

.....

KARTA INFORMACYJNA PRZEDMIOTU

Nazwa przedmiotu	Rozproszone przetwarzanie danych		Distributed data processing				
Kod przedmiotu	WCYKBWSM_DDPBD						
Język wykładowy	polski						
Profil studiów	ogólnoakademicki						
Forma studiów	studia stacjonarne						
Poziom studiów	wojskowe studia jednolite magisterskie						
Rodzaj przedmiotu	obowiązkowy						
Obowiązuje od naboru	2021/2022						
Forma zajęć, liczba godzin/rygor, razem godz., pkt ECTS	semestr	(x egzamin, + zaliczenie, # projekt)					punkty ECTS
		razem	wykłady	ćwiczenia	laboratoria	projekt	seminarium
	IX	44+	14		30+		3.0
	razem		14		30		3.0
Przedmioty wprowadzające	● Brak przedmiotów kształcenia wprowadzających						
Semestr/kierunek studiów	semestr 9 / Kryptologia i cyberbezpieczeństwo / Bezpieczeństwo informacyjne						
Autor	dr inż. Jarosław Koszela						
Jednostka odpowiedzialna za przedmiot	WCY/ISI						
Skrócony opis przedmiotu	- Wykład - z wykorzystaniem klasycznych metod dydaktycznych (tablica, rzutnik, projektor) oraz praktycznym pokazem przykładowych rozwiązań i mechanizmów omawianych na wykładanych. - Ćwiczenia laboratoryjne - z wykorzystaniem środowiska i narzędzi do projektowania, implementacji i zarządzania komponentami do rozproszonego przetwarzania danych - Big Data (m.in. - bazy danych: SQL, NoSQL, NewSQL; ETL/ELT; ekosystem Apache Hadoop - przetwarzania wsadowego i Apache Spark - przetwarzania strumieniowego lub równoważne).						
Pełny opis przedmiotu (treści programowe)	lp.	Semestr IX temat/tematyka zajęć	liczba godzin				
			wkł.	ćw.	lab.	prj.	sem.
	1	Wprowadzenie do tematyki rozproszonego przetwarzania danych. Modele architektury Kappa i Lambda.	2				
	2	Pozyskiwanie i składowanie rozproszone wielkoskalowych zasobów danych - rozwiązania pozyskiwania danych i strumieni danych, buforowanie/kolejkowanie oraz rozproszonego składowania danych: Hadoop/HDFS, bazy danych - SQL, NoSQL, NewSQL. Procesy ETL/ELT.	2		6		
	3	Procesy przetwarzania danych w trybie wsadowym. Wzorzec przetwarzania Map Reduce. Środowiska, narzędzia i języki wykorzystywane do wielkoskalowego rozproszonego przetwarzania danych w trybie wsadowym oraz zarządzanie tym procesem przetwarzania.	4		6		
	4	Procesy przetwarzania danych w trybie strumieniowym. Środowiska, narzędzia i języki wykorzystywane do wielkoskalowego rozproszonego strumieniowego i mikrowśadowego przetwarzania danych oraz zarządzanie procesem przetwarzania.	4		8		
	5	Ekosystem komponentów środowiska Big Data. Stos technologiczny Apache Hadoop i Apache Spark - narzędzia open source dla Big Data. Wykorzystanie narzędzi rozproszonego przetwarzania danych w zadaniach eksploracji danych, w tym z wykorzystaniem modeli i metod uczenia maszynowego oraz przetwarzania języka naturalnego. Zarządzanie procesem wielkoskalowego rozproszonego przetwarzania i analizy danych.	2		10		
	Razem		14		30		
Literatura	podstawowa: - David Stephenson; Big data, nauka o danych i AI bez tajemnic; Helion; 2019 R. Bembienik, L. Skonieczny, G. Protaziuk, M. Kryszkiewicz, H. Rybinski; Intelligent Methods and Big Data in Industrial Applications; 2018; Springer International Publishing AG; 2018 - A Gorelik; Korporacyjne jezioro danych. Wykorzystaj potencjał big data w swojej organizacji; Helion; 2019 G. Harrison; NoSQL, NewSQL i BigData. Bazy danych następnej generacji; Helion; 2019 N. Marz, J. Warren; Big Data. Najlepsze praktyki budowy skalowalnych systemów obsługi danych w czasie						

	rzeczywistym; Helion; 2016 uzupełniająca: ● P. J. Deitel, H. Deitel; Python dla programistów. Big Data i AI. Studia przypadków; Helion; 2020 ● Strony ekosystemu Apache: www.apache.org		
Efekty uczenia się	Symbol	Efekty kształcenia	odniesienie do efektów kształcenia dla kierunku
	W1	Posiada znajomość architektury i środowisk przetwarzania danych masowych Big Data. Znajomość trendów i kierunków rozwoju systemów przetwarzania danych masowych.	K_W03, K_W04, K_W09, K_U03, K_U04
	U1	Posiada umiejętność implementacji procedur transformacji danych masowych na platformie Hadoop w paradygmacie Map Reduce. Umiejętność implementacji procedur transformacji danych wsadowych i strumieniowych.	K_W03, K_W04, K_W09, K_U03, K_U04
	U2	Posiada umiejętność wykorzystania narzędzi rozproszonego przetwarzania danych do rozwiązywania zadań eksploracji danych. Znajomość roli narzędzi w stosie technologicznym Hadoop.	K_W03, K_W04, K_W09, K_U03, K_U04
Metody i kryteria oceniania (sposób sprawdzania osiągnięcia przez studenta zakładanych efektów uczenia się)	● Przedmiot zaliczany jest na podstawie kolokwium w formie pisemnej oraz zaliczenia ćwiczeń laboratoryjnych. Warunkiem zaliczenia przedmiotu jest uzyskanie pozytywnej oceny z zaliczenia ćwiczeń laboratoryjnych oraz z kolokwium z części teoretycznej. Ocena końcowa jest wypadkową pozytywnych ocen z kolokwium i ćwiczeń laboratoryjnych. ● Zaliczenie laboratorium: ocena wystawiona na podstawie wykonanego projektu z wykorzystaniem wybranych narzędzi do projektowania i tworzenia Big Data oraz na podstawie sporządzonego sprawozdania laboratoryjnego i ocen cząstkowych uzyskanych w trakcie zajęć laboratoryjnych.		
Bilans ECTS (nakład pracy studenta)	SEMESTR 9		
	Aktywność	Obciążenie studenta	
		Liczba godzin	Liczba ECTS
	Udział w wykładach	14	0.5
	Udział w laboratoriach	30	1
	Udział w ćwiczeniach	0	0
	Udział w projektach	0	0
	Udział w seminariach	0	0
	Samodzielne studiowanie tematyki wykładów	15	0.5
	Samodzielne przygotowanie do laboratoriów	40	1
	Samodzielne przygotowanie do ćwiczeń		
	Samodzielna realizacja projektu		
	Samodzielne przygotowanie do seminariów		
	Udział w konsultacjach		
	Przygotowanie do egzaminu		
	Przygotowanie do zaliczenia		
	Udział w egzaminie / kolokwium	2	
	Sumaryczne obciążenie pracą studenta	101	3
	Zajęcia z udziałem nauczycieli	46	1.5
	Zajęcia powiązane z działalnością naukową	99	3
	Zajęcia o charakterze praktycznym	70	2

autor

**kierownik jednostki organizacyjnej
odpowiedzialnej za przedmiot**

dr inż. Jarosław Koszela

tytuł, stopień naukowy, imię, NAZWISKO, podpis

dr hab. inż. Nowicki Tadeusz

tytuł, stopień naukowy, imię, NAZWISKO, podpis