

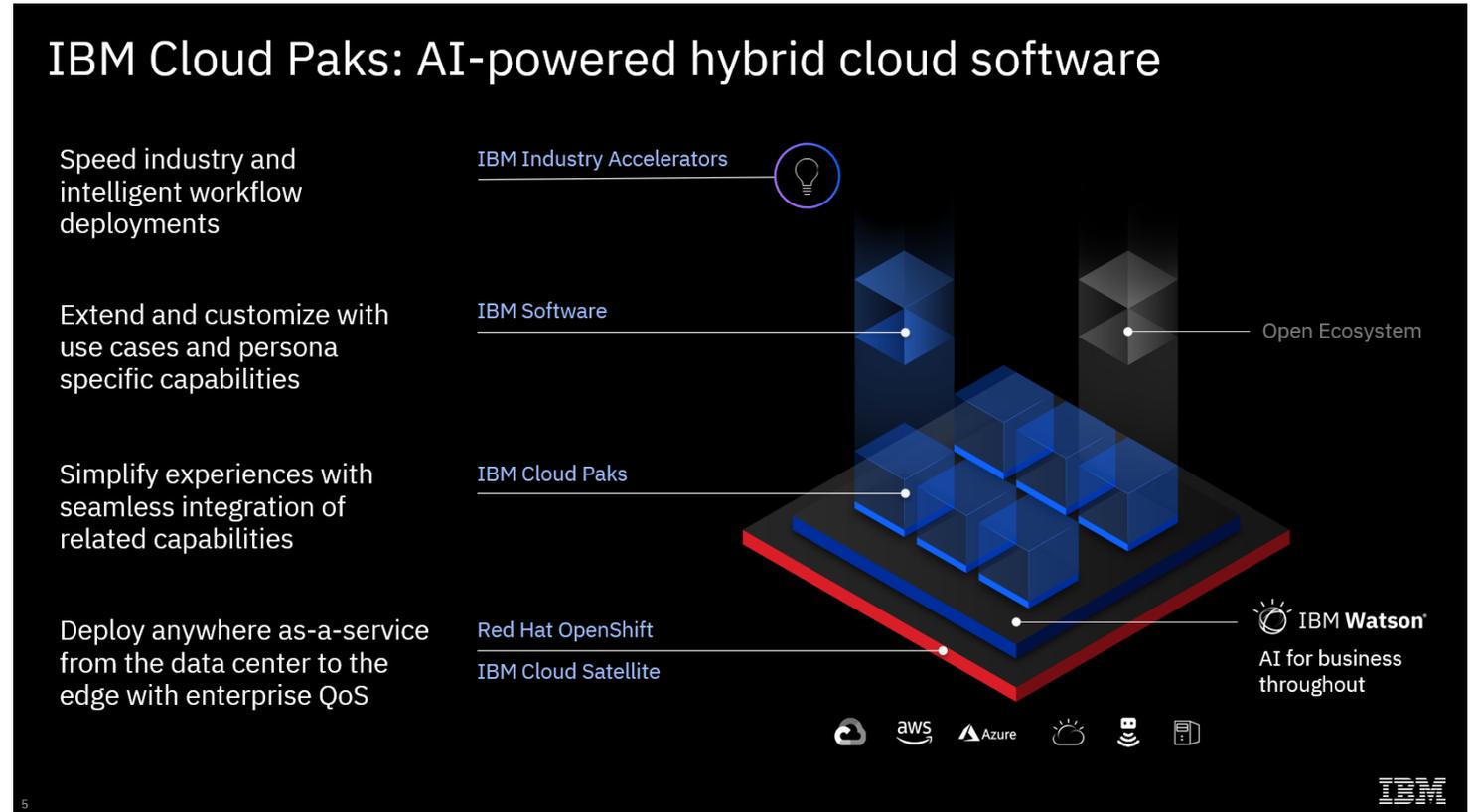
*Datengetriebene  
Prozesse und  
Geschäftsmodelle*

# IBM Cloud Pak for Data – strategische Einordnung Martin Clement



# IBM Architektur-, Plattform- und Produktstrategie

- Hybride Cloud Architekturen
  - On Premise oder auf IBM Cloud, Azure, AWS, Google etc. verfügbar
  - **Containertechnologie** auf Basis IBM Red Hat OpenShift
- IBM Software wird sukzessive „**Containerisiert**“
- Verschiedene **Cloud Paks** for
  - Data (& AI)
  - Automation
  - Watson AIOps
  - Integration
  - Network Automation
  - Security etc.



# IBM Cloud Pak for Data – Plattform für Data Fabric

## ➤ Collect (Sammeln)

- Zugriff auf Daten aller Art
- Speichern von Daten

## ➤ Organize (Organisieren)

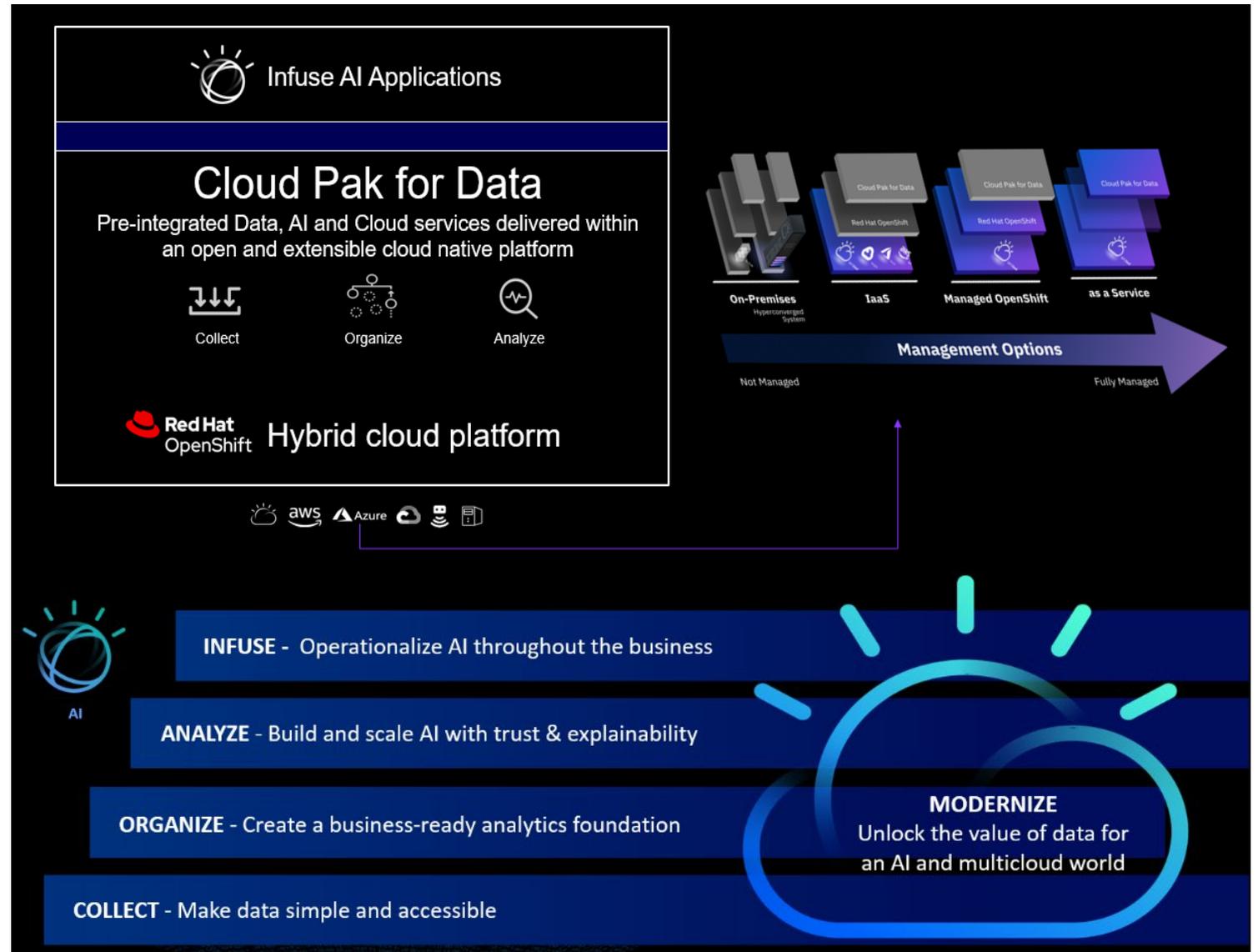
- Dokumentation von Daten
- Governance, fachliche & technische Metadaten
- Datenqualität

## ➤ Analyze (Analysieren)

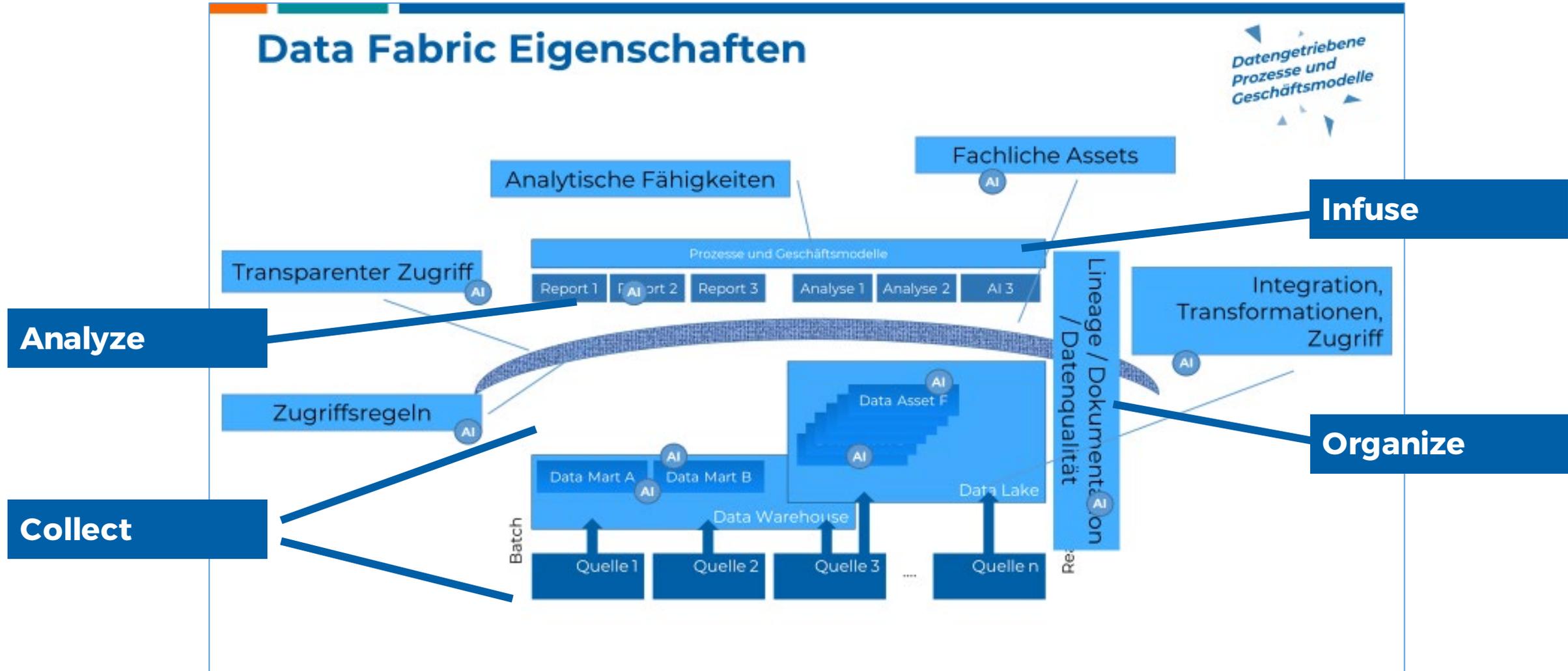
- Dashboard
- Data Science und AI
- Optimierung

## ➤ Infuse (Beschleunigen)

- Operationalisierung des kompletten Prozesses



# Recap zu Data Fabric



# Collect – Sammeln, Aufbereiten, zur Verfügung stellen

- Watson Query
  - Virtualisierung
- DataStage
  - ETL / Datenintegration
- Db2 & Db2 Warehouse
  - Relationale Datenbank
- NPS (Netezza)
  - DWH Appliance
- MongoDB
  - NoSQL Database
- Object Store



# Organize – Dokumentieren, Qualifizieren, Absichern

## ➤ Watson Knowledge Catalog

➤ Datenkatalog

➤ Metadata und Policy Management

## ➤ Data Profiling & Data Discovery

➤ Datenquellen analysieren & dokumentieren

## ➤ Data Quality

➤ Aktive (regelbasierte) Transformation

## ➤ Customer 360

➤ Master Data Management (MDM)

➤ Golden Record

## ➤ IBM Guardium

➤ Security

Dataset   Datenquelle   **Datenanforderungen**

Datenanforderungen ➤ Neue Datenanforderung

ID	Name	Status	Angefordert von	Zugeordnet zu	Priorität	Zuletzt aktualisiert
1	Get AQM datasource from PostgreSQL DB	Canceled	pggreoire	Unassigned		31 Mar 2020, 2:45 AM

Übersicht   **Spalten**   Quelle

### Neue Datenanforderung

Anforderungsname \*

bitte Daten! MBAS\_Proj

Link zu Problemen

Fügen Sie hier die URL eines bestehenden Problems ein Priorität

Beschreibung

Geben Sie eine Kurzbeschreibung der Anforderung ein

Datenzweck

Geben Sie eine Kurzbeschreibung der beabsichtigen Datenverwendung ein

Übersicht   **Spalten**   Quelle

### Erforderliche Spalten

➤ Spalte hinzufügen

	Name	Typ	Beschreibung
1	kunde_id	Int	Kundennummer
2	kunde_desc	String	Kundenname

# Analytics (AI) – Auswerten, Erkennen, Optimieren

## ➤ Watson Studio

### ➤ Data Science Workbench

#### ➤ Jupyter Notebooks

#### ➤ Python, Spark, GPUs etc.

#### ➤ Offener Technologiestack

### ➤ Model Designer

## ➤ Watson Discovery

### ➤ Unstrukturierte Informationen

## ➤ Watson Assistant (Chatbot), NLU, Speech to Text

## ➤ Decision Optimization

### ➤ Mathematische Optimierung (IBM Cplex)

The screenshot displays the IBM Watson Studio interface. The top navigation bar shows 'My Projects / watson-machine-learning / customer-churn-flow'. A search palette is visible on the left. The main workspace features a workflow diagram with nodes for 'Data As...', 'Type', 'Auto De...', 'Filter', 'Partition', 'churn', and 'Predict...'. Below the workflow, a Jupyter notebook is open, showing Python code for model selection and data loading. The code includes imports for sklearn, matplotlib, and pandas, and a function to load data from IBM Cloud Object Storage.

```
In [1]: from sklearn import model_selection
from sklearn import tree
from sklearn import svm
from sklearn import ensemble
from sklearn import neighbors
from sklearn import linear_model
from sklearn import metrics
from sklearn import preprocessing

In [2]: %matplotlib inline
from IPython.display import Image
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np
import os
import pandas as pd
import sklearn
import seaborn as sns
import json

Dataset
The original dataset can be downloaded from https://www.kaggle.com/becksgod/c churn-in-telecoms-dataset/data. Then upload it to IBM Watson Studio and insert the code to read the data using "Insert to code > Insert pandas DataFrame".

In [5]: # Hidden cell
import type
import pandas as pd
from hotoos.client import Config
import ibm_botoc

def _iter_(self): return 0

# Hidden cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove these credentials before you share the notebook.
client_h8h7na700w42f6d142bfff7219e = ibm_botoc.Client(service_name='s3',
ibm_api_key_id='yefb1kxpf9g1th30x8m1vfdj7o4649fc5qbl1ov9sck',
ibm_auth_endpoint='https://iam.cloud.ibm.com/oidc/token',
config=Config(signature_version='oauth'))
```

# Infuse – Beschleunigen, Operationalisieren

## ➤ Watson MLOps

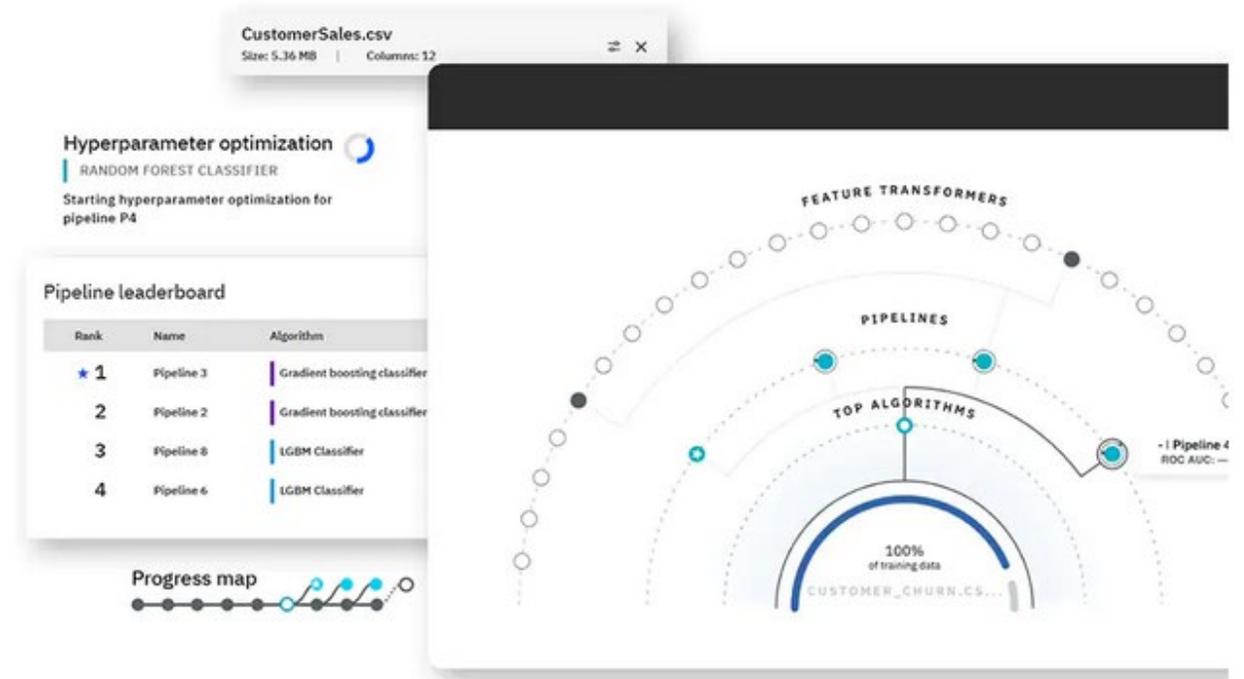
- AI Modelle deployen und operationalisieren

## ➤ Watson Auto AI

- Optimale Modelle ermitteln

## ➤ IBM Watson OpenScale

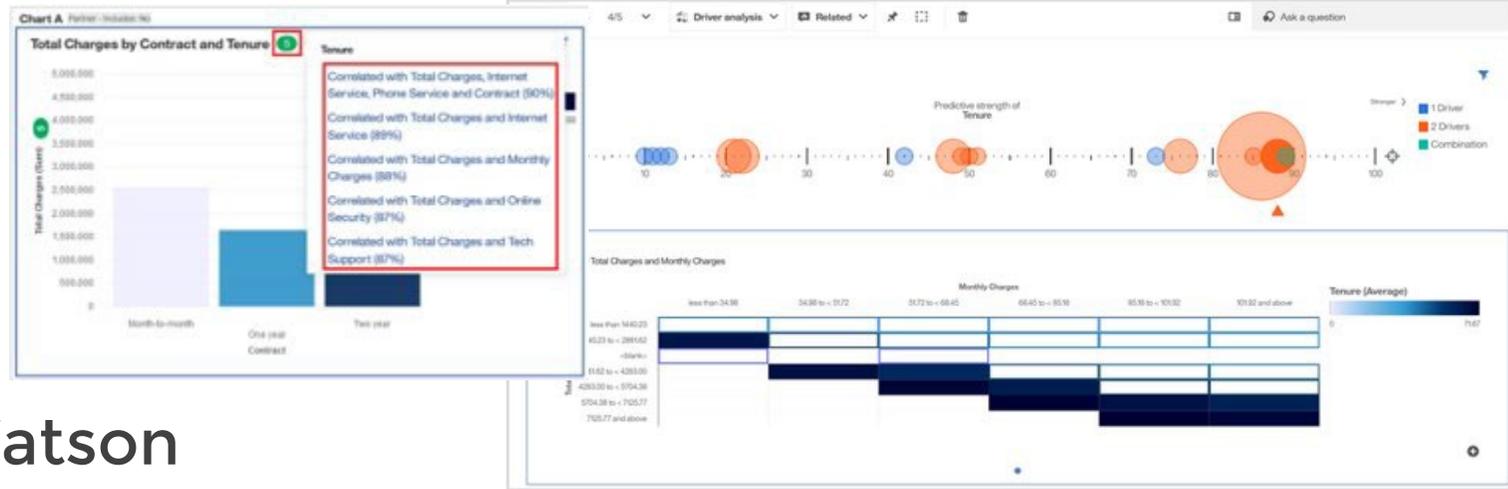
- Vertrauenswürdige KI
- Nachvollziehbarkeit der AI Algorithmen und Prozesse
- Anti-Bias und Privatsphäre



# Analytics (BI) – Auswerten, Planen, Steuern

## ➤ Cognos Analytics with Watson

- Reporting
- Dashboards
- Self Service BI
- Augmented Intelligence



## ➤ Planning Analytics with Watson

- Integrierte Unternehmensplanung
- Multidimensionale Modellierung
- Predictive Planning
- What if, Scenario Analyse, Sandboxing



# Lizenzoptionen

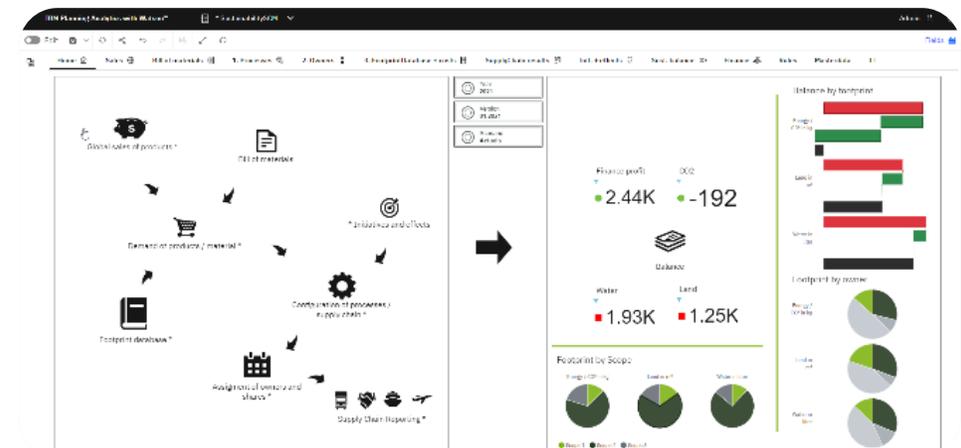
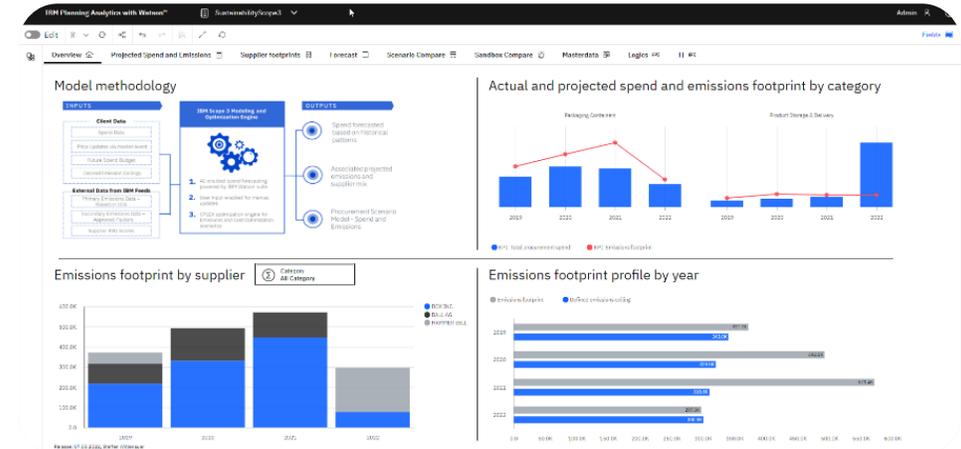
*Datengetriebene  
Prozesse und  
Geschäftsmodelle*

- **Modernization Lizenzen**
  - Upgrade bestehender Lizenzen
  - Definierte Migrationspfade
  - Dual Entitlement
- **Cloud Pak for Data Base & zusätzliche Cartridges**
  - Basisfunktionen in CP4D Base
  - Zusatzfunktionalität in Cartridges
- **Primär Subscription Lizenzen**
  - Software „Mietmodell“

# Anwendungsbeispiel Sustainability

Datengetriebene  
Prozesse und  
Geschäftsmodelle

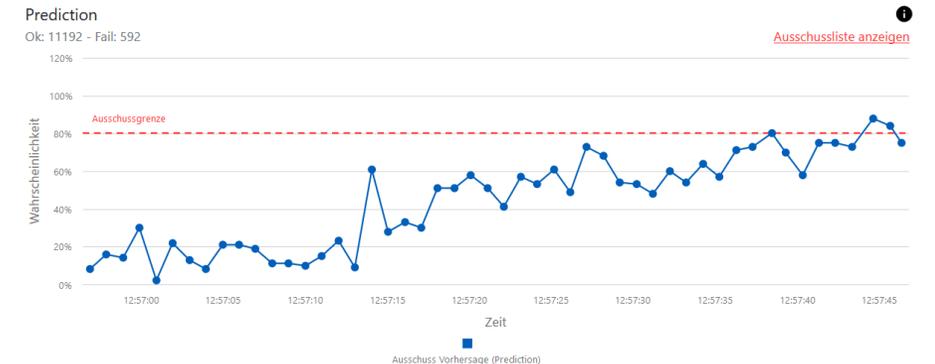
- Modellierung der Nachhaltigkeit
  - Sammlung aller relevanten Daten
  - Reporting, Analyse & Planung
  - Optimierung unter verschiedenen Gesichtspunkten
- Beschaffung / Lieferketten
  - Optimierung der Bestellvorgänge unter Berücksichtigung von CO<sub>2</sub>-Footprint und -Kosten
  - Tourenoptimierung, Transportwege & -zeiten
- Produktion
  - Berechnung und Optimierung des CO<sub>2</sub> Verbrauchs aller Produkte auf unterschiedlichen Fertigungslinien



# Anwendungsbeispiel Produktion

Datengetriebene  
Prozesse und  
Geschäftsmodelle

- Visualisierung der Produktions- und Betriebsdaten
  - Extraktion und Speicherung relevanter Daten aus Maschinen, Sensoren etc.
  - Visualisierung - OEE Dashboard
- Erkennung von Anomalien
  - Predictive Quality / Maintenance
  - verborgene Einflüsse erkennen, um die Produktionsprozesse zu verbessern
- Optimierung von Prozessen
  - Bestände und Auslastung nachfrageabhängig optimieren
  - mathematische Optimierung unter Berücksichtigung aller Einflussfaktoren





## Kundenvortrag:

# Modernisierung der Analytics Infrastruktur bei AGAPLESION auf Basis IBM Cloud Pak for Data System

*Jan Geesen, Teamleiter Business Analytics bei AGAPLESION*