

Доступ к файлам для объектного хранилища

Окончание. Начало в № 5/2020

Том Кофлин

В первой части статьи автор очертил проблемы, связанные с хранением больших объемов цифровых данных, дал краткие определения типам хранилищ, сделал сравнение иерархического и объектного хранения, а также обосновал необходимость в гибридной системе хранения.

Гибридное и публичное облачное хранение

К числу некоторых общих случаев использования объектного хранения относятся такие, как хранение больших данных, доступ к данным через web-приложения и поддержание резервных архивов. Объектное хранилище часто развертывается в больших ЦОД, но его можно интегрировать и в локальные системы хранения. Эти локальные системы объектного хранения могут быть подключены к Интернету, обеспечивая частное облако с ограниченным доступом. Все более широкой практикой становится хранение данных как в частных, так и в публичных облаках с перемещением данных между ними (этот подход часто называют многооблачным).

Гибридный и облачный варианты хранения полезны для многих приложений. Сюда входят рабочие процессы, распределенные между географически удаленными друг от друга комплексами, работающими с одними и теми же данными, и общий дистанционный доступ к таким данным. Доступ к данным в публичном облаке позволяет организовать специализированные сервисы, которые слишком накладно разворачивать на базе локальных средств. Куда выгоднее делать это, арендуя облачные сервисы, такие как AI-обучение, доставка контента и рендеринг видео.

Многие люди также копируют данные или архивируют их в облако на случай восстановления после катастроф или для долговременного размещения. Гибридные облака к тому же позволяют задействовать облачные вычислительные ресурсы для выполнения операций в случаях, когда локальные ресурсы оказываются полностью загруженными.

Есть немало организаций, которые прибегают к аренде публичных облачных сервисов для того, чтобы заменить капитальные вложения

эксплуатационными расходами либо минимизировать необходимость в локальной IT-поддержке. Есть еще и такие сервисы, поддерживаемые гибридной и облачной моделью, как копирование данных, восстановление после катастроф, перенос данных с сервера на сервер и масштабирование без остановки работы, организация многоуровневого хранения, видеоархив систем наблюдения, многоуровневое архивирование в облаке, географическая синхронизация данных, совместный доступ к файлам, перенос данных с ленты в облако и дистанционный доступ к данным из офисов.

Генерирование львиной доли данных будет по-прежнему происходить вне облака – на конечных устройствах, в локальных комплексах и с помощью множества сенсоров, которыми оснащены устройства, относящиеся к Интернету вещей, включая камеры видеонаблюдения, различные датчики и автомобили. Использование гибридных облаков в обозримом будущем продолжится, а полностью облачные рабочие процессы пока будут единичными.

Хотя облако обеспечивает практически безграничное наращивание объемов хранения, оказалось, что перенос в облака имеющихся серверов и приложений проходит медленно, сложно и для многих пользователей проблематично. Перенос столь больших объемов данных в облако требует много времени и больших усилий. К тому же, как только данные перенесены, их извлечение сопровождается большими расходами вследствие оплаты за это. Есть потребность в новой гибридной архитектуре, способной обеспечить интеграцию и перенос крупных облачных хранилищ на базе Windows NTFS при одновременном снижении сложности и стоимости.

Многие приложения были созданы для работы с иерархическими файловыми системами, встроенными в хранилища типа DAS, NAS и SAN. Эти файловые системы имеют иерархические индексы, содержащие папки, подпапки и файлы. Доступ к файлам осуществляется по фиксированному пути. Файловые системы опираются на специфические адреса и места расположения. Следовательно, могут быть сложности с перемещением данных без риска разрыва связей.

Иерархические файловые системы могут также содержать меньше средств собственной защиты от повреждений. Если часть иерархии файловой системы повреждается, доступ к некоторым данным может стать невозможным. Поэтому есть необходимость в методе, дающем

надежный доступ к файлам, находящимся в облачном объектном хранилище.

IT-менеджеры должны найти способ экономически эффективно управлять размещением данных, использованием объемов хранения и расходами. Чтобы сделать это, организации сейчас внедряют разные системы – специализированные облачные шлюзы, синхронизацию файлов и общий доступ к ним, NFS/SMB, уровни абстракции объектов, резервное копирование и архивирование, географическую синхронизацию, восстановление после катастроф и др. Достоинства этих подходов вытекают из особенностей их предназначения, а в числе недостатков зачастую входят отсутствие системной интеграции, множество поставщиков, закрытые форматы, привязка к тому или иному производителю, неэффективный рабочий процесс, несовместимость на уровне приложений и т.д.

Для многих организаций, желающих в полной мере использовать потенциал облачного хранения, выбором является применение облачного шлюза (cloud gateway). Облачный шлюз – это гибридная облачная платформа подключения к облаку и пример платформы как сервиса – PaaS (Platform as a Service). Шлюз защищенно подключает локальные файловые сети к нескольким провайдерам облачных сервисов, создавая гибридное облако, которое содержит хранилища двух типов – локальное и в публичном облаке.

Сейчас уже есть немало облачных шлюзов. Многие серверы-шлюзы представляют объектное хранилище файловой системе как новую точку подключения, которую надо интегрировать в существующие процедуры и рабочие процессы. В этом случае производительность и совместимость с протоколом SMB/CIFS/NFS нередко не оптимальны из-за дополнительной нагрузки и сложности. Такие облачные шлюзы различаются по способности обеспечить быстрый доступ к локальным и дистанционным данным, но обычно требуют дополнительного специализированного оборудования, а в их основу положена идея о том, что данные постоянно перемещаются из хранилища на базе файловой системы в объектное хранилище.

Одним из важных требований к хорошему облачному шлюзу является его способность работать с известными облачными провайдерами (Amazon, Azure, IBM, Google, Backblaze, Wasabi и др.), включая интеграцию с любыми уровнями хранения в этих облачных сервисах и возможности управления облачным хранением. Такой шлюз должен еще способствовать поддержа-

нию синхронизации серверов, как локальных, так и расположенных в любой точке мира, чтобы контент, записанный в одном сервере, появлялся как локальный файл на всех остальных серверах. Облачный шлюз к тому же должен помогать организациям в переносе имеющейся у них инфраструктуры в облако по их собственной графике.

Общие вызовы, связанные с хранением данных

Есть несколько проблем и вызовов, связанных с организацией хранения данных:

- ♦ в большинстве файловых серверов на сегодня используются Windows OS и NTFS, а также поддерживаются существующие приложения и рабочие процессы на базе Windows;
- ♦ увеличение объемов данных требует новых подходов к управлению хранением;
- ♦ стоимость, масштабируемость и гибкость объектного и облачного хранения делает их привлекательными для IT-менеджеров;
- ♦ объектное и облачное хранилища несовместимы с этими имеющимися приложениями и рабочими процессами;
- ♦ полный переход в облако и к инфраструктуре на базе сервисов разрушителен для нынешней бизнес-модели и требует переобучения персонала.

Если бы все эти проблемы можно было решить с помощью единой системы управления хранением, это могло бы помочь преобразовать компанию, дав ей конкурентные преимущества, снизив расходы и уменьшив сложность работы с одновременным расширением возможностей.

Tiger Bridge

Tiger Bridge – это чисто программная система управления данными и хранением, облачный шлюз и HSM-менеджер, позволяющий организовать гибридную и полностью облачную работу, а также обеспечивающий множество других сервисов. Это своего рода соединитель на базе определенных правил, объединяющий NTFS с облаком, NAS и ленточным хранением в общую глобальную файловую систему. Tiger Bridge уникален в том, что он придает объектному хранилищу возможность обычного доступа на уровне файлов. Bridge также открывает постоянный файловый доступ к данным в рамках имеющейся файловой системы, позволяя использовать нынешние рабочие процессы с опорой на существующие инфраструктуры, а также копировать данные и размещать их на разных уровнях хранения в нескольких местах. Tiger Bridge – это единое средство управления хранением, оптимальное для решения разных задач, таких как синхронизация файлов и общий доступ к ним, наращивание емкости, HDM, восстановление после сбоев и других.

Tiger Bridge запускается прямо на Windows-сервере и интегрируется в имеющуюся файловую систему NTFS на уровне ядра. Он гарантирует строгое поддержание защиты активной директории. При подключении к получателю через RESTful/S3 API данные шифруются непосредственно в сервере через https, прежде чем будут переданы в облако. Таким образом обеспечивается защита данных.

Когда выполняется перенос данных из локального сервера в объектное хранилище, создаются так называемые файлы-заглушки (Stub files), представляющие любые перемещенные файлы. Эти файлы-заглушки содержат все метаданные из исходного файла, а доступ к ним для пользователей и приложений выглядит как доступ к любому другому файлу. Во многих случаях файл-заглушка обеспечивает достаточное количество метаданных для нормального функционирования многих процессов несмотря на отсутствие в нем основных данных, благодаря чему уменьшается задержка. Когда происходит обращение к файлу-заглушке, исходные данные автоматически и прозрачно переносятся в локальную файловую систему и передаются в запросившее их приложение.

Tiger Bridge интегрируется с платформами основных провайдеров облачных сервисов, включая внутриоблачные уровни – активный (Hot), неактивный (Cool) и архивный (Archive), и позволяет сэкономить на расходах, связанных с извлечением данных из облака за счет просмотра файлов-заглушек в локальных папках, а не полных файлов в облаке.

Некоторые варианты применения Tiger Bridge

Доступ на уровне файлов к объектному и облачному хранилищу открывает возможности для ряда приложений. Tiger Bridge обеспечивает управление данными и хранением в рамках многооблачного гибридного хранилища, как показано на рис. 2.

Tiger Bridge еще поддерживает работу с DAS, ленточными приводами, локальным объектным хранилищем, а также с разными уровнями в Azure и AWS – активным, неактивным и архивным. Шлюз динамически отслеживает операции с локально сохраненными данными, держа активные файлы в локальном хранении, чтобы

доступ к ним был максимально быстрым и с наивысшей пропускной способностью.

Неактивные файлы автоматически или вручную переносятся в менее дорогостоящие уровни хранения (например, в публичное облако). Небольшие файлы-заглушки остаются в локальном хранилище, где они видны приложениями и файловой системе. Когда происходит обращение к одному из этих файлов-заглушек, данные автоматически извлекаются по запросу.

Благодаря переносу неактивных данных в локальное хранилище высвобождается пространство. Использование файлов-заглушек гарантирует, что файловые рабочие процессы могут получать доступ к файлам, сохраненным в облаке и, как следствие, петабайты данных становятся доступными для локальных и совместных проектов. Tiger Bridge обеспечивает оптимизацию хранения, управление иерархическим хранением, резервное копирование и архивирование, геолокацию, синхронизацию файлов и общий доступ к ним. Также к функциям Tiger Bridge относится быстрое восстановление после сбоев. Благодаря функциям Geosynchronization, имеющимся у шлюза, файлы-заглушки заблаговременно создаются и готовы к запуску на сервере DR (Disaster Recovery).

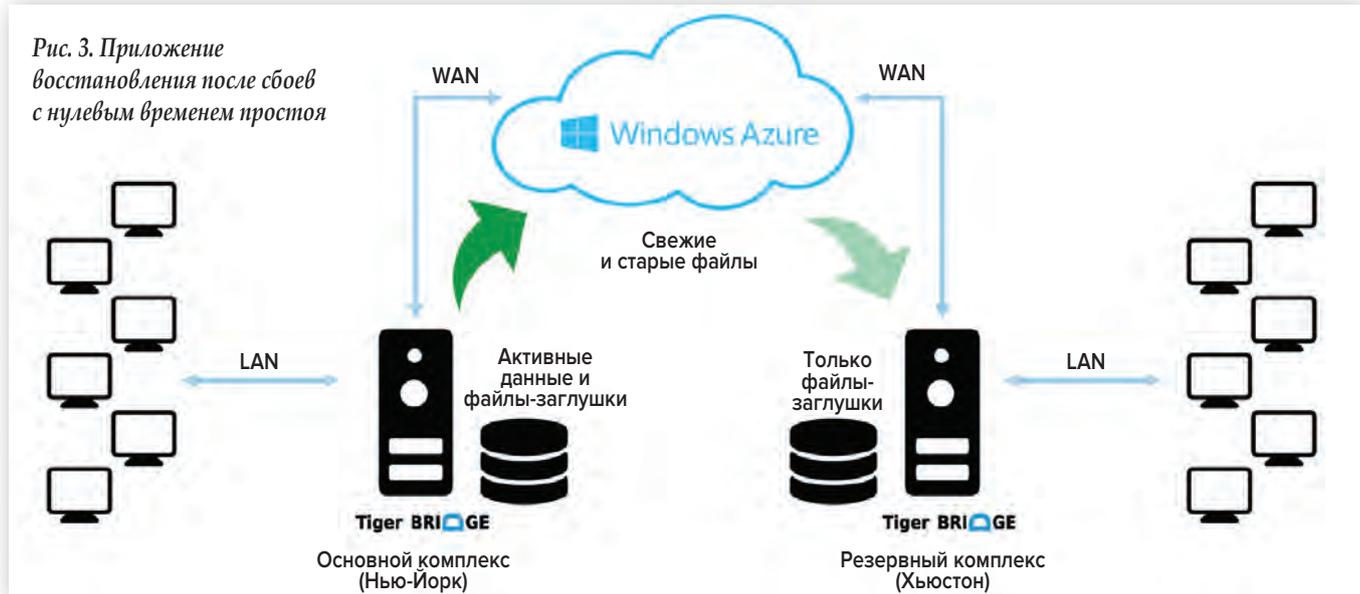
Tiger Bridge, запущенный в основном комплексе, автоматически копирует вновь созданный контент в облако. Соответствующие файлы-заглушки появляются в файловой системе Tiger Bridge, запущенной в резервном комплексе (если оба комплекса подключены к одному и тому же объектному хранилищу). Поскольку резервный комплекс может мгновенно получить доступ к метаданным файла, а значит, и к основным данным файла, полное восстановление данных из облака может вообще не понадобиться в резервном комплексе, за счет чего снижается размер инфраструктуры хранения, требуемой для резервного комплекса (рис. 3).

Tiger Bridge также способен обеспечить недорогое хранение данных видеонаблюдения. Сейчас стало привычным, когда тысячи камер, находящихся в одной и той же зоне, отправляют потоки данных на несколько серверов записи. Вся эта масса видео не имеет никакой ценности до тех пор, пока не потребуется важное свидетельство, содержащееся в видеозаписи. За счет представления объектного хранилища как



Рис. 2. Управление данными и хранением с помощью Tiger Bridge

Рис. 3. Приложение восстановления после сбоев с нулевым временем простоя



расширения для локальной файловой системы NTFS Tiger Bridge позволяет приложениям видеонаблюдения легко взаимодействовать с объектным хранилищем (рис. 4).

Еще Tiger Bridge способен придать традиционным сетевым видеосерверам возможность записывать видеопотоки на локальный диск или в SAN перед тем, как скопировать их в объектное хранилище. Когда это объектное хранилище является дистанционным (в облаке), такой буфер гарантирует, что все видеосвидетельства записаны до того, как будут зашифрованы и перенесены в облако. Поскольку Tiger Bridge действует непосредственно в памяти сервера, https-шифрование происходит до того, как любые критически важные данные попадают в сеть, за счет чего достигается защита данных.

Как только данные оказываются в облаке, виртуальные машины могут оперировать ими, выполняя разные процедуры, такие как распознавание лиц и объектов.

Заключение

Требования к цифровому хранению данных растут, поскольку увеличиваются количество и размер цифровых объектов. Разным отраслям, в том числе и медиаиндустрии, требуются стратегии по экономически эффективному управлению и защите используемых в работе данных. Это привело к разработке объектного хранения, которое можно масштабировать для работы с огромным количеством объектов данных, а само объектное хранение стало основой большинства онлайн-облачных хранилищ.

Тем не менее многие распространенные приложения опираются на использование файлов и связанной с ними иерархической структуры данных. Использование этих приложений в связке с объектными хранилищами требует облачного шлюза, который может обеспечить доступ на уровне файлов к объектному хранилищу в облаке. Tiger Bridge представляет собой систему с функционалом облачного шлюза и управления хранением, которая образует из NTFS, NAS и ленточных систем единую глобальную файловую систему. Кроме того, Tiger Bridge предоставляет удобный подход к резервному копированию, архивированию и восстановлению после сбоев с помощью географической синхронизации и различных уровней переноса данных в облако.

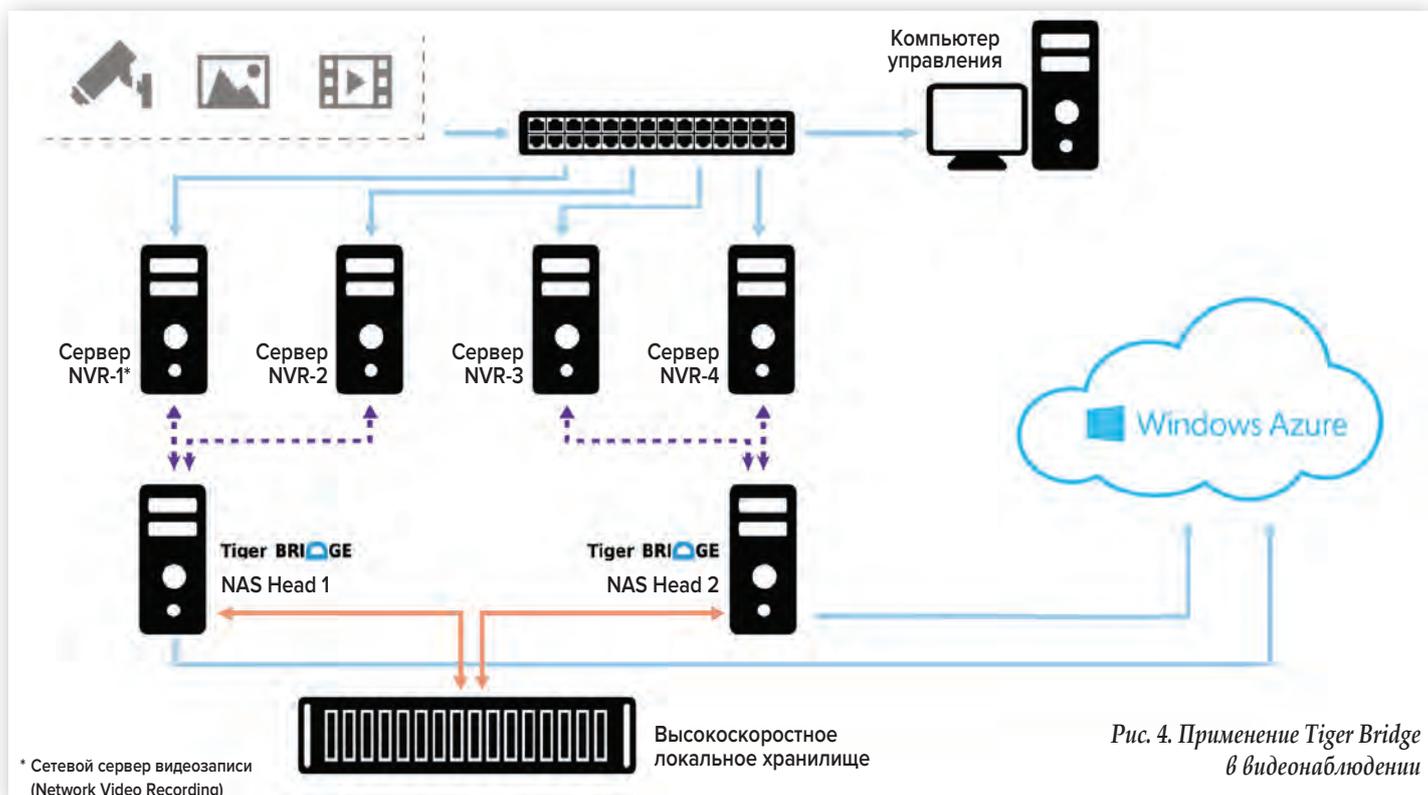


Рис. 4. Применение Tiger Bridge в видеонаблюдении