

<b>KARTA OPISU MODUŁU KSZTAŁCENIA</b>		
Nazwa modułu/przedmiotu <b>Systemy rozproszone dużej skali</b>		Kod <b>1010512321010519247</b>
Kierunek studiów <b>Informatyka</b>	Profil kształcenia (ogólnoakademicki, praktyczny) <b>ogólnoakademicki</b>	Rok / Semestr <b>1 / 2</b>
Ścieżka obieralności/specjalność <b>Systemy rozproszone</b>	Przedmiot oferowany w języku: <b>polski</b>	Kurs (obligatoryjny/obieralny) <b>obligatoryjny</b>
Stopień studiów: <b>II stopień</b>	Forma studiów (stacjonarna/niestacjonarna) <b>stacjonarna</b>	
Godziny Wykłady: <b>30</b> Ćwiczenia: - Laboratoria: <b>30</b> Projekty/seminaria: -		Liczba punktów <b>5</b>
Status przedmiotu w programie studiów (podstawowy, kierunkowy, inny) <b>kierunkowy</b>		(ogólnouczelniany, z innego kierunku) <b>z danego kierunku</b>
Obszar(y) kształcenia i dziedzina(y) nauki i sztuki <b>nauki techniczne</b>		Podział ECTS (liczba i %) <b>5 100%</b>
<b>Odpowiedzialny za przedmiot / wykładowca:</b>		
dr inż. Anna Kobusińska email: Anna.Kobusinska@cs.put.poznan.pl tel. 61 6652964 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań		dr inż. Tadeusz Kobus email: Tadeusz.Kobus@cs.put.poznan.pl tel. 61 6652964 Instytut Informatyki ul. Piotrowo 2, 60-965 Poznań
<b>Wymagania wstępne w zakresie wiedzy, umiejętności, kompetencji społecznych:</b>		
1	<b>Wiedza:</b>	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP. W szczególności student rozpoczynający ten przedmiot powinien posiadać podstawową wiedzę z zakresu systemów operacyjnych, rozproszonych systemów operacyjnych oraz sieci komputerowych.
2	<b>Umiejętności:</b>	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP.
3	<b>Kompetencje społeczne</b>	Efekty kształcenia ze studiów I stopnia zdefiniowane w Uchwale Senatu PP. Ponadto w zakresie kompetencji społecznych student musi prezentować takie postawy jak uczciwość, odpowiedzialność, wytrwałość, ciekawość poznawcza, kreatywność, kultura osobista, szacunek dla innych ludzi.
<b>Cel przedmiotu:</b>		
1. Przekazanie studentom podstawowej wiedzy związanej z wyzwaniami konstrukcji systemów rozproszonych dużej skali, w zakresie prezentacji teoretycznych i praktycznych aspektów konstrukcji rozproszonych systemów dużej skali. 2. Prezentacja problematyki przetwarzania dużych ilości danych (big data) w systemach rozproszonych dużej skali. 3. Rozwijanie u studentów umiejętności rozwiązywania problemów związanych z przetwarzaniem w systemach rozproszonych dużej skali.		
<b>Efekty kształcenia i odniesienie do kierunkowych efektów kształcenia</b>		
<b>Wiedza:</b>		
1. ma zaawansowaną i pogłębioną wiedzę z zakresu szeroko rozumianych systemów informatycznych, podstaw teoretycznych ich budowania oraz metod, narzędzi i środowisk programistycznych w zakresie algorytmów i złożoności, architektury systemów komputerowych, systemów operacyjnych, technologii sieciowych, języków i paradygmatów programowania - [K2st_W1] 2. ma uporządkowaną i podbudowaną teoretycznie wiedzę ogólną związaną z kluczowymi zagadnieniami z zakresu informatyki takimi jak: architektura i klasyfikacja systemów rozproszonych dużej skali i ich środowiska komunikacyjnego, protokołów epidemicznych i plotkujących, systemów przetwarzania dużej ilości danych - [K2st_W2] 3. ma zaawansowaną wiedzę szczegółową dotyczącą wybranych zagadnień z zakresu informatyki - [K2st_W3] 4. ma zaawansowaną i szczegółową wiedzę o procesach zachodzących w cyklu życia systemów informatycznych sprzętowych lub programowych - [K2st_W5]		
<b>Umiejętności:</b>		

<p>1. potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich i prostych problemów badawczych metody analityczne, symulacyjne oraz eksperymentalne - [K2st_U4]</p> <p>2. 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 - [K2st_U5]</p> <p>3. potrafi ocenić przydatność i możliwość wykorzystania nowych osiągnięć (metod i narzędzi) oraz nowych produktów informatycznych - [K2st_U6]</p> <p>4. potrafi dokonać krytycznej analizy istniejących rozwiązań technicznych oraz zaproponować ich ulepszenia (usprawnienia) - [K2st_U8]</p> <p>5. potrafi ocenić przydatność metod i narzędzi służących do rozwiązania zadania inżynierskiego, polegającego na budowie lub ocenie systemu informatycznego lub jego składowych, w tym dostrzec ograniczenia tych metod i narzędzi; - [K2st_U9]</p> <p>6. potrafi - stosując m.in. koncepcyjnie nowe metody - rozwiązywać złożone zadania informatyczne, w tym zadania nietypowe oraz zadania zawierające komponent badawczy - [K2st_U10]</p> <p>7. potrafi ? zgodnie z zadaną specyfikacją, uwzględniającą aspekty pozatechniczne ? zaprojektować złożone urządzenie, system informatyczny lub proces oraz zrealizować ten projekt ? co najmniej w części ? używając właściwych metod, technik i narzędzi, w tym przystosowując do tego celu istniejące lub opracowując nowe narzędzia - [K2st_U11]</p>
<b>Kompetencje społeczne:</b>
<p>1. rozumie, że w informatyce wiedza i umiejętności bardzo szybko stają się przestarzałe - [K2st_K1]</p> <p>2. rozumie znaczenie wykorzystywania najnowszej wiedzy z zakresu informatyki w rozwiązywaniu problemów badawczych i projektowych - [K2st_K2]</p>

<b>Sposoby sprawdzenia efektów kształcenia</b>
<p>Ocena formująca:</p> <p>a)w zakresie wykładów:</p> <ul style="list-style-type: none"><li>- na podstawie odpowiedzi na pytania dotyczące materiału omówionego na poprzednich wykładach,</li></ul> <p>b)w zakresie laboratoriów:</p> <ul style="list-style-type: none"><li>-na podstawie oceny bieżącego postępu realizacji zadań,</li></ul> <p>Ocena podsumowująca:</p> <p>a) w zakresie wykładów weryfikowanie założonych efektów kształcenia realizowane jest przez:</p> <ul style="list-style-type: none"><li>- ocenę wiedzy i umiejętności wykazanych na egzaminie ustnym składającym się z 3 pytań związanych z materiałem przedstawionym na wykładach</li></ul> <p>b) w zakresie laboratoriów weryfikowanie założonych efektów kształcenia realizowane jest przez:</p> <ul style="list-style-type: none"><li>- ocenę przygotowania studenta do poszczególnych sesji zajęć laboratoryjnych (sprawdzian wejściowy) oraz ocenę umiejętności związanych z realizacją ćwiczeń laboratoryjnych,</li><li>- ocenę sprawozdania przygotowywanego częściowo w trakcie zajęć, a częściowo po ich zakończeniu; ocena ta obejmuje także umiejętność pracy w zespole,</li><li>- ocenę wiedzy i umiejętności związanych z realizacją zadania projektowego</li></ul> <p>Uzyskiwanie punktów dodatkowych za aktywność podczas zajęć, a szczególnie za:</p> <ul style="list-style-type: none"><li>- omówienia dodatkowych aspektów zagadnienia,</li></ul>
<b>Treści programowe</b>
<p>Program wykładu obejmuje następujące zagadnienia:</p> <ol style="list-style-type: none"><li>1. Przedstawienie wyzwań związanych z budową i przetwarzaniem w systemach rozproszonych dużej skali: klasyfikacje systemów, idea działania, zastosowania, przykłady implementacji</li><li>2. Architektury systemów P2P (ustrukturalizowane, nieustrukturalizowane, hybrydowe); organizacja węzłów, topologie, skalowalność, równoważenie obciążeń; gossip-based small word networks, Kleinberg's peer sampling</li><li>3. Nieustrukturalizowane systemy P2P: dołączanie/odłączanie węzłów, peer-sampling, strategie wyszukiwania zasobów (flooding-based multicast, tree-based multicast, random walk, expanding ring, rendezvous point, bubblecast) na przykładzie systemów: FreeNet, FastTrack, eDonkey, Gnutella</li><li>4. Ustrukturalizowane systemy P2P: dołączanie/odłączanie węzłów; przestrzeń identyfikatorów, Distributed Hash Table (DHT), utrzymywanie informacji o stanie przetwarzania, redukcja opóźnień, odporność na awarie na przykładzie systemów Pastry, Chord, Tapestry, CAN (ang. Content Addressable Network), Kademlia, Kad.</li><li>5. Monitorowanie systemów P2P (aggregation, polling, slicing, renaming), anonimowość w systemach P2P: miary anonimowości, mechanizmy zapewnienia anonimowości</li><li>6. Protokoły Bittorrent i Blockchain - zasada działania, przeznaczenie</li><li>7. Wprowadzenie do tematyki dużych danych: definicje i charakterystyka ?Big data?, źródła dużych danych, aspekty przetwarzania dużych danych, wyzwania związane z ?Big data?</li><li>8. Architektura systemów Big Data: pozyskiwanie dużych danych, składowanie dużych danych, przetwarzanie</li></ol>

wsadowe/przetwarzanie strumieni danych, analiza dużych danych; algorytmy rozproszone w przetwarzaniu dużych danych; przykłady wykorzystania narzędzi dużych danych w istniejących systemach (m.in. na przykładzie Google, Facebook, Yahoo, LinkedIn, Cloudera, Microsoft)

9. Wprowadzenie do tematyki baz danych NoSQL: klasyfikacja ze względu na modele danych (key value, column-oriented, document-oriented, column-oriented, graph-oriented); budowa i działanie systemów NoSQL (data partitioning, load balancing, replication, data versioning, membership management, failure handling) na przykładzie Amazon Dynamo; Google BigTable, HBase, Cassandra.

10. Przetwarzanie dużych danych z wykorzystaniem platformy Hadoop (MapReduce);

11. Przetwarzanie dużych danych z wykorzystaniem Resilient Distributed Datasets (RDD) i platformy Apache Spark

12. Systemy interaktywnych zapytań w systemach dużych danych (systemy Hive i Shark)

13. Przetwarzanie danych reprezentujących grafy dużej skali ? system Pregel (Bulk Synchronous Parallel model, Vertex-centric, Superstep - sequence of iterations, Master-worker model)

14. Systemy zarządzania zasobami Mesos i YARN (architektura, algorytmy lokowania zasobów)

Program laboratorium obejmuje zagadnienia związane z problematyką przetwarzania dużych ilości danych w systemie Casandra:

1. Instalacja, konfiguracja, interfejs programistyczny, typy danych, podstawowe operacje dostępne w danym systemie.

2. Praktyczne wykorzystanie systemu: modelowanie danych; sortowanie, filtrowanie, opcje tabel, indeksy, kolekcje, liczniki, lightweight transactions i inne zaawansowane funkcje, stronicowanie wyników. Wykorzystanie znaczników czasowych. Poziomy spójności. Awarie węzłów, partycje sieciowe i rozsynchronizowanie zegarów. Konflikty znaczników czasowych.

Metody dydaktyczne:

1. wykład: prezentacja multimedialna, prezentacja ilustrowana przykładami podawanymi na tablicy, rozwiązywanie zadań

2. ćwiczenia laboratoryjne: demonstracja, dyskusja, warsztaty, ćwiczenia praktyczne, praca w zespole

**Literatura podstawowa:**

1. Peer-to-peer systems and applications, R. Steinmetz, K. Wehrle, Springer, 2005

2. NoSQL distilled, P. Sadalage, M. Flower, Addison-Wesley, 2013

**Literatura uzupełniająca:**

1. Spark in Action, Bonać M., Zečević P., Manning, 2015

2. . Large Scale Network-Centric Distributed Systems, H.Sarbazi-Azad, A.Y.Zomaya, Wiley-IEEE Computer Society Press, 2013

3. M. Jelasity, S. Voulgaris, R. Guerraoui, A.-M. Kermarrec, M. Van Steen: Gossip-based peer sampling. ACM Trans. Comput. Syst 25(3) 2007

4. Nitin Sawant, Himanshu Shah, Big data application architecture Q&A, Springer, 2013

5. J.Berman, Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information, Morgan-Kaufman, 2013

**Bilans nakładu pracy przeciętnego studenta**

Czynność	Czas (godz.)
1. udział w zajęciach laboratoryjnych	30
2. przygotowanie do ćwiczeń laboratoryjnych	10
3. dokończenie (w ramach pracy własnej) sprawozdań z ćwiczeń laboratoryjnych	12
4. napisanie programu, uruchomienie i weryfikacja (czas poza zajęciami laboratoryjnymi)	10
5. przygotowanie do sprawdzianów	10
6. udział w konsultacjach związanych z realizacją procesu kształcenia, w szczególności ćwiczeń laboratoryjnych / projektu (część konsultacji może być realizowana drogą elektroniczną)	2 30
7. udział w wykładach	10
8. przygotowanie do egzaminu i obecność na egzaminie: 8 godz. + 2 godz.	10
9. zapoznanie się ze wskazaną literaturą / materiałami dydaktycznymi (10 stron tekstu naukowego = 1 godz.), 100 stron	

**Obciążenie pracą studenta**

forma aktywności	godzin	ECTS
Łączny nakład pracy	124	5
Zajęcia wymagające bezpośredniego kontaktu z nauczycielem	64	3
Zajęcia o charakterze praktycznym	60	2