**Кочовски Петар Разработка технологии обеспечения высокоскоростных вычислений на основе облачной инфраструктуры**

**ОГЛАВЛЕНИЕ ДИССЕРТАЦИИ**

**кандидат наук Кочовски Петар**

**Введение**

**1. Обзор технологий обеспечения высокоскоростных вычисленный на основе облачной инфраструктуры**

**1.1 Обзор парадигм облачных вычислений**

**1.1.1 Хронология вычислений от вычислительной машины до облачных вычислений**

**1.1.2 Вычислительные парадигмы**

**1.1.3 Типы облачных вычислений**

**1.2 Виртуализация вычислительных ресурсов**

**1.2.1 Виртуализация на основе виртуальных машин**

**1.2.2 Виртуализация на основе контейнеров**

**1.2.3 Атрибуты, определяющие качество вычисления**

**1.3 Методы принятия решения и механизмы для выбора оптимальных облачных инфраструктур**

**2. Методика выбора оптимальной облачной инфраструктуры**

**2.1 Вероятностная модель и Марковские процессы**

**2.2 Расчет вероятностей модели вознаграждений**

**2.3 Процесс принятия решений**

**3. Методика сокращения вычислительной сложности**

**3.1 Определение классов эквивалентности**

**3.2 Сокращения вычислительной сложности на основе классов эквивалентности**

**4. Методика оценки корректности результата на основе темпоральной логики**

**4.1 Проверка на модели**

**4.2 Свойства, формулируемые в темпоральной логике PCTL**

**5. Формализованный метод для автоматизации процесса развертывания приложений в облачных инфраструктурах**

**5.1 Концепция процесса автоматического ранжирования инфраструктур и развертывания приложений**

**5.2 Архитектура программного комплекса**

**5.3 Механизм принятия решений о развертывании**

**5.4 Автоматическое развертывание приложений**

**6. Экспериментальная оценка предложенных методик**

**6.1 Экспериментальные сценарии**

**6.1.1 Хранилище документов - &laquo;File Upload&raquo;**

**6.1.2 WebRTC видеоконференции**

**6.1.3 Базы данных**

**6.2 Экспериментальная среда**

**6.3 Экспериментальные результаты**

**6.4 Сравнения и соотнесения**

**Заключение**

**Список работ, опубликованных автором по теме диссертации**

**Список литературы**

**Список рисунков**

**Рисунок 1 Хронология вычислений от вычислительной машины до облачных вычислений**

**Рисунок 2 Вычислительные парадигмы**

**Рисунок 3 Традиционная вычислительная архитектура (без виртуализации)**

**Рисунок 4 Виртуализация на основе виртуальных машин**

**Рисунок 5 Виртуализация на основе контейнеров**

**Рисунок 6 Архитектура Kubemetes кластера**

**Рисунок 7 Пример вероятностной модели, состоящей из облачных инфраструктур**

**развертывания в одном классе эквивалентности**

**Рисунок 8 Пример классификации по эквивалентности**

**Рисунок 9 Концепция процесса принятия решений для развертывания микросервисов**

**Рисунок 10 Системная архитектура высокого уровня**

**Рисунок 11 Диаграмма компонентов**

**Рисунок 12 YAML инструкция по развертыванию приложения на Kubemetes кластере**

**Рисунок 13 Первоначальная вероятностная модель**

**Рисунок 14 Вероятностная модель использующая метод классификации эквивалентности**

**Рисунок 15 МАИ иерархия нефункциональных требований и доступных инфраструктур развертывания для сценариев 1 и**

**Список таблиц**

**Таблица 1 Сравнение виртуализации на основе контейнеров и виртуальных машин**

**Таблица 2 Нефункциональные атрибуты, используемые в этой работе**

**Таблица 3 Инфраструктуры &laquo;Edge-to-Cloud&raquo;, используемые в экспериментальном процессе ... 67 Таблица 4 Средняя сетевая производительность инфраструктур, используемых в**

**экспериментальном процессе**

**Таблица 5 Мягкие ограничения для экспериментального тестирования**

**Таблица 6 Результаты развертывания микросервисов для хранилища документов**

**Таблица 7 Результаты развертывания микросервисов для WebRTC видеоконференции**

**Таблица 8 Результаты развертывания микросервисов для базы данных**

**Таблица 9 Проверка результатов развертывания микросервисов для хранилища документов на**

**основе жестких и мягких ограничений**

**Таблица 10 Проверка результатов развертывания микросервисов для WebRTC**

**видеоконференции на основе жестких и мягких ограничений**

**Таблица 11 Проверка результатов развертывания микросервисов для базы данных на основе**

**жестких и мягких ограничений**

**Таблица 12 Сравнение результатов ранжирования методов по количеству нарушений порогов к**

**качеству обслуживания для сценария хранилища документов**

**Таблица 13 Сравнение результатов ранжирования методов по количеству нарушений порогов к**

**качеству обслуживания для сценария WebRTC видеоконференции**

**Таблица 14 Сравнение результатов ранжирования методов по количеству нарушений порогов к качеству обслуживания для сценария базы данных**

**Заключение**

С появлением IoT сред вычислений на периферии, в тумане и в облаке, концепция распределенных систем непрерывно развивается. Однако это также создает некоторые новые проблемы оптимизации, когда приложения, основанные на контейнерах, должны быть организованы на оптимальной инфраструктуре развертывания. Различные IoT-приложения, например, разработанные в ходе исследовательского и инновационного проекта DECENTER [87], требуют развертывания программных компонентов на нескольких компьютерных уровнях, начиная с реальных датчиков и видеокамер, которые производят потоки данных вплоть до центров обработки данных, где выполняется комплексный анализ данных.

Нефункциональные требования - это основные требования к качеству от разработчиков программного обеспечения на этапе развертывания приложения. В этой работе представлена технология, которая может использоваться для цели выбора оптимальной инфраструктуры развертывания приложений и интегрирована в существующие инструменты разработки программного обеспечения, рабочую среду *SWITCH-Software Workbench for Interactive, Time Critical and Highly self-adaptive Cloud applications* (*Программное обеспечение для разработки интерактивных, критичных ко времени и самоадаптирующихся облачных приложений,* ID проекта: 643963) и проект *DECENTER-Decentralised technologies for orchestrated Cloud-to-Edge intelligence (Децентрализованные технологии для управления интеллектуальными вычислениями от облака до периферии,* ID проекта: 815141). Кроме того, предлагаемый программный комплекс был использован в рамках международного проекта *BI-RU/16-18-043 Internet of Things and Cloud computing as support for the development of new smart approaches in the construction sector (Интернет вещей и облачные вычисления как поддержка разработки новых интеллектуальных подходов в строительном секторе).* Это прокладывает путь для использования таких вероятностных методов в современных практиках разработки программного обеспечения.

В диссертационной работе поставлена и решена проблема обеспечения высокого уровня качества при вычислениях в «Edge-to-Cloud» континууме, путем удовлетворения критерия качества, заданного программным инженером для определения оптимальной вычислительной инфраструктуры. В ходе выполнения диссертационной работы получены следующие результаты:

* методика группировки инфраструктур, удовлетворяющих жестким ограничениям качества, определенным программным инженером;
* методика определения оптимальной инфраструктуры для контейнерного приложения, которая удовлетворяет мягкие ограничения, определенные программным инженером, на основе текущих измерений инфраструктуры и предшествующих знаний об использовании;
* методика проверки оптимального решения с помощью PCTL логики и оценки вероятностей возможных сценариев перераспределения приложений в будущем на основе предыдущих знаний об использовании;
* оценка эффективности предложенных методик на базе использования с различными типами приложений, с различными требованиями к качеству обслуживания;
* многоуровневая архитектура, объединяющая вышеупомянутые методики в систему, позволяющую автоматически разворачивать программные компоненты в «Edge-to-Cloud» континууме, гарантируя высокое качество обслуживания и исключение человеческих ошибок, возникающих при ручном развертывании;
* разработан программный комплекс, прошедший апробацию в рамках трех проектов, автоматизирующий процесс развертывания приложений в инфраструктурах от периферии до облаков с гарантией высокого качества обслуживания.

Результаты, полученные в процессе выполнения проектов, использующих методики данной диссертации, позволили сделать выводы о работоспособности разработанных методов. Разработанное решение принимает необходимые свойства приложений, а также требования к качеству и информацию от поставщиков инфраструктуры, сопоставляет эти свойства и выполняет расчеты, чтобы сформировать список ранжирования доступных инфраструктур развертывания. Результат, который представляет собой список ранжирования инфраструктур развертывания, проверяется с помощью логики PCTL.

В этой работе представлены три микросервиса с различными требованиями к качеству и применение разработанной технологии. Результаты тестирования показывают, что предлагаемая технология является достаточно общей для развертывания микросервисов с различными требованиями к качеству обслуживания. Таким образом, предлагаемая технология не ограничивается атрибутами нефункциональных требований, рассматриваемых в данном исследовании. Она может быть дополнительно расширена за счет реализации дополнительных нефункциональных требований для эффективного удовлетворения требований к качеству любого другого микросервиса.

Дальнейшие исследования целесообразно проводить в области расширения базы знаний информацией о минимальных требованиях к качеству различных типов приложений. В таком случае классы эквивалентности будут группировать доступные инфраструктуры по типу приложения, что позволит разработчику программного обеспечения выбрать тип приложения, а система предоставит ему оптимальную инфраструктуру для развертывания.