

Technologie BI w zarządzaniu

Wykład 1

Marta Juźwin

Dla kogo?!

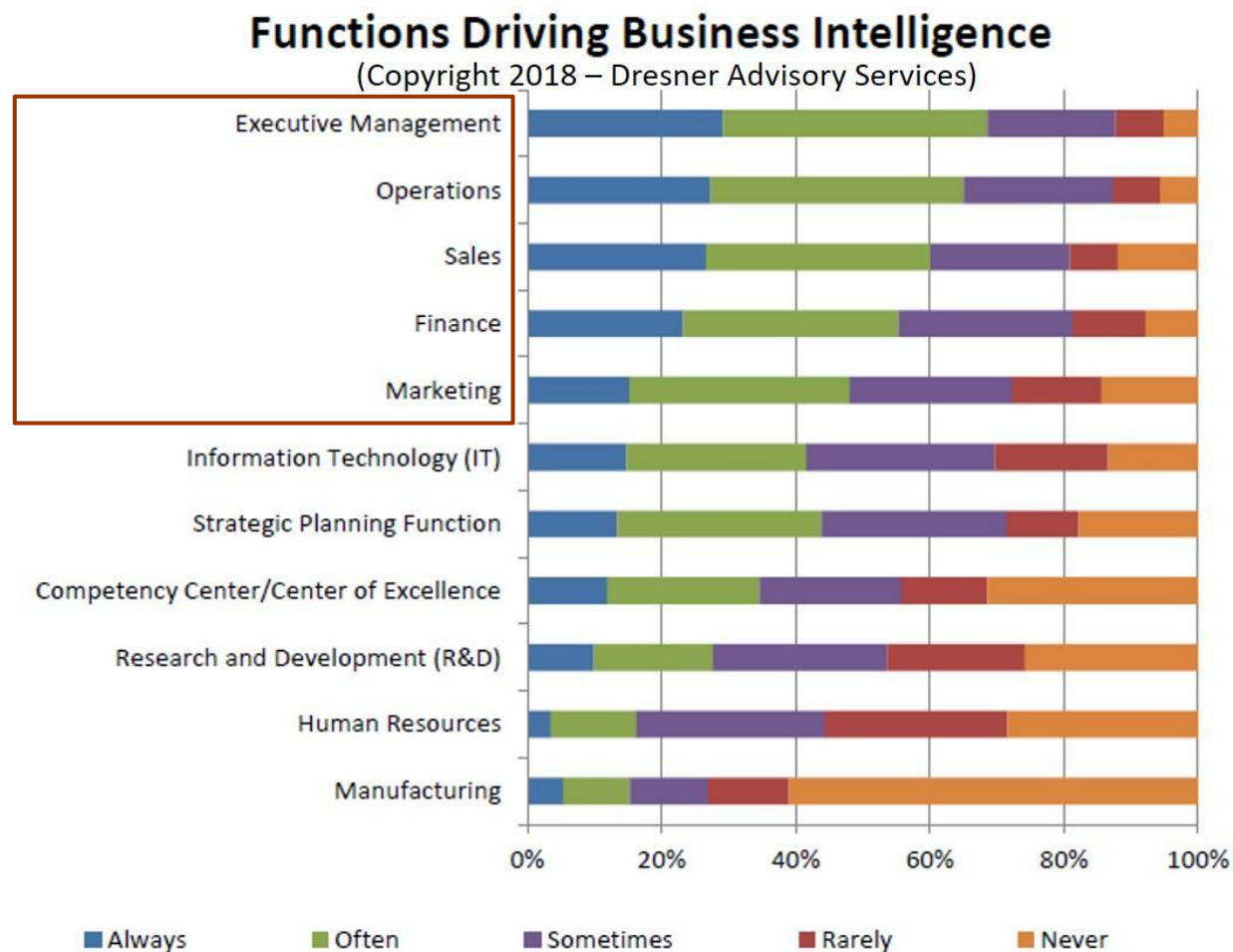


Figure 5 – Functions driving business intelligence

Business Intelligence Objectives

(Copyright 2018 – Dresner Advisory Services)

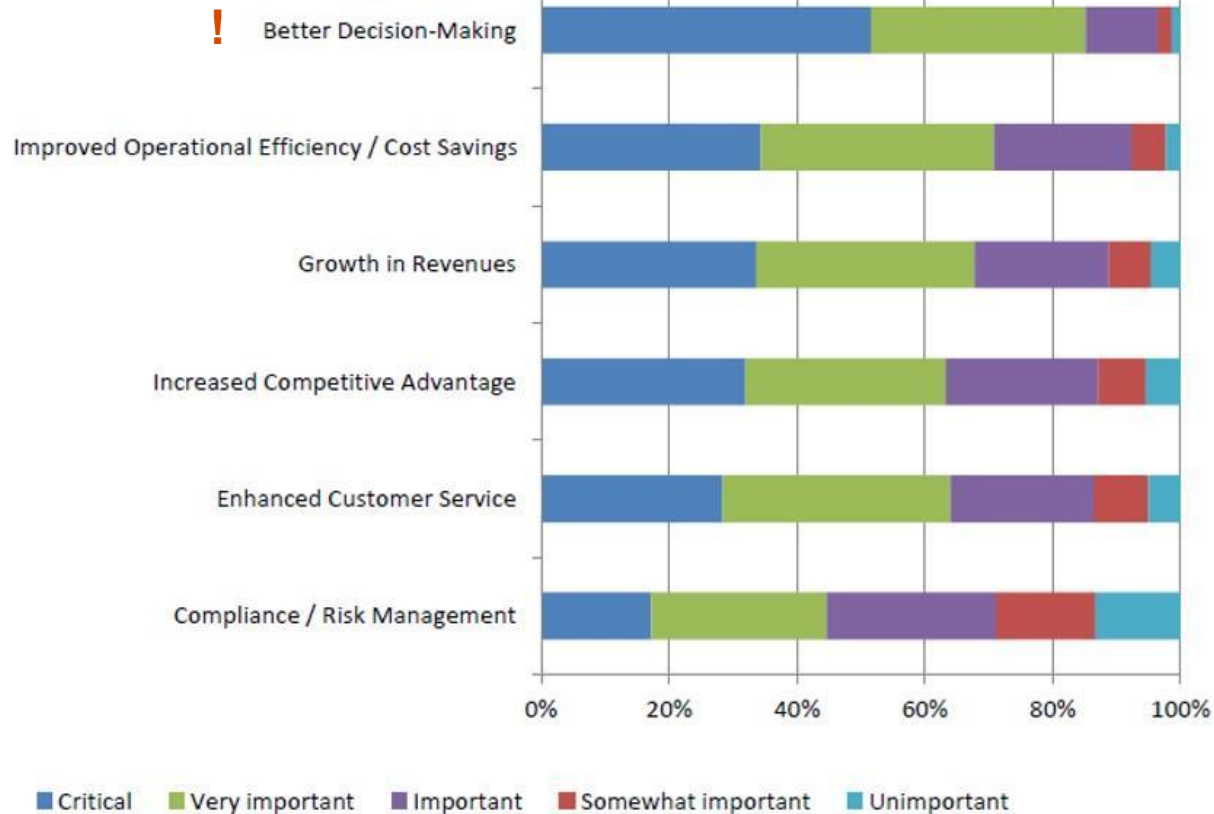
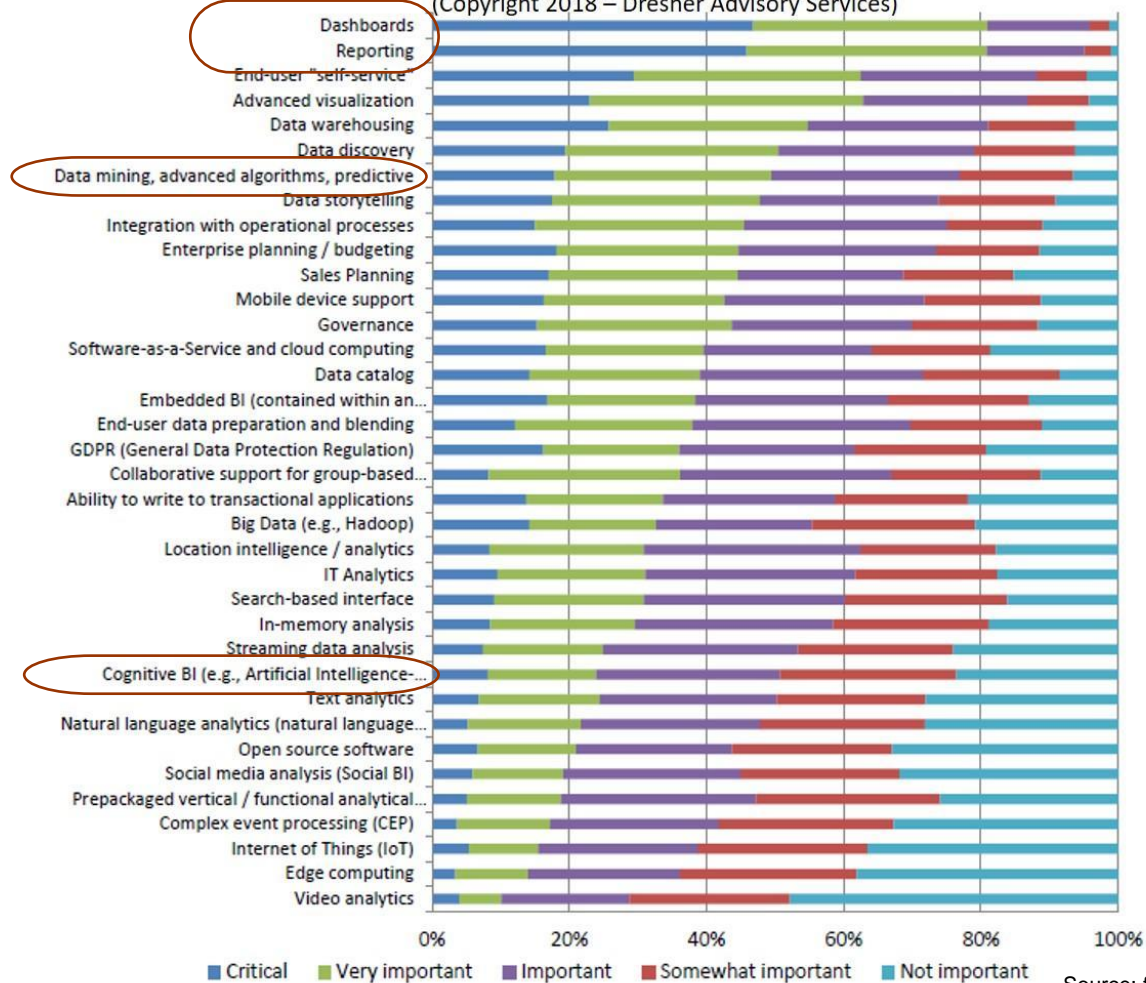


Figure 16 – Business intelligence objectives

Po co?!

Technologies and Initiatives Strategic to Business Intelligence

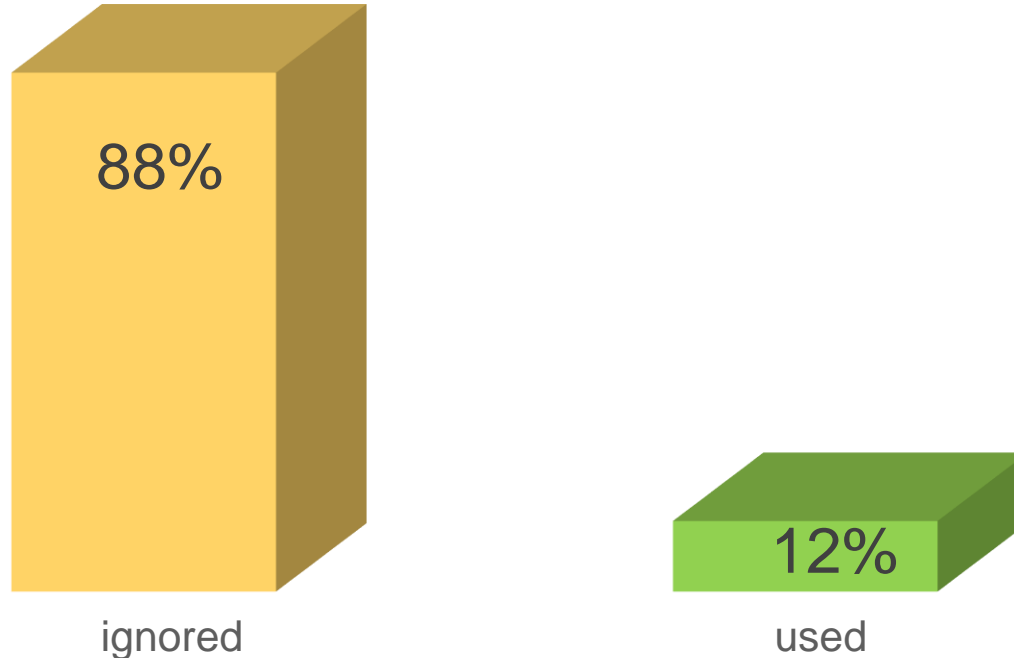
(Copyright 2018 – Dresner Advisory Services)



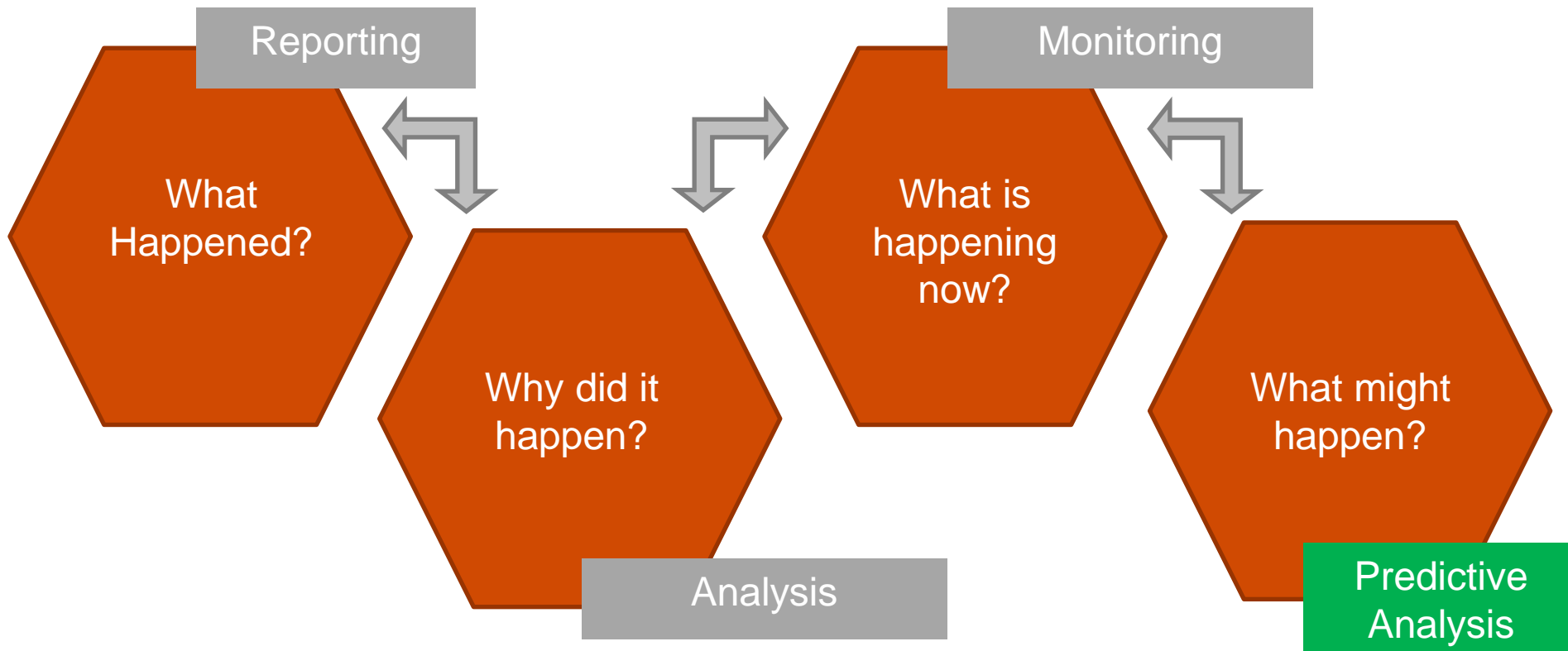
Zasoby

88% data is ignored (in Big Data analytics)

According to a recent study by Forrester Research, most companies analyze a mere amount of 12% of the data they have.



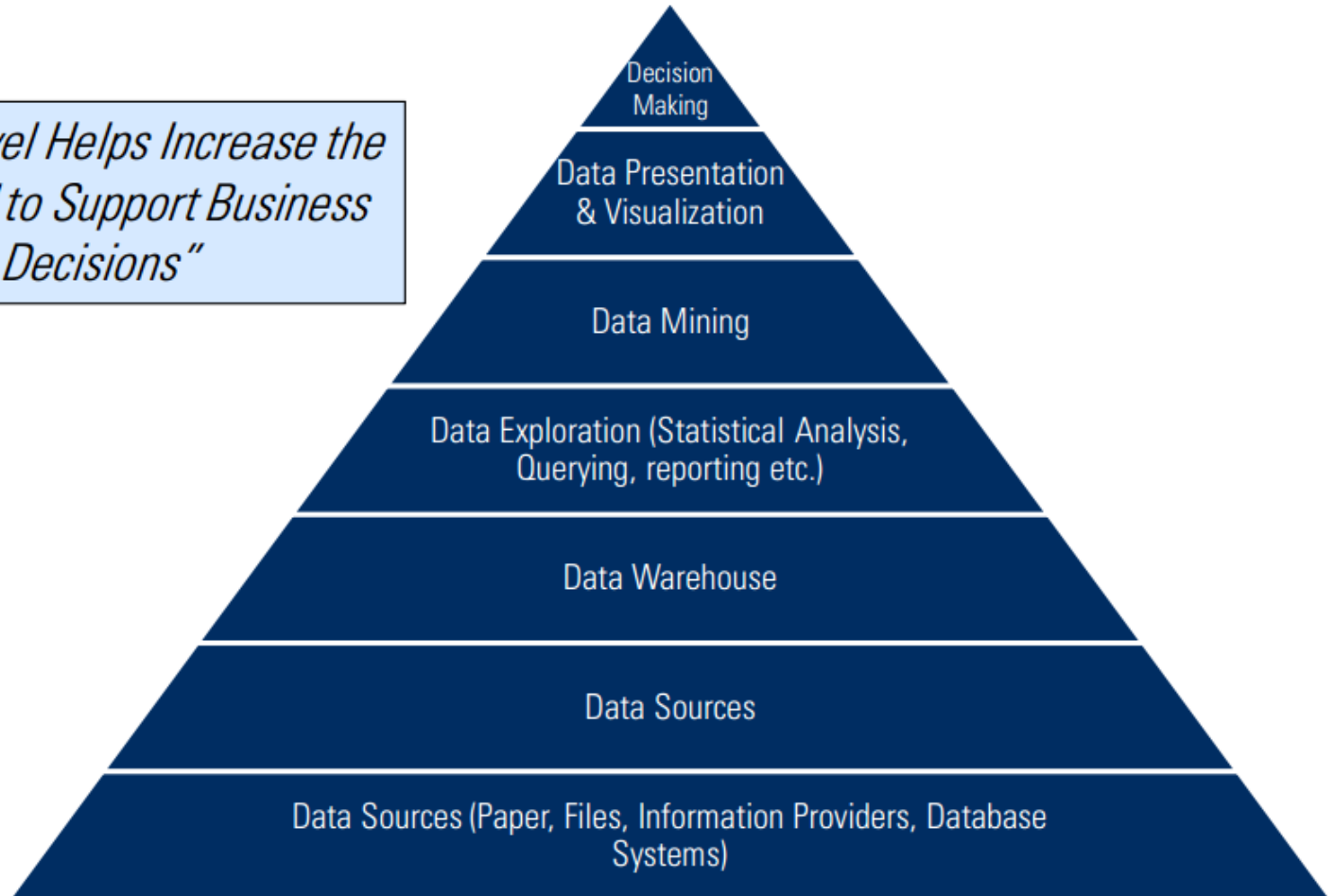
Obszary rozwoju BI



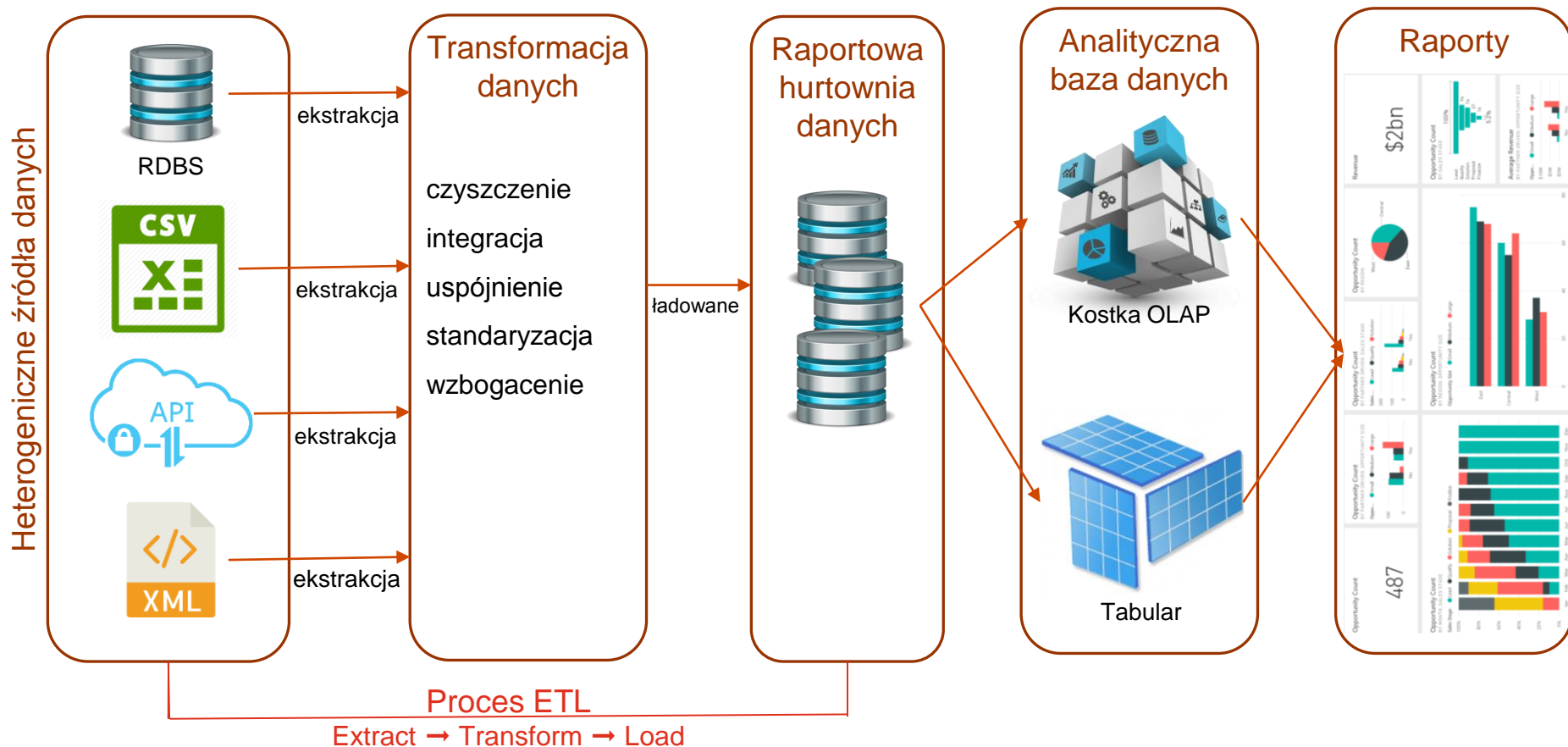
Business Intelligence (BI)

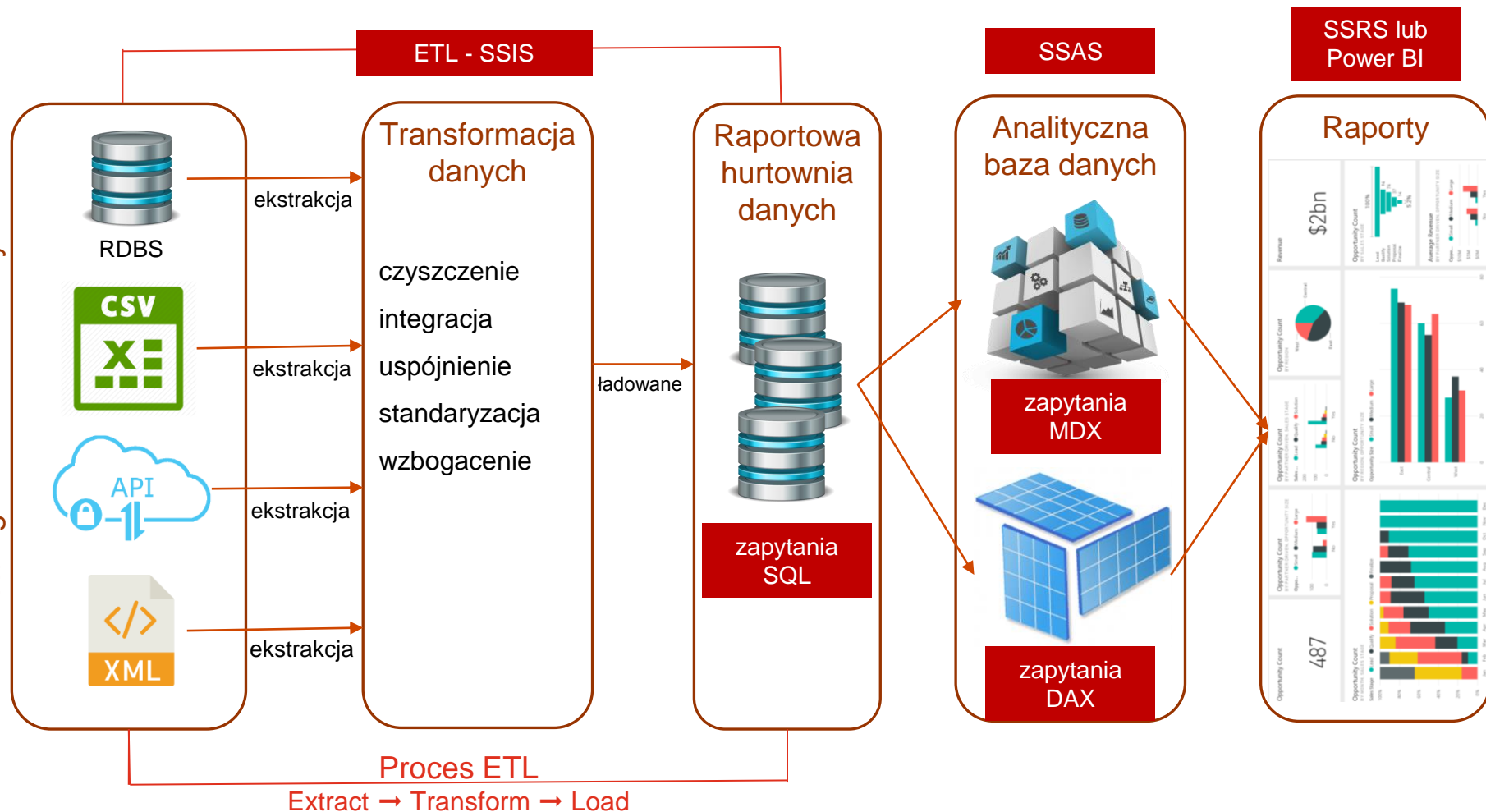
- BI - zbiór technologii służących do przekształcania dużych wolumenów danych w informacje, a następnie przekształcania tej informacji w wiedzę
- BI = OLTP + ETL + HD + OLAP/TABULAR + eksploracja danych + ...
- adresowany do użytkowników biznesowych
- skupiony wokół technologii hurtowni danych
- **Aplikacje BI - głównie raporty, pulpity menadżerskie; algorytmy eksploracji danych**

"Every Level Helps Increase the Potential to Support Business Decisions"

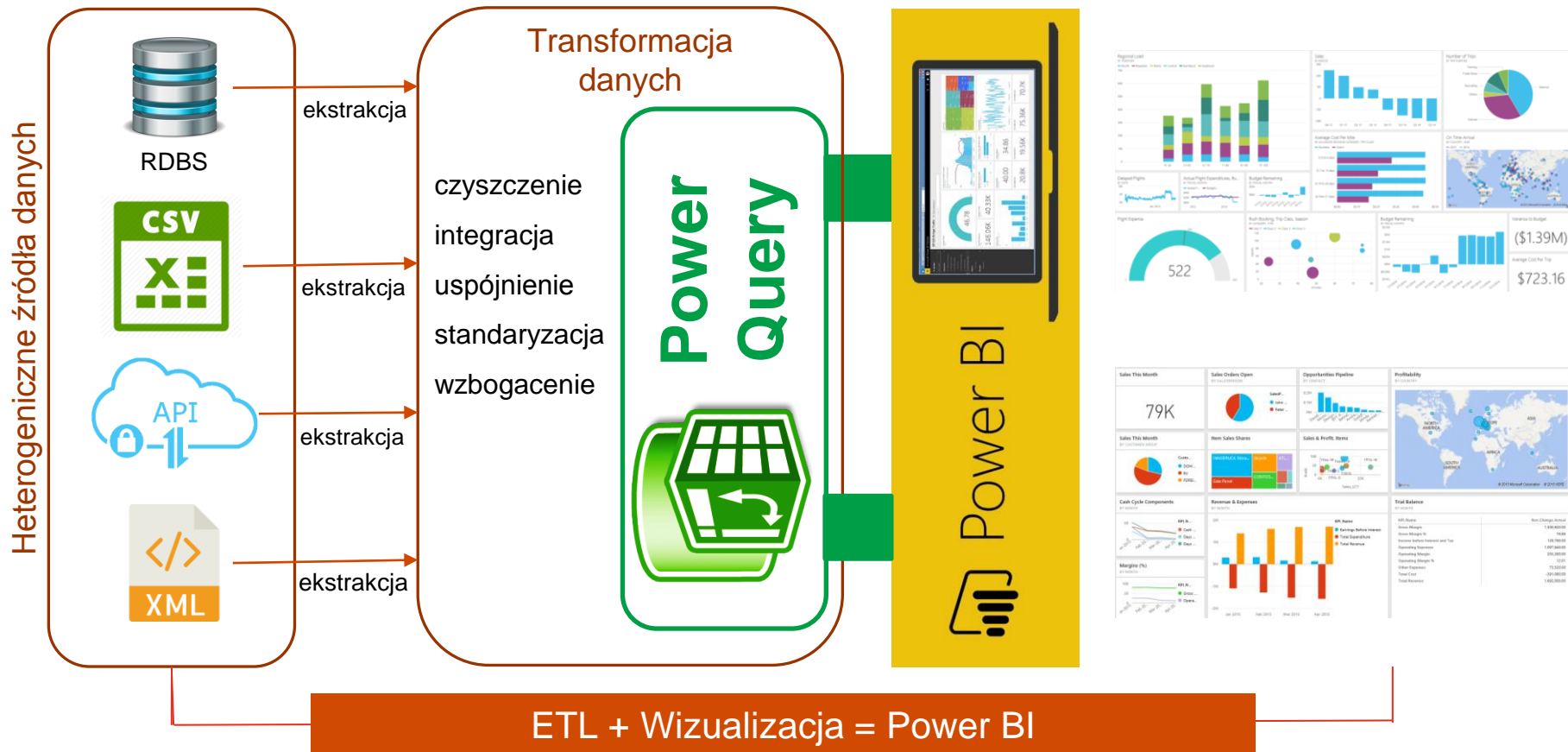


Wewnątrz BI





Wewnątrz BI (uproszczony model dla małych firm)



ETL \ ELT

ETL – Extract, Transform, Load

Ekstrakcja – pozyskanie danych ze źródeł zewnętrznych

- Bazy OLTP, pliki tekstowe, arkusze kalkulacyjne

Transformacja

- Czyszczenie, integracja, redukcja, konwersja,

Ładowanie danych do hurtowni

Proces ETL jest najtrudniejszym etapem implementacji HD – nawet 70% czasu

ELT – Extract, Load, Transform

- bardziej dla Big Data
- gdy można przetwarzać dane w docelowym miejscu (np. klaster Hadoop)

Systemy transakcyjne vs systemy analityczne

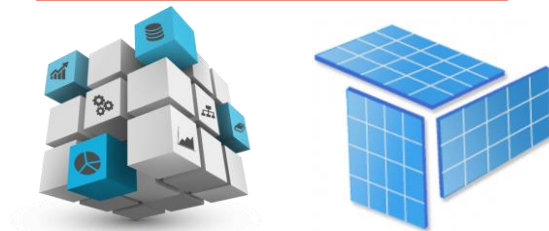
OLTP



Systemy transakcyjne
(operacyjne)

Zbieranie
danych

OLAP/TABULAR



Systemy analityczne
(BI)

Analiza
danych

Te same dane
różne potrzeby

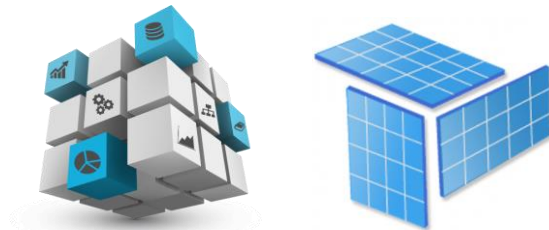
Systemy transakcyjne vs systemy analityczne

OLTP



- Zbieranie zamówień
- Rejestracja klientów
- Monitorowanie działalności
- Przykłady: rozkład jazdy pociągów, bankomaty, sprzedaż biletów, system finansowo-księgowy

OLAP/TABULAR



- Obserwacja liczby zamówień
- Analiza zachowań klientów
- Weryfikacja poprawności działania
- Przykłady: analiza klientów, sprzedaży, analiza budżetowa

Jakie pytania do bazy transakcyjnej, a jakie do analitycznej?

- Jaka grupa klientów generuje 80% obrotu?
- Który klient zalega z płatnością?
- Jaki jest obecny status zamówienia 63ab201046ZC?
- Jaka była wielkość sprzedaży w poszczególnych miesiącach ostatniego roku?
- Jakie są charakterystyczne cechy naszych najlepszych klientów, biorąc pod uwagę te informacje, które o nich gromadzimy?
- Jak automatycznie wykryć, którzy klienci zamierzają prawdopodobnie niedługo zrezygnować z naszych usług?

Jakie pytania do bazy transakcyjnej, a jakie do analitycznej?

OLAP • Jaka grupa klientów generuje 80% obrotu?

OLTP • Który klient zalega z płatnością?
• Jaki jest obecny status zamówienia 63ab201046ZC?

OLAP • Jaka była wielkość sprzedaży w poszczególnych miesiącach ostatniego roku?

DM • Jakie są charakterystyczne cechy naszych najlepszych klientów, biorąc pod uwagę te informacje, które o nich gromadzimy?
• Jak automatycznie wykryć, którzy klienci zamierzają prawdopodobnie niedługo zrezygnować z naszych usług?

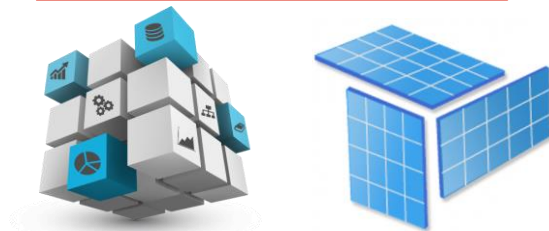
Systemy transakcyjne vs systemy analityczne

OLTP



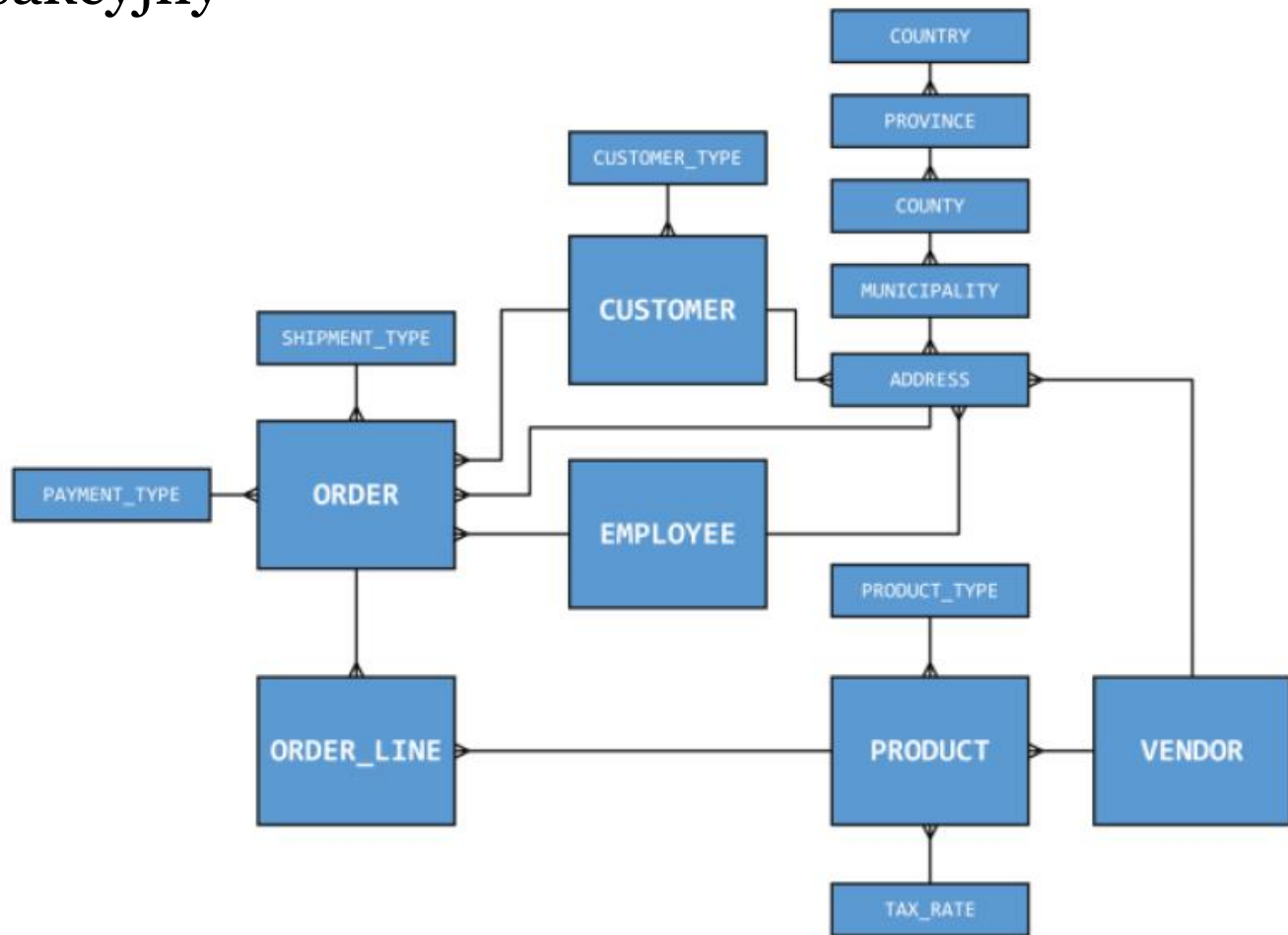
- Wiele krótkich transakcji
- Istotna – przepustowość (trans/sek)
- Powtarzalne operacje
- Często modyfikowana
- Dane odzwierciedlają stan bieżący
- Użytkownicy – bardzo wielu
- Przetwarzanych rekordów – kilka, kilkadziesiąt
- Ukierunkowany na aplikację

OLAP/TABULAR



- Niewiele złożonych zapytań
- Istotny – czas odpowiedzi
- Zapytania ad-hoc
- Głównie do odczytu
- Dane za wybrany okres czasu
- Użytkownicy – kilku analityków
- Przetwarzanych rekordów – miliony
- Ukierunkowany na temat

Model transakcyjny



Model analityczny

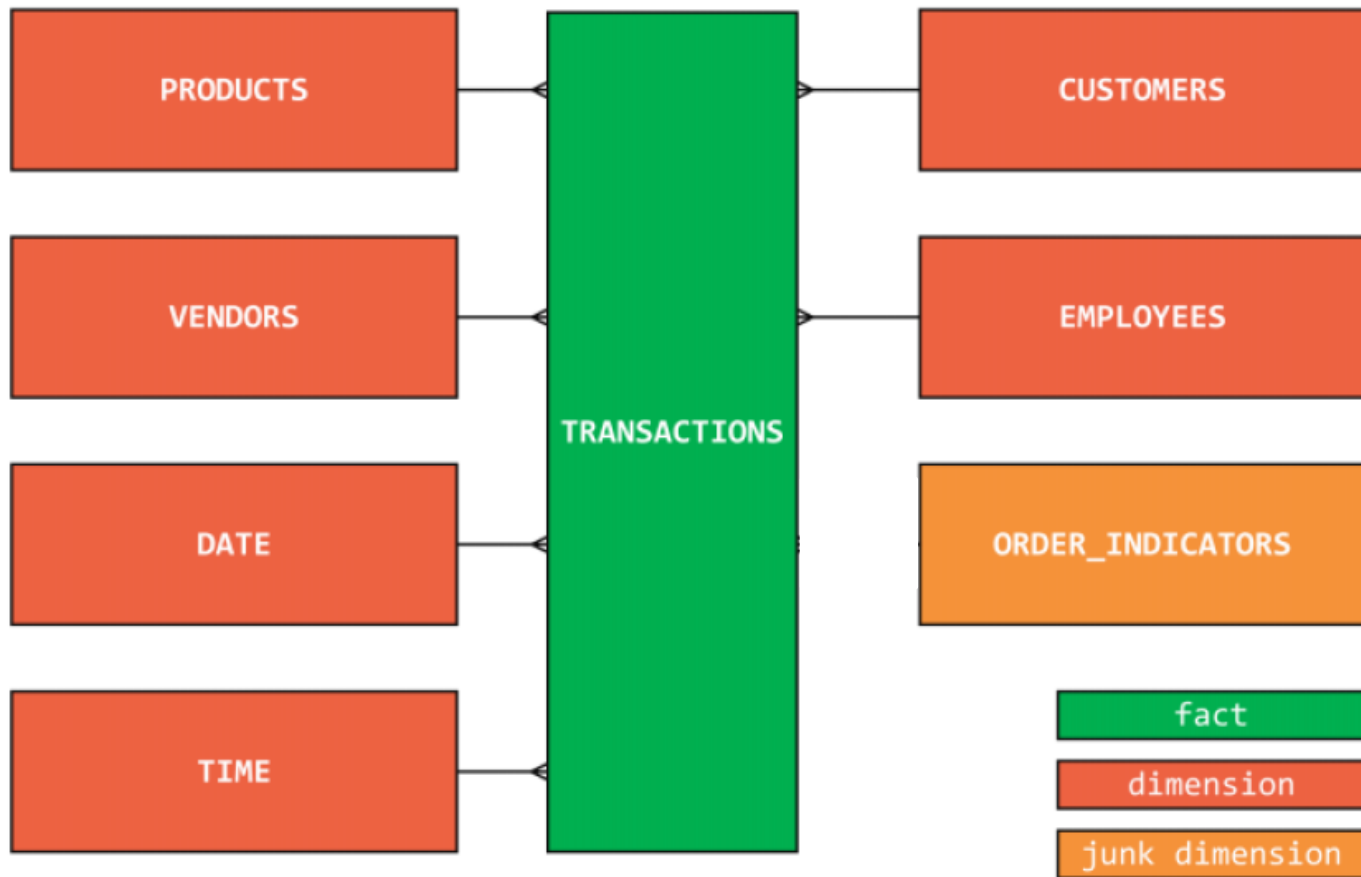


Tabela wymiarów

- Zawiera atrybuty opisowe
- Zawiera dane rzadko podlegające zmianom
- Zawiera dużo kolumn – dostarczających dodatkowych informacji opisowych
- Mogą zawierać wiele hierarchii – umożliwiają organizację danych na różnych poziomach agregacji

	CustomerKey	GeographyKey	FirstName	LastName	BirthDate	MaritalStatus	EmailAddress	TotalChildren	AddressLine1
1	11000	26	Jon	Yang	1971-10-06	M	jon24@adventure-works.com	2	3761 N. 14th St
2	11001	37	Eugene	Huang	1976-05-10	S	eugene10@adventure-works.com	3	2243 W St.
3	11002	31	Ruben	Torres	1971-02-09	M	ruben35@adventure-works.com	3	5844 Linden Land
4	11003	11	Christy	Zhu	1973-08-14	S	christy12@adventure-works.com	0	1825 Village Pl.
5	11004	19	Elizabeth	Johnson	1979-08-05	S	elizabeth5@adventure-works.com	5	7553 Harness Circle
6	11005	22	Julio	Ruiz	1976-08-01	S	julio1@adventure-works.com	0	7305 Humphrey Drive
7	11006	8	Janet	Alvarez	1976-12-02	S	janet9@adventure-works.com	0	2612 Berry Dr
8	11007	40	Marco	Mehta	1969-11-06	M	marco14@adventure-works.com	3	942 Brook Street
9	11008	32	Rob	Verhoff	1975-07-04	S	rob4@adventure-works.com	4	624 Peabody Road
10	11009	25	Shannon	Carlson	1969-09-29	S	shannon38@adventure-works.com	0	3839 Northgate Road
11	11010	22	Jacquelyn	Suarez	1969-08-05	S	jacquelyn20@adventure-works.com	0	7800 Cominne Court
12	11011	22	Curtis	Lu	1969-05-03	M	curtis9@adventure-works.com	4	1224 Shoenic
13	11012	611	Lauren	Walker	1979-01-14	M	lauren41@adventure-works.com	2	4785 Scott Street

Wymiary

Wymiary – nadają znaczenie faktom, stanowią kontekst wykonywanej analizy

- wymiary stałe
- wymiary zmienne

Na przykład: analiza sprzedaży może być wykonywana pod kątem

- czasu (kiedy sprzedano towar?)
- produktu (jaki towar sprzedano?)
- klienta (kim był klient, który dokonał zakupu – płeć, wiek, miejsce zamieszkania?)

Na wymiarach buduje się hierarchie

- Czas: rok – kwartał – miesiąc - dzień
- Produkt: kategoria – podkategoria - nazwa

Hierarchie

Hierarchie (ang. hierarchies) umożliwiają organizację danych na różnych poziomach agregacji

Hierarchia bazująca na poziomach (Level-based dimension)

- data - dzień tygodnia
- data - miesiąc - rok

Hierarchia bazująca na wartościach (Value-based dimension)

- pracownik - kierownik

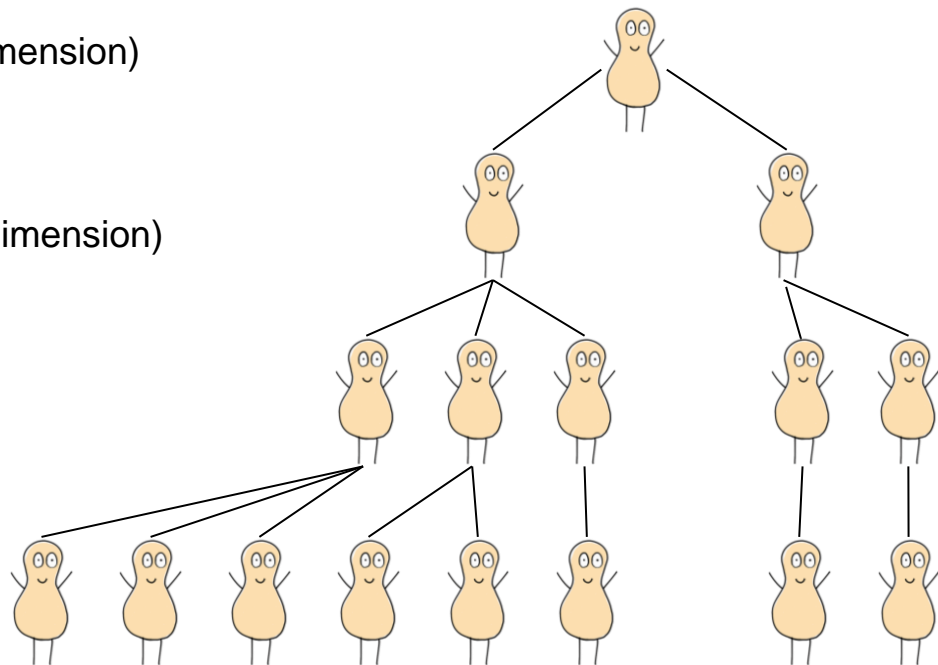
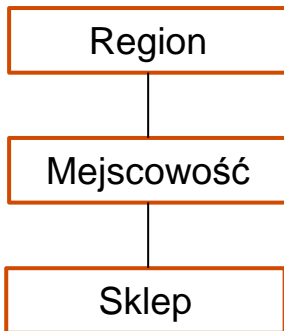
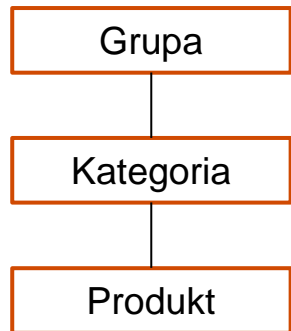


Tabela faktów

Tabela faktów zawiera:

- Klucze obce do tabel wymiarów
- Fakty (miary)

Wszystkie klucze obce stanowią klucz podstawowy

Bardzo duża liczba wierszy (zawiera 90% danych całej hurtowni), szybko rośnie

	intTimeID	dcAmount	intCurrencyID	intScenarioID	nchTransactionType	sdTransactionDate	intFimStructureID	intEntityID	btTransactionID	sintInvestm
1	102	412030.0000	16	1	D	2009-06-01 00:00:00	139	1	0x52781F02958FDE0E20DD7E4ACE100F95E199B3CB	1
2	102	35870.0000	16	1	D	2009-06-01 00:00:00	139	1	0xD9B4B2462A5255C837AADBE2873DDE51DFEC8274	1
3	102	-35870.0000	16	1	C	2009-06-01 00:00:00	139	1	0xCC6704715988467215799686B37722D3AD962376	1
4	102	-423030.0000	16	1	C	2009-06-01 00:00:00	139	1	0x3C2E35371B6A147AA7BEDA7753615524EDC6C4AE	1
5	102	-3000.0000	16	1	C	2009-06-01 00:00:00	139	1	0x9604490670CDAC0C86A0897A596DA9ED2069A34D	1
6	107	-11300.0000	5	3	C	2009-11-30 00:00:00	1059	19	0xB8737EB67B90AF43F7611DFDC0EF9C55668982A8	1
7	107	700.0000	5	3	D	2009-11-30 00:00:00	1059	19	0x22F8AFB220309C182DE429F873C8FAADFDECBDC6	1
8	107	1.0000	5	3	D	2009-11-30 00:00:00	1059	19	0x67C250B61F8EDFD9ACB2EF2420E00ABEC26FCDA8	1
9	107	-21000.0000	5	3	C	2009-11-30 00:00:00	3015	19	0x2FB187C505F1B7055AE41D198C85DFE7A8FBD8B2	1
10	107	2200.0000	5	3	D	2009-11-30 00:00:00	3015	19	0x4B3A882424FF8D3FF8FEB82EFC3020307C69A71A	1
11	107	2000.0000	5	3	D	2009-11-30 00:00:00	3015	19	0x6759564AA9164A932E114021E67E9631B8720983	1
12	107	16500.0000	5	3	D	2009-11-30 00:00:00	1059	19	0x4CA9F8B77E5112B3B2F14225E38D567AED2AC015	1

Klucze w tabelach

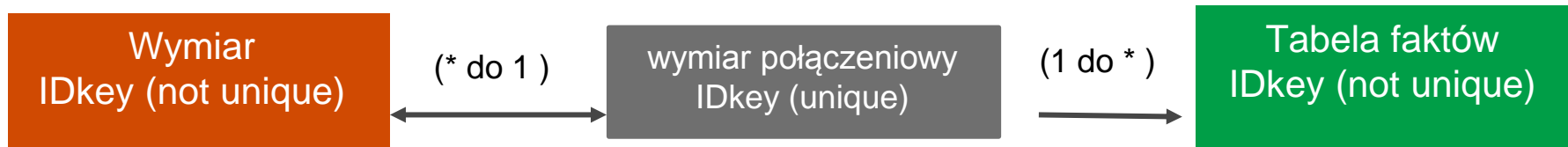
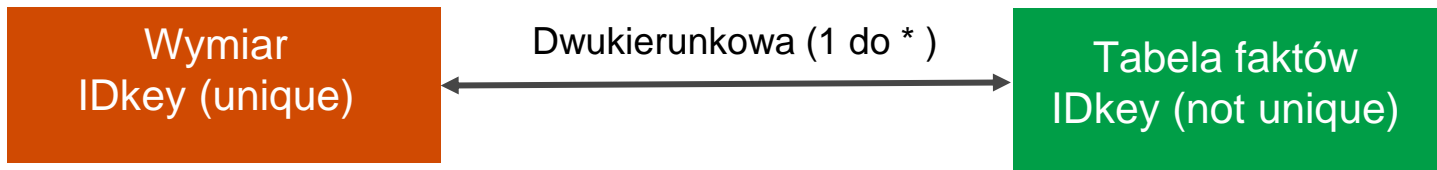
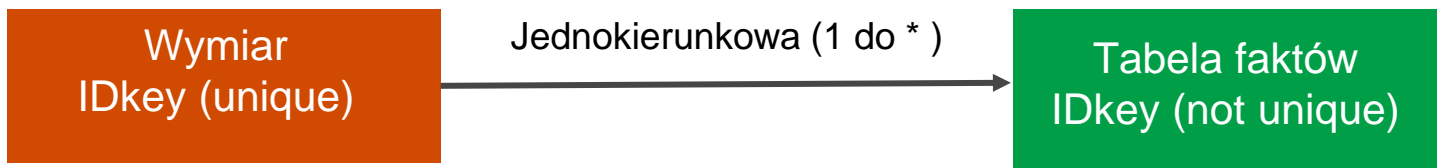
Klucze w tabelach wymiarów

- unikalny w ramach danej bazy operacyjnej (np. PESEL, nr. rejestracyjny pojazdu, numer faktury) - **natural key, business key**
- może się zmienić lub nie być unikalny w systemie BI integrującym różne bazy
Tworzony sztucznie, niezależny od bazy operacyjnej, pod pełną kontrolą systemu BI, niezmienny i unikalny - **durable key**
- może być generowany automatycznie, jest krótszy, nie ma ryzyka dublowania - **surrogate key**

Klucz w tabeli faktów

- Najczęściej klucz podstawowy składa się z kluczy obcych do wymiarów

Relacje między tabelami



Miary

Fakty (miary) to informacje podlegające analizie

Najczęściej – wartości liczbowe

Na przykład:

- liczba sprzedanych sztuk, kwota kredytu, wartość paragonu, liczba klientów

Amount :=

```
CALCULATE (
    SUM ( Finance[dcAmount] ) * AVERAGE ( FactExchangeRate[dcExchangeRate] )
)
```

Cztery kroki modelowania analitycznego

Zidentyfikuj procesy biznesowe

- analiza: przychody netto od klientów zewnętrznych, zysku brutto ze sprzedaży do przychodów, trendy sprzedażowe

Dobierz stopień szczegółowości/ziarnistość

- jak szczegółowe dane przechowywać? Dla każdego dnia, miesiąca? Dla każdej minuty?

Wyróżnij wymiary

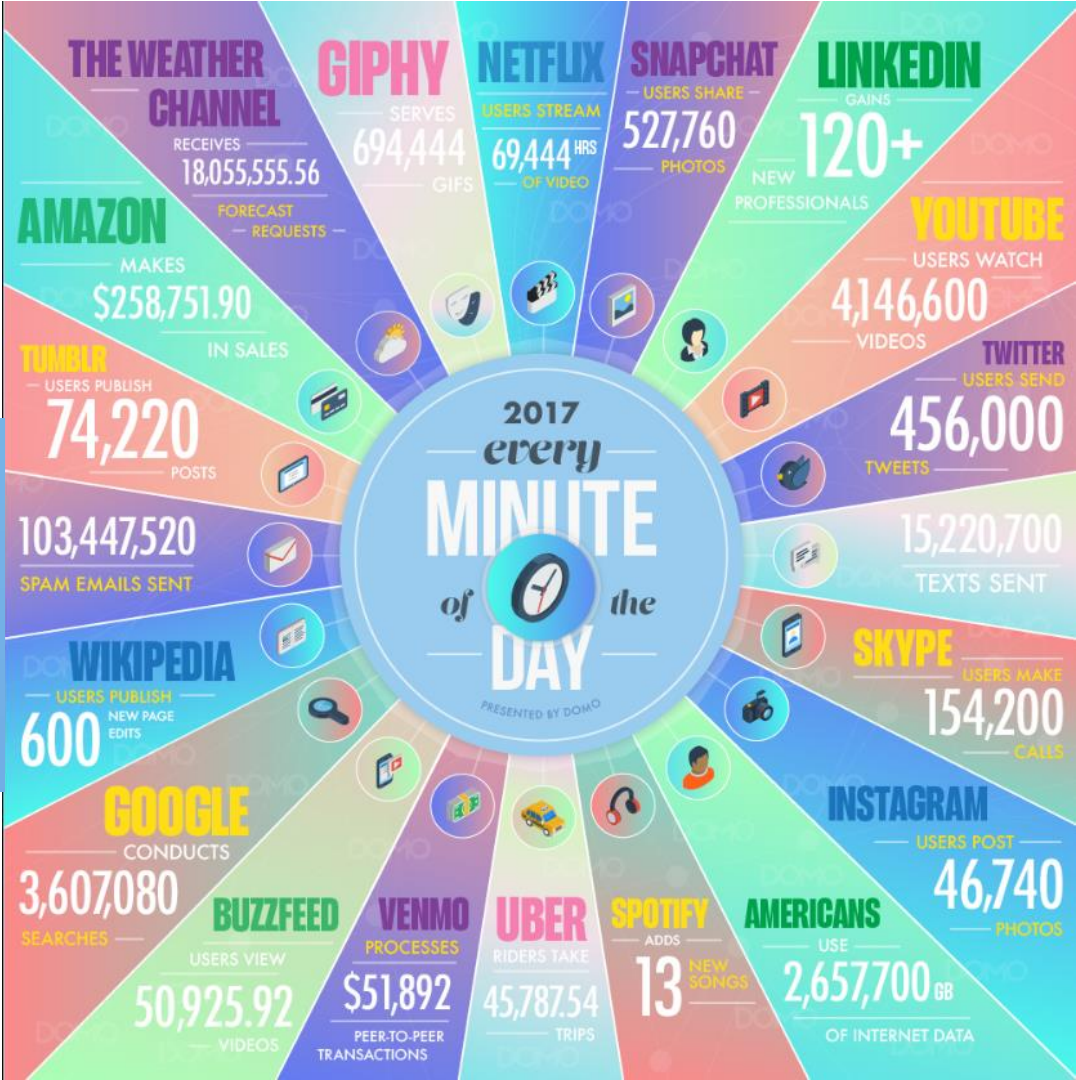
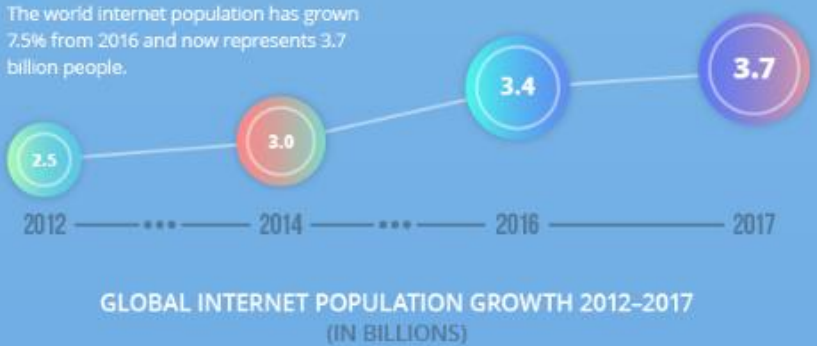
- podział geograficzny
- podział biznesowy (produkty)
- period
- placówka
- promocja
- metoda płatności

Wyróżnij fakty/miary

DATA NEVER SLEEPS

How much data is generated every minutes

90% of all data today was created in last two years – that's 2,5 quintillion bytes of data per day.



Source: https://www.domo.com/learn/data-never-sleeps-5?aid=ogsm072517_1&sf100871281=1