

HR ILM: информация, которая не “лежит”, а работает

В середине марта с.г. HP анонсировала 3 продукта, относящихся к портфелю решений для управления жизненным циклом информации: Reference Information Manager for Messaging v1.1, Reference Information Storage System v1.1 и File System Extender 3.1. Объявление первого из них фактически было презентацией целого класса ILM-решений. Данная публикация продолжает серию материалов по ILM-стратегиям, предлагаемых на рынке, и, если кратко резюмировать усилия основных разработчиков в этом направлении, то в настоящее время HP имеет наиболее понятную концепцию, конкретное видение архитектуры ILM-решений и примеры реализации ILM-стратегии.

Введение

Начнем с определения ILM. “ILM – это комбинация процессов, людских ресурсов и технологий для управления информацией на основании ее значимости для бизнеса с целью предоставления нужной информации в нужное время в нужном месте...” (Meta Group). Хотя трактовки этого термина частично пересекаются у поставщиков решений, акцент, который расставляется, крайне важен для понимания направленности решений и их преимуществ в сравнении с альтернативными.

Итак, что говорит HP: “ILM – это стратегия по отношению к информации на предприятии, которая охватывает и регламентирует обращения с данными, и аппарат-

ные системы, и управляющий программный инструментарий, что подразумевает:

- задание корпоративных правил и процедур, определяющих, где должны храниться данные в зависимости от их значимости для бизнеса;
- автоматизацию процессов создания, размещения, миграции и удаления данных в соответствии с этими правилами;
- высвобождение ресурсов, снижение общей стоимости владения и повышение уровня обслуживания IT-инфраструктуры”.

Это означает, что ILM – не только “правильное” хранение данных, но и возможность активного использования хранимых данных при обеспечении условий снижения затрат на хранение и увеличение прибыльности за счет “правильного” использования хранимых данных (рис. 1). Иными словами, целью реализации ILM-стратегии является не построение хранилища корпоративных данных, а построение такой IT-инфраструктуры, где информация работает на ваш бизнес.

Сегодня многие производители, предлагающие свои продукты для рынка хранения данных, ориентированы на решения для отдельных стадий жизненного цикла данных: соз-

дание и модификация, реплицирование и распределение, защита и восстановление, архивация и повторные обращения к данным, а также удаление информации.

Все эти решения существуют и работают в пределах конкретной стадии жизненного цикла информации, например: ПО реплицирования дисковых систем, решения по резервному копированию/восстановлению данных, архивация данных на ленты, WORM-носители и др.

Концепция адаптивного предприятия HP требует, чтобы информация управлялась в течение всего ее жизненного цикла, т.е. использовать политико-ориентированные сервисы хранения для автоматизации управ-



Рис. 1. Сравнение традиционного подхода к управлению информацией с ILM, где все операции с данными находятся в постоянной взаимосвязи.

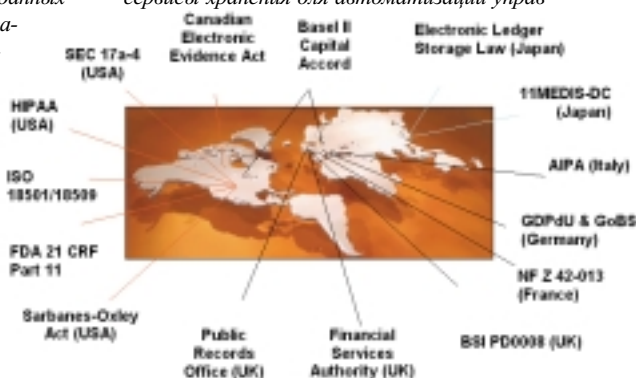


Рис. 2. Законодательные акты и нормативные документы, принятые в мире, регулирующие хранение информации.

ления информацией, основанном на ее бизнес-значении. И важно не только управлять информацией, т.е. доставлять ее в нужное время в нужное место нужным людям, но также доставлять ее за стоимость, соответствующую ее ценности.

Причины актуальности ИЛМ-решений

Причин, которые способствуют активному внедрению ИЛМ-решений, несколько. Среди них можно выделить:

- лавинообразный рост данных;
- необходимость обеспечения доступа к данным в течение длительных периодов времени;
- необходимость управления статусом ценности информации во времени;
- введение законов и нормативных документов, регламентирующих хранение информации;
- появление широкого большого диапазона систем хранения, отличающихся уровнем доступности к данным и стоимостью хранения данных;
- необходимость снижения сложности управления хранением данных, общей стоимости хранения и общих затрат на ИТ.

По оценкам IDC, уже сейчас ежегодный прирост объемов данных составляет 80% и в дальнейшем он будет только возрастать (что опережает темпы увеличения объемов дисковых носителей). Произошло разделение данных на оперативную и справочную (объем которой составляет от 50% до 75%) информацию, требующих разных подходов с точки зрения доступности к данным, надежности и стоимости хранения. Кроме того, за последние несколько лет во многих странах мира приняты законодательные акты и нормативные документы, регулирующие хранение информации (рис. 2).

Архитектура ИЛМ-решений

ИЛМ имеет законченную архитектурную структуру (рис. 3). Фактически все решения строятся на следующих ключевых компонентах: HP ИЛМ Policy Manager (его появление ожидается в ближайшем будущем и в настоящее время его функциональность реализована в программах OV SAM и HP Storage Essentials), API-интерфейсы (RIM) и RISS – Reference Information Storage System (SN №2/20, 2004). Некоторым расширением этого ряда является HP File System Extender, ПО для перемещения файлов между разными уровнями хранения, согласно заданным



Рис. 3. Общее представление архитектуры HP ИЛМ.

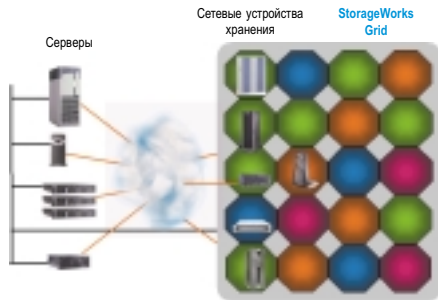


Рис. 4. В перспективе архитектура StorageWorks GRID должна стать общей платформой для корпоративного хранения данных. Самое важное место в этом представлении занимает RISS. В ближайшей перспективе все основные ИЛМ-решения будут строиться на ее основе. А в средней перспективе – архитектура StorageWorks GRID должна стать общей платформой для корпоративного хранения данных (рис. 4).

По сути, сама архитектура RISS и является ответом на основную цель ИЛМ-решений ИЛМ – сделать процесс хранения из затратного, где информация просто лежит, прибыльным, где она активно работает и используется. Это достигается прежде всего за счет того, что исходная неструктурированная информация при перемещении на RISS превращается в хорошо индексируемую, с упрощенным и минимальным в зависимости от заданных политик по времени доступом, высокой степенью сохранности информацию. Это достигается за счет следующих факторов:

- значительного расширения интеллектуальности систем хранения. Каждая ячейка HP RISS это не только единица хранения, но и мощная поисковая подсистема с полнотекстовой индексацией и расширенными сервисами по управлению данными, хранящимися в ней;
- смещения архитектуры данных систем хранения от “пассивной” в “активную” сторону, дающего возможность гораздо более широкого автономного управления данными между различными уровнями иерархии систем хранения, а также жизненным циклом данных без “вмешательства извне”;
- повышенной степени самоорганизации и самовосстановления;
- значительно более высокого уровня интеграции в одном продукте функциональности от одного вендора, ранее поддерживаемой различными компонентами систем хранения от нескольких вендоров.

Единицей хранения в RISS является неизменяемый файл (не запись, например, для БД и не блок данных). Основная задача хранилища – обеспечение управления заявленной доступности к файлам и поддержание в неизменном состоянии совокупности единиц хранения в течение заданного периода времени. Большинство систем хранения класса RISS (считать которые можно по

пальцам одной руки) строятся на IP-интерфейсах. При этом IP-протоколы используются как для внутреннего взаимодействия между компонентами, так и для внешней связи, что необходимо учитывать при оценке общей пропускной способности системы.

Основные преимущества использования RISS определяются за счет экономии средств и времени на вспомогательных операциях, обеспечивающих целостность данных (на текущий момент – не менее нескольких терабайт) в течение многих лет (до 15 и более), а также при необходимости работы поиска необходимой информации среди многих миллионов документов.

Как видит ИЛМ развитие своих ИЛМ-решений? В середине марта с.г. компания HP сделала важный анонс: был выпущен продукт Reference Information Manager (RIM) for Messaging v1.1. и одновременно Reference Information Storage System v1.1 – вторая версия RISS. Кроме того, в рамках данного анонса был представлен Software Development Kit (SDK), который позволит независимым разработчикам ПО (ISV) создать собственный интерфейс к RISS для ISV-приложений. Фактически этим объявлением HP заявила о целой серии своих продуктов. Первая версия RISS была жестко интегрирована (с точки зрения ПО и аппаратного обеспечения) продуктом, ориентированным только на одно применение – поддержка почтовых систем. Появление RIM это прежде всего “отделение” управляющего ПО от RISS в качестве API-интерфейса к приложению. При этом сам RISS уже может позиционироваться как универсальное решение для широкого круга применений как с ПО от HP, так и от других фирм.

В ближайшее время (в течение 2005 г.) к RIM for Messaging (поддержкой почтовых

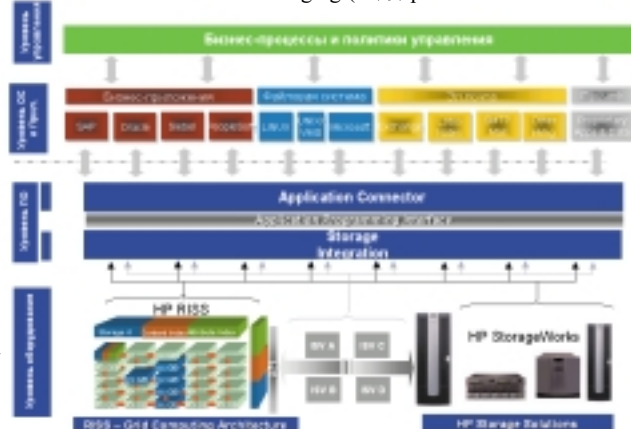


Рис. 5. Расширенное представление архитектуры HP ИЛМ.

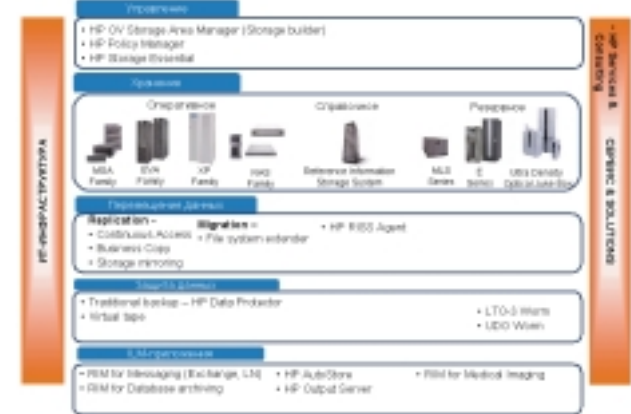


Рис. 6. Портфель решений HP ИЛМ.

серверов MS Exchange и Lotus Domino) добавятся RIM for Database archiving и RIM for Medical Imaging, а также приложения от третьих фирм. В перспективе планируется поддержка наиболее популярных приложений: PeopleSoft, AutoCAD, IP Telephony, ORACLE, SAP, Siebel, Backup Apps и др.

Расширенное представление архитектуры HP ILM и портфель решений HP ILM дано на рис. 5 и 6 соответственно. В концепции HP ILM все хранение делится на 3 группы: оперативное, справочное и резервное. Первая группа представляется семействами устройств: XP, EVA, MSA, NAS; вторая – RISS и третья – ленты (MSL-series, E-series) и оптические системы.

Основной принцип построения ILM-решений HP – их независимость от приложений. Уровень политик управления в основном строится на базе API-интерфейсов или прикладных программных приложений (RIM) к RISS. Часть функциональности расширяется в области правило-ориентированного перемещения данных и осуществляется за счет

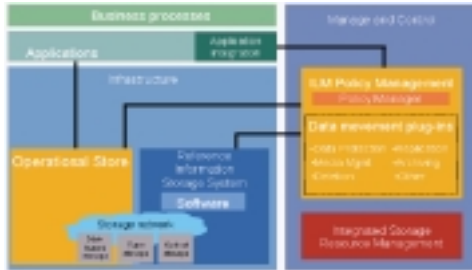


Рис. 7. Основные элементы HP ILM архитектуры. Левая половина отображает логический путь данных, правая – пути управления.

File System Extender. Дальнейшее развитие управления ILM-решениями с точки зрения бизнеса будет расширяться за счет анонсированного HP ILM Policy Manager (PM).

В задачи HP ILM Policy Manager будет входить интеграция всех правил управления информацией/данными в едином центре и обеспечение управления на всех уровнях хранения: оперативном, справочном и резервном. Роль PM – обеспечить анализ данных в контексте приложений/пользователей и помочь администратору в управлении политиками. Например, PM ответственен за понимание, какие приложения управляют данными, на каком уровне хранения они должны находиться и в какой стадии жизненного цикла они находятся. Через PM будут задаваться политики для RIM, FSE и других продуктов, поддерживающих ILM-решения. В целом, PM позиционируется как продукт корпоративного класса, который ориентирован на управление прежде всего оперативными данными (рис. 7) и необходим в развитой распределенной корпоративной IT-инфраструктуре, в отличие от более простых в использовании продуктов, например, FSE.

Помимо вышеназванных RIM-приложений, портфель ILM-решений дополняют HP AutoStore и HP Output Server (см. пример 3). HP AutoStore позволяет отправлять цифровые копии отсканированных или напечатанных документов непосредственно на RISS, минуя оперативное хранилище корпоративной информации.

Примеры реализаций HP ILM-решений

В общем виде разработка ILM-решения состоит из 5 этапов:

- изучение и классификация данных: инвентаризация данных, изучение типов данных, изучение контента;
- распределение данных по уровням хранения: оперативные данные, справочная информация, архивное хранение;
- создание политик перемещения данных – автоматическое, основанное на политиках перемещение данных между уровнями хранения;
- организация непрерывного доступа к информации: индексирование и поиск данных, защита данных, повышение качества обслуживания (QoS), удаление дублирующей информации;
- реализация ILM-решения: для электронной почты, баз данных, ERP/CRM, медицины и др.

Пример 1. Решения HP ILM на основе RIM&RISS для электронной почты.

В списке поддерживаемых данное решение следующие почтовые серверы: Exchange 5.5/2000/2003, Lotus Domino V5/V6, Sendmail, а также почтовые клиенты: Outlook 2000/2002-XP/2003. Из поддерживаемых ОС для управления документами: Windows 2000 и Windows XP.

Основные преимущества этого решения в отличие от традиционных почтовых систем – в том, что любой объект хранения находится в “активном” состоянии, что дает более широкие возможности по “манипулированию” объектом хранения, управлению доступом к нему и поисковым возможностям системы, когда поиск некоего подмножества документов (документа) осуществляется на множестве параллельно работающих серверов. Для этого, применительно к почтовой системе, каждая ячейка содержит индексацию по атрибутам и по ключевым словам содержимого объекта, а также непосредственно сам объект.

Среди файлов, которые автоматически индексируются по ключевым словам, позволяя впоследствии осуществлять поиск по содержанию документа, следующие: MS Office (.doc, .ppt, .xls), MS Access (с различными расширениями), MS Outlooks personal folder (.pst), Adobe Public Distribution Format (.pdf), Rich text format (.rtf), Outlook encapsulated .rtf (TNEF), Hypertext Markup Language (.html), Text files (.txt), ASCII files (.asc). Файлы других типов, не содержащие информации для индексирования, например, TIFF, MPEG, GIF, JPEG, WAV, могут записываться в RISS, но их поиск будет осуществляться только по атрибутам (тип файла, имя файла, время создания файла, автор файла).

Данное решение хорошо подходит для хранения как “регулируемых”, так и “нерегулируемых” сообщений в течение многих лет. В регулируемой среде почтовые сообщения немедленно направляются к системе хранения; в нерегулируемой среде – система хранения сама периодически “перемещает” или “собирает” сообщения от Exchange серверов.

Например, типичная политика может определить системе собирать все почтовые сообщения по истечении 7 дней. В этом заданном режиме архивирования RISS становится расширением сервера Exchange и сгружает всю электронную почту, которая получена 8 дней назад и более, таким образом сокращая размер сервера Exchange. Электронные письма могут также выборочно собираться на основании других критериев (например, отправитель, получатель, тема и т.д.).

Все собранные сообщения (и их вложения) хранятся RISS и удаляются с сервера Exchange. После этого пользователи видят немного изменений: заголовки сообщения появляется в почтовом ящике пользователя; иконка свидетельствует, что данные были успешно сохранены на RISS. Чтобы восстановить сообщения, пользователи просто дважды щелкают на значке: сообщения восстанавливаются с RISS и посылаются Outlook клиенту (не на сервер Exchange).

Почтовые сообщения, к которым предъявляются повышенные требования к неизменности и сохранности в течение многих лет, могут снабжаться цифровыми подписями с целью верификации их целостности.

Правила хранения контента могут определять различные уровни доступа к его содержимому для различных групп или отдельных пользователей.

Среди других преимуществ можно отметить:

- устранение хранения дублированных сообщений;
- ускоренный поиск сообщений (сообщения могут отыскиваться в соответствии со многими критериями поиска, включая: дату сообщения, имя получателя и отправителя, тему или текстовую строку в пределах сообщения);
- обеспечение нескольких уровней защиты данных.

Пример 2. Решения HP ILM на основе RIM&RISS для медицины.

RIM for Medical Imaging позволяет легко хранить и извлекать данные на многоуровневых хранилищах типа HP RISS в соответствии с международными нормами HIPAA. Данное решение предусматривает 2 фазы развертывания: 1) на основе интерфейсов DICOM и CIFS/NFS обеспечивается легкая интеграция в корпоративную систему лечебного учреждения; 2) расширенная интеграция архива с HIS (Hospital Info System) и ERP-системой и поддержкой Electronic Patient Record (EPR).

Пример 3. Решение HP ILM на основе RISS и HP Output Server для хранения документов (рис. 8).



Рис. 8. Решение HP ILM для хранения документов.

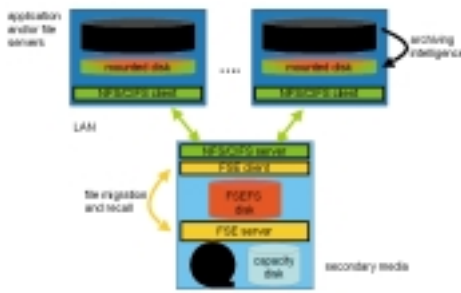


Рис. 9. Конфигурация FSE при консолидированном хранении файлов.

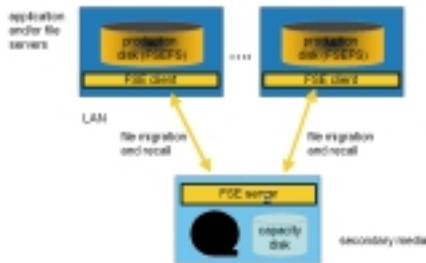


Рис. 10. Конфигурация FSE при распределенном хранении файлов.

RISS в настоящее время объединяется с HP Output Server через почтовую интеграцию. HP Output Server обеспечивает сбор, сохранность на RISS и дальнейшее распространение (через печать, fax, web, FTP и email) отчетов, справок и других документов, генерируемых такими приложениями, как SAP, Oracle Business Suite, Microsoft Business Solutions и др. При этом в дополнение к функциям распространения добавляется возможность повышенной сохранности и неизменности для отдельных групп документов.

Пример 4. Построение ILM-решения на базе FSE.

Идея FSE проста – обеспечивая прозрачный доступ к данным со стороны приложений, оптимизировать их хранение на нескольких уровнях на основе задаваемых политик/правил миграции данных.

Список возможных решений здесь гораздо шире, чем в выше представленных примерах. FSE жестко не привязан к аппаратным платформам от HP и может работать со многими платформами и множественными приложениями одновременно. Среди приложений/сред, протестированных на совместимость с FSE, следующие: почтовые системы – KVS Enterprise Vault, IXOS, COI; системы

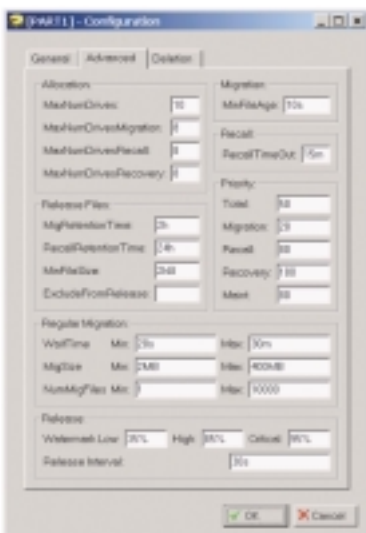


Рис. 11. Параметры управления политиками в FSE.

SAP ERP – IXOS eCONserver, Ceyoniq; архивирование данных на NetApp файлерах – Enigma Data Systems SmartMove; для архивирования в системах документооборота – COI BusinessFlow, Optimal Systems OS:DRT, Dubidot Media Suite, Ceyoniq и ряд др.

Существует 2 схемы (конфигурации) построения решений на базе FSE: для консолидированного хранения файлов (рис. 9) и распределенного хранения файлов (рис. 10). При консолидированном варианте хранения FSE-клиент находится на том же сервере, что и FSE-сервер. Результат – значительное упрощение инсталляции. Файловые системы монтируются/разделяются, используя NFS/CIFS. Вследствие этого – широкая совместимость с операционными/файловыми системами, т.к. в основном они строятся на тех же стандартных протоколах (NFS, CIFS и FTP) для совместного использования файлов.

В случае распределенного хранения требуется установка FSE-клиента на рабочем сервере, который должен существовать со всеми приложениями и утилитами, выполняющимися на нем. Для FSEFS используется любой диск на рабочем сервере. FSE-клиент взаимодействует с FSE-сервером через CORBA по LAN. При этом в данной конфигурации обеспечивается более высокая производительность вызова файлов со второго уровня хранения, чем при консолидированном хранении.



Рис. 12. Roadmap FSE и IMA.

Политики миграции в FSE могут строиться, основываясь, например, на “возрасте” файла или на доступной свободной емкости в оперативном хранилище, или при возникновении каких-либо условий. Меню параметров, которые можно задавать для определения политик, показаны на рис. 11.

В ближайшей перспективе планируется поддержка FSE библиотек HP UDO libraries, LTO WORM таре и появление FSE-клиента для HP-UX. Также скоро должен появиться HP StorageWorks ILM Migration Agent (IMA) 2.0 – простое решение для управления файлами в Windows среде. IMA взаимодействует с FSE-клиентом и технологиями фильтрации файлов. IMA будет поддерживать перемещение неактивных файлов непосредственно на RISS, NAS или на дисковые системы (рис. 12).

Заключение

Сделанная HP серия объявлений в области практической реализации ILM-стратегии после длительной паузы в значительной степени проясняет ее позицию относительно развития цикла управления для жизненным циклом информации, что, безусловно, способствует как развитию данного сектора рынка, так и лучшему пониманию потребителями преимуществ и перспектив ILM-решений от HP.

Катастрофоустойчивые решения от HP среднего класса

Март 2005 г. – Компания HP объявила о выпуске на российский рынок программного пакета HP Cluster Extension для управления автоматическим восстановлением информации в географически распределенных кластерных системах, основанных на системах хранения данных HP StorageWorks Enterprise Virtual Array (EVA). Пакет упрощает администрирование ИТ-инфраструктуры и расширяет возможности совместной работы серверов и систем хранения без ущерба управляемости всей системы при построении катастрофоустойчивых конфигураций и резервных центров.

Предлагаемое решение предназначено для обеспечения целостности и автоматического восстановления данных приложений, критичных для бизнеса, и функционирует под управлением кластеров MSCS для MS Windows и ServiceGuard для HP-UX. Ранее компания HP предлагала своим заказчикам подобный программный продукт, предназначенный для дисковых массивов уровня high-end HP StorageWorks XP.

Решение HP Cluster Extension позволяет управлять автоматической синхронизацией данных между первичной и вторичной системами хранения данных HP StorageWorks EVA, включенных в кластер. В обычном режиме функционирования кластера данные, необходимые для работы приложений пользователей, хранятся на первичной системе хранения HP StorageWorks EVA и в режиме реального времени дублируются на вторичной системе. В случае экстренной ситуации выполняемые приложения автоматически переключаются на работу с вторичной, резервной системой хранения. Причем при устранении экстренной ситуации пакет HP Cluster Extension автоматически инициирует процедуру репликации данных между системами хранения HP StorageWorks EVA, что позволяет восстановить обычный режим функционирования кластера. Этот процесс остается совершенно незаметным для пользователей системы, продолжающих использовать приложения.

Пакет HP Cluster Extension беспрепятственно интегрируется с платформой управления серверами и системами хранения данных, реализующей основные функции управления системами хранения данных, и создает возможность оперативной адаптации систем хранения данных к потребностям корпоративных заказчиков. Помогая виртуализировать и автоматизировать процессы управления системами хранения данных, HP Cluster Extension позволяет заказчикам быстро перераспределять ресурсы и мощности. Такой подход лежит в русле стратегии HP Adaptive Enterprise, которая синхронизирует задачи бизнеса и их информационной инфраструктуры для получения максимальной выгоды от непрерывных изменений, происходящих на рынке.