Эффективная работа с данными на платформе Informatica

Хасанов Василий

Data Integration Software

INFORMATICA®

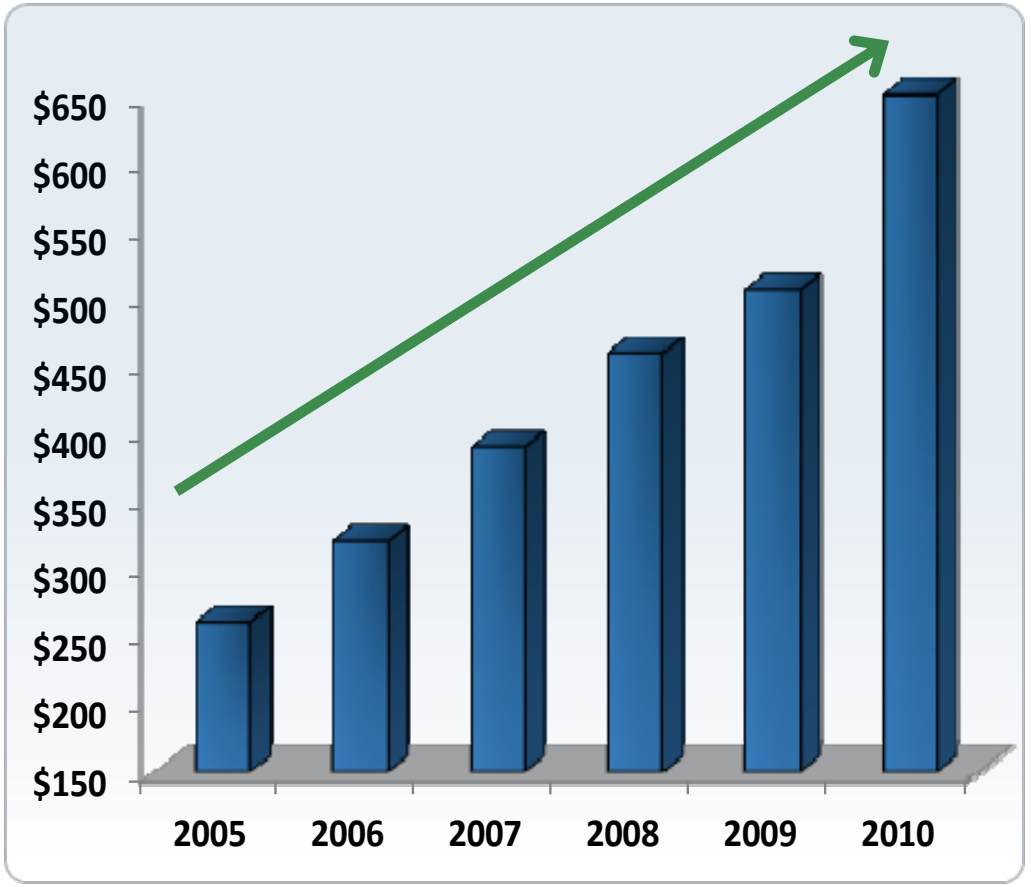
О компании Informatica

#1 Независимый лидер: интеграция данных, качество данных, управление мастер-данными (MDM)

- **Основана** 1993 год
- **Выручка 2010** $600+ млн.
- **Среднегодовой рост за последние 5 лет** 20% per year
- **Заказчики** 4,440+
  - 84 of Fortune 100
  - 87%+ of Dow Jones
  - Правительства 20 стран
- **Сотрудники** 2,340+

Авторизованный дистрибьютор в России и странах СНГ

Data Integration Software



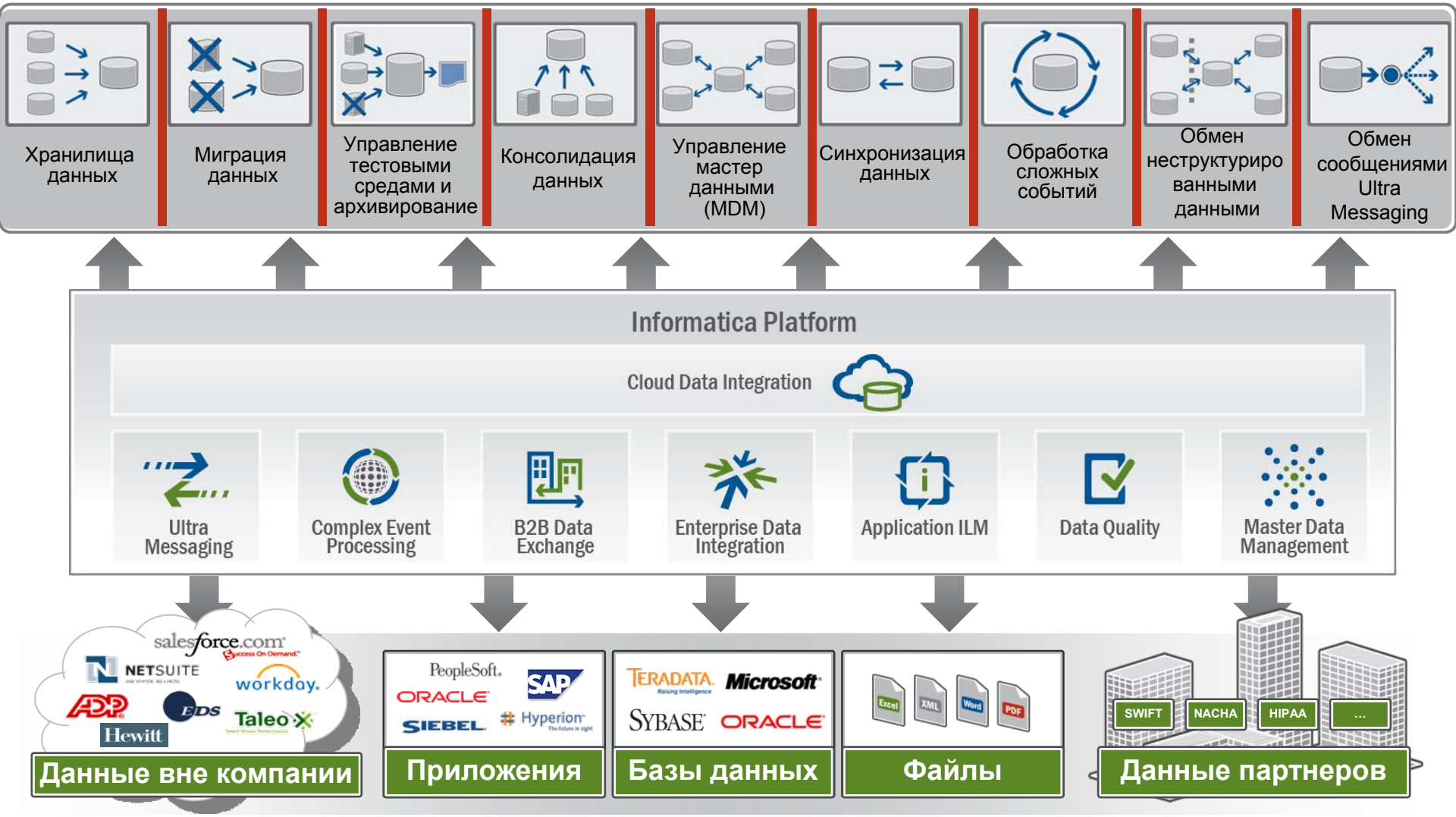
О компании Data Integration Software

- **Data Integration Software (www.data-integration.ru) – авторизованный дистрибьютор корпорации Informatica в России и странах СНГ.**
- **Data Integration Software - мультивендорная компания, с 2005 года предлагающая заказчикам и партнерам в странах СНГ лучшие продукты и решения от крупнейших мировых вендоров.**
  - **Informatica** – мировой лидер в области интеграции и обеспечения качества данных, позволяющий компаниям повысить свою конкурентоспособностью и эффективность бизнеса благодаря надежным, достоверным и своевременным данным.
  - **NICE Systems** -ведущий поставщик решений для управления качеством обслуживания, аналитики на базе речевых технологий и обеспечения физической и информационной безопасности, а также борьбы с мошенничеством в финансовой сфере и отмыванием денег (AML).
- **Заказчиками Data Integration Software являются крупнейшие компании в различных отраслях экономики и госсекторе.**
- **Головной офис компании находится в России, г. Москва.**

Крупнейшие компании СНГ выбирают Informatica

Подход Informatica

Всеобъемлющая, Единая, Открытая, Экономичная платформа



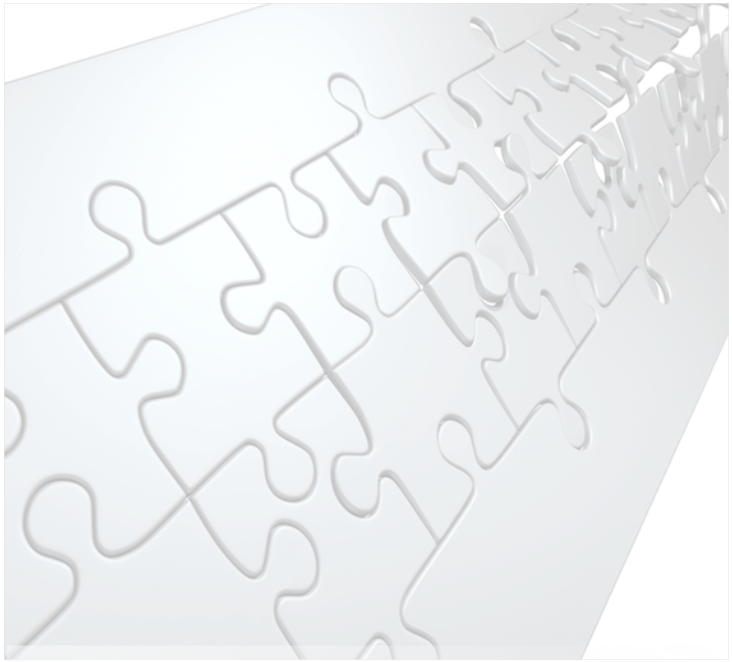
Интеграция данных

INFORMATICA®

Informatica PowerCenter

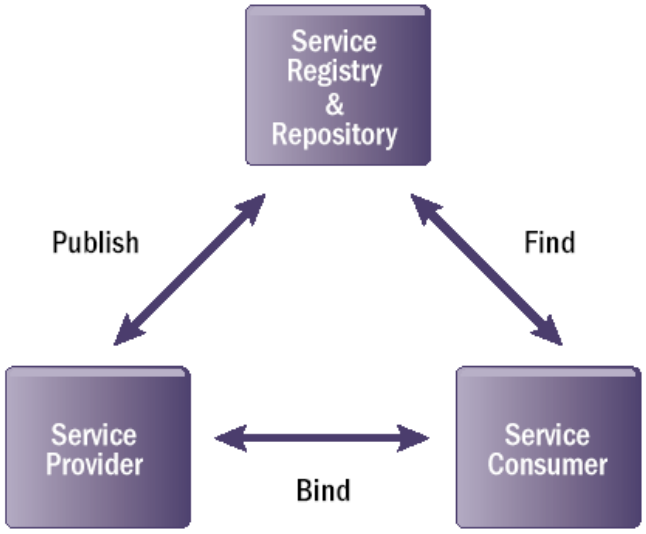
Базовый компонент платформы - интеграция данных

- **Подключения к самым различным источникам и приемникам**
- **Трансформации любой сложности, включающие алгоритмы обеспечения качества**
- **Масштабирование, производительность, отказоустойчивость в промышленных масштабах**
- **Работа в любых режимах, в т.ч. в реальном времени, режиме web-сервиса, пакетном режиме**
- **Управление метаданными и наглядность метаданных**
- **Методология ведения интеграционных проектов**
- **Простота разработки и поддержки**



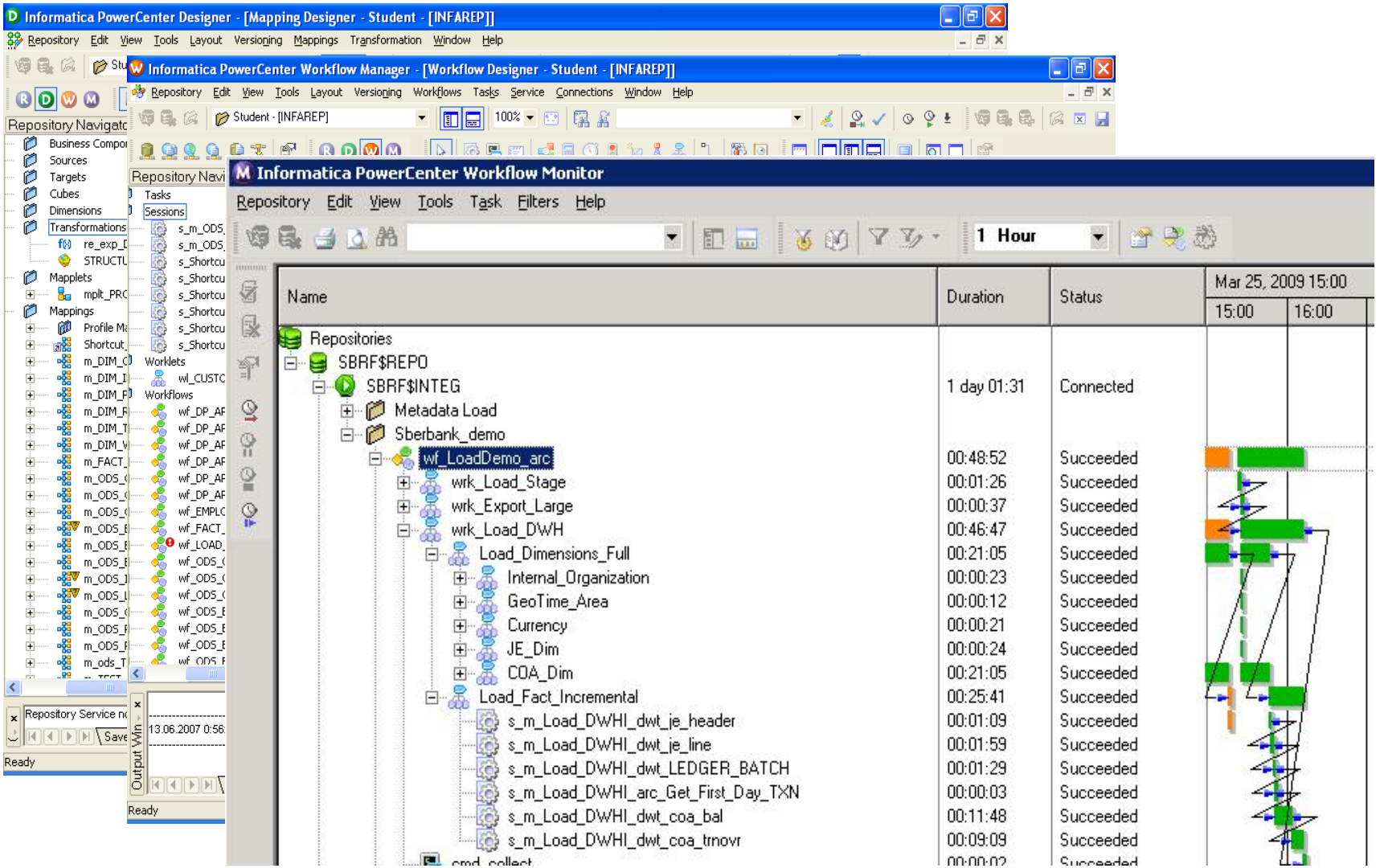
Современная промышленная платформа интеграции

Informatica PowerCenter Real-Time Edition



- Поставщик Web-сервисов (Provider): любой процесс может быть представлен как web-сервис
- Потребитель Web-сервисов (Consumer): возможность взаимодействия с внешними Web-сервисами
- Работа в реальном времени с шинами данных и службами сообщений (IBM MQSeries, TIBCO, Microsoft MQ, webMethods, JMS)
- Работа в реальном времени с изменившимися данными CDC

Эффективная визуальная среда разработки и мониторинга интеграционных процессов



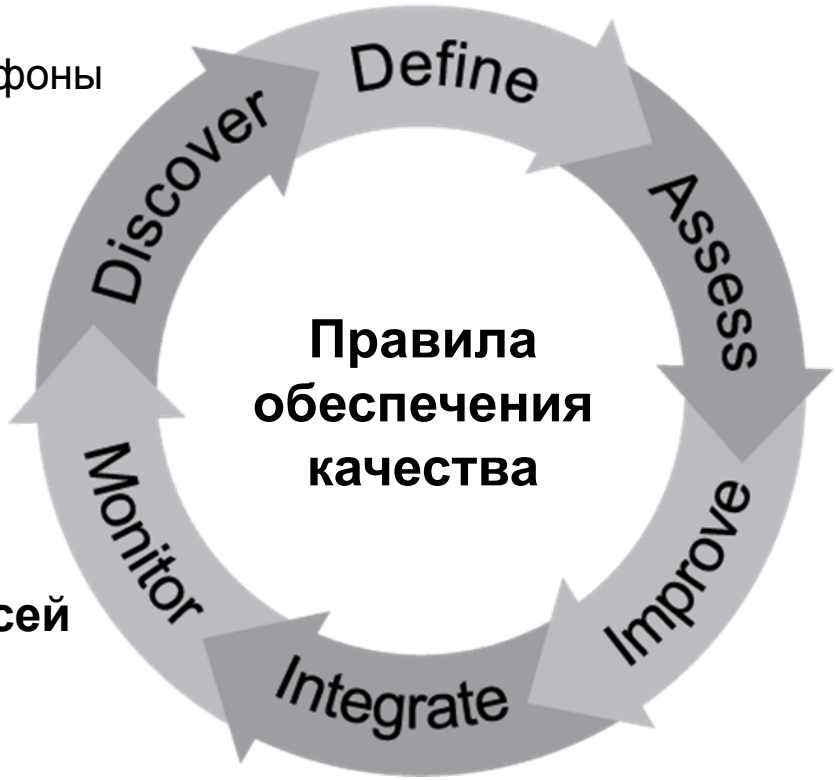
Качество данных

INFORMATICA®

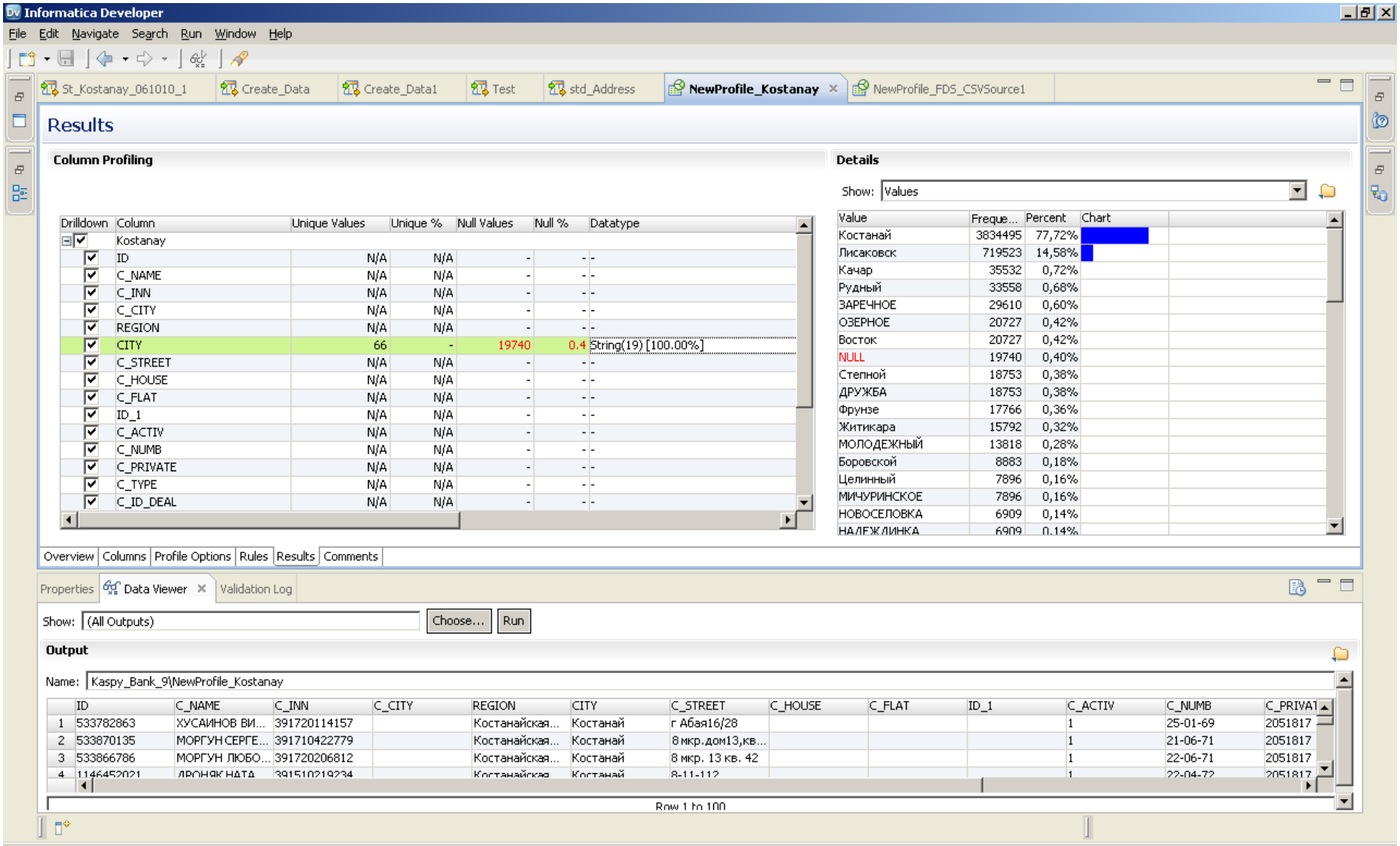
Informatica Data Quality

В пакетном режиме и реальном времени

- **Обработка данных, требующих синтаксического разбора:**
  - ФИО, названия компаний, Адреса,Телефоны
  - Товарные номенклатуры
  - Паспортные данные,Налоговые номера
  - Банковские реквизиты
  - Любая другая информация
- **Стандартизация и проверка данных (в т.ч. по эталонным словарям)**
- **Очистка данных**
- **Сравнение и поиск дубликатов записей вероятностными методами**
- **Профилирование данных**
- **Мониторинг качества данных, в том числе проактивный (оповещение о событиях)**
- **Совместная работа бизнеса и IT**
- **Готовые процессы и словари для стран СНГ**



Эффективная визуальная среда разработки и мониторинга процессов Data Quality



Идентификация и выявление дублей

INFORMATICA®

Проблемы идентификации

ВАСИЛИЙ ХАСАНОВ
В.М.Хасанов
VASILY KHASANOV
MR.V.M.HASANOV

УИЛЬЯМ ВОК
WILLIAM KWOK
W. KWOK KI HOH
MR. BILLY H KWOK

АБДУЛЛА АЛЬ МУСА
ABDULLAH AL MUSA
A.ALLAH ALMOUSA
عبدالاله الموس

Таро Ямада
TARO YAMADA
山田太郎
ﾔﾏﾀﾞ ﾀﾛｳ

АЙ-БИ-ЭМ
ИБМ
КОМПАНИЯ ИБМ
ИБМ КОРПОРЕЙШН
ИБМ КОРПОРЭЙШН
CORPORATION IBM
I B M
I B M CO LTD
I B M CORP
I B M CORPORATION 975
IBM
IBM CO
IBM COPORATION
IBM CORP
INTERNATIONAL BUS MACH CORP
INTERNATIONAL BUS MACHIN
INTERNATIONAL BUS MACHINES
INTERNATIONAL BUS MCHS CORP
INTERNATIONAL BUSI NESS MACHINES CORP
INTERNATIONAL BUSIENSS MACHINES CORPORATION
INTERNATIONAL BUSINEESS MACHINES CORP
INTERNATIONAL BUSINES MACHINES
INTERNATIONAL BUSINES MACHINES CORPORATION
INTERNATIONAL BUSINES MACHINES CORPORATION SO
INTERNATIONAL BUSINESS M CHINES CORPORATION
INTERNATIONAL BUSINESS MAC
INTERNATIONAL BUSINESS MACH CO
INTERNATIONAL BUSINESS MACHIN
INTERNATIONAL BUSINESS MACHINCES CORPORATION
INTERNATIONAL BUSINESS MACHINE
INTERNATIONAL BUSINESS MACHINES
INTERNATIONAL BUSINESS MACHINES AB
INTERNATIONAL BUSINESS MACHINES CO
INTERNATIONAL BUSINESS MACHINES COMPANY
INTERNATIONAL BUSINESS MACHINES COMPANY LIMITED
INTERNATIONAL BUSINESS MACHINES COPORATION
INTERNATIONAL BUSINESS MACHINES COPRORATION
INTERNATIONAL BUSINESS MACHINES COPRPORATION
INTERNATIONAL BUSINESS MACHINES COR
INTERNATIONAL BUSINESS MACHINES CORP
INTERNATIONAL BUSINESS MACHINES CORP (IBM)
INTERNATIONAL BUSINESS MACHINES CORP IBM
INTERNATIONAL BUSINESS MACHINES CORPN
INTERNATIONAL BUSINESS MACHINES CORPOATION
INTERNATIONAL BUSINESS MACHINES CORPORAITON
INTERNATIONAL BUSINESS MACHINES CORPORAT
INTERNATIONAL BUSINESS MACHINES CORPORAT ION
INTERNATIONAL BUSINESS MACHINES CORPORATILON
INTERNATIONAL BUSINESS MACHINES CORPORATIN
INTERNATIONAL BUSINESS MACHINES CORPORATION
INTL BUSINESS MACH CORP
INTL BUSINESS MACHINE
INTL BUSINESS MACHINES
INTL BUSINESS MACHINES CORP
INTNL BUSINESS MACH
INTNL BUSINESS MACHINE
INTNL BUSINESS MACHINES
INTNL BUSINESS MACHINIES
INTNL BUSINESS MAHCINES

Informatica Identity Resolution

Высокоэффективный и максимально быстрый поиск потенциальных дублей с “нечеткой” логикой

- Проверенная во всем мире технология сравнения данных (matching) алгоритмами «нечеткой» логики
- Интеллектуальное индексирование и построение ключей и диапазонов поиска - вне зависимости от синтаксиса, фонетики, транслитерации, специфики национальных языковых данных, типа информации, недостатка слов, опечаток и т.д.
- Гибкость - наличие встроенных и настраиваемых стратегий поиска
- Скорость и масштабируемость – работа с большими объёмами данных в критичных приложениях
- Может работать совместно с Informatica Data Quality, Informatica PowerCenter и Informatica MDM

Примеры

Сбербанк, Сбербанк РФ, СБ РФ, Sberbank, Сбербанк России, Сберегательный Банк РФ

Василий Хасанов,

В. Хасанов,

Хасанов В.М.,

г-н Хасанов,

Mr. Khasanov,

Vasiliy Khasanov

Vasily Hasanov

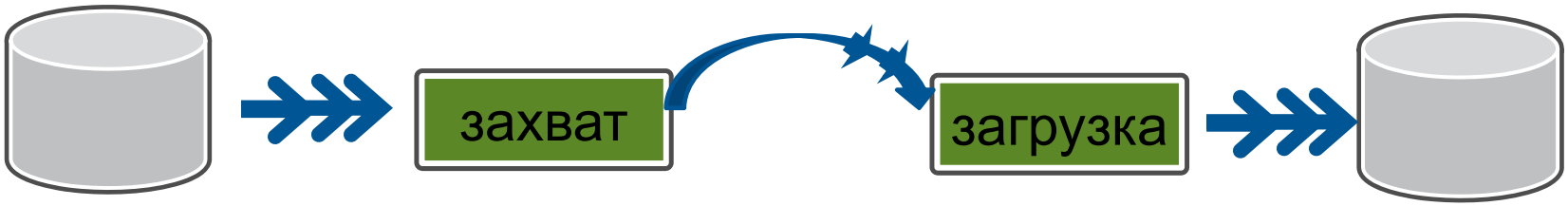
Хасанов Василий Мидехатович…

Отслеживание изменений и репликация транзакций

INFORMATICA®

Informatica Data Replication

Захват, преобразование и доставка транзакционных данных между гетерогенными средами в реальном времени с минимальным влиянием на источники данных



Источник
Oracle,
DB2,
SQL Server,
Netezza

Solaris,
AIX,
Linux,
Windows,

Приёмник
Oracle,
Teradata,
Netezza,
SQL Server,
Greenplum,
DB2,
MySQL,
PostgreSQL,
text files

Informatica Data Replication

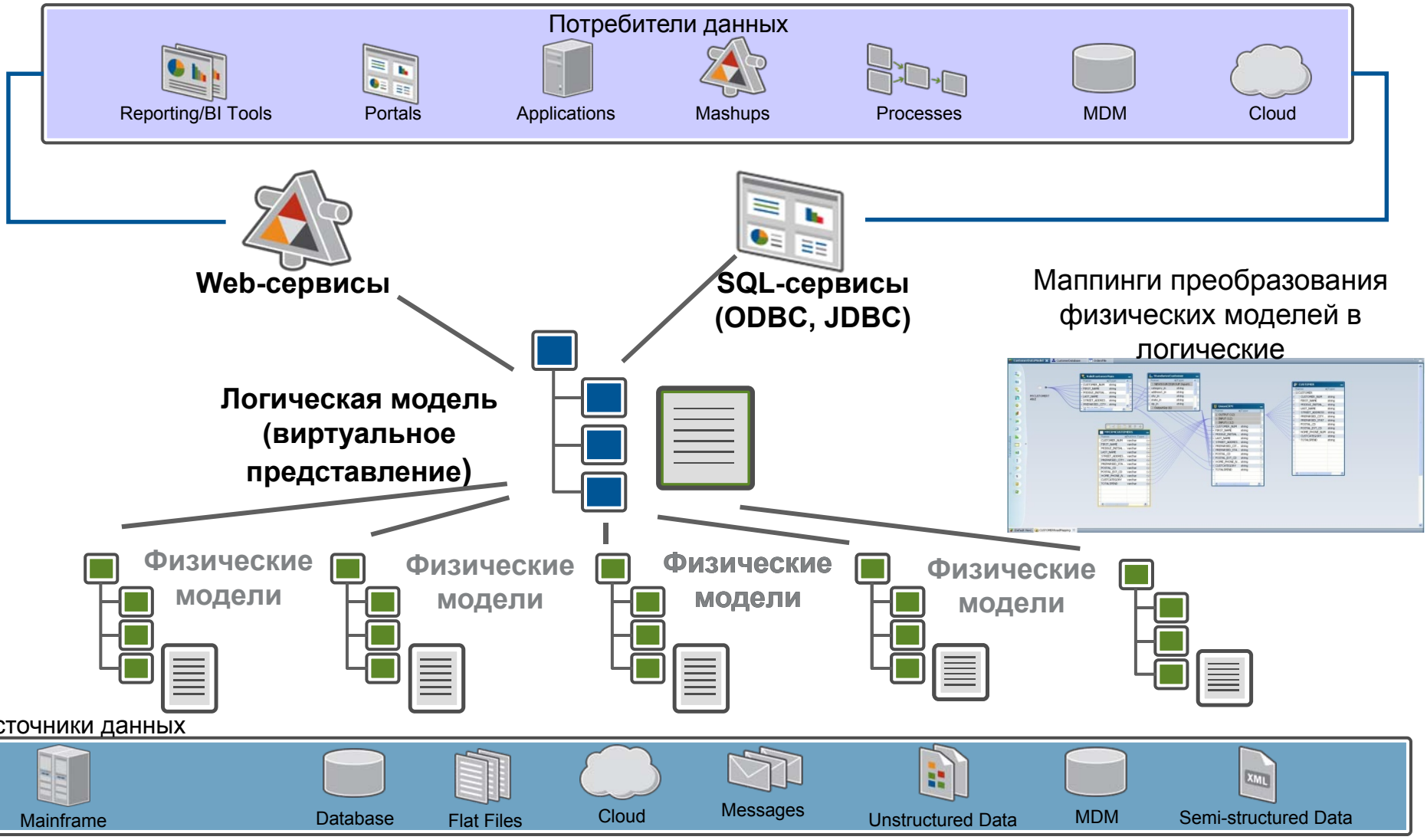
Преимущества использования репликации

Хранилища данных	✓ Актуальные, свежие, критичные для бизнеса данные в режиме 24x7
	✓ Дополнение традиционных способов загрузки (ETL/ELT)
	✓ Использование native способов доступа к БД и appliances
Отчетность	✓ Минимальная нагрузка на оперативные БД
	✓ Изменения в реальном времени
	✓ Актуальная информация для любых критичных приложений
Миграция без простоев	✓ Переход на другую БД или платформу без простоев
Аудит	✓ Отслеживание событий в реальном времени

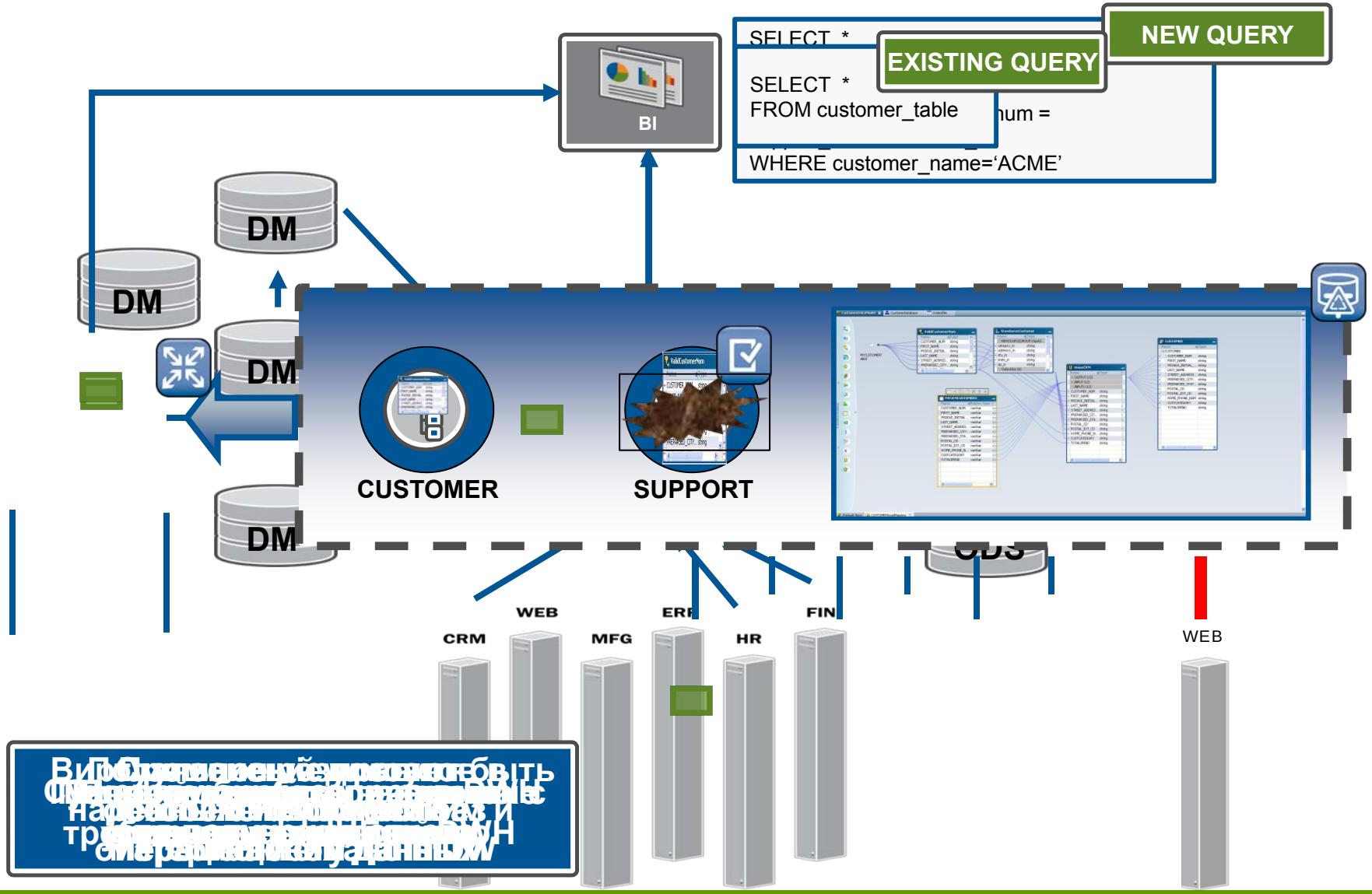
Универсальные сервисы данных

Informatica Data Services

Предоставление универсальных сервисов данных, виртуализация и федерация данных

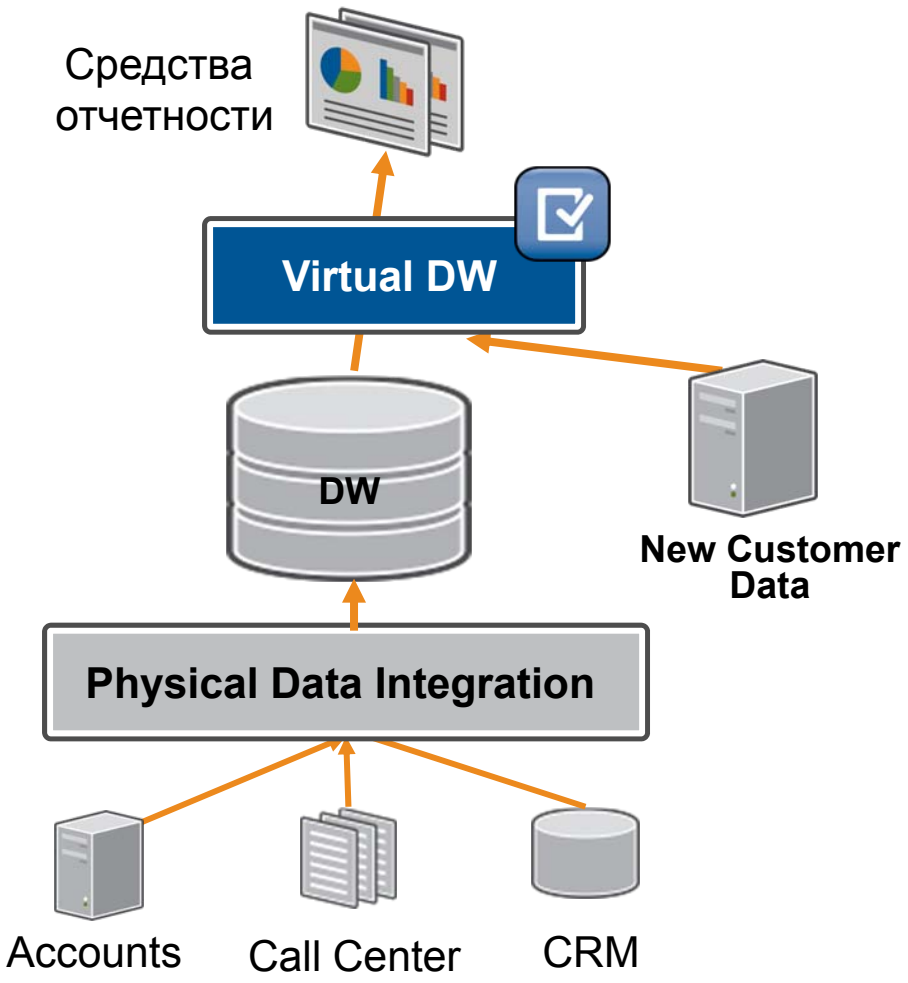


Полнота интеграции с виртуализацией данных в Informatica Data Services



Data Services для BI & Сложных приложений

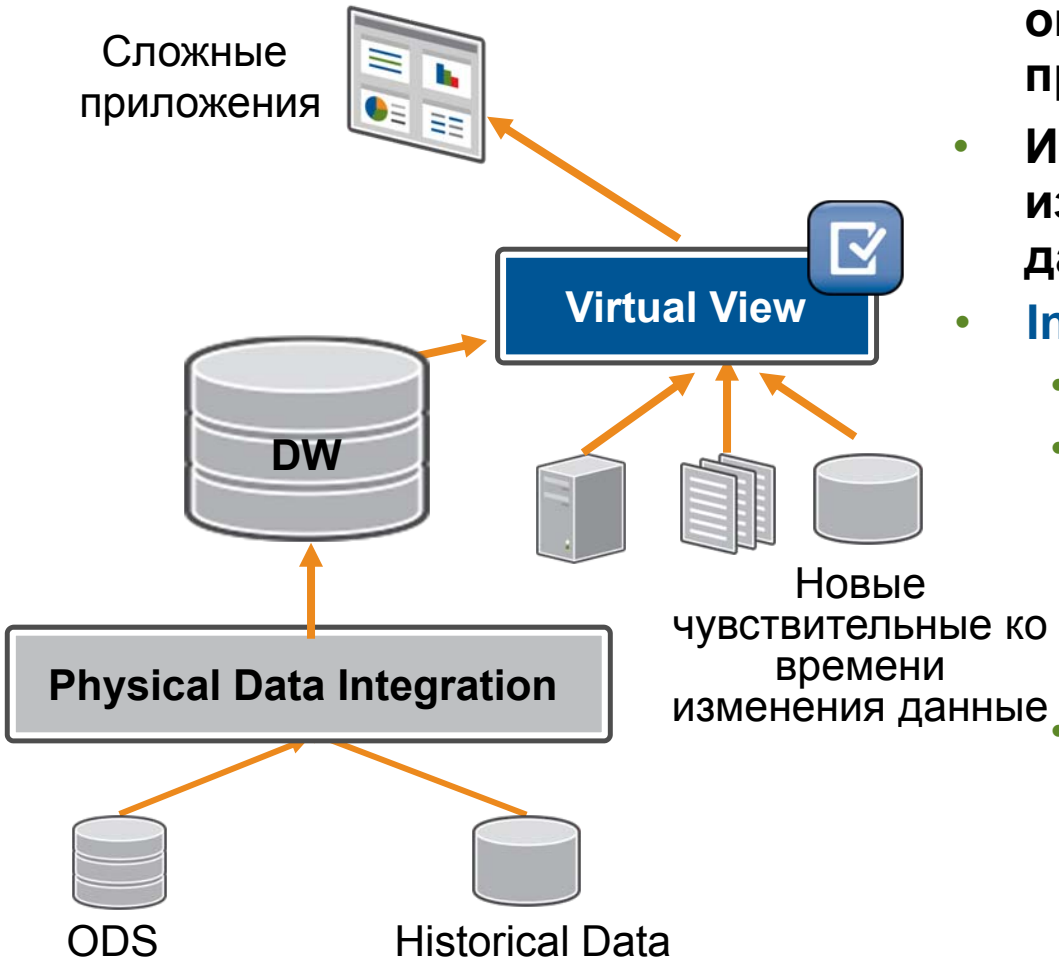
Немедленное включение новых источников



- **Для BI : Возможно быстрое расширение хранилища с появлением новых данных**
- **Для сложных приложений: Возможно немедленное использование новых данных в приложениях**
- **Informatica 9 Data Services:**
  - 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 & Сложных приложений

Получение чувствительных ко времени изменения данных в реальном времени



- **Включение чувствительных ко времени изменения данных пока они еще “свежие” в виртуальное представление**
- **Извлечение немедленной выгоды из виртуально интегрированных данных**
- **Informatica 9 Data Services:**
  - 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

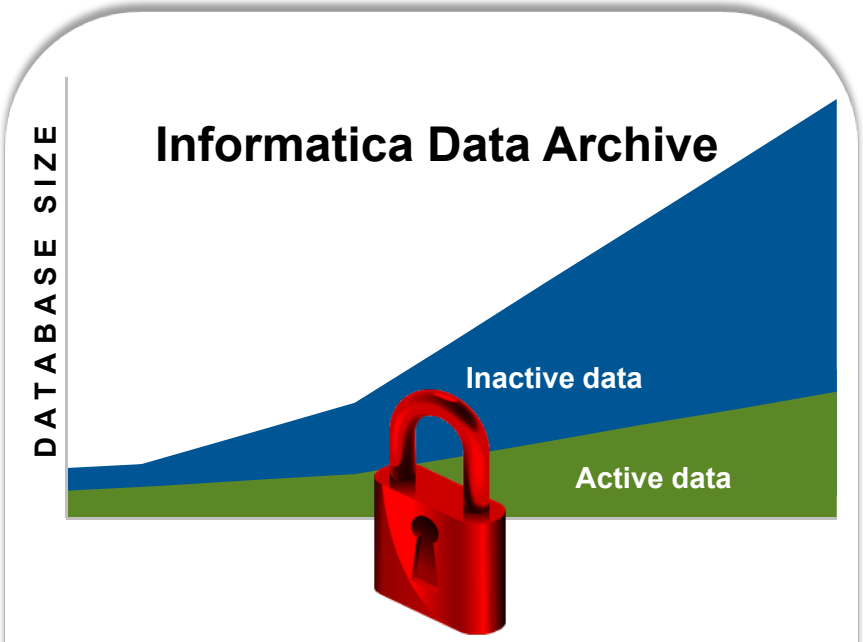
Управление жизненным циклом информации

INFORMATICA®

Informatica ILM

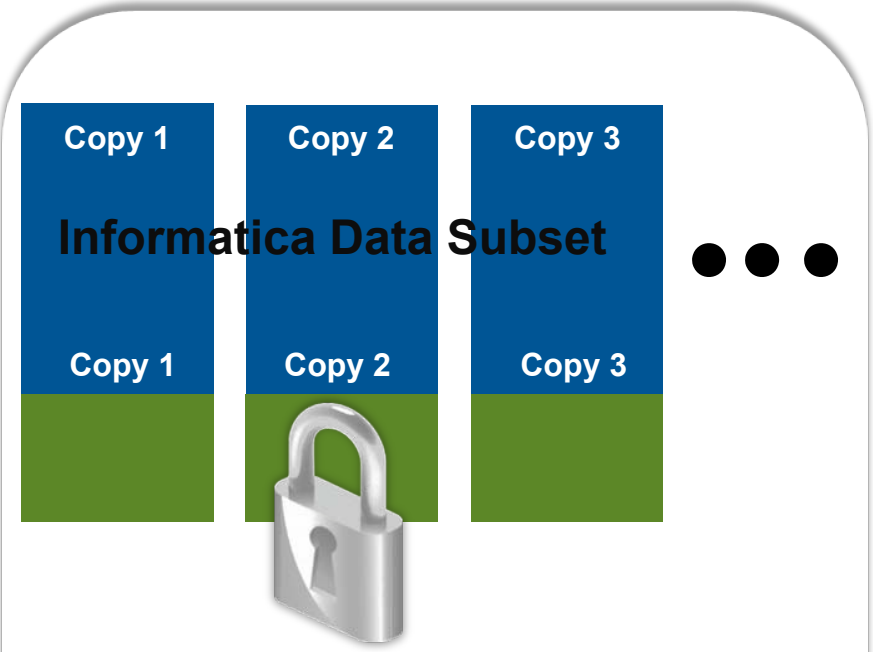
Управление жизненным циклом информации

Production



Informatica Dynamic Data Masking

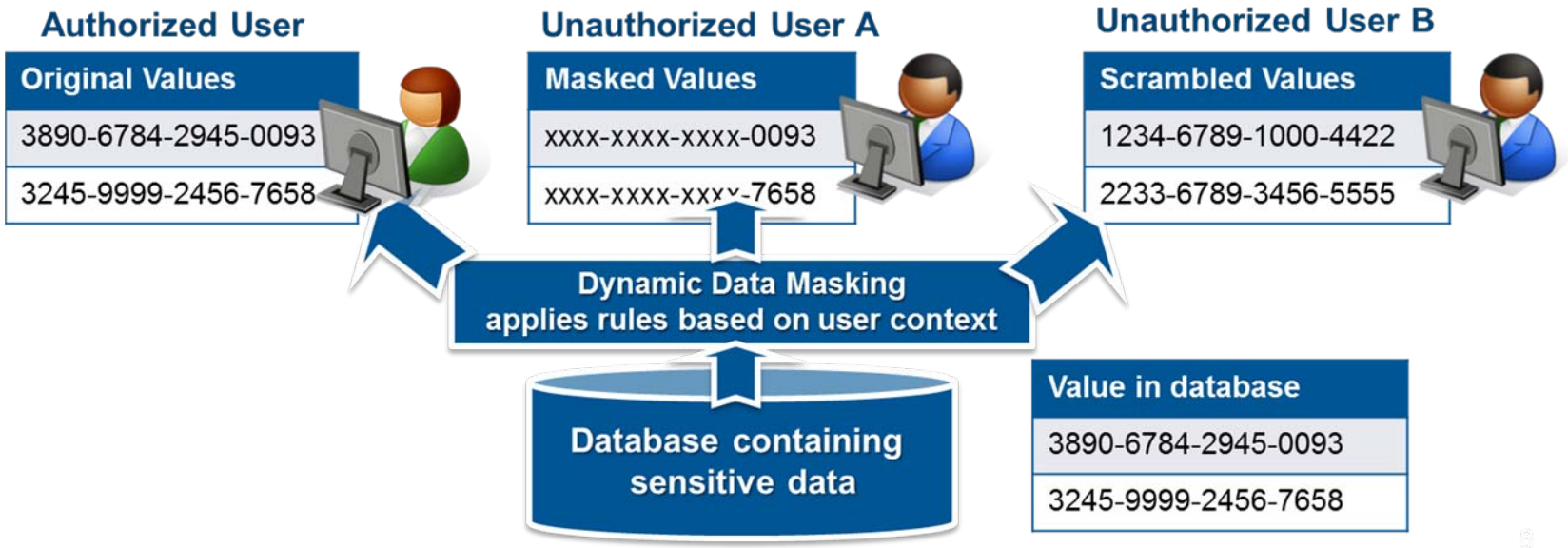
Development and Testing



Informatica Persistent Data Masking

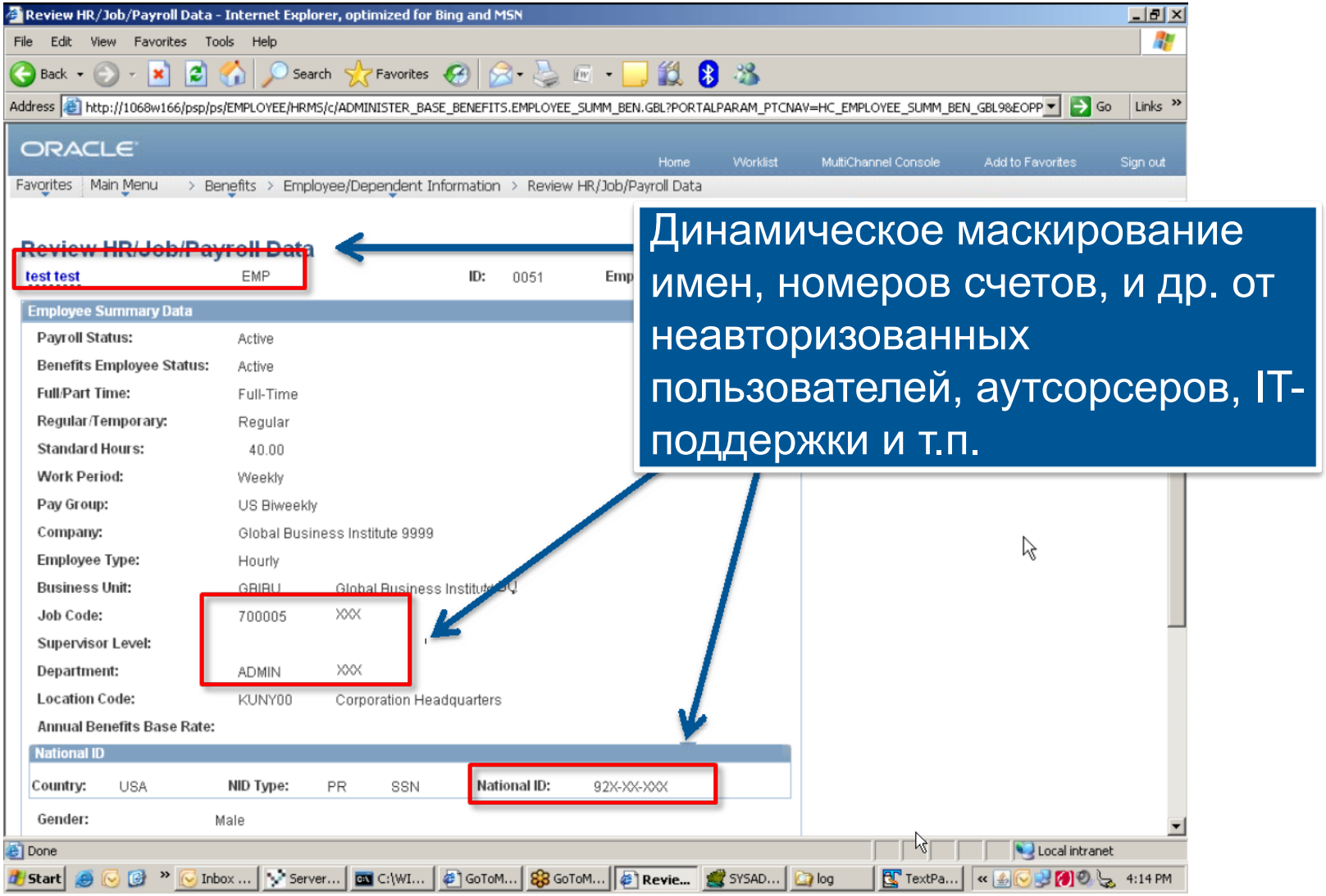
Динамическое маскирование (Dynamic Data Masking) – Защита данных в Production-средах

- Dynamic Data Masking защищает критически важную информацию в Production-средах от тех пользователей, которые не должны иметь доступ к ней для исполнения служебных обязанностей
- Informatica Dynamic Data Masking обеспечивает для каждого пользователя доступ к информации в соответствии с данными об идентификации, ролью и областью ответственности без изменения приложений и баз данных!



Пример динамического маскинга

Приложение Oracle PeopleSoft HR



Высокоскоростной обмен сообщениями

INFORMATICA®

Informatica Ultra Messaging (29 West)

Уникальная технология обмена сообщениями

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

Ultra Messaging Streaming Edition

- Ультра – низкая задержка (~5-10 микросекунд - benchmark)
- Высочайшая производительность (>7 млн. 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 раз и только раз
- Отправка наиболее подходящему получателю
- Балансировка нагрузки - отправка наиболее доступному получателю
- JMS API's

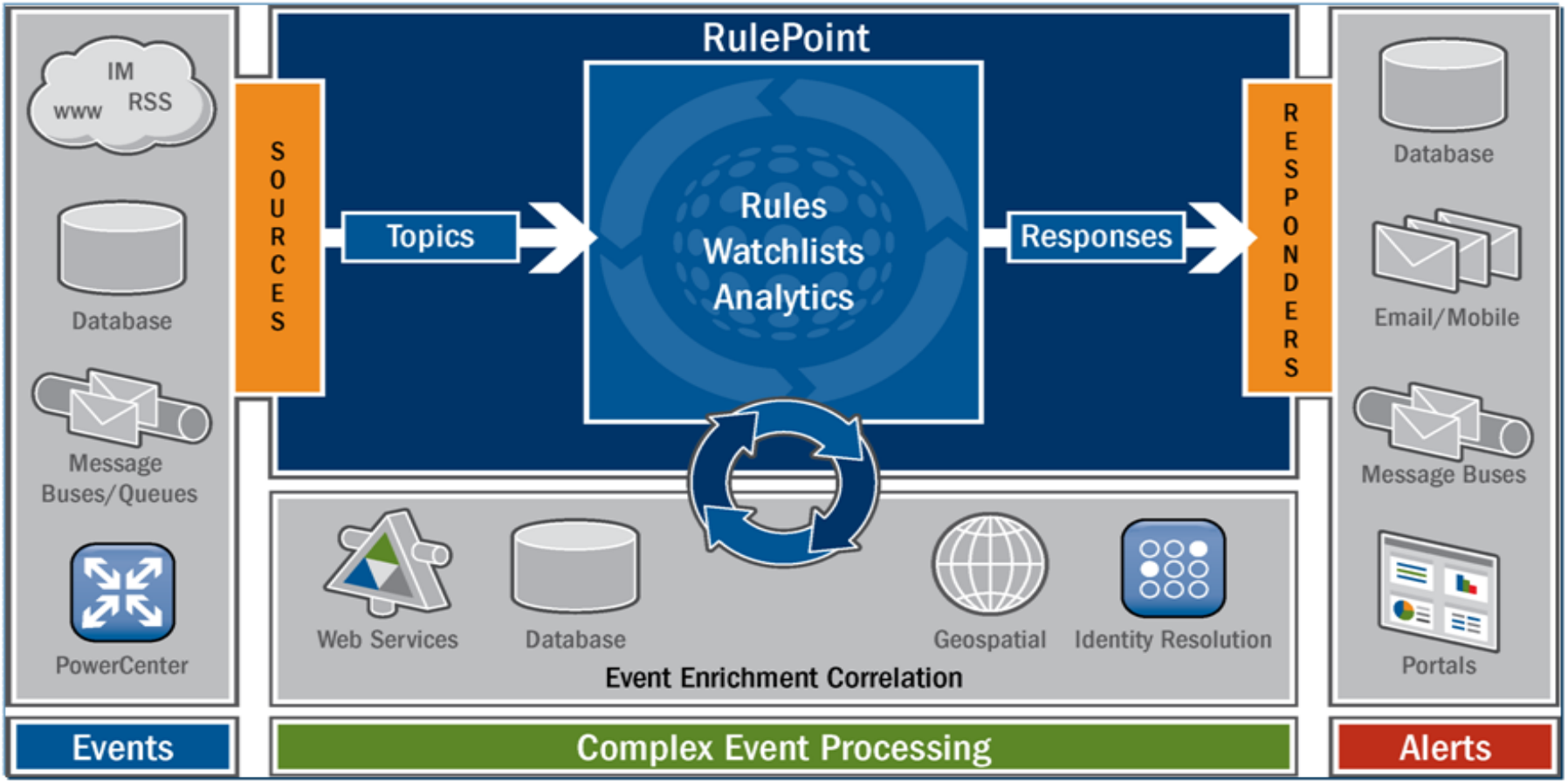
Обработка сложных событий и проактивный мониторинг

Informatica RulePoint

(Complex Event Processing - Обработка сложных событий)

- **Оперативное наблюдение**
  - Отслеживание бизнес-событий по всему предприятию
  - Создание правил бизнес-пользователями
  - Обнаружение скрытых закономерностей
  - Web-интерфейс для создания правил
  - Сокращение времени между обнаружением проблемы и действием
- **Оперативное реагирование**
  - Незамедлительное оповещение пользователей по различным каналам
  - Автоматизированное реагирование на события
  - Автоматизированное взаимодействие со сторонними приложениями

RulePoint Complex Event Processing



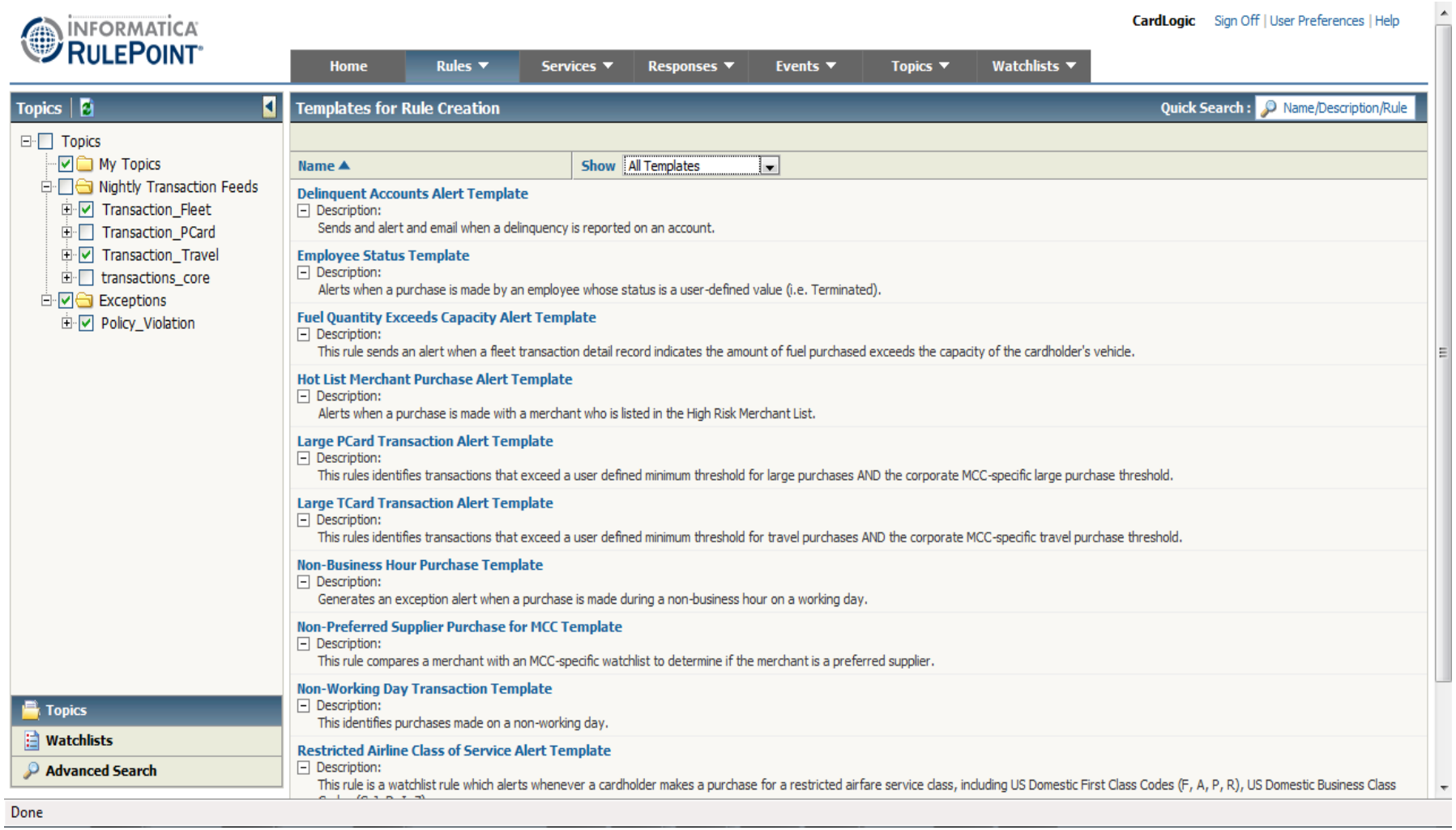
Server Architecture

- 100% Java-based
- Linux, Solaris, Windows
- Runs in Servlet container (e.g. WebLogic, Tomcat, etc.)

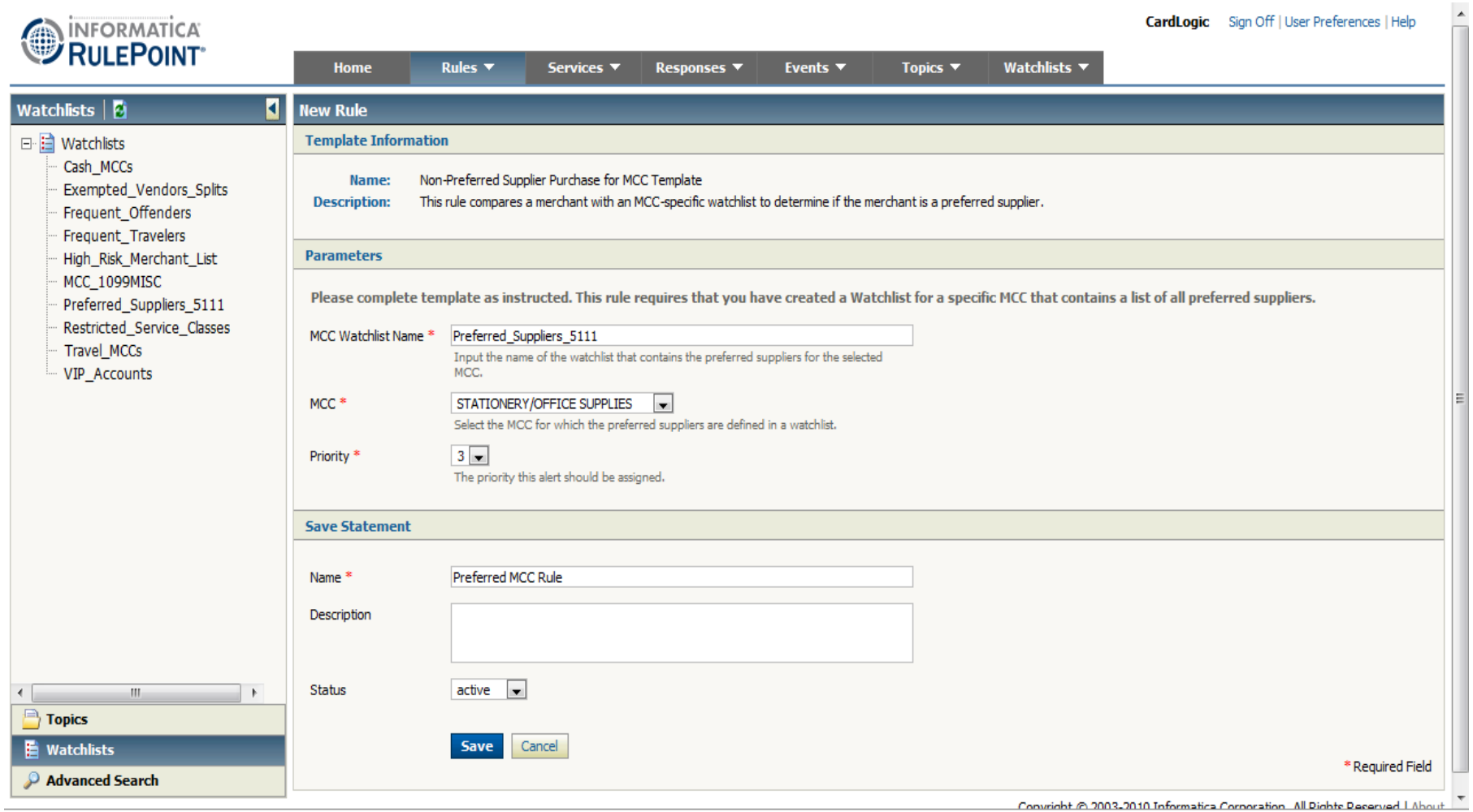
Server Features

- Web browser roles-based user access
- LDAP/Active Directory support
- SDK/API for extensibility and integration

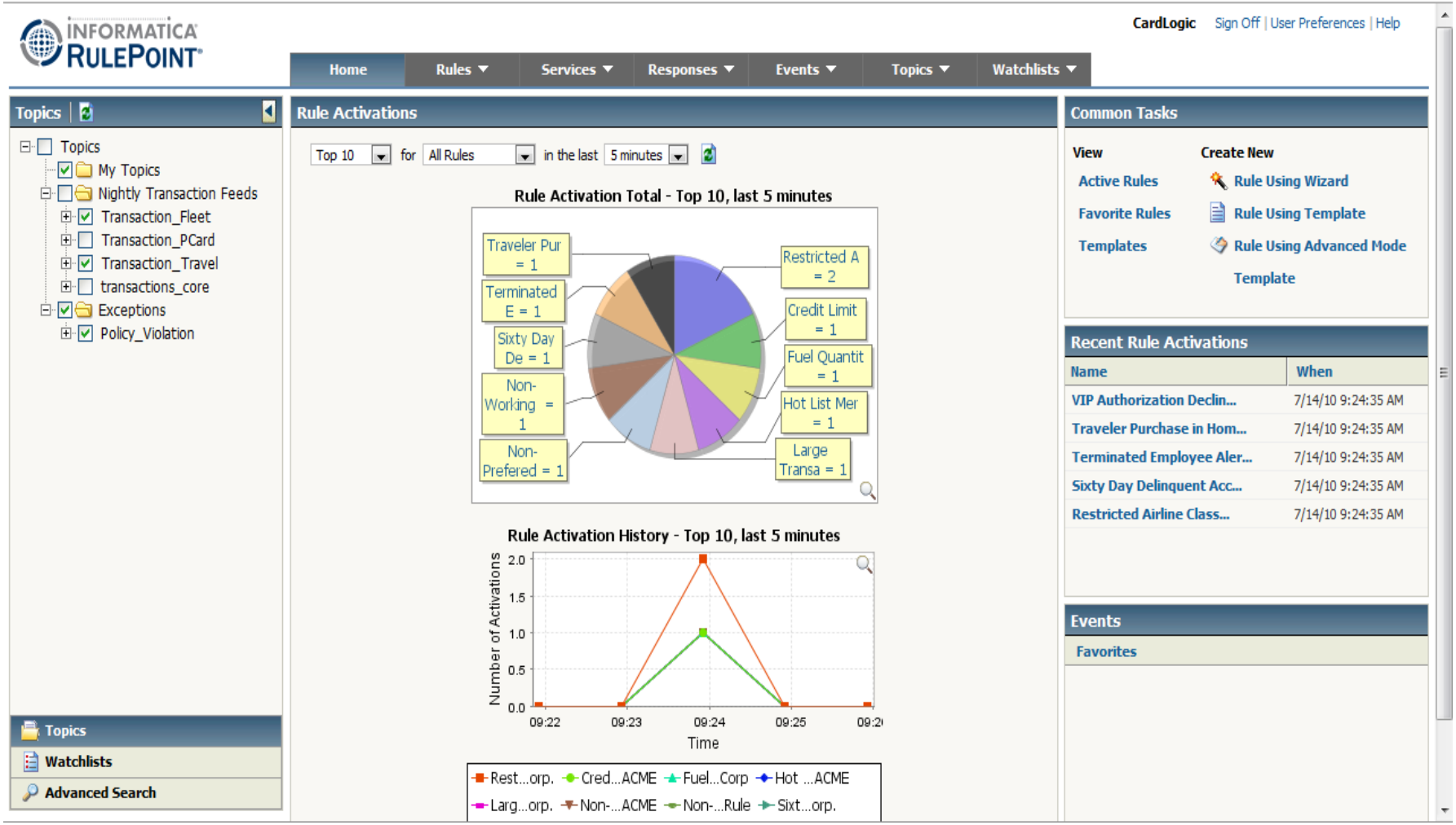
Пользовательские интерфейсы для управления событиями



Пользовательские интерфейсы для управления событиями

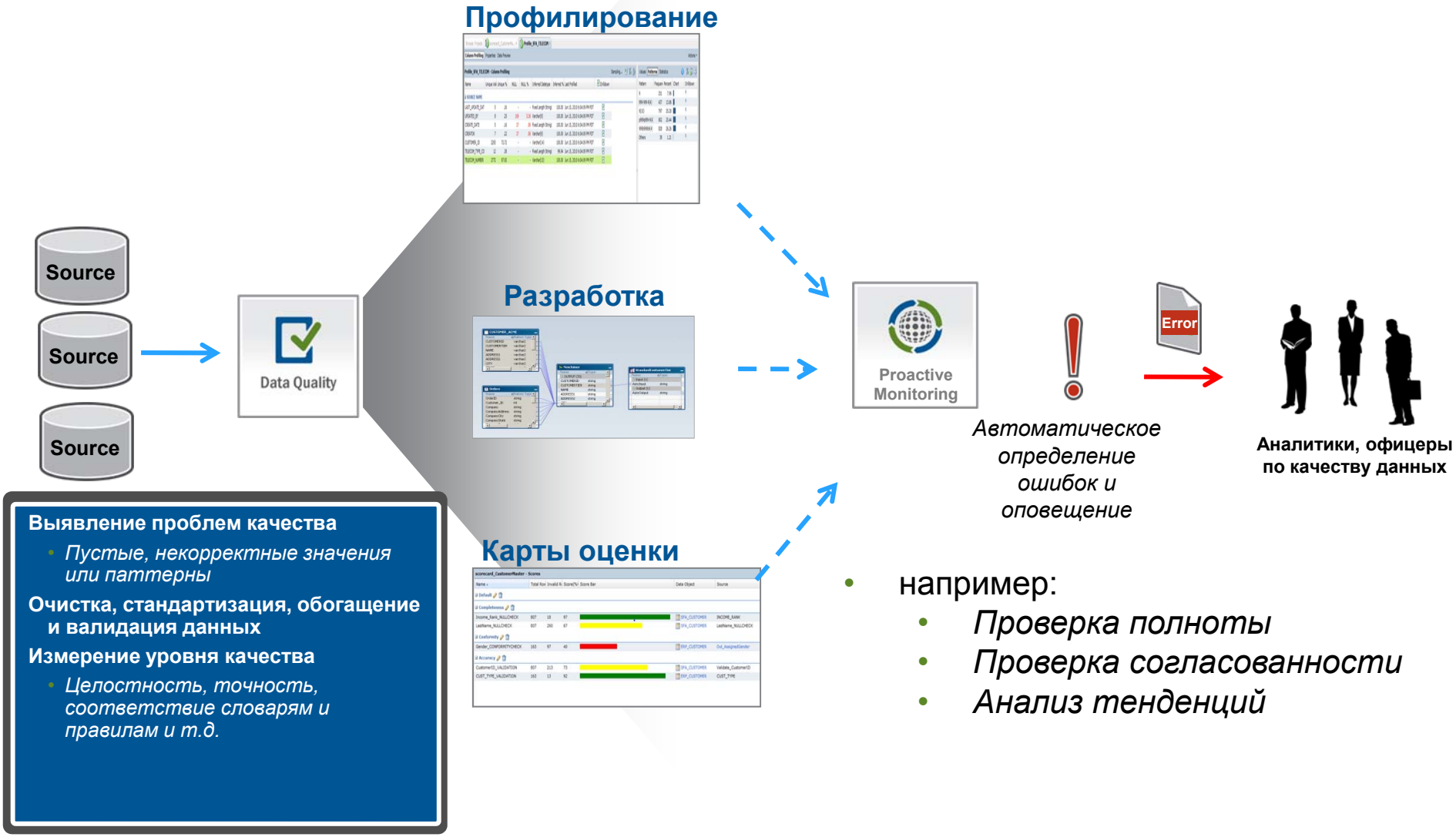


Пользовательские интерфейсы для управления событиями



Проактивный мониторинг качества данных

Непрерывное отслеживание проблем с качеством данных



Заключение

- Проверенное в ведущих мировых компаниях решение
- Комплексный подход и максимальное функциональное покрытие – любые потребности при интеграции лучших в классе решений в рамках одной платформы
- Производительность и обеспечение любых сценариев масштабирования, возможность эффективно обрабатывать любые объемы данных
- Специализированные инструменты для бизнес-пользователей
- Простота конфигурирования (без программирования) → низкая общая стоимость владения и быстрое внедрение

Темы презентации

- Платформа Informatica
- Интеграция данных (Data Integration)
- Качество данных (Data Quality)
- Идентификация и выявление дублей (Identity Resolution)
- Отслеживание изменений и репликация транзакций
- Универсальные сервисы данных (Data Services)
- Динамическое маскирование
- Высокоскоростной обмен сообщениями
- Обработка сложных событий и проактивный мониторинг (Complex Event Processing и Proactive Monitoring)

Спасибо!
Ваши вопросы?

Головной офис компании DIS находится в России:
127473, г. Москва,
ул. Краснопролетарская, д.16/3, офис 8
Tel. : +7 (495) 645-0201

INFORMATICA®