

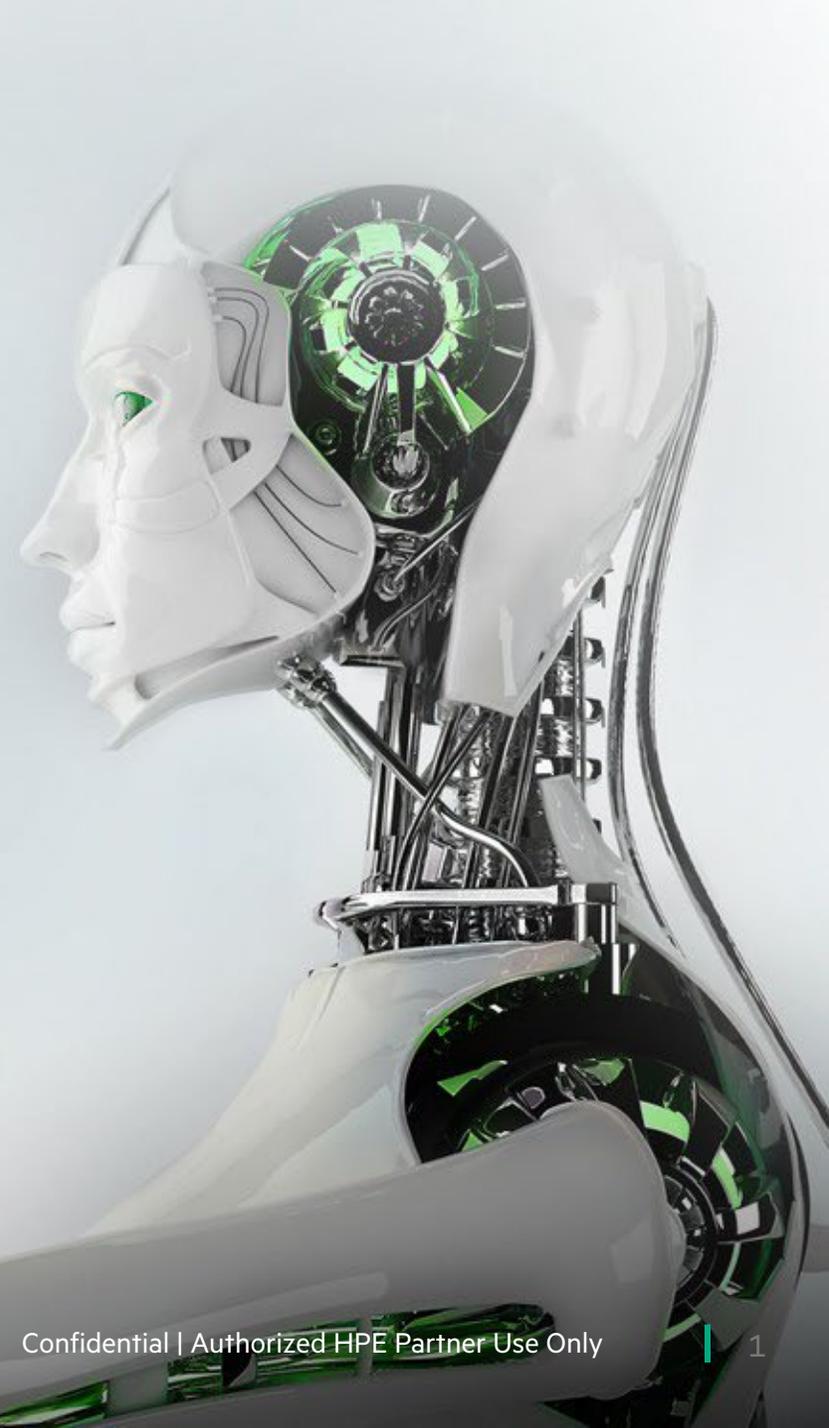


**Hewlett Packard
Enterprise**

Die Begeisterung rund um KI: Von der Konzeption bis zur Anwendung in Ihrem Unternehmen gemeinsam mit HPE und BORGWARE

**Jeremy Massow
Leoni Albers**

04. Juli 2024



Lösung:
As a service

Industrie:
Automotive

Region:
Europa

Ziele

- Entwerfen von Software, um die ersten selbstfahrenden Autos zu ermöglichen
- Reduzieren von Verkehrstoten und -verletzten
- Sichere Autos schneller auf den Markt bringen

Anforderungen

- Erstellen eines KI-Algorithmus, der mit der Zeit lernt
- Erstellen eines Round-Trip-Datenzyklus, der bessere KI-Entscheidungen ermöglicht
- Bereitstellen von zehntausenden Fahrsimulationen pro Sekunde

Lösung

- HPE GreenLake Cloud Services
- HPE GreenLake Management Services
- HPE ProLiant DL Gen10 servers
- Lustre filesystem
- HPE Data Management Framework
- Advisory and Professional Services
- HPE Security and Risk Management Services



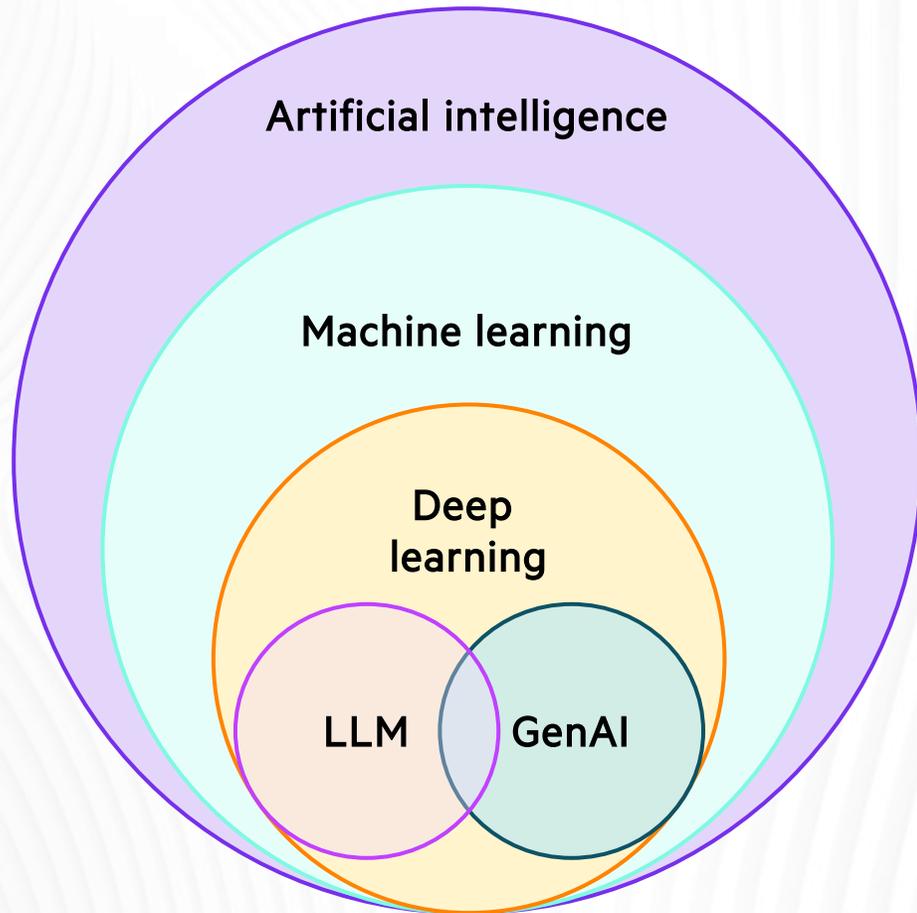
Ergebnisse

- Entwickelt und implementiert KI-Lösungen schneller als die Konkurrenz
- Schafft die Grundlage für weniger Unfälle und eine sicherere Welt
- Agilität gewinnen zu geringeren Kosten mit HPC as a Service

Weitere Ressourcen

- [Case study](#)
- [Video](#)

KI hat viele Formen ...



Generative AI strebt ein Verständnis auf menschlicher Ebene und einen leichten Umgang mit Sprache und Bildern an.

Die neuesten einsatzbereiten Modelle können :



Verstehen und erzeugen
Text, Sprache, Code



Interpretieren
Bilder, Text

Stichworte zum Thema „Künstliche Intelligenz“

Automatisierung

Prozesse wiederholt Ausführen mit vordefinierten Aktionen in einer festgelegten Reihenfolge.

Künstliche Intelligenz

Künstliche Intelligenz ist die Fähigkeit einer Maschine, menschliche Fähigkeiten wie logisches Denken, Lernen, Planen und Kreativität zu imitieren. Die Maschine erkennt Muster und kann darauf aufbauend Entscheidungen treffen

Machine Learning

- Machine Learning ist ein Teilbereich der künstlichen Intelligenz
 - Systeme lernen selbstständig Muster und Zusammenhänge aus großen Datenmengen
 - Maschnielles Lernen wird in drei Methoden unterteilt: Supervised Learning, Unsupervised Learning und Reinforcement Learning
 - Deep Learning ist eine besondere Form von Machine Learning
-

Computer Vision

Technologie, um Computer in die Lage zu versetzen visuellen Input zu verstehen, zu analysieren und darauf basierende automatisierte Maßnahmen zu ergreifen.

Generative AI

Eine Form von Künstlicher Intelligenz, die in der Lage ist, auf Basis vorhandener Informationen und Vorgaben eines Anwenders neue Inhalte zu generieren



Das Rezept für Mehrwerte aus Daten



Daten

Grundzutat, in vielen Unternehmen meist reichlich vorhanden



Ideen

Wie kann ich mit Daten mein Geschäft optimieren oder mit Daten neue Produkte und Services entwickeln?

Grundzutat, idealerweise im Unternehmen vorhanden, können durch

- Inspiration von Erfolgsgeschichten anderer HPE Kunden und
- Gemeinsamen Workshop hinzugefügt werden

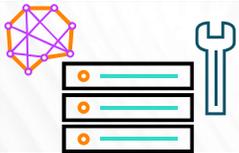


KI-Anwendungsgebiete

Wo könnte eine KI Anwendungsgebiete finden?

Beispiele, wo sich KI einsetzen lässt:

- Bilderkennung; Spracherkennung; Mustererkennung
- Prozessoptimierung.....



Plattformen

HPE hat für alle Bedürfnisse die richtige Lösung:

- Infrastruktur für Daten- und KI-Plattformen
Softwareplattformen für (hybride,) verteilte Data Lakes
Softwareplattformen für Data Scientist / Engineers / Analysts

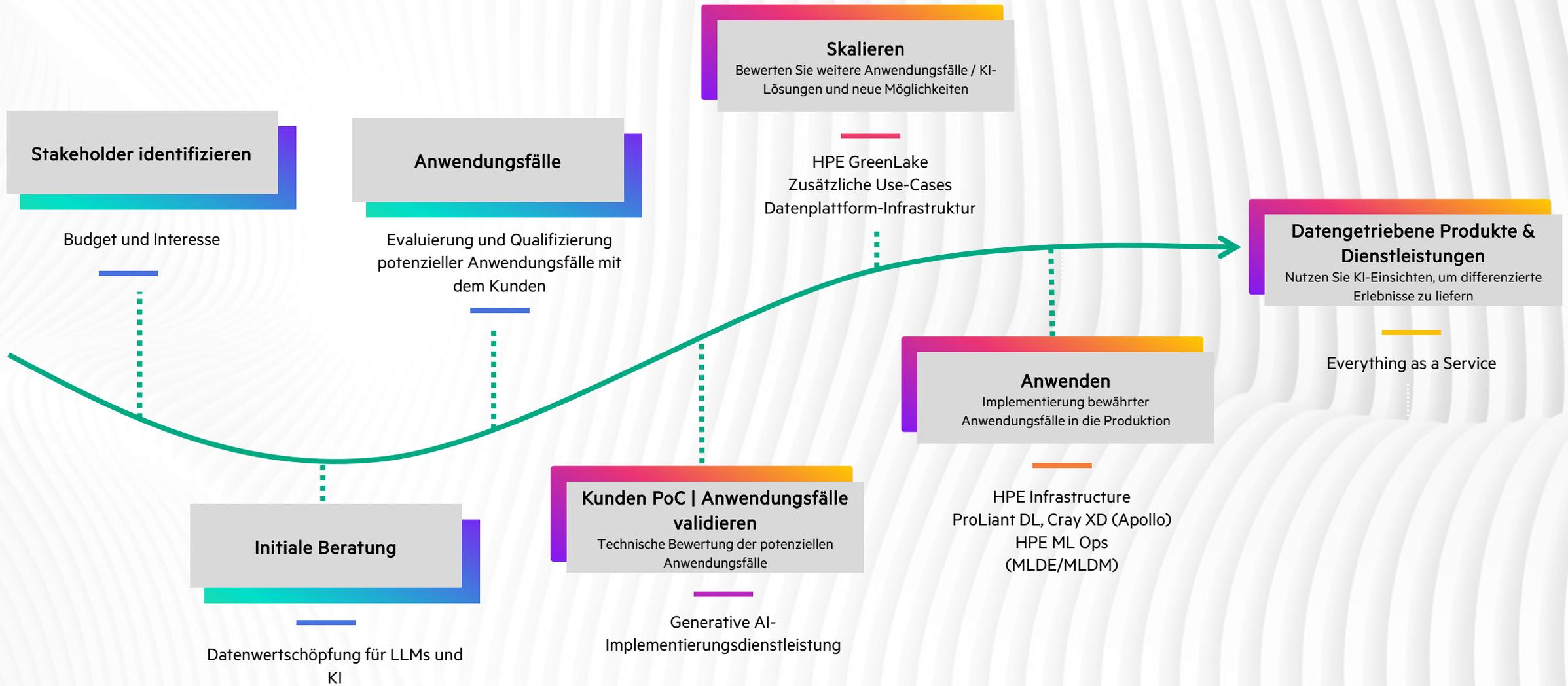


Menschen & Know How

Konzepte, Wissen, Implementierungsleistungen

Kunden, HPE, Partner!!!

KI-Adaptionsstufen



HPE Services für die KI Adaption

Phase	<i>Tag -1</i> Beraten	<i>Tag 0</i> Designen & Planen	<i>Tag 1</i> Implementieren & Integrieren	<i>Tag 2 bis Tag -1</i> Betreiben
Kundenergebnis	KI Strategie & Fähigkeiten	KI Daten & Architektur	KI Platform & Lösungen	KI Leistung & Einfluss
Servicepakete			<ul style="list-style-type: none"> Private Cloud AI Deployment 	<ul style="list-style-type: none"> Private Cloud AI Reactive Support / Incident Management
Services	<ul style="list-style-type: none"> Private Cloud AI Opportunities Conversation (Gen) AI Skills, Technologies & Best Practices Education (Gen) AI Adoption / Readiness Assessment (Gen) AI Business Case Development 	<ul style="list-style-type: none"> Private Cloud AI Full Stack (HW+SW) Technical Design Scalable & Secure Data Foundations for AI Technical Design Private Cloud AI TCO / Sustainability Calculator 	<ul style="list-style-type: none"> Private Cloud AI Integration with Existing IT Eco-System / Multi-Cloud Data Governance Service Security Integration for Private Cloud AI (Gen) AI Solution Tailoring 	<ul style="list-style-type: none"> Private Cloud (Gen) AI Solution Performance Optimization Private Cloud (Gen) AI Solution Compliance & Security Assurance (Gen) AI Model Management MLOps Automation for Private Cloud AI



Möglichkeit, alle Ihre Umgebungen mit KI zu automatisieren

Städte



- Verkehrsregelung
- Verwaltung am Straßenrand

Manufacturing



- Fehlererkennung
- Ertragsoptimierung

Retail stores



- Bestandsverwaltung
- Kassensfreies Einkaufen

Entertainment



- Stau
- Warteschlangenüberwachung

Industrial sites



- Sicherheit am Arbeitsplatz
- Vorausschauende Wartung

Campus



- Zugangskontrolle
- Forensische Suche

Lager



- Lageranalyse
- Paketsortierung

Flughäfen



- Turnaround-Management
- Fluganalyse

sensei

Changing the way the world shops

We make shopping simpler, more enjoyable and more efficient with autonomous store technology. Spend more time on what matters.

Lösung:
Data

Industrie:
IT services

Region:
Portugal

Ziele

- Einkaufserlebnis ohne Warteschlangen bieten
- Filialprodukte für die Bestandsverwaltung digitalisieren
- Umfassende Einblicke in das Verbraucherverhalten erhalten

Anforderungen

- Bereitstellen von Tools zur Entscheidungs- und Aufgabenunterstützung für Werkstatt-Teams
- Genauere Bestandsprognosen ermöglichen
- Sicherheit durch mehrere Aufnahmekameras erhöhen

Lösung

- HPE ProLiant DL380 servers
- NVIDIA® GPUs
- Intel® Xeon® Scalable processors

Schlüsselpartner

- NVIDIA

sensei

Ergebnisse

- Einkaufserlebnis der Kunden durch reibungslose Checkouts verbessern
- Warenkorbgenauigkeit von mehr als 99 % bei vollständiger Transparenz aller Kundenaktivitäten im Geschäft
- Gesteigerte Effizienz des Filialbetriebs durch Echtzeitüberwachung der Lagerbestände im Regal

Additional resources

- [Case study](#)
- [Sensei video](#)



Lösung:
HPE GreenLake

Industrie:
Retail

Region:
Austria

Ziele

- Stärkung der Wettbewerbsfähigkeit und Führung in den Einzelhandelsmärkten Mitteleuropas
- Außergewöhnliches Preis-Leistungs-Verhältnis und unvergleichliche Einkaufserlebnisse
- Beschleunigte Bereitstellung neuer digitaler Innovationen und prognostizieren zukünftiger Präferenzen und Anforderungen

Anforderungen

- Beschleunigte digitale Transformation des Unternehmens im Einzelhandel
- Verbesserte Agilität und Effizienz im IT-Betrieb
- Starkes IT-Betriebsmodell mit skalierbaren Rechen-, Speicher- und Backup-Systemen

Lösung

- HPE GreenLake
- HPE Managed Services
- HPE Complete Care Service
- Microsoft Workspace 365 Advisory and Professional Services
- HPE Synergy
- HPE ProLiant DL380 Servers
- HPE ProLiant DL560 Servers
- HPE Superdome Flex 280
- HPE Apollo
- HPE Primera Storage
- Commvault HyperScale X



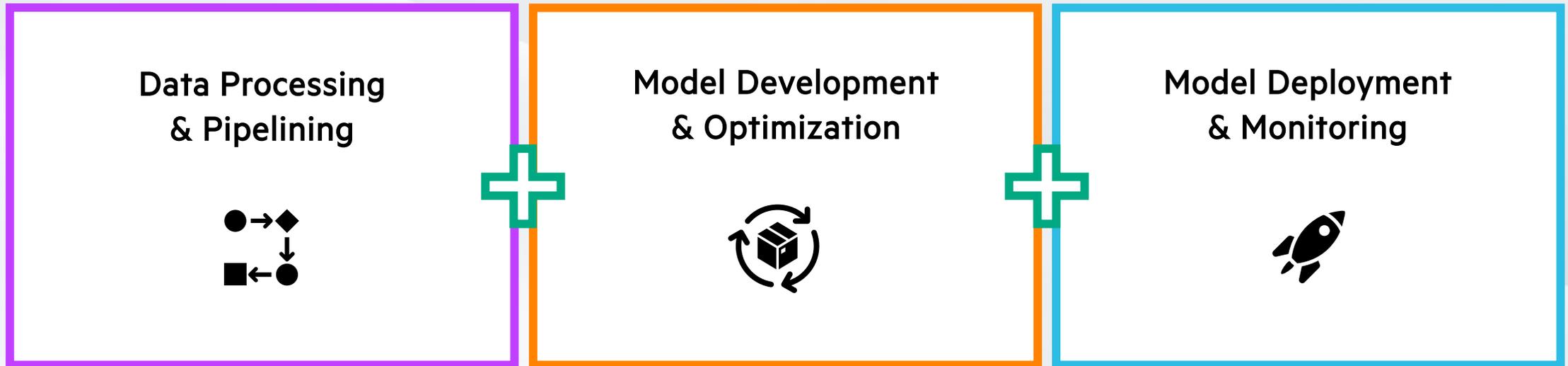
Ergebnisse

- Reduziert die Zeit für die Einführung neuer Einzelhandelslösungen von 6 Monaten auf 1 Woche
- Versteht die Kaufpräferenzen der Kunden und verbessert die Prognose der Produktnachfrage durch KI
- Reduziert Lebensmittelverschwendung durch die Optimierung der Lagerbestände im Geschäft und die Automatisierung von Lagerfunktionen

Additional resources

- [Case study](#)

HPE Machine Learning Platform



Enterprise-grade, End-to-End Solution

Best-of-Breed Components

Flexibility | Scalability | Reproducibility



HPE Machine Learning Platform

Hosted on Kubernetes



Machine Learning
Data Management



Machine Learning
Development Environment



Data Processing & Pipelining

- Any data source, format or type
- Incremental processing optimizes ingestions of massive amounts of data
- Flexible, automated data pipelines
- Immutable data versioning and end-to-end data lineage tracking



Model Development & Optimization

- Distributed training and Hyperparameter optimization
- Experiment tracking and collaboration
- Interactive Jupyter notebooks
- Advanced GPU resource management and monitoring



Model Deployment & Monitoring

- Model serving including shadow and canary rollouts
- Model performance monitoring and auditing
- Model observability including drift detection, explainability, and outlier detection

Distributed File & Object Store

 HPE Ezmeral Data Fabric | Cloud | On-Premises

GPU Accelerated Infrastructure

Delivered by  HPE GreenLake | Cloud | On-Premise



KI-Ankündigungen von HPE auf der Discover im Juni 2024

HPE Private Cloud AI

Branchenweit erste schlüsselfertige Full-Stack-Private Cloud für KI

Small (4 oder 8 L40S GPUs)

Medium (8 oder 16 L40S GPUs)

Large (16 oder 32 H100 NVL GPUs)

- Entwickelt, um **Inferenz, Feinabstimmung und Retrieval Augmented Generation (RAG)-Anwendungsfälle** zu unterstützen, die Daten in Hybrid-Cloud-Umgebungen nutzen.
- Vor Ort, in Rechenzentren, Edges oder Rechenzentren bereitgestellt

HPE ProLiant Compute DL384 Gen12

Enterprise-Funktionen und neueste NVIDIA GH200 NVL2 in 2U, 2P

Bis zu 2 GH200 NVL2

- Optimiert für **KI-Feinabstimmung & -Inferenz** mit RAG für **große Sprachmodelle**, die eine **große Speicherkapazität** erfordern
- **Höchste Leistung** pro mit NVIDIA Dual GH200 NVL2 mit 1,2 TB **schnellem, einheitlichem, kohärentem Speicher**
- **Doppelt so hohe Inferenzleistung***



HPE ProLiant Compute DL380a Gen12

Volles Unternehmenserlebnis und neueste PCIe-GPU in 4U, 2P

Bis zu 8 H200 NVL, B100 GPUs

- Der größere, schnellere Speicher beschleunigt die **Hochleistungs-KI-Feinabstimmung & -Inferenz** mit RAG für **große Sprachmodelle**
 - Für **groß angelegte Produktionsbereitstellungen** entwickelt



* Verglichen mit NVIDIA H100 accelerators.

A person's hands are holding a bright green rectangular frame in front of a server rack. The server rack has a dark, perforated metal mesh. The background is slightly blurred, showing a white light fixture on the right side.

Vielen Dank!

Jeremy Massow

Leoni Albers



AI use cases & vertical industry applications

Horizontal Use Cases

Virtual Assistants

- Chatbots/Q&A for Customer Service or Internal Support
- Code Generation and SW development
- Content Summarization

Automation

- Task and Workflow Automation
- Corporate AI Governance Processes
- AI-Driven Event and Incident Management

Content and Product Creation

- Content Creation for Sales and Marketing
- Internal Communication and Knowledge Management
- AI-driven Product Dev & Prototyping

Improved Productivity

Increased Production

Vertical Industry Applications

Financial Services

- Fraud Detection and Prevention
- Loan Application Review Automation
- Market Research for Investment Strategies

Healthcare and Life Sciences

- Personalized Treatments and Diagnosis
- Drug Discovery
- AI-Powered Diagnosis
- Virtual Assistants for Doctors

Public Sector

- Automated Public Safety
- Chatbots for Citizens Engagement
- Cybersecurity Threat Analysis

Retail

- Personalized Product Recommendations
- Dynamic Forecasting and Inventory Management
- Dynamic Pricing Strategy
- Theft Prevention

Telco

- Edge Computer Vision
- Customer Retention Support
- Network Anomaly Detection

HPE Private Cloud AI

Secure, compliant, sovereign data management

NVIDIA AI Computing by HPE

HPE Private Cloud AI

Instant AI productivity

Accelerate time to value. Self-serve access to essential AI tools speed developer productivity up to 90%.

Unify access to all your data

Allow access to data across the enterprise in a single global namespace via embedded data lakehouse.

Enterprise-grade confidence and control

Private cloud experience to manage, secure, and govern your data, models, and infrastructure.

Cloud experience that keeps data private

Deployed on-premises, designed for hybrid. Flexible and modular with cloud technologies, economics, scalability.

Inferencing

RAG

Model fine-tuning

(Retrieval-augmented generation)

Data Processing

1. Rohstoffe sammeln (Daten sammeln)
2. Qualitätsprüfung (Datenbereinigung)
3. Zerkleinern und Formen (Datenverarbeitung)
4. Zusammenbau (Datenaggregation)
5. Qualitätskontrolle (Datenvalidierung)
6. Verpackung und Versand (Datenoutput)

Private Cloud AI

- Alles was für AI nötig ist
- Small, Medium, Large & XL
- HW (Storage, Compute, Netzwerk)
- NVIDIA AI Enterprise
- Pachyderm, Determined AI
- NVIDIA Modell Libraray
- Hugging Face Library (Open Source Modelle)

Begriffe

Automatisierung

Statisch - Küchengeräte
& Rezepte

Künstliche Intelligenz

Nachahmen menschlichen verhaltens

Machine Learning

System Lernt, welche Menge an
Zutaten kommt gut an

Computer Vision

System Lebensmittelqualität
(Reife, Frische etc.)

GEN AI

wie verschiedene Zutaten auf
molekularer Ebene
interagieren

AI Software

NVIDIA AI Enterprise



HPE AI Essentials



NVIDIA AI Computing by HPE

HPE Private Cloud AI portfolio

Best for

Inferencing

**Inferencing
+ RAG**

**Inferencing + RAG
+ Fine-tuning**

**Inferencing + RAG
+ Fine-tuning**

**AI Software from
NVIDIA and HPE**

Unified experience through HPE GreenLake cloud

HPE Sign In

HPE Private Cloud AI

Tools & Frameworks

NVIDIA AI Enterprise | Data Engineering | Analytics | Data Science

Virtual Assistant
Question Answering Chatbot

In this solution, you build a question-answering system using an open-source Large Language Model (LLM). This system c...

kServe | MLflow | NVIDIA NIM

Deploy

Small

Medium

Large

Extra Large



Single rack

Multi-rack

**Compute
Storage*
Networking**
Power**

4 or 8 L40S GPUs
30 TB to 248 TB
100GbE
up to 8 kW rack

8 or 16 L40S GPUs
109 TB to 390 TB
200GbE
up to 17.7 kW

16 or 32 H100NVL GPUs
670 TB to 1088 TB
400GbE
up to 25 kW x 2

12 or 24 GH200 GPUs
670 TB to 1088 TB
800GbE
up to 25 kW x 2

* Storage Includes 10+ TB Data Fabric Lakehouse capacity in addition to File storage
** Networking based on NVIDIA Spectrum networking technology

Upgraded configurations shown