

The logo consists of the letters 'NTS' in a bold, dark blue, sans-serif font, centered within a white square with rounded corners.

NTS

**RELAX,
WE CARE**



RELAX,
WE CARE

PRIVATE AI

KI SICHER UND UNTER KONTROLLE

WER SIND WIR



Dominik Steiner
Systems Engineer



Christian Schuster
Sales Expert Data Center



1000

Kunden



800+

MitarbeiterInnen



22

Standorte



100%

Eigentümer geführt



341 Mio €

Umsatz 2023



100%

Menschlichkeit

STRATEGIC PARTNER



PARTNER

vmware®

by **Broadcom**

NetApp

PRESTIGE PARTNER

CISCO
Partner

DELLTechnologies
TITANIUM PARTNER

paloalto®
NETWORKS

Commvault®

splunk>

FORTINET.

SOLUTION PARTNER

SUSE

tenable



veeam

GitLab

CLOUD PARTNER

Microsoft
Partner



imagicle

Megaport

Kiteworks

HashiCorp

NTW®

NTS

UNSERE KUNDEN

EIN AUSZUG AUS DEN REFERENZEN



universität
innsbruck

FH JOANNEUM
University of Applied Sciences

MCI
MANAGEMENT CENTER
INNSBRUCK

fh
KufsteinTirol
UNIVERSITY OF APPLIED SCIENCES

MEDIZINISCHE
UNIVERSITÄT
INNSBRUCK

unibz

FACHHOCHSCHULE
KÄRNTEN



MED^{EL}

SVS
Gemeinsam gesünder.

SOS
KINDERDORF

Die Ziegler'schen

GRAWE



ÖSTERREICHISCHES
ROTES KREUZ



BRZ

3 Banken IT

ENERGIEAG
Oberösterreich

Gruppe

brenner.com

HAUFE.Group

Verbund



(STYRIA

HOLDING
GRAZ

ITSMedia
MEDIA IT SOLUTIONS

Red Bull

SPAR

AT&S



REWE
GROUP



PAPPAS

ZEPPELIN^{NT}

EGGER

TYROLIT

HILTI

Berger
LOGISTIK

plansee
GROUP

transped

mondi

blum

JCL
LOGISTICS



IMA SCHELLING
GROUP

KB

MAGNA

sappi

ANDRITZ

KNAPP



OMICRON

KERBL





WARUM BRAUCHT ES PRIVATE AI?

AI VOR GENERATIVE AI



Data Scientists



Vorausschauende Analysen



PREDICTIVE AI

Spezialisiertes AI-Modell

GENERATIVE AI

PREDICTIVE AI



Data Scientists



Vorausschauende Analysen



Spezialisiertes AI-Modell

GENERATIVE AI



Marketing



Lieferketten



Verkauf



Personalwesen



Kunden-Service



Forschung und Entwicklung



Rechtswesen



Software-Entwicklung



Produktion



Verrechnung



IT



Finanz

INTERAKTION MIT NORMALER SPRACHE

APIs



LARGE LANGUAGE MODELLE

GEN AI MEHRWERTE



Finanz

Betrugserkennung
Personalisiertes Banking
Investment Analysen



Gesundheitswesen

Molekular-Simulation
Wirkstoff-Forschung
Analyse klinischer Studien



Einzelhandel

Personalisiertes Einkaufserlebnis
Automatisierte
Produktbeschreibungen
Automatisierte Preis-Anpassung



Telekommunikation

Betrugserkennung
Kundenservice



Unterhaltung

Avatar Entwicklung
Video-Bearbeitung und
Bildgeneration



Produktion

Werkschrittvereinfachung
Betriebsunterstützung



Öffentlicher Sektor

Dokumentenmanagement
Compliance Unterstützung
AI Assistenten



Energiewesen

Q&A mit Wissensdatenbank
Proaktive Wartung
Kundenservice

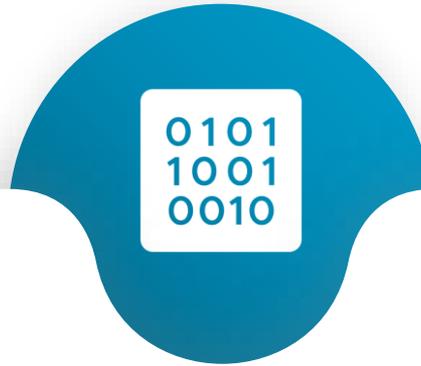
WARUM BRAUCHT ES PRIVATE AI?

Geistiges Eigentum



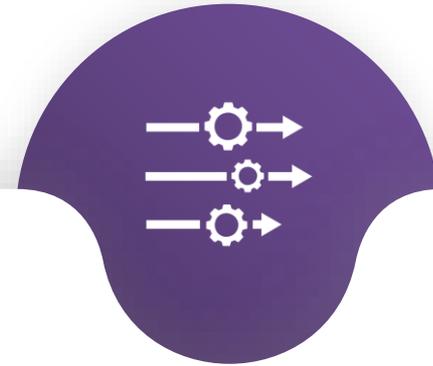
“Ich muss den Verlust meines geistigen Eigentums verhindern”

Geschäftskritische oder sensible Daten



“Ich möchte sicherstellen, dass meine Daten niemals nach außen gehen”

Zugriffskontrolle



“Ich möchte vollständige Kontrolle über den Zugriff auf meine AI-Modelle”

SCHLÜSSEL-ASPEKTE FÜR AI

Wahlfreiheit



Compliance



DATENHOHEIT



Kostenkontrolle



Skalierbarkeit & Effizienz



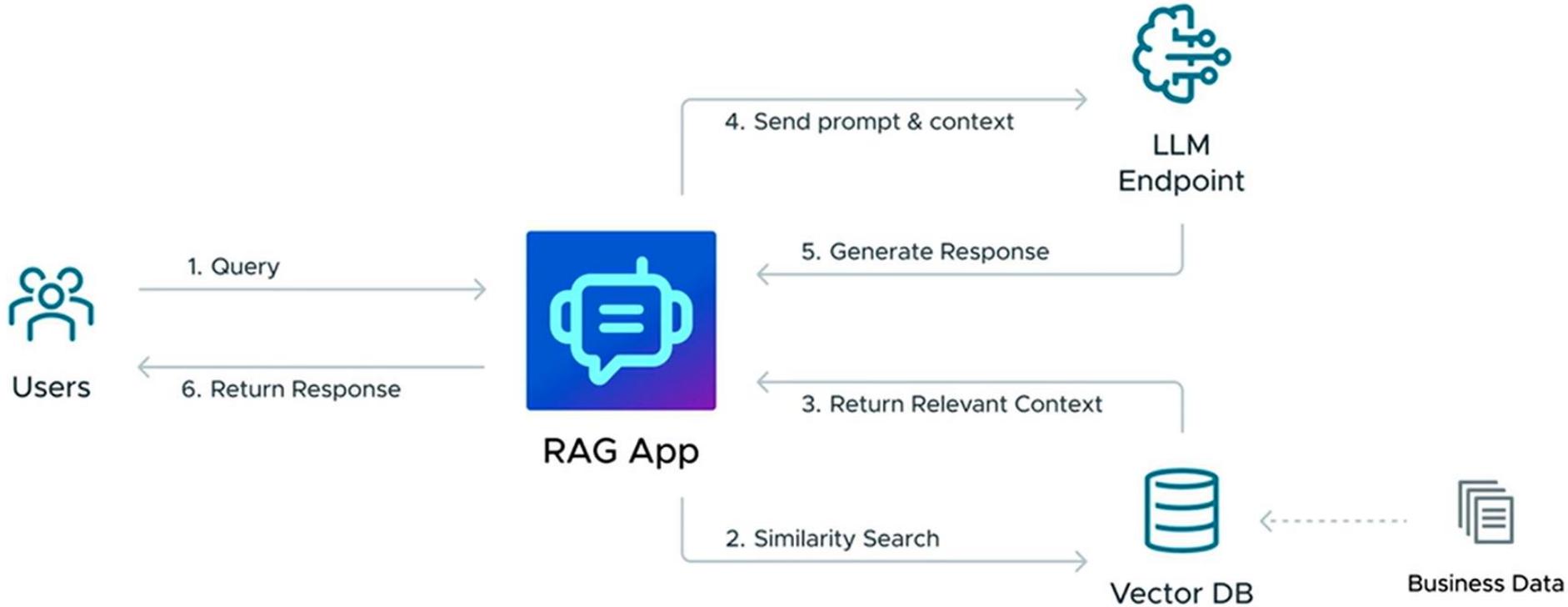
WIE WERDEN LLMS EINGESETZT?

1 Auswahl
des LLM

2 Fine-Tuning
mit kunden-
spezifischen
Daten

3 LLM-
Implementierung
für Inferencing

RETRIEVAL AUGMENTED GENERATION(RAG) FINE-TUNING



LLM-ZUGRIFFSKONTROLLE



Role-Based Access Control

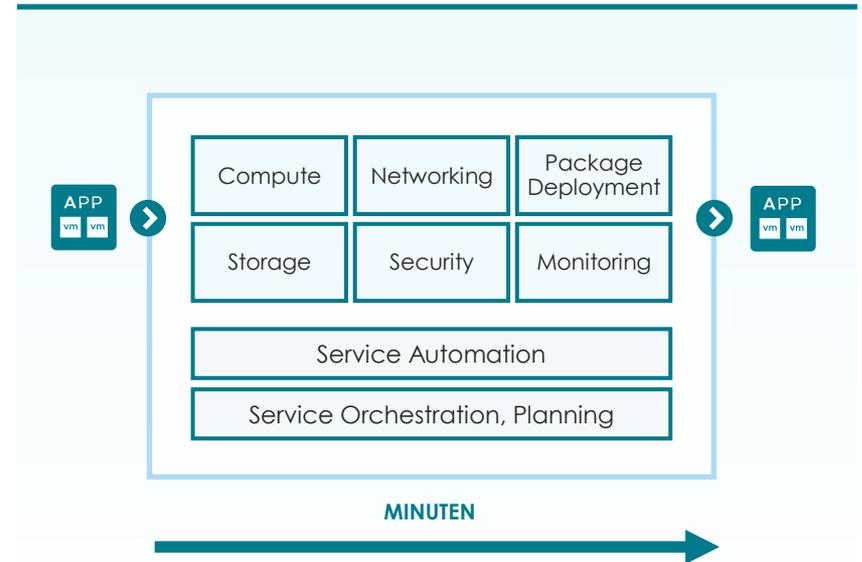
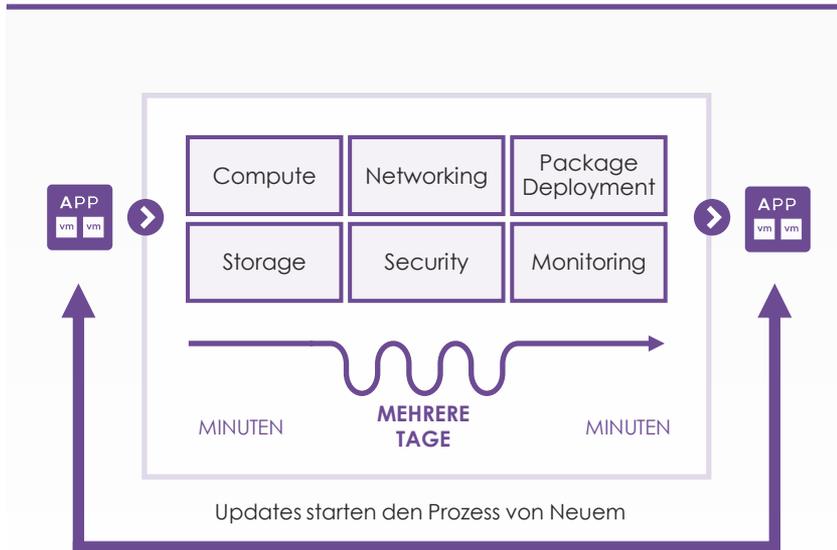


PLATTFORM OHNE NEUE SILOS

ENDE-ZU-ENDE AUTOMATISIERUNG INKLUDIERT

Traditionelle Umgebungen

VMware Cloud Foundation

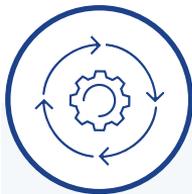


VCF - BAUSTEINE



VMware Cloud Foundation™

SDDC Manager



Auto-matisierter Betrieb

Einfacher Betrieb von Day 0 – Day X

vSphere



Führender Hypervisor

Kern des SDDC
bekannte Tools & Skills

vSAN/ vVOLS



Verteilter sicherer Speicher

verteilte Datendienste

NSX



SDN und Security

verteiltes Switching und Routing

IPS-IDS, LB, SDFW

Tanzu

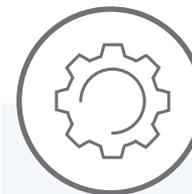


CNCF konformes Kubernetes

Betrieb von VMs und Container

Enterprise K8s Plattform

Aria



Cloud Management

Monitoring und Logging für Stabilität

Self-Service Automatisierung

NTS

Leveraging VCF as a Platform for GenAI Workloads

Solving for Common GenAI Workloads with VMware Cloud Foundation



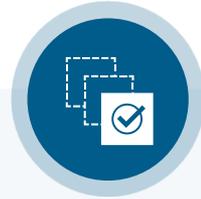
Scale Quickly

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™

WARUM IST VCF DER MATCH FÜR GEN AI



Skalierbarkeit

Compute, GPUs, Netzwerk und Speicher optimiert abgestimmt für wachsende GenAI-Projekte ausrollen



Management und Betrieb

Vereinfachter Betrieb, schnelles Troubleshooting und automatisierte Bereitstellung von KI Umgebungen



Planungssicherheit für Kosten und Kapazität

Kontrolle über kostenintensive GenAI Workloads, um Planung und ROI zu garantieren



Self-Service & Dashboards

Direkte Einblicke in Ressourcen-Management, Engpässe oder weitere KPIs



VMware Cloud Foundation™

PRIVATE AI MIT NVIDIA



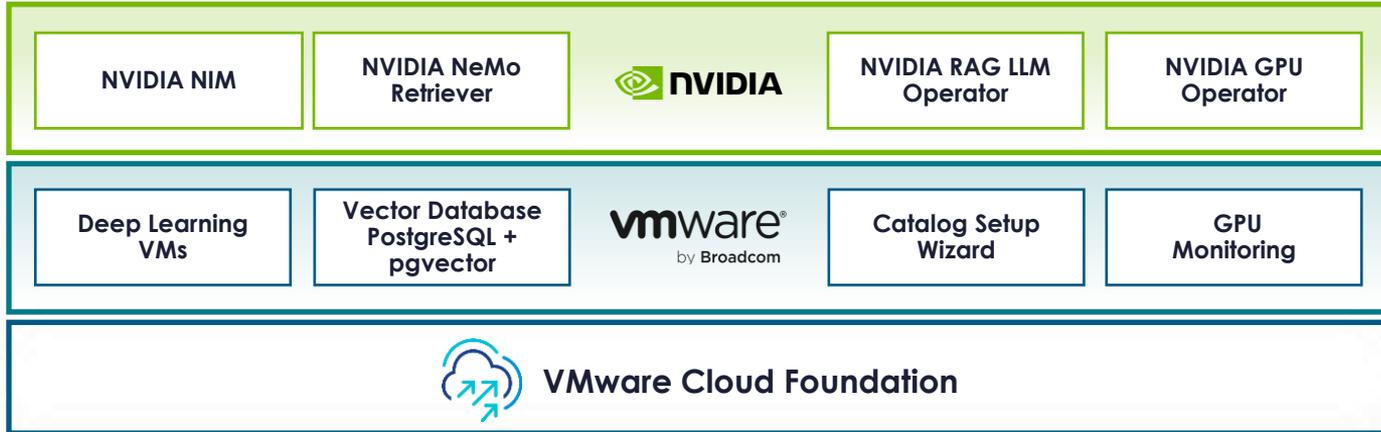
NVIDIA
Foundation Models



NVIDIA Fine-Tuned
Models



Third party &
Community Models



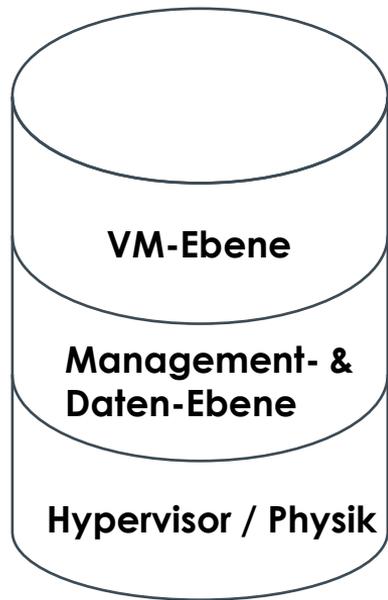
DELL Technologies

CISCO
Partner

AI SECURITY



AI PLATFORM SECURITY AUF BASIS CIA



VM-Verschlüsselung, Mikrosegmentierung, Snapshots

Identity & Access Management, Role Based Access Management, Daten-Verschlüsselung

UEFI Secure Boot, AES-NI
Intel SGX, AMD SEV-ES

PRIVATE AI

GARANTIE FÜR KONROLLE UND SICHERHEIT

verwahrungsmäßige Kontrolle

- sichere und valide Innovationen erlauben, ohne externe Abhängigkeiten
- Innovationen ermöglichen durch einfach Integration neuer Modelle
- Effizienz steigern durch gezieltem Ressourceneinsatz

Security

- Lokale Daten, Integration in bestehende Security-Richtlinien & Compliance-Vorgaben
- Vertrauen und Transparenz durch nachvollziehbare Vorgänge

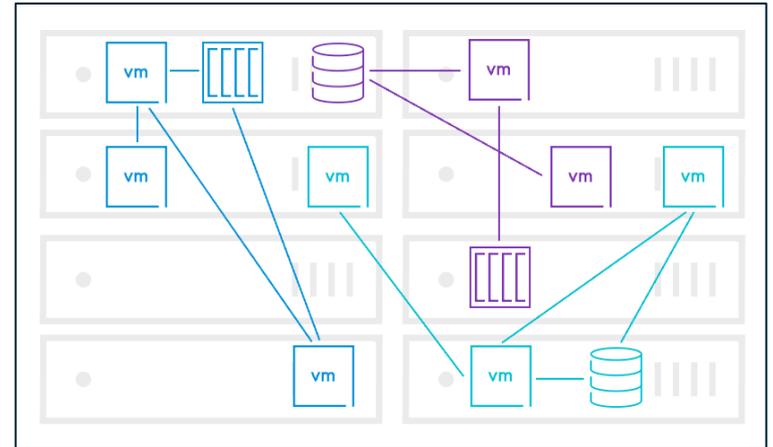
KONTROLLE AUF WORKLOAD-EBENE MIT VMWARE NSX

Was ist NSX?

- Netzwerkvirtualisierung
- Starke Integration mit vSphere & Tanzu
- NSX Application Platform & Advanced Load Balancer(ALB)
- Erweiterungen durch Dritthersteller (AV)
- 2 Bereiche
 - Networking 
 - Overlay, NAT, DHCP, DNS, VPN, Load Balancing
 - Security 
 - Firewall (L2-L7 & IDFW), IDS/IPS, Malware Prevention

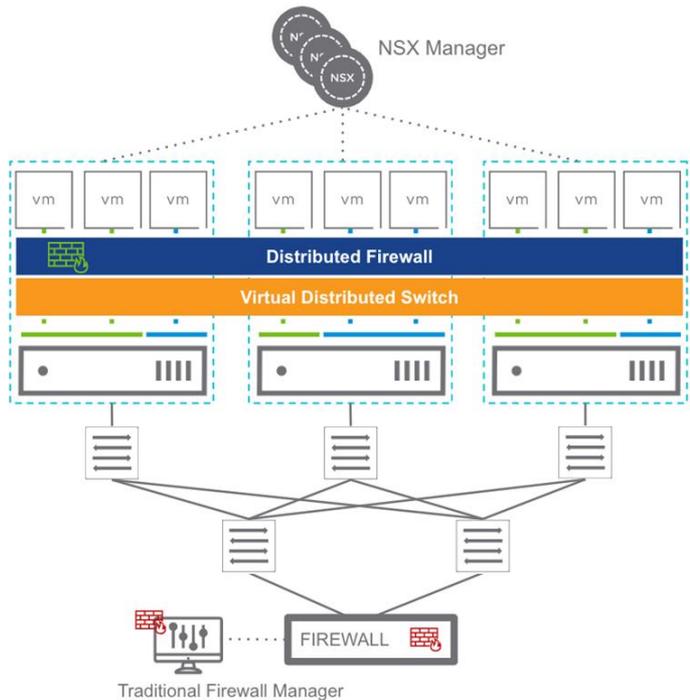
NSX SECURITY USE CASES

- (Mikro-)Segmentierung / Zero Trust 🚫
 - ☐ Applikationen untereinander abschotten
- DMZ 🛡️
 - ☐ Abgeschottete Umgebung für alte Systeme
- VDI / RDSH 🖥️
 - ☐ IDFW (Dienstleister, Abteilungen, etc.)
 - ☐ Malware Prevention
 - ☐ URL Filtering
 - ☐ TLS Inspection
- Compliance ⚖️
 - ☐ PCI-DSS, HIPAA, etc.



Keine Änderungen am Netzwerk!

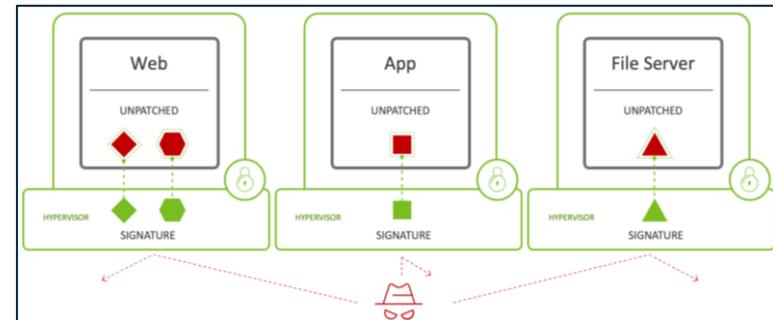
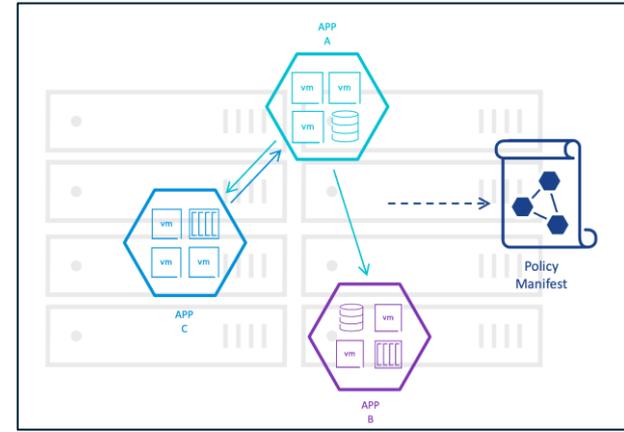
NSX SECURITY: VDEFEND FIREWALL ARCHITEKTUR



- ✓ Verteilt und in den Hypervisor integriert - Skaliert linear mit der Anzahl der Hosts
- ✓ Keine Änderungen am Netzwerk notwendig
- ✓ Jedes Paket passiert die NSX DFW
- ✓ Ein Management (GUI, CLI, API) für die Policies

WARUM NSX?

- Ransomware & Security beschäftigt jede Organisation
 - Lateral Movement
 - Virtual Patching durch NSX IDS/IPS
- Netzwerktopologie muss NICHT verändert werden
 - Subnetze und IPs bleiben gleich
 - VLANs bleiben gleich
 - Portgruppen bleiben gleich
- Keine (oder kleinere) DC Firewall
- Einfach(st)e Option für DC Segmentierung
- VMware vSphere ist allgegenwärtig



VMWARE INTELLIGENT ASSIST

KI-Assistent-LLM für schnelle Analyse

The screenshot displays the VMware Intelligent Assist interface. At the top, there are tabs for 'AI CHAT' and 'GUIDES', along with a 'Clear Chat' button. The chat history shows a user asking, 'Which vms are choking my network?' at 4:55 PM on Jun 15, 2023. The assistant responds with a query: 'vms where network traffic rate >= 100 mbps order by network traffic rate desc' at the same time. Below the chat, a confirmation bar shows the executed query with a green checkmark. The results section indicates 'Showing 6 results for VM at Jun 15, 16:55'. A 'Sort' dropdown is set to 'Dsc ↓'. A summary bar shows '6 entities' and highlights 'APP_TIER_VM1' with a 'TROUBLESHOOT' button. Below this, it specifies 'Network Traffic Rate | Average for selected duration: 9.1 Gbps'.

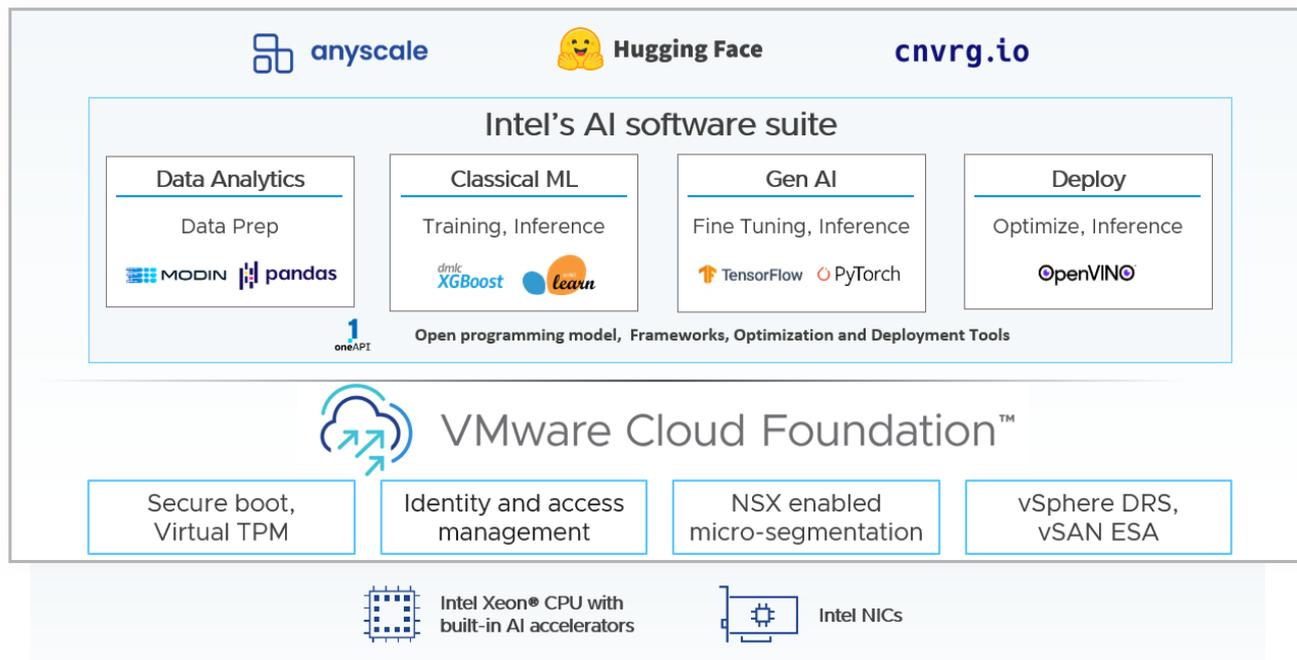


PRIVATE AI OHNE GPUS MIT INTEL CPUS

 NLP  Price Prediction  Fraud Detection  Mistral  Falcon  Llama 2


Data Scientist


Admins



HÄUFIGSTE USE-CASES FÜR GEN AI



Code Erstellung

Effizienzsteigerung
der Entwickler



Kunden Service

Verbessern von
Unterlagen,
Kundenanfragen
& Feedback



IT Betrieb Automatisieren

IT Betriebsteams
signifikant
entlasten



Erweiterte Datenanalyse

Dokumentensuche,
Richtlinien- und
Arbeitsanweisungs-
Bereitstellung

VMWARE PRIVATE AI VORTEILE



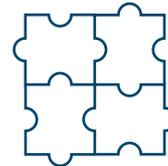
**Skalierbarkeit & Automatisierung
durch VMware Cloud Foundation**



**Kontrolle über Daten-Zugriff und
deren Nutzung**



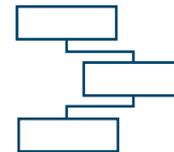
**Effizienzsteigerung durch bessere
Auslastung und leichterem Betrieb
von (GPU-)Ressourcen**



**Deep Learning VM Templates
& Vector Datenbank für
RAG-Workflows**



**kein neuer Silo für Betrieb
von AI & Nutzen von
Partner-Ökosystem**



**Self-Service Plattform für
Data Scientists +
Automatisierung**



FRAGEN?



DANKE

The logo consists of the letters 'NTS' in a bold, dark blue, sans-serif font, centered within a white square with rounded corners.

NTS

**RELAX,
WE CARE**