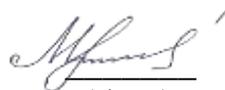


Міністерство освіти і науки України
Національний аерокосмічний університет
«Харківський авіаційний інститут»

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

ЗАТВЕРДЖУЮ

Гарант освітньої програми


(підпис)

Ольга МОРОЗОВА
(ім'я та ПРІЗВИЩЕ)

« 29 » _____ серпня _____ 2025 р.

**СИЛАБУС ОBOB'ЯЗКОВОЇ
НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Технології обробки великих даних
(назва навчальної дисципліни)

Галузь знань: F «Інформаційні технології»
(шифр і найменування галузі знань)

Спеціальність: F7 «Комп'ютерна інженерія»
(код і найменування спеціальності)

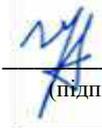
Освітньо-професійна програма: «Системне програмування»
(найменування освітньої програми)

Рівень вищої освіти: **другий (магістерський)**

Силабус введено в дію з 01.09.2025

Харків – 2025 р.

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


(підпис)

Силабус навчальної дисципліни розглянуто на засіданні кафедри _____
комп'ютерних систем, мереж і кібербезпеки
(назва кафедри)

Протокол № 1 від « 29 » 08 2025 р.

Завідувач кафедри _____
(науковий ступінь і вчене звання)


(підпис)

Вячеслав ХАРЧЕНКО
(ім'я та ПРІЗВИЩЕ)

Погоджено з представником здобувачів освіти:


(підпис)

Дмитро ВАСИК
(ім'я та ПРІЗВИЩЕ)

1. Загальна інформація про викладача



ПІБ: Фесенко Герман Вікторович

Посада: професор

Науковий ступінь: доктор технічних наук

Вчене звання: професор

Перелік дисциплін, які викладає:

Технології великих даних,
Технології обробки великих даних,
Великі дані для кібербезпеки,
Обробка та аналіз результатів наукових досліджень з використанням ІТ,
Теорія і методи data science і штучного інтелекту.

Напрями наукових досліджень:

Гарантоздатність безпілотних інтелектуальних систем, моніторинг та кібербезпека об'єктів критичної інфраструктури, великі дані.

Контактна інформація:

h.fesenko@csn.khai.edu

2. Опис навчальної дисципліни

Форма здобуття освіти	Денна
Семестр	1
Мова викладання	Українська
Тип дисципліни	Обов'язкова
Обсяг дисципліни: кредити ЄКТС/ кількість годин	4,5 кредити ЄКТС / 135 годин (48 аудиторних, з яких: лекції – 32, лабораторні – 16; СРЗ – 87)
Види навчальної діяльності	Лекції, практичні заняття, самостійна робота
Види контролю	Поточний контроль, модульний контроль, семестровий контроль – іспит
Пререквізити	Дисципліна є обов'язковим компонентом освітньої програми і базується на знаннях, отриманих під час вивчення дисциплін у циклі загальної і професійної підготовки, передбачених навчальним планом спеціальності

3. Мета та завдання навчальної дисципліни, переліки компетентностей та очікуваних результатів навчання

Мета – формування знань та умінь про застосування технологій розподіленої обробки структурованих та неструктурованих наборів великих даних з використанням сучасних методів та інструментів.

Завдання:

- формування знань і навичок щодо організації зберігання великих даних;
- формування знань і навичок щодо організації розподіленої обробки великих даних;
- отримання знань щодо побудови і функціонування еталонної архітектури великих даних.

Компетентності, які набуваються:

Загальні компетентності (ЗК):

- ЗК2. Здатність до абстрактного мислення, аналізу і синтезу.
- ЗК4. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- ЗК6. Здатність виявляти, ставити та вирішувати проблеми.

Спеціальні компетентності (СК):

- СК7. Здатність досліджувати, розробляти та обирати технології створення великих і надвеликих систем.
- СК11. Здатність обирати ефективні методи розв'язування складних задач комп'ютерної інженерії, критично оцінювати отримані результати та аргументувати прийняті рішення.

Програмні результати навчання (ПРН):

- ПРН2. Знаходити необхідні дані, аналізувати та оцінювати їх.
- ПРН10. Здійснювати пошук інформації в різних джерелах для розв'язання задач комп'ютерної інженерії, аналізувати та оцінювати цю інформацію.

4. Зміст навчальної дисципліни

МОДУЛЬ 1

Змістовний модуль 1. Технології зберігання та потокової обробки великих даних

Тема 1. Розуміння великих даних

Стисла анотація: Термінологія та стандарти у галузі великих даних. Характеристики великих даних. Організаційні передумови, придбання, конфіденційність, безпека даних та походження даних. Особливі вимоги до управління, методологія, хмарні середовища. Життєвий цикл аналітики великих даних. Обробка транзакцій (OLTP) та аналітична обробка (OLAP) в реальному часі, добування, перетворення і завантаження (ETL). Сховища та вітрини даних.

Лабораторна робота №1. Робота зі вбудованими функціями модуля Spark SQL.

Самостійна робота: Опрацювання матеріалу лекцій. Підготовка до захисту лабораторних робіт.

Самостійна робота: Опрацювання матеріалу лекцій.

Тема 2. Концепції та технології зберігання великих даних

Стисла анотація: Кластери, файлові системи та розподілені файлові системи, база даних NoSQL. Шардінг і реплікація. Поєднання шардінга і реплікації. Теорема CAP, принцип проектування бази даних ACID та принцип BASE. Дискові системи зберігання. Системи зберігання в оперативній пам'яті.

Самостійна робота: Опрацювання матеріалу лекцій. Аналіз спільних та відмінних рис технологій IoT та IoT.

Тема 3. Вступ до потокової обробки великих даних. Внесення, транспортування та аналіз поточкових даних

Стисла анотація: Вступ до потокової обробки даних. Отримання даних від клієнтів: внесення даних. Транспортування даних із ланки збору даних: розчленування конвеєра даних. Аналіз поточкових даних.

Лабораторна робота №2. Робота зі Spark Machine Learning Library: Transformers and Estimators.

Самостійна робота: Опрацювання матеріалу лекцій. Підготовка до захисту лабораторних робіт.

Тема 4. Алгоритми аналізу поточкових великих даних. Збереження результатів збору або аналізу даних

Стисла анотація: Обмеження та їх ослаблення. Часові аспекти потокової обробки. Методи узагальнення. Необхідність організації довговічного сховища. Зберігання даних у пам'яті. Сеансова персоналізація.

Самостійна робота: Опрацювання матеріалу лекцій. Підготовка до захисту лабораторних робіт.

Змістовний модуль 2. Особливості потокової обробки великих даних з використанням Kafka, Kafka Streams та Flink.

Тема 5. Отримання доступу до даних. Можливості кінцевих пристроїв та обмеження доступу до даних

Стисла анотація: Патерни взаємодії. Протоколи надсилання даних клієнтам. Фільтрування потоку. Основні концепції: достатня швидкість читання, запам'ятовування стану, пом'якшення наслідків втрати даних, обробка рівно один раз, введення у вебклієнт.

Лабораторна робота №3. Робота зі Spark Machine Learning Library: Supervised Learning (навчання з учителем).

Самостійна робота: Опрацювання матеріалу лекцій. Підготовка до захисту лабораторних робіт.

Тема 6. Потокова обробка з використанням Kafka

Стисла анотація: Екскурсія по Kafka. Різні пакети похідного коду Kafka і як вони працюють. Потокова обробка та термінологія в контексті роботи Kafka. Виробники. Споживачі. Трасування. Маркування місцезнаходження. Стиснені теми та читання з них. Брокери. Темі. Розділи. Kafka як сховище. Архітектури з використанням Kafka. Kafka Streams.

Самостійна робота: Опрацювання матеріалу лекцій.

Тема 7. Потокова обробка з використанням Flink

Стисла анотація: Стислі відомості про Flink. Компоненти набору Flink. Розгортання застосунків Flink. Виконання завдання. Високоступна конфігурація. Збої TaskManager та JobManager. Передача даних у Flink. Водяні знаки. Управління станом.

Лабораторна робота №4. Робота зі Spark Machine Learning Library: Recommendation Engines.

Самостійна робота: Опрацювання матеріалу лекцій. Підготовка до захисту лабораторних робіт.

Тема 8. Функційна архітектура мережі, керованої великими даними

Стисла анотація: Загальні відомості про мережу, керовану великими даними. Функційна архітектура площини великих даних для мережі, керованої великими даними. Функційна архітектура площини мережі для мережі, керованої великими даними. Функційна архітектура площини управління для мережі, керованої великими даними. Приклади застосування мереж, керованих великими даними.

Самостійна робота: Опрацювання матеріалу лекцій.

5. Індивідуальні завдання

Не передбачено

6. Методи навчання

Проведення аудиторних лекцій, лабораторних робіт, консультацій, а також самостійна робота здобувачів з використанням відповідних матеріалів (п.11, 12).

7. Методи контролю

Проведення поточного контролю, електронного тестування, підсумковий контроль у вигляді іспиту.

8. Критерії оцінювання та розподіл балів, які отримують здобувачі освіти

Таблиця 8.1 – Розподіл балів, які отримують здобувачі освіти

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях	0...1	8	0...8
Виконання і захист лабораторних робіт	0...11	2	0...22
Модульний контроль	0...20	1	0...20
Змістовний модуль 2			
Робота на лекціях	0...1	8	0...8
Виконання і захист лабораторних робіт	0...11	2	0...22
Модульний контроль	0...20	1	0...20
Усього за семестр			0...100

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

Білет для іспиту складається із двох теоретичних та одного практичного запитання, максимальна кількість балів за кожне теоретичне запитання, складає 34 балів, а за практичне – 32 балів.

Таблиця 8.2 – Шкали оцінювання: бальна і традиційна

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

Критерії оцінювання роботи здобувача освіти протягом семестру

Задовільно (60-74) – Показати мінімум знань та умінь. Захистити не менше 75% від усіх завдань лабораторних занять. Знати основні концепції обробки та зберігання великих даних. Уміти працювати зі вбудованими функціями модуля Spark SQL та зі Spark Machine Learning Library з використанням Transformers and Estimators.

Добре (75-89) – Твердо знати мінімум, захистити не менше 90% завдань лабораторних занять. Знати ключові принципи організації розподіленої та потокової обробки великих даних, а також принципи побудови еталонної архітектури великих даних. Уміти працювати зі вбудованими функціями модуля Spark SQL та зі Spark Machine Learning Library з використанням Transformers, Estimators та Supervised Learning.

Відмінно (90-100) – Здати всі контрольні точки з оцінкою «відмінно». Досконально знати всі теми та уміти їх застосовувати. Вміти ефективно застосовувати функції модуля Spark SQL та основні функції Spark Machine Learning Library.

9. Політика навчального курсу

Відвідування занять. Інтерактивний характер курсу передбачає обов'язкове відвідування лабораторних занять. Здобувачі освіти, які за певних обставин не можуть відвідувати лабораторні заняття регулярно, повинні протягом тижня узгодити із викладачем графік індивідуального відпрацювання пропущених занять. Окремі пропущені заняття мають бути відпрацьовані на найближчій консультації протягом тижня після їх пропуску. Відпрацювання занять здійснюється усно у формі співбесіди за питаннями, визначеними планом заняття. В окремих випадках дозволяється письмове відпрацювання пропущених занять шляхом виконання індивідуального письмового завдання.

Дотримання вимог академічної доброчесності здобувачами освіти під час вивчення навчальної дисципліни. Під час вивчення навчальної дисципліни здобувачі освіти мають дотримуватися загальноприйнятих морально-етичних норм і правил поведінки, вимог академічної доброчесності, передбачених Положенням про академічну доброчесність Національного аерокосмічного університету «Харківський авіаційний інститут»

(<https://khai.edu/assets/files/polozhennya/polozhennya-pro-akademichnu-dobrochesnist.pdf>). Очікується, що роботи здобувачів освіти будуть їх оригінальними дослідженнями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших здобувачів освіти становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в письмовій роботі здобувача освіти є підставою для її незарахування викладачем.

Вирішення конфліктів. Порядок і процедури врегулювання конфліктів регламентуються Кодексом етичної поведінки в Національному аерокосмічному університеті «Харківський авіаційний інститут» (<https://khai.edu/ua/university/normativna-baza/ustanovchi-dokumenti/kodeks-etichnoi-povedinki/>).

10. Методичне забезпечення

1. Сторінка дисципліни у системі дистанційного навчання «Ментор» [Ел. ресурс]. URL: <https://mentor.khai.edu/course/view.php?id=3702>

11. Рекомендована література

Базова

1. Олещенко Л. М. *Технології оброблення великих даних*. Конспект лекцій. Київ : КПІ ім. Ігоря Сікорського, 2021. 227 с.
2. Shapira C., Palino T., Sivaram R., Petty K. *Kafka: The Definitive Guide*. USA, CA, Sebastopol : O'Reilly Media, 2022. 486 p.
3. Fischer J., Wang N. *Grokking Streaming Systems: Real-time event processing*. USA, NY, Shelter Island : Manning Publications Co., 2022. 312 p.
Hueske F., Kalavri V. *Stream Processing with Apache Flink*. USA, CA, Sebastopol : O'Reilly Media, 2019. 534 p.

Допоміжна

1. Perrin J.-G. *Spark in action*. USA, NY, Shelter Island : Manning Publications Co., 2020, 576 p.
2. Ravi V., Cherukuri A. *Handbook of Big Data Analytics*. Vol. 1. *Methodology*. UK, London : The Institution of Engineering and Technology, 2021. 390 p.
3. 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.

12. Інформаційні ресурси

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