

Autor/Zespół Autorski: *dr inż. Marcin Kowalczyk*

BAZY DANYCH I BIG DATA

Database and Big Data

Poziom kształcenia: *I stopień*

Forma i tryb prowadzenia przedmiotu: *stacjonarna*

Kierunek studiów: Cyberbezpieczeństwo

Specjalność:

Klasy programowe:

Poziom przedmiotu: *podstawowy*

Status przedmiotu: *obowiązkowy*

Język przedmiotu: *polski*

Semestr nominalny (tylko dla przedmiotów obowiązkowych): *3*

Minimalny numer semestru: *dowolny*

Wymagania wstępne, zalecane przedmioty poprzedzające: *Algorytmy i programowanie 1 oraz Algorytmy i programowanie 2*

Limit liczby studentów: *60*

Powód zgłoszenia przedmiotu: program studiów na nowym kierunku Cyberbezpieczeństwo

Cel przedmiotu:

Celem przedmiotu jest zapoznanie studentów z technologią baz danych, stanowiącą podstawę współczesnej inżynierii oprogramowania, ze szczególnym zwróceniem uwagi na aspekt ich bezpieczeństwa. Omawiane są zagadnienia zarówno z obszaru baz relacyjno-obiektowych, jak i nierelacyjnych NoSQL, które stanowią podstawę ekosystemu technologii określanych mianem Big Data.

Treść kształcenia:

W ramach zajęć studenci zostaną zapoznani z rozwiązaniami baz danych, które są wykorzystywane na rynku, w szczególności z metodyką ich projektowania, implementacji oraz zabezpieczania, zarówno w odniesieniu do rozwiązań o ugruntowanej pozycji, takich jak bazy relacyjno-obiektowe, jak też zyskujących coraz bardziej na znaczeniu tzw. baz NoSQL, stanowiących m.in. podstawę dla bazodanowych rozwiązań Big Data. Integralnym elementem przedmiotu jest projekt. Umożliwi on ugruntowanie nabytej w ramach wykładu wiedzy w drodze praktycznej realizacji w pełni funkcjonalnego rozwiązania w oparciu o bazodanowe repozytorium danych, wraz ze współpracującą z nim aplikacją kliencką mającą formę graficznego interfejsu, spełniających odpowiednie wymagania w zakresie bezpieczeństwa i ochrony danych.

Wykład:

- **Wprowadzenie w dziedzinę baz danych – wymagania bezpieczeństwa (2h)**

Inżynieria baz danych – przypadki użycia, znaczenie. Architektura informacyjno-informatyczna przedsiębiorstwa. Baza danych jako kolekcja danych, a system zarządzania bazą danych SZBD. Pojęcie danych oraz metadanych. Rodzaje baz danych i ich charakterystyka. Kierunki ewolucji i specjalizacji baz danych. Bazy OLTP i OLAP –

charakterystyka. Wymogi bezpieczeństwa stawiane tego typu systemom. Uwarunkowania prawne dotyczące gromadzenia danych i ich ochrony – dane wrażliwe. Źródła zagrożeń.

- **Bazy relacyjne i relacyjno-obiektowe (14h)**

Architektura, sposób działania i obsługa relacyjnych i relacyjno-obiektowych silników baz danych na przykładzie wybranych rozwiązań. Architektura ANSI/SPARC. Pojęcie transakcji i ich właściwości ACID. Obsługa współbieżności. Poziomy dostępów. Integralność bazy danych i jej trwałość. Autentykacja oraz autoryzacja, nadawanie uprawnień. Metody wykrywania nieuprawnionego dostępu – audyt aktywności, weryfikacja logów, etc. Kopia zapasowa bazy danych i jej bezpieczeństwo. (4h)

Modelowanie i implementacja baz relacyjnych i relacyjno-obiektowych. Cykl życia bazy danych jako produktu. Opis i modelowanie rzeczywistości – encje, atrybuty, dziedziny, związki. Pułapki wachlarzowe i szczelinowe. Diagram związków ER – notacje UML i Chena. Fazy projektowania: model conceptualny, logiczny oraz fizyczny połączony z jego implementacją. Niekompatybilności modelu conceptualnego i logicznego – eliminacja związków wielu do wielu. Normalizacja oraz denormalizacja – skutki projektowe. Podejścia projektowe: wstępujące i zstępujące. Podeschematy oraz role odbiorców systemów – model bezpieczeństwa. (6h)

Język SQL – polecenia DDL, DML, DQL oraz DCL. Tworzenie i modyfikacja obiektów bazodanowych, wstawianie i modyfikacja danych, czytanie danych oraz zarządzanie uprawnieniami do obiektów i danych na poziomie relacyjno-obiektowego silnika bazy danych, tworzenie ról dostępów. (3h)

Struktury pomocnicze - wyszukiwanie i indeksowanie danych w pamięci zewnętrznej, wykorzystywane algorytmy. Optymalizacja zapytań. (1h)

- **Aplikacja kliencka (8h)**

Architektura klient-serwer, architektura wielowarstwowa. Mapowanie model relacyjny / obiektowy, model MVVM. Interfejsy komunikacji z bazą danych: JDBC, ODBC, inne. Tworzenie graficznej aplikacji wykorzystującej jako podstawę w swoim działaniu bazę danych. Wymogi i zabezpieczanie aplikacji klienckiej przed nieautoryzowanym dostępem. Szyfrowanie oraz anonimizacja danych. Ochrona przed atakami typu wstrzykiwanie kodu SQL (SQL Injection). (6h)

Rozszerzenia języka SQL – programowanie bazy danych, procedury składowane, wyzwalacze, pakiety. (2h)

- **Big Data (6h)**

Architektura i terminologia rozwiązań Big Data. Znaczenie systemów Big Data we współczesnym świecie – przykłady rozwiązań. Dane nieustrukturyzowane i ich przetwarzanie. Bazy NoSQL i ich klasyfikacja oraz charakterystyka. Właściwości BASE. Architektura ekosystemu Apache Hadoop. Podstawy realizacji i obsługi systemów Big Data. (6h)

Projekt:

- **część 1** – projekt i implementacja relacyjno-obiektowej bazy danych z naciskiem na aspekt bezpieczeństwa danych.

Zadaniem studenta (zespołu) jest zaprojektowanie i implementacja w pełni funkcjonalnej bazy danych (około 20 tabel) w nawiązaniu do zadanego tematu projektu. Studenci kolejno

wykonują projekt założeń funkcjonalnych, który obejmuje m.in. definicje zakresu obsługiwanych przez bazę operacji, oraz ról użytkowników z niej korzystających z uwzględnieniem wymagań bezpieczeństwa. Na bazie tych wymagań realizowany jest w dalszym etapie projekt conceptualny bazy danych, czego efektem jest model conceptualny z właściwym mu diagramem związków ER, obrazującym zamodelowane zbiory encji oraz ich wzajemne powiązania w postaci związków. Stanowi on po analizie zagadnienia pułapek wachlarzowych oraz szczelinowych punkt wyjścia do zbudowania modelu logicznego oraz modelu fizycznego, uwzględniającego charakterystykę i możliwości docelowego silnika bazy danych, w tym również jego możliwości w zakresie bezpieczeństwa. Na tym etapie prowadzone są m.in. działania mające na celu usunięcie możliwych niekompatybilności modelu conceptualnego oraz relacyjnego. Przeprowadzana jest normalizacja w celu osiągnięcia właściwej postaci schematów relacji w celu uniknięcia problemów anomalii z tytułu redundancji danych. Podejmowane są też decyzje odnośnie potrzeby ewentualnego obniżenia postaci normalnej na drodze analizy możliwych zysków i strat z tego tytułu (faza denormalizacji). Efektem tych działań jest końcowa postać projektowanej bazy danych, która po dodatkowych zabiegach jej optymalizacji na drodze m.in. wprowadzenia dodatkowych struktur pomocniczych jest implementowana. Dla zaimplementowanych obiektów przypisywane są odpowiednie uprawnienia, tworzone perspektywy dostępu w postaci bazodanowych ról oraz widoków (perspektyw). W celu ochrony danych szczególnie cennych oraz sensytywnych implementowane są dodatkowe mechanizmy ochrony i dostępu, takie jak np. VPD (ang. *Virtual Private Database*), TDE (ang. *Transparent Data Encryption*), czy też składowane procedury wyzwalane, pozwalająca m.in. na ich odpowiednią anonimizację oraz szyfrowanie. Uzupełnieniem tej części projektu jest jego dokumentacja.

- **część 2** – projekt i realizacja „bezpiecznej” aplikacji klienckiej, współpracującej z zaprojektowaną bazą danych w części 1, w formie graficznego interfejsu użytkownika.

Dla bazy danych, stanowiącej wynik pierwszej części projektu, tworzona jest aplikacja kliencka w architekturze wielowarstwowej z nią współpracująca, w formie graficznej. Aplikacja

ta oprócz spełnienia założeń funkcjonalnych, które dotyczą obsługi wymaganych operacji (transakcji) na bazie danych, takich jak: wstawianie, usuwanie, modyfikacja oraz pobieranie danych, musi spełniać odpowiednie wymogi bezpieczeństwa w zakresie m.in. ochrony danych

i niekontrolowanego do nich dostępu. Spełnienie tych ostatnich możliwe jest np. poprzez odpowiednie ograniczenie zakresu zestawu możliwych do realizacji zadań dla różnych ról użytkowników, które obsługuje. Aplikacja powinna być też bezpieczna pod kątem np. możliwości dokonania z jej udziałem nieautoryzowanych, bądź niechcianych operacji na bazie danych leżącej u jej podstawy. Te mogą polegać np. na możliwości niekontrolowanego wstrzykiwania kodu SQL, czy też możliwości osadzenia w niej i uruchamiania nieuprawnionych skryptów (ang. *cross-site scripting*). Zrealizowana aplikacja poddawana jest ocenie jej podatności na zagrożenia, oraz proponowane są rozwiązania pozwalające na jej ewentualne obniżenie.

Tematyka prowadzonych projektów jest bezpośrednio powiązana z treścią prowadzonych wykładów. Zajęcia te mają charakter uzupełniających zajęć praktycznych.

Uwagi realizacyjne:

Wykład

Zajęcia wykładowe mają formę interaktywnych warsztatów, będących połączeniem tradycyjnych form wykładowych z dużą ilością krótkich pokazów na żywo, obrazujących omawiane treści wykładowe. Zakłada się stosowanie podczas wykładów podejścia typowego dla metodologii *Design Thinking*, której celem jest maksymalne aktywizowanie studentów oraz wzmacnianie ich potencjału twórczego. Jest to szczególnie ważne w kontekście przyszłych problemów bezpieczeństwa danych.

Projekt

W zakresie projektu zakłada się połączenie dwóch form pracy, tj.: pracy indywidualnej studentów w małych zespołach nad zadanym tematem projektowym przy odpowiednim nadzorze opiekuna projektu, oraz pracy grupowej w ramach wspólnych spotkań z udziałem dużych grup. Te ostatnie, mające formę pokazów, będą poświęcone prezentacji rozwiązań dla szczególnie trudnych kwestii projektowych, w szczególności aspektu bezpieczeństwa.

Formy weryfikacji efektów kształcenia:

a) treści wykładowe

1. *Kolokwium 1*, po pierwszej części semestru. Sprawdza wiedzę wyniesioną z wykładów oraz poziom realizacji założonych efektów kształcenia w odniesieniu do takich zagadnień jak: rozumienie pojęć: baza danych, silnik bazy danych i jego charakterystyka, metodologia projektowania relacyjnych i relacyjno-obiektowych baz danych, pojęcie transakcji i ich właściwości ACID, bezpieczeństwo danych, źródła zagrożeń.
2. *Kolokwium 2*, po drugiej części semestru. Sprawdza wiedzę wyniesioną z wykładów oraz poziom realizacji założonych efektów kształcenia w odniesieniu do takich zagadnień jak: rozszerzone funkcjonalności baz danych, struktury pomocnicze – indeksy, zagadnienie tworzenie bezpiecznej aplikacji klienckiej, bazy NoSQL, charakterystyka koncepcji Big Data, dane nieustrukturyzowane, sposoby czerpania wiedzy i pozyskiwania informacji z wolumenów danych o dużej różnorodności.

b) projekt

Projekt ma na celu zweryfikowanie umiejętności wykorzystania w praktyce zdobytej na wykładach wiedzy w odniesieniu do technologii baz danych na drodze ich projektowania oraz implementacji z zachowaniem szerokokorozumianych wymogów w zakresie bezpieczeństwa danych, jak też podstaw tworzenia bezpiecznych, graficznych aplikacji klienckich, współpracujących z rozwiązaniami tego typu. Projekty będą realizowane w dwu lub trzy osobowych zespołach.

Treść kształcenia - streszczenie w jęz. angielskim:

The main aim of the course is to present and teach broadly understand database technologies, that form the fundament of modern software engineering, with particular attention to the aspect of their security. Issues in the area of the relational and object-relational databases as well as NoSQL databases, which form the base for technologies ecosystem referred as Big

Wydziałowa Komisja Akredytacji Przedmiotów (WKAP)

Data, are discussed. During the course, students will be familiarized among others with methodology their design, implementation, and programming, as well as securing them appropriately.

An integral part of the course is project, which will strengthen the acquired skills by students on a way of practical implementation of fully functional solutions based on an appropriate data repository as well as cooperating with it a client application in a form of graphical user interface, which meet the assumed security requirements.

Egzamin: *nie*

Literatura i oprogramowanie:

1. Slajdy do wykładu, materiały uzupełniające w postaci zadań interaktywnych oraz filmów obrazujących omawiane zagadnienia.
2. Książki:
 - Garcia-Molina Hector, Ullman Jeffrey D., Widom Jennifer, Systemy baz danych. Kompletny podręcznik, Wydanie II, Helion 2011 (lub nowsze)
 - Connolly Thomas, Begg Carolyn, Database Systems: A Practical Approach to Design, Implementation, and Management, 6th Edition, Addison-Wesley Longman, 2015 (lub nowsze)
 - Marz Nathan, Warren James, Big Data. Najlepsze praktyki budowy skalowalnych systemów obsługi danych w czasie rzeczywistym, Helion, 2016 (lub nowsze),
 - Mayer-Schonberger Viktor, Cukier Kenneth, Big Data: A Revolution That Will Transform How We Live, Work and Think, John Muray Press, 2013 (lub nowsze)
 - White Tom, Hadoop Komplety przewodnik, Analiza i przechowywanie danych, Helion, 2016 (lub nowsze)
 - Knox David C., Maroulis William, Gaetjen Scott, Oracle Database 12c Security, McGraw-Hill Professional, 2015 (lub nowsze)
 - Thuraisingham Bhavani, Database and Applications Security: Integrating Information Security and Data Management, Auerbach Publications, 2005 (lub nowsze)
 - Gertz Michael (Editor), Jajodia Sushil, Handbook of Database Security: Applications and Trends, Springer, 2010
 - Larose Daniel T., Metody i modele eksploracji danych, PWN, 2008 (lub nowsze),
 - Inne.
3. Internet
4. Baza danych Oracle oraz Apache Hadoop plus oprogramowanie współpracujące, narzędzia do modelowania baz danych, narzędzia do komunikacji zespołowej, narzędzie do testów.

Wymiar godzinowy zajęć:

W	C	L	P
30	-	-	30

W ramach projektu co najmniej połowa godzin w ramach zajęć projektowych będzie miała formę zajęć grupowych. W ich ramach będą prezentowane możliwe rozwiązania dla kluczowych elementów realizowanych projektów w formie warsztatowych pokazów z wykorzystaniem realnych narzędzi i rozwiązań, w tym m.in. ze szczególnym zwróceniem uwagi na aspekt bezpieczeństwa.

Wymiar w jednostkach ECTS: 5

Liczba godzin pracy studenta związanych z osiągnięciem efektów kształcenia (opis):

1. liczba godzin kontaktowych – **64** godz., w tym:
 - obecność na wykładach: **30** godz.
 - obecność na zajęciach projektowych: **20** godz.
 - konsultacje projektowe: **10** godz.
 - obecność na kolokwium: **4** godz.
2. praca własna studenta – **66** godz., w tym:
 - analiza literatury, materiałów i przykładów z wykładów: **10** godz.
 - realizacja dwóch projektów oraz przygotowanie ich dokumentacji: **40** godz.
 - przygotowanie do kolokwium: **16** godz.

Łączny nakład pracy studenta wynosi 130 godz., co odpowiada 5 pkt. ECTS.

Liczba punktów ECTS na zajęciach wymagających bezpośredniego udziału nauczycieli akademickich: 2.5 pkt. ECTS, co odpowiada 64 godz. kontaktowym.

Liczba punktów ECTS, którą student uzyskuje w ramach zajęć o charakterze praktycznym: 2.5 pkt. ECTS, co odpowiada 66 godz. pracy własnej przy rozwiązywaniu zagadnienia projektowego

Efekty kształcenia / uczenia się i formy ich weryfikacji:

efekty kształcenia/uczenia się	forma zajęć/ technika kształcenia	sposób weryfikacji (oceny)*	odniesienie do efektów uczenia się dla programu
Wiedza			
w01: ma wiedzę w zakresie potencjalnych możliwości, ograniczeń i zagrożeń wiążących się z wykorzystaniem szerokorozumianych technologii baz danych, w tym rozwiązań klasy Big Data.	wykład + projekt	projekt, kolokwium	W05, W08,W12
w02: ma wiedzę z zakresu projektowania i implementacji bezpiecznych systemów gromadzenia danych różnej postaci, w tym baz relacyjno-obiektowych i NoSQL.	wykład + projekt	projekt, kolokwium	W05,W07
w03: zna koncepcję, możliwości oraz ograniczenia współbieżnego dostępu do danych, zna podstawy efektywnego wyszukiwania danych.	wykład + projekt	projekt, kolokwium	W05,W07
w04: ma wiedzę z zakresu działania silników baz danych, sposobu ich ochrony i zabezpieczania oraz optymalizacji ich wydajności.	wykład + projekt	projekt, kolokwium	W05,W07,W08

Wydziałowa Komisja Akredytacji Przedmiotów (WKAP)

w05: ma fundamentalną wiedzę z zakresu tworzenia bezpiecznych aplikacji klienckich, współpracujących z różnego rodzaju bazami danych.	wykład + projekt	projekt, kolokwium	W05,W07,W08
w06: ma fundamentalną wiedzę w zakresie podstawowych norm i wymogów prawnych w zakresie gromadzenia, przetwarzania oraz udostępniania danych w postaci komputerowych baz danych.	wykład + projekt	projekt, kolokwium	W10,W12
Umiejętności			
u01: potrafi zaprojektować oraz zaimplementować bazę relacyjną z zachowaniem określonych norm bezpieczeństwa danych.	wykład + projekt	projekt	U05,U06,U08
u02: potrafi zaproponować rozwiązanie klasy Big Data w oparciu o stos technologiczny Apache Hadoop.	wykład + projekt	projekt, kolokwium	U04, U05,U09
u03: potrafi zaimplementować graficzną aplikację klienta, współpracującą z wybranym silnikiem bazy danych z zachowaniem określonych wymogów bezpieczeństwa.	wykład + projekt	projekt	U06,U08,U09
u04: potrafi zweryfikować spełnienie podstawowych poziomów bezpieczeństwa w zakresie ochrony zbiorów danych oraz narzędzi wykorzystywanych do tego celu.	wykład + projekt	projekt	U02
u05: potrafi wykorzystać narzędzia do modelowania baz danych oraz implementacji graficznych aplikacji klienckich.	wykład + projekt	projekt	U09
u06: potrafi pozyskiwać informacje z literatury dotyczące wybranych zagadnień z obszaru baz danych i Big Data.	wykład + projekt	projekt, kolokwium	U06,U13
u07: potrafi przygotować i przedstawić prezentację oraz dokumentację dotyczącą zrealizowanego projektu.	projekt	projekt	U10,U12
u08: potrafi pracować w zespole.	projekt	projekt	U09
Kompetencje społeczne			
ks01: ma świadomość konieczności komunikowania się z otoczeniem, także pozazawodowym, w sposób zrozumiały dla odbiorcy.	wykład + projekt	projekt, kolokwium	KS05
ks02: ma świadomość znaczenia i wartości wspólnej pracy w zespole z zachowaniem norm etycznych.	projekt	projekt	KS03

Wydziałowa Komisja Akredytacji Przedmiotów (WKAP)

ks03 ma orientację zawodową w obszarze inżynierii baz danych i ich bezpieczeństwa oraz jest świadomy procesu ciągłego uczenia się w kierunku zwiększania kompetencji w tym obszarze.	wykład + projekt	projekt	KS01
---	------------------	---------	------

Uwagi:

Data i podpis autora (kierownika zespołu autorskiego):