

Платформа хранения неструктурированных данных Hitachi Content Platform

Такие факторы, как быстрый рост объемов цифровой информации, потребность в консолидации различных типов данных и в обеспечении их долгосрочного хранения заставляют искать принципиально новые подходы к самой парадигме хранения редко изменяемой и неизменяемой информации. Возникает необходимость в платформе, на базе которой возможна централизация и консолидация структурированных и неструктурированных данных и организация электронных архивов. При этом платформа должна гарантиро-

вать безопасность, доступность и целостность хранимых данных.

Платформа Hitachi Content Platform (HCP) является центральным компонентом для построения решений по консолидации, хранению и управлению данными. Система представляет собой единый программно-аппаратный комплекс, состоящий из узлов хранения данных, построенных на серверах платформы x86, и внешней системы хранения данных, в случае старшей модели HCP500. В младшей модели HCP300, полезный объем предостав-

ляется внутренними дисками SATA большого объема. Мощный программный функционал позволяет решать широкий спектр задач по хранению, обеспечению безопасности и доступности содержимого. Широчайшие возможности платформы не ограничивают сферу ее применения только созданием электронных архивов, а позволяют использовать ее для создания облачных хранилищ, территориально распределенных файловых репозиторий и многих других задач.

Ключевые возможности

Платформа использует современные возможности виртуализации и может быть разделена на множество виртуальных логических разделов (tenants), каждый из которых будет иметь свои пространства имен, хранить свои группы объектов со своими политиками хранения. Доступ к каждому такому разделу и объектам, так же как управление ими, может осуществляться независимо.

Платформа HCP использует открытые стандарты и протоколы доступа: CIFS и NFS для прямого клиентского доступа, HTTP и WebDAV для взаимодействия и интеграцией с приложениями, NDMP для опциональной возможности резервного копирования архива на ленточные носители.

Каждый хранящийся в системе файл представляет собой объект. Помимо са-

мых данных, в системе хранятся различные метаданные — как системные, например время создания, размер и срок хранения, так и пользовательские метаданные, которые могут формироваться пользователями или приложениями, оперирующими этим объектом.

Подлинность хранящихся данных гарантируется специальными алгоритмами — при записи объекта в систему происходит расчет хэш-функции этого объекта, после чего эта информация записывается в метаданные. Каждый раз при обращении происходит сверка хэш-функций, что и гарантирует подлинность. Механизм расчета хэш-функций может быть сконфигурирован исходя из требований предприятия к стандартам шифрования.

Безопасность и защита данных



Внешний вид платформы Hitachi Content Platform

Платформа HCP обеспечивает безопасность и высокий уровень защиты данных. Диски системы объединены в массив RAID-6, гарантирующий сохранность данных при выходе из строя любых двух дисков в группе. Периодический аудит хранящихся объектов обеспечивает целостность информации. Технологии репликации на уровне объектов через глобальную сеть позволяют решить задачу защиты данных от катастроф.

Срок хранения данных задается специальными политиками, которые позволяют определить критерии для сроков доступности объектов. Политики могут формироваться приложениями или пользователями, например, можно установить, что файлы опреде-

ленного типа должны храниться определенное ограниченное время.

Система обеспечивает хранение нескольких версий одного и того же объекта, что, например, позволяет отслеживать весь жизненный цикл документа. Механизмы NENR (невозможность перезаписи и удаления) и WORM (одна запись, многократное чтение), которые можно задействовать, способны гарантировать неизменяемость объектов. Это дает возможность использовать платформу для хранения информации, подлежащей нормативному регулированию.

Современные технологии обеспечивают высокую эффективность хранения, которая достигается за счет

встроенного механизма исключения дубликатов объектов, функций аппаратной компрессии и дедубликации.

Механизм индексации метаданных позволяет проводить оперативный поиск объектов по всему объему репози-

тория. Интеграция с продуктом Hitachi Data Discovery Suite позволяет обеспечить еще более широкие возможности поиска, в том числе по содержимому документов.

Примеры практического использования Hitachi Content Platform

Архивирование файлов

Создание надежного хранилища для большого количества файлов зачастую является нетривиальной задачей, так как одновременно с прозрачностью доступа к файлам необходимо обеспечить приемлемый уровень безопасности хранения и иметь возможность управления и масштабирования архивного хранилища. Интеграция платформы HCP с программными решениями для архивации (CommVault, Symantec Enterprise Vault) позволяет создать файловый архив, удовлетворяющий современным требованиям по доступности, масштабированию и безопасности. Использование политик миграции данных из разнообразных файловых ресурсов (файловые серверы Windows, файловые устройства HNAS, NetApp и т.д.) позволяют перемещать в зону архивации только те объекты, которые

соответствуют определенным критериям — например, файлы, которые не были изменены последние полгода или к которым не обращались более 3 месяцев. Создание на местах хранения файлов ссылок на новое расположение объекта в архиве обеспечивает полную прозрачность для пользователя — файл будет доступен по этой ссылке в течении всего времени хранения. Дедубликация и компрессия обеспечивают более экономичное использование дисковых ресурсов, что замедляет рост объемов архивов и существенно снижает стоимость хранения данных. Использование продукта HDI совместно с платформой HCP позволяет создавать многоуровневые территориально-распределенные архивы, которые быстро и эффективно масштабируются исходя из потребностей предприятия.

Архивирование почты

Сообщения электронной почты являются неотъемлемым элементом современных коммуникаций как в среде бизнеса, так и при повседневном общении. Почтовые ящики имеют тенденцию к переполнению, что требует от пользователя создания архивных папок и других операций по оптимизации хранения сообщений. С внедрением решения на базе Hitachi Content Platform можно быстро и эффективно решить все проблемы, сопутствующие интенсивному использованию электронной почты. Использование программного интеграционного модуля, например, с Microsoft Exchange Server позволяет обеспечить архивацию сообщений по расписанию, при этом архивированные сообщения будут присутствовать в почтовом ящике

в виде ссылок, что обеспечивает простой и быстрый доступ к ним в случае необходимости. Вложения большого объема (документы, фотографии и т.д.) могут быть аналогичным образом помещены в архив, что существенно уменьшит размер почтового ящика. Использование этих механизмов избавляет пользователя от необходимости хранить отдельные архивные папки сообщений в виде локальных файлов PST и выполнять архивацию вручную, по сути предоставляя почтовый ящик без ограничений по объему. При этом из-за удаления вложений и замены их на прямые ссылки значительно замедляются темпы роста базы данных почтовых сообщений, что существенно уменьшает время резервного копирования почтовых баз.

Платформа хранения неструктурированных данных Hitachi Content Platform

Цифровой архив документов

Тенденция стремительного роста объема информации, представленной в электронном виде и, как следствие, потребность электронных систем для хранения цифровой информации определяет создание надежных решений для централизованного хранения большого количества объектов. Платформа HCP идеально подходит для обеспече-

ния длительного и надежного хранения больших объемов неструктурированных данных отсканированных бумажных документов, фотоотпечатков, оцифрованных аудио- и видео-данных с устаревших носителей. Этот функционал может быть востребован в целом ряде областей, например, в геологической разведке или в публичных библиотеках.

Хранение медицинской информации

Автоматизация процессов в сфере медицинских услуг в данный момент является одним из самых перспективных направлений в области ИТ. К медицинским данным предъявляются высокие требования по безопасности и доступности. При этом очень важным процессом является консолидация всей возможной информации о пациентах, оказанных услугах и результатах, чтобы можно было в несколько простых шагов отследить всю медицинскую историю человека. На платформе Hitachi Content Platform построено специализированное решение Hitachi Clinical Repository (HCR), которое предоставляет уникальные возможности для медицинских учреждений.

HCR — унифицированная платформа для хранения всех типов медицинских данных: медицинских карт, историй болезни, результатов клинических исследований и многого другого. Интеграция с автоматизированными медицинскими информационными системами (МИС) позволяет хранить все востребованные документы в едином хранилище. Индексация содержимого средствами Hitachi Data Discovery Suite делает возможным быстрый поиск по всему репозиторию. Решение HCR поддерживает работу со специализированными протоколами HL7 и DICOM, что обеспечивает глубокую интеграцию с медицинскими приборами (томографами, аппаратами МРТ и т.д.).

Технические характеристики

	Hitachi Content Platform			Hitachi Data Ingestor		
	HCP300	HCP500	HCP500XL	HDI cluster	HDI Single Node	HDI VMware Appliance
Полезный объем	4ТБ - 84ТБ (DPL1) 9ТБ - 140ТБ (DPL2)	18ТБ - 5ПБ	20ТБ - 40ПБ	6ТБ	4ТБ	2ТБ
Количество нод	4 - 20	4 - 80	40 - 80	2	1	Н/Д
Внутренние диски	6 x 3.5" (1ТБ/2ТБ)	-	6 x 3.5" (500ГБ)	8x1ТБ	6x1ТБ	Н/Д
SAN/СХД	Нет	Да*	Да*	Да*	Нет	Н/Д
Возможности HCP и HDI	Полный пакет опций HCP доступен для всех моделей без ограничений			Полный пакет опций HDI доступен для всех моделей без ограничений		

* поддержка продуктов Hitachi VSP/HUS/AMS/USP-V