



Karta przedmiotu

Nazwa i kod przedmiotu	Big Data processing frameworks, PG_00045325						
Kierunek studiów	Inżynieria danych, Inżynieria danych						
Data rozpoczęcia studiów	październik 2026 r.	Rok akademicki realizacji przedmiotu			2028/2029		
Poziom kształcenia	I stopnia - inżynierskie	Grupa zajęć			Grupa zajęć fakultatywnych Grupa zajęć powiązanych z prowadzonymi badaniami naukowymi w dziedzinie nauki związanej z kierunkiem - profil ogólnoakademicki		
Forma studiów	stacjonarne	Sposób realizacji			na uczelni		
Rok studiów	3	Język wykładowy			angielski		
Semestr studiów	6	Liczba punktów ECTS			5.0		
Profil kształcenia	ogólnoakademicki	Forma zaliczenia			egzamin		
Jednostka prowadząca	Wydziały Politechniki Gdańskiej -> Wydział Elektroniki, Telekomunikacji i Informatyki -> Katedra Inżynierii Oprogramowania						
Imię i nazwisko wykładowcy (wykładowców)	Odpowiedzialny za przedmiot		dr Adam Przybyłek				
	Prowadzący zajęcia z przedmiotu		dr Adam Przybyłek				
Formy zajęć	Forma zajęć	Wykład	Ćwiczenia	Laboratorium	Projekt	Seminarium	RAZEM
	Liczba godzin zajęć	15.0	0.0	30.0	15.0	0.0	60
	W tym liczba godzin zajęć na odległość: 0.0						
Aktywność studenta i liczba godzin pracy	Aktywność studenta	Udział w zajęciach dydaktycznych, objętych planem studiów		Udział w konsultacjach		Praca własna studenta	RAZEM
	Liczba godzin pracy studenta	60		6.0		59.0	125
Cel przedmiotu	Celem przedmiotu jest wprowadzenie studentów w tematykę Big Data. W ramach przedmiotu omawiane są 3 nowoczesne frameworki umożliwiające łatwe pisanie aplikacji przetwarzających równoległe ogromne ilości danych na łatwo skalowalnych klastrach						
Efekty uczenia się przedmiotu	Efekt kierunkowy		Efekt z przedmiotu		Sposób weryfikacji i oceny efektu		
	[K6_U05] projektuje innowacyjne rozwiązania analizy i przetwarzania danych, wykorzystując odpowiednie metody i narzędzia		Student potrafi wykorzystać framework Apache Spark oraz Hadoop do przetwarzania rozproszonego masowych danych oraz system HDFS do magazynowania masowych danych.		[SU1] Ocena realizacji zadania [SU4] Ocena umiejętności korzystania z metod i narzędzi		
	[K6_U04] formułuje logiczne rozwiązania złożonych lub nieustrukturyzowanych problemów		Student zna wzorce przetwarzania rozproszonego Map-Reduce.		[SU3] Ocena umiejętności wykorzystania wiedzy uzyskanej w ramach przedmiotu		
	[K6_W02] demonstruje zaawansowane przygotowanie w zakresie metod oraz technik formułowania i rozwiązywania problemów		Student potrafi sformułować problem badawczy, zastosować odpowiednie metody, rozwiązać problem oraz poprawnie zinterpretować wyniki. Ponadto, student potrafi także krytycznie ocenić uzyskane wyniki.		[SW3] Ocena wiedzy zawartej w opracowaniu tekstowym i projektowym		

Treści przedmiotu	Treści przedmiotu - wykład		
	<ol style="list-style-type: none"> 1. Ekosystem Apache Hadoop architektura i główne komponenty 2. HDFS jako rozproszony system składowania danych: zasady działania, replikacja, niezawodność 3. Model obliczeń MapReduce: projektowanie zadań, przepływ danych, typowe wzorce przetwarzania 4. Wprowadzenie do zagadnień Big Data oraz przetwarzania w chmurze 5. Apache Spark przetwarzanie rozproszone in-memory, podstawowe moduły i zastosowania 6. Apache Storm przetwarzanie strumieniowe w czasie rzeczywistym, topologie i scenariusze użycia 7. Process Mining analiza procesów na podstawie dzienników zdarzeń 		
	Treści przedmiotu - laboratoria		
	<p>Laboratoria Hadoop:</p> <ul style="list-style-type: none"> • Lab 1,2: podstawy MapReduce (uruchamianie zadań, praca z danymi, proste analizy) • Lab 3-5: bardziej zaawansowane analizy w MapReduce (łączenia, optymalizacja) • Lab 6,7: algorytmy iteracyjne realizowane w MapReduce <p>Laboratoria Apache Spark:</p> <ul style="list-style-type: none"> • Lab 8,9: praca na RDD • Lab 10,11: DataFrame i Spark SQL • Lab 12,13: MLib budowa i ocena modeli uczenia maszynowego • Lab 14: Structured Streaming przetwarzanie danych strumieniowych i okna czasowe 		
	Treści przedmiotu - projekt		
	Realizacja zadań praktycznych oraz analizy przypadków rozwijających kompetencje z zakresu technologii Big Data. Projektowanie potoków przetwarzania informacji oraz wdrażanie rozwiązań z obszaru inżynierii danych i zaawansowanej analityki danych.		
Wymagania wstępne i dodatkowe	Programowanie w Javie i Pythonie		
Sposoby i kryteria oceniania osiągniętych efektów uczenia się	Sposób oceniania (składowe)	Próg zaliczeniowy	Składowa oceny końcowej
	laboratoria	40.0%	40.0%
	wykład	40.0%	30.0%
	projekt	40.0%	30.0%
Zalecana lista lektur	Podstawowa lista lektur	<ol style="list-style-type: none"> 1. http://hadoop.apache.org/ 2. http://storm.apache.org/ 3. http://spark.apache.org/ 	
	Uzupełniająca lista lektur	<ol style="list-style-type: none"> 1. Hwang, K., Dongarra, J., Fox, G.: Distributed and Cloud Computing: From Parallel Processing to the Internet of Things. Morgan Kaufmann, 2011 2. Karau, H., Konwinski, A., Wendell, P., Zaharia, M.: Learning Spark: Lightning-Fast Big Data Analysis. O'Reilly, 2015 3. Erl, T., Puttini, R., Mahmood, Z.: Cloud Computing: Concepts, Technology, and Architecture. Prentice Hall, 2013 4. Miner, D., Shook, A.: MapReduce Design Patterns: Building Effective Algorithms and Analytics for Hadoop and Other Systems. O'Reilly, 2012 	
	Adresy eZasobów		
Przykładowe zagadnienia/ przykładowe pytania/ realizowane zadania			
Zajęcia praktyczne w ramach przedmiotu	Nie dotyczy		

Dokument wygenerowany elektronicznie. Nie wymaga pieczęci ani podpisu.