Инструкция по планированию ресурсов

ALD Pro

Содержание

От работы доменных служб зависит надежность работы всей ИТ инфраструктуры, поэтому при планировании домена администраторы должны действовать проактивно и применять различные способы вертикального и горизонтального масштабирования, чтобы исключить проблемы недостаточной производительности. В настоящей инструкции представлены рекомендации, которые помогут вам в решении указанной задачи.

1 Вертикальное масштабирование

Вертикальным масштабированием называют повышение производительности системы за счет повышения производительности отдельного узла путем выделения серверу дополнительных вычислительных ресурсов — оперативной памяти, потоков центрального процессора и т.п.

1.1 Объем памяти

В работе службы каталога преобладают операции чтения, поэтому в производительности контроллера решающую роль играет достаточный объем оперативной памяти, чтобы контроллер мог обрабатывать запросы без обращения к медленным дискам.

Файлы каталога расположены в папке /var/lib/dirsrv, после развертывания первого контроллера размер базы составляет порядка 45 Мб и увеличивается по мере создания объектов, в среднем по 30 КБ на каждую дополнительную учетную запись пользователя. Таким образом размер файлов на диске можно рассчитать математически: для нужд самой операционной системы контроллеру следует выделить порядка 30 ГБ, а под хранение каталога для упрощения расчетов возьмем по 1ГБ на каждые 35 тысяч объектов. Расчетное значение должно составлять не более 40% от доступного пространства.

HDD, ГБ = 30 + $N_{объектов}$ / 35 000

Для загрузки каталога требуется больше оперативной памяти, чем на диске, т. к. для ускорения операций поиска служба ns-slapd индексирует данные каталога. Пустая база данных занимает в памяти порядка 65 Мб и это значение растет по мере увеличения числа объектов, в среднем по 50 КБ на каждую дополнительную учетную запись. По ходу работы службы каталога для ускорения операций чтения происходит кеширование запросов, и потребление оперативной памяти возрастает до 100 КБ на каждую учетную запись каталога. Таким образом минимально необходимый объем оперативной памяти можно рассчитать математически: для нужд самой операционной системы контроллеру следует выделить порядка 2-3 ГБ, а для работы с каталогом в целях упрощения расчетов возьмем по 1ГБ на каждые 10 тысяч объектов:

ОЗУ, ГБ = 3 + $N_{объектов}$ / 10 000

Приведем несколько примеров:

- Для работы с каталогом, в котором содержится 10 000 пользователей и 100 групп, контроллеру нужно выделить порядка 4ГБ ОЗУ

ОЗУ = 3 + 10 100 объектов / 10 000 ~ 4 Гб ОЗУ

- Для работы с каталогом, в котором 100 000 пользователей и 30 000 групп, контроллеру нужно выделить до 16 ГБ ОЗУ

ОЗУ = 3 + 130 000 объектов /10 000 ~ 16 Гб ОЗУ.

1.2 Количество потоков центрального процессора

Если требуемый объем оперативной памяти зависит от размера базы данных, то количество ресурсов центрального процессора определяется тем, сколько пользователей должен обслуживать конкретный контроллер. Например, в базе может быть 100 тысяч пользователей, но нагрузка по их обслуживанию может распределяться между 20 репликами, тогда на каждый контроллер будет приходиться не более 5 тысяч обслуживаемых пользователей.

Приложения имеют разный сценарий взаимодействия с каталогом, какие-то нагружают его больше, какие-то меньше, таким образом у каждой организации получается свой уникальный профиль использования ресурсов этой службы, и предложить какой-то

универсальный нормативный показатель на одного сотрудника не представляется возможным. Поэтому администраторам следует рассчитать данный показатель для своей организации самостоятельно, достигая использования ЦПУ на уровне 40% от максимума в пиковые периоды, а отправной точкой может быть выделение одного потока для работы операционной системы и еще по одному потоку на каждую тысячу обслуживаемых пользователей.

ЦПУ, потоков = 1 + N $_{\text{обслуживаемых пользователей}}$ / 1 000

Приведем несколько примеров:

- Для обслуживания 2 - 3 тысяч пользователей контроллеру домена нужно выделить порядка 4 потоков

ЦПУ = 1 + 3 000 / 1 000 = 4

Компания RedHat в своей документации рекомендует использовать контроллеры домена именно такой производительности, но эта рекомендация не подойдет крупным организациям, в штате которых работает несколько сотен тысяч сотрудников, т. к. контроллер, обслуживающий 10 тысяч сотрудников будет в три раза эффективнее использовать оперативную память в пересчете на активного пользователя, чем контроллер, обслуживающий 3 тысячи сотрудников.

- Для обслуживания 10 тысяч пользователей контроллеру домена нужно выделить 11 ядер.

ЦПУ = 1 + 10 000 / 1 000 = 11

Повторимся, что данные оценки сильно зависят от сценария использования службы каталога, поэтому на такое количество пользователей мы рекомендуем выделять порядка 8-12 потоков.

- Если под контроллер домена выделить физический сервер с двумя процессорами Xeon Silver (48 потоков), то теоретически он сможет обслуживать 47 тысяч пользователей

ЦПУ = 1 + 47 000 / 1 000 = 48

При планировании таких контроллеров следует принимать во внимание ограничения по пропускной способности сети. Для обслуживания такого числа пользователей потребуется не менее гигабитного интерфеса.

1.3 Пропускная способность сети

При планировании домена следует учитывать два основных вида трафика:

- трафик клиентских запросов (Kerberos и LDAP запросы к каталогу)
- трафик репликации (по стандартному LDAPS протоколу)

1.3.1 Клиентский трафик

В процессе обслуживания клиентов контроллеры домена, как правило, получают небольшие входящие запросы на предоставление относительно больших объемов данных, поэтому на контроллерах обычно преобладает исходящий трафик.

Любые оценки являются субъективными, но в качестве общих рекомендаций следует исходить из того, что при количестве пользователей до 5 000 будет достаточно интерфейса с пропускной способностью в 100 Мбит/с, а если предполагается обслуживать больше пользователей установите на контроллере интерфейс с пропускной способностью в 1 Гбит/с. В результате оптимизации нужно добиться, чтобы показатель использования сети не выходил за пределы 40% от максимума в пиковые периоды.

Учитывая то, что внешние каналы связи редко достигают гигабитных скоростей и активно используются пользователями для выхода в Интернет, в каждом офисе рекомендуется

создавать отдельный сайт и размещать в нем один-два контроллера для локальной обработки клиенстких запросов. Такой подход к организации домена позволит разгрузить VPN туннели между офисами и использовать эти каналы связи только для репликации.

1.3.2 Трафик репликации

В ходе репликации контроллеры обмениваются только новыми и измененными данными, поэтому объем трафика оказывается небольшим, и его довольно просто прогнозировать, например,

- при добавлении 1000 пользователей трафик репликации окажется в два раза больше, чем при добавлении 500 пользователей;
- если вы устанавливаете новый контроллер в удаленном офисе с подключением в 5Мбит/с, то приготовьтесь, что на репликацию базы данных размером 1.5 ГБ потребуется не менее часа.

Крайне важно исключить долгую работу контроллера в автономном режиме, т. к. после этого может потребоваться повторная инициализация контроллера с загрузкой всего каталога.

2 Горизонтальное масштабирование

Горизонтальным масштабированием называют повышение производительности системы за счет увеличения количества вычислительных узлов, обслуживающих клиентов, без изменения их производительности. Например, если предполагается использовать контроллеры с 8-12 потоками ЦПУ, которые могут обслуживать по 10 тысяч пользователей, то в организации с штатом в 30 тысяч сотрудников потребуется не менее 3 контроллеров такой производительности:

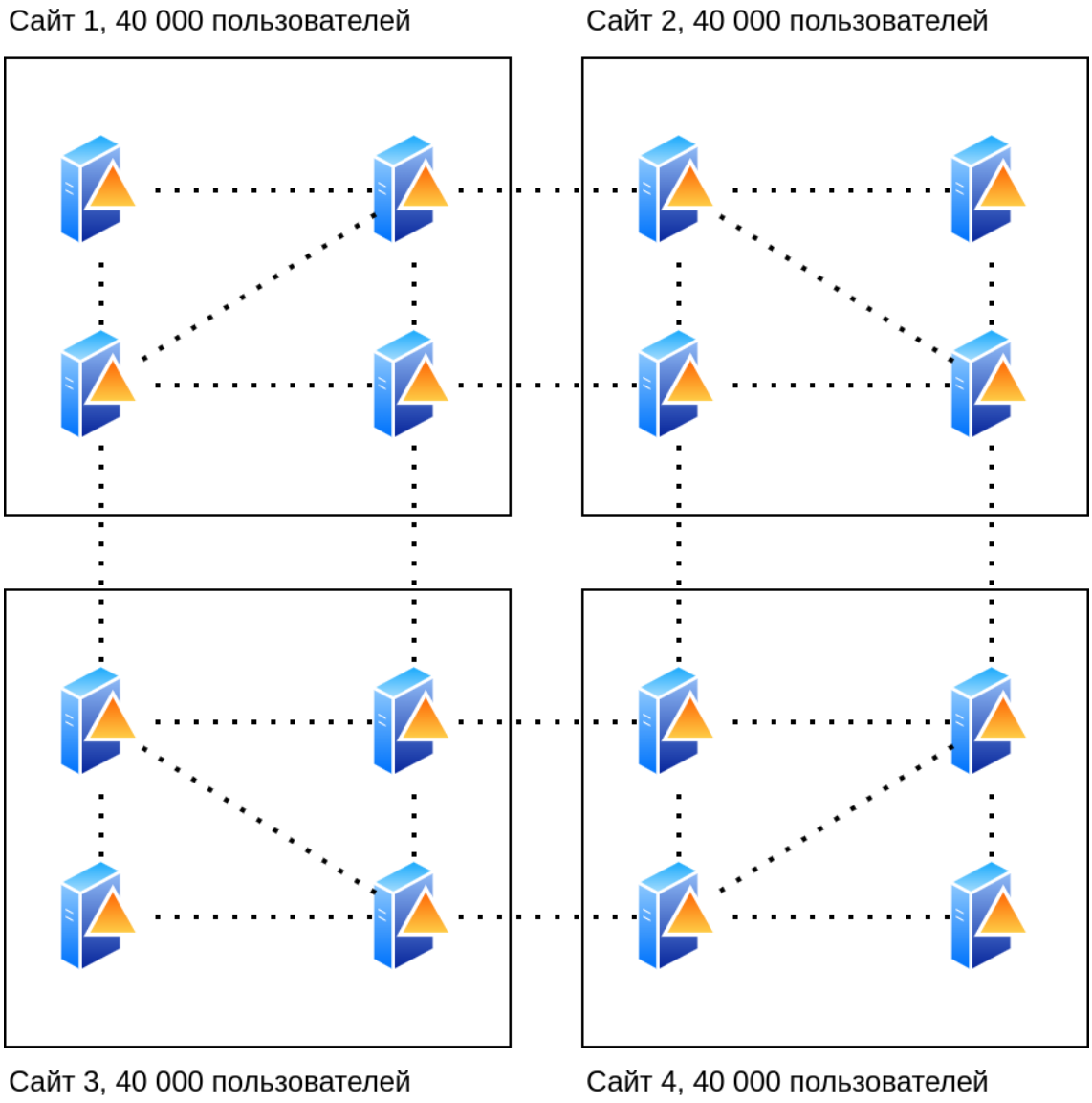
N контроллеров = 30 тыс. сотрудников / 10 тыс. на контроллер ~ 3 контроллера

Синхронизация данных между контроллерами обеспечивается процессом репликации, в FreeIPA она происходит в мультимастер-режиме, т. е. изменения можно вносить на любом контроллере. Балансировка нагрузки и отказоустойчивость обеспечивается через DNS — клиентам выдается до трех DNS серверов, каждый из которых может принадлежать определенному сайту и выдавать SRV записи в соответствии принадлежностью служб к соответствующим сайтам.

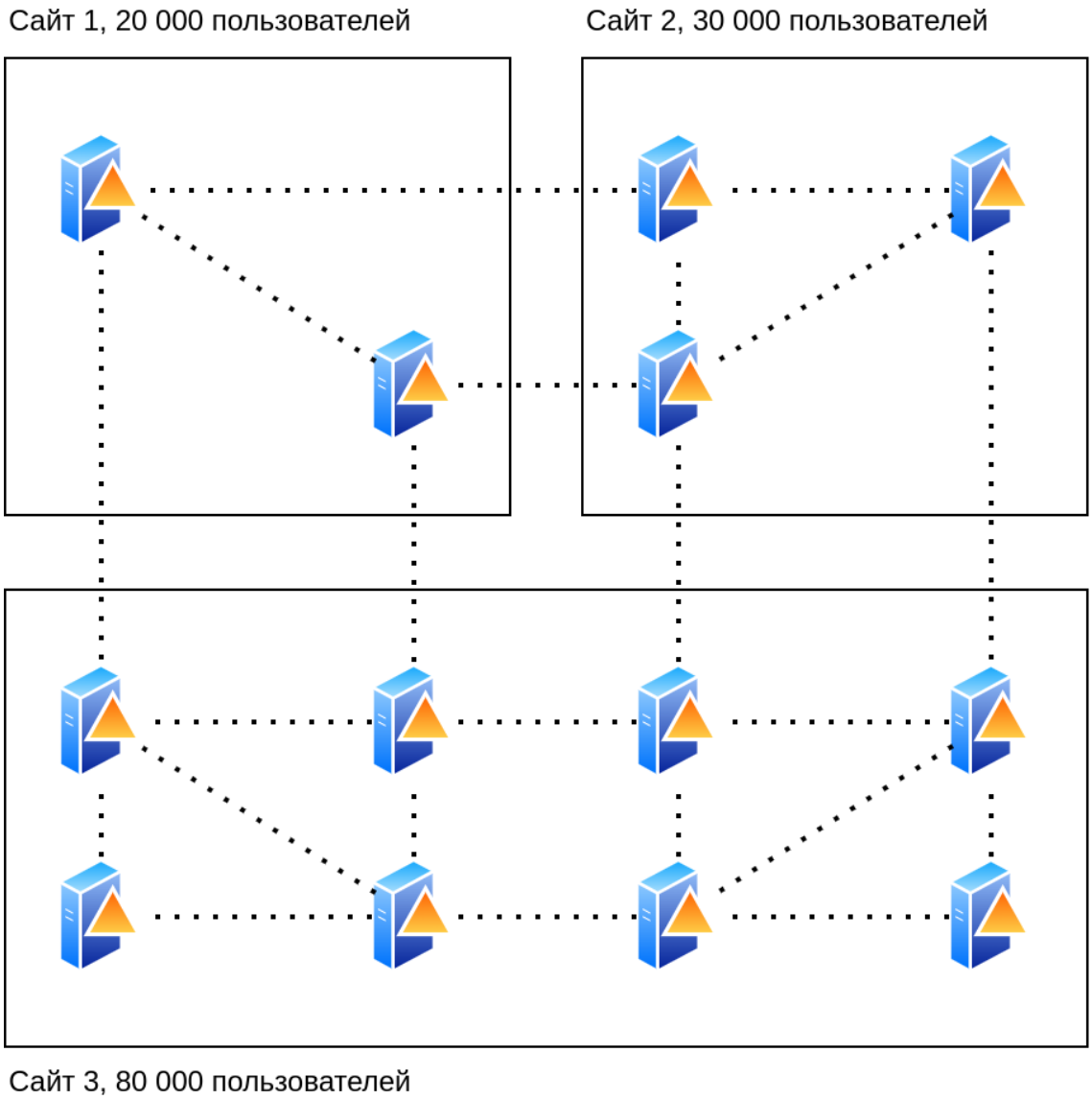
Компания RedHat в документации по своему продукту IdM, построенному на базе FreeIPA, утверждает, что «поддерживает до 60 реплик» в связи с чем довольно часто можно столкнуться с заблуждением, что это число является ограничением сверху (Red Hat supports environments with up to 60 replicas.). На самом деле репликация обеспечивается одним из модулей 389 Directory Server, который не накладывает ограничений по количеству реплик, и в сети Интернет можно найти упоминания об инсталляциях даже на тысячу серверов. При этом вы должны понимать, что увеличение числа серверов будет неминуемо приводить к усложнению топологии, удлинению максимального пути репликации и появлению соответствующих эффектов.

2.1 Топология соглашений о репликации

В качестве примера обычно приводят топологию из четырех сайтов, в каждом из которых по четыре контроллера, в результате чего получается красивая схема, которая вводит администраторов в заблуждение, что в сайте может быть не более четырех контроллеров.



В реальной жизни редко бывает, чтобы численность сотрудников в каждой локации была одинаковой, поэтому количество контроллеров может быть как больше так и меньше. Приведем пример с тремя сайтами численностью 20, 30 и 80 тысяч человек.



2.2 Рекомендации по построению топологии

При планировании топологии следует учитывать несколько рекомендаций:

- Для обеспечения надежности на уровне двух девяток устанавливайте в каждом сайте не менее двух контроллеров домена.

Допустим, сервер работает с надежностью в 90%, тогда при добавлении еще одного такого сервера мы получаем, что в 90% случаев от оставшихся 10% система будет оставаться все так же работоспособной. Итоговая надежность повысится до двух девяток 0.9 + 0.1* 0.9 = 0.99

Максимальное количество контроллеров на сайт не лимитируется, но в настройках рабочих станций вы сможете задать только три DNS сервера, поэтому стандартными средствами можно обеспечить надежность до трех девяток, а дополнительные контроллеры будут использоваться только для распределение нагрузки.

- Для обеспечения надежности репликации внутри сайта у каждого контроллера должно быть от 2 до 4 соглашений о репликации.

Создавать более четырех соглашений не рекомендуется, т. к. это уже не дает существенного прироста надежности, но замедляет репликацию (контроллер не принимает запросы на изменение от нескольких реплик одновременно, в один момент времени работает только одно соглашение, а остальные простаивают, поэтому увеличение количества соглашений в два раза приведет к тому что репликация по этим соглашениям будет происходить в два раза реже). Но создавать более 4 соглашений оправдано, если, например, это дает сокращение максимального пути репликации в используемой топологии.

- Для обеспечения надежности репликации между сайтами на каждом из них рекомендуется создавать от 2 до 4 межсайтовых соглашений о репликации.

При оптимизации топологии репликации вашей задачей является также уменьшение количества прыжков (hops), за которые изменения могут быть распространены между контроллерами. В классическом примере из четырех сайтов можно заметить, что между сайтами 1 и 4 нет прямых соглашений, поэтому для доставки изменений между крайними контроллерами потребуется не менее 5 прыжков.

- Для максимальной производительности репликации под эту функцию можно выделить отдельные контроллеры и снять с них обслуживание обычных пользователей путем переключения в режим скрытой реплики (hidden replica). Данная функция появилась во FreeIPA с версии 4.6

3 Практический кейс

Исходные требования:

- 1 млн. объектов в каталоге (650 тыс. пользователей, 350 тыс. групп)
- 2 тыс. обслуживаемых пользователей

Расчеты

Для размещения файлов операционной системы и каталога на жестком диске потребуется не менее 60 ГБ HDD

HDD = 30 + 1 млн. / 35 000 ~ 60 ГБ

Для того, чтобы контроллер домена смог разместить каталог с 1 млн. объектов в оперативной памяти, ему потребуется не менее 100 ГБ ОЗУ:

ОЗУ = 3 + 1 млн. / 10 000 ~ 100 ГБ

Если к контроллеру будут обращаться 2 тысячи пользователей то ему потребуется не менее 3 ядер:

ЦПУ = 1 + 2 000 / 1 000 = 3

Конкретно к этому серверу будет обращаться 2 тыс. пользователей, поэтому достаточно одной виртуальной машины. Но если нужно будет обеспечить единовременную работу всех 600 тысяч пользователей, то потребуется не менее 300 таких контроллеров и более целесообразным видится увеличение производительности контроллеров до 10 тысяч пользователей (8-12 ЦПУ), тогда количество контроллеров можно будет сократить до 60 единиц.

Следует учесть, что объекты в каталоге создаются в среднем со скоростью один объект в секунду и это значение растет по мере увеличения количества объектов в каталоге, поэтому на создание 1 млн. объектов уйдет не менее 11 дней. Ускорить процесс можно увеличением числа потоков, но зависимость будет нелинейной, т. к. потоки будут конкурировать за общие ресурсы. Дополнительного ускорения процедуры можно достичь путем создания пользователей на нескольких контроллерах одновременно, т. к. при репликации вычислительных ресурсов требуется меньше, чем при создании объектов.

Для тестирования остальных подсистем в соответствии с типовой схемой нужно дополнительно 8 виртуальных машин: 2 CPU, 4 ГБ ОЗУ, 100 ГБ HDD (DHCP, установка ОС по сети, сервер печати, сервер общего доступа к файлам, сервер аудита, сервер мониторинга, сервер репозиториев, клиентская машина).