

РЕКОМЕНДА ЦИИ ПО ОПТИМИЗА ЦИИ IP- СЕТЕЙ ХРАНЕНИЯ

**Массивы хранения
PowerVault MD3000i**

www.dell.com/MD3000i



СОДЕРЖАНИЕ

Содержание

INTRODUCTION	ERROR! BOOKMARK NOT DEFINED.3
OVERVIEW ISCSI	ERROR! BOOKMARK NOT DEFINED.3
IP SAN DESIGN	4
BEST PRACTICE - IMPLEMENTATION	5
REDUNDANCY	5
SECURITY	6
IP SAN NETWORK CONFIGURATION	7
IP SAN OPTIMIZATION	8
SUMMARY	11

ВВЕДЕНИЕ

Настоящий документ является руководством по оптимизации IP-сред сетевого хранения с помощью массива хранения Dell MD3000i. Приведенные ниже сведения являются рекомендациями, которые призваны обеспечить построение отказоустойчивой, высокопроизводительной среды для максимальной реализации потенциала сетей хранения данных на базе массивов MD3000i. Практическое применение данных рекомендаций зависит от требований среды, в которой установлены массивы хранения; в некоторых случаях часть приводимых рекомендаций может быть неприменима. Настоящие рекомендации рассчитаны прежде всего на оптимизацию решений на базе технологий корпорации Dell.

ИНТЕРФЕЙС iSCSI: ОБЗОР

Протокол iSCSI, предназначенный для управления данными на уровне блоков, позволяет создавать сети хранения на основе стандарта Ethernet. При этом Ethernet выступает в роли транспортного протокола для передачи данных от серверов к устройствам или к сетям хранения данных. Поскольку в основе iSCSI лежит стандарт Ethernet, ограничения по сложности и удаленности затрагивают его в меньшей степени, чем другие протоколы хранения.

Благодаря iSCSI стандартные команды SCSI могут передаваться через сети Ethernet по протоколу TCP. Любая сеть хранения iSCSI состоит из серверов (с iSCSI-совместимым адаптером главной шины или сетевой платой), дисковых массивов и ленточных библиотек. В отличие от других технологий сетевого хранения, протокол iSCSI подразумевает применение стандартных коммутаторов, маршрутизаторов и кабелей Ethernet. При этом весь сетевой трафик передается по тому же самому транспортному протоколу Ethernet, что и в локальных сетях (TCP/IP). Это позволяет использовать одни и те же типы кабелей, маршрутизаторов и коммутаторов как в локальных сетях, так и в сетях хранения.

Поскольку в iSCSI используются команды SCSI, а сеть Ethernet используется только для их передачи, операционные системы взаимодействуют с iSCSI-устройствами, как с SCSI-устройствами. Фактическое местоположение устройств и масштаб сети роли не играют.

iSCSI-устройства состоят по большей части из компонентов, которые хорошо знакомы специалистам по сетевым технологиям. Они практически ничем не отличаются от обычных контроллеров RAID-массивов и накопителей SCSI или FC-интерфейса. Единственная разница заключается в поддержке протокола iSCSI, для чего достаточно

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

стандартных сетевых плат с соответствующим ПО либо специализированных микросхем с поддержкой iSCSI или сетевых адаптеров с механизмом разгрузки TCP/IP и iSCSI.

Протокол iSCSI создан на основе двух наиболее распространенных протоколов: SCSI для взаимодействия с устройствами хранения и TCP для сетевых подключений. За долгие годы исследований, доработок и интеграции обе технологии были значительно усовершенствованы. Кроме того, IP-сети традиционно отличаются высокой управляемостью, совместимостью и экономичностью.

ПРОЕКТИРОВАНИЕ IP-СЕТЕЙ ХРАНЕНИЯ

Любая инфраструктура сетевого хранения на базе IP включает в себя один или несколько сетевых коммутаторов или аналогичных сетевых устройств (например, маршрутизаторов). В настоящем документе предполагается, что в сети имеется не менее одного устройства коммутации или маршрутизации. Варианты прямого (не через сеть) подключения массивов хранения MD3000i к управляющим серверам хотя и возможны, но в данном документе не рассматривается. Таким образом, IP-сеть хранения должна состоять из одного или нескольких серверов, которые через IP-сеть подключены к одному или нескольким массивам хранения; при этом в сетевой инфраструктуре должно находиться не менее одного коммутатора.

При проектировании IP-сетей хранения необходимо учитывать несколько факторов, важность которых зависит от конкретного варианта реализации сети. Из всех этих факторов следует выделить следующие.

1. Резервирование. Если доступность данных должна поддерживаться при любых условиях, рекомендуется выбрать отказоустойчивую конфигурацию IP-сети хранения.
2. Безопасность. В зависимости от конкретных требований к IP-сети хранения данных при ее создании могут применяться различные механизмы защиты, например выделенные сети, протокол SHAR, запрос пароля при доступе к массиву и пр.
3. Сетевая инфраструктура. Производительность и обслуживаемость IP-сети хранения в значительной мере зависят от таких компонентов, как сетевые платы, адаптеры главной шины, коммутаторы, кабели, маршрутизаторы и пр.
4. Оптимизация. В зависимости от сферы применения повысить производительность IP-сети хранения можно с помощью «доводки» ее отдельных элементов. К примеру, этого можно добиться, обеспечив поддержку различных аппаратных механизмов разгрузки, расширенных пакетов Jumbo Frame и т.д.

РЕКОМЕНДАЦИИ: ВНЕДРЕНИЕ

В зависимости от потребностей, доступных ресурсов и предполагаемой сферы применения внедрение IP-сети хранения может выполняться различными методами. Так, одним из главных, но зачастую упускаемых из виду способов улучшения управляемости сети является наличие единой и понятной системы наименования массивов хранения. Это особенно важно в сетях с несколькими массивами. Для правильной идентификации каждого физического массива можно использовать функцию «blink array», доступную в ПО MD Storage Manager.

Ниже представлены общие рекомендации по внедрению. При этом следует учитывать, что они носят общий характер и, возможно, требуются не во всех случаях.

Резервирование

Обычно под резервированием понимается наличие второго набора оборудования и коммуникационных каналов, используемых в случае сбоя того или иного сетевого устройства. В IP-сети хранения для этого в массив устанавливается дополнительный контроллер, а для бесперебойной передачи iSCSI-данных используется два разных коммутатора. Ниже приведена упрощенная схема резервирования при помощи Dell PowerVault MD3000i; далее следует описание преимуществ такой топологии.

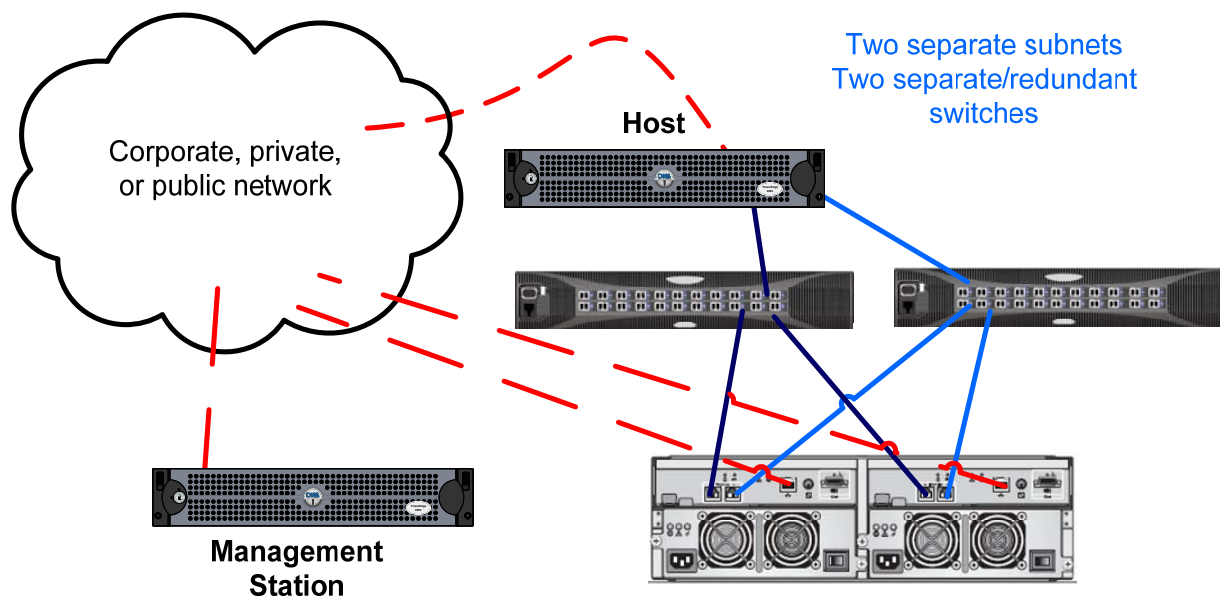


Схема 1. Конфигурация MD3000i с полным резервированием

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

Физическая сетевая инфраструктура. IP-сеть хранения с полным резервированием отличается наличием нескольких независимых физических каналов передачи данных iSCSI между хостом и массивом. Каждый канал должен обеспечиваться с помощью отдельной подсети.

Конфигурация iSCSI. Если целевым iSCSI-устройством является массив MD3000i или его аналог, для каждой подсистемы хранения рекомендуется установить несколько сеансов передачи данных от каждого управляющего сервера. Наиболее надежная конфигурация включает в себя один сеанс передачи от каждого порта каждой сетевой платы к каждому модулю RAID-контроллера. В этом случае при сбое одного канала сеанс будет перезапущен без ущерба для других каналов связи.

RAID. В зависимости от сферы применения необходимо выбрать оптимальный уровень RAID. К применению рекомендуются уровни не ниже RAID 1. Такие конфигурации способны обеспечить некоторый уровень резервирования в качестве страховки от сбоев физических дисков. В зависимости от сферы применения оптимальными оказываются различные уровни RAID. Это следует учитывать при конфигурировании MD3000i.

Электропитание. Каждый резервный канал передачи данных должен быть подключен к отдельному источнику питания. Таким образом, даже в случае сбоя одного из компонентов из-за перебоев в электропитании альтернативный канал продолжит свою работу. Точно так же каждый из двух установленных на MD3000i блоков питания должен получаться к отдельному источнику.

Безопасность

Оптимальный метод защиты данных в IP-сетях хранения состоит в организации изолированной и физически независимой сети для передачи iSCSI-трафика. Помимо преимуществ с точки зрения безопасности, наличие выделенной подсети для передачи трафика систем хранения позволяет избежать перегрузки каналов из-за использования одной сети для передачи данных сразу всех рабочих систем и приложений.

Виртуальная локальная сеть. Если создание физически изолированных сетей iSCSI невозможно, отделить трафик iSCSI от других данных можно при помощи виртуальных локальных сетей (VLAN). При этом рекомендуется активировать функцию маркировки пакетов VLAN, поддерживаемую массивами MD3000i. Через каждый порт смогут проходить либо все маркированные IP-пакеты, либо все IP-пакеты без маркировки.

Примечание. Поддержка виртуальных локальных сетей должна быть включена для всех сетевых плат, коммутаторов и iSCSI-портов, через которые данные поступают в сеть хранения iSCSI. В противном случае работа оборудования может оказаться непредсказуемой. Чтобы упростить поиск и устранение неполадок при начальном

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

развертывании, перед тем как активировать поддержку виртуальных ЛС на всех компонентах существующей инфраструктуры, убедитесь, что все сетевые платы, коммутаторы и массивы хранения MD3000i включены и находятся в рабочем состоянии.

Протокол CHAP. Для защиты взаимного доступа между сервером и массивом хранения на всех серверах и массивах необходимо активировать функцию аутентификации по протоколу CHAP (как взаимную, так и целевого устройства). Достижение максимального уровня безопасности помогает соблюдение стандартных правил по установке пароля CHAP.

Настоятельно рекомендуется установить пароли на всех устройствах в пределах IP-сети хранения. Кроме того, желательно использовать надежные пароли, соответствующие стандартным рекомендациям по защите ИТ-сред.

Инфраструктура IP-сети хранения

Ниже представлены общие рекомендации по внедрению. При этом следует учитывать, что они носят общий характер и, возможно, требуются не во всех случаях.

Общие рекомендации по развертыванию сети. Убедитесь, что используемая категория кабеля совместима с сетями Gigabit Ethernet (CAT5e, CAT6). При проектировании сети рекомендуется минимально сократить число промежуточных участков между массивами и серверами. Это позволит значительно уменьшить количество точек отказа, сократить задержку, упростить управление сетью и сетевую архитектуру (в том числе с точки зрения резервирования). Желательно использовать управляемые коммутаторы, поскольку они обладают расширенными функциями, которые помогают оптимизировать и поддерживать сеть с учетом конкретной сферы применения. Также рекомендуется активировать поддержку автосогласования, так как сети Gigabit Ethernet разрабатывались для использования с этой функцией. Если для какого-то приложения необходима определенная скорость передачи данных или дуплексный режим, следует изменить соответствующие настройки в параметрах объявления о состоянии коммутатора.

Протокол STP. Рекомендуется отключить поддержку протокола STP (Spanning-Tree Protocol) на тех портах коммутатора, к которым подключаются конечные узлы сети (сетевые интерфейсы инициаторов iSCSI и массивов хранения данных). Если все-таки на этих портах необходим протокол STP, то на них требуется включить функцию STP FastPort, которая позволит портам незамедлительно переключаться в режим переадресации. (Примечание. При подключении к сети функция PortFast незамедлительно переводит порт в режим переадресации STP. При этом порт продолжает участвовать в коммуникации по протоколу STP. Это означает, что если порт является частью циклического маршрута, он перейдет в режим блокировки STP.)

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

Примечание. Коммутаторы PowerConnect по умолчанию настроены на поддержку протокола RSTP (Rapid Spanning Tree Protocol), который является усовершенствованной версией STP и обеспечивает более быструю конвергенцию связующего дерева. Предпочтительно использовать протокол RSTP.

Примечание. Для однокабельных соединений между коммутаторами рекомендуется алгоритм связующего дерева, а для многокабельных следует использовать группирование портов.

Алгоритм предотвращения перегрузок TCP. Данный алгоритм представляет собой сквозной протокол управления потоком данных, который ограничивает объем данных, пересылаемых от отправителя к получателю по протоколу TCP. При этом используется технология «скользящего окна», которая позволяет устанавливать размер блоков данных пересылаемых получателю по протоколу TCP. В начале коммуникации данные отправляются малыми сегментами, размер которых постепенно увеличивается, до тех пор пока какой-либо сегмент не будет отброшен. После этого протокол TCP повторяет процесс сначала.

Управление потоком данных Ethernet. Dell рекомендует включить функцию управления потоком данных (Flow Control) на тех портах коммутации, которые назначены для обработки iSCSI-трафика. Кроме того, если на сервере для обработки трафика iSCSI совместно используются программный инициатор iSCSI и сетевая плата, то для достижения максимальной производительности необходимо активировать функцию управления потоком данных на сетевых платах. Во многих случаях в сети может возникнуть дисбаланс по объему трафика между устройствами, передающими данные, и устройствами, которые эти данные принимают. Это особенно характерно для тех конфигураций сетей хранения, в которых с устройствами хранения данных взаимодействует сразу несколько управляющих серверов (инициаторов). При одновременной пересылке данных от отправителей может быть превышена пороговая пропускная способность получателя. В этом случае получатель может отменить прием пакетов, передавая отправителям команду повторить пересылку данных через некоторое время. Это не приводит к потере данных, однако необходимость повторной отправки пакетов может увеличить задержку и снизить пропускную способность интерфейсов ввода-вывода.

Примечание. По умолчанию функция управления потоком данных в коммутаторах PowerConnect отключена.

При ее активации массив MD3000i будет автоматически настроен в соответствии с параметрами коммутатора.

Управление одноадресным «штормовым» трафиком. «Штормовой» трафик возникает в тех случаях, когда увеличение количества исходящих пакетов становится причиной чрезмерного трафика в сети, что снижает ее производительность. В большинстве коммутаторов предусмотрены функции управления штормовым трафиком, которые предотвращают прерывание работы портов из-за возникновения широковещательного,

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

многоадресного или одноадресного штормового трафика на физических интерфейсах. Обычно при достижении трафиком определенного процента от общей загрузки сети (по умолчанию — 80%) эти функции отменяют передачу сетевых пакетов.

Поскольку трафик iSCSI является одноадресным и в большинстве случаев может занимать весь канал, рекомендуется отключить функции управления одноадресным штормовым трафиком на коммутаторах, предназначенных для обработки iSCSI. При этом функции управления ширококестельным и многоадресным штормовым трафиком отключать не следует. Сведения об отключении функций управления одноадресным штормовым трафиком см. в документации к коммутатору.

Пакеты Jumbo Frame. Корпорация Dell рекомендует включить поддержку пакетов Jumbo Frame на тех портах коммутатора, которые предназначены для работы с трафиком iSCSI. Кроме того, если на сервере для обработки трафика iSCSI совместно используются программный инициатор iSCSI и сетевая плата, то для достижения максимальной производительности необходимо активировать поддержку пакетов Jumbo Frame на сетевых платах. Это также позволит избежать дополнительной нагрузки на ЦП и обеспечить согласованное поведение всех компонентов сети.

Примечание. Поддержка расширенных пакетов Jumbo Frame должна быть включена для всех портов сетевых плат, коммутаторов и массивов, через которые данные поступают в сеть хранения iSCSI. В противном случае работа оборудования может оказаться непредсказуемой. Чтобы упростить поиск и устранение неполадок при начальном развертывании, перед тем как активировать поддержку расширенных пакетов Jumbo Frame на всех компонентах существующей инфраструктуры, убедитесь, что все сетевые платы, коммутаторы и массивы хранения MD3000i включены и находятся в рабочем состоянии.

Оптимизация IP-сети хранения

При проектировании IP-сети хранения необходимо учитывать различные особенности сети, а также сферу ее применения. Тем не менее, при этом будет полезным соблюдение нескольких общих правил. Для достижения максимальной пропускной способности в массивах хранения необходимо задействовать все имеющиеся порты передачи данных. Если используемые приложения предъявляют повышенные требования к интерфейсам ввода-вывода, для разгрузки трафика iSCSI рекомендуется использовать сетевые платы. Кроме того, можно выполнить настройку контроллеров-владельцев для виртуальных дисков; при этом ни один из контроллеров не будет вынужден обрабатывать больший объем операций ввода-вывода, чем другой.

Массивы хранения данных MD3000i поддерживают конфигурации контроллеров «активный/активный», в которых оба компонента могут одновременно обрабатывать ввод и вывод данных. Контроллеры имеют ассиметричную конструкцию, то есть контроллер выступает владельцем виртуальных логических устройств (LUN) и доступ к виртуальному

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

диску для операций ввода-вывода возможен только через контроллер-владельца. Чтобы осуществлять доступ для операций ввода-вывода через оба контроллера, виртуальные диски могут быть распределены между ними. Владелец виртуального диска может быть изменен, чтобы сбалансировать доступ для операций ввода-вывода и, следовательно, сбалансировать коэффициент загрузки обоих контроллеров. Если на сервере настроен доступ с резервированием, то при потере доступа к виртуальному диску через контроллер-владельца диск передаст права владельца от одного контроллера к другому и завершит операции ввода-вывода через новый контроллер-владельца.

На следующем рисунке показана организация асимметричной архитектуры «активный/активный» для массивов хранения MD3000i. Эта конфигурация состоит из двух виртуальных дисков (виртуальный диск 0 и виртуальный диск 1), при этом владельцем первого является контроллер 0, а второго — контроллер 1. Виртуальный диск 0 присвоен серверу 1, а виртуальный диск 1 — серверу 2.

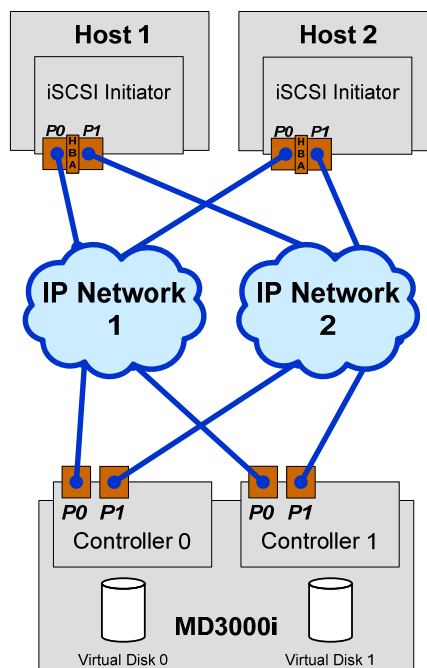


Схема 2. Конфигурация контроллера массива хранения данных MD3000i

Благодаря тому, что в асимметричной архитектуре назначены владельцы виртуальных дисков, сервер 1 получает доступ к виртуальному диску 0 через контроллер 0, а сервер 2 — к виртуальному диску 1 через контроллер 1.

Агрегация полосы пропускания. Массив хранения MD3000i позволяет подключать два порта Ethernet одного сервера к одному контроллеру для агрегации полосы пропускания. Если в драйвере iSCSI MD3000i настроено циклическое обслуживание, массив выполняет агрегацию

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

пакетов, отправляемых на данный контроллер, и помещает их в каждый канал, удваивая тем самым доступную пропускную способность.

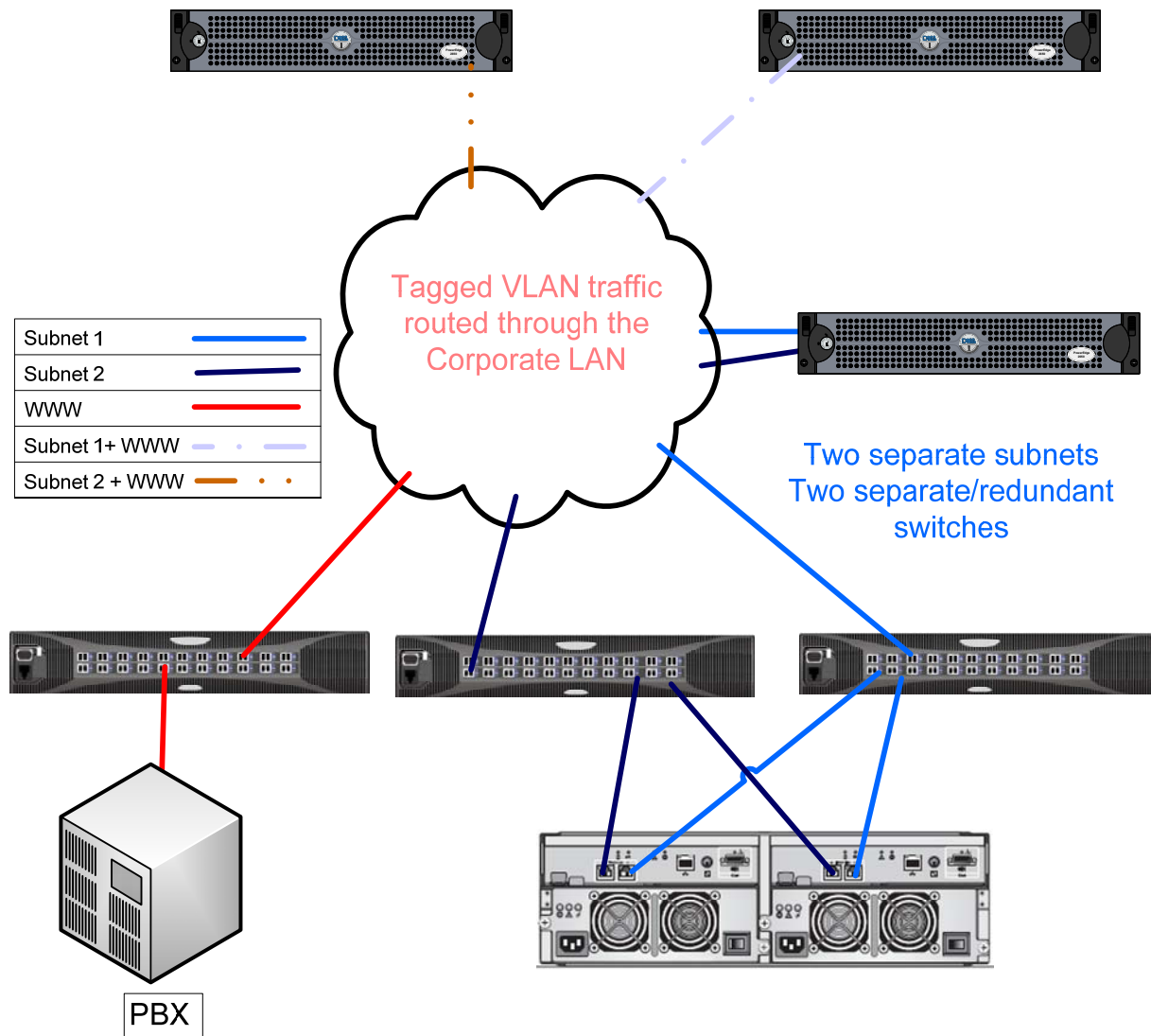


Схема 3. Массив хранения MD3000i в сети

Проверьте архитектуру сети и убедитесь, что на участке между сервером и массивом нет «узких мест». Выше обсуждались некоторые положения относительно безопасности, которые также помогут оптимизировать IP-сеть хранения данных: использование отдельных коммутаторов для физической изоляции iSCSI-трафика и использование виртуальных локальных сетей с включенной функцией FastPort.

Оптимизация сетей на уровне 2. При настройке виртуальной локальной сети можно использовать ее маркировку, которая помогает выполнять маршрутизацию iSCSI-трафика.

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

Затем можно назначить приоритет для трафика, который передается по виртуальной локальной сети, однако для этого необходимо внимательно обследовать весь трафик. Например, если VOIP-трафик проходит по той же виртуальной локальной сети, необходимо убедиться в том, что это не влияет на качество голосовой связи. Также при этом требуется учитывать соотношение общего интернет-трафика по сравнению с трафиком iSCSI и VOIP.

Оптимизация сетей на уровне 3. Надежным способом управления трафиком являются дифференцированные службы (технология DiffServ). В некоторых коммутаторах реализованы проприетарные версии этой технологии, которые имеют наименование «Quality of Service» (QoS), т. е. «качество обслуживания». Технология DiffServ использует протокол DSCP (Differentiated Services Code Point), который дифференцирует уровни обслуживания для каждого IP-подключения. Соглашения об уровне обслуживания реализуются на основании принципа «Per Hop Basis» (PHB). Поэтому потоки трафика в пределах корпоративной сети легко предсказать, однако при отказе одного из каналов глобальной сети соблюдение этих соглашений становится невозможным. Обычно технология DiffServ подразумевает применение четырех уровней.

1. PHB по умолчанию — это, как правило, оптимальный уровень для обработки любого вида трафика.
2. *Expedited Forwarding* (EF) PHB — ускоренная пересылка; применяется для трафика с низким уровнем потерь и задержки.
3. *Assured Forwarding* (AF) — гарантированная пересылка; применяется для указания поведения.
4. *Class Selector* PHB — селекторы классов; при их определении необходимо соблюдать обратную совместимость со значением IP Precedence (Приоритет IP-пакетов).

Чтобы выбрать уровень обслуживания, необходимо изучить требования приложений, которые подключены к конкретному массиву. Например, если для серверов установлена загрузка по протоколу iSCSI, виртуализация применяется для того, чтобы «скрыть» массив, а гостевая ОС загружается с диска C:, который фактически располагается в массиве хранения; этому трафику следует назначить уровень **EF**, поскольку данные должны быть всегда доступны, а задержка может стать причиной зависания сервера. При этом интернет-трафику можно присвоить минимальный класс **AF**, поскольку он не является критически важным.

РЕЗЮМЕ

IP-сеть хранения представляет собой гибкую, простую в развертывании и эксплуатации систему хранения данных для предприятий любого размера. Соблюдение приведенных в этом документе рекомендаций и использование лучших отраслевых наработок позволит создать надежную и гибкую систему хранения данных. Необходимо помнить, что при

РЕКОМЕНДАЦИИ ПО ОПТИМИЗАЦИИ IP-СЕТЕЙ ХРАНЕНИЯ

проектировании и построении корпоративной сети следует учитывать требования IP-сети хранения, поскольку объем данных предприятия будет расти и вместе с ним будет увеличиваться сетевой трафик. Соблюдение рекомендаций, приведенных в этом документе, позволит заблаговременно подготовиться к этим изменениям.