Serverless chatbot in Java

Как пройдет доклад?

- Рассмотрим темы в стиле проблемы и решения или вопроса и ответа
- Вопросы можно задавать в любой момент
- Задеплоим чатбота и потестируем на каждом этапе

Telegram

Spring **Cloud**

Amazon
Lambda

nickolay.laptev@gmail.com

https://laptev.co/

IT Home

Кто я?

- Инженер, архитектор, отец
- Помогаю бизнесу решать сложные технические и организационные проблемы
  https://laptev.co/

В чем проблема?

Появилась идея сэкономить время и собрать в одном месте все развлечения для детей в Санкт-Петербурге.

Нужен мобильный интерфейс для общения пользователя с системой.

nickolay.laptev@gmail.com

https://laptev.co/

Какие рассматривались решения?

1. Мобильное приложение
2. Веб сайт
3. Чат бот

Выбор пал на чат бот.
Быстрая возможность проверить идею.
Не требуется трата времени на интерфейс, этой задачей занимается мессенджер.

nickolay.laptev@gmail.com
https://laptev.co/

Какой мессенджер выбрать?

1. Facebook Messenger.
   Не популярен. Отличная платформа для чатботов.
2. Whatsapp
   Самый популярный, но платформа для чатботов заточена под крупный бизнес и не выглядит разумной.
   - Длительный процесс верификации
   - Работаешь через API партнеров, доступ к API whatsapp есть только у партнеров.
   - Оплачиваешь вызовы API партнеров для отправки сообщений (например 1000 сообщений за $50 в Twillio).

3. Telegram
   Популярен. Отличная платформа для чатботов.

nickolay.laptev@gmail.com
https://laptev.co/

Где запускать чатбота?

Необходимо проверить гипотезу быстро и с минимальными затратами.

Amazon
Lambda

Как быть с cold start?
Не критично для чат бота. Человек готов ждать ответа какое-то время, молниеносный ответ иногда может напугать (подробнее).
Популярный чат бот находится большую часть времени в горячем режиме.

nickolay.laptev@gmail.com

На каком языке написать бота?

- Важна скорость реализации прототипа для проверки идеи
- AWS Lambda поддерживает определенный набор языков. Время cold start для них сопоставимо.

IT Home
nickolay.laptev@gmail.com
https://laptev.co/

Plain old Java или фреймворки?

- Использовать Spring в AWS Lambda не эффективно
  Огромное время cold start с большим потреблением RAM.
  20 секунд cold start с 2GB RAM – это реальность для Hello World приложения со Spring.
- Spring рекламирует проект Cloud Function как оптимизированный под Serverless решения.
- Я не нашел бенчмарков, которые бы опровергали это маркетинговое сообщение

Решил попробовать Spring Cloud Function

Spring **Cloud**

nickolay.laptev@gmail.com

https://laptev.co/

Какой SDK выбрать для работы с Telegram?

- Нет официального SDK для Java
- Официальный HTTP API
- Популярен сторонний SDK [rubenlagus/TelegramBots]

Чтобы сэкономить время выбрал сторонний SDK.

Telegram

nickolay.laptev@gmail.com

https://laptev.co/

Чатбот для сегодняшней сессии

Исходный код https://github.com/laptevn/lambda-telegram-chatbot

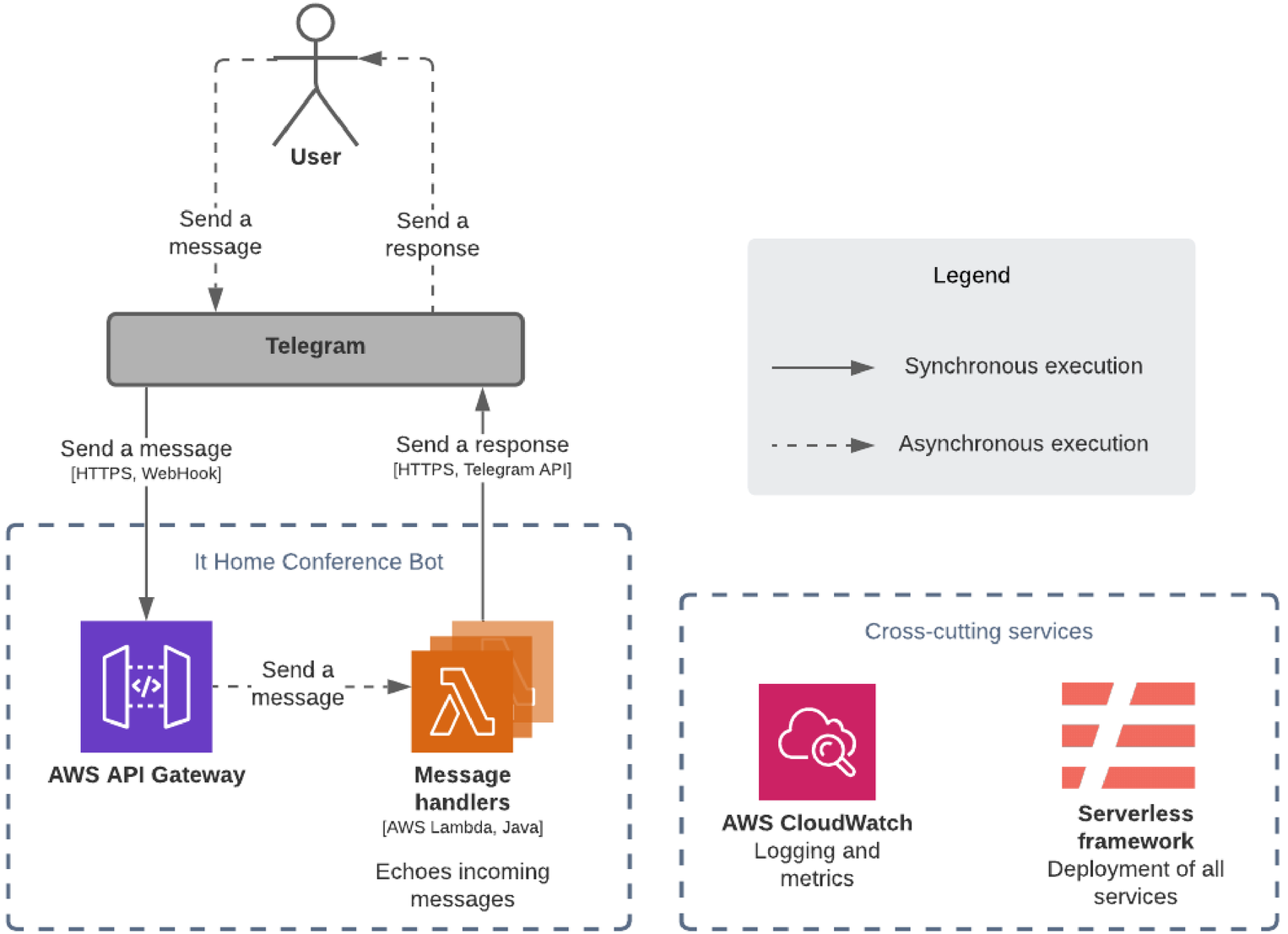
Чатбот https://t.me/it_home_conference_bot

IT Home

nickolay.laptev@gmail.com

https://laptev.co/

Архитектура решения



nickolay.laptev@gmail.com
https://laptev.co/

IT Home

Приступил к разработке и начались проблемы

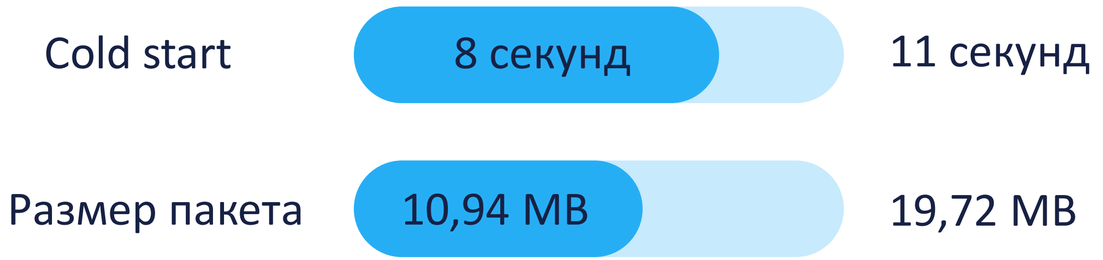
nickolay.laptev@gmail.com

https://laptev.co/

Spring Cloud Function

- Добавляет 8+ секунд к cold start (1GB RAM).
  Зависит от размера объектной модели.
- Сложно заставить его работать.
  Устаревшие и некорректные мануалы, включая официальные.
- Это Spring Cloud проект, поэтому проблемы с интеграцией (например Micrometer и AWS CloudWatch).
- Дает лишь Dependency Injection.

Отказываемся от него в пользу Plain Old Java.

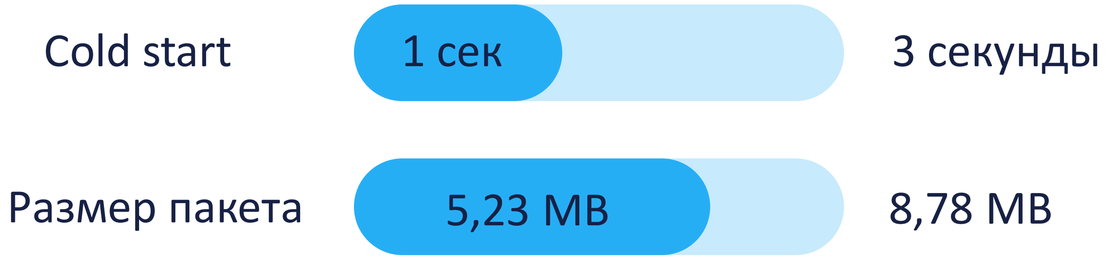


nickolay.laptev@gmail.com
https://laptev.co/

Telegram Java SDK

- Много ненужных зависимостей.
- Добавляет до 1 секунды к cold start (1GB RAM).
- Нелучшая архитектура.
  В одном JAR много несвязанного функционала.
- Нелучший саппорт.

Отказываемся от него в пользу официального HTTP API.



IT Home
nickolay.laptev@gmail.com
https://laptev.co/

Telegram HTTP API

- Непредсказуемо большой latency у отправки сообщений.
  Зачастую в диапазоне 100-400 миллисекунд независимо от размера текстового сообщения и его типа.
- Отправка сообщений синхронна.
  Мы будем платить пока Lambda ждет окончания отправки сообщения.

  Не нашел решения. Занял Lambda полезной активностью пока жду Telegram.

IT Home

C
b
c

nickolay.laptev@gmail.com

https://laptev.co/

Plain Old Java

- **Cold start Java с одной лишь записью в лог – 30 мс (1GB RAM)**
- HTTP Client из Java 11 добавляет 1.5 секунды к cold start.
- JSON сериализация добавляет 130 мс к cold start и 1.8 MB к размеру пакета.
- Cold start для базовой Java - отличный, но фреймворки и некоторые API из JDK сильно на него влияют.

AWS создает альтернативы проблемным API в JDK.

nickolay.laptev@gmail.com

Выводы

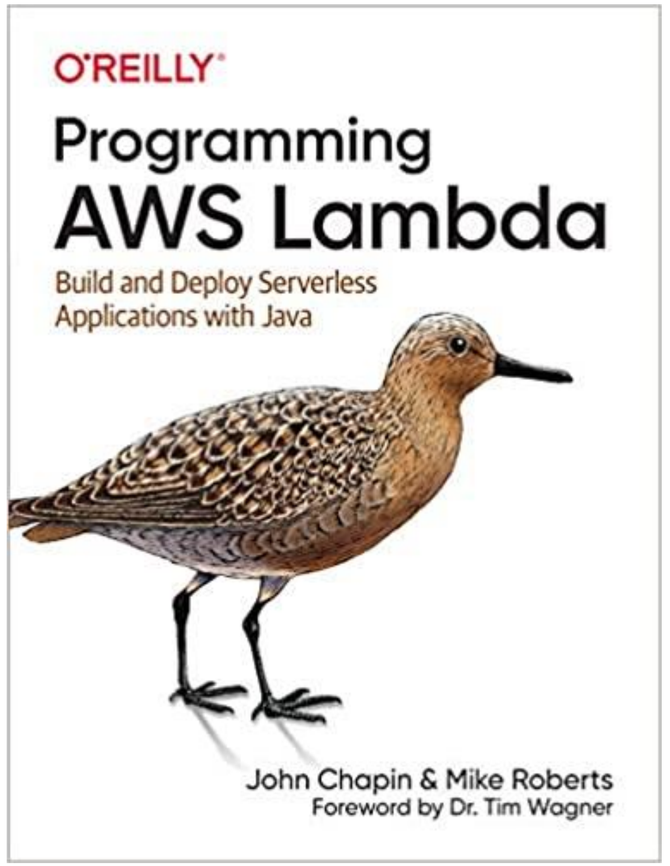
- Whatsapp чатботы очень затратные.
- Serverless решения позволяют минимизировать траты на бота.
- Java и вся инфраструктура фреймворков вокруг нее разработаны для Server мира, не для Serverless.
- Spring Cloud Function значительно улучшил cold start по сравнению со Spring Boot, но он все равно слишком долгий.
- Непредсказуемо большой latency у отправки сообщений в Telegram.
- Отправка сообщений в Telegram синхронна.

nickolay.laptev@gmail.com

https://laptev.co/

Дальнейшее изучение

Если интересна идея писать ботов на Java
https://www.amazon.com/Programming-AWS-Lambda-Serverless-Applications/dp/149204105X



IT Home
C b c
nickolay.laptev@gmail.com
https://laptev.co/

Финальные вопросы