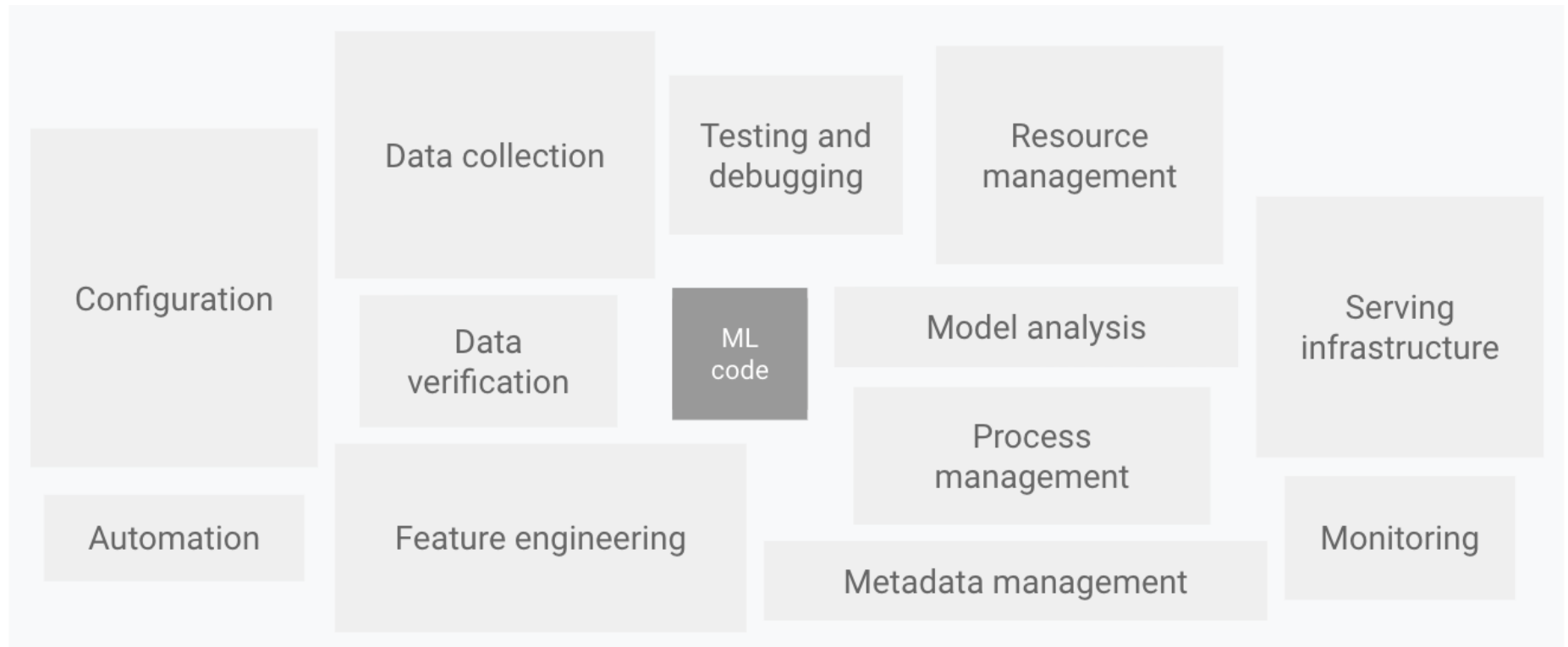


# MLOps: Machine Learning Operations

Гущин Александр

DMIA Production ML  весна 2021

# О чем лекция



“Hidden Technical Debt in Machine Learning Systems” paper

**Технический долг** — накопленные в программном коде или архитектуре проблемы, связанные с пренебрежением к качеству при разработке программного обеспечения и вызывающие дополнительные затраты труда в будущем.

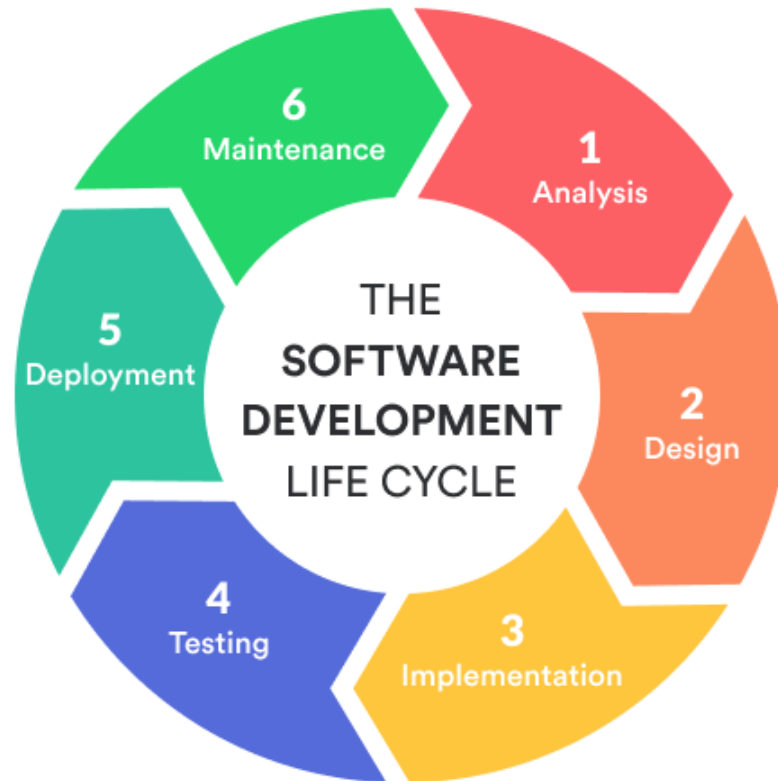
**ML-системы** имеют дополнительные “возможности” для создания технического долга. ML Engineer сталкивается с этим **в первую очередь**.

Например — unstable data dependencies, feedback loops, glue code, pipeline jungles, dead experimental codepaths, fixed thresholds in dynamic systems, entanglement (CACE principle), etc.

ML системы отличаются от прочих программных систем в следующем:

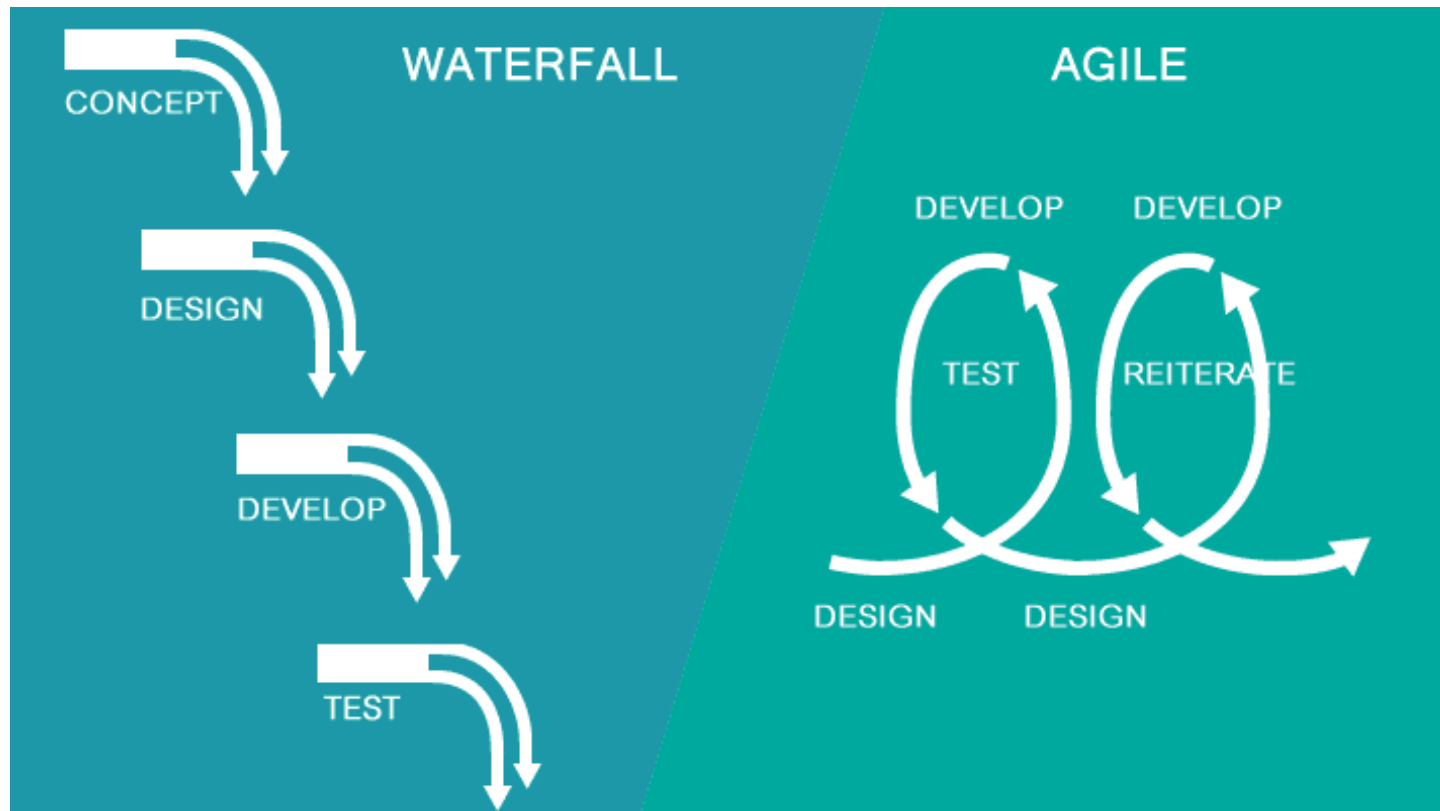
- Квалификация команды: DS и MLRs могут не иметь опыта построения production-сервисов
- Разработка: ML-модели требуют постоянных экспериментов. Сложность состоит в отслеживании результатов и их воспроизводимости, в то же код должно быть легко изменять и переиспользовать.
- Тестирование: помимо обычного тестирования, нам требуется также проверять качество данных, качество обученных моделей и адекватность их работы.
- Деплоймент: требует задеплоить не просто ML модель, а целый ML pipeline, который уже будет переобучать и деплоить модель
- Production: в отличие от обычного ПО, ML модели устаревают, поскольку данные непрерывно меняются

# SDLC



Source

# Методологии управления проектами



Waterfall or Agile

# Devops

DevOps - набор практик, объединяющий разработку ПО (Dev) и управление IT-инфраструктурой и сервисами (Ops). Цель DevOps - сократить цикл разработки и обеспечить непрерывную поставку обновлений (Continuous Delivery) с высоким качеством.



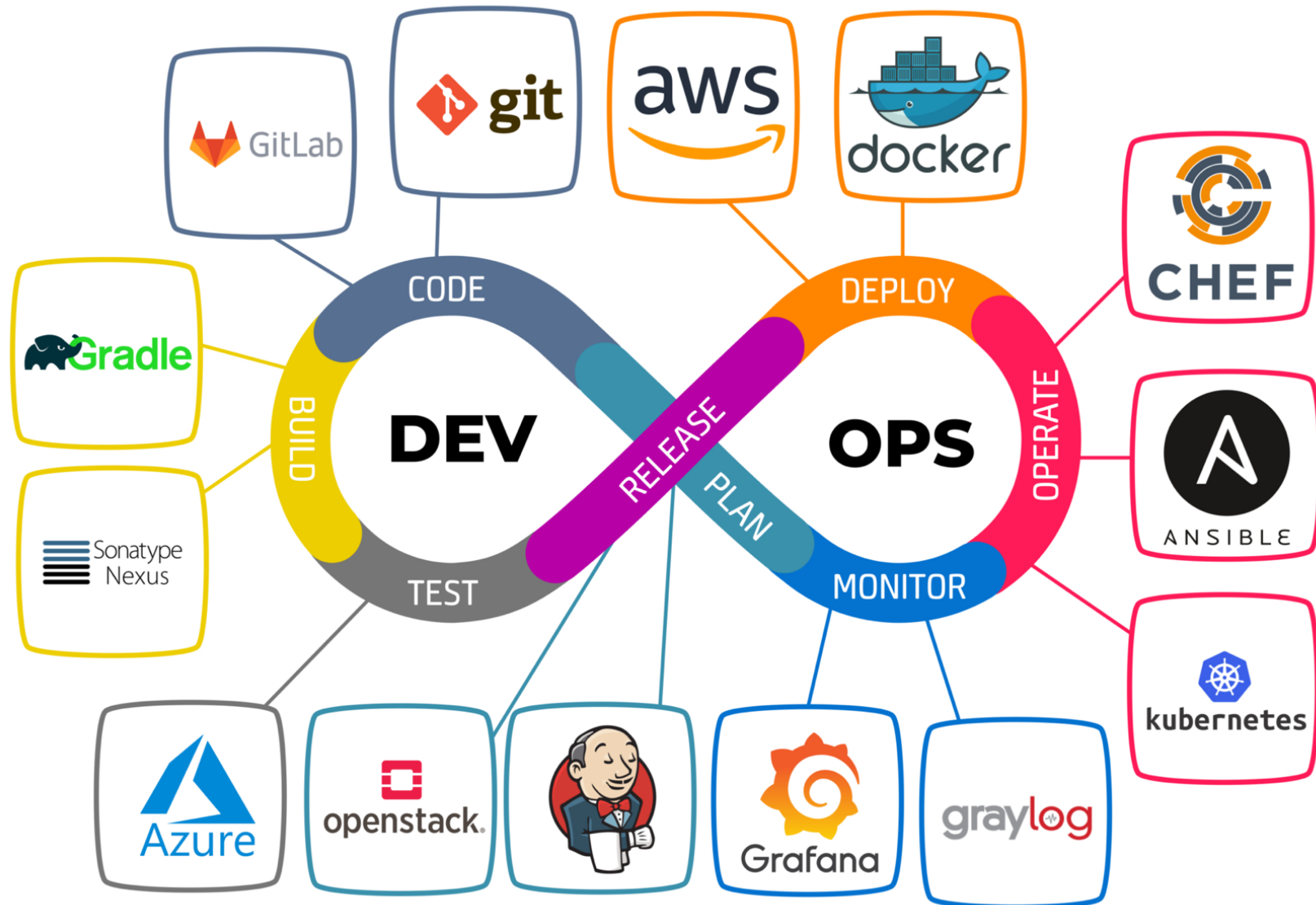
A peek into Agile DevOps

## Key DevOps aspects

- Coding – code development and review, source code management tools, code merging.
- Building – **continuous integration** tools, build status.
- Testing – **continuous testing** tools that provide quick and timely feedback on business risks.
- Packaging – artifact repository, application pre-deployment staging.
- Releasing – change management, release approvals, release automation, **continuous deployment** tools.
- Configuring – infrastructure configuration and management, infrastructure as code tools.
- Monitoring – applications performance monitoring, end-user experience.



# DevOps toolchain



# What is DevOps and where is it applied

ML and other software systems are similar in continuous integration of source control, unit testing, integration testing, and continuous delivery of the software module or the package. However, in ML, there are a few notable differences:

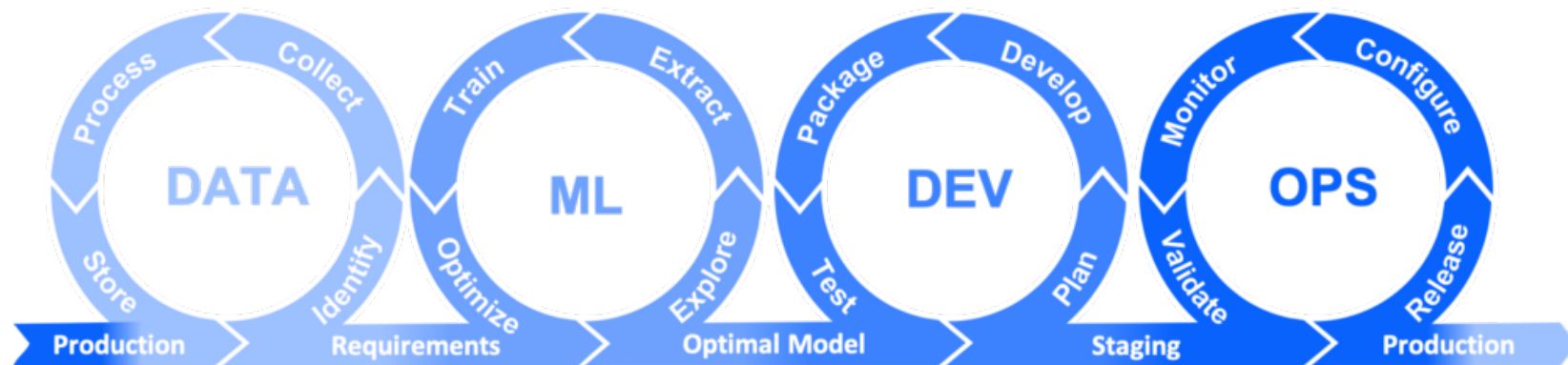
- CI is no longer only about testing and validating code and components, but also testing and validating data, data schemas, and models.
- CD is no longer about a single software package or a service, but a system (an ML training pipeline) that should automatically deploy another service (model prediction service).
- CT is a new property, unique to ML systems, that's concerned with automatically retraining and serving the models.

**MLOps: Continuous delivery and automation pipelines in machine learning**

# MLOps

MLOps is the extension of the DevOps methodology to include Machine Learning and Data Science assets as first-class citizens within the DevOps ecology. [MLOps SIG](#)

Machine Learning Engineering (MLE) is the use of scientific principles, tools, and techniques of machine learning and traditional software engineering to design and build complex computing systems. MLE encompasses all stages from data collection, to model building, to make the model available for use by the product or the consumers." [A.Burkov](#).



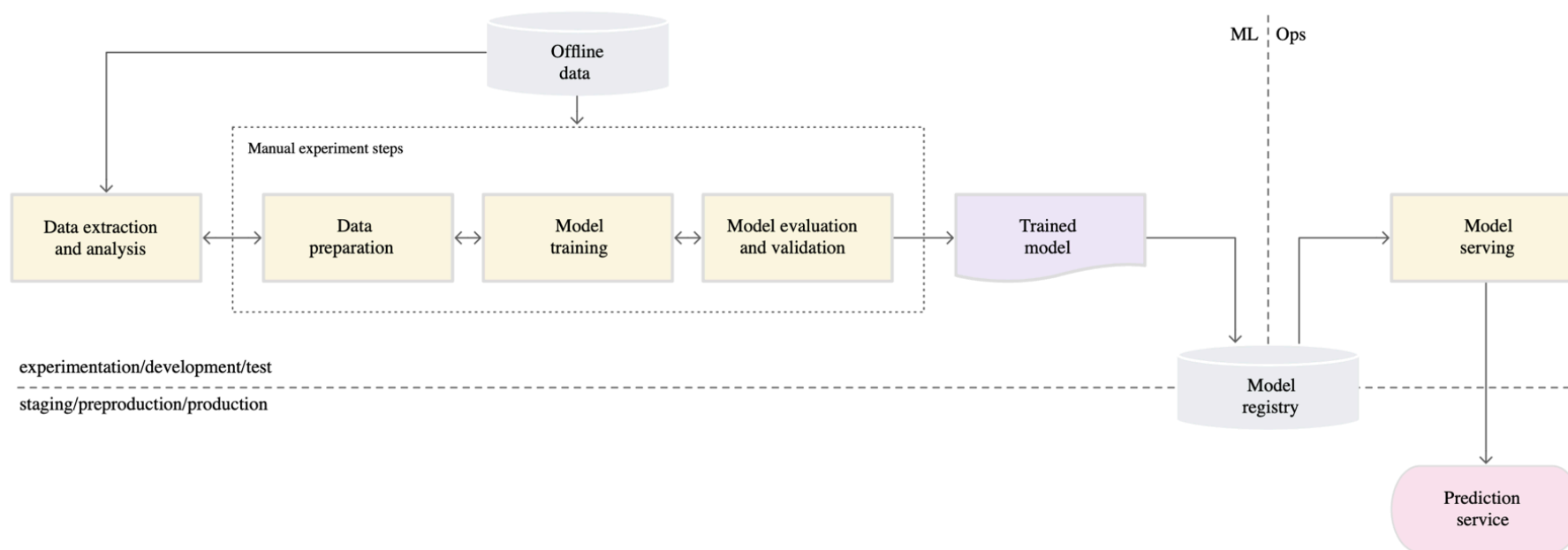
The MLOps Blueprint

## Что должен делать MLOps?

- Стремиться объединить релизный цикл ML и релизный цикл ПО
- Автоматизировать тестирование ML артефактов
- Применять Agile принципы к ML проектам
- Интегрировать ML артефакты в CI/CD системы
- Снижать технический долг, возникающий при использовании ML

SIG MLOps

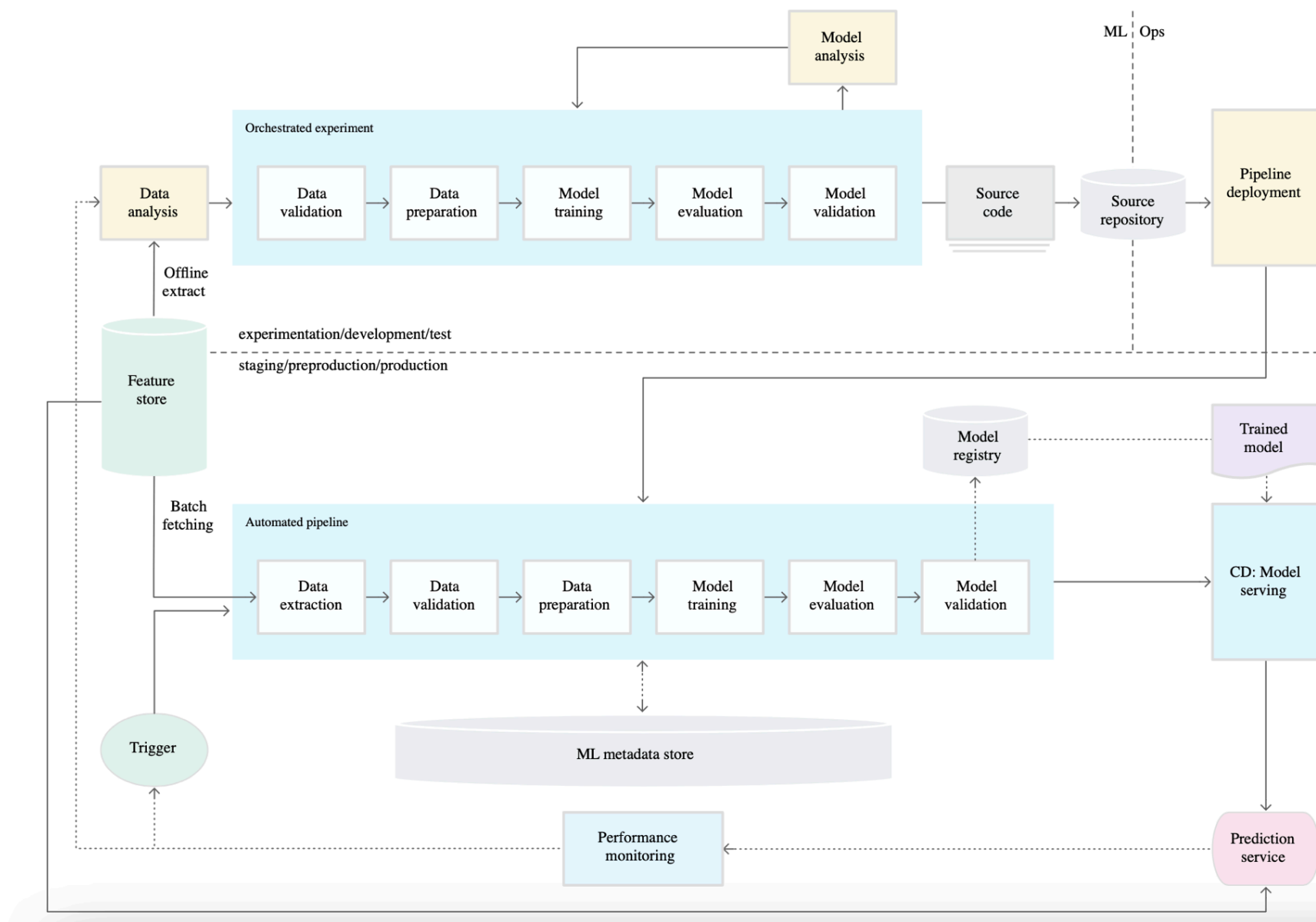
# MLOps Level 0: ручной процесс



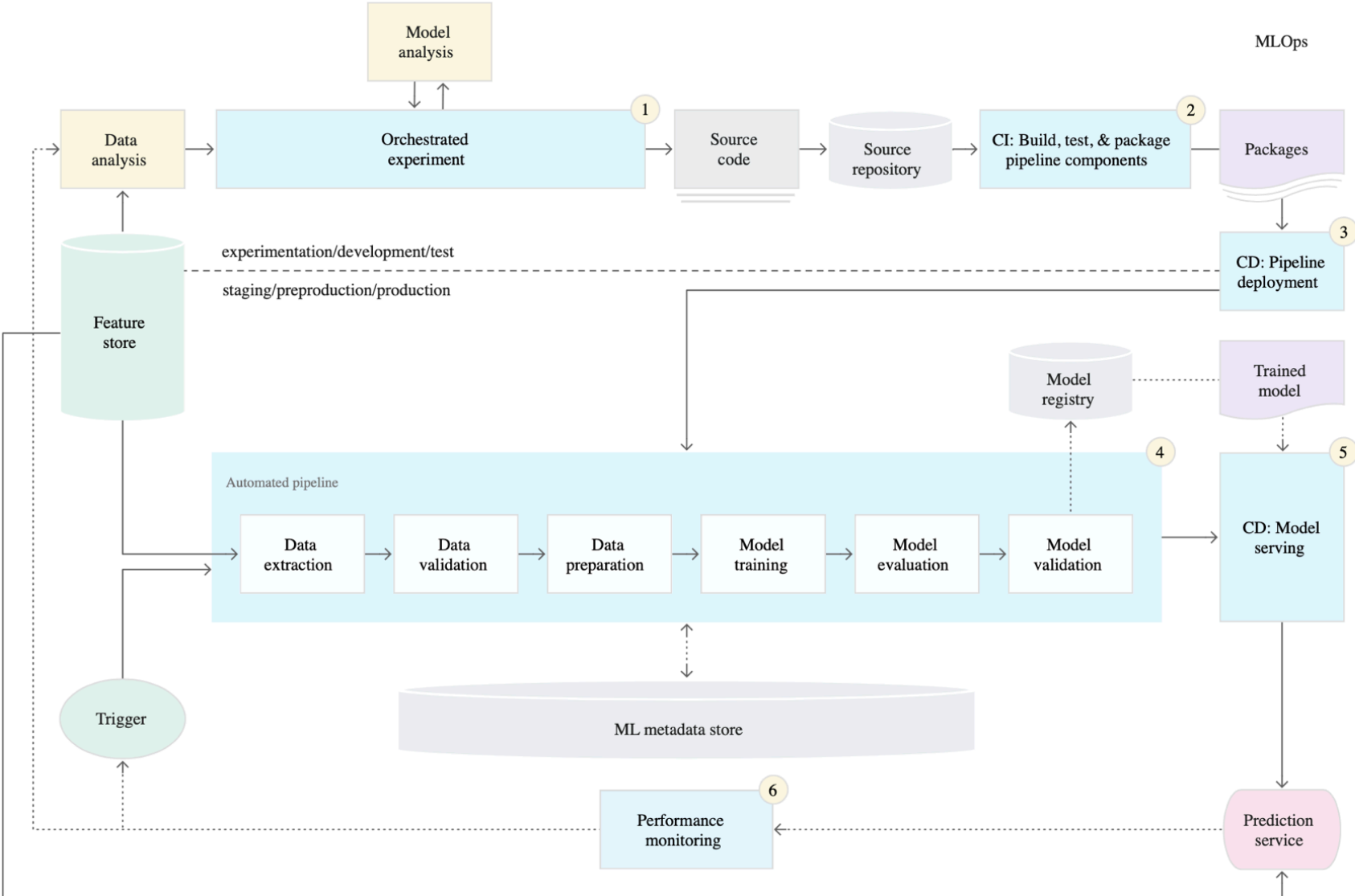
+ Model monitoring

MLOps: Continuous delivery and automation pipelines in machine learning

# MLOps Level 1: автоматизация ML пайплайна



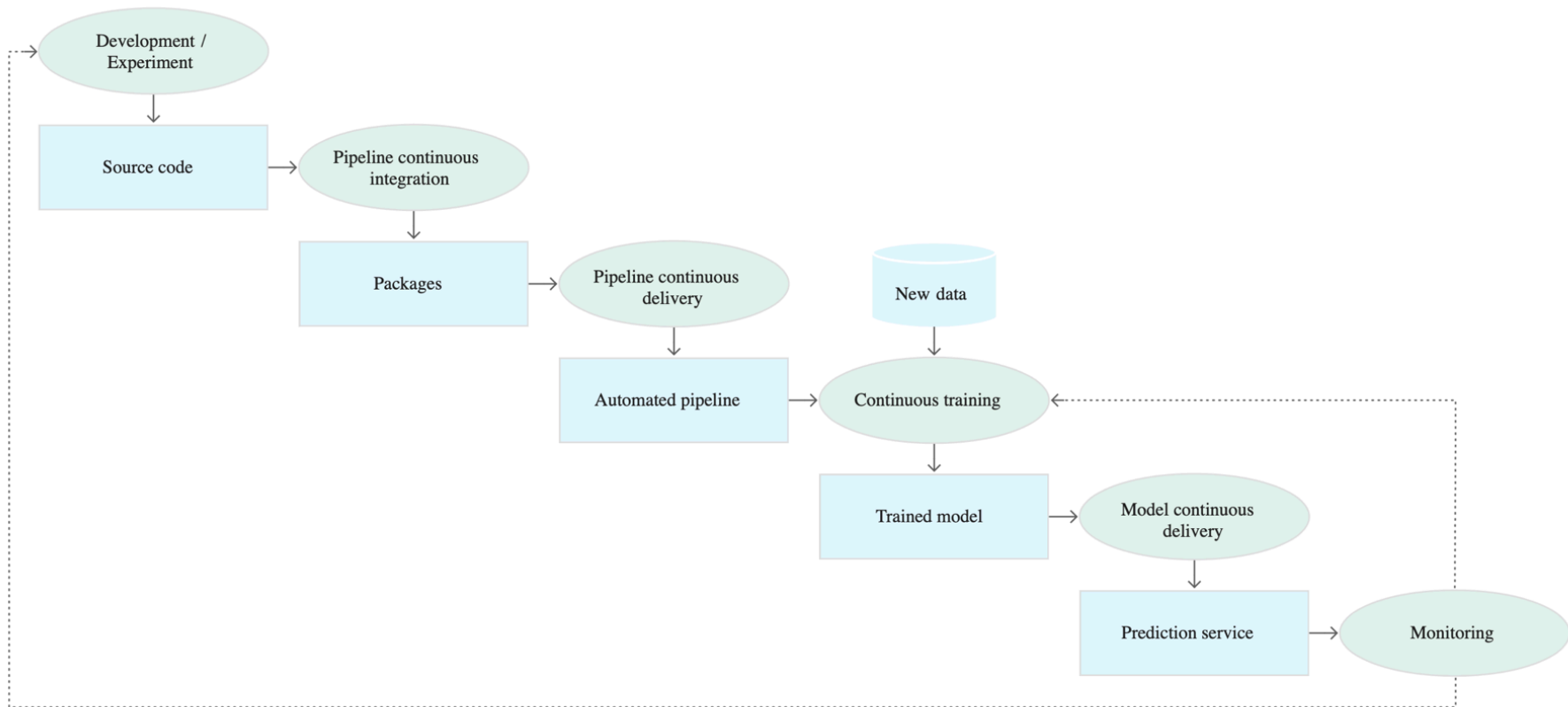
# MLOps Level 2: автоматизация CI/CD пайплайна







# MLOps - итоговый пайплайн



# Саммари

1. DevOps is a set of practices that combines software development (Dev) and IT operations (Ops). It aims to shorten the systems development life cycle and provide continuous delivery with high software quality.
2. Одними из наиболее важных практик DevOps являются Continuous Integration и Continuous Delivery.
3. MLOps применяет подход DevOps для разработки ML-систем с учетом их особенностей. С необходимостью обновлять модели добавляется практика Continuous Training.
4. В нашем курсе мы будем работать с инструментами, позволяющими реализовать подход MLOps в работе над нашим проектом - ботом для Шляпы.