

## Wykorzystanie bazy dokumentowej MongoDB

**Termin oddania:** 9.12.2022, godzina 7:00

**Kara za przekroczenie terminu:** -0,5 za każde rozpoczęte 12h;

**Punkty:** maksymalnie 12 (uzasadnienie wyboru i złożoność zbioru/zbiorów danych – 2, zagnieżdżenia i referencje – 1, logika biznesowa – 3, indeksy – 2, transakcje – 1, omówienie funkcjonalności i wnioski – 4);

**Wymagania minimalne:** Zainstalowana i skonfigurowana baza MongoDB zawierająca co najmniej 100 nietrywialnych dokumentów (więcej niż 3 pola), założone indeksy co najmniej w jednej kolekcji **wraz z praktycznym porównaniem** wydajności zapytań, 3 różne funkcje (co najmniej jedna wykorzystuje **dokumenty zagnieżdżone lub referencje**), zrealizowana autoinkrementacja.

**Zakres:**

- 1) Wybierz zbiór (lub zbiory) danych, który wykorzystasz podczas projektu i **uzasadnij wybór**. Dane powinny być **właściwe do bazy dokumentowej** i odpowiednich rozmiarów (co najmniej 100 dokumentów do spełnienia wymagań minimalnych, dla uzyskania pełnych punktów **zbiory powinny być znacząco większe**, jak również posiadać dokumenty o złożonej strukturze i dużej liczbie pól). **Wybrany zbiór danych wpisz na MS Teams wraz z krótkim opisem problemu i linkiem do 2.12 włącznie**. Z tego samego zbioru danych może korzystać tylko jedna osoba w projekcie. Mile widziane jest wykorzystanie zbiorów danych dotyczącej problematyki, która jeszcze nie została przez nikogo wybrana. **Jeśli chcesz zmienić zbiór danych skontaktuj się najpierw z prowadzącym**.
- 2) Zainstaluj i skonfiguruj MongoDB, zaimportuj i przetwórz potrzebne dane. Zdecyduj w jakich kolekcjach będą się znajdować i **uzasadnij wybór**. Zmodyfikuj dane tak, żeby skorzystać **zarówno z dokumentów zagnieżdżonych, jak i referencji** do powiązania danych ze sobą. Przedstaw różnice i pokaż w jaki sposób można skorzystać z obu rodzajów (**praktycznie**, w postaci złożonych zapytań).
- 3) Przygotuj potrzebne elementy, żeby obsłużyć logikę biznesową po stronie bazy danych – funkcje do dodawania i modyfikacji elementów w kolekcji, sekwencje do autoinkrementacji, agregacje, itd. Liczy się zarówno złożoność, jak i kompletność obsłużenia logiki biznesowej. Jak w poprzednich ćwiczeniach – w przypadku elementów (np. funkcji) o analogicznej konstrukcji wystarczy je wypisać, nie potrzeba implementować wszystkich.
- 4) Stwórz indeksy na odpowiednich polach kolekcji. Przedstaw na **praktycznych przykładach** wydajność zapytań pokrytych różną liczbą indeksów.  
Uwaga: Dla pełnych punktów polecane jest porównanie zapytań pokrytych różną liczbą indeksów oraz indeksami prostymi i kompozytowymi oraz wnioski prezentujące zrozumienie działania indeksów w MongoDB (i czym się różnią od działania indeksów w bazach relacyjnych)
- 5) Pokaż na **praktycznym przykładzie** jak działa mechanizm transakcyjny w używanej wersji MongoDB. W przypadku występowania kilku mechanizmów transakcyjnych krótko omów pozostałe.
- 6) Wybierz, **zaprezentuj w praktyce (przetestuj) i oceń** do trzech zaawansowanych funkcjonalności MongoDB, które pasują do wybranego projektu (np. obsługa Map/Reduce, capped collections, GridFS, zapytania geoprzestrzenne, przetwarzanie grafów, obsługa dużych danych tekstowych, szyfrowanie, zapytania ad hoc, load balancing, ...). Pełne punkty

można otrzymać w przypadku dokładnego przedstawienia, omówienia i przetestowania **bardziej złożonych** funkcjonalności SZBD.

- 7) Krótko i treściwie (max ½ strony A4) przedstaw wnioski dotyczące zastosowania bazy dokumentowej dla wybranego problemu i praktycznych różnic, jakie są w stosunku jakbyśmy mieli zamiast tego skorzystać z bazy relacyjnej.

**Uwaga** – w przypadku zmiany zbioru danych proszę to wyraźnie zaznaczyć na MS Teams w nowym komentarzu, podając zarówno stary zbiór danych, nowy zbiór danych i powód zmiany. Zmiana w innym przypadku (np. usunięcie / edycja starego komentarza) będzie skutkowałą otrzymaniem 0 punktów za wybór zbioru danych.

## Wykorzystanie bazy grafowej Neo4j

**Termin oddania:** 13.01.2023, godzina 7:00

**Kara za przekroczenie terminu:** 0 punktów za dostarczenie pracy po 21.01, w innym przypadku brak kary punktowej. Prace będą sprawdzane zaczynając od terminowych, a potem w kolejności złożenia – więc im później złożona praca, tym późniejsza informacja o zaliczeniu przedmiotu i tym mniej czasu na ewentualną poprawę.

**Punkty:** maksymalnie 13 (uzasadnienie wyboru i złożoność obu zbiorów danych – 2, zapytania – 2, indeksy i optymalizacja zapytań – 2, funkcje przestrzenne - 2, procedura – 1, analiza zbioru danych, modyfikacje i wnioski – 2, APOC - 2);

**Wymagania minimalne:** Zainstalowana i skonfigurowana baza Neo4j zawierająca wczytane dane (**co najmniej 100 połączonych węzłów 5 rodzajów, co najmniej 5 rodzajów krawędzi**), skrypty umożliwiające odtworzenie całej pracy na innej maszynie z zainstalowanym Neo4j, **co najmniej 2 nietrywialne zapytania**, utworzony indeks i **oceniona jego skuteczność, stworzona i omówiona procedura**.

**Zakres:**

- 1) Wybierz zbiór danych, który wykorzystasz podczas projektu i uzasadnij wybór. Zbiór musi być odpowiednich rozmiarów (**co najmniej 100 połączonych węzłów 5 rodzajów, co najmniej 5 rodzajów krawędzi**, jest to wymaganie minimalne, w celu otrzymania punktów zbiory **powinny być znacząco większe i bardziej skomplikowane**). Wybrany zbiór danych wpisz na MS Teams wraz z krótkim opisem, linkiem i informacjami dotyczącymi rozszerzania (**więcej o tym w punkcie 3**). Można skorzystać ze zbioru danych wykorzystanego do MongoDB i takiego, który został wykorzystany już przez innego studenta.  
Te same informacje umieść również w sprawozdaniu.
- 2) Zainstaluj i skonfiguruj Neo4j, zaimportuj i przetwórz potrzebne dane. W razie potrzeby utwórz potrzebne relacje. **Nie dokumentuj całej instalacji** – przedstaw tylko problemy i rozwiązania (jeśli będą) oraz rezultat (prezentując na przykład interesujący wycinek zbioru danych i go omawiając).
- 3) Dołącz drugi, zbliżony zbiór danych i powiąż go z pierwszym zbiorem danych (przykładowo – jeśli nasz pierwszy zbiór danych dotyczył postaci i miejsc w Grze o Tron to drugi może być na przykład mapą Westeros, zbiorem danych o aktorach i filmach czy

informacjami o stoczonych bitwach). Można wykorzystać istniejący zbiór danych lub stworzyć go samemu. Umieść informację o tym, w jaki sposób planujesz rozszerzyć pierwszy zbiór danych i do czego wykorzystać – **zaprezentuj to przy pomocy zestawu co najmniej 5 zapytań (złożoność zapytań jest istotna dla oceny, co najmniej jedno powinno korzystać z UNION, co najmniej jedno powinno korzystać z MERGE)**. Jeśli oba zbiory były istniejące utwórz i powiąż z oboma zbiorami co najmniej 5 nowych, stworzonych przez siebie, węzłów (**udokumentuj i wyjaśnij powiązania**).

- 4) Dla przygotowanych zapytań stwórz indeksy – przedstaw zasadę działania indeksów w Neo4j, **ocień w praktyce ich skuteczność**. Omów jakie jeszcze indeksy warto stworzyć w wykorzystywanym zbiorze danych i dlaczego. Zapoznaj się z **mechanizmami optymalizacji zapytań w Neo4j, przedstaw praktycznie**, jak przygotowane przez ciebie zapytania można wykonać szybciej.  
Uwaga: Dla uzyskania pełnych punktów trzeba przestawić również inne mechanizmy optymalizacji zapytań niż tylko indeksy.
- 5) Przygotuj (jeśli nie ma) zestaw danych przestrzennych dla swojego projektu – **przedstaw w praktyce wykorzystanie, wady i zalety** funkcji przestrzennych w Neo4j. **Wykonaj co najmniej jedno zapytanie** wykorzystujące agregacje oraz co najmniej jedno korzystające z funkcji odległości (np. najkrótsza ścieżka), omów ich działanie i wydajność. **Premiowane jest skorzystanie z większej liczby funkcjonalności przestrzennych**, jeśli to możliwe dla wybranych zbiorów danych.
- 6) Stwórz co najmniej jedną procedurę i wywołaj ją przy pomocy CALL. Omów w jaki sposób procedury w Neo4j różnią się w stosunku do relacyjnych baz danych oraz co czego możemy je zastosować.
- 7) **Przeanalizuj końcowy zbiór danych** – oceń w jaki sposób warto byłoby go umieścić na kilku fizycznych maszynach przy rozwiązaniu produkcyjnym (o znacząco większym rozmiarze danych) w celu uzyskania maksymalnej wydajności, znajdź mosty i węzły przegubowe, wyszczególnij podgrafy. Omów jaki mają wpływ na zbiór danych w projekcie. **Dla pełnej punktacji zaproponuj i wdróż potrzebne modyfikacje, pokaż różnice jakie spowodowały.**
- 8) Zapoznaj się z biblioteką APOC (<https://neo4j.com/labs/apoc/>) – wybierz i omów najciekawsze możliwości jakie daje, pokaż ich zastosowanie **w praktyce** na swoim zbiorze danych (lub jeśli się nie da można na innym), wybór uzasadnij.

**Pamiętaj o przygotowaniu i załączeniu wraz ze sprawozdaniem skryptów, które umożliwią załadowanie obu zbiorów danych i odtworzenie całej wykonanej pracy.**

Przed przystąpieniem do ćwiczenia warto przejrzeć przykładowe zbiory danych i zapytania z nimi powiązane na <https://neo4j.com/developer/example-data/> - zdecydowanie pomaga w zrozumieniu działania Neo4j.