

Методы визуализации и анализа больших данных в облачных информационных системах

С.В. Мещеряков, Д.А. Щемелинин
serg-phd@mail.ru|dshchmel@gmail.com

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

Эффективная эксплуатация глобально распределенных вычислительных комплексов с многофункциональными информационными сервисными системами и большими потоками данных невозможна без интеграции процессов непрерывного мониторинга и визуализации текущего состояния всей облачной инфраструктуры по ключевым критериям производительности. В цели и задачи современного мониторинга входит не только визуализация и оповещение обслуживающего персонала и пользователей о временных перерывах работоспособности всей инфраструктуры или ее отдельных частей, но и по возможности автоматическое восстановление работоспособности информационных систем и сервисов. Все эти вопросы рассмотрены в данной статье на конкретном примере большой глобально распределенной интернациональной IT компании.

Ключевые слова: информационные системы, глобально распределенная инфраструктура, большие данные, интеграция, облачные вычисления, мониторинг, визуализация.

Methods of big data visualization and analysis in cloud computing information systems

S.V. Mescheryakov, D.A. Shchemelinin
serg-phd@mail.ru|dshchmel@gmail.com

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

Effective operations of a globally distributed cloud computing infrastructure with multi-service information systems and big data flows is impossible without integration of continuous monitoring processes and visualization of the current status of the entire infrastructure using key performance indicators (KPIs). Modern goals and initiatives of the monitoring system include not only visualization and alerting to technical support and to customers about temporary service outage in either entire or partial production infrastructure but also auto-restoration of information services as soon as possible. In this paper, all mentioned problems are discussed using particular example of big globally distributed International IT Company.

Keywords: information systems, globally distributed computing infrastructure, big data, integration, cloud computing, monitoring, visualization.

1. Введение

Исследования проведены в крупной интернациональной IT компании RingCentral (RC), США [5], предоставляющей следующие телекоммуникационные сервисы в разных регионах мира – США и Канаде, Северной и Западной Европе, Юго-Восточной Азии и Океании:

1. Сервисные услуги виртуальной автоматической телефонной станции (АТС) посредством голосового Интернет протокола (VoIP), включая факсимильные и короткие текстовые сообщения (SMS).
2. Облачные программные решения для организации центра обработки звонков, так называемых кол-центров.
3. Видео и телеконференции, включая селекторные совещания между тремя и более участниками.
4. Службы мгновенных сообщений, организации групповых текстовых чатов, звуковых сигналов, файлов, изображений, видео, в том числе на мобильные устройства.

База данных клиентов RC насчитывает более 350000 юридических предприятий, каждое из которых может обслуживать тысячи физических лиц, в разных регионах мира.

На рис. 1 приведена облачная архитектура RC, размещенная в 9 центрах обработки данных на западном и восточном побережье США, в странах Западной Европы и Юго-Восточной Азии. Масштабируемость инфраструктуры RC обеспечивает ежегодный рост на 40% клиентской базы,

программного обеспечения и компьютерного оборудования для тестирования. Общий объем физических и виртуальных серверов на сегодняшний день составляет более 20000 вычислительных машин [3]. Все серверы распределены на компоненты и группы, находящиеся под балансировкой нагрузки в зависимости от активности со стороны пользователей и сторонних провайдеров информационных услуг.

2. Система мониторинга ИС

Для мониторинга глобально распределенной облачной инфраструктуры RC использована программная система с открытым исходным кодом Zabbix [6], которая входит в 5 наиболее популярных мониторинговых решений в мире и эффективно используется для контроля системных метрик, таких как доступность сетевых сервисов, использование процессора, наличие памяти, свободное место на внешних дисках и др. Реагирование на возможные неполадки происходит в реальном режиме времени.

Система мониторинга Zabbix построена по стандартной архитектуре «клиент-сервер», состоящей из центральной базы данных MySQL реляционного типа и удаленных серверов (рис. 2). Дополнительные прокси серверы установлены с целью уменьшения нагрузки больших потоков данных со стороны удаленных серверов на базу данных. Интервал времени опроса удаленных серверов составляет для разных системных метрик от 1 до 30 минут, а объем хранимых исторических данных варьируется от месяца до года. Эти параметры заранее определяются конфигурацией Zabbix, которая включает [4]:

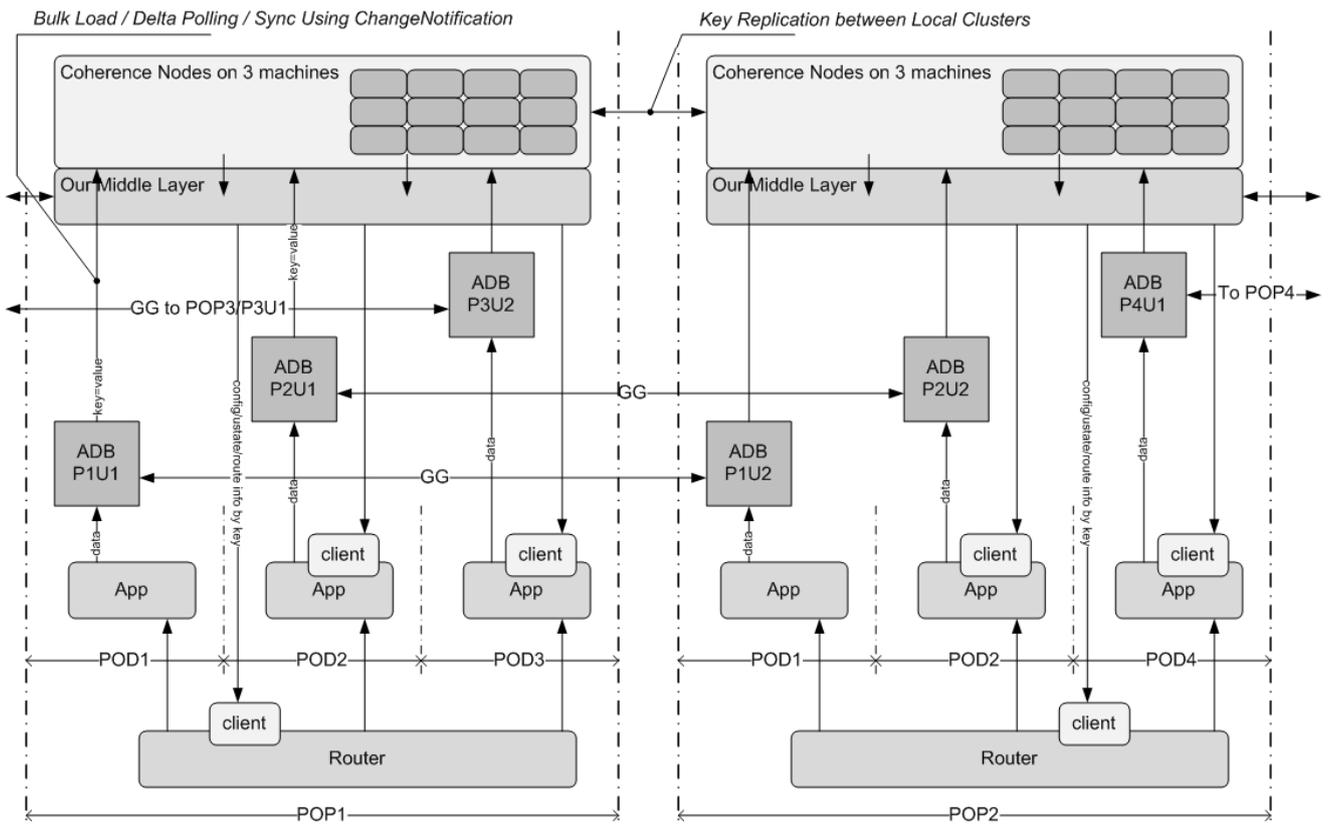


Рис. 1. Масштабируемая архитектура RC.

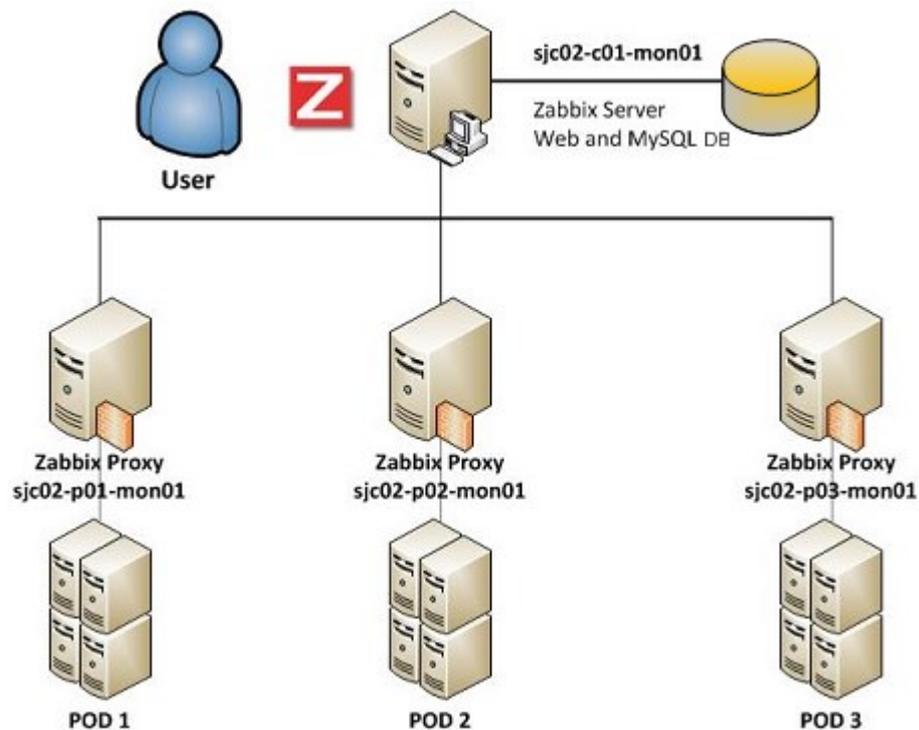


Рис. 2. Архитектура системы мониторинга Zabbix.

- 1) список всех серверов в системе мониторинга;
- 2) описание метрик для мониторинга доступности и производительности серверов, системных служб и приложений, работающих на серверах;
- 3) триггеры различного приоритета (информационный, предупреждающий, критический, аварийный), которые срабатывают на события, когда текущее значение метрики выходит за пределы заранее предопределенной пороговой величины;
- 4) графики анализа данных о производительности и исторических тенденций поведения как отдельных метрик серверов, так и ИС в целом.

3. Интерактивный графический интерфейс мониторинга ИС

Система мониторинга Zabbix не обеспечивает реализацию графического интерфейса визуализации состояния всей облачной инфраструктуры с алгоритмом подавления информационного шума и группировки событий, происходящих в РС. С увеличением количества обслуживаемого оборудования возникла необходимость решения следующих задач:

- 1) уменьшить общее число ложных срабатываний триггеров событий во время отказов;
- 2) улучшить качество средств мониторинга с точки зрения визуализации корневой причины отказа ИС;
- 3) сократить время, затраченное на устранение поломок и автоматическое восстановление сервисов.

Для решения поставленных задач реализована новая технология построения среды мониторинга, основанная на интеграции используемых программно-инструментальных средств, – Zabbix API, Atlassian Jira, vCenter, MySQL, Laravel и др., а также собственного программного обеспечения. Разработан интегрированный графический интерфейс (рис. 3), который формируется как сводный отчет в формате HTML, где весь информационный шум (как, например, триггеры C04 на рис. 3) представлен методами [1, 2] в виде иерархического дерева триггеров, сгруппированных под одним корневым событием.

4. Прогнозирование отказов в ИС

Общий подход к мониторингу состоит в сборе данных с удаленных серверов с заданным интервалом времени опроса, сравнении последних значений с указанным пороговым значением и активации триггера в случае превышения. Цель нового решения – создать процесс мониторинга с прогнозированием и заблаговременным предотвращением аномалий, не допуская отказов в сервисном обслуживании. Для этого предложено создать модели прогнозирования отказов, которые оценивают будущие значения метрик и тенденции поведения компонент ИС на основе исторической статистики.

Для прогнозирования отказов в системах с большими данными предложен следующий алгоритм работы:

NOC Monitoring Console better experience

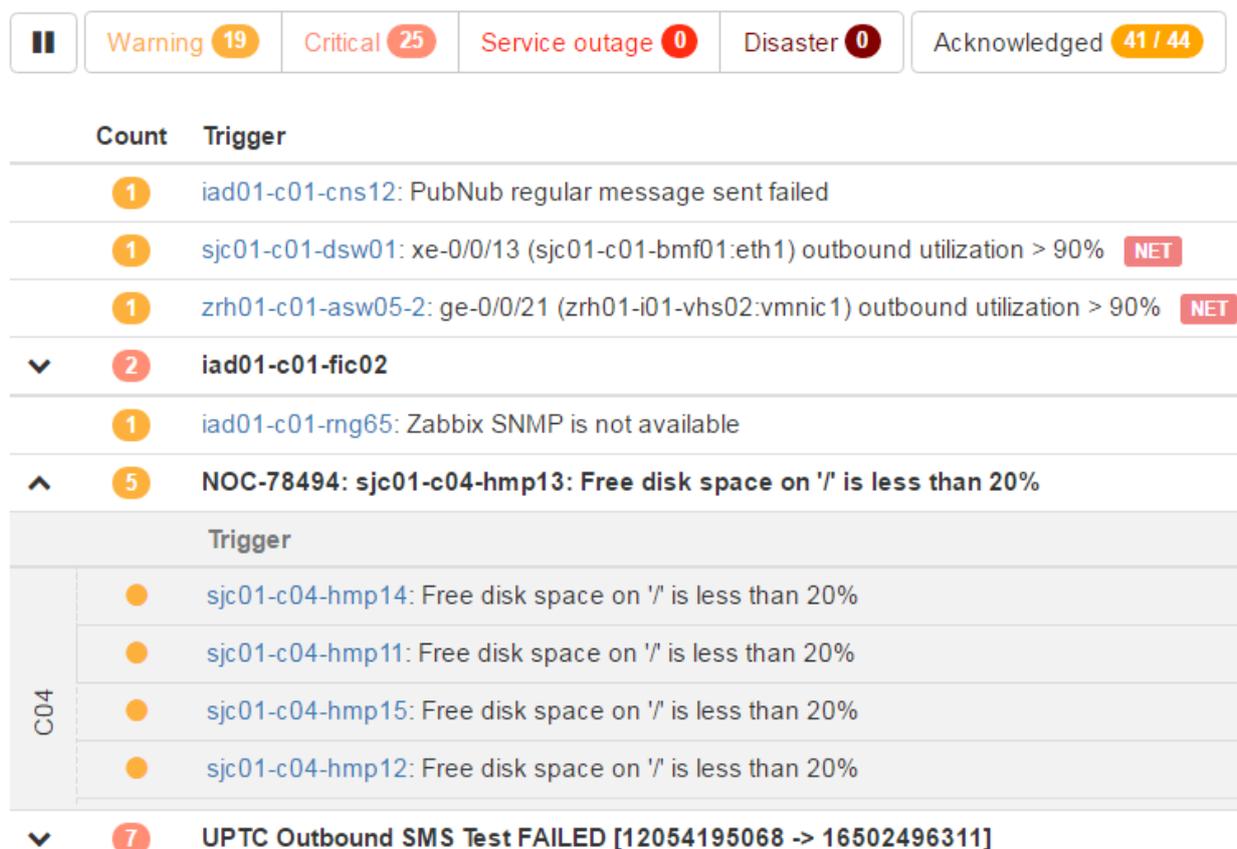


Рис. 3. Интерактивный интерфейс визуализации состояния ИС.

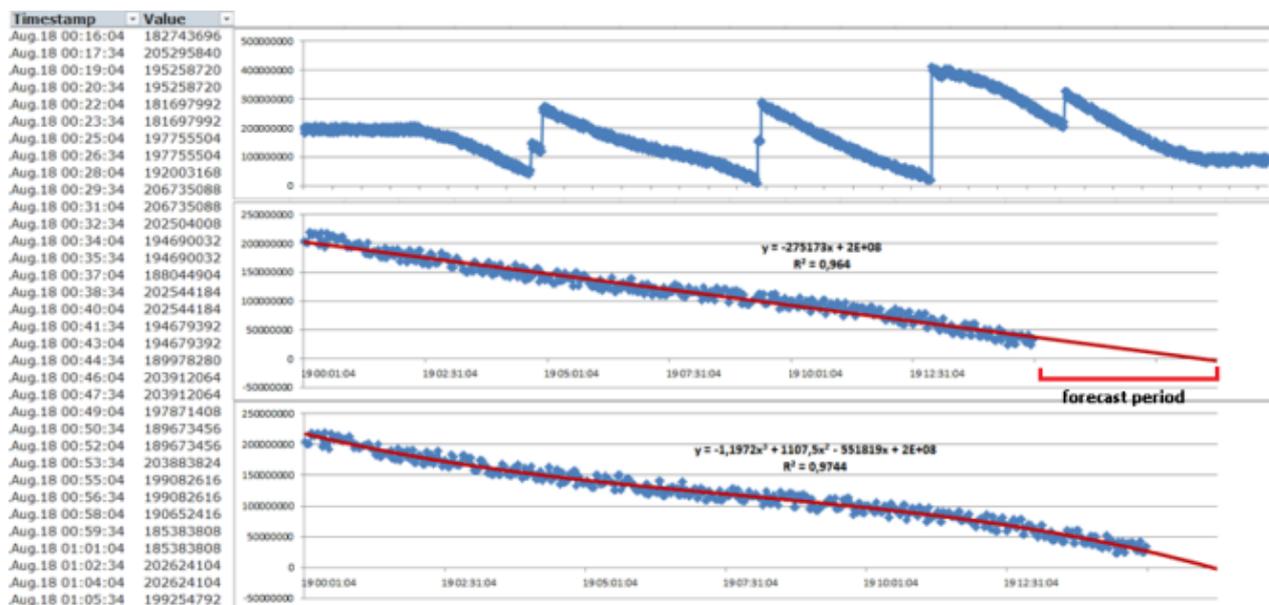


Рис. 4. Пример определения функции прогнозирования поведения элемента ИС.

- 1) экспортировать статистические значения из базы данных мониторинга в одну из математических систем анализа – Mathematica, R, Matlab (рис. 4);
- 2) задать исторический период изменения показателей;
- 3) настроить тип модели данных и горизонт прогноза;
- 4) полученное регрессионное выражение использовать в системе мониторинга.

5. Выводы

Практическая реализация предложенной технологии мониторинга в международной ИТ компании RingCentral позволили повысить доступность облачных сервисов в режиме 24/7 до общепринятого в мире уровня 99,999%.

Улучшение визуализации текущей рабочей обстановки достигнуто за счет разработанного интерактивного графического интерфейса, позволяющего отображать только наиболее значимые события, являющиеся первопричиной аномалий в глобально распределенной системе, а также управлять этими событиями в реальном режиме времени.

Эффективность регламентного обслуживания улучшена за счет практической реализации и внедрения моделей прогнозирования в систему мониторинга. Это позволило получать уведомления о неполадках заблаговременно и выполнять превентивные мероприятия по их устранению в часы плановой профилактики, а не во время пиковой нагрузки со стороны пользователей.

6. Литература

- [1] Мещеряков С.В. Сравнительный анализ вариантов организации иерархии в базах данных // Системы управления и информационные технологии. – 2009. – № 1 (35). – С. 34-37.

- [2] Мещеряков С.В., Иванов В.М. Реализация модели данных для описания иерархических объектов с произвольными атрибутами // Научно-технические ведомости СПбГПУ. Информатика. Телекоммуникации. Управление. – 2009. – № 1 (72). – С. 139-144.
- [3] Bortyakov D.E., Mescheryakov S.V., Shchemelinin D.A. Integrated Management of Big Data Traffic Systems in Distributed Production Environments // St. Petersburg State Polytechnic University Journal. Computer Science. Telecommunications and Control Systems. – 2014. – № 1 (188). – pp. 105-113.
- [4] Hegedus J.C. Perrobix + Zabbix DB Monitoring. Zabbix Annual International Conference, 2013. Available: http://youtu.be/pJCV_ui0orQ. Published: Oct 14, 2013. Last access: 2018.
- [5] RingCentral Inc. Official web site. <http://www.ringcentral.com/>. Last access: 2018.
- [6] Zabbix Enterprise-class Monitoring Solution. Official web site. Available: <http://www.zabbix.com/product.php>. Last access: 2018.

Об авторах

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

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