

Лекция. Концепция «облака»

1. Сущность и задачи облачных технологий
2. История развития облачных технологий

1. Сущность и задачи облачных технологий

Определение *облачных вычислений* (от англ. – *cloud computing*, также используется термин «облачная обработка данных») окончательно не сформировалось. Довольно часто наряду с этим понятием употребляют также *облачные технологии*. При этом, независимо от того или иного варианта, обычно подразумевают технологии обработки данных, в которых компьютерные ресурсы предоставляются *Internet*-пользователю как *on-line*-сервис¹.

Для того чтобы лучше представить, что такое облачные технологии (вычисления), можно привести простой пример. Если раньше пользователь для доступа в электронную почту прибегал к программному обеспечению, установленному на его компьютере, то теперь же он просто заходит на сайт выбранной им организации непосредственно через браузер, минуя посредников.

Облачные технологии позволяют постоянно хранить пользовательскую информацию на *Internet*-серверах и лишь временно кэшировать ее на стороне пользователя. В качестве рабочего инструмента могут использоваться не только стационарные компьютеры, но и ноутбуки, планшеты, смартфоны и т.д.

Облачные технологии как концепция включают в себя много разных понятий, обеспечивающих услуги: программное обеспечение, инфраструктура, платформа, данные, рабочее место и т.п. Самой главной функцией облачных технологий является удовлетворение потребностей пользователей, нуждающихся в удаленной обработке данных.

Таким образом, под облачными технологиями обычно понимается предоставление пользователю компьютерных ресурсов и мощностей в виде *Internet*-сервиса. Вычислительные ресурсы предоставляются пользователю в «чистом» виде, и пользователь может не знать, какие компьютеры обрабатывают его запросы, под управлением какой операционной системы это происходит и т.д.

К основным признакам облачных служб относятся следующие:

- сторонний поставщик услуг, расположенный вне территории потребителя;
- доступ через *Internet*;
- возможность использования сотрудниками со слабой подготовкой в области информационных технологий или даже без нее;
- предоставление ресурсов и услуг по запросам самообслуживания;
- развертывание практически в реальном времени;
- динамическое или детализированное масштабирование;
- детализированная ценовая модель на основе использования ресурсов;
- пользовательский интерфейс – браузер и заменяющие его программы;

¹ Слово «облако» здесь присутствует как метафора, олицетворяющая сложную инфраструктуру, скрывающую за собой все технические детали.

- совместно используемые ресурсы/типовые версии и др.

Использование облачных технологий весьма привлекательно. Во-первых, пользователю не нужно заботиться о производительности своего компьютера, о свободном месте на дисковом пространстве. Во-вторых, пользователь может не тратиться полностью на весь нужный ему продукт, имея в виду, что в дальнейшем большая часть продукта может не понадобиться. Он платит только за услугу, предоставленные возможности и только за конкретные функции. В-третьих, облачные технологии полезны организациям, строящим свою работу периодически, поскольку они сокращают нести затраты на обслуживание огромного функционала, который нужен только в пиковый период.

Основным недостатком облачных технологий является полная зависимость от поставщика услуг. Фактически предприятие (пользователь) оказывается зависимым провайдера сервисов и провайдера доступа в *Internet*. Хотя уровень предоставления услуг поставщиков облачных технологий возрастает, для обеспечения надежности и безопасности данных необходимо приложить немало усилий, например, иметь дублирующие каналы связи, дублирующие мощности для возможности переключения и т.д. Кроме этого, облачные технологии не подходят для организаций, имеющих отношение к государственной тайне.

Целесообразно рассмотреть некоторые задачи, возникающие перед разработчиками облачных сервисов и перед использующими их специалистами.

Задачи распределения и использования ресурсов. При организации вычислительных процессов в сетях с облачной инфраструктурой объектами являются виртуальные машины, сервисы, программы, наборы данных, заявки; позициями – вычислительные узлы, устройства памяти, места в очередях на исполнение. При этом рассматривается ряд количественных характеристик: интенсивность поступающих запросов и степень загрузки центральных устройств, интенсивность межмашинного взаимодействия через сетевые адаптеры и др. Могут также учитываться характеристики индикаторного типа, например, наличие необходимого пакета на данном вычислительном узле. Далее возникают оптимизационные задачи на графах, связанные с распределением и использованием ресурсов и составлением расписаний. Например, необходимо решить заданную прикладную задачу и при этом минимизировать сумму трафика между всеми парами узлов облака. Для решения возникающих задач используют различные, как правило, приближенные, методы: оптимизационные, основанные на теории игр, статистические, машинного обучения.

Модель вычислений MapReduce. Это модель распределенных вычислений, предложенная компанией *Google*, используемая для параллельных вычислений над большими объемами данных. Работа *MapReduce* состоит из двух шагов: *Map* и *Reduce*. На *Map*-шаге происходит предварительная обработка входных данных. Для этого один из компьютеров, называемый главным узлом (*master node*), получает входные данные задачи, разделяет их на части и передает другим компьютерам, называемым рабочими узлами (*worker node*), для предварительной обработки. На *Reduce*-шаге происходит «свёртка» предварительно обработанных данных. Главный узел получает ответы от рабочих узлов и на их основе формирует результат, т.е. решение задачи. На основе этой модели были реализованы алгоритм Дейкстры для

нахождения кратчайшего пути в графе, различные алгоритмы нахождения значимых вершин в графе, метод ближайшего соседа, алгоритм байесовской классификации и др.

Защита облачной инфраструктуры. В облачной информационной среде возникают многочисленные проблемы информационной безопасности: распространение вредоносного программного обеспечения, его обнаружение, выявление программного обеспечения, не являющегося вредоносным, но содержащим в себе ошибки, которые могут привести к возникновению деструктивных процессов. Для решения возникающих задач обычно используются существующие решения: антивирусное программное обеспечение, системы обнаружения и/или предотвращения вторжений.

Однако в силу большого количества вычислительных узлов и больших объемов информации, циркулирующей в среде, а также ввиду неоднородности (например, многоплатформенности) среды все задачи существенно усложняются. Например, сигнатурные методики затруднительно применять для облачных вычислений в средах, использующих различные аппаратно-программные платформы. Даже в случае, когда вредоносные алгоритмы будут одинаковыми, их реализации на различных платформах будут различаться. Таким образом, приходится хранить целый набор сигнатур. Статистические методы требуют слишком больших объемов информации о вычислительных узлах и программных интерфейсах. В частности, для каждой платформы необходимо иметь свой набор статистических данных. Таким образом, возникающие задачи в принципе решаемые, но значительно усложняются в связи с мультиплатформенностью.

Обеспечение надежности работы множества серверов. Очевидно, что в распределенной системе отдельные серверы могут выходить из строя, поэтому необходимо иметь средства, позволяющие восстанавливать потерянную информацию. Простое дублирование (репликация) данных – это далеко не лучший способ, поскольку он связан со значительными потерями времени. Более удобно использовать коды, корректирующие ошибки.

Поиск информации. Еще один пример применения облачных технологий – организация поиска. Например, задача нечеткого поиска по ключевым словам в зашифрованных данных. Основная идея безопасного нечеткого поиска по ключевым словам заключается в использовании двух этапов: построение нечеткого множества ключевых слов и разработка эффективного и безопасного алгоритма поиска.

Появление облачной платформы отражает смену подходов к организации вычислений. Ее начало принято вести от мэйнфреймов – мощных вычислительных комплексов, с которыми группы пользователей работали через терминальные устройства. Затем настало время персональных компьютеров, завоевавших рынок благодаря своим интерактивным возможностям. Однако персональные компьютеры не обладали и не обладают требуемой пользователям производительностью. Поэтому вскоре пришло время сетевых вычислений, где в центре платформы встал сервер – компьютер для решения специализированных задач, подключенный к пользовательским компьютерам локальной сетью. Объединение локальных сетей породило переход на следующий этап – *Internet*. Следом пришло время *grid-*

вычислений – особой формы распределенных вычислений, выполняемых «виртуальным суперкомпьютером», который представляет собой кластер из маломощных, слабо связанных между собой компьютеров, соединенных через сеть и решающих одну общую целевую задачу. Наконец, сегодня фокус развития переходит к облачным вычислениям.

Часто облака сравнивают с мэйнфреймами (*mainframe*), справедливо находя между ними много общего. Тем не менее, существуют и весьма важные различия. Первое принципиальное отличие облака от мэйнфреймов в том, что его вычислительная мощность теоретически не ограничена. Второе принципиальное отличие в том, что терминалы мэйнфреймов служили только для интерактивного взаимодействия пользователя с запущенной на обработку задачей. В облаке же терминал сам является мощным вычислительным устройством, способным не только накапливать промежуточную информацию, но и непосредственно управлять глобальной системой вычислительных ресурсов.

Причины возрастающей популярности облачных технологий понятны: возможности их применения очень разнообразны и позволяют экономить как на обслуживании и персонале, так и на инфраструктуре. Аппаратное обеспечение может быть сильно упрощено при обработке данных и хранении информации в удаленных центрах данных. Все эти проблемы почти полностью переключаются на провайдера услуг. Кроме того, облачные технологии облегчают доступ к данным организации клиентов и собственных сотрудников, находящихся вне ее пределов, но имеющих возможность подключиться через *Internet*.

Современные облачные технологии не только используются в готовом сетевом и серверном оборудовании, но и постепенно проникают на рынок встраиваемых систем и становятся причиной масштабной реструктуризации рынка. Внедрение встраиваемых систем приводит к размещению компьютерных процессоров в таких изделиях, как счетчики учета расхода ресурсов, интеллектуальные датчики, автомобили, бытовая техника и т.д. Это позволяет управлять работой устройств, сбором данных и обеспечением интерактивных возможностей посредством подключения к компьютерной сети.

Технологию подключения различных устройств к глобальной сети называют «Интернет вещей» (*Internet of Things – IoT*). Идея этой технологии была окончательно реализована благодаря «облакам».

Кратко сформулируем достоинства и недостатки облачных технологий.

Достоинства:

- пользователю не требуются собственные большие вычислительные мощности;
- отказоустойчивость;
- высокая скорость обработки данных;
- экономия на программном обеспечении – все необходимые программы есть в сервисе;
- внешняя память компьютера пользователя не переполняется – все данные хранятся в сети;
- доступность – облака доступны всем и из любой точки, где есть *Internet*, с любого компьютера на котором функционирует браузер.

- безопасность – облачные сервисы характеризуются высоким уровнем безопасности при должном её обеспечении, однако при халатном отношении эффект может быть полностью противоположным

- надежность – в основном определяется надежностью специально оборудованных центр обработки данных (ЦОД), которые имеют резервные источники питания, охрану, профессиональных работников, регулярное резервирование данных, высокую пропускную способность *Internet*-каналов.

Недостатки:

- облачная услуга предоставляется всегда какой-либо компанией, соответственно, сохранность пользовательских данных зависит от этой компании;

- появление облачных монополистов;

- необходимость быть всегда в сети для работы;

- высокая стоимость оборудования – для построения собственного «облака» необходимо выделять значительные материальные ресурсы, что может быть выгодно только крупным организациям.

2. История развития облачных технологий

Идея облачных вычислений появилась еще в 1960 году, когда Джон Маккарти высказал предположение, что когда-нибудь компьютерные вычисления будут производиться с помощью «общенародных утилит». Считается, что идеология облачных вычислений получила популярность с 2007 года благодаря быстрому развитию каналов связи и стремительно растущим потребностям пользователей.

Идея того, что сейчас называют облачными вычислениями, впервые была озвучена Джозефом Карлом Робнеттом Ликлайдером (*J.C.R. Licklider*) в 1970 году, когда он был ответственным за разработку ARPANET (*Advanced Research Projects Agency Network*). Идея Ликлайдера заключалась в том, что каждый человек будет подключен к сети, из которой он будет получать не только данные, но и программы. Другой ученый Джон Маккарти (*John McCarthy*) говорил о том, что вычислительные мощности будут предоставляться пользователям как услуга (сервис). На этом развитие облачных технологий было приостановлено до 90-х годов. Ее развитию поспособствовали ряд факторов:

- стремительное развитие сети Интернет, а именно пропускной способности. Хотя в начале 90-х глобальных прорывов в области облачных технологий не произошло, сам факт «ускорения» Интернета дал толчок к скорейшему развитию технологии;

- в 1999 году появилась компания *Salesforce.com*, которая предоставила доступ к своему приложению через сайт. Эта компания стала первой компанией, предоставившей свое программное обеспечение по принципу «программное обеспечение как сервис» (SaaS);

- в 2002 году *Amazon* запустила свой облачный сервис, где пользователи могли хранить информацию и проводить необходимые вычисления.

- в 2006 году *Amazon* запустила сервис *Elastic Compute cloud* (EC2), где пользователи могли запускать свои собственные приложения. Таким образом, сервисы *Amazon EC2* и *Amazon S3* стали первыми сервисами облачных вычислений.

Свой вклад в развитие облачных вычислений внесла компания Google со своей платформой *Google Apps* для *web*-приложений в бизнес секторе.

Развитие аппаратного обеспечения (а именно создание многоядерных процессоров и увеличение емкости накопителей информации) и технологий виртуализации (в частности программного обеспечения для создания виртуальной инфраструктуры, например, *Xen*-виртуализация) способствовало не только развитию, но и большей доступности облачных технологий.

Появление облачной платформы отражает смену парадигмы вычислений. Ее начало принято вести от мэйнфреймов – мощных вычислительных комплексов, с которыми работали через терминальные устройства группы пользователей. Затем настало время персональных компьютеров, победивших рынок благодаря своим отличным интерактивным возможностям. Однако ПК не обладали (и не обладают) требуемой пользователям вычислительной мощностью. Поэтому вскоре пришло время сетевых вычислений, где в центре платформы встал сервер – компьютер для решения специализированных задач, подключенный к пользовательским ПК локальной сетью. Объединение локальных сетей породило переход на следующий этап – Интернет. Следом пришло время *grid*-вычислений – особой формы распределенных вычислений, выполняемых «виртуальным суперкомпьютером», который представляет собой кластер из маломощных, слабо связанных между собой компьютеров, соединенных через сеть и решающих одну общую целевую задачу. Наконец, сегодня фокус развития переходит к новой парадигме – облачным вычислениям.

Часто облака сравнивают с мэйнфреймами (*mainframe*), справедливо находя между ними много общего. Тем не менее, существуют и весьма важные различия. Первое принципиальное отличие облака от мэйнфреймов в том, что его вычислительная мощность теоретически не ограничена. Второе принципиальное отличие в том, что терминалы для мэйнфреймов служили только для интерактивного взаимодействия пользователя с запущенной на обработку задачей. В облаке же терминал сам является мощным вычислительным устройством, способным не только накапливать промежуточную информацию, но и непосредственно управлять глобальной системой вычислительных ресурсов.

Среди ранее возникших (в 1990-х гг.) технологий обработки данных некоторое распространение получили так называемые *grid*-вычисления. Это направление первоначально рассматривалось как возможность использования свободных ресурсов процессоров и развития системы добровольной аренды вычислительных мощностей. Ряд проектов (*GIMPS*, *distributed.net*, *SETI@home*) доказали, что такая модель вычислений достаточно эффективна. Сегодня эта технология применяется для решения научных, математических задач, где требуются значительные вычислительные ресурсы. Известно, что *grid*-вычисления также применяются для коммерческих целей. Например, с их помощью выполняются некоторые трудоемкие задачи, связанные с экономическим прогнозированием, анализом сейсмических данных, разработкой и изучением свойств вакцин и новых лекарств. Действительно, *grid*-вычисления и облака имеют много схожих черт в архитектуре и применяемых принципах. Тем не менее, модель облачных вычислений считается сегодня более перспективной благодаря значительно более гибкой платформе для

работы с удаленными вычислительными ресурсами.

Причины возрастающей популярности облачных технологий понятны: возможности их применения очень разнообразны и позволяют экономить как на обслуживании и персонале, так и на инфраструктуре. Аппаратное обеспечение может быть сильно упрощено при обработке данных и хранении информации в удаленных центрах данных. Все эти проблемы почти полностью перекладываются на провайдера услуг.

К тому же такой подход позволяет стандартизировать программное обеспечение, даже если на компьютерах предприятия установлены разные операционные системы (*Windows, Linux, MacOS* и т.п.). Облачные технологии облегчают обеспечение доступа к данным компании, как для клиентов, так и для собственных сотрудников, находящихся вне офиса, но имеющих возможность подключиться через Интернет.

Современные облачные технологии не только используются в готовом сетевом и серверном оборудовании, но и постепенно проникают на рынок встраиваемых систем (*embedded cloud*) и становятся причиной масштабной реструктуризации рынка. Внедрение встраиваемых систем приводит к размещению компьютерных процессоров в таких изделиях, как счетчики учета расхода ресурсов, интеллектуальные датчики, автомобили, бытовая техника и т.д. Это позволяет управлять работой устройств, сбором данных и обеспечением интерактивных возможностей посредством подключения к компьютерной сети.

Идею подключения всевозможных устройств к глобальной сети называют Интернетом вещей (*Internet of Things – IoT*). Эта идея существует уже много лет, однако для ее реализации не хватало одного звена, чтобы построить такую сеть, – облака.

Так как количество встраиваемых компьютеров увеличивается благодаря снижению цен на процессоры и повсеместному распространению Интернета, растут также и объемы передаваемых данных с последующей их обработкой (часто в режиме реального времени). Поэтому можно предположить, что в ближайшие годы роль Интернета вещей и облачных вычислений будет увеличиваться.