

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
УМАНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ САДІВНИЦТВА

Кафедра інформаційних технологій

"ЗАТВЕРДЖУЮ"

Гарант освітньої програми

 Р.І. Ліщук

" 30 " _____ 08 _____ 2023 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
"Технології BigData"

Освітній рівень: Другий рівень вищої освіти
Галузь знань: 12 "Інформаційні технології"
Спеціальність: 122 "Комп'ютерні науки"
Освітня програма: "Комп'ютерні науки"
Факультет: Економіки і підприємництва

Умань – 2023 р.

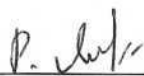
Робоча програма навчальної дисципліни "Технології BigData" для здобувачів вищої освіти спеціальності 122 "Комп'ютерні науки". – Умань: Уманський НУС, 2023 р. – 8 с.

Розробник: д.т.н., професор кафедри
інформаційних технологій _____ В.Ю. Кучерук



Робоча програма затверджена на засіданні кафедри інформаційних технологій
Протокол № 1 від « 30 » _____ 08 2023 року

Завідувач кафедри, к.т.н., доцент _____ Р.І. Лішук
" 30 " _____ 08 2023 року



Схвалено науково-методичною комісією факультету економіки і підприємництва

Протокол № 1 від « 31 » _____ 08 2023 року

Голова _____ Р.П. Мудрак



" 31 " _____ 08 2023 року

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, напрям підготовки, освітній ступінь	Характеристика навчальної дисципліни	
		денна форма навчання	заочна форма навчання
Кількість кредитів – 6	Галузь знань 12 "Інформаційні технології"	Обов'язкова	
Модулів – 1	Спеціальність 122 "Комп'ютерні науки"	Рік підготовки:	
Змістових модулів – 2		1-й	
		Семестр	
Загальна кількість годин – 165 год.		1	
Тижневих годин для денної форми навчання: аудиторних – 3 самостійної роботи студента – 4	Освітній ступінь: магістр	Лекції, год.	
		32	8
		Лабораторні заняття, год.	
		42	4
		Самостійна робота, год.	
		106	168
		Вид контролю: екзамен	

Примітка.

Співвідношення кількості годин аудиторних занять до самостійної роботи становить (%):

для денної форми навчання – 36% / 64%

для заочної форми навчання

2. Мета та завдання навчальної дисципліни

Мета – оволодіння студентами навичок застосування різних методичних підходів до використання технологій та алгоритмів роботи з великими даними.

Завдання – набуття навичок проектування методів та засобів оброблення великих даних, навичок застосування різних технологій проектування систем оброблення великих даних.

Компетентності:

Загальні компетентності:

ЗК01 Здатність до абстрактного мислення, аналізу та синтезу.

ЗК03 Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК06 Здатність бути критичним і самокритичним.

Спеціальні компетентності:

СК03 Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК04 Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проектних рішень.

СК09 Здатність розробляти та адмініструвати бази даних та знань.

Програмні результати навчання:

РН7 Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

РН8 Розробляти математичні моделі та методи аналізу даних (включно з великим).

РН9 Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

3. Програма навчальної дисципліни

Змістовий модуль 1. Основні поняття великих даних

Тема 1. Вступ у великі дані.

Історія виникнення поняття BigData. Джерела великих даних. Інтернет речей. Основні поняття Hadoop, MapReduce. Проблема масштабування великих даних.

Тема 2. Загальний огляд інструментів та технологій великих даних.

Платформа Apache AirFlow. Delta Lake. Механізм запитів Apache Drill. Розподілене середовище Hadoop. Інфраструктура сховища даних Hive. Платформа обробки великих даних HPC Systems. Технології Hudi. Відкритий формат таблиці Iceberg. Платформа розподіленої потокової передачі подій Kafka. Розподілене сховище інформації та аналітична платформа Kylin. Система запитів SQL Presto. Система розподіленої обробки потоків Samza. Механізм обробки та аналізу даних у пам'яті Spark. Розподілена система обчислень у реальному часі Storm. База даних NoSQL.

Тема 3. Розроблення програмного забезпечення для аналізу веб-сайтів.

Можливості інструментів аналізу даних. Роль Python в аналізі даних. Традиційна аналітика великих даних та аналітика нового покоління. Життєвий

цикл аналізу даних. Відкриті дані, їх формати та засоби обробки. Веб-скрепінг. Витягування, перетворення та завантаження даних.

Тема 4. Використання Python для аналізу великих даних.

Форматування даних про час та дату у Python. Читання та запис файлів в Python. Взаємодія із зовнішніми додатками. Програмування Python та SQLite. Призначення утиліти csvsql.

Змістовий модуль 2. Прикладні методи та технології обробки великих даних.

Тема 5. Програмна бібліотека Pandas.

Процедура імпорту даних із файлів у Pandas. Імпорт даних з мережі Інтернет. Засоби для кореляційного аналізу в Pandas. Статистичні підходи до аналітики великих даних. Використання Pandas. Імпорт даних з файлів. Імпорт даних з мережі Інтернет. Описова статистика в Pandas. Засоби для кореляційного аналізу в Pandas.

Тема 6. Технології оброблення великих даних в Python.

Оброблення відсутніх даних. Перетворення типів даних та маніпулювання дата фреймами у Python. Оброблення відсутніх даних. Перетворення типів даних. Маніпулювання дата фреймами. Регресійний аналіз даних в Python. Методи та типи аналізу машинного навчання. Регресійний аналіз. Типи регресійного аналізу. Застосування регресійного аналізу. Помилки в аналізі даних та прогностичній аналітиці. Оцінка помилок регресії засобами Python. Призначення бібліотеки scikit-learn.

Тема 7. Класифікація та візуалізація великих даних.

Проблеми класифікації. Алгоритми класифікації. Візуалізація класифікацій. Застосування та валідація класифікацій. Модуль Rpyplot. Інструмент Plotly. Типи візуалізації даних. Візуалізація аномалій.

Topic 8. Big data analysis using the R programming language.

History of R language development. R language capabilities. Objects, packages, functions. Vectors, matrices and operations on them in R. Factors, lists, frames and operations on them. Exporting, importing, and processing data in R. Exporting and importing data in R. Using R for time series analysis. Data processing in R. Basic tools for data analysis and visualization in R. The plot() function and its parameters. Management of general parameters - arguments of graphic functions. Graph types in R.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин									
	Денна форма					Заочна форма				
	усього	у тому числі				усього	у тому числі			
л		лаб	інд	с.р.	л		лаб	інд	с.р.	
<i>1</i>	2	3	4	5	6	7	8	9	10	11
Змістовий модуль 1. Основні поняття великих даних										
Тема 1. Вступ у великі дані.	20	4	4		12	21	1			20
Тема 2. Загальний огляд	24	4	6		14	21	1			20

інструментів та технологій великих даних.										
Тема 3. Розроблення програмного забезпечення для аналізу веб-сайтів.	20	4	4		12	20	1	1		18
Тема 4. Використання Python для аналізу великих даних.	24	4	6		14	20	1	1		18
Разом за змістовим модулем 1	88	16	20		52	90	4	2		84
Змістовий модуль 2. Прикладні методи та технології обробки великих даних.										
Тема 5. Програмна бібліотека Pandal.	21	4	4		13	21	1			20
Тема 6. Технології оброблення великих даних в Python.	23	4	6		13	23	1			22
Тема 7. Класифікація та візуалізація великих даних.	24	4	6		14	21	1			20
Торік 8. Big data analysis using the R programming language.	24	4	6		14	25	1	2		22
Разом за змістовим модулем 2	92	16	22		54	90	4	2		84
Усього годин	180	32	42		106	180	8	4		168

Теми лабораторних занять

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1.	Основні поняття MapReduce	8	2
2.	Основні поняття Hadoop	8	2
3.	Основні поняття Apache Pig	8	
4.	Основні поняття Apache Hive	8	
5.	Основні поняття Apache Oozie	10	
	Всього	42	4

5. Самостійна робота

№ з/п	Назва теми	Кількість годин	
		денна	заочна
1.	Архітектурні моделі Big Data. Технології віртуалізації. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері Інжиніринг даних.	21	33
2.	Технології Hadoop Big Data. Розподілена обробка MapReduce. Масштабованість за допомогою великих даних. Зберігання та оброблення даних в розподілених файлових системах. Розподілені бази даних. Розподілена файлова система Hadoop (HDFS). MapReduce	21	33
3.	Розподілена потокова платформа Kafka. Переваги Cassandra	21	34
4.	Проблема обчислювальної функції. Технологія Spark.	21	34

	Порівняння Spark та MapReduce. Spark і sparklyr для роботи з великими даними в R.		
5.	Lambda та Карра архітектури оброблення великих даних. Переваги і недоліки Lambda –архітектури. Переваги і недоліки Карра-архітектури	22	34
	Всього	106	168

8. Методи навчання

Під час вивчення дисципліни використовуються такі методи навчання:

- словесні (лекції, пояснення, бесіди, консультації);
- наочні (ілюстрації, презентації, роздатковий матеріал тощо);
- практичні (лабораторні та практичні роботи тощо);
- творчі та проблемно-пошукові (колективне обговорення можливих підходів до вирішення проблемної ситуації).

9. Методи контролю

Оперативний поточний контроль:

- перевірка підготовки студентів до лабораторних занять.

Модульний поточний контроль:

- письмові контрольні роботи за темами лекційного курсу;
- тестування знань студентів з певного розділу (теми) або з певних окремих питань дисципліни;
- розв'язання задач, виконання певних розрахунків, тощо;
- усне опитування;
- виконання та захист лабораторних робіт.

10. Розподіл балів, які отримують студенти

Назва модуля	Змістовий модуль 1				Змістовий модуль 2					Підсумковий контроль	Загальна сума
	Т1	Т2	Т3	Т4	Т5	Т6	Т7	Т8	МК		
Кількість балів за змістовий модуль	20				50						
Теми											
Кількість балів за темами і модульний контроль	5	5	5	5	10	10	10	10	10	30	100

T1, T2 ... Tп - теми змістових модулів.

11. Критерії та шкала оцінювання знань і умінь студентів

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90-100	A	відмінно	зараховано
82-89	B	добре	
74-81	C		
64-73	D	задовільно	
60-63	E		
35-59	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання

0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни
------	---	--	---

12. Рекомендована література

1. Таран В.І., Гордієнко Ю.Г., Стіренко С.Г. Технології BigData. Практикум/ Навчальний посібник, К.: КПІ ім. І. Сікорського, 2022. – 56 с.
2. Akhtar S.M.F. Big Data Architect's Handbook: A Guide to build proficiency in tools and systems used by leading Big Data experts. - Packt Publishing, 2018. — 593 p.
3. Aytas Yusuf. Designing Big Data Platforms: How to Use, Deploy, and Maintain Big Data Systems. - Wiley, 2021. — 326 p.
4. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>
5. IoT Analytics Platform // Електронний ресурс. Режим доступу: <https://blog.codecentric.de/en/2016/07/iot-analytics-platform/>
6. IoT-Analyse-Plattform: Floating Bus Data // Електронний ресурс. Режим доступу: https://www.youtube.com/watch?v=VYxc-3ZRRL4&ab_channel=codecentricAG
7. A brief introduction to two data processing architectures — Lambda and Kappa for Big Data // Електронний ресурс. Режим доступу: <https://towardsdatascience.com/a-brief-introduction-to-two-data-processing-architectures-lambda-and-kappa-for-big-data-4f35c28005bb>