

# AQUARIUS

## Построение сетевых фабрик VXLAN/EVPN с использованием коммутаторов Aquarius

### Валидированный дизайн и внедрение

#### История изменений

Версия	Дата публикации	Описание
1.0	10.10.2024	Первоначальная версия
1.01	29.10.2024	Добавлены примеры конфигураций
1.02	08.11.2024	Уточнены сетевые диаграммы

#### Обзор документа

Данный документ описывает принципы работы сетевых фабрик ЦОД с использованием технологии VXLAN/EVPN, а также рекомендации по планированию, настройке и диагностике фабрик с использованием коммутаторов Aquarius.

Он предназначен для сетевых архитекторов, отвечающих за планирование сетевой инфраструктуры ЦОД, и сетевых инженеров, занимающихся её развертыванием, диагностикой и развитием. Предполагается предварительное знакомство с сетевыми технологиями, включая вопросы коммутации на 2 и 3 уровне. Документ должен использоваться совместно с документацией на программное и аппаратное обеспечение коммутаторов Aquarius, размещённой на посвященном им портале <https://net.aq.ru>

Информация, изложенная в данном руководстве, насколько возможно, отражает текущее представление об оптимальном внедрении в типичном случае корпоративного развертывания. Приведённые варианты топологии, логической структуры и конкретных настроек не являются единственными, и не обязательно являются наилучшими или корректными для всех сценариев.

Данное руководство задумано как живой документ, который будет дополняться и изменяться по мере выпуска новых функций программного обеспечения AqNOS, отработки дополнительных

сценариев использования и получения практического опыта применения. В текущем виде он может не отражать всех аспектов, которые нужно учитывать при внедрении.

Компания Аквариус не берёт на себя ответственность за полноту изложенной информации и отсутствие в ней ошибок и неточностей. Планирование и развертывание ответственных, крупных или технологически сложных внедрений всегда должно сопровождаться стендовой отработкой для подтверждения соответствия физической и логической архитектуры и настроек оборудования предъявляемым требованиям.

## Обзор технологии VXLAN для построения сетей ЦОД

Предметом данного документа являются архитектуры, использующие наложенные сети VXLAN с уровнем управления MP-BGP EVPN. По историческим причинам такие сети часто называют фабриками, и мы будем использовать этот термин в дальнейшем. Рассмотрим каждый из упомянутых элементов (VXLAN и MP-BGP EVPN) по отдельности.

Наложённые сети (также часто используется термин «оверлей» (overlay)) используют принцип «туннелирования» трафика через опорную сеть (underlay), при котором пакет наложенной сети помещается (инкапсулируется) внутрь пакета опорной сети на входном узле, и передаётся по ней до выходного узла, где выполняется декапсуляция. В центрах обработки данных доминирующей технологией оверлейных сетей является VXLAN (Virtual eXtensible Local Area Network, RFC 7348). Применительно к VXLAN речь идёт о передаче кадров Ethernet через опорную сеть IP (IPv4 или IPv6) путём их помещения в датаграмму протокола UDP. Структура VXLAN пакета показана на диаграмме ниже:



Поскольку опорная сеть, в свою очередь, обычно работает на основе Ethernet, фактически по опорной сети происходит передача Ethernet пакета внутри Ethernet пакета. В дальнейшем мы будем называть Ethernet и IP заголовок опорной сети «внешним» (outer), а Ethernet заголовок наложенной сети «внутренним» (inner). Внешний Ethernet заголовок меняется на каждом участке маршрутизации по обычным принципам работы IP сетей, внутренний остаётся неизменным от входного до выходного узла туннеля.

Что даёт использование оверлейных сетей? Поскольку в VXLAN сети Ethernet пакет передаётся в неизменном виде через IP сеть, то, по сути, происходит растягивание сетевых сегментов 2 уровня через маршрутизируемую инфраструктуру. Тем самым, появляется возможность миграции виртуальных машин, работы отказоустойчивых вычислительных кластеров, синхронизации систем, требующих связности на 2 уровне. При этом нижележащая опорная сеть остаётся маршрутизируемой, что обеспечивает все связанные с этим преимущества – быстрая сходимость при отказах, отсутствие потребности в отдельных механизмах для борьбы с петлями внутри сети, возможность эффективной балансировки трафика по альтернативным «путям с одинаковой стоимостью» (Equal Cost Multipathing – ECMP). Кроме того, для идентификации трафика разных логических сегментов в VXLAN используется 24-битное поле VNI (VXLAN Network Identifier), что, теоретически, позволяет транспортировать через единую сеть до 16 миллионов изолированных друг от друга VXLAN сегментов.

Необходимо отметить, что VXLAN как таковой описывает только собственно передачу Ethernet фреймов через туннели поверх IP сети – вопрос о том, как должно происходить установление туннелей, и через какой туннель доступен конкретный MAC адрес, не регламентируется

стандартом. RFC 7348 указывает, как основную возможность, выучивание ассоциаций между MAC адресами и IP адресами конечных узлов на основе прослушивания трафика (data-plane learning), а также упоминает возможность использование плоскости управления (control-plane) для распространения информации об адресах, подключенных к сети.

Использование data-plane learning соответствует той логике, которой следуют обычные коммутаторы Ethernet, и частично наследует от них недостатки: опора на широковещательные пакеты, централизованная маршрутизация, отсутствие контроля и возможности применения политик. Для эффективного внедрения VXLAN фабрик в сетях ЦОД необходимо решение с использованием плоскости управления, что позволит сократить опору на широковещание, сделать работу сети более контролируемой и стабильной.

Основным решением для данной цели является технология EVPN, опирающаяся на распространение служебной информации через протокол MP-BGP. С точки зрения MP-BGP, EVPN является одним из «семейств адресов» (address family, AF), а передаваемая EVPN информация часто называется префиксами или маршрутами, даже если она не является маршрутом в обычном понимании этого слова. Существенно, что использование плоскости управления на основе EVPN позволяет передавать информацию не только о MAC адресах подключений, но и об IP адресах подключений и подсетей. Это, совместно с общими механизмами MP-BGP для управления распространяемой адресной информацией (route targets) позволяет создавать наложенные сети со связностью на 2 и на 3 уровне, то есть с возможностью растягивать не только L2 сегменты, но и VRF – по сути, строить комбинированные L2+L3 VPN. При этом маршрутизация может выполняться на каждом из VTEP на входе в фабрику, без необходимости передавать трафик на какие-то центральные устройства. Поддержка распределённой маршрутизации является ключевым отличием VXLAN/EVPN от более простых оверлейных подходов. Отметим, что, с точки зрения коммутатора, маршрутизация в наложенной сети всегда выполняется в VRF.

## Элементы типичной VXLAN/EVPN фабрики

Поскольку сеть с использованием оверлейных технологий состоит из верхнего (overlay) и нижнего (underlay) слоёв, то удобно обсуждать её организацию на каждом из этих уровней по отдельности.

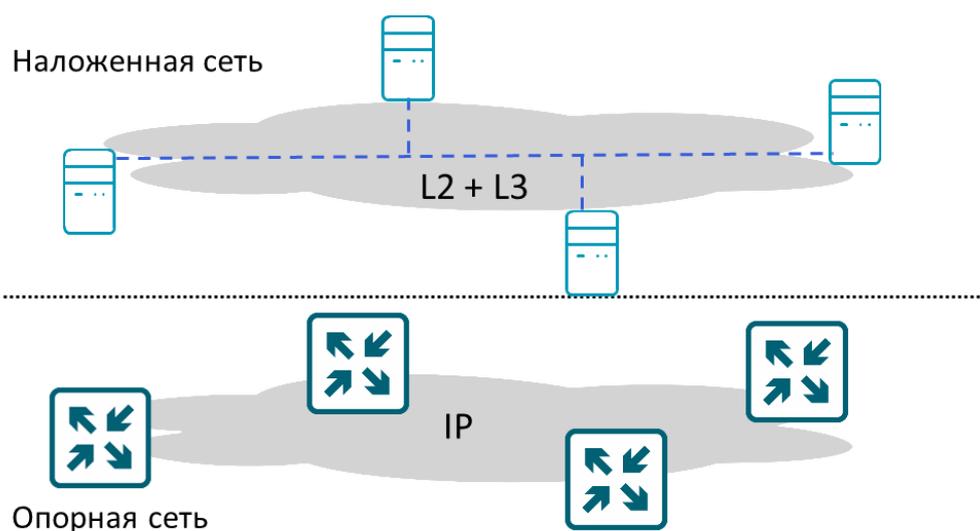


Рисунок 1. Наложённая и опорная сеть

Задачей нижележащей (опорной) сети является организация связности между коммутаторами, необходимой для работы как плоскости передачи данных, так и плоскости управления наложенной сети. Используемые в ней подходы должны обеспечивать высокую производительность и низкое время восстановления при отказе отдельных соединений или промежуточных коммутаторов. Таким требованиям хорошо соответствует топология Spine-Leaf, в которой все внешние подключения выполняются к оконечным коммутаторам (Leaf), а связь между этими коммутаторами осуществляется через набор промежуточных коммутаторов, образующих «хребет» (Spine). При этом все Leaf коммутаторы подключены ко всем Spine, и наоборот. Как правило, Leaf коммутатор реализует тот или иной механизм распределения (балансировки) трафика по альтернативным путям с «одинаковой стоимостью» (ECMP), проходящим через разные Spine коммутаторы.

Преимуществами такого подхода является:

- Высокая масштабируемая производительность. Максимальная общая производительность «хребта» сети с N коммутаторами Spine в N раз больше производительности одного коммутатора, и может наращиваться при необходимости
- Отказоустойчивость. При отказе одного из N Spine коммутаторов, или одного из соединений, теряется лишь 1/N часть общей производительности, при этом переключение трафика, с учетом наличия в таблицах коммутации N параллельных маршрутов, происходит очень быстро.
- Простота проектирования, внедрения и развития сети. Производительность в расчёте на Leaf коммутатор определяется числом Spine коммутаторов и скоростью соединений между коммутаторами, и может быть при необходимости увеличена путём добавления Spine. Максимальное число Spine определяется количеством аплинков на Leaf коммутаторе. Число оконечных (серверных) подключений определяется количеством Leaf коммутаторов (и портов на каждом из них), и может быть увеличено путём их добавления. Максимальное число Leaf коммутаторов определяется числом портов на Spine коммутаторе.

Более подробно требования к опорной сети и варианты её построения обсуждаются ниже.

Наложённая сеть, как уже было отмечено выше, опирается с точки зрения передачи данных (data plane) на VXLAN туннели через IP сеть, с точки зрения плоскости управления (control plane) – на MP-BGP EVPN. Соответственно, работа её элементов включает в себя следующие основные функции:

- Функции терминирования туннеля (инкапсуляции и декапсуляции туннельного трафика). В терминологии VXLAN точка терминирования туннеля называется VTEP (VXLAN Tunnel Endpoint)
- Соотнесение информации о получателе пакета (MAC и/или IP адресе) с туннелем, в который должен быть направлен пакет. Используемые для этого в VXLAN/EVPN сети таблицы заполняются, для локальных подключений к VTEP, за счет прослушивания трафика (изучения MAC и IP), для удалённых – за счёт получения информации по EVPN
- Передача широковещательных (broadcast), групповых (multicast) и адресованных неизвестному получателю (unknown unicast) пакетов наложенной сети. Для обозначения совокупно всех перечисленных типов часто используется термин BUM-трафик (Broadcast, Unknown Unicast, Multicast). Существует два основных способа доставки BUM трафика в VXLAN сетях:
  - Передача его в виде группового (multicast) трафика в IP сети. Достоинством такого подхода является то, что необходимую репликацию (размножение) трафика выполняет сама транспортная сеть. Недостатком является то, что для этого сеть

- должна поддерживать маршрутизацию multicast, что поддерживается не всем сетевым оборудованием, и требует отдельного набора компетенций для планирования, настройки и диагностики сети.
- Передача его в виде необходимого числа обычных unicast пакетов IP в опорной сети, адресованных другим VTEP, относящимся к данному сегменту. При этом репликацию выполняет входной VTEP, через который пакет попадает в наложенную сеть, поэтому такой метод часто называется входной репликацией (ingress replication)
  - Функции взаимодействия по MP-BGP. Помимо конечных устройств, отправляющих и получающих информацию по MP-BGP, логический дизайн обычно включает в себя промежуточные узлы. В случае использования единого номера автономной системы эти узлы называются «отражателями маршрутов» (route reflector), и выполняют распространение информации по MP-BGP между всеми коммутаторами без необходимости организации полносвязного набора сессий. Более подробно варианты логического дизайна MP-BGP обсуждаются ниже.

С учётом сказанного выше, простейший типовой дизайн сети на основе VXLAN/EVPN выглядит следующим образом:

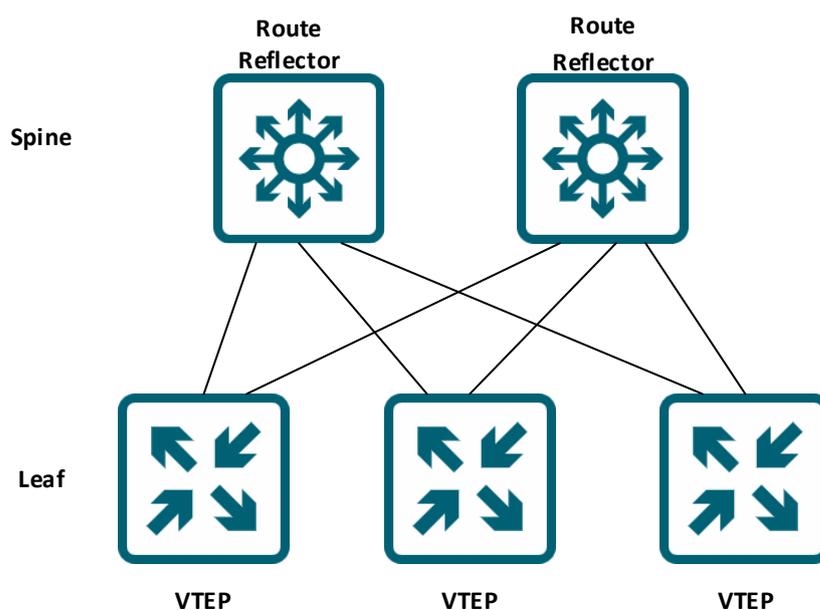


Рисунок 2. Структура VXLAN/EVPN фабрики Spine/Leaf

- Leaf коммутаторы выступают точкой подключения всего конечного оборудования (серверов и т.д), а также внешних сетевых соединений, несущих трафик к серверам. С точки зрения VXLAN сети они являются VTEP-ами (иногда также используется термин «VXLAN шлюз»). С точки зрения MP-BGP они являются конечными узлами, и могут использовать как полносвязный (full mesh) набор сессий для обмена информацией, так и устанавливать сессии только к «отражателям маршрутов» (как правило, двум, из соображения отказоустойчивости). Leaf коммутаторы, как правило, не соединяются между собой напрямую, исключением из этого правила может выступать организация пар по технологии MLAG для обеспечения зарезервированных подключений серверов.
- Spine коммутаторы обеспечивают связь между Leaf коммутаторами. В простейшем сценарии они не выполняют операций инкапсуляции/декапсуляции туннельного трафика, и

могут вообще «не знать» о его существовании, их функцией является только маршрутизация IP. С точки зрения MP-BGP на Spine коммутаторы, как правило, возлагается роль «отражателей маршрутов» (route reflector, RR). К Spine коммутаторам, как правило, подключаются только Leaf коммутаторы (исключением могут являться, например, модульные фабрики, которые предполагается рассмотреть в будущих версиях данного документа), и они не соединяются между собой.

## Опорная сеть (underlay)

В число ключевых требований к опорной сети входят:

- Высокая производительность, как общая, так и на уровне отдельных сетевых соединений. Желательно, чтобы скорость соединений внутри опорной сети превышала скорость окончательных подключений к Leaf коммутаторам.
- Быстрая сходимости, что требует эффективно работающих протоколов динамической маршрутизации, и поддержки быстрого переключения трафика на альтернативные соединения при отказе.
- Поддержка передачи пакетов увеличенного размера (jumbo frames). Поскольку VXLAN инкапсуляция предполагает добавление дополнительного заголовка, то опорная сеть должна поддерживать передачу фреймов размера, увеличенного на 50 байт относительно исходного (если в опорной сети используются пакеты с заголовками 802.1Q, то на 54 байта). Например, если сервер передаёт фреймы размером до 9000 байт, то, при условии использования нетегированных портов, коммутаторы сети должны поддерживать передачу фреймов размером как минимум до 9050 байт.

Реализация перечисленных требований частично зависят от используемого оборудования и схемы сети, частично – от используемых протоколов и настроек.

В качестве базовой сетевой архитектуры для VXLAN/EVPN, как правило, используется топология Spine/Leaf. Принципы её работы и преимущества уже были описаны выше. Текущая версия документа описывает построение сетей, масштаб которых укладывается в возможности одного «блока» Spine/Leaf. Расширение архитектуры для увеличения масштаба внедрения в пределах ЦОД, а также для связи сетей нескольких ЦОД, предполагается осветить в будущих версиях руководства.

Логическая организация опорной сети включает в себя выбор протокола динамической маршрутизации, а также набора его настроек. Основными вариантами здесь являются BGP, с одной стороны, и современные IGP протоколы (OSPF, IS-IS) – с другой. Оба варианта могут успешно использоваться в опорной сети VXLAN/EVPN фабрик, у каждого из них есть свои преимущества – в том числе, и с точки зрения наличия в организации соответствующих эксплуатационных практик. BGP как протокол маршрутизации в крупных ЦОД сейчас достаточно популярен (см, в частности, RFC 7938), но, применительно к задачам построения фабрики VXLAN/EVPN, он обладает рядом особенностей, усложняющих его внедрение. В текущей версии документа мы будем в дальнейшем рассматривать вариант, опирающийся на использование протокола OSPF в единой «бэкбонной» области (area 0), но, где возможно, упомянем отличия и особенности, возникающие при использовании BGP в качестве протокола маршрутизации в опорной сети.

## Наложенная сеть: инкапсуляция и сигнализация

В наложенной сети, как и в опорной, выделяются задачи плоскости данных (инкапсуляция/декапсуляция VXLAN) и плоскости управления (сигнализация с использованием адресного семейства EVPN протокола MP-BGP).

Инкапсуляция VXLAN уже была в общих чертах описана выше, но необходимо дополнительно остановиться на нескольких существенных деталях.

В качестве адресов источника и получателя во внешнем IP заголовке, в принципе, могут использоваться любые адреса, обеспечивающие достижимость между VTEP-ми. Но наилучшим вариантом является использование адреса логического (loopback) интерфейса, поскольку такой адрес достижим независимо от состояния конкретных физических интерфейсов, а, значит, в случае отказа одного из соединений в сети, на уровне наложенной сети (в EVPN) не возникает необходимости в распространении и обработке информации об изменениях – всё обрабатывается протоколом маршрутизации в опорной сети.

В качестве порта назначения протокола UDP всегда используется выделенный для VXLAN номер 4789 (нужно упомянуть, что в некоторых старых реализациях использовался номер 8472). А вот номер порта источника может меняться, и это играет важную роль в построении продуктивных сетей. Дело в том, что, если бы номер порта источника также был бы фиксирован, то, с точки зрения опорной сети, весь трафик между данной парой VTEP представлял бы собой один L4 поток, и, соответственно, был бы направлен логикой хеширования по одному и тому же пути из всех возможных. Чтобы избежать такой ситуации, номер порта источника используется в VXLAN как «метка энтропии» - в него переносится информация из внутреннего заголовка, которая позволяет, по возможности, отобразить разные потоки в наложенной сети в разные потоки в опорной сети.

Сигнализация в VXLAN/EVPN фабрике служит для распространения необходимой информации, прежде всего – о подключениях к фабрике: MAC адресах, IP адресах и IP подсетях. Для этого используются различные типы префиксов в адресном семействе EVPN, набор поддерживаемых типов зависит от особенностей реализации в EVPN фабрике конкретного производителя. Реализация в ОС AqNOS использует следующие типы префиксов:

- Тип 2 (Route Type 2). Используется для передачи информации о подключенных к фабрике конечных адресах. В случае, если в фабрике или конкретном VNI используется только коммутация 2-го уровня, маршруты 2-го типа несут только информацию о MAC адресах. В случае, если фабрикой выполняется маршрутизация IP, для каждого конечного устройства, имеющего IP адрес, передаётся два маршрута 2-го типа: первый только с MAC адресом, второй с MAC адресом и с IP адресом.
- Тип 5 (Route Type 5). Используется для передачи информации о доступных через фабрику IP подсетях. Применяется только в случае, если фабрика выполняет маршрутизацию IP. Данный тип маршрутов используется в двух разных сценариях. Во-первых, они могут соответствовать внешним, по отношению к фабрике, подсетям – например, в корпоративной сети, в Интернете и т.д. Во-вторых, они могут соответствовать подсетям самой фабрики, это необходимо для формирования ARP запросов при обращении к IP адресам в подсетях фабрики, которые пока что ей неизвестны как маршруты 2-го типа
- Тип 3 (Route Type 3). Этот тип маршрутов используется для распространения информации о VTEP-ах, относящихся к конкретному сегменту (VNI), и используется для репликации BUM трафика.

Каждый анонс несёт в себе информацию об адресе, через который доступен соответствующий объект (next hop), логических сегментах (VNI), признаках MAC/IP VRF с которыми связаны анонсы и др. Точный набор атрибутов зависит от типа префикса и сценария его использования. Схема назначения Route Distinguisher (RD) и Route Target (RT) является, вообще говоря, предметом

выбора проектировщика, но стандарты RFC 7432 и RFC 8365 описывают рекомендации по их использованию, и требования к ним в случае автоматического назначения, которым мы стараемся следовать в примерах конфигураций в данном документе.

Передача информации адресного семейства EVPN в фабрике использует общую MP-BGP инфраструктуру сети, и оперирует теми же терминами, что и обычное взаимодействие по BGP, включая AS номера, BGP сессии между устройствами и так далее. Для внедрения MP-BGP в фабрике возможно два принципиальных варианта – IBGP (при котором коммутаторы используют один и тот же номер автономной системы), либо EBGP (при котором номера автономной системы на разных коммутаторах, вообще говоря, отличаются). Оба варианта имеют право на существование, и некоторые соображения по их поводу рассмотрены ниже, но для внедрения в пределах одного блока Spine/Leaf архитектуры проще вариант с использованием IBGP.

По правилам работы BGP информация, полученная из одной IBGP сессии, не передаётся в другую. Поэтому для при использовании IBGP необходимо использование либо полносвязного набора сессий между всеми VTEP (что малопримемлемо даже для небольших внедрений), либо использование «отражателя маршрутов» (route reflector), который, как правило, располагается на Spine коммутаторе (для отказоустойчивости – на двух). Если же используется EBGP, то для передачи EVPN информации через Spine коммутаторы на них необходимо в настройке адресного семейства BGP EVPN указать команду `retain route-targets all`, чтобы обеспечить на них приём и распространение информации, не помещаемой Spine коммутатором в локальные таблицы.

## Подключения к VXLAN/EVPN фабрике

В VXLAN фабрике все внешние подключения осуществляются к Leaf коммутаторам, выполняющим функцию VTEP. Это касается серверов, сервисных устройств (МСЭ, балансировщиков и т.д.) и внешних сетей.

Модель подключения оконечных устройств к VXLAN/EVPN фабрике на коммутаторах Aquarius использует отображение VLAN коммутатора в VNI VXLAN фабрики. Тем самым, управление подключениями сохраняет привычный синтаксис и семантику, и отличия возникают только в разделах конфигурации, связанных с VXLAN. При этом соотношение VLAN и VNI, вообще говоря, является локальным для коммутатора, то есть разные коммутаторы могут отображать один и тот же VNI в разные VLAN, и, наоборот, отображать одни и те же VLAN в разные VNI – хотя, на практике, такой подход следует использовать с осторожностью, чтобы сократить риск ошибок.

Как правило, для повышения надежности сервисов ЦОД используется отказоустойчивое подключение серверов. Часто оно осуществляется с использованием механизма агрегирования портов в единый канал с использованием протокола LACP (либо статической настройки). Для поддержки резервирования на стороне сети используется технология MC-LAG (Multi-Chassis Link AGgregation), часто называемая также MLAG, позволяющая внешним устройствам (например, серверам) строить агрегированные подключения сразу к паре коммутаторов доступа.

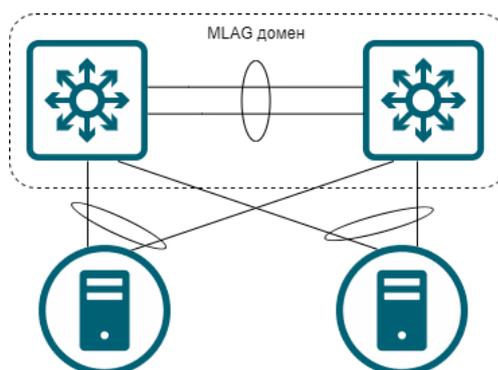


Рисунок 3. MLAG пара и подключения к ней.

Детали работы и настройки MLAG выходят за рамки данного документа, но, применительно к работе фабрики, очевидно, возникает вопрос о том, как соотнести организацию VXLAN туннелей между конкретными коммутаторами с тем, что при использовании MLAG сервер должен быть доступен сразу через два коммутатора, образующих MLAG пару. Реализация в AqNOS решает эту задачу с использованием подхода, называемого anycast VTEP, при котором два коммутатора в MLAG паре используют один и тот же IP адрес VTEP (привязанный к логическому интерфейсу), доступность которого анонсируется обоими устройствами в протокол динамической маршрутизации опорной сети. Пока оба коммутатора в MLAG паре работоспособны, трафик, адресованный подключенным к паре серверам, балансируется сетью между ними за счёт наличия альтернативных маршрутов с одинаковой стоимостью. При отказе одного из них весь трафик идёт через оставшийся в работе коммутатор, при этом, что существенно, это целиком обрабатывается опорной сетью и никак не затрагивает работу уровня наложенной сети. Заметим, что с точки зрения MP-BGP устройства в MLAG паре являются независимыми, и используют для построения сессий разные адреса.

Подключения фабрики к внешнему миру также выполняются Leaf коммутатором, который, в этом случае, иногда называют термином Border Leaf. В его задачи входит передача трафика между VXLAN туннелями и внешними маршрутизируемыми соединениями, а также обмен маршрутной информацией между фабрикой и внешним миром. В простейшем случае реализация функции Border Leaf сводится к настройке в протоколе BGP для соответствующего VRF редистрибуции из протокола динамической маршрутизации, либо из статических маршрутов, возможно - с применением необходимых фильтров (route map).

## Особенности использования коммутаторов Aquarius с ОС AqNOS для построения VXLAN/EVPN фабрики

В номенклатуре коммутаторов Aquarius для Spine уровня рекомендуется модель N6000-32C, обладающая 32 портами, работающими на скорости 100 Гбит/с (16 из них также могут работать на скорости 40 Гбит/с, а также поддерживают разбиение в режиме 4\*10 Гбит/с и 4\*25 Гбит/с). Соответственно, при использовании в фабрике соединений 40/100 Гбит/с обеспечивается подключение до 32 коммутаторов Leaf уровня.



Рисунок 4. Коммутатор AQ-N6000-32C

Основной моделью Aquarius для внедрения на уровне Leaf является коммутатор N6000-48Y8C, имеющий 48 портов SFP28 1/10/25 Гбит/с, используемых для подключения серверов и другого оконечного оборудования, и 8 портов QSFP28 40/100 Гбит/с. Как правило, 2 порта QSFP28 используются для соединения коммутаторов в MLAG паре, что оставляет доступными для подключения к Spine уровню до 6 восходящих портов (аплинков) QSFP28, что, соответственно, позволяет использовать в фабрике до 6 Spine коммутаторов. Заметим, что при задействовании всех 48 SFP28 портов доступа на скорости 25Гбит/с, и 6 QSFP28 аплинков на скорости 100 Гбит/с, коэффициент переподписки от серверов к аплинкам составит 2:1, что является очень хорошим показателем.



Рисунок 5. Коммутатор AQ-N6000-48Y8C

При необходимости подключения к фабрике высокопроизводительных серверных систем, СХД и т.д. в качестве Leaf коммутатора может использоваться упомянутая выше модель N6000-32C, при этом количество портов, которые используются для связи со Spine уровнем, диктуется целевыми показателями переподписки.

Хотя для продуктивного использования в современных ЦОДах скорость 1 Гбит/с, как правило, недостаточна, такие подключения бывают необходимы для портов управления, подключений BMC серверов и т.д. В этом случае в качестве Leaf коммутатора может использоваться модель N3000-48T4Y2C, имеющая 48 витопарных портов 1 Гбит/с, 2 порта QSFP+ 40 Гбит/с и 4 порта SFP28 10/25 Гбит/с. Для поддержки VXLAN функциональности на ней необходим дополнительный заказ лицензии Professional.



Рисунок 6. Коммутатор N3000-48T4Y2C

Все перечисленные модели находятся под управлением единой сетевой ОС AqNOS, и, при наличии соответствующих лицензий, идентичны с точки зрения поддерживаемых функций применительно к задачам VXLAN/EVPN. В дальнейшем обсуждении, мы не будем акцентировать внимание на используемых в примерах моделях, понимая под ними коммутаторы Aquarius семейств 3000-6000 с лицензией Professional.

С точки зрения протоколов опорной сети для использования доступны как IGP протоколы (OSPF, IS-IS), так и BGP. Отметим, что в AqNOS номер автономной системы является единым на коммутаторе и общим для разных адресных семейств, поэтому, если при планировании внедрения выбирается BGP в качестве протокола маршрутизации в опорной сети, то если в опорной сети используется EBGP (то есть на разных коммутаторах используются разные номера автономных систем), то и EVPN будет использовать EBGP.

Рекомендуемым способом организации опорной сети является использование физических L3 интерфейсов («**no switchport**»). На них также нужно выполнить две настройки:

- Разрешить обработку VXLAN трафика командой «**vxlan enable**»
- Разрешить обработку IP пакетов увеличенного размера командой «**mtu 9216**»

Репликация в наложенной сети VXLAN в AqNOS всегда выполняется в режиме «входной репликации» (ingress replication), опирающейся на EVPN анонсы 3-го типа, поэтому предусматривать поддержку группового (multicast) трафика в опорной сети не требуется.

## Планирование и внедрение VXLAN/EVPN фабрики

Прежде, чем начинать проектирование и внедрение VXLAN/EVPN фабрики, следует определить несколько базовых моментов:

- Число и тип используемых коммутаторов уровней Leaf и Spine, определяющиеся требованиями по оконечным подключениям и целевым показателям по переподписке
- Структура MLAG пар, VLAN интерфейсы для синхронизации в MLAG и адресация на них
- Тип протокола динамической маршрутизации в опорной сети. Как уже упоминалось, данная версия валидированного дизайна предполагает использование протокола OSPF
- Тип настройки MP-BGP для распространения EVPN информации. Данная версия валидированного дизайна предполагает использование IBGP, то есть общего номера AS на всех коммутаторах, и размещение route reflector на двух Spine коммутаторах. Если

предполагается взаимодействие по BGP с внешними сетями, номер автономной системы должен выбран с учётом требований такого сопряжения.

- Адресация в опорной сети. Желательно, чтобы под адресное пространство опорной сети был выделен единый суммаризируемый адресный диапазон, что позволит упростить настройку протоколов маршрутизации. Размер адресного пространства должен учитывать, с запасом на будущее расширение, необходимое количество соединений между участниками фабрики (как соединения Spine-Leaf, так и в MLAG парах), а также виртуальные интерфейсы (loopback). Структура адресации, по возможности, должна позволять по виду адреса легко идентифицировать его функцию и расположение в сетевой топологии
- Правила формирования VXLAN VNI. Как уже упоминалось, VXLAN использует в качестве идентификатора L2 и L3 сегментов 24-битное поле VNI, что соответствует более чем 16 миллионам значений. Как правило, для L2 и L3 VNI используются непересекающиеся диапазоны, которые позволяют быстро определить, к какой категории относится номер VNI и (в случае L2 VNI) с каким номером VLAN его предполагается связывать (в простом сценарии, использующем единую схему отображения VXLAN на VLAN). Например, для L2 VNI могут использоваться номера вида 2xxxx, где xxxx – номер VLAN, а для L3 VNI могут использоваться номера, начиная с 30000. При необходимости возможна и более сложная структура выбора VNI – например, с группированием по подразделению, номеру заказчика и т.д.
- Правила формирования идентификаторов, связанных с MP-BGP (Route Distinguisher/Route Target)

Вот модельный пример внедрения, на котором мы будем демонстрировать блоки настройки, и, в следующем разделе, общие конфигурации:

- 3 Leaf коммутатора, из которых Leaf3 является одиночным, а Leaf1 и Leaf2 объединены в MLAG пару
- Используются два Spine, оба являются Route Reflector, на всей фабрике используется номер AS 65000
- Под внутреннее адресное пространство выбран диапазон 172.20.0.0/16 (такой большой диапазон выбран для наглядности структуры адресов)
- VLAN 101 «растянут» между всеми Leaf, а VLAN 102/103 присутствуют только на MLAG паре или только на Leaf3, соответственно. Такая модельная конфигурация позволяет проверить и связность через фабрику на 2 уровне, и маршрутизацию между локализованными сетями
- На фабрике используется один клиентский VRF (VRF1), к которому относятся подсети, связанные с VLAN 101-103

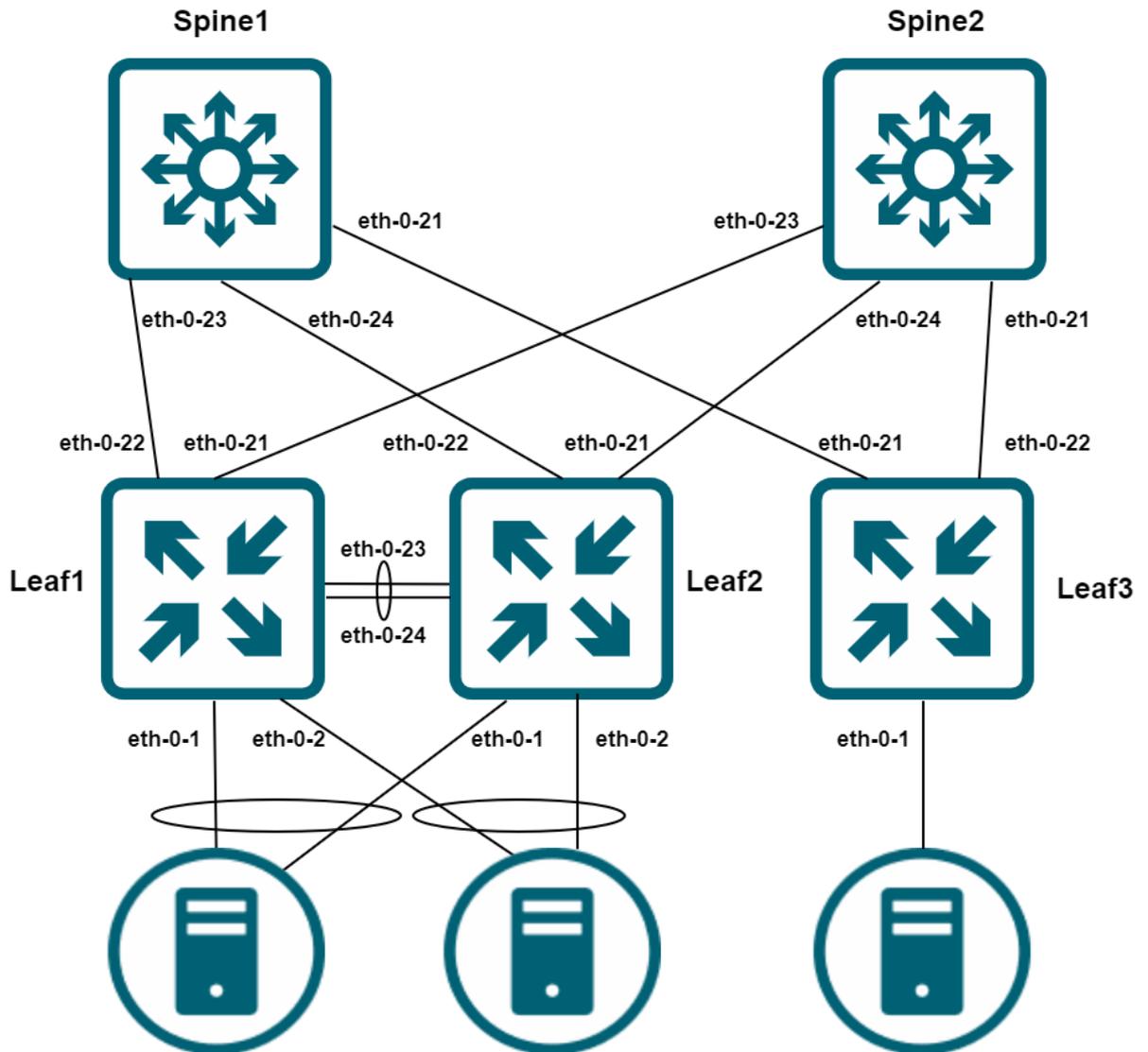


Рисунок 7. Схема тестовой фабрики

## Структура конфигурации Leaf на примере Leaf1:

### L2 сервисы

```
vlan database
vlan 101
vlan 101 overlay enable
vlan 102
vlan 102 overlay enable
vlan 4094
!
overlay
vlan 101 vni 20101
vlan 101 mac-address-tunnel learning-disable
vlan 102 vni 20102
vlan 102 mac-address-tunnel learning-disable
!
evpn
vni 20101
rd 10.0.2.11:20101
route-target both 65000:20101
route-target export 65000:30001
vni 20102
rd 10.0.2.11:20102
route-target both 65000:20102
route-target export 65000:30001
!
```

Разрешение привязки VLAN к оверлею

L2 VNI для VLAN

Выключение data-plane learning

RD для L2 VNI

RT для L2 VNI

RT для связанного L3 VNI

### L3 сервисы

```
ip virtual-router mac 0000.c0ff.ee00
evpn virtual-mac enable
!
ip vrf Tenant1
vxlan vni 30001
rd 10.0.2.11:30001
route-target both 65000:30001 evpn
!
interface vlan 101
ip vrf forwarding Tenant1
ip address 192.168.101.253/24
ip virtual-router address 192.168.101.1
overlay distributed-gateway enable
overlay host-collect enable
!
interface vlan 102
ip vrf forwarding Tenant1
ip address 192.168.102.253/24
ip virtual-router address 192.168.102.1
overlay distributed-gateway enable
overlay host-collect enable
!
```

MAC адрес распределённого шлюза по умолчанию

Описание клиентского VRF

L3 VNI для VRF

Route Distinguisher для L3 VNI

Route Target для L3 VNI

Связь VLAN интерфейса с VRF

IP адрес распределённого шлюза по умолчанию

Распределённый шлюз по умолчанию

и сбор информации для анонса адресов

### VXLAN инкапсуляция

```
interface loopback 0
description VTEP tunnel source
ip address 10.0.1.11/32
!
interface nve 0
source loopback 0
```

IP адрес, к которому привязан туннель

```
member vni 20101
member vni 20102
member vni 30001 associate-vrf
!
```

L2 VNI, привязанные к туннельному интерфейсу  
L3 VNI, привязанные к туннельному интерфейсу

## EVPN пиринг

```
interface loopback 1
description MP-BGP peering loopback
ip address 10.0.2.11/32
!
router bgp 65000
bgp router-id 10.0.2.11
max-paths ebgp 8
max-paths ibgp 8
neighbor 10.0.2.1 remote-as 65000
neighbor 10.0.2.1 update-source loopback 1
neighbor 10.0.2.2 remote-as 65000
neighbor 10.0.2.2 update-source loopback 1
address-family ipv4
no neighbor 10.0.2.1 activate
no neighbor 10.0.2.2 activate
exit
address-family l2vpn evpn
neighbor 10.0.2.1 activate
neighbor 10.0.2.2 activate
neighbor 10.0.2.1 send-community both
neighbor 10.0.2.2 send-community both
exit
address-family ipv4 vrf Tenant1
redistribute connected
advertise l2vpn evpn
exit
```

IP адрес, к которому привязаны BGP сессии  
Соседства на loopback адресах  
Выключаем AF IPv4 для EVPN соседства  
Включаем AF EVPN  
Передача extended community (включая RD/RT)  
Анонс подсетей – нужен для работы ARP glean  
Передача IP информации в BGP EVPN

## Связность в опорной сети

```
interface eth-0-22
no switchport
ip address 10.0.111.11/24
description Core link to Spine1:eth-0-23
ip ospf network point-to-point
vxlan uplink enable
mtu 9216
no shutdown
!
interface eth-0-21
no switchport
ip address 10.0.112.11/24
description Core link to Spine2:eth-0-21
ip ospf network point-to-point
vxlan uplink enable
mtu 9216
no shutdown
!
router ospf 65000
router-id 10.0.2.11
network 10.0.0.0/16 area 0
```

Соединения Leaf-Spine  
OSPF point-to-point для ускорения сходимости  
Включение обработки VXLAN  
Увеличенный IP MTU для VXLAN трафика  
OSPF протокол для связности в опорной сети

## Настройка MLAG домена

```
no spanning-tree enable
!
interface eth-0-23
switchport
channel-group 63 mode active
description Peer link to Leaf2
no shutdown
!
interface eth-0-24
switchport
channel-group 63 mode active
description Peer link to Leaf2
no shutdown
!
interface agg 63
switchport mode trunk
switchport trunk allowed vlan all
spanning-tree port disable
!
interface vlan 4094
ip address 10.0.11.11/24
!
mlag configuration
peer-address 10.0.11.12
peer-link agg 63
!
```

Порты для организации peer-link

Агрегированный канал peer-link

VLAN интерфейс для связи в MLAG паре

MLAG домен

#### Клиентские подключения

```
interface eth-0-1
switchport
switchport mode access
switchport access vlan 101
channel-group 1 mode active
!
interface eth-0-2
switchport
switchport mode access
switchport access vlan 102
channel-group 2 mode active
!
interface agg 1
switchport access vlan 101
mlag 1
!
interface agg 2
switchport access vlan 102
mlag 2
!
```

## Структура конфигурации Spine на примере Spine1

### EVPN пиринг

```
evpn
!
interface loopback 1
  description MP-BGP peering loopback
  ip address 172.20.2.1/32
!
router bgp 65000
  bgp router-id 1.1.1.1
  max-paths ebgp 4
  max-paths ibgp 4
  neighbor 172.20.2.11 remote-as 65000
  neighbor 172.20.2.11 update-source loopback 1
  neighbor 172.20.2.12 remote-as 65000
  neighbor 172.20.2.12 update-source loopback 1
  neighbor 172.20.2.13 remote-as 65000
  neighbor 172.20.2.13 update-source loopback 1
address-family ipv4
  no neighbor 172.20.2.11 activate
  no neighbor 172.20.2.12 activate
  no neighbor 172.20.2.13 activate
exit
address-family l2vpn evpn
  neighbor 172.20.2.11 activate
  neighbor 172.20.2.12 activate
  neighbor 172.20.2.13 activate
  neighbor 172.20.2.11 send-community both
  neighbor 172.20.2.12 send-community both
  neighbor 172.20.2.13 send-community both
  neighbor 172.20.2.11 route-reflector-client
  neighbor 172.20.2.12 route-reflector-client
  neighbor 172.20.2.13 route-reflector-client
exit
```

IP адрес, к которому привязаны BGP сессии

Соседства на loopback адресах

Выключаем AF IPv4 для EVPN соседства

Передача extended community (включая RD/RT)

Используется IBGP, Spine является RR

### Связность в опорной сети

```
interface eth-0-23
  no switchport
  ip address 10.0.111.1/24
  description Core link to Leaf1:eth-0-22
  ip ospf network point-to-point
  mtu 9216
  no shutdown
!
interface eth-0-24
  no switchport
  ip address 10.0.121.1/24
  description Core link to Leaf2:eth-0-22
  ip ospf network point-to-point
  mtu 9216
  no shutdown
!
interface eth-0-21
  no switchport
  ip address 10.0.131.1/24
  description Core link to Leaf3:eth-0-21
  ip ospf network point-to-point
  mtu 9216
  no shutdown
```

Соединения Spine-Leaf

OSPF point-to-point для ускорения сходимости  
Увеличенный IP MTU для VXLAN трафика

!

```
router ospf 65000  
router-id 10.0.2.1  
network 10.0.0.0/16 area 0
```

OSPF протокол для связности в опорной сети

## Примеры целостных конфигураций оборудования

Приведённые ниже примеры конфигураций соответствуют сетевой топологии, описанной в разделе выше. Для краткости в конфигурациях оставлены только значимые части, порядок следования отдельных частей конфигурации не обязательно соответствует тому, в котором коммутатор её отображает.

Leaf1

```
hostname Leaf1
!
vlan database
vlan 101
vlan 101 overlay enable
vlan 102
vlan 102 overlay enable
vlan 4094
exit
!
no spanning-tree enable
ip virtual-router mac 0000.c0ff.ee00
evpn virtual-mac enable
!
ip vrf Tenant1
vxlan vni 30001
rd 10.0.2.11:30001
route-target both 65000:30001 evpn
!
overlay
vlan 101 vni 20101
vlan 101 mac-address-tunnel learning-disable
vlan 102 vni 20102
vlan 102 mac-address-tunnel learning-disable
!
evpn
vni 20101
rd 10.0.2.11:20101
route-target both 65000:20101
route-target export 65000:30001
vni 20102
rd 10.0.2.11:20102
route-target both 65000:20102
route-target export 65000:30001
!
interface loopback 0
description VTEP tunnel source
ip address 10.0.1.11/32
!
interface loopback 1
description MP-BGP peering loopback
ip address 10.0.2.11/32
!
interface eth-0-22
no switchport
ip address 10.0.111.11/24
description Core link to Spine1:eth-0-23
ip ospf network point-to-point
vxlan uplink enable
mtu 9216
no shutdown
!
```

```
interface eth-0-21
  no switchport
  ip address 10.0.112.11/24
  description Core link to Spine2:eth-0-21
  ip ospf network point-to-point
  vxlan uplink enable
  mtu 9216
  no shutdown
!
interface eth-0-23
  switchport
  channel-group 63 mode active
  description Peer link to Leaf2:eth-0-23
  no shutdown
!
interface eth-0-24
  switchport
  channel-group 63 mode active
  description Peer link to Leaf2:eth-0-24
  no shutdown
!
interface nve 0
  source loopback 0
  member vni 20101
  member vni 20102
  member vni 30001 associate-vrf
!
interface vlan 101
  ip vrf forwarding Tenant1
  ip address 192.168.101.253/24
  ip virtual-router address 192.168.101.1
  overlay distributed-gateway enable
  overlay host-collect enable
!
interface vlan 102
  ip vrf forwarding Tenant1
  ip address 192.168.102.253/24
  ip virtual-router address 192.168.102.1
  overlay distributed-gateway enable
  overlay host-collect enable
!
interface agg 63
  switchport mode trunk
  switchport trunk allowed vlan all
  spanning-tree port disable
!
interface vlan 4094
  ip address 10.0.11.11/24
!
mlag configuration
  peer-address 10.0.11.12
  peer-link agg 63
!
interface eth-0-1
  switchport
  switchport mode access
  switchport access vlan 101
  channel-group 1 mode active
!
interface eth-0-2
  switchport
  switchport mode access
  switchport access vlan 102
```

```
channel-group 2 mode active
!
interface agg 1
switchport access vlan 101
mlag 1
!
interface agg 2
switchport access vlan 102
mlag 2
!
router ospf 65000
router-id 10.0.2.11
network 10.0.0.0/16 area 0
!
router bgp 65000
bgp router-id 10.0.2.11
max-paths ebgp 8
max-paths ibgp 8
neighbor 10.0.2.1 remote-as 65000
neighbor 10.0.2.1 update-source loopback 1
neighbor 10.0.2.2 remote-as 65000
neighbor 10.0.2.2 update-source loopback 1
address-family ipv4
no neighbor 10.0.2.1 activate
no neighbor 10.0.2.2 activate
exit
address-family l2vpn evpn
neighbor 10.0.2.1 activate
neighbor 10.0.2.2 activate
neighbor 10.0.2.1 send-community both
neighbor 10.0.2.2 send-community both
exit
address-family ipv4 vrf Tenant1
redistribute connected
advertise l2vpn evpn
exit
```

## Leaf2

```
hostname Leaf2
!
vlan database
vlan 101
vlan 101 overlay enable
vlan 102
vlan 102 overlay enable
vlan 4094
exit
!
no spanning-tree enable
ip virtual-router mac 0000.c0ff.ee00
evpn virtual-mac enable
!
ip vrf Tenant1
vxlan vni 30001
rd 10.0.2.12:30001
route-target both 65000:30001 evpn
!
overlay
vlan 101 vni 20101
vlan 101 mac-address-tunnel learning-disable
vlan 102 vni 20102
```

```
vlan 102 mac-address-tunnel learning-disable
!
evpn
vni 20101
rd 10.0.2.12:20101
route-target both 65000:20101
route-target export 65000:30001
vni 20102
rd 10.0.2.12:20102
route-target both 65000:20102
route-target export 65000:30001
!
interface loopback 0
description VTEP tunnel source
ip address 10.0.1.11/32
!
interface loopback 1
description MP-BGP peering loopback
ip address 10.0.2.12/32
!
interface eth-0-22
no switchport
ip address 10.0.121.12/24
description Core link to Spine1:eth-0-24
ip ospf network point-to-point
vxlan uplink enable
mtu 9216
no shutdown
!
interface eth-0-21
no switchport
ip address 10.0.122.12/24
description Core link to Spine2:eth-0-22
ip ospf network point-to-point
vxlan uplink enable
mtu 9216
no shutdown
!
interface eth-0-23
switchport
channel-group 63 mode active
description Peer link to Leaf1:eth-0-23
no shutdown
!
interface eth-0-24
switchport
channel-group 63 mode active
description Peer link to Leaf1:eth-0-24
no shutdown
!
interface nve 0
source loopback 0
member vni 20101
member vni 20102
member vni 30001 associate-vrf
!
interface vlan 101
ip vrf forwarding Tenant1
ip address 192.168.101.252/24
ip virtual-router address 192.168.101.1
overlay distributed-gateway enable
overlay host-collect enable
!
```

```
interface vlan 102
 ip vrf forwarding Tenant1
 ip address 192.168.102.252/24
 ip virtual-router address 192.168.102.1
 overlay distributed-gateway enable
 overlay host-collect enable
!
interface agg 63
 switchport mode trunk
 switchport trunk allowed vlan all
 spanning-tree port disable
!
interface vlan 4094
 ip address 10.0.11.12/24
!
mlag configuration
 peer-address 10.0.11.11
 peer-link agg 63
!
interface eth-0-1
 switchport
 switchport mode access
 switchport access vlan 101
 channel-group 1 mode active
!
interface eth-0-2
 switchport
 switchport mode access
 switchport access vlan 102
 channel-group 2 mode active
!
interface agg 1
 switchport access vlan 101
 mlag 1
!
interface agg 2
 switchport access vlan 102
 mlag 2
!
router ospf 65000
 router-id 10.0.2.12
 network 10.0.0.0/16 area 0
!
router bgp 65000
 bgp router-id 10.0.2.12
 max-paths ebgp 8
 max-paths ibgp 8
 neighbor 10.0.2.1 remote-as 65000
 neighbor 10.0.2.1 update-source loopback 1
 neighbor 10.0.2.2 remote-as 65000
 neighbor 10.0.2.2 update-source loopback 1
 address-family ipv4
  no neighbor 10.0.2.1 activate
  no neighbor 10.0.2.2 activate
 exit
 address-family l2vpn evpn
  neighbor 10.0.2.1 activate
  neighbor 10.0.2.2 activate
  neighbor 10.0.2.1 send-community both
  neighbor 10.0.2.2 send-community both
 exit
 address-family ipv4 vrf Tenant1
 redistribute connected
```

```
advertise l2vpn evpn
exit
```

## Leaf3

```
hostname Leaf3
!
vlan database
vlan 101
vlan 101 overlay enable
vlan 103
vlan 103 overlay enable
exit
!
no spanning-tree enable
ip virtual-router mac 0000.c0ff.ee00
evpn virtual-mac enable
!
ip vrf Tenant1
vxlan vni 30001
rd 10.0.2.13:30001
route-target both 65000:30001 evpn
!
overlay
vlan 101 vni 20101
vlan 101 mac-address-tunnel learning-disable
vlan 103 vni 20103
vlan 103 mac-address-tunnel learning-disable
!
evpn
vni 20101
rd 10.0.2.13:20101
route-target both 65000:20101
route-target export 65000:30001
vni 20103
rd 10.0.2.13:20103
route-target both 65000:20103
route-target export 65000:30001
!
interface loopback 0
description VTEP tunnel source
ip address 10.0.1.13/32
!
interface loopback 1
description MP-BGP peering loopback
ip address 10.0.2.13/32
!
interface eth-0-21
no switchport
ip address 10.0.131.13/24
description Core link to Spine1:eth-0-21
ip ospf network point-to-point
vxlan uplink enable
mtu 9216
no shutdown
!
interface eth-0-22
no switchport
ip address 10.0.132.13/24
description Core link to Spine2:eth-0-23
ip ospf network point-to-point
vxlan uplink enable
```

```
mtu 9216
no shutdown
!
interface eth-0-1
switchport
switchport mode trunk
switchport trunk allowed vlan add 101,103
!
interface nve 0
source loopback 0
member vni 20101
member vni 20103
member vni 30001 associate-vrf
!
interface vlan 101
ip vrf forwarding Tenant1
ip address 192.168.101.251/24
ip virtual-router address 192.168.101.1
overlay distributed-gateway enable
overlay host-collect enable
!
interface vlan 103
ip vrf forwarding Tenant1
ip address 192.168.103.251/24
ip virtual-router address 192.168.103.1
overlay distributed-gateway enable
overlay host-collect enable
!
router ospf 65000
router-id 10.0.2.13
network 10.0.0.0/16 area 0
!
router bgp 65000
bgp router-id 10.0.2.13
max-paths ebgp 8
max-paths ibgp 8
neighbor 10.0.2.1 remote-as 65000
neighbor 10.0.2.1 update-source loopback 1
neighbor 10.0.2.2 remote-as 65000
neighbor 10.0.2.2 update-source loopback 1
address-family ipv4
no neighbor 10.0.2.1 activate
no neighbor 10.0.2.2 activate
exit
address-family l2vpn evpn
neighbor 10.0.2.1 activate
neighbor 10.0.2.2 activate
neighbor 10.0.2.1 send-community both
neighbor 10.0.2.2 send-community both
exit
address-family ipv4 vrf Tenant1
redistribute connected
advertise l2vpn evpn
exit
```

### Spine1

```
hostname Spine1
!
evpn
!
interface loopback 1
```

```
description MP-BGP peering loopback
ip address 10.0.2.1/32
!
interface eth-0-23
no switchport
ip address 10.0.111.1/24
description Core link to Leaf1:eth-0-22
ip ospf network point-to-point
mtu 9216
no shutdown
!
interface eth-0-24
no switchport
ip address 10.0.121.1/24
description Core link to Leaf2:eth-0-22
ip ospf network point-to-point
mtu 9216
no shutdown
!
interface eth-0-21
no switchport
ip address 10.0.131.1/24
description Core link to Leaf3:eth-0-21
ip ospf network point-to-point
mtu 9216
no shutdown
!
router ospf 65000
router-id 10.0.2.1
network 10.0.0.0/16 area 0
!
router bgp 65000
bgp router-id 10.0.2.1
max-paths ebgp 8
max-paths ibgp 8
neighbor 10.0.2.11 remote-as 65000
neighbor 10.0.2.11 update-source loopback 1
neighbor 10.0.2.12 remote-as 65000
neighbor 10.0.2.12 update-source loopback 1
neighbor 10.0.2.13 remote-as 65000
neighbor 10.0.2.13 update-source loopback 1
address-family ipv4
no neighbor 10.0.2.11 activate
no neighbor 10.0.2.12 activate
no neighbor 10.0.2.13 activate
exit
address-family l2vpn evpn
neighbor 10.0.2.11 activate
neighbor 10.0.2.12 activate
neighbor 10.0.2.13 activate
neighbor 10.0.2.11 send-community both
neighbor 10.0.2.12 send-community both
neighbor 10.0.2.13 send-community both
neighbor 10.0.2.11 route-reflector-client
neighbor 10.0.2.12 route-reflector-client
neighbor 10.0.2.13 route-reflector-client
exit
```

Spine2

```
hostname Spine2
!
```

```
evpn
!
interface loopback 1
  description MP-BGP peering loopback
  ip address 10.0.2.2/32
!
interface eth-0-21
  no switchport
  ip address 10.0.112.2/24
  description Core link to Leaf1:eth-0-21
  ip ospf network point-to-point
  mtu 9216
  no shutdown
!
interface eth-0-22
  no switchport
  ip address 10.0.122.2/24
  description Core link to Leaf2:eth-0-21
  ip ospf network point-to-point
  mtu 9216
  no shutdown
!
interface eth-0-23
  no switchport
  ip address 10.0.132.2/24
  description Core link to Leaf3:eth-0-22
  ip ospf network point-to-point
  mtu 9216
  no shutdown
!
router ospf 65000
  router-id 10.0.2.2
  network 10.0.0.0/16 area 0
!
router bgp 65000
  bgp router-id 10.0.2.2
  max-paths ebgp 8
  max-paths ibgp 8
  neighbor 10.0.2.11 remote-as 65000
  neighbor 10.0.2.11 update-source loopback 1
  neighbor 10.0.2.12 remote-as 65000
  neighbor 10.0.2.12 update-source loopback 1
  neighbor 10.0.2.13 remote-as 65000
  neighbor 10.0.2.13 update-source loopback 1
  address-family ipv4
    no neighbor 10.0.2.11 activate
    no neighbor 10.0.2.12 activate
    no neighbor 10.0.2.13 activate
  exit
  address-family l2vpn evpn
    neighbor 10.0.2.11 activate
    neighbor 10.0.2.12 activate
    neighbor 10.0.2.13 activate
    neighbor 10.0.2.11 send-community both
    neighbor 10.0.2.12 send-community both
    neighbor 10.0.2.13 send-community both
    neighbor 10.0.2.11 route-reflector-client
    neighbor 10.0.2.12 route-reflector-client
    neighbor 10.0.2.13 route-reflector-client
  exit
```

## Источники дополнительной информации: документация и стандарты

[Портал информации о телеком продуктах Aquarius \(net.aq.ru\)](http://net.aq.ru)

[Virtual eXtensible Local Area Network \(VXLAN\) RFC 7348](#)

[BGP MPLS-Based Ethernet VPN \(RFC 7432\)](#)

[A Network Virtualization Overlay Solution Using EVPN \(RFC 8365\)](#)

[Integrated Routing and Bridging in EVPN \(RFC 9135\)](#)

[IP Prefix Advertisement in EVPN \(RFC 9136\)](#)

[EVPN Interoperability Modes \(draft-ietf-bess-evpn-modes-interop-02\)](#)

[Use of BGP for Routing in Large-Scale Data Centers RFC 7938](#)