

# EMC Data Domain Boost: распределенная дедупликация

В середине мая 2010 г. компания EMC анонсировала новое ПО Data Domain® Boost, позволяющее переносить часть нагрузки, связанной с дедупликацией данных с систем хранения Data Domain на сервер резервного копирования. За счет этого скорость резервного копирования в среднем повышается на 50% и значительно сокращается нагрузка на серверы и сетевую инфраструктуру резервного копирования. В настоящее время DD Boost – интерфейс транспортировки данных и управления системами Data Domain – интегрирован с EMC NetWorker, а также Symantec NetBackup и Backup Exec.

## Введение

Приложения резервного копирования – критически важный компонент резервного копирования/восстановления (РКВ) и стратегий подготовки к аварийным ситуациям. Данные приложения часто управляют работой дисковых и ленточных носителей, на которых обычно хранятся в 5–10 раз больше данных, чем находится в первичной СХД того же центра обработки данных. Поскольку непрерывный рост объемов данных продолжается и все больше внимания уделяется эффективной защите данных и аварийному восстановлению, традиционные подходы к резервному копированию больше не приемлемы. Решения по резервному копированию и дедупликации на основе дисков становятся обязательными во многих подходах к управлению данными.

ПО EMC Data Domain Boost построено на основе открытого API, значительно расширяет возможности комплекса сервера резервного копирования, интегрированного с системой EMC Data Domain за счет частичного перераспределения процесса дедупликации на сервер резервного копирования, упрощает процесс аварийного восстановления и выступает в роли надежной основы для дополнительной интеграции между приложениями резервного копирования и системами Data Domain. При этом значительно сокращается нагрузка на серверы и сетевую инфраструктуру резервного копирования.

## Решение EMC Data Domain Boost

При “традиционной” архитектуре (без ПО DD Boost) медиа-серверы резервного копирования будут отправлять все данные, уникальные или избыточные, в систему Data Domain для осуществления дедупликации. При использовании ПО DD Boost (рис. 1), медиа-серверы и DD постоянно обмениваются информацией об уникальности каждого блока, и в систему Data Domain поступают только сжатые уникальные блоки (рис. 2). За счет этого агрегированная пропускная способность систем DD в среднем повышается на 50% – до 8,8 Тбайт/час, загрузка CPU медиа-серверов снижается на 20–40%, сетевой трафик между медиа-серверами и DD уменьшается на 80–99%, сокращая время готовности к аварийному восстановлению.

ПО DD Boost обеспечивает простоту эксплуатации, позволяя серверу резервного копирования управлять соединением между приложением резервного копирования и одной или несколькими системами Data Domain.

ПО DD Boost также позволяет увеличить скорость перезапуска и завершения неудачных операций резервного копирования. Поскольку по сети отправляются только уникальные данные, после перезапуска неудачного задания не нужно снова отправлять данные, которые уже отправлены в систему Data Domain для данного задания резервного копирования. Это приводит не только к существенному сокращению нагрузки на сеть, но также к повышению общей производительности для неудачных операций резервного копирования при повторной попытке.

В общем, ПО DD Boost позволяет увеличить совокупную производительность, существенно уменьшить окно резервного копирования и повысить эффективность использования сервера резервного копирования.

### Управление репликацией из консоли приложения резервного копирования

Технология репликации Data Domain с эффективным использованием сетевых ресурсов позволяет отправлять образы резервных копий из одной системы Data Domain в другую или несколько других

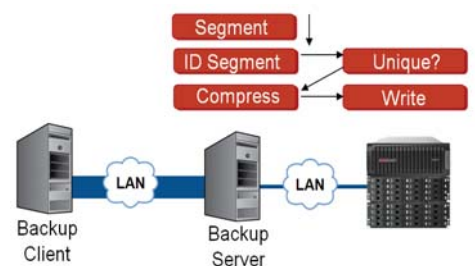


Рис. 2. При использовании ПО DD Boost РКВ-сервер и DD постоянно обмениваются информацией об уникальности каждого блока, и на DD пишутся только сжатые уникальные блоки.

систем, если они все настроены с помощью ПО DD Boost. Такой подход позволяет сократить необходимую полосу пропускания для репликации на 99%. Это существенно уменьшает время и полосу пропускания сети WAN, необходимые для создания нескольких резервных копий с целью консолидации дисков или аварийного восстановления. Использование ресурсов на серверах резервного копирования также уменьшается, поскольку теперь они отсутствуют в маршруте данных при создании и передаче дублированных резервных копий.

При помощи ПО DD Boost приложения резервного копирования могут контролировать репликацию данных между несколькими системами Data Domain и предоставлять администраторам резервного копирования единую точку управ-

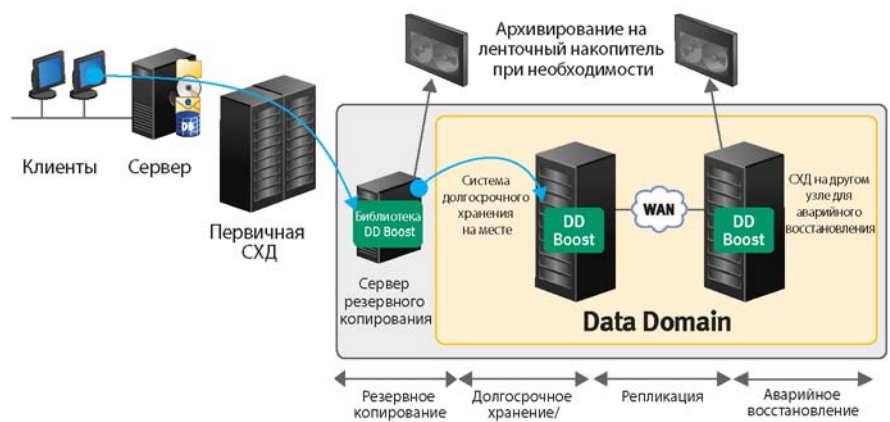


Рис. 1. Общая архитектура РКВ-решения с использованием ПО DD Boost, за счет которого удастся существенно снизить нагрузку на систему Data Domain и на сеть.

ления для отслеживания всех резервных копий и дублированных копий. Данная парадигма позволяет администраторам резервного копирования эффективно создавать копии для аварийного восстановления резервных копий по сети WAN при помощи программного обеспечения EMC Data Domain Replicator, а также следить за всеми копиями в каталоге приложения резервного копирования в целях простого аварийного восстановления. Благодаря этому, администраторы также могут управлять периодами долгосрочного хранения для различных копий по отдельности, что обеспечивает более гибкое внедрение средств аварийного восстановления.

Благодаря применению репликации Data Domain с эффективным использованием сети WAN ПО DD Boost позволяет устранить необходимость процедур резервного копирования и восстановления на основе ленточных накопителей на удаленных узлах. Восстановление из резервных копий на центральной площадке также упрощается, поскольку все копии отслеживаются в каталоге приложения резервного копирования. Поскольку ленточные накопители на удаленных узлах больше не требуются, целесообразно внедрять системы Data Domain с ПО DD Boost в качестве ключевых, основополагающих элементов стратегии консолидации ленточных накопителей с нескольких площадок.

#### Расширенная балансировка нагрузки и аварийное переключение канала на резервный ресурс

Усовершенствованная функция балансировки нагрузки и аварийного переключения канала на резервный ресурс позволяет агрегировать несколько каналов Ethernet в системе Data Domain в одну группу. Благодаря этому, несколько каналов распознаются приложением резервного копирования как один. Затем в системе выполняется прозрачная балансировка между каналами Ethernet в группе. Агрегирование каналов связи упрощает управление несколькими политиками резервного копирования и обеспечивает полное использование всех доступных каналов Ethernet. По сравнению с агрегированием каналов связи на уровне Ethernet в этом случае обеспечивается более легкое и масштабируемое решение с лучшим использованием полосы пропускания сети.

Кроме того, возможности аварийного переключения на резервный ресурс полезны при сбое одного из интерфейсов группы, когда система Data Domain все еще находится в рабочем состоянии. Система Data Domain автоматически направит последующие входящие задания с серверов резервного копирования на работающие интерфейсы. Этот механизм автоматического аварийного переключения канала на резервный ресурс поддерживает системы резервного копирования в рабочем состоянии в случае временных сбоев сети.

Функция расширенной балансировки нагрузки и аварийного переключения канала на резервный ресурс прозрачно взаимодействует с базовой возможностью сетевого подключения и поддерживает как физические, так и виртуальные

интерфейсы. Кроме того, в отличие от агрегирования на уровне Ethernet, отсутствуют специальные требования к настройке коммутаторов Ethernet.

#### СХД с высоким уровнем безопасности для надежного восстановления

В системах с ПО DD Boost используются преимущества архитектуры неуязвимости данных EMC Data Domain Data In-vulnerability Architecture: непрерывная проверка, обнаружение и устранение ошибок, а также другие функции, которые повышают отказоустойчивость и выполняются прозрачно для приложения резервного копирования.

При помощи ПО DD Boost эта защита распространяется на сервер резервного копирования путем создания контрольных сумм данных, которые отправляются в систему Data Domain. Контрольные суммы затем передаются вместе с данными. Система Data Domain, которая получает данные, вычисляет новые контрольные суммы входящих данных, а затем сравнивает их с подсчитанными значениями из ПО DD Boost в целях проверки, что обеспечивает сквозную проверку данных.

#### Прозрачная интеграция с приложениями резервного копирования

В настоящее время ПО DD Boost интегрирован с EMC NetWorker, а также с отдельными системами резервного копирования сторонних разработчиков, например, Symantec NetBackup 6.5 (или более поздней версии) и Backup Exec 2010 (или более поздней версии).

Благодаря интеграции продукта EMC NetWorker с Data Domain, в комбинации с осуществленной ранее интеграцией с EMC Avamar®, решение EMC NetWorker поддерживает прозрачную интеграцию с двумя ведущими в отрасли решениями для дубликации данных — на целевом устройстве и на источнике. Также интеграция NetWorker с ПО DD Boost позволяет:

- управлять программным продуктом DD Replicator и полностью контролировать состояние каталогов всех задач резервного копирования как в основном, так и во вторичном расположении, включая поддержку отдельных политик хранения;
- упростить процесс развертывания систем Data Domain за счет новых мастеров поиска и настройки ресурсов;
- упростить администрирование и увеличить прозрачность системы в процессе всего цикла резервного копирования за счет функции мониторинга и составления отчетов о хранимых данных для систем Data Domain.

В отличие от внедрения виртуальной ленточной библиотеки (VTL), для ПО DD Boost не требуется каких-либо избыточных компонентов в виде ленточных накопителей или их эмуляции.

Подобно реализации в системах Data Domain систем NFS и CIFS, образы резервных копий можно одновременно записывать в систему Data Domain и считывать из нее во встроенном формате образов приложения. Образы резервных копий можно удалять по одному, благодаря чему эффективность хранения выше, чем при вне-

дрении решений только на основе виртуальной ленточной библиотеки (VTL).

ПО DD Boost обеспечивает простой общий доступ к дисковым пулам на нескольких разнородных серверах резервного копирования, а также одновременное использование данных пулов. Это упрощает балансировку нагрузки и использование наиболее подходящих медиа-серверов для выполнения конкретного задания резервного копирования на основе предварительно настроенных политик.

## EMC представила Data Domain DD670

**Июль 2010 г.** — Корпорация EMC объявила о выпуске новой системы хранения среднего класса со встроенной дубликацией — EMC® Data Domain® DD670 с процессором Intel® Xeon® Processor 5500 Series на базе архитектуры Nehalem.

Новая система хранения DD670 обеспечивает внутреннюю дубликацию со скоростью 5,4 Тбайт/час и поддерживает 76 Тбайт неформатированной емкости или до 2,7 Пбайт логической емкости с физическим сжатием данных в 50 раз. Одна система DD670 может поддерживать репликацию данных до 90 удаленных офисов с помощью систем Data Domain меньшего размера, таких как DD140 или DD610, обеспечивая полностью автоматизированное восстановление данных в удаленных офисах в случае сбоя.

Базовая модель снабжена 12 Тбайт дискового пространства и шасси в двухстоечном варианте, а также поддерживает модули расширения с дисками SATA емкостью 1 или 2 Тбайт, что в 2,1 раза экономит пространство в рэковой стойке. Как и все продукты Data Domain, система DD670 проста в установке и достаточно гибка для безболезненного внедрения в существующие пользовательские среды. Data Domain поддерживает широкий круг приложений в любых системах хранения. Сюда входят системы NFS, CIFS, Symantec OpenStorage или EMC Data Domain Boost, которые работают через 1- или 10-гигабитное соединение Ethernet, и / или виртуальные библиотеки магнитных лент (VTL), работающие через Fibre Channel. Для систем DD670 и DD880 теперь доступно 8-гигабитное соединение Fibre Channel.

EMC также объявила о выпуске новой серии дисковых библиотек EMC Disk Library 5000, а также модуля расширения системы хранения функционалом дубликации для Disk Library для мэйнфреймов. Кроме того, в модулях расширения новых систем DD670, EMC Data Domain DD880 и недавно вышедшего массива EMC Data Domain GDA теперь поддерживаются диски емкостью в 2 Тбайт. Диски большой вместимости также доступны для новых библиотек Disk Library.

Система EMC Data Domain DD670, модули расширения с дисками на 2 Тбайт, а также модуль расширения систем хранения с внутренней дубликацией для библиотек DLm960 поступят в продажу в 3 кв. 2010 г. Виртуальные ленточные библиотеки DL5100 и DL5200 можно приобрести уже сейчас.