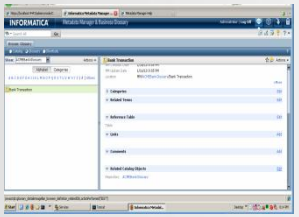
Полезные опции платформы Informatica 9
Новое в версии 9.1

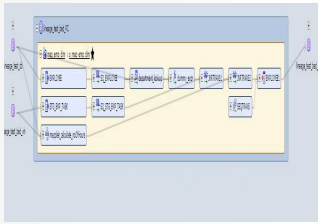
Олег Гиацинтов
Технический Директор
Data Integration Software

Informatica 9.1

Полный цикл интеграции данных

БИЗНЕС-ГЛОССАРИЙ





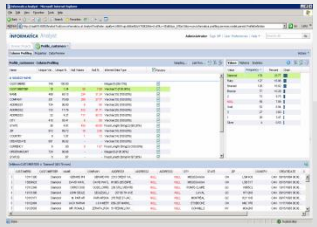
INFORMATICA METADATA MANAGER

Прямой доступ к бизнес-терминам

Анализ зависимостей из любой точки

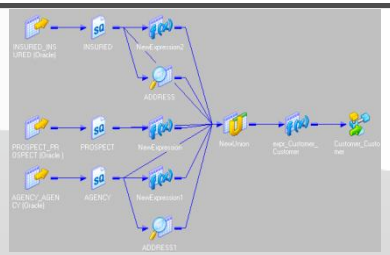
- **Совместная работа бизнеса и ИТ**
- **Ускорение внедрения проекта**

БИЗНЕС-ТРЕБОВАНИЯ



СПЕЦИФИКАЦИЯ

Автоматическая генерация маппингов



МАППИНГ

Mapping Architect for Visio

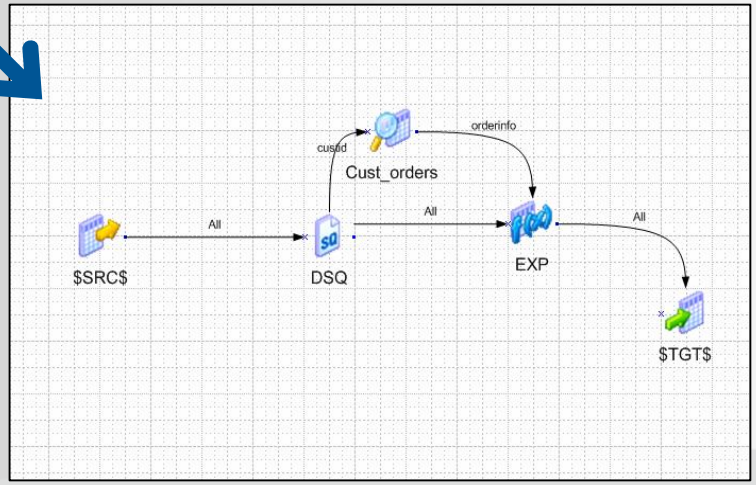
Кросс-функциональная совместная работа

DIS

Архитектор

Разрабатывает шаблон маппинга в Visio, логику процесса

Генерация маппинга



Разработчик

Дорабатывает, настраивает сгенерированный маппинг

- **Поддержка источников PowerExchange** NEW
- **Поддержка большинства типов трансформаций данных** NEW
- **Поддержка ссылок на трансформации** NEW
- **Работа с custom transformations после генерации маппинга** NEW

Data Validation Option

Эффективная валидация данных

DIS

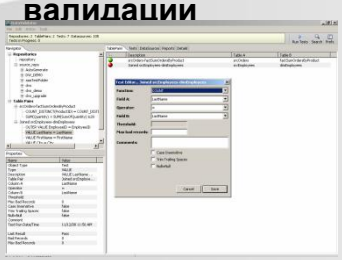
- **Отказ от ручного тестирования**
- **Обеспечение 100% покрытия при тестировании данных – *каждая колонка в каждой строке в каждой таблице***
- **Использование различных определений правил для валидации корректности**
- **Использование простых правил при проведении сравнений без написания SQL-кода**
- **Сравнение обезличенных данных с реальными**
- **Валидация гетерогенных источников и приемников (Oracle, DB2, Teradata и т.д.)**
- **Получение и фиксация результатов на любом этапе**
- **Использование возможностей PowerCenter (масштабируемость, платформы, доступ к данным)**

Informatica Data Validation Option

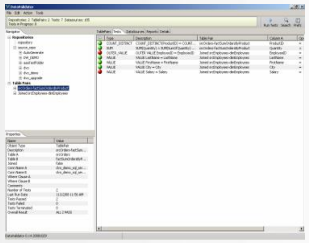
Процесс валидации данных

DIS

1 DataValidator GUI - определение правил валидации

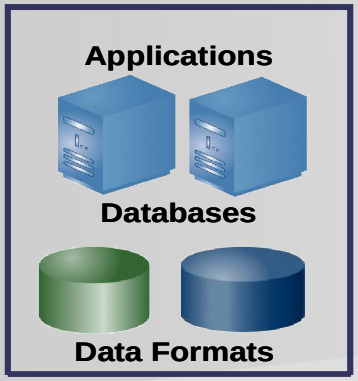


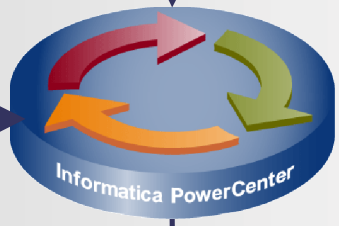
3 DataValidator GUI – показ результатов



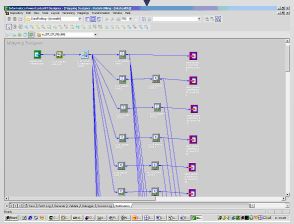
5 Отчетность по результатам

PC API





Приемник данных



2 Генерация и выполнение маппингов и сессий PowerCenter

4 Результаты сохраняются в приемнике данных

Informatica Data Quality 9.1

Новые возможности

DIS

- **Интеграция с MS Excel**
  - **Чтение и запись данных напрямую из/в Excel**
  - **Для пользователей Excel не требуется установка клиентской части Informatica**
  - **Обработка данных ведется через web-сервисы**
- **Различные источники для референсных таблиц**
  - **При создание референсной таблицы данные импортируются непосредственно из источника**
  - **Референсная таблица может быть внешней без импорта данных - используется только ссылка на таблицу**
- **Возможность использовать reusable-логику и создавать shortcuts**

Informatica Data Quality 9.1

Proactive Data Quality Monitoring option

DIS

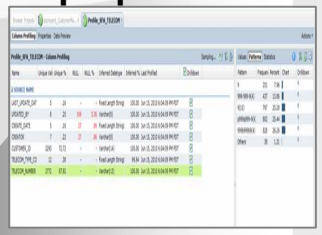
- **Опция для решения Informatica Data Quality**
- **Оповещает пользователей в режиме реального времени при превышение порога качества обрабатываемых данных**
- **Удобство – аналитики и разработчики легко могут менять существующие правила и создавать новые**
- **Уровень правил – от простых до самых сложных**
- **Внедряется в первый же день проекта – в состав входит набор правил для начала работы**

Proactive Data Quality Monitoring

Процесс выявления ошибок

DIS

Профиль



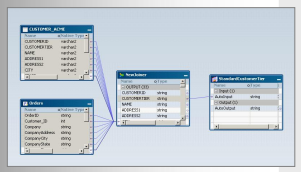
Source

Source

Source

Data Quality

Маппинг



Proactive Monitoring

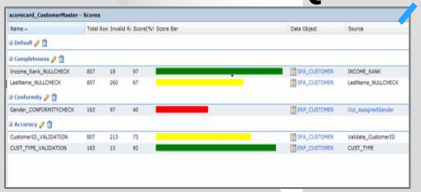
Complex Event Processing

Error

Специалисты от бизнеса и ИТ

Автоматическое определение ошибок в данных на основе правил

Статистика из IDQ



Выявление проблем качества:

- *Пустые, невалидные значения*

Стандартизация, Очистка, Проверка данных:

- *Выявление дубликатов, поиск индексов, проверка адресов*

Оценка метрик качества:

- *Полнота, точность правил и т.п.*

Informatica Identity Resolution

Эффективная идентификация данных

DIS

- **Процесс идентификации данных может приводить к серьезным ошибкам и вариациям**
- **Информация может содержать многоязычные данные**
- **Используемые средства идентификации не находят дубликатов либо находят их слишком много**
- **Используемые техники идентификации чересчур сложны и ресурсоемки**
- **Очистка данных не позволяет полностью решить проблему**

Примеры

Сбербанк, Сбербанк РФ, СБ РФ, СБРФ, SBRF, Sberbank

Павел Лихницкий,

Лихницкий П.,

П. В. Лихнитский,

г-н Лихницкий Павел Викторович,

Mr. Likhnitsky,

P. Likhnitskiy…

Алексей, Алекс, Алеша, Алёша, Леха…

Informatica Identity Resolution

Сравнение данных любого типа, формата и качества

Код	Описание	Размер	Цена
AP-2199	Sailors' Desk Lamp	12 in	27.99
AP2199	Nautical Lamp	12 inch	27.99
PA-2119	Sailors Lamp	12 дюймов	34.99

- Адреса, имена, продукты
- Неверные и потенциально некорректные данные могут быть использованы для сравнения и поиска
- Аббревиатуры
- Надписи на других языках

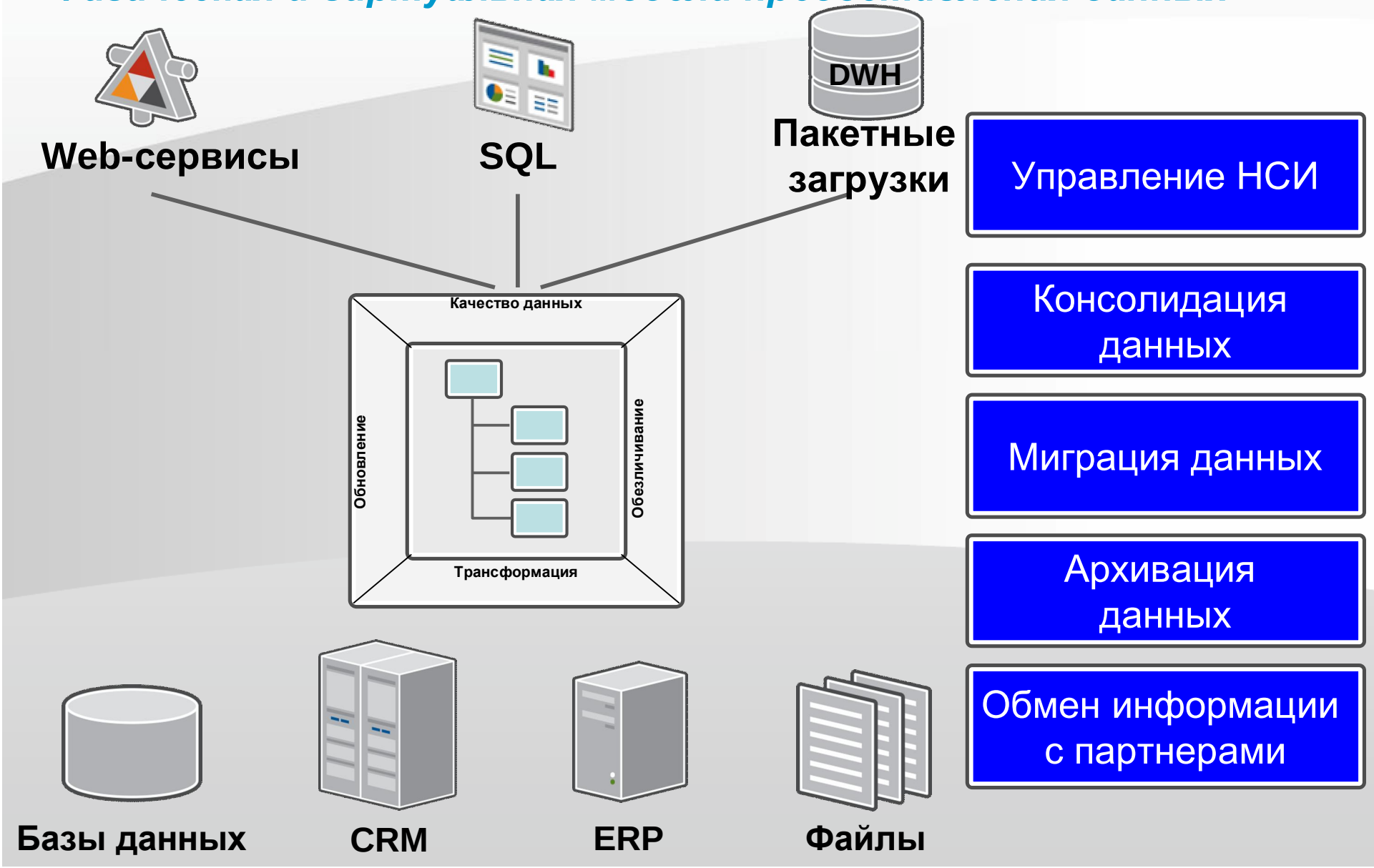
Имя	Дата	Адрес	Город	Индекс
Карманов Сергей	03.01.1967	ул. Советская, д. 16, кор. 4, кв. 32	Ростов-на-Дону	352658
Карманов С.А.	03.01.1967	Советская, 16-4-32	Ростов	352 658
Серега Карманов	9/9/99	Донской район, Советская, 16/3	Р-н-Д	
Карманов Сергей Александрович	13/01/1967	Ростов-Дон, Советская, 16, кв. 32	Ростов-Дон	352-658
Sergey Karmanoff		16-3, app. 32 Sovetskaya st.	Rostov-on-Don	352658

Сравнение с применением нечеткой логики дает высокие результаты по выявлению дубликатов

Informatica Data Services

DIS

Физическая и виртуальная модели предоставления данных



Data Services для BI и сложных приложений

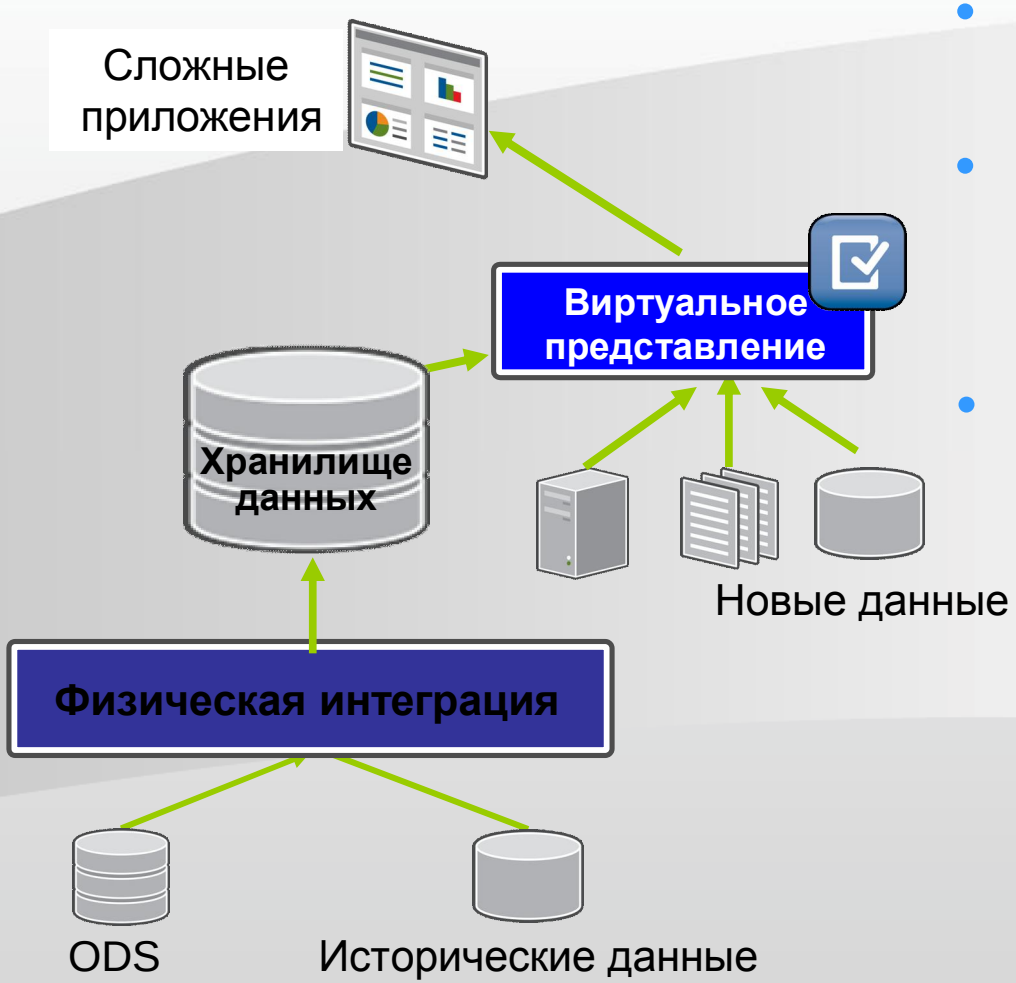
Немедленное подключение новых источников

- Для BI : Возможно быстрое расширение хранилища с появлением новых данных
- Для сложных приложений: Возможно немедленное использование новых данных в приложениях
- Informatica Data Services 9:
  - Немедленная виртуальная интеграция новых данных с существующим DWH и бесшовная конвертация на физический интеграционный уровень
  - Профилирование данных
  - Применение правил обеспечения качества к виртуально интегрированным данным

Data Services для BI и сложных приложений

DIS

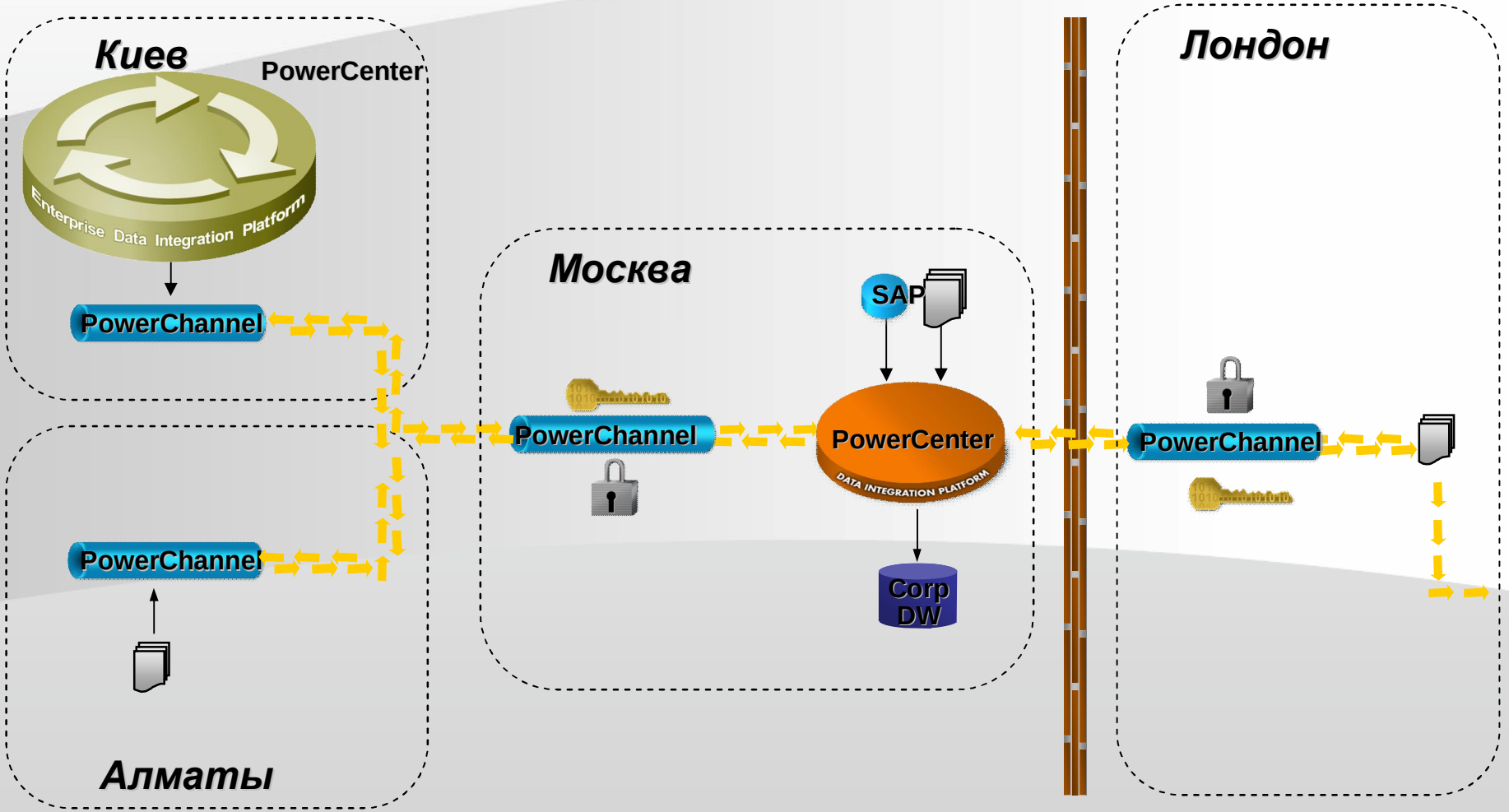
Получение данных в реальном времени



- Подключение “свежих” данных в виртуальное представление
- Извлечение немедленной выгоды из виртуально интегрированных данных
- Informatica Data Services 9:
  - Поддержка всех форматов данных
  - Немедленная интеграция временных или быстроменяющихся данных без задержек физической консолидации
  - Применения правил обеспечения качества к виртуально интегрированным данным даже в случае изменения источников

Informatica PowerChannel

Доступ к данным по низкокачественным каналам связи



Informatica PowerChannel

DIS

Высокая производительность и параллелизация обработки данных

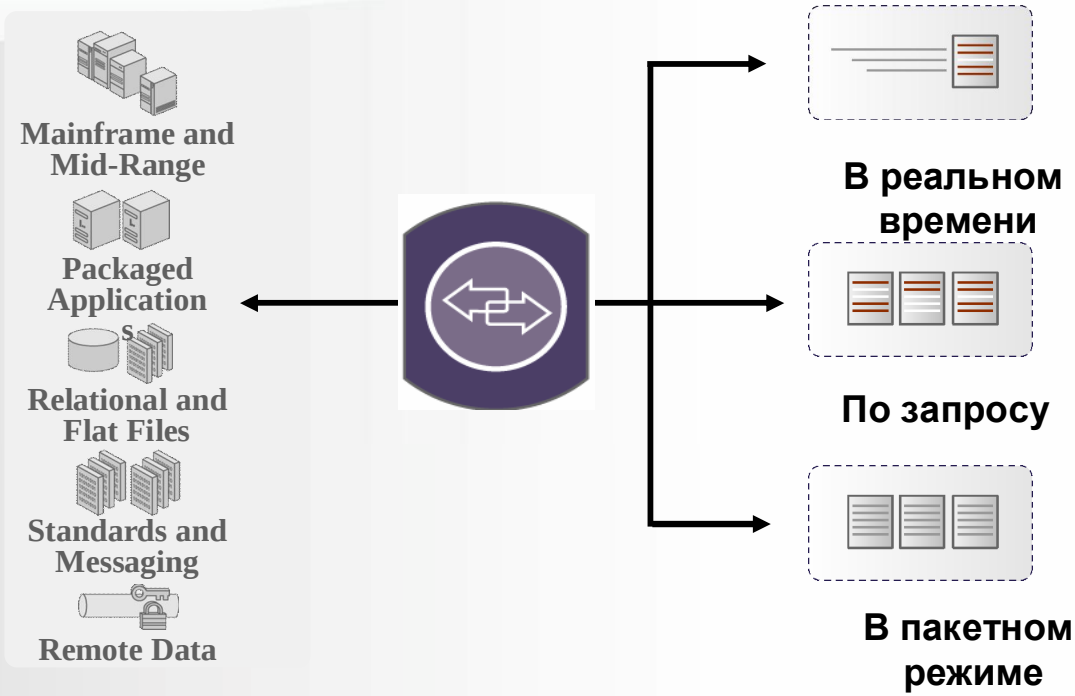
- **Прямой доступ к базам данных и плоским файлам, расположенным вне компании**
- **Высокая степень сжатия данных**
  - До 9 раз снижается объем передаваемых данных
- **Высокая скорость передачи данных**
- **Автоматическое восстановление передачи данных после сбоя в канале связи**
- **Легкое внедрение**
  - Встроенная поддержка PowerCenter

PowerExchange CDC

Доступ к изменившимся данным

DIS

- **Чтение изменившихся данных из логов баз данных в режиме реального времени или по запросу**
- **Любая последующая обработка данных через PowerCenter**
- **Настраивается на требуемые типы записей, данных, виды операций**



- **Доступ ко всем основным СУБД, midrange и мейнфреймам**
- **Визуальный интерфейс для управления и администрирования**
- **Новое в версии 9.1 – высокоскоростная работа с СУБД Oracle для больших и очень больших объемов данных**

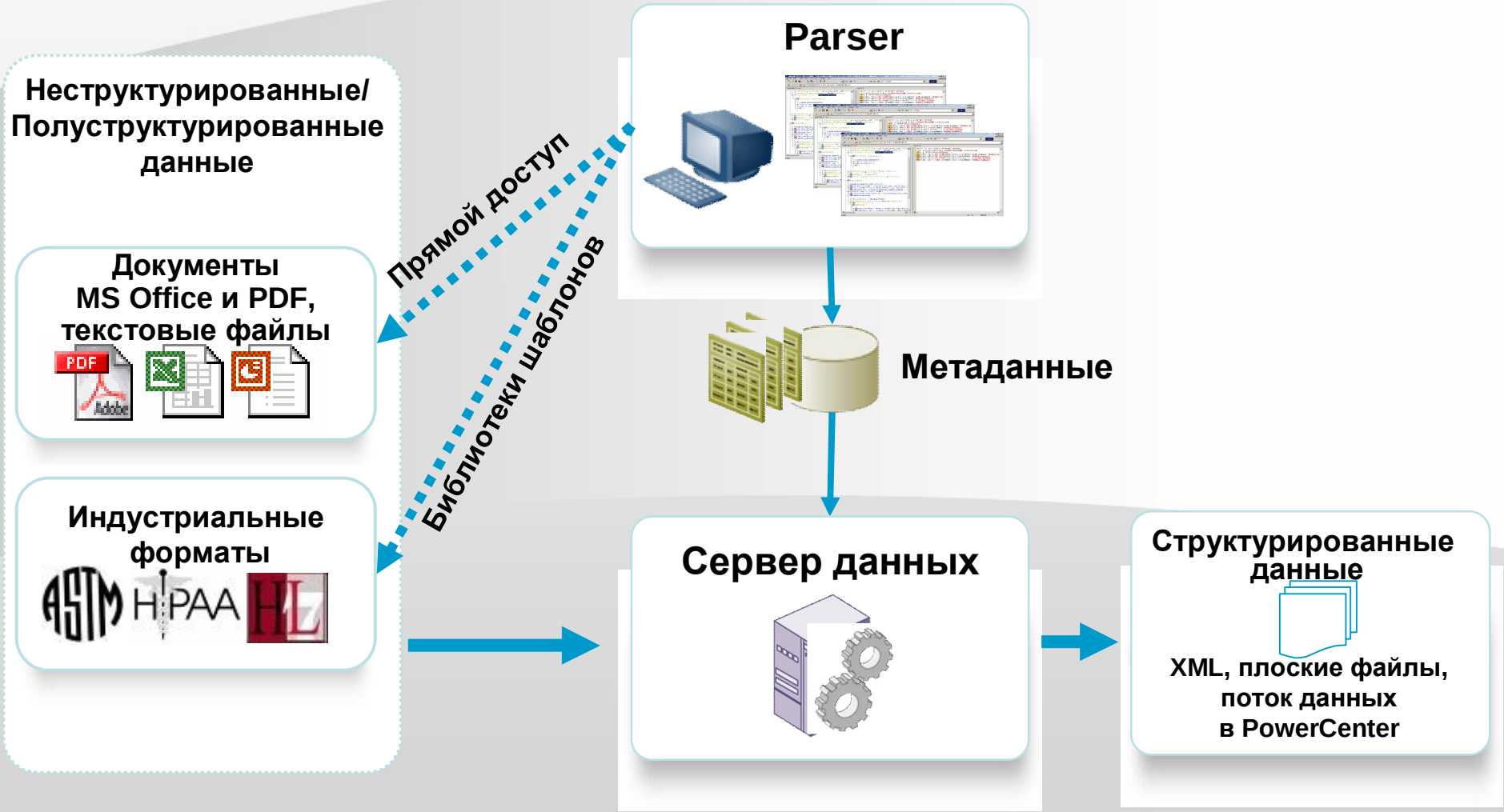
Informatica 9.1

PowerExchange Express CDC for Oracle

- **Новый продукт в версии 9.1**
- **Прямое чтение изменившихся данных из логов Oracle**
- **Не используется Oracle LogMiner для извлечения данных**
- **Производительность выше в 7-10 раз, чем с использованием Oracle LogMiner**
- **Идентичное управление, как и в других продуктах PowerExchange CDC**

Informatica B2B

Доступ к неструктурированным и полуструктурированным данным



Informatica B2B

Обмен информацией с партнерами

- Получение и разбор поступающих файлов от партнеров в их форматах в онлайн-режиме
- Обработка поступающей сложной текстовой информации
- Формирование и отправка данных партнерам в виде файлов их форматов с гарантией доставки
- Справочники (продукты, клиенты и т.д.) и транзакционные данные
- Высокая легкость внесения изменений в процессы при изменении форматов поступающих или отправляемых файлов

Informatica Ultra Messaging (29 West)

Уникальная технология обмена сообщениями

DIS

Modern, No-Limits Design

For the next 20 years
- Ultra low latency
- Efficient
- Unified

Reliable & Powerful

Bet your business
- Bullet proof (plus no storms!)
- Unmatched throughput
- Flexible

Reduce Costs

Do more with less
- Cut / eliminate hardware
- Messaging as a service
- Enterprise standard

JPMorgan · Credit Suisse · RBS (The Royal Bank of Scotland) · Deutsche Bank · HSBC · Lloyds

Morgan Stanley · Goldman Sachs · Barclays Capital · ISE International Securities Exchange · ICAP · Liffe NYSE Euronext · ICE

UBS · Chicago Stock Exchange · DirectEdge · tradefair · CURRENEX

Ultra Messaging

Ключевые возможности

DIS

Ultra Messaging Streaming Edition

- Ультранизкая задержка (~5-10 микросекунд)
- Высочайшая производительность (>7 млн. сообщений в секунду на реальных проектах)
- Различные транспорты/протоколы
- Унифицированный API

Ultra Messaging Persistence Edition

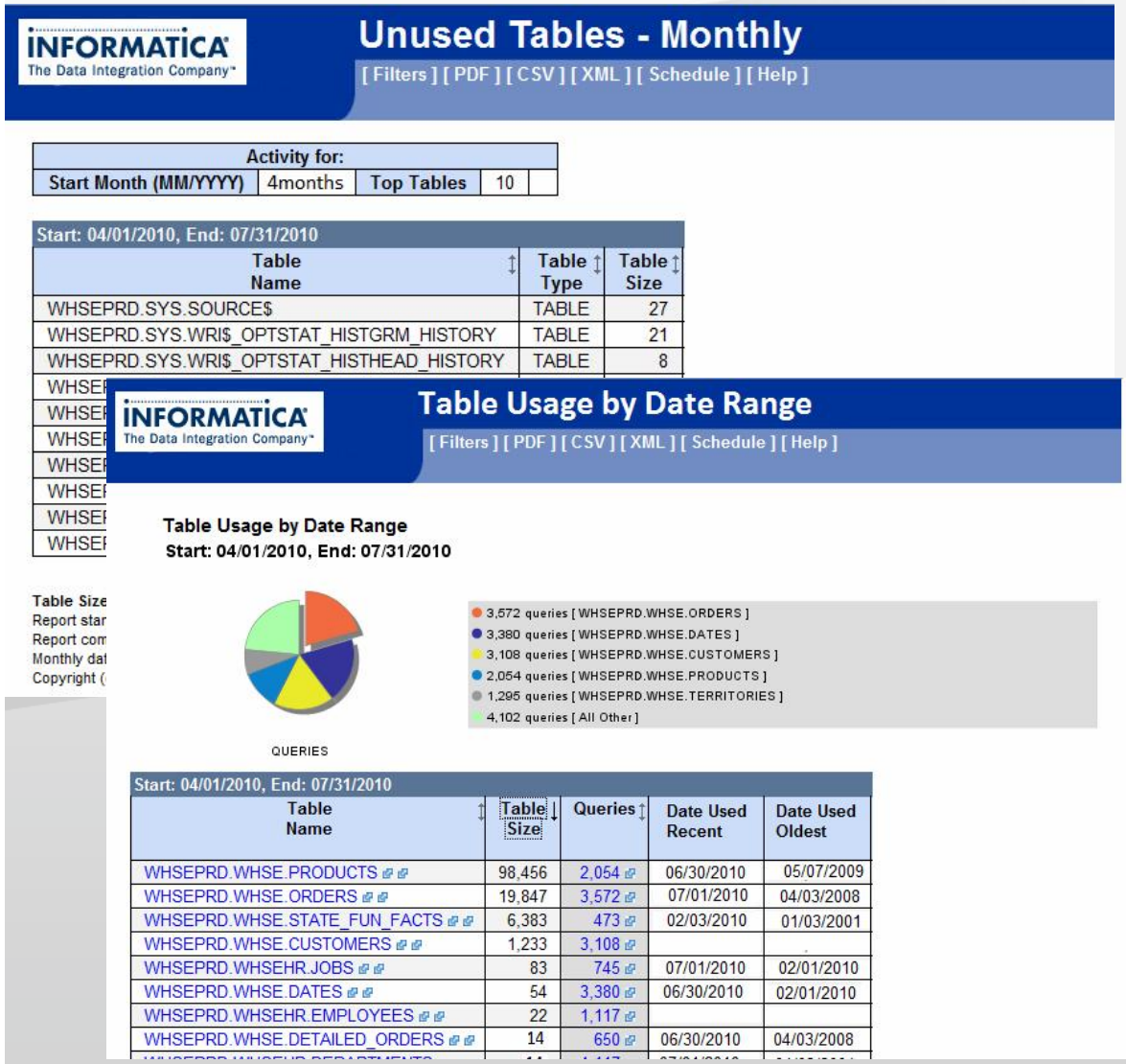
- Гарантированная доставка
- Нулевая задержка при обработке отказов

Ultra Messaging Queuing Edition

- Обработка очередей сообщений
- Доставка только один раз
- Отправка наиболее подходящему получателю
- Балансировка нагрузки - отправка наиболее доступному получателю
- JMS API’s

Informatica Dormant Data Analyzer

Определяет малоиспользуемые таблицы и записи для архивации



Возможности

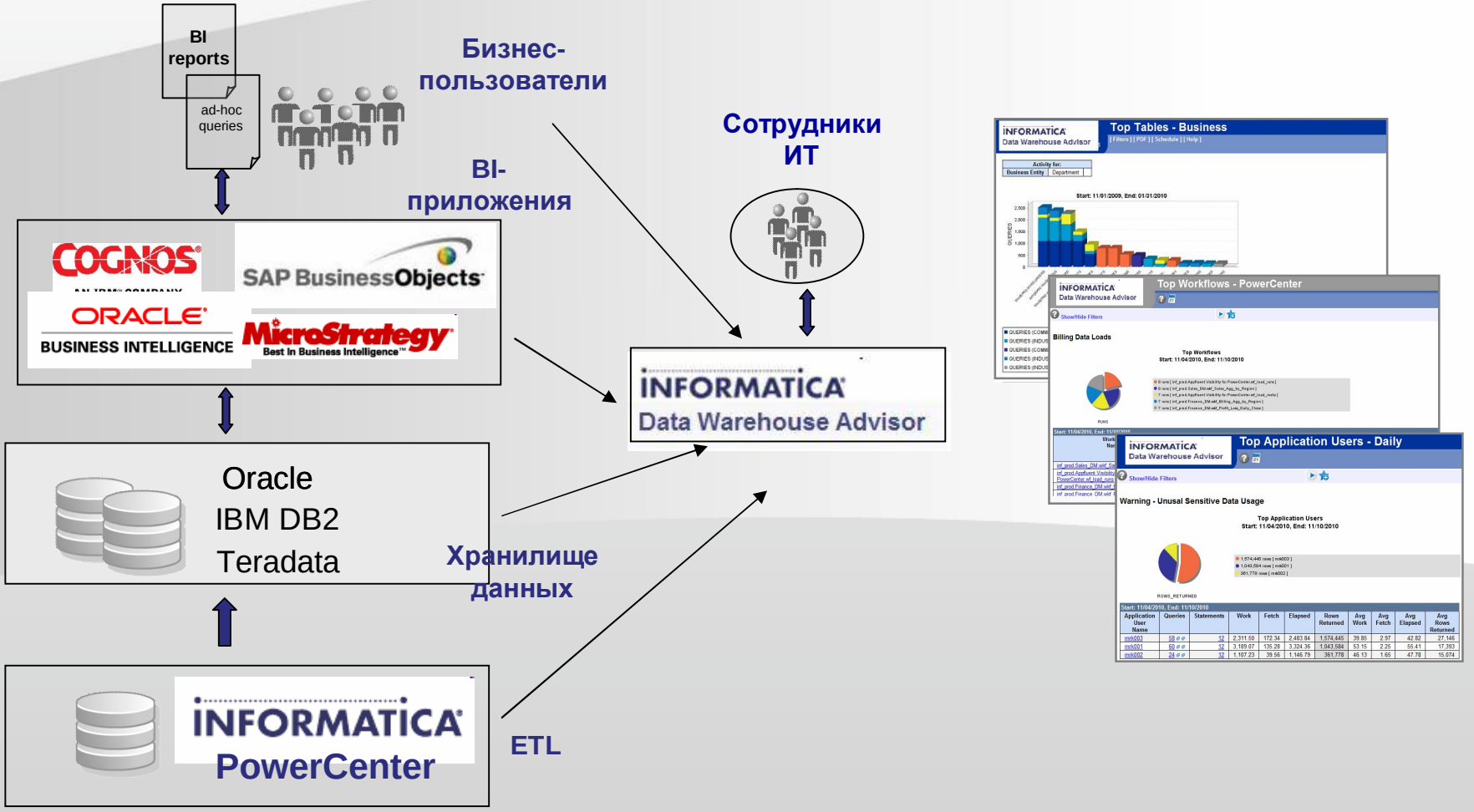
- Отслеживает и отображает неиспользуемые таблицы и записи для последующей архивации средствами Informatica Data Archive (ILM)
- Определяет, кто использует данные и насколько часто

Преимущества

- Снижение размера хранилища данных, а также стоимости используемого ПО и серверов
- Уменьшает время выполнения бэкапов, апгрейдов и т.п.
- Повышает производительность хранилища данных

Informatica Data Warehouse Advisor

Визуализация бизнес-активности и использования данных для оптимизации DWH



Informatica Data Warehouse Advisor

Возможности и Преимущества

DIS

Возможности

- **Метрики используемости:**
  - Понимание, как данные используются бизнесом, какие наиболее востребованы
- **Метрики производительности:**
  - Понимание «узких» мест в производительности
  - Проактивное улучшение производительности процессов загрузки PowerCenter и хранилища данных
- **Метрики аудита:**
  - Отслеживание пользовательской активности по конфиденциальным данным, предоставление отчетов для аудита

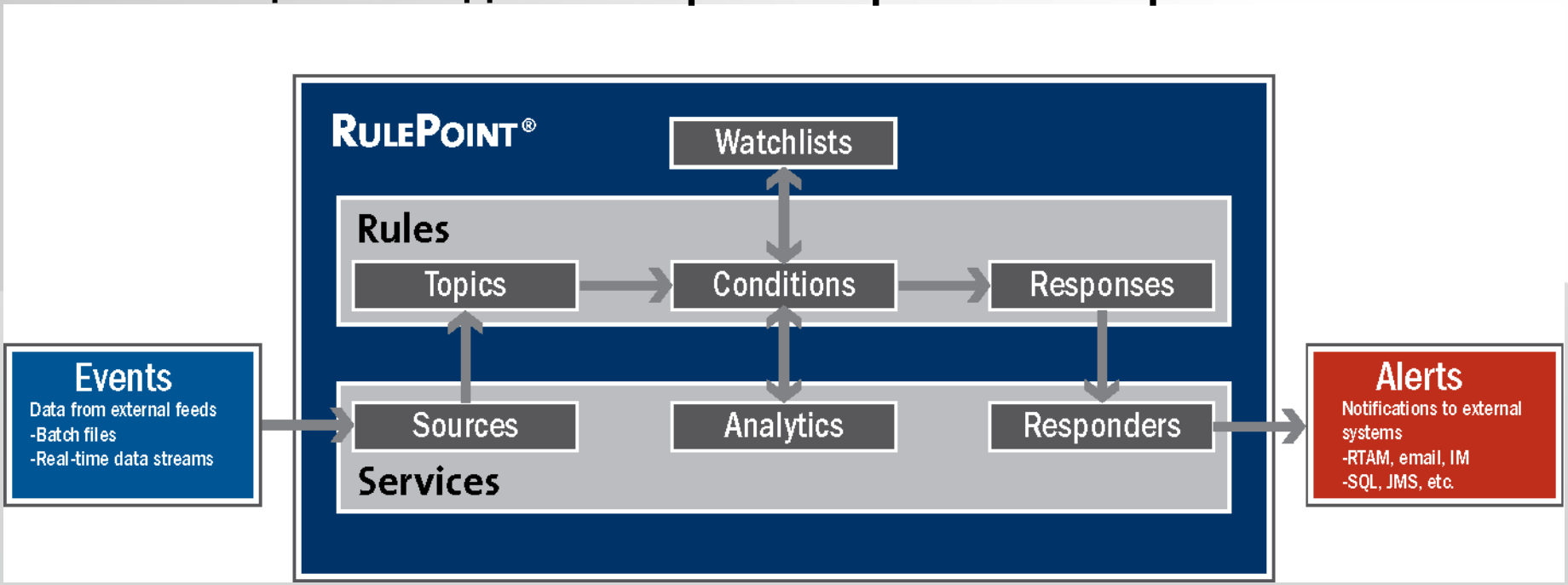
Преимущества

- **Контроль затрат** – идентификация неиспользуемых, малоиспользуемых данных
- **Максимизация приносимой данными ценности –** измерение бизнес-активности
- **Увеличение эффективности –** оптимизация процессов PowerCenter и хранилища данных
- **Обеспечение правомочного использования данных –** отслеживание пользовательской активности

Informatica Complex Event Processing

Концепция работы RulePoint

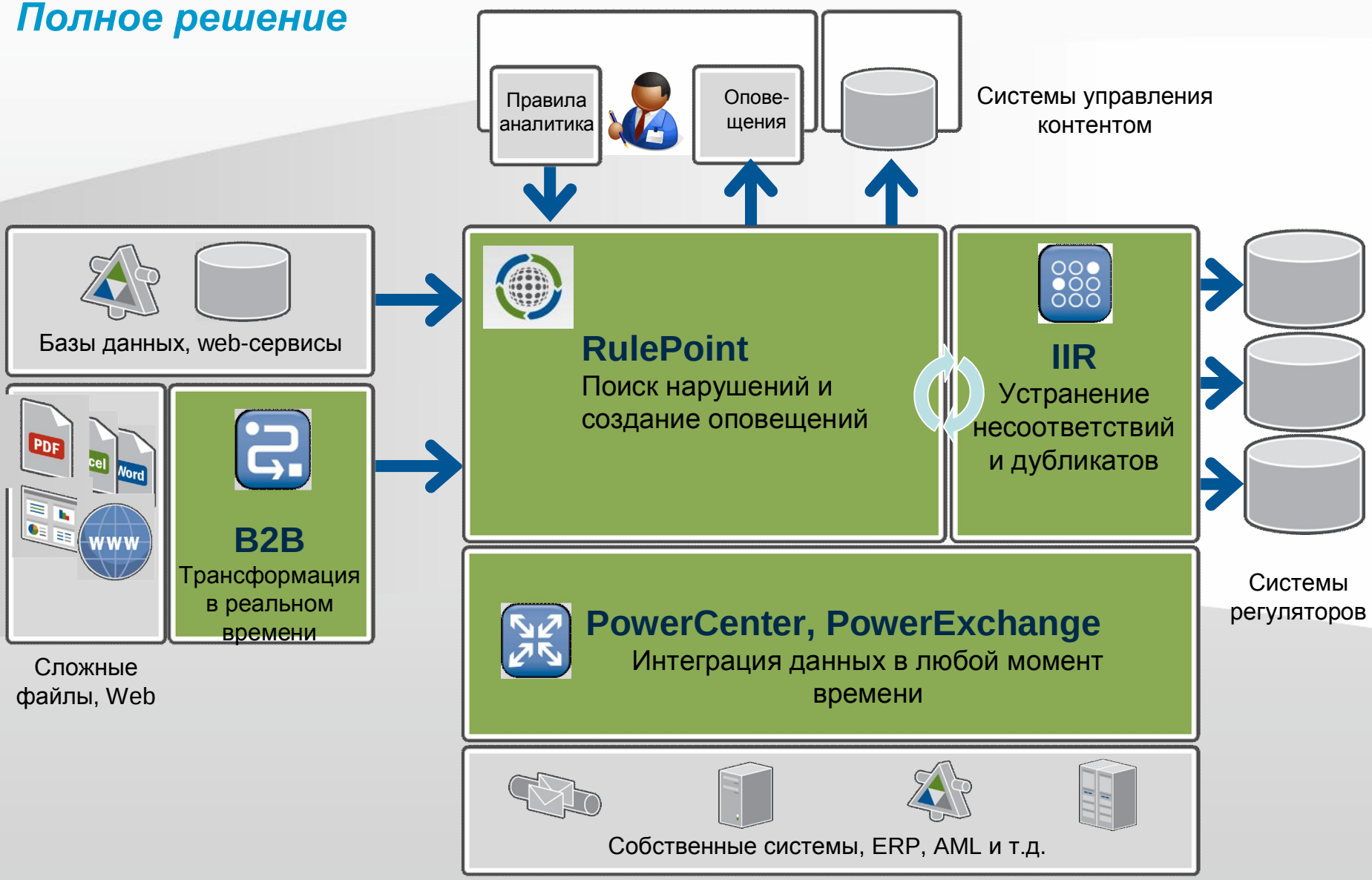
- **Отслеживание любых ситуаций и формирование оповещений**
- **Правила создаются аналитиками в визуальном графическом интерфейсе**
- **Оповещения выдаются в режиме реального времени**



Informatica Complex Event Processing

Полное решение

DIS



DIS
Спасибо!
Ваши вопросы?
Олег Гиацинтов
Data Integration Software