Гипервизоры и СХД

Обзор особенностей интеграции последней версии гипервизора Microsoft Hyper-V с компонентами уровня хранения данных – продолжение темы двух предыдущих номеров SN (№ 1/49, 2012 – Citrix XenServer; № 4/48, 2011 – VMware).

Сергей Платонов – менеджер по продуктам, компания AVRORAID.

Введение

Все взгляды разработчиков компании Microsoft устремлены в облака. Операционную систему Windows Server 8 смело можно называть "облачно-ориентированной".

Традиционно платформа виртуализации считается "сердцем" облачной инфраструктуры. Сегодня бета-версия MS Windows Server 8 (Microsoft уже переименовал свой новый продукт в MS Windows Server 2012) доступна всем, желающим ознакомиться с инновациями софтверного гиганта.

В Windows Server 8 и Hyper-V были радикально переработаны сервисы, отвечающие за хранение данных, что неудивительно – СХД являются одной из основных частей Microsoft Private Cloud Fabric. Разработчики проделали огромную работу для увеличения производительности, отказоустойчивости, масштабируемости, безопасности и управляемости инфраструктуры.

Рассмотрим те изменения, которые появились в сервисах хранения данных новой версии гипервизора Microsoft Hyper-V.

Формат файлов

Все наверняка знакомы с форматом файлов VHD (Virtual Hard Disk – VHD), если работали с продуктами серверной виртуализации от Microsoft или Citrix: это формат файла, у которого полная структура и содержимое аналогичны жесткому диску. Формат VHD был представлен в 2003 г. Изначально формат был создан компанией Connectix и позднее куплен Microsoft вместе с программой виртуализации Connectix Virtual PC. С 2005 г. Microsoft сделала спецификацию формата VHD доступной третьим фирмам в рамках Microsoft Open Specification Promise. В том же году Microsoft обещала, что данный формат будет развиваться и использоваться в следующих продуктах виртуализации компании.

Формат VHD имеет ряд технических ограничений, не позволяющих считать его удовлетворяющим требованиям современного мира.

Новая версия серверной ОС и гипервизора от Microsoft предлагают нам новый формат файлов – VHDX. На рис. 1 представлена архитектура формата VHDX.

Header: заголовок VHDX-диска определяет расположение других структур VHDX диска. На самом деле, в разделе Header хранятся две копии заголовка, одна из которых активна, а вторая – для избыточности.

Intent Log: журнал, в который записываются транзакции обновления метаданных перед их записью непосредственно в таблицу. Сюда записываются лишь метаданные, но не сами данные.

Data region: таблица BAT хранит в себе ссылки на блоки пользовательских данных и секторных битмапов, что является существенным отличием от формата VHD, где секторные битмапы не имели отдельной структуры, но дописывались в конец каждого блока пользовательских данных.

Metadata region: таблица, хранящая файловые метаданные, такие как размер блока, размер физического и логического секторов, а также пользовательские метаданные, о которых мы уже говорили выше.

Максимальный размер блока увеличен до 256 Мбайт. Все внутренние операции выровнены по сектору 4k.

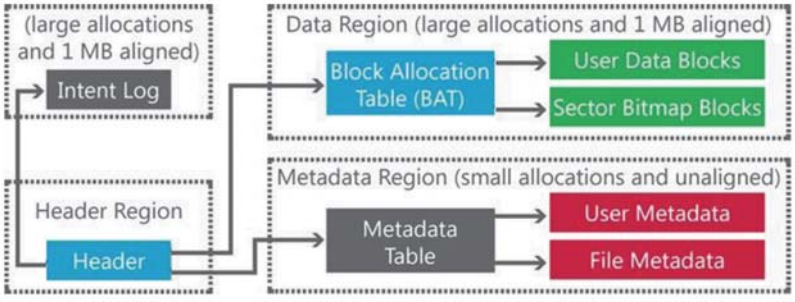


Рис. 1. Структура диска VHDX.

Какие же преимущества принесет новый формат?

Максимальный размер файла был увеличен до 64 Тбайт. Формат VHD не позволял создавать файлы размером более чем 2 Тбайт. Единственной возможностью использования дисков больших размеров в виртуальных машинах было применение Pass-through GPT дисков.

Теперь же Microsoft позволяет использовать образы дисков размером 64ТБ. Причем, переход был выполнен поступательно. В Windows 8 Developer Preview невозможно было создать файл размером более чем 16 Тбайт.

В VHDX появилось журналирование. Любые изменения метаданных сначала записываются в статичную зону журнала и лишь после этого в саму таблицу метаданных. Если во время обновления таблицы метаданных произойдёт сбой записи, то в журнале останутся незавершенные транзакции, которые позволят накатить изменения в таблицу метаданных из журнала и вернуть целостный формат диска.

В Windows Server 2012 появилась также возможность создания виртуальных дисков на устройствах хранения, имеющих размер сектора 4kb и не поддерживающих спецификацию 512e. В том числе, возможно выравнивание по блоку 4k, что избавит от проблем потери производительности на устройствах хранения с большими секторами.

Для сохранения информации об образе диска в формате VHDX используются дополнительные поля: теперь пользователь диска может занести о нем любую информацию, касающуюся его использования. Можно создать до 1024 дополнительных полей размером до 1 Мбайт каждое.

Для возможности возвращения ресурсов в пул хранения при удалении данных в VHDX дисках поддерживается операция unmap (TRIM).

В VHDX поддерживаются операции с метаданным "на лету", что позволит выполнять некоторые операции с хранилищами виртуальных машин и мгновенными снимками виртуальных машин (например, объединение мгновенных снимков) без простоя.

Также, по заверениям разработчиков, в VHDX значительно повышена производительность, что видно на диаграмме, представленной на рис. 2: Microsoft значительно улучшила производительность виртуальных дисков.

VHDX будет использоваться как формат виртуальных дисков по умолчанию в Windows Server 2012 и Hyper-V. При этом будет поддерживаться возможность конвертации в VHD и обратно.

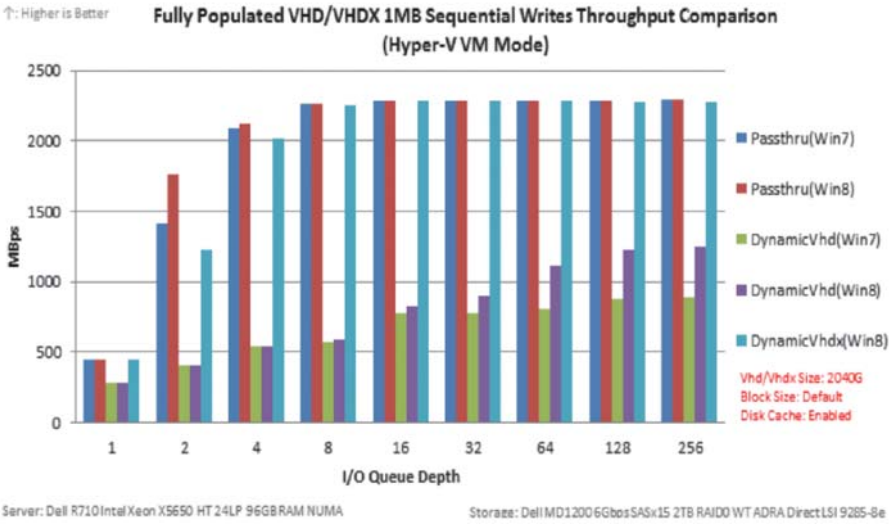


Рис. 2. Сравнение производительности виртуальных дисков. Слайд SNIA Storage Developer Conference 2011.

Для создания образов дисков формата VHDX можно воспользоваться различными способами, среди которых: WMI, Diskpart, Disk2VHD, VHD Tool, Disk Management, Hyper-V Manager, PowerShell 3.0.

VHDX образы монтируются в операционной системе двумя кликами мыши без использования дополнительного ПО.

30 апреля 2012 г. спецификация VHDX стала доступной для всех под лицензией Microsoft Open Source Promise (OSP).

Возможности по размещению образов виртуальных машин

В предыдущих версиях гипервизора было возможно размещать образы виртуальных машин исключительно на внешней блочной системе хранения данных (StandAlone конфигурацию вынесем за рамки текущего разговора). При использовании Windows Server 2012 у пользователей появляются совершенно новые возможности:

- использование файл-сервера, в том числе отказоустойчивого;
- использование Storage Spaces;
- использование PCI-E RAID контроллера в отказоустойчивой конфигурации.

Поддержка файловых хранилищ

Среди всех представителей “большой тройки” серверной виртуализации компания Microsoft была последней, кто предоставил возможность использования файловых систем хранения для организации хранилищ файлов виртуальных машин. Конечно, в отличие от VMware и Citrix, Microsoft будет поддерживать “родной” протокол SMB.

Протокол SMB/CIFS традиционно считался намного менее производительной (но и более функциональной и сложной) альтернативой NFS. В новой версии протокола SMB 2.2 Microsoft представила нововведения, позволяющие решить проблему (во время написания этой статьи появилась информация о переименовании SMB 2.2 в SMB 3.0).

Рассмотрим более подробно изменения в новом протоколе.

При разработке новой версии протокола Microsoft ориентировалась на возможность использования файловых хранилищ для таких приложений, как гипервизоры, Microsoft Exchange Sever, Microsoft SQL Server. Поэтому особое внимание было уделено производительности и отказоустойчивости.

Одним из самых важных нововведений является поддержка RDMA в протоколе SMB 3.0 SMB over RDMA – это новый протокол хранения данных в Windows Server 2012. Он обеспечивает прямую передачу данных между сервером и хранилищем с минимальным использованием ресурсов CPU, используя стандартные RDMA-совместимые сетевые адаптеры. SMB Direct поддерживает все доступные технологии RDMA: iWARP, InfiniBand и RoCE. Минимизация использования ресурсов CPU для операций ввода-вывода означает, что сервер может поддерживать бОльшие рабочие среды (например, Hyper-V может запускать больше виртуальных машин), экономя ресурсы CPU.

На последней конференции SNIA Storage Developer Conference Microsoft показала, что при использовании SMB с поддержкой RDMA достигается производительность 3,2 GB/s и 160 тысяч IOps.

В SMB 3.0 появился механизм мультиканальной передачи, который позволяет задействовать одновременно несколько физических сетевых адаптеров при обмене данными между SMB-клиентом и сервером. Например, в корпорации протестировали СУБД Microsoft SQL Server, передающую данные с использованием SMB 3.0 по четырем каналам 10GbE. При этом была достигнута пропускная способность 6,5 Гбайт/с и 280 тыс. операций записи 8-килобайтных блоков в секунду. Функция SMB 3.0 Multichannel обеспечивает лучшую пропускную способность и несколько резервных путей от сервера (например, Hyper-V или SQL Server) к удаленному хранилищу SMB 3.0 Отказ одного из сетевых путей обрабатывается автоматически и прозрачно, не вызывая прерывания сервиса приложения.

Microsoft позаботилась и над обеспечением отказоустойчивости. Новые сервер и клиент SMB 3.0 взаимодействуют друг с другом для обеспечения прозрачной отработки отказов для переключения на альтернативный кластер для всех операций SMB 3.0 как в случае запланированных операций, так и в случае незапланированных отказов. Одним из новшеств Windows Server 2012 является отказоустойчивый файл-сервер, называемый Scale Out File Server (SOFS) и работающий в режиме active-active.

Горизонтально-масштабируемый файловый сервер позволяет увеличивать производительность путем добавления новых узлов в кластер. Теперь производительность кластера не ограничена производительностью одного узла, как это было в предыдущих версиях.

Значительно упрощено управление кластерным файл-сервером и добавлена возможность прозрачного перенаправления SMB сессий для балансировки нагрузки.

Но, так как SOFS является кластером, нельзя будет обойтись без внешнего блочного СХД, доступного всем узлам.

Microsoft также позаботился и над обеспечением безопасности. SMB 3.0 поддерживает безопасную передачу данных с использованием шифрования.

Storage Spaces

Виртуализация СХД пришла на операционную систему Windows. Наверняка, многие имели дело с менеджерами томов в Unix-подобных операционных системах. Подобные решения позволяют абстрагировать уровень хранения данных от физического размещения, значительно упростить управление и предоставить дополнительные функции.

Итак, каким же путем пошел Microsoft?

Новая функция Storage Spaces позволяет сделать следующее:

- организовать носители информации в пул хранения. В едином пуле хранения возможно использование различных типов носителей одновременно. Пул может быть легко расширен простым добавлением носителей. Одним из возможных применений пула является изоляция нагрузки;
- создавать виртуальные диски, называемые Spaces. Spaces используются ОС точно так же, как обычные жесткие диски, но при этом поддерживают такую функциональность, как Thin Provisioning и устойчивы к сбоям физических носителей.

Для организации отказоустойчивости spaces используют зеркалирование и контроль четности. Если вы создаете space с зеркалированием, то 2 или 3 копии данных хранятся на реальных дисках. При использовании четности – только одна контрольная.

Storage Spaces могут использовать обычные JBOD в качестве физических носите-

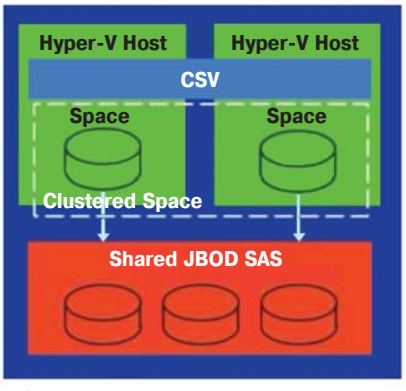


Рис. 3. Организация отказоустойчивого хранилища с использованием Storage Spaces и JBOD.

лей информации при этом создавая отказоустойчивое хранилище (рис. 3).

Средствами операционной системы возможно разграничение прав доступа к определенным пулам.

PCI-E RAID контроллер

Совместно с партнерами Microsoft разрабатывает архитектуру Cluster-in-a-box. Данная архитектура позволяет строить отказоустойчивое решение без использования внешней СХД и размещаемое в едином стандартном шасси. Каждый такой блок состоит из двух серверов, размещенных в едином шасси; PCI-E контроллеров, работающих в режиме высокой доступности и общего для обоих узлов набора жестких дисков (рис. 4).

Offloaded Data Tansfer (ODX)

Облачная инфраструктура повышает требования к производительности передачи данных. Компанией Microsoft был разработан механизм ускоренной передачи данных – Offloaded Data Transfer (ODX) – также известный как SCSI PROXY READ and WRITE. ODX представляет собой реализацию SBC-3 / SPC-4 offloaded copy.

Данная функция позволяет копировать или перемещать данные между ресурсами хранения без необходимости передачи их на хост, что значительно сокращает время передачи, а также экономит ресурсы хоста, такие как процессорное время и сетевые ресурсы.

Для поддержки данной функции СХД должны реализовывать спецификацию T10 11-059r8 XCOPY. ОС определяет поддерживает ли устройство хранение ODX при подключении СХД к хосту.

При копировании или перемещении большого объема данных с одного устройства хранения на другое хост отправляет offload read операцию к системе хранения данных. Затем отправляется команда receive offload read result. В ответ на эту команду СХД предает хосту специальный токен – некое представление данных, которые должны быть скопированы на момент времени. Этот токен передается менеджеру копирования конечного устройства хранения с использованием операции offload write. После этого инициализируется процесс передачи данных напрямую между устройствами хранения. Завершается процесс выполнением команды receive offload write, в ответ на которую приходит статус выполнения передачи данных (рис. 5).

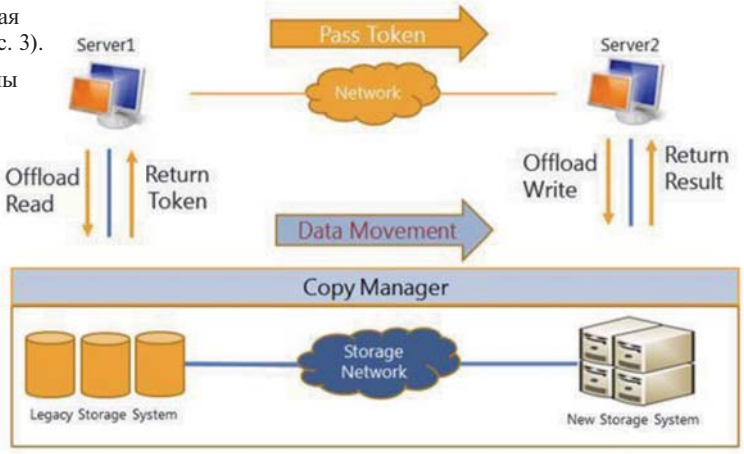


Рис. 5. Копирование большого объема данных с применением ODX.

Всего применяются 4 команды для передачи данных: 1) Offload read, 2) Receive offload read result, 3) Offload write with the token, 4) Receive offload write result.

Offloaded Data Transfer может использоваться при копировании данных между различными системами хранения, при копировании данных внутри одной СХД и в конфигурациях с одним и несколькими хостами. Также одним из возможных сценариев является миграция данных между уровнями хранения, управляемая со стороны хоста.

Для того чтобы избежать проблем с таймаутами SCSI-команд и обеспечить поддержку кластеров и сценариев, в которых используется MPIO, Microsoft выполнил некоторые оптимизации.

Большие запросы на запись разбиваются на меньшие. Обычно оптимальный размер передачи определяется целевым устройством хранения. Если целевое устройство хранения не предоставляет данного параметра, то по умолчанию оптимальный размер передачи устанавливается в значение 64 Мбайт. Если на конечном устройстве оптимальный размер передачи превышает 256 Мбайт, то хост принимает его равным 256 Мбайт. Ожидается, что время выполнения операций чтения/записи в данном случае не будет превышать 4 секунд.

ODX работает на любых томах, отформатированных в NTFS, но не поддерживает зашифрованные и сжатые файлы.

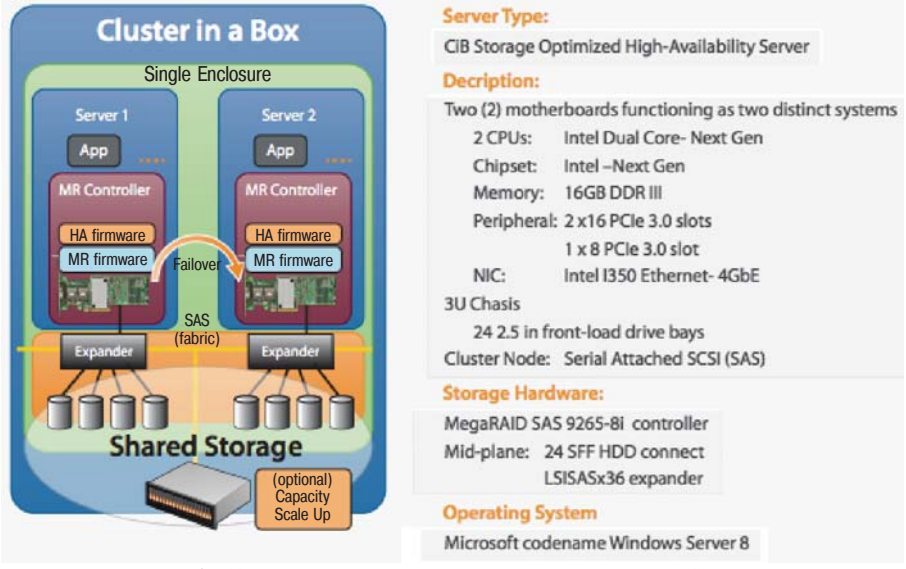


Рис. 4. Cluster in a Box с использованием RAID-контроллеров LSI.

Виртуализация оборудования

Виртуальные адаптеры Fibre Channel

В новой версии гипервизора появилась возможность виртуализации WWN-адаптеров FC для виртуальных машин, подключенных непосредственно к SAN, с использованием стандарта N_port ID virtualization (NPIV). Для каждой машины можно виртуализовать до 4-х портов.

Поддержка виртуальных FC адаптеров значительно расширит возможности для построения HA-кластеров, для которых требуется разделяемое хранилище. Данный механизм позволит эффективно применять зонирование и презентацию LUN на системах хранения данных, а также Virtual SAN. Разработчики Microsoft продумали механизм переключения между WWN с целью предотвращения потери доступа к LUN во время Live Migration. В Microsoft позаботились о том, чтобы ни одна новая функция не препятствовала работе Live Migration

Для обеспечения отказоустойчивости SAN возможно использование на гостевых машинах MPIO.

Также доступна функция загрузки виртуальных машин из сети хранения данных.

Поддержка SR-IOV

Single-Root I/O Virtualization (SR-IOV) – эта технология виртуализации (созданная группой PCI-Special Interest Group, или PCI-SIG), которая обеспечивает виртуализацию устройств в сложных системах с одной корневой системой (случай, когда устройство используется на одном сервере с несколькими виртуальными машинами).

В рамках концепции SR-IOV устройство PCIe может экспортировать не только ряд физических функций шины PCI, но также ряд виртуальных функций, которые совместно используют ресурсы на устройстве ввода/вывода. Это позволяет не-

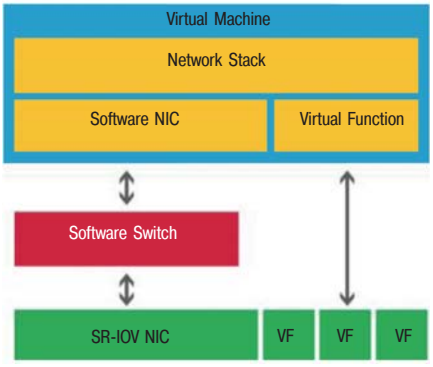


Рис. 6. Принцип работы SR-IOV на примере сетевого адаптера.

скольким драйверам независимо и прозрачно друг для друга подключаться к одному PCIe-адаптеру. Технология SR-IOV достигает такого эффекта, открывая пользователям виртуальные функции, которые отображаются как физические функции PCIe-адаптера, но на самом деле реализованы в адаптере как функции для совместного использования. Так же, как и остальные функции, SR-IOV не мешает миграции виртуальных машин (рис. 6).

Дедупликация данных

В новой версии операционной системы от Microsoft была анонсирована дедупликация данных на NTFS. На первый взгляд, функция выглядит очень полезной, ведь далеко не на всех системах хранения данных реализована Primary Deduplication.

Дедупликация появилась в Microsoft Windows не впервые. Но на этот раз Microsoft перенес функцию дедупликации во все издания серверной ОС, а не только в Storage Server. И, что наиболее важно, дедупликация теперь выполняется на блочном уровне, а не на файловом.

Microsoft реализовала post-process дедупликацию, запускаемую с использованием PowerShell или из планировщика. Такое решение позволяет не оказывать влияния на производительность записи. Производительность чтения для незакэшированных данных ухудшается примерно на 3%. Но при этом сервис дедупликации увеличивает эффективность кэширования (рис. 7).

Microsoft использует “чанки” – блоки, на которые разбивается вся информация, переменного размера, колеблющиеся между 32 и 128 KB. На рисунке блоки A, B и C являются кандидатами на дедупликацию. После завершения процесса дедупликации данных блоки, “попавшие” под дедупликацию, теперь хранятся на диске только как одна копия и только в одном месте. Chunk store, хранящий чанки, расположен в папке System Volume Information в сжатом виде.

Процесс дедупликации данных фактически представляет собой запланированную задачу, которую можно найти в репозитории задач Windows, или задачу запускаемую с использованием PowerShell.

При хранении образов виртуальных дисков на томе с включенной дедупликацией Microsoft обещает более 90% выигрыша пространства.

Microsoft сделала операцию дедупликации устойчивой к сбоям. В случае неожиданного прерывания операции мы застрахованы от потери данных.

Рассмотрим ограничения технологии дедупликации:

- отсутствует поддержка на загрузочных и системных томах, возможно включение только дисков с данными;
- отсутствует поддержка зашифрованных и сжатых томов;
- работает только на дисках с файловой системой NTFS, не поддерживает новую файловую систему ReFS и Clustered Shared Volumes;
- не работает с зашифрованными файлами, файлами меньше 64Кб, файлами с расширенными атрибутами.

Как видно, основной недостаток скрывается в 3-м пункте, который означает невозможность применения дедупликации в большом числе инсталляций Hyper-V.

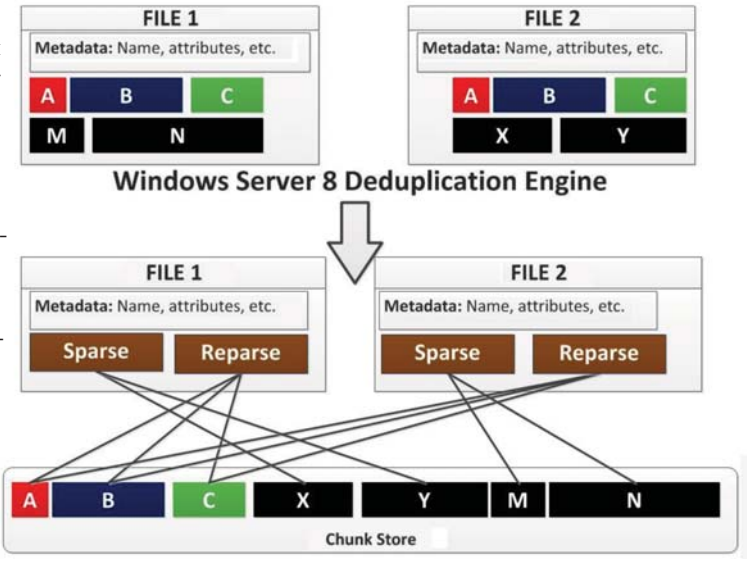


Рис. 7. Принцип работы дедупликации в Windows Server 2012.

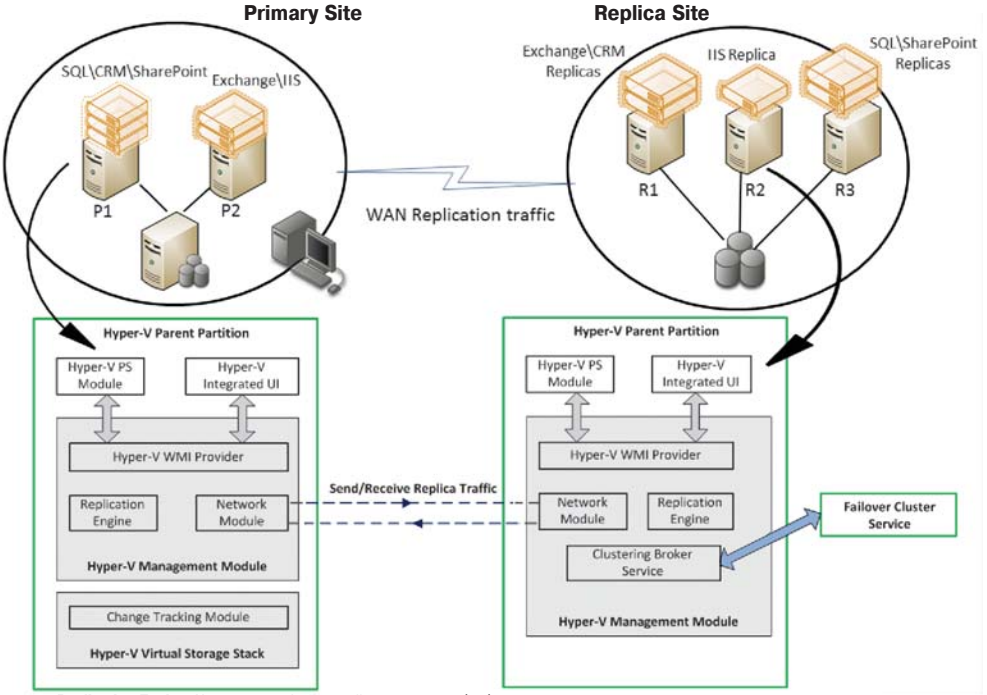
Hyper-V Replica

Hyper-V реплика является одним из новых сервисов, предназначенных для повышения доступности и отказоустойчивости. Hyper-V Replica предназначена для репликации виртуальных машин и конфигурации между узлами и кластерами Hyper-V с целью обеспечения непрерывности бизнеса и восстановления после катастроф.

Репликация возможна в локальной сети в случае отсутствия SAN или в случае георасределенной инфраструктуры между несколькими ЦОД. Репликация виртуальных машин может выполняться независимо от типа используемой СХД. При этом возможно использование совершенно различных СХД на первичном и вторичном сайтах. Доступна только асинхронная репликация. Возможна репликация одной виртуальной машины или нескольких – на выбор, что дает преимущества перед репликацией на уровне СХД, где выполняется репликация всего тома целиком. Поддерживается режим “многие-к-одному”, когда один вторичный сервер (Replica Server) принимает данные от множества первичных. Архитектура Hyper-V replica представлена на рис. 8.

Перед тем, как репликация может быть начата, копия виртуальных дисков должна быть размещена на вторичном сайте. Hyper-V поддерживает 3 возможности размещения начальной копии.

1. Репликация по сети – выполняется передача копии виртуального диска по сети, по которой в дальнейшем будут передаваться реплицируемые данные. При настройке репликации пользователь может назначить копирование по расписанию.



- **Replication Engine**. Управляет конфигурацией репликации, обрабатывает первоначальную и разностную репликации, отслеживает события репликации и при необходимости приостанавливает и возобновляет процесс переноса данных;
- **Change Tracking Module**. Модуль, отслеживающий операции чтения на уровне виртуальной машины на первичном узле или кластере, вне зависимости от типа хранилища ВМ (DAS, SAN LUN, папка SMB на файловом сервере или CSV);
- **Network Module**. Компонент призван обеспечить безопасный канал связи между первичным и принимающим узлами, строящий соединение с использованием HTTP/HTTPS с возможностью использования механизмов шифрования;
- **Hyper-V Replica Broker role**. Роль, обеспечивающая прозрачную миграцию в случае размещения виртуальной машины на кластерных узлах совместно с сетевым модулем и компонентов Failover Clustering;
- **Management Experience**. Включает в себя следующие компоненты для управления процессами репликации:
  - интерфейс Hyper-V Manager;
  - интерфейс Failover Cluster;
  - Scripting – управление функциональностью реплик с помощью PowerShell;
  - Hyper-V Replica APIs – интерфейс может использоваться сторонними управляющими приложениями;
  - Remote Management – включает в себя средства удаленного управления (RSAT).

Рис. 8. Архитектура Hyper-V Replica.

2. Использование резервной копии. Пользователь может выполнить копирование резервной копии виртуального диска и использовать ее.

3. Копирование через внешний источник. Копию образа также можно просто перенести, используя такие средства как USB-HDD.

Как можно видеть, возможностей здесь тоже больше, чем при использовании репликации на стороне СХД.

Все изменения виртуального диска на первичном сайте сохраняются в сжатом файле формата .hrl, который через определенные промежутки времени передается на вторичный сайт.

Вторичный сайт управляет несколькими точками восстановления, которые используются для восстановления виртуальных машин. Каждая такая точка содержит один или несколько мгновенных снимков. Точки восстановления создаются каждый час. Кроме того, существует возможность создания консистентных снэпшотов с использованием сервиса VSS.

Hyper-V реплика поддерживает следующие сценарии развертывания:

– основной офис и филиал;
– ЦОД компании;
– ЦОД провайдера;
– офис клиента и ЦОД провайдера.

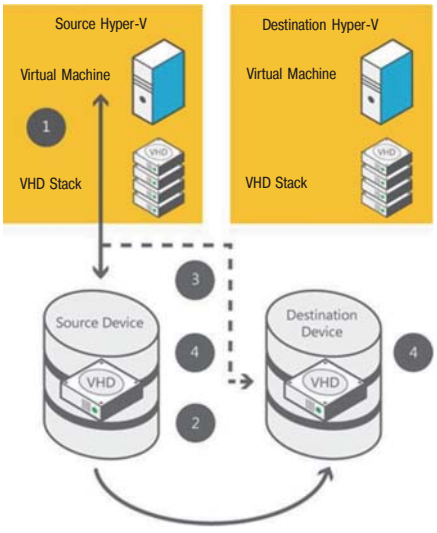
Особое внимание уделено обеспечению безопасности репликации. Безопасность Hyper-V Replica реализована на следующих уровнях:

– используется новая модель авторизации – при установке роли Hyper-V создается локальная группа Hyper-V Administrators, включающая в себя локальных администраторов сервера по умолчанию;
– администраторы могут настроить серверы-реплики на принятие входящих соединений только от определенных серверов;
– администраторы могут настроить на правила брандмауэра серверов-реплик на принятие входящих соединений по настраиваемому порту;
– взаимная аутентификация может осуществляться на основе встроенной проверки подлинности между доверенными доменами. Во внедоменной инфраструктуре могут (и должны) использоваться сертификаты.

Live Migration

Live Migration было значительным обновлением для Windows Server 2008 R2 Hyper-V RTM, и в следующей версии гипервизора технология будет значительно улучшена. Данная технология позволяет выполнять перенос виртуальных машин между различными хостами без приостановки работы. Microsoft убрала ограничение, запрещающее одновременное выполнение нескольких миграций. Количество одновременных миграций является настраиваемым параметром, и администратор сам сможет выбрать его значение, исходя из возможностей своего оборудования.

Microsoft превзошла своих конкурентов и позволяет выполнять Live Migration без кластеров и общих дисков. Теперь Live-Migration могут использовать компании, которые не могут позволить себе внешнее разделяемое СХД, соответствующее размерам их виртуальной инфраструктуры. Кроме того, поддерживается миграция между различными видами СХД (рис. 9).



(1) Вплоть до завершения операции переноса виртуальная машина работает с оригинальным источником виртуальных дисков. В то время как операции чтения/записи происходят с оригинальным диском, он копируется по сети на новый сервер.
(2) По завершении копирования все операции записи дублируются на оригинальное и новое расположение, и в это время реплицируются изменения с оригинального источника, произошедшие за время первоначальной копии диска.
(3) По завершении синхронизации виртуальная машина начинает использовать диск с сетевого SMB места размещения и инициируется Live Migration машины на новый сервер.
(4) После того, как миграция успешно завершится и подтвердится, что машина запущена на новом сервере, оригинальная машина удаляется.

Рис. 9. Процесс Live Migration.

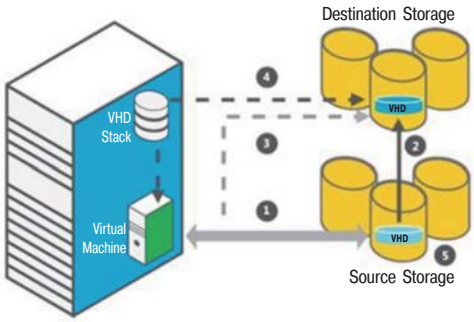
Live Storage Migration

Функция, которую ждали многие, теперь доступна и для новой версии гипервизора от Microsoft. Мобильность данных является важным критерием в современных облачных средах. Live Storage Migration позволяет выполнять перенос виртуальных машин между различными системами хранения без приостановки работы.

Процесс переноса виртуальных машин представлен на рис. 10.

Резервное копирование

В Windows Server 2008 при использовании Windows Backup Вы могли сделать резервную копию только всего тома. Данная проблема касалась и восстановления. В Windows Server 8 Beta была обнаружена возможность создания резервных копий отдельных виртуальных машин.



(1) Вплоть до завершения операции переноса виртуальная машина работает с оригинальным источником виртуальных дисков.
(2) В то время как операции чтения/записи происходят с оригинальным диском, он копируется в новое место.
(3) По завершении копирования все операции записи дублируются на оригинальное и новое расположение, и в это время реплицируются изменения с оригинального источника, произошедшие за время первоначальной копии диска.
(4) По завершении синхронизации виртуальная машина начинает использовать диск с нового места размещения.
(5) Оригинальная копия диска удаляется.

Рис. 10. Процесс Live Storage Migration.

Также приятной новостью является то, что теперь сервис VSS поддерживает не только локальные диски, но и файловые серверы. В Windows Server 2012 Hyper-V появилась возможность инкрементальных бэкапов виртуальных дисков во время работы гостевой ОС.

Управление инфраструктурой

Windows Server 2012 предлагает новый API для управления СХД, основанный на WMI и названный Storage Management API (SMAPI). Он представляет собой набор примитивов, предназначенных для управления как подключенных напрямую СХД, таких как PCI-E RAID-контроллеры, так и внешними системами хранения, подключенных через сеть хранения данных (SAN).

Производители СХД могут поддержать новый API двумя способами:

1. *Реализовать спецификацию SMI-S.* Storage Management Initiative – Specification или SMI-S – это стандарт управления дисковыми хранилищами, разрабатываемый с 2002 года Storage Networking Industry Association. SMI-S является ANSI и ISO стандартом. Актуальная версия SMI-S 1.5. Более 800 различных аппаратных и 75 программных решений поддерживают данный стандарт. Основная идея стандарта – унификация управления дисковыми хранилищами через веб-запросы.

2. *Реализовать новую модель провайдера, называемую Storage Management Provider (SMP).* Управление СХД осуществляется Microsoft Storage Management Service – компонентом операционной системы.

Но необязательно ждать выхода новой версии операционной системы для получения возможности управления СХД

с использованием SMI-S провайдеров. Уже сейчас эта возможность существует в System Center Virtual Machine Manager 2012 (SCVMM 2012).

Являясь ключевым компонентом набора System Center, Virtual Machine Manager 2012 предоставляет гибкое, скоростное и малозатратное частное облако, обеспечивает управление разнородными виртуальными средами с помощью единого средства, оптимизирует существующие приложения для развертывания частного облака.

В VMM 2012 с помощью консоли можно обнаруживать, классифицировать и обеспечивать удаленное хранение на поддерживаемых СХД. VMM полностью автоматизирует назначение хранилищ хостам Hyper-V или кластеру хостов Hyper-V, а также следит за хранилищем, находящимся под управлением VMM. Автоматизация хранения с помощью VMM поддерживается только для хостов Hyper-V.

Чтобы включить новые функции хранения, VMM 2012 использует новую службу Microsoft Storage Management Service (служба управления хранением) для связи с внешними массивами посредством поставщика SMI-S (Storage Management Initiative – Specification – спецификации стандарта управления хранением). Служба Storage Management Service устанавливается по умолчанию при установке VMM 2012. Вы должны установить поддерживаемого поставщика SMI-S на доступный сервер и затем добавить этого поставщика под управление VMM.

SCVMM 2012 использует возможности по управлению СХД для поддержки следующих сценариев использования:

– непрерывное назначение ресурсов. Определение как виртуальные машины, хосты, кластеры связаны с основополагающей инфраструктурой хранения;
– управление ресурсами хранения для хостов и кластеров. Добавление хранилища к хосту или кластеру, включая: назначение тома, инициализацию, создание разделов, форматирование, проверку и создание кластерного ресурса CVS;
– быстрое выделение;
– быстрое создание новых виртуальных машин, используя SAN для копирования VHD(x).

Для удобства управления SCVMM 2012 позволяет выполнять классификацию ресурсов хранения на основе их возможностей.

Обеспечение безопасности

Обеспечение безопасности является одной из основных задач при построении частных и публичных облаков.

Выше было описание того, каким образом осуществляется безопасность при выполнении репликации и при использовании файлового сервера в качестве хранилища образов виртуальных машин. Но Microsoft позаботилась о защите информации не только на уровне каналов связи. Кража носителей злоумышленниками или некорректные действия инженеров при переконфигурировании инфраструктуры или списании оборудования могут скомпрометировать данные компании или ее клиентов.

Начиная с Windows Server 2012, Microsoft позволяет использовать технологию

Сравнение функциональности гипервизоров VMware vSphere 5, Citrix XenServer 6.0 и Windows Server 2012 Hyper-V.

Критерий	VMware vSphere 5	Citrix XenServer 6.0	Windows Server 2012 Hyper-V	Примеч.
Поддерживаемые интерфейсы подключения СХД	FC, iSCSI(аппаратная и программная поддержка), FCoE(аппаратная и программная поддержка), SAS, Infiniband (SRP)	FC, iSCSI (аппаратная и программная поддержка), FCoE (только аппаратная поддержка), SAS	iSCSI(аппаратная и программная поддержка), FC, SAS, FCoE(аппаратная и программная поддержка), Infiniband(SRP, iSER)	1)
Использование файловых хранилищ	NFSv3	NFSv3	SMB 3.0	2)
Возможность использования локального хранилища	Да	Да	Да	3)
Возможность использования программного RAID, в том числе для зеркалирования СХД	Нет	Да, неофициально	Да	4)
Поддержка MPIO	Да	Да	Да	
Использование SSD	Как swap	Нет	Нет	
Загрузка из сети	Да	Да	Да	
Формат виртуальных дисков	vmdk поверх VMFS-5 и NFS	vhd поверх ext3 и NFS, vhd поверх LVM	VHDX	
Максимальный размер виртуального диска	2ТБ	2ТБ, 15ТБ для некоторых СХД	64ТБ	5)
Расширение виртуальных дисков "на лету"	Да	Нет	Да	
Возможность расширения хранилища	Добавление LUN в имеющееся хранилище без перебоя в работе работы Расширение размера LUN без перебоя в работе	Расширение размера LUN требует отключение от СХД или перезагрузку хоста	Расширение CSV 2 возможно при увеличении емкости носителя без простоя в работе виртуальных машин	6)
Возможность классифицировать хранилища и объеденять их в кластера	Да	Нет	Возможность классификации хранилищ заложена в SCVMM 2012	
Возможность "прокинуть" LUN	Да, 2 режима	Да	Да	
Возможность "прокинуть" HBA	Да	Ограниченно	Да	
Поддержка SR-IOV		Да	Да	
Поддержка NPIV	Да, для RDM	Нет	Да	
Поддержка MPIO для FC на гостевых ОС	Нет	Нет	Да	
Поддержка "тонких" дисков	Да, в том числе thin reclamation	Только для файловых хранилищ	Да	
"Горячая" миграция виртуальных машин между хранилищами	Да, в том числе между хранилищами разных типов	Только "холодная" миграция	Да, в том числе между хранилищами разных типов. При этои количество одновремен-ных миграций не ограничено.	
QoS для СХД	SIOC обеспечивает QoS, в том числе для NFS DRS обеспечивает перемещение и правильное размещение гостей на хранилищах	Только для блочных хранилищ Только для гостей, доступных с одного хоста	Нет, но возможно появление сторонних средств, в том числе бесплатных	7)
Обеспечение отказоустойчивости без внешней СХД	VSA Прототип CloudFS	Неофициальная возможность использования DRBD	Да, разрабатывается архитектура Cluster-in-a-Box Storage Spaces можно использовать на JBOD	
Возможность перемещения ВМ между хостами при отсутсвии общего хранилища	Нет	Нет	Да	8)
Дополнительные средсва обеспечения безопасности SAN	Да, изоляция LUN	Нет	Нет прямых средсв, позволяющих дополнительно изоляцию LUN на стороне хостов	9)
Интеграция с СХД	VAAI – API для взаимодействия с СХД и перемещения ресурсоемких операций на сторону СХД, частично стандартизован VASA – API, позволяющее определять параметры СХД VAMP – позволяет разрабатывать собственные плагины для MPIO	StorageLink – позволяет использовать функции СХД, такие как Thin Provisioning, мнгновенные снимки и копии	ODX для поддержки передачи данных между хранилищами Storage Management API (SMAPI) использует SMI-S для управлением СХД SCVMM 2012 использует SMI-S для получения информации о характеристиках СХД и для управления СХД	10)
Связанные образы	Нет	Да, при клонировании из шаблонов, PVS	Да, но есть рекомендации не использовать их при интен-сивных нагрузках ввода-вывода	11)
Обеспечение катастрофоустойчивости	SRM	Site Recovery	Hyper-V Replica	
Инкрементальные резервные копии виртуальных дисков	Да	Нет	Да	
Дедупликация данных на хосте	Нет	Нет	Да, но не доступна для CVS	

1) Также возможна установка драйверов для таких протоколов, как ATAoE.
2) Несомненно, SMB 3.0 является наиболее зрелым и функциональным протоколом. Кроме того, благодаря использованию RDMA, этот протокол будет самым производительным. Но стоит дождаться результатов SPECsfs2008. На начальной стадии очень малое количество СХД будут поддерживать SMB 3.0.
3) Все продукты позволяют использовать локальные диски для хранения образов виртуальных машин. Наиболее широкие возможности по использованию локальных дисков, пожалуй, у Microsoft.
4) Благодаря возможности использования Storage Spaces в Hyper-V можно организовывать избыточность зеркалирования на 1 или 2 носителя или с использованием четности
5) Microsoft предоставляет возможность создания виртуального диска наибольшего объема.
6) VMware предоставляет наиболее удобные способы изменения размеров хранилища, но большинство инженеров не рекомендуют их использовать.
7) VMware предлагает пользователям наиболее интересные функции предоставления QoS для СХД. На наш взгляд, QoS является must-have функциональностью при построении облачных инфраструктур.
8) Данная функция вкупе с использованием Cluster-in-box или подобными решениями будет интересна хостерам.
9) Только VMware позволяет выполнять изоляцию LUN на стороне хостов, опытные инженеры не рекомендуют использовать данную функцию. На платформе Microsoft есть возможность разграничения доступа через дополнительные сервисы.
10) VMware VAAI уже частично стал стандартом, количество SCSI примитивов, используемых для передачи функций на сторону СХД, значительно больше, чем у конкурента. ODX основан на стандартной спецификации. StorageLinks позволяет выполнить наиболее тесную интеграцию с СХД, но технологии проприетарны, и малое количество производителей СХД поддерживает их. Использование SMI-S для определения характеристик массивов и для управления ими является предпочтительным, т.к. данный стандарт открыт и поддерживается большим числом производителей СХД.
11) Самое зрелое решение предложено компанией Citrix.

BitLocker (впервые появилась в Windows Vista) в кластерной конфигурации для шифрования Cluster Shared Volumes версии 2.0. BitLocker позволяет защищать данные путем полного шифрования тома.

В новой версии серверной ОС от Microsoft BitLocker доступен в виде опционального компонента. Кроме того, в Windows Server 2012 будут доступны еще два компонента: Enhanced Storage, предназначенный для поддержки устройств с аппаратным шифрованием, и BitLocker Network Unlock.

Максимальную функциональность BitLocker обеспечивает, работая с Trusted Platform Module версии 1.2 и выше – криптопроцессором, предназначенным для хранения ключей и сертификатов.

Рассмотрим изменения, появившиеся в BitLocker, для операционной системы Windows Server 8 Beta:

- шифрование только использованного пространства тома;
- активация BitLocker до начала установки ОС;
- возможность изменения пароля и стандартным пользователем;
- Network Unlock. Автоматическое разблокирование системного тома при перезагрузке системы при подключении к сети в доменной среде. Функция доступна только для систем, имеющих прошивку UEFI с модулем UEFI DHCP driver;
- возможности шифрования кластерных томов. Основная, на наш взляд, функция добавленная в Windows Server следующего поколения;
- поддержка носителей с встроенным аппаратным шифрованием;
- управление BitLocker в распределенных средах в соответствии с групповыми политиками.

Выводы

Разработчики Microsoft проделали большую работу по развитию сервисов хранения в новых версиях своей операционной системы и гипервизора. Хотя мы и не видим здесь такого разнообразия функций, которое существует в продуктах компании VMware, все вновь добавленные или улучшенные функции действительно являются необходимыми для построения частных и публичных облаков и долгое время ожидались пользователями, которые ориентируются на продукты компании Microsoft при построении своей инфраструктуры.

Сергей Платонов,
компания AVRORAID