

ПОЛТАВСЬКИЙ УНІВЕРСИТЕТ ЕКОНОМІКИ І ТОРГІВЛІ
Навчально-науковий інститут денної освіти
Кафедра комп'ютерних наук та інформаційних технологій

ЗАТВЕРДЖУЮ

Завідувач кафедри КНІТ

 Олена ОЛЬХОВСЬКА

« 28 » __06__ 2024 р.

РОБОЧА ПРОГРАМА

навчальної дисципліни
освітня програма
спеціальність
галузь знань
ступінь вищої освіти

«Технології обробки та аналізу даних»
Комп'ютерні науки
122 Комп'ютерні науки
12 Інформаційні технології
магістр

Робоча програма навчальної дисципліни «Технології обробки та аналізу даних»
рекомендована до використання в освітньому процесі на засіданні кафедри комп'ютерних
наук та інформаційних технологій
Протокол від 28.06.2024 року, №13

Полтава 2024

Укладач: Олексійчук Юрій Федорович, доцент кафедри комп'ютерних наук та інформаційних технологій, к. ф.-м. н.

ПОГОДЖЕНО:

Гарант освітньої програми «Комп'ютерні науки»
спеціальності 122 "Комп'ютерні науки"

ступеня магістр

 Олена ОЛЬХОВСЬКА

підпис

ініціали, прізвище

«_28_» ____ 06 ____ 2024 року

Розділ 1. Опис навчальної дисципліни

Таблиця 1. Опис навчальної дисципліни "Технології обробки та аналізу даних"

Місце у структурно-логічній схемі підготовки	<i>Пререквізити:</i> – Спеціальні (фахові) компетентності зі спеціальності 122 Комп'ютерні науки галузі знань 12 Інформаційні технології для першого (бакалаврського) рівня вищої освіти. <i>Постреквізити:</i> – Хмарні технології та великі дані, Якість програмного забезпечення, Проектування та програмування робото технічних систем, Машинне навчання, Кваліфікаційна робота
Мова викладання	Українська
Статус дисципліни	Обов'язкова
Курс/семестр вивчення	1/1
Кількість кредитів ECTS/кількість модулів	5/2
Денна форма навчання:	
Кількість годин: 150 год	
- Лекції: 20 год	
- Практичні (семінарські, лабораторні) заняття: 40 год	
- Самостійна робота: 90 год.	
- Вид підсумкового контролю (ПМК, екзамен): ПМК	
Заочна форма навчання:	
Кількість годин: 150 год	
- Лекції: 8 год	
- Практичні (семінарські, лабораторні) заняття: 4 год	
- Самостійна робота: 138 год	
- Вид підсумкового контролю (ПМК, екзамен): ПМК	

Розділ 2. Перелік компетентностей, які забезпечує дана навчальна дисципліна, програмні результати навчання

Метою вивчення дисципліни «Технології обробки та аналізу даних» є формування у студентів системного мислення, навичок роботи з фреймворками Spring Data, Spring Boot, використання реляційних та NoSQL баз даних та аналізу даних, в тому числі великих.

Таблиця 2. Перелік компетентностей та програмні результати навчання, які забезпечує навчальна дисципліна «ТОАД»

Програмні результати навчання	Компетентності, якими повинен
-------------------------------	-------------------------------

	оволодіти здобувач
<p>РН8. Розробляти математичні моделі та методи аналізу даних (включно з великим).</p> <p>РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).</p> <p>РН10. Проектувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення</p> <p>РН12. Проектувати та супроводжувати бази даних та знань.</p> <p>РН14. Тестувати програмне забезпечення</p>	<p>ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.</p> <p>ЗК2. Здатність застосовувати знання у практичних ситуаціях.</p> <p>ЗК3. Здатність спілкуватися державною мовою як усно, так і письмово.</p> <p>ЗК5. Здатність вчитися й оволодівати сучасними знаннями.</p> <p>ЗК6. Здатність бути критичним і самокритичним.</p> <p>ЗК7. Здатність генерувати нові ідеї (креативність).</p> <p>СК4. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проектних рішень.</p> <p>СК6. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.</p> <p>СК7. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.</p> <p>СК8. Здатність розробляти і реалізовувати проекти зі створення програмного забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом.</p> <p>СК9. Здатність розробляти та адмініструвати бази даних та знань.</p>

Розділ 3. Програма навчальної дисципліни

Модуль 1. Сучасні Java-фреймворки для роботи з реляційними базами даних

Тема 1. ORM-фреймворки

Object-relational mapping (ORM). Spring Framework. Spring Boot. Spring Data. Spring Data JPA. Hibernate. HQL. Н2. Проектування та супроводжування бази даних та знань. Управління міграцією баз даних. Liquibase. Тестування програмного забезпечення. Тестування SpringBoot-проектів. Тестування REST API. Профілювання.

Тема 2. Мікросервісна архітектура

Проектування архітектурних рішень інформаційних та комп'ютерних систем різного призначення. Мікросервісна архітектура. Патерни проектування мікросервісів. Spring Cloud. Eureka Server. Spring Cloud Gateway. Управління транзакціями в розподілених системах. Вирішення проблеми консистентності даних у розподілених системах.

Модуль 2. NoSQL-бази даних та аналіз даних

Тема 3. NoSQL-бази даних

Недоліки реляційних баз даних. Теорема CAP. Реплікація та сегментація. Пост-реляційні бази даних. NoSQL. Типи NoSQL баз даних. MongoDB. Графова СУБД Neo4J. Spring Data Neo4J. Моделювання даних за допомогою графових баз даних.

Тема 4. Робота з великими даними

Великі дані. Розробка математичних моделей та методів аналізу даних (включно з великим). Розробка алгоритмічного та програмного забезпечення для аналізу даних (включно з великими). MapReduce. Apache Hadoop. HDFS (Hadoop Distributed File System). Apache Spark. RDD (Resilient Distributed Datasets). DataFrames. Аналіз і візуалізація великих даних. Безпека та етика при роботі з великими даними. Майбутні тенденції та технології у великих даних.

Тема 5. Бібліотеки для аналізу потокових даних

Потокові дані. Spring Cloud Stream. Apache Kafka. RabbitMQ. Apache Flink

Розділ 4. Тематичний план навчальної дисципліни

Таблиця 4. Тематичний план навчальної дисципліни для студентів денної форми навчання

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	кількість годин	Назва теми та питання семінарського, практичного або лабораторного заняття	кількість годин	Завдання самостійної роботи в розрізі тем	кількість годин
Модуль 1 Сучасні Java-фреймворки для роботи з реляційними базами даних					
Тема 1. ORM-фреймворки					
Лекція 1. Spring Data 1. Spring Framework та Hibernate 2. Компоненти Spring Framework 3. Налаштування Spring-проєкту 4. Особливості тестування SpringBoot-проєктів	2	Лабораторна робота №1. Spring Data JPA Лабораторна робота №2. Spring Data JPA. Embedded Лабораторна робота №3. Spring Data JPA. Зв'язки між таблицями Лабораторна робота №4. Spring Data JPA. DTO. HQL. Управління міграцією БД. Liquibase Лабораторна робота №5. Spring Data JPA. Тестування REST API Лабораторна робота	2 2 2 2 2	опрацювати лекційний матеріал до теми 1, готуватись до практичних занять, опрацювати питання: 1. Hibernate 2. HQL 3. Сімейство фреймворків Spring	10

1	2	3	4	5	6
<p>5. Типи NoSQL баз даних</p> <p>6. Бази даних в розподілених системах</p> <p>Лекція 5. MongoDB</p> <p>1. Основні поняття</p> <p>2. Типи даних MongoDB</p> <p>3. Програмне забезпечення</p> <p>4. Базові команди для роботи з базами даних</p> <p>Лекція 6. Графова СУБД Neo4J</p> <p>1. Моделювання даних за допомогою графів</p> <p>2. Особливості графової бази даних Neo4J</p> <p>3. Spring Data та Neo4J</p>	2	<p>MongoDB Compass. Аналіз даних в MongoDB</p> <p>Лабораторна робота №13. MongoDB. MongoDB Data</p> <p>Лабораторна робота №14. Графова СУБД Neo4J</p>	2	<p>можливості MongoDB.</p> <p>2. Програмне забезпечення для роботи з MongoDB.</p>	
Тема 4. Робота з великими даними					
<p>Лекція 7. MapReduce</p> <p>1. Функції map() і reduce()</p> <p>2. Розпаралелювання функцій map() і reduce()</p> <p>3. Технологія MapReduce</p> <p>4. Apache Hadoop</p> <p>5. Недоліки Apache Hadoop</p> <p>6. MapReduce в MongoDB</p> <p>Лекція 8. Apache Spark</p> <p>1. Особливості роботи з Apache Spark</p> <p>2. Архітектура Apache Spark</p> <p>3. Spark Core і RDD</p> <p>4. DataFrame</p> <p>5. MLlib</p>	2	<p>Лабораторна робота №14-15. Apache Spark</p>	2	<p>опрацювати лекційний матеріал до теми 4, готуватись до практичних занять</p> <p>Опрацювати теми:</p> <p>1. Робота з Apache Hadoop</p> <p>2. Альтернативні підходи</p>	20
Тема 5. Бібліотеки для аналізу потокових даних					
<p>Лекція 9. Потокові дані та їх обробка</p> <p>1. Основні задачі аналізу даних</p> <p>2. Потокові дані та їх особливості</p> <p>3. Spring Cloud Stream</p>	2	<p>Лабораторна робота №16. Аналіз даних з різних джерел</p> <p>Лабораторна робота №17. Обробка потоків з Apache</p>	2	<p>опрацювати лекційний матеріал до теми 5, готуватись до практичних занять, пройти тести в ДК</p> <p>Опрацювати теми:</p>	20

1	2	3	4	5	6
для обробки потоків 4. Пакетна обробка Лекція 10. Технології для потокового аналізу 1. Apache Kafka 2. RabbitMQ 3. Apache Flink 4. Інтеграція Spring Cloud Stream з Kafka та RabbitMQ	2	Kafka Лабораторна робота №18. Аналіз потокових даних з Apache Flink МКР №2. NoSQL- бази даних та аналіз даних	2 2	1. Spring Cloud Data Flow	
Всього, годин	20		40		90

Таблиця 4. Тематичний план навчальної дисципліни для студентів заочної форми навчання

1	2	3	4	5	6
Назва теми (лекції) та питання теми (лекції)	кіль кіст ь год ин	Назва теми та питання семінарського, практичного або лабораторного заняття	кіль кіст ь год ин	Завдання самостійної роботи в розрізі тем	кіль кіст ь годи н
Модуль 1 Сучасні Java-фреймворки для роботи з реляційними базами даних					
<u>Тема 1. ORM-фреймворки</u>					
Лекція 1. Spring Data 1. Spring Framework та Hibernate 2. Компоненти Spring Framework 3. Налаштування Spring-проєкту 4. Особливості тестування SpringBoot- проєктів	2	Лабораторна робота №1. Spring Data JPA Лабораторна робота №2. Spring Data JPA. Embedded Лабораторна робота №3. Spring Data JPA. Зв'язки між таблицями Лабораторна робота №4. Spring Data JPA. DTO. HQL. Управління міграцією БД. Liquibase Лабораторна робота №5. Spring Data JPA. Тестування REST API Лабораторна робота №6. Тестування SpringBoot-проєктів.	2 2	опрацювати лекційний матеріал до теми 1, готуватись до практичних занять, опрацювати питання: 1. Hibernate 2. HQL 3. Сімейство фреймворків Spring	20
<u>Тема 2. Мікросервісна архітектура</u>					

1	2	3	4	5	6
<p>Лекція 2. Мікросервісна архітектура</p> <ol style="list-style-type: none"> 1. Вступ до мікросервісної архітектури 2. Переваги та недоліки мікросервісів 3. Основні патерни проектування мікросервісів 4. Комунікація між мікросервісами 5. Вирішення проблеми консистентності даних 6. Управління транзакціями в розподілених системах 7. Приклади реалізації мікросервісної архітектури <p>Лекція 3. Spring Cloud</p> <ol style="list-style-type: none"> 1. Spring Cloud та його можливості. 2. Роль Eureka Server у мікросервісах 3. Використання Spring Cloud Gateway для маршрутизації 4. Інструменти для управління конфігураціями 5. Аутентифікація та авторизація в мікросервісах 6. Моніторинг і управління мікросервісами 	2	<p>Лабораторна робота №7. Мікросервіси.</p> <p>Лабораторна робота №8. Eureka Server.</p> <p>Лабораторна робота №9. Spring Cloud Gateway.</p> <p>Лабораторна робота 10. МКР№1. Сучасні Java-фреймворки для роботи з реляційними базами даних</p>		<p>опрацювати лекційний матеріал до теми 2, готуватись до практичних занять, пройти тести в ДК</p> <p>Опрацювати теми:</p> <ol style="list-style-type: none"> 1. Prometheus 2. Grafana 3. Spring Cloud Config 	30
Модуль 2. NoSQL-бази даних та аналіз даних					
<u>Тема 3. NoSQL-бази даних</u>					
<p>Лекція 4. NoSQL бази даних</p> <ol style="list-style-type: none"> 1. NoSQL та Big Data 2. NoSQL та Web 3. NoSQL та мікросервіси 4. Теорема CAP 5. Типи NoSQL баз даних 6. Бази даних в розподілених системах 	2	<p>Лабораторна робота №11. MongoDB. MongoDB Atlas</p> <p>Лабораторна робота №12. MongoDB. MongoDB Compass. Аналіз даних в MongoDB</p>		<p>опрацювати лекційний матеріал до теми 3, готуватись до практичних занять</p> <p>Опрацювати теми:</p> <ol style="list-style-type: none"> 1. Додаткові можливості MongoDB. 2. Програмне забезпечення для 	30

1	2	3	4	5	6
<p>Лекція 5. MongoDB</p> <ol style="list-style-type: none"> 1. Основні поняття 2. Типи даних MongoDB 3. Програмне забезпечення 4. Базові команди для роботи з базами даних <p>Лекція 6. Графова СУБД Neo4J</p> <ol style="list-style-type: none"> 1. Моделювання даних за допомогою графів 2. Особливості графової бази даних Neo4J 3. Spring Data та Neo4J 		<p>Лабораторна робота №13. MongoDB. MongoDB Data</p> <p>Лабораторна робота №14. Графова СУБД Neo4J</p>		роботи з MongoDB.	
<u>Тема 4. Робота з великими даними</u>					
<p>Лекція 7. MapReduce</p> <ol style="list-style-type: none"> 1. Функції map() і reduce() 2. Розпаралелювання функцій map() і reduce() 3. Технологія MapReduce 4. Apache Hadoop 5. Недоліки Apache Hadoop 6. MapReduce в MongoDB <p>Лекція 8. Apache Spark</p> <ol style="list-style-type: none"> 1. Особливості роботи з Apache Spark 2. Архітектура Apache Spark 3. Spark Core і RDD 4. DataFrame 5. MLlib 		Лабораторна робота №14-15. Apache Spark		опрацювати лекційний матеріал до теми 4, готуватись до практичних занять Опрацювати теми: 1. Робота з Apache Hadoop 2. Альтернативні підходи	30
<u>Тема 5. Бібліотеки для аналізу потокових даних</u>					
<p>Лекція 9. Потокові дані та їх обробка</p> <ol style="list-style-type: none"> 1. Основні задачі аналізу даних 2. Потокові дані та їх особливості 3. Spring Cloud Stream для обробки потоків 4. Пакетна обробка <p>Лекція 10. Технології для потокового аналізу</p>	2	<p>Лабораторна робота №16. Аналіз даних з різних джерел</p> <p>Лабораторна робота №17. Обробка потоків з Apache Kafka</p> <p>Лабораторна робота №18. Аналіз</p>		опрацювати лекційний матеріал до теми 5, готуватись до практичних занять, пройти тести в ДК Опрацювати теми: 1. Spring Cloud Data Flow	28

1	2	3	4	5	6
1. Apache Kafka 2. RabbitMQ 3. Apache Flink 4. Інтеграція Spring Cloud Stream з Kafka та RabbitMQ		потокових даних з Apache Flink МКР №2. NoSQL-бази даних та аналіз даних			
Всього, годин	8		4		138

Розділ 5. Система оцінювання знань студентів

Таблиця 5.1 - Розподіл балів за результатами вивчення навчальної дисципліни

Вид робіт	Максимальна кількість балів
Модуль 1 Сучасні Java-фреймворки для роботи з реляційними базами даних	
<i>Практичні роботи 1-2 (тема 1)</i>	6 практичних роботи x 4 балів = 24
<i>Практичні роботи 3-8 (тема 2)</i>	3 практичних робіт x 4 балів = 12
<i>Практична робота 9</i> (Контрольна робота №1)	14
Всього за модулем 1	50
Модуль 2. NoSQL-бази даних та аналіз даних	
<i>Практичні роботи 10-13 (тема 3)</i>	4 практичних роботи x 4 балів = 16
<i>Практичні роботи 14-15 (тема 4)</i>	2 практичних роботи x 4 балів = 8
<i>Практичні роботи 16-19 (тема 5)</i>	3 практичних роботи x 4 балів = 12
<i>Практична робота 20</i> (Контрольна робота №2)	14
Всього за модулем 2	50
Всього по курсу	<u>100</u>

Система нарахування додаткових балів за видами робіт з вивчення навчальної дисципліни

Форма роботи	Вид роботи	Бали
Навчальна робота	Опрацювання додаткових навчальних курсів, написання та публічний захист наукового реферату на тему, що відповідає дисципліні	10*

За додаткові види навчальних робіт студент може отримати не більше 10 балів. Додаткові бали додаються до загальної підсумкової оцінки за вивчення навчальної дисципліни, але загальна підсумкова оцінка не може перевищувати 100 балів.

Таблиця 6 – Шкала оцінювання знань здобувачів вищої освіти за результатами вивчення навчальної дисципліни

Сума балів за всі види навчальної діяльності	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою
90-100	А	Відмінно
82-89	В	Дуже добре

74-81	C	Добре
64-73	D	Задовільно
60-63	E	Задовільно достатньо
35-59	FX	Незадовільно з можливістю повторного складання
0-34	F	Незадовільно з обов'язковим повторним вивченням навчальної дисципліни

Розділ 6. Інформаційні джерела

1. Davis, Adam L., and Adam L. Davis. "Spring data." Spring Quick Reference Guide: A Pocket Handbook for Spring Framework, Spring Boot, and More (2020): 43-59.
2. Yang, Herong. JDBC for SQL Server-Herong's Tutorial Examples. HerongYang. Com, 2019.
3. Siahaan, Vivian, and Rismon Hasiholan Sianipar. SQLite For Beginners: Learn Fundamentals of Queries and Implement PyQt-Based Projects Easily. SPARTA PUBLISHING, 2019.
4. Pollack, Mark, Oliver Gierke, Thomas Risberg, Jon Brisbin, and Michael Hunger. Spring Data: modern data access for enterprise Java. "O'Reilly Media, Inc.", 2012.
5. Siahaan, Vivian, and Rismon Hasiholan Sianipar. Java In Practice: JDBC And Database Applications. Sparta Publishing, 2019.
6. Добролюбова М. В. Програмування баз даних: конспект лекцій [Електронний ресурс] : навч. посіб. для студ. спеціальності 152 «Метрологія та інформаційно-вимірвальна техніка» / М. В. Добролюбова; КПІ ім. Ігоря Сікорського. – Київ : КПІ ім. Ігоря Сікорського, 2021. – 275 с.
7. Олещенко Л.М. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем») / Л.М. Олещенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с.
8. Bradshaw, Shannon, Eoin Brazil, and Kristina Chodorow. MongoDB: the definitive guide: powerful and scalable data storage. O'Reilly Media, 2019.
9. Bierer, Doug. Learn MongoDB 4. x: A guide to understanding MongoDB development and administration for NoSQL developers. Packt Publishing Ltd, 2020.
10. Giamas, Alex. Mastering MongoDB 6. x: Expert techniques to run high-volume and fault-tolerant database solutions using MongoDB 6. x. Packt Publishing Ltd, 2022.
11. Singh, Chanchal, and Manish Kumar. Mastering Hadoop 3: Big data processing at scale to unlock unique business insights. Packt Publishing Ltd, 2019.
12. Wiktorski, Tomasz. Data-intensive systems: Principles and fundamentals using Hadoop and spark. Springer International Publishing, 2019.
13. Jeyaraj, Rathinaraja, Ganeshkumar Pugalendhi, and Anand Paul. Big Data with Hadoop MapReduce: A Classroom Approach. Apple Academic Press, 2020.
14. Gelman, Andrew, Jennifer Hill, and Aki Vehtari. Regression and other stories. Cambridge University Press, 2020.
15. Pardoe, Iain. Applied regression modeling. John Wiley & Sons, 2020.
16. Олексійчук Ю. Ф., Ольховська О. В., Ольховський Д. М., Орлова Д. І. Проектування та розробка web-сервісу для генерування та розсилки pdf-документів. / Системи та технології, 65(1), 2023 – С. 39-45. <https://doi.org/10.32782/2521-6643-2023.1-65.5>

Розділ 7. Програмне забезпечення навчальної дисципліни

Використовується наступне програмне забезпечення:

1. Персональний комп'ютер або ноутбук зі сталим доступом до мережі Інтернет.
2. Система дистанційного навчання ПУЕТ та його засоби відео-конференц зв'язку, дистанційний курс з ОК «Технології обробки та аналізу даних».

3. Середовище розробки для мови програмування Java IntelliJ IDEA Community.
4. Maven.
5. MongoDB Compass.
6. Apache Spark.
7. СУБД MySQL
8. Postman
9. Docker