

Centera

— хранилище распределенной архитектуры для постоянного контента

В конце весны 2002 г. компания EMC анонсировала новый класс устройств – EMC Centera, ориентированный прежде всего на работу с неизменным (фиксированным/постоянным) контентом в составе справочно-информационных и архивных систем. Отношение к новому продукту у многих было не совсем понятным – поддержат ли новую разработку основные поставщики ПО, работающие в сфере электронного документооборота. Сейчас ситуация стала более ясной – уже 50 поставщиков ПО (Documentum, Hummingbird, Legato, Kodak Medical Imaging, KPMG, FileTek, Fujitsu – и этот список увеличивается) поддержали новую разработку, а 20 из них уже выпустили свои продукты на рынок.

И фактически данная разработка открыла новый тип хранения на основе промышленно поставляемых устройств – объектно-ориентированное хранение.

Введение

Стремительное развитие современного информационного общества приводит к росту информации в геометрической прогрессии ежегодно. Вместе с тем, структура хранимой информации и ее состав также претерпевают существенные изменения. Возможности существующих технологий, с одной стороны, и потребности самого общества, с другой, приводят к тому, что все в большей степени начинают сохранять для последующего использования информацию, накопленную в предшествующие годы. По данным исследований, проведенных университетом Berkeley (USA, California, School of Information Management & Systems Research Study) к 2010 г. основной объем хранимой человечеством информации будет использоваться в информационно-поисковых и справочных системах (рис. 1).

В другом обзоре – ESG (Enterprise Storage Group) – среди самых больших американских предприятий и правительственных объектов установили, что транзакционные данные (ориентированные на базы дан-

ных, брокерские транзакции и т.п.) возрастают на 62 % в год. Примерный рост в справочной информации прогнозируется в размере 92 % в год. Прогнозом этого отчета стало то, что к 2004 г. эти два типа данных будут приблизительно равны в требуемом необходимом объеме хранения, и справочная информация будет составлять 54 % общего количества в 2005 г. При этом общий объем рынка “постоянной” информации увеличивается на 36% в год, несмотря на снижение цен.

Фиксированный/постоянный контент

Фиксированный контент (Fixed Content), или “ссылочная информация” (Reference Information) – это цифровые данные, которые были созданы в результате какой-либо деятельности, но теперь являются статическими (то есть, никогда не будут изменяться). Корпорации и правительства хранят эти цифровые активы как справочные и как основы для возможного развития бизнеса в будущем.

Фиксированный контент имеет отношение к цифровым активам, которые сохраняются в течение многих лет или десятилетий в качестве ссылок/справочников и архивных данных. Примерами являются: цифровые документы (финансовые отчеты, контракты, платежные документы), фотографии, медицинские снимки (рентгеновские, MRI и др.), медианосители (включая видеонаблюдения), аудио (включая речевую почту или MP3 файлы), геофизические данные (включая сейсмические, астрономические или географиче-

ские), другие электронные активы (типа книг, руководств, писем и вложений, презентаций или электронных таблиц) и CAD/CAM проекты. Все они имеют общность в том, что после того как контент получен (например, видео произведено, отредактировано и подготовлено для распространения), его хранят очень длительное время. К нему могут обращаться тысячи раз тысячи пользователей, но его содержимое остается неизменным.

Фиксированный контент хранится в течение длительного периода (многих лет или десятилетий), который зависит от типа требований к данным документам, установленных: правительством, рамками исследований или политикой компании. Большинство информации, хранящейся сегодня в виде фиксированного контента, доступна в режиме off-line или near-line. Запросы, диктуемые временем, выдвигают требования по доступу к нему в режиме on-line – предоставление нового уровня сервиса пользователям. Например, большинство клиентов готовы платить за то, чтобы иметь возможность онлайн-доступа к их платежным документам.

Контентно-адресуемое хранение

Centera в дополнение к существующим типам систем хранения (SAN or NAS, DAS) создает новую категорию, которую EMC называет контентно-адресуемым хранением – Content Addressed Storage (CAS). Проще говоря, это программно-аппаратное решение, специально разработанное для хранения и управления “ссылочной информацией”, или фиксированным контентом.

Требования к CAS отличаются от требований для NAS, SAN или DAS. CAS отличается от них следующим (рис. 2). SAN строится на основе блочного доступа к данным,

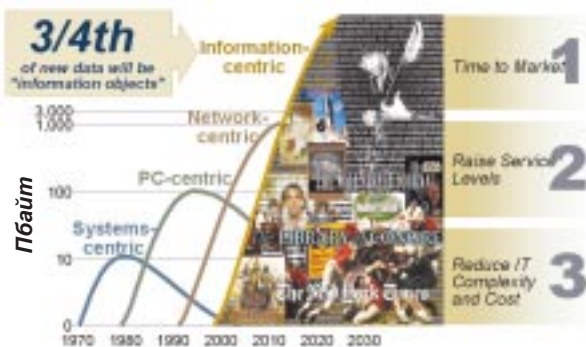


Рис. 1. Прогноз накапливаемой человечеством информации.

используемым в OLTP, хранилищах данных (data warehousing) и других транзакционных приложениях. В SAN окружении данные являются обычно переменными, и адрес хранения данных связан с их местоположением. NAS ориентирован на совместное раздельное использование файлов, содержимое которых является переменным (например, CAD файл постоянно редактируется несколькими пользователями), и его адрес, определяемый файловой системой, зависит от его местоположения. CAS имеет дело с объектами, их адрес, создаваемый самим объектом, и есть независимое его местоположение. Вместо того, чтобы сосредотачиваться на транзакционной производительности, CAS хранение ориентировано на масштабирование в режиме онлайн (без останова приложений), сохранение подлинности и целостности данных в течение длительного периода времени. При этом, как показали исследования EMC, адресация к содержимому, а не к его месторасположению, может значительно повысить объем

	SAN	NAS	CAS
Typical Applications	OLTP, DW	File Sharing, Collaboration	Content Management
Fixed or Updateable?	Updateable		Fixed
Key Technology Issue	Transaction/Performance	Edit/Share	Scalability, Longevity, Authenticity
Type of Address	Location-based		Content-based
Type of Data Stored	Volume	File	Object + Metadata

Рис. 2. Новая категория сетевого информационного хранения – CAS.

хранимых данных при тех же ресурсах и снизить сложность управления.

Когда объект впервые записывается на Centera приложением, Centera создает адрес, используя контентно-адресуемый генерирующий алгоритм. Адрес вместе со связанными с объектом метаданными (дата создания, размер и первоначальное – оригинальное – имя файла), в формате XML, помещается в файл метаданных, названный файлом описания контента (Content Description File – CDF), размером 128 бит. И этот CDF возвращается приложению в качестве идентификатора объекта или его слепка (или “отпечатка пальца”) для сохраненного объекта. Независимо от того, где объект расположен, к нему можно всегда обратиться по этому уникальному адресу. Это то, почему контентная адресация Centera не зависит от местоположения. Никакой другой объект не будет иметь тот же самый CDF и все уникальные объекты сохраняются только однажды.

Архитектура Centera

Архитектура Centera представляется как связанное IP сетью множество узлов хранения и узлов доступа (рис. 3) и называется – RAIN (Redundant Arrays of Independent Nodes – избыточное множество независимых узлов). Надежность – основное концептуальное ее свойство – отдельные отказы накопителей, узлов

RAIN Storage Architecture (Redundant Arrays of Independent Nodes)

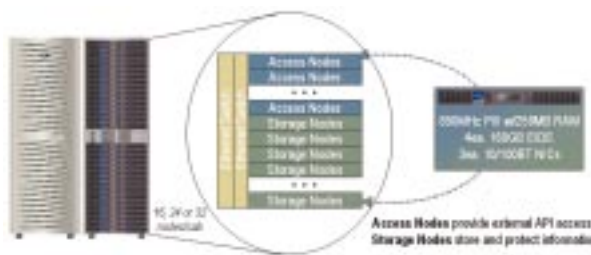


Рис. 3. Архитектура Centera: связанное множество узлов хранения и узлов доступа – RAIN.

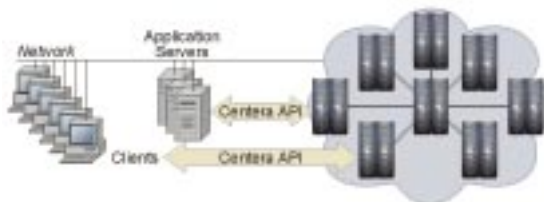
или коммутаторов незаметны пользователям. В зависимости от отказавшего элемента происходит его автоматическая регенерация (или обход коммутатора) для продолжения процесса. Каждая стойка укомплектовывается двумя дополнительными резервными (аккумуляторными) источниками питания, на каждый из которых подсоединена половина узлов. И в целом такая архитектура не имеет какой-либо непродублированной точки отказа.

Centera самоконфигурируется и самоуправляется. При инсталляции она автоматически конфигурирует все внутренние IP адреса. Независимо от размера комплектации Centera всегда представляет только один IP адрес прикладным серверам. Любой запрос на обслуживание может идти в любой узел доступа и получить обслуживание.

Программное обеспечение Centera – результат более чем 13-летней работы. Это связанное программно-дисковое решение обеспечивает простое использование и управление хранимого контента. Программное обеспечение Centera “наполняет” ее узлы необходимой функциональностью, на которую ориентирована Centera. Это, однако, не распределенная файловая система, как это может показаться. Centera не имеет никакого в обычном понимании дополнительного программного обеспечения для обеспечения взаимодействия между узлами. Размещение объектов определяется программным обеспечением Centera и всегда делается “поверх” узлов и коммутаторов. Администратор системы не должен определять путь файловой системы или основную структуру LUN.

Базовая комплектация Centera – 16 узлов (высотой 1U), каждый с 850 MHz Pentium III процессором с 256 MB RAM, 4x160 GB IDE накопителями, тремя 10/100 Mb Ethernet адаптерами и программным обеспечением Centera – CentraStar, с общим объемом хранения приблизительно 10 Тбайт. Каждый модуль может конфигурироваться как

Архитектура для петабайтной масштабируемости



- From ~5 to 10TB usable per cabinet (information mirroring)
 - 10/100BT Ethernet links to external LAN
- Up to 16 cabinets per Centera Cluster (~160TB usable)
 - GbE interconnects between cabinets
- Multiple clusters will be used to achieve (~1PB usable)

Рис. 4. Уровни масштабируемости Centera.

узел доступа или узел хранения. Узел доступа используется для того, чтобы взаимодействовать с прикладным сервером через программный интерфейс приложения Centera (Centera API), а узел хранения – чтобы фактически хранить объекты. В узлах доступа диски используются для кэширования объектов. До 32 узлов может быть установлено в 72” стойку вместе со двойными Ethernet коммутаторами, которые соединяют все узлы хранения друг с другом 10/100 Mb кабелем, и поддерживают двойные 10 Gb Ethernet связи для соединения стоек. Полная стойка имеет общую емкость 19,2 Тбайт или защищенных 9,6 Тбайт, состоящую из разделенных заменяемых (защищенных) блоков по 2,5 Тбайт.

Centera копирует объекты или локально, или удаленно для обеспечения высокой готовности и катастрофоустойчивости. Когда администратор устанавливает политику реплицирования, Centera перемещает все объекты автоматически.

До 16 стоек может конфигурироваться последовательно, используя сдвоенные GE порты с общей емкостью 320 Тбайт (160 Тбайт защищенной) в кластере (рис. 4). До семи кластеров может быть связано в Centera домен, обеспечивая более чем 1 Пбайт защищенного хранения (около 10 Пбайт требуется для хранения всей напечатанной информации в американской Библиотеке Конгресса).

Простое управление хранением делает возможным управление полным кластером Centera не более чем одним администратором, тем самым значительно сокращая ТСО. Любая дополнительная емкость, которая добавляется, самоконфигурируется.

Программное обеспечение Centera проверяет объекты на целостность в течение периодов простоя. Взаимодействие между узлом доступа и узлами хранения происходит через внутреннюю LAN (100 Mbps Ethernet), тогда как взаимодействие между узлами доступа и прикладным сервером – через простой программный интерфейс API Centera (на основе C-библиотеки, приобретаемой дополнительно).

Centera поддерживает платформы: Windows NT/2000, Solaris 8, HP/UX 11i и Linux. В 2003 г. запланирована поддержка AIX.

Мониторинг Centera осуществляется через интерфейс командной строки (CLI), обеспечивая информацией о: доступной емкости (Гбайт) и как процент от общей защищенной емкости; интенсивности операций read и write (Мбайт/с) на входе Centera; времени реакции при обслуживании запроса к Centera.

EMC предлагает минимальную конфигурацию (из 16 узлов) приблизительно за \$204 тыс. (10 Тбайт – общая емкость, что соответствует 2 центам за 1 Мбайт). Стартовый модуль может наращиваться по 5 Тбайт приблизительно за \$102 тыс. Полная стойка с 20 Тбайт (общая емкость) поставляется приблизительно за \$408 тыс. (цены даны без учета НДС и таможенных пошлин – прим. ред.).

Преимущества, предоставляемые Centera

CAS технология хранения данных обеспечивает следующие преимущества:

- **идентификация** — поскольку “отпечаток” контента получен из его содержания, любая попытка просто записать поверх или изменить содержание, создает новое содержание с новым ключом CDF, оставляя оригинал неизменным. Ни ошибка, ни преступное намерение не могут уничтожить или изменить содержание и запретить доступ к объекту в любое время. Подлинность объекта гарантируется, и необходимость в его копировании отсутствует, потому что нет никаких изменяемых объектов, а появляются только новые объекты;
- **независимость от месторасположения** — с наличием самоадресуемого контента отпадает необходимость в управлении содержанием в пределах файловой системы или логического тома в дисковом массиве. Поэтому, требуемые ресурсы сопровождения системы, сосредоточенные на этом типе информационной поддержки внутри IT подразделений, значительно уменьшаются;
- **устранение дублирования** — способность Centera точно идентифицировать уникальность содержимого объекта позволяет ей устранять любую попытку, связанную с возможностью повторной записи уже существующего контента, что обеспечивает значительную экономию ресурсов хранения;
- **снижение эксплуатационных затрат.** Базовая стоимость Centera превышает (в пересчете на единицу хранения) цену решений на основе ленточных и оптических носителей, однако за счет снижения затрат на администрирование и текущую эксплуатацию (в течение уже трех лет) позволяют сделать этот показатель более низким в сравнении с другими решениями*);

*) по мнению представителей компании EMC (прим.ред.).

E-Mail Archiving Example



- 800,000 e-mails / day at ~10KB each
- Three-plus years of e-mail history indexed and “archived”
- High access density: hundreds of simultaneous threads (# writes > # reads)
- 5% read retrievals at random intervals
- Bandwidth: < 5MB/s
- Additional challenge: Archiving FAX, voicemail, web transactions, etc.

Рис. 5. Архив электронной почты на базе Centera.

Check Imaging Example



- Take care of long-term retention and disaster recovery requirements
- Minimize management costs
- Distribute information as new service
- Meet regulatory requirements

Рис. 6. Архив платежных документов на базе Centera.

• **практически неограниченная масштабируемость.** Если производительность/емкость становятся узким местом, добавляются узлы доступа/хранения в режиме on-line.

Примеры реализаций

1. **Архив электронной почты.** Пример реализации Centera в компании “Global 1000 Financial Services” (рис. 5).
2. **Архив платежных документов.** Реализация архива на базе Centera как альтернатива архиву на основе иерархической системы управления (рис. 6).

Неполный список продуктов – независимых поставщиков ПО, поддержавших новую разработку, – приведен в таблице.

Заключение

Разработчики Centera не ставят своей целью заменить ею все существующие на рынке решения для архивного хранения, но при правильном применении, в ряде случаев, ее потенциальные возможности способны произвести небольшую революцию при организации массовых хранилищ данных.

Список выпущенных на рынок продуктов, поддерживающих Centera

Партнер	Продукт	Применение	Поддерживаемые ОС
Amicas	ForeverPriors	Medical Imaging	Windows NT/2000
Artesia	Artesia TEAMS	Digital Asset Management	Solaris 2.8
BancTec	eFirst Archive	Content and workflow management	Solaris 2.8, Windows 2000, HP-UX 11
Computer Science Corp.	Check Vision	Check Imaging Processing	Solaris 2.8
Connected	TLM	Backup and restore of PC/Laptop data	Windows NT
Daa Systemhaus	ScanView	Document Mgmt and Archiving	Windows 2000/NT
Digital Concepts Inc. (DCI)	Melasoft	Distributed Content Management	Solaris 2.8, Windows NT
Documentum	4i	Enterprise Content Management	Solaris 2.8, Windows 2000/NT
EMC Avalon	AVALONidm	Intelligent Data Management	Solaris 2.8 with VxFS, SGI IRIX with XFS
Enigma Data Systems	PARS SmartMove	Project Archiving and RetrievalSystem for Seismic information	Solaris 2.8 Solaris 2.8
EPSIIA	DX Server	Intelligent Data Management	Windows 2000/NT
H.A.Thomson	StorageSwitch	Document Management (for Financial Services Statement Output)	Windows 2000
HyperTrust	Send2 Services	Application Gateway	Windows NT
IXOS	eCONserver	Trusted Internet transactions	Windows NT
KVS	Enterprise Vault	Archiving for Microsoft Exchange, Lotus Notes, SAP, and Siebel	Windows 2000/NT
Legato	Xtender (DiskXtender, ApplicationXtender, EmailXtender)	Archiving for MS Exchange and SharePoint Portal Server	Windows 2000/NT, Solaris 2.8, HP-UX 11i, AIX 5.1
Medi-Face	PiView	Data Management/HSM	Windows 2000/NT
Merge	eFilm	Medical Imaging	Windows NT
Scientific Software Inc.	CyberLab	Medical Imaging	Windows 2000/NT
Tower Technology	IDM and WebCapture	Biotech Content Management	Windows 2000/NT
TrueArc	ForeMost Enterprise 3	Content management/Web capture	Windows NT
Unisys	Payment Information Archive (PIA)	Record Mgmt.	Windows 2000/NT
		Check Imaging	Windows 2000/NT

HP лидирует в продажах систем класса HPC

2 декабря 2002 г. – По данным проводимого аналитической компанией IDC ежеквартального исследования Technical Computing Quarterly Update за III кв. 2002 года (<http://idctracker.com>), HP вновь стала лидером по доходам от продаж высокопроизводительных вычислительных систем (High Performance Computing, HPC), заняв 35% рынка. Компания HP показала самый высокий уровень доходов в трех из четырех выделяемых IDC сегментов рынка систем для инженерно-технических расчетов: 46% в сегменте систем центров обработки данных (Technical

Capability Computers – системы для решения самых неотложных задач), 62% в сегменте систем масштаба предприятия (Technical Enterprise Computers – системы стоимостью от 1 млн дол.) и 27% в сегменте систем масштаба отдела (Technical Departmental Computers – системы стоимостью до 250 тыс. дол.).

По данным IDC, общий объем продаж HPC в III кв. составил 1 млрд дол.

“Sun Storage Forum”

В середине ноября 2002 г. компания Sun Microsystems провела Sun Storage Forum, посвященный вопросам создания, эксплуатации и развития систем хранения данных. На форуме были затронуты как технические, так и бизнес-аспекты внедрения и эксплуатации

крупных информационных систем, а также представлены стратегические альянсы в области технологий сис-



тем хранения данных с лидерами отрасли. На форуме выступили: Сергей Тарасов (Глава представительства, Sun

Microsystems, СНГ), Пол Галлет (Paul Gullett, Директор департамента решений по управлению информацией, Sun Microsystems, EEA), Фил Джонс (Phil Jones, Менеджер по развитию бизнеса, HDS, Европа), Найджел Уильямс (Nigel William, Директор по развитию бизнеса, Legato, Европа) и др.