

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

УДК 004.5+004.051+004.75

Забродец
Дарья Артуровна

Исследование существующих средств анализа производительности
распределенных систем

АВТОРЕФЕРАТ

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
Неборский С.Н.
к.т.н.

Минск 2017

КРАТКОЕ ВВЕДЕНИЕ

На сегодняшний день, в условиях расширения сфер деятельности людей и увеличения объема необходимой к анализу и использованию информации, наиболее популярной средой как обработки, так и обмена данными является World Wide Web. Ряд задач, выполняемых пользователями Web-браузеров ежедневно, требуют больших калькуляционных мощностей, требуя принимать сложные архитектурные и технические решения, позволяющие соответствовать современным бизнес требованиям и являющихся гибкими для расширения новыми возможностями.

Таким образом, наиболее популярный тип разрабатываемых систем сегодня – распределенные системы – приложения, которое используют несколько компьютеров, расположенных в разных местах, но, подключенных к сети, чтобы выполнить одну задачу или набор связанных задач.

Поле параллельных вычислений восходит к середине пятидесятих годов, когда исследовательские лаборатории приступили к разработке так называемых суперкомпьютеров с целью значительно увеличить производительность, в основном подразумевая увеличение числа операций (с плавающей запятой), которое машина способна выполнить за единицу времени. С тех пор значительные успехи в области аппаратного и программного обеспечения технологии довели возможности производительности вычислений до одного терафлопа к 1998 году.

В то время как увеличение производительности по-прежнему является движущим фактором в параллельной и распределенной обработке данных, есть также многие другие проблемы, которые необходимо решать в этой области. Достижения в области беспроводных сетей расширяют возможности различных мобильных устройств (в том числе ноутбуков, КПК или мобильных телефонов), что способствует распределению информации по своей природе; пользователи требуют немедленного доступа к источникам информации, к вычислительной мощности и средствам связи. В то время как производительность в определенном выше смысле по-прежнему является важным критерием такого рода систем, другие вопросы, в том числе правильность, надежность, безопасность, простота использования, повсеместный доступ, интеллектуальные услуги и т.д., должны быть рассмотрены уже в самом процессе разработки таких систем.

Основываясь на вышесказанном, применение средств анализа производительности распределенных систем является актуальной задачей. На сегодняшний день, доступные и принятые методы и средства анализа распределенных систем разнообразны и находятся в стадии прикладных исследований и доработок.

Диссертационная работа посвящена исследованию существующих профилирующих инструментов анализа распределенных систем, а также методов и моделей оценки производительности производственных сред; разработке улучшения к существующим методам; проработке достигнутых в области анализа распределенных систем результатов.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Цель магистерской диссертации – провести исследование существующих профилирующих инструментов анализа распределенных систем, а также методов и моделей оценки производительности производственных сред; выяснить и предложить улучшения к существующим методам; проработать достигнутые в области анализа распределённых систем результаты; предложить оценку производительности распределённых систем исходя из недостатков существующих аналогов.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести исследование существующих средств анализа производительности распределенных систем.
2. Произвести выбор основных мер оценки производительности распределенных систем
3. Провести исследование моделей и методов анализа производительности распределенных систем.
4. Определить зависимость производительности от структуры анализируемой системы и ее рабочей нагрузки.
5. Предложить улучшения к существующим методам и метрикам анализа.

Объектом магистерской диссертации являются методы и способы оценки производительности систем в условиях распределенной среды.

Предметом магистерской диссертации является нахождение оптимальных методов анализа данных и метрик производительности распределенных систем. Основная *гипотеза*, положенная в основу работы: наиболее эффективные методы оценки производительности распределенных систем, строятся на моделировании и прототипировании реальной системы с последующим снятием показателей с наибольшими значениями. Особенностью исследуемой темы является то, что среди методов оценки производительности той или иной системы, наблюдается различия в значительной погрешности и требуется унификация процесса для достижения абстракции. Таким образом экспериментальное исследование становится невозможным без программных компонентов – адаптеров, разработка которых является неотъемлемой частью данной работы.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя С. Н. Неборского, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на Международной научно-практической конференции «Наука и образование в XXI веке» (Москва, Россия, 2016).

Опубликованность результатов диссертации

По теме диссертации опубликована 1 печатная работа в сборниках трудов и материалов международных конференций.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, четырех глав, заключения, библиографического списка и пяти приложений. В главе 1 приводится исследование инструментов анализа производительности распределенных систем. Глава 2 посвящена исследованию существующих моделей оценки производительности систем. В главе 3 анализируются существующие методы анализа и оценки систем, их достоинства и недостатки. В главе 4 подводятся итоги, выделяются факторы, влияющие на показатель производительности системы и предлагается модель анализа производительности распределенных систем.

Общий объем работы составляет 74 страниц, из которых основного текста – 74 страниц, 7 рисунков на 7 страницах, 29 таблиц на 32 страницах, список использованных источников из 31 наименование на 3 страницах.

ОСНОВНОЕ СОДЕРЖАНИЕ

Во **введении** определена область и указаны основные направления исследования, показана актуальность темы диссертационной работы, дана краткая характеристика исследуемых вопросов, обозначена практическая ценность работы.

В **первой главе** проведен анализ применяемых инструментов оценки производительности распределенных систем, среди которых: встроенный ASP.NET профилировщик; монитор приложений Azure; AppDynamics. Проанализирована и определена проблемная область. Рассмотрено построение групп метрик, собираемых системами о производительности, выделены сходства и различия. Описаны проблемы по работе с рассмотренными системами.

Распределенные приложения специально предназначены для исполнения в распределенной системе. Поскольку многие распределенные системы включают большое количество модулей, дизайн и конфигурация распределенных приложений значительно сложнее, чем разработка и настройка приложений для исполнения на одной компьютерной системе. Каждая часть распределенного приложения, в дополнение к конфигурированию для правильной работы в автономном режиме, должны также быть сконфигурированы соответствующим образом, чтобы поддерживать связь с другими частями распределенного приложения, а также другими модулями, связанными с распределенными системами. Примером архитектуры распределенной системы, можно привести серверную архитектуру

туру, приведенную на рисунке 1.1. Из-за этой сложности, связь между частями распределенных приложений (даже те, которые настроены правильно) может работать нежелательным образом, время от времени. Таким образом, часто желательно собрать и информацию от посреднических модулей распределенной системы, которые облегчают связь между частями распределенного приложения. Сбор такой информации часто упоминается как "трассировка" или "профилирование".

Один общий метод профилирования, используемый в распределенных системах, состоит в присоединении профилирующего процесса к части распределенного приложения и слежении за общением, как внутренним, так и внешним относительно частей распределенного приложения. Процесс профилирования записывает данные, как происходит коммуникация, в лог-файл. В некоторых случаях процессы профилирования присоединены к ряду различных участков распределенного приложения и каждый процесс профилирования записывает данные в отдельный файл журнала. Отдельные файлы журналов затем стягиваются и коррелируются, чтобы дать некоторое представление о том, что может вызывать нежелательное взаимодействие между частями распределенного приложения.

Проанализировано применяемое ПО и определена значимость различных параметров ОС при сборе мер и оценок производительности распределенной системы в реальном времени. Выполнен анализ применяемых инструментов анализа производительности распределенных систем, выявлены их достоинства и недостатки.

Расширение функциональных возможностей инструментов анализа производительности, помимо расширения базовых функций, достигается путем разработки новых методов, алгоритмов и моделей анализа данных, в том числе на основе вычислительных экспериментов и математических моделей. Ведется активная разработка методов и алгоритмов выявления диагностических признаков, чувствительных к конкретным видам рабочей нагрузки исследуемых систем.

Результаты исследований, проведенных в этих направлениях, отражены в работах А.С. Хританкова, Л. Клейнрок, Е.С. Вентцель, К. К. Васильев, М. Н. Служивый, Д.А. Рид (Daniel A. Reed), Р.А. Аудт (Ruth A. Aydt), Л. ДеРоз (Luiz DeRose), К. Фарэл (C. Farrell), Н. Харисон (N. Harrison) и др.

Вторая глава посвящена исследованию и анализу наиболее часто применяемых моделей оценки производительности системы. Среди них рассмотрены более детально такие модели производительности как: эмпирическая, эталоны и аналитическая.

Была дана общая характеристика моделей, определены их достоинства и недостатки, принципы и подходы к построению таких моделей. Также представлены сводные результаты оценки моделей с дальнейшей детализацией о целесообразности применения той или иной модели для производства анализа и оценки производительности распределенных систем.

В третьей главе представляются методы анализа производительности систем на основании рассмотренных во второй главе моделей. Рассмотрена общая система классификации методов моделирования. Дана общая характеристика ме-

тодов, представлена общая классификация методов анализа и моделирования систем.

Более детально проанализирован и рассмотрен в качестве возможно применяемого для оценки производительности распределенных систем методы моделирования с помощью систем массового обслуживания.

Использование данного метода позволяет предсказывать производительность систем, их масштабируемость и поведение в различных условиях. Также, усложняя устройство каналов обработки, происходит усложнение базовой аналитической модели, позволяя применение рекурсивного подхода для расчета итоговых значений. [1]

В четвертой главе предложена математическая модель оценки производительности распределенной системы и определены информативно значимые влияющие факторы работы системы. Обозначены пределы максимальных и минимальных значений влияющих факторов. Представлены результаты экспериментального исследования модели, произведена оценка адекватности модели.

Модель построена на основе регрессионного анализа с решением уравнения регрессии для экземпляра узла системы с последующим объединением результатов и анализа общесистемного показателя. Общий вид уравнения регрессии представлен ниже.

$$y = k_0 + k_1 * x_1 + k_2 * x_2 + k_3 * x_3 + k_4 * x_4 + k_5 * x_5, \quad (1)$$

где x_i – значение загрузки i -ого параметра,

y – загруженность системы,

k_i – значение i -ого коэффициента уравнения регрессии.

Так же алгоритм кластерного анализа для принятия решения по отнесению исследуемого объекта к определенной группе показателей производительности системы с формированием начального положения ядер кластеров на основе экспертных и экспериментальных оценок.

Тогда для комплексной системы уравнение принимает следующий вид:

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N}. \quad (2)$$

Сущность метода формирования диагностических признаков для системы поддержки принятия решений по оценке технического состояния систем на основе регрессионного анализа экспериментальных данных. Эксперименты были проведены на базе идентичных виртуальных машин с разбегом рабочей нагрузки от 0 до 100% занятости системы.

На первом шаге алгоритма анализа выбирается ряд параметров, наиболее полно характеризующих состояние системы. На втором шаге строится характеристический вектор. Координатами вектора являются значения выбранных на первом шаге параметров. Третий шаг заключается в нахождении зависимости. Четвертый шаг уже является практическим применением модели для кластеризации полученных характеристических показателей.

Для реализации процедуры кластеризации используется итерационный метод динамических ядер. Начальное положение ядер определяется одним из следующих способов:

1. В качестве ядер на основе экспертных или экспериментальных оценок выбираются определенные характеристические векторы.

2. При условии первоначальной неопределенности ядра формируются путем обработки экспериментально полученных для группы однотипных объектов характеристических векторов, заключающейся в определении наилучшего, наихудшего и среднего значений для каждой координаты вектора.

Результатом кластеризации является отнесение характеристических показателей к одному из сформированных кластеров. Каждому из полученных кластеров ставится в соответствие определенное техническое состояние входящих в него объектов. Таким образом, отнесение исследуемого объекта или системы к какому-либо из кластеров позволяет сопоставить его с одной из возможных категорий технического состояния.

Экспериментальные исследования разработанной аналитической модели выполнены на базе опытного образца вычислительного комплекса для размещения распределенной системы web-ресурсов.

ЗАКЛЮЧЕНИЕ

Основные научные результаты диссертации

1. Проведено исследование следующих инструментов анализа производительности систем: встроенный ASP.NET профилировщик; монитор приложений Azure; AppDynamics. Рассмотрено построение групп метрик, собираемых системами о производительности, выделены сходства и различия. Описаны проблемы по работе с рассмотренными системами.

2. Предложено общее описание моделей для оценки данных о производительности распределенных систем. Выделены признаки полноты моделей. Описана типовая структура модели. Описаны особенности применения моделей.

3. Обобщены и описаны итоги применения моделей оценки производительности распределенных систем. Выделены достоинства и недостатки.

4. Предложено общее описание методов сбора и оценки данных о производительности распределенных систем. Выделены признаки полноты методов. Описана типовая структура и шаги методов. Описаны особенности применения методов.

5. Предложен комбинированный метод, основанный на объединении простейших, рассмотренных выше методов моделирования и оценки производительности систем.

6. Обобщены и описаны итоги применения методов оценки производительности распределенных систем. Выделены достоинства и недостатки.

7. Произведена оценка адекватности разработанной модели оценки производительности как комплекса простейших моделей, входящих в состав.

Рекомендации по практическому использованию результатов

1. Полученные результаты формируют теоретическую и практическую базу для внедрения оценки характеристики производительности распределенных систем.

2. Предложена модель комплексной оценки производительности систем может служить для долгосрочного процесса стандартизации высокоуровневого принятия решения о необходимости горизонтального или вертикального масштабирования оцениваемой системы.

3. Результаты могут использоваться для реализации оценочного механизма схожих систем, а также как база для дальнейшего усовершенствования и применения оценочной характеристики к более сложным системам.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ

1. Забродец, Д. А. Метод оценки производительности распределенных систем как систем массового обслуживания / Д. А. Забродец // Наука и образование в XXI веке. Электронный научный журнал по материалам междунар. науч.–практ. конф., Москва, 30 ноября 2016 г. – Москва, 2016. – с. 48-51. – ISSN 2414-5041