

## **BI: przegląd, ETL, raportowanie i analizy**

**2013-04-12**

**Krzysztof Bokiej, Łukasz Limanowski, Mariusz Pyka**

**Roche**



## **Wstęp**

**2013-04-12**

**Krzysztof Bokiej, Roche**

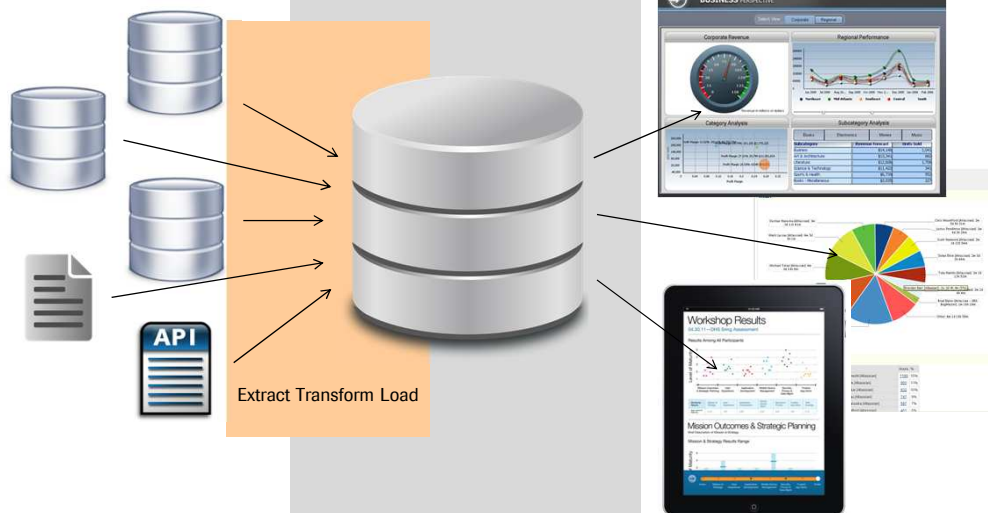
## Architektura typowego środowiska DW/BI



Systemy Źródłowe

Hurtownia Danych

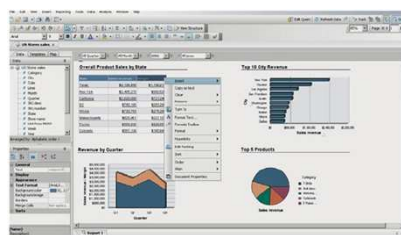
Warstwa prezentacji



## Trendy w DW / BI



- Self Service BI i warstwa semantyczna
- Urządzenia mobilne
- HTML5 i JavaScript
- In Memory Analysis (SAP Hana, Oracle Times ten, Qlick View, )
- Big Data i Cloud Computing (Hadoop, Google Big Querv)
- Exadata i bazy kolumnowe



## Innowacja i nowe technologie



- Monitorowanie rynku i „trzymanie ręki na pulsie”
- Testowanie nowych narzędzi
- Proof of Concept
- Przydatność z punktu widzenia korporacji



QlikView



## BI – BACKEND: ETL czyli Extract – Transform – Load w praktyce

2013-04-12

*Lukasz Limanowski, Roche*



# GOD W

czyli **G**lobal **O**racle **D**ata **W**arehouse

- hurtownia danych w Roche

## Hurtownia – zawartość ( wybrane przykłady )

- badania kliniczne ( historia, stan aktualny + prognozy )
- planowanie projektów farmaceutycznych, projekty związane z developmem, badaniami klinicznymi itp., ( historia, stan aktualny )
- finanse: zmienne koszty badań klinicznych ( historia, stan aktualny + prognozy )
- CRM – m.in. aktywność przedstawicieli handlowych
- dane o lekach (użycie, percepcja itp. )
- i inne ...

DECYZJE



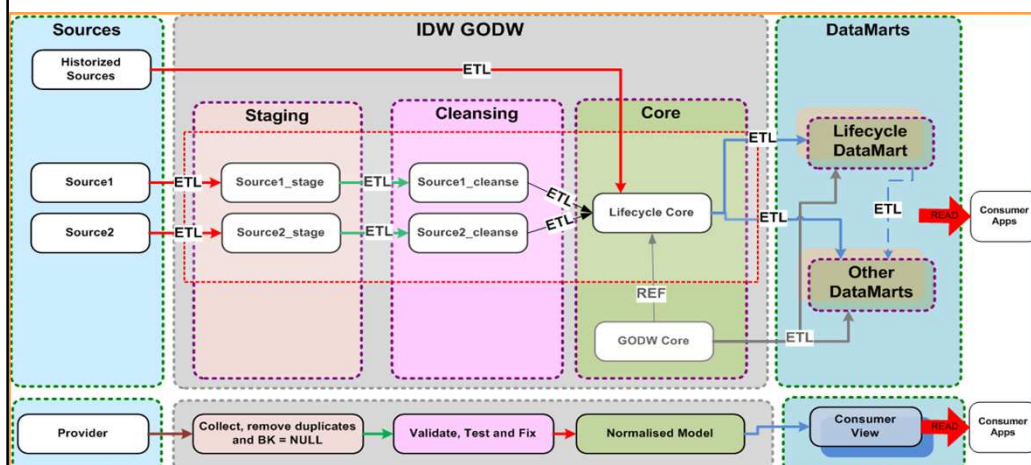
## Hurtownia – liczby (ETL)



- > **50** data martów
- > **90** folderów tematycznych
- > **5 TB** danych
- > **32 000** dziennych sesji ETL na PROD

9

## Proces zasilania: SRC - MART (1)



10

## ETL – dostępne narzędzia (1)



1. Rozwiązania **hardcore**: skrypty / procedury SQL

**Zalety:** dostępność, brak dodatkowych kosztów

**Wady:** czasochłonna, prymitywna metoda, kody w postaci tekstowej, trudne w serwisowaniu, analizie, bardzo utrudniony dostęp do funkcji dodatkowych (parsowanie XML, FTP, Java, bash itp.), bardzo trudne monitorowanie, współdzielenie kodu itp.

**Uwagi:** raz zaimplementowane i ,wygrzane’ rozwiązanie cechuje się wyjątkową żywotnością i często wielokrotnie przekracza planowany czas migracji na nowo zakupione i bardzo drogie narzędzie ETL. Sama migracja dostarcza wielu zmartwień osobom, które się jej podjęły.

11

## ETL – dostępne narzędzia (2)



2. Profesjonalne narzędzia ETL, m.in.: Informatica PowerCenter, AbInitio, Oracle Warehouse Builder, MS Integration Services

**Wady:** koszt

**Zalety:** szybkość implementacji, kod ,obrazkowy’ – łatwy w serwisowaniu i analizie, mnogość dodatkowych funkcji, predefiniowane **połączenia do różnych źródeł**, np.: Facebook, Twitter, LinkedIn, SAP, DB2, pliki binarne itd.

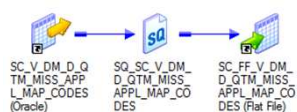
Gotowe rozwiązania dla wersjonowania, monitorowania, migracji pomiędzy środowiskami, współdzielenia kodu.

12

## ETL – dostępne narzędzia (2 - c.d.)



**Uwagi:** u niektórych osób zaobserwowano nieodpartą pokusę implementacji rozwiązań hardcore w narzędziach ETL. Najczęściej takie przypadki są bezlitośnie zwalczane przez architektów.



13

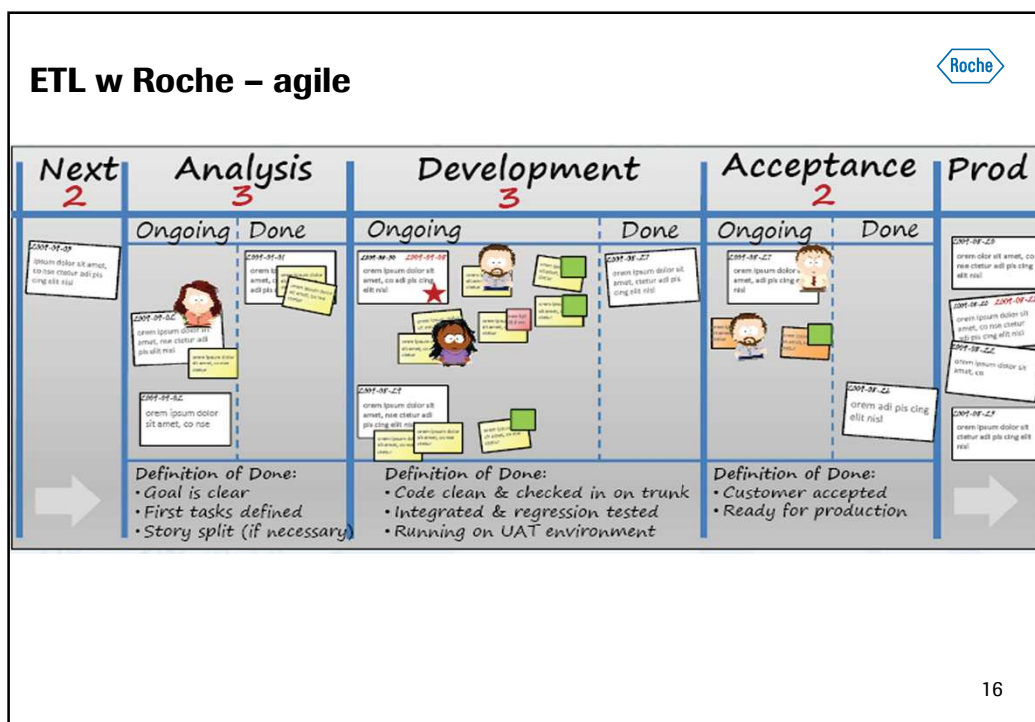
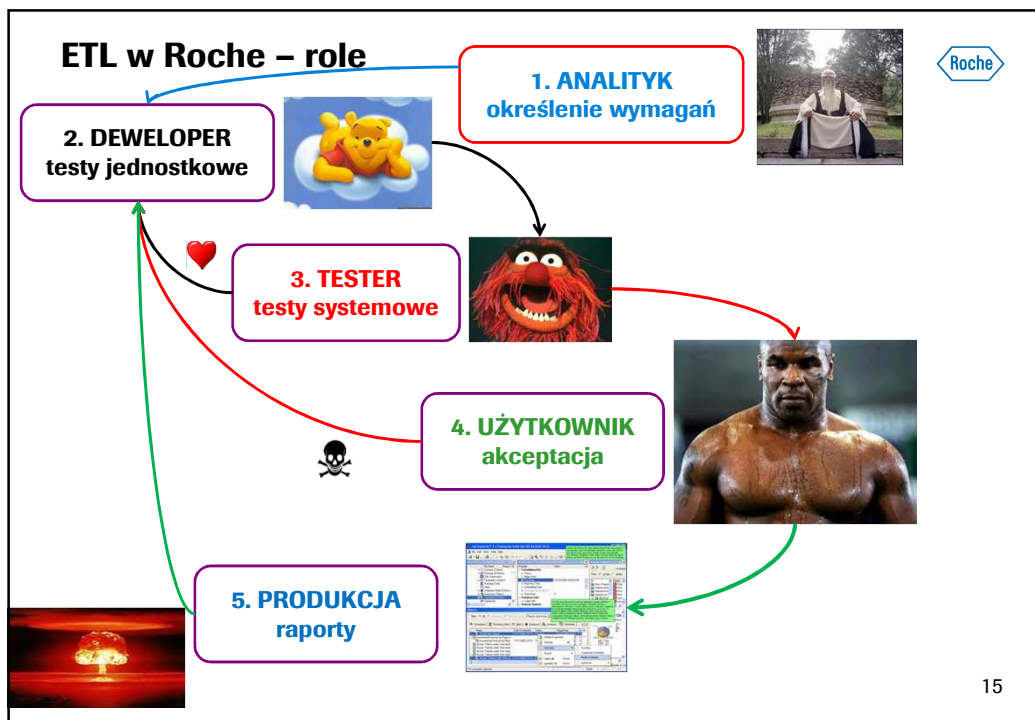
## ETL w Roche – Informatica Power Center



Obecnie niezwykle popularne narzędzie (w przeciwieństwie np. do AbInitio):

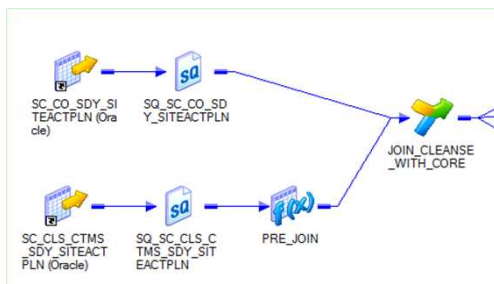
- 2000+ klientów (80% z listy Fortune 100),
- partnerstwo strategiczne z IBM, HP, Accenture, SAP
- stale rozwijane, często kosztem konkurencyjnych firm (przejęcia), szybko reagujące na najnowsze trendy – np. Big Data

14





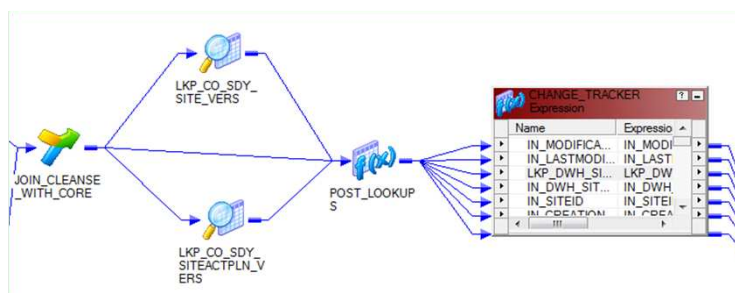
## Przykładowa implementacja ETL: SCD2



1. full outer join porównujący nowy wsad (CLEANSE) ze stanem z poprzedniego zasilania (CORE)

17

## Przykładowa implementacja ETL: SCD2



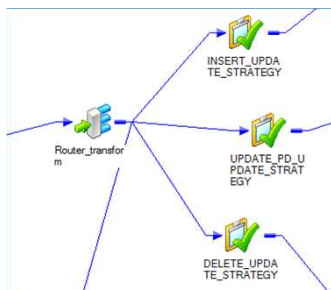
2. Lookup'y dodają dane referencyjne oraz numer aktualnej wersji - będzie nam potrzebny do ewentualnej aktualizacji danych w tabeli wersji.

CHANGE\_TRACKER – w nim wyliczane są flagi:

insert / update / delete na podstawie danych otrzymanych z full outer join.

18

## Przykładowa implementacja ETL: SCD2

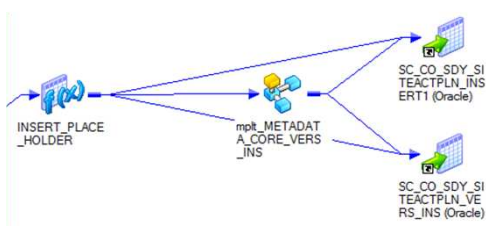


3. Router na podstawie flag wyliczonych w kroku 2 rozdziela strumień danych na 3 ścieżki:

insert / update / delete.

19

## Przykładowa implementacja ETL: SCD2

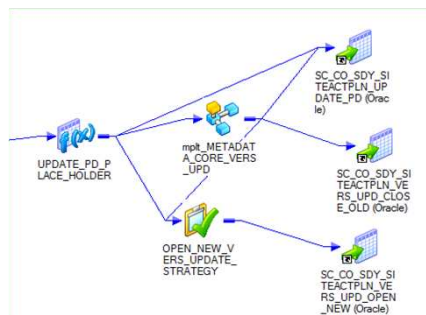


4. Ścieżka insert (rekord istnieje w CLEANSE, brak w CORE):

rekord zostaje dodany do tabeli CURR (dane bieżące) oraz VERS (dane historyczne). Mplt\_% dodaje do rekordu klucze hurtowniane (surrogate keys), które odtąd będą używane do złączeń z wykorzystaniem tych tabel.

20

## Przykładowa implementacja ETL: SCD2



5. Ścieżka update (rekord istnieje w CLEANSE i CORE, porównanie rekordów CLS-CO wykazało różnicę – dane zostały zaktualizowane w systemie źródłowym):

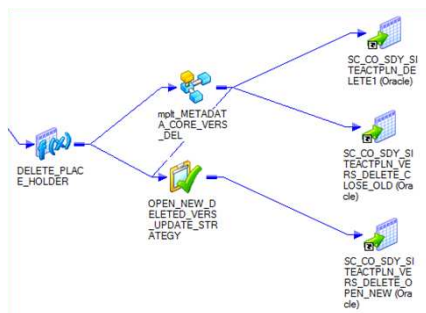
Tabela CURR zostaje zaktualizowana.

Tabela VERS: stara wersja zostaje zamknięta przez aktualizację pola VALID\_TO(date), pozostałe wartości bez zmian.

Zostaje otwarta nowa wersja (jak insert) z VALID\_TO=9999-12-31

21

## Przykładowa implementacja ETL: SCD2



6. Ścieżka delete (rekord istnieje w CORE, brak w CLEANSE, rekord został usunięty z systemu źródłowego):

Aktualizacja flagi: deleted=Y w tabeli CURR.

Tabela VERS: stara wersja zostaje zamknięta przez aktualizację pola VALID\_TO(date), pozostałe wartości bez zmian.

Zostaje otwarta nowa wersja (jak insert) z VALID\_TO=9999-12-31 oraz deleted=Y

22

## Raportowanie i analizy

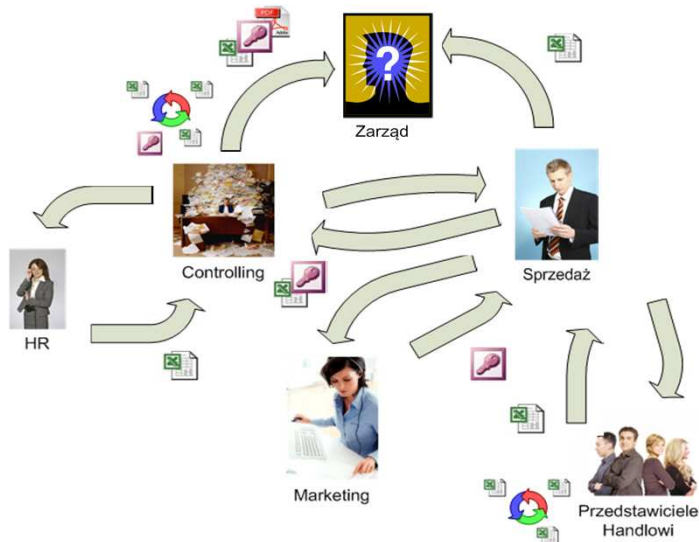
*2013-04-12*

*Mariusz Pyka, Roche*

### Agenda

- Typowy model dystrybucji raportów i analiz
- Zdefiniowanie problemów
- Docelowy model dystrybucji raportów i analiz
- Rola warstwy semantycznej
- Komponenty platformy analitycznej (SAP Business Objects)
- Scenariusze wykorzystania (Prezentacja Live)
- Raportowanie i analizy - trendy

## Model dystrybucji raportów i analiz



## Zdefiniowanie problemów

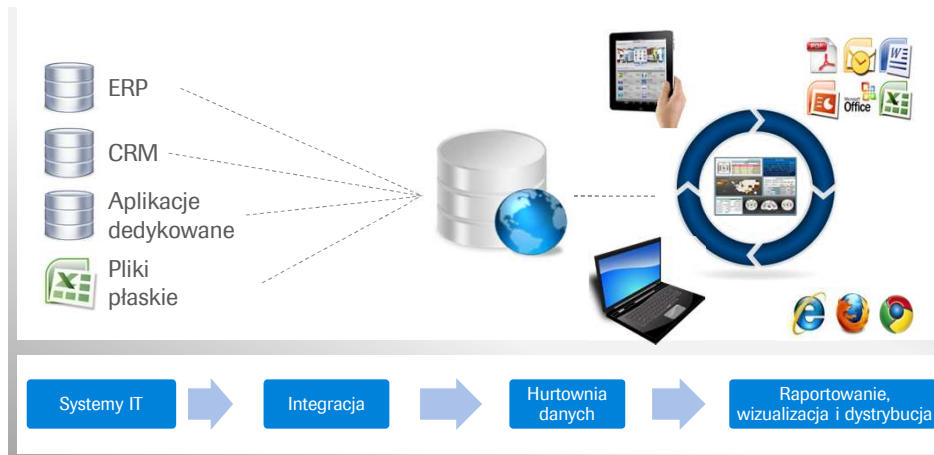


- Wykorzystanie wielu **niespójnych** narzędzi / formatów raportów
- Konieczność **manualnej** integracji / agregacji danych na raportach
- Brak zunifikowanego modelu **dystrybucji raportów**
- **Uzależnienie wiedzy** na temat procesów od wiedzy użytkownika
- Wykorzystanie metod podatnych na **błędy**
- Przywiązanie użytkowników do narzędzi

„47% użytkowników systemów IT nie ma zaufania do otrzymywanych informacji.”

**PriceWaterhouseCoopers**

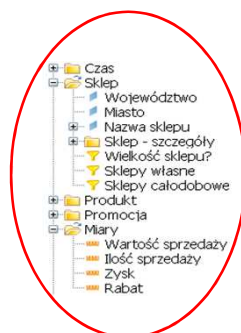
## Model dystrybucji raportów i analiz



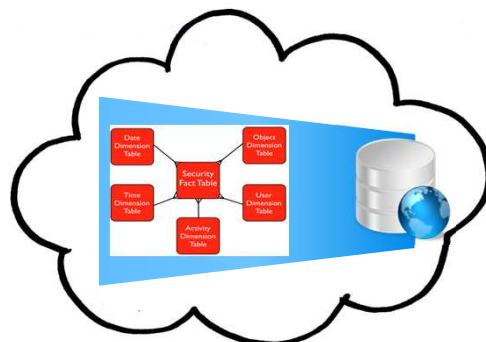
## Koncepcja świata obiektów



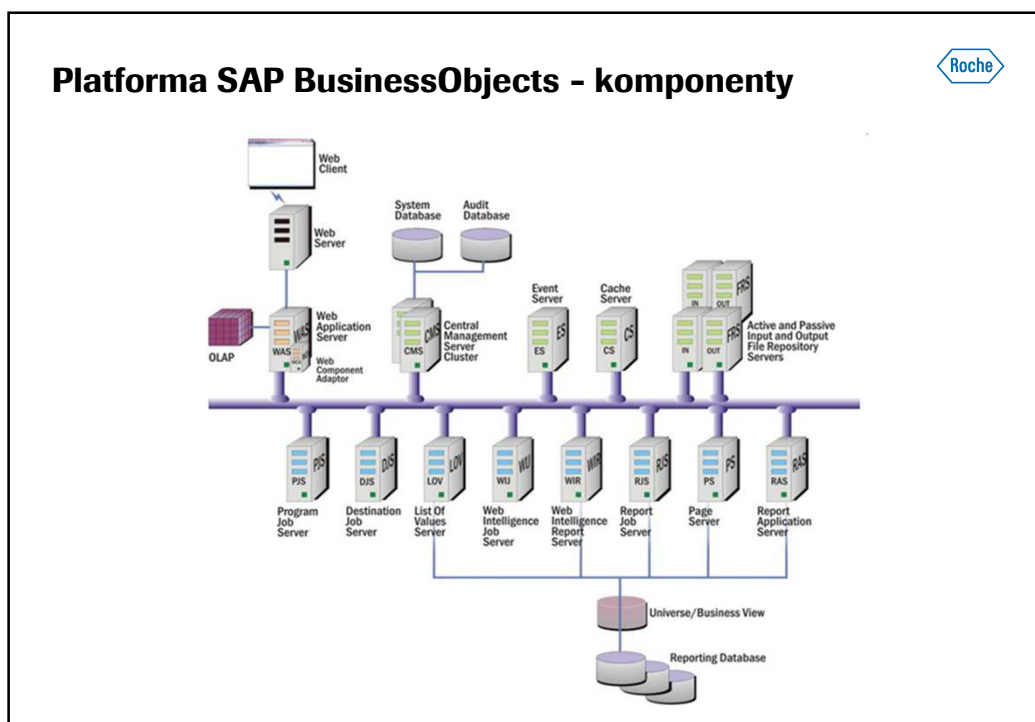
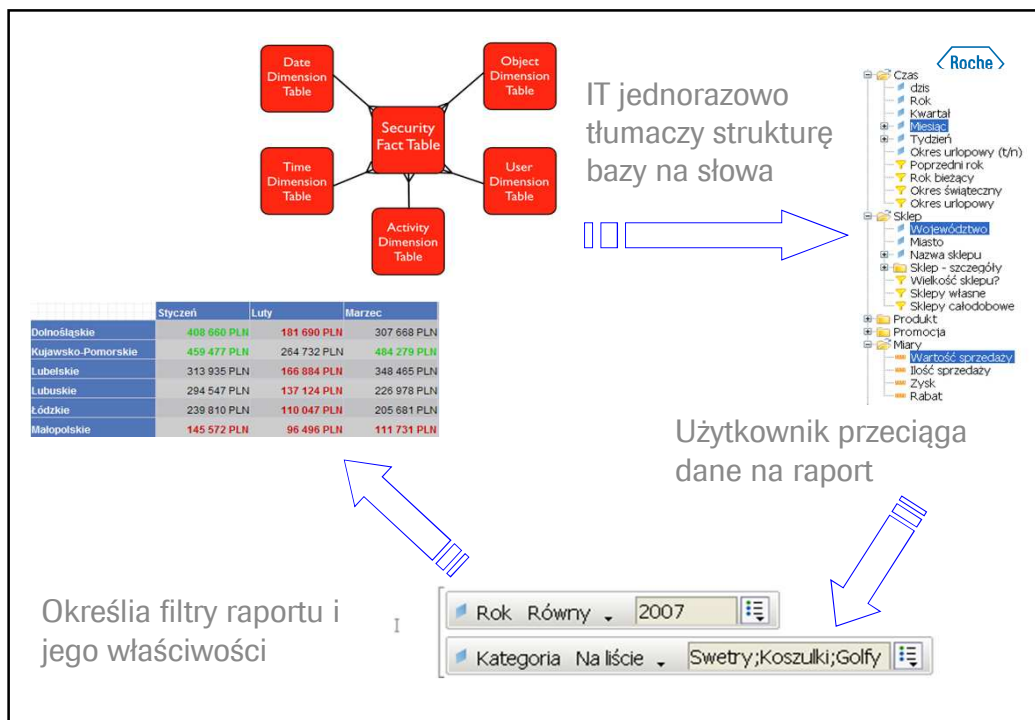
Użytkownicy  
biznesowi



Świat  
obiektów



Hurtownia  
danych



## Prezentacja Live



Przedstawiciel Handlowy

### Odpowiedzialność

- Prowadzenie działań terenowych
- Szybki dostęp do bieżącej informacji na temat własnych działań/wyników

### Systemy ewidencyjne

- Sprzedaż
- CRM

### Metody dostępu

- Urządzenia mobilne
- E-mail
- Wersje offline



Dyrektor Regionalny

### Odpowiedzialność

- Zarządzanie codziennym funkcjonowaniem działu handlowego
- Szybki dostęp do bieżącej skonsolidowanej informacji na temat działań/wyników działu

### Systemy ewidencyjne

- Sprzedaż
- Planowanie/budżetowanie
- CRM

### Metody dostępu

- Urządzenia mobilne
- Portal BI
- E-mail



Kadra Zarządzająca

### Odpowiedzialność

- Ustalanie strategii przedsiębiorstwa
- Szybki dostęp do skonsolidowanej informacji na temat wyników przedsiębiorstwa

### Systemy ewidencyjne

- Finance/Controlling
- Sprzedaż
- Planowanie/budżetowanie
- CRM

### Metody dostępu

- Urządzenia mobilne
- E-mail

## Raportowanie i analizy - trendy



Urządzenia mobilne

Wizualizacja na mapach

Self-service BI

Analizy predykcyjne

Rozwiązania „w chmurze”



## Projekt: Standaryzacja rozwiązania raportowego do wizualizacji na mapach



- Cel
  - Stworzenie uniwersalnego rozwiązania raportowego pozwalającego na wizualizację danych biznesowych na interaktywnych mapach geograficznych



## Projekt: Standaryzacja rozwiązania raportowego do wizualizacji na mapach



- Zadania
  - Zdefiniowanie konfiguracji POI w MDMS i/lub Salesforce.com
  - Stworzenie analitycznego modelu danych
  - Stworzenie modułu integracji (zasilenie modelu danych na podstawie konfiguracji)
  - Integracja z wybranym procesem biznesowym
  - Stworzenie modułu raportowego (świat obiektów, raporty i kokpity)
- Narzędzia/Technologie
  - Oracle DB, Salesforce.com, MDMS
  - SAP Business Objects (Designer, Web Intelligence, SAP Mobile, Dashboard Design)



*Doing now what patients need next*