

# Кібернетика та системний аналіз

УДК 004.89

С.О. Варганян, В.В. Сокол, Н.С. Лесная

*Харьковский национальный университет радиоэлектроники, Харьков*

## МЕТОДЫ БАЛАНСИРОВКИ НАГРУЗКИ В CLOUD СИСТЕМАХ

*Проведен анализ современного состояния облачных вычислений. Сформулирована задача балансировки нагрузки в Cloud системах. Исследованы основные методы решения поставленной задачи: Least connection, Robin Round, The Least Response Time Method, The Least Bandwidth Method. Для каждого метода произведены оценки его преимуществ и недостатков. Обоснован выбор комбинации методов для балансировки нагрузки Cloud системы. Сформулирован общий подход для решения задачи балансировки нагрузки серверов Cloud систем.*

**Ключевые слова:** Cloud системы, балансировка нагрузки, Citrix Netscaler.

### Введение

Введение и постановка задачи. Облачные вычисления, в информатике – это модель обеспечения повсеместного и удобного сетевого доступа по требованию к общему пулу конфигурируемых вычислительных ресурсов (например, сетям передачи данных, серверам, устройствам хранения данных, приложениям и сервисам – как вместе, так и по отдельности), которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами и/или обращениями к провайдеру.

Потребители облачных вычислений могут значительно уменьшить расходы на инфраструктуру информационных технологий (в краткосрочном и среднесрочном планах) и гибко реагировать на изменения вычислительных потребностей, используя свойства вычислительной эластичности облачных услуг.

Отличительные особенности облачных вычислений (cloud computing) – быстрое предоставление услуг и доступ к ресурсам в любом месте и в любое время. При использовании модели доступа к ИТ-ресурсам Cloud Computing (облачные вычисления), информационные сервисы предоставляются таким образом, что обеспечивающие технологии становятся практически «невидимыми» за пределами пользовательского устройства. А поскольку это позволяет отделить приложения и сервисы от обеспечивающей инфраструктуры и тем самым позволяет бизнесу быстрее адаптироваться к изменениям, облачные вычисления могут являться частью стратегии по повышению динамичности предприятия. Кроме того, облачные вычисления способны поддержать инициативы в таких областях, как внедрение сервис-ориентированных архитектур (SOA), оптимизация центров обработки данных и управление информа-

цией. Облачные вычисления неразрывно связаны с управлением ИТ-услугами, что позволяет снизить совокупную стоимость владения и повысить степень безопасности, качество обслуживания и «эластичность» сервисов до требуемого уровня. Кроме того, по мере добавления всё новых ИТ-ресурсов использование модели Cloud Computing способно свести увеличение сопутствующих затрат на рабочую силу почти к нулю. Облачные вычисления одна из самых быстро развиваемых технологий в ИТ-индустрии.

По данным Gartner 2010 года, показали, что облачные вычисления занимают второе место после виртуализации. Это произошло благодаря тому, что у облачных вычислениях есть потенциал для уменьшения затрат, увеличение отзывчивости системы, улучшения гибкости, так же облачные вычисления полностью переворачивает мир ИТ технологий, позволяя фокусировать внимание на получения результатов, а не заботиться инфраструктурой проекта.

Однако с появлением новой технологии, появляются и новые проблемы, связанные с новой моделью технологии. Одной из проблем является балансировка нагрузки между серверами.

### Основная часть

Однако с появлением облачных вычислений, появляются и новые проблемы, связанные с новой моделью технологии. Все это привело к множественным изменениям скорости, и в какой степени организации решают инициативы облачных вычислений. Одной из главных аспектов в ИТ индустрии, это готовность работы вашего продукта в cloud-системах и поддержка всех возможностей облачных вычислений.

Одним из таких продуктов является Citrix Netscaler, который готов полностью работать в облачных вычислениях, поддерживая все принципы.

Для данной статьи было выбрана продукция Citrix.

Система балансировки нагрузки Web-серверов – это инструментальное средство, предназначенное для переадресации клиентских запросов на наименее загруженный или наиболее подходящий Web-сервер из группы машин, на которых хранятся зеркальные копии информационного ресурса. Клиент не подозревает о том, что обращается к целой группе серверов: все они представляются ему в виде некоего единого виртуального сервера.

Системы балансировки нагрузок делятся на три категории: аппаратные устройства (hardware appliances), сетевые коммутаторы и программные решения. Системы балансировки на базе аппаратного устройства можно рассматривать как "черный ящик"; обычно это функционирующая под управлением UNIX или фирменной ОС машина с процессором Intel, на которой установлена разработанная поставщиком система балансировки нагрузки. Такие системы соответствуют спецификации Plug and Play (PnP), что облегчает работу администраторов узлов Web. Для реализации систем балансировки на базе сетевых коммутаторов используются коммутаторы второго и третьего уровня.

В отличие от аппаратных решений эти системы не предусматривают установки дополнительных устройств, через которые Web-серверы подключаются к коммутатору. Если при организации службы распределения нагрузки в пуле Web-серверов используются программные продукты, то можно обойтись без модификации имеющихся сетевых средств и оборудования. Программные пакеты устанавливаются на существующих Web-серверах или на специальных серверах выравнивания нагрузки.

Вне зависимости от того, к какой категории относится та или иная система балансировки нагрузки, она выполняет следующие задачи: контроль за нагрузкой и состоянием серверов, правильный выбор сервера, который будет отвечать на запрос клиента, и управление трафиком между клиентом и сервером. Сейчас я подробнее расскажу о каждой из этих функций и о том, как их реализуют средства балансировки.

В распределители нагрузки, есть возможность определения виртуального сервера который является фронт-эндом для нескольких серверов. Виртуальный сервер определяет критерий распределения нагрузки, и перенаправляет на один из бек-энд серверов. Когда клиент инициализирует соединение с сервером, виртуальный сервер уничтожает соединение и создаёт своё, что и осуществляет распределение нагрузки.

Распределитель нагрузки использует огромное количество алгоритмов для балансировки, по умолчанию в Citrix Netscaler'e стоит Least Connection Method.

## Методы распределение нагрузки

**Least connection.** Когда виртуальный сервер настроен для использования Least Connection Method, тогда сервер выбирает сервис с наименьшим количеством активных соединений, это стандартный метод, т.к. в большинстве случаев предоставляет лучшую производительность.

**Robin Round.** Самый простой и самый распространённый метод для выборки, используется не только в балансировки, а в таких технологиях как DNS-routing. И так при настройке виртуальной машины на использование метода Робин Раунд, сервер выбирает сервисы циклическим образом.

**The Least Response Time Method.** Когда виртуальный сервер настроен для использования The Least Response Time Method, тогда сервер выбирает сервис с наименьшим количеством активных соединений и наименьшим средним временем отклика. Этот метод может быть сконфигурирован только для HTTP и сервисом Secure Sockets Layer (SSL). Время отклика (также называется Time to First Byte, или TTFB) – это интервал времени между отправленным пакетом запросов к серверу и полученным откликом от сервиса.

**The Least Bandwidth Method.** Если вы сконфигурировали виртуальный сервер на использование метода The Least Bandwidth, то система балансировки выбирает сервис, на котором находится наименьшее количество трафика исчисляемого в мегабайтах в секунду (Mbps). Виртуальная машина выбирает сервис по значению пропускной способности(N), которая получается в результате суммы количеством байт отправленных и полученных в предыдущие 14 секунд.

**Выбор метода.** Как было описано выше Citrix Netscaler поддерживает огромное количество различных методов, начиная от Раунд Робина и заканчивая методом выборки по значению URL.

По данным опроса самой фирма Citrix большинство людей пользовавшиеся распределителем нагрузки используют два метода least connection method, round robin.

Метод Раунд Робин, этот метод довольно простой, используется в случаях когда, все бек-энд сервера по характеристикам равны, но, к сожалению даже одинаковые характеристики серверов, могут быть недостаточными для нормальной работы этого метода, т.к. запросы к бек-энд серверам могут отличаться по нагрузке запрашиваемых данных.

Второй же метод Least Connection Method, даёт в большинстве случаев максимальную производительность, но в некоторых случаях может привести к сбоям, один из таких случаев это отключение бек-энд сервера.

К счастью Netscaler имеет возможность использовать два метода одновременно, что и решает про-

блему между выбором метода, использовать связку двух популярных методов. Появляется вопрос, почему же стоит сразу же два метода одновременно. Ответ на этот вопрос довольно простой: мониторинг серверов у Netscaler'a позволяет на лету удалять сервер из списка доступных бэк-энд серверов, когда сервер стал не доступным, или вышел из строя, так и наоборот добавлять сервер, когда он заработал. И так представьте ситуацию в когда из существующих двух серверов, один перестал работать, вся нагрузка переходит на оставшийся, т.к. он у вас один, далее по каким то причинам второй сервер заработал, и при использовании least connection, все последующие запросы будет получать, только что заработавший сервер, и по этому, что бы избежать перегрузки только что заработавшего сервера используется второй дополнительный метод Раунд Робина.

### Выводы

Проблема распределение нагрузки остро стоит в таких технологиях как облачные вычисление, поскольку скорость развития технологии и требования к ее использованию уже сейчас не дают возможности на долгие исследования.

Не так давно поставщики наладили выпуск систем балансировки нагрузки – программных или аппаратных продуктов, которые выравнивают нагрузку, распределяя ее по нескольким серверам. Кроме того, они повышают отказоустойчивость Web-серверов: в случае отказа одной машины направляют пакеты данных на другой сервер или сайт. Таким образом, время ожидания сокращается, а число необработанных запросов сводится к минимуму. Системы балансировки нагрузки можно использовать как при наличии лишь одного Web-сайта,

так и при работе с целым рядом узлов. Получив представление о том, что такое системы балансировки нагрузки и как они работают, можно определить наиболее важные их характеристики, которые следует учитывать при выборе средства выравнивания нагрузки, а именно:

- скорость работы,
- равномерность загрузки серверов,
- отказоустойчивость.

Одним из лучших аппаратных решений является Citrix Netscaler, который полностью удовлетворяет всем спецификациям облачных вычислений, и при этом поддерживает различные методы балансировки нагрузки, что позволяет использовать его в широком спектре задач.

### Список литературы

1. Chang, J. Dean. *A distributed storage system for structured data*. – Seattle, USA, 2006. – 205-218 с.
2. Dean, J. and Ghemawat, S. *MapReduce: Simplified data processing on large clusters*. – San Francisco, USA, 2004. – Pp. 137-150.
3. Barroso, L. A., Dean, J., and Urs Hölzle, U. *Web search for a planet: The Google cluster architecture*. – Seattle, USA, 2003. – Pp. 22-28.
4. Hadoop Map/Reduce Tutorial [Электронный ресурс]. – Электрон. дан. – Режим доступа: [http://hadoop.apache.org/common/docs/current/mapred\\_tutorial.html](http://hadoop.apache.org/common/docs/current/mapred_tutorial.html), свободный.
5. CloudEra [Электронный ресурс]. – Электрон. дан. – Режим доступа: <http://www.cloudera.com/>, свободный.

Поступила в редколлегию 21.03.2012

Рецензент: д-р. техн. наук, проф. Е.П. Пуятин, Харьковский национальный университет радиоэлектроники, Харьков.

### МЕТОДИ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ В CLOUD СИСТЕМАХ

С.О. Варталян, В.В. Сокол, Н.С. Лесна

*Проведено аналіз сучасного стану хмарних обчислень. Сформульовано задачу балансування навантаження в Cloud системах. Досліджено основні методи вирішення поставленої задачі: Least connection, Robin Round, The Least Response Time Method, The Least Bandwidth Method. Для кожного методу зроблено оцінки його переваг і недоліків. Обґрунтовано вибір комбінації методів для балансування навантаження Cloud системи. Сформульовано загальний похід для вирішення завдання балансування навантаження серверів Cloud систем.*

**Ключові слова:** Cloud системи, балансування навантаження, Citrix Netscaler.

### METHODS OF LOADING BALANCING IN CLOUD SYSTEMS

S.O. Vartanyan, V.V. Sokol, N.S. Lesna

*The analysis of the current state of cloud computing was given. The problem of load balancing in the Cloud systems was described. The basic methods of solving this problem were compared. For each method its advantages and disadvantages were described. The choice of a combination of methods for load balancing in Cloud system was explained. A general approach to solution of the problem of load balancing servers, Cloud systems.*

**Keywords:** Cloud systems, load balancing, Citrix Netscaler.