

Міністерство освіти і науки України
Національний аерокосмічний університет ім. М. Є. Жуковського
«Харківський авіаційний інститут»

Кафедра комп'ютерних систем, мереж і кібербезпеки (№503)

ЗАТВЕРДЖУЮ

Голова НМК

 Д.М. Крицький
(підпис) (ініціали та прізвище)

«31» серпня 2021р.

**РОБОЧА ПРОГРАМА ОБОВ'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Інженерія програмного забезпечення. Частина 2
(назва навчальної дисципліни)

Галузь знань: 12 "Інформаційні технології"
(шифр і найменування галузі знань)

Спеціальність: 123 "Комп'ютерна інженерія"
(код та найменування спеціальності)

Освітня програма: Комп'ютерні системи та мережі
(найменування освітньої програми)

Освітня програма: Системне програмування
(найменування освітньої програми)

Освітня програма: Програмовні мобільні системи та Інтернет речей
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2021 рік

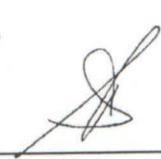
Розробник: Фесенко Г. В., доцент, д.т.н., доцент
(прізвище та ініціали, посада, науковий ступінь та вчене звання)


(підпис)

Робочу програму розглянуто на засіданні кафедри _____
комп'ютерних систем, мереж і кібербезпеки
(назва кафедри)

Протокол № 1 від «30» 08 2021 р.

Завідувач кафедри д.т.н., професор
(науковий ступінь та вчене звання)


(підпис)

В. С. Харченко
(ініціали та прізвище)

1. Опис навчальної дисципліни

Найменування показника	Галузь знань, спеціальність, освітня програма, рівень вищої освіти	Характеристика навчальної дисципліни (денна форма навчання)
Кількість кредитів – 4	Галузь знань <u>12 «Інформаційні технології»</u> (шифр і найменування) Спеціальність <u>123 «Комп’ютерна інженерія»</u> (код і найменування) Освітня програма <u>Комп’ютерні системи та мережі, Системне програмування, Програмовні мобільні системи та Інтернет речей</u> (найменування) Рівень вищої освіти: перший (бакалаврський)	Обов’язкова
Кількість модулів – 1		Навчальний рік
Кількість змістовних модулів – 2		2021/2022
<u>Індивідуальне завдання</u> - немає (назва)		Семестр
Загальна кількість годин – 48/120		<u>8-й</u>
		Лекції*
		<u>24</u> години
		Практичні, семінарські*
		<u>0</u> годин
Кількість тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 4,5	Лабораторні*	
	<u>24</u> години	
	Самостійна робота	
	<u>72</u> години	
	Вид контролю	
	іспит	

Співвідношення кількості годин аудиторних занять до самостійної роботи становить: **48/72**.

*Аудиторне навантаження може бути зменшене або збільшене на одну годину залежно від розкладу занять.

2. Мета та завдання навчальної дисципліни

Мета: формування у студентів сучасного рівня інформаційної та програмістської культури, оволодіння основними принципами інженерії програмного забезпечення, набуття ними практичних навичок самостійної розробки програмного забезпечення і використання сучасних інформаційних технологій для розв'язання практичних задач, а також надання знань та навичок щодо застосування технологій, концепцій зберігання, обробки та методів аналізу великих даних під час організації корпоративних сховищ даних та розробки програмного забезпечення бізнес-аналітики.

Завдання: формування теоретичних знань та практичних умінь у сфері розробки програмного забезпечення на всіх етапах життєвого циклу; використання інструментарію інженерії програмного забезпечення для роботи в великими даними

Компетентності, які набуваються:

- здатність до абстрактного мислення, аналізу і синтезу;
- здатність вчитися і оволодівати сучасними знаннями;
- здатність застосовувати знання у практичних ситуаціях;
- здатність спілкуватися державною мовою як усно, так і письмово;
- здатність використовувати сучасні методи і мови програмування для розроблення алгоритмічного та програмного забезпечення;
- здатність створювати системне та прикладне програмне забезпечення комп'ютерних систем та мереж;
- здатність використовувати та впроваджувати нові технології, включаючи технології розумних, мобільних, зелених і безпечних обчислень, брати участь в модернізації та реконструкції комп'ютерних систем та мереж, різноманітних вбудованих і розподілених додатків, зокрема з метою підвищення їх ефективності;
- здатність системно адмініструвати, використовувати, адаптувати та експлуатувати наявні інформаційні технології та системи;
- здатність проектувати системи та їхні компоненти з урахуванням усіх аспектів їх життєвого циклу та поставленої задачі, включаючи створення, налаштування, експлуатацію, технічне обслуговування та утилізацію.

Очікувані результати навчання:

- знати новітні технології в галузі комп'ютерної інженерії;
- вміти системно мислити та застосовувати творчі здібності до формування нових ідей.

Крім того студенти повинні вміти організовувати зберігання, обробку та аналіз великі даних, використовуючи сучасні технології паралельної та розподіленої обробки великих даних в пакетному та реальному режимах.

Пререквізити – «Бази даних», «Паралельні та розподілені обчислення», «Інженерія програмного забезпечення», «Моделі та структури даних».

Кореквізити – Дипломне проектування».

3. Програма навчальної дисципліни

Модуль 1.

Змістовий модуль 1. Основи великих даних

Тема 1. Розуміння великих даних.

Набори та аналіз даних. Дескриптивна, діагностична, прогностична та прескриптивна аналітика з використанням великих даних. Бізнес аналітика. Ключові показники ефективності. Характеристики великих даних (об'єм, швидкість, різноманіття, достовірність, цінність). Різні типи даних (структуровані, неструктуровані, слабо структуровані та метадані).

Тема 2. Перехід до великих даних та питання планування.

Організаційні передумови. Набуття даних. Конфіденційність. Безпека. Походження. Обмеження підтримки у реальному часі. Особливі проблеми продуктивності. Особливі вимоги до керівництва. Методологія. Великі дані і хмарні технології. Життєвий цикл великих даних (оцінка ситуації, ідентифікація даних, збирання та фільтрація даних, діставання даних, перевірка та очищення даних, агрегування та представлення даних, аналіз даних, візуалізація даних та використання результатів аналізу).

Тема 3. Корпоративні технології і бізнес-аналітика для великих даних.

Обробка транзакцій в режимі реального часу (OLTP). Аналітична обробка у реальному часі (OLAP). Вилучення, перетворення та завантаження (ETL). Сховища даних. Вітрини даних. Традиційна бізнес-аналітика (спеціалізовані звіти, інформаційні панелі). Бізнес-аналітика для великих даних (традиційна візуалізація даних, візуалізація даних під час обробки великих даних).

Модульний контроль.

Змістовий модуль 2. Зберігання та аналіз великих даних

Тема 4. Концепції зберігання великих даних.

Кластери. Файлові системи та розподілені файлові системи. NoSQL. Шардінг. Реплікація (режим реплікації «ведучий-ведений», режим реплікації «одноранговий»). Шардінг і реплікація (об'єднання шардінга і реплікації у режимі «ведучий-ведений», об'єднання шардінга і реплікації у режимі «одноранговий»). Теорема CAP. ACID. BASE.

Тема 5. Концепції обробки великих даних.

Паралельна обробка даних. Розподілена обробка даних. Hadoop. Обробка робочих завдань (пакетна обробка). Транзакційна обробка. Кластер. Обробка у пакетному режимі (Пакетна обробка за допомогою MapReduce. Map. Об'єднання. Розбиття. Перетасування та сортування. Reduce. Інтерпретація

алгоритмів MapReduce). Обробка в режимі реального часу (об'єм, узгодженість, швидкість (ОУШ), обробка потоку подій, обробка складних подій в режимі реального часу і ОУШ), обробка великих даних в режимі реального часу і MapReduce. Apache Spark.

Тема 6. Технології зберігання великих даних.

Дискові пристрої зберігання (розподілені файлові системи, системи управління базами даних (СУБД), бази даних NoSQL: характеристики, логічне обґрунтування, об'єм, швидкість, різноманіття, типи, ключ-значення, документ, об'єднання стовпців, граф, база даних NewSQL). Системи зберігання у оперативній пам'яті (In-memory Data Grid, наскрізне читання, наскрізний запис, відкладений запис, оновлення, In-memory Databases).

Тема 7. Основні методи аналізу великих даних.

Кількісний аналіз. Якісний аналіз. Data Mining. Статистичний аналіз (A/B тестування, кореляція, регресія). Машинне навчання (класифікація, кластеризація, виявлення викидів, фільтрація). Семантичний аналіз (обробка природної мови, обробка тексту, аналіз емоціонального забарвлення висловлювань). Візуальний аналіз (кольорові карти, часові ряди, мережеві графи). Співвідношення просторових даних.

Модульний контроль.

4. Структура навчальної дисципліни

Назва змістового модуля і тем	Кількість годин				
	Денна форма				
	Усього	У тому числі			
		л	п	лаб.	с.р.
1	2	3	4	5	6
Модуль 1					
Змістовний модуль 1. Основи великих даних					
Тема 1. Розуміння великих даних	20	4		4	12
Тема 2. Перехід до великих даних та питання планування	20	4		4	12
Тема 3. Корпоративні технології і бізнес-аналітика для великих даних	19	4		3	12
Модульний контроль	1			1	
Разом за змістовним модулем 1	60	12		12	36
Змістовний модуль 2. Зберігання та аналіз великих даних					
Тема 4. Концепції зберігання великих даних	13	2		2	9
Тема 5. Концепції обробки великих даних	17	4		4	9
Тема 6. Технології зберігання великих даних	15	4		2	9
Тема 7. Основні методи аналізу великих даних	14	2		3	9
Модульний контроль	1			1	
Разом за змістовним модулем 2	60	12		12	36
Усього годин	120	24		24	72

5. Теми семінарських занять

№ п/п	Назва теми	Кількість годин
1	<i>Не передбачено</i>	
2		
	Разом	

6. Теми практичних занять

№ п/п	Назва теми	Кількість годин
1	<i>Не передбачено</i>	
2		
	Разом	

7. Теми лабораторних занять

№ п/п	Назва теми	Кількість годин
1	Програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD) з використанням Spark	4
2	Створення та робота з DataFrames з використанням Spark	4
3	Робота з графами з використанням Spark GraphX	4
4	Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: статистичні обчислення і перевірка гіпотез	4
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
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	2
7	Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: кластеризація і колаборативна фільтрація	2
	Разом	24

8. Самостійна робота

№ п/п	Назва теми	Кількість годин
1	Загальні принципи аналізу даних за допомогою Spark	12
2	Особливості роботи зі стійкими розподіленими наборами даних (RDD)	12
3	Особливості роботи с парами «ключ-значення»	12
4	Пакет для роботи зі структурованими даними Spark SQL	9
5	Бібліотека для обробки графів та виконання паралельних обчислень Spark GraphX	9
6	Компонент для роботи з потоковими даними Spark Streaming	9
7	Бібліотека Spark Mlib для реалізації машинного навчання	9
	Разом	72

9. Індивідуальні завдання

Не передбачено

10. Методи навчання

Проведення аудиторних лекцій, лабораторних занять, консультацій, а також самостійна робота студентів за відповідними матеріалами.

11. Методи контролю

Проведення поточного контролю, тестового модульного контролю, підсумковий контроль у вигляді екзамену.

12. Критерії оцінювання та розподіл балів, які отримують студенти

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях	0...2	6	0...12
Виконання і захист лабораторних робіт	0...6	3	0...18
Модульний контроль	0...20	1	0...20
Змістовний модуль 2			
Робота на лекціях	0...2	6	0...12
Виконання і захист лабораторних робіт	0...6	4	0...24
Модульний контроль	0...24	1	0...24
Усього за семестр			0...100

Контроль знань при проведенні занять оцінюється за такими шкалами:

1) активність на лекції під час відповідей на питання:

- повна відповідь на питання – 2 бали;
- неповна відповідь – 1 бал;
- відсутність на лекції – 0 балів;

2) виконання і захист лабораторних робіт:

- виконані всі завдання лабораторної роботи, оформлені відповідно до вимог, надано вичерпні правильні відповіді на всі запитання – 6 балів;
- виконані всі завдання лабораторної роботи, мають місце незначні відхилення від оформлення відповідно до вимог, надано правильні відповіді на всі запитання – 5 балів;
- виконані всі завдання лабораторної роботи, мають місце незначні відхилення від оформлення відповідно до вимог, надано правильні відповіді не на всі запитання – 4 бали;
- виконано 85-99% завдань лабораторної роботи, мають місце незначні відхилення від оформлення відповідно до вимог, надано правильні відповіді не на всі запитання – 3 бали;
- виконано 65-84% завдань лабораторної роботи, мають місце значні відхилення від оформлення відповідно до вимог, надано правильні відповіді не на всі запитання – 2 бали;
- виконано 50-64% завдань лабораторної роботи, мають місце значні відхилення від оформлення відповідно до вимог, надано правильні відповіді не на всі запитання – 1 бал;
- студент не допущений до захисту роботи (виконано менше 50% її завдань) або робота не захищена – 0 балів.

На модульний контроль (всього 20 балів для змістовного модуля 1 та всього 24 бали для змістовного модуля 2) виносяться всі пройдені за контрольований період теми, які включаються у варіанти тестових завдань, що містять по 20 питань для змістовного модуля 1 та 24 питання для змістовного модуля 2. Максимальна кількість балів за кожне тестове питання – 1.

Семестровий контроль у вигляді іспиту проводиться у разі відмови студента від балів поточного тестування й за наявності допуску до заліку. Під час складання семестрового іспиту студент має можливість отримати максимум 100 балів.

Білет для іспиту складається з одного теоретичного та одного практичного запитань, максимальна кількість за кожне із запитань, складає 50 балів.

Критерії оцінювання роботи здобувача протягом семестру

Задовільно (60-74). Мати мінімум знань та умінь. Відпрацювати та захистити всі лабораторні роботи та домашні завдання. Вміти самостійно давати характеристику використовуваній технології зберігання, обробки та аналізу великих даних, вміти програмування операції зі стійкими розподіленими наборами даних та DataFrames з використанням Spark.

Добре (75 - 89). Твердо знати мінімум знань, виконати усі завдання. Показати вміння виконувати та захищати всі лабораторні роботи в обумовлений викладачем строк з обґрунтуванням рішень та заходів, які запропоновано у роботах. Вміти пояснювати вибір технологій зберігання, обробки та аналізу великих даних, вміти програмувати операції зі

стійкими розподіленими наборами даних, DataFrames, обробляти графи з використанням Spark. Розуміти корпоративні технології і бізнес-аналітику для великих даних та вміти застосовувати їх для вирішення практичних завдань.

Відмінно (90 - 100). Повно знати основний та додатковий матеріал. Знати усі теми. Орієнтуватися у підручниках та посібниках. Досконально знати усі основні технології та фреймворки, які використовуються для зберігання, обробки та аналізу великих даних. Вміти удосконалювати корпоративну технологію та бізнес аналітику для великих даних. Вміти програмувати операції зі стійкими розподіленими наборами даних, DataFrames, обробляти графи з використанням Spark, реалізовувати алгоритми машинного навчання з використанням бібліотеки Spark MLlib. Безпомилково виконувати та захищати всі лабораторні роботи в обумовлений викладачем строк з докладним обґрунтуванням рішень та заходів, які запропоновано у роботах.

Шкала оцінювання: бальна і традиційна

Сума балів	Оцінка за традиційною шкалою	
	Іспит, диференційований залік	Залік
90 – 100	Відмінно	Зараховано
75 – 89	Добре	
60 – 74	Задовільно	
0 – 59	Незадовільно	Не зараховано

13. Методичне забезпечення

Дистанційний курс в системі дистанційного навчання Moodle, розташований за адресою: <https://moodle.csn.khai.edu/course/view.php?id=37>.

14. Рекомендована література

Базова

1. ISO/IEC TR 20547-1:2020. Information technology – Big data reference architecture – Part 1: Framework and application process.
2. ISO/IEC TR 20547-2:2020. Information technology – Big data reference architecture – Part 2: Use cases and derived requirements.
3. ISO/IEC DIS 20547-3:2020. Information technology – Big data reference architecture – Part 3: Reference architecture.
4. ISO/IEC CD 20547-4:2020. Information technology – Big data reference architecture – Part 4: Security and privacy.
5. ISO/IEC TR 20547-5:2020. Information technology – Big data reference architecture – Part 5: Standards roadmap.
6. Balusamy B., Abirami N., Kadry S., Gandomi A. *Big Data: Concepts, Technology, and Architecture*. USA, NJ, Hoboken : Wiley, 2021. 371 p.
7. Pop F., Neagu G. *Big Data Platforms and Applications: Case Studies, Methods, Techniques, and Performance Evaluation*. Switzerland, Cham : Springer, 2021. 300 p.

8. Ravi V., Cherukuri A. *Handbook of Big Data Analytics*. Vol. 1. *Methodology*. UK, London : The Institution of Engineering and Technology, 2021. 390 p.
9. Ravi V., Cherukuri A. *Handbook of Big Data Analytics*. Vol. 2. *Applications in ICT, security and business analytics*. UK, London : The Institution of Engineering and Technology, 2021. 420 p.
10. Ерл Т., Хаттак В., Булер П. *Основи Big Data: концепції, алгоритми й технології* : пер. з англ. Дніпро: Баланс Бізнес Букс, 2018. 320 с.

Допоміжна

1. Akhtar S. *Big Data Architect's Handbook*. Birmingham – Mumbai : Packt Publ., 2018. 477 p.
2. Chellappan S., Ganesan D. *Practical Apache Spark*. USA, New York : Apress, 2018. 288 p.
3. Chinnici M., Pop F., Neagu G. *Data Science and Big Data Analytics in Smart Environments* . USA, Boca Raton, FL : CRC Press, 2021. 305 p.

15. Інформаційні ресурси

1. The 9 Best Free Online Big Data And Data Science Courses. URL: <https://bernardmarr.com/the-9-best-free-online-big-data-and-data-science-courses>.
 2. Big Data And Data Science Courses. URL: <https://www.edx.org/learn/big-data>.
- Mastering Big Data Analytics. URL: <https://www.mygreatlearning.com/academy/learn-for-free/courses/mastering-big-data-analytics>.