

# Trendy rozwojowe w inżynierii danych

Załącznik Nr 5 do Zarz. Nr 33/11/12

(pieczęć wydziału)

## KARTA PRZEDMIOTU

Z1-PU7

WYDANIE N1

Strona 1 z 4

<b>1. Nazwa przedmiotu: TRENDY ROZWOJOWE W INŻYNIERII DANYCH</b>	<b>2. Kod przedmiotu:</b>
<b>3. Karta przedmiotu ważna od roku akademickiego: 2018/19</b>	
<b>4. Forma kształcenia: studia drugiego stopnia</b>	
<b>5. Forma studiów: STUDIA STACJONARNE</b>	
<b>6. Kierunek studiów: INFORMATYKA, WYDZIAŁ AEII</b>	
<b>7. Profil studiów: ogólnoakademicki</b>	
<b>8. Specjalność: BAZY DANYCH I INŻYNIERIA SYSTEMÓW</b>	
<b>9. Semestr: 1, 2</b>	
<b>10. Jednostka prowadząca przedmiot: Instytut Informatyki, RAU2</b>	
<b>11. Prowadzący przedmiot: dr inż. Henryk Josiński</b>	
<b>12. Przynależność do grupy przedmiotów: przedmioty specjalnościowe</b>	
<b>13. Status przedmiotu: obowiązkowy</b>	
<b>14. Język prowadzenia zajęć: polski</b>	
<b>15. Przedmioty wprowadzające oraz wymagania wstępne:</b> Podstawy baz danych, Tworzenie aplikacji bazodanowych. Zakłada się, że przed rozpoczęciem niniejszego kursu student posiada przygotowanie w zakresie znajomości relacyjnego modelu danych, architektury systemu zarządzania bazą danych, umiejętności posługiwania się językiem SQL, projektowania schematu relacyjnej bazy danych oraz tworzenia aplikacji współpracującej z bazą danych.	
<b>16. Cel przedmiotu:</b> Przedmiot realizowany jest w formie wykładu oraz ćwiczeń laboratoryjnych. Celem wykładu jest przekazanie studentom wiedzy na temat zagadnień inżynierii danych, związanych z przetwarzaniem dużych, zmiennych i różnorodnych zbiorów danych: modeli i form opisu danych, organizacji fizycznej danych, obszarów zastosowań, wybranych mechanizmów systemu zarządzania bazą danych oraz języków zapytań. Celem ćwiczeń laboratoryjnych jest nabycie przez studentów umiejętności w zakresie przetwarzania różnorodnych zbiorów danych, których organizacja oparta jest na różnych modelach, w tym składowania danych, ich aktualizacji i realizacji zadań wyszukiwania, oraz stosowania wybranych mechanizmów systemu zarządzania bazą danych.	

<b>17. Efekty kształcenia:</b>				
Nr	Opis efektu kształcenia	Metoda sprawdzenia efektu kształcenia	Forma prowadzenia zajęć	Odniesienie do efektów dla kierunku studiów
w1	Student ma wiedzę w zakresie modeli danych, na których opierają się bazy danych, rozumie ideę nierelacyjnych baz danych oraz zna zasady organizacji specyficznych zbiorów danych i umie wskazać obszary zastosowań poznanych rozwiązań bazodanowych.	ET, PS	WM, L	K2A_W07, K2A_W08, K2A_W09
w2	Student zna różnice między kolumnowym i wierszowym sposobem przechowywania danych i rozumie konsekwencje wynikające z tych różnic.	ET, PS	WM, L	K2A_W08, K2A_W09, K2A_U11
w3	Student ma wiedzę na temat konieczności rezygnacji w bazach NoSQL z niektórymi własnościami ACID na rzecz kompromisów CAP.	ET, PS	WM, L	K2A_W05, K2A_W06, K2A_W07
U1	Student potrafi tworzyć bazy różnych typów, wprowadzać do nich dane oraz je modyfikować, a także rozwiązywać zadania wyszukiwania danych w różnych typach baz danych.	ET, PS	L	K2A_U01, K2A_U11, K2A_U13
U2	Student potrafi posługiwać się językiem XML.	ET, PS	WM, L	K2A_U01, K2A_U10, K2A_U13
U3	Student ma świadomość wpływu właściwego doboru rozwiązań bazodanowych na efektywność przetwarzania.	ET, PS	WM, L	K2A_U11, K2A_U12, K2A_U13, K2A_K02
<p>Znaczenie skrótów stosowanych w punkcie 17:</p> <ul style="list-style-type: none"> <li>• ET – egzamin testowy</li> <li>• PS – przygotowanie sprawozdania</li> <li>• WM – wykład multimedialny</li> <li>• L – laboratorium</li> </ul>				
<p><b>18. Formy zajęć dydaktycznych i ich wymiar (liczba godzin)</b></p> <p><b>30 W. 0 Ćw. 45 L. 0 P. 0 Sem.</b></p>				

## 19. Treści kształcenia:

### Wykład:

*Big Data*. Określenie „duże dane” odnosi się do dużych, zmiennych i różnorodnych zbiorów danych, których przetwarzanie jest trudne i znacznie wykracza poza możliwości tradycyjnych systemów zarządzania bazami danych. Tematyka poświęconego problemom *Big Data* wykładu (a także ćwiczenia laboratoryjnego) obejmuje model 3/5V, platformy przetwarzania dużych danych, sposoby składowania danych oraz model programowania i przetwarzania na wybranej platformie.

Bazy *NoSQL* (wykład podwójny). W ramach pierwszego z dwóch wykładów przedstawiony zostanie stosunkowo młody i obecnie modny nurt znany pod nazwą *NoSQL* (*Not only SQL*) – stojący w pewnej opozycji do modelu relacyjnego. Główną przesłanką powstania *NoSQL* są problemy skalowalności relacyjnych baz danych, które wynikają z faktu, że nie były one projektowane do wydajnego przetwarzania ogromnych ilości danych w czasie rzeczywistym. Na zajęciach omówiona zostanie idea baz typu „klucz-wartość” (na przykładzie rozwiązania *Amazon Dynamo*), baz implementujących „rodzinę kolumn” (rozwiązania *Google BigTable* i *Cassandra*) oraz baz zorientowanych dokumentowo (*MongoDB*, *CouchDB*). Drugi wykład jest wprowadzeniem do tematyki grafowego modelu danych oraz zagadnień związanych z mechanizmami dostępu do danych grafowych. Omawiane zagadnienia będą prezentowane w praktyce na przykładzie grafowej bazy danych *Neo4j* i jej języka zapytań *Cypher* oraz narzędzi *Apache TinkerPop*.

Bazy danych rezydujące w pamięci operacyjnej (*In-memory databases*). Wykorzystanie pamięci operacyjnej do przyspieszenia zarządzania danymi na poziomie systemów zarządzania bazami danych. Wykorzystanie dysków RAM. Systemy zarządzające bazami danych rezydującymi w pamięci operacyjnej (*IBM solidDB*). Sugerowane konfiguracje oraz zastosowania.

Rozproszone bazy danych (wykład podwójny). Architektura systemu zarządzania rozproszoną bazą danych. Charakterystyka typów rozproszenia danych. Operacja połączenia. Realizacja transakcji rozproszonej. Protokół wypełnienia i protokół wspólnego dokończenia transakcji rozproszonej. Propagacja zmian między danymi powielonymi.

Optymalizacja zapytań w relacyjnych systemach zarządzania bazą danych (wykład podwójny). Optymalizacja logiczna i fizyczna. Reguły transformacji wyrażeń algebry relacji. Algorytmy realizacji złączenia. Mechanizmy użytkownika. Przykłady i wskazówki praktyczne.

Rozmyte bazy danych. Operatory rozmyte. Funkcja przynależności. Formułowanie zadań wyszukiwania danych.

Wyszukiwanie pełnotekstowe (*Full-Text Search*). Indeksowanie danych binarnych. Tryby wyszukiwania danych. Predykaty i operatory stosowane podczas wyszukiwania. FTS w wybranych systemach baz danych.

Język XML. XML w bazach danych – podział dokumentów XML ze względu na ich strukturę. Przechowywanie danych z dokumentów XML a przechowywanie dokumentów XML. Przejście ze struktury dokumentu XML na strukturę relacyjnej bazy danych. Bazy danych wspierające XML. Bazy danych dedykowane dla XML.

Języki zapytań dla XML – *XPath*, *XQuery*.

Architektura, zarządzanie, bezpieczeństwo systemu zarządzania bazą danych na przykładzie SZBD *Oracle*.

Mechanizm *Flashback Queries* (*Flashback Database*, *Flashback Table*, *Flashback Drop*) w systemie *Oracle*. Konstruowanie zapytań w celu przeszukiwania danych archiwalnych (zapytania historyczne).

### Zajęcia laboratoryjne:

Zestaw ćwiczeń obejmuje 12 tematów realizowanych w wymiarze 3 godzin laboratoryjnych tygodniowo:

Model, składowanie, programowanie i przetwarzanie *Big Data*.

Dokumentowe, grafowe i kolumnowe bazy *NoSQL* (3 ćwiczenia).

Pamięciowa baza danych *solidDB*.

Rozmyte bazy danych.

Rozproszona baza danych i replikacje (2 ćwiczenia).

Architektura, zarządzanie i bezpieczeństwo w systemie *Oracle*.

Zapytania historyczne w systemie *Oracle*.

Język XML – mapowanie relacyjnych baz danych do XML, XML w systemach baz danych (2 ćwiczenia).

Podczas laboratorium wykorzystywane są systemy zarządzania bazami danych *Oracle*, *solidDB*, *Cassandra*, *MongoDB*, *CouchDB*, *Neo4j*, *MS SQL Server*, *MySQL*, *PostgreSQL*.

## 20. Egzamin: tak

**21. Literatura podstawowa:**

1. Elmasri R., Navathe S.B.: Wprowadzenie do systemów baz danych. Helion, 2005.
2. Garcia-Molina H., Ullman J.D., Widom J.: Implementacja systemów baz danych. Wydawnictwa Naukowo-Techniczne, 2003.
3. Graves M.: Projektowanie baz danych XML. Vademecum profesjonalisty. Helion, 2002.
4. Majors Ch., Campbell L.: Inżynieria niezawodnych baz danych. Helion, 2018 (seria O'Reilly)
5. Mayer-Schönberger V., Cukier K.: Big Data – efektywna analiza danych. MT Biznes, 2015. (audiobook)
6. Sadalage P. D., Fowler M.: NoSQL. Kompendium wiedzy. Helion, 2014.

**22. Literatura uzupełniająca:**

1. Carpenter J., Hewitt E.: Cassandra: The Definitive Guide. Distributed Data at Web Scale. Helion, 2016. (ebook)
2. Kazienko P., Gwiazda K.: XML na poważnie. Helion, 2002.
3. Lee H., Sohn I.: Big Data w przemyśle. Jak wykorzystać analizę danych do optymalizacji kosztów procesów? Wydawnictwo Naukowe PWN, 2016.
4. McLaughlin M.: Oracle Database 12c. Programowanie w języku PL/SQL. Helion, 2015.
5. Sullivan D.: NoSQL. Przyjazny przewodnik, Helion, 2015.
6. Szeliga M.: Data Science i uczenie maszynowe. Wydawnictwo Naukowe PWN, 2018.
7. Tow D.: SQL. Optymalizacja. Helion, 2004.
8. Wrembel R., Bębel B.: Projektowanie rozproszonych baz danych. Helion, 2003.

**23. Nakład pracy studenta potrzebny do osiągnięcia efektów kształcenia**

Lp.	Forma zajęć	Liczba godzin kontaktowych / pracy studenta
1	Wykład	30/30
2	Ćwiczenia	0/0
3	Laboratorium	45/45
4	Projekt	0/0
5	Seminarium	0/0
6	Inne	15/0
	Suma godzin	<b>90/75</b>

**24. Suma wszystkich godzin: 165****25. Liczba punktów ECTS: 3****26. Liczba punktów ECTS uzyskanych na zajęciach z bezpośrednim udziałem nauczyciela akademickiego: 3****27. Liczba punktów ECTS uzyskanych na zajęciach o charakterze praktycznym (laboratoria, projekty): 2****26. Uwagi:**

Zatwierdzono:

.....  
(data i podpis prowadzącego).....  
(data i podpis dyrektora instytutu/kierownika katedry/  
Dyrektora Kolegium Języków Obcych/kierownika lub  
dyrektora jednostki międzywydziałowej)