

## BIG DATA В АВТОМАТИЗАЦИИ ИТ-РЕШЕНИЙ



**А.А. Навроцкий**

*Заведующий кафедрой информационных технологий автоматизированных систем  
БГУИР, кандидат физико-математических наук, доцент*



**С.Ю. Крюков**

*Старший инженер-программист EPAM Systems,  
аспирант БГУИР*

*Белорусский государственный университет информатики и радиоэлектроники, Республика Беларусь*

*E-mail: siarhei.krukau@gmail.com*

*Abstract.* Big Data has already established itself as an important part of many data-driven businesses. Industries that make use of Big Data include science, health care, finances and marketing, to name a few. But not only scientists and managers, who need to analyze big amounts of data today. A wide range of IT specialists could win from applying Big Data to their workflows: the complexity of present-day IT solutions leaves no room for manual management and human supervising.

Бурное развитие информационных технологий в последние годы изменило подход многих компаний к созданию и сохранению конкурентных преимуществ. Даже такие традиционно консервативные отрасли, как отельно-ресторанный бизнес, вынуждены идти в ногу со временем для продвижения услуг через сеть Internet и растущий рынок мобильных устройств. Появляются новые компании, предоставляющие привычные услуги, но при этом делающие упор на разработку соответствующих программных продуктов. Uber — яркий пример такого бизнеса. На прошедшей в 2015 г. конференции разработчиков Microsoft глава корпорации Сатья Наделла высказал мнение, что в скором будущем любая компания будет в той или иной степени компанией по разработке программного обеспечения [1].

Это порождает повышенный спрос на инструментарий по управлению и поддержке ИТ-инфраструктуры, ведь сегодня ей отводится важная роль практически любого крупного бизнеса. От такого инструментария требуется:

- возможность постоянного наблюдения за состоянием парка серверов и систем хранения данных;
- контроль целостности сетевой инфраструктуры;
- контроль работоспособности программного обеспечения;
- автоматизация рутинных задач.

Постоянное наблюдение за составными частями ИТ-инфраструктуры

генерирует огромный поток данных, таких как системные метрики, лог-файлы, телеметрия, информация о различных событиях. Потенциал этого потока чаще всего остается нераскрыт. При этом сложность применяемых IT-решений со временем лишь растёт — системы становятся более динамичными, широкое распространение получает виртуализация и контейнеризация приложений — что повлечёт за собой рост объёма генерируемых данных. Описанная ситуация характерна для отрасли Big Data. Впервые на этот факт обратили внимание инженеры “F5 Networks” и ввели новый термин: Big Operational Data [2].

Big Operational Data — это “сырые” данные, требующие соответствующих средств аналитики и обработки, к которым предъявляются следующие требования:

- агрегирование данных из нескольких источников “на лету”;
- контроль за производительностью и работоспособностью;
- выявление сбоев и их причин;
- проактивное предсказывание сбоев;
- выявление зависимостей и корреляций;
- прогнозирование потребности в различных ресурсах.

Класс таких средств получил название “операционная IT-аналитика” (IT Operations Analytics, ИТОА). Издание Gartner прогнозирует широкий рост рынка ИТОА в будущем [3].

Операционная IT-аналитика заимствует некоторые подходы и инструменты из бизнес-аналитики и других областей (теоретическая физика, финансы, медицина), где давно существует развитый инструментарий анализа. Можно выделить следующие аналитические алгоритмы, применимые для ИТОА:

- составление корреляций;
- выявление топологических зависимостей;
- текстовая аналитика;
- экономическое моделирование.

Сопоставление событий — классический подход к анализу данных, подразумевает одновременную обработку множества потоков данных и выявление зависимости между ними. Корреляционный анализ — основа ИТОА, его результаты используются в других алгоритмах.

Анализ топологических зависимостей подразумевает построение визуальной карты взаимодействия различных частей IT-инфраструктуры. Такая карта может ограничиваться визуализацией сетевой топологии, либо раскрывать логические взаимосвязи между сервисами, приложениями, базами данных. Анализ топологических зависимостей позволяет находить “точки отказа” в критических ситуациях или проблемные места с низкой пропускной способностью (“бутылочные горлышки”) в инфраструктуре.

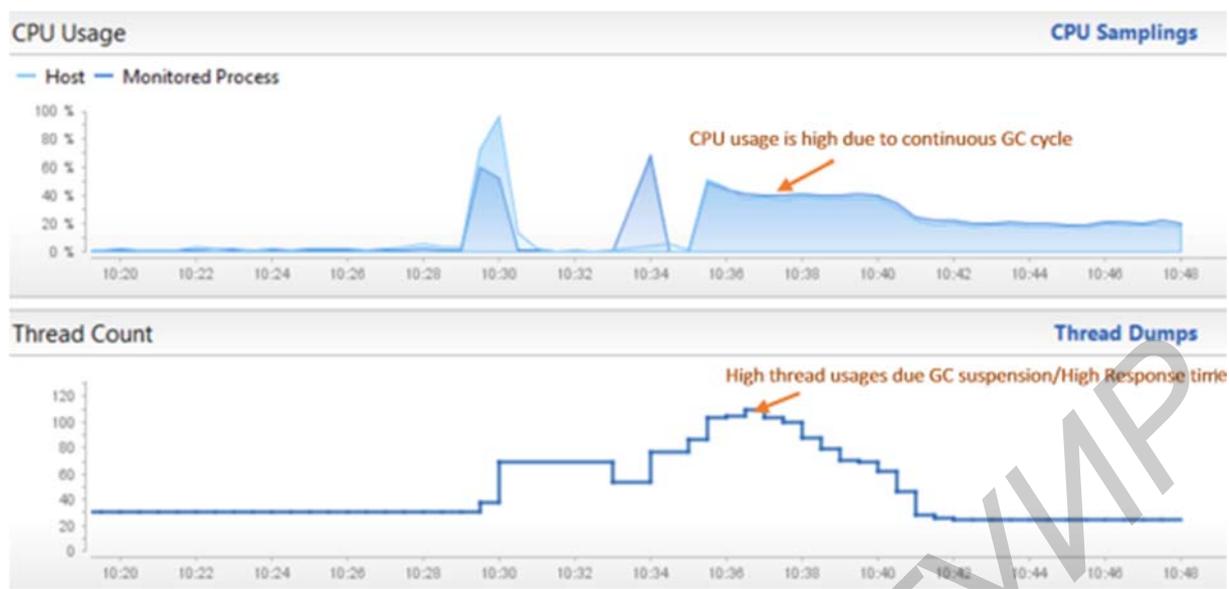


Рис. 1. Пример корреляции между количеством потоков и загрузкой CPU

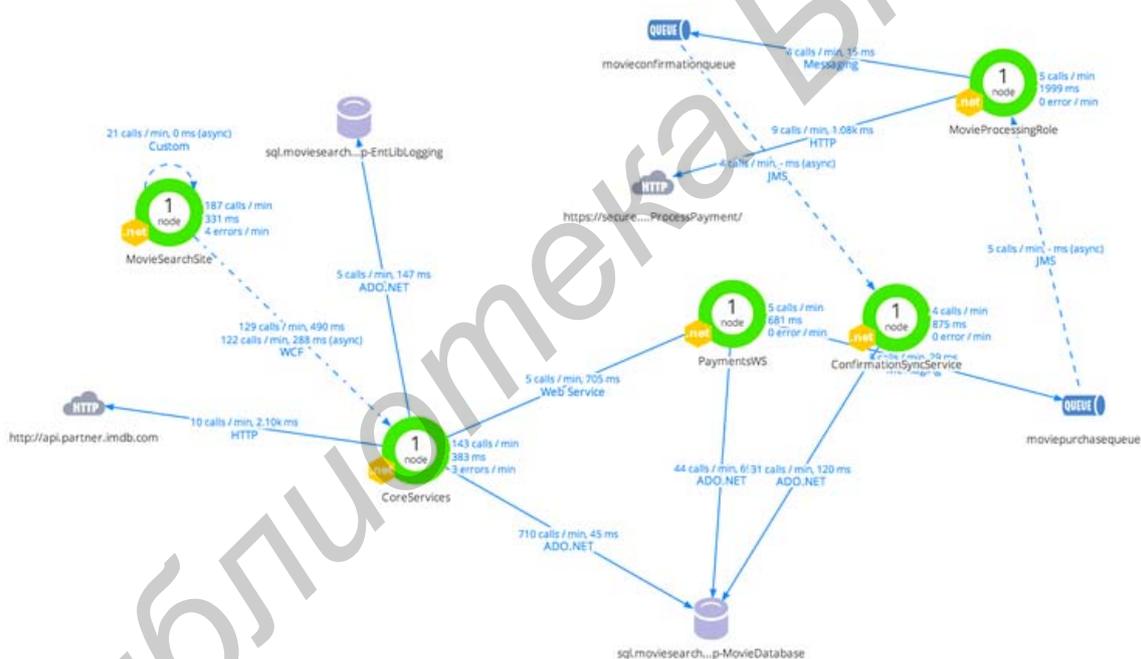


Рис. 2. Пример визуализации топологии IT-инфраструктуры

Текстовая аналитика используется для анализа лог-файлов и журналов операционных систем. Эти файлы протоколируют все действия, происходящие с системой. Традиционно, они использовались для анализа событий постфактум, новый подход подразумевает возможность выделения и анализа записей в непрерывном потоке в режиме реального времени.

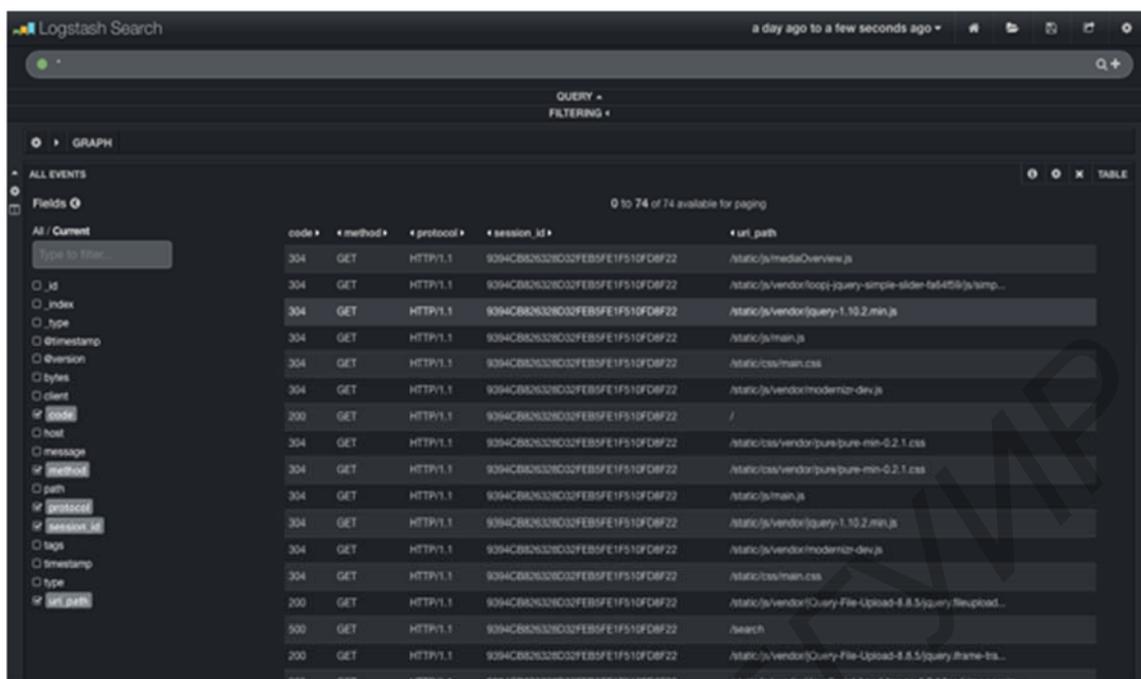


Рис. 3. Анализ лог-файлов HTTP-трафика в LogStash

Экономическое моделирование – относительно новая отрасль IT-аналитики – позволяет наиболее эффективно использовать ресурсы IT-инфраструктуры, масштабируя их выделение оптимальным образом. Экономические модели получили широкое распространение в виртуализированных облачных средах, предоставляющих ресурсы на повременной основе, таких как Amazon Web Services.

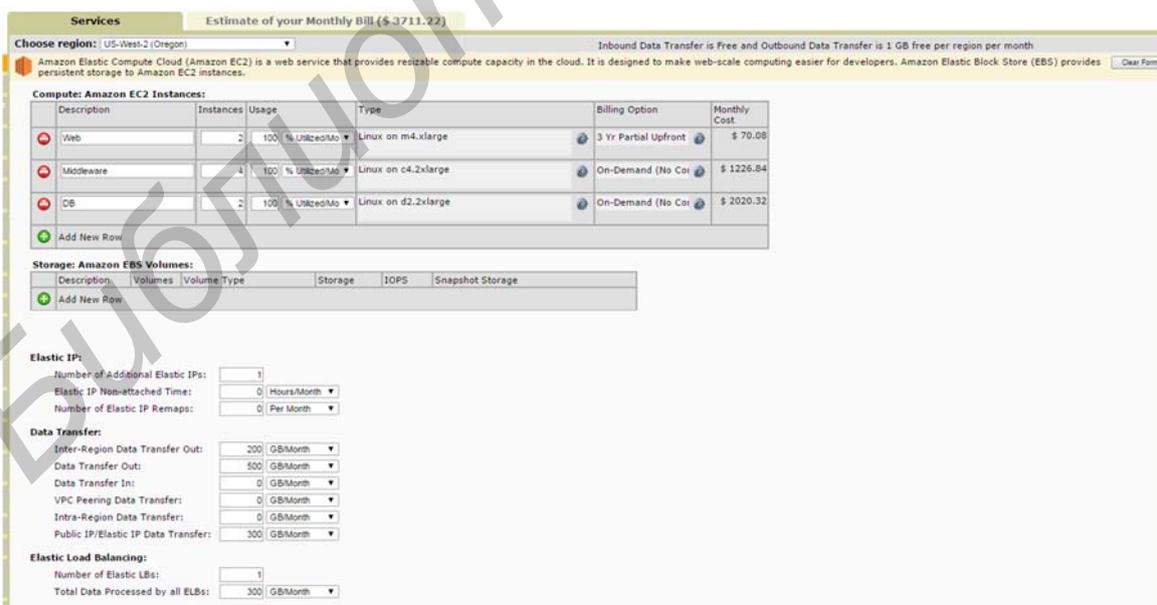


Рис. 4. Моделирование денежных затрат при использовании AWS

Различают три уровня ИТ-аналитики по сложности решаемых задач:

- анализ доступности
- анализ поведения
- анализ масштабируемости

Анализ доступности даёт представление о текущем состоянии ИТ-инфраструктуры без каких-либо прогнозов. Анализ поведения использует алгоритмы проактивного анализа для заблаговременного выявления проблем и путей их решения. Анализ масштабируемости используется для долгосрочного прогнозирования потребностей ИТ-инфраструктуры в ресурсах.

Таким образом, ИТОА-решения помогают получить общую картину использования ИТ-инфраструктуры, отследить потоки данных в ней, обнаружить узкие места и точки отказа. С их помощью можно повысить оперативность внесения изменений в конфигурацию сервисов, выявлять причины неполадок и значительно сократить время восстановления. Средства ИТ-аналитики позволяют предупреждать проблемы с помощью анализа поведения компонентов инфраструктуры, изучения и распознавания "нормальных" и "аномальных" шаблонов их поведения. Большое значение аналитика приобретает в процессах управления изменениями, предоставляя возможность на основе метрик производительности, анализа конфигураций и модели сервиса быстро оценить последствия внесения изменения в компоненты инфраструктуры для сервиса в целом.

#### *Литература*

[1]. Satya Nadella: Every business will be a software business. [Электронный ресурс]. – Режим доступа: <http://www.computerweekly.com/news/2240242478/Satya-Nadella-Every-business> – Дата доступа: 23.04.2016

[2]. С

l  
o  
u [3]. IT operations analytics: Changing the IT perspective. [Электронный ресурс]. – Режим  
доступа: <http://www.information-age.com/technology/data-centre-and-it-perspective> – Дата доступа: 23.04.2016

a  
n  
d

T  
h  
i  
n  
g  
s

a  
n  
d

В  
ig Operational Data. [Электронный ресурс]. – Режим доступа:  
– Дата доступа: 23.04.2016