toil in order to enable our customers to go fast, forever.

MEASUREMENTS

- ❑ Time between identifying and patching a CVE
- ❑ Cost in person-hours or dollars of leaked credential
- ❑ Fraction of operator time spent on security configuration
- ❑ # of disruptions/suspensions due to security concerns

Savings

IT must meet the needs of thousands of developers within tight budgetary constraints.

We provide a platform that simultaneously reduces complexity and sprawl and improves the ops:dev ratio.

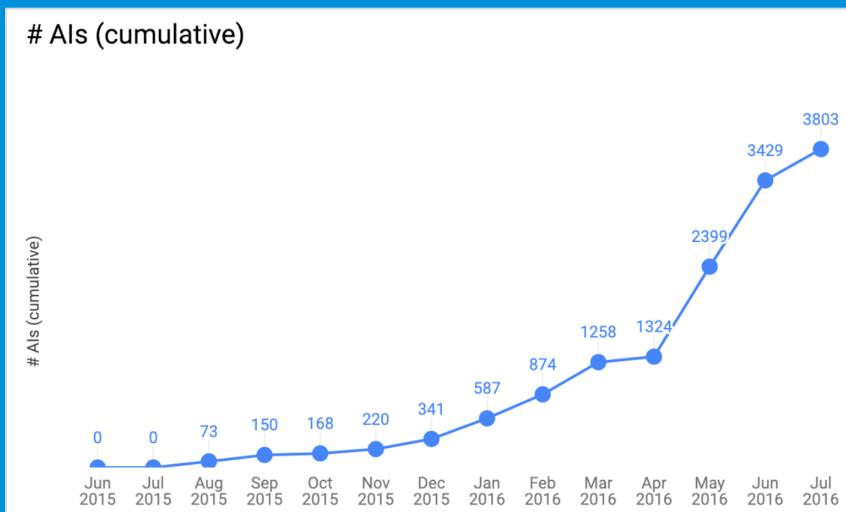
MEASUREMENTS

- ❑ Operator:developer ratio
- ❑ # of apps per operator
- ❑ # of foundations per operator
- ❑ Degree of automation for provisioning, build, test, change approval governance, deployment, perf

Scaling Phase – Pairing & Seeding to build trust & training



1. Create platform marketing program.
2. Find two to five more apps.
3. Pair & seed from first dev & platform team to new teams.
4. "Shift Left" – build golden paths for governance, security, etc.
5. Add more infrastructure staff with pairing & seeding.
6. Do this for three months.
7. Repeat, growing number of apps as pairing & seeding allows.



What about AI?

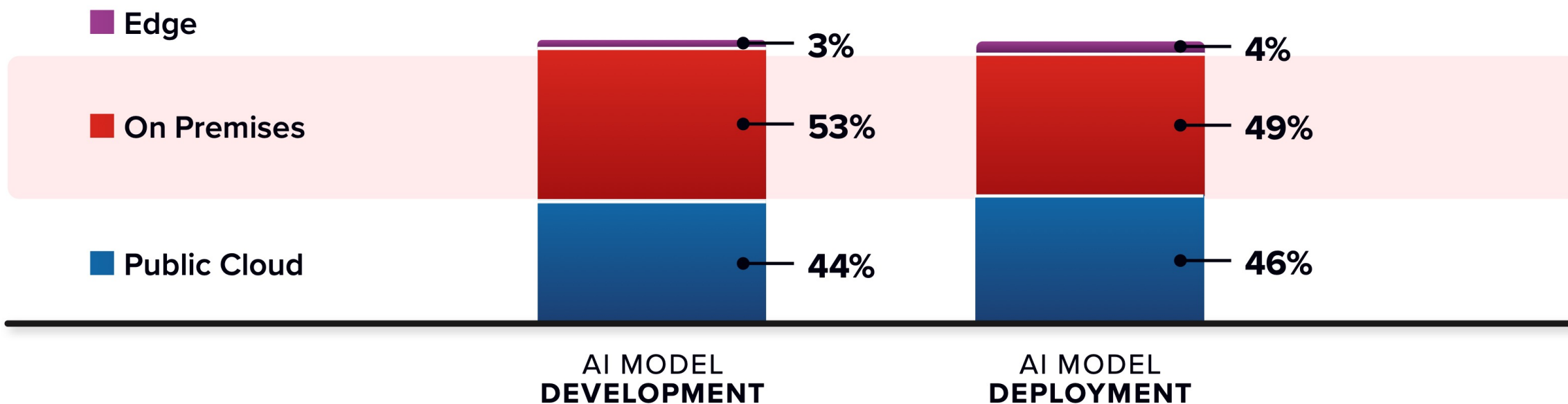


FIGURE 5

Deployment Location for the Development and Deployment of AI Models

Where does your organization primarily develop and deploy AI models?

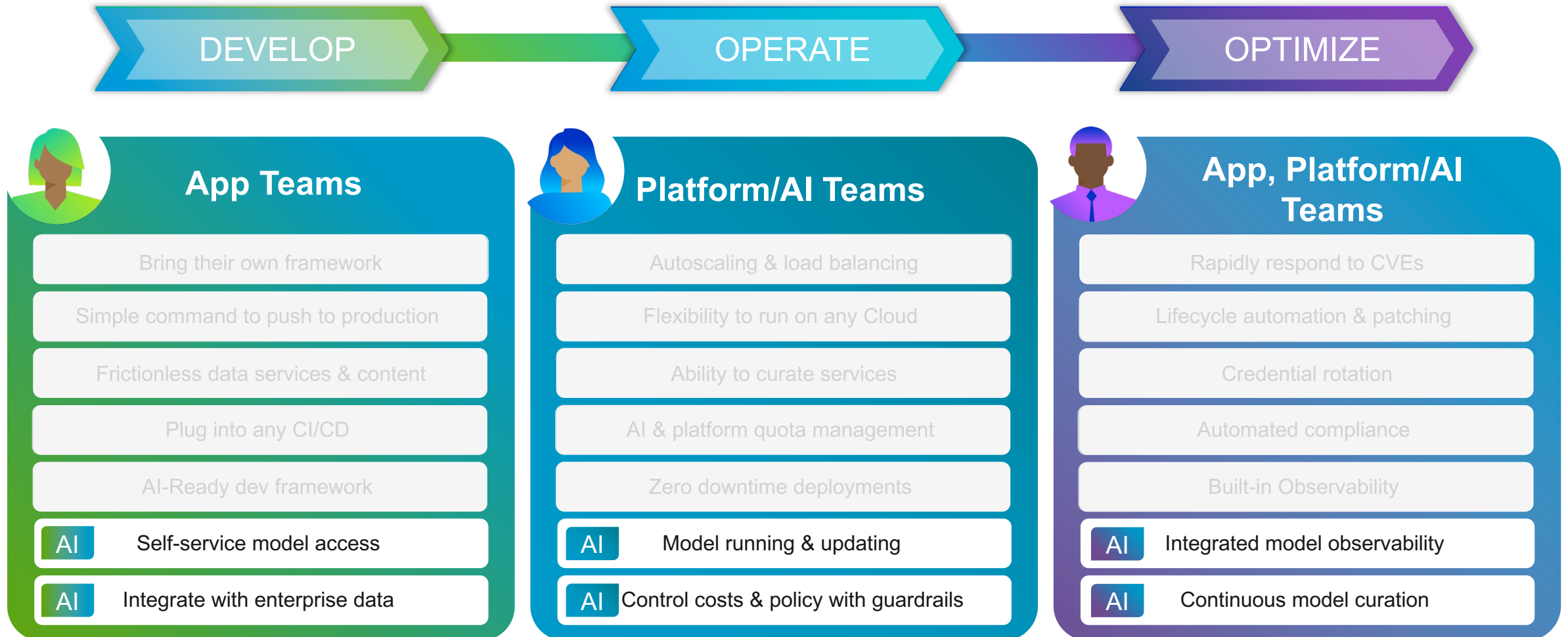
(Percentage of respondents)



Note: Public cloud includes public cloud provider, managed services provider, and special-purpose cloud. On premises includes private cloud, traditional infrastructure, and colocation provider. n = 411; Source: IDC's *AI Infrastructure Survey*, July 2024.

For an accessible version of the data in this figure, see [Figure 5 Supplemental Data](#) in Appendix 1.

A platform treats AI like any other service, adding AI middleware & focusing on new models & frameworks



Thanks!



<https://newsletter.cote.io/>



cote@broadcom.com



Slides

