



IBM Information Management

## IBM Dynamic Warehouse

Николай Куликов  
Nikolay\_Kulikov@ru.ibm.com



March 2007

*Ясная, необходимая информация в Любом месте в Любое время*

**Information On Demand**

- Оптимизация каждой транзакции
- Call Центры, Сотрудники в полях



*Помощь в раскрытии преступлений - предоставление необходимой информации в момент прибытия офицеров на место происшествия*

**OLAP & Data Mining**

- Запасы, Товары, Движение



*Оптимизация работы полицейских*

- Запросы и Отчеты
- Финансы, Продажи



*Статистика преступлений*



# Dynamic Warehousing

## *A New Approach to Leveraging Information*

### Information On Demand

Оптимизация реальных процессов



*Dynamic Warehousing*

### **Dynamic Warehousing Требуется:**

1. Встроенная аналитика которая используется как часть бизнес процесса
2. Возможность включения знаний из неструктурированной информации
3. Возможность быстрого доступа к агрегированной и очищенной информации в любом контексте
4. Полный набор средств для интеграции информации



# Больше примеров Dynamic Warehousing в действии

## Традиционные Хранилища

Аналитика и отчетность мошенничеству в страховании

Отчетность по претензиям клиентов

Анализ и отчетность по продажам за прошедший период

Статистика и отчетность по преступлениям

## Dynamic warehousing

Определение потенциально мошеннических заявок до согласования и оплаты

**Преобразование страхования**



Выявление потенциальных похожих заявок, снижение ухода клиента, возможность кросс продаж

**Преобразование клиентской службы**



Выявление необходимой информации о клиенте для идентификации возможности дополнительных продаж

**Преобразование эффективности продаж**



Выявление похожих инцидентов до прибытия на место происшествия

**Преобразование борьбы с преступниками**



## Вызовы стоящие перед традиционными ХД

### ХД должны сейчас:

- **Адресованы** к расширяющимся потребностям аналитики по запросу
- **Использовать** ВСЕ типы информации включая неструктурированную
- **Обслуживать** увеличивающиеся количество и типы приложений и пользователей с различными потребностями уровня обслуживания



Увеличивающаяся Смешанная нагрузка и  
Постоянно изменяющиеся требования различных бизнес направлений  
требует более Динамических Возможностей ХД

# IBM предоставляет больше чем хранилище

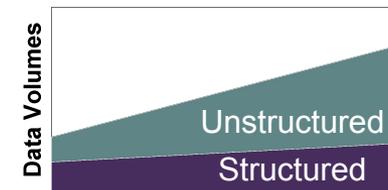
## Встроенная Аналитика

(встроенная и как сервис)

Многомерный анализ  
Data mining, Визуализация



За пределами традиционных структурированных данных



*“As a direct effect of the mixed workload, with continuous loading and the increase in automated transactions from the functional analytics in OLTP, the transactional DBMSs have an edge that challenges the DW DBMSs (such as Teradata)”*

Gartner Data Warehouse Magic Quadrant, 2006

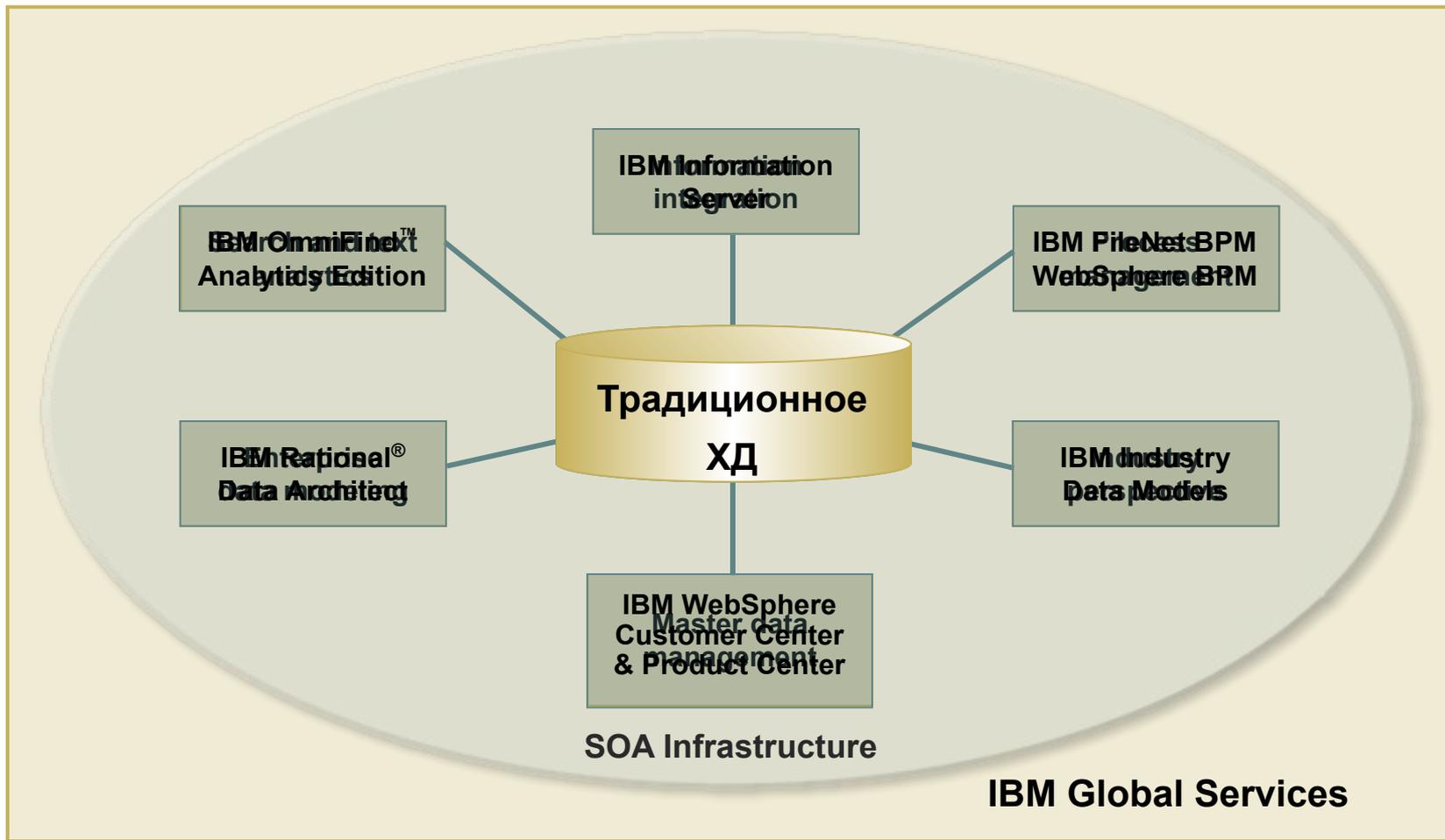
Преимущества транзакционных систем:  
Оптимизация для быстрого доступа, высокая доступность и надежность,  
Масштабируемость, безопасность, аудируемость

Преимущества традиционных систем:  
дешевые диски, лучшая утилизация дисков, повышение производительности запросов

Преимущества современных систем:  
Shared-nothing архитектура  
Партиционирование  
Управление нагрузкой

# IBM Dynamic Warehousing

*Интегрированное предложение*



# Решение от IBM: IBM Information Server

*Предоставление информации которой вы можете доверять*

## IBM Information Server

### Общее Внедрение

#### Понимание



Обнаружение,  
моделирование, управление  
информацией

#### Очистка



Стандартизация,  
объединение,  
и корректировка  
информации

#### Трансформация



Объединение и  
реструктуризация  
информации для нового  
пользования

#### Доставка

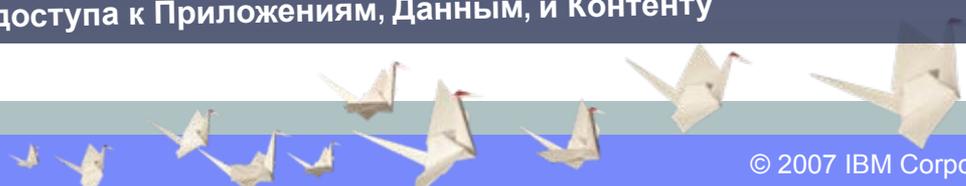


Синхронизация,  
виртуализация,  
перемещение информации

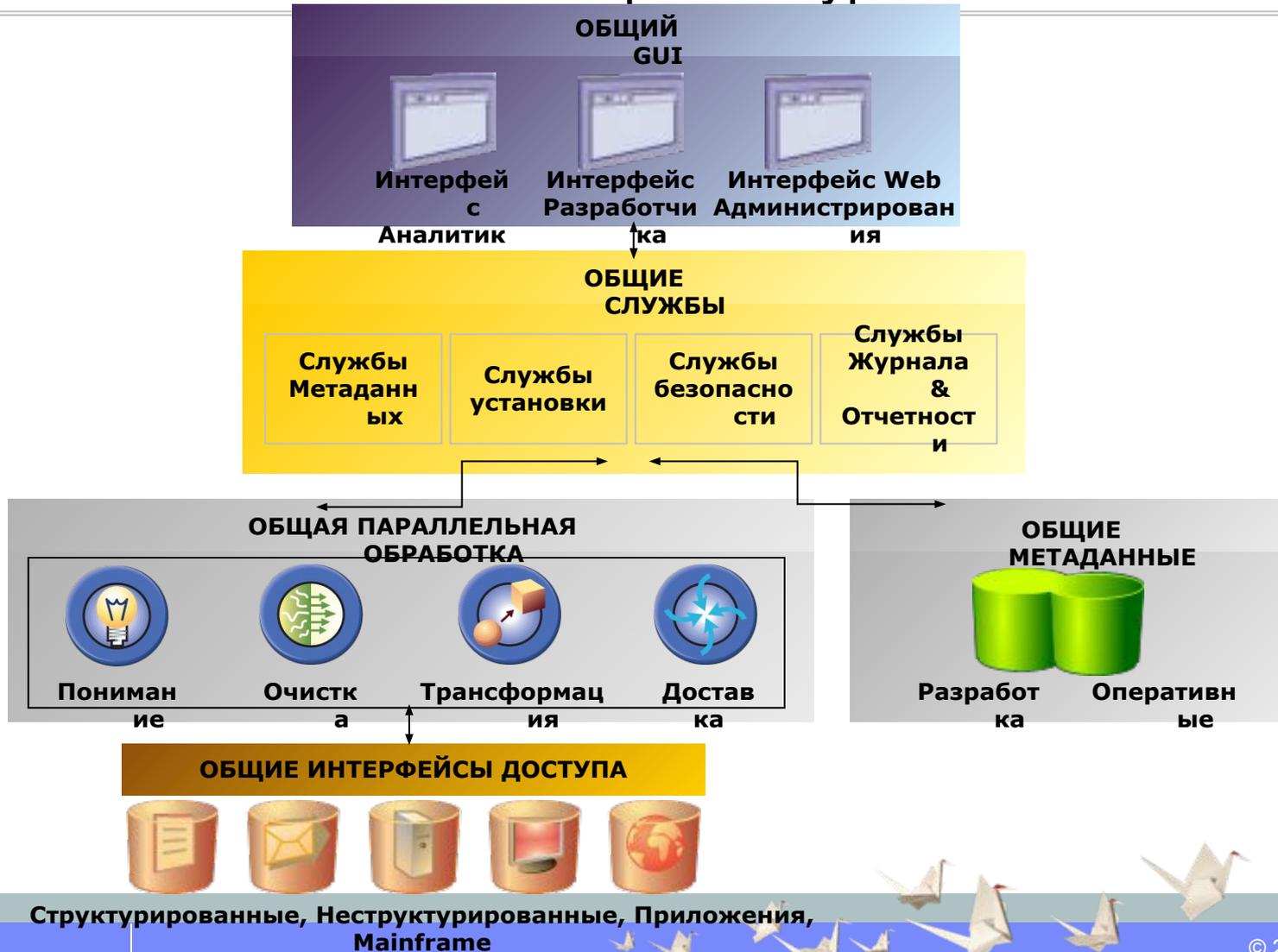
### Общее управление метаданными

#### Параллельная обработка

Широкие возможности доступа к Приложениям, Данным, и Контенту

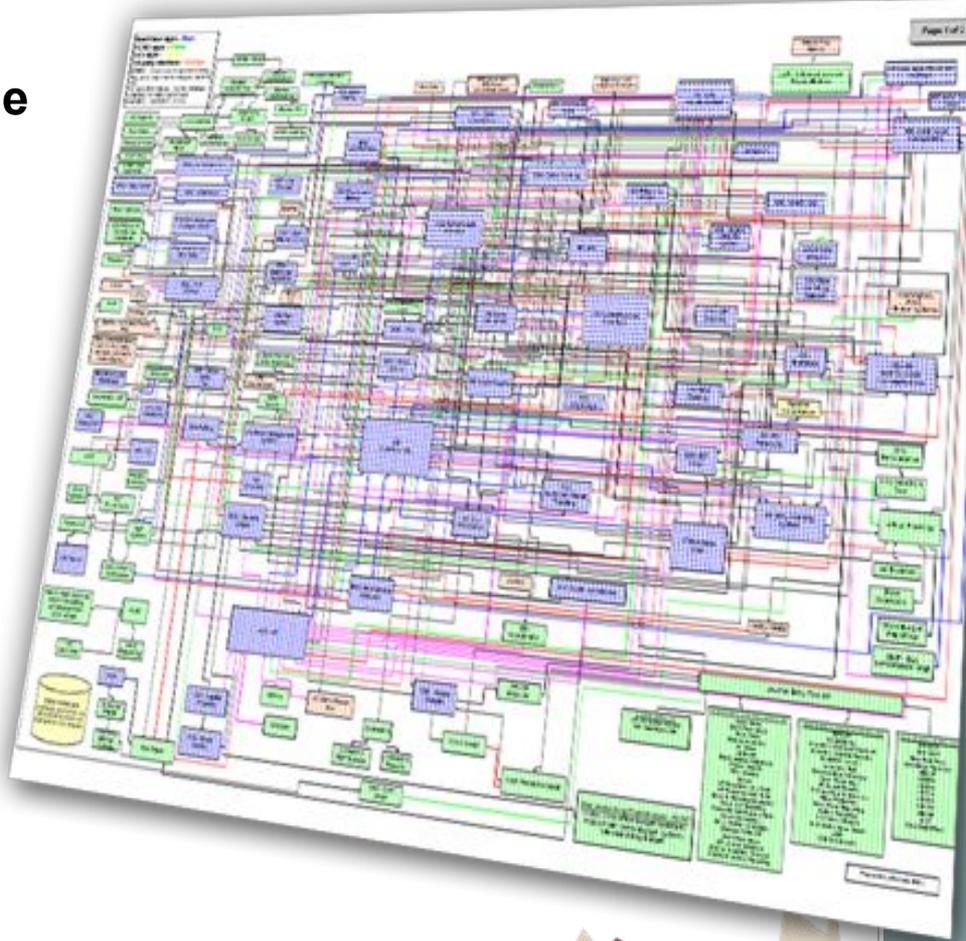


# IBM Information Server - Архитектура



## Почему важно начинать с Понимания?

- Где необходимая информация?
- Как я могу ее получить если она мне нужна?
- Что она означает?
- Могу ли я ей доверять?
- Когда я ее получаю в какой форме она мне нужна?
- Как я могу ее контролировать?

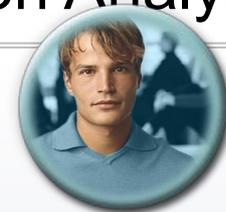


# Физические метаданные: IBM WebSphere Information Analyzer

- Основанный на данных анализ приложений, БД и файловых источников данных
- Безопасная, детальная профилировка полей, между полями и между источниками
- Создание метаданных как результатов профилирования
- Результаты постоянно доставляются посредством IBM Information Server



Subject Matter Experts



Data Analysts

Understand



## IBM WebSphere Information Analyzer

Анализ структур исходных данных и мониторинг их качества и соответствия правилам

Frequency Distribution | Data Class | Properties | Domain & Completeness\* | Format | Pattern

Required Review Not Complete | Reviewed

**Data Type**

Original:  VarChar  Inferred:  VarChar  New:  Select...

View Summary

**Inferred Summary:**

Integer - 25%  
 Decimal - 50%  
 Char - 12.5%  
 Big int - 12.5%

**Inferred Frequency Distribution**

Data Value	Data Type	#	%
efdcve	Char	45	1.00
vfrc	Char	384	8.53
eve	Char	769	17.09
efvefrg	Char	444	9.87
vfrefvev	Char	252	5.60
dfdfd	Char	444	9.87
fdf	Char	252	5.60

Length

Precision

Scale

Nullability

**Physical View**

# Бизнес метаданные: IBM WebSphere Business Glossary

- Построенная на Web система для описания, управления & общего доступа к бизнес метаданным
- Выравнивание действий IT с целями бизнеса
- Предоставление бизнес контекста к информационным активам



Subject Matter Experts



Business Users

Understand



## IBM WebSphere Business Glossary

Create and manage business vocabulary and relationships, while linking to physical sources

Database = DB2

Schema = NAACCT

Table = DLYTRANS

Column = ACCT\_NO

data type = char(11)



Technical



Business

GL Account Number

The ten digit account number. Sometimes referred to as the account ID. This value is of the form L-FIIIIVVVV.

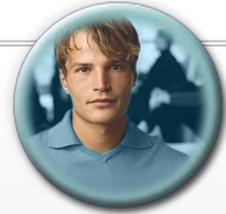


# Очистка данных: IBM WebSphere QualityStage

- Специализированные функции очистки данных прозрачно интегрированы с DataStage
- Визуальные инструменты для определения комплексных отображений и логики сохранения полей
- Проверка очистки, стандартизации, дедубликации информации
- Единая версия правды



Subject Matter Experts



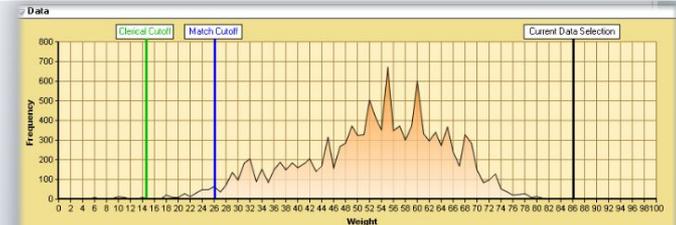
Data Analysts

Cleanse



IBM WebSphere QualityStage™

Standardize and correct source data fields, and match records together across sources to create a single view

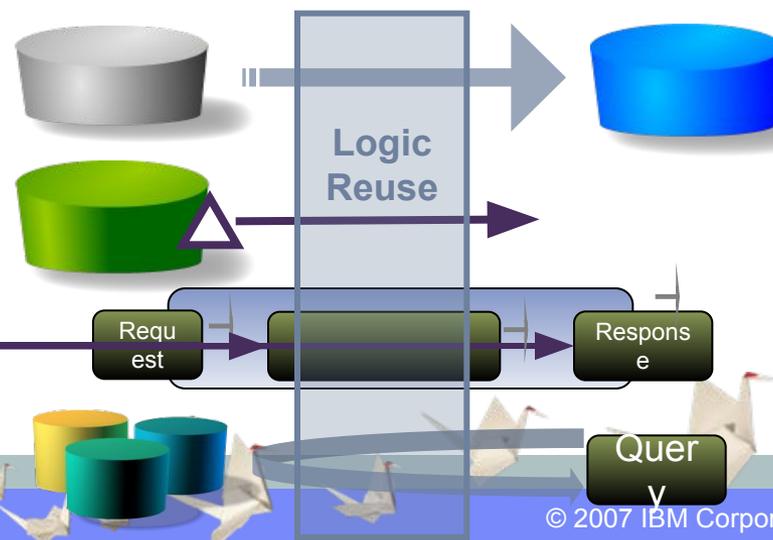
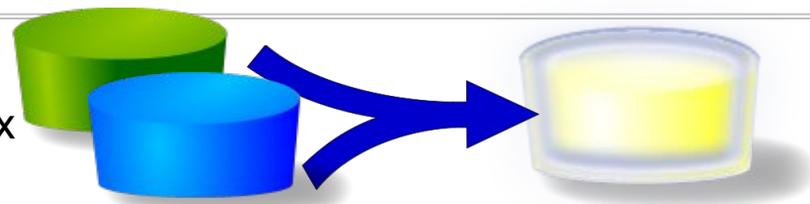


SellD	RecordType	PassNumber	Weight	GenderCode	FirstName	MiddleName
4111	XA	1	85.57	IM	OTIS	GARLAND
4111	DA	1	85.57	IM	OTIS	GARLAND
5311	XA	1	84.08	IM	NICHOLAS	T
5311	DA	1	84.08	IM	NICHOLAS	T
3420	XA	1	81.68	IM	CLAUD	LADALE
3420	DA	1	81.68	IM	CLAUD	LADALE
3420	XA	1	81.68	IM	CLAUD	LADALE
3420	DA	1	81.68	IM	CLAUD	LADALE
4826	XA		80.20	IM	CLIFF	H
4826	DA		80.20	IM	CLIFF	H
7328	XA		79.75	I	HSHU	CHIN
7328	DA		79.75	I	HSHU	CHIN
2282	XA		79.47	IF	GILLIS	GILLIS
2282	DA		79.47	IF	GILLIS	GILLIS
6879	XA		79.47	IF	MICHELLE	MICHELLE
6879	DA		79.47	IF	MICHELLE	MICHELLE

Visual Match Rule Design

# Что важно в трансформации и доставке данных?

- Трансформация это ключ для предоставления информации для использования ее в новых бизнес контекстах – это требует что бы эта информация была основанная на метаданных
- Разработана для использования экспертами по информации и управляется посредством метаданных
- Трансформация и доставка может быть использована множеством механизмов
  - Перемещение больших объемов заданиями (batch)
  - Ответ на событие в реальном времени
  - SOA
  - Федеративные запросы



# Трансформация & Перемещение: IBM WebSphere DataStage

- Визуальная разработка потоков данных с сотнями встроенных трансформаций
- Оптимизированное использование объектов интеграционных
- Возможность параллельной обработки без необходимости изменять процессы обработки
- Возможность работы как с batch так real-time операциями



Developers



Architects

Transform

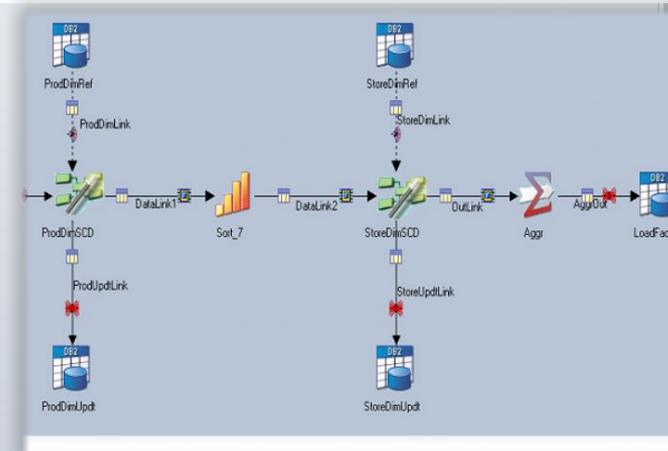


Deliver



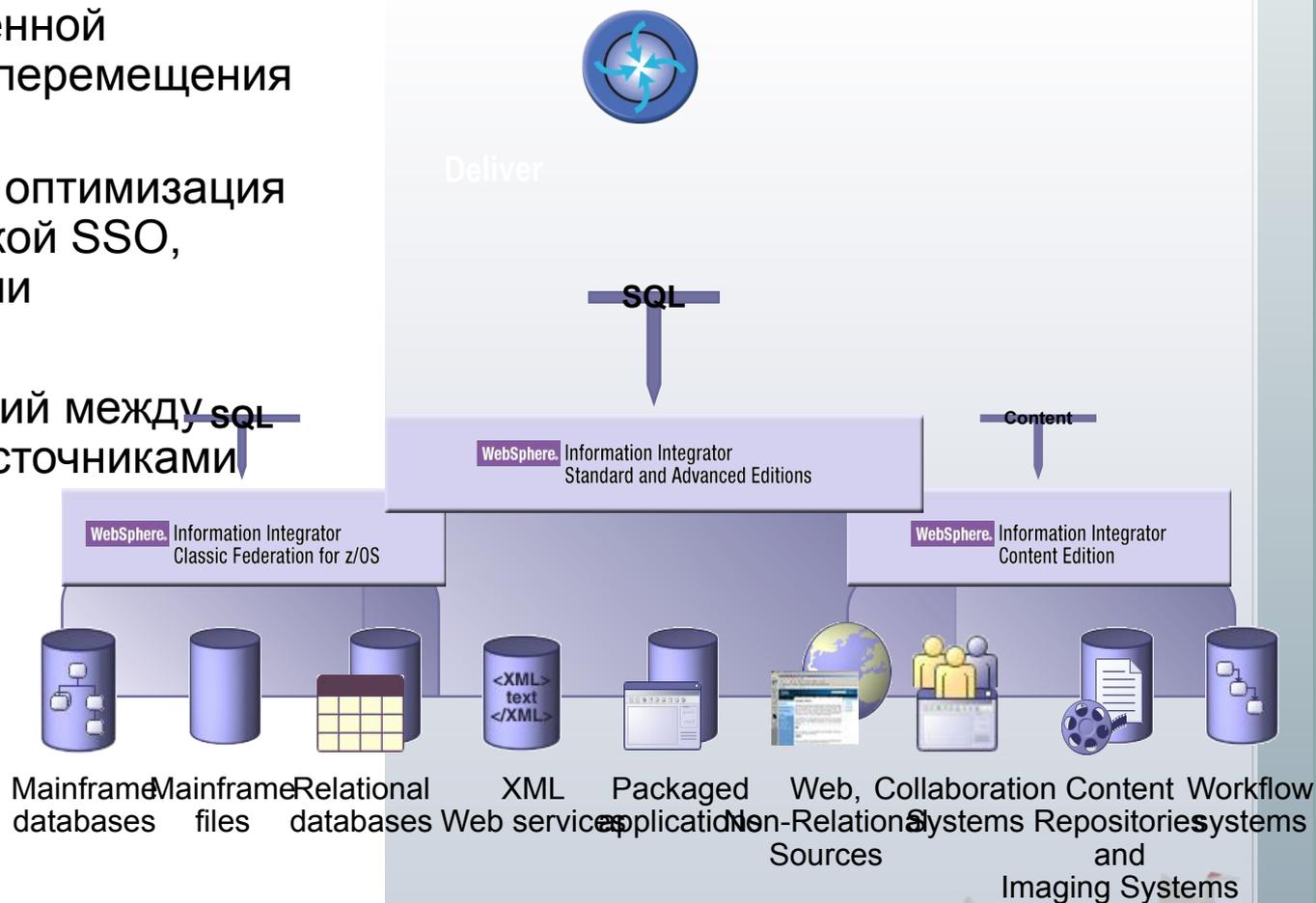
## IBM WebSphere DataStage®

Трансформация и агрегация любых объемов информации



# Федеративные данные: IBM WebSphere Federation Server

- Доступ к распределенной информации без ее перемещения
- Лучшая в индустрии оптимизация запросов с поддержкой SSO, функции компенсации
- Поддержка транзакций между ~~SQL~~ распределенными источниками



# Федеративные запросы делают интеграцию простой как SQL

```
parameters_return_billto_key as BILL_TO_KEY
```

```
parameters_return_shipto_key as SHIP_TO_KEY
```

```
CASES_SHIPPED
```

```
GROSS_SALES
```

```
URL
```

```
GETKEYSSOAP_GETKEYSREALTIME_NN
```

```
GLOBAL_SALES_TRAN_NN
```

```
URL_INVOICES
```

```
getkeysrealtime_ship_to_number
```

```
getkeysrealtime_ship_to_number URL_INVOICES.shipno
```

```
ltrim(rtrim(translate(ship_to_number, ' ', x'0a')))
```

```
getkeysrealtime_ship_to_number
```

```
parameters_return_billto_key
```

```
ltrim(rtrim(translate(sales_order_number, ' ', x'0a')))
```

```
URL_INVOICES.orderno
```

Один запрос SQL :

- Web Service
- XML Documents
- Data Warehouse
- Unstructured Data

# Интегрированный набор инструментов для Dynamic Warehousing

The screenshot displays the IBM Business Intelligence (BI) Design Studio V9.1 interface. The main window shows a data flow diagram with several data sources and transformation nodes. A SQL script window is open on the right, showing the following code:

```

CODE_UNIT:SQLSCRIPT
($RESOURCE = $(grpl/eqldb1)
)
{
SET CURRENT SCHEMA="$(grpl/eqldb1)";
DECLARE GLOBAL TEMPORARY TABLE INPUT_02SCD_TEMP_02
INSERT INTO SESSION.INPUT_02SCD_TEMP_02 (SELECT
DELETE FROM INPUT_02 WHERE NOT EXISTS (SELECT
UPDATE HUIILIAO.SCD_FACT SET (AMT) = (SELECT
UPDATE HUIILIAO.SCD_FACT SET (AMT, FSK2, SERVE
INSERT INTO HUIILIAO.SCD_FACT (FK1, AMT, FSK2)
    
```

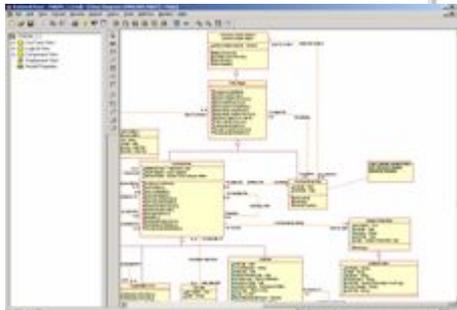
The interface includes a Navigator pane on the left showing a project tree with folders like 'ComplexOptTests', 'control-flows', 'databases', and 'data-flows'. A Palette on the left provides various tools for data flow, including 'Table Source', 'Order By', 'Distinct', 'Where Condition', and 'Fact Key Replace'.

**IBM Information Server**

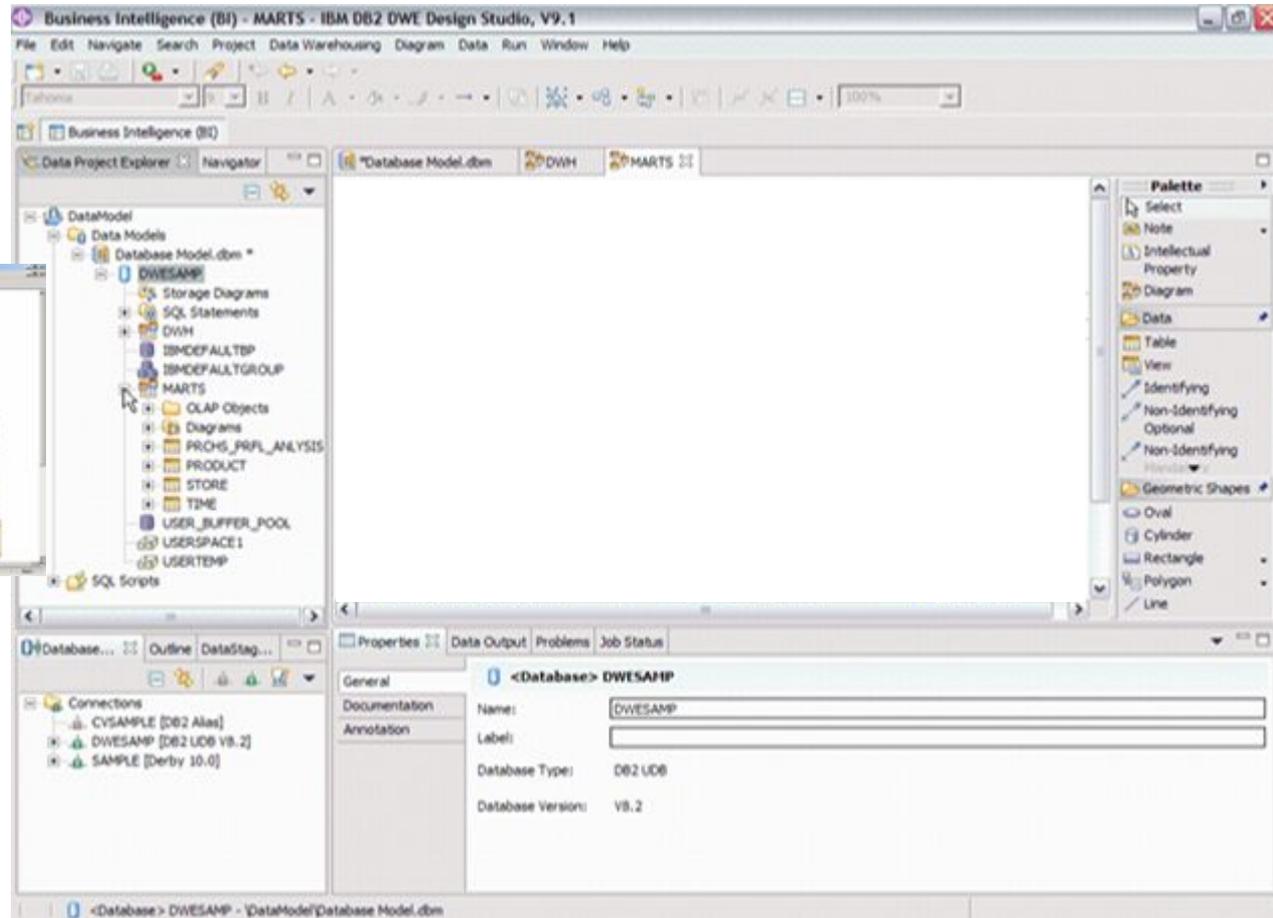


# Интегрированный набор инструментов для Dynamic Warehousing

Rational software



Data Architect



# Introducing IBM Balanced Warehouse™

*A fast track to warehousing*



Balanced Warehouse



## Простота

- Предопределенная конфигурация для

## Balanced Configuration Unit (BCU)

- Один номер для поддержки решения
- Преконфигурированный,

Претестированный, масштабируемый

## Гибкость при росте

Набор Программного и Аппаратного Обеспечения

Обеспечение BCU для

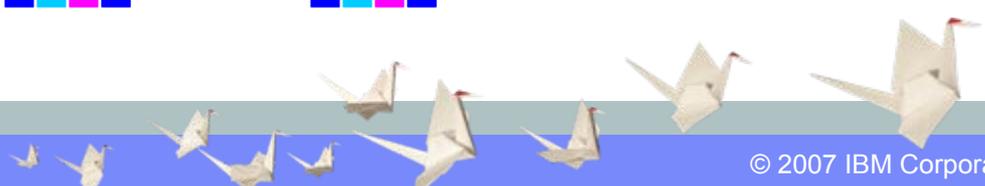
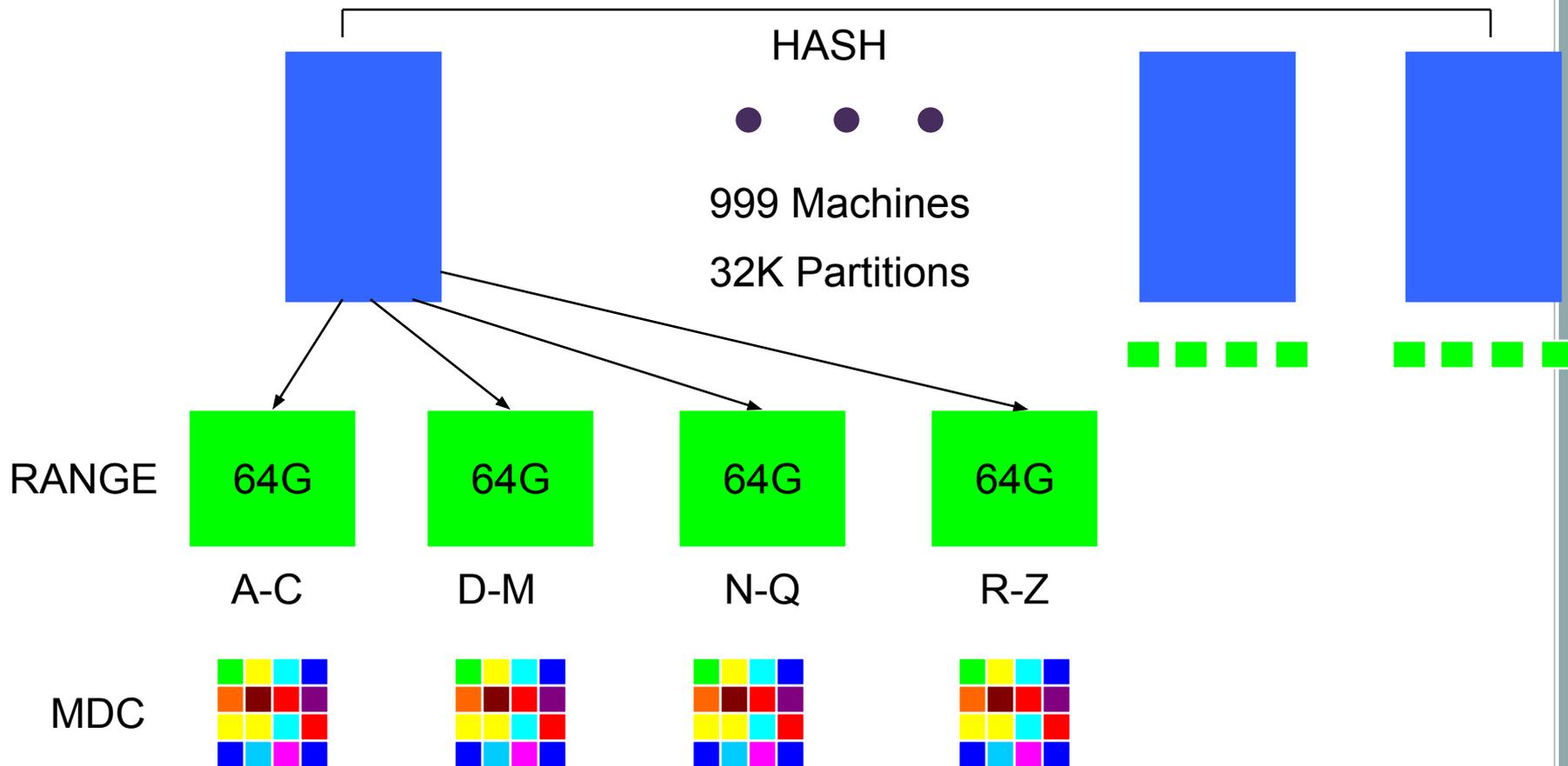
- Надежно, возможность повторного использования для других задач

## Оптимизирован для

## Производительности

- Преконфигурирован и сертифицирован для предоставления гарантированной производительности
- Основан на лучших практиках для уменьшения риска

# Гибридное партиционирование

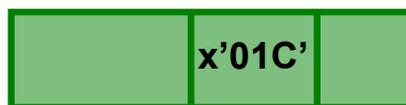


# Компрессия строк

Без Компрессии

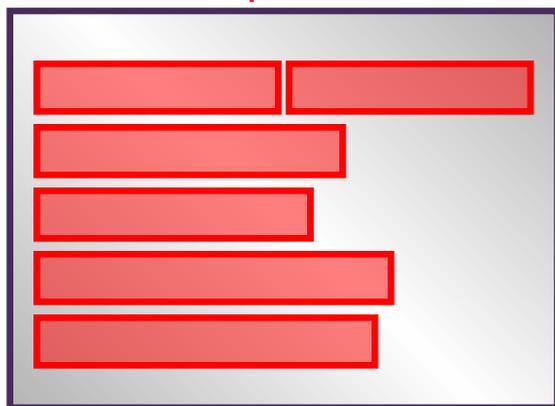


С Компрессией

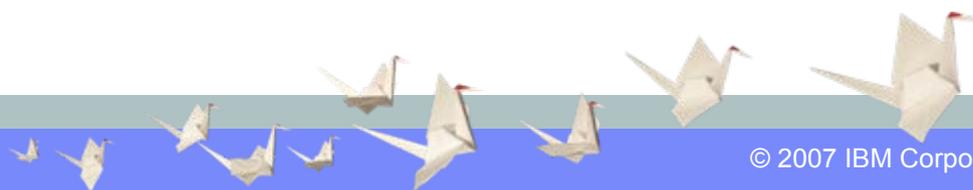
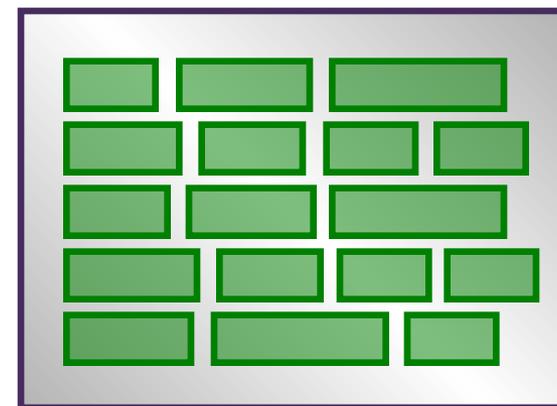


Общая  
последовательность  
заменяется на 12 байт

Страница без  
Компрессии



Страница с  
компрессией



## DB2 - More Compression Ratios (Customer Data)

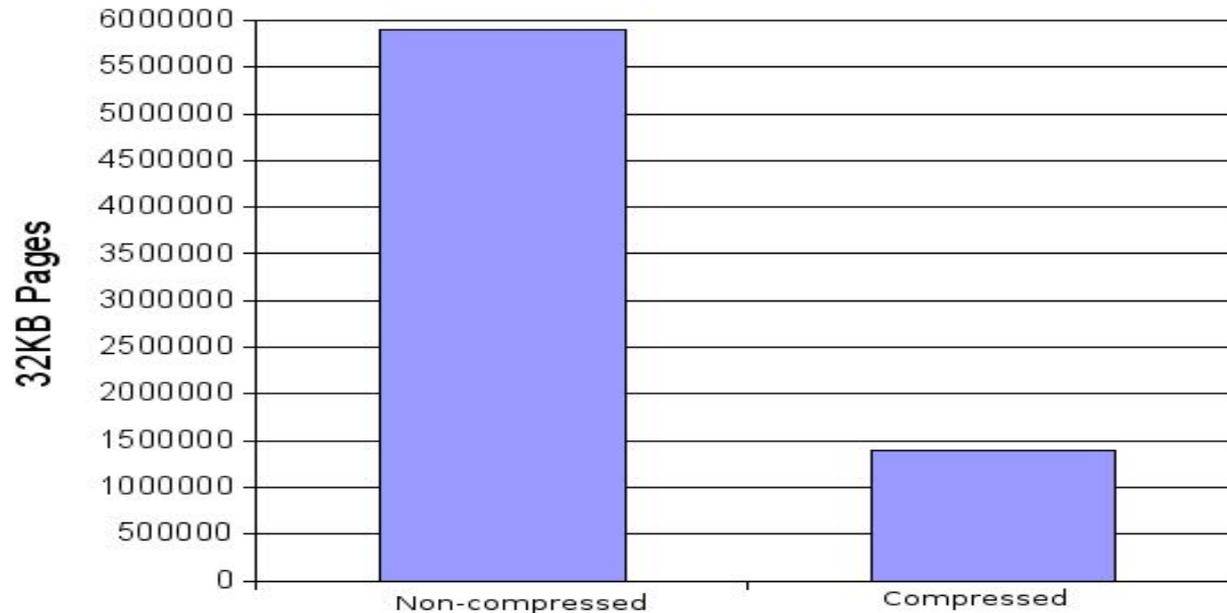
### Compression Type 32KB Page Count Space Required on Disk

No compression 5893888 179.9GB

Row compression 1392446 42.5GB

% Pages Saved: 76.4%

### T1 Compression - 179.9GB Initial Size



# Индустриальные Модели данных

## BDW

(Banking Data Warehouse)

- Прибыльность
- Маркетинг
- Управление рисками
- Управление активами пассивами
- Соответствие Требованиям

## Финансовые Рынки

(Financial Markets Data Warehouse)

- Управление Рисками
- Управление активами пассивами
- Соответствие требованиям

## Управление Здоровьем

(Health Plan Data Warehouse)

- Заявки
- Medical management
- Управление сетью
- Продажи, маркетинг and membership
- Финансы

## Страхование

(Insurance Information Warehouse)

- Клиенты
- Заявки
- Прибыльность
- Соответствия требованиям
- Управления рисками

## Розница

(Retail Data Warehouse)

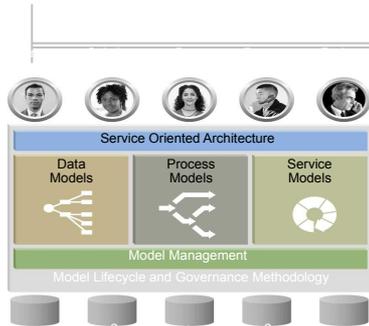
- Клиенты
- Управление закупками
- Оптимизация складов
- Управление цепочкой
- Соответствие требованиям

## Телекоммуникации

(Telecommunications Data Warehouse)

- Управление оттоком
- Управление взаимоотношениями
- Продажи и маркетинг
- Качество обслуживания
- Профайл пользователя

# IBM Industry Data Models: Ключевые компоненты



## ▪ Модель данных для терминов

- Словарь бизнес определений для единых определений в рамках организации и индустрии

## ▪ Модель EDW

- Эта модель определяет как множество источников могут быть консолидированы в единую логическую структуру

## ▪ Проектные представления

- Логическое подмножество EDW и BSTs для решения конкретной задачи

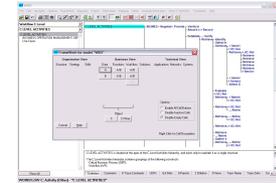
## ▪ Шаблоны бизнес решений (BST)

- Набор Key Performance Indicators для создания OLAP кубов, витрин, шаблонов отчетности

# Фаза планирования

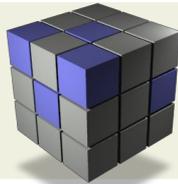


	Revenue	Operating Profit	Operating Assets	Capital Expenditure	Debt	Equity	Operating Income	Operating Expenses	Operating Assets	Capital Expenditure	Debt	Equity
Q1-12	1000	200	500	100	100	100	100	100	100	100	100	100
Q2-12	1000	200	500	100	100	100	100	100	100	100	100	100
Q3-12	1000	200	500	100	100	100	100	100	100	100	100	100
Q4-12	1000	200	500	100	100	100	100	100	100	100	100	100
Q1-13	1000	200	500	100	100	100	100	100	100	100	100	100
Q2-13	1000	200	500	100	100	100	100	100	100	100	100	100
Q3-13	1000	200	500	100	100	100	100	100	100	100	100	100
Q4-13	1000	200	500	100	100	100	100	100	100	100	100	100
Q1-14	1000	200	500	100	100	100	100	100	100	100	100	100
Q2-14	1000	200	500	100	100	100	100	100	100	100	100	100
Q3-14	1000	200	500	100	100	100	100	100	100	100	100	100
Q4-14	1000	200	500	100	100	100	100	100	100	100	100	100

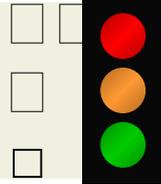


- Создание стандартов на данные и их управление
- Бизнес планирование и развития IT
- Стабильность проектов

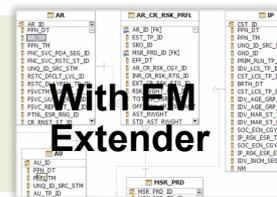
# Фаза анализа



Business Solution Templates



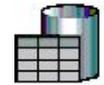
Project Views



Rational Data Architect



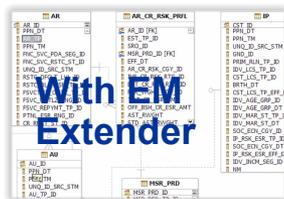
Information Analyzer



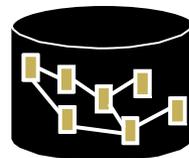
Analyze  
Source Data

- RDA (Enterprise Model Extender)
  - Выявление Бизнес требований
  - Структурированная по области проблем
  - Определяет область для реализации IT реализации
- Дополнение Information Analyzer
  - Предоставление необходимого анализа данных

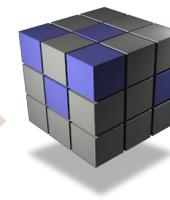
## Создание дизайна



Create Database Schema

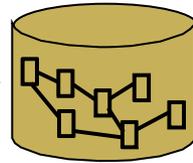
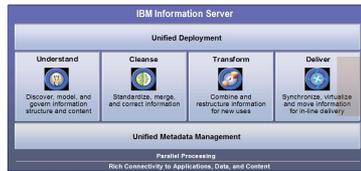


Create OLAP or Data Marts



- RDA (plus Enterprise Model Extender)
  - Генерация ХД, витрин/кубов
  - Инкрементальная, проектная, корпоративная разработка
  - Трассировка бизнес требований до уровня кода
  - Сравнение, объединение моделей

# Фаза реализации



- Определение целевых структур
- Общий глоссарий
- Инкрементальная разработка без переписывания

# Анализ требований

Model Set Explorer

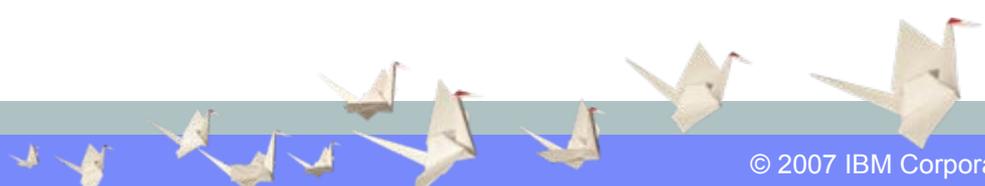
- Aggregate Measures
- Analysis
- Core Business Concepts
- Focus Areas
  - BDW Asset & Liability Management
  - BDW Profitability
    - Activity Based Costing Analysis
    - Business Procedure Performance Measurement
    - Channel Profitability
    - Customer Lifetime Value Analysis
      - Asset Dealing Segment Type
      - Communication Form
      - Customer Market Segment
      - Customer Relationship Age Segment
      - Individual Age Group
      - Line Of Business Reporting Group
      - Measurement Currency
      - Organization Unit Function
      - Reporting Currency
      - Scenario
      - Time Period
      - Customer Lifetime Value
      - Total Acquisition Cost
      - Total Customer Net Present Value**
      - Total Number Of Credit Bureau Inquiries
      - Total Projected Revenue Growth

Properties

Asset Explorer Report Explorer

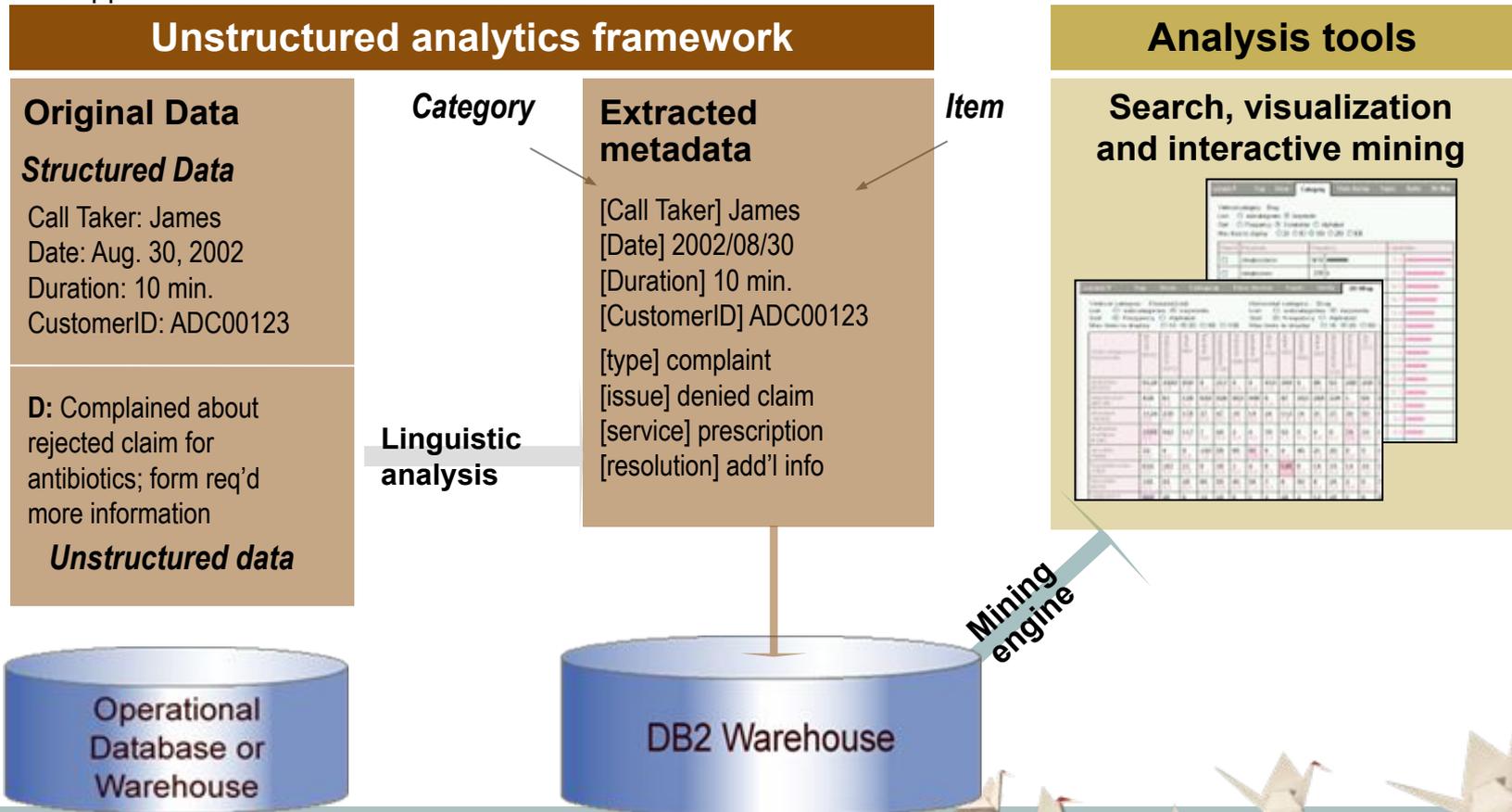
<Measure> Total Customer Net Present Value

Dependency	Reason
Total Customer Net Present Value	
Customer Lifetime Value Analysis	Container
Total Customer Net Present Value	Model semantics
Aggregate Descriptors	Container
Net Present Value	Model semantics

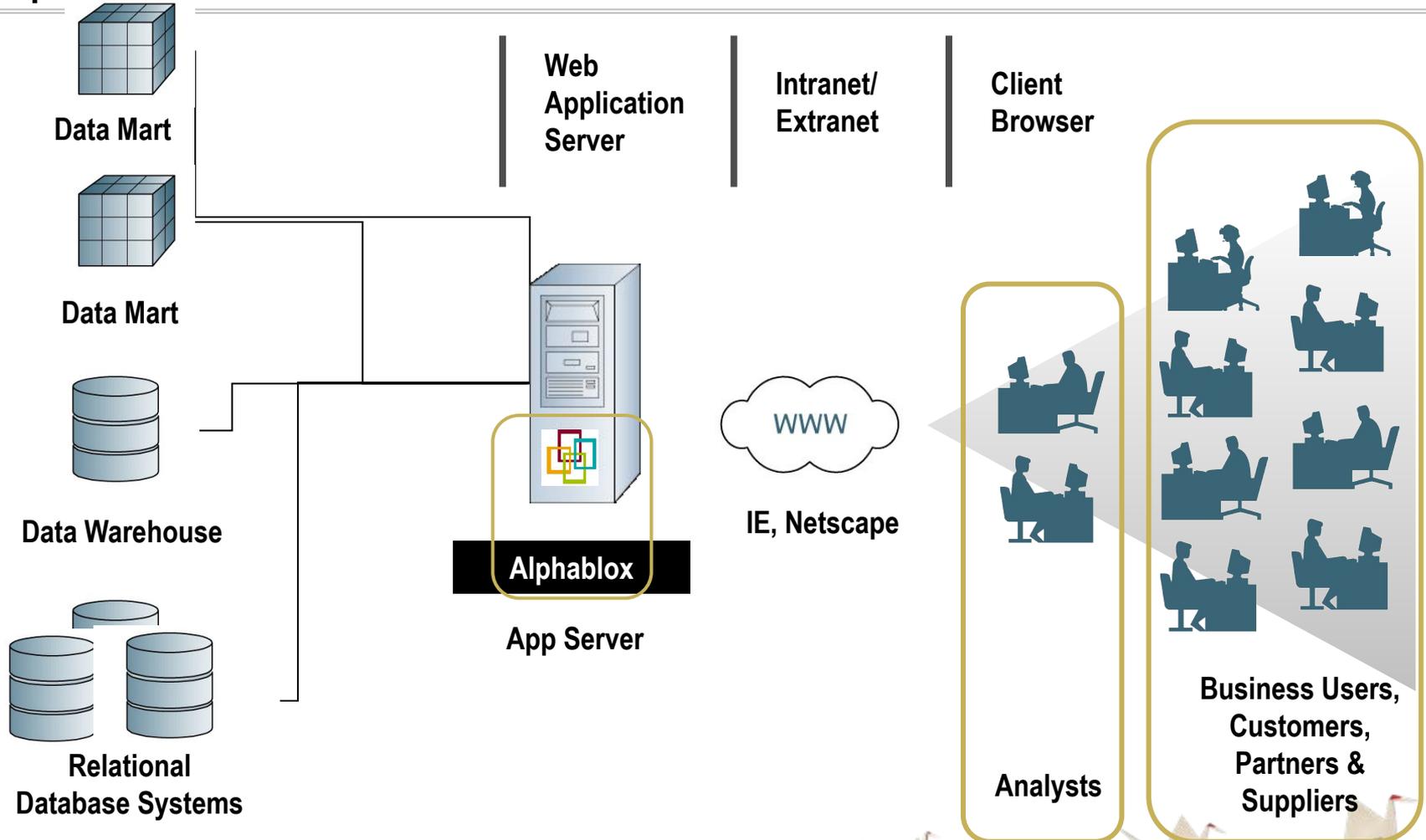


# Introducing IBM OmniFind Analytics Edition

- Обширные интерфейсы для объединения структурированных и не структурированных данных
- Объединение поиска и текстовой аналитики



# Alphablox

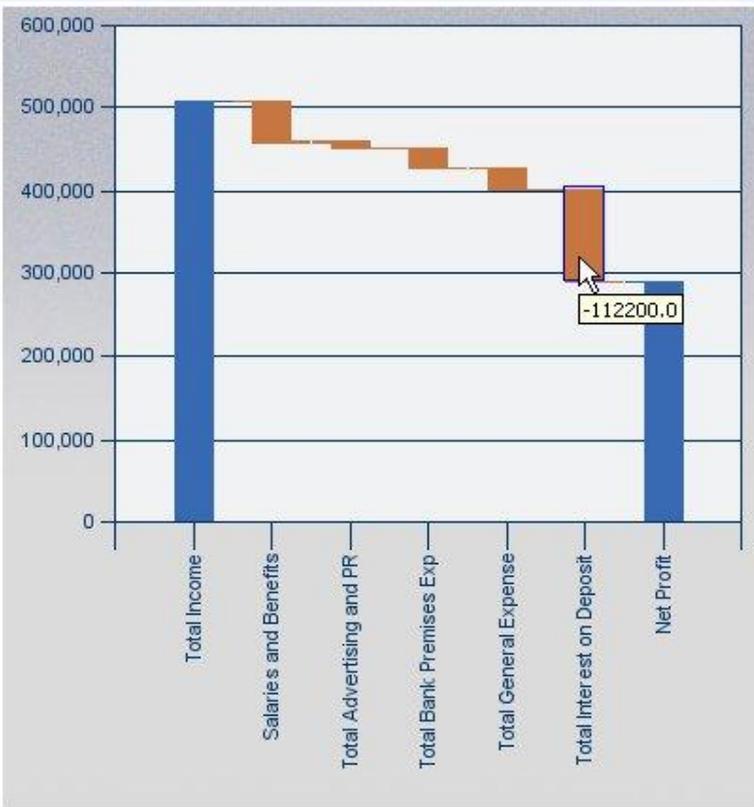




Welcome Betty!

- Home
- Branch View
  - Financials
    - Income Statement
  - Metrics
    - New Accounts
    - New Deposits
    - Ranking
  - Budget
    - Adjust Budget

	Actual	Budget	Variance
October			
Income	\$61,000	\$80,000	\$19,000
Expenses	\$64,000	\$80,000	\$16,000
Net Income	\$49,000	\$80,000	\$31,000
Operating	\$85,000	\$90,000	\$5,000
Non-Operating	\$76,500	\$80,000	\$3,500
Other	\$60,000	\$105,000	\$45,000
Total	\$111,500	\$130,000	\$18,500
Operating	\$507,000	\$645,000	\$138,000
Non-Operating	\$48,000	\$80,000	\$32,000
Other	\$5,800	\$55,100	\$49,300
Total	\$24,000	\$110,000	\$86,000
Operating	\$27,900	\$70,000	\$42,100
Non-Operating	\$55,000	\$45,000	-\$10,000
Other	\$56,000	\$45,000	-\$11,000
Total	\$400	\$800	\$400
Operating	\$800	\$1,200	\$400
Non-Operating	\$112,200	\$92,000	-\$20,200
Other	\$217,900	\$407,100	\$189,200
Total	\$289,100	\$237,900	-\$51,200
Operating	\$1,019,000	\$1,290,000	\$271,000



User Details

Country: US  
 Region: West  
 State: CA  
 City: San Francisco  
 Branch: 012  
 User: Branch, Betty A  
[Logout](#)

AF News

9:02am EST  
**AF Reports - Rev Increases 76% - WW**  
**Trades Grow 62%**

**NEW!** You have **1** NSF items for review. [Go...](#)



IBM Information Management

Thank You

