

CDP-решения

Обзор наиболее распространенных в России CDP-решений, появившихся на рынке 2 года назад и позволяющих добиться уровня непрерывности бизнеса, практически равного кластерным решениям.

Введение

Направление CDP-решений (Continuous Data Protection), или решений, обеспечивающих непрерывную защиту данных, активно стало развиваться в виде законченных продуктов с середины 2005 г. Его суть, в отличие от других методов по защите данных (мгновенные копии, реплицирование, виртуальные ленты и др.), – возможность восстановить состояние защищаемого “объекта” с точностью до нескольких секунд. CDP-решения могут быть файл-, блок-ориентированными и ориентированными на приложения. Следует учитывать также число поддерживаемых ОС/приложений/“гранулярность” восстановления и др.

В регионе Россия&СНГ решения данного класса доступны от вендоров: **HP** (HP StorageWorks Continuous Information Capture – CIC) Solution под управлением ПО Mendocino Software Recovery One v1.4), **EMC** (EMC RecoverPoint), **IBM** (IBM Tivoli Continuous Data Protection for files), **CA** (линейка XOsafi), а также компаний **Symantec Corp.** (в составе Backup Exec 10d и ряде новых продуктов, анонсы которых запланированы на конец с.г.), **FalconStor Software Inc.**, **Microsoft** и ряд других.

Позиционирование CDP-решений

На настоящий момент у каждого вендора существует достаточное множество решений, обеспечивающих защиту данных с разными уровнями доступности данных/приложений. Общая классификация подобных решений была разработана еще в начале 90-х. Например, в классификации HP все решения по поддержанию доступности данных делятся на 6 уровней (рис. 1). Первый уровень представлен самыми простейшими системами с резервированием данных на ленту/дополнительный диск. “Апофеоз” (или шестой уровень) классификационной модели – кластерные географически удаленные системы с синхронной репликацией.

Несмотря на все многообразие решений по обеспечению целостности данных, все они в основном строились на двух базовых технологиях: 1) создание ре-

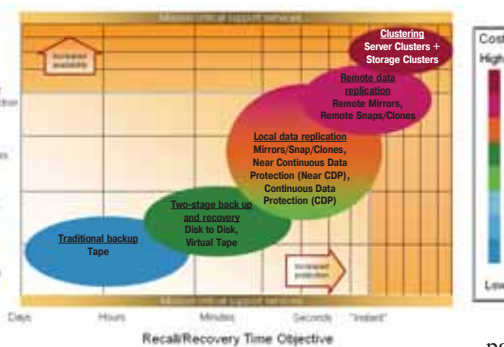


Рис. 1. Таксономия решений по поддержанию доступности данных/приложений, классифицируемых по показателям RPO и RTO в представлении HP.

зервной копии всех/выделенных файлов в течение “окна” для резервного копирования на ленту (в этот период приложения, как правило, не доступны); 2) поддержание целостности данных на основе создаваемых мгновенных снимков/копий томов (приложения остаются доступными). К этим технологиям добавлялась еще одна технологическая компонента – Continuous Remote Replication (CRR), обеспечивающая удаленное реплицирование томов, и вместе с различными техническими вариациями формировался рынок решений по поддержанию сохранности данных.

В середине 2005 г. SNIA стандартизован новый класс решений по защите данных – CDP, который был определен как “технология, которая непрерывно фиксирует или прослеживает модификации данных и сохраняет изменения независимо от первичных данных, позволяя осуществлять восстановление с любого места в прошлом. CDP-системы могут быть ориентированы на блок данных, файл или на приложение”.

CDP – новый вид технологии защиты данных, обеспечивающей мгновенный доступ к любой “точке восстановления” с неограниченной гранулярностью. Данные перехватываются при любом изменении, а “точки восстановления” могут быть определены произвольно, что обеспечивает быстрое восстановление поврежденных данных на момент, предшествующий сбою. CDP-решения в общем семействе технологий резервного копирования/восстановления занимают среднее положение как по временным

показателям RTO (Recovery Time Objectives – требуемое время восстановления объекта) и RPO (Recovery Point Objectives – периодичность сохранения данных объекта, подлежащего восстановлению), так и с точки зрения стоимости (см. рис. 1).

Помимо первоочередной задачи – поддержание доступности данных/приложений производственных серверов в случае сбоев/отказов – на основе CDP-решений могут решаться ряд дополнительных, например, создание вторичных образов данных (без влияния на работу производственных серверов) на любой момент времени с возможностью использования любыми другими серверами в целях аудита, подготовки отчетов, тестирования, миграции данных и т.п.

При использовании CDP-технологии в качестве базовой удается в большинстве случаев существенно уменьшить потребности в ресурсах хранения за счет минимизации использования других технологий по защите данных. Так, по данным HP, за счет оптимизации использования ресурсов хранения на основе CDP-решений удается на 20–40% уменьшить потребности в системах хранения и на 10–30% снизить затраты на администрирование.

При этом следует помнить, что все CDP-решения представленные на рынке, в большинстве случаев существенно отличаются друг от друга по функциональности, позиционированию, производительности и др., вследствие чего преимущества их применения в каждом случае необходимо рассматривать отдельно, а все вышесказанное относится ко всему CDP-направлению в целом.

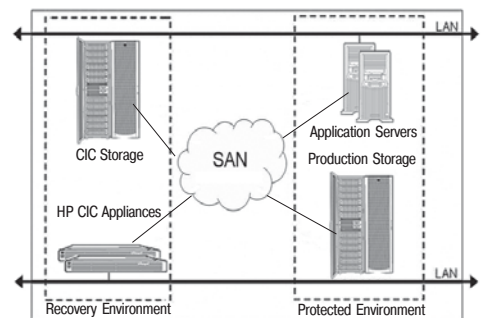


Рис. 2. Структурная схема HP CIC.

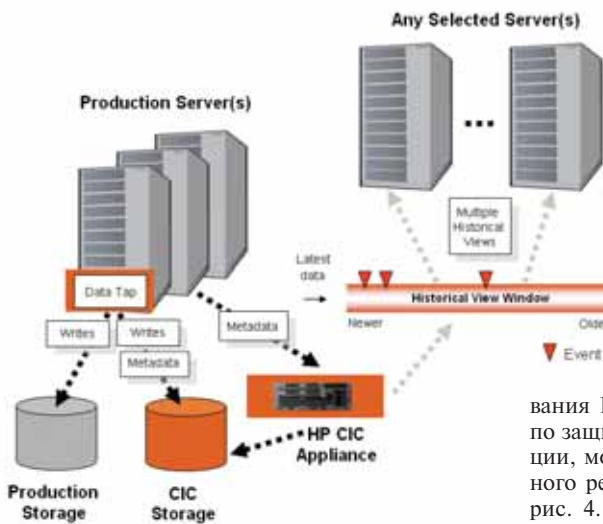


Рис. 3. Расширенная архитектура HP CIC с возможностью предоставления виртуального образа данных любого из обслуживаемых производственных серверов любому другому серверу на любой момент времени.

Обзор рынка CDP-решений

CDP-решения от HP

В семействе решений HP по обеспечению защиты данных CDP-технология представлена продуктом HP StorageWorks Continuous Information Capture (CIC) Solution. Данное решение строится на основе ПО от компании Mendocino Software. Его структурная схема CIC представлена на рис. 2. HP CIC представляет собой CIC Appliance и CIC Storage с установленным на прикладном сервере агентом – CIC Data Tap. CIC 1.4 recovery appliance устанавливается на сервер типа HP ProLiant DL585 G2 с предустановленной версией Enterprise Linux, Advanced Server и работает под управлением ПО Mendocino Software Recovery One v1.4. В качестве CIC Storage могут использоваться большинство систем хранения как high-end класса – HP XP (StorageWorks XP12000/10000), HDS, EMC Symmetrix, IBM, Sun Microsystems, Dell, так и среднего уровня – HP EVA (3000/5000/4000/6000/8000), HP MSA (1000/1500cs/1510i), EMC Clariion, IBM DS и др. Более полная информация о совместимости и поддерживаемых платформах в документе – CIC Support Matrix.

В общем случае число поддерживаемых CIC Appliance и прикладных серверов не ограничено. Кроме одиночных прикладных серверов, поддерживаются и

кластерные системы. Среди платформ прикладных серверов – Windows, Solaris, AIX и HP-UX.

Расширенная архитектура HP CIC с возможностью предоставления виртуального образа данных любого из обслуживаемых производственных серверов любому другому серверу на любой момент времени дана на рис. 3.

Сравнение позиционирования HP CIC с другими технологиями по защите данных, в частности, репликации, моментальных снимков и ежедневного резервного копирования дано на рис. 4.

В общем случае HP CIC, например, в сравнении с технологиями моментальных снимков позволяют сократить вре-

Traditional Steps	Recovery from Tape (Minutes)	Recovery from Disk Snapshot (Minutes)	Roll-back with HP CIC (Minutes)	HP CIC Steps
Troubleshooting	60	60	15	Troubleshooting
Find Backup & Archive Logs	15	15	10	Present Historical Views
Restore	180	5	15	Verify Recovery Point
Replay Logs	120	80	10	Production Rollback
Verify Recovery Point	15	15		
Example Time	390	175	50	Example Time

↑
Entire recovery process before recovery point validation can significantly increase recovery times

Рис. 5. Сравнение CIC с другими методами при восстановлении 2 Тбайт БД.

мя восстановления в несколько раз. Эффективность использования данного решения в сравнении с другими при восстановлении БД объемом 2 Тбайт дана на рис. 5.

Первые версии HP CIC позволяли создавать только локальные решения. Теперь с помощью HP Continuous Access или HP Storage Mirroring возможно создание географически удаленных CIC-решений.

Основные применения HP CIC: поддержка критических баз данных (ORACLE, SAP), корпоративных систем обмена сообщениями (Microsoft Exchange Server) и критических файловых систем.

EMC RecoverPoint

Первая версия продукта EMC RecoverPoint (реализация CDP-технологии) анонсирована в конце 2005 г. и активно продвигалась до середины 2007 г. По сути это была OEM-версия от компании Mendocino, которая имела возможность только локального развертывания, но уже была полной реализацией определения CDP, т.е. включала как поддержку файлов, так и поддержку БД.

После приобретения EMC компании Kashya (в мае

2006 г.) в августе того же года была анонсирована вторая версия EMC RecoverPoint – полностью реализованная только на технологии Kashya и уже с поддержкой удаленного (синхр./асинхр.) CDP.

Технологии, разработанные Kashya, достаточно хорошо апробированы и отработаны. Во-первых, они давно используются в решениях вместе с т.н. Cisco Fibre Channel MDS 9000 Storage Services Module. Эти решения одни из первых появились на рынке в составе SSM и уже с осени 2005 г. давали возможность:

- осуществлять репликацию с использованием политик и оптимизацией либо полосы пропускания, либо расогласования систем в заданных пределах и возможностью задания разных политик для разных приложений (consistency group);
- получать наивысший уровень компрессии данных за счет распознавания многих типов приложений;
- обеспечивать работу на любом расстоянии по IP (шлюзовые порты Ethernet включены в устройства Kashya);
- обеспечивать откат томов на любую точку в прошлом.

EMC RecoverPoint позиционируется как предпочтительное решение для инфраструктур с более чем одним критическим приложением и большим объемом быстро изменяющихся критически важных данных. Почему такая оговорка?

Прежде всего, вследствие того, что и сами приложения могут поддерживать свою доступность собственными средствами, например, на основе log'ов, но когда таких средств множество, могут существенно возрастать накладные расходы, увеличиваться сложность их поддержания и управления, и переход к универсальным консолидированным технологиям может оказаться более предпочтительным.

По данным исследований, проведенных Enterprise Strategy Group ("The Evolution of Enterprise Data Protection", December, 2004) среди более чем 400 storage-профессионалов и IT-менеджеров в США, специализирующихся на защите данных, около одной трети (31%) респондентов из организаций с годовым оборотом более \$100 млн сообщили, что они понесут существенные убытки в случае, если система будет недоступна от 1 минуты до 1 часа, а для 58% допустим простой менее 4 часов.

Эти исследования дают возможность оценить рынок CDP-решений. Если учесть, что мгновенные копии/snapshots/реплики обычно производят с частотой не более одной в течение часа, то среди компаний с оборотом > \$100 млн, бизнес которых жестко завязан на IT-инфраструктуру, более трети из них (за 2 прошедшие года требования к доступности только возрастали и будут расти) могут испытывать острую потребность в таких решениях.

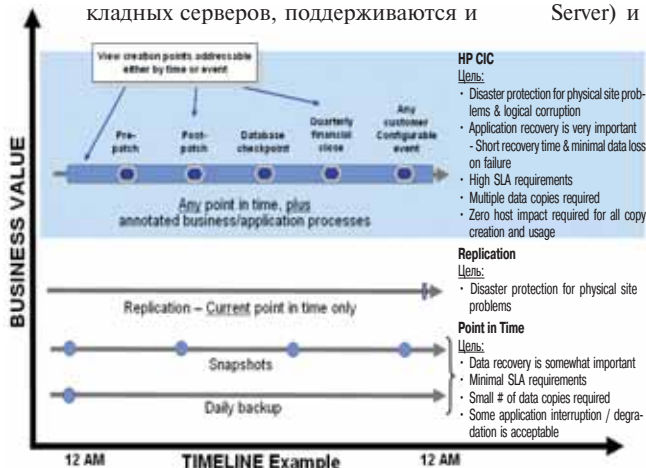


Рис. 4. Позиционирование HP CIC в сравнении с технологиями репликации, моментальных снимков и ежедневного резервного копирования.

К позиционированию CDP необходимо сделать еще ряд замечаний. По мнению многих аналитиков, развитие рынка CDP-решений даст мощный толчок и сектору т.н. “интеллектуальных” коммутаторов SAN, активно продвигаемых, например, компаниями Brocade и Cisco в составе своих продуктов (представляют собой appliance, устанавливаемые в SAN-коммутаторы; в спецификации EMC – Connectrix AP-7420B от Brocade и Cisco Storage Services Module). Это связано с тем, что, хотя реализация CDP-решений возможна и на базе драйверов на серверах, использование “интеллектуальных” коммутаторов значительно упрощает их реализацию как локальных, так и удаленных CDP-инсталляций, а также до минимума снижает накладные расходы и сложность управление решениями. В данном контексте EMC RecoverPoint является мощным дополнением к решению EMC Invista (SN № 3/24, 2005), а в перспективе возможна их интеграция.

В общем семействе EMC-решений RecoverPoint дополняет аналогичные технологии, реализованные в составе массивов – SRDF и MirrorView, и сервер-ориентированные – RepliStor.

Особенности реализации EMC RecoverPoint

EMC RecoverPoint appliance это двухузловой Intel-сервер, который имеет не менее 4 FC-портов. В минимальной конфигурации решение RecoverPoint должно содержать 2 (с целью обеспечения надежности и осуществления upgrade без останова системы) RecoverPoint appliance с внутренней IP-связью, которые управляются как жестко связанный кластер и каждый из которых подключается к своей фабрике, при этом имеется возможность распределения нагрузки между appliance.

Перечислим основные особенности решений на базе EMC RecoverPoint:

- гетерогенность – защита любого типа массива/сервера/операционной системы/приложения;
- возможность локального и удаленного восстановления независимо от расстояния, типа интерфейса (FC/IP);
- восстановление на любой момент времени (практически) или событие;
- поддержание синхронизированных по времени копий нескольких томов одного приложения (Consistency group);
- управляемые политиками показатели RPO и RTO;
- решение класса out-of-band (управление “вне” потока данных), масштабируемое, с высоконадежной архитектурой;
- интеграция с решениями (“интеллектуальными” модулями для FC-коммутаторов) Cisco (в ближайшем будущем и с Brocade);

- обеспечение максимума преимуществ для поддержки различных уровней обслуживания (резервирование и восстановление/тестирование, разработка и обучение/параллельная обработка, отчеты и запросы).

EMC предлагает две базовые конфигурации решения RecoverPoint на основе технологии CDP/Kashya, с использованием которых уже могут строиться другие реализации CDP-решений:

- решение с постоянной защитой данных – обеспечение постоянной локальной защиты любых приложений с гранулярностью во времени до нескольких секунд;
- решение с постоянной защитой данных на базе удаленной репликации (CDP + Continuous Remote Replication). Данное решение географически расширяет возможности предыдущего, делая его и катастрофоустойчивым.

Особенности управления EMC RecoverPoint

Управление RecoverPoint осуществляется на основе создаваемых политик управления ресурсами, которые позволяют оптимизировать точки восстановления и время восстановления объектов.

С помощью групповых политик можно минимизировать задержки между сайтами (локальным и удаленным) за счет группирования пакетов данных и контроля его прохождения в целом, а не каждого пакета. RecoverPoint самостоятельно выполняет оптимизацию ресурсов с целью поддержания задаваемых политик, а при превышении пороговых значений политик выдаются соответствующие предупреждения.

Одна из возможностей RecoverPoint – его способность регулировать поведение ввода/вывода приложения, гарантируя выполнение соответствующей политики. Например, если установленная политика минимизирует задержки между сайтами (при максимальной задержке 10 сек.), то операции записи приложения могут быть замедлены, чтобы гарантировать выполнение задержки. Но, как правило, RecoverPoint самостоятельно динамически изменяет процесс реплицирования с целью оптимизации политик сразу для всех активных групп.

Множественные политики дают гибкость в управлении RPO и RTO среди различных групп, а расширенные возможности политик позволяют подстраивать эти политики в группе. Кроме того, политика может быть динамически изменена, учитывая данные, получаемые от операций в реальном масштабе времени.

RecoverPoint поддерживает большинство открытых систем. Поддерживаются Microsoft Windows, IBM AIX, Sun Solaris и HP-UX. Поддерживаемые массивы хранения (для производственных данных и восстановления): EMC CLARiON и Symmetrix, массивы от сторонних по-

ставщиков, подключаемые к SAN, в том числе: ZPAR, Apple, Dot Hill, Engenio, Hitachi, HP, IBM, JetStor, NexSAN, Pillar, Sun, StorageTek, XIOTech.

В целом, решение EMC RecoverPoint позиционируется для среднего и корпоративного рынка с ценовым диапазоном от \$100 тыс.

СА: линейка решений XOsoft

Технологии CDP в составе решений компании СА появились год назад – с момента приобретения (в июле 2006 г.) компании XOsoft. Это целая группа продуктов, полностью интегрированная с ARCserv. В отличие от других решений, продукты XOsoft с самого начала были ориентированы для работы в географически распределенной среде. В составе семейства XOsoft – следующие решения с интегрированной технологией CDP:

- **WANSync** – обеспечение катастрофоустойчивости основного сайта (резервная площадка географически удалена) на основе репликации (в реальном времени) файлов и баз данных. Ориентировано на низкоскоростные каналы связи;
- **WANSyncHA** – полная поддержка функциональности предыдущего решения, но с возможностью автоматического переключения на резервный сервер. В сравнении с чисто кластерным решением, в основном, дешевле и проще в обслуживании;
- **WANSyncCD** – решение для обмена данными (распространение и консолидация; консолидация эквивалента централизованному бекапу) с удаленными офисами;
- **Enterprise Rewinder** – реализация технологии CDP для локальных серверов;
- **InMotion** – автоматизация задач агрейда, обслуживания и миграции для серверов Exchange на основе технологии CDP.

Помимо этих решений есть еще продукт Assured Recovery, позволяющий проводить тестовый запуск резервного сервера (для решения WANSyncHA) без разрушения реплицированных данных.

Среди поддерживаемых ОС: Windows 2000/2003 (32/64bit)/MSCS, Sun Solaris, Linux (Red Hat), IBM AIX. Поддерживаемые файловые серверы, приложения/базы данных: MS Exchange, MS SQL, IIS (Internet Information Services), Oracle.

IBM Tivoli CDP for File (CDPF)

CDP-технология от IBM в виде решения IBM Tivoli CDP for File стала доступна на рынке с августа 2005 г. Из трех типов CDP-технологий – блок-уровневая, файл-уровневая и на уровне приложений – данное решение поддерживает только файл-уровневую CDP-технологию с ограниченным количеством платформ. IBM позиционирует данный продукт для двух применений:

- поддержание доступности данных конечных пользователей (настольные и мобильные компьютеры);

- поддержание доступности данных файловых серверов на уровне компании и департаментов.

Сейчас доступна версия 3.1 CDPF. Для инициализации CDPF устанавливается программный агент на компьютер или сервер с указанием целевого устройства для сохраняемых данных, которое может располагаться на том же компьютере/сервере, сетевом устройстве хранения или удаленном устройстве.

CDPF представляет собой альтернативу традиционным подходам к резервному копированию “конечных точек” и ориентирован на применение технологии использования жестких дисков в качестве хранилища резервных копий данных. В момент сохранения файла Tivoli CDP for Files создает копии этого файла для отправки в различные местоположения или пункты назначения.

CDPF благодаря непрерывной защите наиболее важных файлов пользователя, включая документы, над которыми ведется активная работа, позволяет уменьшить или устранить “пробелы” в процессах резервного копирования данных и обеспечить эффективную и действенную защиту данных в “конечных точках” и помогает осуществлять восстановление данных после сбоев, вызванных вирусными атаками, повреждением данных и ошибками пользователей.

В версии 3.1 CDPF сделаны ряд расширений, в частности:

- CDPF интегрирован с Lotus Notes с поддержкой клиента восстановления файлов .nsf до состояния на конкретный момент времени;
- реализована поддержка Microsoft Vista;

- сделан ряд усовершенствований, благодаря которым настраивать и управлять CDPF стало проще.

CDPF поддерживает все Windows операционные системы и файловые серверы.

Приобретение лицензии на использование Tivoli CDP for Files V3.1 для обеспечения непрерывной защиты данных на сервере и на ПК стоит \$321 и \$45 без НДС соответственно.

Вместо заключения

Все представленные в обзоре CDP-решения существенно отличаются друг от друга. Так, предложения EMC, HP в большей степени ориентированы на корпоративный сектор.

Продукты CA и IBM – на средний и малый бизнес. Первые 2 CDP-решения реализуются на выделенных специализированных серверах, которые в целях повышения надежности могут быть кластеризованы. Соответственно, и их производимость, с точки зрения поддерживаемых производственных серверов, несопоставима, например, с предложением IBM, которое работает как агент на том же компьютере/сервере, который поддерживает.

Принципиальное значение имеет уровень CDP-решений – блочный или файловый, а также список поддерживаемых платформ. В совокупности эти 2 показателя могут значительно сужать или расширять возможные области применения соответствующих CDP-решений.

Также следует обращать внимание на сопутствующую интегрированную в CDP-решения функциональность. Например, возможность создания моментальных снимков производственных томов в целях тестирования/аудита/подготовки отчетов

тов (от HP) или возможности как для тестирования CDP-решений, так и по созданию удаленных распределенных решений с различной функциональностью на основе CDP-технологий (от CA).

Решения от EMC и HP, в отличие от решений CA и IBM, не являются “коробочными” и требуют определенных усилий по установке и созданию специальных скриптов.

CDP-решения являются определенной альтернативой кластерным высокодоступным системам – как локальным, так и распределенным, но с меньшей стоимостью реализации.

Уже в ближайшей перспективе, по нашим оценкам, этот класс продуктов поддержания высокой доступности займет свою долю на рынке. Прежде всего, это связано с теми убытками, которые несут компании из-за недоступности к своим системам продаж/услуг.

Группа компаний ЛАНИТ использует все вышеперечисленные решения, в зависимости от требований, предъявленных заказчиком. Огромный опыт и высокая квалификация сотрудников ЛАНИТ позволяют выполнять проекты по внедрению CDP-решений любой сложности и масштаба, а также гарантировать отличное качество создаваемых систем.

В ближайшее время на рынке должны появиться специализированные платформы расширений для установки в SAN-директора, которые можно использовать совместно/вместе со специализированными appliance, что позволит еще больше упростить установку корпоративных CDP-решений.

Евгений Фохт,
системный инженер,
ДСИ ЛАНИТ