

BAZY DANYCH 2020

UPROSZCZONY MODEL HURTOWNI DANYCH LIBRARY

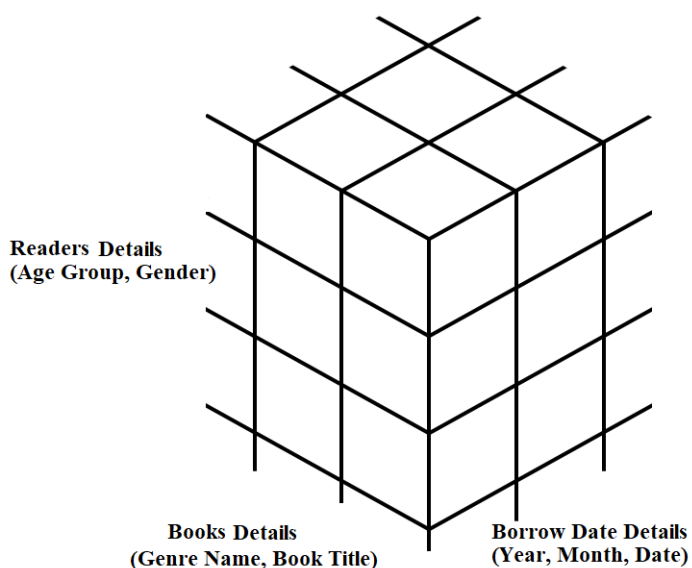
7 czerwca 2020

Elżbieta Jowik

1 Koncepcja

Generalnym zadaniem hurtowni danych jest uporządkowanie tematyczne krytycznych, z punktu widzenia organizacji, obszarów analitycznych, ujednolicenie informacji oraz udostępnienie ich do analizy zagadnień decyzyjnych. Przygotowany projekt nosi znamiona uproszczonego modelu, ponieważ agreguje dane pochodzące z pojedynczej bazy danych, a nie z wielu różnych systemów. Istotnym czynnikiem decyzyjnym z punktu widzenia działalności biblioteki jest badanie gatunkowych trendów czytelniczych w poszczególnych grupach wiekowych, czy też dla poszczególnych płci. W oparciu o takie dane można na przykład podejmować decyzje pozwalające na dostosowanie księgozbioru do bieżących oraz prognozowanych potrzeb czytelników, a także dokonywać zlecenia kupna takich książek, które pozwolą na powiększenie grona wypożyczających.

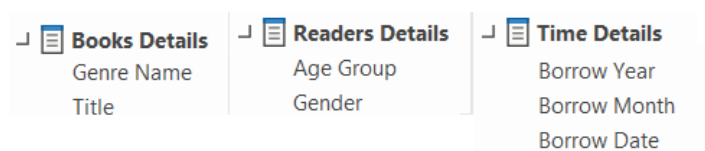
Logiczny model wielowymiarowy pozwalający na realizację powyższej koncepcji ma następującą postać:



Rysunek 1: Proponowana organizacja wybranych danych w kostce.

W sugerowanym modelu jedyną analizowaną **miarą** jest liczba wypożyczeń, obliczona w oparciu o pojedyncze zdarzenia przechowywane w tabeli faktów.

Wymiary natomiast są definiowane przez atrybuty ujęte na Rysunku 1. w następujących hierarchiach:

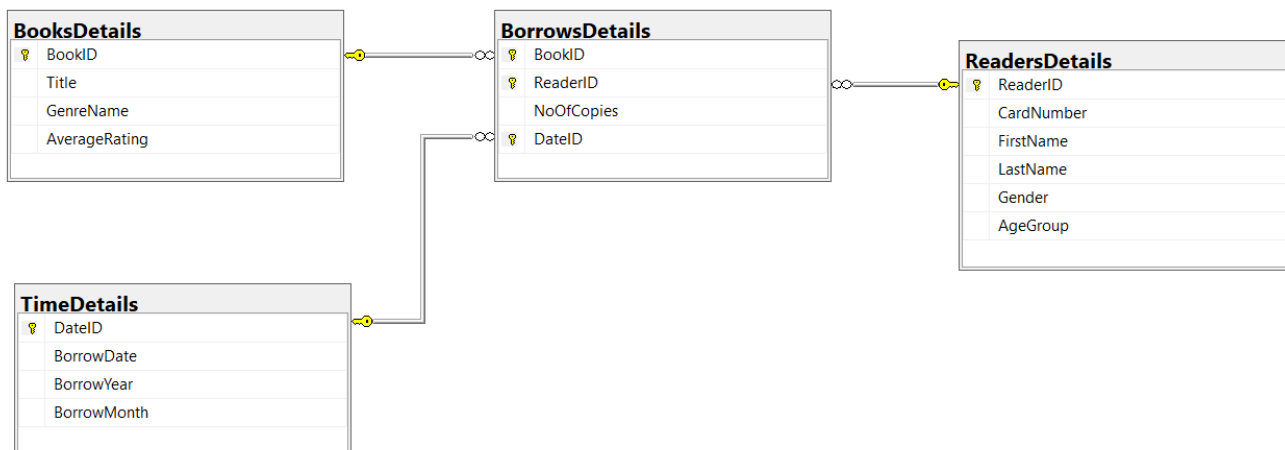


Rysunek 2: Hierarchia zdefiniowanych wymiarów

2 Praktyczna realizacja głównych założeń

W rozważanym projekcie zastosowana została prosta struktura, model gwiazdy.

Jako centralną tabelę faktów przyjęto tabelę *BorrowsDetails*, przechowującą informacje o poszczególnych wypożyczeniach. Jest to tabela, która z punktu widzenia praktycznych zastosowań podlega częstym aktualizacjom i szybko się powiększa. Charakteryzuje się ona wieloatrybutowym kluczem głównym złożonym z kluczy obcych odwołujących się do wymiarów. Wymiary te ustalają kontekst analizy i definiowane są przez atrybuty tabel *BooksDetails*, *ReadersDetails* i *TimeDetails*.



Rysunek 3: Diagram bazy

Dane w poszczególnych tabelach zostały w znacznym stopniu ograniczone. W celu zachowania prostoty struktury wyeliminowano wszystkie klucze obce odwołujące się do tabel nieujętych w proponowanym uproszczonym modelu hurtowni.

Zaproponowana struktura pozwala na stworzenia zestawienia, w którym wybory gatunkowe czytelników są prezentowane jako funkcja czasu z możliwością ograniczenia prezentowanych danych do dotyczących poszczególnych grup wiekowych tudzież płci.

Age Group	All	
Gender	All	
Borrows Count	Etykiety kolumn	
	+ 2018	+ 2019 - 2020
		+ 1 + 2 + 3 - 4
Etykiety wierszy		+ 5
	2020-04-04	2020-04-09 2020-04-11 2020-04-13 2020-04-16
+ Fantasy	4	3 2 9
+ Horror		2 1 2 5
- Mystery		
Gone Girl		1 1 2
In the Woods	2	1 3
Still Life		1 1
The Adventures of Sherlock Holmes	1	1 1 4
The Devil in the White City		1 1
+ Romance		1 1
+ Science Fiction		1 1
+ Thriller	3	3 2 1 1 1 4 3
Suma końcowa	9	4 1 1 3 3 2 1 1 1 4 3

Rysunek 4: Końcowe przykładowe zestawienie, w którym analizowaną miarą jest liczność wypożyczeń. W kolumnach znajduje się pełna hierarchia wymiaru *Time Details* (atr. *Year, Month, Date*), w wierszach hierarchia wymiaru *Books Details* (atr. *Genre Name, Title*). Atrybuty tożsame z grupą wiekową i płcią czytelnika są czynnikami filtrującymi.