



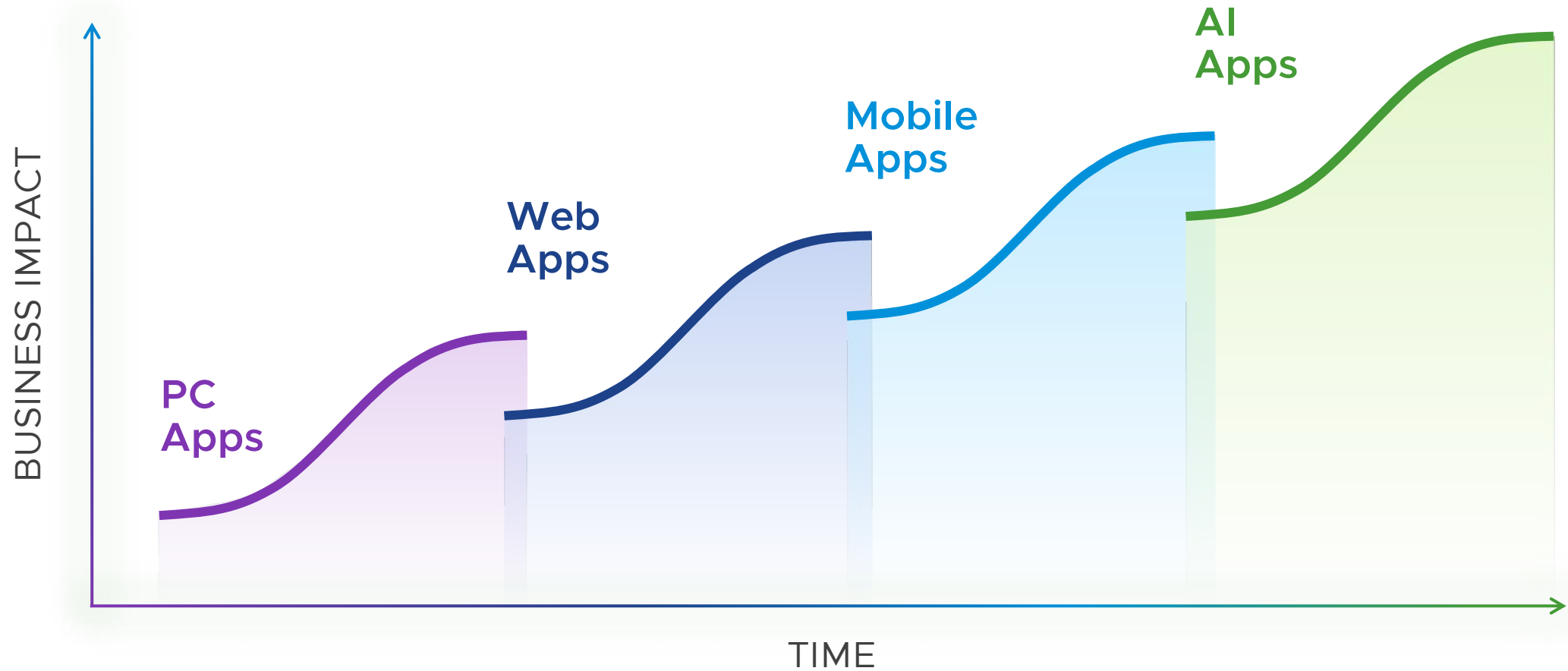
Accelerate
AI Innovation

VMware Private AI

Ruggero Citterio – Giuliano Copparoni
Tanzu Solution Engineers

Nov. 2023

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

Industries Racing to Adopt Generative AI

FINANCE

Fraud Detection
Personalized Banking
Investment Insights

HEALTHCARE

Molecule Simulation
Drug Discovery
Clinical Trial Data Analysis

RETAIL

Personalized Shopping
Automated Catalog Descriptions
Automatic Price Optimization

TELECOMMUNICATIONS

AI Virtual Assistants
Network Performance Tuning
Remote Support Capabilities

MEDIA & ENTERTAINMENT

Character Development
Video Editing & Image Creation
Style Augmentation

MANUFACTURING

Factory Simulation
Product Design
Predictive Maintenance

FEDERAL

Document Summarization
Audit Compliance
AI Virtual Assistants

ENERGY

Knowledge Base Q&A
Predictive Maintenance
Customer Service

Enterprises that adopt next-generation AI like LLMs and Generative AI are 2.6X more likely to increase revenue by 10% or more but must invest in their AI infrastructure to fully reap the benefits.

– Accenture Research. Breakthrough Innovation: Is your organization equipped for breakthrough innovation? WEF 2023.

Intelligent Assist

Ask anything or try one of the following prompts

🗨️ # ⓘ

Welcome to VMware Tanzu Hub, jelzein

My Accounts

Total	227
AWS	121
Azure	17
GCP	89

Top 5 Service Instances

- Kubernetes.Workload (508893)
- AWS.IAM (334718)
- AWS.EC2 (314031)
- Kubernetes.Resource (77653)
- AWS.SNS (74261)

Daily

ONBOARD ACCOUNT

Active Insights

Last 48 Hours

Critical: 23 | Warning: 38

61 Insights

VIEW ALL INSIGHTS

Governance

Open findings

273,330 ▲ 2% last 7 days

Critical	1
High	63K
Medium	112K
Low	98K

Types of findings

Drifts	721	Violations	272K	Threats	88
Vulnerabilities	6	Errors	122		

VIEW SUMMARY

Cost

Projected Cost: **\$3,715,139** ▼ 1% last month

Potential Savings: **\$0**

Rightsizing Savings Opportunities: 0

- AWS EC2 \$0
- AWS EBS \$0
- AWS RDS \$0

Cost per month by provider for 10 accounts

My Applications

Business: 415 | Potential: 1487

Recently modified applications

- appForE2eTesting - 2023-07-06 16:42:12.41761100
- time-series-service/time-series-service
- vss-policy-service/vss-policy-service-nlb
- appForE2eTesting - 2023-09-01 11:25:36.714833500
- appForE2eTesting - 2023-08-03 12:56:46.790254300

MODEL APPLICATIONS

Anomaly Detection

Total Anomalies: 7 | Total Cost Impact: \$689.10

Top Anomalies	Cost Impact
CloudHealth NG Staging	\$2.84
CloudHealth EDS Production	\$8.75
CloudHealth EDS Production	\$2.06

Provider: All (1667841)

Region: AWS: Africa (Cape ...)

Top 5 services by region

- AWS.EC2 (557)
- AWS.IAM (468)
- AWS.KMS (379)
- AWS.SNS (137)

Hi. I'm new here. Can you help me learn how to use Tanzu Hub?

✓

To learn how to use Tanzu Hub (formerly Aria Hub) as a beginner, you can follow these essential resources and steps:

1. VMware Tanzu Tech Zone: This technical content portal provides practical guidance for evaluating, implementing, and operating VMware Tanzu solutions. It offers demos, hands-on labs, and documents to help you learn. You can filter by product, such as Tanzu Kubernetes Grid, and explore the latest featured content.

2. Tanzu Academy: This on-demand learning hub offers curated material, including tutorials, guides, and hands-on labs, for products like VMware Tanzu Application Platform and VMware Tanzu for Kubernetes Operations. It helps improve your time to value with Tanzu products and grow skills for platform operations teams.

3. Spring Academy: If you're interested in learning about Spring, VMware offers Spring Academy. It provides accessible learning

jelzein

Top 5 Service Instances

- Kubernetes.Workload (508893)
- AWS.IAM (334718)
- AWS.EC2 (314031)
- Kubernetes.Resource (77653)
- AWS.SNS (74261)

Daily



Cost

Projected Cost
\$3,715,139
▼ 1% last month

Potential Savings
\$0

Rightizing Savings Opportunities
AWS EC2 [\\$0](#)
AWS EBS [\\$0](#)
AWS RDS [\\$0](#)

Active Insights

Last 48 Hours

Critical
23

Warning
38



Critical
Warning

VIEW ALL INSIGHTS

My Applications

Business
415

Potential
1487

Recently modified applications

- appForE2eTesting - 2023-07-06 16:42:12.41761100
- time-series-service/time-series-service
- vss-policy-service/vss-policy-service-nib
- appForE2eTesting - 2023-09-01 11:25:36.714833500
- appForE2eTesting - 2023-08-03 12:56:46.790254300

MODEL APPLICATIONS

Anomaly Detection

Total Anomalies
7

Total Cost Impact
\$689.10

Top Anomalies

- CloudHealth NG Staging \$2.84
- CloudHealth EDS Production \$8.75
- CloudHealth EDS Production \$2.06

Cost Impact

Modern Infrastructure

Clusters 183 Nodes 4475

Can you provide more information about Tanzu Academy and its offerings?

How can I access the latest featured content on Tanzu Tech Zone?

What are some recommended resources for learning about Tanzu Kubernetes

Display all applications with security vulnerabilities

Graph resources by region as pie chart

How many apps do I have by region?

How many apps do I have running across all regions?

List all potential business applications

Prompts

Type to chat



Intelligent Assist

4. KubeAcademy: While not directly related to Tanzu Hub, KubeAcademy is a valuable resource for learning Kubernetes, a core component of Tanzu Hub. It offers courses and tutorials on Kubernetes, including Docker, Kubernetes basics, and Kubernetes security best practices.

By utilizing these resources, you can gain the necessary skills and technical proficiency to effectively utilize Tanzu Hub as a beginner.

Welcome to VMware Tanzu Hub, jelzein

My Accounts

Total	227
AWS	121
Azure	17
GCP	89

Top 5 Service Instances

- Kubernetes.Workload (508893)
- AWS.IAM (334718)
- AWS.EC2 (314031)
- Kubernetes.Resource (77653)
- AWS.SNS (74261)

Daily

Active Insights

Last 48 Hours

Critical: 23, Warning: 38

61 Insights

VIEW ALL INSIGHTS

Governance

Open findings: 273,330 (▲ 2% last 7 days)

Threats: 88

Cost

Projected Cost: \$3,715,139 (▼ 1% last month)

Potential Savings: \$0

Rightsizing Savings Opportunities: 0

Cost per month by provider for 10 accounts

My Applications

Business: 415, Potential: 1487

Recently modified applications:

- appForE2eTesting - 2023-07-06 16:42:12.41761100
- time-series-service/time-series-service
- vss-policy-service/vss-policy-service-nlb
- appForE2eTesting - 2023-09-01 11:25:36.714833500
- appForE2eTesting - 2023-08-03 12:56:46.790254300

MODEL APPLICATIONS

regions. The region with the most resources is 'US' with 148 resources, and the region with the least resources is '' with 194 resources.

Provider: All (1667841)

Region: AWS: Africa (Cape ...)

Top 5 services by region

- AWS.EC2 (557)
- AWS.IAM (468)
- AWS.KMS (379)
- AWS.SNS (137)

Anomaly Detection

Total Anomalies: 7, Total Cost Impact: \$689.10

Top Anomalies

CloudHealth NG Staging	\$2.84
CloudHealth EDS Production	\$8.75
CloudHealth EDS Production	\$2.06

Here is the code

```

query countEntitiesByRegion {
  entityQuery {
    queryAggregations(
      aggregations: [
        {
          fieldName: "region"
        }
      ]
    )
  }
}

```

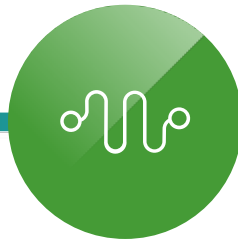
Loading...

Type to chat

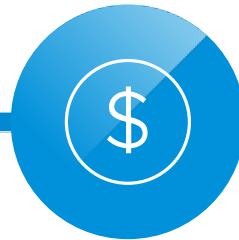
Key Challenges in Generative AI today



PRIVACY



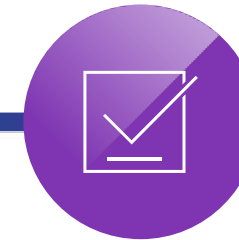
CHOICE



COST



PERFORMANCE



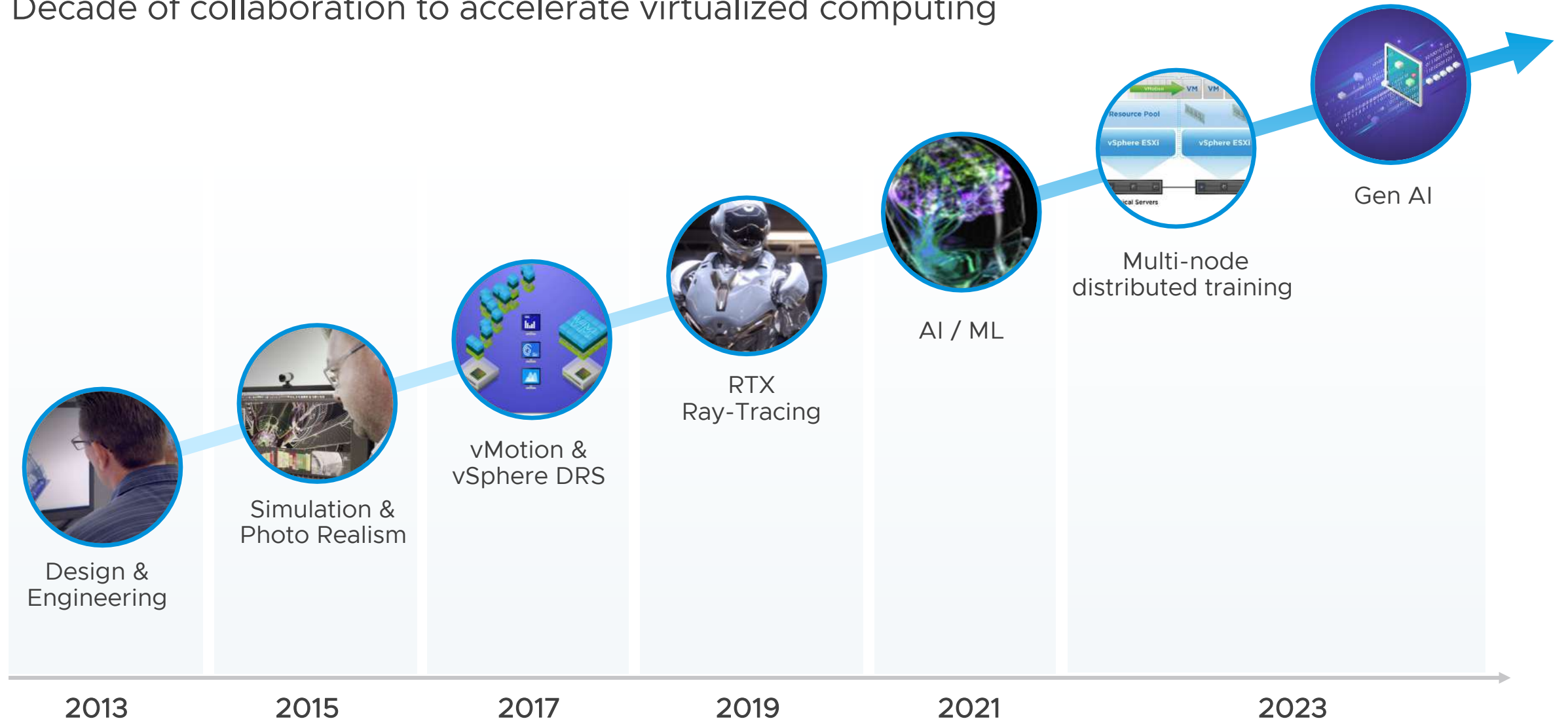
COMPLIANCE

INTRODUCING

VMware Private AI Foundation WITH NVIDIA

VMware & NVIDIA: 10 Years of Innovation

Decade of collaboration to accelerate virtualized computing



Integrated Gen-AI Platform: Available Early 2024


Falcon LLM


Llama 2

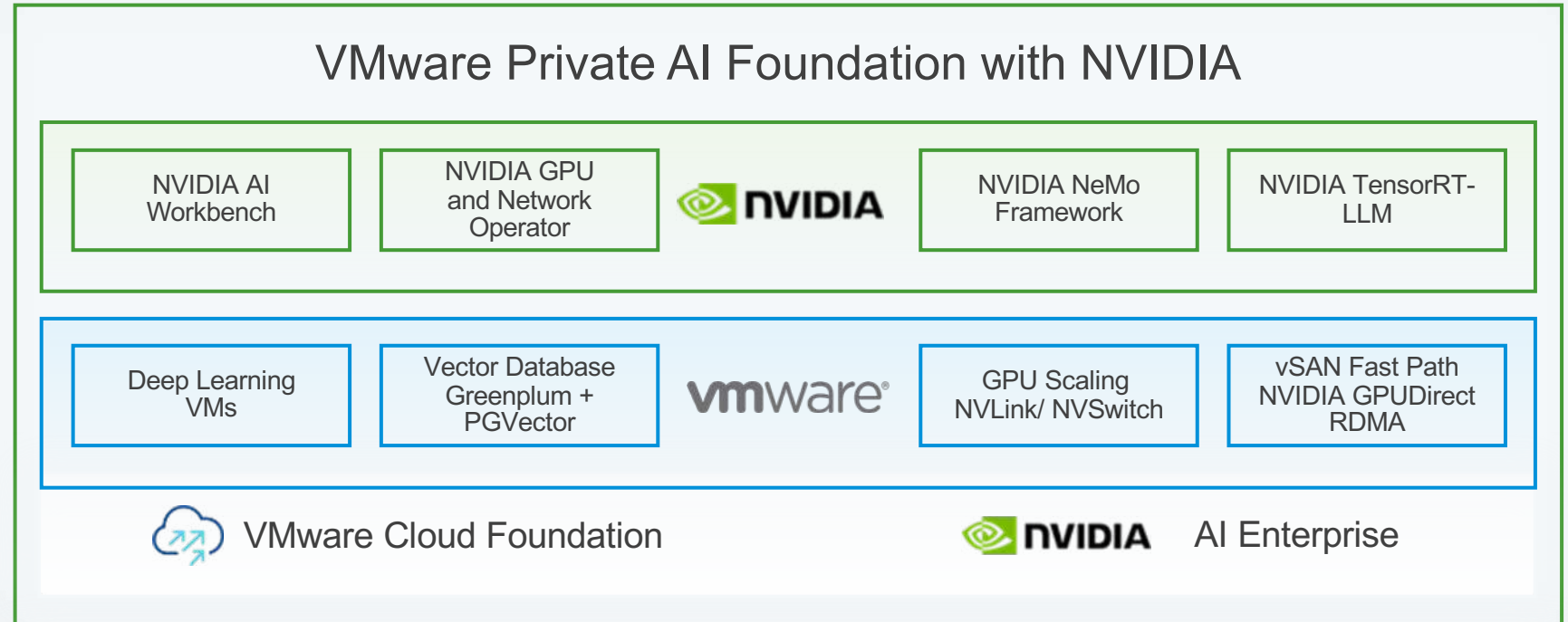

MPT


NVIDIA NeMo

INTRODUCING

VMware Private AI Foundation

WITH NVIDIA









NEW

VMware Private AI with Intel

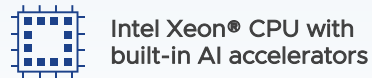
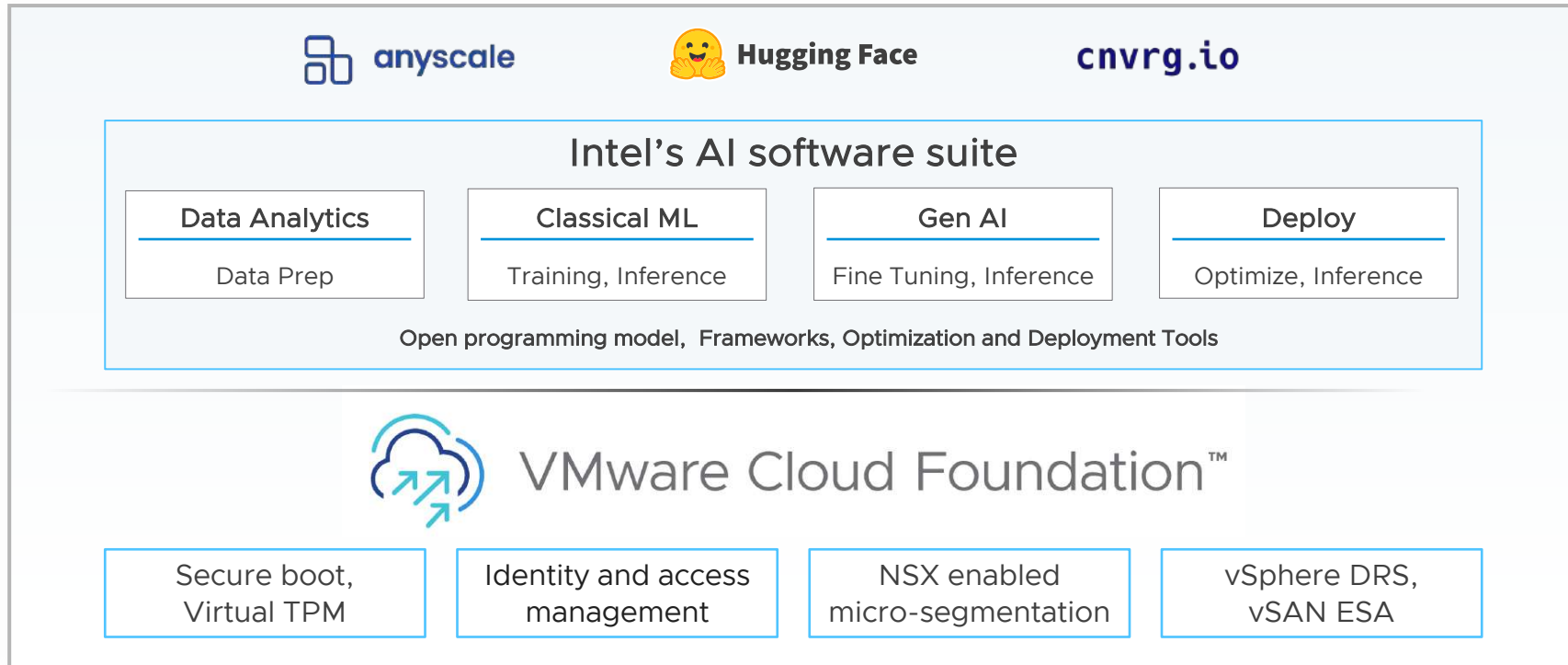
VMware Private AI with Intel



Data Scientist



Admins



DELL Technologies

Hewlett Packard Enterprise

Lenovo

NEW

VMware Private AI with IBM

VMware Private AI with IBM



Data Scientist



Admins



Hugging Face



PyTorch

watsonx



Red Hat
OpenShift



VMware Cloud Foundation™

Open Source

Open Source AI

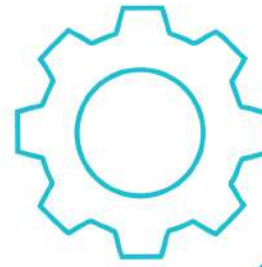


INTRODUCING

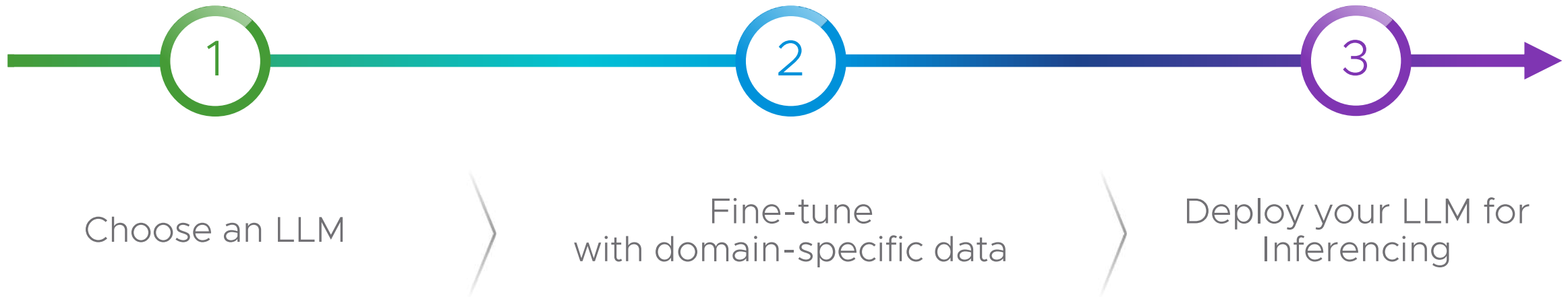
VMware Private AI

Democratize generative AI and ignite business innovation for all enterprises.

An architectural approach that balances the business gains from AI with the privacy and compliance needs of the organization.



Key Value Provided by VMware Private AI



VMware Private AI Open Ecosystem



Dolly



Falcon



Llama 2



MPT



Platypus 2



anyscale

cnvrg.io



DOMINO



FedML



Hugging Face



Kubeflow



NVIDIA
NEMO



One Convergence™



PyTorch



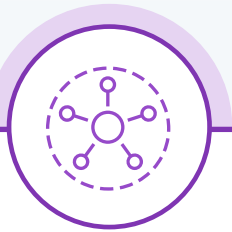
run:ai



Weights & Biases



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

Reference Architecture for VMware Private AI



Falcon LLM



Llama 2



Hugging Face



MPT



NVIDIA NeMo



Deep Learning VMs

Greenplum Vector DB

GPU Scaling

vSAN Fast Path



VMware Cloud Foundation



Compute



Storage

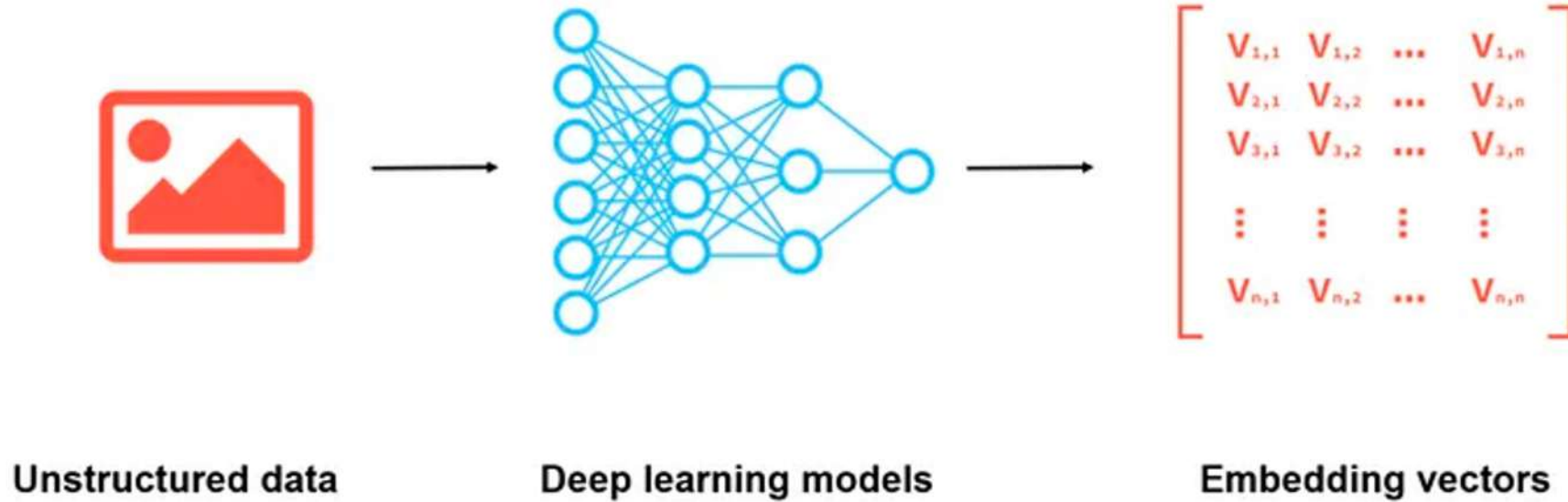


Network/DPU



GPU

The Rise of Vector Database



Vector database startups raise over \$350M to build generative AI infrastructure

April 27, 2023



[Artificial Intelligence](#)

[Client Intelligence](#)

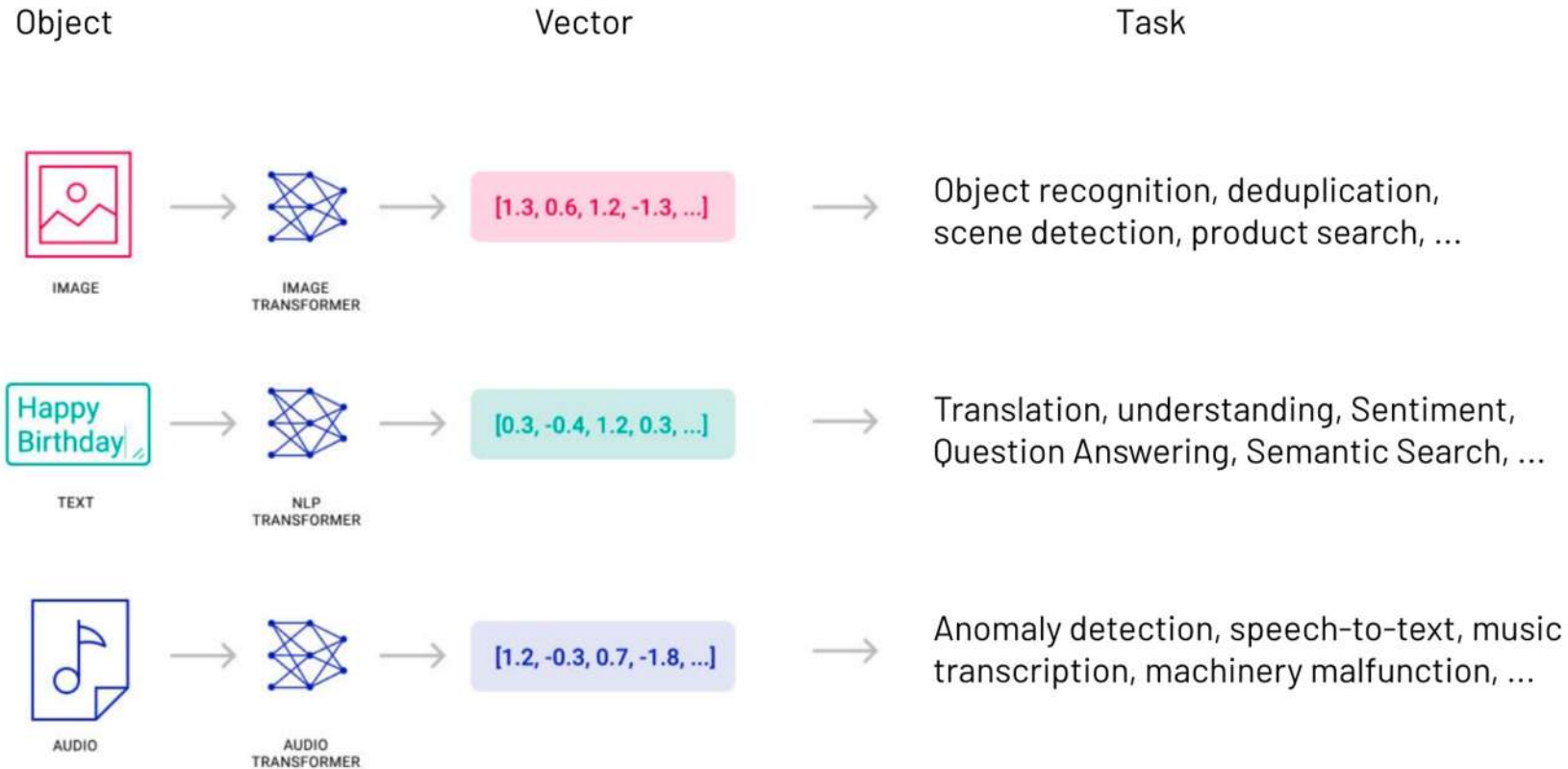
[Core Intelligence](#)

[Data Management](#)

[Enterprise Tech](#)

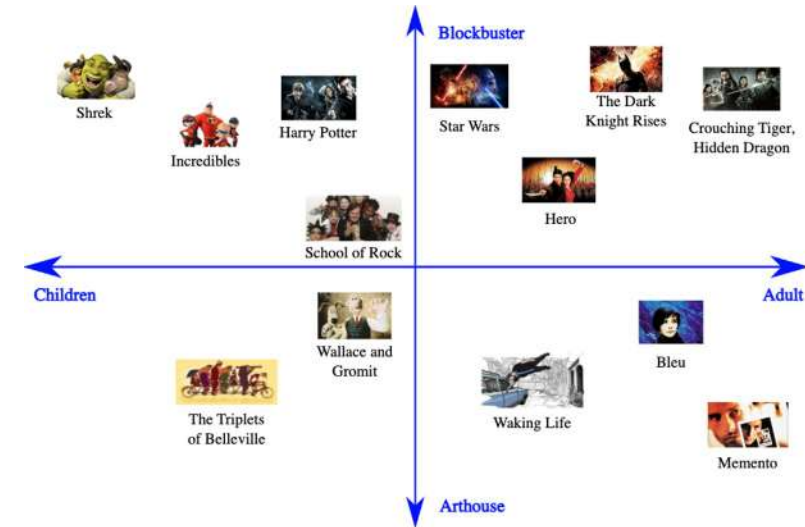
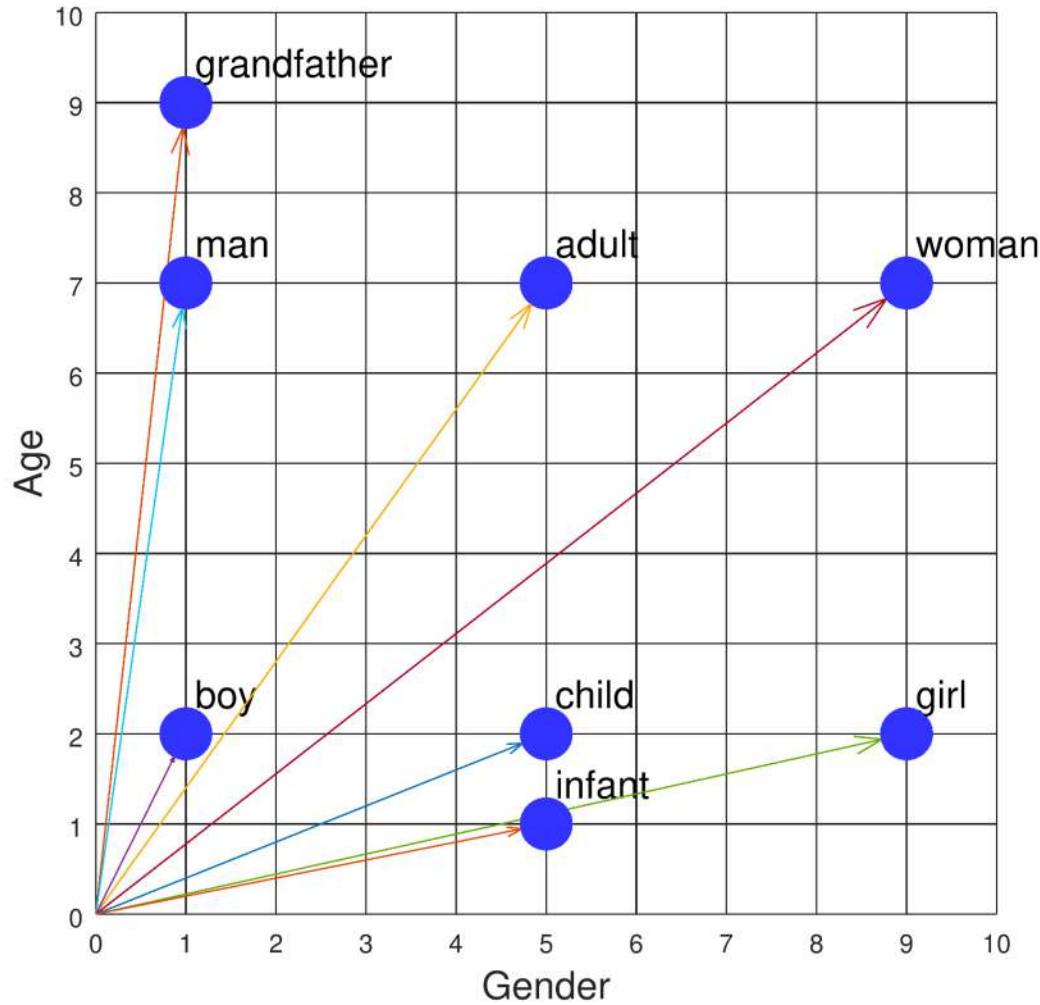
[TMT](#)

How to Represent Complex Data in a Standardized Way



Text / Image Embeddings - Semantic Similarity Search

Words As Vectors



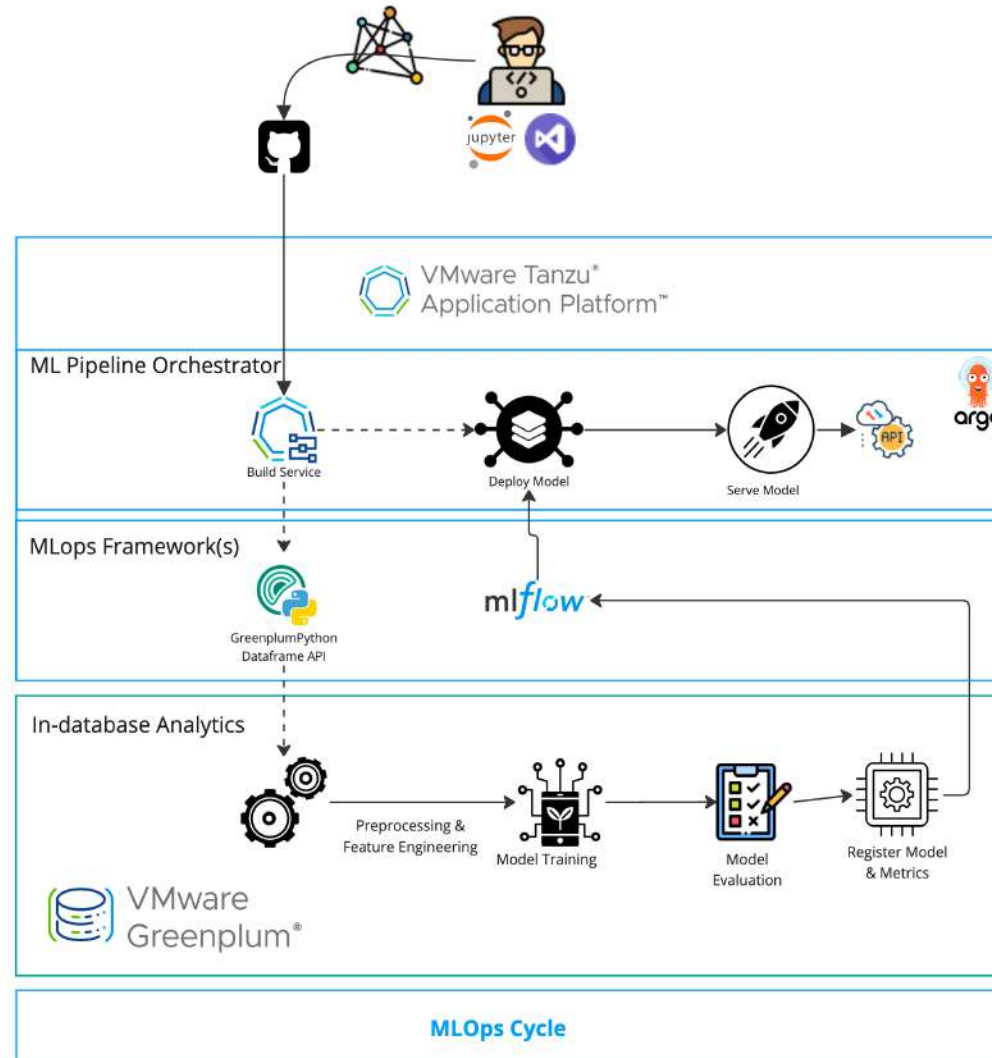
Tanzu Application Platform

Development, deployment and management for modern ML/AI applications

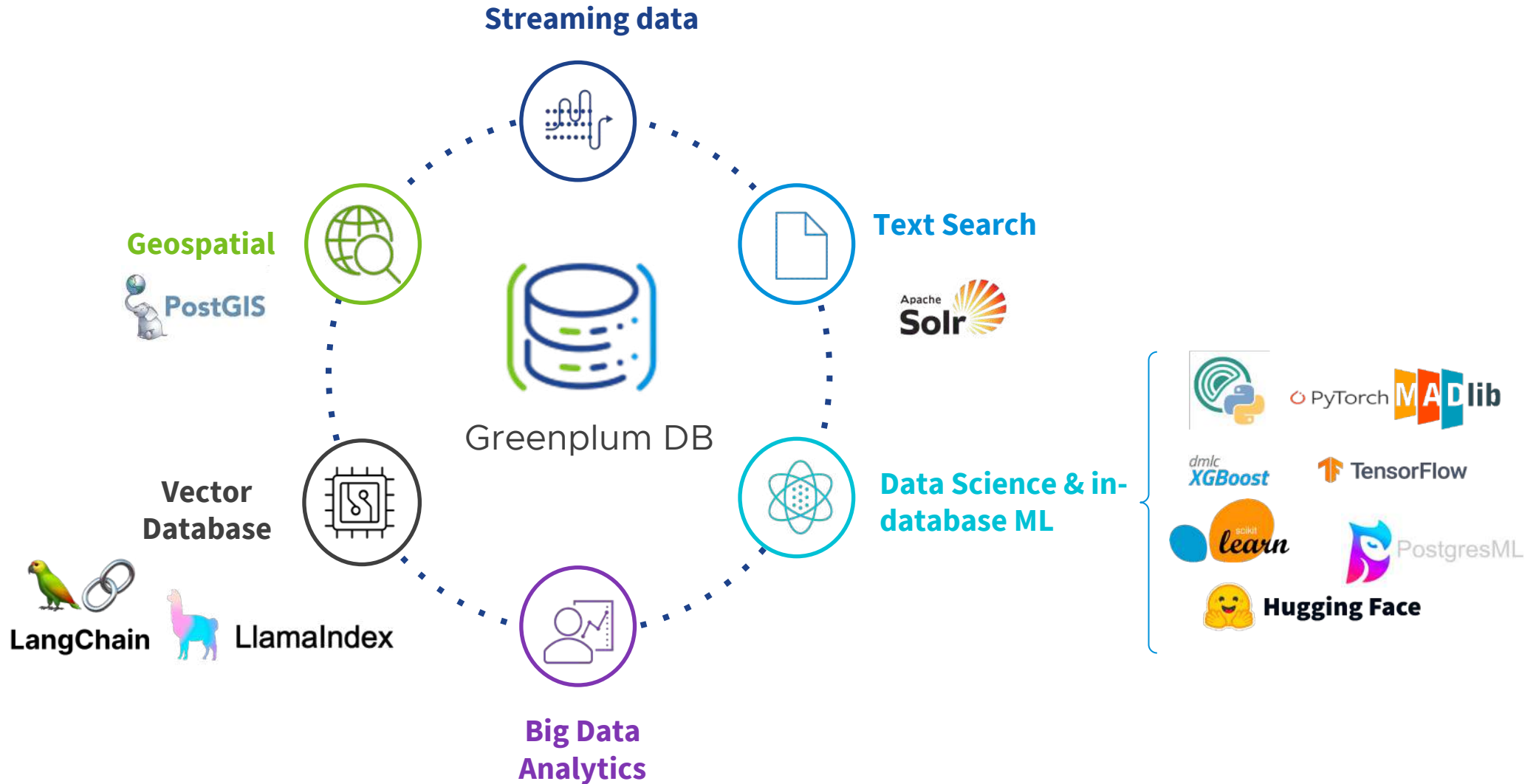
Tanzu Application Platform

GreenplumPython Dataframe API

Greenplum Data Lake



ONE PLATFORM • MANY WORKLOADS • ANY DATA • ANY SIZE • ANYWHERE



Demo

Learn More

Private AI Resources

- [VMware Private AI – Press Release](#)
- [VMware Private AI Launch Blog](#)
- [VMware AI and ML page](#)
- [Private AI Reference Architecture White Paper](#)
- [Joint webinar with NVIDIA](#)
- [Dell support blog](#)



Thank You