

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

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

ЗАТВЕРДЖУЮ



_____ Д.М. Крицький

(підпис)

(ініціали та прізвище)

« 31 » серпня 2022 р.

**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

Освітня програма: _____ Системне програмування
(найменування освітньої програми)

Форма навчання: денна

Рівень вищої освіти: перший (бакалаврський)

Харків 2022 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 2022 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 – 4,0	<p style="text-align: center;">Галузь знань <u>12 «Інформаційні технології»</u> (шифр і найменування)</p> <p style="text-align: center;">Спеціальність <u>123 «Комп'ютерна інженерія»</u> (код і найменування)</p> <p style="text-align: center;">Освітня програма <u>Комп'ютерні системи та мережі</u></p> <p style="text-align: center;"><u>Системне програмування</u> (найменування)</p> <p style="text-align: center;">Рівень вищої освіти: перший (бакалаврський)</p>	Вибіркова
Кількість модулів – 1		Навчальний рік
Кількість змістовних модулів – 2		2022/2023
Індивідуальне завдання _____ (назва)		Семестр
		<u>8-й</u>
Загальна кількість годин – 48/120		Лекції*
Кількість тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 4,5		<u>24</u> години
		Практичні, семінарські*
		<u>0</u> годин
		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

1. Мета вивчення: 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.

2. 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, а також:

- придбання знань про методи зберігання та обробки великих даних з використанням сучасних фреймворків;
- придбання знань про основні методи аналізу великих даних та технології використання нереляційних баз даних;
- придбання знань про технології паралельної та розподіленої обробки великих даних в пакетному та реальному режимах.

3. Програмні компетентності. Дисципліна має допомогти сформувати у студентів такі компетентності

- (ЗК1) здатність до абстрактного мислення, аналізу і синтез;
- (ЗК2) здатність вчитися і оволодівати сучасними знаннями;
- (ЗК3) здатність застосовувати знання у практичних ситуаціях;
- (ЗК4) здатність спілкуватися державною мовою як усно, так і письмово;
- (ЗК7) вміння виявляти, ставити та вирішувати проблеми;
- (ФК2) здатність використовувати сучасні методи і мови програмування для розроблення алгоритмічного та програмного забезпечення;
- (ФК3) здатність створювати системне та прикладне програмне забезпечення комп'ютерних систем та мереж;
- (ФК7) 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;
- (ФК9) здатність системно адмініструвати, використовувати, адаптувати та експлуатувати наявні інформаційні технології та системи;
- (ФК14) здатність проектувати системи та їхні компоненти з урахуванням усіх аспектів їх життєвого циклу та поставленої задачі, включаючи створення, налаштування, експлуатацію, технічне обслуговування та утилізацію.

4. Програмні результати навчання. В результаті вивчення дисципліни студенти мають досягти такі програмні результати навчання:

- (ПРНЗ) знати новітні технології в галузі комп'ютерної інженерії;
- (ПРН8) 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.

Міждисциплінарні зв'язки: Дисципліна базується на знаннях, отриманих під час вивчення дисциплін у циклі загальної і професійної підготовки, передбачених навчальним планом спеціальності.

Матеріал дисципліни базується на знаннях, отриманих під час вивчення дисциплін із циклу загальної підготовки, зокрема «Гуманітарна дисципліна за вибором студента», «Теорія ймовірностей та математична статистика».

Матеріал дисципліни базується на знаннях, отриманих під час вивчення дисциплін із циклу професійної підготовки, а саме «Інженерія програмного забезпечення. Частина 1», «Кросплатформенні технології».

Матеріал, засвоєний під час вивчення цієї дисципліни, є базою, необхідною для виконання дипломної роботи (проекту) бакалавра.

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					
Змістовий 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 змістовим модулем 1	60	12		12	36
Змістовий 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 змістовим 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	4

	використанням Spark	
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 навчальним планом

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

12.1. Розподіл балів, які отримують студенти (кількісні критерії оцінювання)

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях	0...2	3	0...6
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	4	0...8
Виконання і захист лабораторних робіт	0...6	4	0...24
Модульний контроль	0...24	1	0...24
Усього за семестр			0...100

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

Білет для іспиту/заліку складається з двох теоретичних і одного практичного запитання. За перше та друге запитання студент отримує по 30 балів, за практичне – 40 балів.

12.2. Якісні критерії оцінювання

Необхідний обсяг знань для одержання позитивної оцінки:

- знати види аналітики з використанням великих даних, основні характеристики та типи вихідних даних.
- знати життєвий цикл великих даних.
- знати основні концепції та технології зберігання великих даних.
- знати основні концепції обробки великих даних.

Необхідний обсяг вмінь для одержання позитивної оцінки:

- вміти здійснювати програмування операцій зі стійкими розподіленими наборами даних.
- вміти працювати з DataFrames та графами з використанням Spark GraphX.
- вміти здійснювати статистичні обчислення і перевірку гіпотез з використанням Spark MLlib і пакета spark.ml.

12.3 Критерії оцінювання роботи студента протягом семестру

Задовільно (60 - 74). Показати необхідний обсяг знань та вмінь для одержання позитивної оцінки відповідно до п.12.2. Захистити не менше 80% від усіх завдань лабораторних занять. Вміти самостійно давати характеристику видам аналітики з використанням великих даних, знати основні характеристики та типи вихідних даних. Вміти програмувати операції зі стійкими розподіленими наборами даних.

Добре (75 - 89). Твердо знати мінімум знань, виконати не менше 90% завдань лабораторних занять. Вміти пояснювати основні концепції та технології зберігання великих даних, вміти працювати з DataFrames та графами, використовуючи Spark. Вміти здійснювати статистичні обчислення і перевірку гіпотез з використанням Spark MLlib і пакета spark.ml.

Відмінно (90 - 100). Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та вміти їх застосовувати.

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

1. Навчально-методичний комплекс дисципліни розміщений на кафедральному сервері у відповідному каталозі

2. 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. Эрл Т. Основы Big Data: концепции, алгоритмы и технологии [Текст] : пер с англ. / Т. Эрл, В. Хаттак, П. Булер. – Днепр : Баланс Бизнес Букс, 2018. – 320 с.

2. Радченко И. А. Технологии и инфраструктура Big Data [Текст] / И. А. Радченко, И. Н. Николаев. – СПб : Университет ИТМО, 2018. – 52 с.

3. Просто о больших данных [текст] / Дж. Гурвиц, А. Ньюджент, Ф. Халпер, М. Кауфман. – Москва : Эксмо, 2015. – 400 с.

4. Furht В. Big Data technologies and applications [Text] / В. Furht, F. Villanustre. – Cham : Springer Publ., 2016. – 400 p.

5. Big Data : a primer [Text] / Н. Mohanty, Р. Bhuyan, D. Chenthati (Eds.) . – New Delhi : Springer India Publ., 2016. – 400 p.

Допоміжна

1. Изучаем Spark: молниеносный анализ данных [Текст] : пер с англ. / Х. Карау, Э. Конвински, П. Венделл, М. Захария. – М. : ДМК Пресс, 2015. – 304 с.

2. Spark для профессионалов: современные паттерны обработки больших данных [Текст] : пер с англ. / С. Риза, У. Лезерсон, Ш. Оуэн, Дж. Уиллс. – СПб.: Питер, 2017. – 272 с.

3. Duvvuri S. Spark for data science [Text] / S. Duvvuri, B. Singhal. – Birmingham – Mumbai : Packt Publ., 2016. – 322 p.

4. Zečević P. Spark-in-action [Text] / P. Zečević M. Bonaći. – Shelter Island : Manning Publ., 2017. – 443 p.

5. Malak M.S. Spark GraphX in action [Text] / M. S. Malak, R. East. – Shelter Island : Manning Publ., 2016. – 262 p.

15. Інформаційні ресурси

1. Як великі дані допомагають потрапити в точку призначення [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/wayray/blog/407681/>

2. Огляд кейсів цікавих впроваджень Big Data в компаніях фінансового сектора [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/post/353008/>

3. Як великим операторам заробляти великі гроші на великих даних? [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/huawei/blog/280980/>

4. Курси по великим даним від Coursera [Електронний ресурс]. – Режим доступу: <https://www.coursera.org/learn/real-time-streaming-big-data?specialization=big-data-engineering>

5. Великі дані [Електронний ресурс]. – Режим доступу: [http://www.tadviser.ru/index.php/Статья:Большие_данные_\(Big_Data\)](http://www.tadviser.ru/index.php/Статья:Большие_данные_(Big_Data)).

6. 12 кейсів по великим даним [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/newprolab/blog/318208/>

7. Big Data від А до Я. Частина 1: Принципи роботи з великими даними, парадигма MapReduce [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/dca/blog/267361/>

8. Big Data від А до Я. Частина 2: Hadoop [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/dca/blog/268277/>

9. Big Data від А до Я. Частина 3: Прийоми і стратегії розробки MapReduce-додатків [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/dca/blog/270453/>

10. Big Data від А до Я. Частина 4: Hbase [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/dca/blog/280700/>

11. Apache Spark - переваги, недоліки, побажання [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/post/329838/>
12. Введення в Apache Spark [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/mlclass/blog/250811/>
13. Знайомство з Apache Spark [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/piter/blog/276675/>
14. Apache Spark: що там під капотом? [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/post/251507/>
15. Навіщо вивчати Spark [Електронний ресурс]. – Режим доступу: https://habr.com/ru/company/epam_systems/blog/336090/
16. Spark SQL. Трохи про оптимізатор запитів [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/neoflex/blog/417103/>
17. Apache Spark: з open source в індустрію [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/netologyru/blog/331728/>
18. Навчання на великих даних: Spark MLlib [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/mlclass/blog/251471/>
19. Класифікація великих обсягів даних на Apache Spark з використанням довільних моделей машинного навчання [Електронний ресурс]. – Режим доступу: <https://habr.com/ru/company/lanit/blog/413137/>
20. Документація по Apache Spark [Електронний ресурс]. – Режим доступу: <https://spark.apache.org/>
21. Big Data і бізнес-аналітика [Електронний ресурс]. – Режим доступу: <https://www.croc.ru/solution/business-solutions/applications/analytics/>
22. Методи і техніки аналізу великих даних [Електронний ресурс]. – Режим доступу: <https://www.it.ua/ru/knowledge-base/technology-innovation/big-data-bolshie-dannye>