

УДК 004.896(06)

А.А. Мельник, Е.С. Сакало, Т.С. Ткачева

Харьковский национальный университет радиоэлектроники, Харьков

МОДЕЛЬ ОНТОЛОГИИ ДЛЯ ПОИСКОВЫХ СИСТЕМ В CLOUD-СРЕДЕ

В статье анализируется и рассматривается модель онтологии для поисковых систем в облаке с помощью распределённых вычислений, используемая для параллельных вычислений больших наборов данных в поисковых системах.

Ключевые слова: *AWS Management Console, Amazon CloudSearch, CloudComputing.*

Введение и постановка задачи

CloudComputing (облачные вычисления) – это гибкая, экономически эффективная и проверенная технология для доставки ИТ-сервисов организациям или индивидуальным потребителям с использованием сети Интернет. Ресурсы Cloud-среды быстро развертываются и легко масштабируются, при этом инициализация всех процессов, приложений и сервисов осуществляется по требованию, вне зависимости от местоположения пользователя или используемого компьютерного устройства.

Облачные вычисления представляют собой динамически масштабируемый способ доступа к внешним вычислительным ресурсам в виде сервиса, предоставляемого посредством Интернета, при этом пользователю не требуется никаких особых знаний об инфраструктуре "облака" или навыков управления этой "облачной" технологией.

Под "облаком", разумеется, подразумевается не сам Интернет, а весь тот набор аппаратного и программного обеспечения, который обеспечивает обработку и исполнение клиентских заявок "с той стороны". Если не мудрить с определениями и вглядеться в суть, даже самое простое действие посредством Интернета – запрос страницы сайта, в какой-то степени представляет собой пример облачного вычисления. Интернет, по сути, и возник именно для того, чтобы предоставлять удалённый доступ к различным вычислительным ресурсам.

Таким образом, распространившееся в последнее время мнение об облачных вычислениях как о чём-то специфическом, имеющем отношение к научным исследованиям или мощным корпоративным системам как минимум необъективно, а с точки зрения клиента – конечного пользователя, и вовсе не соответствует действительности.

Как результат, концепция CloudComputing – облачных вычислений – позволяет организациям повысить эффективность предоставления услуг, упростить управление и улучшить согласование ИТ-сервисов с меняющимися потребностями бизнеса.

Использование преимуществ технологии CloudComputing в средах разработки и тестирования прило-

жений может помочь наполовину снизить расходы на рабочую силу, повысить качество приложений и резко сократить сроки вывода программных продуктов на рынок. Применяя самые разнообразные методики, платформа CloudComputing обеспечивает надёжную поддержку базовым функциям бизнеса и одновременно с этим помогает создавать инновационные сервисы. В настоящее время применяются как публичные (public), так и частные (private) модели Cloud-среды.

Публичными моделями Cloud-решений может пользоваться любой человек с доступом в Интернет. Аспект аутсорсинга затрудняет поддержание целостности и конфиденциальности данных, обеспечение доступности данных и готовности сервиса, а также демонстрацию соответствия нормативным требованиям.

Фактически концепция CloudComputing передаёт значительную часть контроля над данными и операциями от клиентской организации к её поставщику Cloud-сервисов подобно тому, как организации доверяет часть своих ИТ-операций аутсорсинговой компании. Даже такие базовые задачи, как установка пакетов с исправлениями и конфигурирование межсетевых экранов, могут стать обязанностью поставщика Cloud-сервисов, а не конечного пользователя.

При использовании модели доступа к ИТ-ресурсам CloudComputing (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 позволить бизнесу быстрее адаптироваться к изменениям, облачные вычисления могут являться частью стратегии по повышению динамичности предприятия. При разработке поисковой системы в Cloud-среде за основу была взята поисковая система Amazon CloudSearch. Amazon CloudSearch представляет собой полностью управляемый поисковый сервис в облаке, что позволяет заказчикам легко интегрировать быстро и масштабируемые функции поиска в своих приложениях. Несколькими щелчками мыши в консоли управления AWS, разработчики просто задают области поиска, загрузки данных, которые они хотят найти.

Amazon CloudSearch легко масштабируется, когда объем поиска данных увеличивается, а разработчики могут изменить параметры поиска и применить новые настройки в любое время без необходимости загружать данные еще раз. Эта технология направлена на упрощение процесса для веб-разработчиков, желающих интегрировать поиск в своих размещенных сайтах. Поисковая система позволяет клиентам разгрузить административную нагрузку на эксплуатацию и масштабирование поисковой платформы. Клиентам не придется беспокоиться об аппаратном обеспечении предоставления данных разделов или исправлений программного обеспечения.

Основная часть

Далее будет рассмотрена реализация поисковой системы в Cloud-среде. Концептуально технологию облачных вычислений делят на предоставление инфраструктуры в качестве сервиса – IaaS (Infrastructure as a service), платформы в качестве сервиса – PaaS (Platform as a service), или программного обеспечения в виде сервиса – SaaS (Software as a service), а также многие придуманные и ещё не придуманные интернет-технологии для удалённых пользовательских вычислений.

При поиске домена для каждого набора данных, которые необходимо сделать доступными, поиск домена имеет одну или несколько поисковых случаев, каждый из которых имеет конечный объем оперативной памяти и ресурсов процессора для индексирования данных и обработки запросов. Количество поисковых случаев домена зависит от количества документов, а также объема и сложности поисковых запросов. В управляемых услугах поисковой системы в облачной среде система определяет размер и количество поисковых случаев, необходимых для достижения низкой задержки, высокой производительности поиска пропускной способности. Когда пользователь загружает данные и настройки индексов, система поиска строит индекс и выбирает соответствующий начальный экземпляр типа поиска, чтобы индекс клиента мог храниться в оперативной памяти.

Поисковая система будет масштабироваться до поиска области (или раздела индекса по несколько экземпляров, если клиент уже на самом крупном). И наоборот, когда объем данных сокращается, система масштабируется до поиска меньших случаев или меньший тип поиска, например, если индекс помещается на одном разделе. Базовая архитектурная модель CloudComputing состоит из нескольких уровней сервисов. Слой физической системы описывает обычные требования к центру обработки данных, обуславливающие необходимость в средствах управления доступом, мониторинга ресурсов и т.д. Слой системных ресурсов описывает управление ИТ-инфраструктурой – сетью, серверами и хранилищами. Слой виртуальных ресурсов представляет строгую изоляцию как базовое свойство безопасной виртуализации, включая

изоляция процессов.

Следующие слои – сервисы поддержки бизнеса (BSS) и сервисы поддержки операционной деятельности (OSS) – описывают платформу управления Cloud-средой. Выше находятся различные сервисы, предоставляемые Cloud-средой.

На каждом слое этой архитектуры существуют определенные требования по безопасности, при этом необходимо поддерживать согласованность между этими слоями. Например, если политика безопасности на самом высоком уровне описанного стека определяет, что клиентская информация не должна покидать пределов государства, то на более низком уровне физических ресурсов должно быть обеспечено выделение дискового пространства внутри страны, на котором и будут храниться соответствующие данные.

Архитектура системы онтологии основана на доминировании веб-приложений над обычными функциями ОС. Ключевая роль при этом отводится браузеру. Стратегия создания нового продукта подразумевает архитектуру, нетребовательную к аппаратным ресурсам персонального компьютера, используемого для выхода в сеть Интернет. Эта тенденция переноса центра тяжести с ПК пользователя на Интернет-ресурсы и соответствует идеологии облачных вычислений. Основными особенностями системы будут: скорость загрузки, выхода в Интернет, получения электронной почты, тесная интеграция с интернет-сервисами, надёжность работы, обеспечение безопасности в автоматическом режиме, автоматическое обновление, простота. Как и объем данных, система автоматически масштабирует поиск домена в соответствии с требованиями. Например, если коллекция настолько велика, что она требует трех разделов, поиск домена будет иметь три поиска экземпляров (по одному для каждого раздела). Трафик увеличивается при каждом поиске, например, раздел реплицируется предоставить дополнительные мощности. После этого у пользователя есть в общей сложности шесть поисковых случаев, поддержка трех разделов на домене.

При управлении облачными сервисами, прежде всего, необходимо четко понимать, из каких компонентов, в каких инфраструктурах состоит тот или иной сервис, на какие элементы частного облака опирается его реализация, какие внешние сервисы он привлекает, если речь идет о гибридной инфраструктуре. Поэтому ключевым элементом платформы являются средства автоматического обнаружения, способные выделить компоненты сервиса и их изменения, включая виртуальные и физические серверы, сетевые соединения и приложения.

Для реализации динамических сервисов, помимо технологий виртуализации и автоматизации необходимы процессы и инструменты управления производительностью (performance management) и управления мощностями (capacity management). Параметры производительности сервиса устанавливаются в соглашении об уровне обслуживания, а программные средства

должны дать возможность отслеживать их выполнение в реальном времени, обеспечивая контроль производительности сервиса с точки зрения его потребителя. Управление мощностями позволяет оптимизировать выделение и распределение облачных ресурсов, исходя из данных об исходных потребностях пользователей и их текущих изменениях. Управление мощностями позволяет также планировать необходимые ресурсы под создаваемые облачные сервисы.

Важная характеристика облака – возможность самообслуживания, и оптимальным инструментом для этого оказывается такой известный элемент методологии, как каталог сервисов, на котором базируется процесс управления запросами (request management) для облачной среды. Каталог сервисов предоставляет меню облачных сервисов и должен сопровождаться реализацией специального пользовательского портала самообслуживания.

Рассматривая систему Amazon CloudSearch всегда можно увидеть ресурсы через страницу аккаунта, активность на сайте AWS, консоль AWS управления CloudSearch инструменты командной строки, или CloudSearch API. Каждое из полей в образце документа должно быть настроено с несколькими вариантами индексации. Каждый из этих вариантов непосредственно влияет на возможности поиска, например, в терминах количества документов.

Конфигурация службы позволяет создать и настроить поиск доменов. Каждый домен инкапсулирует набор данных, которые необходимо найти. Для создания нового домена поиска необходимо ввести имя для обозначения поиска домена. Поиск доменов может быть настроен с указанием индексирования.

В случае облака каталог сервисов является триггером для автоматического выделения ресурсов облака при заказе сервиса и потому должен быть интегрирован с программными средствами реализации процесса управления изменениями. Также необходима интеграция с механизмом управления жизненным циклом сервиса для своевременного удаления сервиса из каталога, если он выводится из употребления, и соответствующего высвобождения облачных ресурсов.

Процесс управления изменениями в случае работы с облаком должен быть по возможности сведен к стандартному набору запросов на изменения, чтобы автоматически получать заранее запрограммированное подтверждение таких изменений и безотлагательно

инициировать операции выделения, изменения или удаления ресурсов. Обращение пользователя к каталогу сервисов может трансформироваться в один или несколько более детальных запросов, например, на создание новой виртуальной машины или установку приложения. Далее технологии автоматизации должны обеспечить изменение облачной инфраструктуры «на лету» для выполнения запроса пользователя.

Индексацию параметров определяют поля, которые вы нужно включить в индекс. Параметры текста позволяют определить предметно-ориентированные словари, игнорировать определенные слова в процессе индексации, определить общие синонимы терминов и карты вариаций слов в общий ствол с тем, чтобы соответствовать по всем вариантам. Ранг выражения математических функций, которые можно использовать, чтобы изменить результаты поиска ранжируются. В системе поиска можно использовать звание выражений на другие факторы в рейтинге.

Выводы

В результате построения и исследования онтологической модели для поисковых систем в Cloud-среде был проведен детальный анализ работы нескольких поисковых сред, разработанных на основе облачных вычислений. Был проведен анализ и сравнение поисковой системы AmazonCloudSearch и на основе этой системе была построена модель онтологии, основанная на архитектуре поисковой системы Amazon с использованием облачных вычислений

Список литературы

1. Chang. *A distributed storage system for structured data* / Chang, J. Dean. – Seattle, USA, 2006. – P. 205-218.
2. Dean J. *MapReduce: Simplified data processing on large clusters* / J. Dean, S. Ghemawat. – San Francisco, USA, 2004. – P. 137-150.
3. Barroso L.A. *Websearchfor a planet: The Google cluster architecture* / L.A. Barroso, J. Dean, U. UrsHölzle. – Seattle, USA, 2003. – P. 22-28.
4. *HadoopMap/ReduceTutorial [Электронный ресурс]* – Режим доступа к ресурсу: http://hadoop.apache.org/ common/docs/current/mapred_tutorial.html, свободный.

Поступила в редколлегию 8.05.2012

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

МОДЕЛЬ ОНТОЛОГІЇ ДЛЯ ПОШУКОВИХ СИСТЕМ В CLOUD-СЕРЕДОВИЩІ

А.А. Мельник, Е.С. Сакало, Т.С. Ткачова

У статті аналізується і розглядається модель онтології для пошукових систем в хмарі за допомогою розподілені обчислень, використовується для паралельних обчислень великих наборів даних у пошукових системах.

Ключові слова: AWS Management Console, Amazon CloudSearch, CloudComputing.

A MODEL FOR THE ONTOLOGY SEARCHENGINES IN CLOUD-MEDIA

A.A. Melnik, E.S. Sakalo, T.S. Tkachova

The given paper analyzes and discusses a model of the ontology for these archeng inesinthe cloud using distributed computing, isused for parallel computation of large datasets in search engines.

Keywords: AWS Management Console, Amazon CloudSearch, CloudComputing.