Einführung in MLOps

15 FTI Pipeline Architektur

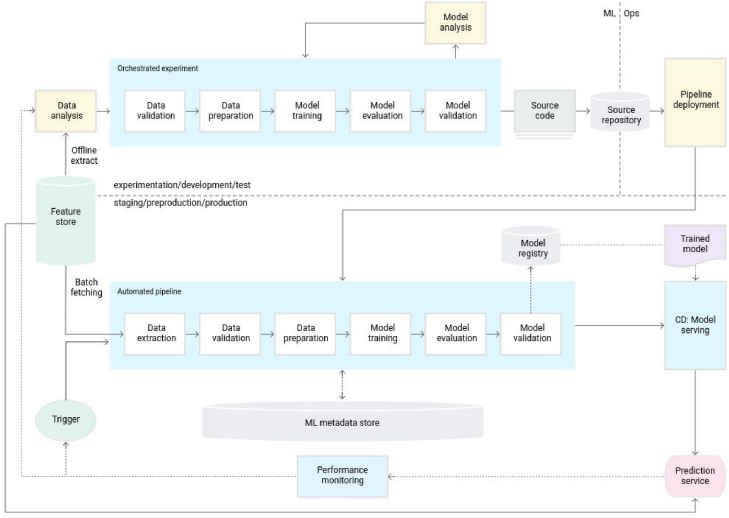
Tobias Mérinat
teaching2025@fsck.ch

Lucerne University of
Applied Sciences and Arts
HOCHSCHULE
LUZERN

Department of Information Technology
Lucerne University of Applied Sciences and Arts
6343 Rotkreuz, Switzerland

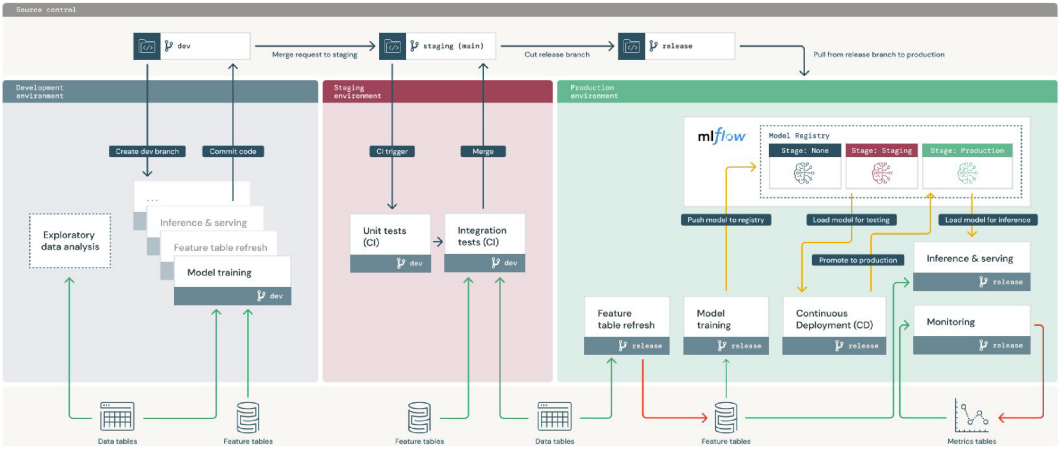
14. und 15. Februar 2025

Architectur Overview 1/2



Quelle: Google

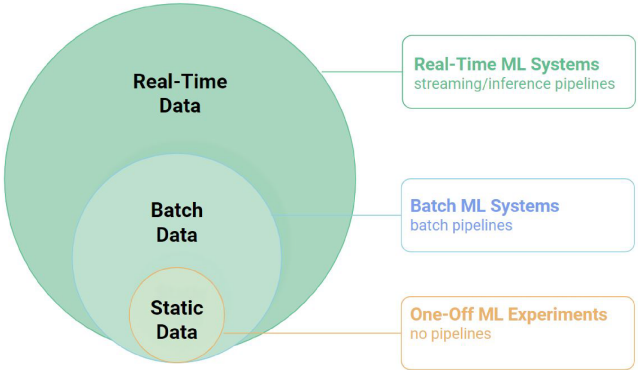
Architectur Overview 2/2



Quelle: Databricks

Experiment vs. System

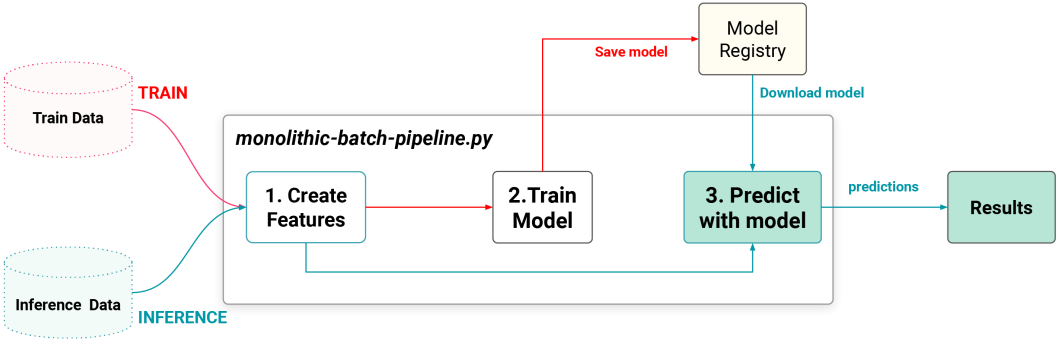
- Notebook und ein trainiertes Modell =! ML System
- Erst, wenn mit neuen Daten predicted wird
- Für ein System benötigen wir Pipelines



Quelle: Hopsworks

Erster Versuch: Monolithische Pipeline

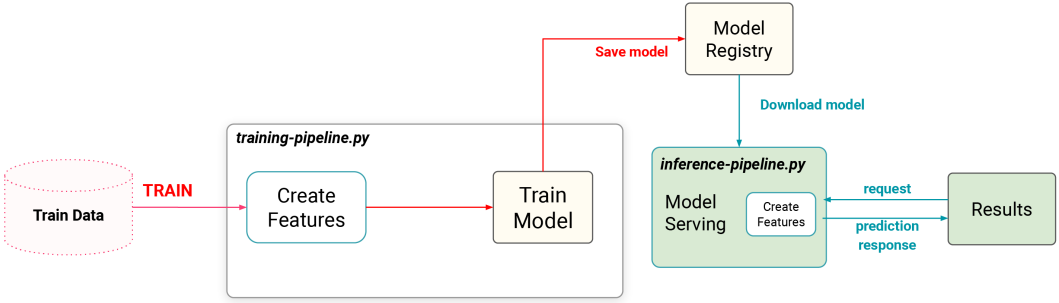
- Alles in einer Pipeline
- Ein Toggle wählt *Training* oder *Inferenz*
- Vorteil: Featureberechnung erfolgt nur einmal
- Nachteil: Keine Modularität, skaliert schlecht da eine Technologie für alles



Quelle: Hopsworks

Training und Inferenz separieren

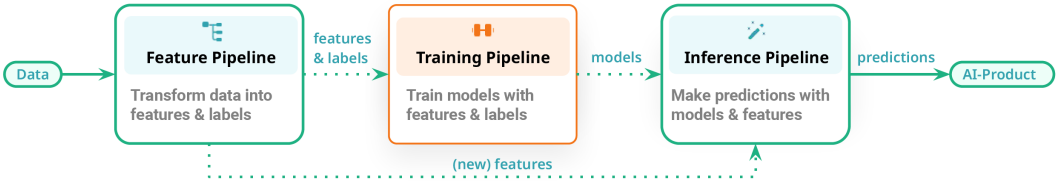
- Natürlicher Schritt: Training und Inferenz separieren
- Vorteil: Mehr Modularisierung
- Nachteil: Feature Engineering muss dupliziert werden



Quelle: Hopsworks

FTI or 3 Pipeline Architektur

- **Feature Pipeline**: Input = Rohdaten, output = Features (und Labels)
- **Training Pipeline**: Input = Features und Labels, output = trainiertes Modell
- **Inference Pipeline**: Input = Aktuelle Features+Modell, output = prediction

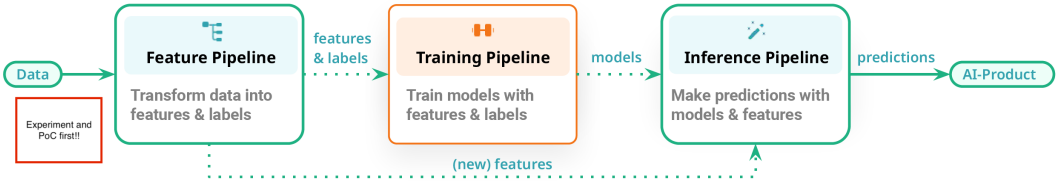


FTI Vorteile

- Technologie-offen
  - Python/Pandas für Training, Spark/dbt für Feature-Berechnung
  - Nicht abhängig von einer bestimmten ML Registry oder eines Feature Stores
  - Cloud oder on-premise möglich
- Einfach ausbaubar
  - Feature Store / Model Registry von einfacher Fileablage bis Full-Featured Produkt
- Modular / Reusable / Separat optimier- und skalierbar

F Pipeline vs. Experimentierphase

- F vor T vor I nicht effizient
- Experiment first, prove that it works or fail fast



FTI(M) Architektur

- Mehr zu Monitoring später
- Es braucht jedoch noch einen vierten Typ von Pipeline: Die Monitoring-Pipeline
- Nicht das ganze Monitoring kann und sollte in den FTI Pipelines gemacht werden
  - Referenzdaten nur bei Monitoring benötigt
  - Gewisse Datenmenge notwendig -> funktioniert nicht innerhalb von Streaming Feature Pipelines
  - Wenn Natural Labels stark verzögert -> funktioniert nicht in I Pipeline