

# Руководство для покупателей балансировщика нагрузки в 2020 году

Современные программно-определяемые архитектуры меняют представление о стратегиях балансировки нагрузки, стимулируя переход от аппаратных платформ и виртуальных устройств к матрице балансировки нагрузки на основе контроллеров

## Содержание

ЭВОЛЮЦИЯ ПРИЛОЖЕНИЙ И ИНФРАСТРУКТУРЫ	3
ПОТРЕБНОСТЬ В НОВЫХ СТРАТЕГИЯХ БАЛАНСИРОВКИ НАГРУЗКИ	3
ОРГАНИЗАЦИИ, ОРИЕНТИРОВАННЫЕ НА ПРИЛОЖЕНИЯ	4
ИСПОЛЬЗОВАНИЕ РУКОВОДСТВА	4
ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ	4
Службы приложений	4
Балансировка нагрузки	4
Многооблачные среды	4
ПРЕДВАРИТЕЛЬНАЯ ОЦЕНКА ТЕКУЩИХ ВОЗМОЖНОСТЕЙ БАЛАНСИРОВКИ НАГРУЗКИ	5
КОНЦЕПЦИЯ СОВРЕМЕННЫХ СЛУЖБ БАЛАНСИРОВКИ НАГРУЗКИ	5
Плоскость данных служб приложений	5
Контроллер служб приложений и плоскость управления	6
Диспетчер жизненного цикла: автоматизированное создание, размещение, масштабирование и высокая доступность служб	6
Модуль анализа: мониторинг производительности приложений, поиск и устранение неполадок, аналитика	6
Репозиторий политик: конфигурации и политики	6
Облачные коннекторы: интеграция и возможности многооблачных сред	6
API-интерфейсы REST: автоматизация для ИТ-отделов и самообслуживание для разработчиков	7

## ЭВОЛЮЦИЯ ПРИЛОЖЕНИЙ И ИНФРАСТРУКТУРЫ

Стремительно растущий спрос на приложения и современные архитектуры приложений, такие как микрослужбы, заставляет пересматривать методы управления корпоративной инфраструктурой. Организации выбирают подходы, ориентированные на приложения, выделяя значительные средства и ресурсы для непрерывного предоставления услуг, внедрения методов DevOps и автоматизации рутинных ИТ-процессов и задач. Потребность в приложениях продолжает расти, поэтому важно еще раз проанализировать предоставление служб приложений в различных центрах обработки данных и облачных средах. См. рис. 1. Кроме того, сложность и масштаб микрослужб затрудняют управление трафиком. Традиционные балансировщики нагрузки не справляются с новыми задачами, включая входящий вертикальный трафик и шлюзы между кластерами контейнеров.

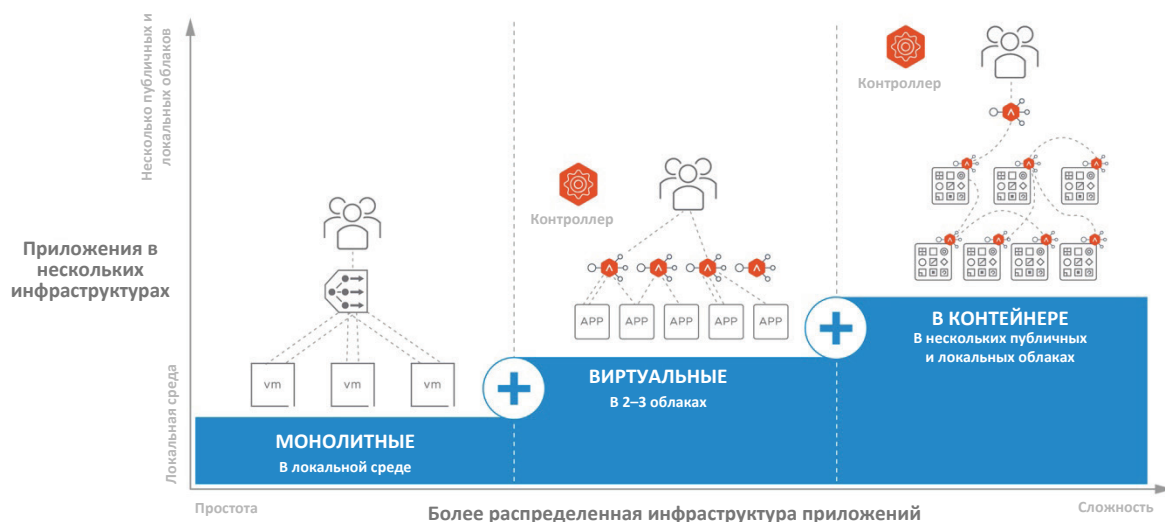


Рис 1. Современные вычислительные среды: эволюция архитектур и инфраструктуры приложений

## ПОТРЕБНОСТЬ В НОВЫХ СТРАТЕГИЯХ БАЛАНСИРОВКИ НАГРУЗКИ

Стратегии балансировки нагрузки, основанные на устройствах, не позволяют адаптироваться к быстрым изменениям в динамичных средах. Устаревшие балансировщики нагрузки, также называемые контроллерами предоставления приложений (ADC), не могут подстроиться под меняющиеся требования, что во многих случаях приводит к задержкам развертывания приложений, чрезмерным расходам и избыточной инициализации. Если говорить о развертывании новых приложений и установке обновлений, то в большинстве организаций наблюдается ситуация, показанная на рис. 2. Организациям необходимо соблюдать цели по непрерывному предоставлению, поэтому они не хотят ждать инициализации приложений или виртуальных устройств и ручной настройки конфигураций, без которых невозможно провести развертывание или обновление приложений. Организации хотят использовать балансировку нагрузки как услугу. Кроме того, аппаратные контроллеры предоставления приложений не отвечают потребностям в масштабировании и многооблачных средах, что приводит к переосмыслению стратегий управления трафиком в различных средах.

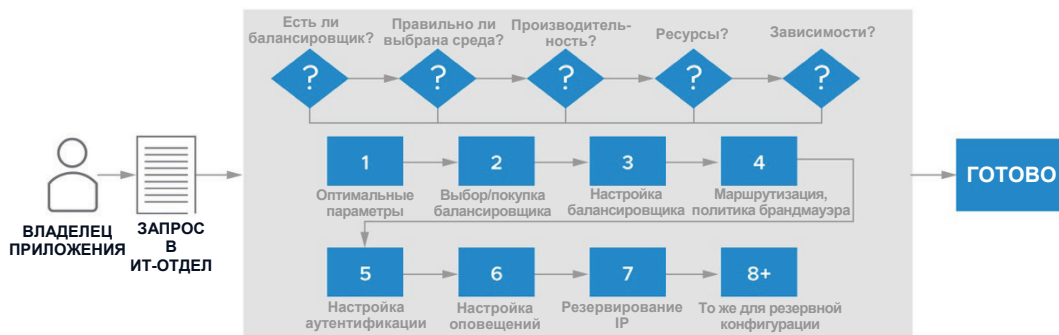


Рис. 2. Службы приложений отстают от темпов развития бизнеса

## ОРГАНИЗАЦИИ, ОРИЕНТИРОВАННЫЕ НА ПРИЛОЖЕНИЯ

Организации, ориентированные на приложения, выбирают сервисы предоставления приложений, обеспечивающие гибкость эксплуатации и возможность непрерывного предоставления в различных средах. Им требуются сервисы на основе данных, поддерживающие возможности самообслуживания для владельцев приложений и автоматизацию для ИТ-отделов.

Эксперты IDC Брэд Кейсмор (Brad Cascemore) и Мэттью Марден (Matthew Marden) представили свой взгляд на эту тенденцию балансировки нагрузки в тематической статье «*A Study of Enterprises Using Next-Generation Application Delivery*».

«В связи с насущной необходимостью цифровой трансформации крупные компании и другие организации внедряют стратегии многооблачных сред, значительно повышающие адаптивность бизнеса и усиливающие конкурентные преимущества.

К сожалению, существующая инфраструктура предоставления приложений была разработана для систем "клиент/сервер" и не отвечает новым потребностям. Физические аппаратные контроллеры предоставления приложений (ADC) имели смысл, когда между приложениями и серверами в корпоративном центре обработки данных существовала взаимосвязь "один к одному". Однако они плохо приспособлены к облачным средам, где часто требуется гибкое вертикальное масштабирование в соответствии с меняющимися требованиями приложений и моделями трафика. Даже появление виртуальных контроллеров предоставления приложений (vADC), которые с точки зрения архитектуры по-прежнему являлись устройствами, не решило все эти проблемы.

Следовательно, возникла потребность в более адаптивной, эластичной, распределенной и согласованной архитектуре предоставления приложений. Наступил период, когда контроллеры ADC должны перейти в категорию служб приложений, эластичных пулов сетевых служб и служб безопасности (так называемых служб уровня 4–7), которые обеспечивают оптимальное развертывание приложений, а также динамическую поддержку доступности, производительности и безопасности приложений».

## ИСПОЛЬЗОВАНИЕ РУКОВОДСТВА

Данное руководство покупателя балансировщиков нагрузки основано на исследованиях, лучших методиках и беседах с администраторами сети и специалистами по эксплуатации ИТ-систем из более чем 500 крупнейших компаний мира. В руководстве представлены рекомендации по выбору служб приложений, соответствующих требованиям современных центров обработки данных и инициативам по внедрению публичного облака.

Информация сгруппирована по разделам, начиная с предварительной оценки текущих возможностей предоставления приложений. Данные, собранные на этапе предварительной оценки, можно использовать при разработке программно-определяемой стратегии служб приложений, которая позволит улучшить автоматизацию служб приложений и оптимизировать процессы. Контрольный список в конце руководства поможет определить главные приоритеты при выборе решений для корпоративных служб приложений.

## ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

### Службы приложений

Службы балансировки нагрузки, безопасности, анализа, мониторинга производительности и визуализации, предоставляемые традиционным корпоративным приложениям и приложениям микрослужб.

### Балансировка нагрузки

Процесс распределения сетевого трафика между несколькими серверами. Этот процесс предотвращает перегрузку отдельных серверов. Балансировка нагрузки повышает время отклика и отказоустойчивость приложений. Кроме того, она повышает доступность приложений и веб-сайтов для пользователей. Современные приложения не могут работать без балансировщиков нагрузки. Балансировка нагрузки обеспечивается либо устройствами (аппаратными или виртуальными), либо новыми программно-определяемыми решениями.

### Многооблачные среды

Использование нескольких облачных вычислительных ресурсов и служб хранения данных в одной сетевой архитектуре. Этот термин относится к распространению облачных ресурсов, программного обеспечения, приложений и других ресурсов в нескольких облачных средах. Многооблачная среда призвана устранить зависимость от одного поставщика или экземпляра облака.

## ПРЕДВАРИТЕЛЬНАЯ ОЦЕНКА ТЕКУЩИХ ВОЗМОЖНОСТЕЙ БАЛАНСИРОВКИ НАГРУЗКИ

Оцените текущие возможности корпоративных служб приложений. Запишите ответы на следующие вопросы, чтобы понять свои потребности и выявить аспекты, требующие улучшения.

1. Что используется в качестве основного решения для балансировки нагрузки: аппаратное устройство ADC, виртуальное устройство ADC, балансировщик нагрузки поставщика облачных услуг или решение с открытым исходным кодом?
2. Соответствует ли текущее решение ADC вашим потребностям по следующим критериям?
  - а) автоматизация
  - б) поддержка многооблачных сред
  - в) применимость к сценариям разработки/тестирования
  - г) производительность и масштабируемость
  - д) совокупная стоимость владения
3. Как часто вам приходится вносить изменения в конфигурацию ADC? Сколько дней или недель в среднем занимает выполнение запросов на изменение?
4. Какой подход применяется для поиска и устранения неполадок приложений и выявления таких проблем, как задержки, снижение производительности и проблемы безопасности?
5. Используются ли в текущем решении ADC специализированные аппаратные устройства? Позволяет ли это решение использовать высокопроизводительные серверы x86?
6. Как выполняется масштабирование балансировщиков нагрузки? Сколько времени занимает этот процесс? Требуется ли приобретение и инициализация дополнительных аппаратных устройств?
7. Как часто наблюдается избыточная инициализация балансировщиков нагрузки? Насколько балансировщики нагрузки соответствуют объемам трафика в режиме реального времени?
8. Есть ли у вас возможность централизованного администрирования всех ресурсов балансировки нагрузки в центре обработки данных или в облаках?
9. Требуются ли разные решения балансировки нагрузки для облака и центра обработки данных?
10. Позволяет ли ADC автоматически масштабировать приложения с учетом пороговых значений трафика?

## КОНЦЕПЦИЯ СОВРЕМЕННЫХ СЛУЖБ БАЛАНСИРОВКИ НАГРУЗКИ

Службы приложений, разработанные с учетом программно-определяемых принципов, должны отражать цели автоматизации и программируемости для центра обработки данных и инициатив по внедрению облака. Всё больше организаций внедряют стратегии масштабирования веб-приложений, поэтому службы приложений, основанные на специализированном оборудовании или привязанные к одной среде (центру обработки данных или облаку), становятся ограничением. Чтобы удовлетворить как текущие, так и новые требования организаций, ориентированных на приложения, рекомендуется изучить представленную ниже программно-определяемую структуру служб приложений (см. рис. 3). Главный архитектурный принцип этой концепции заключается в разделении плоскости управления и плоскости предоставления служб. Позднее вы сможете убедиться, что это предоставляет для системы гораздо больше возможностей по сравнению с исходными службами балансировки нагрузки. Рассмотрим компоненты этой архитектуры.

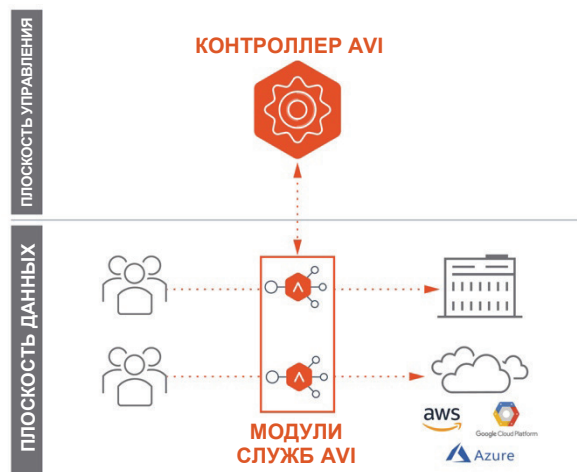
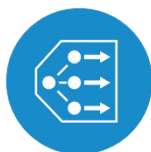


Рис. 3.



### Плоскость данных служб приложений

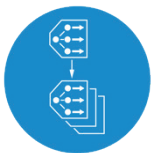
Предоставление служб в программно-определяемых архитектурах служб приложений осуществляется на распределенной плоскости данных, состоящей из программных балансировщиков нагрузки, которые работают на любых стандартных серверах x86, виртуальных машинах, в контейнерах и в публичных облачных средах. Балансировщики нагрузки в этой плоскости данных располагаются в соответствии с трафиком приложений. Они непрерывно собирают и передают данные телеметрии приложений в плоскость управления. Балансировщики нагрузки могут быть распределены по центру обработки данных или облачной среде и представлять собой пул ресурсов, которые можно развернуть везде, где требуются службы приложений. Благодаря более высокой вычислительной мощности серверов x86 эти программные балансировщики нагрузки позволяют снизить затраты и

обеспечивают эластичность, высокую производительность и доступность служб. Программные балансировщики нагрузки можно развернуть для предоставления служб, приближенных к фактическому приложению, или даже отдельных служб для каждого приложения. Благодаря этому устраняется необходимость в развороте трафика на периметр сети, как в случае с аппаратными устройствами. Более современные архитектуры приложений, включая микрослужбы, выигрывают от этого подхода благодаря динамичным и точным службам приложений, которые предоставляются программно-определяемой архитектурой.



### Контроллер служб приложений и плоскость управления

Контроллер представляет плоскость управления и является центром управления и визуализации всей матрицы. Он автоматизирует администрирование и масштабирование распределенных программно-определяемых служб приложений на плоскости данных. В отличие от диспетчеров экземпляров, используемых в традиционных решениях и требующих ручной настройки и администрирования, контроллер централизованно управляет всеми службами балансировки нагрузки, которые предоставляются на всей плоскости данных. Контроллер принимает и анализирует непрерывный поток данных телеметрии приложений, отправляемых распределенными балансировщиками нагрузки из различных сред. На основе полученных данных принимаются решения о размещении служб, автоматическом масштабировании и высокой доступности для каждого приложения. Балансировщики нагрузки можно развернуть в нескольких средах под управлением контроллера (в центре обработки данных или публичном облаке). Ниже перечислены основные компоненты контроллера.



### Диспетчер жизненного цикла: автоматизированное создание, размещение, масштабирование и высокая доступность служб

Поскольку элементы плоскости данных размещаются на типовом оборудовании, их можно динамически развертывать или сокращать везде, где требуются службы или горизонтальное масштабирование служб. Управление жизненным циклом балансировщиков нагрузки в плоскости данных осуществляется с помощью диспетчера жизненного цикла, в соответствии с политиками, настроенными администраторами в плоскости управления. Плоскость управления также помогает отслеживать «работоспособность» системы и реагирует на сбои компонентов и изменения приложений в плоскости данных. Благодаря расширенным возможностям программных балансировщиков нагрузки можно настроить конфигурации высокой доступности «активный-активный» с минимальными затратами по сравнению с традиционными балансировщиками нагрузки. Кроме того, когда требуется масштабирование балансировщиков нагрузки в соответствии с потребностями приложений, в плоскости управления добавляются и автоматически настраиваются дополнительные ресурсы.



### Модуль анализа: мониторинг производительности приложений, поиск и устранение неполадок, аналитика

Одно из главных преимуществ программно-определяемой платформы служб приложений — это возможность получать ценные данные о приложениях с помощью средств анализа в плоскости управления. Модуль больших данных в составе контроллера анализирует данные телеметрии приложений, полученные из распределенной плоскости данных, и предоставляет подробную информацию о транзакциях, включая время полного цикла передачи данных от конечного пользователя до серверного приложения. Система записывает все данные транзакций и воспроизводит их по требованию администратора или владельца приложения. Центральная консоль управления обеспечивает точный анализ проблем и сбоев приложений в течение определенных периодов времени. Модуль анализа также предоставляет подробные сведения об уровне безопасности (включая состояние DDoS-атак, используемые сертификаты SSL и показатели работоспособности системы безопасности), а также сведения о клиентах, включая подключение устройств, информацию о географическом положении конечных пользователей и длительности транзакций. Благодаря вышеперечисленным возможностям эти службы представляют больше преимуществ администраторам, позволяя сэкономить время и усилия. Администраторам больше не нужно использовать несколько специализированных решений для поиска и устранения неполадок или мониторинга производительности приложений.



### Репозиторий политик: конфигурации и политики

В репозитории политик хранятся все конфигурации системы, включая сведения обо всех виртуальных службах и элементах пула. Многим администраторам сети приходится тщательно документировать и поддерживать конфигурации виртуальных служб (иногда с помощью электронных таблиц), чтобы обеспечить точное представление развернутых служб. Репозиторий политик и конфигураций позволяет отказаться от действий, выполняемых вручную, и администраторы могут видеть всю карту развертывания. Репозиторий автоматически обновляется при перемещении рабочих нагрузок приложений или изменении IP-адресов.



### Облачные коннекторы: интеграция и возможности многооблачных сред

Программно-определяемая платформа служб приложений должна поддерживать многооблачные среды с единой архитектурой приложений. Облачные коннекторы в плоскости управления упрощают развертывание готовых служб приложений в любом центре обработки данных, а также в частном или публичном облаке. Облачные коннекторы изначально интегрированы с точками управления облачных платформ, таких как



Microsoft Azure, AWS, Google Cloud и OpenStack, контроллерами SDN, такими как VMware NSX Manager, а также диспетчерами виртуальных машин, такими как VMware vCenter. В более современных средах микрослужб облачные коннекторы интегрируются с платформами оркестрации контейнеров, обеспечивая такие возможности, как обнаружение служб, служебный прокси-сервер, микросегментация и сопоставление служб во всей среде. Облачные коннекторы также позволяют использовать ресурсы балансировки нагрузки для принятия упреждающих решений по автоматическому масштабированию компонентов приложений. Плоскость управления инициирует соответствующий сценарий или вызов API-интерфейса к платформе оркестрации облака или центра обработки данных. При этом создаются экземпляры дополнительных рабочих нагрузок. Облачные коннекторы помогают разработать стратегию ADC, ориентированную на будущее и адаптированную к непрерывной эволюции архитектуры приложений и вариантов инфраструктуры.



#### API-интерфейсы REST: автоматизация для ИТ-отделов и самообслуживание для разработчиков

Программно-определяемые платформы служб приложений должны обеспечивать доступ ко всем службам и поддерживать весь набор API-интерфейсов REST. Команды DevOps ценят гибкость и программируемость публичных облачных сред, таких как Amazon AWS. API-интерфейсы REST позволяют разработчикам приложений встраивать службы и политики, включая пороговые значения автоматического масштабирования, уже в процессе разработки приложений. Разработчики и тестировщики могут инициализировать балансировщики нагрузки корпоративного класса совместно с мощными средствами анализа, чтобы понять, как приложения будут работать в производственной среде. Благодаря этому повышается гибкость развертывания приложений, а ИТ-специалисты могут автоматизировать повторяющиеся задачи и настраивать политики самообслуживания, чтобы расширить возможности специалистов по приложениям.

Обратите внимание на рис. 4 со сравнением архитектур и объяснением его важности. В случае с монолитными приложениями выбор ограничивается статичными неэффективными решениями — аппаратными ADC, развернутыми в каждом местоположении. Для масштабирования среды с приложениями необходимо дождаться развертывания и ручной настройки каждого устройства. Это невозможно при использовании современных распределенных приложений, для которых создаются экземпляры в различных средах. Службы приложений должны автоматически масштабироваться по требованию и контролироваться с помощью централизованных политик, совместно используемых несколькими арендаторами.

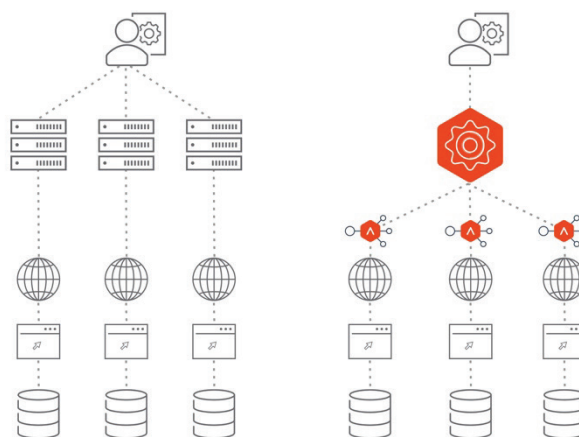


Рис. 4. Сравнение монолитной и современной архитектуры

См. табл. 1 со сравнительным обзором устаревших решений ADC и современных решений для предоставления служб приложений в многооблачных средах. Используйте эту таблицу вместе с подробным контрольным списком.

УСТАРЕВШИЙ БАЛАНСИРОВЩИК НАГРУЗКИ Оборудование или виртуальное устройство	МАТРИЦА БАЛАНСИРОВЩИКОВ НАГРУЗКИ Только программное обеспечение
Специализированное оборудование	Стандартные серверы x86
Большие капитальные расходы	Операционные расходы на подписки
Избыточная инициализация	Упреждающая эластичность
Отсутствие аналитических данных о приложениях	Глубокий анализ приложений
Ручные процессы, неприспособленность к изменениям	Динамическая среда: автоматизация для ИТ-отделов, самообслуживание для разработчиков
Несколько точек управления	Единая точка управления

Таблица 1. Сравнение устаревших и программно-определяемых подходов

Опираясь на представленную концепцию служб приложений, оцените следующие атрибуты. Присвойте каждому атрибуту приоритет по трем критериям (см. табл. 2):

**ВЫСОКИЙ** Насущная потребность, для удовлетворения которой потребуются значительные улучшения центра обработки данных или облачной стратегии

**СРЕДНИЙ** Хороший вариант, но не насущная потребность.

**НИЗКИЙ** Не приоритетно, так как текущее решение удовлетворяет все потребности

Этот контрольный список поможет понять ваши приоритеты и потребности в службах приложений.

АТРИБУТ СЛУЖБ ПРИЛОЖЕНИЙ ДЛЯ МНОГООБЛАЧНЫХ СРЕД	ВЫСОКИЙ	СРЕДНИЙ	НИЗКИЙ
Предоставление согласованной архитектуры служб приложений в облаке или центре обработки данных			
Централизованное управление распределенным пулом балансировщиков нагрузки			
Динамическое масштабирование балансировщиков нагрузки с учетом трафика в режиме реального времени			
Быстрое устранение неполадок в работе приложений без журналов, дампа TCP и т. д.			
Получение аналитической информации о производительности, безопасности и конечных пользователях			
Ускорение развертывания приложений, автоматизация задач балансировки нагрузки и инициализации			
Балансировка нагрузки, WAF и GSLB как услуги			
Запись и воспроизведение транзакций для просмотра времени цикла передачи данных и выявления проблем производительности			
Поддержка многоарендного режима для внутренних групп без приобретения дополнительных устройств			
Возможности самообслуживания на основе политик для бизнес-подразделений, использующих службы приложений			
Повышение уровня безопасности: динамическая защита от DDoS-атак, изоляция приложений, микросегментация			
Возможности самообслуживания для балансировщиков нагрузки в сценариях разработки и тестирования			
Динамическая балансировка нагрузки по требованию для каждого приложения			
Возможности самообслуживания и управления на основе ролей (RBAC) для внутренних пользователей, выполняющих развертывание и мониторинг приложений			
Полная интеграция с VMware, OpenStack, SDN и API-интерфейсами публичного облака			
Интеграция с дистрибутивами Kubernetes, включая PKS, AKS, EKS, GKE, OpenShift, а также другими платформами оркестрации контейнеров			
Обнаружение служб и отображение карт приложений для микрослужб			
Интеграция с инструментами DevOps, включая комплект SDK Python/GO, Ansible, Terraform и SaltStack			
Снижение совокупной стоимости владения решений для балансировки нагрузки			

Таблица 2. Контрольный список для покупателей балансировщиков нагрузки





VMware, Inc. 3401 Hillview Avenue Palo Alto CA 94304 USA Tel (877) 486-9273 Fax (650) 427-5001 [www.vmware.com](http://www.vmware.com)  
125284, Россия, Москва, ул. Беговая, д. 3/1. Тел.: +7 (495) 212-2900 Факс: +7 (495) 212-2901 [www.vmware.com/ru](http://www.vmware.com/ru)

© VMware, Inc., 2019–2020. Все права защищены. Этот продукт защищен законами США и международными законами об авторских правах и интеллектуальной собственности. Продукты VMware защищены одним или несколькими патентами, перечисленными по адресу [vmware.com/go/patents](http://vmware.com/go/patents). VMware является зарегистрированным товарным знаком компании VMware, Inc. и ее дочерних компаний в США и других странах. Все остальные знаки и наименования, упомянутые в этом документе, могут быть товарными знаками соответствующих компаний. Номер по каталогу: vmw-wp-temp-word-104-proof 5/19