

ПОЛТАВСЬКИЙ ДЕРЖАВНИЙ АГРАРНИЙ УНІВЕРСИТЕТ
Кафедра інформаційних систем та технологій

ЗАТВЕРДЖУЮ

Завідувач кафедри

Юрій УТКІН

«_03_» вересня 2024 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
(обов'язкова фахова навчальна дисципліна)

ТЕХНОЛОГІЇ ОБРОБКИ ВЕЛИКИХ ДАНИХ

освітньо-професійна програма Інформаційні управляючі системи та технології

спеціальність 126 Інформаційні системи та технології

галузь знань 12 Інформаційні технології

освітній ступінь магістр

навчально-науковий інститут економіки, управління, права та інформаційних технологій

Полтава
2024/2025 н. р.

Робоча програма навчальної дисципліни «Технології обробки великих даних»
для здобувачів вищої освіти за освітньо-професійною програмою
Інформаційні управляючі системи та технології спеціальності 126
Інформаційні системи та технології

Мова викладання українська

Розробник: Юрій Поночовний, професор кафедри інформаційних систем та
технологій, д.т.н., професор

«03» вересня 2024 року

Розробник  Юрій ПОНОЧОВНИЙ

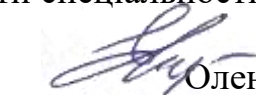
Схвалено на засіданні кафедри інформаційних систем та технологій
протокол від 03.09.2024 р. № 2

Погоджено гарантом освітньої програми Інформаційні управляючі системи та
технології

«03» вересня 2024 року

 Юрій УТКІН

Схвалено головою ради з якості вищої освіти спеціальності «Інформаційні
системи та технології»

 Олена КОПШИНСЬКА

протокол від 03.09.2024 р. № 1

1. Опис навчальної дисципліни

Елементи характеристики	Денна форма навчання 126ICT_мд_2024
Загальна кількість годин	150
Кількість кредитів	5
Місце в індивідуальному навчальному плані здобувача вищої освіти	обов'язкова
Рік навчання (шифр курс)	1
Семестр	1
Лекції (годин)	24
Лабораторні роботи (годин)	26
Самостійна робота (годин)	100
Вид семестрового контролю	екзамен

2. Мета вивчення навчальної дисципліни

Надання теоретичних і практичних знань щодо встановлення, обслуговування та користування системами обробки великих даних та відповідним програмним забезпеченням у професійній та повсякденній діяльності, а також формування логічного мислення.

3. Передумови для вивчення навчальної дисципліни

Перелік базових дисциплін що включені до вступного фахового екзамену: «Інформаційні системи», «Проектування інформаційних систем», «Об'єктно-орієнтоване програмування», «Системи баз даних».

4. Компетентності

загальні:

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

фахові (спеціальні):

СК04. Здатність розробляти математичні, інформаційні та комп'ютерні моделі об'єктів і процесів інформатизації.

СК05. Здатність використовувати сучасні технології аналізу даних для оптимізації процесів в інформаційних системах.

5. Результати навчання

РН01. Відшукувати необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, аналізувати та оцінювати цю інформацію.

РН09. Розробляти і використовувати сховища даних, здійснювати аналіз даних для підтримки прийняття рішень.

Співвідношення програмних результатів навчання із очікуваними результатами навчання

Програмний результат навчання (визначений освітньою програмою)	Очікувані результати навчання навчальної дисципліни
РН01. Відшукувати необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, аналізувати та оцінювати цю інформацію.	вміти знаходити, оцінювати та відбирати релевантні джерела інформації для вивчення сучасних технологій великих даних, таких як Hadoop, Spark, методи розподіленої обробки та паралельної обробки даних
РН09. Розробляти і використовувати сховища даних, здійснювати аналіз даних для підтримки прийняття рішень.	демонструвати здатність проектувати і впроваджувати сховища даних з використанням сучасних технологій, таких як Hadoop, Spark, та методи шардінгу і реплікації, що дозволить ефективно зберігати й обробляти великі дані

6. Методи навчання і викладання

1. Методи організації та здійснення навчально-пізнавальної діяльності:
 - лекція, інструктаж;
 - демонстрування;
 - лабораторні роботи.
2. Методи стимулювання і мотивації навчально-пізнавальної діяльності:
 - метод використання життєвого досвіду; метод відповідей на запитання і опитування думок;
 - роз'яснення мети навчальної дисципліни, висування вимог до вивчення дисципліни, заохочення і покарання, оперативний контроль, вказування на недоліки, зауваження.
3. Інноваційні та інтерактивні методи навчання:
 - використання мультимедійних презентацій, використання комп'ютерних навчальних програм.
4. Методи контролю і самоконтролю за ефективністю навчально-пізнавальної діяльності:
 - перевірка виконання завдань лабораторних робіт; самостійний пошук помилок.

7. Програма навчальної дисципліни:

Тема 1. Предмет і завдання дисципліни. Розуміння концепції великих даних. Мета та завдання дисципліни. Розуміння великих даних. Концепції та термінологія. Характеристики великих даних. Чим технології великих даних відрізняються від традиційної бізнес-аналітики. Проблеми використання великих даних. Визначення еталонної архітектури. Обґрунтування використання архітектурних патернів

Тема 2. Перехід до великих даних і питання планування. Організаційні передумови. Придбання даних. Конфіденційність. Безпека. Походження. Обмеження підтримки в реальному часі. Окремі проблеми з продуктивністю. Особливі вимоги до управління. Методологія. Хмари. Життєвий цикл аналітики великих даних.

Тема 3. Корпоративні технології та бізнес-аналітика для великих даних. Обробка транзакцій в реальному часі (OLTP). Аналітична обробка в реальному часі (OLAP). Вилучення, перетворення і завантаження (ETL). Сховища даних. Вітрини даних. Традиційна бізнес-аналітика. Бізнес-аналітика для великих даних

Тема 4. Концепції зберігання великих даних. Кластери. Файлові системи та розподілені файлові системи. База даних NoSQL. Шардінг (Sharding). Реплікація. Поєднання шардінга і реплікації. Теорема CAP. Принцип проектування бази даних ACID. Принцип BASE

Тема 5. Концепції обробки великих даних. Паралельна обробка даних. Розподілена обробка даних. Hadoop. Обробка робочих завдань. Кластер. Обробка в пакетному режимі. Обробка в режимі реального часу.

Тема 6. Технології зберігання великих даних. Дискові системи зберігання. Системи зберігання в оперативній пам'яті

Тема 7. Основні методи аналізу великих даних. Кількісний аналіз. Якісний аналіз. Data Mining. Статистичний аналіз. Машинне навчання. Семантичний аналіз. Візуальний аналіз

Структура (тематичний план) навчальної дисципліни

Назви тем	Кількість годин			
	денна форма 126ICT_мд_2024			
	усього	у тому числі		
		л	лаб	с.р.
Тема 1. Предмет і завдання дисципліни. Розуміння концепції великих даних	18	2	2	14
Тема 2. Перехід до великих даних і питання планування	23	4	4	15
Тема 3. Корпоративні технології та бізнес-аналітика для великих даних	20	2	4	14
Тема 4. Концепції зберігання великих даних	23	4	4	15
Тема 5. Концепції обробки великих даних	22	4	4	14
Тема 6. Технології зберігання великих даних	22	4	4	14
Тема 7. Основні методи аналізу великих даних	22	4	4	14
Усього годин	150	24	26	100

8. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
		денна форма навчання 126ICT_мд_2024
	Тема 1. Предмет і завдання дисципліни. Розуміння концепції великих даних	
1.	Програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD) з використанням Spark	2
	Тема 2. Перехід до великих даних і питання планування	
2.	Створення та робота з DataFrames з використанням Spark	4
	Тема 3. Корпоративні технології та бізнес-аналітика для великих даних	
3.	Робота з графами з використанням Spark GraphX	4
	Тема 4. Концепції зберігання великих даних	
4.	Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: статистичні обчислення і перевірка гіпотез	4
	Тема 5. Концепції обробки великих даних	
5.	Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: вилучення (TF-IDF, Word2Vec, CountVectorizer і CountVectorizerModel) і перетворення (Tokenizer, StopWordsRemover, n-gram, Binarizer, StringIndexer, IndexToString) ознак.	4
	Тема 6. Технології зберігання великих даних	
6.	Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: перетворення (Interaction, MinMaxScaler, MaxAbsScaler, Bucketizer, ElementwiseProduct, SQLTransformer, VectorAssembler, Imputer) і вибір (VectorSlicer, MinHash) ознак.	4
	Тема 7. Основні методи аналізу великих даних	
7.	Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: кластеризація і колаборативна фільтрація	4
	Разом	26

9. Теми самостійної роботи

Назви тем	Кількість годин
	денна форма навчання 126ICT мд 2024
Тема 1. Предмет і завдання дисципліни. Розуміння концепції великих даних	14
Тема 2. Перехід до великих даних і питання планування	15
Тема 3. Корпоративні технології та бізнес-аналітика для великих даних	14
Тема 4. Концепції зберігання великих даних	15
Тема 5. Концепції обробки великих даних	14
Тема 6. Технології зберігання великих даних	14
Тема 7. Основні методи аналізу великих даних	14
Усього годин	100

10. Індивідуальні завдання

Не передбачено навчальним планом

11. Оцінювання результатів навчання

Програмні результати навчання	Форми контролю програмних результатів навчання
РН01. Відшуковувати необхідну інформацію в науковій і технічній літературі, базах даних, інших джерелах, аналізувати та оцінювати цю інформацію.	Для поточного контролю: - розв'язування тестів; - виконання лабораторних робіт; - виконання завдань самостійної роботи; Для семестрового контролю: - екзамен;
РН09. Розробляти і використовувати сховища даних, здійснювати аналіз даних для підтримки прийняття рішень.	

Критерієм успішного навчання є досягнення здобувачем вищої освіти мінімальних порогових рівнів оцінок за кожним результатом навчання. Мінімальний пороговий рівень оцінки за кожним результатом навчання становить 60 % від максимально можливої кількості балів. Мінімальний пороговий рівень оцінки з освітнього компонента є єдиним в Університеті і не залежить від форм контролю і методів оцінювання результатів навчання.

Схема нарахування балів з навчальної дисципліни

Денна форма навчання 126ICT мд 2024

Назва теми	Форми контролю результатів навчання ЗВО				Разом
	Розв'язування тестів	Виконання лабораторних робіт	Виконання завдань самостійної роботи	Екзамен	
Тема 1. Предмет і завдання дисципліни. Розуміння концепції великих даних		7	1		9
Тема 2. Перехід до великих даних і питання планування		7	1		9
Тема 3. Корпоративні технології та бізнес-аналітика для великих даних		7	1		9
Тема 4. Концепції зберігання великих даних		7	1		9
Тема 5. Концепції обробки великих даних		7	1		9
Тема 6. Технології зберігання великих даних		7	1		9
Тема 7. Основні методи аналізу великих даних	24	7	1		26
Екзамен				20	20
Разом	24	49	7		100

Шкала та критерії оцінювання результатів навчання при проведенні поточного контролю успішності здобувачів вищої освіти

Денна форма навчання І26ІСТ_мд_2024

Виконання лабораторних робіт

Кількість балів	Критерії оцінювання
7 балів (максимальна)	Досягнення мети лабораторної роботи у повному обсязі запланованих результатів навчання, здобувач навів правильні відповіді на всі контрольні питання та продемонстрував вміння роботи на ПК із відповідним програмним продуктом
6 балів	завдання лабораторної роботи виконано самостійно та правильно, поставлений результат та мету досягнуто в повному обсязі, здобувач навів правильні відповіді на половину контрольних питань і продемонстрував посередні вміння роботи на ПК із відповідним програмним продуктом
5 балів	завдання лабораторної роботи виконано самостійно та правильно, поставлений результат та мету досягнуто в повному обсязі, здобувач не навів правильні відповіді контрольні питання, але продемонстрував посередні вміння роботи на ПК із відповідним програмним продуктом
4 бали	завдання лабораторної роботи виконано не в повному обсязі, або ж не самостійно, поставлений результат та мету досягнуто частково, здобувач навів правильні відповіді на всі контрольні питання та продемонстрував вміння роботи на ПК із відповідним програмним продуктом
3 бали	завдання лабораторної роботи виконано не в повному обсязі, або ж не самостійно, здобувач навів правильні відповіді на половину контрольних питань і продемонстрував посередні вміння роботи на ПК із відповідним програмним продуктом
2 бали	завдання лабораторної роботи виконано не в повному обсязі, або ж не самостійно, здобувач не навів правильні відповіді контрольні питання, але продемонстрував посередні вміння роботи на ПК із відповідним програмним продуктом
1 бал	завдання лабораторної роботи виконано не в повному обсязі, або ж не самостійно, здобувач не навів жодної правильної відповіді на контрольні питання і продемонстрував відсутність вмінь роботи на ПК із відповідним програмним продуктом
0 балів (мінімальна)	завдання лабораторної роботи не виконано, поставлений результат та мету не досягнуто, здобувач не навів жодної правильної відповіді на контрольні питання і продемонстрував відсутність вмінь роботи на ПК із відповідним програмним продуктом, що не дає можливість оцінити формування компетентностей і досягнення програмних результатів.

Самостійна робота

1 бал (максимальна)	здобувач навів правильні відповіді на всі контрольні питання та продемонстрував знання, що підтверджують високий рівень опанування результату навчання
0 балів (мінімальна)	здобувач не навів жодної правильної відповіді на контрольні питання і не продемонстрував відсутність опанування результату навчання

Розв'язування тестів

Розв'язування тестів: (до 24 питань) 0-24 бали	24 бали – 100 % правильних відповідей; - за кожен правильну відповідь на питання тесту здобувач отримує 1 бал 0 балів – 0% правильних відповідей, що не дає можливість оцінити формування компетентностей і досягнення програмних результатів
--	--

Шкала та критерії оцінювання результатів навчання здобувачів вищої освіти на екзамені

Вид завдання, кількість балів	Критерії оцінювання тестового завдання екзаменаційного білету в межах зазначеної кількості балів
Відповіді на теоретичні питання у вигляді тестів (відсоток правильних відповідей розраховується в програмі тестування автоматично), 20 балів (максимум) 0 балів (мінімум)	20 балів – більше 95 % правильних відповідей 19 балів – більше 90 % правильних відповідей 18 балів – більше 85 % правильних відповідей 17 балів – більше 80 % правильних відповідей 16 балів – більше 75 % правильних відповідей 15 балів – більше 70 % правильних відповідей 14 балів – більше 65 % правильних відповідей 13 балів – більше 60 % правильних відповідей 12 балів – більше 55 % правильних відповідей 11 балів – більше 50 % правильних відповідей 10 балів – більше 45 % правильних відповідей 9 балів – більше 40 % правильних відповідей 8 балів – більше 35 % правильних відповідей 7 балів – більше 30 % правильних відповідей 6 балів – більше 25 % правильних відповідей 5 балів – більше 20 % правильних відповідей 4 бали – більше 15 % правильних відповідей 3 бали – більше 10 % правильних відповідей 2 бали – більше 5 % правильних відповідей 1 бал – більше 2 % правильних відповідей 0 балів – правильних відповідей немає, що не дає можливість оцінити формування компетентностей і досягнення програмних результатів

12. Інструменти, обладнання та програмне забезпечення, використання яких передбачає навчальна дисципліна (за потреби)

Засоби навчання: персональний комп'ютер (14 шт. – 2017 р., 15 шт. – 2019 р., 15 шт. – 2021 р.), платформа MS Windows 10 Pro (43 ліцензій), Windows 10 Edu (15 ліцензій), MS Office 365 (58 ліцензій) або Libre Office, Google Docs, Internet-браузери, мережа Wi-fi, мультимедійне забезпечення (проектор), проекційний екран, презентації, дошка аудиторна, електронна бібліотека ПДАУ (<https://lib.pdau.edu.ua>), Електронний репозитарій ПДАУ (<http://dspace.pdau.edu.ua>).

Інструменти, обладнання та програмне забезпечення, необхідне для навчальної дисципліни, забезпечують центр підготовки користувачів інформаційної системи «Soft.Farm» – спеціалізована комп'ютерна лабораторія (аудиторія 212), навчально-наукова лабораторія вебтехнологій та хмарних обчислень (аудиторія 203), спеціалізована комп'ютерна лабораторія (аудиторія 213), спеціалізована комп'ютерна лабораторія (аудиторія 202).

13. Політика навчальної дисципліни

Політика навчальної дисципліни визначається системою вимог, які викладач висуває до здобувача вищої освіти при вивченні дисципліни та ґрунтується на засадах справедливості, об'єктивного оцінювання роботи кожного студента і дотримання академічної доброчесності.

Вимоги можуть стосуватися:

1. Термінів виконання та перескладання:

- обов'язковість виконання завдань лабораторних робіт, самостійної роботи і захист

результатів у відведений термін;

- за активну участь у науковій роботі за тематикою кафедри, дисципліни, участь у творчих конкурсах і т. ін. можуть нараховуватися додаткові бали;
- обов'язковість виконання завдань лабораторних робіт, самостійної роботи і захист результатів у відведений термін (за несвоєчасне подання звітів про виконання лабораторної роботи без поважних причин оцінка може бути знижена на 10%).

2. Академічної доброчесності:

Здобувач вищої освіти повинен дотримуватись Кодексу академічної доброчесності та Кодексу про етику викладача та здобувача вищої освіти Полтавського державного аграрного університету. Дотримання академічної доброчесності здобувачами вищої освіти передбачає: самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей); посилення на джерела інформації у разі використання ідей, розробок, тверджень, відомостей; дотримання норм законодавства про авторське право і суміжні права; надання достовірної інформації про результати власної навчальної (наукової, творчої) діяльності, використанні методики досліджень і джерела інформації.

При виявленні академічного плагіату під час виконання запланованих видів робіт такі роботи не зараховуються та повертаються на доопрацювання зі зниженням загальної оцінки мінімум на 20 %.

3. Відвідування занять:

обов'язковість відвідування занять (неприпустимість пропусків без поважних причин, запізнь і т. ін.).

4. Зарахування результатів неформальної/інформальної освіти:

Врахування результатів навчання, отриманих під час неформальної/інформальної освіти та зарахування результатів відбувається згідно Положення про порядок визнання результатів навчання, здобутих у неформальній та інформальній освіті здобувачами вищої освіти Полтавського державного аграрного університету.

5. Оскарження результатів оцінювання:

Порядок оскарження результатів оцінювання здійснюється згідно процедур, затверджених у Положенні про оцінювання результатів навчання здобувачів вищої освіти в Полтавському державному аграрному університеті

14. Рекомендовані джерела інформації

Основні

1. Основи Big Data: Концепції алгоритми та технології / пер.з англ. Анатолія Гладуна; За наук.ред. Олексія Найдю. Дніпро: «Баланс Бізнес Букс», 2018. 320 с.
2. Data Science для бізнесу. Як збирати, аналізувати і використовувати дані / Фостер Провост, Том Фоусет. Видавництво: "Наш Формат" 2019, 400 с.
3. Організація баз даних: навч. посіб. / О. Г. Трофименко, Ю. В. Прокоп [та ін.]. 2-ге вид., випр. і допов. Одеса : Фенікс, 2019. 246 с.

Допоміжні

1. Zaharia M. Spark: The Definitive Guide / M. Zaharia, B. Chambers. Sebastopol: O'Reilly Media Publ., 2017. 630 p.
2. Duvvuri S. Spark for data science / S. Duvvuri, B. Singhal. Birmingham-Mumbai : Packt Publ., 2016. 322 p.
3. Zečević P. Spark-in-action / P. Zečević M. Bonaći. – Shelter Island : Manning Publ., 2017. – 443 p.
4. Malak M.S. Spark GraphX in action / M. S. Malak, R. East. Shelter Island : Manning Publ., 2016. 262 p.

5. Ivanchenko, O., Kharchenko, V., Moroz, B., Kabak, L., Smoktii, K., Ponochovnyi, Y. Availability as a Cloud Service for Control System of Critical Energy Infrastructure // ICT in Education, Research and Industrial Applications: Integration, Harmonization and Knowledge Transfer 2018 / Vadim Ermolayev. et al (edits), CEUR Workshop Proceedings, Volume 2104, 2018, pp. 571-582. http://ceur-ws.org/Vol-2104/paper_262.pdf
6. Поночовний Ю.Л., Пряда О. В., Сорока Ю. О., Дикун Ю. С. Модель пулу серверів для оцінювання енергоспоживання при обробці великих даних. IT Synergy. 2021. № 1. С. 26-31. URL: <http://dspace.pdaa.edu.ua:8080/handle/123456789/12215>
7. Kharchenko V, Ponochovnyi Y, Ivanchenko O, Fesenko H, Illiashenko O. Combining Markov and Semi-Markov Modelling for Assessing Availability and Cybersecurity of Cloud and IoT Systems. Cryptography. 2022; 6(3):44. <https://doi.org/10.3390/cryptography6030044>

Інформаційні ресурси мережі Інтернет

1. Apache Spark (a project managed by the Apache Spark Committee). URL: <https://projects.apache.org/project.html?spark> (дата звернення: 30.08.2023).
2. Ubuntu. Comparing Red Hat OpenStack Platform and Canonical's Charmed OpenStack URL: <https://ubuntu.com> (дата звернення: 30.08.2023).
3. Prometheus: каталог курсів. URL: <https://prometheus.org.ua/courses-catalog/it>
4. Coursera. URL: <https://www.coursera.org/>