

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Rozproszone bazy danych		Kod 1010512321010510947
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) (brak)	Rok / Semestr 1 / 2
Ścieżka obieralności/specjalność Systemy rozproszone	Przedmiot oferowany w języku: polski	Kurs (obligatoryjny/obieralny) obieralny
Stopień studiów: II stopień	Forma studiów (stacjonarna/niestacjonarna) stacjonarna	
Godziny Wykłady: 30 Ćwiczenia: - Laboratoria: 30 Projekty/seminaria: -		Liczba punktów 5
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) (brak)		(ogólnouczelniany, z innego kierunku) (brak)
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne nauki techniczne		Podział ECTS (liczba i %) 5 100% 5 100%
Odpowiedzialny za przedmiot / wykładowca:		
prof. dr hab. inż. Tadeusz Morzy email: tadeusz.morzy@put.poznan.pl tel. (0-61) 665-2906 Informatyki ul. Piotrowo 2, 60-965 Poznań		dr inż. Juliusz Jezierski email: juliusz.jezierski@cs.put.poznan.pl tel. (0-61) 665-2961 Informatyki ul. Piotrowo 2, 60-965 Poznań
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Efekty kształcenia ze studiów I stopnia: K1st_W1-8, K1st_U2-14, weryfikowane w procesie rekrutacji na studia 2 stopnia ? efekty te prezentowane są w serwisie internetowym wydziału www.fc.put.poznan.pl Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu systemów baz danych ze studiów I stopnia.
2	Umiejętności:	Do realizacji zajęć laboratoryjnych konieczna jest podstawowa znajomość języków programowania Java oraz Python. Student powinien posiadać umiejętność rozwiązywania podstawowych problemów z zakresu przetwarzania i analizy danych oraz umiejętność pozyskiwania informacji ze wskazanych źródeł. Powinien również rozumieć konieczność poszerzania swoich kompetencji / mieć gotowość do podjęcia współpracy w ramach zespołu.
3	Kompetencje społeczne	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		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		
1. 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] 2. posiada zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień z zakresu problematyki systemów rozproszonych baz danych - [K2st_W3] 3. posiada szczegółową i zaawansowaną wiedzę o procesach zachodzących w cyklu życia systemu informatycznego wykorzystującego system rozproszonej bazy danych - [K2st_W5] 4. zna zaawansowane metody, techniki i narzędzia stosowane przy rozwiązywaniu problemów inżynierskich i prac badawczych w zakresie budowy systemów rozproszonych baz danych - [K2st_W6]		

Umiejętności:
1. potrafi, przy formułowaniu i rozwiązywaniu zadań inżynierskich, integrować wiedzę z różnych obszarów informatyki - [K2st_U5]
2. potrafi ocenić przydatność i możliwość wykorzystania nowych metod, technik i narzędzi informatycznych - [K2st_U6]
3. potrafi rozwiązywać złożone zadania informatyczne w zakresie systemów rozproszonych, w tym zadania nietypowe lub zawierające komponent badawczy - [K2st_U10]
4. potrafi, zgodnie z podaną specyfikacją, zaprojektować i zaimplementować złożony system rozproszonej bazy danych używając właściwych metod, technik i narzędzi - [K2st_U11]
5. potrafi określić i zrealizować kierunki dalszego samokształcenia - [K2st_U16]
Kompetencje społeczne:
1. rozumie, że we współczesnej informatyce wiedza i umiejętności stają się szybko przestarzałe i wymagają ciągłej aktualizacji - [K2st_K1]
2. rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i praktycznych - [K2st_K2]

Sposoby sprawdzenia efektów kształcenia
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 / ćwiczeń: - 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 / ćwiczeń 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: 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 muliti-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. Integracja danych rozproszonych: podejście wielobazowe i systemy mediacyjne. 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 kworum. Zajęcia są prowadzone z wykorzystaniem oprogramowania dwóch post-relacyjnych baz danych: Postgres i Oracle oraz dwóch noSQL-owych baz danych: Cassandra i Mongo.		
Literatura podstawowa:		
1. M. T. Ozsú, P. Valduriez, Principles of Distributed Database Systems, Springer, 3rd ed., 2011.		
Literatura 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		
Czynność		Czas (godz.)
1. udział w zajęciach laboratoryjnych		30
2. przygotowanie do ćwiczeń laboratoryjnych		10
3. udział w konsultacjach związanych z realizacją procesu kształcenia		2
4. przygotowanie projektów		5
5. przygotowanie do sprawdzianów i kolokwium		10
6. udział w wykładach		30
7. przygotowanie do zaliczenia wykładu i udział w kolokwium zaliczeniowym		20
Obciążenie pracą studenta		
forma aktywności	godzin	ECTS
Łączny nakład pracy	107	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	64	3
Zajęcia o charakterze praktycznym	45	2