

# Rozwiązania typu Business Intelligence

*Katedra Inżynierii Oprogramowania  
Wydział Elektroniki, Telekomunikacji i Informatyki  
Politechnika Gdańska*

**Teresa Zawadzka**

[togra@eti.pg.edu.pl](mailto:togra@eti.pg.edu.pl)

**Krzysztof Goczyła**

[kris@eti.pg.edu.pl](mailto:kris@eti.pg.edu.pl)

Wydział Elektroniki, Telekomunikacji i Informatyki

- 1 -

Dzień dobry,

Gdybym te wykłady Państwu prezentowała osobiście byłby teraz czas na przedstawienie siebie. Jednak moje imię i nazwisko mogą Państwo przeczytać na slajdzie – ja to Teresa Zawadzka. Wszystkie informacje dodatkowe mogą Państwo znaleźć na stronie Wydziału – między innymi jak mnie znaleźć osobiście i telefonicznie.

## Data Warehouses

### Introduction to *Business Intelligence* systems



- 2 -

Dzień dobry,

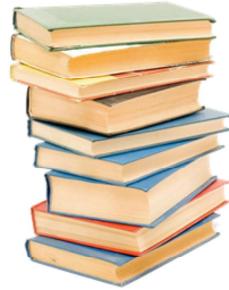
Gdybym te wykłady Państwu prezentowała osobiście byłby teraz czas na przedstawienie siebie. Jednak moje imię i nazwisko mogą Państwo przeczytać na slajdzie – ja to Teresa Zawadzka. Wszystkie informacje dodatkowe mogą Państwo znaleźć na stronie Wydziału – między innymi jak mnie znaleźć osobiście i telefonicznie.

## Literature

**W. H. Inmon:** „Building the Data Warehouse”. J. Wiley&Sons, 2002.

**R. Kimball:** „The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling, Third Edition”. J. Wiley&Sons, 2013.

P. Ponniah: „Data Warehousing”. J. Wiley&Sons, 2001.



- 3 -

Jeśli wykład dla Państwa nie jest jasny lub chcą Państwo rozszerzyć swoje horyzonty zawsze można zająrzyć do literatury. Ja w swoim wykładzie bazuję przede wszystkim na pozycji R. Kimball'a.

## Introduction to *Business Intelligence* systems

**What is the BI system?**

**The wizard, or not the wizard?**



- 4 -

Lubię Państwo książki J.R.R. Tolkiena? Ja bardzo. Moje ulubione postacie z tej opowieści zmieniały się, jak stawałam się coraz starsza, ale Gandalfa lubię do dzisiaj. Zastanawiają się pewnie Państwo, o czym to ja właściwie piszę. Otóż to. Czy Gandalf może mieć coś wspólnego z systemami typu Business Intelligence (BI). Systemy BI, mówiąc skrótnie, mają za cel (jak sama nazwa wskazuje) pomagać w podejmowaniu decyzji biznesowych. A Gandalf? No właśnie... co robi Gandalf?

## Advisor, informant, executor?



5

- 5 -

Jaka jest jego rola?

Przypomnijmy sobie kilka zdarzeń z książki.

Pierwsze: Gandalf przyjeżdża do władcy Rohanu - Theodena, wyzwala go spod wpływu Grimy (a właściwie Sarumana) i prosi o pomoc dla Gondoru. Nie żąda tej pomocy, lecz decyzję o jej udzieleniu pozostawia królowi.

Drugie: Faramir z rozbitymi wojskami powraca z Osgiliath, atakują go Nazgule. Gandalf ratuje Faramira.

Trzecie: Gandalf rozmawia z Denetorem o obronie Minas Tirith.

Kim zatem jest Gandalf: doradcą, informatorem, egzekutorem? Czy umieją Państwo określić jego rolę w każdym z tych wydarzeń? Czy jest ona niezmienna? Pokuśmy się o zdefiniowanie jego najważniejszej roli. Ja stawiam na doradcę (informatora) zdecydowanie bardziej niż egzekutora. A Państwo?

Jeśli ktoś nie czytał książki niech się nie martwi, nie trzeba jej znać, aby zaliczyć kurs. Przeczytać jednak polecam - warto.

## Who makes the decision?



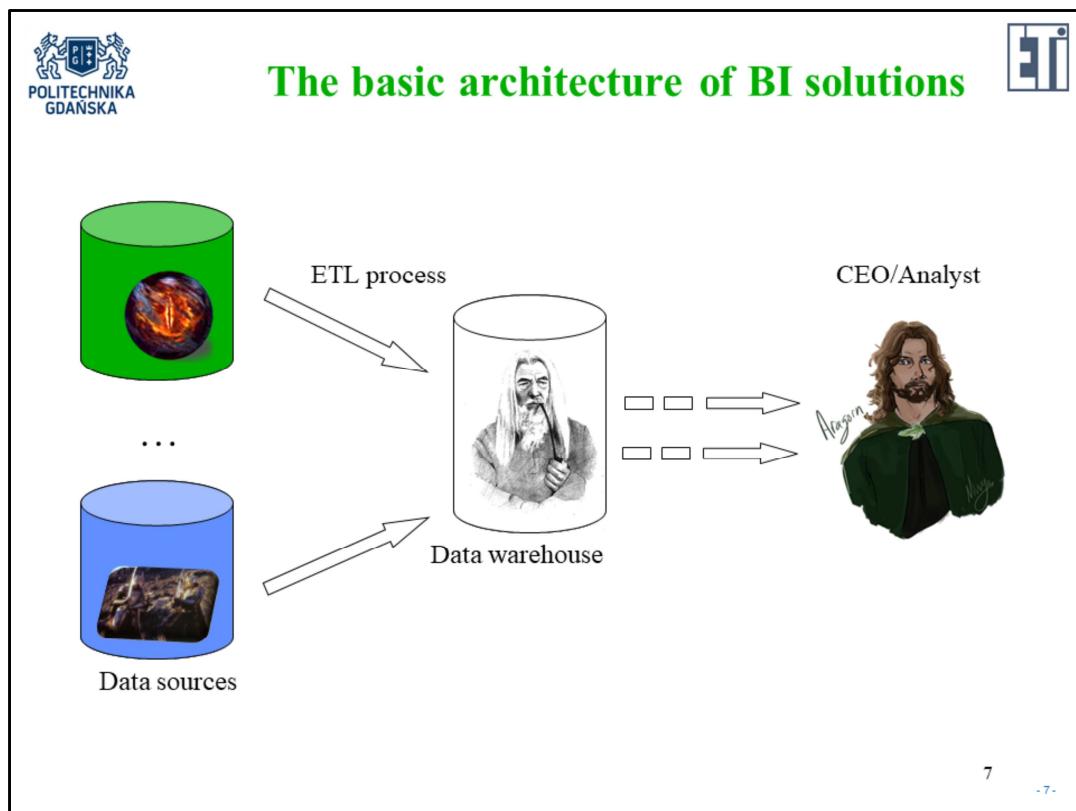
6  
- 6 -

Jeśli już zgodziliśmy się, że Gandalf to przede wszystkim doradca (jeśli ze mną się Państwo nie zgodzili to trudno, niestety muszą Państwo przyjąć mój punkt widzenia :-) to spróbujmy odpowiedzieć sobie na pytanie. Skąd Gandalf tyle wie? Na pewno wiele czytał o dawnych wydarzeniach, poszukiwał faktów i tych aktualnych i tych, które wydarzyły się w zamierzchłej przeszłości. Czy zaglądał do palantira? Nie pamiętam... Aby móc doradzać, informować, zbierał informacje z różnych źródeł i na ich podstawie doradzał. Doradzał niziołkom, władcom i królom.

Proszę spojrzeć na obrazek. W jego centrum znajduje się Gandalf. Zbiera on informacje w czasie (co symbolizuje zegar słoneczny) na temat różnych wydarzeń i osób. Zbiera je i przekazuje innym, tutaj wybierzymy najważniejszą (rangą) postać - samego króla Aragorna.

Może Państwo nie zauważycie, ale właśnie zdefiniowaliśmy system typu BI.

## The basic architecture of BI solutions



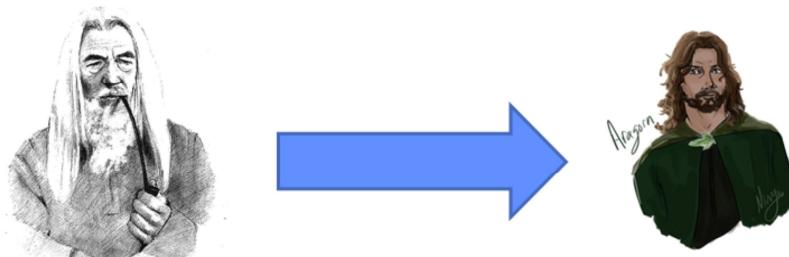
7

- 7 -

Składa się on z różnych źródeł danych, które następnie są pobierane, transformowane i ładowane (procesem ETL – *Extract, Transform, Load*) do hurtowni danych – Gandalfa. Hurtownia danych zaś za pomocą raportów, czy pulpitów menadżerskich przedstawia dane prezesowi czy analitykom (ogólnie mówiąc osobom decyzyjnym – tutaj Aragornowi).

## Who knows?

- The events?
- The connections between events?



8

- 8 -

W naszym toku rozumowania Gandalf reprezentuje hurtownię danych, ale ... No właśnie jest jedno ale. Z czym kojarzy się Państwu słowo hurtownia? Mi z faktem, że czegoś jest dużo, a nawet bardzo dużo? Ale czego tak naprawdę w hurtowni danych jest dużo? Danych? Tak – danych, ale jakich? Czy książka telefoniczna z danymi wszystkich abonentów na świecie to już hurtownia danych, czy tylko duża baza danych? Słownem kluczem jest tutaj słowo „zdarzenie”. R. Kimball pewnie by się ze mną zgodził, ale czy B. Inmon również (dla zainteresowanych Ci dwaj panowie to ojcowie hurtowni danych i każdy z nich hurtownię danych postrzega trochę inaczej, ale o tym później). Podsumowując hurtownię danych to coś takiego co zna zdarzenia i powiązania między tymi zdarzeniami.

## The warehouse of data or events?



9

- 9 -

Może wtedy lepiej byłoby nazwać hurtownię danych, hurtownią zdarzeń? Ciekawi mnie na ten temat Państwa zdanie, ale dopiero wtedy, kiedy ukończą Państwo kurs.

NO

**Gandalf can think!!!**

10

- 10 -

Podobała się Państwu analogia? Mi bardzo, tylko że jest zupełnie niewłaściwa. Dlaczego? Hurtownia danych nie potrafi myśleć i wyciągać wniosków. Zawsze musi zrobić to człowiek. Chociaż, kiedy pomyślę o coraz bardziej rozwijających się systemach eksploracji danych, czy systemach rekomendacyjnych to już tak bardzo do swojego stwierdzenia nie jestem przekonana.

Systems for supporting strategic decisions for the company  
**(DSS, Decision Support System; BI, Business Intelligence):**

- easy access to the information
- powerful analytical tools
- efficient processing of vast amounts of data
- user friendly presentation of consistent results
- ready for changes
- focused on business processes



- 11 -

A teraz skupmy się na cechach systemu typu BI.

Osoby decyzyjne wymagają **łatwego dostępu do informacji**, przekazanej w zrozumiałym dla nich sposób. Zresztą nie tylko osoby decyzyjne. Czy lubią Państwo zajęcia, na których niewiele rozumieją, a każdy slajd to niezrozumiałe i niewytłumaczone pojęcia? System typu BI to **potężne narzędzie analityczne**. Oznacza to, że musi być ukierunkowane na analizowanie danych (co to dokładnie oznacza opowiem na następnym slajdzie).

Ze względu na to, że systemy te bazują na ogromnej ilości danych – zbieranych z różnych źródeł w długim okresie czasowym (Gandalf sięgał nawet do informacji z czasów pierwszej wojny z Sauronem) - muszą wydajnie przetwarzać **duże zbiory danych**. Ponadto, dane muszą być prezentowane w sposób **czytelny i łatwy do zinterpretowania**. Co łatwiej odczytać tabelkę z milionem danych, czy dobrze przygotowany wykres? (Wizualizacja danych to bardzo ważny temat w hurtowniach danych, ale potrzeba oddzielnego kursu, aby pokazać jak dobrze prezentować dane. Może kiedyś go przygotuję. :-). System musi być **gotowy na zmiany**. Wymagania od strony biznesu są nieprzewidywalne. Dzisiaj prezes rozwiązuje problem jednego typu, a jutro zupełnie innego. System typu BI jest **zorientowany na procesy biznesowe** zachodzące w firmach czy instytucjach, w których jest wdrożony.

1. **Checking for existence**, e.g. “whether given product was sold to given customer”?
2. **Comparison**, e.g. “compare purchases of two customers within the last 6 months” or “compare number of items sold based on category” .
3. **Trend analysis**, e.g. “whether within the last 12 months the trend of sales of given product is rising?”.
4. **Rankings**, e.g. “give the 10 bestsellers within last year”.
5. **Statistical analysis**, e.g. “give the average income for the last 6 months”.



- 12 -

Obiecałam, że na tym slajdzie wytłumaczę, co oznacza przetwarzanie danych ukierunkowane na ich analizę. W hurtowniach danych jest to 5 podstawowych klas zapytań, na które narzędzie analityczne musi potrafić odpowiedzieć. Świadomie zaczniemy od punktu 2. To zapytania typu porównawczego. Analiza zakłada porównywanie: pracowników, jakości kształcenia, zarobków, ocen, itd. Punkt 3 to analiza trendu w czasie: czy uczę się coraz lepiej, czy coraz gorzej? Rankingi – punkt 3 to: najlepsze filie, najlepsi pracownicy, dzieci najlepiej uczące się. Analiza statystyczna (punkt 5) dotyczy różnego rodzaju średnich, kwantylów i innych narzędzi statystycznych. Wracając teraz do pominiętego punktu 1. Hurtownia danych potrafi również sprawdzić, czy dane zdarzenie zaistniało. Jednak, ze względu na fakt, że na takie pytanie odpowiadają wydajnie również relacyjne bazy danych, to ta klasa zapytań nie jest charakterystyczna dla systemów typu BI.

Data processing for trend analysis, cross-sectional analysis and other strategic analysis.

Examples:

<https://www.sisense.com/dashboard-examples/>

<https://www.datapine.com/live-dashboards>

**A Key Performance Indicator (KPI)** is a measurable value that demonstrates how effectively a company is achieving key business objectives.

- 13 -

Podsumowując przetwarzanie analityczne (z angielskiego *Online Anlaytical Processing – OLAP*) to przetwarzanie danych, którego celem są analizy trendów, analizy przekrojowe i inne analizy o charakterze strategicznym.

Jak takie analizy wyglądają w rzeczywistości? Warto podpatrzeć kilka przykładowych pulpitów menadżerskich/analitycznych (można spotkać również określenie kokpit menadżerski/analityczny). Pulpit menadżerski/anallityczny (ang. *dashbord*) jest specyficzny rodzajem raportu, na którym najważniejsze informacje i wskaźniki powiązane z celami firmy lub organizacji przedstawione są w formie wizualnej. Informacje te są umieszczone na jednym ekranie lub jednej kartce. Pulpit menadżerski jest sposobem obrazowania i prezentacji danych (w tym danych biznesowych), a nie specyficzny rodzajem informacji lub technologii [www.skuteczneraporty.pl, wrzesień 2017 r.].

Państwa zadanie: proszę przejrzeć podane linki z przykładowymi pulpitami menadżerskimi. Czy udało się Państwo zidentyfikować 5 przykładowych wykresów, jakiego są one typu? Czy znaleźli Państwo przynajmniej na jednym pulpicie menadżerskim wizualizację opisaną skrótem KPI?

Proszę zapoznać się z kilkoma przykładami:

<https://www.klipfolio.com/resources/kpi-examples/sales/sales-growth>

<https://www.klipfolio.com/resources/kpi-examples/healthcare>

Business process is a collection of related, structured activities or tasks that produce a specific service or product.

1. Business processes are typically expressed by verbs because they represent activities within organization.
2. Business processes are often supported by systems e.g. billing system.
3. Business processes generate numerical metrics, which are being analysed.

- 14 -

Inną cechą przetwarzania analitycznego jest jego ukierunkowanie na procesy biznesowe. Na temat samych procesów biznesowych więcej ode mnie wiedzą pewnie menadżerowie i ekonomiści. To, czego potrzebujemy do celów tego wykładu to fakt, że każdy proces biznesowy to zbiór powiązanych ze sobą aktywności, których celem jest osiągnięcie jakiegoś celu, często dostarczenia jakieś usługi lubtworzenia produktu.

	OLTP	OLAP
Volume		
History		
Number of users		
Query complexity		
Normalization		
Addition		
Removal		
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 15 -

Zdefiniowałam, czym jest przetwarzanie OLAP, a czy Państwo pamiętają, czym jest przetwarzanie transakcyjne? Jeśli nie, to odsyłam do wykładów z Baz danych. Oczekuję, że poniżej wymienione zagadnienia są Państwu bardzo dobrze znane:

1. Relacja i model relacyjny
2. Normalizacja
3. Transakcje, ACID

Na pewno, zanim przystąpią Państwo do rozwiązywania quizu warto zainwestować czas w krótkie przypomnienie.

Na kolejnych slajdach porównamy przetwarzanie analityczne z przetwarzaniem transakcyjnym. Porównanie to zrobimy w oparciu o takie metryki jak:

Wielkość (ang. *volume*) – jaka wielkość danych jest przetwarzana?

Historia (ang. *history*) – czy dane historyczne są przetwarzane podczas wykonywania zapytań?

Liczba użytkowników (ang. *number of users*) – ile użytkowników przetwarza dane?

Złożoność zapytań (ang. *query complexity*) – jak bardzo złożone są zapytania?

Normalizacja (ang. *normalization*) – czy dane są znormalizowane?

Dodawanie (ang. *addition*) – czy dane są dodawane?

Usuwanie (ang. *removal*) – czy dane są usuwane?

Aktualizacja (ang. *updates*) – czy dane są aktualizowane?

Tranzakcyjność (ang. *transactions*) – czy dane są przetwarzane w ramach transakcji?

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History		
Number of users		
Query complexity		
Normalization		
Addition		
Removal		
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 16 -

Dane przygotowane do przetwarzania analitycznego są większe aniżeli dane przetwarzane transakcyjnie. Analizy przekrojowe dotyczą wielu metryk i różnych zagadnień, często w długich przedziałach czasowych.

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users		
Query complexity		
Normalization		
Addition		
Removal		
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 17 -

Historia jest jądrem przetwarzania analitycznego, które przede wszystkim opiera się na analizach statystycznych i wyznaczaniu trendów w czasie. Przetwarzanie transakcyjne czasami przechowuje historię, a czasami jedynie stan aktualny.

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users	Big	Small
Query complexity		
Normalization		
Addition		
Removal		
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 18 -

Użytkownicy systemów typu BI, to przede wszystkim analitycy i prezesie firm/organizacji. Użytkownikami systemów transakcyjnych mogą być różnorodni udziałowcy biorący udział w danym procesie biznesowym. Przykładowo klienci sklepów internetowych są użytkownikami systemów transakcyjnych i w ogólności nie są użytkownikami systemów typu BI.

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users	Big	Small
Query complexity	Small or medium	Big
Normalization		
Addition		
Removal		
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 19 -

Złożoność zapytań analitycznych jest dużo większa niż zapytań w przetwarzaniu transakcyjnym. Te drugie dotyczą przede wszystkim prostych operacji wstawiania, usuwania i aktualizacji.

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users	Big	Small
Query complexity	Small or medium	Big
Normalization	Yes	No
Addition		
Removal		
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 20 -

Dane w systemach transakcyjnych są znormalizowane, w systemach analitycznych zaś, są zorganizowane w sposób ułatwiający wydajne wykonywanie zapytań analitycznych (co oznacza, że nie muszą być znormalizowane).

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users	Big	Small
Query complexity	Small or medium	Big
Normalization	Yes	No
Addition	Yes	Yes (BULK)
Removal		
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 21 -

Dane w standardowych systemach typu BI są dodawane w postaci paczek (zazwyczaj raz na dobę), inaczej niż w systemach transakcyjnych. Należy pamiętać jednak, że coraz bardziej rozwijają się systemy, które analizują dane w czasie rzeczywistym (tzw. *Real Time Analytical Processing*). Systemy te są jednak poza zakresem tego kursu.

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users	Big	Small
Query complexity	Small or medium	Big
Normalization	Yes	No
Addition	Yes	Yes (BULK)
Removal	Yes	No
Updates		
Transactions		

**OLTP (On-Line Transactional Processing)**

- 22 -

Dane z systemów analitycznych nie są usuwane (chyba, że dojdzie do wpisania do systemu błędnych danych). Danych się nie usuwa, gdyż są potrzebne do analiz historycznych.

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users	Big	Small
Query complexity	Small or medium	Big
Normalization	Yes	No
Addition	Yes	Yes (BULK)
Removal	Yes	No
Updates	Yes	Rarely
Transactions		

**OLTP (On-Line Transactional Processing)**

- 23 -

W systemach typu BI zazwyczaj nie aktualizujemy danych przez nadpisanie. Informacje historyczne są potrzebne!

## OLAP features

	OLTP	OLAP
Volume	From small to big	Large
History	Sometimes	Always
Number of users	Big	Small
Query complexity	Small or medium	Big
Normalization	Yes	No
Addition	Yes	Yes (BULK)
Removal	Yes	No
Updates	Yes	Rarely
Transactions	Yes	No

**OLTP (On-Line Transactional Processing)**

- 24 -

Ze względu na fakt, że dane są ładowane w postaci paczek danych nie ma przetwarzania transakcyjnego.

1. What is BI system?
2. Describe the overall architecture of BI systems.
3. What dashboard is?
4. What KPI is and how is it presented?
5. What OLAP is?
6. Compare OLAP with OLTP.



- 25 -