

Einführung in MLOps

20 ORCHESTRIERUNG

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

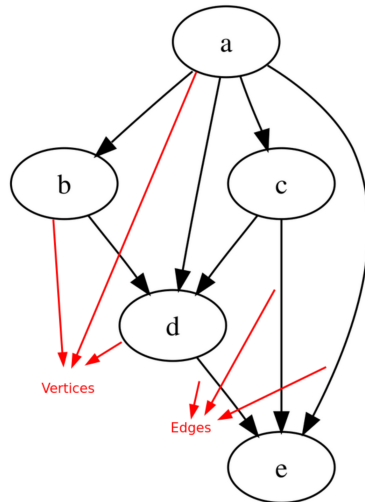
Generell: *In system administration, orchestration is the automated configuration, coordination, deployment, development, and management of computer systems and software. - Wikipedia*

Data Orchestration: *Data orchestration automates the movement and transformation of data between various systems, ensuring data is accurate, up-to-date, and ready for analysis. - ascend.io*

- Automatisierung
- Konsistenz
- Minimierung von Fehlern
- Parallelisierung und Skalierbarkeit

- Batch Pipelines müssen getriggert werden
 - zeitlich
 - externen Trigger

- Reihenfolge der auszuführenden Aufgaben
- Abhängigkeiten
- Parallelität
- Dispatching
- Terminologie
 - Directed Acyclic Graph (DAG)
 - DAG = Data Pipeline
 - Task = 1 Node im DAG
 - Dependency
 - Upstream: was vorher laufen muss
 - Downstream: was danach abhängig ist



Observability is a measure of how well internal states of a system can be inferred from knowledge of its external outputs. - Wikipedia

Zugriff auf alle Metadaten eines Pipeline-Runs

- Live Progress
- Logs
- Timestamps, Durchlaufzeit
- CPU, Memory Verbrauch

- Apache Airflow
- Prefect
- Dagster
- Mage

Ausführlicher Vergleich der ersten drei auf [ZenML](#).