



IBM East Europe / Asia Ltd.

Электронные цифровые архивы :
долговременное сохранение данных и
обеспечение их аутентичности.

Модель сохранения данных OAIS (ISO 14721:2003).

Валерий Елизов
IBM Systems & Technology Group

Электронные цифровые архивы : долговременное сохранение данных и обеспечение их аутентичности.

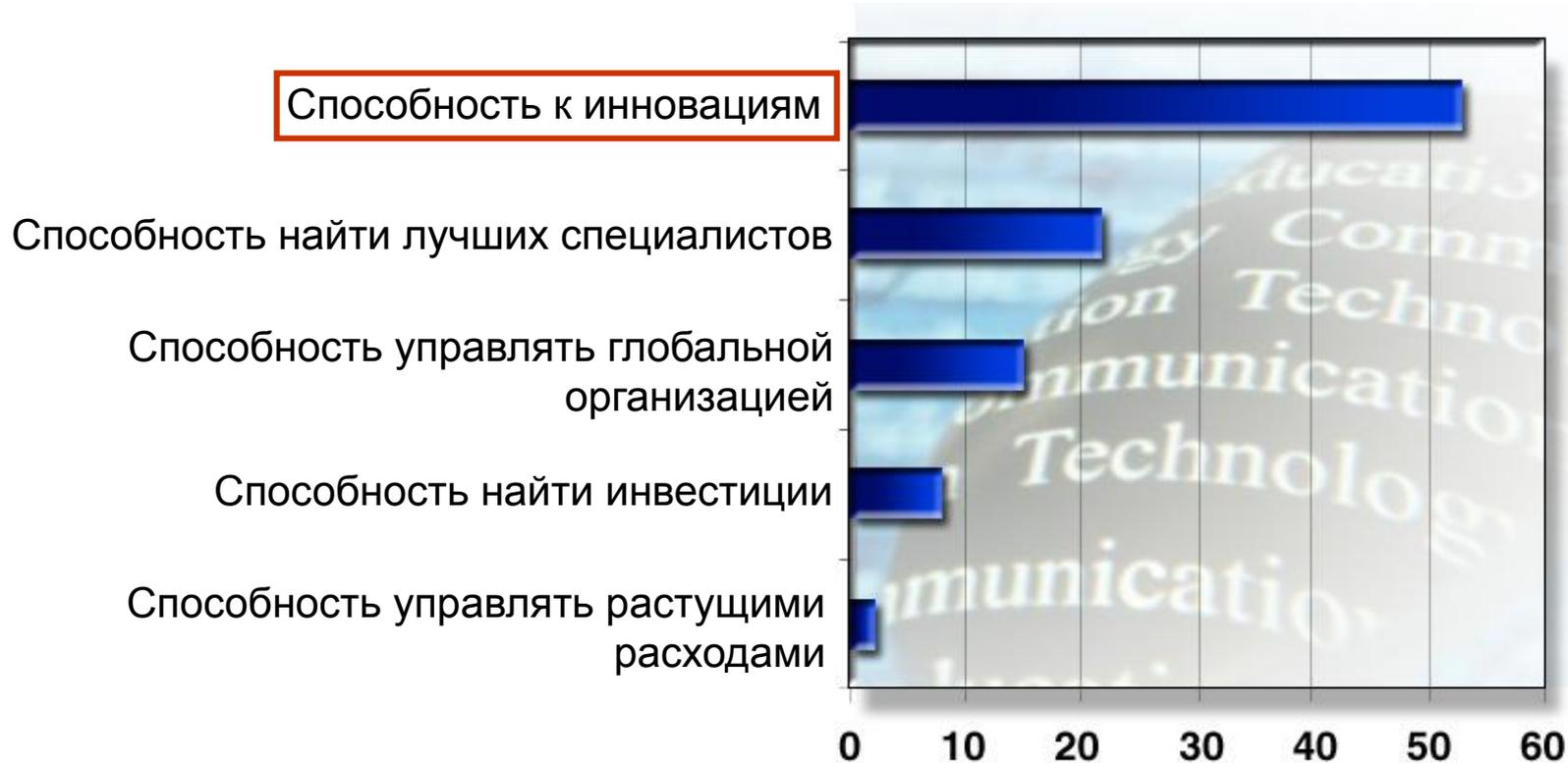
- Инновации IBM;
- Стандарты в области долговременного сохранения данных:
 - международные
 - отечественные;
- Проекта CASPAR и участие IBM;
- Модель цифрового архива DIAS (Digital Information Archival System);
- Обеспечение фактической и юридической значимости объекта хранения;
- Функциональная архитектура;
- Примеры реализации;
- Программно-аппаратное обеспечение.



Преамбула



Руководители уверены, что ключевым фактором успеха и конкурентного преимущества являются инновации

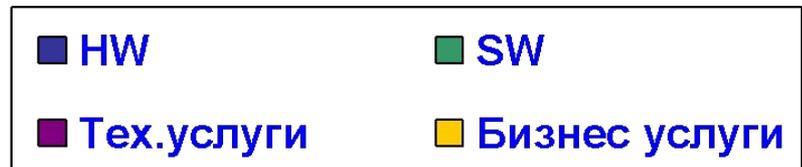
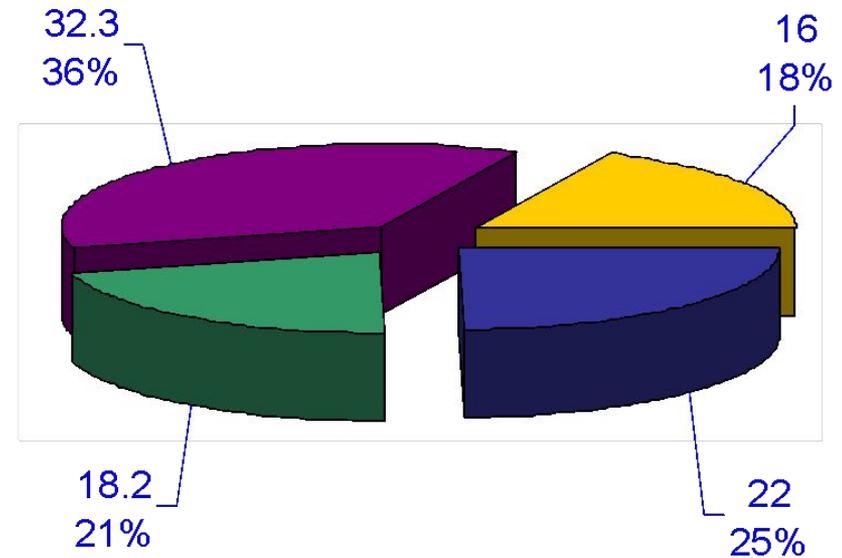


Source: McKinsey Quarterly survey of 9,345 global executives

Инновации как основа успеха IBM

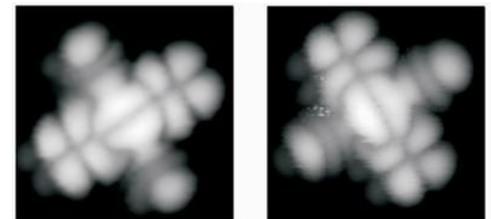
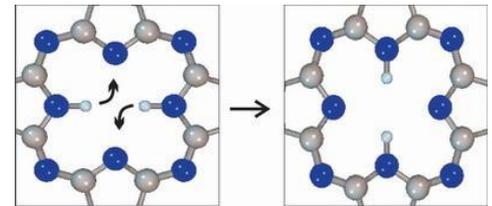
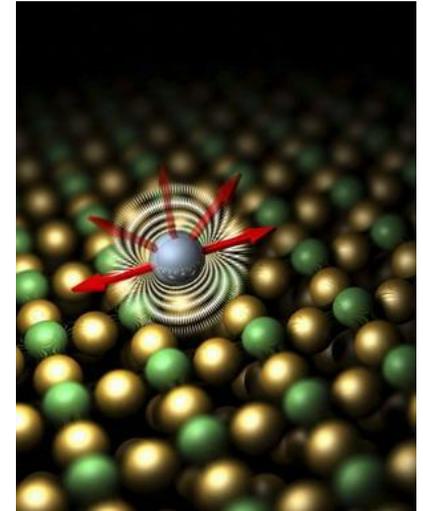
2006 год:

- ДОХОДЫ -\$91.4млрд
- Лидер рынка ПО промежуточного слоя;
- №1 На рынке ИТ-услуг и консалтинга;
- №1 В списке TOP100 суперкомпьютеров;
- Лидер рынка серверов - 32.8%;
- 100,000 Бизнес-партнеров;
- 3621 патент;
- \$6.1млрд Инвестиции в исследования и разработку.



Нанотехнологии и цифровые архивы

- **Открытие магнитной анизотропии атома:**
 - Будущее технологий хранения данных: атом железа на специальной медной подложке сохраняет магнитный момент и ориентацию оси в зависимости от заряда. Перемена магнитного момента атома может интерпретироваться как «0» или «1».
- **Молекулярный вентиль (элементарный переключатель):**
 - В качестве элементарного переключателя (элемента логического вентиля) может использоваться отдельная молекула (нафталинцианин), которая может переходить из состояния «включено» в состояние «выключено» без изменения формы.





Ситуация в мире

2

CASPAR – сохранение культурного, художественного и научного наследия.

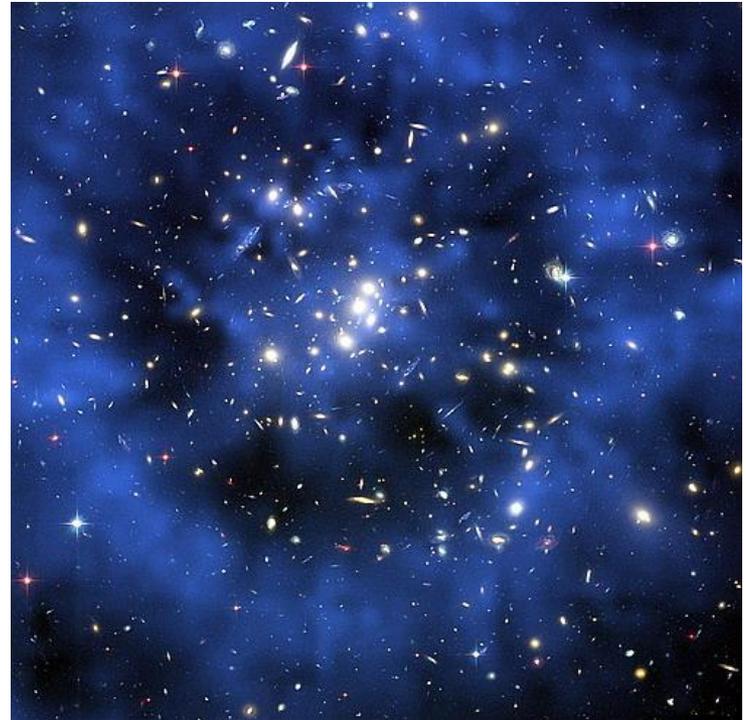
- **CASPAR** (Cultural, Artistic, and Scientific knowledge for Preservation, Access and Retrieval)
 - международный проект ЕС и UNESCO, основной целью которого является разработка стандартов и механизмов долговременного (неограниченного) срока сохранения культурного, научного и художественного наследия мира.
 - Preservation DataStore – инновационная система хранения данных, в основе которой - модель OAIS (ISO:14721:2002). Разрабатывается IBM.
 - <http://www.casparpreserves.eu/>



Расширяющаяся цифровая вселенная

- *В 2006 г. объем оцифрованной информации – созданной и сохраненной для дальнейшего воспроизведения, составил 161 экзабайт (161 миллиард гигабайт) – объем, примерно в 3 миллиона раз превышающий объем всех написанных за историю человечества книг!*
- *До конца 2010 г. этот объем увеличится в 6 раз и составит 998 экзабайт.*

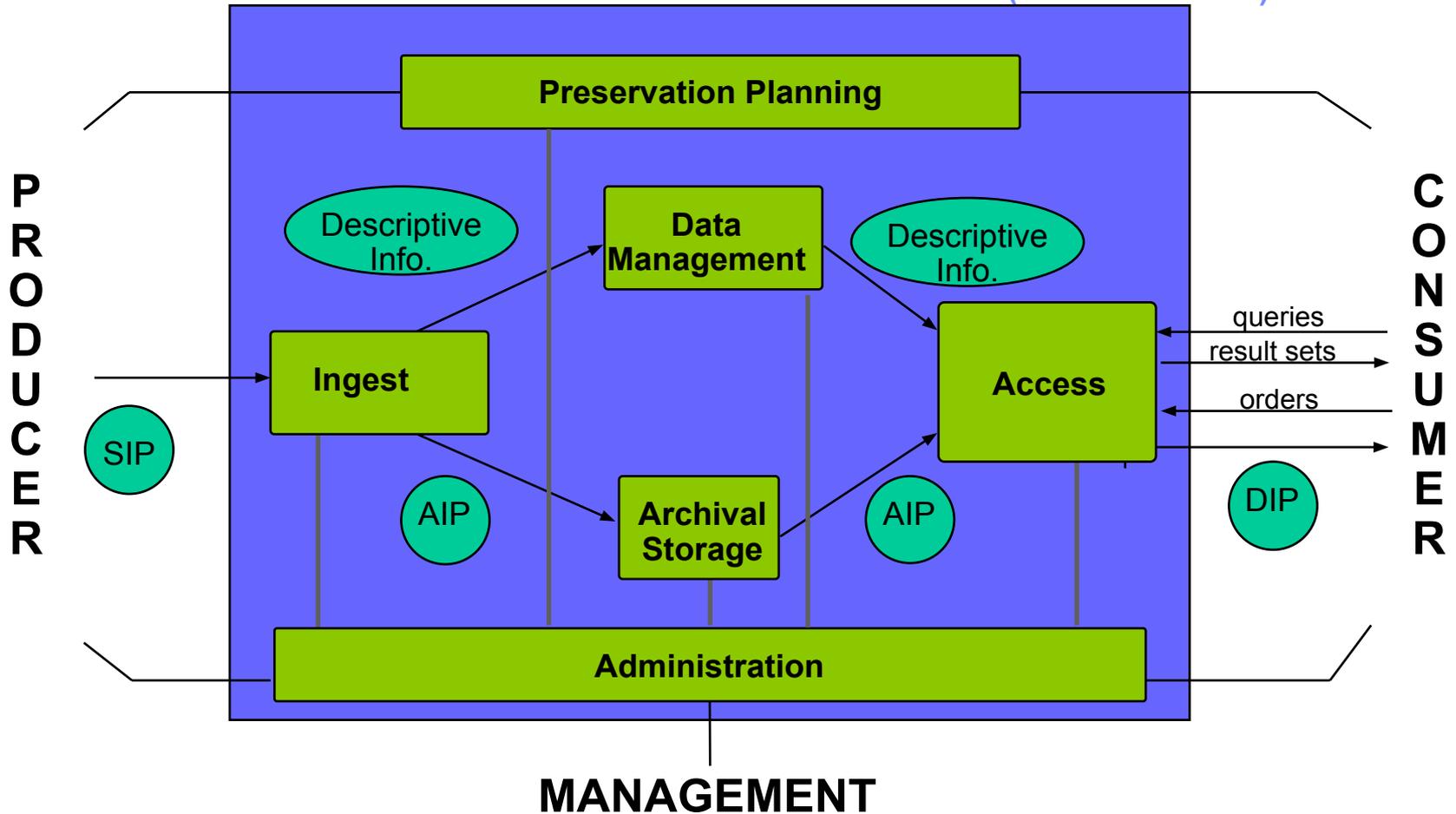
IDC (2007) The Expanding Digital Universe



Основные требования к цифровым архивам

- **Централизация;**
- **Структурность;**
- **Гарантированная сохранность данных;**
- **Гибкость ;**
- **Универсальность;**
- **Масштабируемость;**
- **Соответствие (международным) стандартам** сохранения данных и создания электронных цифровых фондов;
- **Эффективная модель управления;**
- **Производительность;**
- **«Читабельность»** объектов сохранения **в течение всего срока** сохранения.

Open Archival Information System Reference Model - OAIS (ISO:14721:2002)

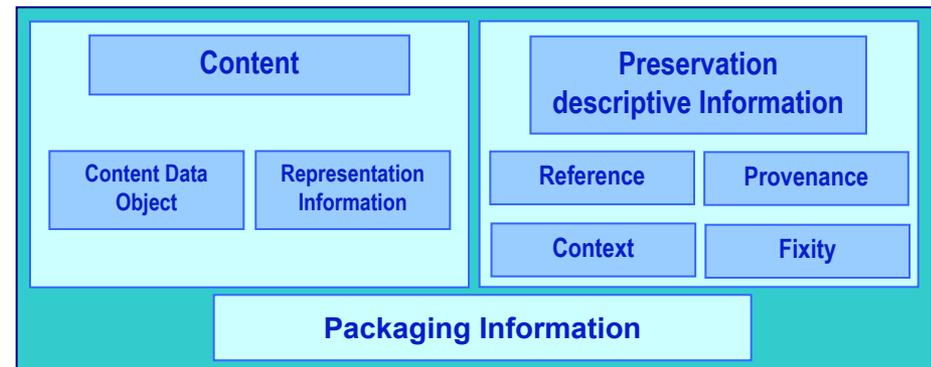


SIP = Submission Information Package
 AIP = Archival Information Package
 DIP = Dissemination Information Package

Открытая система архивной информации - OAIS

- Простое хранение цифровой копии – нонсенс.
- Как обеспечить воспроизводимость и понимаемость в будущем?
- Описательные метаданные.
- Определение фокусной аудитории и принятых в этой аудитории форматов.
- ...

Представление объектов архивного хранения (Archival Information Package) в соответствии со стандартом OAIS (ISO:14721:2002)



Представление информации в строго заданных параметрах воспроизведения, подходящих для передачи и обмена, интерпретации или дальнейшей обработки

Модели сохранения

▪ Музей

- Исходные данные и средства просмотра сохраняются в оригинальном виде и поддерживаются в рабочем состоянии;
 - Запрет на интерпретацию данных, высокие затраты на обслуживание программно-аппаратных средств;

▪ Эмулятор

- Исходные данные сохраняются в оригинальном виде;
- Эмуляция средств просмотра (комплекса ПО и АО) – IBM UVC (Universal Virtual Computer);
 - Необходимость поддержки виртуальной платформы UVC;

▪ Миграция

- Миграция основных компонент контентного содержания;
- Сохранение основных характеристик идентичности и целостности объектов хранения;
 - Возможно появление «цифрового шума»;

▪ Описание

- Создание описательного содержания, позволяющего воспроизведение оригинала;
- Нет необходимости поддержки исходных данных и оригинальных средств просмотра;
 - Потеря исходных данных (оригинального контента).



IBM Research Lab in Haifa, Israel

<http://www.haifa.il.ibm.com/>

Модель IBM Preservation DataStore поддерживает все модели сохранения.

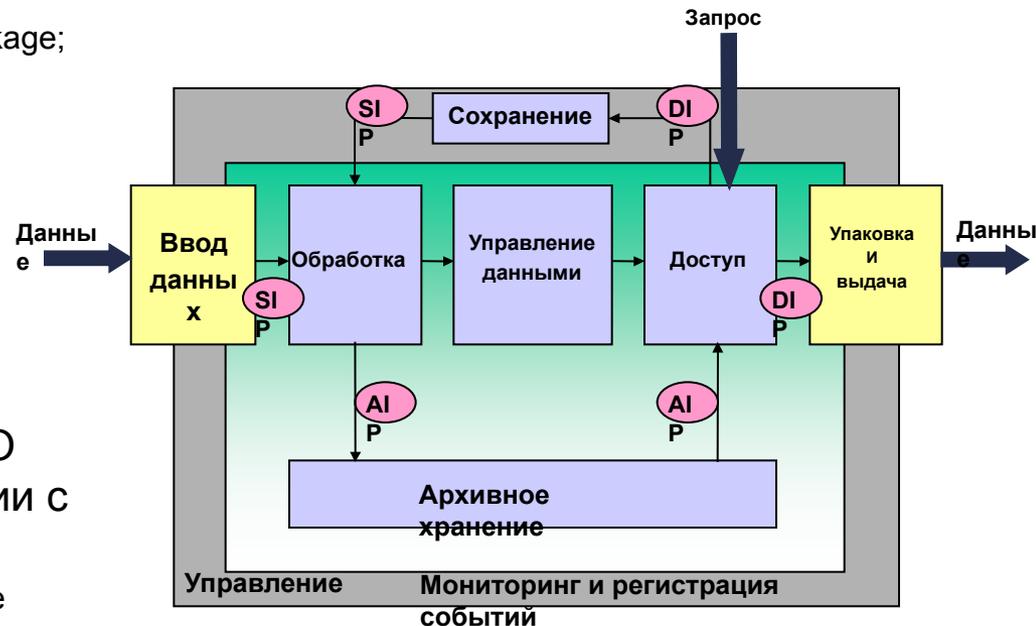


Позиция и решения IBM



Модель IBM DIAS основана на результатах проекта Networked European Deposit Library (EU-Projekt 2000 – 2002)

- **Ввод данных:**
обработка входящей информации или оцифровка объекта:
 - SIP – Submission Information Package;
- **Упаковка и выдача:**
интерфейс вывода данных из депозитария:
 - проверка запросов;
 - подготовка и выдача электронного документа (при необходимости – вместе с ПО для просмотра) в соответствии с метаданными:
 - Dissemination Information Package (DIP).
- Основа – IBM Content Manager;
- Digital Media Center.



Решение «Электронная библиотека» IBM DIAS:



Обеспечение фактической и юридической значимости объекта хранения

- Верификация аутентичности документа через применение ЭЦП необходима на всех значимых этапах преобразования объекта:
 - Создание эталонной копии для архивного хранения;
 - Создание рабочей копии для оперативной деятельности;
 - Любое перемещение эталонной копии внутри архива;
 - Любое преобразование во время миграции объекта;
 - При необходимости – при выдаче заверенной копии клиенту;
 - И т.д.
- Услуги ДТС – для обеспечения как фактической, так и юридической значимости.



Производственные службы электронной библиотеки

- Лаборатория оцифровки объектов;
- Прием и обработка объектов на цифровых носителях;
- Подготовка объектов архивного хранения;
- Каталогизация;
- Электронного хранилище;
- Сохранение и миграция данных;
- Обеспечение фактической и юридической значимости;
- Служба экспорта цифровых ОАХ;

- Научно-исследовательская служба (НИС);
- Служба обеспечения Web-портала;
- Служба информационной безопасности.

Вычислительная инфраструктура ЭБ

Централизованные ИТ-сервисы:

- Электронный документооборот;
- Средства коллективной работы;
- Электронная почта, ERP, биллинг...
- Информационная безопасность;
- Обеспечение аутентичности и юридической значимости;
- Интернет-портал;
- Библиографический каталог.



Хранилище электронной библиотеки

- Главный сервер управления хранением документов;
- Индекс поиска и буфер активных документов;
- Долговременный архив документов.



Функциональные подразделения:

- Поддержка клубной деятельности;
- Межбиблиотечный обмен;
- Общественный Совет;
- Экспозиционная деятельность;
- Научно-просветительская деятельность;
- Организационно-методическая поддержка исследований;
- Электронный читальный зал;
- Конференц-зал, Видео-зал, Издательский отдел (10-15 подразделений Библиотеки).



- Серверы рабочих групп;
- Рабочие места пользователей;
- Локальные функциональные задачи.

Технологические службы ЭБ:

- Подготовка объектов архивного хранения;
- Приём и обработка объектов на цифровых носителях;
- Каталогизация;
- Электронное хранилище;
- Сохранение и миграция объектов;
- Экспорт цифровых ОАХ;
- Научно-исследовательская служба (НИС).



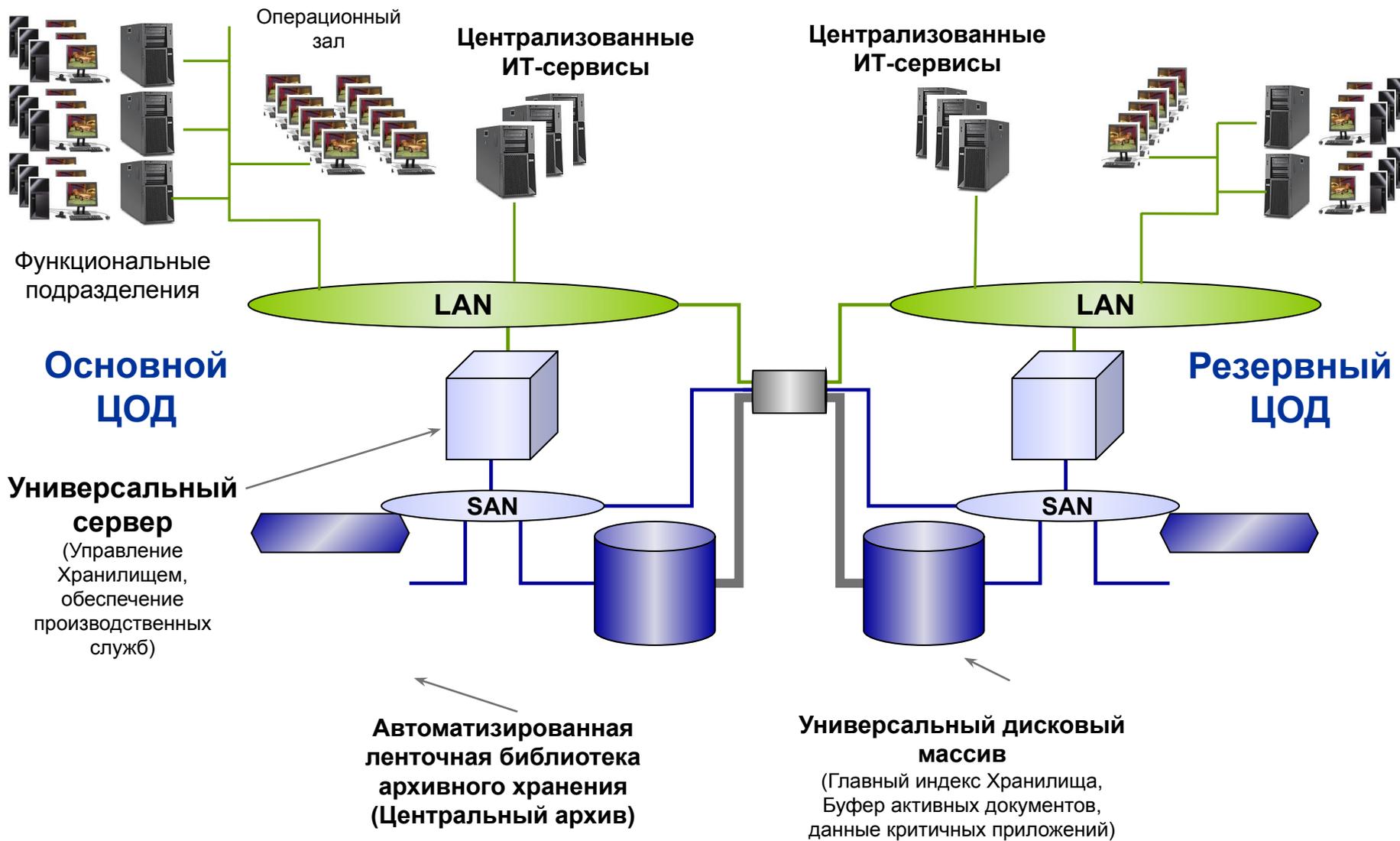
- Технологические кластеры
- Специализированное оборудование
- Автоматизация функций DIAS



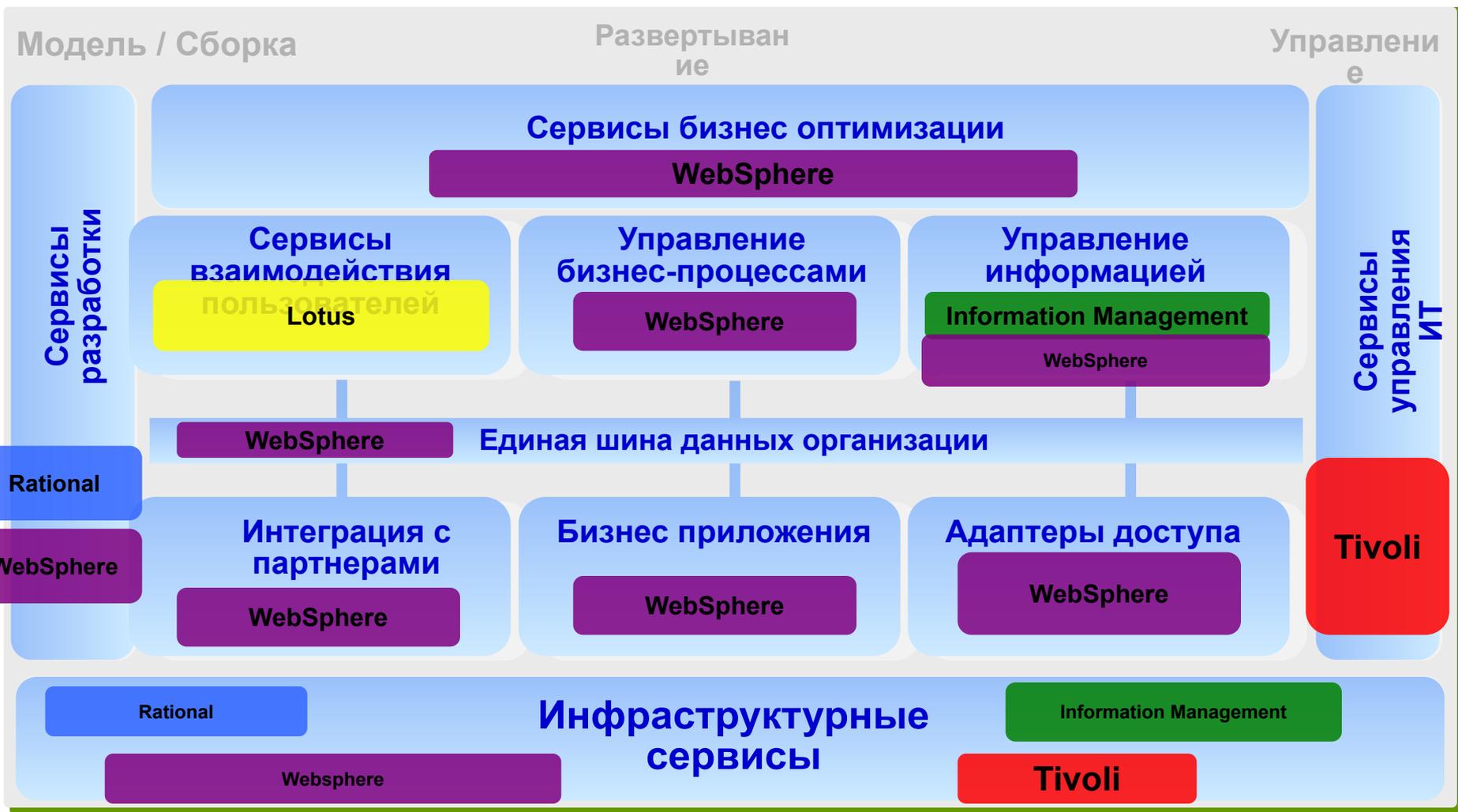
Общая архитектура



Общая архитектура электронно-цифрового архива



Референсная модель программного обеспечения





Примеры реализации

5

Государственный Эрмитаж

- Благотворительный проект IBM II-й половины 90-х гг.
- Продолжение проекта – 2007 г.
- Особенности
 - Решение «под ключ»;
 - Введение новых административно-производственных процедур в жизнь музея;
 - Решение ряда важнейших технологических вопросов с нуля.
 - В течение нескольких лет сайт Эрмитажа – среди победителей международных интернет-конкурсов на лучший электронный музей.

СПРАВКА КАРТА САЙТА ОТЗЫВ СЛУЖБМ КАЛЕНДАРЬ ЗАКАЗ БИЛЕТОВ SHOP

ГОСУДАРСТВЕННЫЙ ЭРМИТАЖ
САНКТ-ПЕТЕРБУРГ, РОССИЯ

БЫСТРЫЙ ПОИСК

ИНТЕРНЕТ МАГАЗИН

ЗАКАЗ БИЛЕТОВ

ГЛАВНЫЕ НОВОСТИ

20 | 21
ГОСУДАРСТВЕННЫЙ ЭРМИТАЖ
The State Hermitage Museum

новости

новое на сайте

◆ ENGLISH LANGUAGE

◆ ИНФОРМАЦИЯ
◆ ШЕДЕВРЫ КОЛЛЕКЦИИ
◆ ВЫСТАВКИ
◆ ИСТОРИЯ ЭРМИТАЖА
◆ ОБУЧЕНИЕ И ОБРАЗОВАНИЕ
◆ ЦИФРОВАЯ КОЛЛЕКЦИЯ

◆ Виртуальный визит

◆ Виртуальные выставки

◆ Виртуальная академия

IBM

Государственный Эрмитаж занимает шесть величественных зданий, расположенных вдоль набережной Невы в самом центре Санкт-Петербурга. Ведущее место в этом неповторимом архитектурном ансамбле, сложившемся в XVIII - XIX веках, занимает Зимний дворец - резиденция русских царей, построенная в 1754 - 1762 годах по проекту Ф.Б. Растрелли. В музейный комплекс входят также восточное крыло здания Главного штаба, Меншиковский дворец и недавно построенное Фондохранилище.

Почти за два с половиной столетия в Эрмитаже собрана одна из крупнейших коллекций, насчитывающая около трех миллионов произведений искусства и памятников мировой культуры, начиная с каменного века и до нашего столетия. Сегодня, при помощи современных технологий, музей создает свой цифровой автопортрет, который смогут увидеть во всем мире.

<http://hermitagemuseum.org>

Королевская библиотека Нидерландов

■ Особенности проекта:

- Центральная часть европейского проекта;
- Первые в Европе, опыт реальной производственной эксплуатации в течение нескольких лет;
 - Конкурс: Siemens, Wang, CAP Gemini, Getronics/Wang, Pink Roccade.

■ Комплекс:

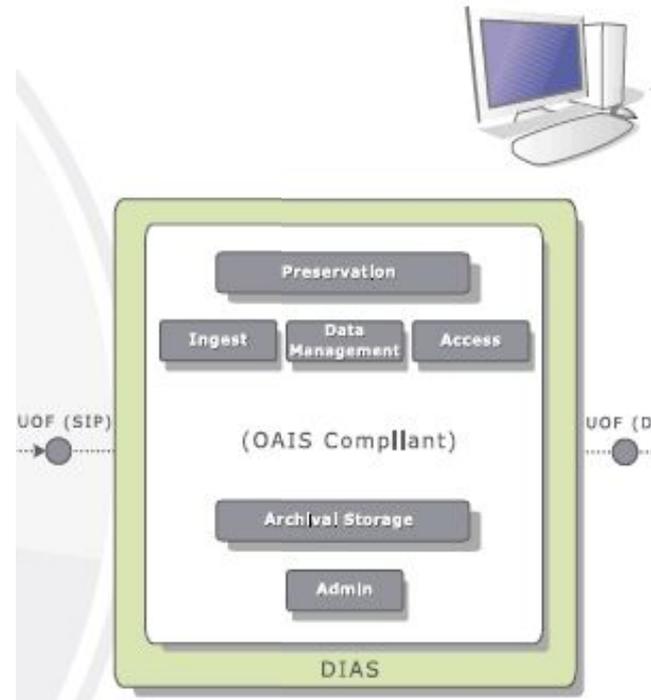
- Ядро системы – DIAS:
 - WebSphere;
 - Content Manager,
 - Tivoli Storage Manager (TSM),
 - DB2.
- 7-node RS/6000 SP server (processing of the document input/output);
- RS/6000 SP 375 MHz POWER3 SMP Wide server - processing software;
- RS/6000 F50 server - Control Workstation for the RS/6000 SP;
- Enterprise Storage Server,
- 3 x 3494 Enterprise Tape Libraries
- 3995 Model C68 Optical Library.





Kopal Project

- **Национальная библиотека Германии**
 - Deutsche Nationalbibliothek
(The German National Library)
- **Научное общество по обработке информации**
 - Gesellschaft für wissenschaftliche
Datenverarbeitung mbH Göttingen
- **Государственная библиотека
Университета Гёттингена**
 - Niedersaechsische Staats- und
Universitaetsbibliothek Goettingen (Goettingen
State and University Library)
- IBM Deutschland GmbH
- **Ядро: DIAS+**



<http://kopal.langzeitarchivierung.de/>

kopal 
Data into the future



Государственная библиотека Тяньцзинь

- Особенности проекта:
 - Государственная библиотека в КНР;
 - Первый подобный проект;
 - Конкурс: Computer Associates, Dell, EMC, Legend, Oracle + местные компании.
- Комплекс:
 - 7 x eServer pSeries 630 security control servers;
 - 8 x Server pSeries 650 digital content metadata servers and resource managers;
 - eServer xSeries 235 - library automation server;
 - eServer xSeries 335 - team development server;
 - eServer xSeries 345 - staging server;
 - eServer xSeries 360 - a portal server;
 - 2 SAN 2109 Fibre Channel Switches;
 - IBM 3583 L18 Ultrium Scalable Tape Library.
- Information Management:
 - DB2 Content Manager for AIX, DB2 Content Manager VideoCharger;
 - Lotus: WebSphere Portal Extend;
 - Rational Software: Rational Web Developer for WebSphere Software;
 - Tivoli Storage Management: Tivoli SANergy, Tivoli Storage Manager;
 - Tivoli Security Solutions: Tivoli Access Manager for e-business;
 - WebSphere: WebSphere Application Server – Enterprise.



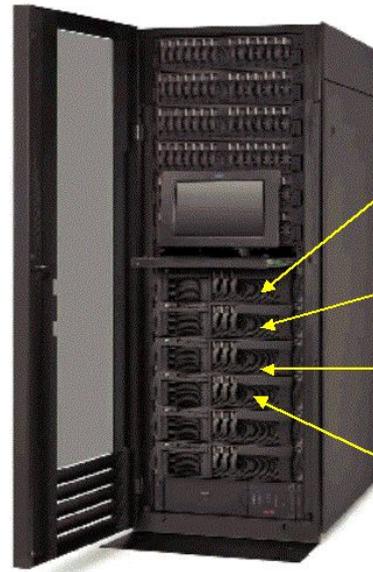
Библиотека университета Mashiakhat Al Azhar

Особенности проекта:

- Оцифровка и каталогизация древних манускриптов.

Комплекс:

- IBM RS/6000 SP2;
- IBM RS/6000 B50;
- IBM Netfinity 4500R;
- Storage:
 - Optical Storage;
 - Tape Storage.
- Программные платформы:
 - IBM Content Manager;
 - IBM Enterprise Information Portal (EIP);
 - IBM DB2;
 - IBM WebSphere Application Sever and Edge Server.



Server name: DC
Function: Domain Controller
IP: 10.1.1.200
Platform: Windows 2000 Server

Server name: XCHANG
Function: Mail Server
IP: 10.1.1.201
Platform: Windows 2000 Server



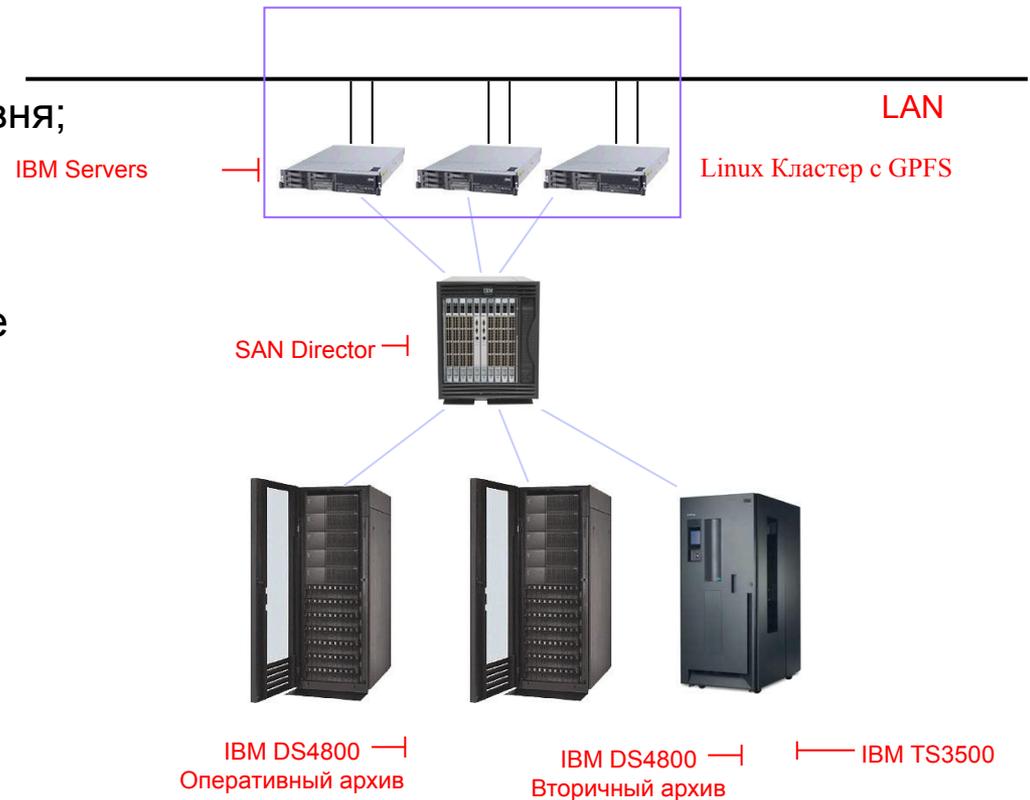
Высшая школа менеджмента СПбГУ



- **Особенности проекта:**
 - Головной ВУЗ СЗФО;
 - Школа бизнеса мирового уровня;
 - Электронная библиотека;
 - Ядро системы – Digital Media Center;
 - Проектирование – российские бизнес-партнеры.

- **Сроки:**
 - Проектирование – 6 мес.
 - Исполнение (архив) – 12 мес.

- **Контроль за исполнением:**
 - Приоритетный национальный проект «Образование».



Высшая школа менеджмента
 Санкт-Петербургского государственного университета

Аппаратная платформа

LAN

■ Центральный кластер из 3-х серверов:

- 4-CPU Power5+ 2.2 GHz;
- 16 GB RAM;
- 1x 2-Port Ethernet x 1 GB;
- 2 x 2-Port Fiber-Channel x 2 GB;
- 4 HDD x 73GB;
- Administrator Console.

IBM System p570
Кластер с GPFS

IBM
BladeCenter

SAN
Director

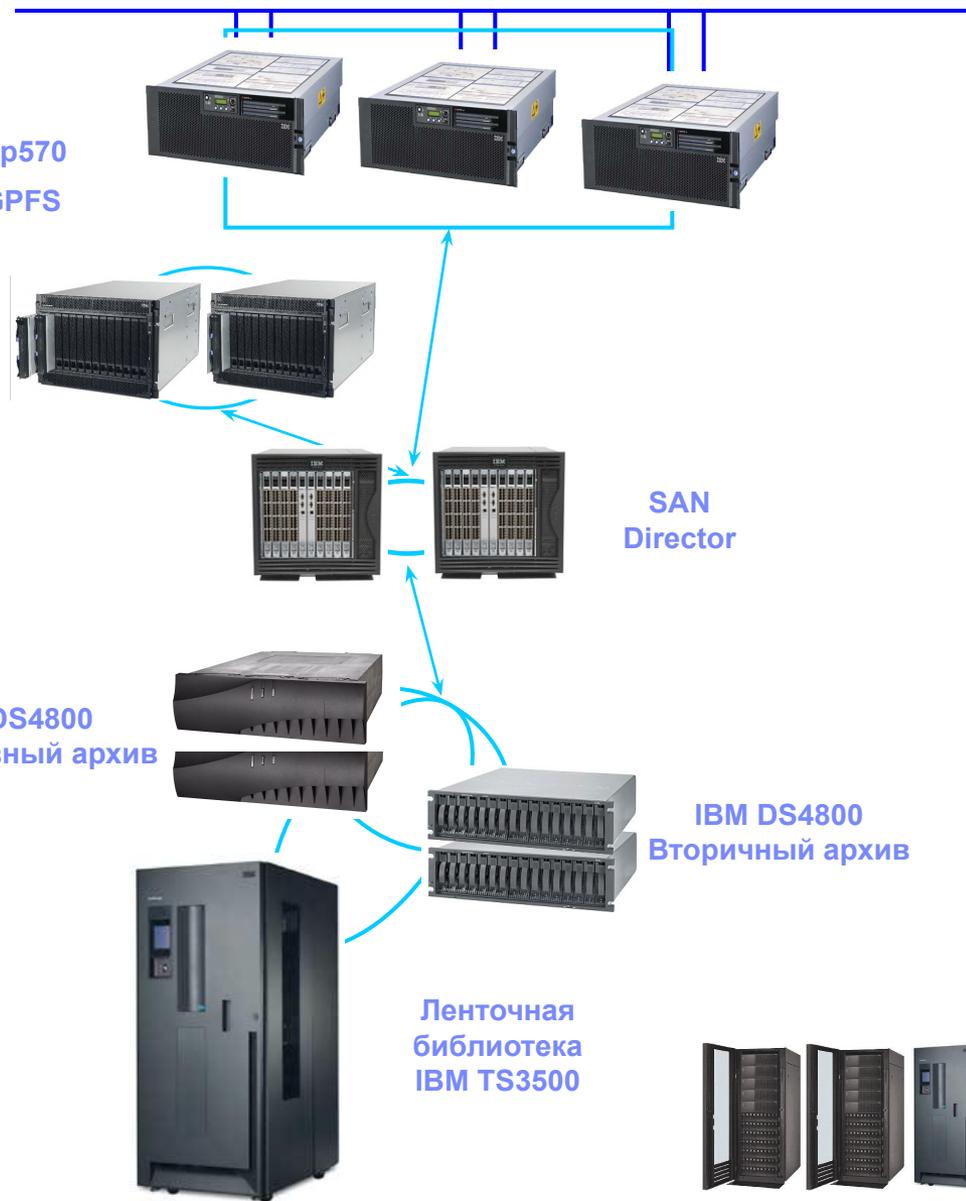
■ Цифровой архив:

- Оперативный архив: DS4800 **6TB**
 - 48 FC HDD x 146 GB;
- Вторичный архив: DS4800 **24 TB**
 - 48 SATA HDD x 500GB;
- Доловременной архив: TS 3500 **100TB**
 - 4 LTO-3 привода чтения / записи;
 - 260 кассет LTO3.

IBM DS4800
Оперативный архив

IBM DS4800
Вторичный архив

Ленточная
библиотека
IBM TS3500





Вопросы и ответы

Валерий Елизов

R90423@ru.ibm.com

+7 (985) 761-9108





Дополнительные материалы

**Системы хранения данных
IBM**



Семейство дисковых систем IBM System Storage DS

Наиболее широкая продуктовая линейка в индустрии



DS3200/DS3400

DS4000

DS6000

DS8000

Начальный уровень

Средний уровень

Уровень предприятия

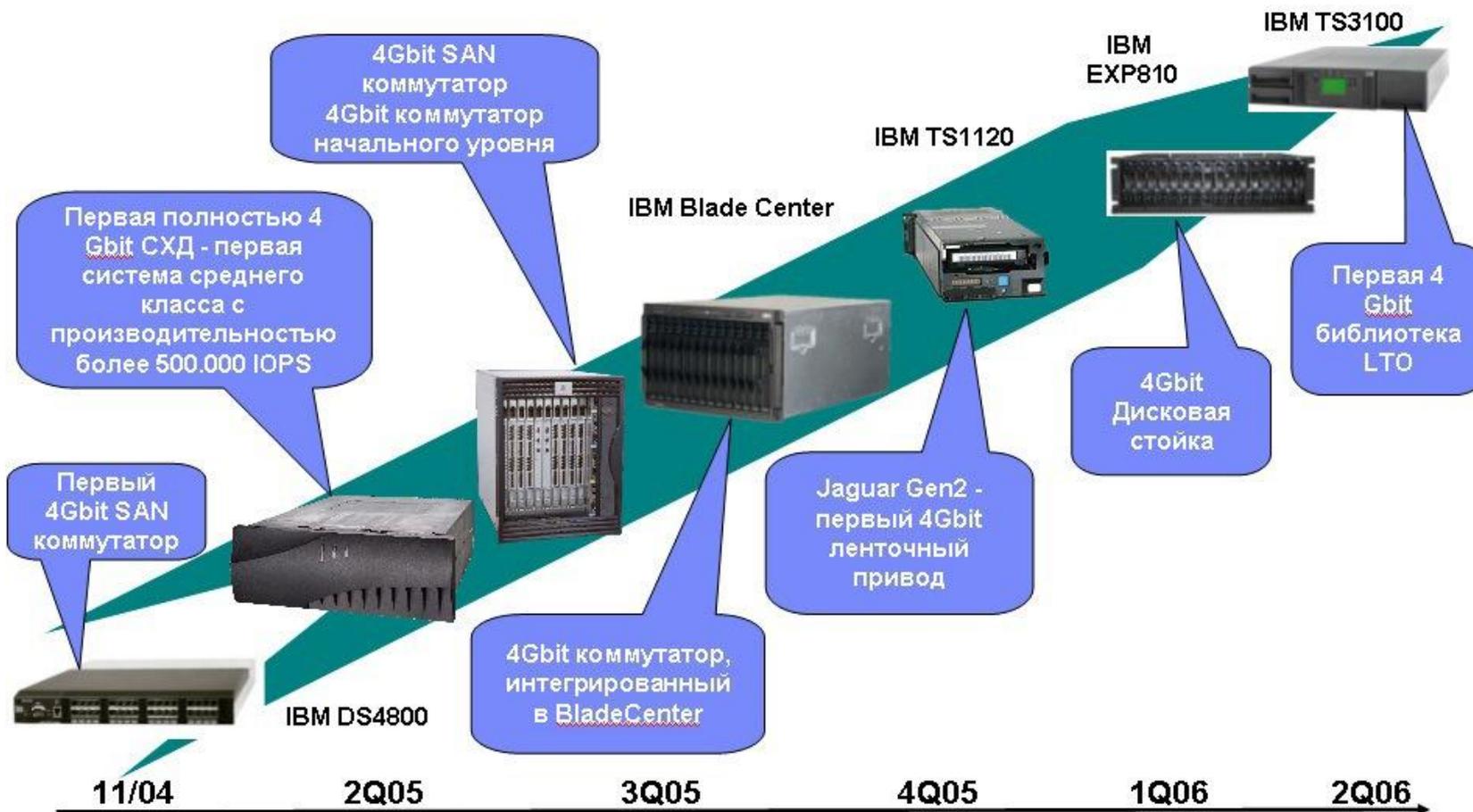
Инновации семейства IBM System Storage DS помогают:

Упростить инфраструктуру и уменьшить затраты на ее управление, внося изменения по мере необходимости.

Обеспечить непрерывность бизнеса и защиту данных.

Эффективно управлять информацией на протяжении всего жизненного цикла.

IBM - авторитет индустрии в 4Gbit сетях хранения



Пример инфраструктуры для ILM



Тип СХ



Основная
система
(FC диски)



Промежуточное
хранилище
(SATA диски)



Архив
(Лента)

Класс данных

Критичные:

- Высокая производительность;
- Постоянно и активно используемые.

Актуальные:

- Умеренная производительность;
- Периодическое использование.

Архивные данные:

- Низкая производительность;
- Редкое использование и хранение на всякий случай.

Задачи:

- Снизить издержки хранения;
- Уменьшить время восстановления.

Решение:

- Несколько уровней хранения;
- Производительная основная СХ;
- Промежуточное хранилище;
- (например: DS4200 или EXP810 с SATA дисками);
- Долговременный архив на ленте.

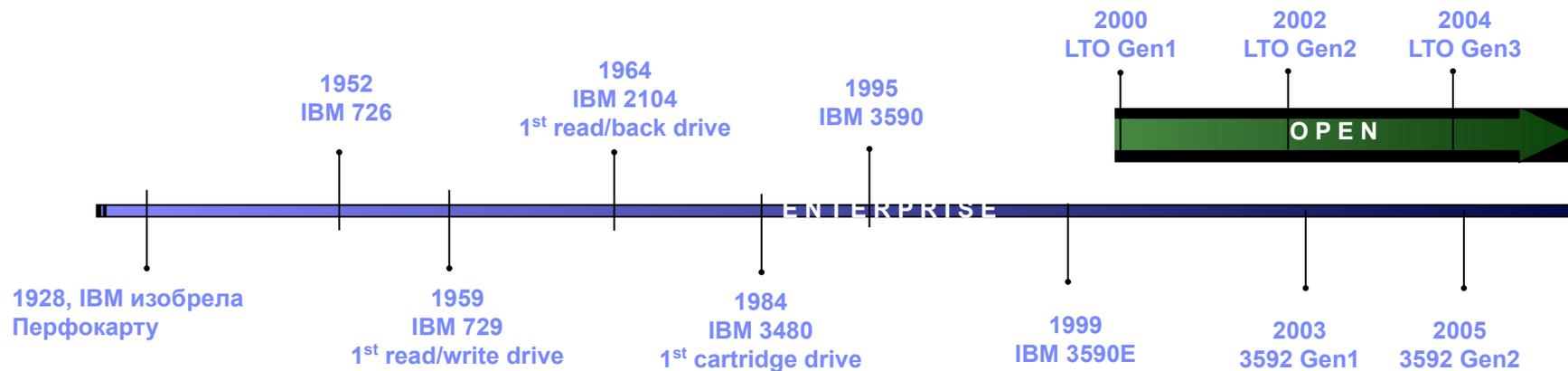
Преимущества:

- Соотношение стоимости информации с издержками на хранение;
- Быстрое восстановление;
- Лучшее использование ресурсов.

IBM: более 50 лет в разработке лент

- **Начало - 1952 год:**
 - IBM 726 Tape Unit
 - 7,500 cps;
 - 100 bpi.

- **И продолжение в 2005 году:**
 - IBM TS1120 Tape Drive
 - 100 MB/sec;
 - 500 GB.



Ленточные системы IBM System Storage

Наиболее широкая продуктовая линейка в индустрии



TS3100

TS3200 / TS3310

TS3500

3494

Virtual Tape

LTO

3592/TS1120

Начальный уровень

Средний уровень

Уровень предприятия

Инновации семейства IBM System Storage:

- Обеспечивают непрерывность бизнеса и защиту данных.
- Эффективно управляют информацией на протяжении всего жизненного цикла.
- Обеспечивают гибкость построения и расширяемость систем.
- Обеспечивают высокую рентабельность систем хранения

UltraScalable Tape Library TS3500

- Поддерживает до 16 стоек, до 192 приводов, до 6887 лент, максимальная емкость до 2754 ТБ;
- Восстановление канала передачи данных после сбоя;
- Балансировка нагрузки;
- Поддерживает Ultrium 3 и IBM 3592 приводы;
- Multi-Path архитектура;
- Возможность чтения/записи Ultrium 2 кассет с емкостью Ultrium 2, но с улучшенной производительностью;
- Восстановление пути управления под AIX;
- Улучшение управления/функциональности.



