**МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,**

**СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ ИМ. ПРОФ. М.А. БОНЧ-БРУЕВИЧА»**

**(СПбГУТ)**

Факультет Инфокоммуникационных сетей и систем

Кафедра Защищенных систем связи

Дисциплина Защита информации в центрах обработки данных.

**Пояснительная записка к курсовой работе**

Внедрение Openstack в виртуальную инфраструктуру VMware vSphere

*(тема отчета)*

11.03.02 Инфокоммуникационные технологии и системы связи

*(код и наименование направления/специальности)*

Студент:

Громов А.А. ИКТЗ - 83

Преподаватель:

Цветков А.Ю

Оглавление

[Исходные данные 3](#_Toc89351013)

[Постановка задач 4](#_Toc89351014)

[Алгоритм решения задачи 5](#_Toc89351015)

[Структура OpenStack. 5](#_Toc89351016)

[Обзор VMware Integrated OpenStack 7](#_Toc89351017)

[Основные возможности VMware Integrated OpenStack 8](#_Toc89351018)

[Основные сценарии использования VMware Integrated OpenStack 11](#_Toc89351019)

[VMware vCenter driver for OpenStack Compute. 12](#_Toc89351020)

[Высокоуровневое представление архитектуры драйвера VMware. 14](#_Toc89351021)

[Применённое решение. 15](#_Toc89351022)

[Развёртывание vAPP шаблона OpenStack 15](#_Toc89351023)

[Автоматическое развёртывание OpenStack 22](#_Toc89351024)

[Вывод 31](#_Toc89351025)

## Исходные данные

Исследование принципов работы протоколов SAN-сетей: FC, iSCSI и FCoE.

## Постановка задач

Познакомиться с типами систем хранения данных DAS, NAS и SAN.

Рассмотреть разницу между различными СХД: DAS, NAS и SAN.

Подробнее рассказать про SAN

Провести сравнение FC и ISCSI

Рассказать про NVME.

Составить сравнительную таблицу протоколов SAN (скорость, стоимость).**Алгоритм решения задачи.**

**Системы хранения данных.**

В случае отдельного ПК под системой хранения данных (СХД) можно понимать отдельный внутренний жесткий диск или систему дисков.

Традиционно можно выделить три технологии организации хранения данных: Direct Attached Storage (DAS), Network Attach Storage (NAS) и Storage Area Network (SAN). Они различаются порядком организации доступа к системам хранения:

* DAS - система хранения, непосредственно подключаемая к серверу;
* NAS - система хранения, подсоединяемая к сети;
* SAN - сеть хранения данных. Основой SAN является выделенная специализированная сеть, которая служит исключительно для организации доступа к данным.

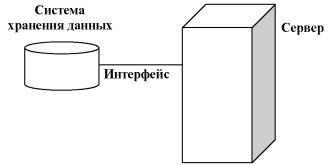
Основное отличие технологий связано с сетевыми интерфейсами. В случае с DAS все три компонента (приложения, файловая система и диски) размещены в одной системе.

Трафик данных в NAS-решениях идет по сети Ethernet, что с точки зрения безопасности является не безупречным вариантом. А поскольку по локальной сети идет и другой трафик, кроме NAS, то скорость передачи данных ниже, чем у DAS, и при обращении приложений к данным возможны заметные потери производительности. Дополнительные нагрузки, связанные с протоколом TCP/IP, также негативно влияют на быстродействие. Однако, будущие реализации iSCSI (SCSI over Ethernet) обещают улучшенное быстродействие.

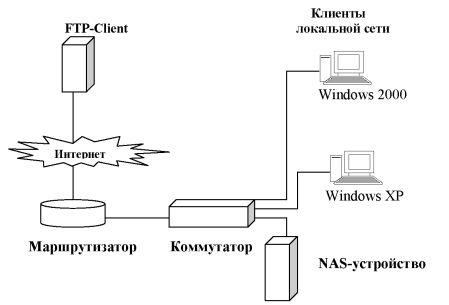
В сети SAN физические диски сосредоточены в единой сети. Они могут консолидироваться в огромные фермы, которые могут предоставлять необходимое дисковое пространство по требованию.

Каждая из трех технологий хранения имеет свои плюсы и минусы, однако часто невозможно найти компромисс между высокой готовностью и низкими ценами, поэтому при построении систем с заделом на будущее необходимо тщательно проработать вопросы масштабируемости и производительности. В таблице 1 отражены особенности технологий хранения.

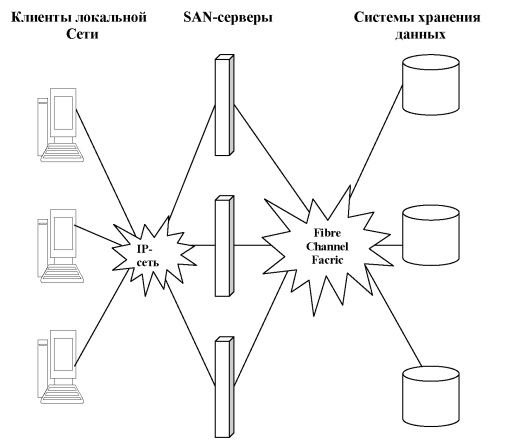
|  |  |  |  |
| --- | --- | --- | --- |
| Технологии | DAS | NAS | SAN |
| Сетевая технология | Нет | Да | Да |
| Защита от внешних помех | Нет (медь) | Нет (медь) / да (оптика) | Да (оптика) |
| Уровень цен | Низкий | Высокий | Очень высокий |
| Масштабируемость | Плохая | Хорошая | Очень хорошая |
| Максимальное расстояние до сервера | 25м | - | 10км |
| Основной используемый интерфейс | SCSI | Ethernet | FCP/ISCI |



Пример инфраструктуру DAS



Пример инфраструктуры NAS



Пример инфраструктуры SAN

FC vs ISCSI

В данный момент существует 2 основных протокола передачи информации в san-сетях:

* Fiber Channel Protocol(FCP)
* Internet Small Computer System Interface (ISCSI)

FC-SAN

Fibre Channel — популярный протокол хранения, обеспечивающий низкие задержки и высокую пропускную способность за счёт своих архитектурных особенностей. Fibre Channel не требователен к ресурсам и отлично подходит для передачи большого объёма данных, так как все операции FC выполняются на стороне HBA, разгружая центральный процессор.

Новые версии протокола Fibre Channel обратно совместимы с прошлыми редакциями, что открывает хорошие перспективы для модернизации и масштабирования. Например, если внедрять FC 32Гб/с, то всё ещё можно будет использовать FC 8Гб/с и 16Гб/с, т.е. можно поэтапно менять FC-коммутаторы и FC адаптеры.

В ближайшее время FC будет обновлён до 64Гб/с и 128Гб/с (уже сейчас есть коммутаторы, поддерживающие агрегацию 4-х портов 32Гб/с в один канал 128Гб/с для соединения коммутаторов).

Простота настройки и удобство в администрировании позволили FC стать одним из наиболее распространенных протоколов хранения. Большинство администраторов SAN-сетей во всем мире знает, как он устроен и какие преимущества обеспечивает при решении различных задач. При этом FC всё ещё сложнее, чем Ethernet, хотя и обладает большим количеством средств управления и мониторинга.

|  |  |  |  |
| --- | --- | --- | --- |
| Название | Скорость линии, Гбод | Пропускная способность, Мбайт/с | Год |
| 1GFC | 1 | 100 | 1997 |
| 2GFC | 2 | 200 | 2001 |
| 4GFC | 4 | 400 | 2004 |
| 8GFC | 8 | 800 | 2005 |
| 10GFC | 10 | 1,200 | 2008 |
| 16GFC | 14 | 1,600 | 2011 |
| 32GFC "Gen 6" | 28 | 3,200 | 2016[[10]](https://en.wikipedia.org/wiki/Fibre_Channel#cite_note-g620release-10) |
| 64GFC "Gen 7" | 29 | 6,400 | 2019[[11]](https://en.wikipedia.org/wiki/Fibre_Channel#cite_note-fcpi7-11) |
| 128GFC "Gen 6" | 28.05 ×4 | 12,800 | 2016[[10]](https://en.wikipedia.org/wiki/Fibre_Channel#cite_note-g620release-10) |
| 256GFC "Gen 7" | 28.9 ×4 | 25,600 | 2019[[12]](https://en.wikipedia.org/wiki/Fibre_Channel#cite_note-Petrilla256GFC-12) |
| 128GFC "Gen 8" | 57 | 12,800 | Planned 2022 |

В Parallel FC на 128 Гбит/с используются четыре оптических волокна для передачи данных в прямом направлении и четыре — в обратном. Какие кабели и соединители потребуются для FC на 128 Гбит/с? В качестве портов могут использоваться модули QSFP28, CFP2, CFP4 или какие-то будущие четырехканальные интерфейсы, а для подключения оборудования — 12-волоконные кабели с соединителями MPO, активные волоконно-оптические кабели (AOC) длиной до 50 м или кабели прямого подключения (DAC) до 5 м.

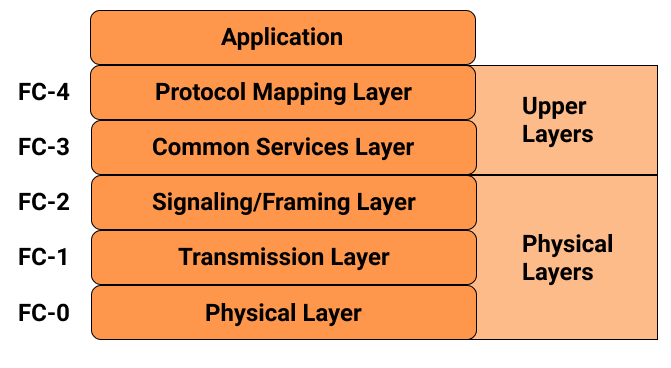
The anatomy of a Fibre Channel SAN

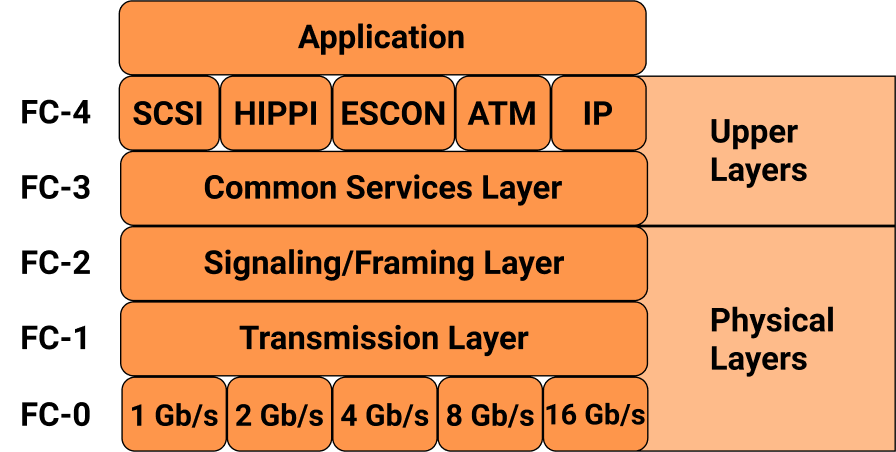
Fibre Channel technology is one SAN technology you can choose for your SMB. Fibre Channel is an extremely stable mature layer-2 switching SAN technology, and the predominant SAN variant today in the enterprise. It preceded the iSCSI standard by nine years, and the Fibre Channel protocol (FCP) is the only commercially utilized protocol on Fibre Channel. FCP is the standard storage SCSI storage protocol wrapped in Fibre Channel frames. There are three physical components to a Fibre Channel SAN. The first piece is the Fibre Channel host bus adapter (HBA), otherwise known as the initiator. An HBA is either a PCIe card or chip residing on the host motherboard. It embeds the FC standard into silicon, minimizing host resources. A software driver resides on the host as well. The second piece is the Fibre Channel switch. The FC switch is a non-blocking (no internal oversubscription of bandwidth) layer-2 switch with extensive management capabilities that conform to the FC standard. The variation of the FC switch is the FC director. The FC director is a large port count non-blocking switch; it has high-availability (HA) with no single point of failure and enterprise class feature functions (such as encryption, virtualized SANs, gateways to other networks, etc.). The third piece is the objective of the HBAs called the FC storage target port, or target. It is the external storage connection to the Fibre Channel SAN (FC fabric.) The target storage port provides the address (which is built into the silicon) or worldwide name (WWN) address of each target storage port. The hosts each have their own unique WWN address that allows them to connect to target storage for each read or write session. One key piece of software is Fibre Channel SAN multipathing code that resides on each host accessing the external storage. This allows each host to have multiple paths to its target storage for primarily path failures and in some cases, load balancing and performance aggregation. FC initiator, target, and switch port bandwidth are currently available as 8 Gbps or 4 Gbps interoperable variants. They're also interoperable with previous 2 Gbps but not 1 Gbps variants. Page 5 of 8 In the next part of our series on Fibre Channel vs. iSCSI SANs, we'll look at the components of an iSCSI storage area network.

**Тут начал редактирование**

**Fibre Channel Protocol**

Как и сетевая модель TCP/IP, Fibre Channel состоит из пяти уровней. Каждый уровень выполняет определённый набор функций.





**Краткий обзор каждого уровня**

**FC-0** — уровень физических интерфейсов и носителей. Описывает физическую среду: коннекторы, кабели, трансиверы, HBA, электрические и оптические параметры.

**FC-1** — уровень передачи и кодирования. Описывает кодирование и декодирование данных перед передачей и после получения. Уровень имеет три основные функции:

* Кодирование / декодирование
* Ordered sets
* Инициализация соединения (link initialization)

**FC-2** — уровень кадрирования и сигналов. Создает FC кадры и управляет потоком. Контроль потока нужен нам для проверки того, сколько данных мы можем отправить в определенный момент времени, чтобы принимающая сторона могла обработать все запросы. Функции, осуществляемые на этом уровне:

* Кадрирование (определение структуры кадра).
* Управление последовательностями (Sequence management)
* Управление обменом (Exchange management)
* Класс обслуживания (Class of Service)
* Управление потоком (Flow control)

**FC-3** — уровень базовых служб. Заложен, для новых функций, которые могут быть внедрены в Fibre Channel. На этом уровне обеспечивается шифрование и сжатие данных перед отправкой.

**FC-4** — уровень отображения протоколов. Обеспечивает сопоставление возможностей Fibre Channel с протоколами верхнего уровня. Наиболее популярным является SCSI. Благодаря FCP, есть возможность использовать протокол передачи данных SCSI с Fibre Channel инфраструктурой.

Рассмотрим физические уровни Fibre Channel Protocol.

**FC-0**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Тип волокна | Скорость (MB/s) | Трансмиттеры | Модификация | Расстояние |
| Одно- модовое волокно (SMF) | 12,8 | 1,310 nm Длинноволновой лазер | 128GFC-PSM4 | 0.5m - 0.5km |
| 1,270, 1,290, 1,310 and 1,330 nm Длинноволновой лазер | 128GFC-CWDM4 | 0.5m - 2km |
| 6,4 | 1,310 nm Длинноволновой лазер | 64GFC-LW | 0.5m - 10km |
| 3,2 | 1,310 nm Длинноволновой лазер | 3200-SM-LC-L | 0.5m - 10km |
| 1,6 | 1,310 nm Длинноволновой лазер | 1600-SM-LC-L | 0.5m - 10km |
| 1,490 nm Длинноволновой лазер | 1600-SM-LZ-I | 0.5m - 2km |
| 800 | 1,310 nm Длинноволновой лазер | 800-SM-LC-L | 2m - 10km |
| 800-SM-LC-I | 2m - 1.4km |
| 400 | 1,310 nm Длинноволновой лазер | 400-SM-LC-L | 2m - 10km |
| 400-SM-LC-M | 2m - 4km |
| 400-SM-LL-I | 2m - 2km |
| 200 | 1,550 nm Длинноволновой лазер | 200-SM-LL-V | 2m - 50km |
| 1,310 nm Длинноволновой лазер | 200-SM-LC-L | 2m - 10km |
| 200-SM-LL-I | 2m - 2km |
| 100 | 1,550 nm Длинноволновой лазер | 100-SM-LL-V | 2m - 50km |
| 1,310 nm Длинноволновой лазер | 100-SM-LL-L | 2m - 10km |
| 100-SM-LC-L |
| 100-SM-LL-I | 2m - 2km |
| Много- модовое волокно (MMF) | 12,8 | 850 nm Коротковолновой лазер | 128GFC-SW4 | 0m - 100m |
| 6,4 | 64GFC-SW | 0m - 100m |
| 3,2 | 3200-SN | 0m - 100m |
| 1,6 | 1600-M5F-SN-I | 0.5m - 125m |
| 1600-M5E-SN-I | 0.5m - 100m |
| 1600-M5-SN-S | 0.5m - 35m |
| 1600-M6-SN-S | 0.5m - 15m |
| 800 | 800-M5F-SN-I | 0.5m - 190m |
| 800-M5E-SN-I | 0.5m - 150m |
| 800-M5-SN-S | 0.5m - 50m |
| 800-M6-SN-S | 0.5m - 21m |
| 400 | 400-M5F-SN-I | 0.5m - 400m |
| 400-M5E-SN-I | 0.5m - 380m |
| 400-M5-SN-I | 0.5m - 150m |
| 400-M6-SN-I | 0.5m - 70m |
| 200 | 200-M5E-SN-I | 0.5m - 500m |
| 200-M5-SN-I | 0.5m - 300m |
| 200-M6-SN-I | 0.5m - 150m |
| 100 | 100-M5E-SN-I | 0.5m – 860m |
| 100-M5-SN-I | 0.5m - 500m |
| 100-M6-SN-I | 0.5m - 300m |
| 100-M5-SL-I | 2m - 500m |
| 100-M6-SL-I | 2m - 175m |

В таблице Х представлены актуальные на данный момент типы волокна, скорости, модификации и расстояния.

Существует 2 типа волокна: многомодовое и одномодовое

Многомодовое волокно (Multimode Fiber, MMF) —широкое в сечении (50-62,5 микрон) относительно одномодового, предназначено для коротковолновых лазерных лучей. Слово многомодовое означает, что по волокну может идти сразу несколько лучей света, отражаясь от стенок волокна. Это позволяет быть кабелю менее чувствительным к перегибам, однако снижает силу и ухудшает качество сигнала, таким образом накладывая ограничение на дистанцию передачи данных до 500 м.

Одномодовое волокно (Singlemode Fiber, SMF) — тонкое волокно (8-10 микрон), через которое передается, не видимый человеческому глазу, сигнал длинноволновым лазером. Свет в таком волокне передается по прямой, без отражений от стенок волокна. По этой причине сигнал передаётся быстрее и с меньшими потерями. Однако из-за высоких цен на оборудование, передача данных по одномодовому волокну производится только для больших расстояний (до 50км без усилителей).

Для подключения двух устройств и работы дуплексного режима, необходимо использовать два кабеля: один для передачи, второй для приема.

**Тут закончил редактирование**

Плавный переход от FC-0 к FC-1 и обратно обеспечивает ASIC — элемент таких устройств как HBA, дисковых массивов и коммутаторов.

Из Википедии:

ASIC (аббревиатура от англ. application-specific integrated circuit, «интегральная схема специального назначения») — интегральная схема, специализированная для решения конкретной задачи. В отличие от интегральных схем общего назначения, специализированные интегральные схемы применяются в конкретном устройстве и выполняют строго ограниченные функции, характерные только для данного устройства; вследствие этого выполнение функций происходит быстрее и, в конечном счёте, дешевле. Примером ASIC может являться микросхема, разработанная исключительно для управления мобильным телефоном, микросхемы аппаратного кодирования/декодирования аудио- и видео-сигналов (сигнальные процессоры).

В оборудовании Fibre Channel ASIC состоит из следующих функциональных элементов:

Encoder / Decoder — обеспечивает кодирование каждых 8 бит передаваемых данных в 10-битное представление. И декодирование обратно принимаемых данных.

SERDES (Serializer / Deserializer) — преобразует параллельный поток 10-битных порций данных в последовательный поток 10-битных порций данных.

Transceiver — преобразует электрические импульсы в световые сигналы.

ASIC

Transceivers, трансиверы или SFP — в случае FC-коммутаторов это отдельные модули, необходимые для подключения кабеля к порту. Различаются на коротковолновые (Short Wave, SW, SX) и длинноволновые (Long Wave, LW, LX). LW-трансиверы совместимы с многомодовым и одномодовым волокном. SW-трансиверы — только с многомодовым. И к тем и к другим кабель подключается разъёмом LC.

Есть ещё SFP xWDM (Wavelenght Division Multiplexing), предназначенные для передачи данных из нескольких источников на дальние расстояния единым световым пучком. Для подключения кабеля к ним используется разъём SC.

FC-1

8/10 и 64/66

Первое, что происходит на этом уровне — кодирование / декодирование информации. Это довольно мудрёный процесс, в ходе которого каждые 8 бит поступающей информации преобразуются в 10-битное представление. Делается это с целью повышения контроля целостности данных, отделения данных от служебных сигналов и возможности восстановления тактового сигнала из потока данных (сохранение баланса нулей и единиц).

Это ведёт к заметному снижению полезной пропускной способности, ибо как можно подсчитать, 20% потока данных является избыточной служебной информацией. А ведь кроме всего прочего, немалую часть этого потока может занимать служебный трафик.

Однако хорошая новость в том, что кодирование 8/10 используется в оборудовании 1G, 2G, 4G и 8G. В части реализаций 10G и начиная с 16G кодирование осуществляется по принципу 64/66, что существенно увеличивает полезную нагрузку (до 97% против 80% в случае 8/10).

Ordered sets

В русской википедии этот термин переведён как "упорядоченные наборы". В то время как на мой взгляд, слово order тут стоит понимать не в значении «порядок», а в значении «приказ, команда».

Для начала стоит упомянуть ещё один термин, используемый в контексте FC — transmission word — минимальная порция данных для передачи, равная 4 байтам. Если передаваемая информация меньше по объёму, то transmission word дополняется специальными заполняющими байтами (fill bytes), которые вырезаются на приёмнике.

Так вот, ordered sets — это специальные служебные transmission words. Делятся на три категории:

Разделители фреймов (Start-of-Frame, SOF и End-of-Frame, EOF).

Два базовых сигнала — IDLE (порт готов принимать или передавать данные) и R\_RDY (receiver ready — порт освободил буфер для приёма очередной порции данных)

Базовые последовательности (primitive sequences):

NOS (Not Operational) — порт обнаружил разрыв / отсутствие соединения

OLS (Offline State) — порт инициирует установление соединения, или порт получил NOS, или порт переходит в состояние off-line

LR (Link Reset) — инициализация сброса соединения. Отправляется в случае получения OLS или каких-то ошибок приёма-передачи (как правило, на уровне Flow Control). Отправивший порт очищает свои буферы и их счётчики

LRR (Link Reset Response) — ответ на LR. Отправивший порт очищает свои буферы и их счётчики

Инициализация соединения (Link initialization)

При установлении физического соединения между портами A и B, между ними происходит следующий «обмен веществ»:

FC-2

Фреймы (Кадры, Frames)

Все данные, передаваемые в среде Fiber Channel разбиваются на фреймы (кадры). Структура фрейма следующая:

SoF — 4 байта (1 tw) — идентификатор начала фрейма.

Header — 24 байта (6 tw) — заголовок. Содержит такую информацию как адрес источника и приёмника, тип фрейма (FT-0 — управляющий или FT-1 — данные), номер последовательности и порядковый номер фрейма в ней и прочая служебно-контрольная информация.

Data — 0-2112 байт (0-528 tw) — непосредственно данные (например, SCSI-команды).

CRC — 4 байта (1 tw) — контрольная сумма.

EoF — 4 байта (1 tw) — идентификатор конца фрейма.

Промежутки между фреймами заполняются специальными «заполняющими словами» — fill word. Как правило, это IDLE, хотя начиная с FC 8G было стандартизовано использование ARB(FF) вместо IDLE, в целях снижения электрических помех в медном оборудовании (но по-умолчанию коммутаторами используется IDLE).

Последовательности (Sequences)

Чаще всего источник стремится передать приёмнику гораздо больше информации, чем 2112 байт (максимальный объём данных одного фрейма). В этом случае информация разбивается на несколько фреймов, а набор этих фреймов называется последовательностью (sequence). Чтобы в логическую последовательность фреймов не вклинилось что-то лишнее при параллельной передаче, заголовок каждого фрейма имеет поля SEQ\_ID (идентификатор последовательности) и SEQ\_CNT (номер фрейма в последовательности).

Обмен (Exchange)

Одна или несколько последовательностей, отвечающих за какую-то одиночную операцию, называется обменом. Источник и приёмник могут иметь несколько параллельных обменов, но каждый обмен в единицу времени может содержать только одну последовательность. Пример обмена: инициатор запрашивает данные (последовательность 1), таргет возвращает данные инициатору (последовательность 2) и затем сообщает статус (последовательность 3). В этот набор последовательностей не может вклиниться какой-то посторонний набор фреймов.

Для контроля этого процесса заголовок каждого фрейма включает поля OX\_ID (Originator Exchange ID — заполняется инициатором обмена) и RX\_ID (Responder Exchange ID — заполняется получателем в ответных фреймах, путём копирования значения OX\_ID).

Классы обслуживания (Classes of Services)

Различные приложения предъявляют разные требования к уровню сервиса, гарантии доставки, продолжительности соединения и пропускной способности. Некоторым приложениям требуется гарантированная пропускная способность в течение их работы (бэкап). Другие имеют переменную активность и не требуют постоянной гарантированной пропускной способности канала, но им нужно подтверждение в получении каждого отправленного пакета. Для удовлетворения таких потребностей и обеспечения гибкости, FC определяет следующие 6 классов обслуживания.

Class 1

Для этого класса устанавливается выделенное соединение, которое резервирует максимальную полосу пропускания между двумя устройствами. Требует подтверждения о получении. Требует чтобы фреймы попадали на приёмник в том же порядке, что вышли из источника. Ввиду того, что не даёт другим устройствам использовать среду передачи, используется крайне редко.

Class 2

Без постоянного соединения, но с подтверждением доставки. Не требует соответствия порядка отправленных и доставленных фреймов, так что они могут проходить через фабрику разными путями. Менее требователен к ресурсам, чем класс 1, но подтверждение доставки приводит к повышенной утилизации пропускной способности.

Class 3

Без постоянного соединения и без подтверждения доставки. Самый оптимальный с точки зрения использования ресурсов фабрики, но предполагает, что протоколы верхних уровней смогут собрать фреймы в нужном порядке и перезапросить передачу пропавших фреймов. Наиболее часто используемый.

Class 4

Требует постоянного соединения, подтверждение и порядок фреймов как и класс 1. Главное отличие — он резервирует не всю полосу пропускания, а только её часть. Это гарантирует определённое QoS. Подходит для мультимедиа и Enterprise-приложений, требующих гарантированного качества соединения.

Class 5

Ещё до конца не описан и не включен в стандарт. Предварительно, класс, не требующий соединения, но требующий немедленной доставки данных по мере их появления, без буферизации на устройствах.

Class 6

Вариант класса 1, но мультикастовый. То есть от одного порта к нескольким источникам.

Class F

Класс F определён в стандарте FC-SW для использования в межкоммутаторных соединениях (Interswitch Link, ISL). Это сервис без постоянного соединения с уведомлениями о сбое доставки, использующийся для контроля, управления и конфигурирования фабрики. Принцип похож на класс 2, но тот используется для взаимодейтсвия между N-портами (порты нод), а класс F — для общения E-портов (межкоммутаторных).

Flow Control

В целях предотвращения ситуации, когда отправитель перегрузит получателя избыточным количеством фреймов так, что они начнут отбрасываться получателем, FC использует механизмы управления потоком передаваемых данных (Flow Control). Их два — Buffer-to-Buffer flow control и End-to-End flow control. Их использование регламентируется классом обслуживания. Например, класс 1 использует только механизм End-to-End, класс 3 — Buffer-to-Buffer, а класс 2 — оба эти механизма.

Buffer-to-Buffer flow control

Принцип технологии — кредит в каждый дом отправка любого фрейма должна быть обеспечена наличием кредита на отправку.

Все поступающие на вход порта фреймы помещаются в специальную очередь — буферы. Количество этих буферов определяется физическими характеристиками порта. Один буфер (место в очереди) соответствует одному кредиту. Каждый порт имеет два счётчика кредитов:

TX BB\_Credit — счётчик кредитов передачи. После отправки каждого фрейма, уменьшается на 1. Если значение счётчика стало равным нулю — передача невозможна. Как только от порта-приёмника получено R\_RDY, счётчик увеличивается на 1.

RX BB\_Credit — счётчик кредитов приёма. Как только фрейм принят и помещён в буфер, уменьшается на 1. Когда фрейм обрабатывается или пересылается дальше, счётчик увеличивается на 1, а отправителю отправляется R\_RDY. Если значение счётчика падает до 0, то в принципе, приём новых фреймов должен быть прекращён. На практике, из-за ошибок синхронизации кредитов может возникнуть ситуация, что источник прислал ещё один-несколько фреймов уже после того как RX BB\_credit стал равен нулю. Данная ситуация называется buffer overflow. В большинстве реализаций порт обрабатывает такие ситуации «по-доброму» — за счёт резервных буферов. Хотя некоторое оборудование в таких случаях может сынициировать Link Reset.

Отсюда исходит сильное влияние расстояния между портами на производительность. Чем выше расстояние и больше пропускная способность, тем больше фреймов будет отправлено (читай кредитов передачи потрачено) ещё до того как получатель получит хотя бы первый. Ситуацию облегчает особенность архитектуры FC-коммутаторов. Дело в том, что количество буферов не закреплено жёстко за каждым портом (кроме восьми обязательных), а является общим для всех. И в случае определения «дальних линков» (автоматически или вручную) количество выделяемых коммутатором буферов для этого порта увеличивается. Другой плюс общей памяти — не требуется гонять буферы от одного порта к другому внутри коммутатора.

End-to-End flow control

Реализуется счётчиком EE\_Credit, который определяет максимум фреймов, которые источник может отправить приёмнику без получения подтверждения (Acknowledge, ACK). В отличие от BB\_Credit распространяется только на фреймы с данными, а обмен/учёт происходит между конечными нодами.

Конец

Изначально мне казалось, что статья будет раза в два меньше, но в ходе написания всплыло много деталей, без которых счастье казалось не полным. Ещё кучу вещей, которые хотелось бы осветить, пришлось пока отбросить — процесс написания грозил стать бесконечным. Если у кого-то возникнут замечания, предложения и пожелания к тому, про что ещё стоит написать, буду признателен. И спасибо всем, кто дочитал до этого места.

Были использованы материалы из следующих источников:

IBM Redbook «Introduction to SAN and System Networking»

EMC «Network Storage Concepts and Protocols»

Brocade «SAN Fabric Administration Best Practices Guide»

ISCSI-SAN

iSCSI строится на двух наиболее часто используемых протоколах:

SCSI — протоколе обмена блоками данных между компьютером и хранилищем

IP — сетевом транспортном протоколе, широко применяемом в корпоративных сетях Ethernet.

iSCSI — это низкобюджетное решение для внедрения. Администрирование таких инсталляций очень простое, хотя для обеспечения отказоустойчивости необходимо строить выделенную сеть для iSCSI, что приближает нас к сетевой реализации, очень похожей на FC SAN.

Считается, что iSCSI 10Гбит обеспечивает такое же количество IOps и пропускную способность, как и сопоставимый ему FC 8Гбит, но это не совсем так. Хотя пропускная способность iSCSI и выше, но его эффективность ниже, чем у FC за счёт дополнительных накладных расходов.

Производительность iSCSI зависит от существующий инфраструктуры Ethernet (на сегодняшний день минимально рекомендованная сеть для iSCSI – 10Гбит). В ближайшем будущем (по данным Gartner, 10–12 месяцев) стоит планировать переход на 25/40/50GbE, если будет необходимость использовать высокопроизводительные all-flash СХД.

In part one of our series, we discussed today's background of storage area network (SAN) technology choices as well as the features of Fibre Channel (FC) SANs. In this part of our series, we will discuss the features of an iSCSI storage area network and why that might be suitable for your small- to medium-sized business (SMB) data storage environment.

An iSCSI SAN utilizes layer-3 TCP/IP Ethernet switching. It maps the storage SCSI protocol to TCP/IP and wraps it in Ethernet packets. Essentially, iSCSI allows an initiator and a target to negotiate and exchange SCSI commands using well-known and understood TCP/IP networks. The iSCSI SAN emulates direct connected SCSI targets over a local area network.

Just like Fibre Channel, there are three primary iSCSI SAN hardware components. These include iSCSI storage NICs (dedicated network interface cards for iSCSI communications) or initiators; standard Ethernet network switches; and target storage NICs or targets on the external storage arrays.

iSCSI SANs are functionally similar to Fibre Channel, but instead of Fibre Channel's manually intensive deterministic layer-2 protocol (meaning all connections must be manually determined and mapped out ahead of time), iSCSI piggybacks on TCP/IP non-deterministic automated discovery, routing and switching.

Addresses of the hosts and target storage ports are based on TCP/IP. There are three types of iSCSI special name formats to identify initiators and targets. These include the iSCSI qualified name (IQN), extended unique identifier (EUI), and T11 network address authority (NASA). The switches are the same as Ethernet network TCP/IP switches.

To make it a bit more confusing, the iSCSI initiator and target ports come in three flavors as well. First, the most common implementation option utilizes the software initiator where all of the protocol handling is managed by the host CPU. The second option is the split TCP/IP stack-offload where approximately 80% of the standard TCP/IP packet handling is performed in the storage NIC. Only exception handling is performed by the host TCP/IP stack. The third and final option is the full TCP/IP stack-offload, where all TCP/IP protocol handling is performed by the NIC. As CPU cores have proliferated, the requirement for either NIC-based TCP/IP offload engine (TOE) has declined precipitously. The standard software iSCSI initiator is by far and away the most common iSCSI implementation.

There is no direct iSCSI equivalent for a Fibre Channel director class switches even though there are plenty of high port count switches. This is because layer-3 networks do not require the non-blocking of deterministic networks.

Port bandwidth within an iSCSI SAN is available in 1 Gbps and 10 Gbps variants that are not interoperable. However, since iSCSI leverages TCP/IP on Ethernet, multiple ports can be aggregated to provide virtual bandwidths that are much greater than that of an individual port.

In the next part of our series on Fibre Channel vs. iSCSI SANs, we'll look a comparison of Fibre Channel vs. iSCSI SANs.