

26 Maja 2020

Antropologia

BD PD 2

Tomek Karwowski, 284522

1 Oświadczenie o samodzielności pracy

Oświadczam że pracę wykonałem samodzielnie, nie korzystając z żadnych niedozwolonych źródeł. Przy informacjach zaczerpniętych z internetu zawsze podałem źródła.

2 Opis zrealizowanego zadania z uwzględnieniem wygenerowanego modelu bazy

Diagram ER w pliku Anthropology.png

Rozwiązanie składa się z 6 tabel 'głównych' (Post Mortem (PM), Ante Mortem (AM), Antropolodzy, Instytucje, Przedmioty i Znaki szczególne), 6 tabel typu 'enum' (Płeć, Rasa, Kolor oczu, Status, Kraj, Rodzaj wydarzenia) oraz 4 tabel pozwalających zapisać relacje M2M (PM-Przedmioty, AM-Przedmioty, PM-ZnakiSzczególne, AM-ZnakiSzczególne).

Tablice enum argumentuję potrzebą ograniczenia literówek ('gren' nie zmaczuje 'green') i utrzymania możliwie małej liczby wartości (np. 'light green' nie zmaczuje 'green'). Jednocześnie oddanie nowych wartości jeśli jest konieczne jest możliwe.

Mimo dużego podobieństwa tabel PM i AM zdecydowałem się ich nie łączyć, ze względu na to że reprezentują one różne obiekty (oraz ponieważ wynikało tak z kontekstu zadania). Relacje 1-to-1 reprezentuję kluczem obcym w tabeli PM oraz customowym indeksem unique (unique albo null). Relację zwrtoną można otrzymać za pomocą prostej kwerendy, dodatkowo przyspieszam ją użyciem indeksu.

3 Lista indeksów i przyczyn wyboru

Nie wiedziałem czy indeksy tworzyć na podstawie moich przypuszczeń jakie zapytania będą wykonywane na bazie, czy przykładowych kwerend z zadania. Wybrane indeksy są w pewien sposób połączeniem tych dwóch.

Oczywiście każdy klucz główny automatycznie generuje indeks zgrupowany dla swojej tabeli. Dodatkowo mamy jeden indeks unikalny dla kolumny 'ante-mortem-id' tabeli PostMortems który wspomaga relację 1-to-1 między tabelami PM i AM.

Dodatkowe indeksy niezgrupowane, które wybrałem w celu optymalizacji zapytań to:

- Anthropologists (affiliation-id) - wykorzystywane w kwerendzie 3. oraz sensowny klucz do grupowania antropologów w innych zapytaniach, liczba antropologów nie będzie rosła tak drastycznie jak np. w tabelach PM i AM ani nie będzie często modyfikowana.
- Events (type) oraz Events (country) - dla obu kolumn wyobrażam sobie wiele zastosowań. Liczenie ilości katastrof w danym kraju (tablica Event łączy kraje z ofiarami), liczenie globalnych statystyk światowej ilości ofiar z podziałem na powody, etc. Tak samo, ilość wydarzeń nie będzie rosła szybko (miejmy nadzieję).
- PostMortems (ante-mortem-id) - starałem się uniknąć indeksów na tabelach PM i AM ze względu na szybko rosnącą liczbę rekordów oraz fakt, że większość indeksów na kolumny byłaby wykorzystana tylko raz, w procedurze znajdowania par AM-PM. Jednak połączenie 1-to-1 AM-PM będzie wykorzystywane wyjątkowo często i dlatego indeks niezgrupowany na tym polu wydaje mi się konieczny.

4 Komplet testów potwierdzających poprawność rozwiązania (w tym wyniki wykonania kwerend SQL)

Przerosło mnie dodawanie zdjęć do LaTeX'a, wszystkie zdjęcia w załączniku

- Kwerenda 1 - query1.png
- Kwerenda 2 - query2.png
- Kwerenda 3 - query3.png
- Kwerenda 4 - query4.png
- Kwerenda 5 - query5.png

5 Lista załączników z kodem źródłowym

- Skrypt tworzenia bazy - anthropology-create-db.sql
- Populacja bazy - anthropology-insert-fixtures.sql
- Zadane kwerendy - anthropology-queries.sql
- Procedura parowania AM z PM - anthropology-matching-procedure.sql
- (Bonus) Procedura parowania (muszą się zgadzać dokładnie wszystkie pola, ale za to jest prostsza i bardziej kwerendowalna) - anthropology-matching-procedure-querycentric.sql