

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

УДК 004.915

Оков

Кирилл Александрович

Распределенные технологии обработки данных

АВТОРЕФЕРАТ

на соискание степени магистра технических наук
по специальности 1-40 80 04 «Математическое моделирование, численные
методы и комплексы программ»

Научный руководитель

Волорова Наталья Алексеевна

кандидат технических наук, доцент

Минск 2016

ВВЕДЕНИЕ

Информационный бум, сопровождавший современное общество на протяжении последних нескольких десятков лет, дал толчок для развития прикладных устройств и их доступность широким слоям населения стран мира. В соответствии с этим, количество информации, накапливаемой на носителях, а также необходимой для различного рода анализов существующей статистики, возросло экспоненциально и продолжает расти. Для корректной, а главное быстрой обработки данных, необходимы подходы, позволяющие обеспечивать устойчивую работу систем, а также применимые к данным, поступающим в режиме реального времени. В качестве решения широко используется применение так называемых распределенных технологий обработки информации, которые являются естественным путем развития архитектурных решений.

Подход к применению распределенных технологий может и не быть универсальным для всех приложений, ведь в каждом присутствует своя специфика. Однако, есть ряд требований, предъявляемых к разработчикам таких систем, которые позволяют соблюдать основные принципы, а также гарантировать преимущества распределенных систем. Кроме того, существует набор методик, которые применяются для достижения наилучшего результата при выборе данного подхода.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель работы: применение распределенного подхода к разработке архитектурных решений и реализация такой системы на базе существующего приложения.

Задачи исследования:

- Выявить основные характеристики распределенных систем, а также общие затруднения, возникаемые при их реализации;
- Описание основных методик, применяемых к таким системам для достижения лучшего результата;
- Качественная оценка улучшения балансировки нагрузки исходя из ключевых метрик;

Объектом исследования является комплекс модулей по обработке показателей эффективности рекламы.

Предмет исследования: параметры осуществления коммуникации между узлами в распределенной системе.

Актуальность темы исследования заключается в том, что показатели эффективности рекламы характеризуются все меньшим и меньшим временным интервалом, и, соответственно, общий объем таких данных возрастает. Принимая во внимание особенность компонентов обработки, необходимо выявить и качественно оценить подход для удовлетворения требований по отказоустойчивости системы, улучшая подход к обеспечению коммуникации.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В данной магистерской диссертации рассматривается актуальность и проблематика комплексной обработки данных. Для этого предлагается на основе имеющегося приложения реализовать комплекс независимых слабосвязанных модулей, т.е. провести миграцию на распределенную систему обработки. Кроме того, выявляются несколько методик по устранению узких мест системы и собственный подход к обеспечению отказоустойчивости.

Во введении рассмотрено положение дел в обработке информации в настоящем времени, сферы применения таких технологий, а также обосновывается актуальность, определяются цели и задачи магистерской диссертации.

В первой главе дается определение понятия распределенной системы, приводится классификация архитектурных решений по их созданию, выделяются ключевые характеристика распределенного подхода. Кроме того, выявляются основные заблуждения при таком подходе и сложности реализации таких систем, приведен пример классической распределенной системы.

Во второй главе показывается на примере существующего приложения процесс распределения компонентов обработки данных в соответствии с требованиями, предъявляемыми к распределенной технологии обработки данных. В связи с этим, описываются наиболее распространенные методики по созданию компонентов быстрого и масштабируемого доступа к данным с целью увеличения производительности системы.

В третьей главе описывается непосредственно исследование по выбору технологии коммуникации, их сравнительный анализ, возможности конфигурации средств коммуникации. Помимо этого, описывается

возможное решение проблемы загруженности сетевых средств сообщения между отдельными компонентами. Так, в качестве решения для улучшения балансировки нагрузки, предлагается подход по изменению размера запросов. Приводится описание создания отдельного модуля для подсчета оптимального размера запросов, а также подходящие методики для подсчета размера запроса. Получены показатели по ключевым метрикам при статическом и динамическом размера запросов.

В приложениях приводятся основные этапы перехода к распределенной технологии, сравнительные характеристики для компонентов доступа к данным, а также средств обеспечения связи между узлами, корреляция ключевых показателей при различных балансировщиках нагрузки.

ЗАКЛЮЧЕНИЕ

В данной работе приводится доказательство необходимости и важности существования распределенных систем обработки данных. Показано, что распределенные системы являются более эффективными по сравнению с централизованными системами, а также обосновывается их надежность и отказоустойчивость.

Кроме того, был рассмотрен ряд подходов, при которых достигается быстроедействие системы, несмотря на распределенность узлов. Реализация таких компонентов может носить специфический характер, однако прослеживается общий принцип.

Для решения проблемы высокой нагрузки системы был приведен набор технологий, корректное конфигурирование которых позволяет добиться улучшенных показателей по задержке отклика, а также уменьшить трафик сети. В свою очередь приведенные результаты исследования по варьированию размером запросов, а также алгоритма определения оптимального размера запроса позволяют сделать вывод о том, что адаптивное изменение общего подхода под нужды определенной задачи помогают достигнуть показателей, которые подтверждают преимущества распределенного подхода.

Результаты исследований магистерской диссертации можно применить при проектировании архитектуры распределенной системы, а также при выборе средств связи между слабо связанными компонентами. Применение материала диссертации на практике в существующем программном продукте позволяет непосредственному пользователю получать запрашиваемые данные с необходимой точностью и за корректный временной интервал.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Оков, К.А. Модель обработки данных с изменяемым размером пакета данных / К.А. Оков // 51-я научная конференция аспирантов, магистрантов и студентов по направлению 4: Компьютерные системы и сети: материалы конф., Минск, 13–17 апреля 2015 г. / БГУИР; редкол.: В. А. Прытков [и др.]. – Минск, 2015. – С. 151–152.

Библиотека БГУИР