

HP Autonomy Consolidated Archive

— консолидированный архив

С сентября 2012 г. (спустя 8 месяцев после приобретения HP компании Autonomy) в России стало продвигаться решение HP Autonomy Consolidated Archive, занимающее, по оценкам ряда аналитических агентств, на рынке архивных систем одну из лидирующих позиций.



Иван Тимофеев — менеджер канала продаж в HP России и СНГ, департамент решений по Управлению информацией и Autonomy.

Введение

Рост требований информационного надзора и взрывной рост объемов и сложности информации требует от каждой организации применять новые подходы, помогающие управлять информационными процессами и формировать 100%-ное понимание информации, воспринимаемой человеком, в режиме реального времени. Эта информация включает в себя все — начиная от текстовых, аудио-, видеоданных, информации социальных сетей, баз данных и заканчивая данными, порожденными машинами и датчиками. Это разнообразие информации выносит значимость архивирования за пределы традиционных ИТ-преимуществ: консолидации данных, устранения дублирования данных и управления их жизненным циклом. Сегодня основной причиной организации архива является получение контроля над информацией для обеспечения соблюдения правовых, нормативных, ИТ- и внутренних, бизнес-ориентированных политик. Успешные внедрения решений для архивирования позволяют организациям находить, удалять, исследовать и содержать в неизменном состоянии информацию требуемое время (legalhold).

После приобретения компании Autonomy портфолио HP очень удачно было дополнено решением ACA (Autonomy Consolidated Archive), которое позволило совершить эволюционный шаг вперед решению HP IAP, обеспечив бесшовную миграцию и богатый набор функционала. В настоящее время HP

ACA является флагманским решением по архивированию файлов и электронной почты. Решение Консолидированного Архива — HP ACA, занимая на рынке решений архивирования одни из лидирующих позиций (по исследованиям Gartner и Forrester), позволяет в полной мере реализовать вышеназванные задачи.

Традиционно выделяют три уровня хранения продуктивных данных (в основном, БД): онлайн-овый, резервных данных (резервирование) — для быстрого восстановления продуктивных данных в случае отказов/сбоев/катастроф и архивных данных (как правило, эти данные выведены из состава продуктивных и постоянно уже не резервируются) — для длительного (многолетнего) дальнейшего использования.

В традиционных ИТ-инфраструктурах архивные данные, в основном, “привязываются” к конкретным приложениям, между которыми они плохо интегрированы.

HP ACA не заменяет решения по резервному копированию. Особенно это касается систем резервного копирования/восстановления (РКВ) данных, связанных с блочными данными. HP ACA работает, в основном, только с файловыми данными и может лишь дополнять системы РКВ.

Для бизнес-анализа структурированных (продуктивных) данных используются отдельные хранилища данных, т.н. DataWarehouse, которые интегрируются с продуктивными СХД через механизмы ETL (Extraction, Transformation and Loading) и др. Для ряда задач HP ACA может частично заменять DataWarehouse, предлагая как возможности единой точки сосредоточения смыслового хранения и извлечения информации, так и надежное архивное хранение.

Некоторое время назад в составе семейств HP для архивного хранения было представлено решение HP RISS (в основном, для полуструктурированных и неструктурированных данных). Оно в значительной степени ускорило поиск нужной информации (по ограниченному признакам), но не имело необходимых масштабируемости, функциональности и производительности. HP ACA полностью перекрывает и в значительной степени расширяет возможности HP RISS (в настоящее время HP RISS не продвигается на рынке, прим. ред.).

HP ACA позволяет полностью интегрировать все архивные файловые данные организации в одном хранилище, обеспечив при этом не только гораздо более широкие по-

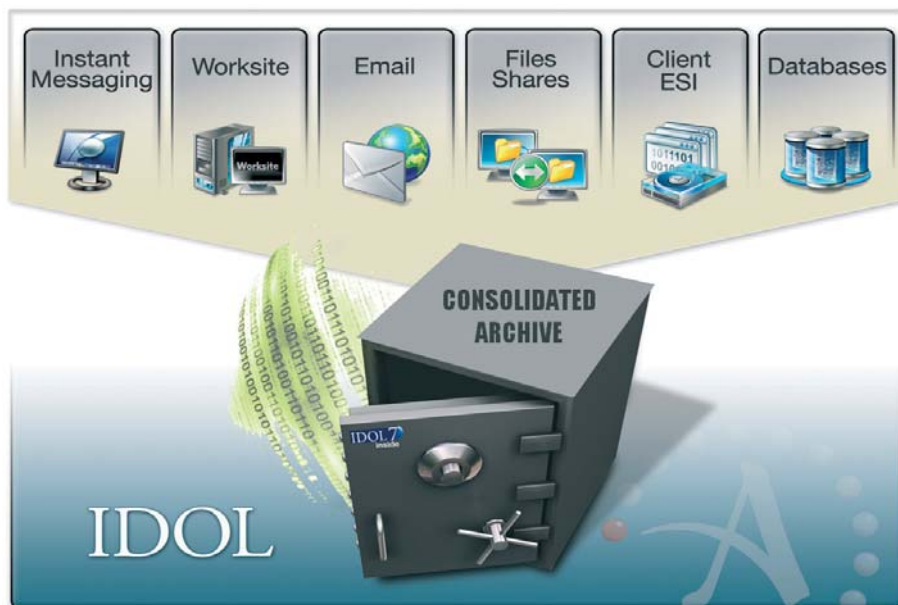


Рис. 1. HP ACA позволяет консолидировать файловые данные всей организации в одном хранилище.

исковые функции, но и возможности управления данными (в соответствии с требованиями регуляторов), восстановления отдельных типов данных первичных источников при необходимости, а также функции анализа данных — и это все в рамках одной платформы.

Интеграция данных в АСА происходит на базе единого индекса, охватывающего весь массив информации, объединяя все ее формы в единую цельную структуру, доступную для общих правил категоризации.

Следующее поколение систем Архивного хранения

Наиболее существенное значение для архива помимо традиционных ИТ-преимуществ (консолидация данных, устранение дублирования и управления жизненным циклом), является способность удовлетворять законодательным и нормативным требованиям (инициативы Юридического соответствия, eGovernance и Электронного документооборота).

АСА предоставляет единственный в отрасли слой интеллектуального управления, из которого предприятия могут обеспечить их соблюдение, раскрытие электронной документации записей, инициативы в области управления непосредственно из архива данных. Все модули Защиты данных Autonomy работают непосредственно с архивными данными, а также взаимодействуют с настольными компьютерами, приложениями, базами данных и общими папками на предприятии (рис. 1), которые охватывают весь спектр доступных информационных источников, в едином надежном хранилище с рядом новаторских функций, столь необходимых современным предприятиям:

- модули *Legal Hold, Investigator & Early Case Assessment* и *Introspect* используются для раскрытия электронной документации (eDiscovery);
- модуль *Supervisor* используется для рассмотрения соблюдения политики предварительного рассмотрения несоответствий;
- модуль *Control Point* используется для управления всеми данными предприятия;
- модуль *Records Manager* используется для управления корпоративными, юридическими и государственными записями.

Все модули управления Autonomy Meaning Based Governance работают в непосредственном взаимодействии с архивными данными, а также со всеми настольными компьютерами, приложениями, базами данных и корпоративными файловыми ресурсами. АСА позволяет автоматически регулировать жизненный цикл любого организационного актива в зависимости от его содержания, и вне зависимости от того, где он находится. С использованием АСА в качестве основной технологии организации получают более широкий и полный процесс взаимодействия с данными для управления рисками, работой с электронной доку-

ментацией (eDiscovery), управлением записями, корпоративным управлением базой знаний.

Работающий на IDOL10 (интеллектуальный слой управления данными) АСА использует алгоритмы управления на основе содержания для распознавания концепций и контекста в рамках всех форм информации. С пониманием содержания АСА автоматически применяет политики управления, основанные на понимании смысла того, что находится в анализируемой информации. Заказчики получают существенный выигрыш от широких возможностей масштабируемости, мирового класса безопасности, информационного индексирования и передовых аналитических средств.

Информационная платформа следующего поколения Autonomy создана для эпохи "человеческой" информации. IDOL 10 является единым слоем обработки, что позволяет организации извлечь смысл и взаимодействовать со всеми формами информации, в том числе аудио-, видео-, социальными сетями, электронной почтой и веб-содержимым, а также структурированными данными, такими как логи пользовательских транзакций и сведения на основе данных датчиков. Платформа сочетает в себе программное обеспечение Autonomy для автоматической обработки и понимания неструктурированных данных с аналитическим движком высокой производительности в реальном времени для высокоструктурированных данных от Vertica компании HP, а также имеет более 400 коннекторов к различным БД и экспертным системам.

Делая все данные доступными, коннекторы IDOL 10 управляют данными, метаданными и безопасностью информации в целевых архивах из более чем 400 источников информации, включая электронную почту, IM, файловые ресурсы, системы управления контентом, SharePoint, аудио-, видео-, веб-сайты и базы данных. АСА использует инновационную архитектуру сплит-ячеек: держит

две безопасные копии на каждую единицу данных с единым дизайном хранения экземпляра для управления распределенными источниками и вложениями, а также и гибридной архитектурой, что позволяет дополнить on-site доступ возможностями облачного хранения для долгосрочного управления данными.

Прозрачный доступ

АСА предоставляет пользователям прозрачный мгновенный доступ к данным, что является серьезным преимуществом для организаций, которые могут быстро внедрить стратегии мобильного управления и облачных вычислений. АСА Anywhere Access позволяет пользователям мгновенно найти архивную информацию с мобильных устройств, планшетов и настольных компьютеров без установки клиентского программного обеспечения. IDOL обеспечивает мощный, концептуальный поиск по всем типам архивных данных. Конечные пользователи могут взаимодействовать с архивной информацией, находясь в то же время в привычном приложении. Сообщения электронной почты просматриваются в родном формате (PST для Outlook и NSF для Lotus Notes). С консолидированным архивом Autonomy организации могут выйти за рамки подходов хранения, таких как stub/ folder-based access, чтобы поднять значение архива от простого хранилища до активной базы знаний предприятия, которая плавно соединит пользователей с бизнес-аналитикой (рис. 2). Например, появляется возможность анализа деятельности каждого работника с целью более точного формирования профиля его специализации и интересов. На основе данных внешних медийных источников и социальных сетей можно в реальном времени оценить успешность проводимых маркетинговых кампаний и конкурентов.

Расширенный поиск и аналитика

IDOL10 обеспечивает формат, хранение и независимые от языка поисковые возможности, далеко превосходящие тради-

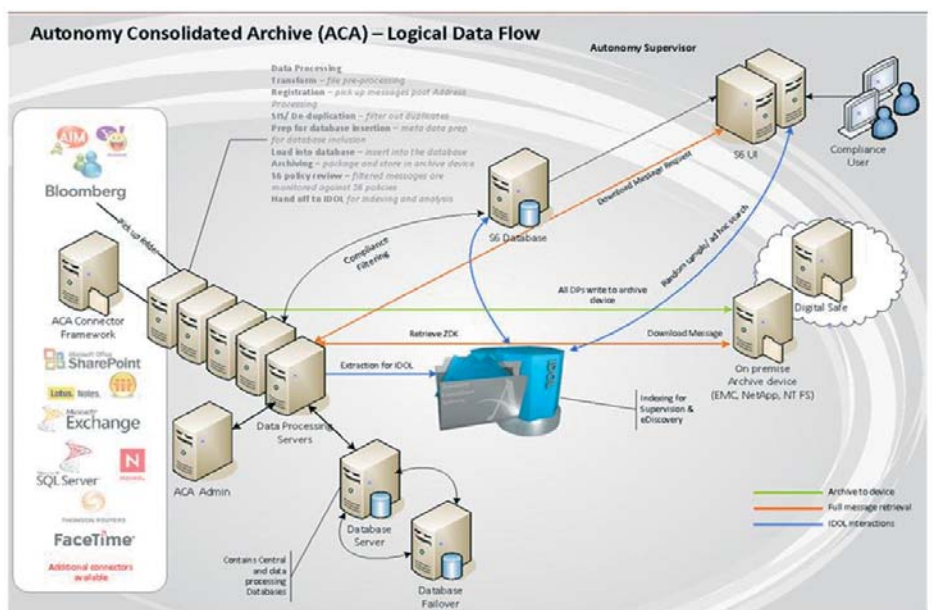


Рис. 2. Логические потоки данных, связанные с HP АСА.

ционные поисковые методы по ключевому слову и логическим маскам, которые используются до сих пор большинством традиционных технологий архива. Расширенный поиск и аналитика приводят к значительной экономии средств по сравнению с альтернативными подходами, так как значение понимания смысла информации ведет к снижению издержек, точности схемы записи и упрощению поиска необходимой информации (eDiscovery).

Autonomy устраняет барьеры между "силосными хранилищами" информации путем консолидации электронных сообщений, вложений и файлов в управляемое, масштабируемое хранилище для непрерывного соблюдения внутренних корпоративных политик и процедур управления, а также внешне утвержденных законов и правил. С IDOL, способность понимать смысл содержания поднимает значение архива из простого репозитория для защиты и хранения данных до интеллектуальной системы, способной выявлять закономерности в режиме реального времени для деловых целей всей организации. Эти возможности помогают организации обеспечить прозрачность, подотчетность и информационную защищенность на корпоративном уровне.

Функциональные возможности:

- захват больших объемов электронной почты в режиме реального времени (Outlook и Lotus Notes) и других электронных файлов;
- быстрая и эффективная обработка, индексирование, дедупликация, виртуализация и единые техники совместного хранения наряду с поддержкой ведущих устройств хранения данных, включая защищенные совместимые медиа;
- высокая гибкость с более чем 400 коннекторами для внутренних, внешних и публичных источников данных;
- проверка соответствия уровням безопасности гарантирует, что все права доступа учтены и эффективно соблюдены;
- автоматическая категоризация контента для более разумного расположения схемы архивации;
- возможности независимости от языка позволяют адаптироваться к конкретному языку и словарю, необходимому пользователю, поддерживая практически неограниченный размер словаря для любого языка;
- безопасное хранение варьируется по времени: от нескольких месяцев до десятков лет с определением политики содержания документа, исходя из его категории и безопасного расположения;
- интеграция с Desktop позволяет автономно синхронизировать, собирать PST, вводить локальные политики юридического удержания документа и пользовательского поиска как для локального индекса, так и для архива.

Преимущества:

- поддержка смешанных сред электронной почты, которая обеспечивает защиту инвестиций и гибкость управления, сокращает время реализации и затраты, а также ускоряет возврат инвестиций;
- интеграция корпоративного поиска необходимой информации (eDiscovery) с исполнительной панелью и отчетностью для обеспечения соблюдения Законов и Правил по всей организации;
- уменьшение зависимости от ИТ и конечных пользователей для поиска, сохранения и получения доказательств злонамеренной активности;
- хранение и сохранение критически важной бизнес-информации при одновременном сокращении объема аппаратных средств и систем хранения;
- минимизация бизнес-рисков путем обеспечения надлежащего использования чувствительной и ценной бизнес-информации в масштабе предприятия, в том числе личной информации, коммерческой тайны и другой непубличной информации;
- проведение интеллектуального наблюдения и мониторинга электронной почты;
- перенос устаревших данных приложений в архив.

Аппаратная реализация HP ASA

Решение может быть развернуто на стандартной индустриальной X86 серверной архитектуре и гибко масштабировано в зависимости от задач. Линейное расширение позволяет осуществлять эффективное планирование развития системы.

Для ASA небольшого объема необходимо всего два сервера: IDOL и сервер непосредственно приложения плюс необходимый объем хранилища. По мере роста добавляются серверы приложений и IDOL.

Заключение

HP ASA значительно увеличивает ценность хранимой информации при управлении критическими бизнес-процессами, консолидируя всю архивную корпоративную информацию из разных источников в едином хранилище, при этом имея возможность управления этой информацией в соответствии с требованиями регуляторов, а также поиска необходимого контента по смысловому содержанию.

IDOL, занимая лидирующую позицию на рынке решений корпоративного поиска (по оценкам IDC), в интеграции с HP ASA позволяет корпорациям, правительственным агентствам и юридическим фирмам полностью отвечать всем требованиям, предъявляемым к современным корпоративным архивам, и проактивно реагировать на все события, связанные с предоставлением информации по запросам регуляторов.

*Иван Тимофеев,
HP Россия и СНГ*

IDC: Цифровая Вселенная в цифрах

Декабрь 2012 г. — Корпорация EMC (NYSE:EMC) обнародовала результаты исследования "Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East", проведенного IDC при поддержке компании EMC (исследование проводится уже в шестой раз). Объем сгенерированных данных в 2012 г. составил 2,8 зеттабайта, а к 2020 г. эта цифра, как прогнозируется, увеличится до 40 зеттабайт, что превосходит прежние прогнозы на 14%.

Основные результаты исследования:

- объемы информации будут удваиваться каждые два года в течение следующих восьми лет. Одним из основных факторов этого роста является увеличение доли автоматически генерируемых данных: с 11% от общего объема в 2005 г. до более 40% в 2020 г.;
- большие объемы полезных данных теряются. На сегодняшний день используется менее 3% из 23% потенциально полезных данных, которые могли бы найти применение с технологиями Big Data;
- большая часть информации плохо защищена. В 2010 г. в защите нуждалось менее трети информации, а к 2020 г. доля такой информации может превысить 40%. Уровень защиты варьируется в зависимости от региона — для развивающихся рынков он гораздо ниже;
- развивающиеся рынки будут генерировать все большую долю информации в общем объеме: в 2010 г. их доля была 23% Цифровой Вселенной, но уже к 2012 г. их доля выросла до 36%. По прогнозам IDC, к 2020 г. 62% данных будет связано с развивающимися рынками;
- продолжает расти роль облачных вычислений в управлении Большими Данными (Big Data): количество серверов в мире должно вырасти в 10 раз, объем данных, управляемых напрямую корпоративными ЦОД — в 14 раз;
- в последующие несколько лет радикально изменятся типы хранившихся в облаках данных. По прогнозам IDC, к 2020 г. 46,7% данных в облаках будут относиться не к корпоративной информации, а к сфере развлечений. Основная часть — данные систем видеонаблюдения, встроенных и медицинских систем, информация с компьютеров, телефонов, бытовой электроники;
- основная часть инвестиций в управление Цифровой Вселенной сейчас приходится на Западную Европу — \$2,49 на 1 Гбайт данных. На втором месте США с инвестициями в \$1,77 на Гбайт, за которыми следуют Китай (\$1,31 на 1 Гбайт) и Индия (\$0,87 на 1 Гбайт).
- инфраструктура Цифровой Вселенной становится все более взаимосвязанной: информация уже не концентрируется в регионе, где с ней работают, да и необходимости в этом нет. По оценкам IDC, к 2020 г. облачными вычислениями будет "затронуто" почти 40% данных.