



Технология ускорения веб-приложений FastView.
Когда дорога каждая секунда
Технологический обзор



Оглавление

Основные положения.....	3
Почему производительность веб-сайта имеет критическое значение.....	4
Факторы, влияющие на производительность веб-сайтов	5
Резкое повышение производительности веб-сайта с помощью Radware FastView™.....	8
Функционал и преимущества технологии FastView	8
Оптимизация на транспортном уровне.....	8
Сокращение передаваемого контента.....	9
Практическое значение FastView	11
Мониторинг производительности приложений	11
Заключение	13

Основные положения

Вне зависимости от того, какая бизнес-метрика используется для оценки успеха и эффективности вашего сайта или критически значимых корпоративных веб-приложений – коэффициент конверсии, количество просмотров страниц, удовлетворенность пользователей, коэффициент снижения количества посетителей (abandonment rate) или продуктивности труда сотрудников – производительность вашего сайта непосредственно влияет на каждый из перечисленных показателей.

Многие компании затрачивают большое количество времени, усилий и ресурсов на повышение производительности своего веб-сайта. Сюда относится сокращение общего объема страницы (page weight) и количества объектов на ней, оптимизация кода или увеличение пропускной способности интернет-соединения. Тем не менее, производительность сайта зависит от полной цепи доставки приложения, где должно быть оптимизировано каждое звено, начиная с веб-сервера и сетей, соединяющих сайт с конечным пользователем, и заканчивая различными браузерами и клиентским ПО конечных пользователей.

Для того, чтобы предоставить пользователям наилучшее потребительское качество при работе с онлайн-приложениями, IT отдел должен обратить внимание на различные факторы, ухудшающие производительность сайта. Для того, чтобы эффективно проанализировать производительность веб-сайта с различных сторон, учитывая различные технологии браузеров, недостаточную производительность соединений глобальной сети, а также постоянно развивающиеся онлайн-приложения, требуется применять целостный подход. Такой подход позволяет достичь наилучшей оптимизации веб-приложений (WPO, web performance optimization) для любого типа клиентского ПО, расположенного в любой точке сети.

В данном документе особое внимание уделяется бизнес-мотивации, которая побуждает организации непрерывно оптимизировать производительность своего сайта, а также тому, как технология Radware FastView™ и модуль мониторинга производительности помогут эффективно ускорить работу корпоративного сайта.

Radware FastView™ – это ориентированная на результат технология ускорения веб-приложений, встроенная в контроллеры доставки приложений ADC Atleon, которая предлагает целостный подход к исследованию различных факторов, влияющих на производительность веб-приложений. Согласно результатам тестирования на различных веб-сайтах и корпоративных веб-приложениях, технология FastView уменьшает критический важный показатель – время отклика – для максимальной оптимизации бизнес-услуг. Путем оптимизации работы различных элементов, относящихся к цепи доставки приложений, а также разгрузки проблемных мест сервера веб-приложений, сети и клиентских браузеров, технология FastView обеспечивает повышение QoE (качества восприятия) для всех конечных пользователей.

Чтобы позволить заказчикам измерить производительность веб-приложения и гарантировать, что она в любой момент времени соответствует ожидаемому уровню предоставления услуг (SLA), Radware также предлагает новый модуль мониторинга производительности приложений (Application Performance Module, APM), который позволяет использовать проактивный подход к мониторингу SLA веб-сайта и веб-приложений и поддержанию его на высоком уровне.

Почему производительность веб-сайта имеет критическое значение

Коммерческие организации понимают, что производительность сайта имеет решающее значение для успеха их дела. Каждый может вспомнить свой опыт, когда при попытке открыть сайт после нескольких секунд ожидания вы теряли терпение и переходили к другим занятиям. В данной главе будет показана связь между производительностью веб-сайта и различными показателями успеха бизнес-предприятия, такими, как количество просмотров страниц (page view), доход с одного пользователя (revenue per user), а также удовлетворенность пользователей (user satisfaction).

Многие ведущие онлайн-бизнесы измерили различные факторы, показывающие, как производительность веб-сайта влияет на бизнес-метрики. Например:

- Представители компании Amazon обнаружили, что уменьшение времени загрузки страницы на 100мс привело к увеличению дохода от продаж их онлайн-магазина на 1%.
- Исследование компании Google показало, что повышение времени загрузки страницы на 500мс привело к сокращению количества поисковых запросов на 25%. Сокращение количества поисковых запросов на поисковой машине Google приравнивается к сокращению количества показов объявлений (и кликов по ним), что оказывает непосредственное влияние на их бизнес.
- Еще одно интересное исследование от компании Microsoft Bing показало, что повышение времени загрузки страницы на 1 секунду привело к падению среднего уровня дохода с пользователя на 2.8%, а также сокращению удовлетворенности конечных пользователей на 1.6%.



Рисунок 1 – Взаимосвязь задержки при загрузке страницы и таких показателей, как доход с одного пользователя и удовлетворенность пользователей (источник: BING)

- Еще больший интерес вызывает тот факт, что компания Microsoft проводила данное исследование в течение 3 недель, показав, что отрицательное влияние на коммерческую деятельность увеличилось в конце тестового периода по сравнению с начальным этапом, поскольку росло недовольство пользователей.
- Согласно исследованию, проведенному исследовательской группой Aberdeen, задержка времени отклика на 1 секунду не только сократила количество просмотров страниц на 11% и коэффициент конверсии на 7%, но также сократило показатель удовлетворенности пользователей на ошеломляющие 16%!
- Проанализировав показатель уходов со страницы (page abandonment) на более чем 150 веб-сайтах и 150 миллионах страниц, компания Gomez выяснила, что повышение времени отклика страницы с 2 до 10 секунд увеличило коэффициент ухода со страницы на 38%.



Рисунок 2 – Усредненные показатели влияния задержки времени отклика на 1 секунду (источник: Aberdeen Group)

Такие наблюдения позволяют сделать вывод, что существует прямая взаимосвязь между временем отклика веб-сайта и успешностью коммерческой деятельности. Вне зависимости от того, ведем ли мы разговор конкретно о таких показателях, как количество просмотров страницы, доход от онлайн-услуг, количество уходов со страницы или удовлетворенность пользователей, на все перечисленные показатели оказывает количественное влияние время отклика веб-сайта.

Факторы, влияющие на производительность веб-сайтов

Говоря о производительности веб-сайта, организации признают важность его доступности, безопасности и времени отклика.

Тем не менее, существует широкий спектр факторов, которые негативно сказываются на производительности веб-сайта и восприятии конечного пользователя. Данный спектр охватывает все этапы доставки приложения, начиная с клиентского ПО и используемого браузера, географического местоположения пользователя, способа получения

доступа к сети, оснащения центра обработки данных (ЦОД), типа контента, способа его использования, и заканчивая качеством приложения, обслуживающего пользователя. Ниже приведены основные тенденции, оказывающие влияние на производительность веб-сайтов:

Изменение веб-контента: формы веб-контента развились от текстовых страниц к более продвинутым приложениям с богатым динамическим контентом, который часто включает объемные объекты: изображения, звук, видеоклипы и объемные графические рекламные объявления.

HTTP Archive, сайт, который отслеживает тенденции развития более ста тысяч веб-сайтов, демонстрирует, что растет не только средний объем веб-страницы, но также и количество элементов на одной странице. Это вынуждает клиентское ПО поддерживать открытыми большее количество соединений на более длительный период времени. В конечном счете, это перегружает инфраструктуру приложения.



Рисунок 3 - Общий объем передаваемого контента и общее количество запросов страницы (источник HTTP archive, Май 2012)

Изменение способа потребления веб-контента: пользователи больше не привязаны к настольному ПК, подключенному к сети посредством проводного соединения по телефонной линии из дома или из офиса. Потребитель может использовать новый тип браузера, запущенного на мобильном устройстве последней модели, для подключения к интернету через сеть оператора мобильной связи с более низкой пропускной способностью и более высоким временем приема-передачи трафика (round trip times - RTT), а также через беспроводную сеть офиса с высокой пропускной способностью и низким временем приема-передачи трафика (RTT). Кроме того, предприятия сталкиваются с новой тенденцией, когда сотрудники используют собственные устройства (концепция Bring Your Own Device - BYOD), что обязывает их адаптировать и оптимизировать веб-приложения к растущему количеству различных устройств и браузеров.

Увеличенное время ожидания вследствие наличия узких мест протокола: используемый веб-приложением протокол TCP или UDP в качестве протокола транспортного уровня оказывает прямое влияние на его производительность. Производительность TCP, протокола транспортного уровня, который используется

приложениями, работающими на основе протокола HTTP, значительно отличается при работе в сетях с низким RTT, таких, как проводные DSL соединения, от производительности в беспроводных мобильных сетях с высоким RTT, таких как 3G, 4G мобильные сети. Пользователь, обращающийся к своему любимому сайту заказа товаров с портативного компьютера дома, ожидает точно такое же качество восприятия (QoE) при обращении к тому же самому сайту с мобильного устройства. Смена устройства не должна влиять на время отклика. Тем не менее, даже если веб-сайт оптимизирован для мобильных устройств, использование протокола TCP может привести к повышению времени отклика и низкой пропускной способности при работе в мобильной сети.

Недостаточное кэширование контента: общим методом, который широко используется для улучшения потребительского качества, является кэширование контента. Кэширование представляет собой расположение информации ближе к пользователю – на близлежащем узле или в кэш-памяти браузера пользователя. По данным сервиса HTTP Archive, 58% контента на протестированных сайтах и страницах до сих пор должным образом не промаркировано для кэширования. Следовательно, вместо передачи такого контента от более быстрого источника, такого, как ADC-сервер или кэш-память браузера; большее количество контента передается от сервера к пользователю. В результате мы получаем увеличенное время отклика веб-приложения и более высокую нагрузку на сервер.

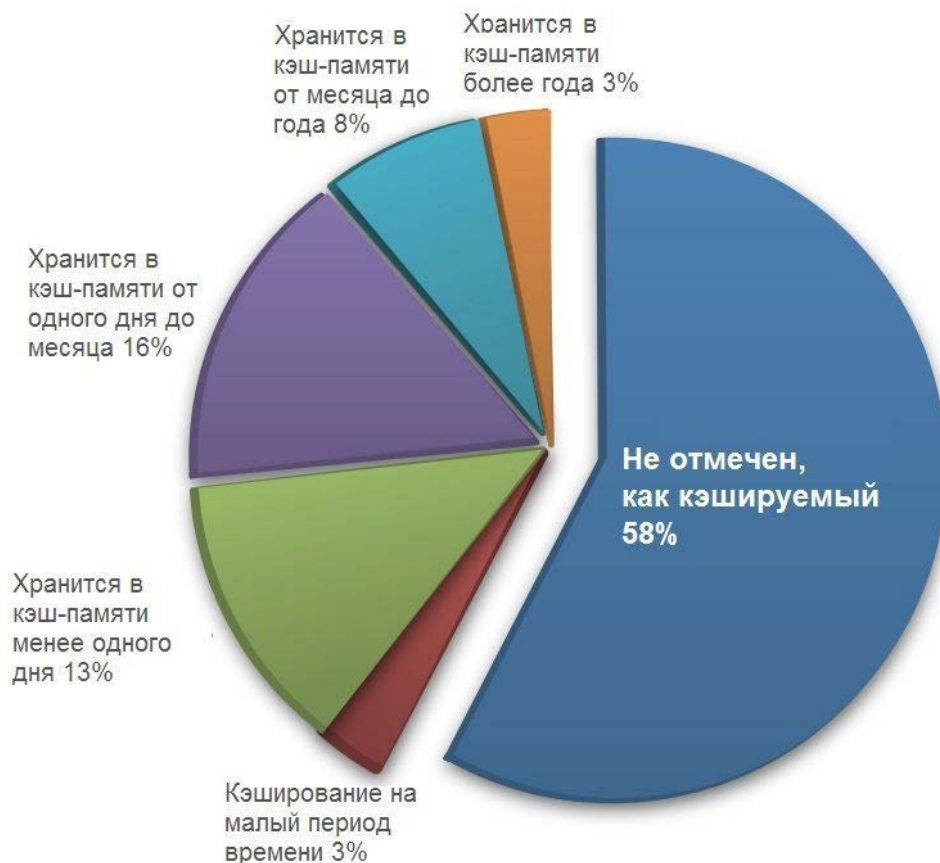


Рисунок 4 - маркировка контента для кэширования (источник: HTTP Archive)

Оптимизация кода веб-приложений: оптимизация кода может значительно повысить QoE конечных пользователей. Например, веб-страницы, разработанные таким образом, что информация вместе с запрашиваемыми объектами уже на начальных этапах загрузки предоставляется в удобном для чтения виде, по ощущениям пользователя уменьшит время отклика и QoE. Тем не менее, большинство организаций не могут позволить себе открыть код веб-приложений из-за существующих рисков и его высокой стоимости.

Помимо этого, к хорошим практикам написания кода относится построение разделенных CSS файлов в секции страницы, а также добавление как можно большего количества комментариев к коду. Однако такая практика увеличивает число объектов на странице и, как показано выше, повышает время задержки при загрузке сайта. Оптимизация производительности веб-приложений потребует решения каждой из вышеуказанных проблем одновременно, для каждого профиля пользователя (то есть, типа устройства, используемого браузера и т.д.) и приложения.

Резкое повышение производительности веб-сайта с Radware FastView™

Традиционно, повышение QoE является одной из функций ADC. Это решение обеспечивает высокую доступность приложения путем интеллектуальной балансировки нагрузки серверов и проверки работы приложений, чтобы автоматически обойти возможные неисправности, а также Глобальной балансировки нагрузки (Global ServerLoad Balancing - GSLB) для резервирования сайтов. В сочетании с запатентованной технологией Radware Proximity она гарантирует, что конечный пользователь будет обслуживаться наиболее близко расположенным и наименее загруженным ЦОД для уменьшения времени отклика. Radware ADC также обеспечивает SSL-разгрузку и сжатие контента для дальнейшей акселерации производительности веб-сайтов. С развитием техник оптимизации веб-приложений (WPO) у ADC появляются дополнительные возможности. Radware FastView™ представляет собой технологию веб-оптимизации (WPO), встроенную в платформы Alteon ADC, которая расширяет возможности ADC по ускорению приложений. FastView предлагает согласованный набор возможностей, которые значительно ускоряют производительность веб-сайтов и веб-приложений при одновременном снижении расходов на их развитие и накладных расходов на поддержание такой оптимизации.

Стив Саудерс, всемирно признанный специалист в области оптимизации производительности веб-сайтов, опубликовал набор из **14 золотых правил ускорения веб-сайтов**¹ Технология Radware FastView охватывает 8 из 14 правил, которые применимы для всех пользователей, не требуя усилий, необходимых для осуществления и поддержания этих наилучших практик в каждом веб-приложении.

Функционал и преимущества технологии FastView

Набор функций технологии FastView можно разбить на две покрываемые проблемные области в производительности веб-сайта: она оптимизирует передачу трафика для сокращения времени задержки при передаче по WAN, а также она сокращает количество контента, передаваемого конечному пользователю.

Оптимизация на транспортном уровне

Всем WAN-соединениям присуще наличие времени задержки при обмене данными между пользователем и сайтом, которое основывается на сетевой задержке, количестве участков сети (на которое влияет географическая

¹ Источник Steve Souders 14 golden rules for accelerating websites and web applications: <http://jstesousders.com/jhpwsjrules.php>

удаленность объектов друг от друга, а также используемые каналы связи), а также качестве пересекаемых сетей. Такая задержка умножается на количество циклов приема-передачи, необходимых для доставки веб-приложения пользователю. Существует два способа, с помощью которых FastView решает эту проблему: сокращение количества циклов приема-передачи и сокращение времени задержки WAN-соединения.

Объединение объектов – количество объектов на веб-странице постоянно растет. Двусторонняя TCP транзакция (для одного объекта) в сочетании с присущей WAN-соединениям задержкой приводит к появлению коэффициента задержки, который умножается на количество объектов на веб-странице.

Radware FastView уменьшает количество запросов к серверу, которые требуются для загрузки веб-страницы, путем группирования нескольких CSS объектов и Java Script объектов в один. Дальнейшее сокращение достигается за счет обработки небольших CSS и Java-скрипт объектов, где потери при транзакции превышают размер объекта. В таких случаях FastView внедряет эти объекты в HTML-документ, еще больше сокращая количество транзакций и пропускную способность, которые требуются для доставки этих объектов. Путем объединения CSS и JS объектов можно уменьшить количество запросов на страницу в среднем на 20 - 30%, ускоряя при этом загрузку страниц, как при первом, так и при повторном обращении. Пользователи мобильных устройств в большей степени ощущают пользу от сокращения количества соединений, поскольку для мобильных сетей характерно большее время задержки.

Оптимизация TCP-протокола – FastView оптимизирует работу используемого протокола и делает ее более эффективной путем уменьшения времени задержки при передаче данных по WAN-соединению. Помимо существующего набора мер по оптимизации работы TCP и мультиплексирования TCP-соединения (которое применяется к пользовательским TCP-соединениям и существующим открытым TCP-соединениям между ADC и сервером) FastView использует усовершенствованный алгоритм для предотвращения перегрузки TCP. Любая из данных мер может быть применена ко всем пользователям и не требует взаимодействия пользователей. Меры по оптимизации протокола TCP, используемые платформой Alton, включают в себя:

- TCP Selective Acknowledgments Options (RFC 2018, 2883)
- TCP Congestion Control (RFC 5681)
- Enhanced Loss Recovery using Limited Transmit and New Remo Algorithm (RFC 3042, 3782)

Оптимизация мер по предотвращению возникновения перегрузки - FastView использует **алгоритм улучшения работы TCP Hybla**, который предлагает значительное улучшение производительности, особенно для пользователей мобильных устройств, используя сети с более высоким уровнем круговой задержки и более высоким уровнем потери пакетов. Этот новый механизм оптимизации TCP протокола обеспечивает более высокую и согласованную пропускную способность, повышая общую производительность веб-сайтов, которые используют данный механизм.

Сокращение передаваемого контента

Другой набор возможностей, предлагаемый технологией FastView для ускорения производительности веб-приложений, работает с количеством информации, передаваемой от серверов к пользователю.

Сокращение контента – удаление любого ненужного контента, такого как, например, комментарии и разделители,

используемые программистами в Java-скрипт и CSS коде. URL-ссылки, встроенные в HTML-документ или код, часто представлены в виде абсолютных ссылок, которые занимают больше места, чем относительные URL-ссылки. FastView обрезает URL, уменьшая размер передаваемого контента. Сокращение контента влияет на время загрузки при первом и повторном обращении к странице.

Динамическое кэширование контента – мощный способ сокращения количества контента, передаваемого от сервера пользователю – это не передавать его вовсе. FastView включает новую технологию распознавания контента, который может быть кэширован, даже если он не отмечен как таковой, или не обладает некоторыми атрибутами для эффективного кэширования. Механизм кэширования FastView включает два уровня динамического кэширования объектов:

Первый уровень представляет собой кэширование контента в Alteon ADC – благодаря кэшированию объектов ADC отвечает на запросы браузера гораздо быстрее, поскольку такие запросы не отправляются на сервер, что позволяет увеличить скорость обработки других запросов. Более того, встроенная в ADC функция глобальной балансировки нагрузки (GSLB) гарантирует, что пользователь будет перенаправлен на ближайший и наименее загруженный ЦОД и будет обслуживаться посредством данных, сохраненных в кэш-памяти ADC. Динамическое кэширование в ADC в сочетании с функцией GSLB непосредственно уменьшает время отклика страниц, запрашиваемых пользователем в первый раз ("первое обращение к странице").

Второй уровень представляет собой кэширование контента в браузере конечного пользователя: FastView добавляет флаги к объектам, которые он определяет, как кэшируемые, и передает команды браузеру сохранить их для будущего использования – такого, как повторное обращение к той же самой странице или даже повторное использование объектов на других страницах одного и того же сайта.

Преимущества динамического кэширования объектов FastView представляют двойную ценность: они снижают время отклика веб-приложения, а также разгружают сервер веб-приложения.

Сжатие веб-трафика – сокращение размера передаваемого контента для уменьшения времени отклика веб-страниц, особенно для ответов на запросы внутренней базы данных, которая включает в себя большое количество данных. Сжатие не только уменьшает общее время отклика на запросы конечного пользователя, но также сокращает размер ответов вплоть до 90%, что приводит к уменьшению потребляемой пропускной способности. Сжатие контента влияет на время отклика как страниц, просматриваемых впервые, так и страниц, к которым пользователь обратился повторно.

Разгрузка Secure Sockets Layer (SSL) – терминация SSL сессий в ADC и снятие этой функции с сервера имеет критическое значение для ускорения веб-приложений. Это позволяет выполнить все из перечисленных выше методов (например, кэширование, объединение CSS и JS и др.) на зашифрованном трафике. Это означает, что ускорится загрузка не только домашней страницы коммерческого сайта, но также и страницы оплаты, которая шифруется посредством SSL. Терминация SSL сессий в аппаратном акселераторе SSL, который встроен в ADC, также позволяет осуществлять быструю обработку SSL-транзакций, в то же время разгружая ресурсы сервера, и тем самым уменьшая количество серверов, необходимых для обслуживания одного и того же количества пользователей.

Практическое значение FastView

Путем реализации передового опыта, FastView покрывает всеобъемлющее количество различных узких мест в работе веб-приложений, что позволяет уменьшить время отклика веб-приложений для всех пользователей, начиная с 1-го посещения страницы. Кроме того, эта технология гарантирует реальные результаты в рамках более быстрых продаж на рынке и ускорения веб-приложений для всех пользователей в любом месте и в любое время.

Путем внедрения технологии FastView в Alteon ADC предприятия могут мгновенно использовать ее преимущества для всех приложений, которые они контролируют посредством ЦОД. Не требуется производить каких-либо изменений в развертывании веб-приложений и какой-либо интеграции, что, в конечном счете, приводит к значительному сокращению усилий, затрат и рисков.

Мониторинг ускорения приложений

Для того, чтобы обеспечить определенную производительность веб-приложений в соответствии с SLA, который обещан пользователям, требуется настроить сбор информации о производительности приложения, а также о фактическом качестве с точки зрения конечных пользователей (QoE). Модуль Radware APM дает администраторам веб-приложений больший контроль над действиями, которые они предпринимают с целью улучшения производительности приложений. Возможности модуля заключаются в следующем:

1. Измерение влияния технологий акселерации на производительность веб-приложения. Обладая возможностью просмотреть подробные отчеты о прошлых показателях производительности, можно сопоставить результаты коммерческой деятельности с различными показателями производительности приложения.
2. Отслеживание фактического QoE конечных пользователей с подробной классификацией реальных пользователей согласно их географическому местоположению, а также с просмотром реальных транзакций приложения, статистикой возникновения ошибок в реальном времени и отправкой уведомлений для проактивного реагирования на ухудшение производительности приложения.

Модуль Radware APM предоставляется как часть APSolute Vision, системы централизованного управления Radware. Он не требует какой-либо интеграции в инфраструктуру приложений; поэтому он не приносит каких-либо дополнительных затрат, не требует создания скриптов для приложения, имитации искусственных транзакций и т.д.

Измерения производительности основаны на реальных пользовательских транзакциях (в отличие от мониторинга искусственно созданных базовых транзакций с правами только на чтение), а также реальных ошибках, которые регистрируются продвинутой системой сбора данных и обработаны аналитическим механизмом; являющимся частью , системы центрального управления APSolute Vision Radware. Механизм анализа предлагает широкий спектр подробных отчетов, а также настраиваемые отчеты для настройки измерений производительности в соответствии с конкретными целями и бизнес-метриками.



Рисунок 5: Методология Radware для измерения производительности приложений

Измерения производительности веб-приложений Radware APM охватывают все приложения и все сервера, трафик которых обрабатывает Alteon ADC. Кроме того, такие измерения могут быть собраны из нескольких устройств Alteon, обеспечивая обзор производительности приложения и QoE конечных пользователей между несколькими ЦОД. Это позволяет быстрее устранить неисправности и проанализировать причины задержки в работе приложений.

Любые измерения транзакции разбиваются на 3-этапную цепочку доставки приложения:

1. Время отклика между сервером и ADC – время, затрачиваемое в ЦОД.
2. Сетевое время – время, затрачиваемое на то, чтобы ответ сервера достиг клиента
3. Время на отображение в браузере – время, затрачиваемое браузером на загрузку и отображение страницы



Рисунок 6 - Подробные отчеты о производительности приложения и QoE пользователя

Решение для сквозного мониторинга производительности приложения фиксирует фактическое QoE пользователя при фактической задержке на выполнение транзакции и обнаружении ошибок в реальном времени. Решение обеспечивает:

- Наглядность, требуемую для непрерывного поддержания веб-приложением коммерческих условий SLA
- Эффективный инструмент устранения сбоев, а также точной локализации проблем с производительностью. Все выполняется в реальном времени.
- Упрощенное внедрение, для которого не требуется специально созданных скриптов, что снижает общие затраты.

Заключение

Вне зависимости от того, используют ли сегодняшние пользователи сети Интернет общедоступный веб-сайт или корпоративное веб-приложение, при разговоре о производительности счет идет на каждую секунду. Корпоративные веб-приложения должны соответствовать высоким стандартам, чтобы обеспечить высокий уровень удовлетворенности пользователей.

Технология Radware FastView предлагает комплексное решение, ориентированное на применение в ЦОД, которое покрывает множество проблемных мест, относящихся к производительности приложений. Целью FastView является достижение наиболее эффективного ускорения веб-приложений для всех типов устройств, браузеров и WAN-соединений.

Используя встроенный модуль мониторинга производительности приложений (Application Performance Monitoring), администраторы веб-приложений могут отслеживать производительность своих приложений, чтобы проактивно обеспечивать наилучшие условия ведения бизнеса.