

POLITECHNIKA WARSZAWSKA
WYDZIAŁ ELEKTRONIKI I TECHNIK INFORMACYJNYCH

Zarządzenie nr 6 /2023

Dziekana Wydziału Elektroniki i Technik Informatycznych

z dnia 21. 12. 2023 2023 roku

w sprawie wprowadzenia zmiany w Projekcie Organizacyjnym Studiów Podyplomowych „Big Data – przetwarzanie i analiza dużych zbiorów danych”, prowadzonych na Wydziale Elektroniki i Technik Informatycznych Politechniki Warszawskiej.

Działając na podstawie § 11, ust. 4, pkt.5 Regulaminu Organizacyjnego Politechniki Warszawskiej oraz w oparciu o § 5, ust. 1 oraz § 6, ust. 1 pkt 2 zarządzenia nr 36/2020 Rektora PW z dnia 17 czerwca 2020 r. w sprawie zasad tworzenia i organizacji studiów podyplomowych z późn. zm., w powiązaniu z § 6, ust. 2, pkt 9 Regulaminu studiów podyplomowych w Politechnice Warszawskiej, stanowiącego załącznik do uchwały nr 428/XLIX/2019 Senatu PW z dnia 20 listopada 2019 r. z późn. zm.– postanawiam co następuje:

§ 1

1. W Projekcie Organizacyjnym Studiów Podyplomowych „Big Data – przetwarzanie i analiza dużych zbiorów danych”, utworzonych na Wydziale Elektroniki i Technik Informatycznych Politechniki Warszawskiej decyzją nr 115/2016 Rektora PW z dnia 14 lipca 2016 r., wprowadzam następującą zmianę:
 - 1) w pkt. 5.7 Projektu Organizacyjnego w miejsce tekstu „- pozytywna ocena z obrony pracy końcowej” wprowadza się tekst „- pozytywna ocena z egzaminu końcowego”;
 - 2) w załączniku 1b „Opis zakładanych efektów kształcenia” stanowiącym część Programu Studiów Podyplomowych „Big Data – przetwarzanie i analiza dużych zbiorów danych”, wprowadzam następującą zmianę w §3 ust. 3:

w miejsce tekstu: „Ostatnim elementem sprawdzenia osiągniętych wyników kształcenia jest obrona pracy końcowej przygotowywanej częściowo w trakcie poszczególnych laboratoriów. Celem obrony pracy końcowej będzie sprawdzenie zastosowania nabytej wiedzy oraz umiejętności do przygotowania scenariusza postępowania prowadzącego do rozwiązania postawionego problemu z obszaru zagadnień big data.” wprowadza się tekst: „Ostatnim elementem sprawdzenia osiągniętych wyników kształcenia jest egzamin końcowy. Egzamin końcowy będzie miał formę ustną. Komisja egzaminacyjna zadaje przystępującym do egzaminu słuchaczom pytania problemowe. Pozytywna ocena zaproponowanych przez osobę egzaminowaną rozwiązań będzie skutkować zdaniem egzaminu końcowego i ukończeniem studiów podyplomowych Big Data.”
2. Pozostałe zapisy w Projekcie Organizacyjnym nie ulegają zmianie.
3. Projekt Organizacyjny z wprowadzoną zmianą stanowi załącznik do niniejszego zarządzenia,

§ 2

Zmiana w Projekcie Organizacyjnym Studiów Podyplomowych „Big Data – przetwarzanie i analiza dużych zbiorów danych” uzyskała pozytywną opinię Rady Programowej Studiów oraz Rady Wydziału Elektroniki i Technik Informacyjnych (Uchwała nr 19/2023 z dnia 26 września 2023 r.).

§ 3

Zarządzenie wchodzi w życie z dniem podjęcia i obowiązuje do odwołania.

DZIEKAN
Wydziału Elektroniki i Technik Informacyjnych
prof. dr hab. inż. Michał Malinowski

Prof. dr hab. inż. Michał Malinowski
Dziekan Wydziału Elektroniki i Technik Informacyjnych

Zaopiniowano, BOP-5765, radca prawny Tomasz Szołucha (Wa-9027), dnia 12.12.2023 r.

Big Data - przetwarzanie i analiza dużych zbiorów danych

- 1. Nazwa SP**
Big Data - przetwarzanie i analiza dużych zbiorów danych
- 2. Kierunek studiów, z którym związany jest zakres SP**
informatyka
- 3. Wydział prowadzący SP**
Wydział Elektroniki i Technik Informacyjnych
- 4. Program SP**
W załącznikach nr 1a (program) i 1b (opis efektów kształcenia)
- 5. Zasady studiowania na SP**
 - 5.1. Nominalny czas trwania studiów**
1 rok (2 semestry)
 - 5.2. Plan studiów**
W załączniku nr 2
 - 5.3. Zasady przyjmowania na studia**

Udział w SP mogą brać absolwenci studiów wyższych 1. i 2. stopnia.
Przyjęcia w kolejności zgłoszeń, do wyczerpania limitu miejsc.
Limit miejsc: 60 osób
Minimalna liczba zapisanych konieczna do uruchomienia edycji SP: 24 osób
 - 5.4. Zalecana liczebność grup zajęciowych**

Wykłady: wszyscy studenci SP
Zajęcia laboratoryjne: grupy do 12 osób
 - 5.5. Termin dokonywania opłat**

Opłaty są wnoszone w dwóch ratach. Pierwszą ratę należy wpłacić przed rozpoczęciem pierwszego semestru zajęć, drugą ratę należy wpłacić przed rozpoczęciem zajęć drugiego semestru.
 - 5.6. Forma kontroli bieżących postępów studiowania**

Postępy w nauce są oceniane w ramach każdego przedmiotu w skali od 2 do 5 (tzn. 2, 3, 3½, 4, 4½, 5). Ocenami pozytywnymi są oceny ≥ 3, 5 jest oceną bardzo dobrą. Sprawdzian z przedmiotu może mieć charakter kolokwium, egzaminu pisemnego bądź ustnego, prezentacji nabytych wiadomości, pracy domowej lub pokazowego wykonania zadań postawionych przez wykładowcę.
 - 5.7. Warunki otrzymania świadectwa ukończenia**

Zaliczenie wszystkich przedmiotów, w tym:

 - zaliczenie wszystkich zajęć laboratoryjnych i projektowych,
 - zaliczenie wszystkich sprawdzianów,
 - pozytywna ocena z egzaminu końcowego.
- 6. Terminarz i miejsce najbliższej edycji SP**

Zajęcia odbywają się w ramach zjazdów, nominalnie w weekendy (sobota i niedziela), średnio co dwa tygodnie.

Miejsce: Wydział Elektroniki i Technik Informacyjnych PW

7. Okres, na który będą utworzone SP

Początek: październik 2016

Koniec: nieokreślony

8. Wykaz osób przewidzianych do prowadzenia zajęć w pierwszej edycji SP

W załączniku nr 1

9. Opis systemu zapewnienia jakości kształcenia na SP

- Ankietowanie uczestników SP, obejmujące każdorazowo wszystkie nauczane przedmioty.
- Nadzór Rady Programowej nad przygotowaniem materiałów dydaktycznych.

10. Kalkulacja planowanych kosztów SP

Źródło finansowania: Wpłaty uczestników studiów

Wysokość opłat: Koszt Studiów wynosi 10.000 PLN

11. Umowa z podmiotem zewnętrznym

Sages

12. Materiały informacyjne SP na stronę WWW

W załączniku nr 3.

Dodatkowe informacje dostępne pod adresem www.ii.pw.edu.pl/DataScience

Załączniki:

1. Program Studiów Podyplomowych (1a – program, 1b – opis efektów kształcenia)
2. Plan Studiów Podyplomowych
3. Materiały informacyjne na stronę WWW

Program studiów podyplomowych
Tytuł studiów: Big Data - przetwarzanie i analiza dużych zbiorów danych

Przedmiot Konspekt	Kod	Zal.	Odpowiedzialny	Wymiar	
				W	L
Wprowadzenie do technologii Big Data 1. Wprowadzenie do Big Data 2. BI, Big Data a Hurtownie danych 3. Geneza i historia, BigTable, MapReduce, GFS 4. Klasyfikacja problemu 5. Pojęcia real time, batch w kontekście przetwarzania danych 6. Przechowywanie danych - pliki, bazy danych NoSQL 7. Przegląd systemów i platform Big Data 8. Wprowadzenie do MapReduce na przykładzie platformy Hadoop	WBD	Z	mgr Patryk Pilarski	8	
Programowanie zorientowane na dane 9. Praca w systemie Linux 10. Podstawy struktury i składni języka programowania 11. Charakterystyka działania środowiska uruchomieniowego 12. Wykorzystanie kolekcji danych 13. Obsługa wątków 14. Problemy współbieżnego przetwarzania danych 15. Klasyczne problemy synchronizacji 16. Programowanie współbieżne 17. Rozwój współczesnych języków programowania 18. Przydatne biblioteki języka w przetwarzaniu danych	PZD	Z	dr inż. Piotr Nazimek	8	16
Wprowadzenie do baz NoSQL 1. Wprowadzenie do baz NoSQL 2. Przyczyna powstania baz chmurowych 3. Spójność, dostępność, odporność na partycjonowanie 4. Twierdzenie CAP 5. Co różni bazy NoSQL od baz relacyjnych 6. Klasyfikacja i przegląd baz NoSQL 7. Apache Cassandra 7.1. Struktura danych 7.2. Praca z danymi i interfejsy programistyczne 7.3. Wewnętrzna struktura danych i mechanizmy 7.4. Konfiguracja i optymalizacja 8. HBase 8.1. Wprowadzenie do HBase 8.2. Model danych 8.3. Korzystanie i komunikacja 8.4. Integracja z MapReduce	WNS	Z	mgr inż. Tomasz Fortuna, mgr Rafał Kaszczuk	12	20
Projektowanie rozwiązań Big Data 1. Definiowanie problemów Big Data 2. Wymagania w projekcie Big Data 3. Przegląd systemów Big Data 4. Przechowywanie danych 5. Dystrybucje Big Data 6. Rozwiązania chmurowe 7. Rozwiązania integracyjne	PBD	Z	mgr Patryk Pilarski	10	6

8. Wyszukiwanie pełnotekstowe (ElasticSearch, Solr) a rozwiązania Big Data					
Przetwarzanie Big Data za pomocą Apache Hadoop i Spark 1. Paradygmat MapReduce 2. Apache Hadoop 3. Apache Hive 4. Tworzenie aplikacji przetwarzania danych 5. Zarządzanie i monitoring infrastrukturą Apache Hadoop & Family 6. Rozproszone kolekcje obiektów Resilient Distributed Datasets (RDDs) 7. Spark Core 8. Spark SQL 9. Spark Streaming 10. Pozostałe komponenty rodziny Apache Spark	HBD	Z	mgr Patryk Pilarski, mgr inż. Arkadiusz Cacko	16	24
Przetwarzanie Big Data z wykorzystaniem chmur obliczeniowych 1. Podstawowe typy usług na chmurach obliczeniowych 2. Wykorzystanie wirtualnych instancji 3. Wykorzystanie rozwiązań platformowych do składowania i przetwarzania danych 4. Infrastruktura jako kod 5. Projektowanie architektury w oparciu o chmurę 6. Analityka i uczenie maszynowe w chmurach	CBD	Z	mgr inż. Radosław Szmit	8	16
Studium przypadku Prezentacje w ramach przedmiotu obejmują przegląd komercyjnego wykorzystania wybranych metod z obszaru Data Science i Big Data	SP	Z	dr inż. Robert Bembenik	16	-
Pozyskiwanie danych oraz przetwarzanie strumieniowe 1. Źródła danych Big Data <ol style="list-style-type: none"> bazy danych pliki płaski logi usług internet i serwisy społecznościowe aplikacje mobilne 2. Narzędzia do wprowadzania danych do systemów Big Data <ol style="list-style-type: none"> Programy i skrypty własne Apache Flume Apache NiFi Apache MiNiFi Apache Kafka Inne Porównanie 3. Narzędzia do przetwarzania danych strumieniowych <ol style="list-style-type: none"> Apache NiFi Apache Kafka Stream Apache Flink Apache Spark Streaming 	PDPS	Z	mgr inż. Piotr Florczyk	3	12

e. Inne					
f. Porównanie					
lowa aplikacji pozyskującej i przetwarzającej dane za pomocą poznanych narzędzi					
Uczenie maszynowe w rozwiązaniach Big Data 1. Wstęp do uczenia maszynowego w kontekście Big Data 2. Teoria 2.1. Czym jest uczenie maszynowe? 2.2. Regresja i klasyfikacja 2.3. Miary jakości 3. Praktyka 3.1. Podstawy Spark MLlib 3.2. Podstawy Mahout 3.3. Otwarte źródła danych 3.4. Praca z danymi z użyciem Spark MLlib i Mahout 4. Algorytmy uczenia maszynowego 4.1. Uczenie z nadzorem 4.2. Uczenie bez nadzoru 4.3. Metody doboru modelu i poprawiania jakości 5. Przygotowanie pracy końcowej	UMBD	Z	mgr inż. Waldemar Kołodziejczyk	10	16
	Razem			91	110

Opis zakładanych efektów kształcenia**§1**

Określa się następujące obszary kształcenia związane ze studiami podyplomowymi:

1. Obszar nauk technicznych

§2

Ustala się następujące efekty kształcenia programu studiów podyplomowych

Big Data - przetwarzanie i analiza dużych zbiorów danych

w odniesieniu do obszaru kształcenia:

Tabela nr 1. Efekty kształcenia dla programu

Lp.	Obszar nauki	Symbol	Nr	Efekty
1	Obszar nauk technicznych	T1A_W01	BD_1	Ma wiedzę z zakresu statystyki matematycznej i algorytmiki przydatną do formułowania i rozwiązywania zadań z zakresu big data
		T1A_W03	BD_2	Ma uporządkowaną, podbudowaną teoretycznie wiedzę ogólną obejmującą kluczowe zagadnienia z zakresu big data
		T1A_W04	BD_3	Ma szczegółową wiedzę związaną z wybranymi zagadnieniami z zakresu big data
		T1A_W07	BD_4	Zna podstawowe metody, techniki i narzędzia stosowane przy rozwiązywaniu złożonych zadań inżynierskich z zakresu big data
		T1A_U09	BD_5	Potrafi wykorzystać do formułowania i rozwiązywania zadań inżynierskich w dziedzinie big data metody analityczne i eksperymentalne
		T1A_U10	BD_6	Potrafi — przy formułowaniu i rozwiązywaniu zadań inżynierskich — dostrzegać ich aspekty systemowe i pozatechniczne
		T1A_U13	BD_7	Potrafi dokonać krytycznej analizy sposobu funkcjonowania i ocenić istniejące rozwiązania techniczne w zakresie big data
		T1A_U14	BD_8	Potrafi dokonać identyfikacji i sformułować specyfikację zadań inżynierskich o charakterze praktycznym, charakterystycznych dla obszaru big data
		T1A_U15	BD_9	Potrafi ocenić przydatność rutynowych metod i narzędzi służących do rozwiązania prostego zadania inżynierskiego o charakterze praktycznym, charakterystycznego dla obszaru big data oraz wybrać i zastosować właściwą metodę i narzędzia
		T1A_U03	BD_10	Potrafi przygotować w języku polskim opracowanie z zakresu big data
		T1A_U04	BD_11	Potrafi przedstawić w języku polskim prezentację, dotyczącą zagadnień z zakresu big data
		T1A_K02	BD_12	Ma świadomość ważności i rozumie pozatechniczne aspekty i skutki działalności inżynierskiej

Ustala się opis modułów kształcenia obejmujący:

1. Efekty kształcenia i ich odniesienie do efektów kształcenia dla programu oraz przypisanie punktów ECTS

Tabela nr 2. Przedmioty/moduły kształcenia

Lp.	Przedmiot	Liczba godzin zajęć dydaktycznych		Punkty ECTS	Efekty kształcenia (kod, opis)	Odniesienie do efektów kształcenia dla programu
		Teoretycznych	Praktycznych		Sposób weryfikacji efektów	
1	Wprowadzenie do technologii Big Data (WBD)	8	0	1	[1] Zna podstawowe pojęcia dotyczące analizy i przetwarzania dużych zbiorów danych [2] Potrafi wskazać obszary zastosowań podejść typu big data [3] Potrafi opisać podstawowe algorytmy stosowane w obszarze big data Efekty kształcenia będą weryfikowane sprawdzianem ([1], [2], [3])	BD_1 BD_2 BD_4 BD_6
2	Programowanie zorientowane na dane (PZD)	12	20	4	[1] Zna teoretyczne podstawy modelu programowania współbieżnego [2] Zna modele programowania współbieżnego i potrafi je zastosować odpowiednio do problemu [3] Rozwiązuje problemy wykorzystując najlepsze praktyki Efekty kształcenia będą weryfikowane sprawdzianem ([1], [2]) oraz oceną sposobu wykonania zadań laboratoryjnych ([2], [3])	BD_6 BD_8 BD_10
3	Wprowadzenie do baz NoSQL (WNS)	12	20	6	[1] Zna podstawowe twierdzenia dotyczące relacyjnych i nierelacyjnych baz danych [2] Zna typologię baz nierelacyjnych i potrafi dobrać odpowiednią implementację do problemu [3] Potrafi wykorzystać w praktyce jedną baz nierelacyjnych Efekty kształcenia będą weryfikowane sprawdzianem i oceną wykonania zadań laboratoryjnych ([1], [2], [3])	BD_1 BD_2 BD_3 BD_4 BD_7 BD_9
4	Projektowanie rozwiązań Big Data (PBD)	10	6	2	[1] Ma wiedzę o istniejących metodach i narzędziach do przetwarzania dużych zbiorów danych [2] Potrafi zaprojektować system typu big data, dobierając odpowiednie narzędzia do zadanego problemu [3] Potrafi przewidzieć i zaplanować problemy integracji pomiędzy heterogenicznymi systemami i technologiami [4] Ma wiedzę z zakresu zarządzania zasobami, danymi, integracji i bezpieczeństwa Efekty kształcenia będą weryfikowane sprawdzianem ([1], [4]) oraz oceną wykonania zadań laboratoryjnych i projektowych ([2], [3], [4])	BD_2 BD_3 BD_4 BD_6 BD_7 BD_8 BD_9 BD_10 BD_11 BD_12
5	Przetwarzanie Big Data za pomocą Apache Hadoop i Spark (HBD)	16	24	8	[1] Ma wiedzę szczegółową na temat paradygmatu MapReduce i potrafi formułować zadania algorytmiczne w tym paradygmacie [2] Potrafi zaprojektować i zaimplementować aplikację z wykorzystaniem Apache Hadoop i Spark [3] Potrafi analizować i interpretować dane za pomocą Apache Spark i Apache Hive [4] Potrafi przetwarzać dane wsadowe i strumieniowe	BD_3 BD_4 BD_5 BD_7 BD_8 BD_9

					[5] Zna podstawowe aspekty zarządzania infrastrukturą niezbędną do realizacji zadań przetwarzania dużych zbiorów danych Efekty kształcenia będą weryfikowane sprawdzianem ([1], [2], [3], [4]) oraz oceną zaplanowania i wykonania zadań laboratoryjnych ([2], [3], [4])	
6	Przetwarzanie Big Data z wykorzystaniem chmur obliczeniowych (CBD)	8	16	4	[1] Zna podstawowe różnice pomiędzy usługami chmurowymi [2] Potrafi zaimplementować infrastrukturę jako kod [3] Potrafi przetwarzać dane wsadowe i strumieniowe używając usług chmurowych [4] Zna podstawowe techniki projektowania architektury z użyciem usług chmurowych Efekty kształcenia będą weryfikowane sprawdzianem ([1], [2]) oraz oceną wykonania zadań laboratoryjnych ([2], [3], [4])	BD_3 BD_4 BD_5 BD_7 BD_8 BD_9
7	Pozyskiwanie danych oraz przetwarzanie strumieniowe (PDPS)	3	12	2	[1] Posiada wiedzę o źródłach danych Big Data [2] Zna narzędzia do wprowadzania danych do systemów Big Data [3] Zna narzędzia do przetwarzania danych strumieniowych i potrafi używać tych narzędzi w celu przetwarzania takich danych Efekty kształcenia będą weryfikowane oceną wykonania praktycznych zadań warsztatowych ([1], [2], [3])	BD_3 BD_4 BD_5 BD_7 BD_8 BD_9
8	Case studies (CS)	16	0	2	[1] Posiada wiedzę nt. specyfiki wykorzystania metod Data Science i w środowisku komercyjnym [2] Posiada wiedzę o komercyjnym podejściu do rekomendacji i personalizacji w systemach internetowych [3] Posiada wiedzę o wykorzystaniu analityki Big Data w bankowości [4] Posiada wiedzę o wykorzystaniu zasobów informacji niestrukturalnej w dużych przedsiębiorstwach Efekty kształcenia będą weryfikowane sprawdzianem ([1], [2], [3], [4])	BD_3 BD_4 BD_5 BD_7 BD_8 BD_9 BD_12
9	Uczenie maszynowe w rozwiązaniach Big Data (UMBD)	10	16	3	[1] Zna podstawowe pojęcia uczenia maszynowego [2] Zna typowe algorytmy i potrafi dokonać klasyfikacji danych [3] Zna typowe algorytmy i potrafi dokonać grupowania danych [4] Potrafi zastosować metody uczenia maszynowego w kontekście dużych zbiorów danych z wykorzystaniem Spark MLlib i Mahout Efekty kształcenia będą weryfikowane sprawdzianem ([1], [2], [3]) oraz oceną wykonania zadań laboratoryjnych ([2], [3], [4])	BD_3 BD_4 BD_5 BD_7 BD_8 BD_9 BD_10
Razem		91	110	32		

2. Forma prowadzenia zajęć (z odniesieniem do efektów kształcenia)

Efekty kształcenia są osiągnięte w wyniku prowadzonych zajęć teoretycznych (forma wykładów) i praktycznych (forma laboratoriów komputerowych i projektów z wykorzystaniem specjalistycznego oprogramowania) w zakresie poszczególnych przedmiotów. Zajęcia będą bogato ilustrowane przykładami praktycznych zastosowań z dyskusją na temat przyjętych w nich rozwiązań. Zajęcia będą prowadzone z wykorzystaniem nowoczesnych technik nauczania.

3. Sposób sprawdzenia, czy założone efekty dla programu zostały osiągnięte przez studenta

Osiągnięcie założonych efektów dla programu jest weryfikowane w formie sprawdzianów przeprowadzanych w ramach każdego z przedmiotów – dotyczy to przede wszystkim efektów dotyczących wiedzy i, w mniejszym stopniu, umiejętności. Drugim elementem sprawdzania efektów kształcenia jest kontrola rozwiązań zadań praktycznych (wyników i sposobów dojścia do nich) realizowanych na zajęciach laboratoryjnych – w tym przypadku weryfikacja dotyczy przede wszystkim efektów dotyczących umiejętności. Ostatnim elementem sprawdzenia osiągniętych wyników kształcenia jest egzamin końcowy. Egzamin końcowy będzie miał formę ustną. Komisja egzaminacyjna zada przystępującym do egzaminu słuchaczom pytania problemowe. Pozytywna ocena zaproponowanych przez osobę egzaminowaną rozwiązań będzie skutkować zdaniem egzaminu końcowego i ukończeniem studiów podyplomowych Big Data.

4. Matrycę efektów kształcenia (zamierzone efekty kształcenia dla – przedmioty, w których osiągnięty jest efekt)

Tabela nr 3

Lp.	Przedmiot	Efekty											
		BD_1	BD_2	BD_3	BD_4	BD_5	BD_6	BD_7	BD_8	BD_9	BD_10	BD_11	BD_12
1	WBD	x	x		x		x						
2	PZD						x		x		x		
3	WNS	x	x	x	x			x		x			
4	PBD		x	x	x		x	x	x	x	x	x	x
5	HBD			x	x	x		x	x	x			
6	CBD			x	x	x		x	x	x			
7	PDPS			x	x	x		x	x	x			
8	CS			x	x	x		x	x	x			
9	UMBD			x	x	x		x	x	x			

Zaopiniowano, BOP-5765, radca prawny Tomasz Szolucha (Wa-9027), dnia 12.12.2023 r.