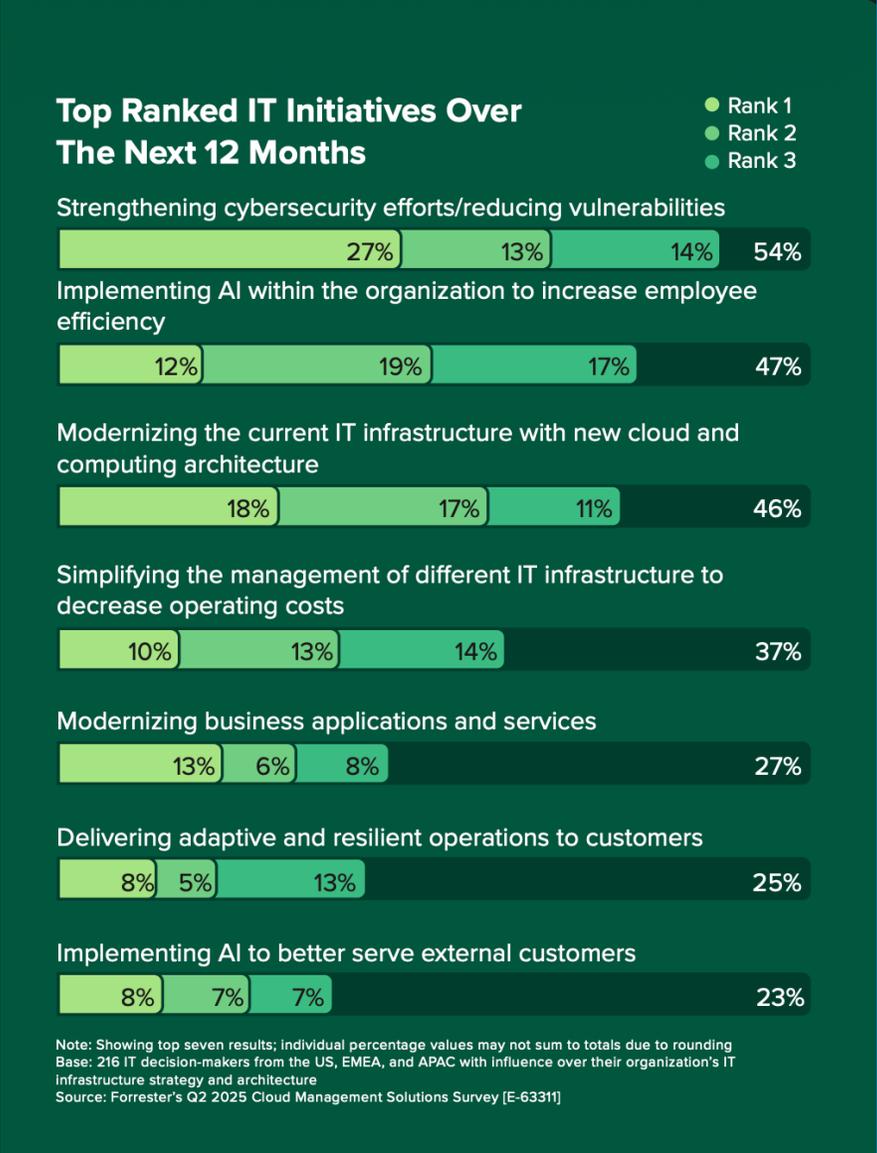


Private Cloud AI Platform Engineering

What little we right know now, maybe

Coté – SREDay Amsterdam, November 7th, 2025

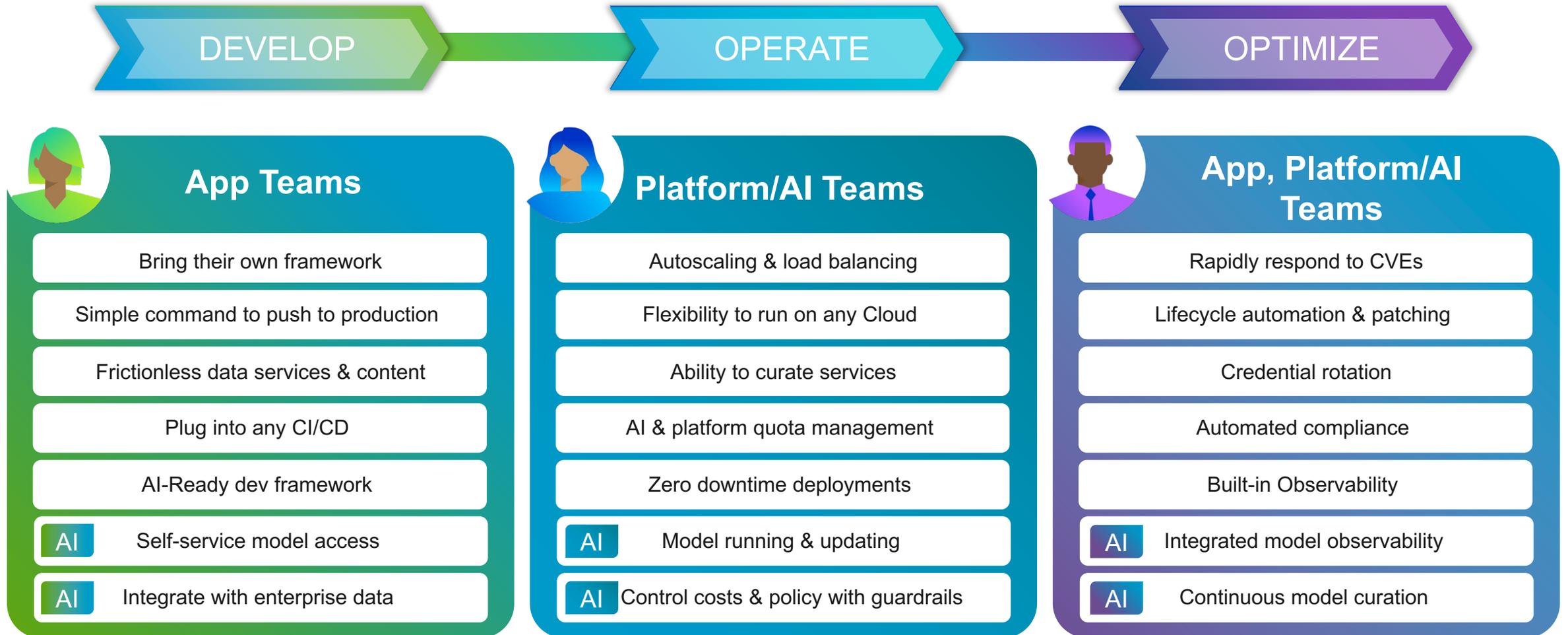
What's cool is not always what's the highest priority



New things platform engineers will likely do with and for AI

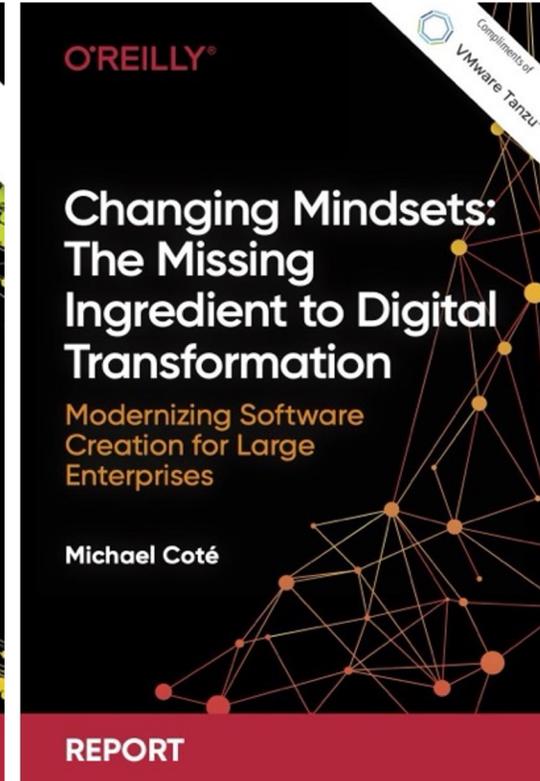
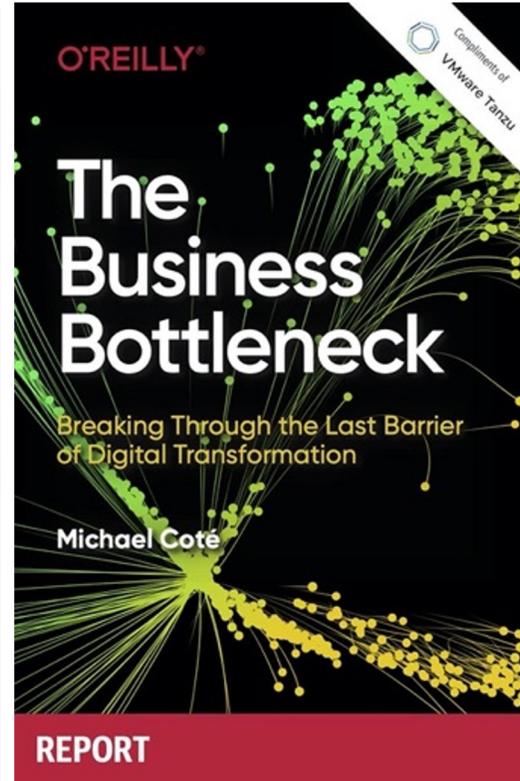
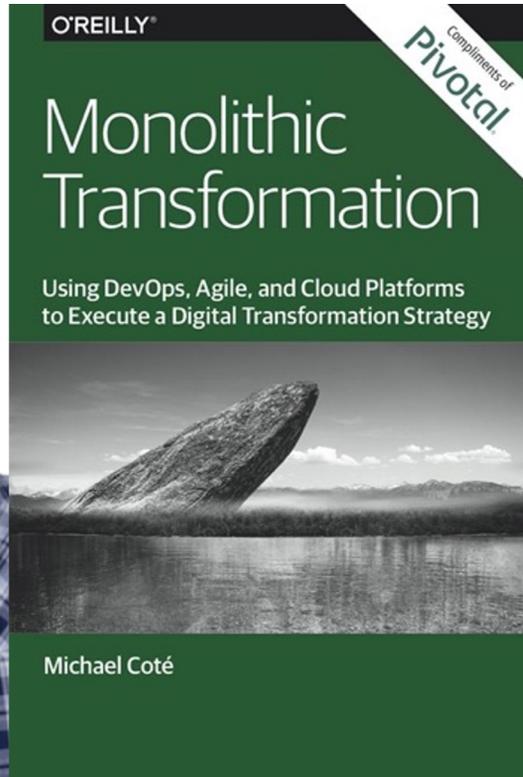
- Hosting models.
- Gateways, brokers.
- Application Frameworks
- ¿Curating models?
- Hosting models with self-service access.
- AI FinOps BizSRE DevOps Eng.
- ¿Eval, testing, safety?
- Audit and compliance.
- Data access.
- Registries – MCP, prompts, integrations.

A platform treats AI like any other service: security, reliability, middleware, new models, & dev frameworks



Coté

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



Home work

Or, acknowledgements and references

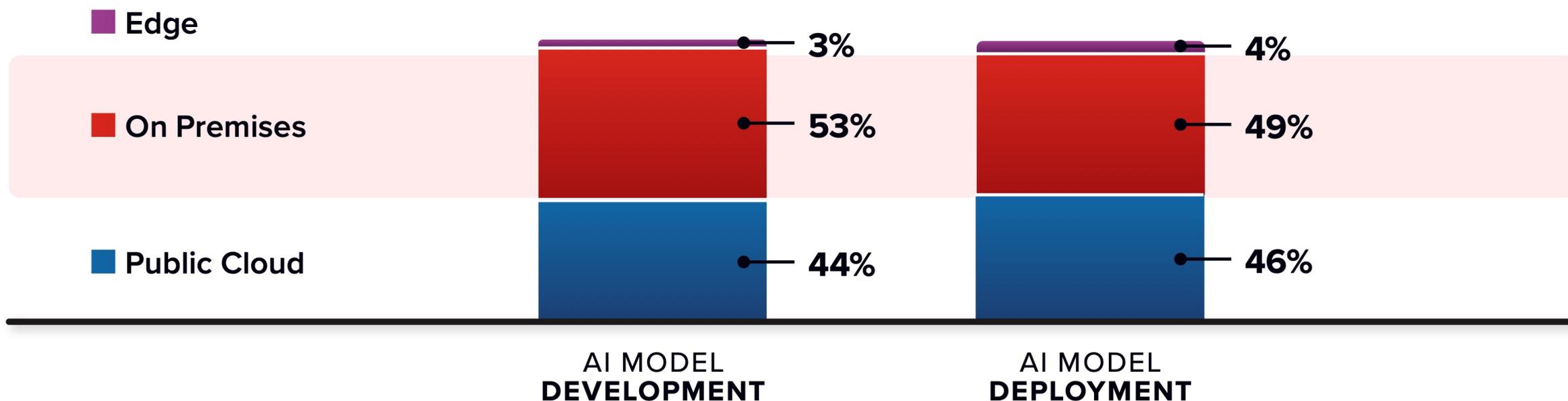
- Tanzu customers and internal use.
- ["Tales from Production - Debugging LLMs and GenAI Apps on VMware Tanzu Platform,"](#) Nick Kuhn VMware Explore 2025, No. CODEQT1641LV, August, 2025.
- ["How platform teams can help scale generative AI application delivery,"](#) Manjunath Bhat, Gartner, June 23rd, 2025.
- ["Why AI needs a platform team,"](#) Patrick Debois, PlatformCon 2025, June 23rd, 2025.
- ["AI Platform Engineering,"](#) Patrick Debois, Dec 31, 2024.
- ["From Experiment to Enterprise: How Block Operationalized MCP at Scale,"](#) Angie Jones, Block, MCP Developers Summit, May, 2025.
- ["The evolving role of Platform teams in the AI era,"](#) Abi Noda and Laura Tacho, DX, August 28th, 2025.
- ["The Emperor's New GPT Why Your 'Custom AI' Is a Demo, Not a Product,"](#) Keith Townsend, August 18th, 2025.

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.

The platform



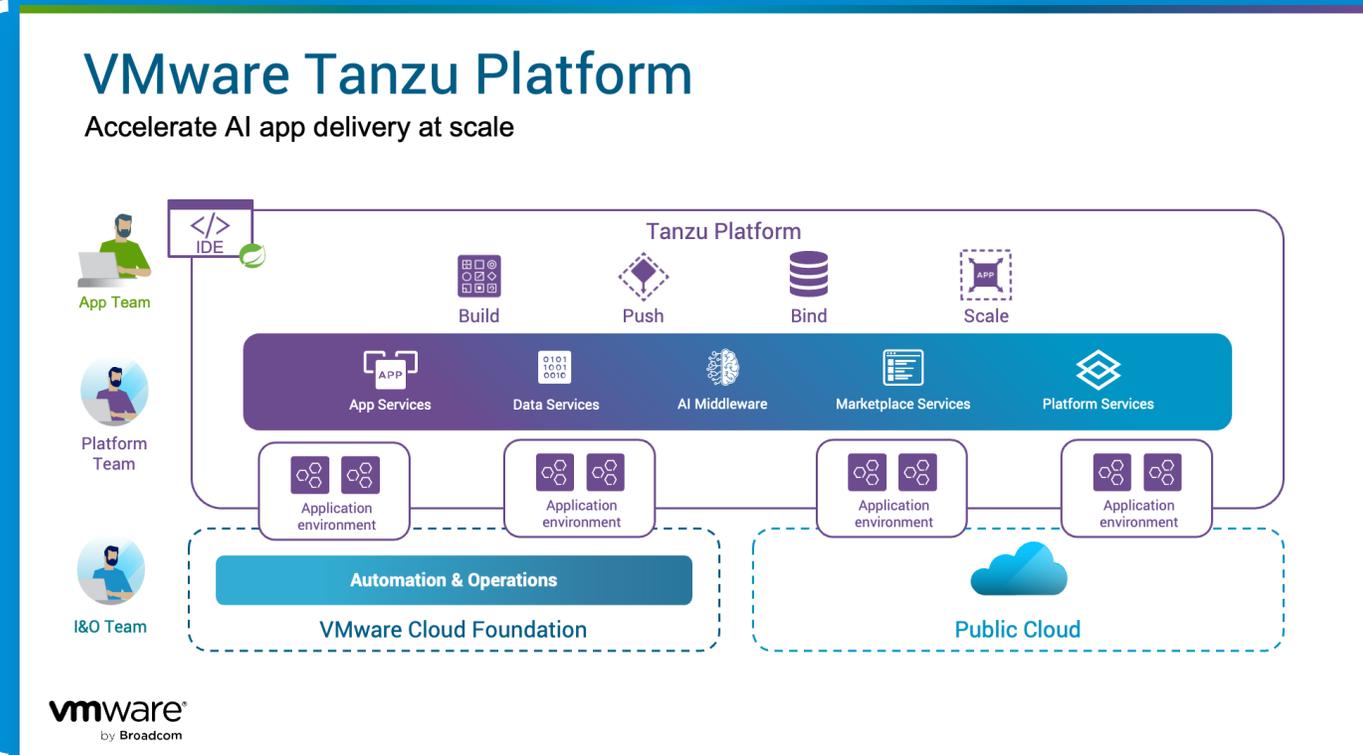
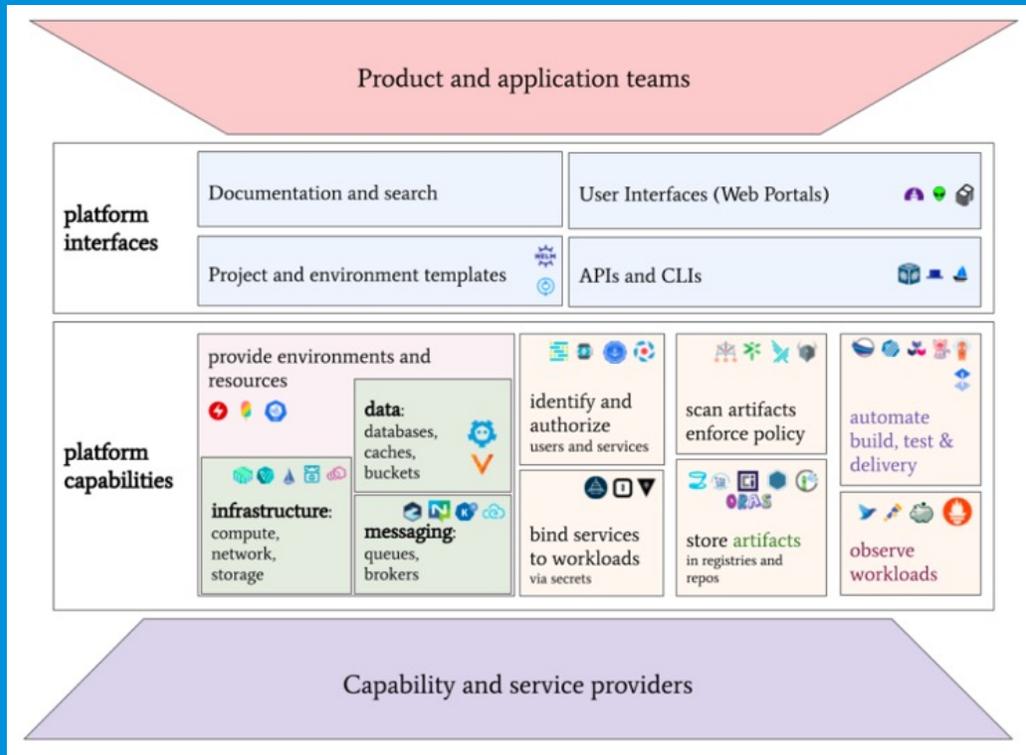
A digital platform is a foundation of self-service APIs, tools, services, knowledge and support which are arranged as a compelling internal product.

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

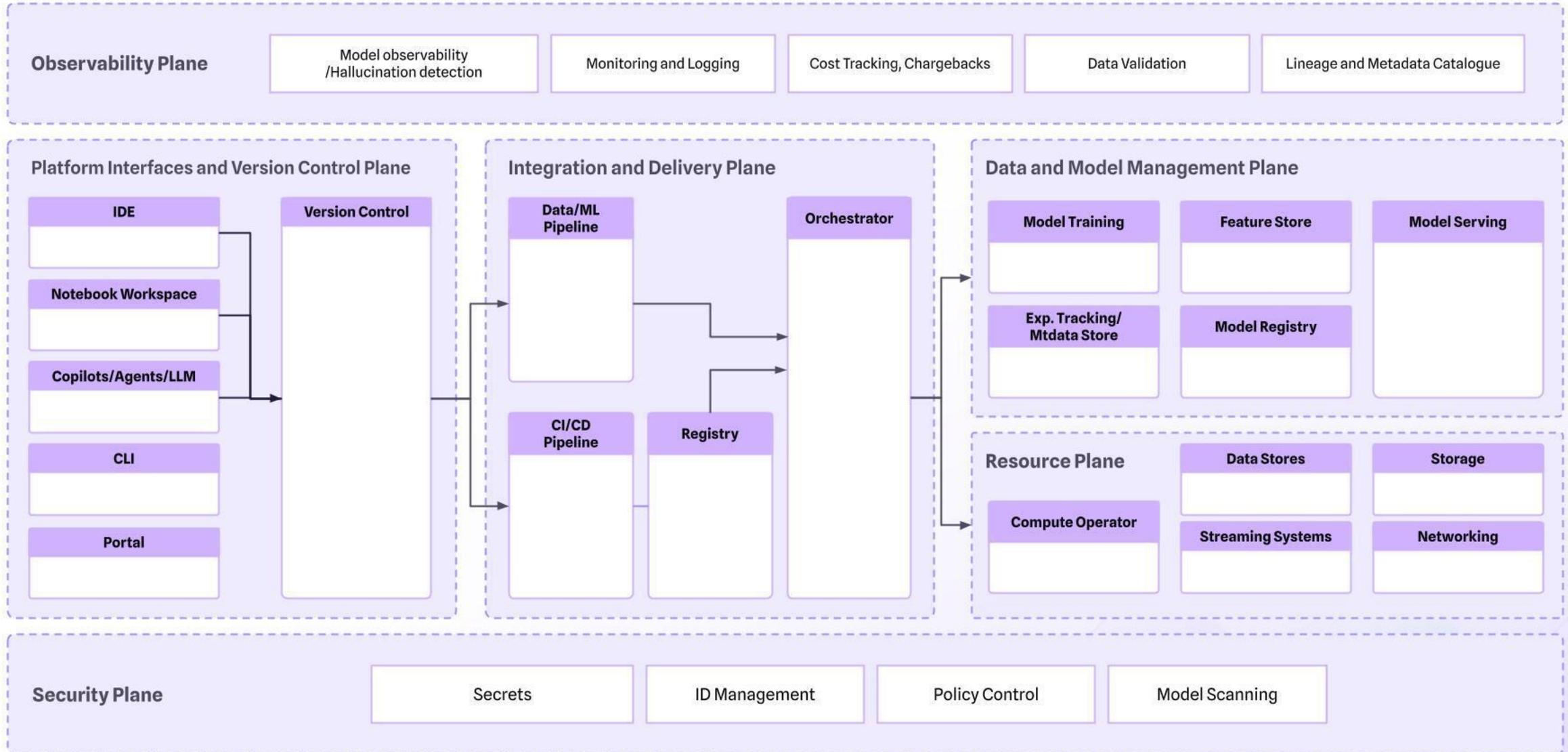
Evan Bottcher, March, 2018

What is a platform?

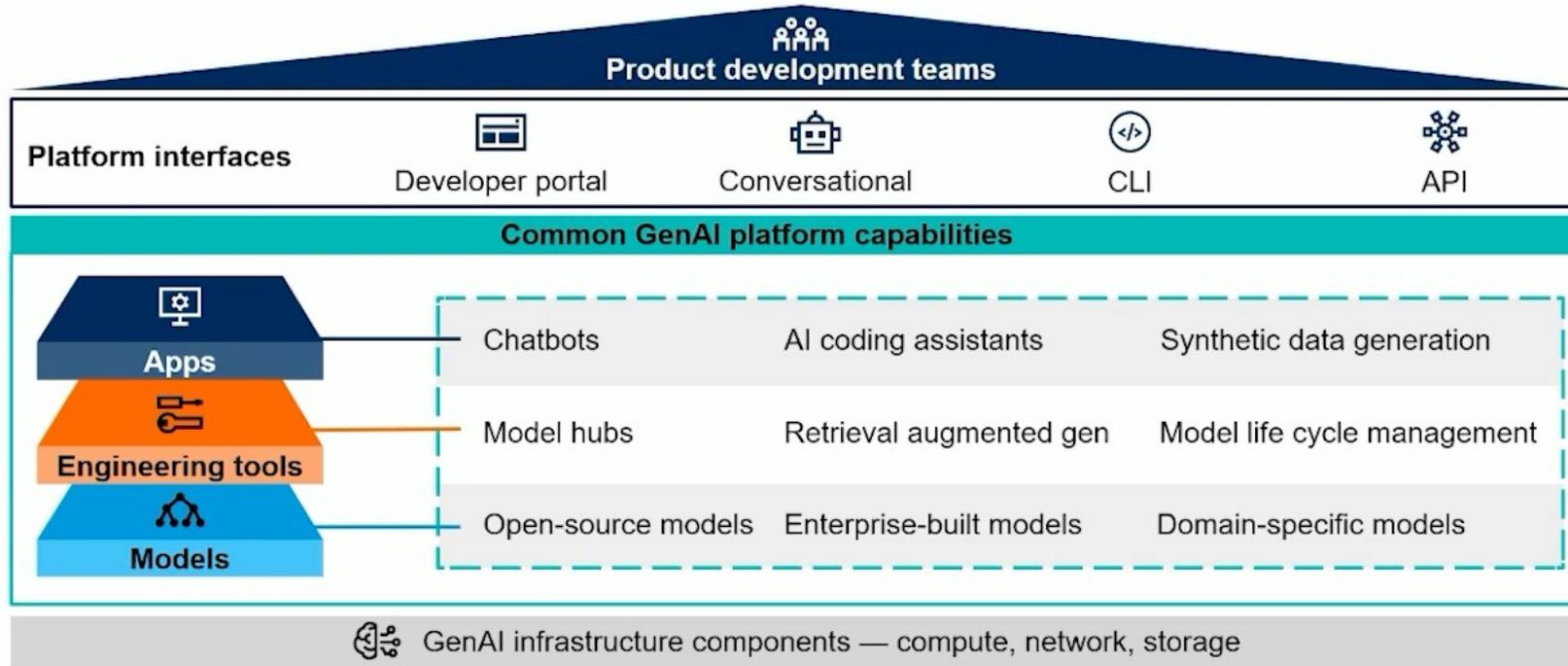
Centralized, standardized stack for building, running, and managing in-house apps.



New reference architecture for AI



Platform Teams Provide Internal Platform Services to Support Common GenAI Needs



*TRISM — AI trust, risk and security management
Source: Gartner

20 © 2025 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner is a registered trademark of Gartner, Inc. and its affiliates.

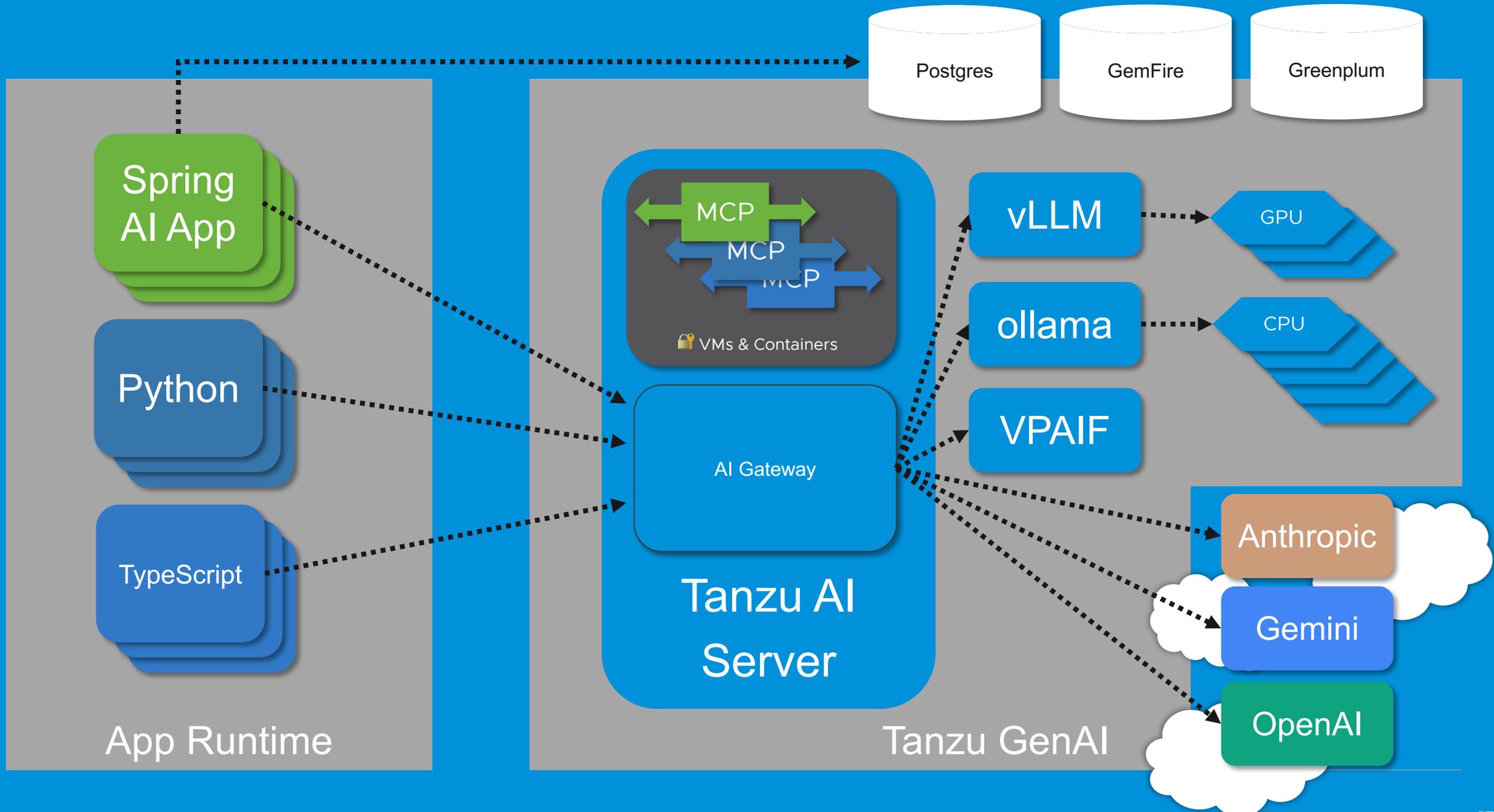


Manjunath Bhat

Research VP, Gartner

The two thing you must do right now





Last login: Thu May 24 16:34:03 on ttys001

PicklesFromTheJar:~ cote\$ █

Meet the developers where they are

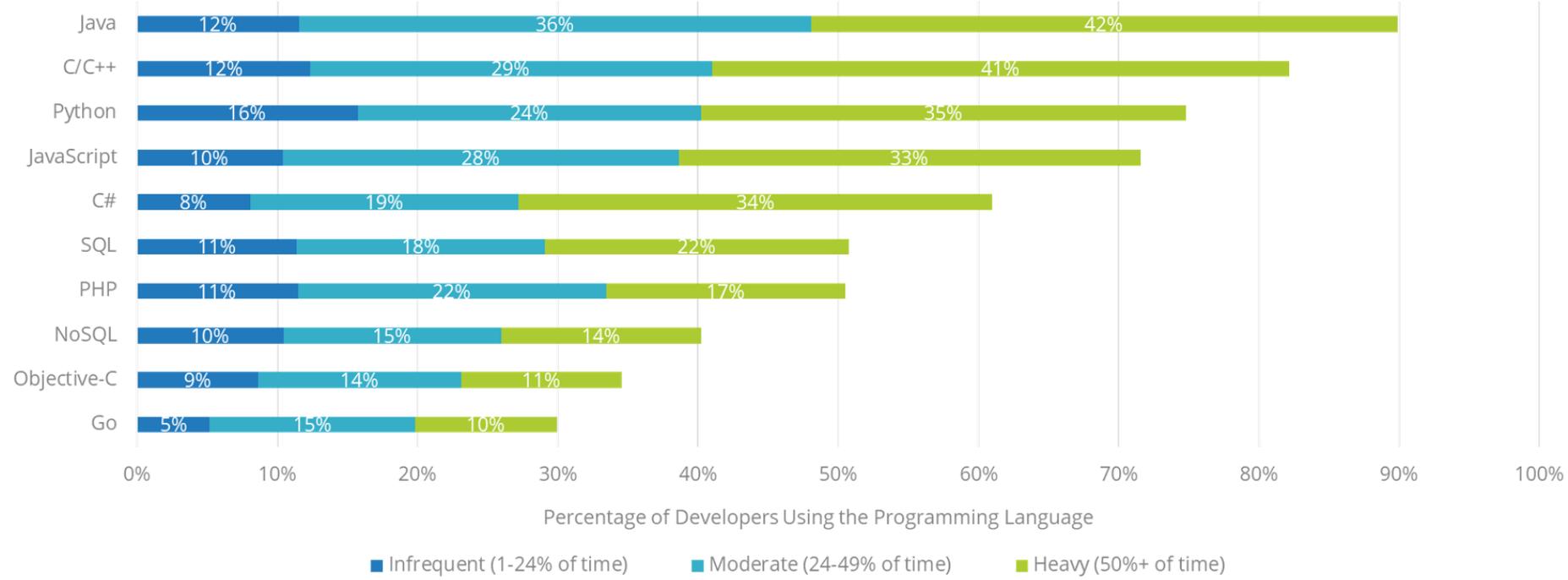
IDC Survey Spotlight

What are the top ten languages used by software developers?



Adam Resnick

QB1. Please indicate how much you (or the team you are responsible for) used the following languages as a software developer in the last 12 months.



IDC # US52257024 (May 2024)
Source: IDC, Developer View 2023, N = 2257, Base=All respondents that answered this section

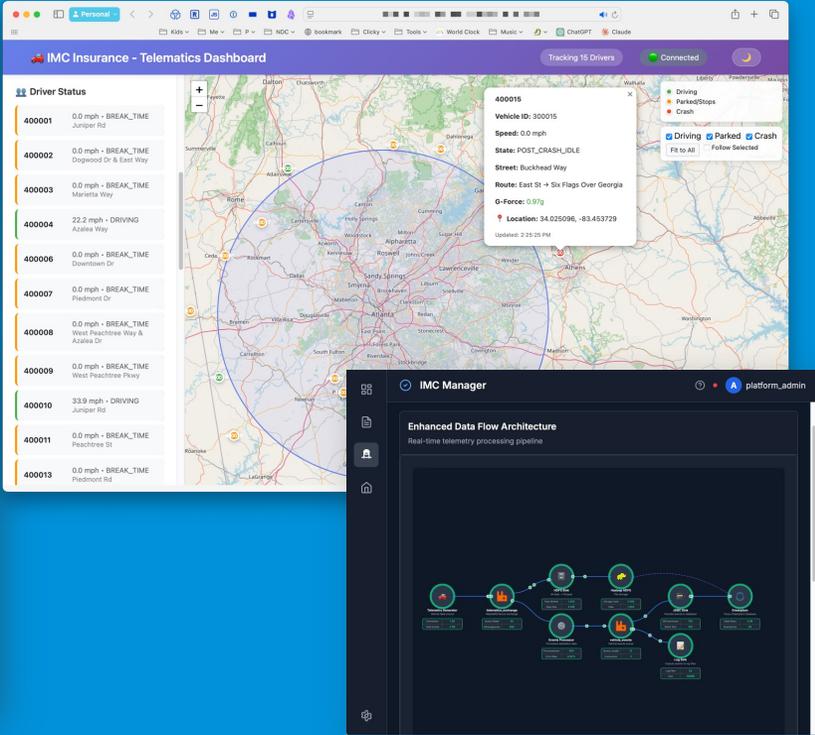
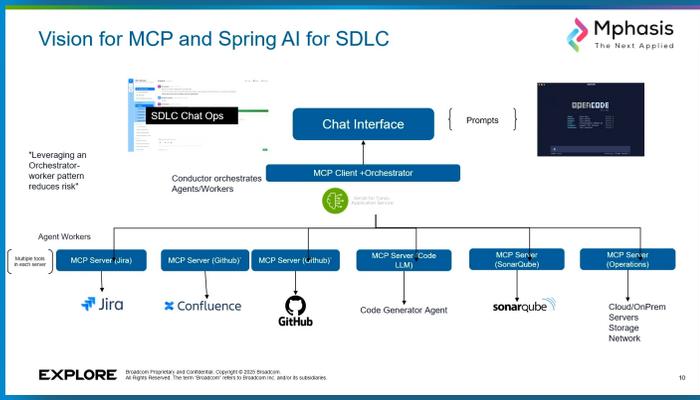
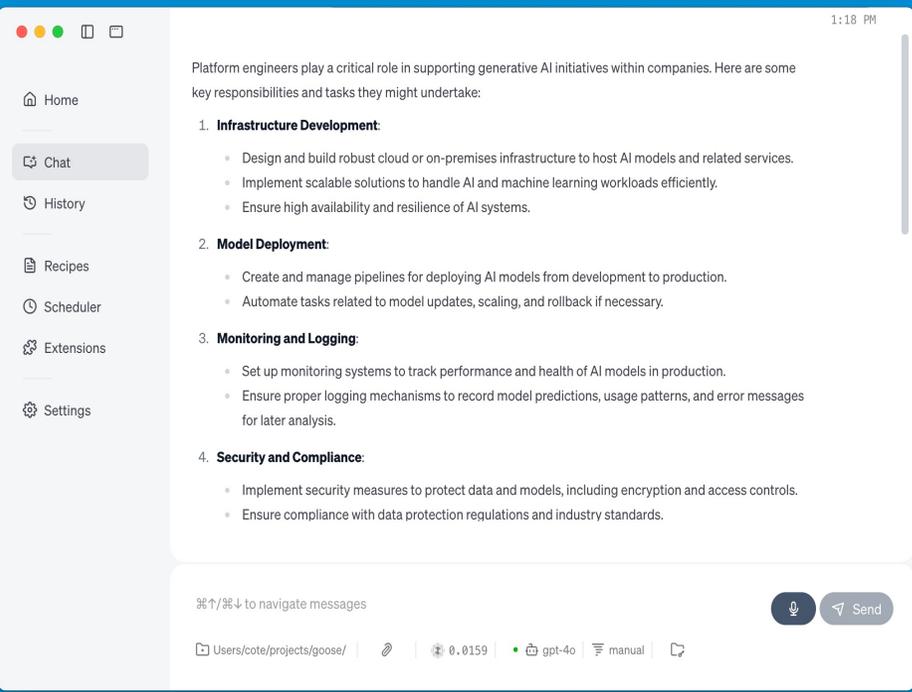


Practices so far

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

Thomas Müller, Mercedes-Benz





👤 Chat (for normals)...

- Your own ChatGPT.
- Customer service.
- Better search.
- Chat-as-UI

👤 Programming...

- New & old code.
- SDLC juicing.
- Trad'l "data science."
- Making pptx?

📁 AI in apps...

- Sales assistants.
- Sloppy integration.
- Science-ing.
- ???

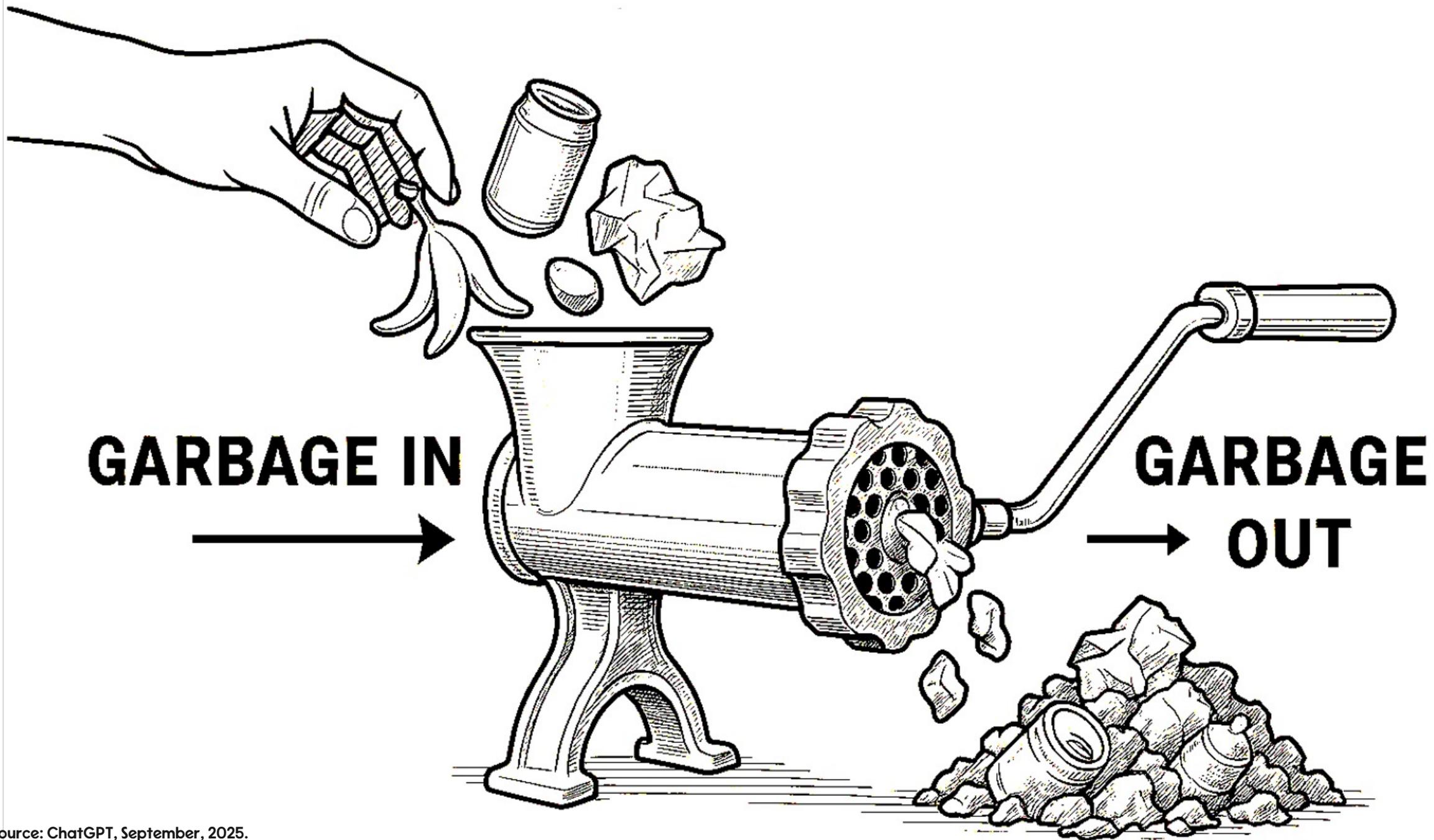
Note: audio, pictures, video are omitted.

Sources: Taznu customers; [AI at Goldman](#), FT, September 14th, 2025; ["Leverage Generative AI to Streamline the Software Development Lifecycle,"](#) Banu Parasuraman, Andrew Berenato, Explore 2025, August, 2025.

Avoiding the Center of Excellent Bottlenecks

Tales from the Customer...

| | |
|-------------------------|---|
| Meta (llama family) | Denied |
| Mistral Family | Approved (<i>only apache2 licensed models</i>) |
| Gemma | Denied |
| Qwen3 + DeepSeek | Denied |
| SaaS (OpenAI, Azure AI) | Partial (<i>only public data</i>) |
| OpenAI (gpt-oss) | <i>Under Review</i> |

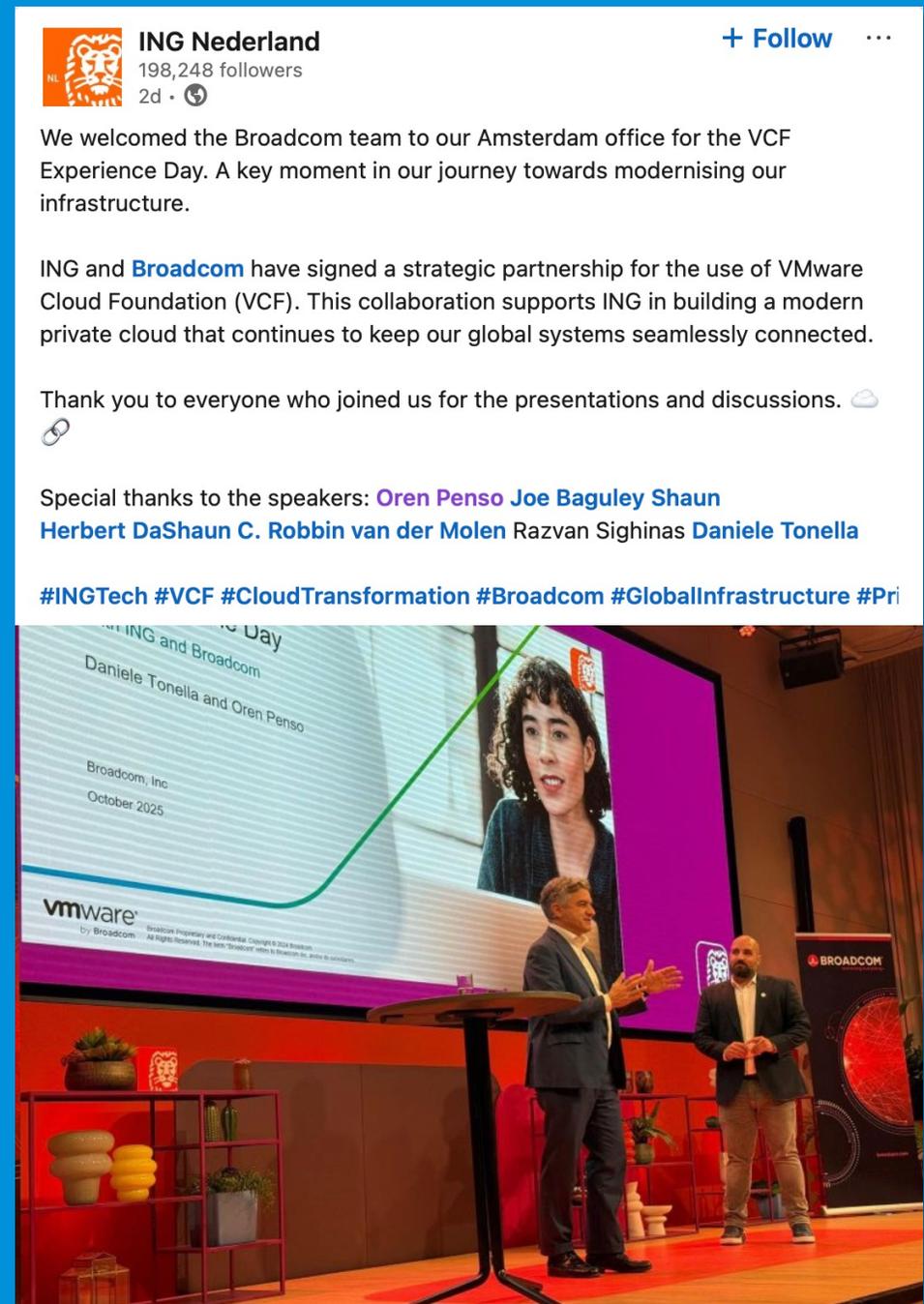


**When you don't know what you're doing,
do a lot of it quickly.**

Lucky enough to have it already?

...if not, 90 day trial:

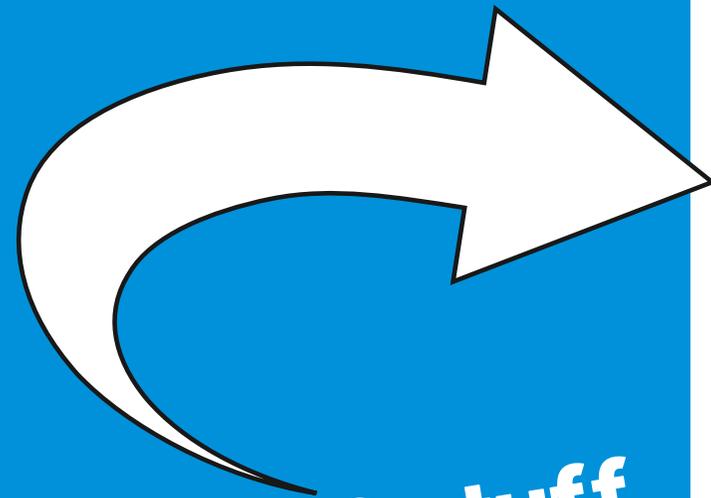
TryTanzu.ai



Thanks!

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

 cote@broadcom.com



Slides & stuff

