

EMC ENTERPRISE CONTENT DIVISION

# Архивирование информации в XXI веке. Почему вам нужен электронный архив, даже если вы считаете, что он у вас есть.

Вячеслав Кадников,  
Региональный директор по  
Восточной Европе, России и СНГ



EMC<sup>3</sup>

# Почему мы вновь говорим об архивировании?

## Предпосылки



Во многих организациях наблюдается "информационный взрыв" в сочетании с необходимостью соблюдения требований регулирующих органов к работе с информацией. Организации содержат все больше и больше приложений, работоспособность которых доведена до предела, как результат - повышенные затраты на инфраструктуру (включая сервера, СХД, техническую поддержку, ПО, лицензии и операционные расходы).

## Цели



- Обеспечить единое решение, которое охватывает ключевые тенденции;
  - Перенос статического контента на более дешевый носитель;
  - Готовность к использованию облачных технологий;
  - Хранение структурированных данных в простом и доступном формате;
  - Готовность к экспоненциальному росту данных, предназначенных для архивирования;
  - Трансформация данных.
- Определить бизнес-заказчика и связать примеры использования с требованиями компании

## Заказчики



- Генеральный директор
- Директор по информационным технологиям
- Технический директор
- Директор по информационной безопасности
- Руководитель отдела корпоративного регулирования и контроля



85%

данных эксплуатируемых систем находится в статическом состоянии<sup>1</sup>



65%

Новых архивных решений будет базироваться на облачных технологиях<sup>2</sup>



20%

ЕСМ бюджета будет тратиться на архивирование структурированных данных<sup>3</sup>



100%

Увеличение объема данных каждые 18 месяцев<sup>4</sup>

<sup>1</sup> Source: "Database Archiving Remains An Important Part Of Enterprise DBMS Strategy" Forrester Research, Noel Yuhanna, 13 August 2007,

<sup>2</sup> <http://www.gartner.com/document/2951217?toggle=1;>

<sup>3</sup> [http://www.gartner.com/Reprints?param=reprint&resId=2717018&usrid=0,](http://www.gartner.com/Reprints?param=reprint&resId=2717018&usrid=0;)

<sup>4</sup> <http://www.ibm.com/software/data/optim/manage-data-growth>

# Архив

## Что такое архив?

Архив надежно и эффективно сохраняет информацию с соблюдением требований организации и регуляторов для дальнейшего использования

5 основных преимуществ архивирования:



**1. Целостность данных**

Сохранение данных для использования в будущем



**2. Эффективность бизнеса**

Миграция данных в архив снижает нагрузку с исходного приложения и увеличивает его производительность



**3. Конфиденциальность данных**

Обеспечение контроля доступа конечных пользователей



**4. Доступность данных**

Обеспечение постоянного доступа конечных пользователей к архивным данным



**5. Соблюдение регламентов**

Применение политик хранения и контроля данных

### Общие заблуждения

‘Мы используем резервное копирование, и нам не нужен архив’

‘Архив - это свалка информации’

‘Архивирование не связано с текущим бизнесом’

‘Архивирование - это риск’

‘Архивирование - это дорого’

‘Нам не нужен архив, у нас уже есть ECM-система’

‘На нас не распространяются требования регуляторов, архив нам не нужен’

# Заблуждение об архиве №1

## Заблуждение первое



**Мы используем резервное копирование данных и нам не нужен архив**  
Резервное копирование и архив часто путают и рассматривают как одно и то же решение. Однако это разные решения, предназначенные для разных целей;

**24**

На данный момент, это среднее количество резервных копий для одного приложения

### Архив

Архив имеет три основные цели:

- **Сохранение:** обеспечение целостности данных для использования в будущем
- **Соответствие требованиям:** данные хранятся в соответствии с требованиями регуляторов
- **Обнаружение:** поиск информации в большом объеме данных

### Резервное копирование

Резервное копирование имеет две основные цели:

- Используется для восстановления информации
- Повышает доступность данных, делая возможным восстановление информации на определенный момент времени

**10%**

Типичный размер резервной копии, по сравнению с первоначальными данными

**VS**

### Преимущества



**Сохранение данных**  
Сохранение данных для использования в будущем



**Доступность данных**  
Структурирование и индексирование



**Регламентирование**  
Применение политик к архивным данным



**Конфиденциальность**  
Доступ к данным и шифрование



**Эффективность данных**  
Миграция данных в архив снижает нагрузку на исходное приложение и потенциально увеличивает его производительность

### Преимущества



**Полная репликация**  
Полная репликация данных для обеспечения резервного копирования и восстановления



**Множественность копий**  
Одни и те же данные присутствуют в нескольких резервных копиях для обеспечения восстановления



**Фиксированный момент времени**  
Данные резервируются на определенный момент времени и не обновляются до следующего резервного копирования

**7 лет**

Среднее время хранения резервной копии

**80%**

Потенциальное сокращение затрат на резервное копирование<sup>1</sup>

<sup>1</sup> <http://www.oracle.com/us/products/servers-storage/storage/tape-storage/backupnotarchivefinal-120513gc-2083314.pdf>

# Заблуждение об архиве №2

## Заблуждение второе

### Архив - это свалка информации

Архив часто ассоциируется с офлайн-хранением, а не с активным источником данных для поиска информации;

### Архивирование не связано с текущим бизнесом

Архив содержит статичные, неизменяемые данные. Соответственно, архив рассматривается как хроника прошедших событий.

## Мнение аналитиков

75%

- Gartner

- Все большее число организаций стремится увеличить преимущества от использования существующей информации<sup>1</sup>
- К 2017 году Gartner прогнозирует, что 75% организаций будут обращаться к архиву как к **первоначальному** источнику информации, по сравнению с сегодняшними 10%<sup>1</sup>

84%

- Deloitte IDO 2014

- 84% клиентов, которые приобрели Аналитику Deloitte IDO, усилили свою конкурентную позицию<sup>2</sup>.
- **IDO использует современный анализ разнообразных срезов данных (озер данных), источником которых может служить архив<sup>2</sup>.**



EMC<sup>2</sup>

<sup>1</sup> <http://www.gartner.com/document/2951217?toggle=1>;

<sup>2</sup> Deloitte IDO <http://www2.deloitte.com/>

# Заблуждение об архиве №3

## Заблуждение третье

### Архивирование - это риск



Многие считают, что при архивировании риск возрастает, так как в архиве проще обнаружить информацию. Однако в случае невозможности предоставить запрашиваемую информацию потери могут быть еще более значительными.

#### Обнаружение

Проведение обследования текущих источников данных, для выявления статичной информации

#### Классификация

Классификация обнаруженных данных:

1. Применение политик
2. Перемещение в архив
3. Удаление на основании политик

#### Снижение риска

Ранее скрытые данные становятся доступны, и могут быть классифицированы в соответствии с требованиями

#### Примеры

**\$15 млн**

Morgan Stanley & Co. оштрафован в 2006 г за непредоставление информации **за счет перезаписи лент с резервными копиями**<sup>1</sup>

**\$10 млн**

The Bank of America Securities LLC (BAS) был оштрафован за недостаточную сохранность электронной почты и документации<sup>1</sup>

#### Метрики

**100,000+**

Правил, касающихся сохранности информации, в мире<sup>2</sup>

**3.5x**

Прогнозируемый рост размера корпоративного архива к 2017 году<sup>2</sup>

Архив обеспечивает управление **рисками с информативной точки зрения**

- Полный и точный набор данных
- Безопасность данных и электронное уничтожение

<sup>1</sup> <http://www.nationalarchives.gov.uk/services/publications/information-risk.pdf>,

<sup>2</sup> [http://www.gartner.com/resources/267500/267518/selecting\\_the\\_best\\_archival\\_\\_267518.pdf](http://www.gartner.com/resources/267500/267518/selecting_the_best_archival__267518.pdf)

# Заблуждение об архиве №4

## Заблуждение четвертое



### Архивирование - это дорого

Архивирование помогает снизить затраты за счет хранения на более дешевых носителях, минимальной стоимости ресурсов поддержки и потенциально увеличивая производительность активно эксплуатируемых систем. Кроме того, общая стоимость владения архива существенно меньше по сравнению с затратами на репутационные и судебные издержки в случае **несоблюдения норм по хранению информации**.

**\$1млн**

Компания Boehringer Ingelheim была оштрафована за несоблюдение норм хранения электронных документов<sup>4</sup>

**Репутационный  
ущерб**

**\$1.2млн**

Филиалы ING были оштрафованы за несоблюдение норм хранения электронной почты<sup>2</sup>

**Имиджевые  
потери**

**\$7.5млн**

Компания LPL Financial оштрафована за несоблюдение норм хранения электронных писем содержащих 7 летнюю переписку<sup>3</sup>

**Доверие  
клиента**

**\$3.75млн**

Банк Barclays был оштрафован за несоблюдение регламентов хранения транзакций<sup>1</sup>

**Доверие  
инвесторов**

**2 рубля за Гб**

Ориентировочная стоимость архивного хранения<sup>5</sup>

**Продуктивность**

Увеличение продуктивности поиска пользователем

**0.15 FTE**

Необходимое количество ресурсов на сопровождение

<sup>1</sup> <https://www.finra.org/newsroom/2013/finra-fines-barclays-375-million-systemic-record-and-email-retention-failures>, <sup>2</sup> <http://www.finra.org/newsroom/2013/finra-fines-five-ing-firms-12-million-email-retention-and-review-violations>, <sup>3</sup> <https://www.finra.org/newsroom/2013/lpl-pay-9-million-systemic-email-failures-and-making-misstatements-finra>, <sup>4</sup> <http://www.law360.com/articles/494446/boehringer-fined-1m-for-withholding-docs-in-mdl>  
<sup>5</sup> [http://www.gartner.com/resources/267500/267518/selecting\\_the\\_best\\_archival\\_\\_267518.pdf](http://www.gartner.com/resources/267500/267518/selecting_the_best_archival__267518.pdf)

# Заблуждение об архиве №5

## Заблуждение пятое



Нам не нужен архив, у нас уже есть ECM  
Электронный архив – часть EMC, но есть нюансы.

### Основы ECM

#### Сбор информации

Прием данных/контента в систему



#### Управление

Версии, ограничение доступа/доступ по запросу



#### Хранение

Определение местоположения информации в ECM



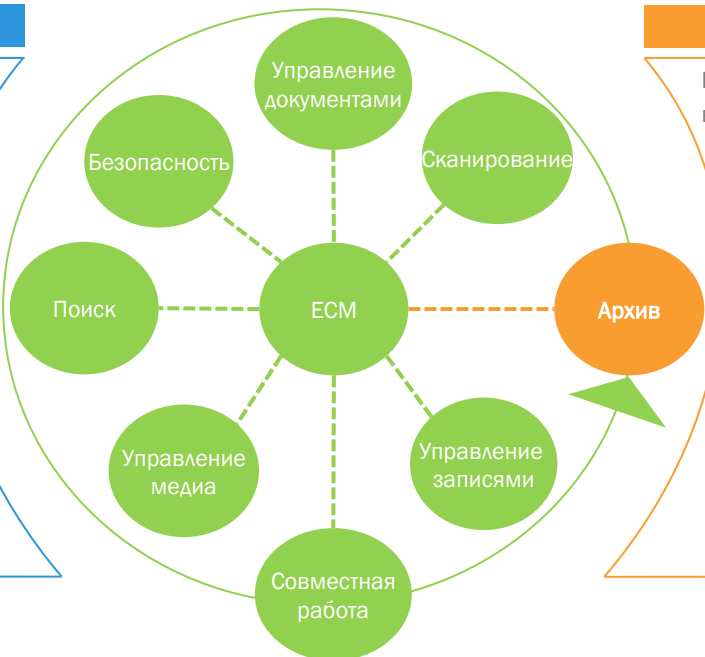
#### Защита

Защита целостности информации для будущего использования



#### Доставка

Доступность информации по требованию



### Основные отличия

Как часть ECM, архив предоставляет все основные возможности ECM с некоторыми отличиями;

#### Масштабирование

Архив оптимизирован на хранение миллиардов записей и документов;

- Архив поглощает больше данных;
- Поиск настраивается на разнообразные наборы данных и время отклика

#### “Агностицизм”

Архивирование структурированных и неструктурированных данных. Другие решения ориентированы на неструктурированные данные.

#### Экономическая эффективность

Перемещение данных с дорогих носителей на более дешевые, а также уменьшение размера и стоимости оперативной резервной копии





Решение EMC InfoArchive

# Что такое EMC InfoArchive?

## Определение

### EMC InfoArchive это:

- Единый унифицированный архив;
- Расширяемое корпоративное решение по архивированию информации, поддерживающее открытые технологии.

## Описание

InfoArchive - это решение, обеспечивающее хранение структурированных данных и неструктурированного контента в едином консолидированном архиве.



В результате организации могут обеспечивать целостность информации, не затрудняя при этом доступ к ней.



Открытые стандарты, используемые в EMC InfoArchive, позволяют гарантировать доступ к информации спустя десятилетия, вне зависимости от изменения программной и аппаратной инфраструктуры.



# Почему EMC InfoArchive?

## Технологический нейтралитет

- Доступность данных в непроприетарном формате
- Хранение структурированных и неструктурированных данных

## Ядро

- Все необходимые функции для пользователей
- Стоимость привязана к объему архива

## Управление

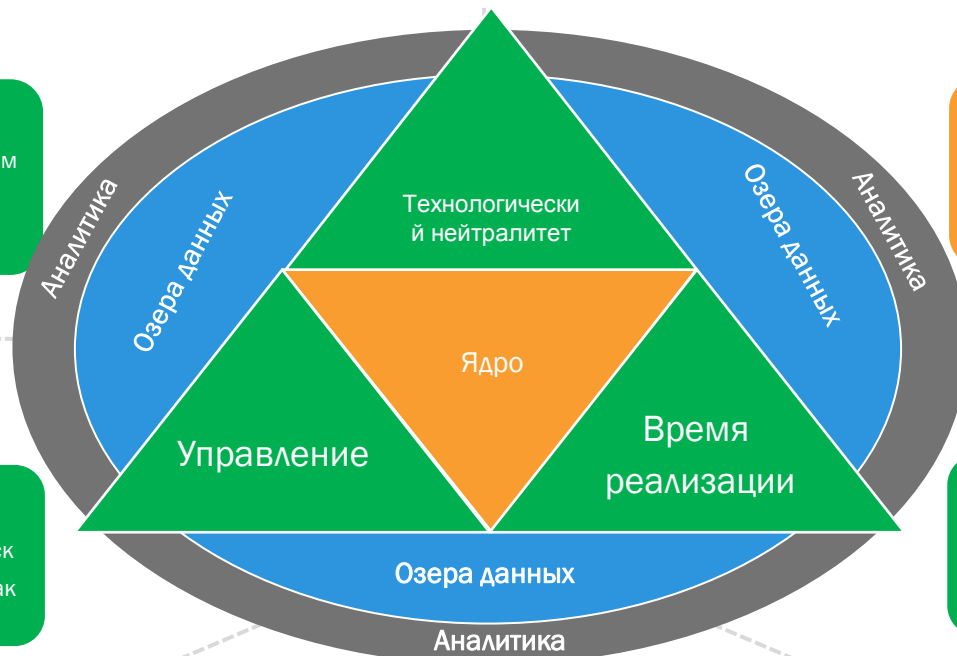
- Быстрый и точный нормативный поиск
- Классификация нескольких записей как единого целого

## Время реализации

- Получение результата в течении недель
- Эффективный возврат инвестиций в IT

## Озера данных

- Целостное хранилище с единообразным набором данных
- Предоставление данных как сервиса



# Возможности InfoArchive

## Критичные кратковременные преимущества

- Быстрое сохранение статических данных для уменьшения стоимости хранения, резервного копирования и расходов на обслуживание
- Применение политик контроля доступа для обеспечения конфиденциальности
- Быстрая реакция на запросы регуляторов

## Критичные долговременные преимущества

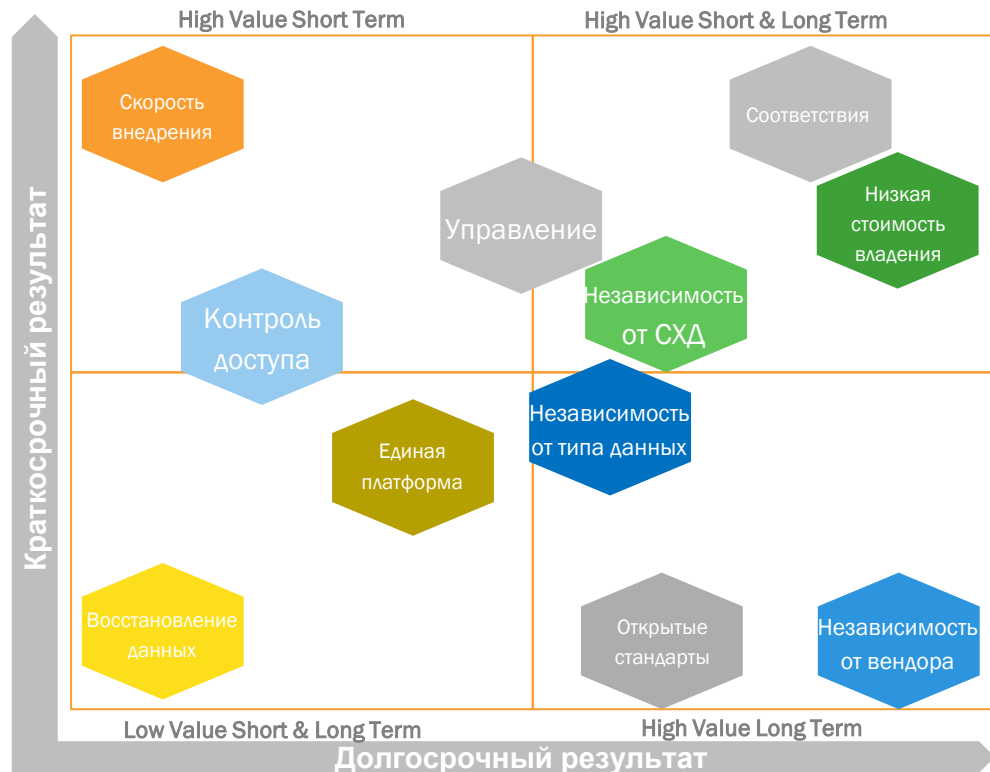
- Массовое применение политик при архивации данных
- Снижение совокупной стоимости владения (ТСО) для существующих систем
- Переход на более дешевые носители для снижения затрат на хранение

## Средне критичные долговременные преимущества

- Переход от проприетарных форматов в сторону открытых данных
- Защита данных и применение международных стандартов регуляторов
- Архивирование любых данных, независимо от формата для упрощения соответствия требованиям регуляторов

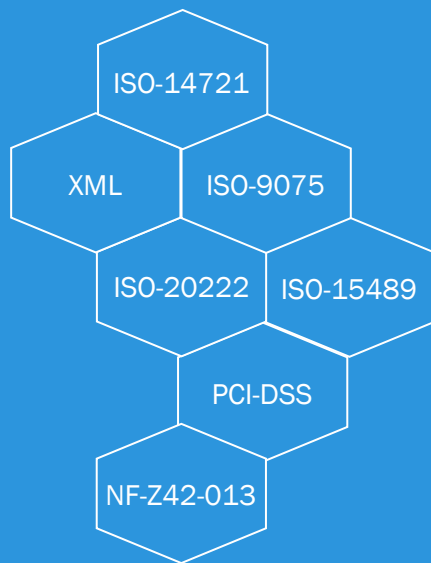
## Средне критичные кратковременные преимущества

- Восстановление данных обратно в приложение в случае необходимости
- Применение единых политик прав доступа в архиве



EMC<sup>2</sup>

# Открытые стандарты



Открытые стандарты дают совместимость между продуктами различных поставщиков

- Эталонная модель открытой информационной системы для архивирования (OAIS), разработанная изначально для нужд освоения космоса
- Стандарт кодировки и обработки документов
- Стандарт для финансовых транзакций
- Стандарт для управления записями
- Стандарт безопасности данных платежных карт
- Стандарт по обеспечению целостности и сохранности электронных документов
- Использование для поддержки web-сервисов
- Использование для сжатия данных

Преимущества

Гибкая и расширяемая схема данных

Данные архивируются в соответствии с требованиями регуляторов

Архивирование и защита финансовых транзакций

Гарантия целостности архивных документов

Поиск по структурированным и неструктурированным данным

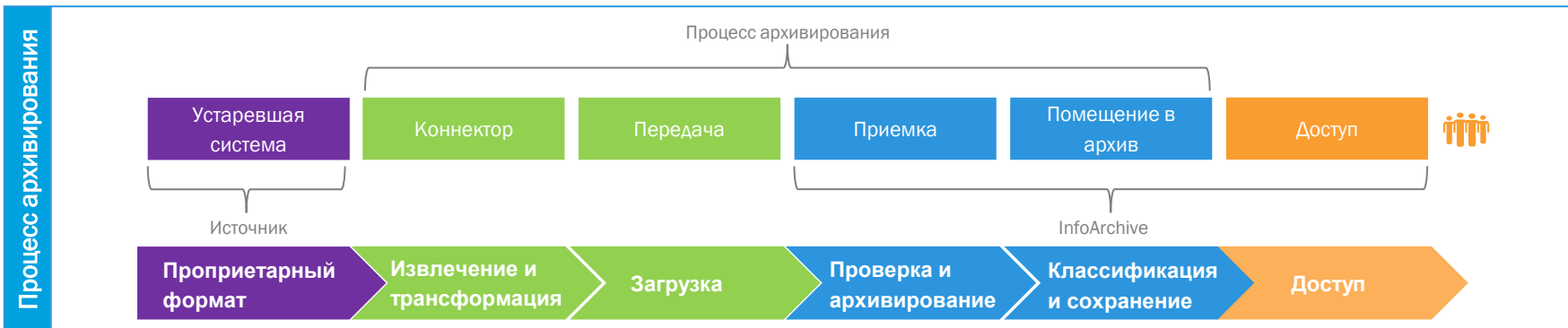
Технологический нейтралитет

EMC<sup>2</sup>

# Процесс архивирования

## Обзор

Процесс архивирования прост и легко повторяем для различных приложений. Тем не менее, архивирование в ряде случаев требует настройки или доработки интеграционного слоя.



## Вывод из эксплуатации устаревших приложений

Mobius

Mainframe

Oracle

SharePoint

Бумажные документы

FranFinance

Streamserve

Microfilm

DB2

Asterion

SQL Server

Lotus Notes

Documentum

PeopleSoft

EMC<sup>2</sup>

# Процессы управления

## Обзор

Процессы управления информацией должны быть установлены в соответствии с процессом архивирования в EMC InfoArchive для соблюдения требований регуляторов и снижения рисков

## Процесс управления

### Обнаружение

- Определить системы, которые нуждаются в извлечении статических данных

### Категоризация

- Категоризация данных;
  1. Архив
  2. Сохранение
  3. Удаление

### Применение политик

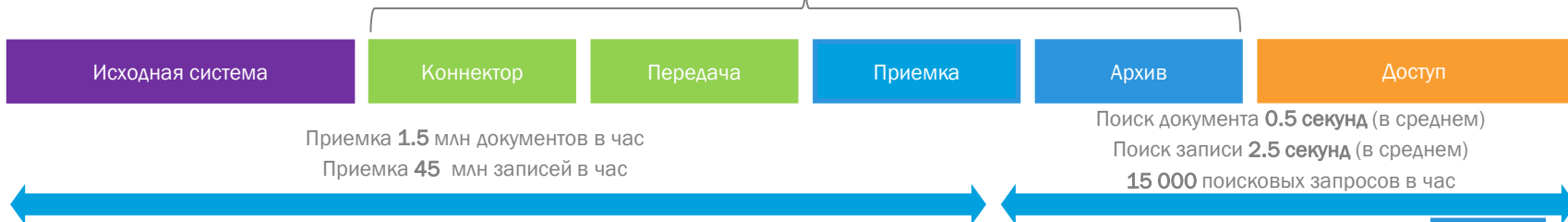
- Применение различных классификаторов в зависимости от конкретных требований

### Управление

- Непрерывный пересмотр и обновление политик хранения в соответствии с изменяемым законодательством

## Метрики управление и архивирования

### Процесс архивирования



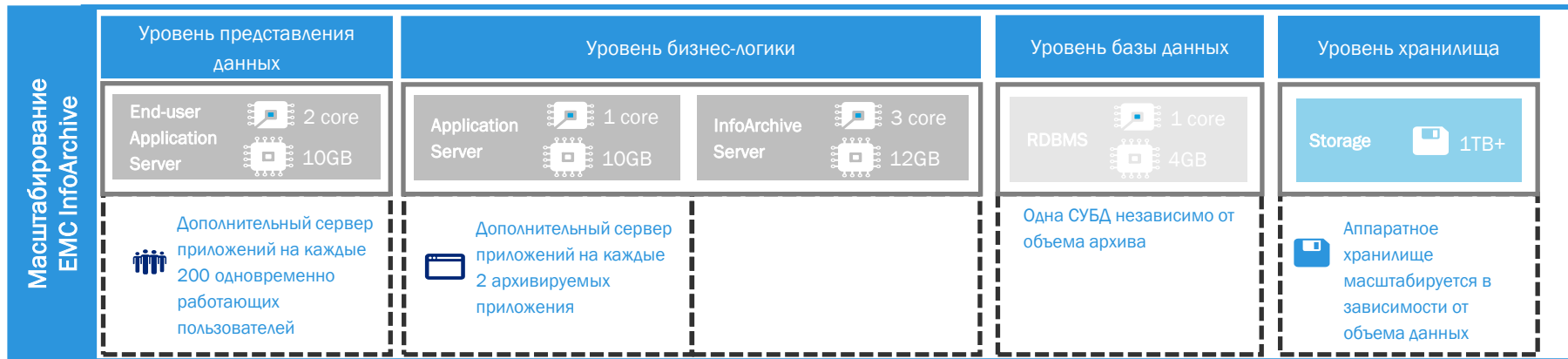
EMC<sup>2</sup>

# Масштабируемость

## Обзор

### Пример: 2 миллиарда документов и 40 миллиардов записей

- Масштабирование до миллиардов записей и документов
- Независимость от аппаратной технологии хранения
- Кастомизируемый интерфейс доступа к данным
- Минимальное использование СУБД





# Примеры использования

Рационализация приложений

1

**NOKIA**



Консолидация ЦОД

2



Расширение ERP / SAP

3



Слияния и поглощения

4

**NOKIA**



# Электронный архив XXI века

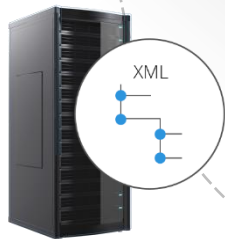
## Архивирование таблиц

Идеальное решение для декомиссии приложений со структурированными файлами



## Архивирование транзакций

Представляет данные в качестве единой транзакции - идеально для архивирования в соответствии с требованиями регуляторов



ЕДИНОЕ РЕШЕНИЕ

## Архивирование файлов

Обеспечивает архивирование неструктурированного содержания, файлов и потоков печати



## Архивирование гибридных записей

Увязывание содержания и структурированных данных в единой записи



EMC<sup>2</sup>®