

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

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

ЗАТВЕРДЖУЮ

Гарант освітньої програми


Анатолій ШОСТАК
(підпис) ім'я та ПРІЗВИЩЕ

« 29 » серпня 2025 р.

СИЛАБУС ОBOB'ЯЗКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Технології великих даних»
(назва навчальної дисципліни)

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

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

Освітня програма: Системне програмування
(найменування освітньої програми)

Рівень вищої освіти: перший (бакалаврський)

Силабус введено в дію з 01.09.2025 року

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 д.т.н., професор
(науковий ступінь і вчене звання)


(підпис)

Вячеслав ХАРЧЕНКО
(ім'я та ПРІЗВИЩЕ)

Представник здобувачів освіти:



(підпис)

(ім'я та ПРІЗВИЩЕ)

Поліна ОГАРКО

2. Опис навчальної дисципліни

Форма здобуття освіти	<i>Денна, заочна</i>
Семестр	7 семестр
Мова викладання	Українська
Тип дисципліни	<i>Обов'язкова</i>
Обсяг дисципліни: кредити ЄКТС/ кількість годин	<i>денна: 4 кредити ЄКТС / 120 годин (64 аудиторних, з яких: лекції - 32, лабораторні - 32; СРЗ - 56) заочна: 4 кредити ЄКТС / 120 годин (8 аудиторних, з яких: лекції - 4, лабораторні - 4; СРЗ - 112)</i>
Види навчальної діяльності	Лекції та лабораторні заняття, самостійна робота
Види контролю	Поточний контроль, модульний контроль, семестровий контроль – залік
Пререквізити	Основи функціонування комп'ютерів, Технології програмування, Мобільне програмування, Технології Java
Кореквізити	Тестування та забезпечення якості

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

Мета – надання студентам знань та навичок щодо застосування технологій великих даних та розробки прикладного програмного забезпечення з їх використанням.

Завдання:

- придбання студентами знань про методи зберігання та обробки великих даних з використанням сучасних фреймворків та вмінь щодо їхнього застосування;

- придбання студентами знань про основні методи аналізу великих даних та технології використання нереляційних баз даних та вмінь щодо їхнього застосування;

- придбання студентами знань про технології паралельної та розподіленої обробки великих даних в пакетному та реальному режимах та вмінь щодо їхнього застосування.

Компетентності, які набуваються:

Загальні компетентності:

ЗК2. Здатність вчитися і оволодівати сучасними знаннями.

ЗК3. Здатність застосовувати знання у практичних ситуаціях.

Фахові компетентності:

ФК5. Здатність використовувати засоби і системи автоматизації проектування до розроблення компонентів комп'ютерних систем та мереж, Інтернет додатків, кіберфізичних систем тощо.

ФК9. Здатність системно адмініструвати, використовувати, адаптувати та експлуатувати наявні інформаційні технології та системи.

Програмні результати навчання:

ПРН2. Мати навички проведення експериментів, збирання даних та моделювання в комп'ютерних системах.

ПРН3. Знати новітні технології в галузі комп'ютерної інженерії.

ПРН4. Знати та розуміти вплив технічних рішень в суспільному, економічному, соціальному і екологічному контексті.

ПРН6. Вміти застосовувати знання для ідентифікації, формулювання і розв'язування технічних задач спеціальності, використовуючи методи, що є найбільш придатними для досягнення поставлених цілей.

ПРН8. Вміти системно мислити та застосовувати творчі здібності до формування нових ідей.

ПРН11. Вміти здійснювати пошук інформації в різних джерелах для розв'язання задач комп'ютерної інженерії.

ПРН15. Вміти виконувати експериментальні дослідження за професійною

тематикою.

ПРН19. Здатність адаптуватись до нових ситуацій, обґрунтовувати, приймати та реалізовувати у межах компетенції рішення.

ПРН20. Усвідомлювати необхідність навчання впродовж усього життя з метою поглиблення набутих та здобуття нових фахових знань, удосконалення креативного мислення.

4. Зміст навчальної дисципліни

МОДУЛЬ 1

Змістовний модуль 1. Основи 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 (Online Transaction Processing, OLTP) та аналітична обробка (Online Analytical Processing, OLAP) в реальному часі, добування, перетворення і завантаження (Extract, Transform, Load, ETL). Сховища та вітрини даних.

Лекція 1. Розуміння великих даних. Частина 1.

Лекція 2. Розуміння великих даних. Частина 2.

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.

Лекція 3. Концепції та технології зберігання великих даних. Частина 1.

Лекція 4. Концепції та технології зберігання великих даних. Частина 2.

Лабораторна робота 1. Програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD), робота з Data Frame та таблицями з використанням Spark.

Самостійна робота: Опрацювання матеріалу лекції. Підготовка до захисту лабораторної роботи.

Тема 3. Екосистема Hadoop.

Стисла анотація: Основні відомості про Hadoop та його екосистему. Взаємодія користувача з кластером Hadoop. Порівняння систем управління базами даних на основі SQL з Hadoop. Модель розподілених обчислень великих масивів даних MapReduce. Розподілена файлова система Hadoop. Менеджер ресурсів YARN.

Лекція 5. Екосистема Hadoop. Частина 1.

Лекція 6. Екосистема Hadoop. Частина 2.

Лекція 7. Екосистема Hadoop. Частина 3.

Самостійна робота: Опрацювання матеріалу лекції.

Тема 4. Spark.

Стисла анотація: Загальні відомості про Spark. Spark у процесі обробки та інженерії даних. Spark у наукових дослідженнях у галузі обробки даних. Фрейм даних. Ментальна модель Spark. Особливості використання фрейму даних. Взаємодія зі Spark. Стислий огляд компонент та взаємодій між ними. Порівняння Hadoop і Spark.

Лекція 8. Spark.

Лабораторна робота 2. Робота з графами з використанням Spark GraphX.

Самостійна робота: Опрацювання матеріалу лекції. Підготовка до захисту лабораторної роботи. Підготовка до модульного контролю.

Модульний контроль 1.

Змістовний модуль 2. Особливості побудови та застосування архітектур великих даних.

Тема 5. Архітектури зберігання та обробки великих даних.

Стисла анотація: Архітектура сховища великих даних тільки для читання (read-only data stores, RDS): знайомство з архітектурою RDS; розподіл відповідальності команд та запитів; служби і компоненти сховища великих даних тільки для читання; рекомендації щодо використання RDS. Парадигма data lakehouse. Розроблення системного контексту для data lakehouse. Розроблення логічної архітектури data lakehouse. Розроблення принципів архітектури data lakehouse. Поглинання та обробка даних в data lakehouse.

Зберігання та обслуговування даних у data lakehouse. Забезпечення можливостей аналітики в data lakehouse. Впровадження управління даними в data lakehouse.

Лекція 9. Архітектури зберігання та обробки великих даних. Частина 1.

Лекція 10. Архітектури зберігання та обробки великих даних. Частина 2.

Лекція 11. Архітектури зберігання та обробки великих даних. Частина 3.

Лекція 12. Архітектури зберігання та обробки великих даних. Частина 4.

Лабораторна робота 3. Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml.

Самостійна робота: Опрацювання матеріалу лекції. Підготовка до захисту лабораторної роботи.

Тема 6. Соціотехнічна архітектурна парадигма data mesh.

Стисла анотація: Основи data mesh. Сховища й озера даних усередині data mesh. Переваги data mesh. Принципи data mesh. Особливості соціотехнічної архітектури. Труднощі впровадження data mesh. Фактори прийняття рішення про реалізацію data mesh. Альтернативні і додаткові рішення для data mesh. Оцінка зусиль по реалізації data mesh.

Лекція 13. Соціотехнічна архітектурна парадигма data mesh. Частина 1.

Лекція 14. Соціотехнічна архітектурна парадигма data mesh. Частина 2.

Самостійна робота: Опрацювання матеріалу лекції.

Тема 7. Еталонна архітектура великих даних.

Стисла анотація: Загальні відомості про еталонну архітектуру великих даних. Концепція еталонної архітектури великих даних. Представлення користувача. Наскрізні аспекти. Функційні компоненти еталонної архітектури великих даних.

Лекція 15. Еталонна архітектура великих даних. Частина 1.

Лекція 16. Еталонна архітектура великих даних. Частина 2.

Лабораторна робота 4. Практичні приклади застосування алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml в авіації.

Самостійна робота: Опрацювання матеріалу лекції. Підготовка до захисту лабораторної роботи. Підготовка до модульного контролю.

Модульний контроль 2.

5. Індивідуальні завдання

Індивідуальні завдання не передбачені.

6. Методи навчання

Словесні (пояснення, розповідь, проблемний виклад), наочні (ілюстрування, демонстрація, презентація), лабораторні.

7. Методи контролю

Проведення поточного контролю під час проведення лабораторних занять, модульний контроль у вигляді тесту, підсумковий контроль у вигляді тесту, семестровий контроль у письмово-усній формі під час заліку.

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
Змістовий модуль 1			
Активність під час аудиторної роботи	0...1	8	0...8
Виконання і захист лабораторних робіт	0...9	2	0...18
Модульний контроль	0...20	1	0...20
Змістовий модуль 2			
Активність під час аудиторної роботи	0...1	8	0...8
Виконання і захист лабораторних робіт	0...12	2	0...24
Модульний контроль	0...20	1	0...22
Усього за семестр			0...100

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

Білет для заліку 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 та 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

Критерії оцінювання роботи здобувач освіти протягом семестру

Задовільно (60-74). Показати мінімум знань та умінь. Захистити не менше 80% від усіх завдань лабораторних занять. 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 не менше 90% завдань лабораторних занять. 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. Виконати усі завдання лабораторних занять в обумовлені викладачем строки. 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.

9. Політика навчального курсу

Відвідування занять. Регуляція пропусків. Характер курсу передбачає необхідність відвідування занять. Здобувачі освіти, які за певних обставин не можуть відвідувати 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 здобувані 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 або міркуваннями. 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 незалежно від масштабів плагіату чи обману.

Вирішення конфліктів. Порядок і процедури врегулювання конфліктів, пов'язаних із корупційними діями, зіткненням інтересів, різними формами дискримінації, сексуальними домаганнями, міжособистісними стосунками та іншими ситуаціями, що можуть виникнути під час навчання, а також правила етичної поведінки 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=1600>.

11. Рекомендована література

Базова

1. Олещенко Л. М. *Технології оброблення великих даних. Конспект лекцій*. Київ : КПІ ім. Ігоря Сікорського, 2021. 227 с.
2. Balusamy B., Abirami N., Kadry S., Gandomi A. *Big Data: Concepts, Technology, and Architecture*. USA, NJ, Hoboken : Wiley, 2021. 371 p.
3. Pop F., Neagu G. *Big Data Platforms and Applications: Case Studies, Methods, Techniques, and Performance Evaluation*. Switzerland, Cham : Springer, 2021. 300 p.
4. Perrin J.-G. *Spark in action*. USA, NY, Shelter Island : Manning Publications Co., 2020. 576 p.

Допоміжна

1. Ravi V., Cherukuri A. *Handbook of Big Data Analytics*. Vol. 1. Methodology. UK, London : The Institution of Engineering and Technology, 2021. 390 p.

2. 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.

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.

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>.