

# Bazy danych

## *przed 14 wykładem*

Andrzej Giniewicz

21.06.2024

Wykład będzie w dużej mierze opowiadany, ponieważ opiszemy sobie sporo technologii oraz technik. Z tego powodu notatki są jedynie niemal wypunktowaniem tematów, o których będziemy rozmawiać. Z niniejszego materiału nie ma kartkówki. Osoby, które poniższe tematy zainteresują, zapraszam do dyskusji podczas wykładu. Szczególny nacisk podczas dyskusji położymy na bazę danych DuckDB, która w tym miesiącu miała wydanie stabilne.

## **Technologie związane z bazami danych**

Podczas zajęć zajmowaliśmy się pojedynczą bazą danych. W środowisku firmowym często mamy ich więcej i potrzebna jest koordynacja pomiędzy różnymi źródłami danych. Wyróżnia się trzy istotne cechy źródeł danych, gdy zamierzamy koordynować pracę pomiędzy nimi:

1. integracja, czyli kompatybilność schematów baz danych,
2. trwałość, czyli możliwość przechowania w celu archiwizacji przez długi czas, oraz
3. ukierunkowanie, czyli optymalizację pod kątem konkretnego typu operacji wykonywanej na bazie danych.

Zbiór baz danych, który spełnia te trzy cechy, nazywamy zwykle hurtownią danych. Z hurtownią danych bardzo często możemy spotkać się w większych firmach, które rozwijają swoją infrastrukturę przez długie lata. Zapewne posiadają dużo różnych baz danych, które niekoniecznie są zintegrowane, ale dodatkowo konfiguruje hurtownię, w której w sposób spójny można z nich korzystać.

Hurtownie danych są przeznaczone do przetwarzania OLAP (ang. *On-Line Analytical Processing*), czyli zoptymalizowanego pod kątem analityki i tworzenia raportów, a nie zapisywania i przechowywania danych, jak ma to miejsce w OLTP (ang. *On-Line Transaction Processing*). Najczęściej, gdy pojawiają się nowe obserwacje w źródłach danych, aktualizacją analiz i raportów zajmuje się system ETL (ang. *Extract, Transform, Load*), automatyzujący proces wydobywania danych ze źródeł, ich przekształcania do formy, w której

mogą zostać przechowane oraz załadowania do hurtowni danych, aby były dostępne. Automatyzacja raportów jest podstawą większości działów nazywających się BI (ang. *Business Intelligence*), którego rolą jest sporządzanie raportów mających na celu wspomaganie decyzji w przedsiębiorstwie.

Systemy BI oprócz analityki dostarczają również metod raportowania. Mamy zarazem raportowanie offline w formie generowanych dokumentów, co możemy zaimplementować samodzielnie bez dodatkowego oprogramowania w KnitR (pakiecie do R pozwalającym na generowanie dokumentów na podstawie kodu R) lub online, w formie tak zwanych paneli (ang. *Dashboard*). Na panelu widzimy na żywo stan obecny oraz często historię wybranych statystyk. Oprogramowanie takie możemy samodzielnie przygotować w HTML z wykorzystaniem ECMAScript (JavaScript), jednakże mamy też możliwość wykorzystania gotowego rozwiązania — darmowego jak Graphana<sup>1</sup> lub płatnego, jak Power BI<sup>2</sup>.

---

<sup>1</sup>Patrz <https://grafana.com/>.

<sup>2</sup>Patrz <https://powerbi.microsoft.com/pl-pl/>.