

MLOps Systems Architecture

Easy process integration for AI Solutions is best enabled by a platform approach

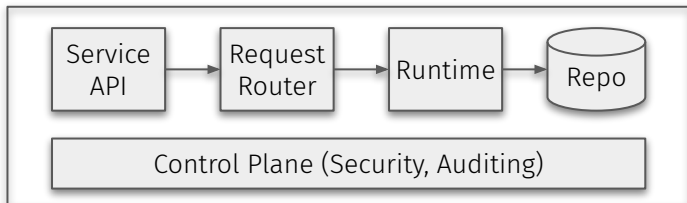
Success Factor: Process Integration

AI solutions only generate added value if integrated into operational processes. Studies show that companies that have integrated AI into their strategy are significantly more successful than those that view AI as merely a matter of efficiency.

AI solutions must be developed such that their integration into operational (IT) processes is quick, efficient and technically straightforward. This requirement is best met by a standardised approach, supported by an MLOps platform designed for this purpose.

MLOps Systems Architecture

The omega-ml MLOps platform consists of the following elements. This architecture is implemented using standard technology, enabling flexibility and cost-efficiency. Thanks to standardisation, deployment is straightforward and with flexible dependencies. omega-ml is horizontally and vertically scalable to adapt to a wide range of business scenarios.



This proven architecture covers all functions:

- **Service API** is the central interface for third-party applications, e.g. via REST API
- **Request Router** ensures each request is authenticated and assigns it to an appropriate runtime for execution
- **Runtime** provides all required resources, depending on the model type (e.g. Python libraries, memory and CPU for classic ML, possibly GPU for Generative AI) and is horizontally and automatically scalable
- **Repository** stores models and enables protected access to data from third-party sources (e.g. data lake, DWH, operational systems)
- **Control Plane** enables management of all platform resources, access protection, monitoring and ensures the traceability of all activities.

omega-ml uses Flask or Django for the service API, RabbitMQ as a request router, Python Celery for the runtime and MongoDB for the repository. The control plane is integrated into standard IT components such as Kubernetes and Keycloak.

Weshalb MLOps als Plattform?

The complexity of AI solutions is often underestimated because the requirement seems unspectacular at first glance: a simple IT service - 'predict' - needs to be packaged and integrated into an operational process as a REST-API interface. However, an MLOps platform offers many advantages.

Perspective	Single AI Service	MLOps Platform
Capabilities	Data Science, DevOps, Software-Engineering in distributed systems	Data Science
Technology	REST API using Flask, FastAPI, Django, Docker, Kubernetes, etc.	Standard Python
Scalability	Low, hard to achieve	High, automated
Security	Complex, specific to each service	Standardized, same for all AI services
Governance	Time-consuming, per-service	Singular, achieved at platform level

Umsetzung in 5 Schritten

Our tried-and-tested steps for the successful introduction of MLOps are suitable for all typical data science teams of 1 to 10 people:

(1) Make or Buy?

In-house development offers a high degree of flexibility, but is complex and, in view of the availability of standardised platforms, comparatively not cost-efficient.



omega-ml

(2) Wahl einer integrierten MLOps Plattform

Many MLOps offerings do not provide an integrated platform, but only partial components.



(3) Proof of Concept

A technical test based on a realistic scenario quickly shows where the benefits and challenges lie.



(4) Operationalisation

By integration into your existing IT architecture, DevOps processes and security environment, the MLOps platform is available for productive operation.



(5) Training and scaling as needed

Um die Plattform effizient zu nutzen, ist die Einführung und Schulung von Data Scientists und DevOps Spezialisten ein effizientes Mittel.

