



Тема 6


Обзор средств обеспечения надёжности



**Материал подготовил
Студент группы ИВ-622:
Евтушенко Никита**



Надежность хранения данных



Типы RAID массивов

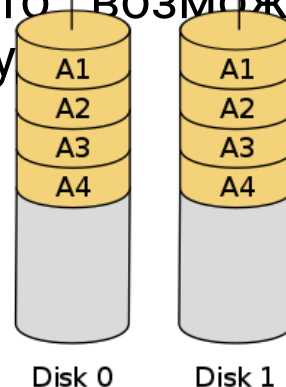
В работе с дисковыми подсистемами IT-специалисты часто сталкиваются с двумя основными проблемами:

- Первая – это низкая скорость чтения / записи, иногда даже скоростей SSD-диска бывает недостаточно.
- Вторая – выход дисков из строя, а значит и потеря данных, восстановление которых бывает невозможно.

- Обе эти проблемы решаются с помощью технологии RAID (*redundant array of independent disks* — избыточный массив независимых дисков) – технологии виртуального хранения данных, объединяющей несколько физических дисков в один логический элемент.
- В зависимости от выбранной спецификации RAID, могут быть повышены скорость чтения / записи и/или уровень защищенности от потери данных.
- Существуют следующие уровни спецификации RAID: 1,2,3,4,5,6,0. Кроме того, существуют комбинации: 01,10,50,05,60,06. Рассмотрим самые распространенные типы RAID-Массивов.

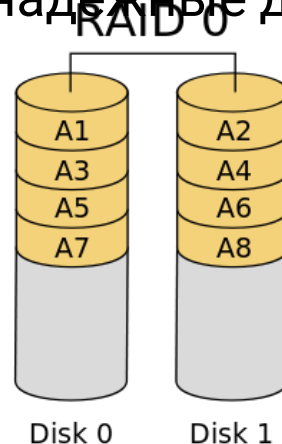
RAID 1

- RAID 1 (также называют «Mirror» – Зеркало) предполагает полное дублирование данных с одного физического диска на другой.
- К недостаткам RAID 1 можно отнести то, что вы получаете в два раза меньше дискового пространства. Т.е. если вы используете ДВА диска по 250 Гб, то система будет видеть всего ОДИН размером 250 Гб. Данный вид RAID не дает выигрыша в скорости, но значительно повышает уровень отказоустойчивости, ведь если один диск выйдет из строя, всегда есть его полная копия. Запись и стирание с дисков происходит одновременно. Если информация была намеренно удалена, то возможности восстановить её с другого диска уже не бу



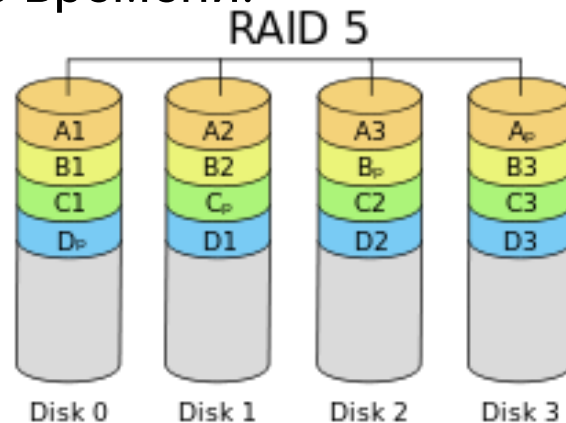
RAID 0

- RAID 0 (также называют «Striping» – Чередование) предполагает разделение информации на блоки и одновременная запись разных блоков на разные диски.
- Такая технология повышает скорость чтения/записи, позволяет пользователю использовать полный суммарный объем дисков, однако понижает отказоустойчивость, вернее сводит её на ноль. Так, в случае выхода из строя одного из дисков, восстановить информацию будет практически невозможно. Для сборки RAID 0 рекомендуется использовать исключительно высоконадежные диски.



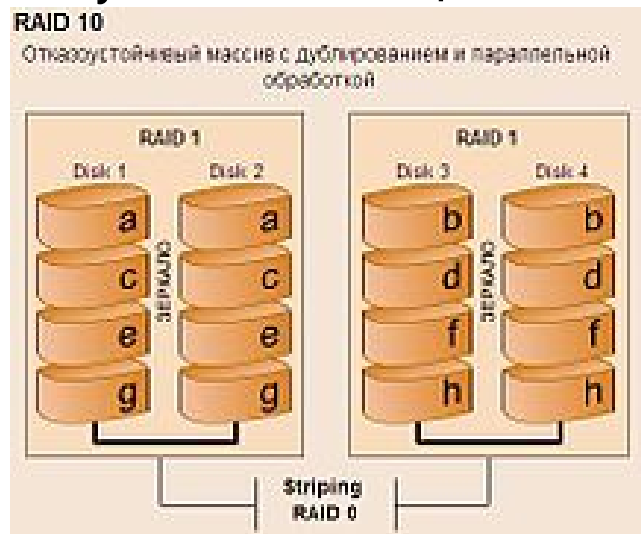
RAID 5

- RAID 5 можно назвать более усовершенствованным RAID 0. Можно использовать от 3 жестких дисков. На все, кроме одного записывается рейд 0, а на последний специальная контрольная сумма, что позволяет сохранить информацию на винчестерах в случае «смерти» одного из них (но не более одного). Скорость работы такого массива высокая. На восстановление информации в случае замены диска потребуется много времени.



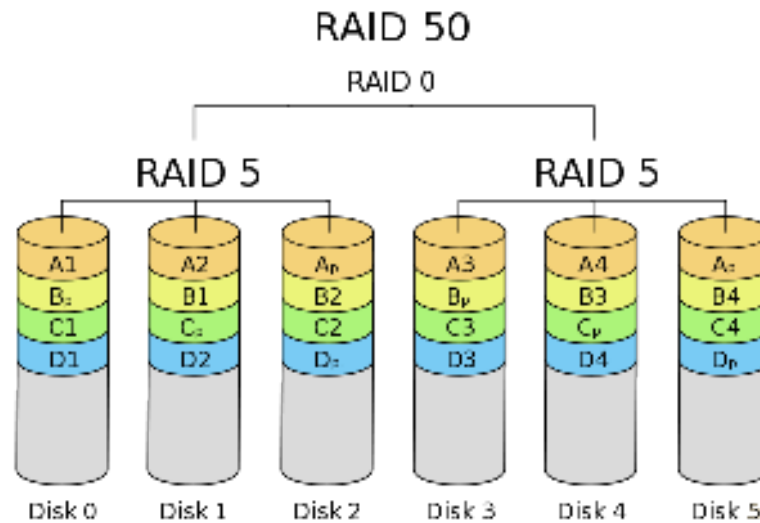
RAID 10

- **Является миксом RAID массивов 1 и 0.** И объединяет в себе плюсы от каждого: высокая производительность и высокая отказоустойчивость.
- Массив обязательно содержит четное количество дисков (минимум 4) и является самым надежным вариантом сохранения информации. Недостатком является высокая стоимость дискового массива: эффективная емкость составит половину от общей емкости дискового пространства.



RAID 50

- **Является миксом RAID массивов 5 и 0.** Строится RAID 5, но его составляющими будут не самостоятельные жесткие диски, а массивы RAID 0.




RAID 2, 3, 4

- Это способы распределенного хранения информации с использованием дисков, выделенных под коды четности. Отличаются друг от друга только размерами блока. На практике практически не используются в связи с необходимостью отдавать большую долю дисковой емкости под хранение кодов ЕСС и/или четности, а также в связи с невысокой производительностью.



На практике чаще всего используют только три вида RAID-массивов. Это RAID 1, RAID 10 и RAID 5.

С точки зрения соотношения стоимость / производительность / отказоустойчивость рекомендуется использовать:

- **RAID 1** (зеркалирование) для формирования дисковой подсистемы для пользовательских операционных систем.
 - **RAID 10** для данных, имеющих высокие требования к скорости записи и чтения. Например, для хранения баз 1С:Предприятие, почтового сервера, AD.
 - **RAID 5** используют для хранения файловых данных.
- 

- Когда поставщик облачных услуг хранит данные пользователя, он должен иметь возможность вернуть эти данные пользователю по требованию. С учетом простоев сети, ошибок пользователей и других обстоятельств выполнение этого условия надежным и детерминированным способом может оказаться затруднительным.
- Существуют интересные новые схемы обеспечения высокой готовности, такие как рассредоточение информации. Компания Cleversafe, которая предоставляет услуги хранения данных в частном облаке, использует алгоритм рассредоточения информации (Information Dispersal Algorithm - IDA) для повышения доступности данных перед лицом физических отказов и простоев сети. Алгоритм IDA позволяет "нарезать" данные для их восстановления в случае потери части данных. Кроме того, IDA позволяет настраивать количество долей данных, так чтобы заданный объект данных можно было разрезать на четыре доли при одном допустимом сбое или на 20 долей при восьми допустимых сбоях. Как и RAID, IDA позволяет восстанавливать данные из подмножества исходных данных при некоторых накладных расходах на коды ошибок (в зависимости от количества допустимых сбоев)



Горячая замена/подключение



HotSwap


- **Hotswap** - (Горячая Перестановка / Горячая Замена) - это возможность замены вышедших из строя дисков без прерывания работы системы. Если в системе используется должным образом сконфигурированный [RAID](#) контроллер, управляющий избыточной дисковой системой ([RAID](#) массивом), то отказ одного дисководов не приводит к прерыванию функционирования системы. В этом случае системой генерируется соответствующее сообщение для системного оператора. Через некоторое время, когда активизируется замещающий диск, системный оператор может удалить отказавший диск, установить новый диск, и дать контроллеру команду "восстановить" данные на новом диске, причем все это происходит без прерывания системных операций и выключения системы.

HotSpare

- Hot Spare - ("Горячее резервирование") - Одна из наиболее важных особенностей, которую обеспечивает RAID контроллер, с целью достичь безостановочное обслуживание с высокой степенью отказоустойчивости. В случае выхода из строя диска, восстанавливающая операция будет выполнена RAID контроллером автоматически, если выполняются оба из следующих условий:
 1. Имеется "резервный" диск идентичного объема, подключенный к тому же контроллеру и назначенный в качестве резервного, именно он и называется Hotspare;
 2. Отказавший диск входит в состав избыточной дисковой системы, например RAID 1, RAID 3, RAID 5 или RAID 0+1.

- В течение процесса автоматического восстановления система продолжает нормально функционировать, однако производительность системы может слегка ухудшиться.
- Для того, что бы использовать особенность резервирования, Вы должны всегда иметь резервный диск ([Hotspare](#)) в вашей системе. В случае сбоя диска, резервный автоматически заменит неисправный диск, и данные будут восстановлены.

Пример реализации Spare ДИСКОВ

 **Storage Management Utility**

System Status

System Time 2020-04-24 17:22:38

System Events 0 0 100

Configuration View

-- ++

Uninitialized Name (P2000 G3 iSCSI)

Logical

Vdisks

vd01 (RAID5)

vd02 (RAID10)

Hosts

Physical

Enclosure 1

vd02 (RAID10)

View Provisioning Configuration Tools Help

vd02 (RAID10) > View > Overview

Vdisk Overview

Details about a specific vdisk

Vdisk Overview					
	Health	Component	Count	Capacity	Storage Space
	OK	Vdisk		599.4GB	<div><div>449.4GB</div><div>150.0GB</div></div>
	OK	Disks	6	599.4GB	<div><div>449.4GB</div><div>150.0GB</div></div>
	OK	Volumes	3	449.4GB	<div><div>449.4GB</div></div>
		Snap Pools	0	0.0KB	

Disk Sets, Total Space: 1.8TB:

599.4GB

600.0GB

600.0GB

Type	Disk Type	Disks	Size
RAID1-1	SAS	1 2	600GB
RAID1-2	SAS	1 2	600GB
SPARE	SAS	1 2	600GB


Tabular

Graphical


Enclosures Front View

	Health	Name	Type	State	Size	Enclosure	Serial Number	Status
	OK	Disk-1.6	SAS	RAID1-2	300.0GB	Enclosure-1	S0M1APX30000B416QZ76	Up
	OK	Disk-1.7	SAS	RAID1-2	300.0GB	Enclosure-1	S0K1CAPD0000B416DFF6	Up
	OK	Disk-1.8	SAS	RAID1-1	300.0GB	Enclosure-1	S0K19QK00000B415J8PK	Up
	OK	Disk-1.9	SAS	RAID1-1	300.0GB	Enclosure-1	S0K1BX4T0000B416FT28	Up
	OK	Disk-1.10	SAS	SPARE	300.0GB	Enclosure-1	S0K1BXCZ0000B416BWV0	Up
	OK	Disk-1.11	SAS	SPARE	300.0GB	Enclosure-1	S0K1BXGJ0000B416BNLC	Up


В примере показан RAID 10 построенный из 6 физических дисков, два из которых RAID 1 по два диска в каждом +2 диска (Spare) для замены в случае отказа одного из RAID 1.




Надежность сети передачи данных



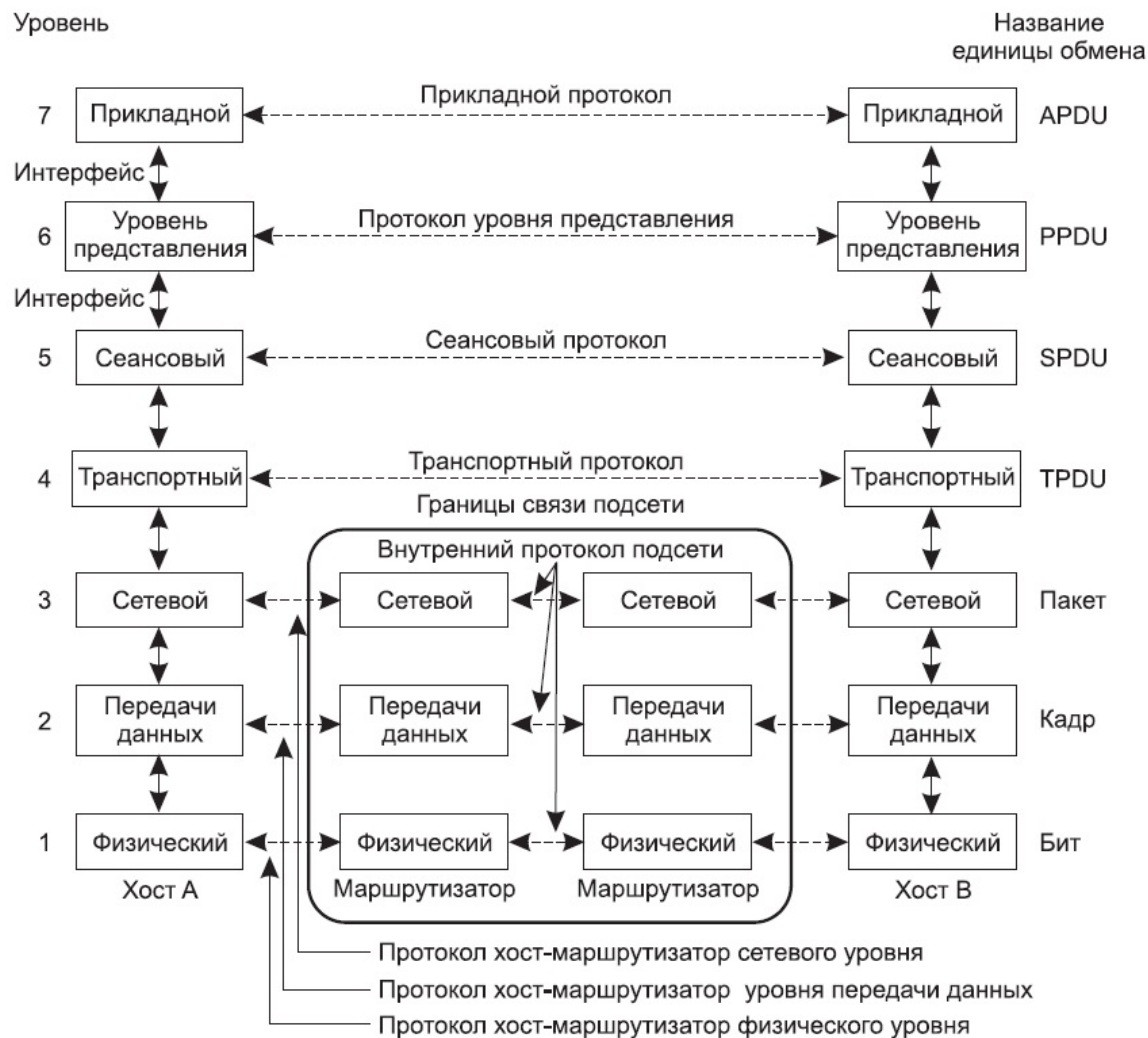
- Эталонная модель OSI (за исключением физической среды). Эта модель основана на разработке Международной организации по стандартизации (International Organization for Standardization, ISO) и является первым шагом к международной стандартизации протоколов, используемых на различных уровнях.
- Называется эта структура эталонной моделью взаимодействия открытых систем (Open System Interconnection), поскольку она связывает открытые системы, то есть системы, открытые для связи с другими системами. Для краткости мы будем называть эту модель просто «модель OSI».



Модель OSI имеет семь уровней. Появление именно такой структуры было обусловлено следующими соображениями.

1. Уровень должен создаваться по мере необходимости отдельного уровня абстракции.
 2. Каждый уровень должен выполнять строго определенную функцию.
 3. Выбор функций для каждого уровня должен осуществляться с учетом создания стандартизированных международных протоколов.
 4. Границы между уровнями должны выбираться так, чтобы поток данных между интерфейсами был минимальным.
 5. Количество уровней должно быть достаточно большим, чтобы различные функции не объединялись в одном уровне без необходимости, но не слишком высоким, чтобы архитектура не становилась громоздкой.
- 

Протоколы



- Облачная архитектура хранения данных — это прежде всего доставка ресурсов хранения данных по требованию в высокомасштабируемой и мультитенантной среде. Обобщенно облачная архитектура хранения данных представляет собой внешний интерфейс, который предоставляет API для доступа к накопителям. В традиционных системах хранения данных это протокол SCSI, но в облаке появляются новые протоколы. Среди них можно найти внешние протоколы Web-сервисов, файловые протоколы и даже более традиционные внешние интерфейсы (Internet SCSI, iSCSI и др.).

- Одним из самых ярких различий между облачной и традиционными системами хранения являются средства доступа к ним. Большинство поставщиков предлагает разные методы доступа, однако общепринятыми являются API Web-сервисов. Многие из них реализованы на принципах REST, что подразумевает объектно-ориентированную схему, разработанную поверх HTTP (с использованием HTTP в качестве транспорта). REST-API без запоминания состояния просты и эффективны. REST-API реализуют многие поставщики облачных услуг хранения, включая Amazon Simple Storage Service (Amazon S3), Windows Azure™ и Mezeo Cloud Storage Platform.
- Одна проблема API Web-сервисов заключается в том, что для того чтобы воспользоваться преимуществами облачной системы хранения, они требуют интеграции с приложением. Поэтому с облачными системами хранения для обеспечения непосредственной интеграции используются также общие методы доступа. Например, протоколы на основе файлов, такие как NFS/Common Internet File System(CIFS) или FTP, или протоколы на основе блоков, такие как iSCSI. Такие методы доступа предоставляют Nirvanix, Zetta, Cleversafe и другие поставщики услуг облачного хранения.

- Вышеупомянутые протоколы наиболее распространены, но для облачного хранения подходят и другие. Один из самых интересных — Web-based Distributed Authoring and Versioning (WebDAV). WebDAV также основан на HTTP и позволяет использовать Web в качестве ресурса для чтения и записи. В число поставщиков, использующих WebDAV, входят Zetta, Cleversafe и другие.
- Можно найти и такие решения, которые поддерживают несколько протоколов доступа. Например, IBM® Smart Business Storage Cloud позволяет использовать протоколы на основе файлов (NFS и CIFS) и протоколы на основе SAN в одной и той же инфраструктуре виртуализации систем хранения данных.

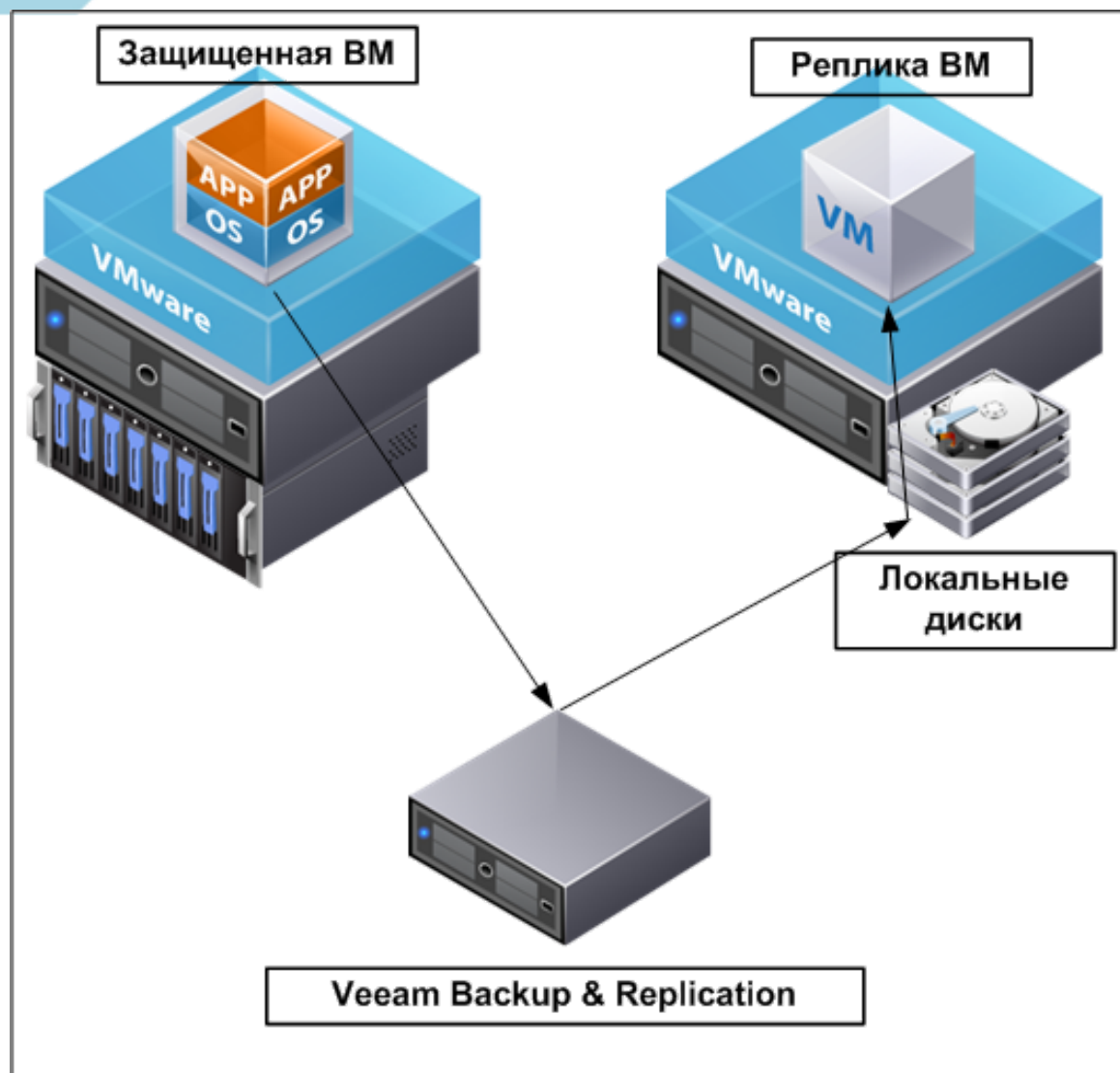


Отказоустойчивость ВМ



Veeam репликация

- Очень полезной в виртуальной инфраструктуре может оказаться функция репликации виртуальных машин.
- Представьте критичную для бизнеса ВМ или несколько ВМ, для которых не подходит вариант с резервным копированием раз в сутки. Нужно всегда иметь ее актуальную резервную копию, готовую к немедленному запуску в случае аварии (выходит из строя гостевая ОС или СХД). Репликация от Veeam сможет обеспечить необходимый уровень отказоустойчивости без приобретения дорогих СХД с функцией репликации.



В случае аварии

Ситуация, при которой виртуальную машину нужно будет восстанавливать из реплики всего две:

- Перестала функционировать гостевая ОС (например, BSOD) – в этом случае если после перезагрузки VM не заработала корректно, придется, либо попытаться починить ее, либо воспользоваться репликой.
- Вышла из строя система хранения данных – файлы VM становятся недоступны и запустить виртуальную машину даже на другом ESX хосте нельзя.

Из консоли vCenter запускаем реплику виртуальной машины, и работа продолжается.

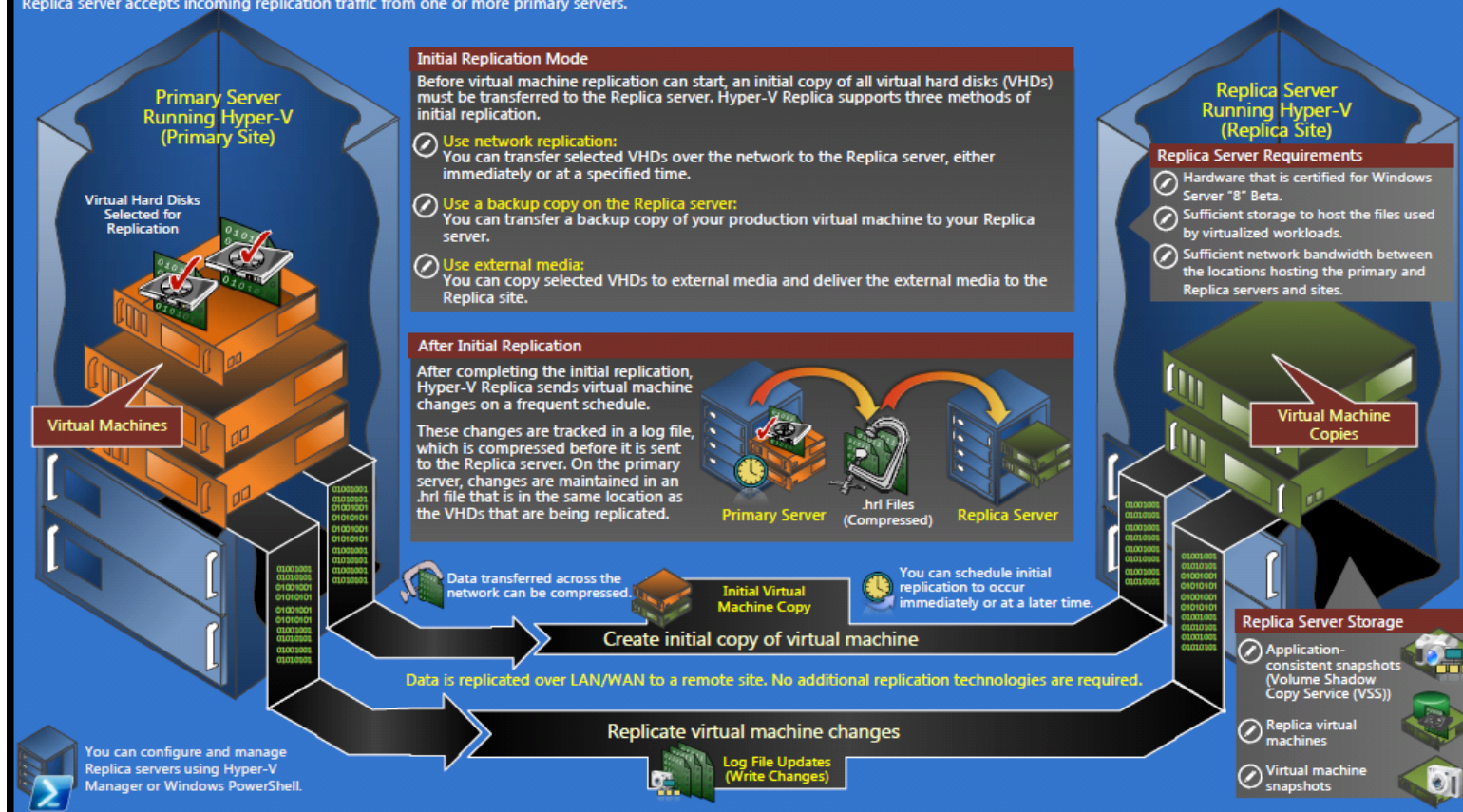
Репликацию можно производить не только внутри локального офиса, но и в резервный ЦОД или ЦОД-ы. Можно расценивать функцию репликации Veeam как не автоматический SRM от VMware (Site Recovery Manager)

Репликация в Hyper V 3

Hyper-V Replica

Virtual Machine Replication

Hyper-V Replica is an asynchronous virtual machine replication technology that is included in Windows Server "8" Beta. It is designed for business continuity and disaster recovery. It works with any server, network, or storage vendor. It does not require any shared storage. It enables you to replicate individual or multiple virtual machines. Hyper-V Replica is tightly integrated with Hyper-V and Failover Clustering. You can replicate virtual machines from one computer running Hyper-V at a primary site (the primary server) to another computer running Hyper-V at a Replica site (the Replica server). The Replica server accepts incoming replication traffic from one or more primary servers.

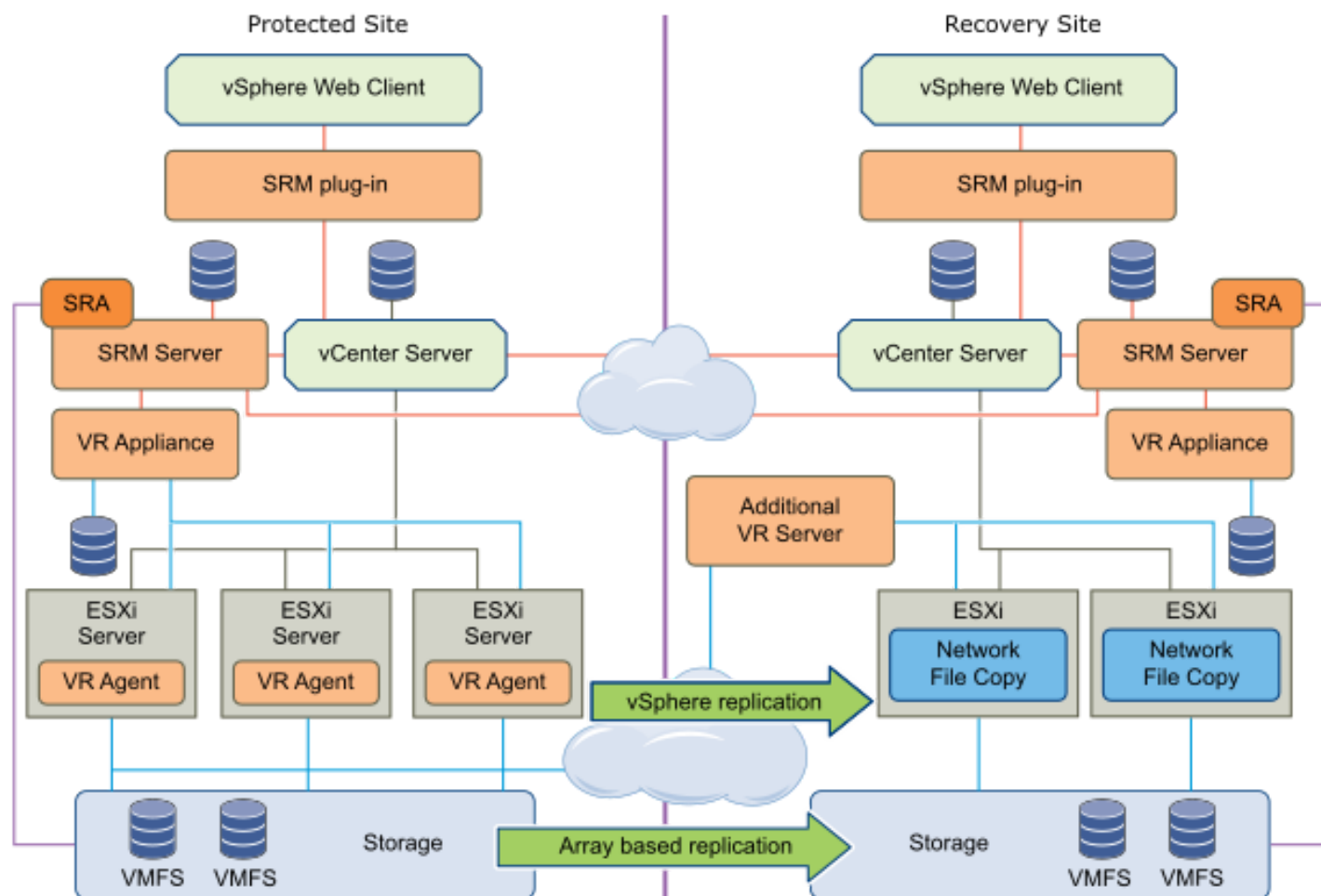


- Во время работы виртуальной машины на одном из серверов (primary server), на соседнем сервере (replica server) создается ее полная копия, всегда готовая к запуску. Причем можно сохранять сразу несколько реплик, за счет мгновенных снимков (snapshot) и запускать любую из них. Например, на стенде репликация происходит каждый час и сохраняются 4 последние реплики. Для репликации не требуется внешнее хранилище, она легко может делаться с сервера на сервер по локальной сети, естественно, что по расписанию передаются только изменения, а не вся виртуальная машин целиком.

VMware Site Recovery Manager

- Это ведущее программное обеспечение для аварийного восстановления, разработанное компанией *VMware*. Данное решение способно обеспечить доступность приложений и возможность их переноса между средами в частных облаках с поддержкой управления на основе политик, тестирования без прерывания рабочих процессов, а также автоматизированного управления. Является расширением *VMware vCenter*, предоставляя защиту от аварий, миграцию сайтов и другие возможности пользователям *VMware*. Полностью интегрируется с *VMware vCenter Server* и веб-клиентом *VMware vSphere*.

- SRM позволяет осуществить быстрое и надежное аварийное восстановление ИТ-инфраструктуры, а также выполнять регулярное тестирование без прерывания рабочих процессов, чтобы обеспечить предсказуемость функционирования ПО для аварийного восстановления ИТ-инфраструктуры и соответствие нормативным требованиям и обеспечить быстрое и надежное восстановление с помощью полностью автоматизированных рабочих процессов и дополнительных решений для программного ЦОД. Кроме того, с помощью SRM возможен перенос приложений без простоев, что позволяет выполнить перенос масштабируемого количества работающих виртуальных машин между средами в пределах региона, используя планы аварийного восстановления для управления переносом приложений между серверами **vCenter** с помощью **VMware vSphere vMotion** без простоев. Также, в SRM присутствует упрощенное управление на основе политик, которое позволяет организовать защиту тысяч виртуальных машин с помощью централизованных планов восстановления, управляемых через веб-клиент **VMware vSphere** и использовать средства автоматизации на основе политик и архитектуру программных ЦОД для упрощения текущего управления.



VMware High Availability

	Avoid Planned Outages	Quick Recovery from Unplanned Outages
Component	NIC Teaming, Multipathing	
Server	VMotion, DRS + Maintenance Mode	VMware HA
Storage	Storage VMotion	Encapsulation, VMware Consolidated Backup
Data	NA	Encapsulation, VMware Consolidated Backup
Site	VMware Site Recovery Manager	

- На платформе виртуализации VMware vSphere можно построить 2 разновидности кластеров: High-availability кластер (HA) и Distributed Resource Scheduler кластер (DRS).
- **HA-кластер** будет означать, что определенное количество физических серверов объединяется в кластер и на них запускаются виртуальные машины. В случае выхода из строя одного из хостов, виртуальные машины запускаются на других серверах из группы, на которых предварительно было выделено для этого место. В итоге время простоя равно времени загрузки операционной системы «виртуалки».

- Если необходимо сократить время простоя до минимального времени рекомендуется использовать технологию **VMware Fault Tolerance**. Основную идею опции можно описать как создание синхронно работающей реплики виртуальной машины на другом сервере и мгновенное переключение на неё при выходе из строя основного хоста.

VMware High Availability



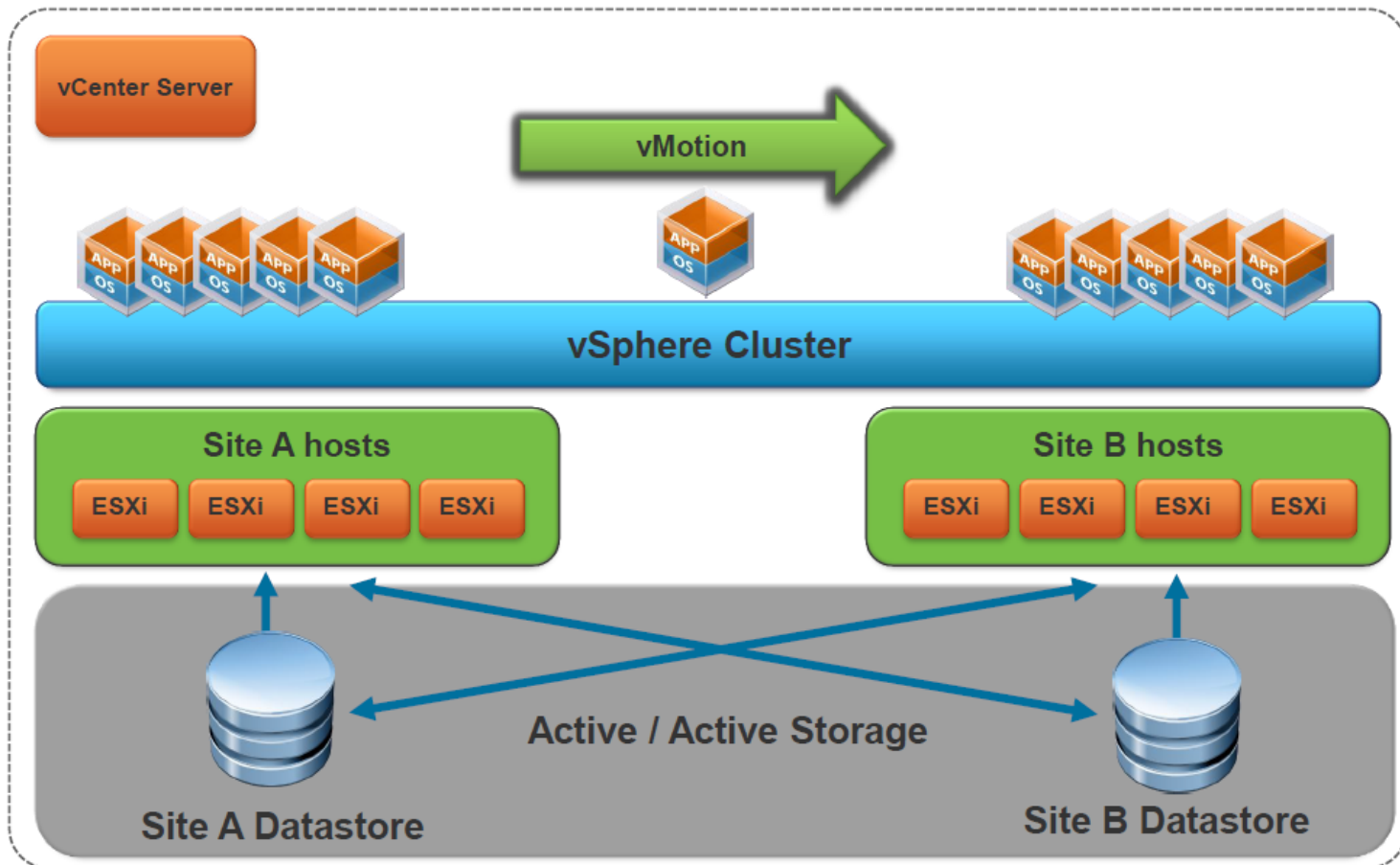
VMware vMotion

Функциональная возможность vSphere, которая позволяет переместить VM с одного хоста на другой в режиме реального времени без прерывания работы.

Назначение:

- автоматическая оптимизация с целью максимальной доступности, масштабирования и эффективного использования пула ресурсов;
- выполнение технического обслуживания оборудования без плановых простоев;
- упреждающая миграция виртуальных машин от отказавших или неэффективных серверов.

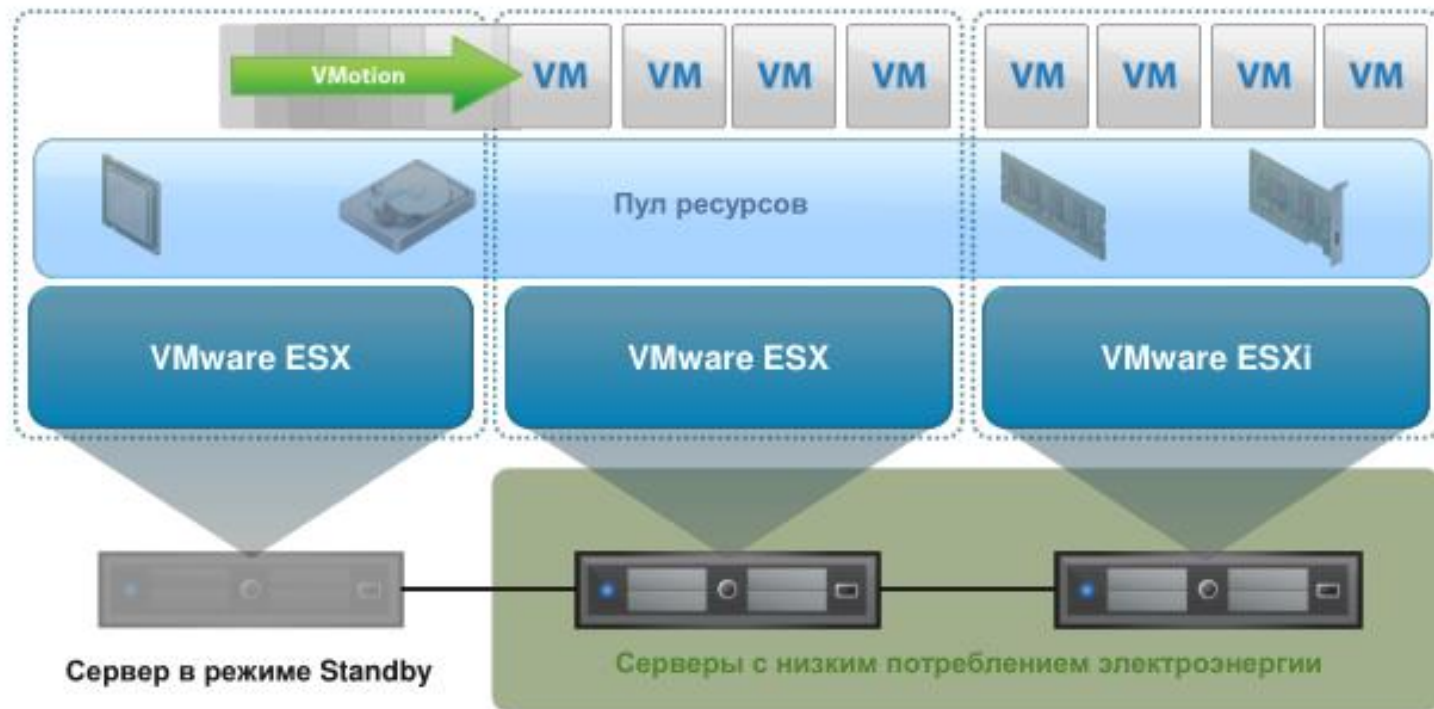
VMware vMotion



VMware Distributed Power Management

- Функции по распределенному управлению питанием VMware Distributed Power Management позволяют экономить электроэнергию для серверов VMware ESX в пределах виртуальной инфраструктуры. Это достигается за счет автоматического перевода в Standby-режим хостов ESX Server при падении загрузки виртуального центра обработки данных. Функции автоматического управления питанием DPM позволяют также автоматически включать серверы ESX при возрастании нагрузки.
- DPM использует механизм автоматической балансировки нагрузки DRS для «горячей» миграции виртуальных машин с хостов ESX, которые можно отключить для минимизации потребления электричества. На данный момент возможности DPM в VMware Virtual Infrastructure 3.5 поддерживаются экспериментально, и не рекомендуется использовать их в производственной среде.

VMware Distributed Power Management



- Анализ DPM основан на тех данных, что собирает и анализирует DRS. DRS выполняет свою работу с пятиминутным интервалом, таким образом, DRS пересчитывает свою аналитику раз в пять минут. При анализе DPM руководствуется данными за 40-минутный интервал.
- По умолчанию DPM считает нормальной нагрузку на сервер $63 \pm 18\%$. Когда нагрузка на серверы превышает 81%, начинается включение серверов. Когда нагрузка падает ниже 45%, DPM начинает консолидировать VM и выключать серверы.



Конец

Источники

- <https://integrus.ru/blog/typy-raid-massivov.html> – Типы RAID массивов
- <https://lawbooks.news/telekommunikatsionnyie-sistemyi-kompyuternye/etalonnaya-model-osi-65020.html> – Модель OSI
- <https://www.ibm.com/developerworks/ru/library/cl-cloudstorage/> – Анатомия облачной инфраструктуры хранения данных
- <https://docs.microsoft.com/ru-ru/azure/virtual-network/virtual-machine-network-throughput> - Пропускная способность сетевых ресурсов виртуальной машины
- http://window.edu.ru/catalog/pdf2txt/802/58802/28664?p_page=2 – Технологии виртуализации
- <http://www.vmstart.ru/resheniya/277-replikaciya-virtualnyx-mashin> – Veeam репликация
- https://ru.bmstu.wiki/VMware_Site_Recovery_Manager – VMware Site Recovery Manager
- <https://www.cloud4y.ru/about/news/chto-takoe-klaster-na-vmware-i-kak-on-ustroen/> - Как устроен кластер на VMware
- <https://www.vmguru.ru/articles/vmware-dpm-drs-settings> – Настройка VMware DM
- <https://www.delphiplus.org/administirovanie-vmware-vsphere/vmware-distributed-power-management.html> – VMware DM
- <https://habr.com/ru/post/97003/> – Hyper-V и устройства хранения данных
- <https://selectel.ru/blog/vsan-v-oblake-na-baze-vmware/> – vSAN в облаке на базе VMware
- <https://habr.com/ru/post/80971/> – О сетях хранения данных
- https://www.bestor.spb.ru/v3/faq?cat_id=87&subcat_id=85&faq_id=482 – Что такое Hotswap?