

PRAXISBERICHT

MLOPS FÜR CONDITION-MONITORING ANWENDUNGEN

PRESENTATION

AGENDA

- Wer sind wir und was machen wir?
- Kontext-Einordnung: Condition Monitoring bei der KSB
- Pain Points: Der Schrei nach MLOps
- Unsere Lösung
 - Architektur-Sicht
 - Organisations-Sicht
 - Technische Details
- Fazit

PRESENTATION

AGENDA

- **Wer sind wir und was machen wir?**
- Kontext-Einordnung: Condition Monitoring bei der KSB
- Pain Points: Der Schrei nach MLOps
- Unsere Lösung
 - Architektur-Sicht
 - Organisations-Sicht
 - Technische Details
- Fazit

WER SIND WIR **UND WAS MACHEN WIR?**

esentri AG



STRATEGY & BUSINESS

Twin
Transformation

Business
Agility

Business
Innovation



DIGITAL INNOVATION

Legacy
Modernization

Connectivity
& Integration

Digital Products
& Applications

Cloud
Migration

Low Code /
No Code



DATA & AI

Data
Platforms

Analytics &
Reporting

Industrial
Analytics & IoT

Artificial Intelligence /
Generative AI



SUSTAINABILITY

Sustainable
Strategies

AI Driven CSR
Reporting

Data-Led
Innovation

WER SIND WIR **UND WAS MACHEN WIR?**

Industrial Analytics & IoT



01

CONNECTIVITY & EDGE

Cloudanbindung von IoT Geräten & Edge Deployments und deren Steuerung aus der Cloud heraus

02

DATA PLATTFORM

Aufbau von Datenplattformen von rohen Sensorwerten bis zu hoch aggregierten KPIs

03

KI SERVICES

Umsetzung von Machine Learning Services mit den Fokusthemen Vibration/ Zeitreihen & Visual Inspection

04

ML OPS

Sicherer Betrieb von ML basierten Anwendungen sowie Sicherstellung von Wartbarkeit, Erweiterbarkeit und Sicherheit

VERTIEFENDE QUELLEN

Konzeptionelle Sicht: [Der Weg vom Sensor zum Digitalen Produkt](#)

Praxisbeispiel: [coffAI - Klassifizierung von Vibrationsdaten](#)

PRESENTATION

AGENDA

- Wer sind wir und was machen wir?
- **Kontext-Einordnung: Condition Monitoring bei der KSB**
- Pain Points: Der Schrei nach MLOps
- Unsere Lösung
 - Architektur-Sicht
 - Organisations-Sicht
 - Technische Details
- Fazit

KONTEXT-EINORDNUNG

Condition Monitoring bei der KSB



- Plug & Play Überwachungslösung für Pumpen (<https://www.ksb.com/de-ch/guard>).
- Neben Realtime Dashboards und digitaler Asset Verwaltung werden auch Machine Learning basierte Services wie Condition Monitoring und Anomaly Detection angeboten.
- Diese intelligenten Services gilt es effizient in Betrieb zu bringen und zu halten und gleichzeitig kontinuierlich weiterzuentwickeln.

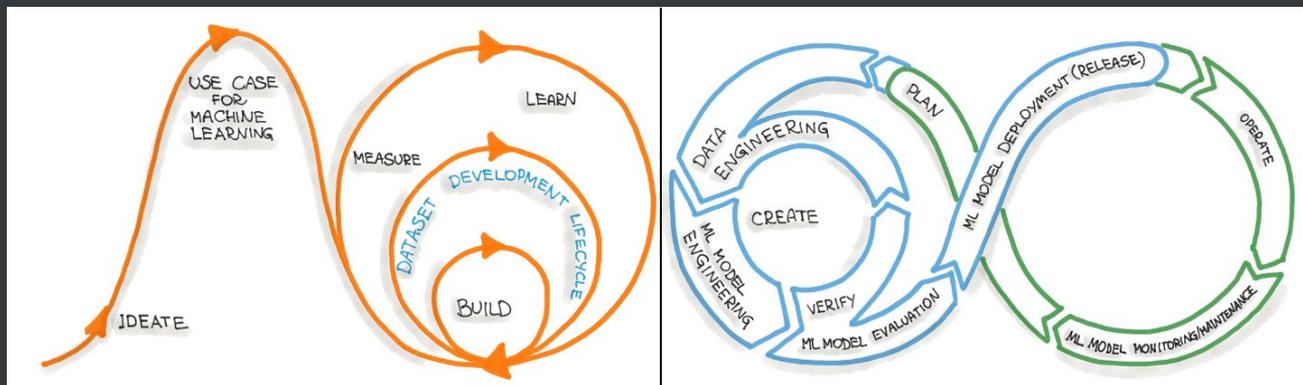
PRESENTATION

AGENDA

- Wer sind wir und was machen wir?
- Kontext-Einordnung: Condition Monitoring bei der KSB
- **Pain Points: Der Schrei nach MLOps**
- Unsere Lösung
 - Architektur-Sicht
 - Organisations-Sicht
 - Technische Details
- Fazit

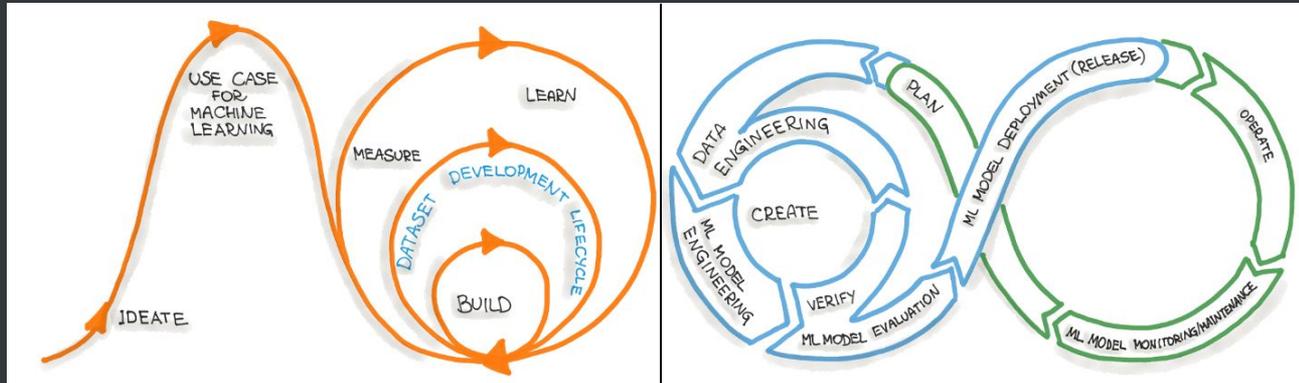
PAIN-POINTS

Der Schrei nach MLOps



PAIN-POINTS

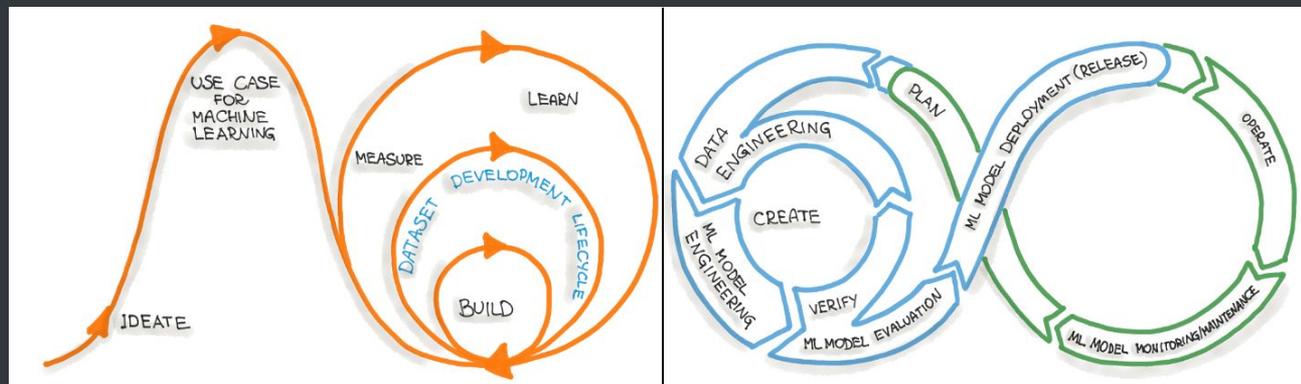
Der Schrei nach MLOps



Clean Architecture, IaC,
CI/CD, Elastic Container
Service....

PAIN-POINTS

Der Schrei nach MLOps



Clean Architecture, IaC,
CI/CD, Elastic Container
Service....

- DS: Wir müssen unsere Modelle updaten!
Eure Repo Struktur ist völlig undurchsichtig,
das fasse ich nicht an.
- DEV: Wie wild ist denn bitte diese library
Struktur, wie wende ich das überhaupt an?
Mit der Logik will ich nichts zu tun haben!
- DS: Wo liegen denn jetzt überhaupt die
Modelle? Wir müssen die kontinuierlich
überwachen und brauchen sie in unserem
Account!
- DEV: So wir haben alle libraries hochgezogen!
Warum lassen sich denn jetzt plötzlich die
Modelle nicht mehr lesen?
- ...

PRESENTATION

AGENDA

- Wer sind wir und was machen wir?
- Kontext-Einordnung: Condition Monitoring bei der KSB
- Pain Points: Der Schrei nach MLOps
- **Unsere Lösung**
 - Architektur-Sicht
 - Organisations-Sicht
 - Technische Details
- Fazit

UNSERE LÖSUNG

Architektur- und Organisationsicht

✓ KLARE VERANTWORTLICHKEITEN

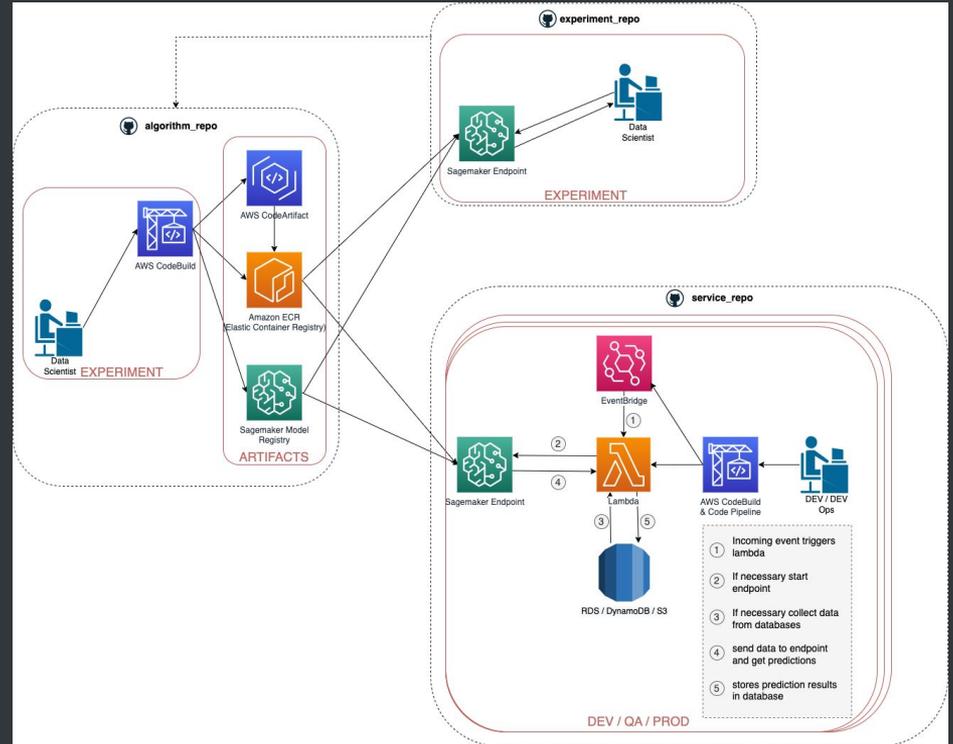
Das Data Science Team stellt sicher, dass die Sagemaker Endpunkte zu jeder Zeit nutzbar sind. Es kümmert sich nicht um Scaling, veränderte Inputs, etc.

✓ MODELL-ÜBERWACHUNG

Gerade wenn ein Service live ist, kann das Data Science Team über die Model registry die Modellnutzung, Model Drifts, Data Drifts etc. überwachen.

✓ MODELL-WEITERENTWICKLUNG

Da die Hoheit über die Modelle und deren Interaktion zu keinem Zeitpunkt abgegeben wird, sind Weiterentwicklungen problemlos möglich. Aus DEV Sicht muss dann nur der neue Endpunkt verwendet werden



UNSERE LÖSUNG

Architektur- und Organisationsicht



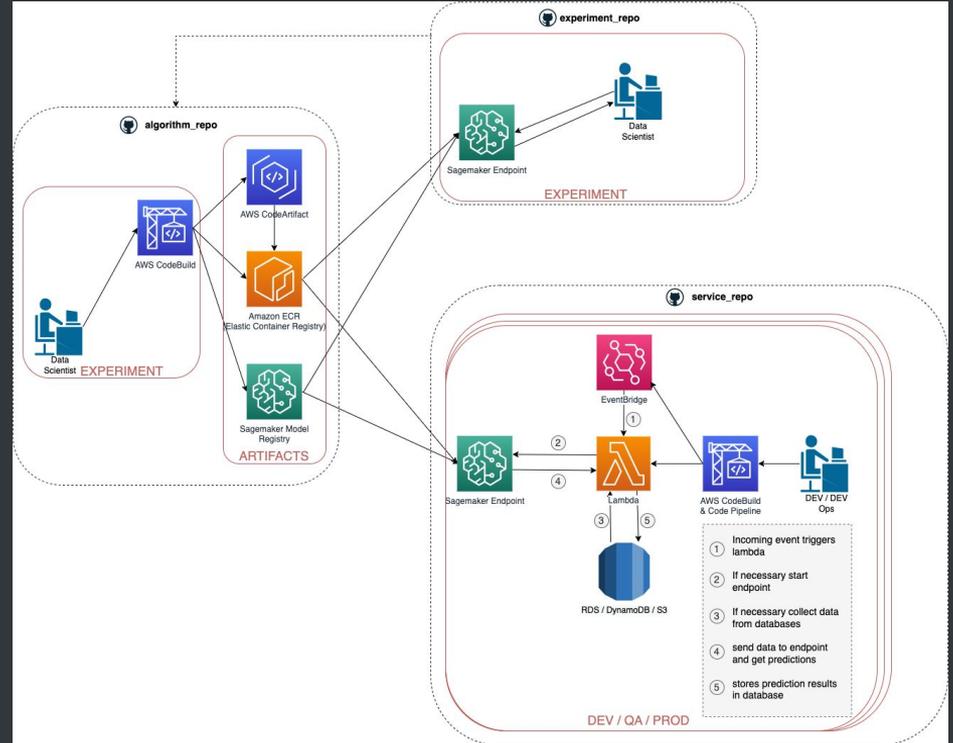
KOMPETENZERWEITERUNG NOTWENDIG

Das Data Science Team muss sich mit Themen wie CI/CD, Containerisierung, Endpunktversionierung etc. auseinandersetzen.



ERHÖHTE INFRASTRUKTURKOSTEN

Die lose Kopplung führt zu einer Erhöhung der Infrastrukturkosten. Nach unserer Erfahrung überwiegt der Effizienzgewinn um eine Vielfaches.



UNSERE LÖSUNG

Technische Details: Sagemaker Model Registry



Was ist eine Model Registry?

- Plattform & (G)UI, um Machine Learning Modelle zu
 - verwalten (Versionierung) & organisieren sowie deren Performance zu überwachen
- Schnittstelle zwischen Data Science & DEV



Deep Dive: Sagemaker Model Registry Spezifika

- Hierarchische Struktur unterteilt in Model Collections, Model Groups und Versionen
- Modelle sind (z.B. als tar-file) in S3 abgelegt
- Model Registry enthält ModelPackage = Metadaten zu Modell, u.a.:
 - Ressourcen-Identifizier (ARN) des Modells
 - Versionsnummer
 - Approval Status
 - Infos zur Inferenz
 - Metriken
 - Custom(er) Metadata
- Cross Account Access auf Model Registry



Model registry Show introduction

Model Groups Collections

Search...

Name	Collection	Description	Status	Created on	Created by
			Completed	2 months ago	arn:aws:sts:...
			Completed	2 months ago	arn:aws:sts:...
			Completed	2 months ago	arn:aws:sts:...
			Completed	2 months ago	arn:aws:sts:...
			Completed	5 months ago	arn:aws:sts:...

Model Group

Remove Create model version

OVERVIEW

Versions 74

BETAILS

Information

Search...

Overview

Versions 74

Version	Status	Description	Updated by	Updated on
Version 74	Approved			2 months a...
Version 73	Approved			--
Version 72	Rejected			--
Version 71	Approved			--

UNSERE LÖSUNG

Technische Details: Sagemaker Endpoints



Was ist ein Sagemaker Endpoint?

- Abstrakte Instanz, auf der ein oder mehrere ML Modelle laufen
- kann über API aufgerufen werden
- Rechenleistung kann individuell angepasst und bei Bedarf (automatisch) hochskaliert werden



Deep Dive: Was brauche ich zum Starten eines Endpunktes?

- Spezifikation: Singlemodel oder Multimodel Endpoint
- (Docker)Image aus der ECR (public oder custom)
- Wahl der Instanz (Rechenleistung)
 - Möglichkeiten für automatisches Scaling



Deep Dive: Welche Art von Endpoint brauche ich?

Single Model Endpoint

→ fixes Modell

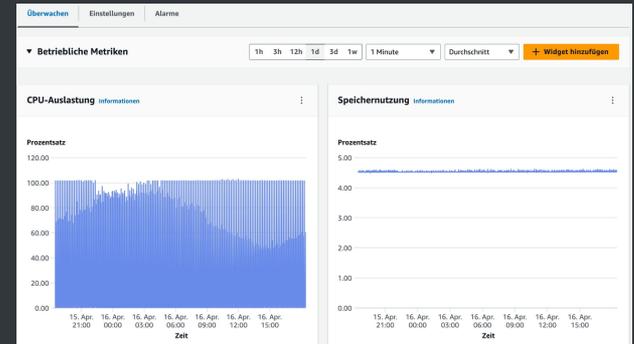
Multimodel Endpoint:

- verschiedene Modelle
- Modell muss bei Aufruf spezifiziert werden

The screenshot shows the 'Endpunktzusammenfassung' (Endpoint Summary) page in the Amazon SageMaker console. It displays the following information:

- Name:** [Redacted]
- Status:** InService
- Typ:** Endpoint
- ARN:** arn:aws:sagemaker:eu-central-1:123456789012:endpoint/production/endpoint-1
- Erstellungzeitpunkt:** Thu Mar 28 2024 13:56:41 GMT+0100 (Mittleuropäische Normalzeit)
- Zuletzt aktualisiert:** Thu Mar 28 2024 13:59:07 GMT+0100 (Mittleuropäische Normalzeit)
- URL:** https://runtime.sagemaker.eu-central-1.amazonaws.com/production/endpoint-1
- Containerprotokolle modellieren:** [Redacted]
- Alarme:** 0 alarme

There is a link for 'Weitere Informationen zur API'.



PRESENTATION

AGENDA

- Wer sind wir und was machen wir?
- Kontext-Einordnung: Condition Monitoring bei der KSB
- Pain Points: Der Schrei nach MLOps
- Unsere Lösung
 - Architektur-Sicht
 - Organisations-Sicht
 - Technische Details
- **Fazit**

FAZIT

Was solltet ihr aus diesem Vortrag mitnehmen?



wachsender **Weiterentwicklungsbedarf** von ML Modellen und Trennung von DEV und Data Science

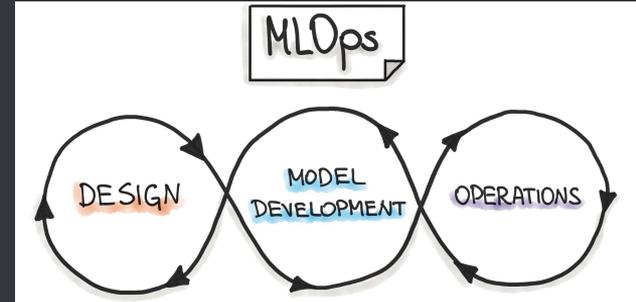
- Einführung von **MLOps Vorgehensmodell**
 - ◆ verringert Reibung
 - ◆ bringt Geschwindigkeit und **Effizienz**



Abgrenzung sowie Schaffen von **Schnittstellen** zwischen DEV und Data Science

- Model Registry
- Endpunkte

AWS Sagemaker ist zuweilen unintuitiv (**“der Teufel steckt im Detail”**), bietet aber gute Funktionalität und Skalierbarkeit



<https://ml-ops.org/content/ml-ops-principles>

IHRE KÖPFE FÜR MLOPS



SIMON KNELLER
Head of Industrial
Analytics & IoT



+49 160 967 648 04



simon.kneller@esentri.com



DR. JOHANNES EHLERT
Senior Consultant
Data & AI



+49 151 704 281 77



johannes.ehlert@esentri.com

