



IBM Software Group

DB2 Information Integrator Software

Марк Ривкин IBM EE/A

DB2 Information Management Software



@.business on demand software

Виды интеграции

- **Интеграция интерфейсов - Portal**
- **Интеграция приложений - WS BI Server**
- **Интеграция данных**

- **Хранилища данных**
- **Information Integrator**

Два подхода к интеграции данных

■ Консолидация данных для локального доступа (DW)

- 4 Высокая производительность и доступность (HA) при централизации данных.
- 4 Нужны не самые свежие, а согласованные на данный момент времени данные
- 4 Для семантической согласованности данных нужны сложные преобразования (ETL)
- 4 Production applications, data warehouses, operational data stores
- 4 Обычно используется ETL (Extract, Transform, and Load) или репликация
- 4 Нужны исторические данные
- 4 Большие объемы данных
- 4 Структурированные данные

■ Интегрированный доступ к распределенным источникам (II)

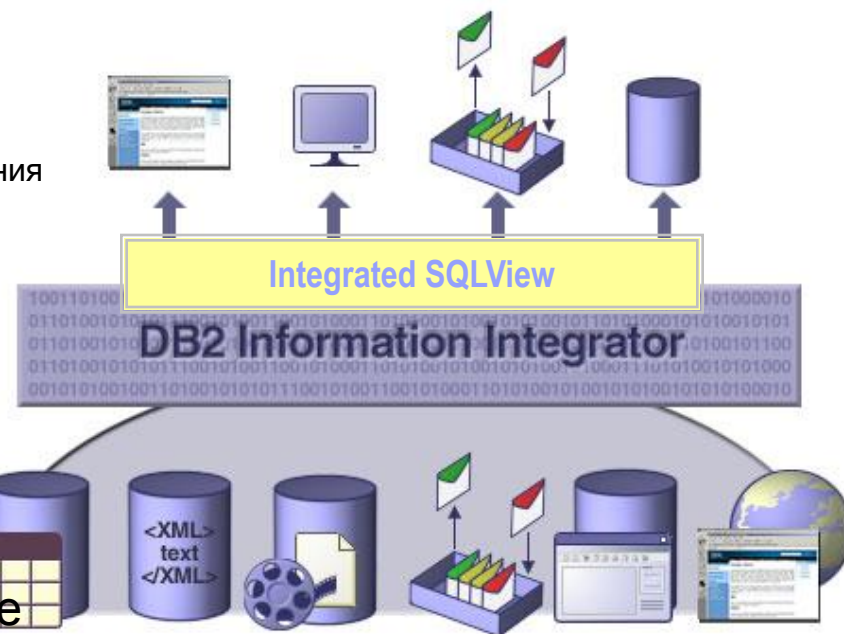
- 4 Производительностью доступа и загрузки можно пожертвовать в угоду простоте и снижению цены
- 4 Нужны самые свежие данные
- 4 Политики безопасности, лицензирования или внутренние правила ограничивают возможность перемещения данных
- 4 Смешение форматов данных, например customer ODS и контрактные документы или изображения
- 4 Данные нужны редко и нет смысла их хранить в DW

■ Наибольшие выгоды – при комбинации обоих подходов

DB2 Information Integrator - “Зонтик”

Федеративный сервер данных – Работать с распределенными данными так, как будто они находятся в одном источнике данных

- Описать интегрированный взгляд на разнотипные распределенные данные
 - 4 Широкий диапазон источников данных и контента
 - 4 Расширяемость на любые источники данных
- Запросить как бы из единого источника
 - 4 Использовать стандартные SQL запросы и SQL выражения
 - 4 Использовать контекстный поиск
 - 4 Реализовать специализированные функции в SQL
 - 4 Использовать оптимизацию запросов и хэширование
- Работать с XML документами
 - 4 Комбинировать различные источники
 - 4 Использовать DTDs или XML схемы
- Единый источник, реляционное обновление



DB2, Oracle, SQL Server, Sybase, Teradata, OLE DB, ODBC, Excel, XML, message queues, Web services, flat files, document repositories, content repositories, LDAP directories, WWW, email databases, and more.

Сценарии использования II

Business Intelligence

Расширение данных хранилища свежими данными

Взгляд на все данные предприятия

Непрерывное пополнение DW свежими данными

Portal

Единый интерфейс к разным источникам

Доступ к старым данным

Единые портлеты

Единый поиск

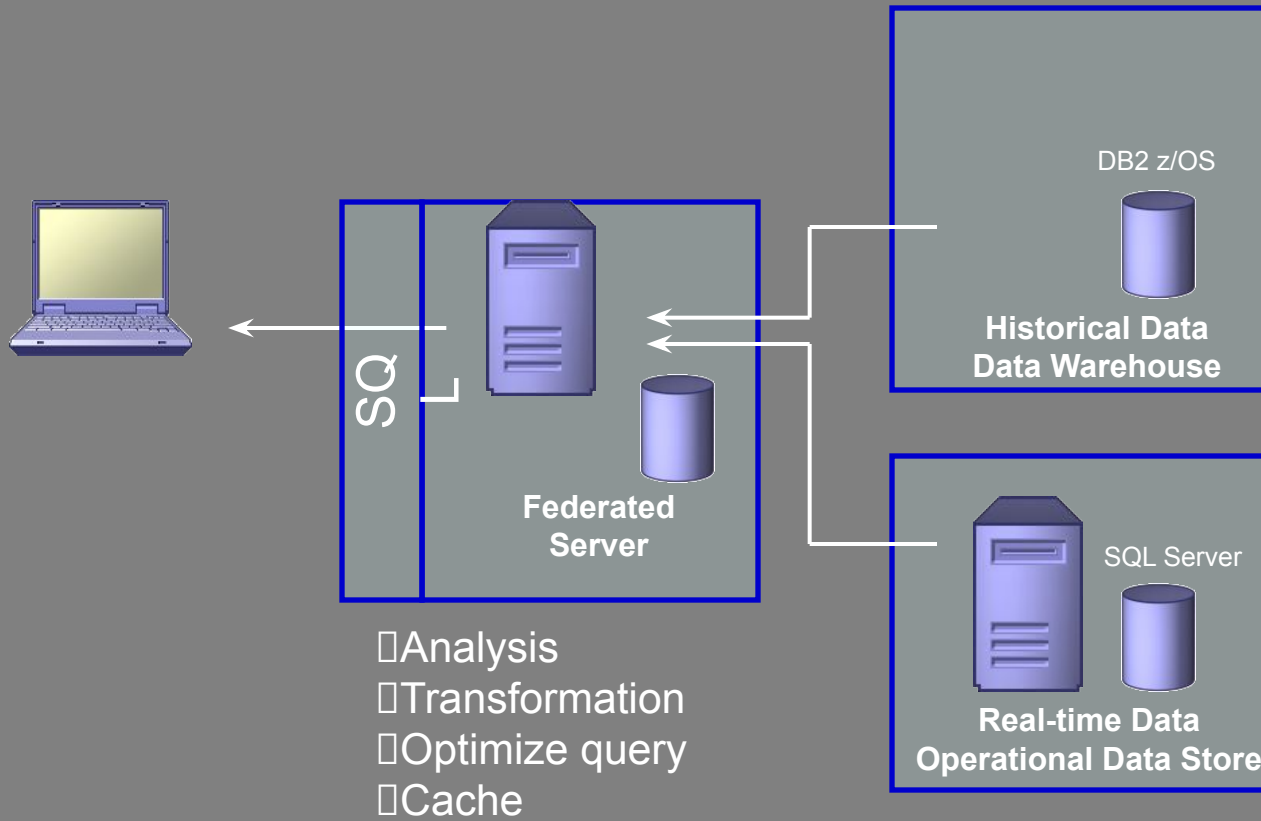
Migration/Coexist

Слияние и приобретение

Миграция между БД

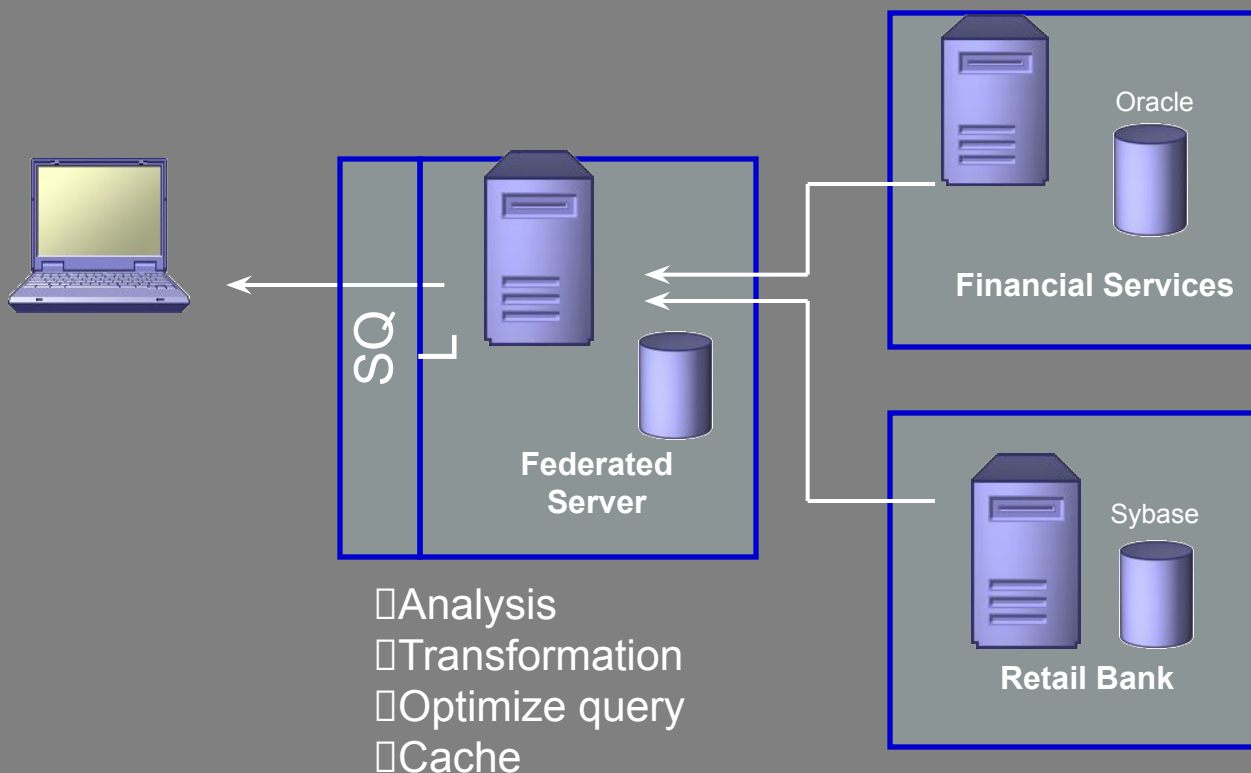
Синхронизация БД

Расширение Data Warehouse оперативными данными



Обеспечение единого взгляда на данные компании

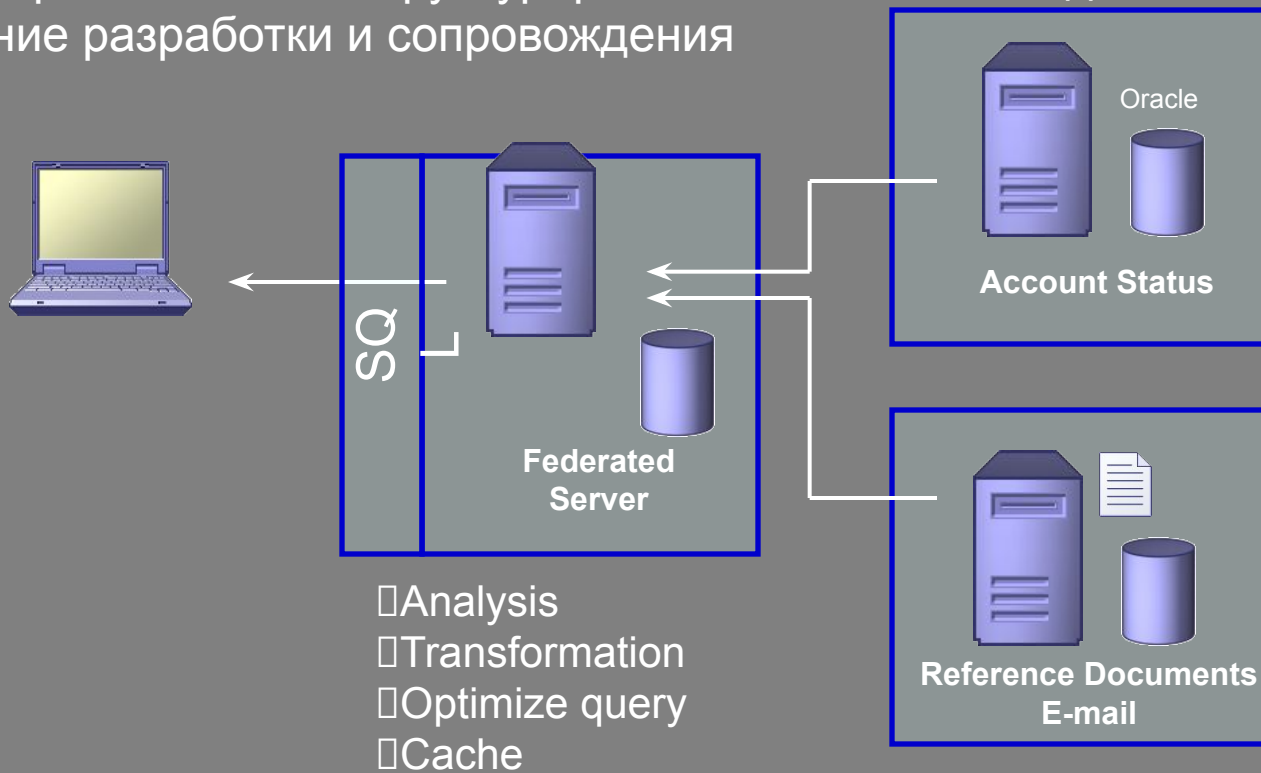
- ✓ Обеспечение единого взгляда на клиентов и контракты, хранящиеся в разных системах
- ✓ Доступ к данным на месте хранения – не надо их копировать



Документы, контракты, текущая инф

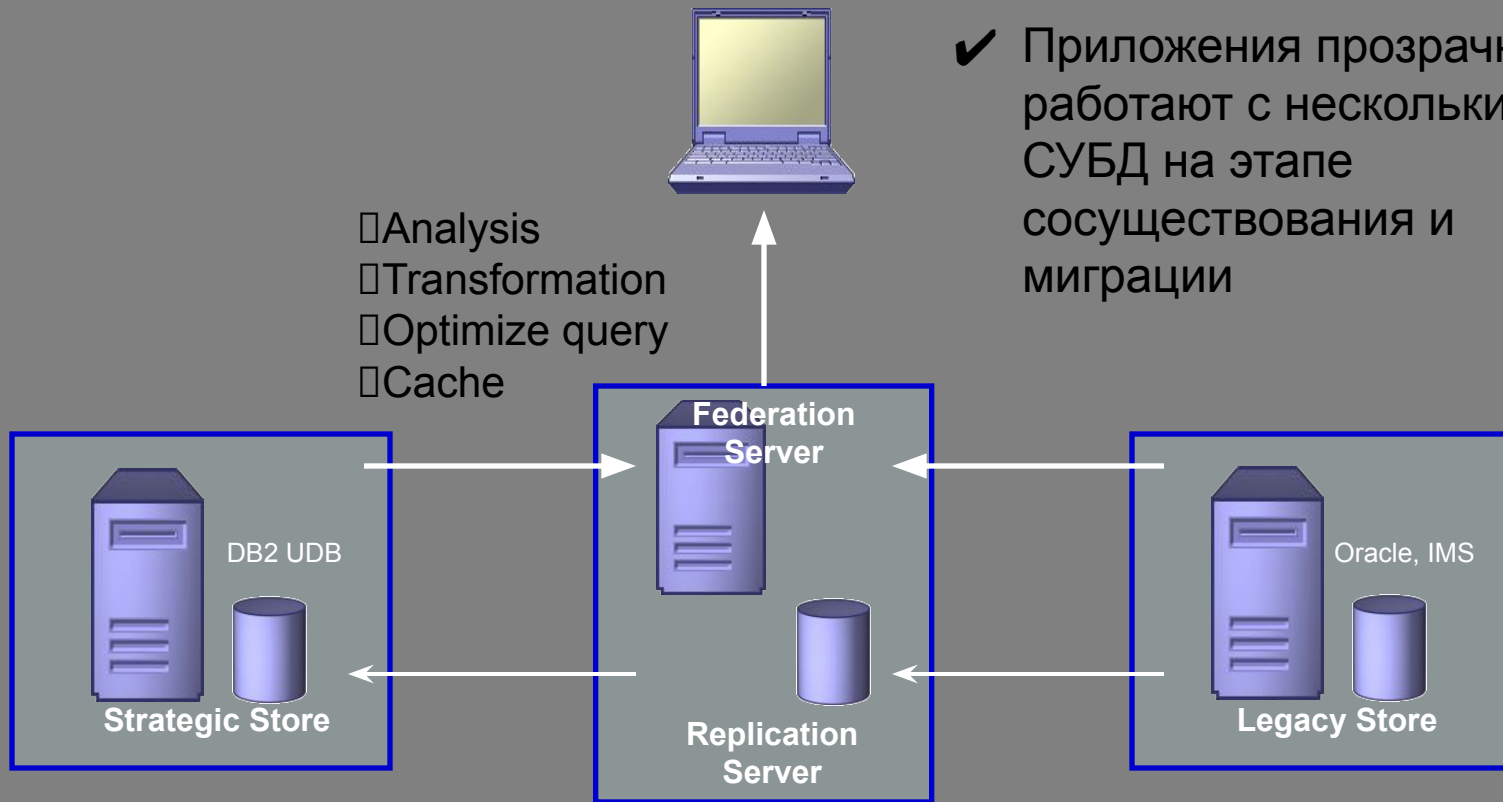
Ускорение разработки портала

- ✓ Обеспечение доступа приложениям портала к множеству структурированных и неструктурированных источников данных
- ✓ Упрощение разработки и сопровождения



Логика в одном месте

Сосуществование различных СУБД



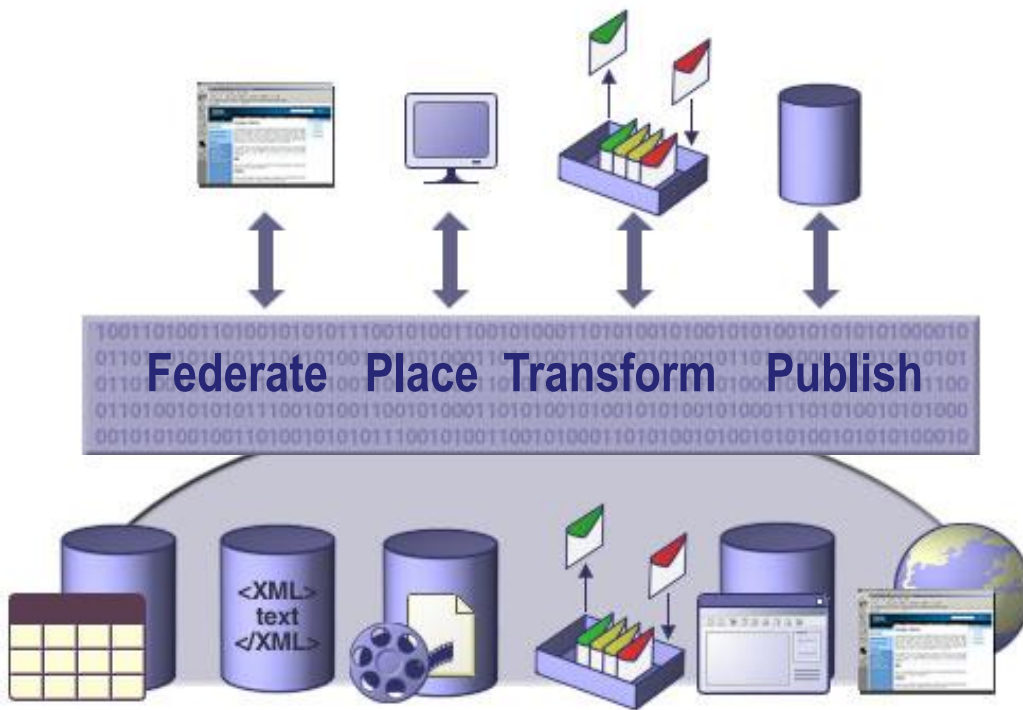
- Analysis
- Transformation
- Optimize query
- Cache

✓ Приложения прозрачно работают с несколькими СУБД на этапе сосуществования и миграции

- Consolidation
- Transformation
- Manage Copies

✓ Репликация для синхронизации данных

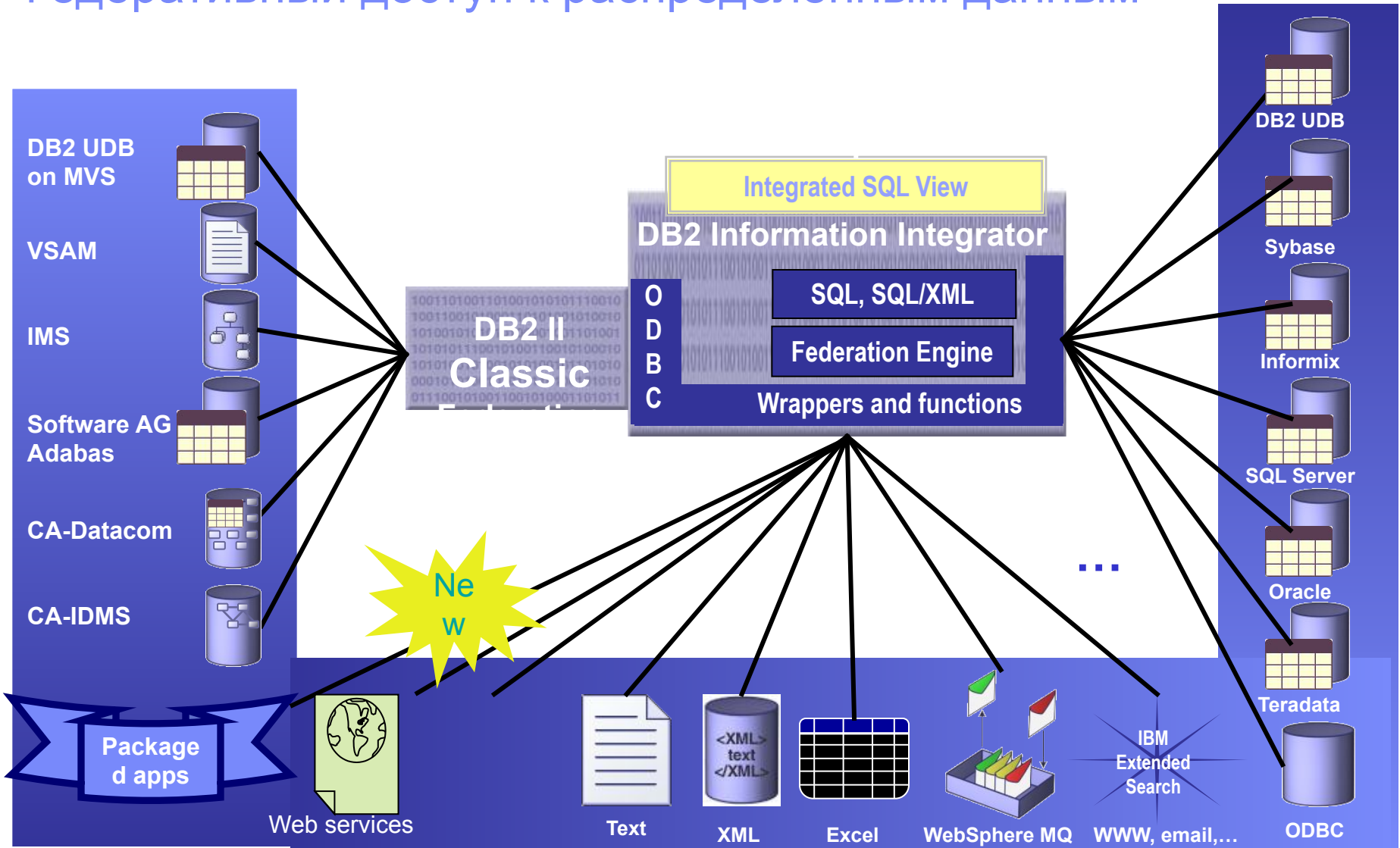
IBM DB2 Information Integrator обеспечивает:



Metadata management and interchange

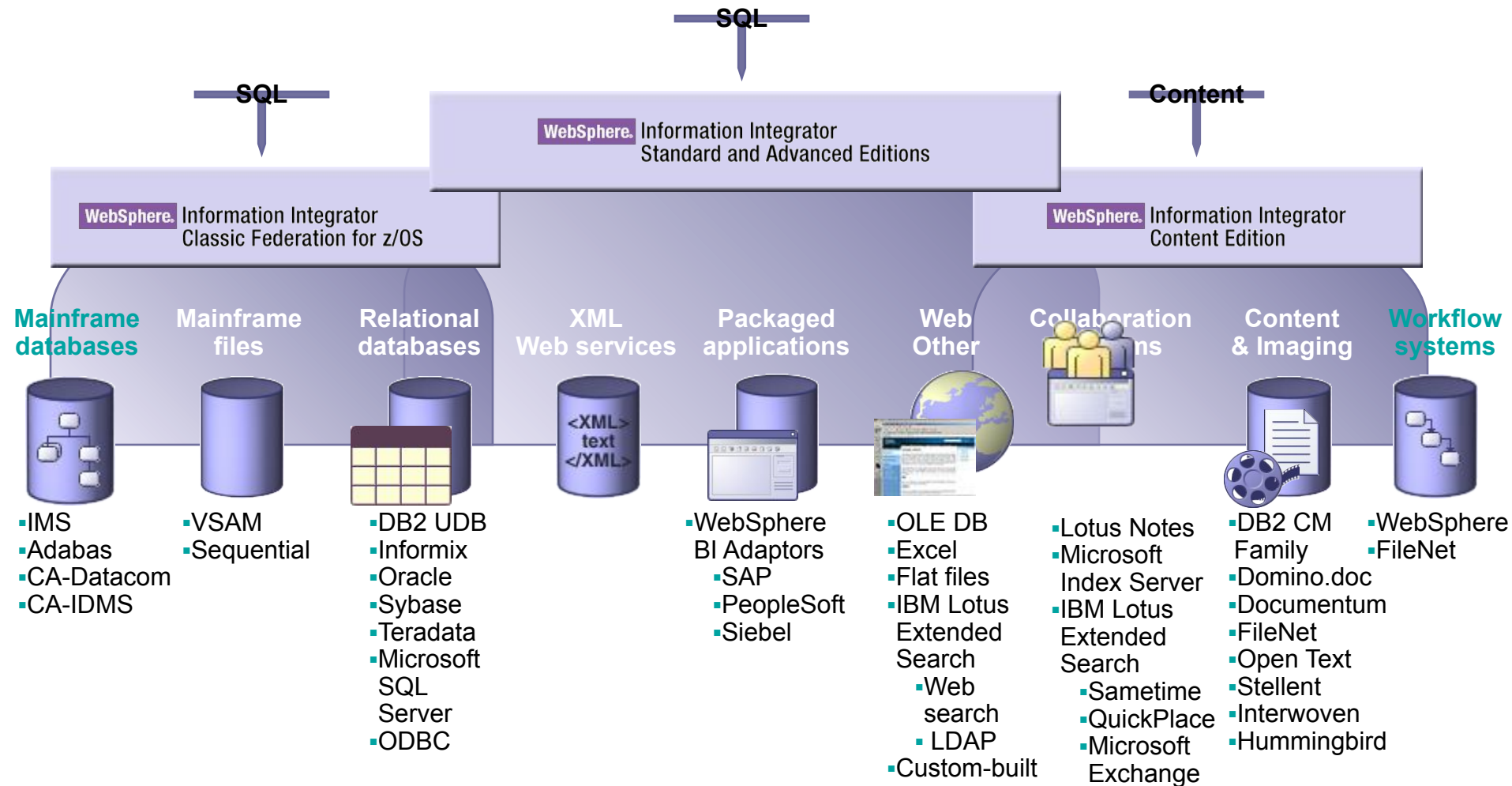
- **Виртуализацию данных**
 - 4 Стандартные языки доступа
 - SQL, XQuery, Content, Search
 - 4 Доступ на чтение/запись к различным источникам данных и контента
 - 4 Расширение доступа
- **Data placement**
 - 4 Кэширование, извлечение и загрузка, репликация разнородной информации
- **Преобразование данных**
 - 4 Контроль качества данных
 - 4 Data analysis and mining
 - 4 SQL, XML, Web services
 - 4 Advanced search
 - 4 Управление метаданными
- **Data publishing**
 - 4 Publishing changed data events from diverse sources
 - 4 MQ интеграция
 - 4 WebSphere business integration

Федеративный доступ к распределенным данным



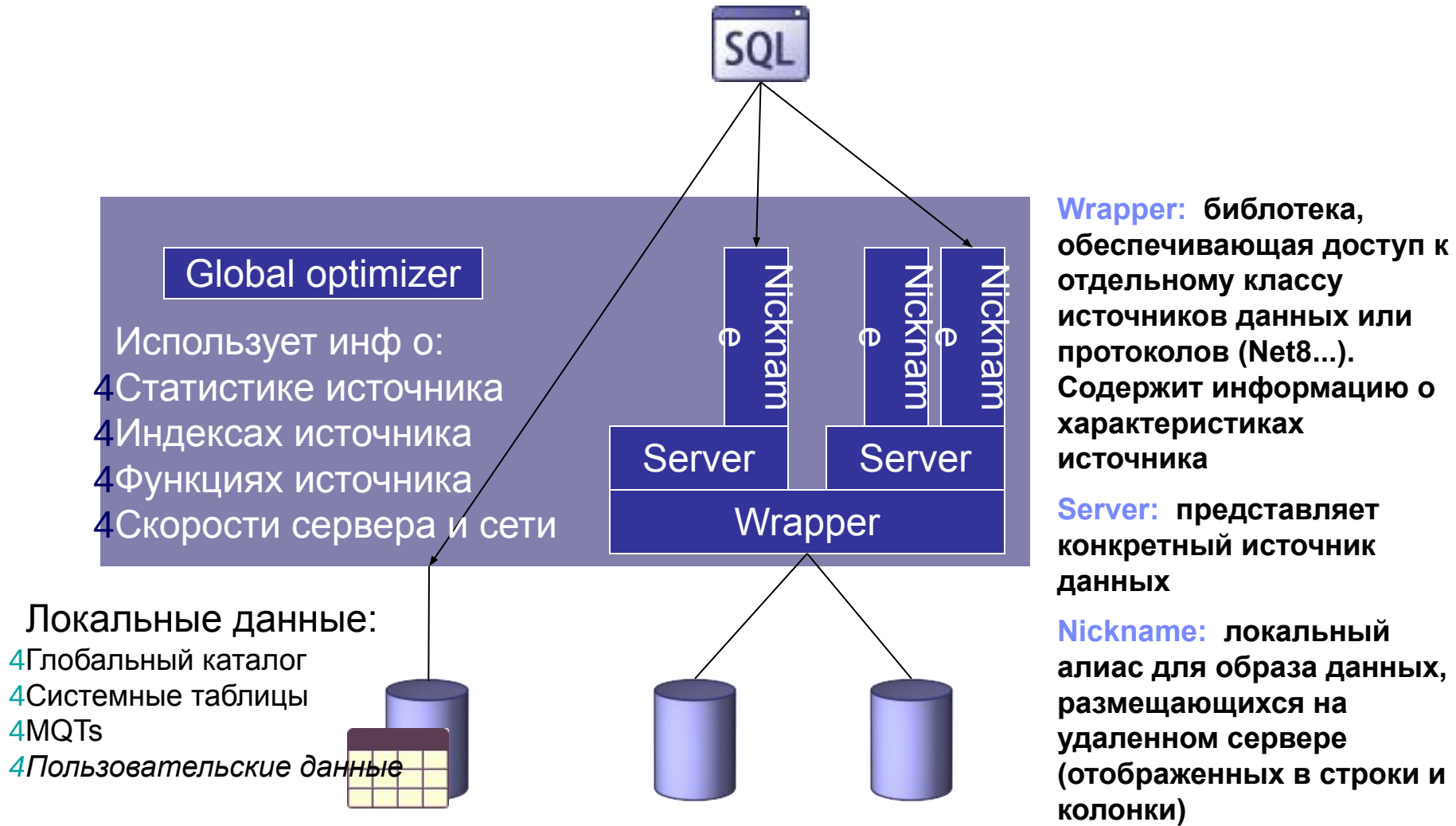
- Data sources accessible by IBM Lotus Extended Search:
Lotus Notes, Lotus Domino.doc, Lotus Discovery Server, Lotus QuickPlace, Lotus Sametime
- IBM WebSphere Portal Search Engine, Microsoft Index Server, Microsoft Site Server, Microsoft Exchange, LDAP Directories,
- 18 Web search engines (Yahoo!, Lycos, Excite, HotBot, Google, Alta Vista, ABCNews.com, AOL.com Search, Business Wire, CNN, Canada.com, EuroSeek, FAST Search, GoTo.com, NBCi, PR Newswire WebCrawler, and Yahoo News)
- Documentum, FileNet, DB2 Content Manager

Federated Sources



Дополнительно инструменты ВР предоставляют доступ к большому кол-ву источников

Концепции



Wrappers

- Выполняют 4 основные задачи:

- f Моделирование данных**

- Отображают исходную модель данных в реляционную модель данных (таблицы с строками и столбцами)
 - Отображают функции в SQL операции

- f Планирование запроса**

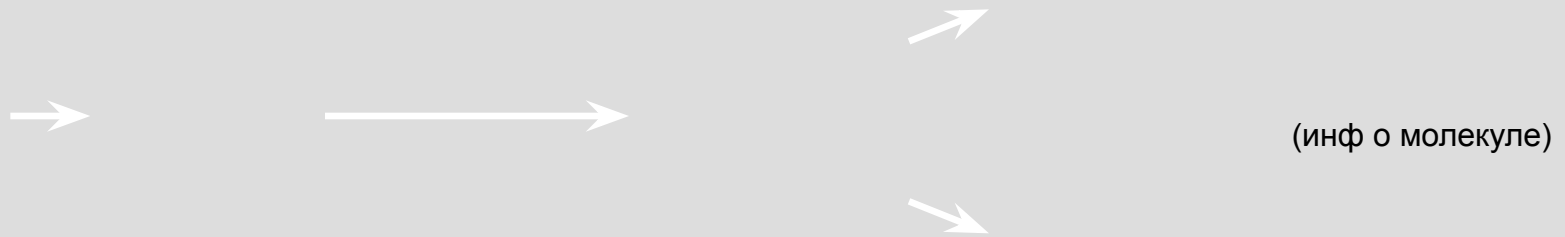
- Предоставляют информацию об источнике
 - Выталкивают на источники как можно больше работы
 - Определяют функции, отсутствующие на источнике (их компенсирует II)
 - Обеспечивают информацию для оптимизатора (cost and cardinality)

- f Connection and transaction management**

- f Выполнение запросов и извлечение данных**

- Выполняют части запроса пользователя на различных источниках данных

Nicknames, Push down and Compensation

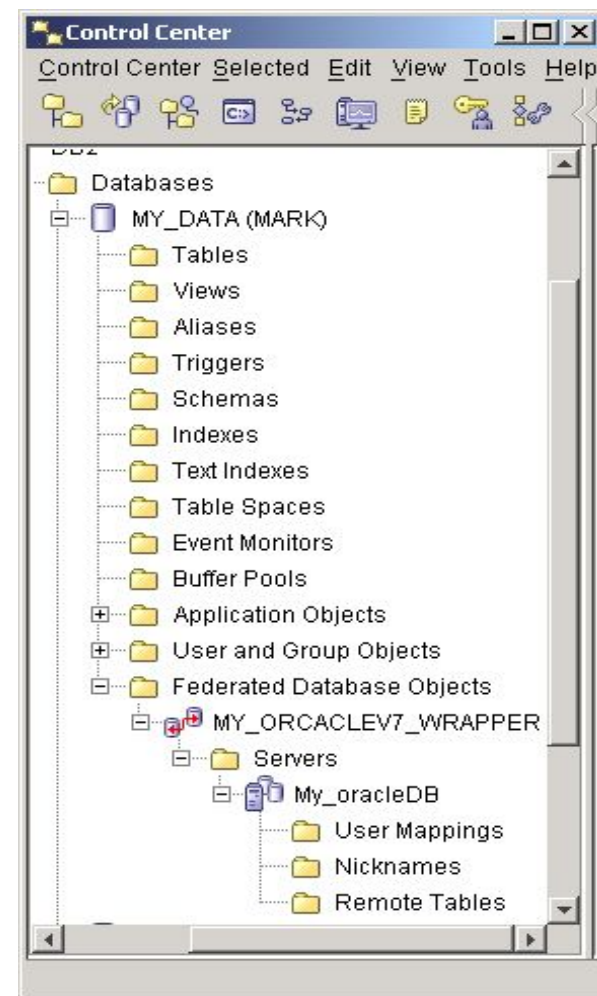


(результаты тестов)

- Push down обработку данных, чтоб улучшить производительность
- Компенсация для пропущенной функции (sort)

Администрирование – Control Center

- Create nickname wizard
 - 4 Помогает создать nickname
 - 4 Выбирает/создает wrapper, server, nickname
- Create cache table wizard
 - 4 Помогает создать кэш таблицы.
 - 4 Выбирает/создает nickname, cache table и конфигурирует репликацию
- Export/ Deploy
 - 4 Запоминает действия control center и записывает их в скрипт. Позволяет редактировать скрипты и выполнять их в Task Center
- Monitoring
 - 4 Мониторит набор nicknames, чтобы определить изменения в схеме данных, которые могут повлиять на эти описания nickname
 - 4 Определяет состояние серверов (работает или недоступен)
 - 4 Обновляет статистику
 - 4 Запускает по расписанию обновление статистики для nickname
- Server discovery
- Конфигурирование II

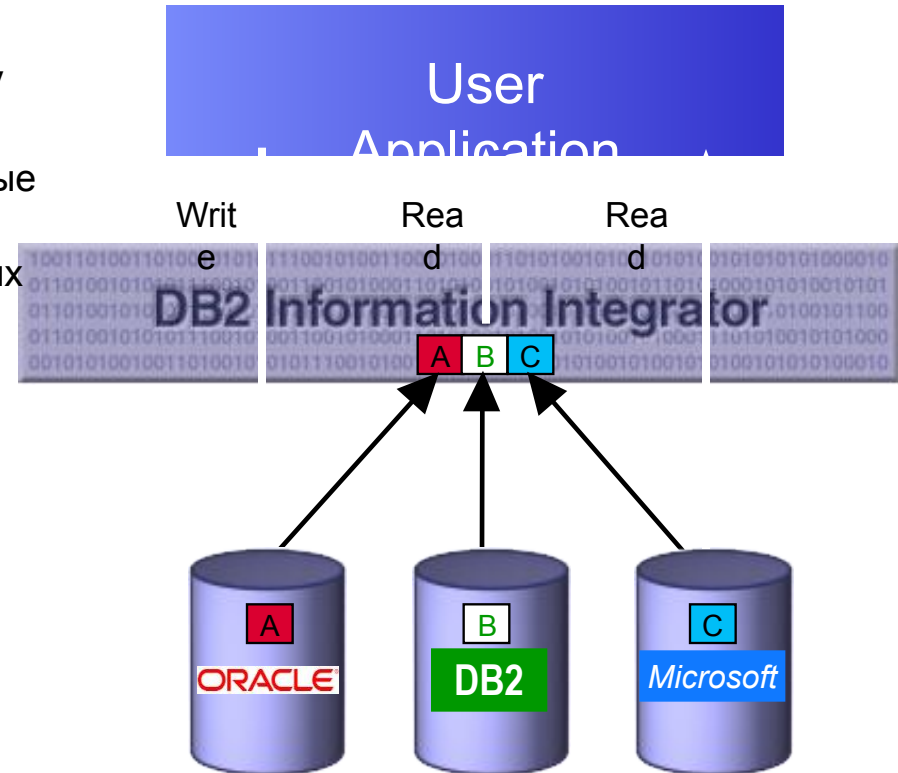


Обработка запросов к федеративным данным

- Разбор и семантика
- Query Rewrite
 - f* Трансформация запроса на основе эвристик и знаний об источнике
- Pushdown Анализ
 - f* Анализирует как декомпонировать запрос
- Cost-Based Optimization
 - f* Генерирует оптимальный план выполнения запроса, используя знания об источнике: **database statistics, indexes, source functions, server capacity, network capacity**
- Генерация SQL
 - f* Производит эффективный для конкретного источника данных SQL (для SQL источников)
- Выполняет запрос над локальными и распределенными данными
- Выполняет компенсацию функций и **virtual database view**

Гетерогенное кэширование

- Улучшает производительность и надежность выполнения запроса
- Администратор описывает Materialized Query Table
 - 4 Заранее вычисленные или часто используемые значения
 - 4 Любые данные из федеративных реляционных источников
 - 4 Неявное или явное использование (automatic routing)
- Разработчик включает использование кэша
 - 4 Если включено – читает из кэша, пишет в источник
 - 4 Если выключено – работает напрямую с источником
- Обновление кэша:
 - 4 Вручную
 - 4 Репликация
 - 4 Различные стратегии обновления задаются при проектировании



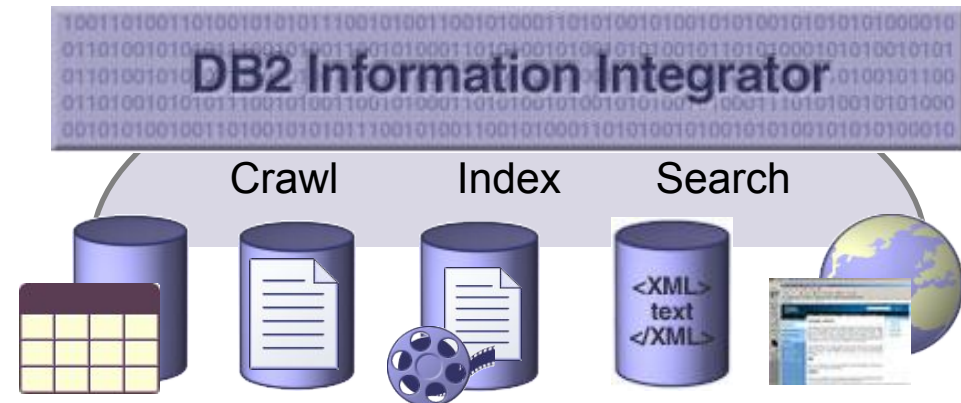
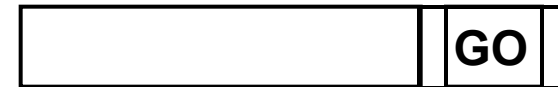
DB2 Information Integrator OmniFind Edition



- Быстрый ответ
 - 4 Сложные алгоритмы для корпоративного контента
 - Intranet, file systems, collaboration systems, content managers, and databases
 - 4 Stemming, lexical analysis, anchor point analysis, dynamic summary, highlighted text

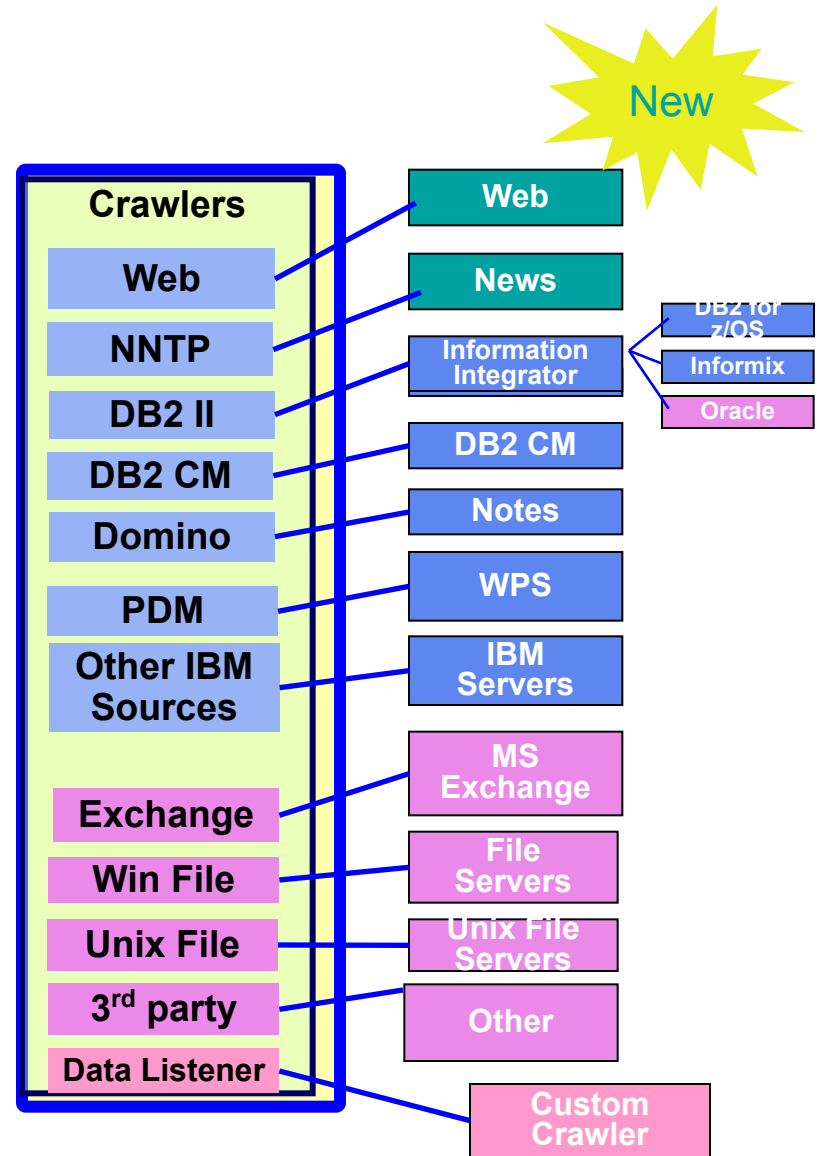
- Легко встраивается в приложения
 - 4 Java APIs
 - 4 Document level security

- Может работать с большими объемами
 - 4 500К документов и более
 - 4 Тысячи одновременных пользователей



OmniFind Crawlers

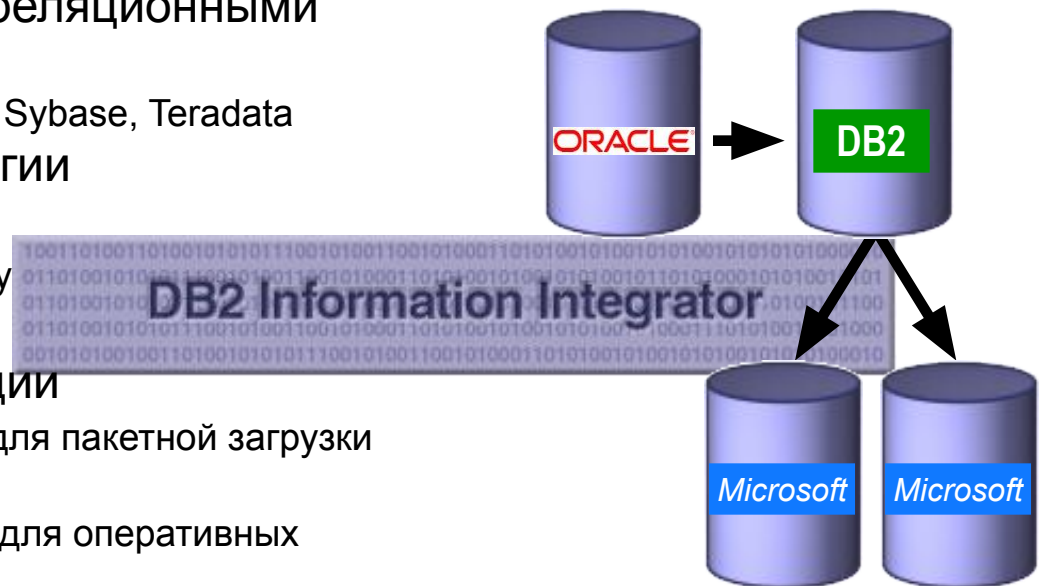
- Web content
 - 4 HTTP/HTTPS
 - 4 News groups (NNTP)
 - 4 WebSphere Portal web content sources
- Collaboration
 - 4 Lotus Notes databases
 - 4 MS Exchange public folders
- File systems
- Content management systems
 - 4 DB2 Content Manager
- Relational Data sources
 - 4 DB2 family (DB2, Informix, DB2 for z/OS)
- DB2 Information Integrator relational data sources (Oracle)



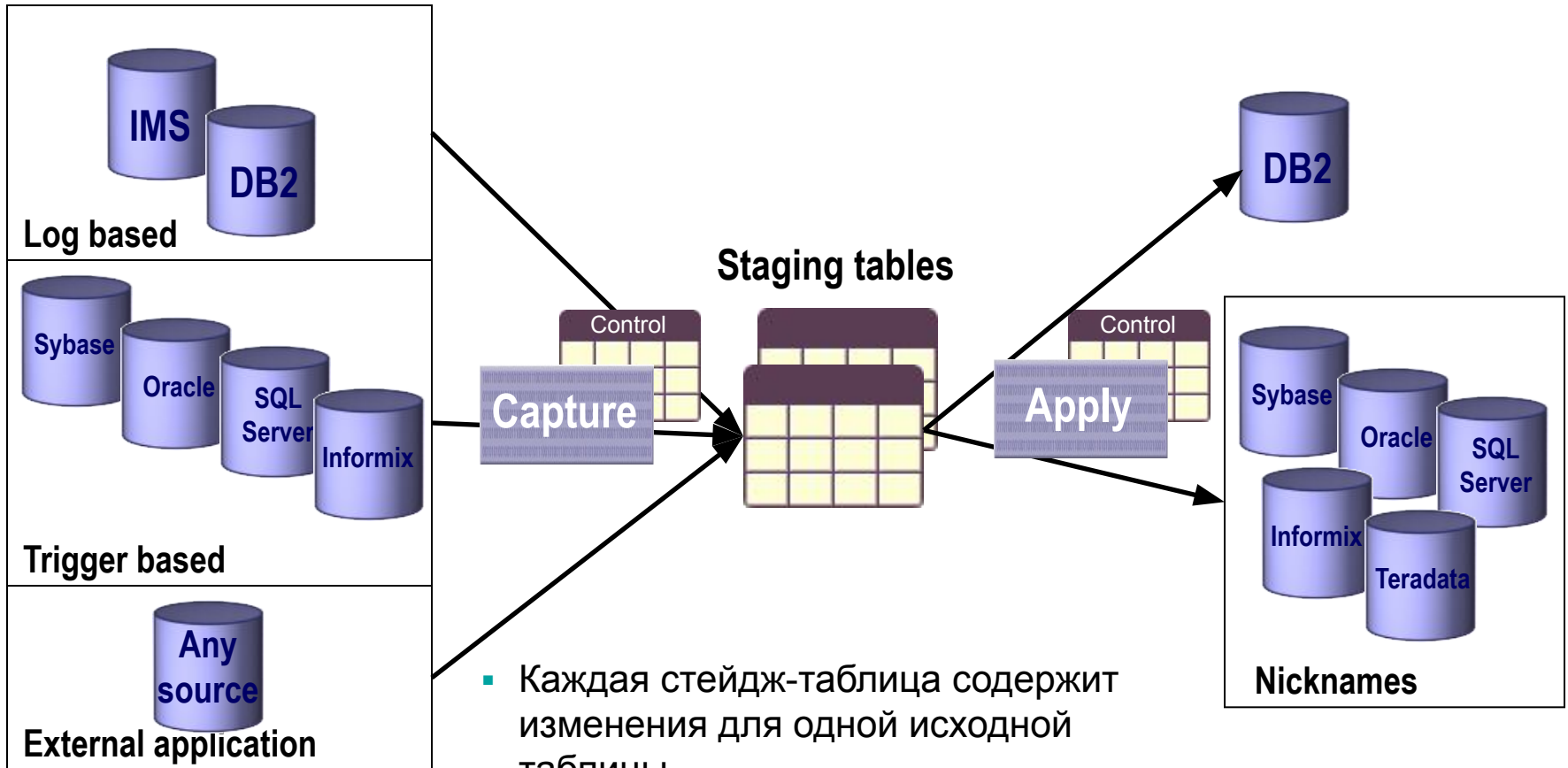
DB2 Information Integrator 8.1

A Replication Server

- Реплицирует данные между реляционными СУБД
 - 4 DB2, Informix, Microsoft, Oracle, Sybase, Teradata
- Поддерживает гибкие топологии
 - 4 Дистрибуция: один ко многим
 - 4 Консолидация: многие к одному
- Различные режимы репликации
 - 4 Table-at-a-time (потабличный) для пакетной загрузки DW в ограниченное время
 - 4 Транзакционно-консистентный для оперативных данных
- Выбор характеристик задержки
 - 4 По расписанию, через интервал, непрерывный
- Применение преобразований на лету
 - 4 Стандартные SQL выражения или выполнение хранимых процедур



Концепции гетерогенной репликации



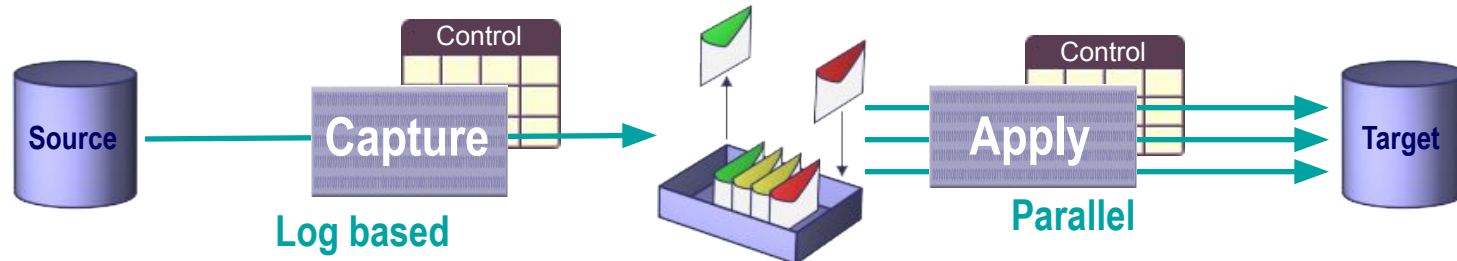
- Каждая стейдж-таблица содержит изменения для одной исходной таблицы
- Формат стейдж-таблицы опубликован для внешнего использования

Расширение архитектуры репликации



New

- Малая задержка, высокая пропускная способность
- Высокая надежность
- Database event publishing



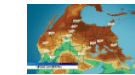
WebSphere MQ

- Каждое сообщение представляет транзакцию
- XML формат сообщения
- Высоко параллельный apply process
- Определение и разрешение конфликтов репликации
- Гетерогенные источники и подписчики
- Wizard для конфигурирования

Доступ к Web Services



```
SELECT l.city, t.degreesFahr
FROM location l, temperature t
WHERE t.city=l.city
```



Temperature

Stock
Quote

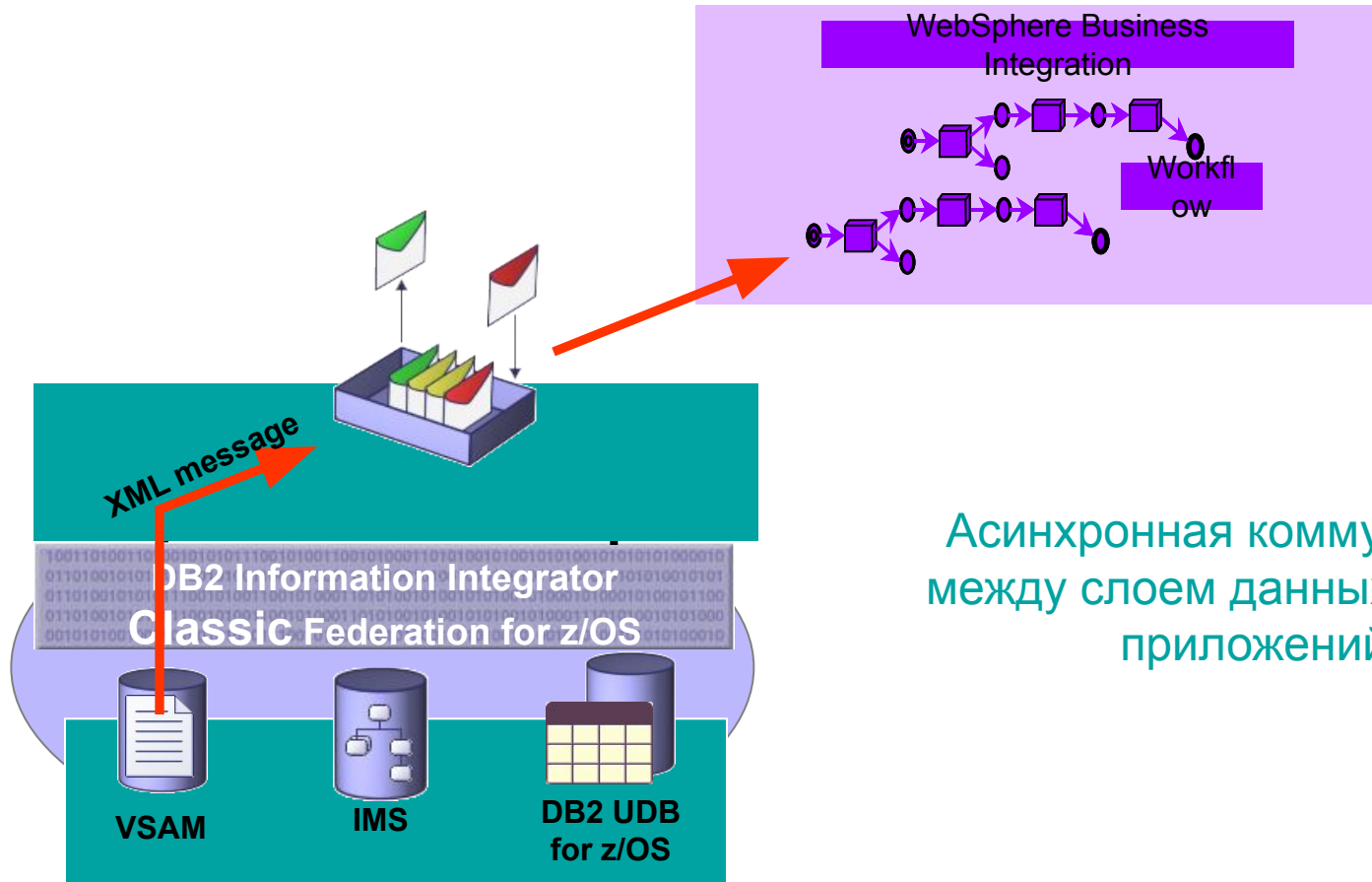
- Интегрирует SQL предложения и вызов Web Service
- Один оператор может работать с локальными и удаленными данными и с web services
- Поддержка генерации SQL функций и табличных UDFs, на основе WSDL web service description (WebSphere Studio)

Использование Web Services

- 4 Любой Web Service становится функцией или источником данных

Публикация событий (или захват изменений в данные)

New



Асинхронная коммуникация между слоем данных и слоем приложений

Связь с пакетными приложениями

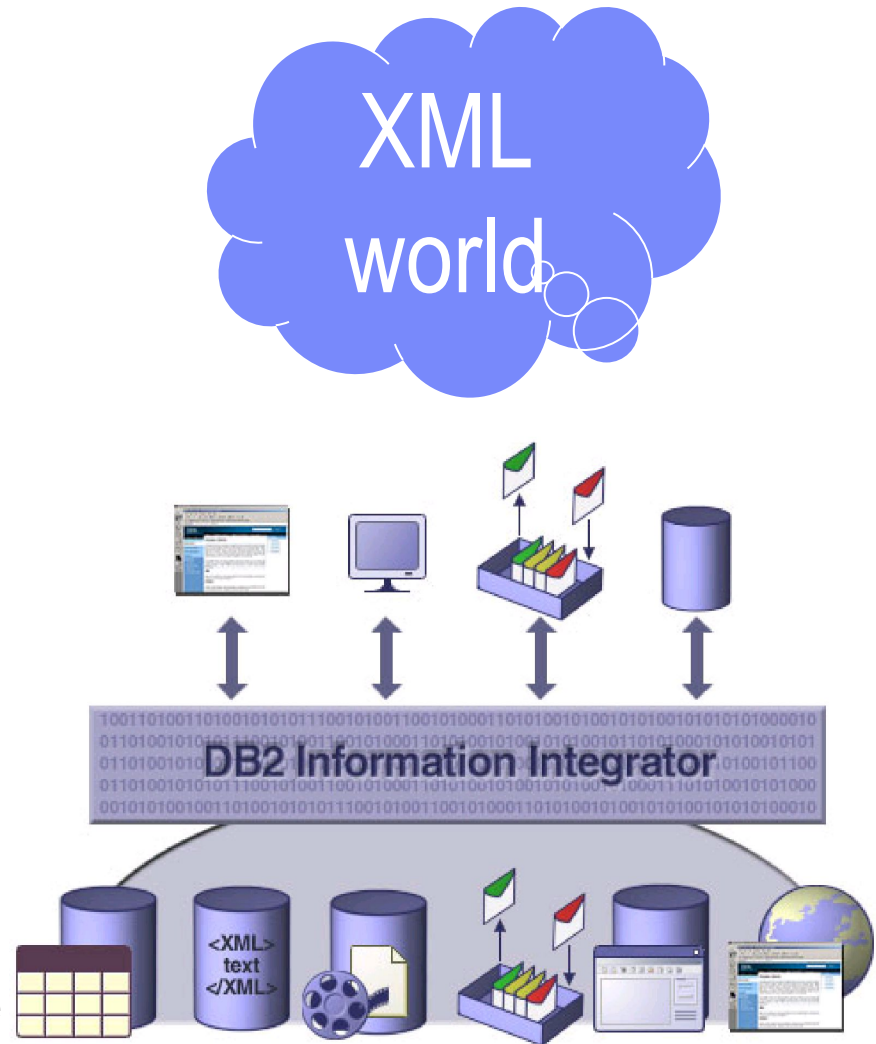


- Генерируются связанные иерархии реляционных схем для бизнес объектов пакетных приложений.
- Запрос через SQL.
- Можно соединиться с другими приложениями и БД.



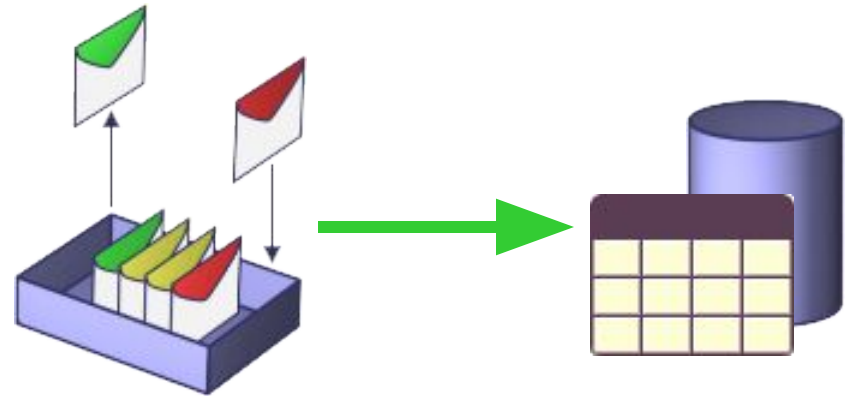
Мост в мир XML

- XML wrapper
 - Доступ к XML документам без их загрузки
 - Соединение XML с другими данными
- XML Extender
 - Объектно-реляционное исполнение
 - Хранить, извлекать, разбирать, собирать XML
 - Оценивать по DTD или Схеме
 - Извлекать, преобразовывать
 - Хранение
 - куском
 - Разбросать по таблицам
- XML расширения в SQL
 - Публиковать федеративные данные как XML документы



Интеграция с WebSphere MQ

- Доступ из приложений к очередям WebSphere MQ с помощью SQL
 - 4 Читать очередь (destructively or non-destructively)
 - 4 Послать сообщение
 - 4 Опубликовать сообщение на определенную тему
 - 4 Зарегистрировать интерес к определенной теме
 - 4 Two-phase commit integrity



```
Insert into PENDING_ORDERS  
select t.msg from table(MQRECEIVEALL());
```

Инструментарий

- Фокус: Сделать DB2 Information Integrator прозрачным для средств разработки
 - 4 DB2 Centers (Control Center, Development Center, etc)
 - Поддержка для ДБА и разработчиков серверных приложений
 - 4 WebSphere Studio (Eclipse Platform)
 - Web разработчики, J2EE разработчики, Data разработчики, XML разработчики и т д
 - 4 DB2 Microsoft Visual Studio
 - Поддержка разработчиков приложений для DB2
 - 4 Business Intelligence Tools
 - Протестировано с. Crystal Decisions, Business Objects, MicroStrategy, Brio, Cognos...

Ценность DB2 Information Integrator

- **Расширяет отдачу от сделанных инвестиций**
 - 4 Работает с существующей инфраструктурой
 - 4 Консолидирует данные и обеспечивает доступ к распределенным данным, включая старые системы, так, как будто они находятся в одном источнике данных
 - 4 Объединяет существующие данные и контент
 - 4 Использует знакомую SQL программную модель и существующие средства разработки
 - 4 Построен на стандартной платформе
- **Уменьшает время разработки сложных приложений**
 - 4 Уменьшает ручное кодирование 40%-65%
 - 4 Уменьшает требования к квалификации разработчиков
 - 4 Наполовину сокращает время разработки