WHITE PAPER

ПОДКЛЮЧЕНИЕ РЕСУРСОВ К СЕРВЕРНОЙ ОС: IP SAN

О WHITE PAPER

Данный документ описывает архитектурные принципы и подходы, а также рекомендации по их применению при построении корпоративных решений на основе продуктов производства компании YADRO.

Документ предназначен для широкой аудитории, но также может включать материалы, требующие специальной профессиональной подготовки, ориентированные на системных архитекторов, администраторов, сервисных инженеров и других IT-специалистов, участвующих в планировании, проектировании и внедрении комплексных решений.

! Представленная информация основана на результатах исследований, а также экспертизе специалистов YADRO. Документ носит не предписывающий, а информационный характер, целью которого является информирование получателей документа о предмете произведенного исследования и о его результатах. При этом конкретные результаты и эффективность конечных решений зависят от множества факторов, которые могут различаться в зависимости от особенностей инфраструктуры и условий эксплуатации. Ни при каких обстоятельствах YADRO не несет ответственность за последствия применения получателем документа рекомендаций, инструкций, подходов и решений, указанных в документе.

Запрещается копировать, воспроизводить, изменять, публиковать, передавать, распространять, продавать информацию (часть информации), содержащуюся в документе, или создавать производные работы на основе содержания документа без предварительного письменного разрешения YADRO.

Любая информация или утверждения относительно возможностей, емкости, производительности или других характеристик продукции, указанной в документе, предоставляются исключительно на условиях «как есть» и приведена исключительно в целях описания произведенного исследования. YADRO оставляет за собой право вносить изменения/обновлять информация в документе без уведомления.

«YADRO» является зарегистрированным товарным знаком ООО «КНС ГРУПП». Другие названия продуктов и компаний, упомянутые здесь, могут быть товарными знаками или торговыми наименованиями их соответствующих владельцев.

Актуальные версии документации публикуются на информационном ресурсе YADRO по адресу service.yadro.com.

ОГЛАВЛЕНИЕ

1. ОБ ЭКСПЕРТАХ 4
2. ВВЕДЕНИЕ 5
3. ПРОТОКОЛ ISCSI 6
   - 3.1. Компоненты iSCSI 6
   - 3.2. Способы подключения iSCSI хоста 6
   - 3.3. Топологии подключений iSCSI 7
   - 3.4. Прямое iSCSI-подключение 7
   - 3.5. Подключение через iSCSI шлюз 7
   - 3.6. Сочетание FC- и прямого iSCSI-подключения 8
   - 3.7. Стек протоколов iSCSI 8
   - 3.8. Блоки данных iSCSI протокола — PDU 9
   - 3.9. Обнаружение устройств в iSCSI 10
   - 3.10. iSCSI-имена 11
   - 3.11. Сеанс связи iSCSI 12
   - 3.12. Выстраивание командных последовательностей iSCSI 13
4. ПРАКТИЧЕСКИЕ АСПЕКТЫ ПРОЕКТИРОВАНИЯ IP SAN СЕТЕЙ 14

ОБ ЭКСПЕРТАХ

Дмитрий Боровской — специалист с более чем 30-летним опытом работы в сфере ИТ, включая работу с системами хранения данных различных вендоров (EMC, Dell, Hitachi) и отечественных производителей (YADRO). В YADRO занимает должность ведущего специалиста отдела Мультивендорной Сервисной Экспертизы. Специализируется на администрировании Linux, конфигурировании и поддержке систем хранения, а также интеграции продуктов TATLIN в инфраструктуру заказчиков. Эксперт в области сетей хранения и передачи данных (SAN, IPSAN).

ВВЕДЕНИЕ

Традиционные сети хранения данных (SAN) обеспечивают блочный доступ к данным через оптоволоконные соединения, демонстрируя высокую производительность и хорошую масштабируемость. Однако для реализации этих преимуществ FC SAN требуется дополнительное оборудование, включая HBA-адаптеры и коммутаторы Fibre Channel. Многие организации уже обладают IP-инфраструктурой, пригодной для использования в сетях хранения. Современные технологии позволяют передавать данные хранилищ по IP-сетям, что получило название IP SAN. IP как зрелая и стабильная технология предлагает преимущества при использовании для построения сетей хранения.

Передача блоков данных через IP позволяет использовать существующую сетевую инфраструктуру, снижая затраты на развертывание новой SAN. Кроме того, IP-сети обладают проверенными механизмами информационной безопасности. IP-сети давно поддерживают технологии аварийного восстановления (disaster recovery, DR) для географически распределенных систем. Технология IP SAN позволила организациям расширить географические границы своих СХД.

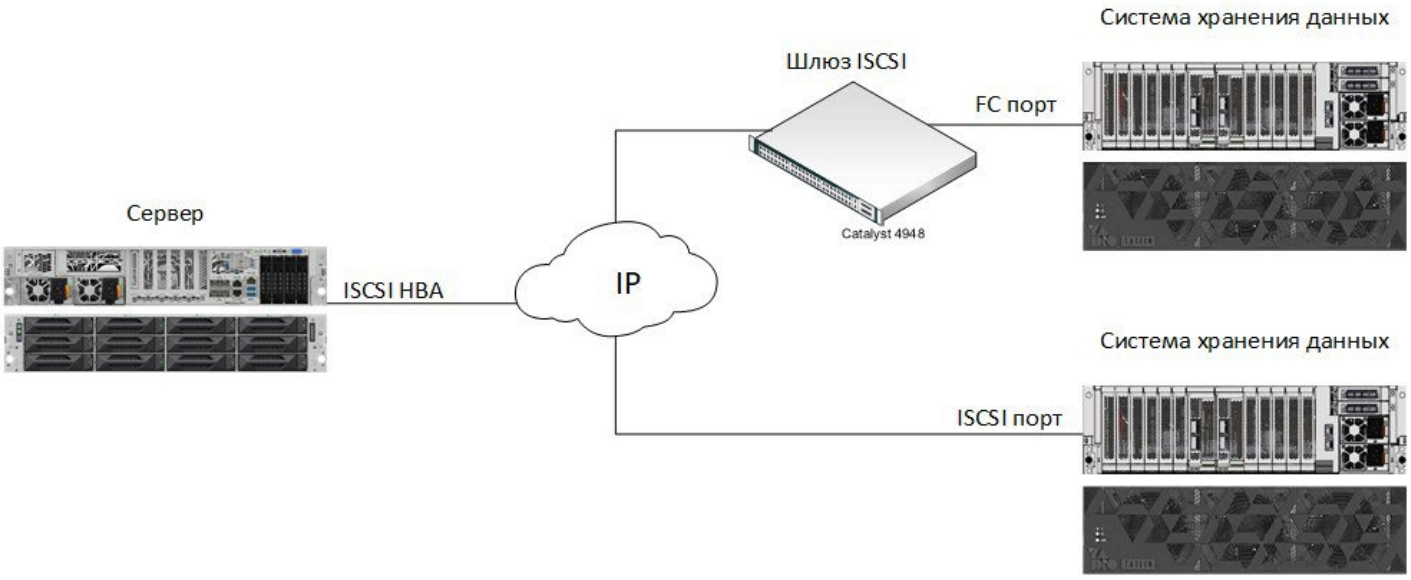
Для передачи данных по IP используются два основных протокола: Internet SCSI (iSCSI) и Fibre Channel over IP (FCIP). iSCSI обеспечивает передачу блоков SCSI через IP-сети. FCIP инкапсулирует трафик Fibre Channel в IP-пакеты для передачи по IP-сети с последующей деинкапсуляцией и передачей в FC SAN на принимающей стороне.

FCIP применяется для соединения распределенных SAN через IP-сети (WAN/MAN), что позволяет создавать географически распределенные СХД на основе существующей IP-инфраструктуры без необходимости в выделенных волоконно-оптических каналах. Кроме того, FCIP обеспечивает маршрутизацию между виртуальными сетями хранения (Inter-VSAN Routing, IVR) и реализует службы маршрутизации Fibre Channel (FCRS), позволяя направлять трафик между конкретными узлами без физического объединения коммутационных систем.

	ВЕРСИИ ИСПОЛЬЗУЕМОГО ПО
TATLIN.UNIFIED Gen1/Gen2	2.7 – 3.1.x
Linux Debian based	10 – 12
Linux RedHat based	7 – 9
FabricOS	8.1.x – 9.2x

ПРОТОКОЛ ISCSI

iSCSI представляет собой IP-протокол, обеспечивающий подключение и управление соединениями между серверами и системами хранения данных (СХД) через IP-сети. Протокол iSCSI инкапсулирует SCSI-команды и данные в IP-пакеты для передачи по TCP/IP. iSCSI активно используется для подключения серверов к СХД благодаря своей экономичности и простоте развертывания, особенно при отсутствии инфраструктуры FC SAN.



Компоненты iSCSI

Ключевыми компонентами iSCSI являются:

- инициатор (сервер);
- таргет — целевое устройство (СХД или iSCSI-шлюз).
- IP-сеть.

При использовании дискового массива с поддержкой iSCSI, хост-инициатор может напрямую взаимодействовать с СХД через IP-сеть. При подключении к существующему массиву Fibre Channel требуется iSCSI-шлюз. iSCSI-шлюз преобразует IP-пакеты в кадры Fibre Channel и обратно, обеспечивая взаимодействие между IP- и FC-средами.

Способы подключения iSCSI хоста

Подключение по iSCSI может быть выполнено тремя способами:

- Стандартная сетевая карта (NIC) с программным iSCSI-инициатором.
- Сетевой адаптер с TCP Offload Engine (TOE) и программным iSCSI-инициатором.
- Аппаратный адаптер iSCSI (iSCSI HBA).

NIC представляют собой наиболее простое и экономичное решение. Большинство серверов оснащены одной или несколькими встроенными NIC, что требует только установки программного iSCSI-инициатора. Так как NIC обеспечивают только базовые IP-функции, обработка протокола iSCSI (инкапсуляция/декапсуляция SCSI) выполняется центральным процессором, создавая дополнительную нагрузку на систему. При интенсивных операциях ввода-вывода процессор может стать узким местом. Адаптеры с TOE снижают нагрузку на процессор, беря на себя обработку TCP, оставляя CPU только функции iSCSI. Данные передаются через TOE-адаптер на СХД по TCP/IP. Однако обработка iSCSI остается задачей программного инициатора, продолжая использовать ресурсы процессора.

iSCSI HBA обеспечивают максимальную производительность, полностью разгружая процессор от обработки стека протоколов iSCSI и TCP/IP. Они также упрощают загрузку сервера с IP SAN. В отсутствие iSCSI HBA требуются изменения в ОС, так как сетевая карта должна получить IP-адрес до загрузки системы. По функциональности iSCSI HBA аналогичны FC HBA.

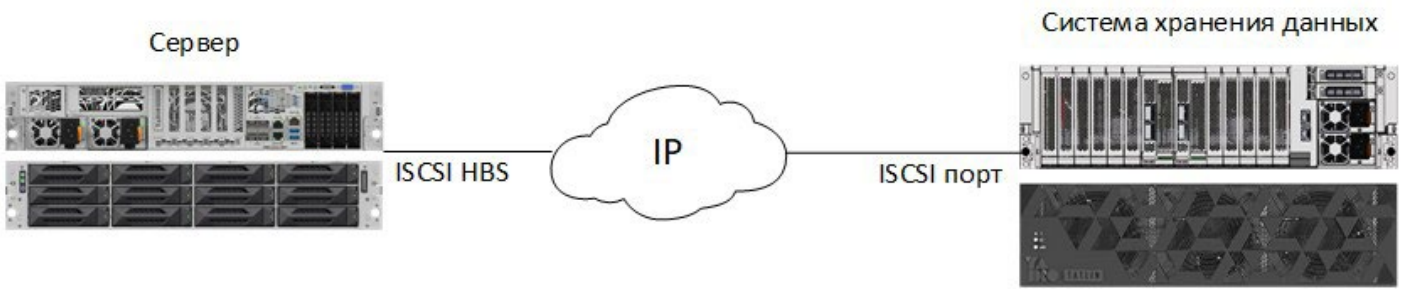
Топологии подключений iSCSI

Топологии подключения iSCSI подразделяются на два основных типа: прямое соединение и подключение через шлюз.

Прямое подключение не требует компонентов Fibre Channel. В такой топологии инициаторы соединяются с целевыми устройствами напрямую или через стандартную IP-сеть.

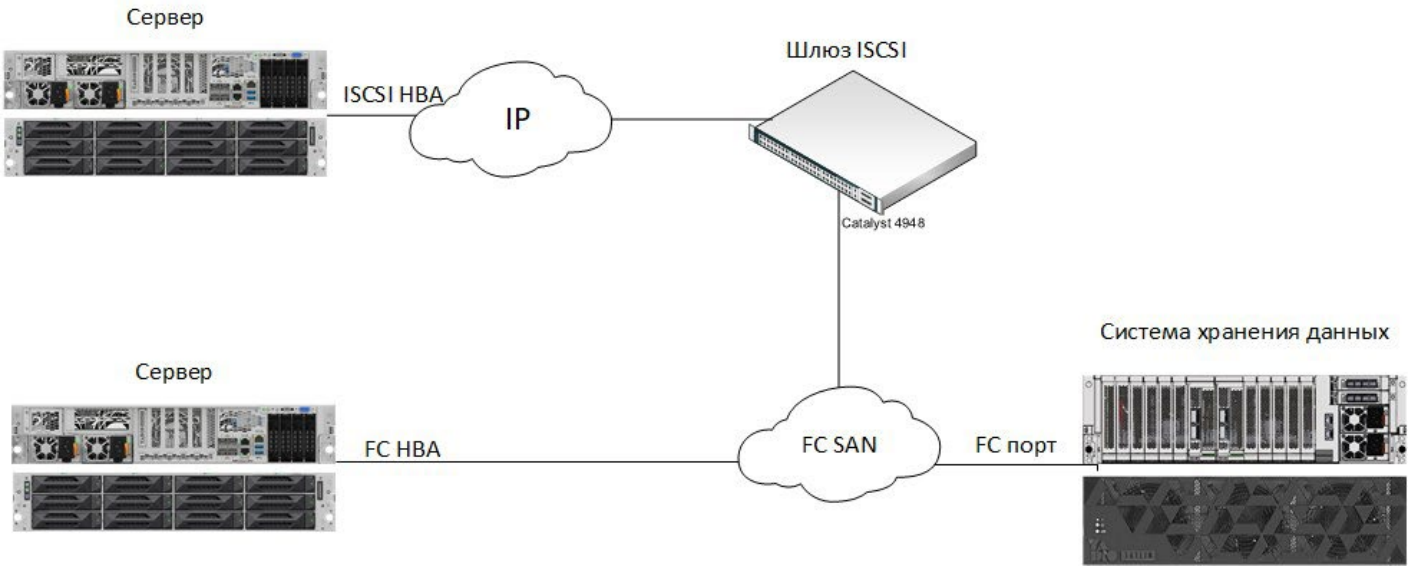
Подключение через шлюз обеспечивает взаимодействие между Fibre Channel и IP-средами, выполняя функции моста между iSCSI и FC. В такой конфигурации инициаторы работают в IP-сети, а СХД остаются в среде Fibre Channel.

Прямое iSCSI-подключение



При использовании массива с поддержкой iSCSI компоненты Fibre Channel не требуются. Массив оснащен одним или несколькими iSCSI-портами, каждый из которых имеет IP-адрес и подключен к стандартному Ethernet-коммутатору. После регистрации инициатора в СХД он получает доступ к доступным логическим устройствам (LUN, Logical Unit Number). Один порт массива способен обслуживать несколько хост-систем.

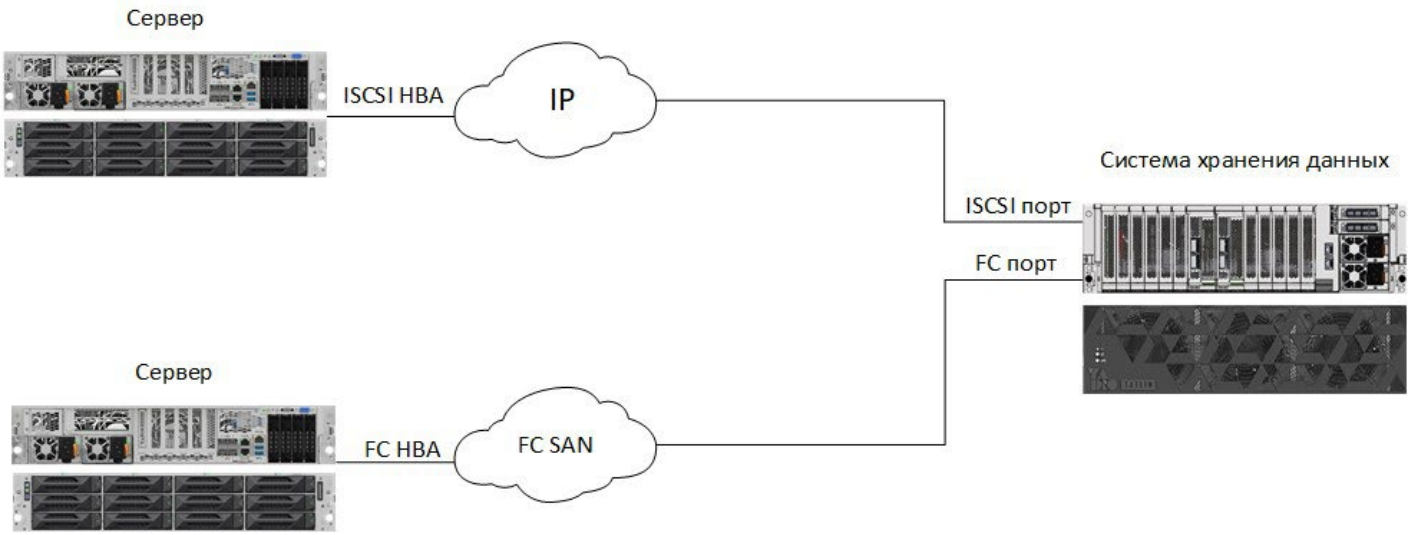
Подключение через iSCSI шлюз



Если массив не имеет встроенных iSCSI-портов, требуется внешнее устройство (шлюз или многопротокольный маршрутизатор) для соединения iSCSI-серверов с СХД Fibre Channel. Шлюз выполняет преобразование между IP-пакетами и кадрами Fibre Channel, а также оснащен портами обоих типов (Ethernet и Fibre Channel) для взаимодействия между разными средами.

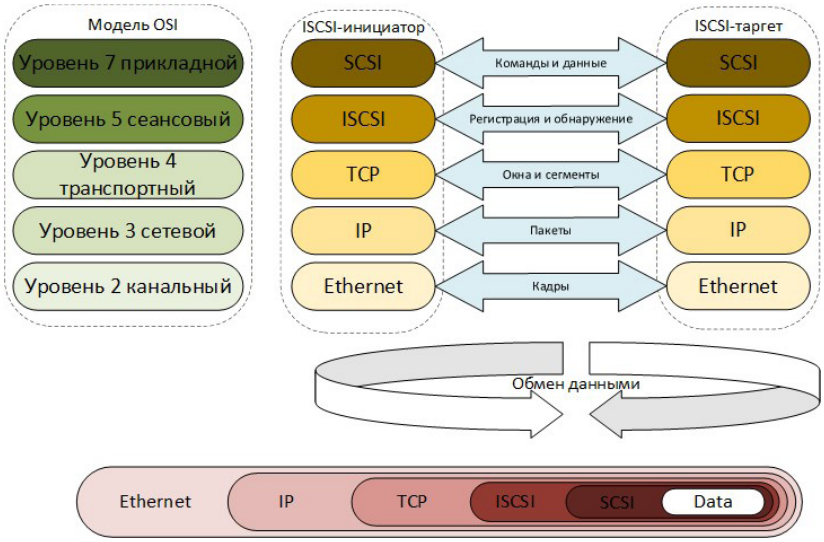
В такой конфигурации iSCSI-инициатор взаимодействует с IP-адресом шлюза (выступающего в роли таргета), тогда как сам шлюз функционирует как инициатор Fibre Channel для СХД.

Сочетание FC- и прямого iSCSI-подключения



Широкое применение получила гибридная топология, объединяющая подключения Fibre Channel и iSCSI. Многие СХД оснащаются портами как Fibre Channel, так и iSCSI, обеспечивая возможность одновременного использования обоих типов подключения к единому массиву.

Стек протоколов iSCSI



Стек протоколов iSCSI определяет уровни инкапсуляции SCSI-команд для их передачи через физическую сеть.

SCSI — протокол прикладного уровня модели OSI (Open Systems Interconnection). Взаимодействие между инициаторами и целевыми устройствами (таргетами) осуществляется посредством SCSI-команд и ответов. Блоки дескрипторов команд SCSI (CDB), данные и статусные сообщения инкапсулируются в TCP/IP-пакеты для передачи между инициаторами и целевыми устройствами.

iSCSI — протокол сеансового уровня модели OSI, устанавливающий надежные соединения между устройствами, поддерживающими SCSI и TCP/IP. На сеансовом уровне iSCSI реализует функции регистрации, аутентификации, обнаружения целевых устройств и управления соединением. На транспортном уровне iSCSI использует TCP для гарантированной доставки данных.

TCP обеспечивает управление потоком, сегментацию данных, контроль ошибок и механизмы повторной передачи. Сетевой уровень модели OSI отвечает за глобальную адресацию и установку соединений. Протоколы канального уровня (уровень 2) модели OSI обеспечивают обмен данными между узлами в физической сети.

Блоки данных iSCSI протокола — PDU



PDU -блок ISCSI, встроенный в пакет IP

Блоки данных протокола iSCSI (Protocol Data Unit, PDU) служат базовой единицей передачи информации в среде iSCSI. Взаимодействие между iSCSI-инициаторами и целевыми устройствами осуществляется посредством PDU.

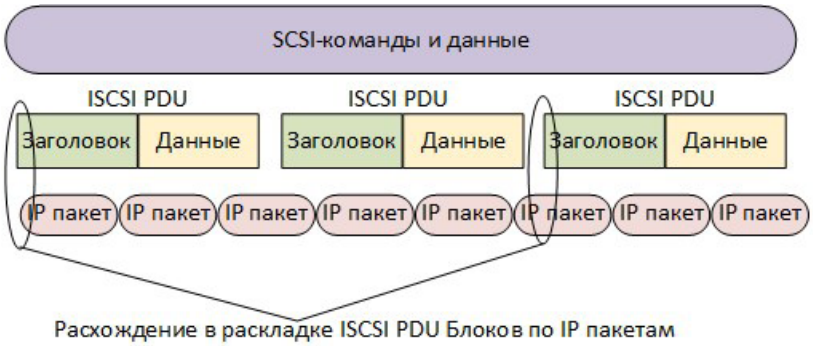
PDU используются для:

- Установления подключений и сеансов iSCSI.
- Обнаружения устройств.
- Передачи SCSI-команд и данных.
- Получения статусной информации.

PDU iSCSI состоят из одного или нескольких заголовочных сегментов, за которыми могут следовать сегменты данных (опционально). PDU инкапсулируются в IP-пакеты для передачи по сети.

Особенности передачи:

- Сообщения сегментируются на пакеты.
- Пакеты могут передаваться разными маршрутами.
- Порядок получения может отличаться от исходного.
- IP обеспечивает доставку, TCP — правильную последовательность.



Расхождение в раскладке ISCSI PDU Блоков по IP пакетам

Раскладка PDU блоков ISCSI по IP пакетам

Таргет извлекает SCSI-команды и данные, используя информацию из заголовков iSCSI. Соотношение 1:1 между PDU iSCSI и IP-пакетами не является обязательным. В зависимости от размера, один PDU может занимать целый IP-пакет или совместно использоваться с другими PDU в одном пакете.

Для достижения соотношения 1:1 между IP-пакетами и PDU выполняется настройка MTU (Maximum Transmission Unit), что исключает фрагментацию IP-пакетов и повышает эффективность передачи.

Обнаружение устройств в iSCSI

Перед установкой сеанса связи iSCSI-инициатору необходимо определить расположение и идентификаторы доступных целевых устройств (таргетов) в сети.

Основные способы обнаружения:

- метод SendTargets;
- служба iSNS (Internet Storage Name Service).

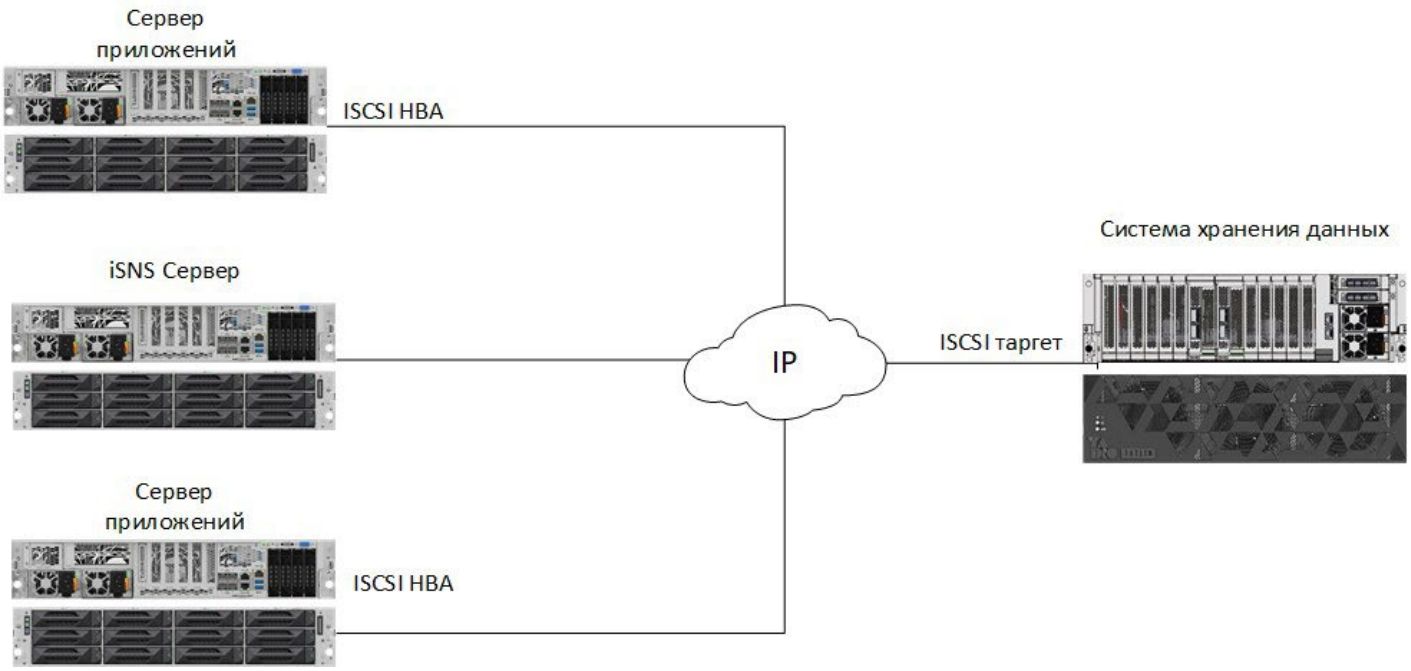
Особенности метода SendTargets:

- Адрес iSCSI-портала задается вручную
- Инициатор отправляет команду SendTargets
- Портал возвращает идентификаторы и адреса доступных целевых устройств

Особенности службы iSNS:

- iSCSI-устройства автоматически регистрируются на iSNS-сервере
- Инициатор запрашивает у сервера список доступных целевых устройств
- Не требует ручной настройки адресов

Альтернативный метод, имеющий ограниченное применение — Service Location Protocol (SLP).



iSCSI-имена

Для идентификации инициаторов и целевых устройств в iSCSI-сетях используются уникальные глобальные идентификаторы — iSCSI-имена.
Идентификатор может включать: название отдела, приложения, производителя, серийный номер или другие уникальные метки для управления устройствами.

Основные форматы iSCSI-имен:

- iSCSI Qualified Name (IQN) — полное iSCSI-имя;
- Extended Unique Identifier (EUI) — расширенный уникальный идентификатор.

Общие правила для основных форматов:

- Допустимые символы: латинские буквы, цифры, точка (.), дефис (-), пробел ( ).
- Длина: до 223 байт (UTF-8 кодировка).
- Регистр букв не учитывается.

Дополнительный специализированный формат iSCSI-имен: NAA (Network Address Authority) — универсальный сетевой идентификатор.

Формат IQN
Структура: iqn.<дата>-<месяц>.<домен>:<дополнительный_идентификатор>.
Пример: iqn.2023-12.com.example:storage-array-1.

Дополнительный идентификатор может содержать:

- серийный номер;
- идентификатор ресурса;
- другие уникальные метки устройства.

Особенности:

- Организация должна владеть зарегистрированным доменным именем (не обязательно активным).
- Доменное имя должно быть уникальным и защищенным от использования другими организациями.
- Включает дату для предотвращения конфликтов.
- Позволяет создавать осмысленные идентификаторы.

Формат EUI
Структура: eui.<16-значный_шестнадцатеричный_код>.
Пример: eui.0300732A32598D26.

Особенности:

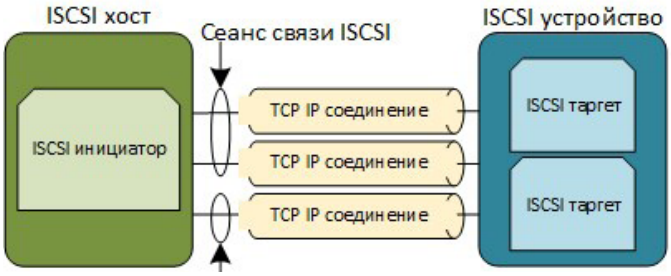
- Основан на стандарте IEEE EUI-64.
- Обеспечивает глобальную уникальность.
- Часто используется для аппаратных идентификаторов.

Формат NAA
Структура: Зависит от подтипа (NAA 1-6).
Пример (naa.<hex_код>): naa.6001405DCBA32598.

Особенности:

- Определен стандартами T11/INCITS.
- Универсальный формат для устройств с разными интерфейсами.
- Особенно полезен для гибридных систем (iSCSI+SAS/FC).

Сеанс связи iSCSI



Сеанс iSCSI представляет собой соединение между инициатором и целевым устройством (таргетом), идентифицируемое уникальным идентификатором сеанса (SSID). SSID формируется из идентификатора инициатора (ISID) и идентификатора таргета (TSID).

Сеанс iSCSI выполняет две основные функции:

- Обнаружение доступных таргетов и определение их сетевого расположения.
- Передача данных между инициатором и таргетом (стандартные операции iSCSI).

Установление сеанса происходит через процесс регистрации iSCSI, который начинается с установки TCP-соединения между инициатором и таргетом. Соединение может использовать стандартный порт 3260 или специально указанный порт таргета.

В процессе регистрации выполняются два ключевых действия:

- Взаимная аутентификация инициатора и таргета.
- Согласование параметров соединения.

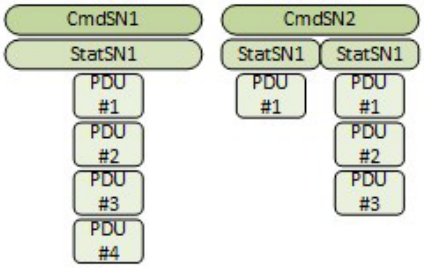
После успешной регистрации сеанс переходит в рабочую фазу, где:

- Выполняются стандартные SCSI-транзакции.
- Инициатор отправляет команды и данные к LUN таргета.
- Данные инкапсулируются в PDU iSCSI.
- Передача осуществляется через установленное TCP-соединение.

Завершение сеанса включает два этапа:

1. Инициация окончания:
   - Обычно выполняется инициатором.
   - Таргет может инициировать при возникновении ошибки.
2. Прекращение обмена:
   - После подтверждения окончания сеанса.
   - Все передачи данных прекращаются.

Выстраивание командных последовательностей iSCSI



Нумерация последовательности команд и состояний.

Протокол iSCSI использует механизм последовательностей "запрос-ответ" для взаимодействия между инициаторами и таргетами. Каждая такая последовательность может быть разбита на несколько PDU-блоков. Для обеспечения упорядоченной доставки команд применяется номер последовательности команд (CmdSN, Command Sequence Number), который:

- Нумерует все PDU-блоки команды.
- Гарантирует сохранение порядка команд.
- Работает независимо от используемого TCP-соединения.

Особенности работы CmdSN:

- Начинается с команды регистрации.
- Инкрементируется на 1 для каждой новой команды.
- Контроль порядка выполнения осуществляется на стороне таргета.

Для ответов применяется аналогичный механизм — номер последовательности состояния (StatSN, Status Sequence Number). StatSN обеспечивает порядковую нумерацию статусных сообщений.

Механизм передачи данных включает:

- Request-to-Transfer (R2T) — запросы на передачу данных.
- Номер последовательности данных (DataSN, Data Sequence Number) — обеспечивает порядок данных в команде.

Технические характеристики счетчиков:

- DataSN и R2TSN - 32-битные беззнаковые счетчики.
- Определяются протоколом iSCSI.
- Хранятся локально на каждом узле.

Правила нумерации DataSN:

1. Для операций чтения:
   - Начальное значение: 0.
   - Инкремент: +1 для каждого PDU.
2. Для операций записи:
   - Начальное значение: 0 (для первого PDU).
   - Инкремент: +1 для каждого последующего PDU.

Правила работы R2TSN:

- Начальное значение: 0.
- Инкрементируется на 1 для каждого нового R2T-запроса.
- Применяется в рамках одной команды.

ПРАКТИЧЕСКИЕ АСПЕКТЫ ПРОЕКТИРОВАНИЯ IP SAN СЕТЕЙ

IP SAN набирают популярность благодаря следующим преимуществам:

- Скорость 10/25 Гбит/с сопоставима с FC 16 Гбит/с при меньших затратах.
- Использование существующей Ethernet-инфраструктуры.
- Гибкость и широкие возможности мониторинга/администрирования.

Принципы проектирования IP SAN аналогичны требованиям для FC SAN.

Требования к подключению:

- Каждый хост должен иметь доступ по минимум двум путям до СХД.
- Пути должны проходить через разные подсети.

Оптимизация производительности:

1. Минимальный уровень: приоритезация iSCSI-трафика.
2. Рекомендуемый уровень: выделенные VLAN для iSCSI.
3. Для высоконагруженных систем:
   - Физически изолированная Ethernet-сеть.
   - Выделенные сетевые адаптеры под iSCSI.

Стандартные практики отказоустойчивости для Ethernet полностью применимы к IP SAN.

Особенности iSCSI:

- Не использовать агрегацию каналов.
- Обеспечивать отказоустойчивость через multipathing-решения.

Настройки MTU:

- Рекомендуемое значение: 9000 байт.
- Должно быть единообразно на всех компонентах сети.

Ограничения по переподписке (over-subscription):

- Следует избегать высокой степени переподписки.
- Риск возникновения узких мест при продолжительной нагрузке.