

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.383.2+004.4`6

Новик
Игорь Петрович

Оркестрация Docker контейнеров:
концепция Zero Downtime Deployment

АВТОРЕФЕРАТ

диссертации на соискание академической степени
магистра технических наук

по специальности 1-40 80 05 – Математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

Научный руководитель
Лапицкая Н.В.
к.т.н., доцент

Минск 2020

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью диссертационной работы является улучшение характеристик процесса развертывания приложений, снижение ресурсных и временных издержек при развертывании приложений, использующих микросервисную архитектуру.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Проанализировать текущие проблемы развертывания новых версий приложений при микросервисной архитектуре.

2. Разработать архитектуру программной системы автоматизированного развертывания новых версий приложения реализующую концепцию Zero Downtime Deployment на основе оркестрации Docker контейнеров.

3. Провести экспериментальные исследования разработанной системы в сравнении с традиционными методами развертывания.

4. Разработать методы минимизации негативных последствий при доставке новых версий приложения при использовании концепции Zero Downtime Deployment.

5. Реализовать алгоритм предсказания пользовательской нагрузки для планирования оптимального времени развертывания.

Объектом исследования является концепция Zero Downtime Deployment и технология контейнеризации. Идея концепции заключается в достижении минимального времени недоступности сервисов при развертывании новой версии приложения.

Предметом исследования является процесс развертывания приложений, реализованных с использованием микросервисного подхода.

Основной *гипотезой*, положенной в основу диссертационной работы, является возможность бесшовного обновления приложения основанного на микросервисной архитектуре, не ведущих к потере реквестов в режиме реального времени. А также потенциальные возможности минимизации негативных эффектов возникающих при развертывании новых версий приложения путем горизонтального масштабирования экземпляров приложения, а также предварительного планирования периодов минимальной загруженности приложения, что приведет к возможности построения систем непрерывного развертывания соответствующей системам критичным к качеству предоставляемых сервисов (QoS).

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Работа выполнялась в соответствии с научно-техническим заданием и планом работ кафедры «Программное обеспечение информационных технологий» по теме «Разработка моделей, методов, алгоритмов, повышающих показатели проектирования, внедрения и эксплуатации программных средств для перспективных платформ обработки информации, решения интеллектуальных задач, работы с большими массивами данных и внедрение в современные обучающие

комплексы» (ГБ № 16-2004, № ГР 20163588, научный руководитель НИР – Н. В. Лапицкая.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя Н. В. Лапицкой, заключается в формулировке целей и задач исследований, а также наставлении по методикам проведения исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на 55-й юбилейной научной конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2019); Международной научно-технической конференции «Современные средства связи» (Минск, Беларусь, 2019).

Опубликованность результатов диссертации

По теме диссертации опубликовано 2 печатных работы в сборниках трудов и материалов международных конференций.

Список публикаций соискателя

1-А. Новик, И.П. Оркестрация Docker контейнеров: концепция Zero Downtime Deployment / И.П. Новик // Материалы международной научной конференции Информационные технологии и системы – 2019.– с. 72–73.

2-А. Новик, И.П. Концепция Zero Downtime Deployment на основе оркестрации Docker контейнеров / И.П. Новик // Современные средства связи. / БГАС; – Минск, 2019. – с. 139–140.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора и приложений.

Общий объем работы составляет 73 страниц, из которых основного текста – 62 страницы, 43 рисунка на 33 страницах, 3 таблицы на 2 страницах, список использованных источников из 49 наименований на 3 страницах и 5 приложений на 7 страницах.

ВВЕДЕНИЕ

В современном мире при разработке программного обеспечения все чаще и чаще возникает ситуация, в которой требования к системе меняются уже на конечных этапах разработки. К тому же, наблюдается постоянный прирост функциональности в уже существующих системах. Помимо этого, к современному ПО предъявляются крайне высокие требования по производительности и отказоустойчивости. Классическая монолитная архитектура приложений не отвечала требованиям современного мира, и на смену ей пришел сервисно-ориентированный подход. Такой подход имеет значительные преимущества по сравнению с монолитными приложениями, такие, например, как лучшая масштабируемость, меньшая связанность между модулями, лучший контроль на этапах разработки, тестирования и развертывания [1].

Системы, построенные с помощью сервисно-ориентированной архитектуры, также являются распределенными. Взаимодействие между сервисами осуществляется с использованием различных протоколов. Микросервисная архитектура является одним из видов сервисно-ориентированной архитектуры. Это означает что основной архитектурный компонент, из которого строится система, это – сервис. К сожалению, ценой указанных выше преимуществ является значительное усложнение системы. Требуется поддерживать контракты протоколов, бороться с недоступностью сервисов, управлять распределенными транзакциями. Также к этим проблемам относится необходимость развертывания большого количества сервисов в различных окружениях на всех этапах разработки.

В современной разработке встречается множество подходов развертывания, которые традиционно используются в индустрии. Экспертиза в данной сфере нарабатывалась годами и привела к формированию практик непрерывной интеграции и непрерывного развертывания (CI/CD). Признанными экспертами в данном направлении являются Humble Jez и Farley David, они заложили основные понятия в концепцию Continuous Integration (непрерывная интеграция) и Continuous Deployment (непрерывная доставка) [2]. В рамках данной концепции, так же рассматриваются вопросы доставки новых версий приложения с минимальными издержками. В диссертационной работе подходы рассмотрены в порядке эволюции технологий развертывания приложений: от прямого развертывания на сервере приложений до достижения zero downtime развертывания на основе оркестрации контейнеров.

КРАТКОЕ СОДЕРЖАНИЕ

В первой главе диссертации описывается понятие микросервисной архитектуры, стратегий масштабируемости приложений. Также рассказывается о проблемах, возникающих во время развертывания микросервисных приложений и о путях решения этих проблем.

Во второй главе диссертации описываются требования к инфраструктуре, при разработке систем, использующих микросервисный подход. Освещаются вопросы, связанные с требованием быстрого выделения ресурсов, на которых будет разворачиваться микросервисная система, а также вопросы, связанные со сборкой и проверкой разрабатываемых сервисов. Необходимость непрерывной интеграции и непрерывной поставки. Рассматривается необходимость быстрого развертывания сервисов. Также в этом разделе идет речь о виртуализации и автоматизации инфраструктуры. Рассматривается как традиционная виртуализация, так и технология контейнеризации, в части применимости этих технологий в решении проблемы быстрого развертывания микросервисов. Рассматривается концепция Zero Downtime Deployment и возможности её применения.

В третьей главе диссертации описываются механизмы оркестрации контейнеров, как способ оптимизации микросервисной инфраструктуры. Рассматриваются вопросы, связанные с управлением кластером, а также технологии, которые обеспечивают возможность централизованного управления развертыванием множества сервисов одновременно. Разрабатываются алгоритмы развертывания обеспечивающих реализацию концепции Zero Downtime Deployment на основе оркестрации Docker контейнеров. Путем эксперимента сравниваются классические методологии развертывания с реализуемой концепцией. Предлагаются пути минимизации негативных эффектов при использовании концепции. Рассматривается возможность планирования развертывания на основе предсказания пользовательской нагрузки с применением модели ARIMA.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

В дипломном проекте проведен анализ современного состояния в области процессов непрерывной интеграции и поставки программного обеспечения, рассмотрены: модели развертывания.

Таким образом в ходе данной работы мы исследовали возможность обрабатывать все пользовательские запросы на протяжении развертывания новой версии приложения путем использования оркестрации Docker контейнеров в Kubernetes кластере, реализовав концепцию Zero Downtime Deployment. Но при этом мы выявили, что при развёртывания новой версий приложения у нас все еще есть эффект «просадки» скорости обработки запросов, после выкатывания новой версии, обусловленный переключением на контекст нового приложения. Дальнейшие усилия для улучшения концепции стоит направить на нивелирование данного эффекта.

Одним из способов может быть рассмотрено включение в процесс развертывания процесс «прогрева приложения» (Warm Startup). В рамках данного процесса предполагается «прогреть» контекст новой версии приложения, что должно положительно сказаться на времени обработки запросов новой версией приложения. Но данное решение может быть применено только с рассмотрением каждого приложения по отдельности.

Также было рассмотрена возможность снизить пиковые скачки с помощью увеличения числа обрабатывающих узлов. Однако данное решение имеет весомый минус в виде ресурсных расходов на обслуживание дополнительных экземпляров приложения.

Еще одним способом минимизации пиковых скачков обработки запросов во время развертывания новой версии была рассмотрена модель построения прогноза пользовательской активности на основании ARIMA модели.

Таким образом в работе были рассмотрены и даны советы к применению практик непрерывного развертывания, которые позволяют минимизировать негативные эффекты при доставке новых версий приложения до клиентов.

Рекомендации по практическому использованию результатов

1. Полученные результаты формируют теоретическую и практическую базу для разработки компьютерных систем для решения задач непрерывного развертывания новых версий микросервисных приложений, функционирующих в режиме реального времени. Они могут быть использованы для модернизации и дальнейшего развития существующих систем развертывания, построенных на технологии Docker контейнеров.

2. Разработанные методы и алгоритмы анализа прогнозирования могут применяться в автоматизированных системах развертывания для определения качества оптимального времени развертывания, с минимальной

пользовательской нагрузкой, тем самым минимизируя негативные эффекты при старте новой версии приложения.

Библиотека БГУИР