

АНАЛИЗ И РАЗРАБОТКА РАСПРЕДЕЛЁННОЙ АРХИТЕКТУРЫ ЭКСПЛУАТАЦИИ НЕЙРОННЫХ СЕТЕЙ

Дипломная работа

Ларин Егор Сергеевич

Белорусский государственный университет ФПМИ, КТС, 4 курс руководитель: старший преподаватель Шолтанюк С. В.

Минск, 2024

Введение



- В последние годы микросервисная архитектура значительно приобретает популярность в области разработки программного обеспечения.
- Использование микросервисной помогает решить вопросы масштабируемости.
- Внедрение микросервисов порождает другие проблемы, которые необходимо решать с помощью современных инструментов.

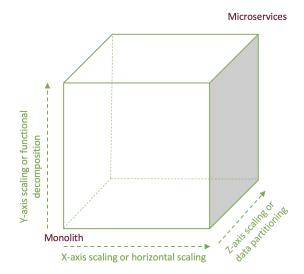
Микросервисы



- 1. Нет формального определения.
- Одним из подходов является проверка соответствия критериям масштабируемости.
- 3. Низкая связность.
- 4. Зачастую начинают использовать слишком рано.

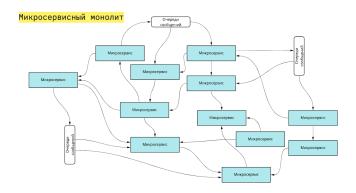
Куб масштабируемости





Распределенный монолит





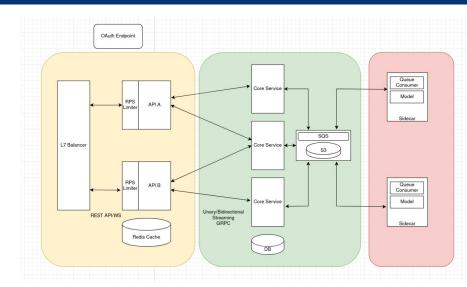
Сервис для генерации изображений



- Gateway.
- Internal Backend.
- Sidecar.

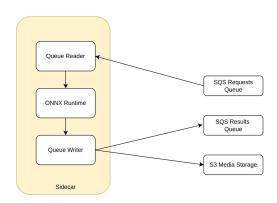
Схема архитектуры





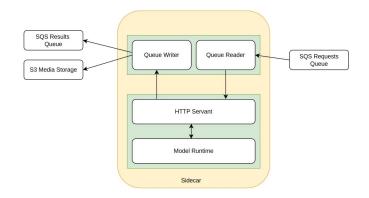
Native Sidecar





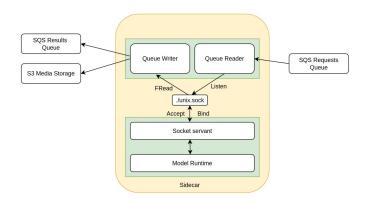
HTTP Sidecar





UDS Sidecar





Заключение



- В ходе работы было рассмотрено понятие микросервисной архитектуры и произведен обзор имеющихся средств и методологий разработки, применяющихся для коммуникации веб-сервисов.
- Результатом работы стала разработка программного обеспечения для генерации изображений с помощью нейронной сети с сетевым интерфейсом.
- Рассмотрены подходы общения разных процессов и проанализированы достоинства и недостатки каждого из методов.

Источники



- 1. Микросервисы: паттерны разработки и рефакторинга / Крис Ричардсон. Санкт-Петербург [и др.] : Питер, Прогресс книга, 2020. 542 с. (Библиотека программиста).
- 2. Высоконагруженные приложения: программирование, масштабирование, поддержка: [перевод с английского] / Мартин Клеппман. Санкт-Петербург [и др.] : Питер, Прогресс книга, 2018. 637 с. (Бестселлеры O'Reilly).
- Marek Bolanowski, Kamil Zak, Andrzej Paszkiewicz, Maria Ganzha, Marcin Paprzycki, Piotr Sowinski, Ignacio Lacalle, and Carlos E. Palau. Eficiency of REST and gRPC Realizing Communication Tasks in Microservice-Based Ecosystems. IOS Press, September 2022..