



# VMware Private AI Foundation with NVIDIA

VMware Private AI Foundation with NVIDIA  
*Democratize generative AI and ignite service innovation for government*

April 4, 2024

Broadcom Proprietary and Confidential. Copyright © 2024 Broadcom. All Rights Reserved. The term "Broadcom" refers to Broadcom Inc. and/or its subsidiaries.

# VMware Private AI Foundation with NVIDIA

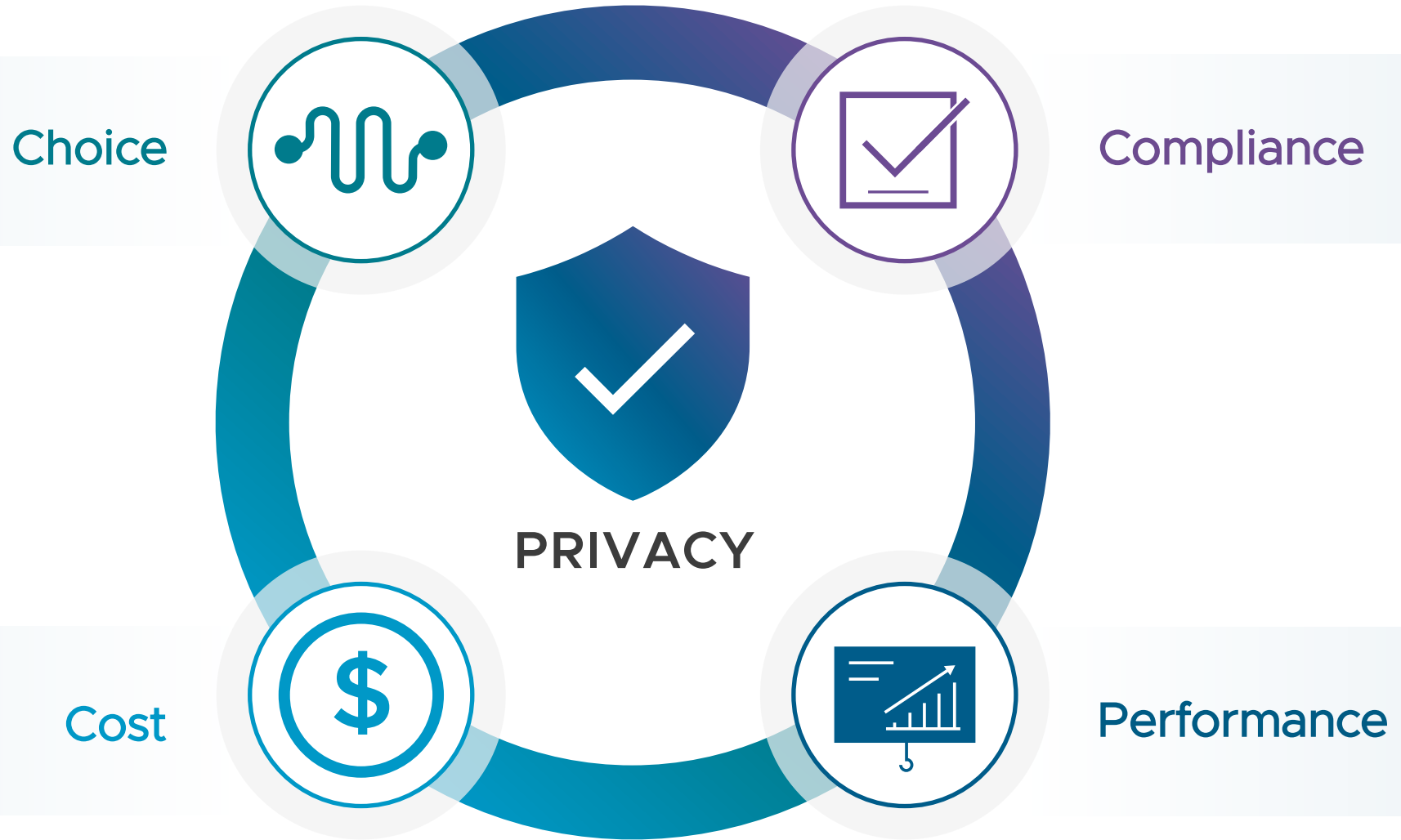
*Democratize generative AI and ignite service innovation for government*

April 4, 2024

# Agenda

- Introduction NVIDIA and VMware by Broadcom
- Demystifying Generative AI
- Use Cases and Perspective
- Solution Overview – VMware Private AI Foundation with NVIDIA
- Q&A

# Key Enterprise Challenges



# Core Challenge

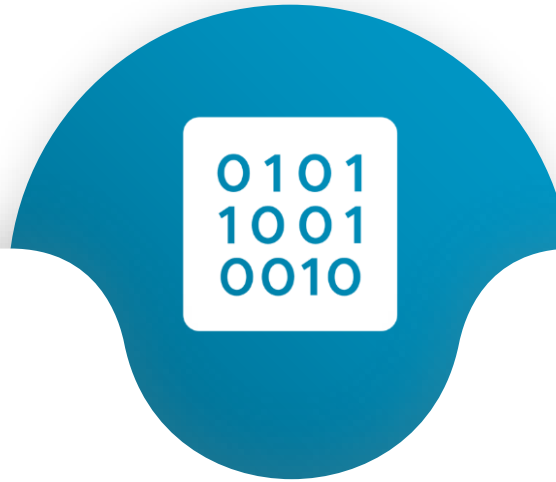
## Privacy

### Private Intellectual Property



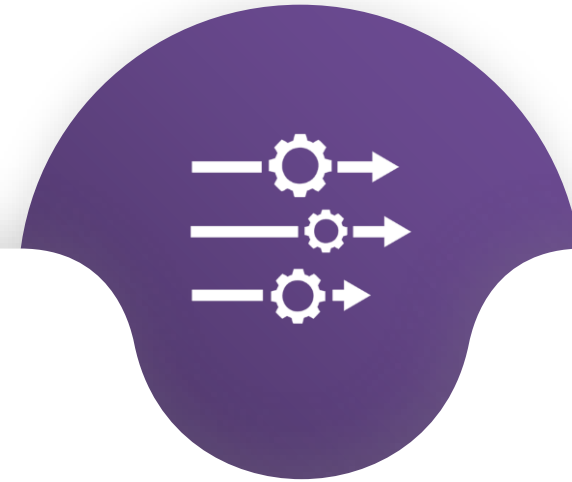
“I need to minimize my intellectual-property risk.”

### Private Data



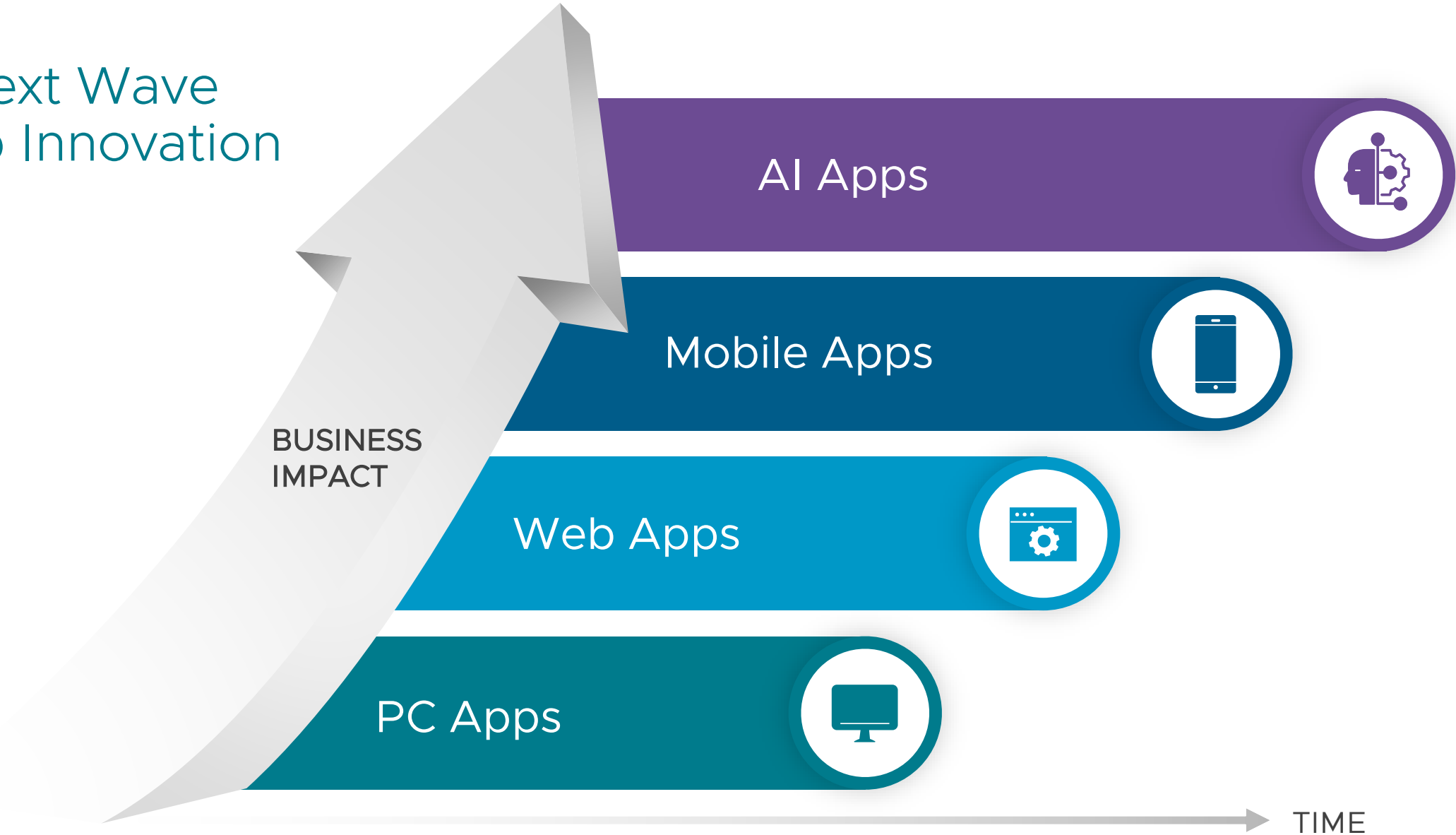
“I need to ensure my private data will never be shared externally.”

### Private Access



“I need complete control over access to my AI models.”

# The Next Wave of App Innovation



# Impact of Generative AI in the Enterprise

## PREDICTIVE AI



Data Scientists



Predictive Analytics



SPECIALIZED AI MODEL

## GENERATIVE AI



Marketing



Supply Chain



Sales



Human Resources



Customer Operations



Research and Development



Legal



Software Development



Manufacturing



Procurement



IT



Finance

NATURAL LANGUAGE INTERACTION

APIs

LARGE LANGUAGE MODELS





ANNOUNCED 2023

# VMware Private AI

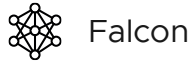
---

Democratize generative AI and ignite service innovation for Government

*An architectural approach that balances the innovation gains from AI with the privacy and compliance needs of government.*



# VMware Private AI Open Ecosystem



Falcon



Llama 2



Mistral



MPT



StarCoder



WizardML



cnvrg.io



VMware Cloud Foundation™



INTRODUCING

# VMware Private AI Foundation with NVIDIA





NVIDIA Foundation Models



NVIDIA Fine-Tuned Models

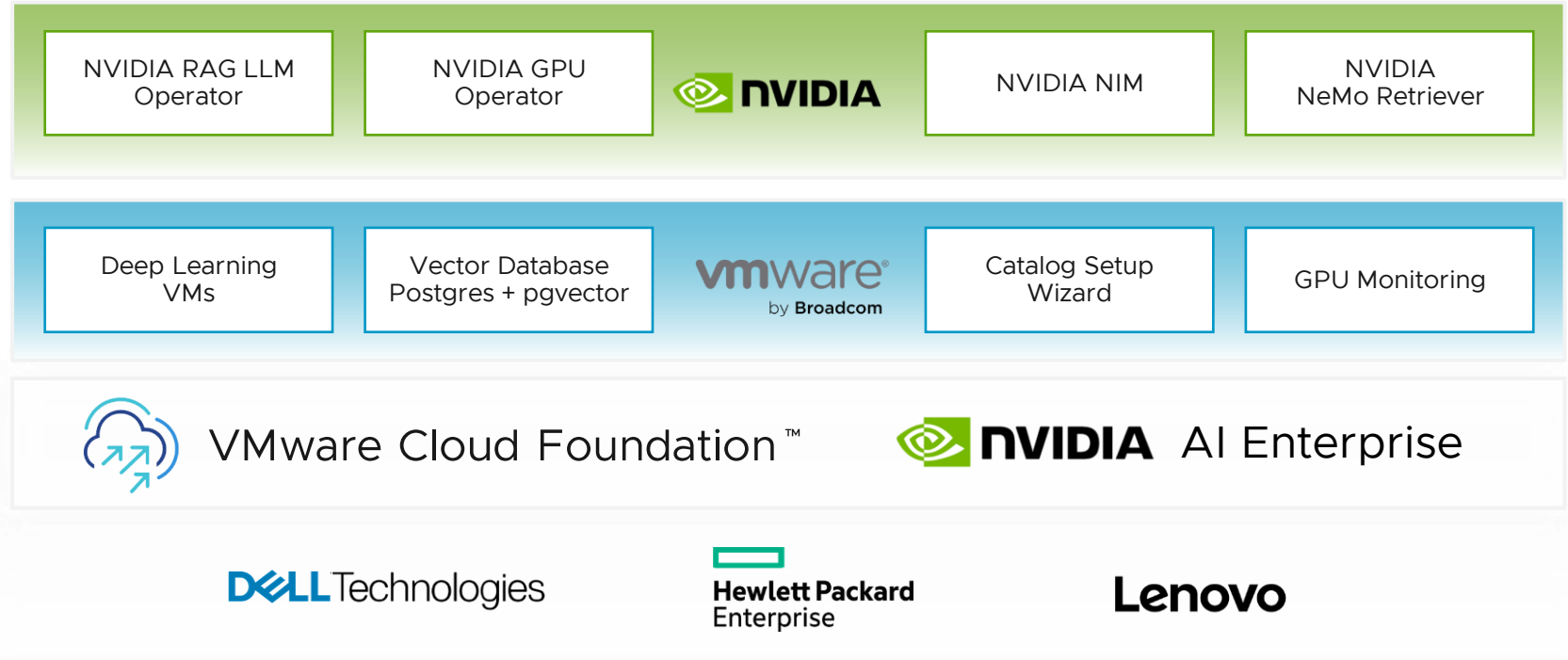


Third party & Community Models

## INTRODUCING

# VMware Private AI Foundation with NVIDIA

Joint Gen AI Platform:  
Initially Available



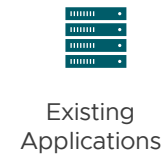
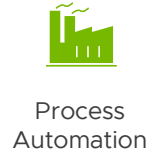
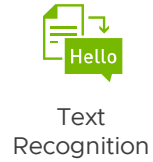
Choice of LLMs

Bare-Metal performance

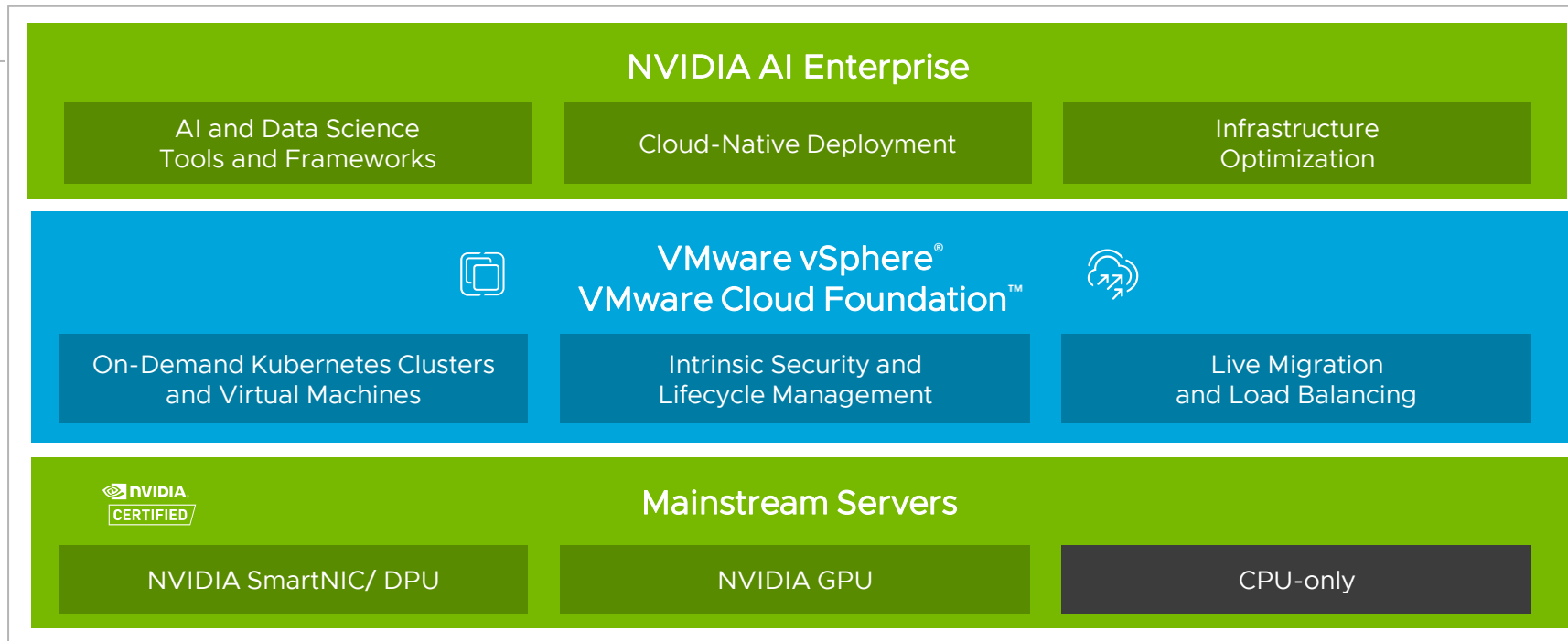
Faster Time-To-Value

AVAILABLE NOW

# VMware + NVIDIA AI-Ready Enterprise Platform



Data Scientist/  
Developer/  
AI Researcher



IT  
Administrator  
MLOps



Multi-Cloud



Hybrid Cloud



Private Cloud

# Leveraging VCF as a Platform for GenAI Workloads

## Solving for Common GenAI Workloads with VMware Cloud Foundation



### Scale Quickly

Scale compute and storage in tandem with your GenAI needs, and provide low-latency, optimized networking



### Management and Operations

Quick actions, quick troubleshooting and ability to bring-up and tear-down environments with a few clicks



### Cost and Capacity Planning

GenAI workloads can be expensive to host and resource demand can bloat quickly. Stay in control and predict growth



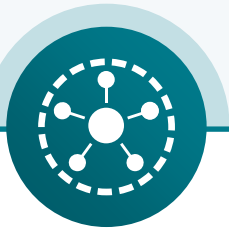
### Custom Dashboards

Track what's important whether it's capacity management, bottlenecks or other advanced KPIs



VMware Cloud Foundation™

# Key Benefits for Customers



## Flexibility of Choice

Broad AI ecosystem of partners



## Deploy with Confidence

Mitigate risks: data/IP leakage, model accuracy.



## Optimized Performance

Support for major NVIDIA accelerated computing & software



## Boost Productivity

Time-saving toolkit speeds developer workflows

# Common Customer Use Cases Addressed



## Code Generation

Accelerate  
developer velocity

**Contact Centers  
Resolution**  
Improve content  
quality and feedback



## IT Operations Automation

Significantly reduce IT  
operations time



**Advanced Information  
Retrieval**  
Document search, policy,  
and procedure search





# Resources

VMware Private AI Foundation with NVIDIA Solution Brief

<https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/docs/vmware-privateai-foundation-with-nvidia-solutions-brief.pdf>

Learn more on VMware Cloud Foundation AI/ML Solutions

<https://www.vmware.com/products/vsphere/ai-ml.html>

Deploying Enterprise-Ready Generative AI on VMware Private AI

<https://core.vmware.com/resource/deploying-enterprise-ready-generative-ai-vmware-vmware-cloud-foundation#executive-summary>



# Thank You



---

Thank you for downloading this VMware presentation! Carahsoft is the distributor for VMware solutions available via several federal, state and local, education and other contract vehicles.

To learn how to take the next step toward acquiring VMware's solutions, please check out the following resources and information:



For additional resources:  
[carah.io/vmwareresources](https://carah.io/vmwareresources)



For additional VMware solutions:  
[carah.io/vmware](https://carah.io/vmware)



To purchase, check out the contract vehicles available for procurement:  
[carah.io/vmwarecontracts](https://carah.io/vmwarecontracts)



For upcoming events:  
[carah.io/vmwareevents](https://carah.io/vmwareevents)



For additional solutions:  
[carah.io/vmware](https://carah.io/vmware)



To set up a meeting:  
[vmware@carahsoft.com](mailto:vmware@carahsoft.com) or 888-6VMWARE