

# Wprowadzenie do Hurtowni Danych

# Organizacyjnie

**Prowadzący:**

mgr. Mariusz Rafało

[mrafalo@sgh.waw.pl](mailto:mrafalo@sgh.waw.pl)

<http://mariuszrafalo.pl> (hasło HD2)

# Literatura

1. Inmon, W., Linstedt, D. (2014). Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault, Morgan Kaufmann
2. Kimball, R., Ross, M. (2013). The Data Warehouse Lifecycle Toolkit, The Data Warehouse Lifecycle. Wiley.
3. Todman, C. (2011), Projektowanie hurtowni danych. Wspomaganie zarządzania relacjami z klientami, Helion
4. Moss, L. T., & Atre, S. (2003). Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley Professional.

# WPROWADZENIE DO BAZ DANYCH

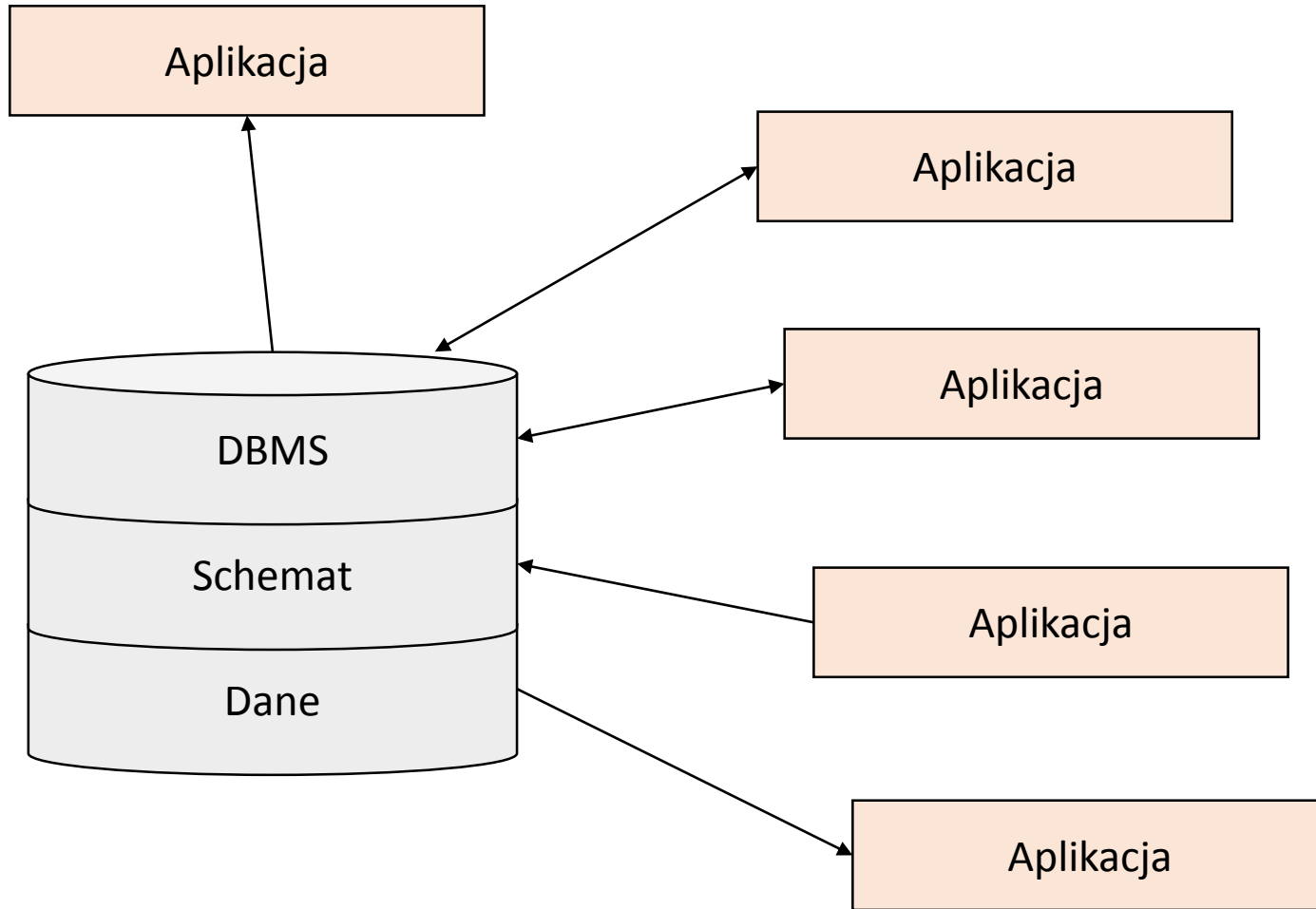
# Baza danych

- Baza danych (*database*) – zbiór danych powiązanych tematycznie, zorganizowany w sposób umożliwiający ich przeszukiwanie według określonych kryteriów
- System bazy danych: ***DataBase Management System*** (DBMS) – oprogramowanie komputerowe służące gromadzeniu danych w bazie danych. Celem systemu DBMS jest:
  - Wprowadzanie danych do bazy danych
  - Pobieranie danych z bazy danych
  - Modyfikacja danych w bazie danych

# Obiekty w bazie danych

- Schematy
- Tabele
- Widoki
- Ograniczenia (*constraints*)
- Sekwencje
- Procedury składowane
- Role
- Użytkownicy
- Wyzwalacze (*triggers*)
- Inne

# System bazy danych



# Język SQL

- SQL (ang. Structured Query Language) – deklaratywny język programowania, pozwalający na komunikowanie się z bazą danych
- SQL pozwala na modyfikowanie struktury bazy danych (DDL - *Data Definition Language*)
- SQL pozwala na dodawanie, modyfikowanie i usuwanie rekordów z bazy danych (DML - *Data Manipulation Language*)



# Typy baz danych

Ze względu na ułożenie danych:

- Relacyjna baza danych
- Kolumnowa baza danych
- Hierarchiczne bazy danych
- Obiektowe bazy danych
- Baza klucz-wartość
- Baza noSQL

Ze względu technologię przetwarzania:

- Bazy „klasyczne”
- Bazy *in memory*
- Bazy rozproszone
- Bazy wielowymiarowe

# WŁAŚCIWOŚCI BAZY DANYCH

# Funkcje bazy danych

- Możliwość współdzielenia danych
- Integralność danych
- Trwałość danych
- Transakcyjność
- Bezpieczeństwo danych
- Niezależność danych

# Transakcyjność bazy danych

- Transakcja to zbiór operacji na danych, stanowiący logiczną całość i realizowany w całości (lub w ogóle)
- Transakcja w bazie danych powinna spełniać zasady ACID:
  - **A**tomicity
  - **C**onsistency
  - **I**solation
  - **D**urability

# RELACYJNA BAZA DANYCH

# Tabele – organizacja danych

- Kolumny
- Wiersze
- Typy danych
- Relacje
- Klucze
- Ograniczenia
- Rekord
- Uprawnienia

# Relacje

KLIENT		
ID	NAZWISKO	PESEL
1	Kowalski Piotr	88121222344
2	Nowak Jan	78121222933

ADRES			
ID	MIASTO	ULICA	KLIENT_ID
1	Warszawa	Kasztanowa 2	1
2	Poznań	Nowa 1	1

Relacja – związek ustanowiony pomiędzy kolumnami w dwóch tabelach. Kolumny muszą być tego samego typu.

# Relacja jeden do jednego

KLIENT		
ID	NAZWISKO	PESEL
1	Kowalski Piotr	88121222344
2	Nowak Jan	78121222933

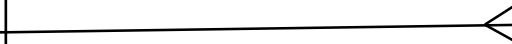
ADRES			
ID	MIASTO	ULICA	KLIENT_ID
1	Warszawa	Kasztanowa 2	1
2	Poznań	Nowa 1	2



# Relacja jeden do wielu

FAKTURA		
ID	NUMER	DATA
1	FV 12/1213/11	2017-01-12
2	FV 12/343/23	2017-02-13

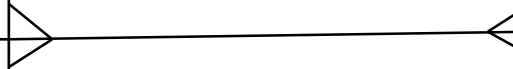
POZYCJA				
ID	PRODUKT	ILOSC	WARTOSC	FAKTURA
1	Krzesło XXX	1	120.00	1
2	Stolik	23	12123.00	1
2	Lampa	2	123.22	1
3	Szafa kokos	2	1299.12	1
4	Stolik XXX	1	83.00	2
5	Lampa A	3	320.00	2



# Relacja wiele do wielu

PRZEDSIĘBIORSTWO		
ID	NAZWA	MIASTO
1	Impex S.A	Bytom
2	Metalex	Gdynia

PRACOWNIK		
ID	PRACOWNIK	WYKSZTAŁCENIE
1	Jan Nowak	Wyższe
2	Adam Wiśniewski	Podstawowe
2	Aldona Kowalska	Średnie
3	Marian Kozłowski	Wyższe
4	Ireneusz Stopa	Wyższe
5	Jan Dzban	Średnie



# Relacja wiele do wielu

PRZEDSIĘBIORSTWO		
ID	NAZWA	MIASTO
1	Impex S.A	Bytom
2	Metalex	Gdynia

?

PRACOWNIK		
ID	PRACOWNIK	WYKSZTAŁCENIE
1	Jan Nowak	Wyższe
2	Adam Wiśniewski	Podstawowe
3	Aldona Kowalska	Średnie
4	Marian Kozłowski	Wyższe
5	Ireneusz Stopa	Wyższe
6	Jan Dzban	Średnie

# Relacja wiele do wielu

PRZEDSIĘBIORSTWO		
ID	NAZWA	MIASTO
1	Impex S.A	Bytom
2	Metalex	Gdynia

PRZEDS_PRAC	
PRZEDS	PRAC
1	1
1	2
1	5
2	3
2	4
2	1

PRACOWNIK		
ID	PRACOWNIK	WYKSZTAŁCENIE
1	Jan Nowak	Wyższe
2	Adam Wiśniewski	Podstawowe
2	Aldona Kowalska	Średnie
3	Marian Kozłowski	Wyższe
4	Ireneusz Stopa	Wyższe
5	Jan Dzban	Średnie

# INNE BAZY DANYCH

# Kolumnowa baza danych

1. Vertica (HP)
2. HBase (open source)
3. Greenplum (EMC)
4. SybaseIQ (SAP)
5. Teradata (Teradata)\*
6. Microsoft (Microsoft)\*

\* Baza może pracować zarówno w trybie wierszowym jak i kolumnowym

# Baza danych typu klucz-wartość

1. Redis
2. Ignite
3. Memcached

# Baza danych *in memory*

1. HANA (SAP)
2. TimesTen (Oracle)
3. VoltDB (VoltDB)
4. Informix (IBM)



# Rozproszona baza danych (MPP)

1. Hadoop (open source)
2. Teradata (Teradata)
3. Pure Data (IBM)
4. Greenplum (EMC)

# BAZY DANYCH A HURTOWNIA DANYCH

# Hurtownia danych – wybór bazy danych

1. Baza relacyjna vs kolumnowa
2. Baza dyskowa vs in-memory
3. Baza macierzowa vs rozproszona
4. Jedna baza vs wiele baz

Dziękuję za uwagę