Мега-ЦОДы – пионеры инноваций

Обзор тенденций развития архитектур ЦОД и мега-ЦОД.

Александр Зейников – представитель компании LSI в России.

Введение

"Великий потоп" данных – экспоненциальный, продолжительный рост объемов информации во всем мире – обуславливает масштабную и ускоряющуюся эволюцию ЦОДов. Несмотря на свою недолгую историю, на протяжении последних лет мега-ЦОДы, обеспечивающие веб-хостинг, соцсети и порталы для онлайн-торговли, демонстрировали экспоненциальный рост и теперь составляют около 25% мирового рынка серверов. В процессе своего развития они стали первопроходцами в деле внедрения ИТ-инноваций и продолжают эволюционировать, чтобы соответствовать постоянному росту объемов данных.

"Великий потоп", изменения в технологиях, динамика ведения бизнеса и финансовые потрясения последних лет заставляют корпорации пересматривать, что и как они покупают. Компании ищут ответы на свои вопросы, имея перед глазами пример мега-ЦОДов, а также стараются воспроизвести архитектуры мега-ЦОДов в своих частных облаках, крупных вычислительных кластерах и приложениях для аналитики "больших данных". Мега-ЦОДы стали своеобразными полигонами для испытания приемов повышения эффективности, экономичности, масштабирования и монетизации данных.

Анатомия мега-ЦОДа

Мега-ЦОДы, которые используют такие компании, как Facebook, Amazon, Google, а также китайские гиганты вроде Tencent и Baidu, объединяют в себе несколько различных платформ для выполнения ряда задач, включая хранение данных, управление базами данных, аналитику, анализ поисковых запросов или графиков, а также обеспечение работы веб-серверов. Масштабы таких ЦОДов поражают: мега-ЦОД обычно состоит из порядка 200 000 – 1 000 000 серверов, в которых от 1,5 до 10 млн накопителей.

Самые крупные мега-ЦОДы используют решения LSI для флеш-технологий, адаптеры HBA, инфраструктуру на основе SAS и RAID-решения, чтобы соединить все эти накопители, благодаря чему компания LSI имеет уникальный практический опыт, позволяющий ей оценить, с какими проблемами сталкиваются такие организации, какие архитектурные решения они пробуют внедрить для решения общеизвестных проблем.

Сервера в мега-ЦОДе обычно объединены в кластеры по 20–2000 узлов на кластер. В зависимости от специфики своих задач, сервер может содержать только загрузочные накопители, незащищенные накопители прямого подключения для дублирования данных из различных географических локаций или защищенные RAID-массивы для баз данных и данных транзакций. Так как все эти приложения разнесены по кластерам, сбой в одном узле может спровоцировать сбой целого кластера. Поэтому отключить один сервер и распределить полную нагрузку между 99% оставшихся серверов более эффективно, чем позволить одному проблемному узлу снизить производительность 200 или 2000 других узлов.

Операционные системы инфраструктуры мега-ЦОДов основаны на открытых технологиях – и большая часть улучшений в мега-ЦОДах пошли на пользу открытому сообществу. Приложения можно разрабатывать самим, и многие из них также передаются открытому сообществу. Аппаратная часть также разрабатывается собственными силами или, в крайнем случае, строится согласно самостоятельно определенным спецификациям.

В мега-ЦОДах редко используется виртуализация. В противоположность традиционным ЦОДам, где многие приложения исполняются на одном сервере, приложения мега-ЦОДа исполняются на тысячах и сотнях тысяч серверных узлов. Из-за того, что приложения настолько распределены, время задержки между узлами – важный фактор производительности приложений. Если в мега-ЦОДе используется виртуализация, она основана на открытых стандартах и ее средства используется как некий "контейнер" для упрощения внедрения и дублирования образов. Создание новых образов или обновление приложения ежедневно, еженедельно или ежемесячно – это распространенная практика, ввиду которой управления образами загрузочных дисков крайне трудно.

Мега-ЦОДы широко используют технологию 10GbE, а также инфраструктуру на основе стандарта 40GbE. Так как сети мега-ЦОДов зачастую характеризуются статичными конфигурациями, целью которых является снижение времени задержки при обработке транзакций, специалисты, конфигурирующие сети, часто используют инфраструктуру SND (программно определяемая сеть), чтобы повысить производительность и снизить издержки.

Некоторые люди считают, что мега-ЦОДы – это дешевка, так как нет value-added сервисов, которые имеются в решениях от вендоров. Это не так. Ввиду своего масштаба мега-ЦОДы должны основываться на максимально автоматизированной инфраструктуре, работа которой обеспечивается за счет программных скриптов и требует только минимального базового обслуживания техническим персоналом. Основная цель мега-ЦОДов – снижение стоимости инфраструктуры и использование сэкономленных средств для масштабирования и оптимизации затрат на обслуживание в пересчете на потраченный доллар. Суть в том, что мега-ЦОДы очень стараются упразднить все, что не является критичным для ключевых приложений, даже если это предлагается бесплатно, потому что в итоге это все равно может привести к росту сопутствующих затрат. Микросхемы, коммутаторы, освещение, кнопки, используемые металлические элементы, кабели, винты, заплаты, уровни ПО и средства климатизации, которые не способствуют увеличению производительности, обуславливают рост затрат, потребность в электропитании и препятствуют эффективному обслуживанию. Если добавить ненужную LED-лампочку в каждый из 200 000 серверов, затраты на LED составят 10 000 долларов, а потребность в энергии возрастет на 26 000 Ватт – столько же потребуется для электропитания 26 ручных фенов, работающих в режиме нон-стоп.

Проблемы мега-ЦОДов

Разработчики архитектуры мега-ЦОДов сталкиваются с такими же основными задачами, какие характерны для традиционных ЦОДов: оптимизация инвестиций с учетом растущих объемов данных и обработка более высоких нагрузок при меньших бюджетах. Оба типа ЦОДов вынуждены справляться с растущими объемами данных и должны обеспечивать исполнение сложных приложений в крупных масштабах. Мега-ЦОДы имеют одно существенное отличие: их размер преумножает даже мелкие проблемы или случаи неэффективности. В парадигме мега-ЦОДов весь ЦОД нужно оценивать как пул ресурсов, которые нужно оптимизировать в глобальном масштабе с учетом того, что мега-ЦОДы постоянно работают с целью предоставление бОльшего объема сервисов и поддержания бОльшего количества пользователей на высоком уровне обслуживания.

Простые проблемы при таком масштабе могут стать значительными. Одна из наиболее серьезных проблем в мега-ЦОДах – это массированные отказы жестких дисков, которые провоцируют серьезные сбои в работе кластеров и всего ЦОДа, несмотря на низкую стоимость замены. Архивные хранилища потребляют много

энергии, даже если данные на них используются редко, то есть становятся причиной большого количества проблем в условиях роста объемов информации, которые теперь исчисляются не петабайтами, а экзабайтами. В корпоративных инфраструктурах, где необходимо расширять имеющиеся ресурсы резервного копирования, нужно будет эмулировать архитектурные решения мега-ЦОДов.

Чему можно научиться у сегодняшних мега-ЦОДов

Изменяющаяся динамика бизнеса и непростая финансовая обстановка заставляют традиционные корпорации переосмысливать типы внедряемых ИТ-инфраструктур и программных приложений, а также способы их покупки и внедрения. Из-за низкой стоимости облачных сервисов в мега-ЦОДах финансовые директора корпораций требуют от технических директоров и специалистов в области ИТ-архитектур обеспечения более высокой емкости при меньших затратах. ИТ-отдел, в свою очередь, не имеет другого выбора, кроме воспроизведения архитектуры мега-ЦОДа для выполнения задач, не относящихся к приоритетным, в условиях корпоративных инфраструктур.

Один из уроков мега-ЦОДа, который полезно усвоить, – это использование гомогенной инфраструктуры: задачи поддержки и управления такой инфраструктуры упрощены. Распределение затрат, связанных с инфраструктурой, с целью минимизировать расходы там, где это не критично, и потратить их там, где это необходимо, высвобождает капитальные средства, необходимые для инвестирования в более совершенные архитектурные решения. Инвестиции необходимо концентрировать на оптимизацию и повышение эффективности, чтобы снизить требования к инфраструктуре, связанным с ней процедурам управления, технической поддержке, электропитанию и охлаждению, а также на внедрение техник обслуживания с минимальным вмешательством, чтобы поддерживать увеличение емкости при сокращении необходимых ресурсов.

Второй урок: признать, что попытки поддерживать надежность даже на 5/9 – это дорого и практически невозможно с точки зрения архитектуры в крупных масштабах. Более рациональным решением станет проектирование устойчивого ЦОДа, где подсистемы могут подвергаться сбоям, но вся система продолжит работать даже при таких условиях. Все программные и аппаратные решения уже доступны на рынке, но они не характерны для корпоративных инфраструктур.

Одна из наиболее важных подсистем – это СХД, которые напрямую влияют на производительность приложений и использования серверов. Мега-ЦОДы – лидеры в оптимизации эффективности СХД, так как они управляют огромным объемом данных и беспрецедентным потоком информации, в то же время обладая высокой доступностью и соответствуя юридическим требованиям относительно удержания сохранения целостности данных и обеспечивая безопасность, определяемую законодательством соответствующих стран. Все мега-ЦОДы основаны на СХД прямого подключения (DAS): такие хранилища проще и дешевле купить и поддерживать в дальнейшем, их процессор обеспечивает более низкое время задержки, а также они обеспечивают более высокий уровень производительности, чем SAN- или NAS-хранилища. Несмотря на то, что многие мега-ЦОДы в своих DAS используют обычные потребительские жесткие диски и твердотельные накопители с интерфейсом SATA, они почти всегда также полагаются на архитектуру на основе Serial-Attached SCSI (SAS), которая поддерживает подключаемые SATA-устройства, повышает общую производительность СХД и упрощает процессы управления. Все чаще мега-ЦОДы мигрируют на SAS-диски для обеспечения более высокой надежности и производительности, по мере того, как SAS-накопители мигрируют на интерфейс с пропускной способностью 12 Гбит/с.

Оценивая СХД, корпоративные инфраструктуры довольно долгое время концентрировались на показатель количества операций ввода-вывода в секунду (IOPS) и скорости интерфейса в Мбайт/с. Практика мега-ЦОДов показала, что приложения, которые обрабатывают операции на SSD, довольно быстро достигают максимальных внутренних значений производительности (зачастую до 200 000 IOPS), и скорость передачи данных по интерфейсу оказывает весьма скромное влияние на результаты работы инфраструктуры. Что на самом деле имеет отношение к производительности приложений, эффективности работы инфраструктуры, степени использования сервера – это время задержки. Например, время задержки операций ввода-вывода серьезно влияет на производительность баз данных. Мега-ЦОДы увеличивают показатель совершенных операций на потраченный доллар, внедряя твердотельные накопители, твердотельное кэширование или обе технологии. Значение времени задержки операций чтения-записи в обычном жестком диске составляет 10 миллисекунд. Сравните: время задержки операций чтения в обычном SSD равно 100 микросекундам, а операций записи – около 200 микросекунд. Специализированная интерфейсная карта PCIe® может снизить показатель задержки до десятков микросекунд. SSD могут дополнять или заменять жесткие диски, чтобы повысить производительность приложений, увеличить количество поддерживаемых пользователей, а также увеличить объем проводимых операций на потраченный доллар, благодаря чему серверы и приложения могут выполнять о 4 до 10 раз больший объем работы.

Корпоративные SAN-инфраструктуры могут добиться даже большего прироста производительности – до 30 раз. Как и в случае с DAS, твердотельное кэширование обеспечивает обычно самые малые значения задержки при условии прямого подключения к шине PCIe на сервере. Технология интеллектуального кэширования размещает "горячие" данные (наиболее часто используемые или временно критичные данные) на твердотельные накопители с самым малым временем задержки, где они находятся в легком доступе для приложений. Некоторые карты ускорения кэширования способны поддерживать несколько терабайт СХД на основе твердотельных накопителей и хранят целые базы данных или набор данных для работы приложений в качестве "горячих" данных. Такие данные легкодоступны в условиях любой нагрузки, так как между приложением и данными нет препятствий в виде сетевой инфраструктуры, где могут произойти "заторы" трафика или задержка доставки данных. И даже более того, более малые объемы трафика данных из СХД, достигающие SAN-массива, получают более быстрый отклик. Внедрения "нулевого уровня" СХД на основе твердотельной технологии для некоторых приложений также возможно, и, по крайней мере, один мега-ЦОД может использовать исключительно SSD, совершенно не используя жесткие диски.

В корпоративных инфраструктурах при принятии решения об использовании SSD подразумевают обычно только уровень хранения данных и основываются на факторе стоимости за гигабайт или стоимости за IOPS, противопоставляя жесткие диски твердотельным накопителям в отношении цены. Мега-ЦОДы показали, что, даже используя более дорогие SSD, корпорации все же могут сэкономить на общей стоимости инфраструктуры, сделав ее более эффективной, увеличив производительность и снизив и затраты на техническую поддержку. SSD-накопители также более надежны, менее подвержены сбоям, более просты в обслуживании, более удобны для дублирования и использования в массивах, а также менее прожорливы в отношении электроэнергии, чем жесткие диски – благодаря таким преимуществам SSD-накопителям легче удовлетворить требования SLA. Более высокая производительность SSD позволяет обрабатывать больше операций при меньшем количестве серверов, лицензий ПО и контрактов на сервисное обслуживание, предоставляя возможность снизить общую стоимость владения инфраструктурой.

Представляя ЦОДы будущего

Мега-ЦОДы используют открытые решения в масштабных архитектурах, обеспечивая стабильные показатели производительности, надежности и масштабируемости. В некоторых случаях мега-ЦОДы первыми использовали приложения, способные масштабироваться намного больше, чем любые известные коммерческие продукты. Примерами могут послужить аналитика Hadoop и производные приложения, а также кластерные решения очередности операций и управления базами данных, включая Cassandra™ и Google Dremel. Природа таких решений очень быстро меняется и эволюционирует, в прямом смысле каждый месяц. И это приложения не только внедряются в средах предприятий – они вдохновляют на создание новых коммерческих решений.

Две достаточно молодые инициативы смогут реализовать такие преимущества мега-ЦОДов, как архитектуры, дешевизна и эффективность управления, и на рынке корпоративных решений, как это сделало в свое время ПО Linux®. Open Compute-инициатива, обеспечивающая минималистичную, экономически выгод-

ную и легко масштабируемую аппаратную инфраструктуру для кластерных вычислительных ЦОДов. Аналогичная этой инициатива в области программного обеспечения, OpenStack®, способна обеспечить автоматизированное управления кластерами, как в мега-ЦОДах, в корпоративных ЦОДах за счет создания пула ресурсов обработки и хранения данных и сетевой инфраструктуры, которым можно управлять автоматически – это настоящий Священный Грааль программно-опеределяемого ЦОДа. Инициатива Open Compute может способствовать внедрению еще большего количества инноваций, включая использование бизнес-модели открытого обслуживания аппаратных устройств, аналогичной модели использования открытого ПО. Некоторые специалисты в области архитектуры ЦОДов оценивают потенциал экономии от внедрения этих решений на уровне немыслимых 70%.

Также уже на походе возможность дезагрегации серверов на уровне стойки – то есть отделение процессора от памяти, хранилища, сетевых коммуникаций и источника питания и управление жизненным циклом каждого устройства по отдельности. Этот ход также позволить увеличить объем работы ЦОДа на потраченный доллар.

Заключение

Современные мега-ЦОДы, являясь полигоном для испытания корпоративных ЦОДов будущего, реализуют на практике инновационные идеи и способы повышения эффективности, линейно масштабируясь по требованию по мере увеличения объемов информации. И при разработке архитектуры ЦОД, в ряде случаев, наиболее целесообразной целью может оказаться оптимизация соотношения обьема работы ЦОДа на потраченный доллар на уровне одиночного сервера или всего ЦОДа вместо оптимизации стоимости компонентов (включая управление и техническую поддержку) – традиционном критерии при построении ЦОДа.

Александр Зейников,
компания LSI

Домашнее облако для медиа

С марта 2013 г. компания Seagate стала продавать в России новое поколение сетевых устройств Seagate® Central для совместного хранения данных. Это первые устройства хранения данных с приложением для «умных» ТВ (Smart TV), обеспечивающим доступ к файлам на большом экране. Лауреат CES 2013 Innovations, Seagate Central осуществляет автоматическое резервное копирование данных со всех компьютеров в доме, обеспечивает доступ к цифровым фильмам и музыке через сетевые устройства, а также организует удаленный доступ к домашней сети.

Передача и резервное копирование фотографий и видео с планшетных ПК и смартфонов может осуществляться везде, где есть подключение Wi-Fi или 3G/4G. Автоматическое и непрерывное резервное копирование доступно в смешанной среде платформ Windows и Mac OS X. В дополнение Seagate Central также создает резервные копии фотографий и видео напрямую из Facebook.

Seagate Central также работает в режиме потоковой передачи мультимедийного контента по Wi-Fi. Бесплатные приложения к «умным» телевизорам Samsung Smart TV, проигрывателям Blu-ray, устройствам на ОС Apple® iOS , Android и планшетам Amazon Kindle HD обеспечивают простейший доступ к медиаконтенту и документам. Seagate Central сертифицирован для работы в стандарте DLNA и работает с Apple® AirPlay® – это означает, что практически любое подключенное к сети устройство совместимо с этим накопителем.

После загрузки бесплатного приложения Seagate Media можно просматривать содержимое по типу, размеру и даже совместимости, независимо от того, является ли оно документом PDF, документом Word, музыкальным файлом, фотографией или видеороликом.

За счет совместимости приложения Seagate Media с технологией Apple® AirPlay пользователи ОС iOS получают максимальную гибкость при воспроизведении мультимедийных файлов:

- смотреть воспроизводимый на вашем iPad® фильм на большом экране с помощью Apple TV®;
- слушать музыку с диска Central, используя свой iPhone, на колонках с поддержкой AirPlay.

Seagate Central с поддерживающими приложениями позволяет создать свое собственное личное облако для просмотра и воспроизведения содержимого на подключенном к Интернету мобильном устройстве, как дома, так и в пути, а также предоставлять родным и друзьям безопасный общий доступ к вашим медиафайлам.

Требования к системе:

- маршрутизатор со свободным портом Ethernet (для беспроводного доступа к файлам и их резервного копирования требуется Wi-Fi-маршрутизатор);
- подключение к Интернету для активации и общего доступа к файлам через Интернет;
- браузер Internet Explorer® 7, Firefox® 3.x, Chrome 4.x, Safari® 3 или более новой версии;
- ОС Windows® 7, Windows Vista®, Windows® XP или Mac OS® X 10.4.9 или более новой версии.

Требования к приложениям для удаленного доступа:

- смартфон или планшетный ПК с ОС Android (Android 2.2, Adobe® AIR);
- доступные приложения: Seagate Media, iTunes App Store, Android Market, Google Play, Amazon® Appstore, Samsung Smart Hub (приложение Seagate Media для телевизоров и проигрывателей дисков Blu-ray производства компании Samsung с поддержкой Smart Hub предназначено для моделей выпуска 2012 года или более новых).

Samsung начинает выпуск 3D V-NAND-памяти

Август 2013 г. – Компания Samsung Electronics объявила о начале массового производства первой в индустрии NAND флэш-памяти с трехмерной структурой упаковки чипа и вертикальным расположением ячеек. Новинка позволяет преодолеть существующие ограничения масштабируемости энергонезависимой памяти. 3D V-NAND с выгодным соотношением производительности и размера будет применяться в широком спектре потребительской электроники и корпоративных приложений, в том числе встраиваемой NAND-памяти и твердотельных накопителях (SSD).

Первый чип Samsung V-NAND объемом 128 Гбит использует технологию 3D Charge Trap Flash (CTF), обеспечивающую большую надежность работы и одновременно позволяющую снизить стоимость чипов флэш-памяти по сравнению с классическими ячейками с плавающим затвором, а также технологию многослойной компоновки с вертикальными внутричиповыми соединениями.

"Новая технология 3D V-NAND является результатом многолетних разработок наших специалистов. Целью Samsung было выйти за рамки традиционных способов мышления и преодолеть существующие ограничения в создании упаковки полупроводниковой памяти с помощью инновационных решений, – говорит Чжон-Хек Чой (Jeong-Hyuk Choi), старший вице-президент направления флэш-памяти и технологий компании Samsung Electronics. – Наряду с запуском производства первой в мире 3D V-NAND флэш-памяти мы продолжим представлять новые 3D V-NAND устройства с улучшенной производительностью и более высокой плотностью. Это поспособствует дальнейшему росту мировой индустрии продуктов памяти".

В течение последних 40 лет флэш-память создавалась на основе плоских однослойных структур с ячейками с плавающим затвором. Однако после освоения процесса производства с топологическим размером 10-нм масштаба возникла необходимость в новых технологических