

Ключевые инновации Documentum 6.5:
Documentum High-Volume Server – новая
платформа для приложений, работающих с
повышенными объемами содержания

Андрей Николаев
Старший технический консультант
EMC СМА Россия и СНГ

EMC Documentum High-Volume Server

EMC²
where information lives[®]

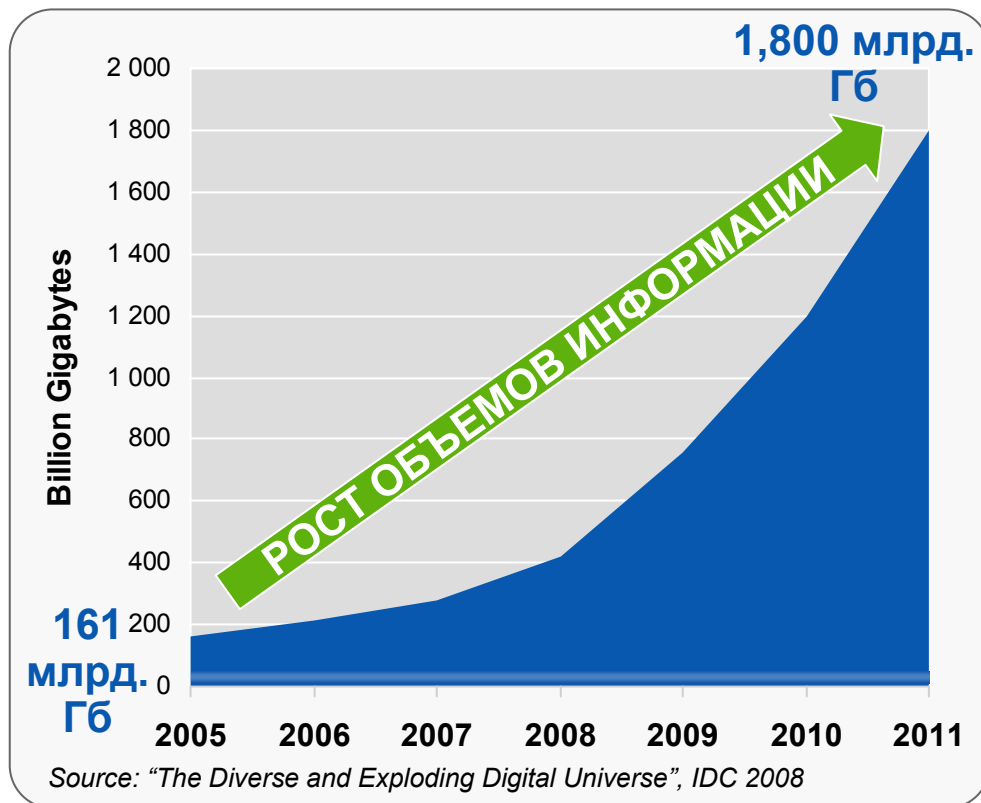


Анонс

High-Volume Server

Новая платформа для приложений с повышенными объемами содержания

Расширяющаяся цифровая вселенная



1.8 Зетабайт!

High-Volume Server (HVS) – крупным планом



Для управления большими объемами содержания необходима масштабируемая, легко управляемая и высокопроизводительная платформа

Новая платформа для приложений высокой производительности:

- Расширяемая, открытая и защищенная
- Способная быстро загружать большие объемы информации
- Эффективное хранение баз данных
- Надежный доступ к содержанию
- Низкая стоимость владения

High-Volume Server (HVS) – основные характеристики



Увеличение отдачи от ЕСМ для организаций с большими объемами содержания

- Снижение затрат на хранение и резервное копирование
 - Снижение размера базы данных
 - Уменьшение потребностей памяти и ЦПУ для базы данных
 - Более быстрое восстановление
 - Для восстановления требуется меньшая пропускная способность сети
 - Дифференцирует активное и заархивированное содержание
- Быстрая загрузка и миграция
 - Загрузка по несколько млн. объектов в день
 - Миграция – два млрд. объектов в неделю

Пример1 – архивирование изображений



Сценарий 1:

Асте Medical Imaging требуется хранить большое количество изображений

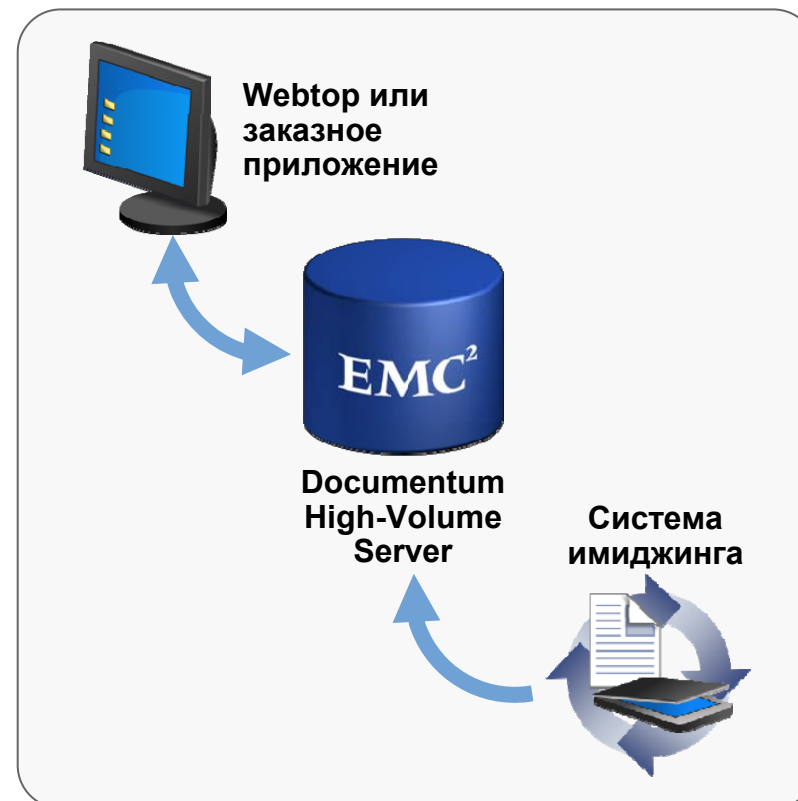
- Нужна система, способная быстро загружать большие объемы содержания
- Изображения нужно хранить в защищенной репозитории с мощными функциями контроля доступа
- Система должна быть масштабируемой и при этом экономичной
- Требуется возможность быстрого поиска изображений с простыми и расширенными поисковыми возможностями

Пример 1 – архивирование изображений

Решение:

- High-Volume Server
 - Быстрая загрузка изображений
 - Изображения хранятся в защищенном, масштабируемом репозитории
 - Списки контроля доступа на все объекты определяют права для пользователей, групп и ролей
 - Эффективное хранение и управления метаданными контролирует затраты на базу данных

- Клиент
 - Для доступа к изображениям можно использовать как клиенты Documentum (Webtop или DAM), так и заказное приложение, построенное на основе Documentum Web Services



Пример 1 – архивирование изображений



Сценарий 2:

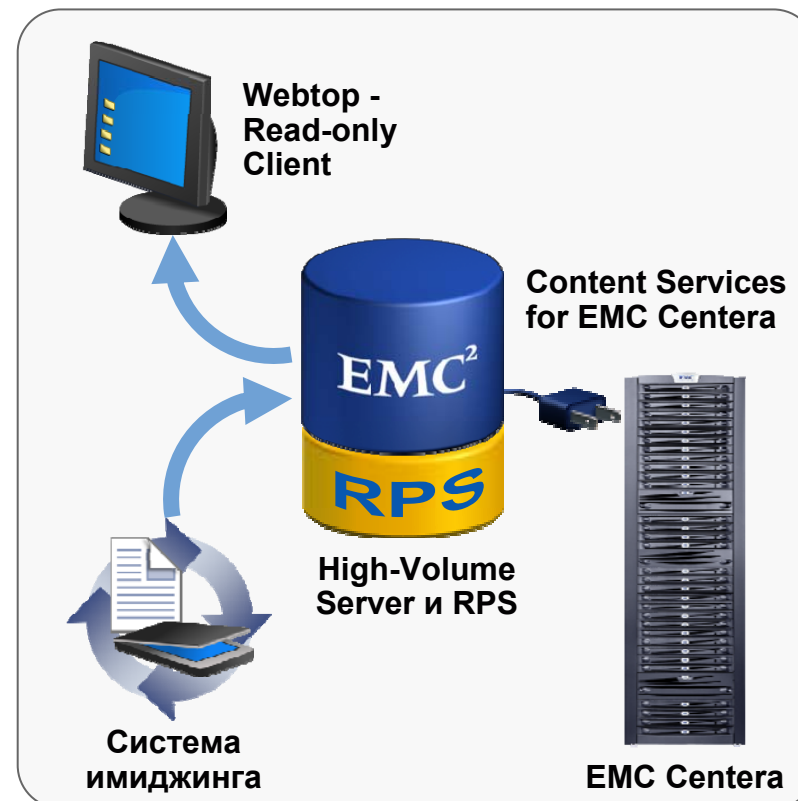
Асте Банк требуется хранить образы отозванных чеков

- Скорость загрузки должна быть достаточной для соответствия большому количеству ежедневно поступающих отозванных чеков
- Образы чекой необходимо хранить семь лет, затем они должны подвергаться цифровому уничтожению
- Для минимизации расходов требуется хранилище с адресацией по содержанию
- Требуется высоко-производительный доступ для географически распределенных представительств

High-Volume Server Пример 1

Решение:

- High-Volume Server
 - Быстрая загрузка
 - Защищенный масштабируемый репозиторий
 - Эффективное хранение метаданных и управление ими
- Retention Policy Services (RPS)
 - Хранение на основе регламентов и автоматическое удаление по истечению сроков хранения
- EMC Services for Centera
 - Соединение с системой хранения с адресацией по содержанию EMC
- CenterStage Essentials
 - Удобный поиск через навигацию с подсказками или визуальный поиск



Пример 2 – вывод систем из эксплуатации



Сценарий:

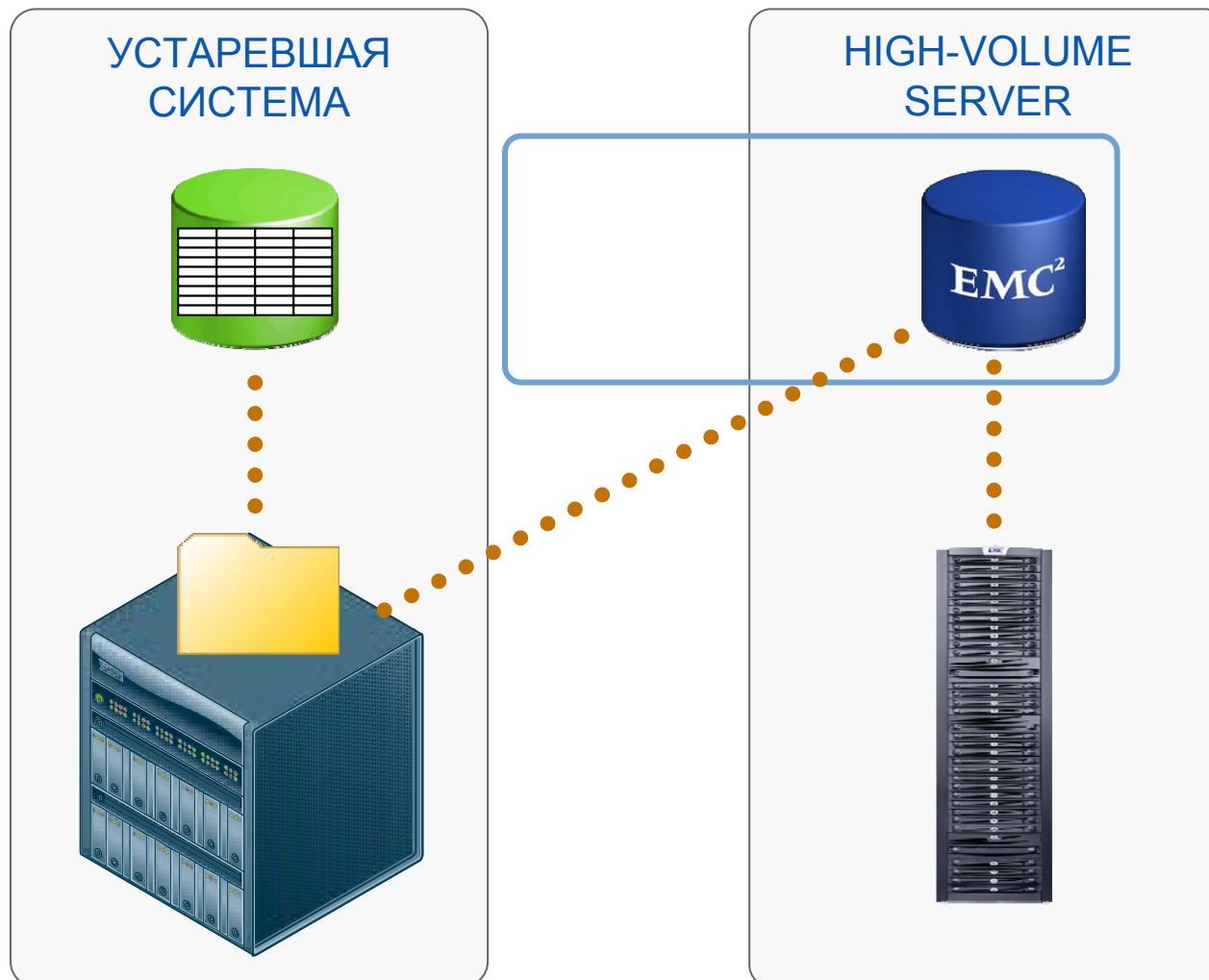
Асme Amalgamated Inc. требуется вывести из эксплуатации одну из своих устаревших систем, сохранив при этом большое количество связанных с ней данных

- Данные находятся на старом мейнфрейме
- Они содержат млрд. записей
- Содержание находится на различных файловых серверах
- Миграция данных не проводилась, т.к. было определено, что для переноса данных в новую систему понадобится несколько месяцев

Пример 2 – вывод систем из эксплуатации

Решение:

1. Создание офф-лайн раздела
2. Массовая загрузка унаследованных данных
3. Перевод раздела в он-лайн
4. Файлы содержания оставляют на месте или переносят в EMC Centera с использованием Content Storage Services
5. Полная замена устаревшего оборудования



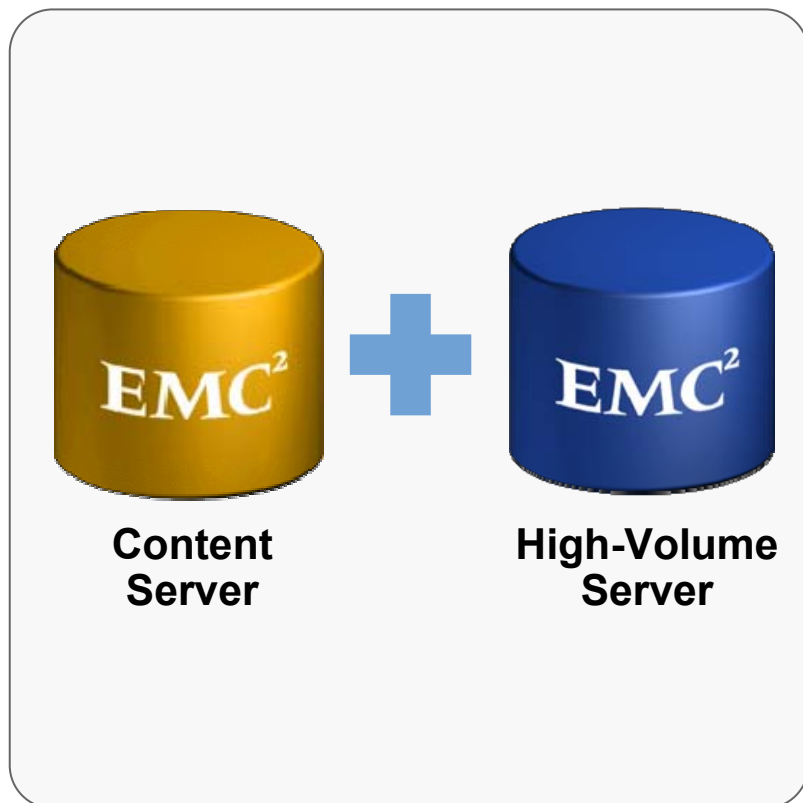


Сценарий:

Acme Shipping каждый день сохраняет в своем репозитории более 500 000 новых инвойсов и товарно-транспортных накладных и желает сократить расходы на базы данных

- Метаданные документов хранятся в СУБД, объем которой возрастает экспоненциально
- Поскольку данные являются жизненно важными для деятельности компании, резервное копирование СУБД осуществляется дважды в день, что еще увеличивает требования к хранению
- Затраты на хранение данных, ЦПУ для базы данных и ассоциированные расходы на ИТ-инфраструктуру огромны

High-Volume Server Пример 2



Сценарий:

За счет добавления High-Volume Server к Content Server, Acme Shipping сможет:

- Увеличить производительность и масштабируемость
- Уменьшить расходы
 - **Легкие объекты** существенно снижают расходы на базу данных
 - **Дробление данных** позволяет переместить более старые данные в «холодные» разделы
 - **Пакетная обработка** обеспечивает более эффективную обработку загруженного содержания

Documentum High-volume Server

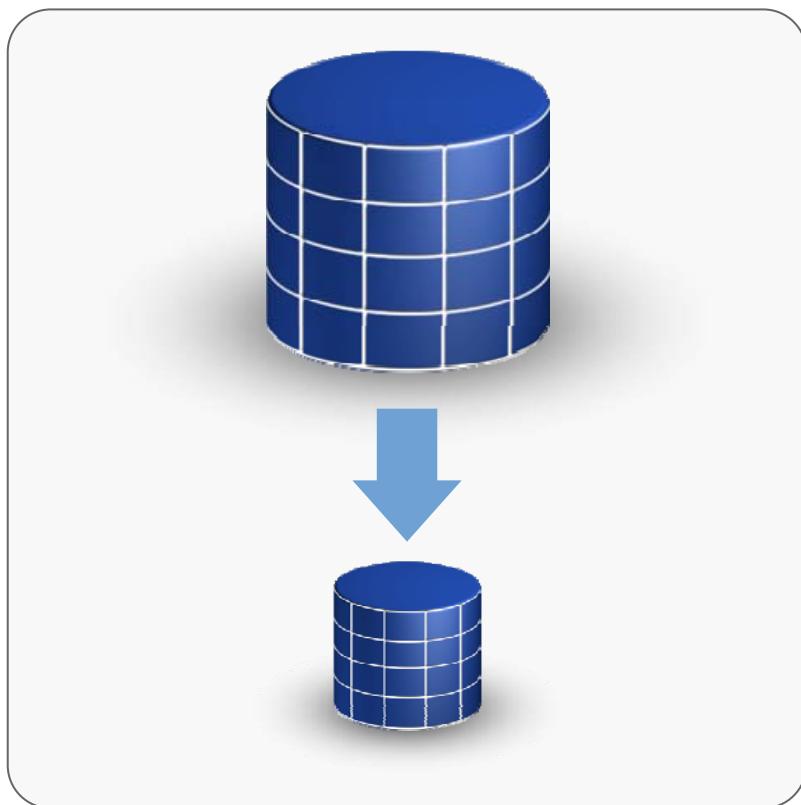


Расширяемая, открытая, защищенная архитектура с масштабируемостью и производительностью, необходимой для архивирования и транзакционно-ориентированных приложений. Обеспечивает более эффективное управление большими объемами содержания.

Основные характеристики:

- Объектная модель на основе «легких» объектов
- Пакетная обработка
- Системная валидация с использованием кэш-памяти
- Дробление данных
- Пакетная загрузка офф-лайн и скопинг
- Улучшение в области передачи содержания и использования ЦПУ

Преимущества – «легких» системные объекты



Система «легких» объектов позволяет определенным типам приложений получить преимущества от системных объектов меньшего размера

Меньший размер объекта
x Миллионы объектов

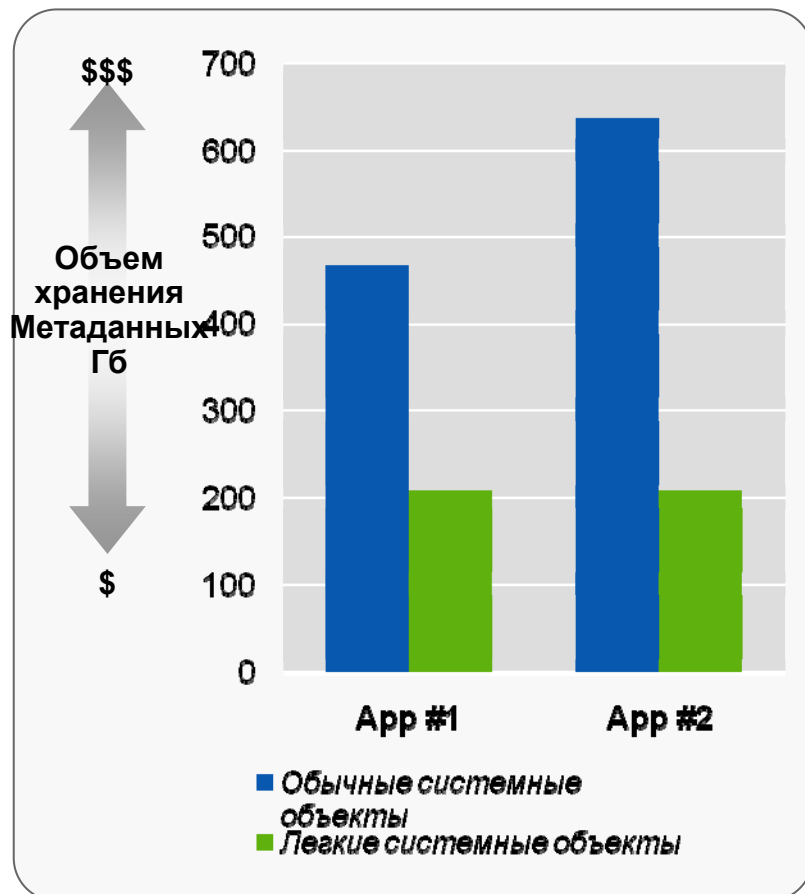
= **Экономия пространства!**

+ Экономия на резервировании
+ Экономия на готовности

= **Еще больше экономии!**

Приложения должны быть специально спроектированы для работы с «легкими» системными объектами

Экономия времени и средств с легкими системными объектами



Приложения, такие как архивирование эл. почты, обработка образов чеков, проверка отгрузки спроектированы, чтобы урезать метаданные до фрагментов от их первоначального размера

Обычно экономия составляет в уменьшении базы данных до 25%-50% первоначального объема!

Пакетная обработка и кэш для откликов на системные коммуникации



Пакетная обработка и кэширование откликов снижает излишнюю нагрузку при загрузке похожих объектов

- Пакетная обработка

- Обеспечивает пакетную обработку множества объектов
- Снижает количество круговых операций между High Volume Server и СУБД

- Кэш для откликов на системные коммуникации (скопинг)

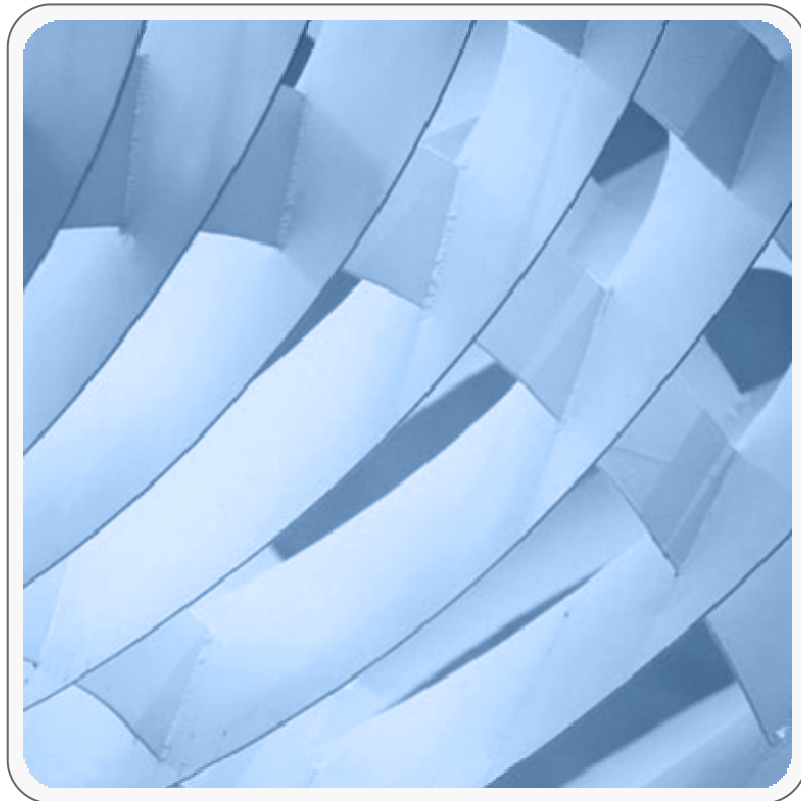
- Снижает количество рутинных проверок таких вопросов, как существование папки или списки контроля доступа, вовремя загрузки
- Обычно используется в комбинации с круговой обработкой

Загрузка свыше 7½ млн. объектов в день!



Благодаря использованию легких системных объектов + пакетированию + скопингу загрузку содержания можно осуществить в пять раз быстрее!

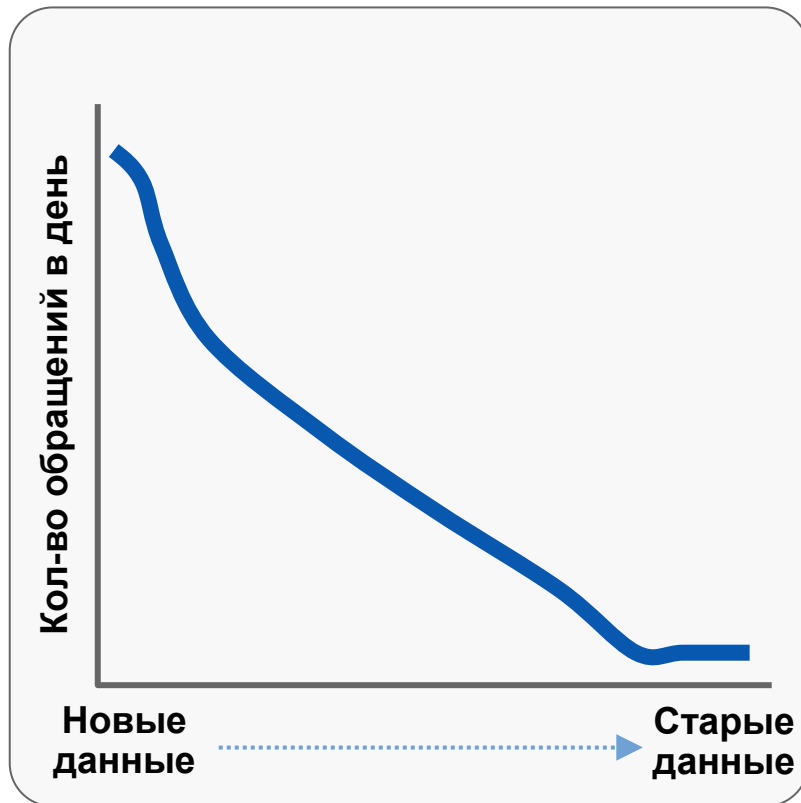
Важное замечание: повышение производительности может варьироваться – терабайт малых объектов покажет большую степень повышения, нежели терабайт больших объектов



Зачем нужно дробление?

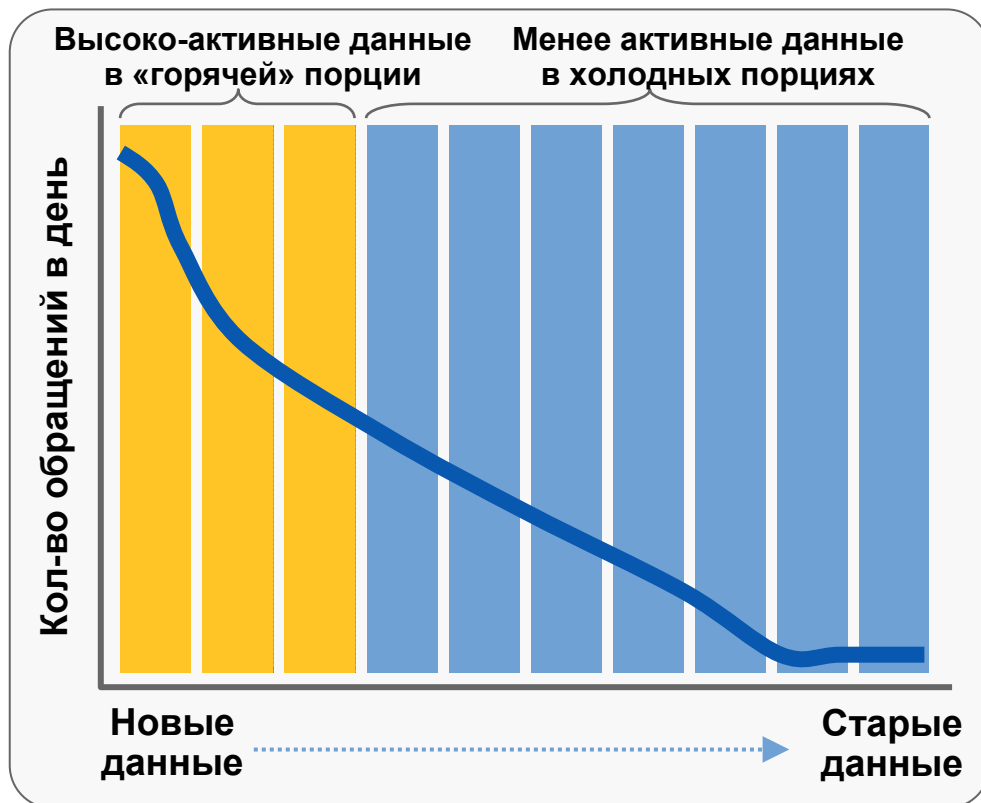
Дробление позволяет вам группировать большие объемы данных в партии управлять каждой партией как единым целым. Это обеспечивает лучшую производительность и управляемость, а также позволяет масштабировать систему целиком

- Поддержка сверхбольших баз данных (VLDB)
- Логическая группировка объектов
- Эффективные запросы
- Пакетные операции (выпуск партии – применение ограничений)
- Загрузка офф-лайн
- Дифференцирование между активным содержанием и менее активным или замороженным содержанием для сохранения времени и уменьшения затрат на такие операции, как резервное копирование



Чем дальше хранится содержание, тем меньше его бизнес-ценность и тем реже к нему обращаются. Частота обращения к связанным с ним метаданным тоже падает.

Дробление данных



Ранжированное дробление

позволяет отделить старые данные от новых

и обращаться со старыми данными иначе, чем с новыми

Миграция - 2 млрд. объектов в неделю!

Данные внутренних тестов EMC D6.5

	Объектов в час
Порционная загрузка в режиме офф-лайн, единым потоком объектов	20,000,000

Порционная загрузка офф-лайн в сочетании с легкими системными объектами позволяет очень быстро загрузить объекты в СУБД.

- 1 **Создание партий офф-лайн**
- 2 **Загрузка метаданных**
несколько миллионов файлов за считанные часы или миллиарды файлов за несколько дней
- 3 **Перевод партий в он-лайн**
в соответствии с расписанием, за несколько секунд
- 4 **Оставить содержание на месте**
или переместить файлы, используя Content Storage Services, обеспечивающих миграцию на основе политик

Применение пакетной загрузки офф-лайн



Быстрая и простая замена устаревших систем или консолидация разнородных репозиториев

Примеры:

- Приложения для управления транзакционным содержанием
- Миграция данных с устаревших мейнфреймов
- Консолидация систем управления содержанием
- Замена других систем управления содержанием (FileNet и т.п.)



Сокращение времени ожидания для пользователей

- Уменьшилось время передачи файла – для High-Volume Server и Content Server
- Сокращение времени ожидания зависит от выполняемых операций и задержек сети, например, в случае простого просмотра документа в сети без задержек сокращение времени составляет 55% по сравнению с Documentum 5.3

Показатели повышения производительности

ЗАДЕРЖКА СЕТИ

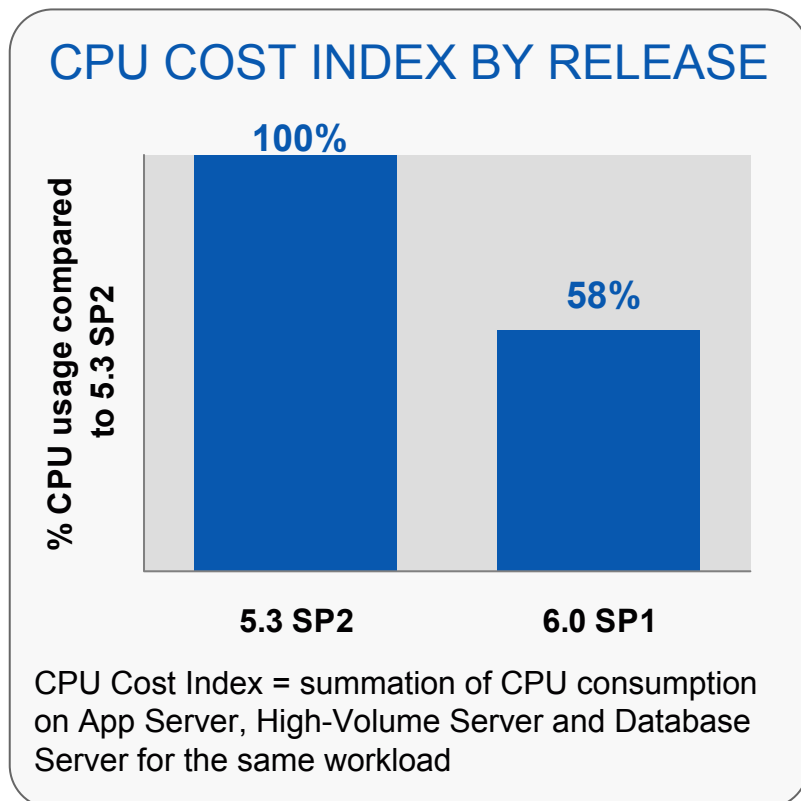
	0мс	100мс	400мс
Импорт	38%	27%–34%	27%–30%
Экспорт	63%	38%–47%	39%–43%
Checkout	22%	37%–38%	43%–44%
Checkin	27%	23%–30%	25%–29%
Просмотр	55%	38%–44%	46%–47%

Источник: EMC Documentum Performance Engineering

Уменьшение времени передачи содержания в Documentum ECM 6.5 по сравнению с Documentum ECM 5.3

Содержание по сети передается на 55% быстрее!

Улучшения использования ЦПУ



Снижение коэффициента загрузки ЦПУ означает лучшее использование ресурсов и меньшие затраты на инфраструктуру

Снижение коэффициента загрузки ЦПУ на 42%!

High-Volume Server



- Управляет возрастающими объемами данных
- Контролирует затраты на администрирование базы данных
- Увеличивает загрузку и масштабируемость
- Уменьшает риски в работе с огромными объемами содержания или данных

V Международный форум по технологиям хранения и управления информацией

