

# UDV Summer School 2021

## Архитектура ПО и DevOps/MLOps

> Чернышов Юрий

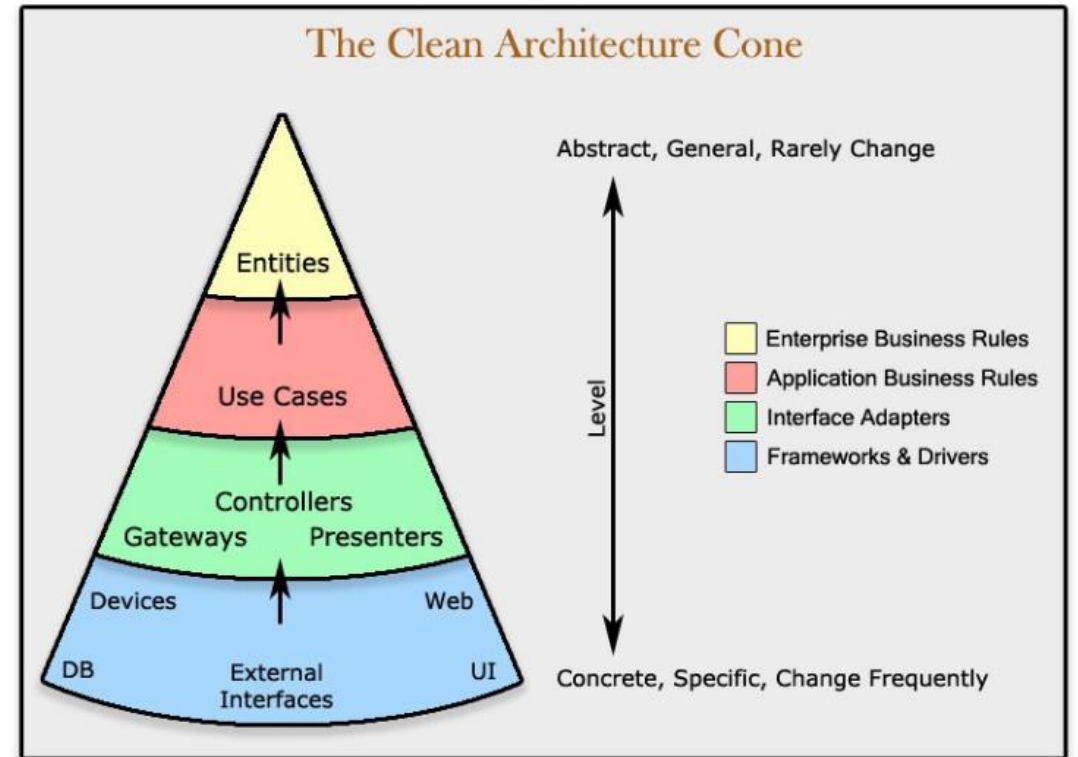
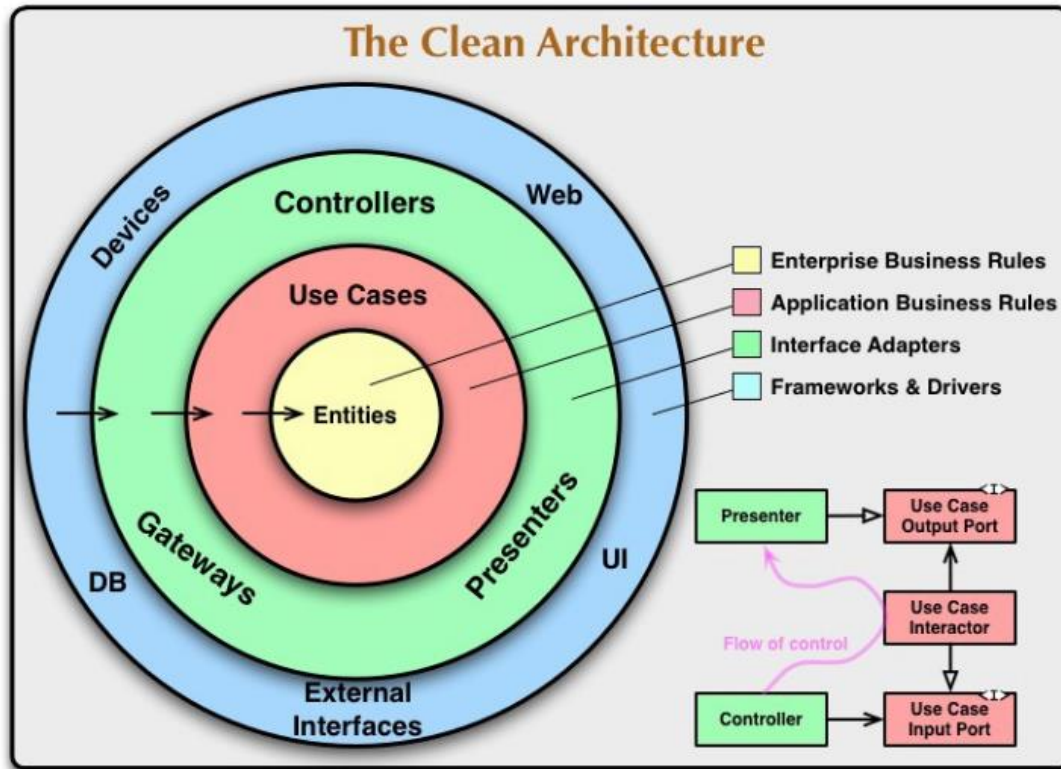


# **Архитектура программного обеспечения**

## > Почему меняются подходы к разработке ПО

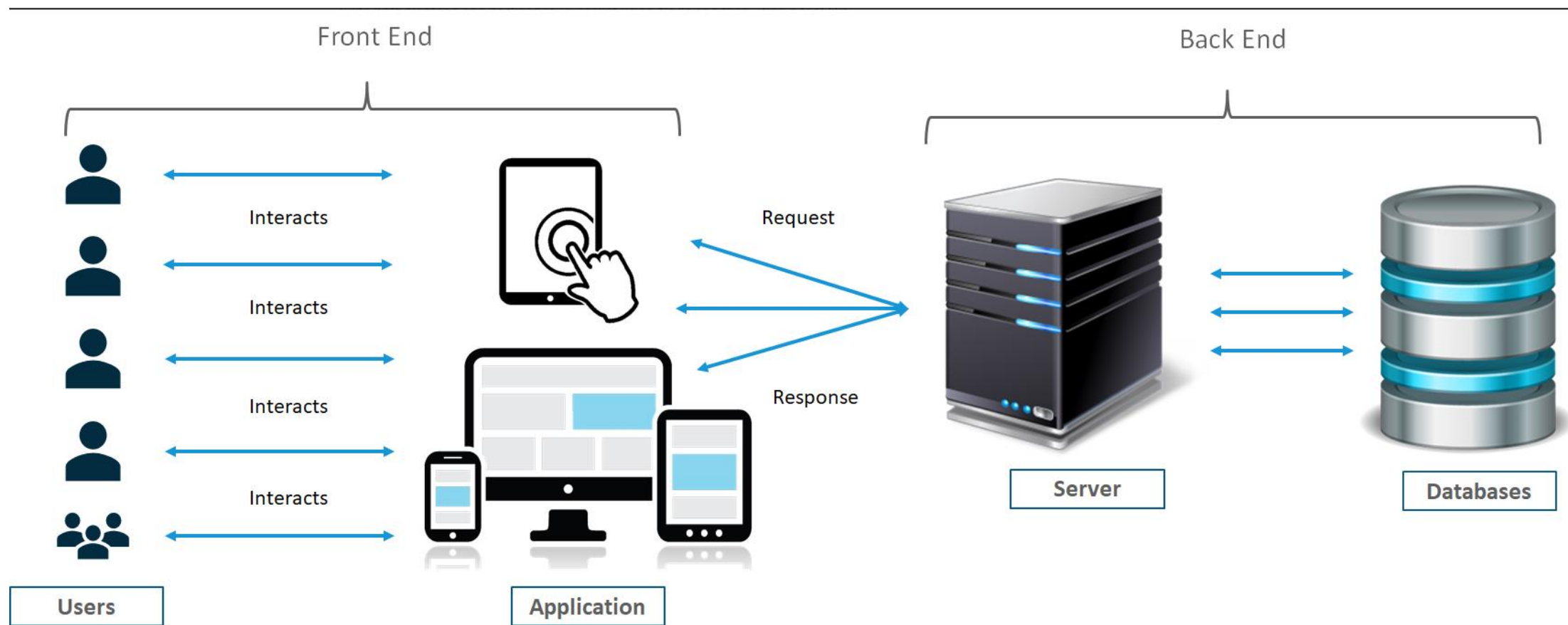
- Сокращается жизненный цикл продуктов и проектов, ускорение процессов
- Многофункциональное программное обеспечение становится все более сложным
- Усиливается специализация разработчиков и компаний
- Важны вопросы производительности, масштабируемости, планирования и распределения ресурсов
- Существует множество технологий
- Для создания надежных и производительных архитектур необходим большой опыт и широкие знания: программирование, инструменты разработки, аппаратное обеспечение, технологии, ...

## ➤ Лучшие практики архитектуры ПО



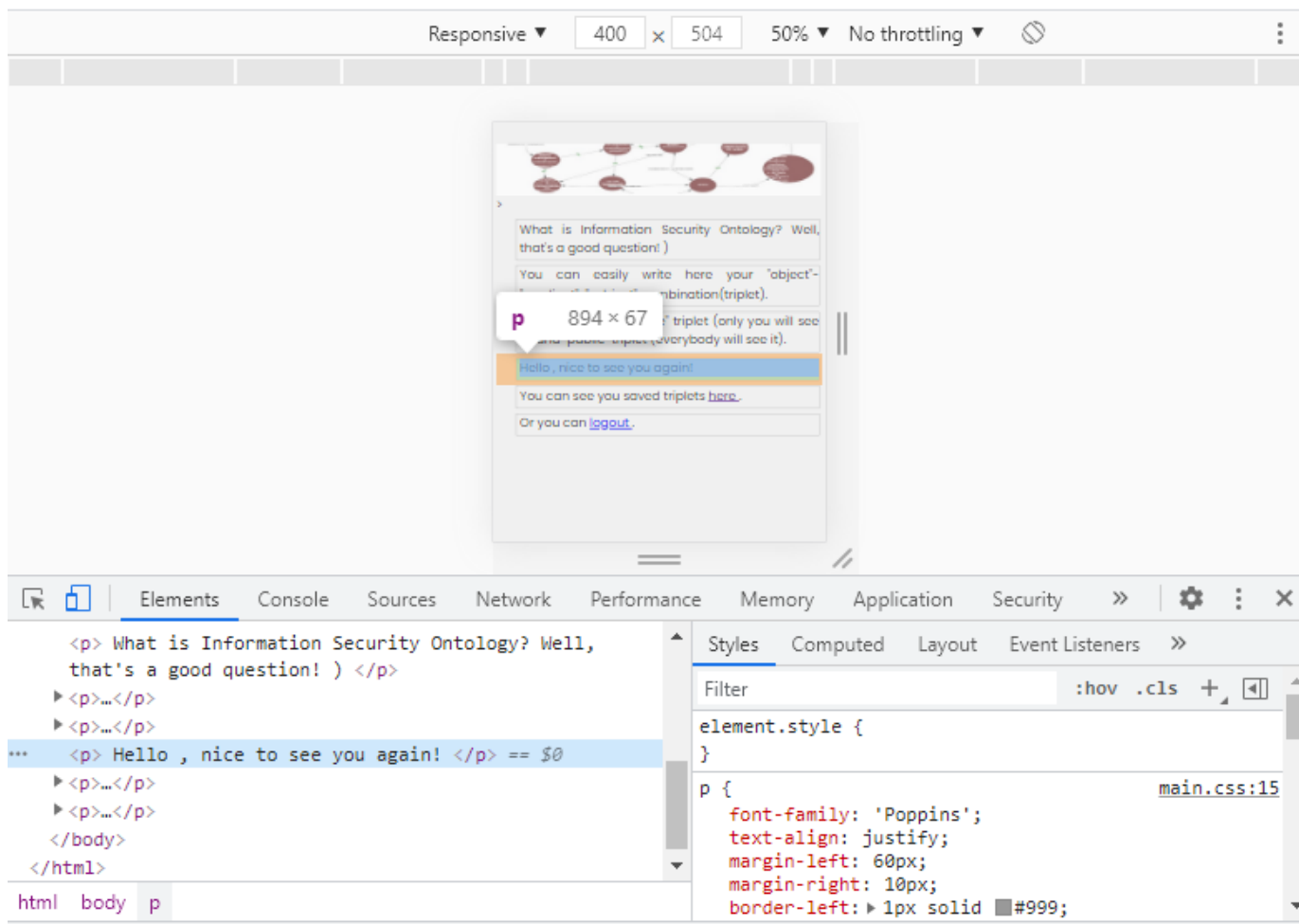
- <https://www.freecodecamp.org/news/a-quick-introduction-to-clean-architecture-990c014448d2/>

## > Архитектура серверного приложения

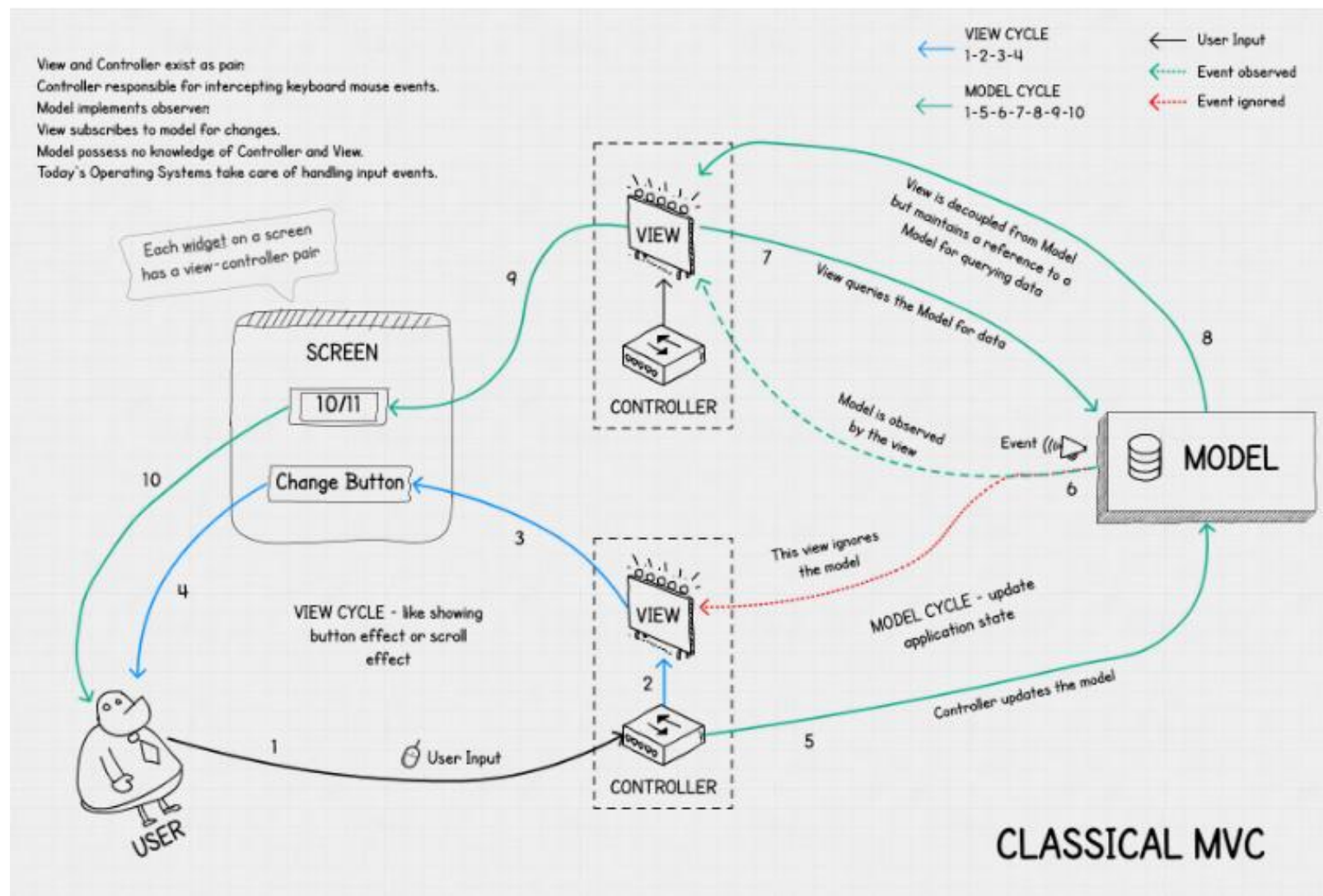


# > Фронтэнд

- Функции
  - Графический интерфейс
  - Обработка входящих запросов
- Технологии
  - HTML+CSS
  - JavaScript, TypeScript, Angular (динамические функции веб-страниц)



## ➤ Архитектура фронтенда

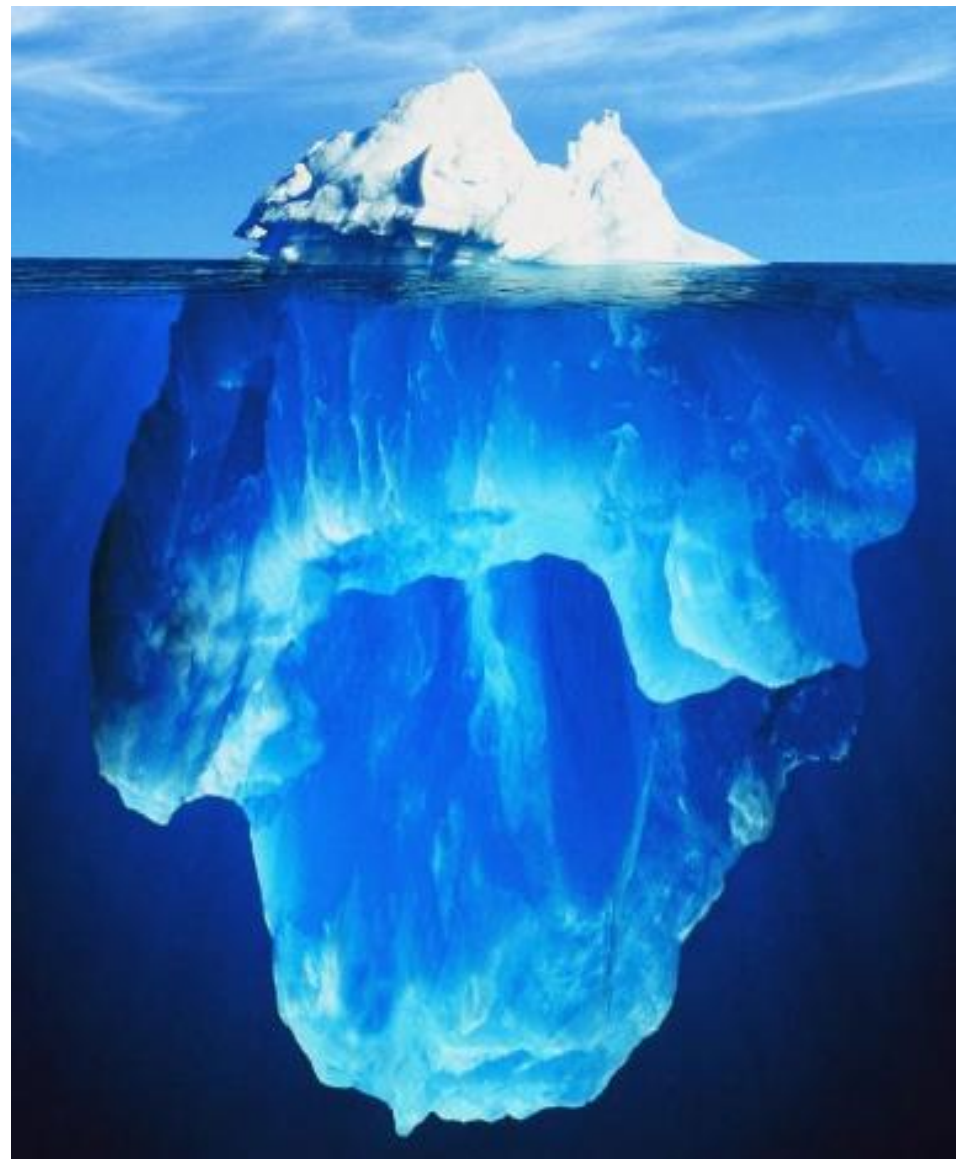






## Бэкэнд

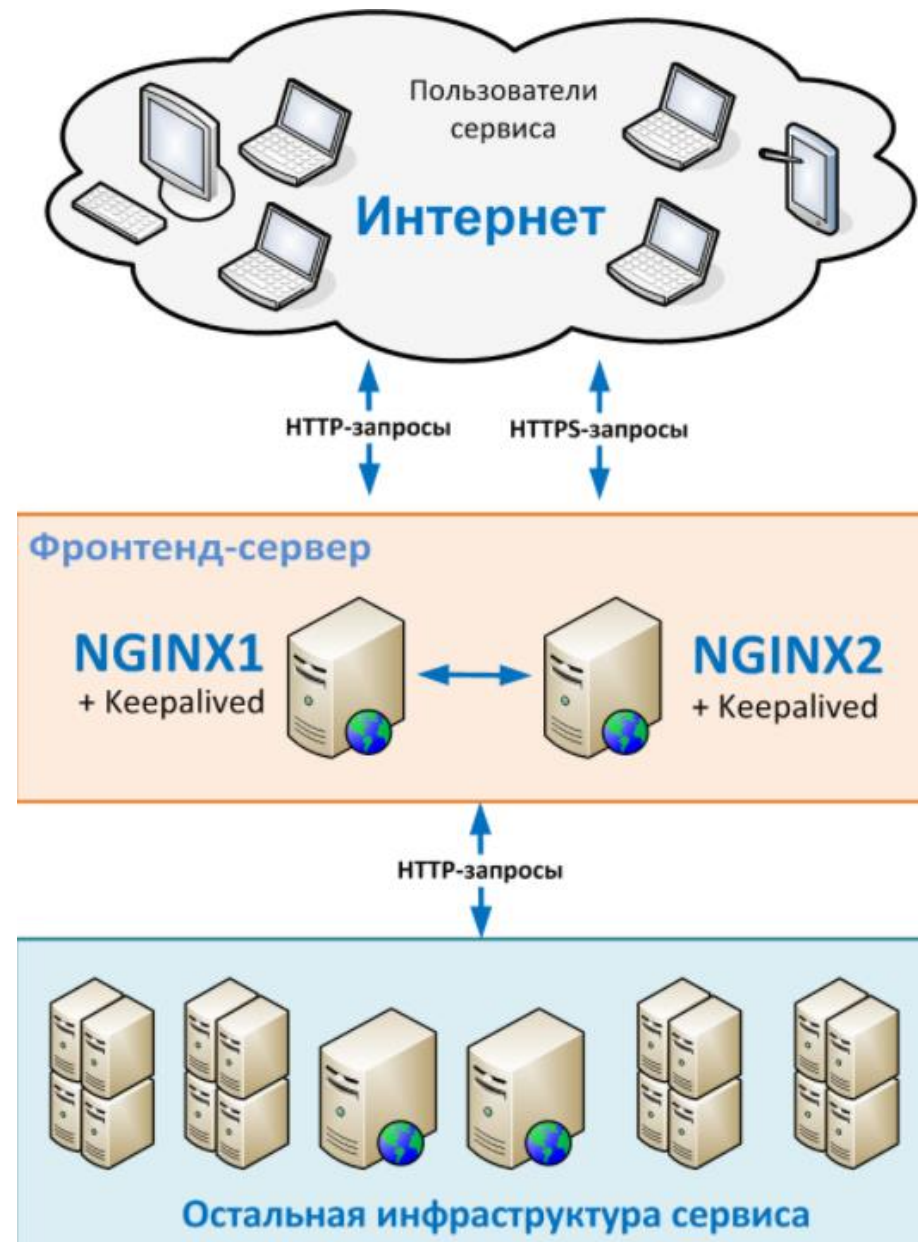
- Функции
  - Серверные функции (обработка запросов, подготовка ответов)
  - Вычисления
  - Работа с БД (ORM)
  - Взаимодействие через API
- Технологии
  - Python
  - Java
  - JS (Node.js)
  - PHP
  - C, C++





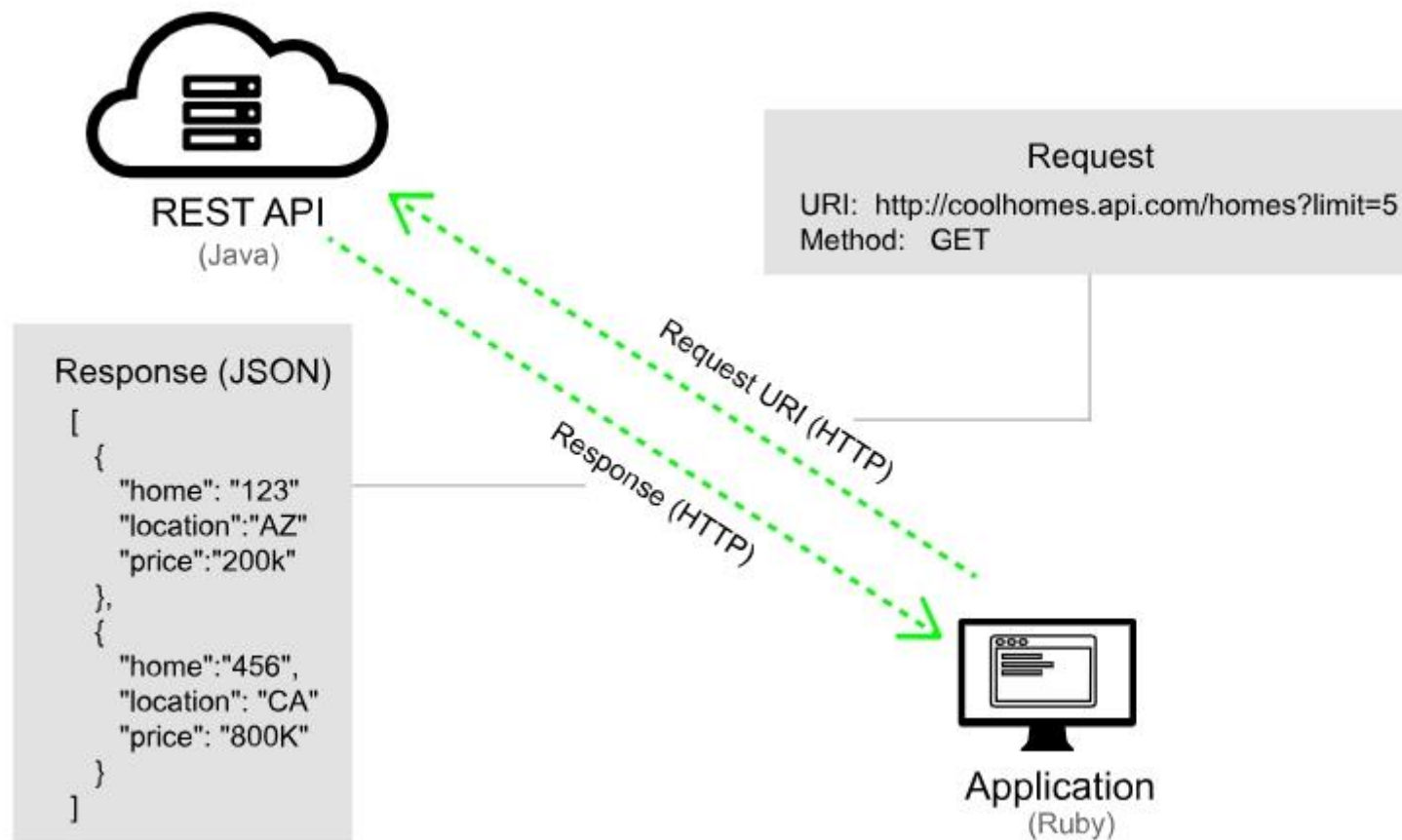
## > Web сервер

- Функции
  - Взаимодействие с пользователями через web
  - Проксирование (сокрытие топологии приложений, распределение запросов)
  - Load Balance (распределение нагрузки)
  - Доступ к статическим файлам (графика, web)
  - Защита
- Программы
  - Nginx, gunicorn, apache



## > API

- Стандартный язык взаимодействия процессов и приложений
- Возможность получения доступа из внешних приложений (yandex.map, yandex.speechkit, openstreetmap, openweather, ... )
- Соответствие микросервисной архитектуре



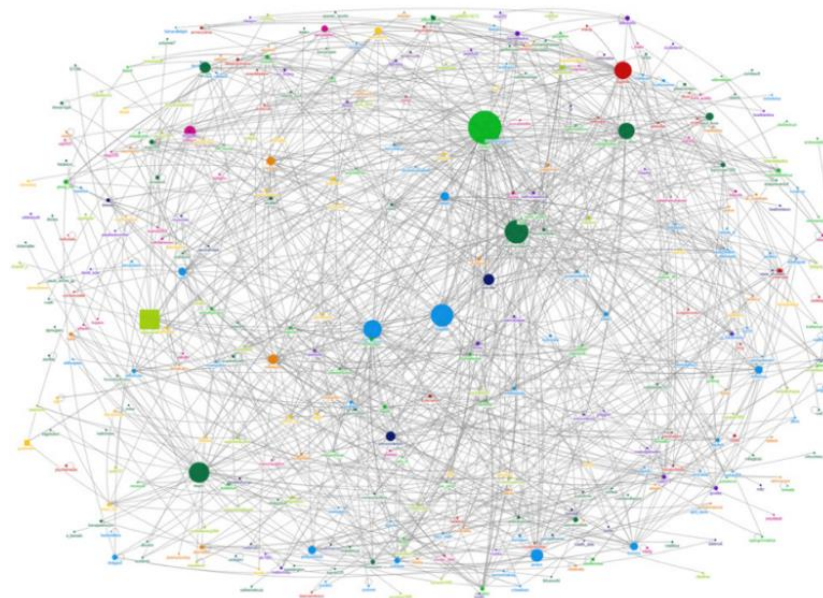
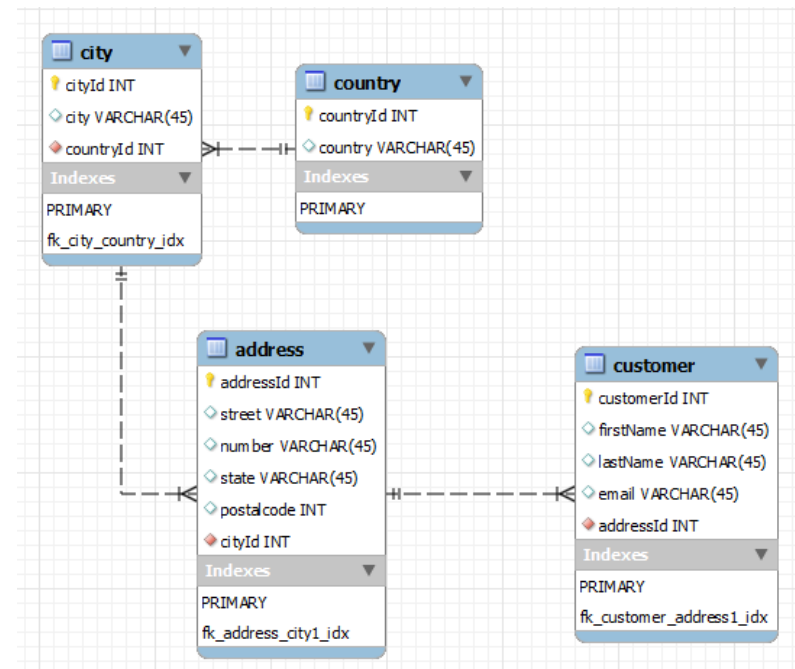
# > Базы данных

- Типы

- Реляционные
- Графовые
- Иерархические БД
- Ключ-значение
- Временные ряды
- Колоночные

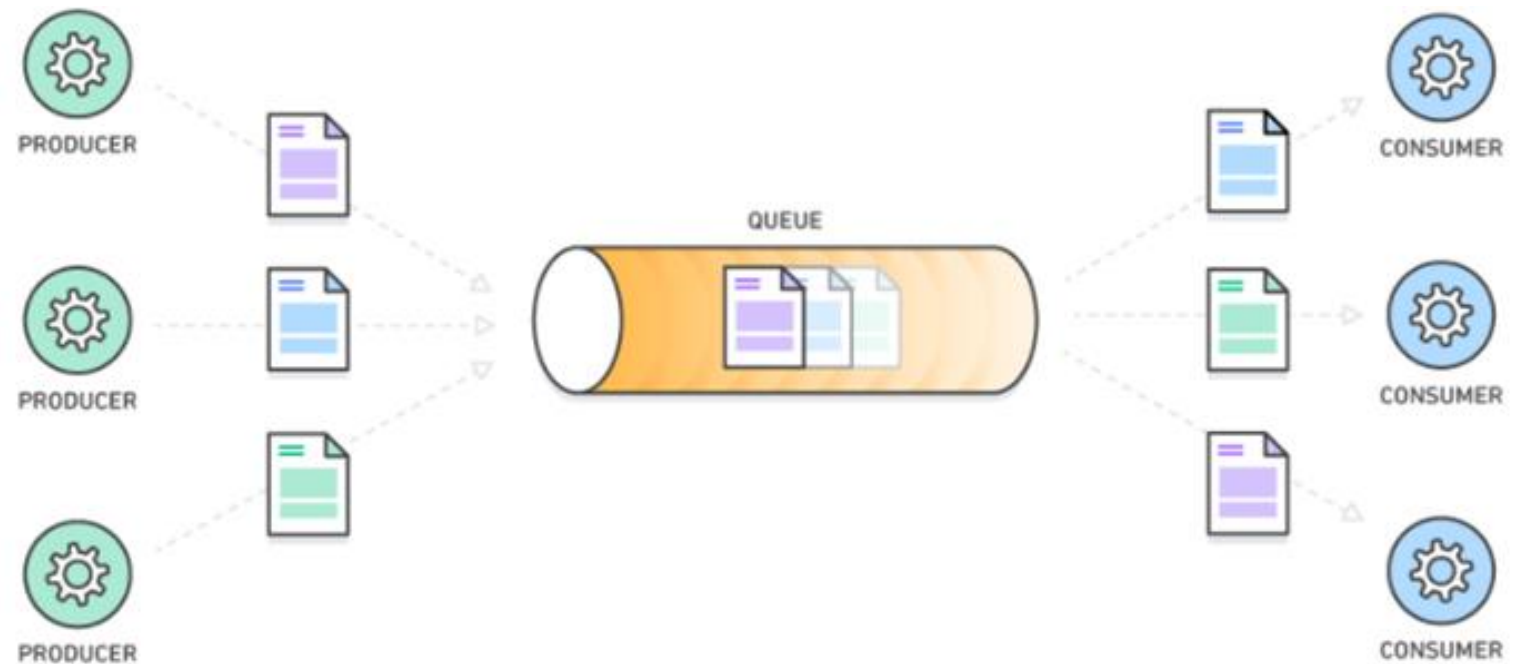
- ПО

- xml
- SQLite, MySQL, MS SQL
- PostgreSQL
- MongoDB
- InfluxDB

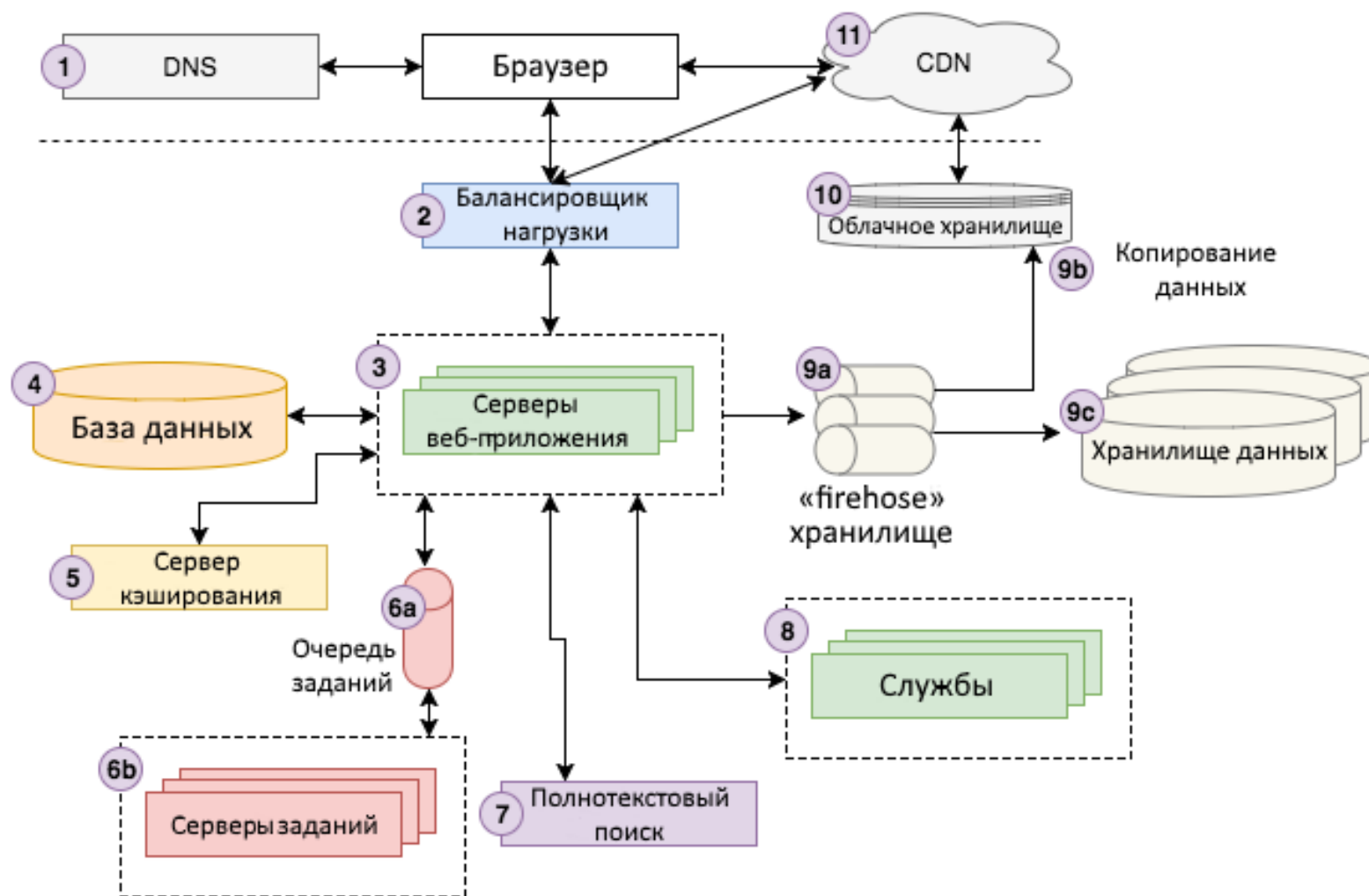


## > Очереди сообщений

- RabbitMQ, Apache Kafka
- Обмен информацией
- Контроль выполнения



## > Микросервисная архитектура (пример)



# DevOps

## > Что такое DevOps

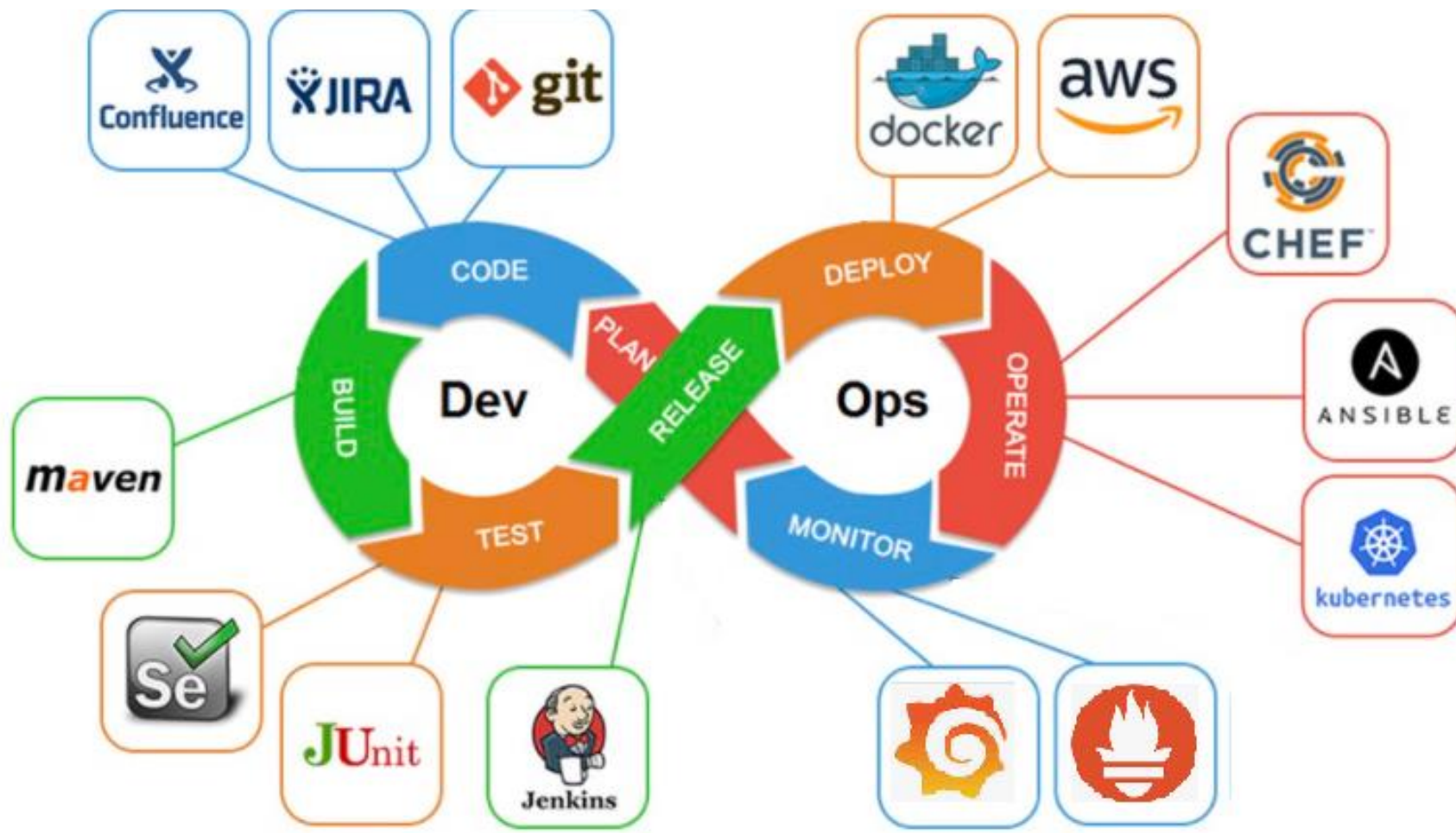
- **DevOps (Development and Operations)** – практика, направленная на объединение команд разработки и эксплуатации для быстрой и надежной сборки, тестирования и выпуска релизов программного обеспечения. Тесно связано с Agile (SCRUM, Kanban)
- **Основная цель DevOps** – ускорить вывод продукта на рынок (за счет применения постепенных улучшений в ответ на меняющуюся среду) и создать более упорядоченный процесс разработки (сокращение продолжительности этапов, автоматизация, контроль версий, ...).



## > Этапы разработки ПО



## > Инструменты DevOps

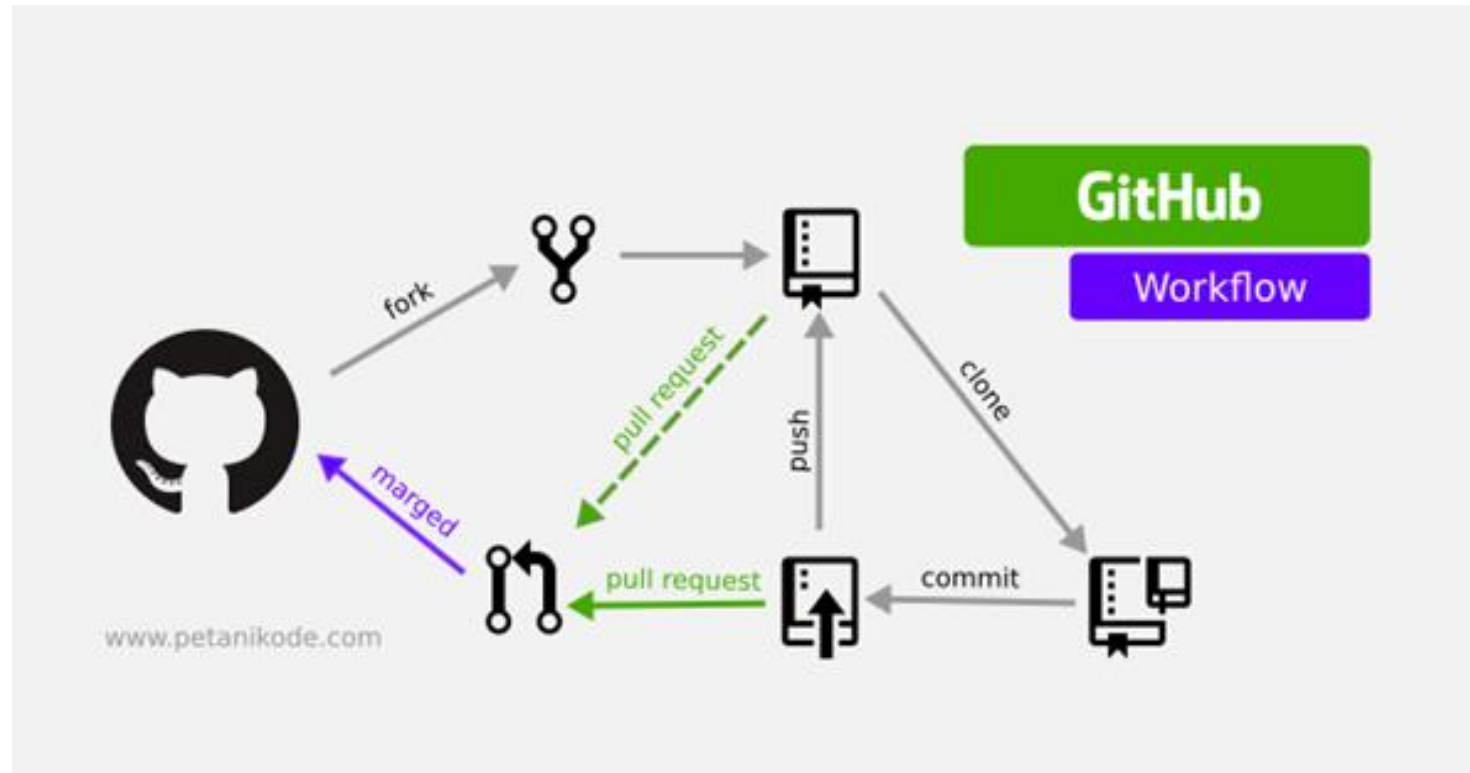


## > Все можно автоматизировать

- **Разработка:** системы управления версиями (git) для контроля процесса создания кода и изменений.
- **Сборка и доставка:** CI/CD (GitLab, Jenkins), скрипты для автоматизации операций (тестирование, сборка, конфигурирование, вывод в production) и доставка протестированных частей по назначению
- **Развёртывание:** создание инфраструктуры и управление приложениями в «облаке», управление инфраструктурой как кодом (IaC) (Terraform, Ansible)
- **Настройка сред выполнения:** оркестраторы контейнеров (docker-compose, Kubernetes)

## > Управление изменениями (версиями)

- Github
- Bitbucket
- GitLab



# > Jira, confluence

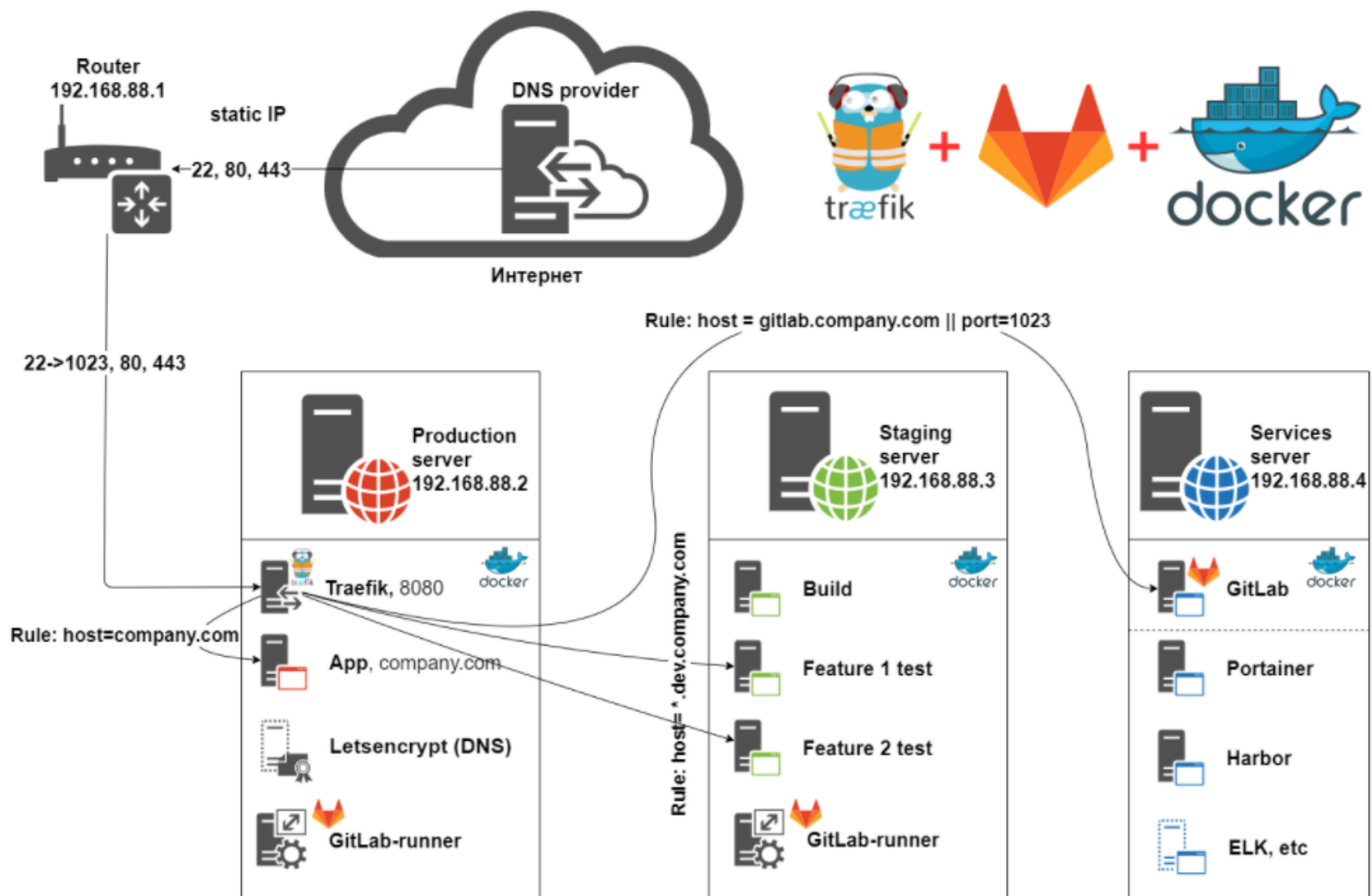
The screenshot displays the Jira interface for a project named "Teams in Space". The top navigation bar includes "Jira", "Your work", "Projects", "Filters", "Dashboards", "People", "Plans", "Apps", and a "Create" button. A search bar and user avatars are on the right.

The left sidebar contains a navigation menu with the following items: "Teams in Space" (Classic software project), "Scrum: Teams in S..." (Board), "Roadmap", "Backlog", "Active sprints" (highlighted), "Reports", "Issues", "Components", "Releases", "Project pages", "Add item", and "Project settings".

The main area is titled "Board" and features a "Release" button and a "Quick Filters" dropdown. The board is organized into four columns representing different stages of the workflow:

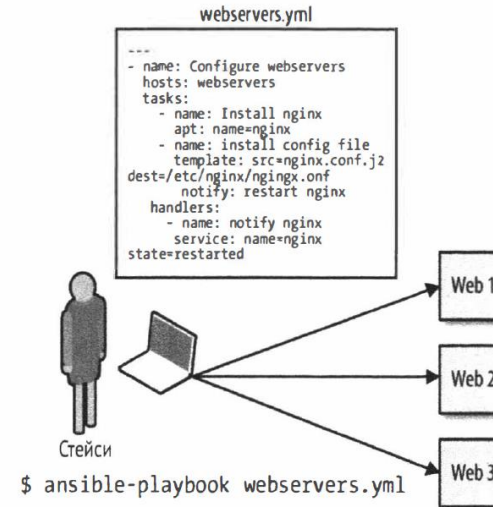
- TO DO 5**:
  - Engage Jupiter Express for outer solar system travel (SPACE TRAVEL PARTNERS, 5 tasks, TIS-26)
  - Create 90 day plans for all departments in the Mars Office (LOCAL MARS OFFICE, 9 tasks, TIS-12)
  - Engage Saturn's Rings Resort as a preferred provider (SPACE TRAVEL PARTNERS, 3 tasks, TIS-17)
  - Enable Speedy SpaceCraft as the preferred
- IN PROGRESS 5**:
  - Requesting available flights is now taking > 5 seconds (SEESPACEEZ PLUS, 3 tasks, TIS-8)
  - Engage Saturn Shuttle Lines for group tours (SPACE TRAVEL PARTNERS, 4 tasks, TIS-15)
  - Establish a catering vendor to provide meal service (LOCAL MARS OFFICE, 4 tasks, TIS-15)
  - Engage Saturn Shuttle Lines for group tours
- CODE REVIEW 2**:
  - Register with the Mars Ministry of Revenue (LOCAL MARS OFFICE, 3 tasks, TIS-11)
  - Draft network plan for Mars Office (LOCAL MARS OFFICE, 3 tasks, TIS-15)
- DONE 8**:
  - Homepage footer uses an inline style- should use a class (LARGE TEAM SUPPORT, 8 tasks, TIS-68)
  - Engage JetShuttle SpaceWays for travel (SPACE TRAVEL PARTNERS, 5 tasks, TIS-23)
  - Engage Saturn Shuttle Lines for group tours (SPACE TRAVEL PARTNERS, 1 task, TIS-15)
  - Establish a catering vendor to provide meal service (LOCAL MARS OFFICE)

## ➤ Инфраструктура как код (IaC)



# > Ansible

- Разработан в 2004, свободное ПО, Ansible Inc. принадлежит Red Hat
- Скрипты для автоматизации: развертывание и управление сложными конфигурациями
- OpenSSH
- Предназначен для: системных администраторов, DevOps
- Аналоги: chef, puppet, salt
- ПО: на хостах SSH + python2.5+
- Режимы принудительной и добровольной настройки



```
- name: Configure webserver with nginx
  hosts: webservers
  become: True
  tasks:
    - name: install nginx
      apt: name=nginx update_cache=yes

    - name: copy nginx config file
      copy: src=files/nginx.conf dest=/etc/nginx/sites-available/default

    - name: enable configuration
      file: >
        dest=/etc/nginx/sites-enabled/default
        src=/etc/nginx/sites-available/default
        state=link

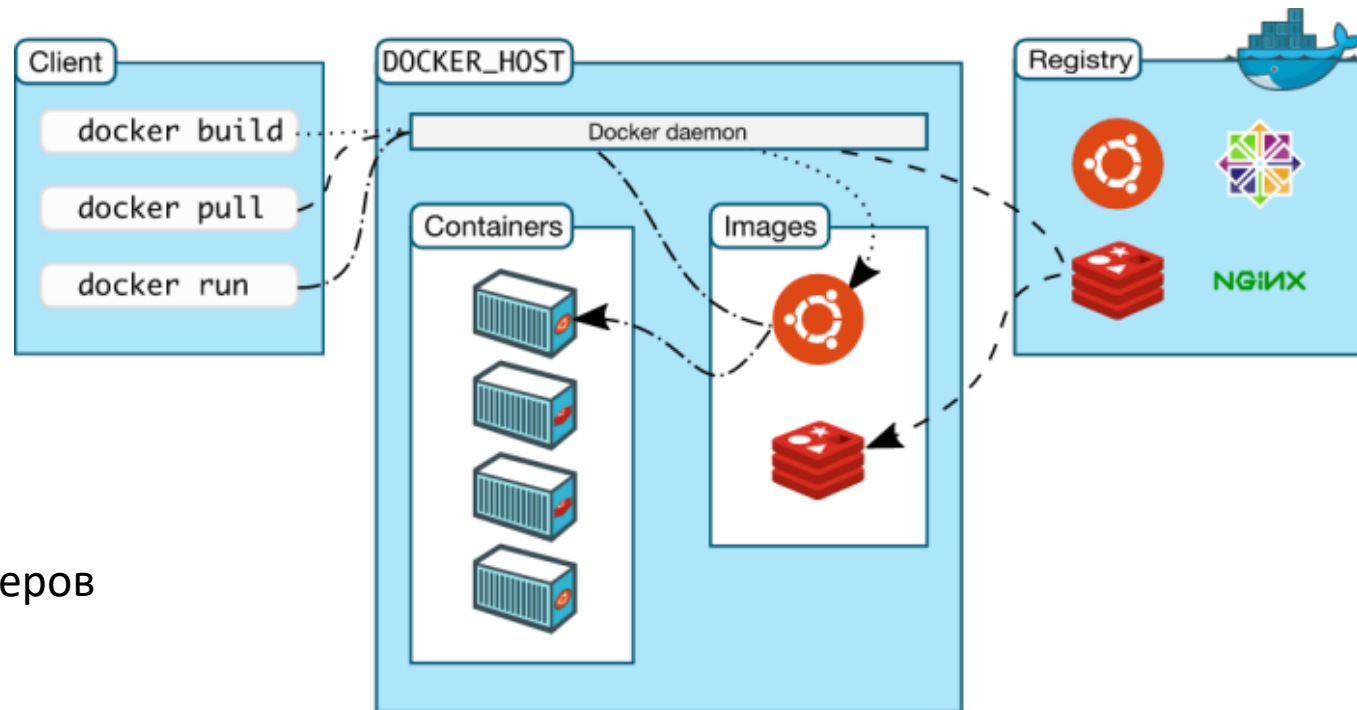
    - name: copy index.html
      template: src=templates/index.html.j2 dest=/usr/share/nginx/html/index.html
      mode=0644

    - name: restart nginx
      service: name=nginx state=restarted
```



# > Контейнеризация

- Docker
  - Альтернатива виртуальным машинам
  - Использует драйверы, функции ОС
  - Быстрота развертывания
  - Большое количество стандартных контейнеров
- Docker-compose
  - Управление большим количеством контейнеров
  - Возможность управлять перезагрузкой, ресурсами
- Kubernetes

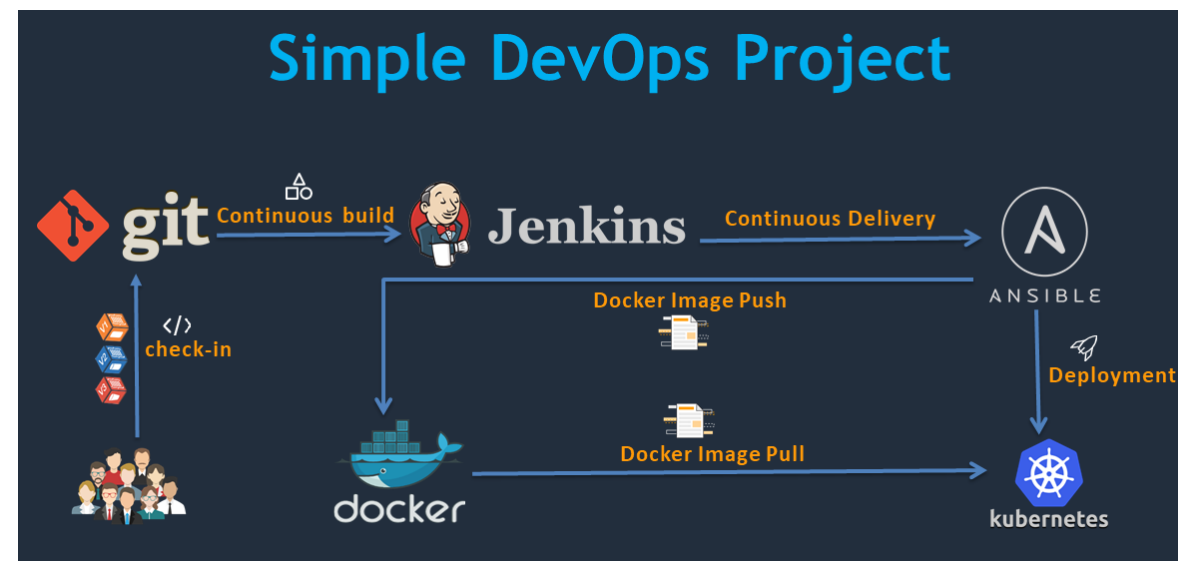
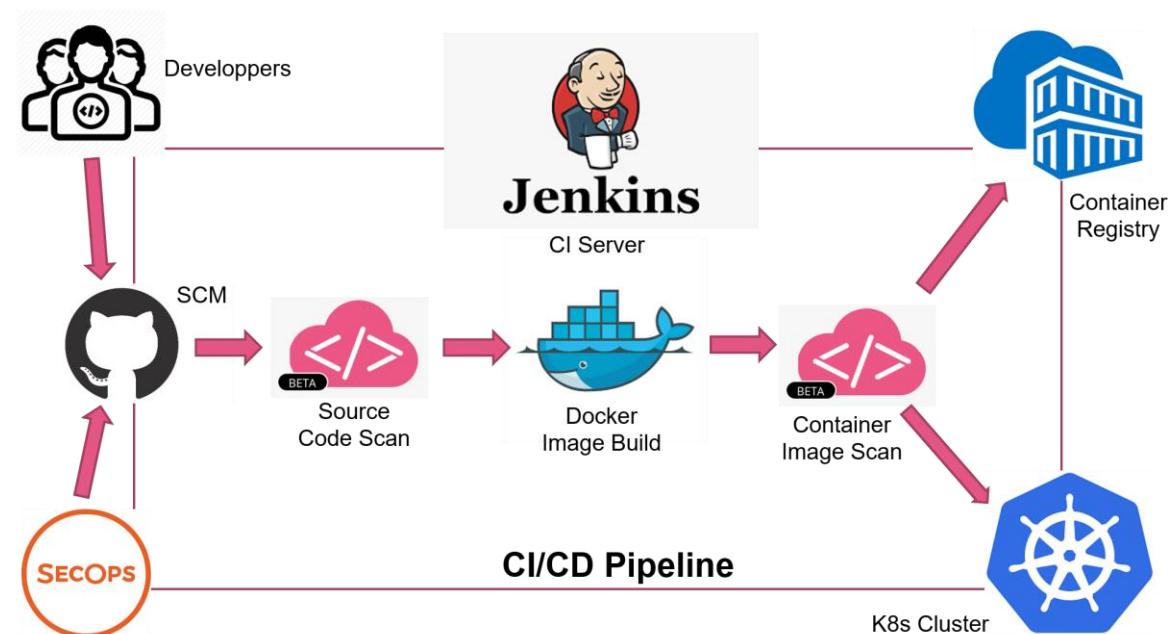


# > Jenkins

обеспечивает непрерывную интеграцию

```
pipeline {
  agent none
  stages{
    stage('Run scripts'){
      parallel{
        stage('First script'){
          agent any
          options{
            retry ( 2 )
          }
          steps{
            bat "C:\\Windows\\test.bat"
          }
        }
        stage('Second script'){
          agent any
          options{
            retry ( 2 )
          }
          steps{
            sh 'python3 C:\\Windows\\test.py'
          }
        }
      }
    }
  }
  post {
    failure {
      echo 'Some kind of process, if everything is broken ...'
    }
  }
}
```

<https://www.jenkins.io/doc/book/pipeline/syntax/>

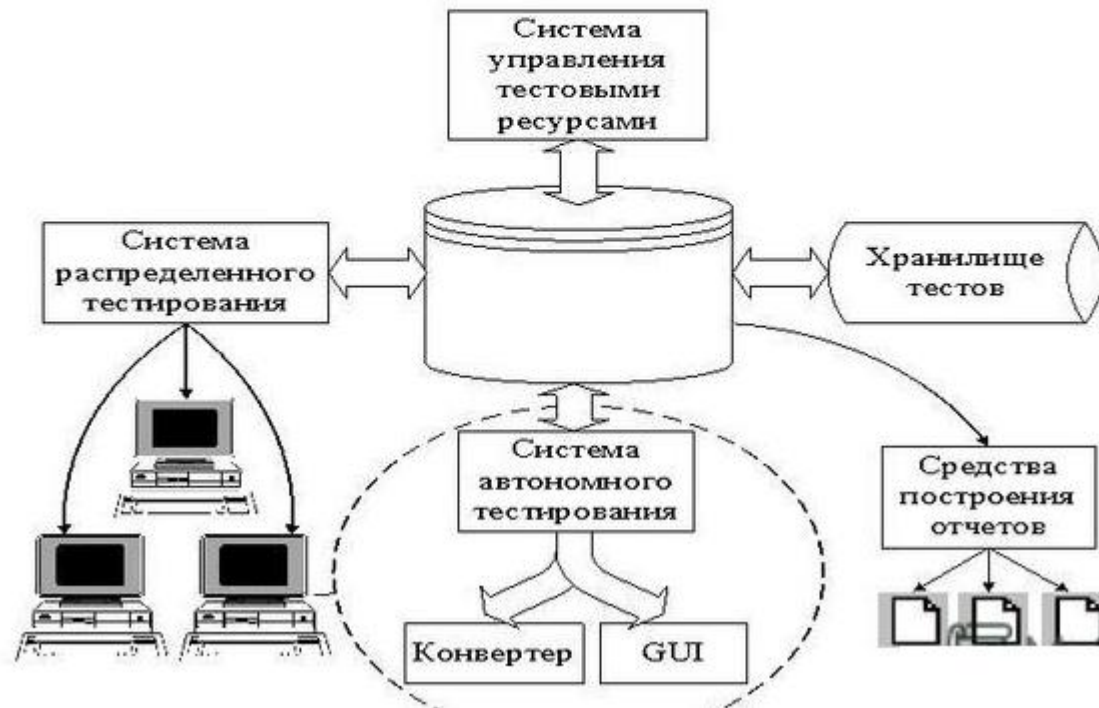


## > Системы автоматизации тестирования

- pytest
- JUnit

### Основные компоненты системы автоматизированного тестирования

Основные компоненты системы автоматизированного тестирования



# Специфика для BigData и ML ("MLOps")

## > Управление окружениями

- pip, pyenv, venv, virtualenv, poetry, anaconda

```
ychernyshov@bruteforce:~/git/test$ ls -la
total 8
drwxr-xr-x  2 ychernyshov cyberthymus 4096 Jul 14 08:57 .
drwxr-xr-x 17 ychernyshov cyberthymus 4096 Jul 14 08:57 ..
ychernyshov@bruteforce:~/git/test$ python3 -m venv venv
ychernyshov@bruteforce:~/git/test$ ls -la
total 12
drwxr-xr-x  3 ychernyshov cyberthymus 4096 Jul 14 09:00 .
drwxr-xr-x 17 ychernyshov cyberthymus 4096 Jul 14 08:57 ..
drwxr-xr-x  6 ychernyshov cyberthymus 4096 Jul 14 09:00 venv
ychernyshov@bruteforce:~/git/test$ source ./venv/bin/activate
(venv) ychernyshov@bruteforce:~/git/test$ pip freeze
(venv) ychernyshov@bruteforce:~/git/test$ pip install numpy
Collecting numpy
  Using cached numpy-1.21.0-cp39-cp39-manylinux_2_12_x86_64.manylinux2010_x86_64.whl (15.7 MB)
Installing collected packages: numpy
Successfully installed numpy-1.21.0
(venv) ychernyshov@bruteforce:~/git/test$ pip freeze
numpy==1.21.0
(venv) ychernyshov@bruteforce:~/git/test$
```

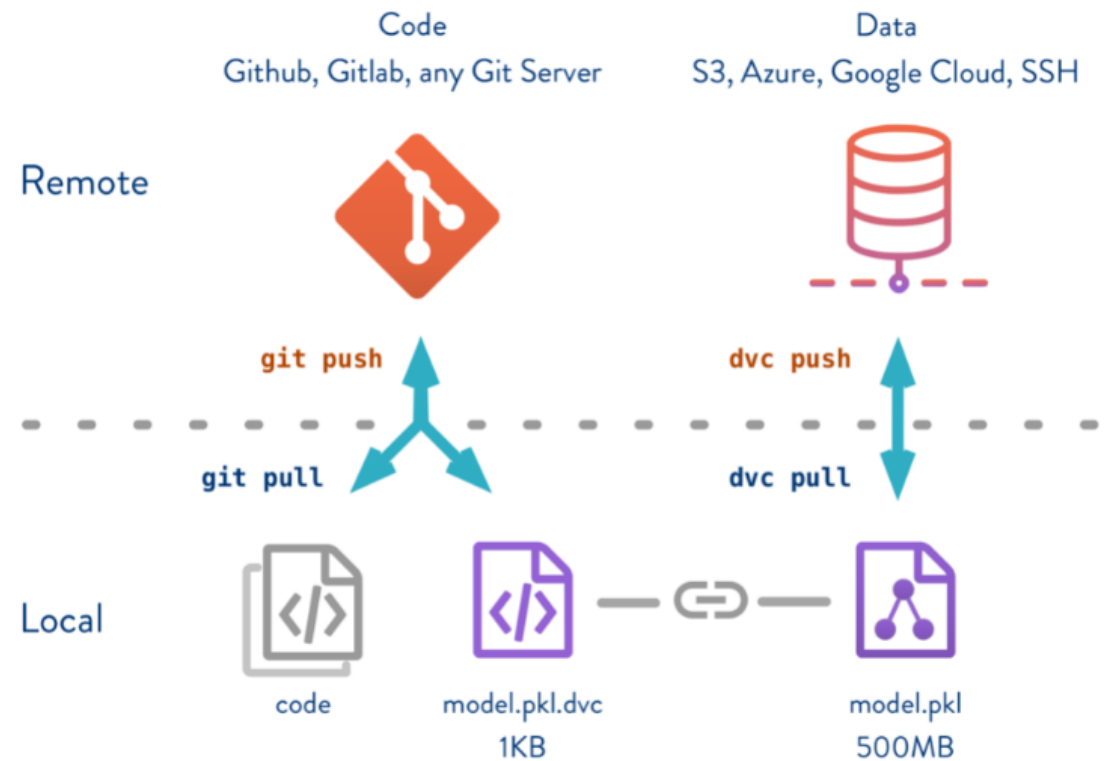
## > Управление данными

- dvc (Data Version Control, <https://dvc.org/>)
- Работает поверх git, имеет схожую структуру

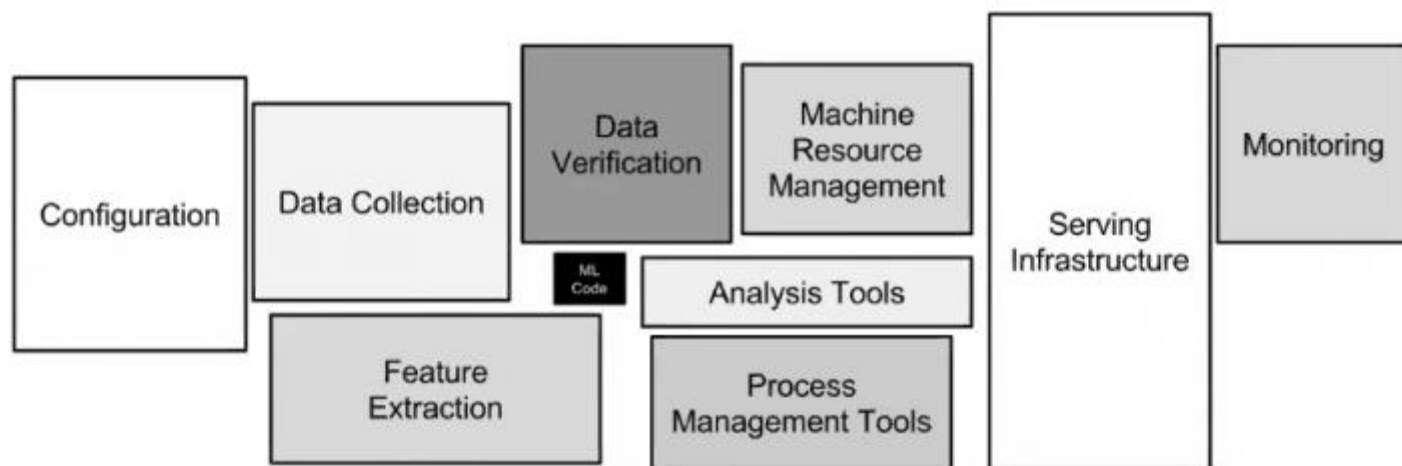
**dvc init**

**commit dvc init**

**git commit -m "Initialize DVC"**



## > Архитектура ML



Все это вместе образует **Систему**.

**Архитектура** - это устройство таких систем.

["Hidden Technical Debt in Machine Learning Systems" paper](#)

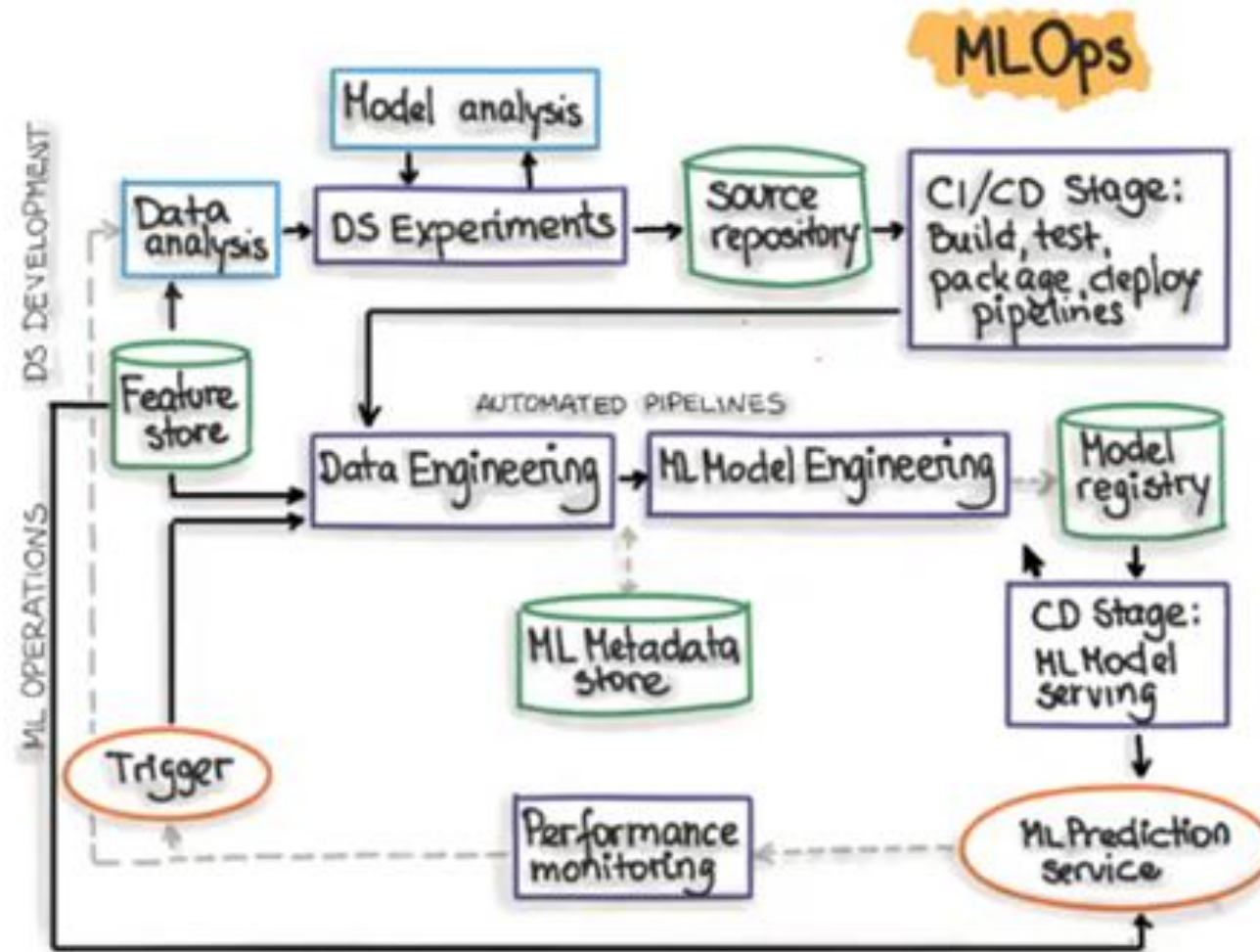


## > Пример репозитория пайплайна

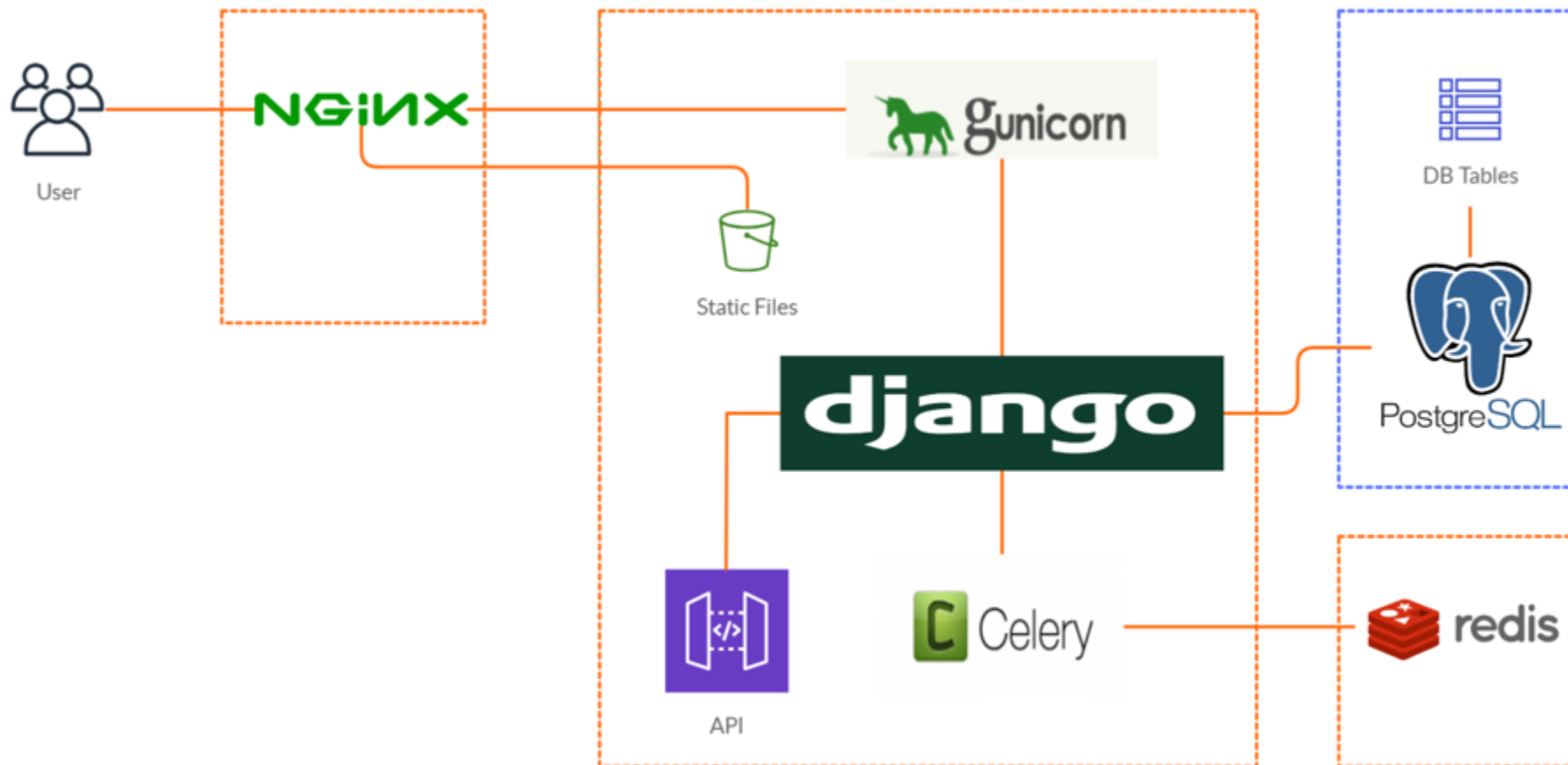
```
.  
├── Dockerfile  
├── Pipfile  
├── Pipfile.lock  
├── config.yaml  
├── scripts  
│   ├── get_data.py  
│   ├── generate_features.py  
│   ├── split_train_val.py  
│   ├── train.py  
│   ├── validate.py  
│   └── save_results.py  
├── run_all.sh  
└── README.md
```

```
1 | ~ bash ./run_all.sh
```

## ➤ Этапы пайплайна (ml-ops.org)





## > Пример архитектуры информационной системы




## > Ссылки

- Ф. Брукс «Мифический человеко-месяц»
- <https://www.freecodecamp.org/news/a-quick-introduction-to-clean-architecture-990c014448d2/>
- «Machine learning systems» Jeff Smith
- Hidden technical debts in ML <https://papers.nips.cc/paper/2015/file/86df7dcfd896fcdf2674f757a2463eba-Paper.pdf>
- DVC <https://habr.com/ru/company/raiffeisenbank/blog/461803/>
- ml-ops.org



**Спасибо за внимание!**  
Вопросы?



Чернышов Юрий, [ychernyshov@ussc.ru](mailto:ychernyshov@ussc.ru), @yuchernyshov  
Астафьева Анна, @astafevaanny

