

KARTA OPISU MODUŁU KSZTAŁCENIA		
Nazwa modułu/przedmiotu Rozproszone systemy baz danych		Kod 1010512331010510729
Kierunek studiów Informatyka	Profil kształcenia (ogólnoakademicki, praktyczny) ogólnoakademicki	Rok / Semestr 2 / 3
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 4
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) kierunkowy		(ogólnouczelniany, z innego kierunku) z danego kierunku
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki nauki techniczne		Podział ECTS (liczba i %) 4 100%
Odpowiedzialny za przedmiot / wykładowca:		
<p>prof. dr hab. inż. Tadeusz Morzy email: Tadeusz.Morzy@put.poznan.pl tel. 61 6652906 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań</p>		
Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:		
1	Wiedza:	Student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu systemów baz danych.
2	Umiejętności:	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.
3	Kompetencje społeczne	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:		
<p>1. Wskazanie praktycznych problemów jakie rozwiązuje się zarządzając rozproszonymi bazami danych w dużej instytucji lub przedsiębiorstwie.</p> <p>2. Przedstawienie problematyki budowy systemu rozproszonych baz danych, w zakresie:</p> <ul style="list-style-type: none"> - tworzenia połączeń między rozproszonymi bazami danych, - zadawania zapytań w rozproszonych systemach baz danych, - architektur replikacji danych, - mechanizmów odświeżania zdalnych replik danych, - algorytmów zatwierdzania rozproszonych transakcji (2PC, 3PC), - optymalizacji zapytań rozproszonych. <p>3. Rozwijanie u studentów umiejętności rozwiązywania problemów, w zakresie:</p> <ul style="list-style-type: none"> - projektowania i implementowania rozproszonych baz danych, - replikacji danych, - zarządzania transakcjami rozproszonymi w przypadku awarii systemu, - optymalizacji zapytań rozproszonych. 		
Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia		
Wiedza:		

1. ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną w zakresie architektur systemu rozproszonych baz danych, algorytmów zatwierdzania transakcji rozproszonych, replikacji danych, optymalizacji zapytań rozproszonych - [K_W4]
2. ma podbudowaną teoretycznie szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu informatyki, takimi jak: replikacja danych, algorytm 2PC, optymalizacja zapytań rozproszonych - [K_W5]
3. ma wiedzę o trendach rozwojowych i najistotniejszych nowych osiągnięciach w zakresie systemów rozproszonych bazach danych - [K_W6]
4. zna podstawowe techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich z optymalizacji zapytań rozproszonych, automatycznej replikacji danych - [K_W8]

Umiejętności:

1. potrafi pozyskiwać informacje z literatury, baz danych oraz innych źródeł (w języku ojczystym i angielskim), integrować je, dokonywać ich interpretacji i krytycznej oceny, wyciągać wnioski oraz formułować i wyczerpująco uzasadniać opinie - [K_U1]
2. potrafi określić kierunki dalszego uczenia się i zrealizować proces samokształcenia z zakresu systemu rozproszonych baz danych - [K_U5]
3. potrafi wykorzystać metody analityczne, symulacyjne oraz eksperymentalne do formułowania i rozwiązywania zadań z zakresu automatycznej replikacji baz danych, optymalizacji zapytań rozproszonych i usuwania awarii transakcji rozproszonych - [K_U9]
4. potrafi - przy formułowaniu i rozwiązywaniu zadań inżynierskich - integrować wiedzę z różnych obszarów informatyki (a w razie potrzeby także wiedzę z innych dyscyplin naukowych) oraz zastosować podejście systemowe, uwzględniające także aspekty pozatechniczne - [K_U10]
5. potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych - [K_U13]

Kompetencje społeczne:

1. rozumie, że w rozproszonych systemach baz danych wiedza i umiejętności bardzo szybko stają się przestarzałe i potrafi systematycznie zdobywać nową wiedzę i umiejętności z tej dziedziny - [K_K1]
2. potrafi odpowiednio określić priorytety służące realizacji określonego przez siebie lub innych zadania - [K_K6]

Sposoby sprawdzenia efektów kształcenia

Ocena formująca:

- a) w zakresie wykładów:
 - ze względu na zakres materiału, ocena formująca nie jest dokonywana.
- 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 kolokwium pisemnym otwartym o charakterze problemowym (student może korzystać z dowolnych materiałów dydaktycznych). Kolokwium składa się z 6-8 zadań problemowych, za które można uzyskać 10 pkt. Łącznie można uzyskać od 60-80 pkt. Zaliczenie kolokwium na ocenę 3.0 wymaga uzyskania 50% maksymalnej liczby punktów.
 - omówienie wyników kolokwium;
 - b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:
 - ocenę wiedzy i umiejętności w formie testu wielokrotnego wyboru
- Uzyskiwanie punktów dodatkowych w zakresie zajęć laboratoryjnych za aktywność podczas zajęć, a szczególnie za:
- omówienia dodatkowych aspektów zagadnienia,
 - uwagi związane z udoskonaleniem materiałów dydaktycznych,
 - wskazywanie trudności percepcyjnych studentów umożliwiające bieżące doskonalenia procesu dydaktycznego.

Treści programowe

Program wykładu obejmuje następujące zagadnienia:

Problemy integracji danych w środowisku rozproszonych heterogenicznych źródeł danych. Zagadnienie niespójności danych, różnej ziarnistości danych, różnych schematów źródeł danych. Podstawowe podejścia w zakresie integracji heterogenicznych źródeł danych: architektura sfederowanych baz danych, systemów mediacyjnych i systemów hurtowni danych. Wprowadzenie do problematyki rozproszonych baz danych (definicja rozproszonej bazy danych, zalety i wady rozproszonych baz danych). Schematy fragmentacji danych: poziomej, pionowej i mieszanej. Algorytm fragmentacji wywiedzionej. Algorytm fragmentacji danych. Schemat alokacji i metody alokacji danych w środowisku rozproszonej bazy danych. Zarządzanie transakcjami rozproszonymi (w tym: transakcja rozproszona, architektura zarządzania transakcjami rozproszonymi, protokół 2PC i jego odmiany)). Poprawność zarządzania współbieżnym wykonywaniem transakcji w środowisku homogenicznej i heterogenicznej rozproszonej bazy danych. Algorytmy zarządzania współbieżnym wykonywaniem transakcji w środowisku homogenicznej i heterogenicznej rozproszonej bazy danych. Algorytm zarządzania współbieżnym wykonywaniem transakcji w systemie współdzielenia danych (DCS). Replikacja danych: poprawność replikacji, typu replikacji (synchroniczna i asynchroniczna), odświeżanie replik. Optymalizacja wykonywania zapytań w środowisku rozproszonej bazy danych (metody szacowania kosztu wykonywania operacji). Zapytania cykliczne, acykliczne i reduktory. Algorytm optymalizacji wykonywania zapytań w środowisku rozproszonej bazy danych. Algorytm optymalizacji wykonywania zapytań w środowisku rozproszonej bazy danych z uwzględnieniem fragmentacji i replikacji danych. Wprowadzenie do problematyki hurtowni danych. Architektury hurtowni danych. Modelowanie pojęciowe hurtowni danych (metodyka projektowania). Wielowymiarowy model danych. Wielowymiarowa kostka danych. Struktury danych i indeksów w hurtowniach danych.

Program laboratorium obejmuje następujące zagadnienia:

Podstawy zarządzania rozproszonymi systemami baz danych: ogólna architektura rozproszonych systemów baz danych, zarządzanie zdalnymi połączeniami, problem uwierzytelniania użytkowników i szyfrowania danych, nazwy globalne baz danych, łączniki bazy danych, zapewnianie przezroczystości lokalizacji. Replikacja danych: architektura i mechanizm tworzenia migawek (perspektyw zmaterializowanych), typy replikacji standardowa (jednokierunkowa) pełna, standardowa przyrostowa, dwukierunkowa pełna, dwukierunkowa przyrostowa, problemy zdalnego odświeżania, warunki automatycznego odświeżania zdalnych migawek. Zbieranie informacji potrzebnych do replikacji migawek opartych o skomplikowane zapytania SQL i warunki ich odświeżania, grupy odświeżania migawek. Problem zatwierdzania transakcji rozproszonych: algorytm 2PC, symulowanie awarii procesu zatwierdzania, ćwiczenia praktyczne z naprawiania poszczególnych typów awarii transakcji rozproszonej. Optymalizacja zapytań rozproszonych: odpowiedzi do zapytań rozproszonych w różnych systemach zarządzania bazą danych, wskazywanie węzła przetwarzającego, warunki połączeń tabel w rozproszonych bazach danych, analiza kosztów zapytań w zależności od liczby i kierunku przesyłanych danych.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy, rozwiązywanie zadań.
2. ćwiczenia laboratoryjne: rozwiązywanie zadań, ćwiczenia praktyczne, wykonywanie eksperymentów, burza mózgów, pokaz multimedialny, studium przypadków

Literatura podstawowa:

1. T. Conolly, C. Begg, Systemy baz danych, Tom 1/Tom 2, Wydawnictwo READ ME, 2004.
2. S. K. Rahimi, F. S. Haug: Distributed Database Management Systems. A Practical Approach. IEEE Computer Society, Wiley, 2010
3. M. T. Özsu, P. Valduriez: Principles of Distributed Database Systems. 3rd edition, Springer Verlag, 2011

Literatura uzupełniająca:

1. K. Stocker, D. Kossmann, R. Braumandi, A. Kemper: Integrating semi-join-reducers into state-of-the-art query processors. Proc. of ICDE, 2001
2. Y-F. Huang, J-H. Chen, Fragment Allocation in Distributed Database Design, July 2000, pp. 491-506
3. A. Brunstrom, S. T. Leutenegger, R. Simha, Experimental Evaluation of Dynamic Data Allocation Strategies in a Distributed Database With Changing Workloads, no. TR-95-2, 1995, pp. 1-15
4. P. M. G. Apers, Data Allocation in Distributed Database Systems, ACM Transactions on Database Systems, vol. 13, no. 3, Sep-tember 1988, pp. 263-304

Bilans nakładu pracy przeciętnego studenta

Czynność	Czas (godz.)
1. udział w zajęciach laboratoryjnych / ćwiczeniach:	15
2. przygotowanie do ćwiczeń laboratoryjnych:	7
3. dokończenie (w ramach pracy własnej) sprawozdań z ćwiczeń laboratoryjnych:	7
4. udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności ćwiczeń laboratoryjnych / projektu (część konsultacji może być realizowana drogą elektroniczną)	6
5. napisanie programu / programów, uruchomienie i weryfikacja (czas poza zajęciami laboratoryjnymi)	5
6. przygotowanie do sprawdzianów / kolokwium	5
7. udział w wykładach	30
8. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.), 100 stron	10
9. przygotowanie do zaliczenia wykładu i obecność na zaliczeniu: 15 godz. + 2 godz.	17

Obciążenie pracą studenta

forma aktywności	godzin	ECTS
Łączny nakład pracy	102	4
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	53	2
Zajęcia o charakterze praktycznym	34	1