

CLOUDERA



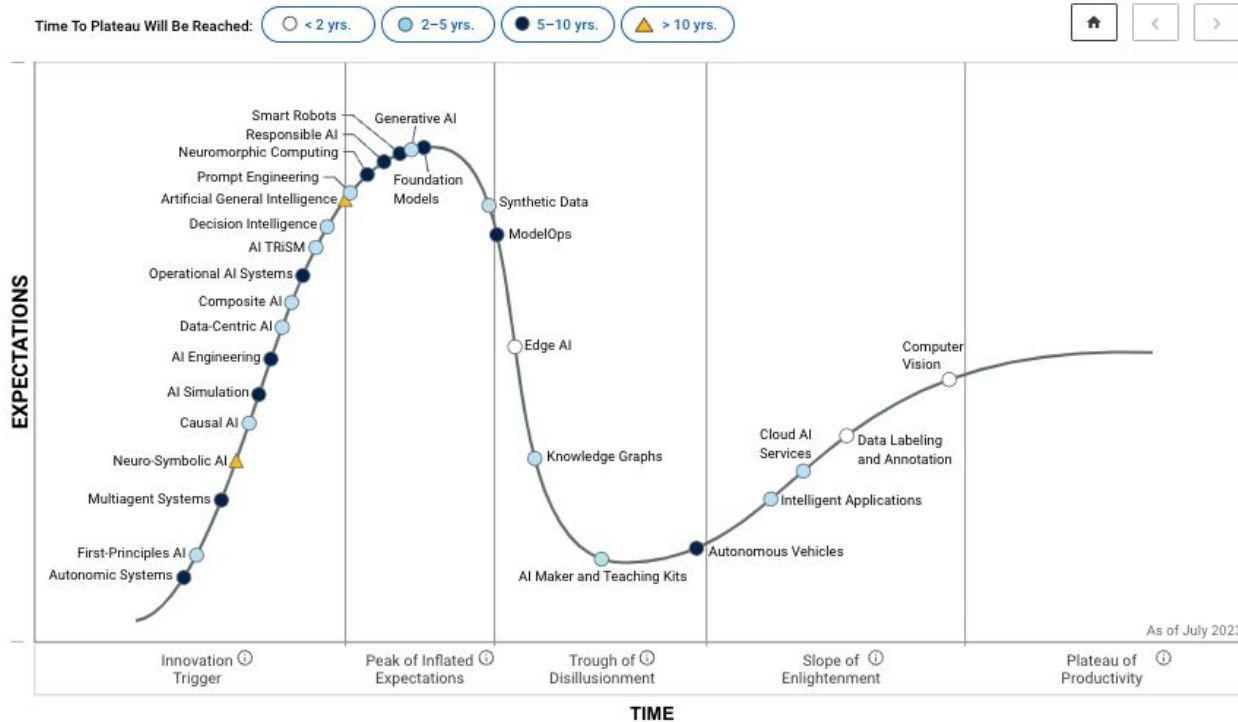
WIENER LINIEN

KI & DATENSTRATEGIE DER WIENER LINIEN

Dominik Michel | Account Executive, Cloudera

Bernhard Ortner | Lead Data Architect, WIENER LINIEN

GARTNER HYPE CYCLE FOR ARTIFICIAL INTELLIGENCE, 2023



[Afraz Jaffri, 19 July 2023, ID G00791179](#)

WHY ARE ORGANISATIONS EXPLORING GENERATIVE AI?

McKinsey and Company's View

“Generative AI’s impact on **productivity** could add trillions of dollars in value to the **global economy**. Our latest research estimates that generative AI could add the equivalent of **\$2.6 to 4.4 trillion annually** across the 63 use cases we analyzed.”

[The economic potential of generative AI](#), McKinsey & Company

AI OPPORTUNITY

Vertical use case categories



R&D



SALES &
MARKETING



CUSTOMER



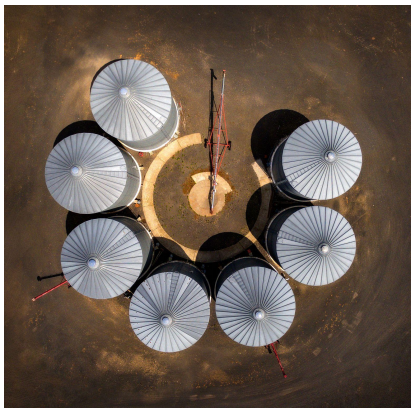
LEGAL



HR

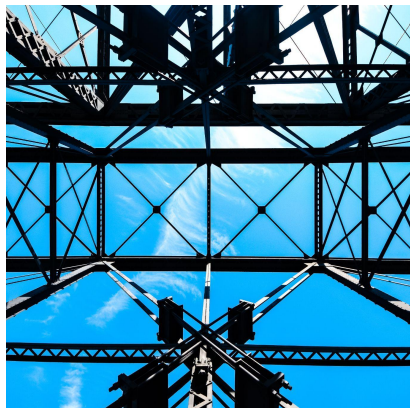
THE CHALLENGES AND RISKS OF ADOPTION

Public SaaS LLMs Presents Challenges, Risks and Implications for Enterprise Organisations



USE YOUR DATA

Trained on your data to provide critical context



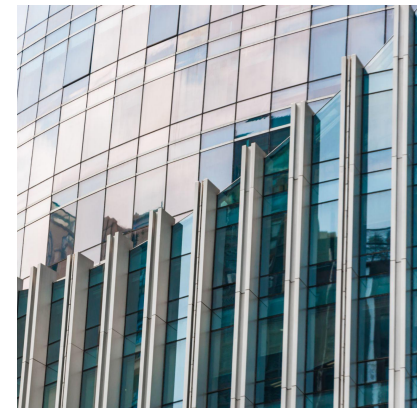
RELEVANCE & ACCURACY

Erroneous responses can have serious consequences



TRUST & SAFETY

Ethical and responsible AI together with explainability and transparency



RISK & COMPLIANCE

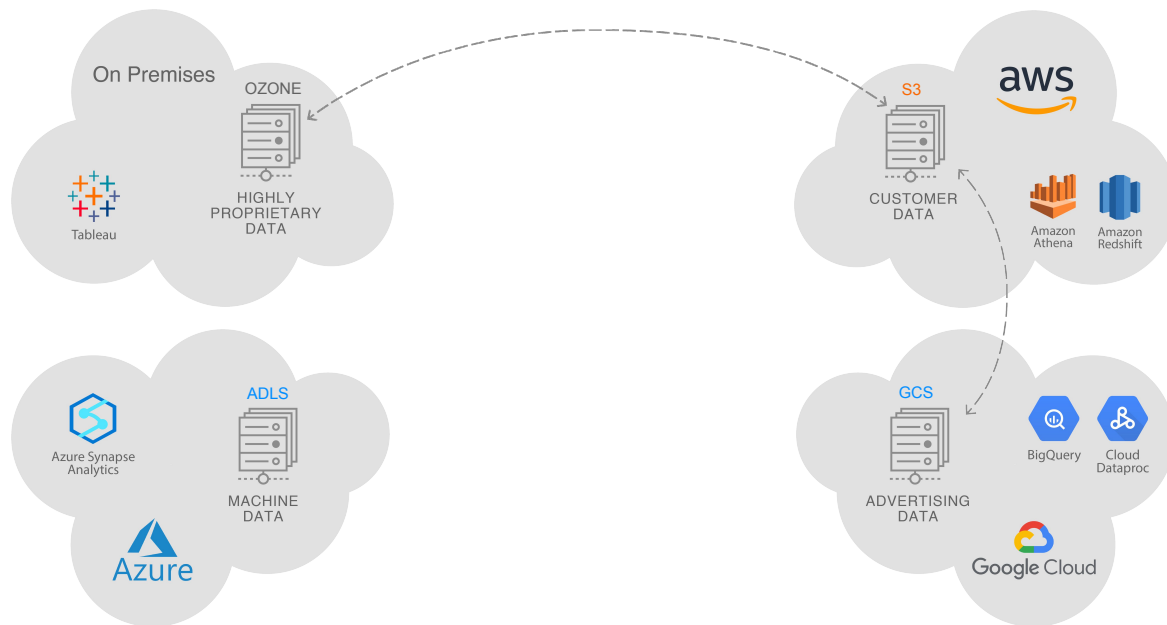
Authorisation, lineage, governance and auditability

ANALYTICS AND AI ARE ONLY AS
GOOD AS THE DATA THEY ARE
BASED ON

TASK 0 – DELIVER TRUSTED DATA

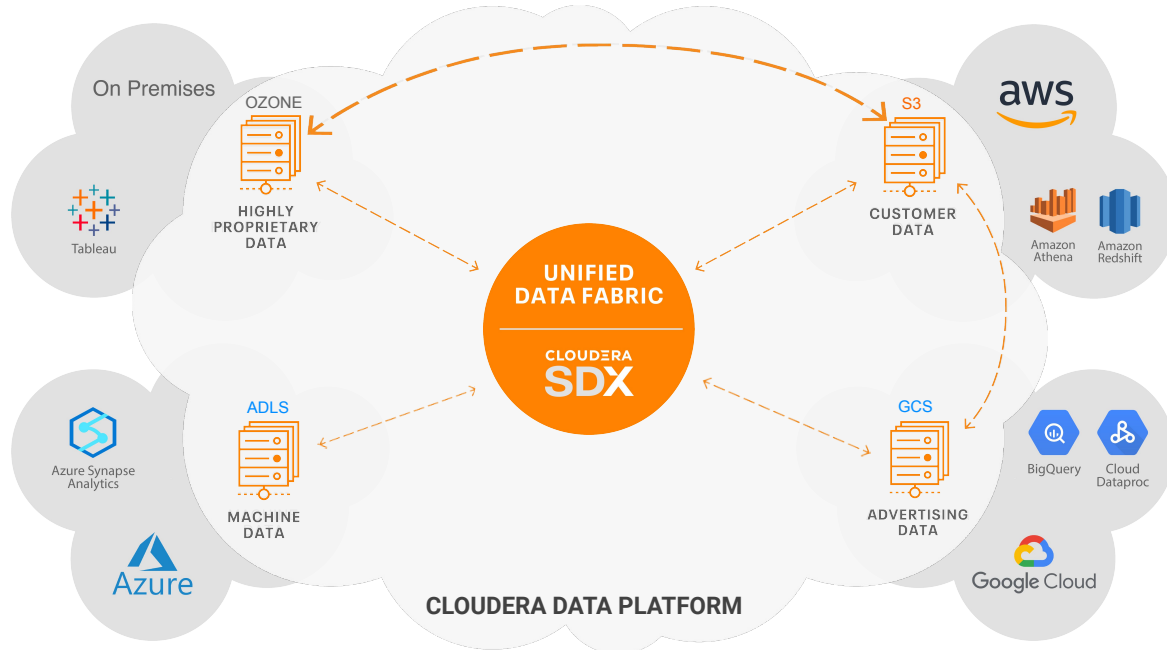
DISTRIBUTED DATA MANAGEMENT COMPONENTS

Which may run on multiple clouds and/or on-premises



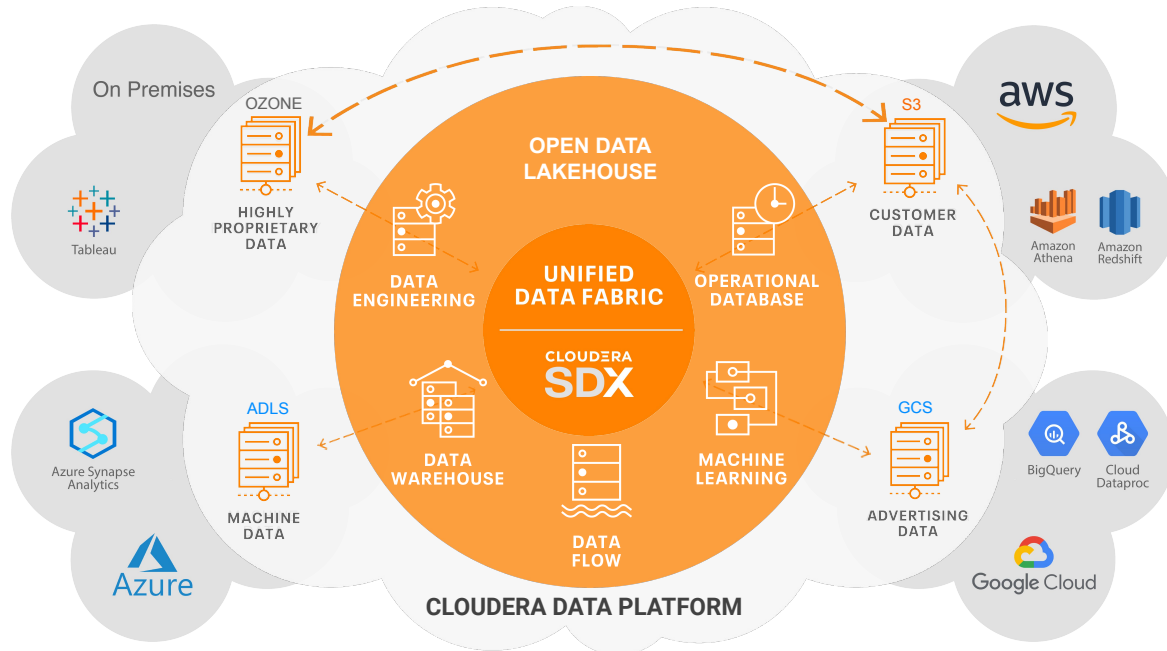
CLOUDERA'S HYBRID DATA PLATFORM

Portable. Interoperable. Open. Secure.



CLUDERA'S HYBRID DATA PLATFORM

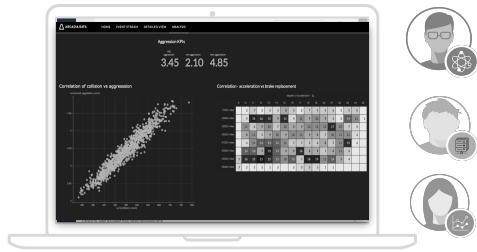
Portable. Interoperable. Open. Secure.



LEVERAGE TRUSTED DATA FOR AI

CLOUDERA MACHINE LEARNING (CML)

An Integrated Development and Deployment Platform for Trusted AI/ML



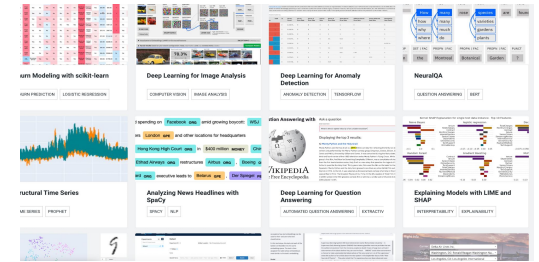
Data Science & Analytics

Collaborative and integrated analytic studio for **data analysts** to **data scientists**



Production ML

ML Platform with capabilities to deploy, monitor and govern **1000s of ML models**



AI Applications

Flexible application hosting platform to connect ML value to **business users**



BERNHARD ORTNER
LEAD DATA ARCHITECT
WIENER LINIEN

Wer bin ich?

Bernhard Ortner

- Data Architect Lead
- Teaching Big Data @ DHBW Stuttgart
- Habe mehrere Bücher geschrieben
- Focus Big Data, Legal, Security, Business



Assets



Weitere Assets

Störungsticker: <https://f59.at/stoerungen/>

Open Data: <https://www.wienerlinien.at/open-data> (data.gv.at)

WienMobil: <https://shorturl.at/grEY5>

API Portal der Wiener Stadtwerke:

<https://api-portal.wienerstadtwerke.at/portal/>

datenbezogen:

Mobilitäts- & Vorfalldaten

Kundendaten

Sensor & IoT-Daten

Q1

WARUM IST DATENMANAGEMENT
WICHTIG FÜR DIE WIENER LINIEN?

Nutzen für diverse Fachabteilungen

Effizienzsteigerung - Wir müssen schneller mit weniger Ressourcen einen Mehrwert generieren, der sich in (Zeit / Geld) niederschlägt.

Anforderungen der Use-Cases sind komplexer geworden

Gleichzeitig sind wir Betreiber wesentlicher Dienste – regulatorische Anforderungen steigen (DSGVO, NIS-2, AI-Act,....)

Erwartungshaltung: Es funktioniert einfach jederzeit und überall.

Q2

WARUM SIND VERTRAUENSWÜRDIGE
& QUALITATIV HOCHWERTIGE DATEN
WICHTIG?

Was bedeuten Daten (für WL)?

Daten sind ein Maßstab, wie gut etwas funktioniert. Je schneller die Daten zur Verfügung stehen, desto eher können sie einen Mehrwert erzielen.

Gleichzeitig werden die Daten „industrialisiert“, d.h. einem geordneten Prozess „Ende-zu-Ende“ unterworfen. Dieses Produkt wird wie jedes Produkt, permanent überwacht, verbessert und effizienter gemacht.

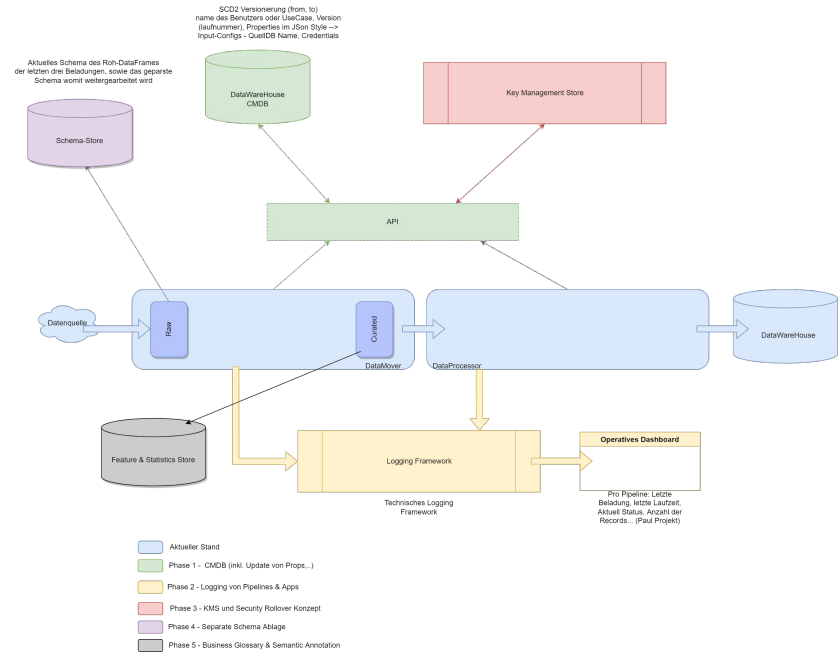
Fachdaten-Modellierung und Standardisierung ist grundlegend! Ebenso der Kontext (Domäne) der Datenanwendung **UND** deren Verwendung.

Wie wird mit Daten gearbeitet?

Die **Stabilität** der Datenverarbeitung von der tatsächlichen Umgebung ist essentiell **unabhängig**, idealerweise eingebettet in eine **standardisierte Arbeitsweise**

Wir benutzen nur „Werkzeuge“ die eine gewisse Verbreitung haben (Spark, Hive, ML-Flow) und unabhängig sind

Die Entwicklung erfolgt über mehrere Iterationen.



Q3

WIE WIRD DIE ROLLE VON KI BEI
DEN WIENER LINIEN GESEHEN?

Szenarien der Hybrid-Cloud

Szenario: Mögliche Anforderungen von mehreren Anwendungsfällen

- Temporäre oder Spezielle Ressourcen der Cloud nutzen
- Kontinuierliche Verarbeitungen von Daten onPrem
- Sensitive Daten dürfen das Unternehmen nicht verlassen
 - Cloudera kann hier helfen
- „andere“ System: Disaster Recovery, aber onPrem führt
- Etablieren von Standards
 - Trennung nach Kritikalität (atomarer Ansatz)

Cloud ist Wegbereiter für KI onPrem - Cloud ist die Entwicklungsumgebung bis zur Produktreife, die schnell bereit gestellt wird. Das Produkt wird dann allerdings im eigenen Rechenzentrum verwendet.

Szenarien der Hybrid-Cloud

KI ist jene Technologie, deren Auswirkungen noch nicht abschätzbar sind. Es wird erwartet, dass gewisse Szenarien zu >80% verbessert werden.

KI unterstützt dort wo „Ungenauigkeit“ kein Problem ist wie z.B. Support, Datengenerierung & Qualitätssicherung,...

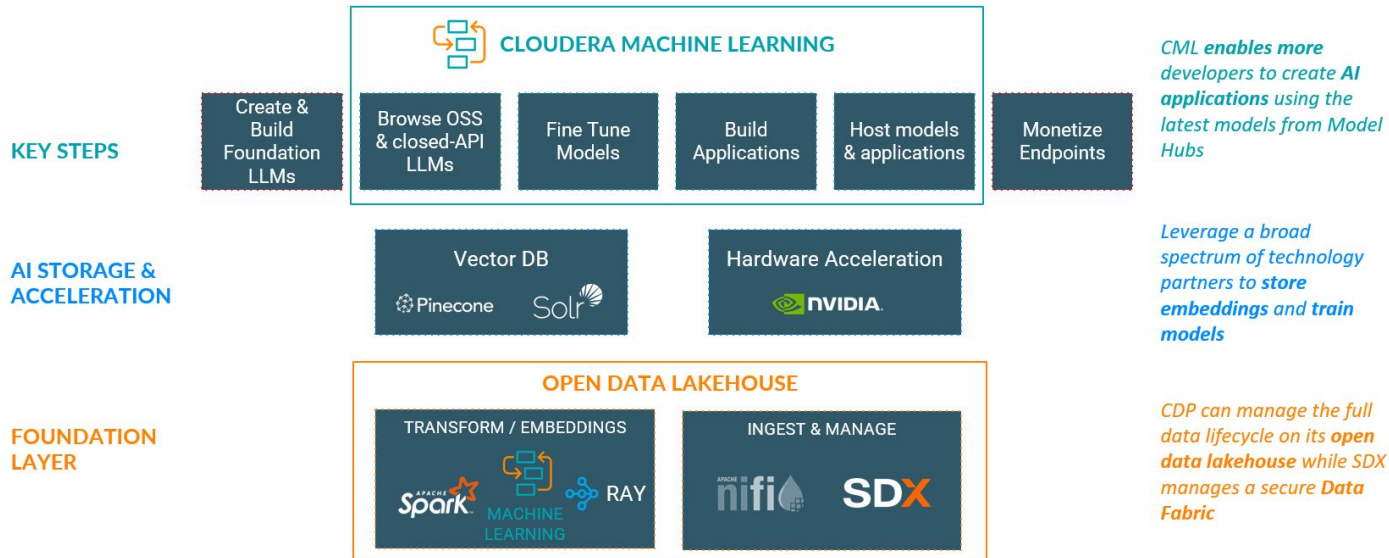
ABER:

1. Sich nicht damit zu beschäftigen, ist problematisch.
2. Es wird eine Konsolidierungsphase erwartet.

Wie lösen wir das Integrationsproblem ?

BUILD AI WITH CDP & CML

Enable the full Enterprise AI Application Lifecycle



Q4

WIE SEHEN KONKRETE
KI-ANWENDUNGSFÄLLE BEI DEN
WIENER LINIEN AUS?

Hybride Cloud der WIENER LINIEN

Open Data Lake-house, Evolutionsrichtung: Data Meshes

Sensors & IoT:

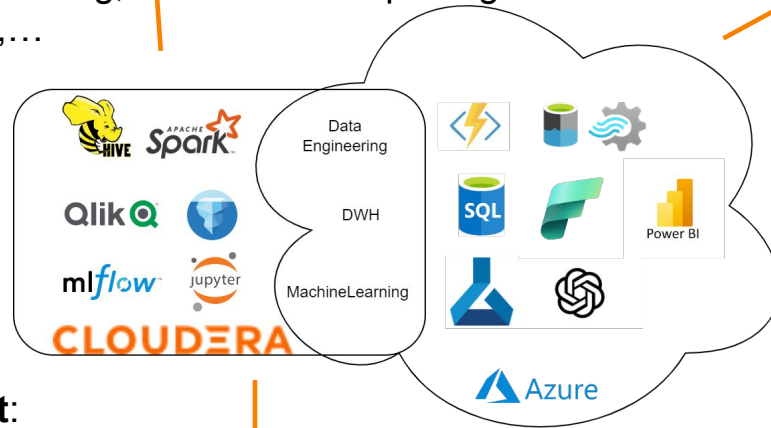
Predictive Maintenance,
Disposition & Route planning,
Verschleiß-Simulation,...

Finance & Controlling:

Materialmanagement, ESG, Risk,
Vendor- Reporting

Shared Mobility:

Mehrere & unterschiedliche
Quellen, diverse Consumer &
Producer, Geschwindigkeit
(Datenübertragung),
Dezentralität der Lösung,...



Prozessmanagement:

Multi-Sprachunterstützung, KI &
Entscheidungssupport,
„Predictive“ Prozesssteuerung

Customer360: Data-driven Experience, Pay per
Distanz/Zeit, „Cross & Upselling“ Möglichkeiten,
Passagierfluss Modellierung, Demand-Forecast

CLUDERA DATA PLATFORM

The only hybrid data platform for modern data architectures with data anywhere



Data Fabric



Data Lakehouse



Data Mesh



AWS, Azure,
Google Cloud



On-Premises



Data Engineering



Data Flow



Data Warehouse



Operational Database



Machine Learning



Open data fabrics,
lakehouses and
mesh with data
anywhere at scale

Multi-cloud &
on-premises data
management
and analytics

“Write once,
run anywhere”
cloud-native data
analytics portability

Unified security &
governance with
open cloud-native
storage formats

THANK YOU

CLOUDERA