

Анализ и визуализация многомерных данных в глобально распределенных облачных мониторинговых системах

К.Н. Кучерова¹, С.В. Мещеряков¹, Д.А. Щемелинин¹

kristina.mylife@gmail.com|serg-phd@mail.ru|dshchmel@gmail.com

¹Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург, Россия

Качественное оказание информационных ИТ услуг, предоставляемых во всех регионах мира через глобальную сеть Интернет, предполагает применение современных глобально распределенных облачных ИТ технологий с большими потоками многомерных данных и поэтому является актуальной проблемой. В данной статье кратко описаны методы анализа и визуализации многомерных данных мониторинга, основанных на ключевых критериях производительности облачной ИТ системы, на примере конкретной глобально распределенной инфраструктуры международной ИТ компании. Внедрение предложенных методов визуальной аналитики в ведущих мировых ИТ компаниях – RingCentral (США) и Zabbix (Латвия), – позволило повысить качество и доступность предоставляемых ИТ услуг до общемирового уровня 99,999% в режиме 24/7. Использование новых решений в ИТ компаниях подтверждено актами о внедрении и публикациями в диссертациях соавторов.

Ключевые слова: информационная система, глобально распределенная инфраструктура, многомерные данные, облачные вычисления, мониторинг событий, системный анализ, графическая визуализация.

Analysis and visualization of big data in globally distributed cloud computing monitoring systems

K.N. Kucherova¹, S.V. Mescheryakov¹, D.A. Shchemelinin¹

kristina.mylife@gmail.com|serg-phd@mail.ru|dshchmel@gmail.com

¹Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

Quality of IT Services (QoS), providing across all globally distributed regions via Internet, use modern cloud computing IT technologies having big data flow and, therefore, is actual. This paper briefly describes the methods of analysis and visualization of monitoring big data based on Key Performance Indicators (KPIs) of a cloud computing IT system using real world example of globally distributed infrastructure of the International IT Company. Implementation of proposed methods of visual analytics in worldwide leading IT companies – RingCentral (USA) and Zabbix (Latvia), – allowed improving of QoS and availability of IT services up to a worldwide level of 99.999% in 24/7 mode. Implementation of new solutions in IT companies is confirmed by corresponding documents and by publications in PhD and DSc thesis of the coauthors.

Keywords: information system, globally distributed computing infrastructure, big data, cloud computing, monitoring events, system analysis, graphics visualization.

1. Введение

Объектом исследования является международная ИТ компания RingCentral (RC), США [9], предоставляющая информационные услуги [1] во всех регионах мира (Северная Америка, Западная Европа, Юго-Восточная Азия) с использованием облачных технологий на основе построенного глобально распределенного вычислительного комплекса (ГРВК) [3].

Цель исследования состоит в повышении качества информационных услуг (Quality of Service, QoS) и доступности Интернет сервисов на основе внедрения методов визуализации и анализа ключевых показателей производительности (Key Performance Indicators, KPIs), таких как [2]:

1. Оценка качества соединения (Mean Opinion Score, MOS) [7], предоставленная системой Palladion [8].
2. Число успешно завершённых вызовов.
3. Число безуспешных запросов на соединение.
4. Число потерянных/оборванных соединений, и т.п.

2. Мониторинг информационной системы

Информационная система (ИС) мониторинга активности клиентов компании RC поддерживает более чем 350 тыс. юридических компаний в различных регионах мира, каждая из которых может обслуживать тысячи физических лиц. Большой поток данных с удаленных клиентов и серверов в единую централизованную ИС мониторинга составляет до

500 значений в секунду (Values per Second, VPS) для каждого сервера по каждой измеряемой метрике.

Поэтому в компании RC установлена система Zabbix [12] как одно из 5 наиболее популярных и эффективных мониторинговых решений в мире для контроля основных системных метрик:

- 1) доступность сетевых ресурсов;
- 2) загрузка процессора;
- 3) наличие свободной оперативной памяти;
- 4) наличие свободного места на внешних носителях;
- 5) возможность своевременного прогнозирования утечки системных ресурсов [5], и др.

Интервал времени опроса в зависимости от системной метрики составляет от 1 секунды и более с возможностью реагирования на возможные неполадки в реальном времени. Объем хранимых исторических данных мониторинга варьируется от 1 недели и более в зависимости от наличия ресурсов базы данных мониторинга и заранее определяется конфигурацией Zabbix, которая включает [4]:

- 1) общее количество серверов в системе мониторинга;
- 2) тип данных каждой измеряемой метрики;
- 3) интервал времени опроса по каждой измеряемой метрике;
- 4) наличие триггеров срабатывания в случае аномалии различного приоритета;
- 5) графики и диаграммы изменения исторических данных о производительности ИС в целом и тенденций поведения отдельных метрик.



Рис. 1. Визуализация мониторинговых данных ИС.

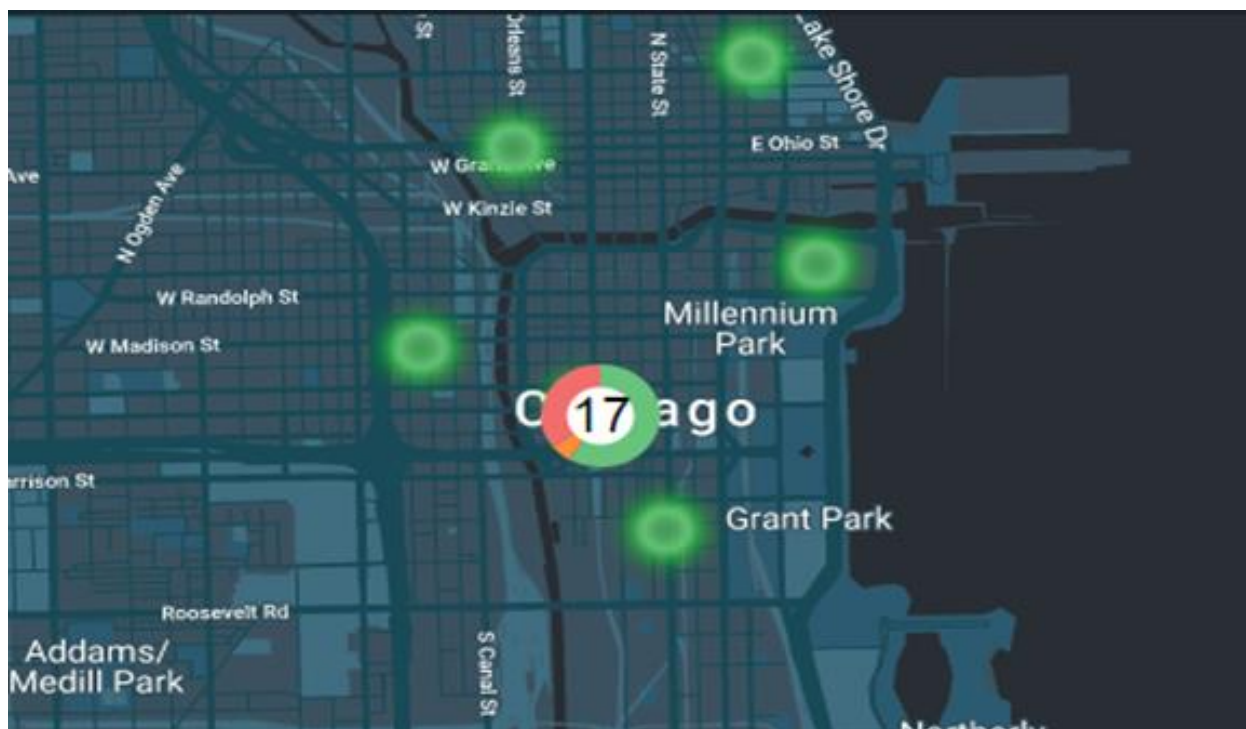


Рис. 2. Детализация мониторинговых данных ИС.

3. Визуализация мониторинговых данных ИС

Пример графического интерфейса визуализации мониторинговых событий ИС и QoS предоставляемых информационных услуг в ГРВК компании РС представлен на рис. 1.

С одной стороны, он свидетельствует о хорошей читабельности общего состояния ИС в целом, а с другой, о низкой степени детализации предоставляемых информационных сервисов и в этой связи плохой читабельности интерфейса как для оперативного дежурного, так и для системного аналитика.

На графическом интерфейсе рис. 1 выделены 5 основных рабочих областей:

1. График QoS соединений – содержит информацию об общем количестве активных запросов и количестве успешных вызовов за период времени с процентом умеренных и плохих соединений, а также сравнением с базовым значением MOS. KPI индикаторы зеленого цвета отображают соединения с высоким MOS. Если индикатор оранжевый или красный, это означает, что значение MOS хуже заданного.
2. Диаграмма качества звонков «Погодная карта» (Weather map) – представляет собой географическое распределение информации о качестве соединений на основе IP-адресов, взятых из мониторинговых данных от терминальных устройств пользователей. Чтобы увидеть качество соединений для конкретной географической области, в которой находится конечный пользователь, предусмотрена функциональность масштабирования (рис. 2), которая позволяет обслуживающему персоналу определять проблемы в каналах связи и в кратчайшие сроки уведомлять об обнаруженных аномалиях.
3. Графики процентов потери Интернет пакетов – эти две диаграммы визуализируют тенденции потери пакетов для восходящего и нисходящего тренда. В этом примере данные представлены за последние 24 часа, где каждый столбец – это процентное соотношение количества потерянных пакетов по сравнению с базовым уровнем для этого часа.
4. Агрегированные данные по 10 самым проблемным географическим зонам с возможностью фильтрации по 5 различным категориям KPI – Интернет провайдер, контроллер сетевой сессии, раздел данных клиентов, пользователь, его местоположение.
5. Исторические данные – представляют собой более детализированные диаграммы выбранной категории KPI.

4. Анализ мониторинговых данных ИС

Графический интерфейс оперативных данных мониторинга облачной инфраструктуры RC (рис. 1 и 2) обеспечивает визуализацию общего состояния всего ГРВК с возможностью быстрого реагирования обслуживающего персонала на аномальные события в ИС. Однако в процессе объективного роста числа Интернет пользователей RC и соответственно рабочей нагрузки на ИС, адекватное увеличение количества обслуживающего персонала и компьютерного оборудования диктует необходимость решения следующих новых задач:

- 1) улучшить QoS предоставляемых IT сервисов за счет сокращения времени отказов ИС;
- 2) повысить надежность средств мониторинга с точки зрения аналитики и визуализации причины и тенденции потенциального отказа ИС;
- 3) сократить время устранения неисправностей и автоматизировать процедуру восстановления IT сервисов без участия человека.

Для решения этих задач разработана технология мониторинга с использованием визуальной аналитики и прогнозирования событий в ГРВК на основе объективных исторических данных. Предложенная технология реализована программно-инструментальными средствами и внедрена в международной IT компании Zabbix (Рига, Латвия), которая в настоящее время входит в 5 наиболее популярных мониторинговых решений в мировой IT индустрии [6].

Предложенные математические модели прогнозирования основаны на 5 основных параметрических функциях, которые могут вычисляться в системе мониторинга в реальном режиме времени для каждого удаленного сервера по каждой измеряемой метрике:

- 1) линейная
$$f(t) = a + b t;$$
- 2) полиномиальная порядка n
$$f(t) = a_0 + a_1 t + a_2 t^2 + \dots + a_n t^n;$$
- 3) экспоненциальная
$$f(t) = a \exp(b t);$$
- 4) логарифмическая
$$f(t) = a + b \log(t);$$
- 5) степенная
$$f(t) = a t^b.$$

Параметры $n, a, b, a_0, a_1, a_2, \dots, a_n$, прогностических моделей вычисляются мониторинговой системой Zabbix автоматически на основе исторически накопленных статистических данных классическими методами решения систем линейных алгебраических уравнений Гаусса-Джордана, Вейерштрасса (Джордан-Кернер), Ньютона, Гершгорина [10].

Критерием соответствия вычисленных прогностических результатов исторически реальным статистическим данным является традиционно общепринятый коэффициент корреляции (детерминации) $R^2 \rightarrow 1$ – это доля дисперсии случайной ошибки прогностической модели σ^2 относительно дисперсии статистической выборки данных σ_i^2 :

$$R^2 = 1 - \sigma^2 / \sigma_i^2;$$

$$0,9 < R^2 < 1.$$

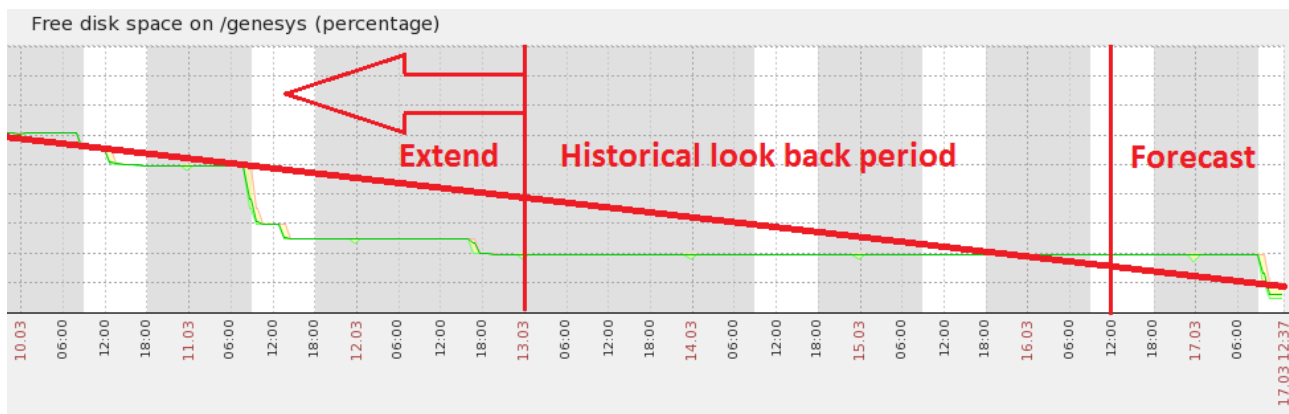


Рис. 3. Прогнозирование свободного места на диске.

В практической реализации мониторинговой системы Zabbix тип прогностической модели по каждой измеряемой метрике выбирается заранее и конфигурируется вручную исходя из визуального анализа исторических графиков. В дальнейшем в процессе эксплуатации коэффициент R^2 автоматически оценивается для различных наборов параметров выбранной модели и оптимизируется с использованием известных математических методов и моделей временных рядов Лагранжа, Фишера, Акаике [11]. Тип модели прогнозирования, так же как исторический интервал времени для анализа статистики и горизонт прогнозирования, можно поменять в любой момент времени в случае недостоверного прогноза.

По результатам прогнозирования в мониторинговой системе Zabbix предусмотрены триггеры, которые срабатывают и посылают уведомление заблаговременно в случае, если прогнозируемое значение метрики может выйти за пределы заданного порога в период установленного горизонта времени прогнозирования.

Графическая иллюстрация визуального аналитического подхода к прогнозированию показана на рис. 3 на примере мониторинга свободного места на диске с применением линейной математической модели. Важным условием достоверности прогноза является соблюдение правильных пропорций от 5:1 до 10:1 между интервалом времени исторической статистики и горизонтом прогнозирования как показано на рис. 3.

Дальнейшей задачей развития системы мониторинга удаленных серверов является разработка моделей прогнозирования для волнообразных циклических рабочих нагрузок, при которых перечисленные функции мало эффективны по причине многочисленных ложных срабатываний триггеров в системе мониторинга в начале и конце рабочего дня, в выходные и праздники [4].

5. Выводы

Предложенный подход к анализу и визуализации многомерных данных мониторинга глобально распределенной ИС реализован в международной IT компании RingCentral [9], что позволило, наряду с другими средствами автоматизации, улучшить QoS облачных сервисов и повысить уровень их доступности до лучших мировых показателей 99,999% в режиме обслуживания 24/7.

Традиционный подход к мониторингу многомерных данных ГРВК (сбор статистики с удаленных серверов с заданным интервалом времени опроса, сравнение полученных данных с заданным пороговым значением и уведомление в случае аномалии) дополнен новым решением, цель которого – прогнозирование событий в ИС на основе объективной исторической статистики. Для этого разработаны математические модели прогнозирования и внедрены в одной из ведущих в мире мониторинговых систем Zabbix [12].

Превентивное обнаружение аномалий в системе позволяет избежать отказов в сервисном обслуживании и выполнять мероприятия по их устранению в часы плановой профилактики, что особенно актуально для современных облачных технологий ГРВК.

Внедрение моделей прогнозирования в систему мониторинга также позволяет регламентировать рабочую нагрузку на ИС, получать уведомления о тенденции роста пользовательских запросов и масштабировать вычислительные ресурсы заблаговременно, а не во время пиковой нагрузки со стороны пользователей.

6. Литература

[1] Мещеряков С.В., Щемелинин Д.А. Методы визуализации и анализа больших данных в облачных

информационных системах // Труды 28-й Международной конференции по компьютерной графике и машинному зрению (Графикон). – Томск, ТПУ, 2018. – С. 151-154, <http://2018.graphicon.ru/>

- [2] Efimov V. Performance and Quality Evaluation of Cloud Information Services // Communications in Computer and Information Science. – Springer International Publishing Switzerland, 2016, V. 601. – P. 230-237, <http://www.springer.com/gp/book/9783319308425>
- [3] Efimov V.V., Yakovlev K.A. Integration and Continuous Service Delivery in Globally Distributed Computing System // Humanities and Science University Journal. – 2017. – №30. – С. 13-20, <http://uni-journal.ru/technics/archive/>
- [4] Kucherova K., Mescheryakov S., Shchemelinin D. Prediction Experience and New Model // The 7th Annual International IT Zabbix Conference, Riga, Latvia, 2017, http://www.zabbix.com/conf2017_agenda.php, <https://youtu.be/4TY4eIB7tiw>
- [5] Mescheryakov S., Shchemelinin D. Capacity Management of Java Based Business Applications Running on Virtualized Environment // The 39th International IT Conference on Computer Performance Evaluation and Management by Computer Measurement Group (CMG), La Jolla, CA, USA, 2013, <http://www.cmg.org/publications/conference-proceedings/conference-proceedings-2013/>
- [6] Mescheryakov S., Shchemelinin D. Proactive and Reactive Monitoring // The 4th Annual International IT Zabbix Conference, Riga, Latvia, 2014, http://www.zabbix.com/conf2014_agenda.php, <http://www.youtube.com/watch?v=PbUwr7tcikw>
- [7] Nadeem Unuth. Mean Opinion Score (MOS): A Measure of Voice Quality, <https://www.lifewire.com/measure-voice-quality-3426718>
- [8] Palladion. User Manual, https://serviceassurance.yourservices.co.uk/help/_static/palladion-user-manual.pdf
- [9] RingCentral Inc. Official web site, <http://www.ringcentral.com/>
- [10] Verschelde J. The Method of Weierstrass (also known as the Durand–Kerner method), 2003.
- [11] Ward M.D., Ahlquist J.S. Maximum Likelihood for Social Science: Strategies for Analysis. Cambridge University Press, 2018. ISBN 978-1-316-63682-4.
- [12] Zabbix Enterprise-class Monitoring Solution. Official web site, <http://www.zabbix.com/product.php>

Об авторах

Кучерова Кристина Николаевна, аспирант кафедры «Инженерная графика и дизайн» Санкт-Петербургского политехнического университета Петра Великого. E-mail kristina.mylife@gmail.com.

Мещеряков Сергей Владимирович, д.т.н., профессор кафедры «Инженерная графика и дизайн» Санкт-Петербургского политехнического университета Петра Великого. E-mail serg-phd@mail.ru.

Щемелинин Дмитрий Александрович, к.т.н., докторант кафедры «Инженерная графика и дизайн» Санкт-Петербургского политехнического университета Петра Великого. E-mail dshchmel@gmail.com.