



KARTA OPISU PRZEDMIOTU - SYLABUS

Nazwa przedmiotu

Rozproszone bazy danych [S2Inf1>RBD]

Przedmiot

Kierunek studiów
Informatyka

Rok/Semestr
1/2

Studia w zakresie (specjalność)
Technologie przetwarzania danych

Profil studiów
ogólnoakademicki

Poziom studiów
drugiego stopnia

Język oferowanego przedmiotu
polski

Forma studiów
stacjonarne

Wymagalność
obligatoryjny

Liczba godzin

Wykład
30

Laboratorium
30

Inne
0

Ćwiczenia
0

Projekty/seminaria
0

Liczba punktów ECTS

4,00

Koordynatorzy

prof. dr hab. inż. Tadeusz Morzy
dr inż. Juliusz Jezierski
juliusz.jezierski@put.poznan.pl

Wykładowcy

Wymagania wstępne

Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu systemów baz danych. Niezbędna jest umiejętność formułowania zapytań w języku SQL oraz programowania w języku PL/SQL. Powinien posiadać umiejętność rozwiązywania podstawowych problemów z zakresu administracji i obsługi systemów baz danych oraz umiejętność pozyskiwania informacji ze wskazanych źródeł. Powinien również rozumieć konieczność poszerzania swoich kompetencji. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.

Cel przedmiotu

1. Przekazanie studentom podstawowej wiedzy dotyczącej technologii i metod stosowanych w systemach rozproszonych baz danych. 2. Rozwijanie u studentów umiejętności rozwiązywania problemów związanych z projektowaniem i eksploatacją systemów rozproszonych baz danych. 3. Kształtowanie u studentów umiejętności pracy zespołowej oraz integracji wiedzy z różnych obszarów informatyki. 4. Rozwijanie u studentów umiejętności formułowania i testowania hipotez związanych z problemami inżynierskimi i prostymi problemami badawczymi w zakresie systemów rozproszonych baz danych. 5. Przekazanie studentom praktycznej wiedzy dotyczącej technologii z wykorzystaniem wiodących dostawców systemów rozproszonych baz danych.

Przedmiotowe efekty uczenia się

Wiedza:

posiada zaawansowaną i pogłębioną wiedzę z zakresu systemów rozproszonych baz danych, ich podstaw teoretycznych oraz metod, narzędzi i środowisk programistycznych wykorzystywanych do ich implementacji (k2st_w1)

posiada zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień z zakresu problematyki systemów rozproszonych baz danych (k2st_w3)

posiada szczegółową i zaawansowaną wiedzę o procesach zachodzących w cyklu życia systemu informatycznego wykorzystującego system rozproszonej bazy danych (k2st_w5)

zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu problemów inżynierskich i prac badawczych w zakresie budowy systemów rozproszonych baz danych (k2st_w6)

Umiejętności:

potrafi, przy formułowaniu i rozwiązywaniu zadań inżynierskich, integrować wiedzę z różnych obszarów informatyki (k2st_u5)

potrafi ocenić przydatność i możliwość wykorzystania nowych metod, technik i narzędzi informatycznych (k2st_u6)

potrafi rozwiązywać złożone zadania informatyczne w zakresie systemów rozproszonych, w tym zadania nietypowe lub zawierające komponent badawczy (k2st_u10)

potrafi, zgodnie z podaną specyfikacją, zaprojektować i zaimplementować złożony system rozproszonej bazy danych używając właściwych metod, technik i narzędzi (k2st_u11)

potrafi określić i zrealizować kierunki dalszego samokształcenia (k2st_u16)

Kompetencje społeczne:

rozumie, że we współczesnej informatyce wiedza i umiejętności stają się szybko przestarzałe i wymagają ciągłej aktualizacji (k2st_k1)

rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych (k2st_k2)

Metody weryfikacji efektów uczenia się i kryteria oceny

Efekty uczenia się przedstawione wyżej weryfikowane są w następujący sposób:

Efekty kształcenia przedstawione wyżej weryfikowane są w następujący sposób:

Ocena formująca:

a) w zakresie wykładów:

- na podstawie ocen realizowanych ćwiczeń/zadań przy tablicy

b) w zakresie laboratoriów:

- na podstawie oceny bieżącego postępu realizacji zadań,

Ocena podsumowująca:

a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę wiedzy i umiejętności wykazanych na otwartym kolokwium pisemnym o charakterze problemowym (student może korzystać z dowolnych materiałów dydaktycznych), Kolokwium składa się z 5-6 zadań problemowych, za które można uzyskać 10 pkt. Łącznie można uzyskać od 50-60 pkt. Zaliczenie na ocenę 3.0 wymaga uzyskania 50% maksymalnej liczby punktów.

- omówienie wyników egzaminu,

b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:

- ocenę stopnia przyswojenia wiedzy prezentowanej podczas laboratorium poprzez pisemne raporty z wykonanych ćwiczeń

- test wielokrotnego wyboru oceniający umiejętność syntezy zdobytej wiedzy

- omówienie wyników

Treści programowe

Wprowadzenie do systemów rozproszonych baz danych.
Pozioma i pionowa fragmentacja danych.
Alokacja danych.
Dekompozycja, lokalizacja i optymalizacja zapytań rozproszonych.
Zarządzanie transakcjami rozproszonymi.
Synchroniczna i asynchroniczna replikacja danych.

Tematyka zajęć

Wprowadzenie do systemów rozproszonych baz danych: systemy rozproszonych baz danych, systemy wielobazowe, systemy równoległych baz danych. Architektury systemów rozproszonych. Schematy fragmentacji poziomej: dezyderaty, fragmentacja podstawowa i wywiedziona, algorytm definiowania schematu fragmentacji poziomej. Schemat fragmentacji pionowej: dezyderaty, algorytm BEA definiowania schematu fragmentacji pionowej. Schemat alokacji danych: sformułowanie problemu, wymagania informacyjne odnośnie schematu alokacji danych, podstawowe heurystyki alokacji danych w systemie rozproszonej bazy danych. Rozproszone przetwarzanie zapytań: wprowadzenie, semantyczna analiza zapytania, dekompozycja zapytania względem fragmentów, lokalizacja zapytania względem fragmentów, optymalizacja zapytań rozproszonych, algorytm półpołączeń, algorytm SDD-1. Zarządzanie rozproszonymi transakcjami: koncepcje i modele transakcji rozproszonych. Zarządzanie współbieżnym wykonywaniem transakcji rozproszonych: kryterium globalnej uszeregowalności, kryterium quasi-uszeregowalności, lokalne warunki poprawności współbieżnego wykonywania transakcji. Algorytmy zarządzania współbieżnym wykonywaniem transakcji rozproszonych: blokowanie, algorytm porządkowania transakcji wg. etykiet czasowych, zagadnienie zakleszczenia i metody jego rozwiązywania. Replikacja danych: sformułowanie problemu. Replikacja synchroniczna i asynchroniczna. Kryteria spójności. Kryterium poprawności współbieżnego wykonywania transakcji (1-copy serializability). Algorytm weryfikacji kryterium. Reguła Thomasa. Podstawowe algorytmy replikacji. Problem replikacji asynchronicznej typu multi-master: sformułowanie problemu, twierdzenie CAP, kryteria i modele spójności (eventual consistency, casual consistency, read committed, snapshot isolation), koncepcja „vector clocks”. Niezawodność w systemach rozproszonych baz danych: algorytm 2PC. Rozproszone bazy danych NoSQL: koncepcje, wady i zalety, model BASE. Zajęcia laboratoryjne prowadzone są w formie 15 2-godzinnych ćwiczeń, odbywających się w laboratorium, poprzedzonych 1-godzinną sesją instruktażową na początku semestru. Ćwiczenia realizowane są indywidualnie. Program laboratorium obejmuje następujące zagadnienia: instalowanie oprogramowania i tworzenie rozproszonej bazy danych. Transparenty dostęp do zdalnych danych. Rozproszone transakcje. Symulowanie awarii i odtwarzanie protokołu 2PC. Obsługa rozproszonych zakleszczeń. Podstawowa replikacja danych. Fragmentaryzacja pozioma obiektów danych (ang. sharding). Optymalizacja zapytań kierowanych do fragmentaryzowanych poziomo tabel. Replikacja fragmentów tabel. Fizyczna i logiczna replikacja strumieniowa. Asynchroniczna replikacja dwukierunkowa. Rodzaje konfliktów aktualizacji danych i metody ich rozstrzygnięcia. Replikacja operacji DDL. Materializowane perspektywy i metody ich odświeżania. Konfigurowanie i testowanie mechanizmu przełączenia awaryjnego. Rekonfiguracja replikacji po trwałym utraceniu quorum. Zajęcia są prowadzone z wykorzystaniem oprogramowania dwóch post-relacyjnych baz danych: Postgres i Oracle oraz dwóch noSQL-owych baz danych: Cassandra i Mongo.

Metody dydaktyczne

Wykład: prezentacja multimedialna, ilustrowana przykładami podawanymi na tablicy.
Laboratorium: prezentacja multimedialna prezentacja ilustrowana przykładami podawanymi na tablicy oraz wykonanie zadań podanych przez prowadzącego - ćwiczenia praktyczne.

Literatura

Podstawowa

1. M. T. Ozsú, P. Valduriez, Principles of Distributed Database Systems, Springer, 3rd ed., 2011.

Uzupełniająca

1. S. K. Rahimi, F. S. Haug, Distributed database management Systems: A Practical Approach, Wiley, 2010.

2. P.A. Bernstein, E. Newcomer, Principles of Transaction processing, Morgan Kaufmann, 2nd ed., 2009.

Bilans nakładu pracy przeciętnego studenta

	Godzin	ECTS
Łączny nakład pracy	100	4,00
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	60	2,00
Praca własna studenta (studia literaturowe, przygotowanie do zajęć laboratoryjnych/ćwiczeń, przygotowanie do kolokwiiw/egzaminu, wykonanie projektu)	40	2,00