

Вечные архивы на оптике



Антон Петроченко — ведущий инженер по СХД компании Panasonic Россия.

Половина данных — «холодные»

По оценкам аналитических агентств, объем данных к 2020 году вырастет в 10 раз по сравнению с 2013 годом. При этом половина объема этих данных будет «холодной» — это означает, что данные необходимо хранить, но доступ к ним очень редкий — раз в несколько лет. Примером могут служить «сырые» оцифрованные данные картин, фильмов, инженерная и строительная документация заводов и фабрик, архивные документы. Такие данные экономически нецелесообразно хранить на быстрых носителях, срок жизни которых — несколько лет, а стоимость владения высока.

Поэтому продолжают развиваться архивные технологии, для которых характерны очень большие объемы и низкая цена за бит. Одной из таких технологий являются оптические диски. Хотя они и ушли из компьютеров и ноутбуков пользователей, для крупных организаций они по-прежнему актуальны с выходом новых дисков повышенной плотности. На смену Blu-ray™ пришел новый формат — Archival Disc (рис. 1), разработанный совместно Panasonic и Sony. В 2015 году вышел диск на 300 гигабайт в том же форм-факторе — привычного нам компакт-диска. Скоро должен выйти диск на 500 Гбайт, а затем — на 1 Тбайт. Таким образом, плотность возрастет в 10 раз по сравнению с 100-ги-

габайтным Blu-ray™ диском, используемым в настоящее время.

Инженеры Panasonic и Sony смогли найти потенциал развития оптической записи Blu-ray, увеличив линейную плотность в одном слое данных и устранив появившиеся помехи, что позволило сохранить приемлемый уровень соотношения сигнал-шум и создать новый формат — Archival Disc с большей емкостью и той же длиной лазера, той же оптической системой. Новая технология Archival Disc является технологией однократной записи, которая позволяет защитить данные от ошибочной перезаписи, фальсификации, сбоя и вирусов, что как раз и нужно в хранении архивов.

Оптика более выносливая

Если речь идет о десятках лет хранения данных, то при планировании цифровых архивов необходим также риск-менеджмент. И это не только противопожарная система. В течение десятков лет могут происходить и менее вероятные события — длительное отключение электроэнергии и отопления, стихийные бедствия, разрушение здания ЦОД. Можно застраховаться от этих событий в страховых компаниях, но утерянные данные уже не восстановить. О сбережении данных техническими средствами и должен думать ИТ-профессионал, учитывая возможные риски и нивелируя их. Механизмы известны: резервное копирование, избыточность данных, хранение минимум 3 копий на 2 различных технологиях и с 1 внешней копией (стратегия 3-2-1).

Хранение на жестких дисках и магнитной ленте является одной технологией — магнитной, соответственно, одинаковый риск может их уничтожить. Поэтому для одной из копий рекомендуется использовать оптику, которая более «выносливая»: выдерживает более широкий диапазон температур, не боится воды, электромагнитного излучения. Проблемы первых поколений — боязнь царапин, ультрафиолетового излучения, малое время хранения — успешно решены, и теперь данные

могут храниться 100 лет (согласно расчетным методам ускоренного старения). Предполагаемый срок службы Archival Disc составляет 50 лет при температуре окружающей среды 50° С, и до 1000 лет при комнатной температуре.

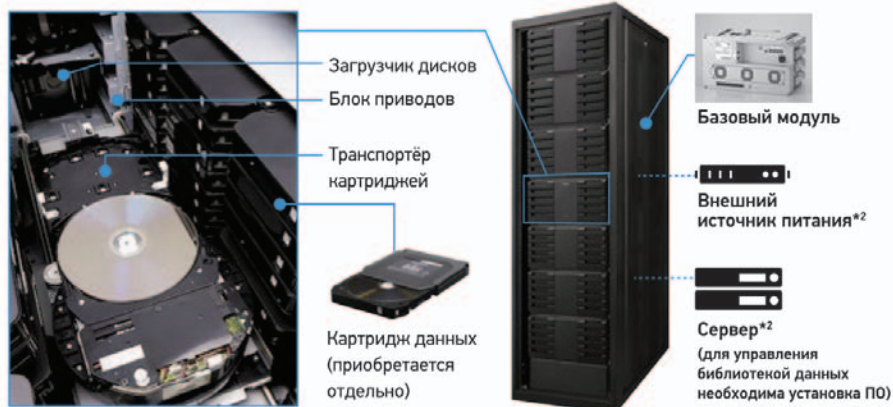
С помощью чего можно будет через 100 лет прочесть этот диск? В оптическом хранении это всегда учитывалось: любой современный привод оптических дисков может читать все предыдущие поколения: Blu-ray™ читает и DVD, и CD, который появился в 1982 году. Новый Archival Disc привод будет читать все предыдущие поколения. И эта тенденция заложена в основу концепции развития оптических технологий. Другие технологии не обладают длительной обратной совместимостью — сейчас трудно прочитать первую кассету LTO1 и подключить 20-летний жесткий диск к современному компьютеру.

Всё больше компаний выбирают, по возможности, стратегию «хранить всё», ничего не удаляя. При этом архивная часть данных компании, как правило, достаточно велика, и хранить их на жестких дисках, не говоря уже о твердотельных носителях, дорого. Гарантия на лучшие жесткие диски составляет 5 лет, после чего их рекомендуется заменять. Кроме того, они часто выходят из строя раньше этого срока. За 20 лет одну и ту же информацию можно переписать с одного жесткого диска на другой несколько раз, при этом вполне возможно, что к этим данным никто ни разу не обращался. Общая стоимость владения будет выше, чем стоимость хранения на архивных СХД на основе оптики. Кроме того, СХД на оптике менее энергозатратны, не требуют принудительного охлаждения, что также позволяет существенно экономить, особенно когда рассматривается период, исчисляемый десятками лет.

Для управления архивными данными необходимо выделить их из остальных данных: оперативные данные должны храниться на оперативных хранилищах, архивные — на архивных. Это обусловлено тем, что при



Рис. 1. Внешний вид картриджа с 12 оптическими дисками, изготовленными по технологии Archival Disc.



*2 Приобретается отдельно (возможна установка в стойку).

Рис. 2. Конструкция библиотеки.

длительном периоде хранения, измеряемом десятками лет, эксплуатационные издержки хранения данных в оперативных хранилищах значительно больше, чем в архивных.

История технологии freeze-ray™

Система freeze-ray™ была создана компанией Panasonic в сотрудничестве с Facebook. При этом использовались передовые разработки Panasonic — технология оптических дисков высокой плотности, оборудование, а также программное обеспечение, облегчающее контроль над библиотеками данных в ЦОДах. Специалисты Facebook поделились непревзойденным опытом в проектировании, развертывании, управлении и обслуживании систем хранения в дата-центрах, а также предоставляли подробные технические и пользовательские отзывы на каждом этапе внедрения продукта.

Анонс новой роботизированной библиотеки данных Panasonic freeze-ray™ на оптических дисках состоялся в январе 2016 г., доступность в России — апрель 2017 г.

Архитектура библиотеки

Минимальная конфигурация библиотеки включает (рис. 2) в себя один нижний модуль, один базовый модуль и один модуль расширения (с блоком приводов). В базовый модуль можно установить 76 картриджей по 3,6 Тбайт общим объемом 273,6 Тбайт. При установке 7 модулей (макс.) в 19" стойку система достигает максимальной емкости (1,9 Пбайт с RAID 0) при низкой стоимости за единицу хранения. Общая скорость передачи данных достигает 360–1080 Мбайт/с (фактическая скорость варьируется в зависимости от внешних условий, например, от производительности сервера) посредством добавления 6 модулей расширения с блоками приводов.

Картридж состоит из 12 оптических дисков. Обслуживающему персоналу не нужно складывать диски в картриджи — они будут поставяться с завода в собранном виде. Картридж не предполагает доступ человека — всё делается роботом внутри СХД. Поэтому нет отпечатков пальцев, мусора, царапин, и важная последовательность дисков внутри не нарушается.

Библиотека данных*1 серии LB-DH7

Название продукта	Базовый модуль	Нижний модуль	Модуль расширения (без блока приводов)	Модуль расширения (с блоком приводов)
Номер модели	LB-DH70A0R	LB-DF81Z1R	LB-DH82Z1R	LB-DF72A0R
	Минимальная конфигурация		Максимальная конфигурация (на стойку)	
Количество устанавливаемых картриджей	152		532	
Объем*2	0,5 ПБ		1,9 ПБ	
Число блоков приводов (на стойку)	2 шт.		6 шт.	
Общая скорость передачи данных*3	До 360 МБ/с		До 1080 МБ/с	
Основной интерфейс	SAS/iSCSI/FC (на сервере с управляющим ПО)			
Протокол управления	SCSI (MMC, SMC)			
При установке в 19-дюймовую стойку с панелями EIA (высота)	16U		46U	
Электроснабжение	+24 В, +12 В постоянного тока			

*1 Данной системе требуется отдельно поставляемый сервер (для управления библиотекой данных необходима установка программного обеспечения) и внешний источник питания.

*2 После физического форматирования. При использовании RAID 0.

*3 Значения получены в стандартном тесте Panasonic. Фактическая скорость записи зависит от внешних факторов, таких как производительность сервера.

Картридж данных (приобретается отдельно)

Номер модели	LM-BM36XB	LM-BM12LB
Емкость	3,6 ТБ (12 шт. 300 ГБ дисков Archival Disc)	1,2 ТБ (12 шт. 100 ГБ дисков Blu-ray Disc™)

Программное обеспечение «Data Archiver Manager» (приобретается отдельно)

Поддерживаемая ОС	Red Hat Enterprise Linux7, CentOS7
Интерфейсы	REST API (SWIFT, S3) NAS (NFS, CIFS)

Эти 12 дисков раскладываются по 12 приводам, которые подключаются в один контроллер. Таким образом, эти приводы можно собрать в RAID: 0, 5 или 6.

Соответственно, диски в картридже будут в такой же версии RAID. Файлы всегда находятся на 12 дисках, записанные небольшими блоками.

RAID-0 означает, что 1 файл записан на все 12 дисков небольшими блоками без избыточности.

RAID-5 означает, что ёмкость одного из 12 дисков выделена под избыточность (8,3%).

RAID-6 означает, что ёмкость двух из 12 дисков выделена под избыточность (16,6%).

При RAID-6 потеря двух полных дисков будет означать, что у нас по-прежнему сохранены все 100% данных.

Загрузка 1 картриджа для библиотеки LB-DH8 составляет чуть больше минуты, что приемлемо для большинства архивов: если данные хранились 10 лет, то подождать минуту это нормально. Скорость чтения и записи на картридж составляет 216 Мбайт/с для LB-DH8 и 720 Мбайт/с для LB-DH7 за счет одновременной работы 12 приводов.

По сравнению с ЛТО (накопители на магнитной ленте) оптические приводы позволяют быстро переключаться к нужному участку носителя, поэтому доступ к первому байту составляет всего 200 мс в том же слое данных и 1100 мс, если необходимо сменить слой (100-гигабайтный Blu-ray™ диск использует 3 слоя, 300-гигабайтный Archival Disc — 3 слоя на каждую из 2 сторон). В случае получения нескольких файлов с носителя это быстрее будет выполнено на оптике, т.к. ленту необходимо перематывать до нужного участка. Частый запрос небольших файлов с ЛТО сильно расходует ресурс кассеты, в отличие от бесконтактной оптической технологии, которая даже после 1 млн прочтений не ухудшает состояние оптического диска.

Интеллектуальное управление данными

При использовании специализированного ПО "Data Archiver Manager" (приобретается отдельно) систему freeze-ray™ можно легко подключить к имеющейся ИТ-инфраструктуре по локальной сети. Несколькими модулями системы и всеми картриджами

можно управлять как одним пространством имен, при этом знания, на каком картридже находится нужный файл, не требуется. Система freeze-ray™ поддерживает Linux, что обеспечивает возможность создания объектной системы хранения, подходящей для архивирования большого объема неструктурированных данных.

В 2016 г. компания Panasonic заключила соглашение с российским производителем программно-определяемого хранилища — компанией «Рэйдикс», в соответствии с которым в существующее решение RAIDIX для сверхбыстрого и надежного оперативного хранилища будет интегрирован модуль для управления и хранения холодных данных на СХД Panasonic. По правилам, определенным пользователем, данные будут отправляться в архив на оптику. Таким образом, для ИТ-архитекторов и пользователей будет предоставлен единый продукт для оперативного хранилища и архива с гибкими правилами автоматизации.

Антон Петроченко,
компания Panasonic Россия.

“Распределенные файловые системы и объектные хранилища”

Октябрь 2016 г. — В опубликованном Gartner исследовании: Magic Quadrant for Distributed File Systems and Object Storage (20 October 2016, ID: G00307798, Analyst(s): Julia Palmer, Arun Chandrasekaran, Raj Bala) в группе лидеров три компании: Dell EMC (с продуктами Dell EMC Isilon и Dell EMC Elastic Cloud Storage, ECS), IBM (с продуктами IBM Spectrum Scale и IBM Cloud Object Storage — ребрендинг Cleversafe, куплена IBM в 4-м кв. 2015 г.) и Scality (в России продает HPE и Dell EMC, см. http://www.storagenews.ru/64/HPE_Scality_%2064.pdf). Dell EMC занимает первое место как “по видению”, так и “по присутствию на рынке”.

“По видению” к группе лидеров максимально приблизилась Red Hat с продуктами Red Hat Ceph Storage и Red Hat Gluster Storage. Red Hat Ceph Storage представляет собой унифицированную платформу хранения для объектных хранилищ и облачных инфраструктур с массивами данных петабайтных размеров. Red Hat Gluster Storage является распределенной файловой системой для организации гибких служб хранения данных для датацентров на основе физических серверов, виртуальных машин, частных и общедоступных облачных сред, а также инфраструктур контейнерных приложений. Оба эти решения не требуют специализированного оборудования для развертывания, активно поддерживаются сообществом разработчиков ПО с открытым кодом и компаниями-партнерами, а также широко применяются в корпоративном секторе по всему миру.

На третьем месте “по присутствию на рынке” — HDS — с решением Hitachi Content Platform (HCP).