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 2021 р.

СИЛАБУС ВИБІРКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Великі дані для кібербезпеки

(назва навчальної дисципліни)

Галузь знань: 10 Природничі науки, 11 Математика та статистика, 12 Інформаційні технології, 15 Автоматизація та приладобудування, 16 Хімічна та біоінженерія, 17 Електроніка та телекомунікації, 19 Архітектура та будівництво

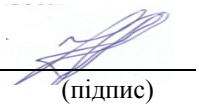
Спеціальність: 101 Екологія, 103 Науки про Землю, 113 Прикладна математика, 121 Інженерія програмного забезпечення, 122 Комп'ютерні науки, 123 Комп'ютерна інженерія, 124 Системний аналіз, 125 Кібербезпека, 151 Автоматизація та комп'ютерно-інтегровані технології, 152 Метрологія та інформаційно-вимірювальна техніка, 163 Біомедична інженерія, 172 Телекомунікації та радіотехніка, 173 Авіоніка, 193 Геодезія та землеустрій

Освітня програма: усі освітні програми за відповідними спеціальностями

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

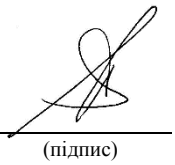
Силабус введено в дію з 01.09.2021 року

Харків – 2021 р.

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

Силабус навчальної дисципліни розглянуто на засіданні кафедри комп'ютерних систем, мереж і кібербезпеки (№ 503)

Протокол № 1 від « 30 » серпня 2021 р.

Завідувач кафедри д.т.н., професор  В. С. Харченко
(науковий ступінь та вчене звання) (підпис) (ініціали та прізвище)

Погоджено з представником здобувачів освіти:

_____ _____ _____
(підпис) (ініціали та прізвище)

1. Загальна інформація про викладача

Фесенко Герман Вікторович, д.т.н., доцент. З 2018 з року викладає в університеті наступні дисципліни:
Інженерія програмного забезпечення;
Технології обробки великих даних;
Великі дані для кібербезпеки;
Теорія і методи data science і штучного інтелекту.

Напрями наукових досліджень: великі дані, гарантоздатні інтелектуальні системи моніторингу критичних інфраструктур, Інтернет літаючих речей.

2. Опис навчальної дисципліни

Семестр, в якому викладається дисципліна – 1 семестр.

Обсяг дисципліни:

5 кредитів ЄКТС (150 годин), у тому числі аудиторних – 64 годин, самостійної роботи здобувачів – 86 годин.

Форми здобуття освіти

Денна, дистанційна, дуальна.

Дисципліна – вибіркова.

Види навчальної діяльності – лекції, лабораторні роботи, самостійна робота здобувача.

Види контролю – поточний, модульний та підсумковий (семестровий) контроль (іспит).

Мова викладання – українська.

Необхідні обов'язкові попередні дисципліни (пререквізити) – архітектура комп'ютерів; моделі та структури даних.

Необхідні обов'язкові супутні дисципліни (кореквізити) – дипломне проєктування.

3. Мета та завдання навчальної дисципліни

Мета

Вивчення навчальної дисципліни «Великі дані для кібербезпеки» полягає в формуванні системи знань, способів діяльності та творчих здібностей з основних теоретичних положень про обробку великих даних та платформи для

їх реалізації з урахуванням стратегій, політик та стандартів щодо регулювання безпеки кіберпростору, а також засвоєння вмінь які б дозволяли на практиці реалізувати ці знання.

Завдання

Вивчення інструментальних матеріалів та їх ефективного використання, фізичної сутності процесу різання, різних видів механічної обробки, різальних верстатів, конструкцій інструментів та основ їх конструювання, а також засвоєння методики розрахунку і призначення раціональних режимів різання.

Після опанування дисципліни здобувач набуде наступні **компетентності**:

– здатність застосовувати знання у практичних ситуаціях;

– здатність до абстрактного мислення, аналізу та синтезу;

– здатність обґрунтовано застосовувати, інтегрувати, розробляти та удосконалювати сучасні інформаційні технології, фізичні та математичні моделі, а також технології створення та використання прикладного і спеціалізованого програмного забезпечення для вирішення професійних задач у сфері інформаційної безпеки та/або кібербезпеки.

Очікується, що після опанування дисципліни здобувач будуть досягнуті наступні **результати навчання** і він буде:

– інтегрувати фундаментальні та спеціальні знання для розв’язування складних задач інформаційної безпеки та/або кібербезпеки у широких або мультидисциплінарних контекстах;

– застосовувати, інтегрувати, розробляти, впроваджувати та удосконалювати сучасні інформаційні технології, фізичні та математичні методи і моделі в сфері інформаційної безпеки та/або кібербезпеки;

– критично осмислювати проблеми інформаційної безпеки та/або кібербезпеки, у тому числі на міжгалузевому та міждисциплінарному рівні, зокрема на основі розуміння нових результатів інженерних і фізико-математичних наук, а також розвитку технологій створення та використання спеціалізованого програмного забезпечення.

4. Зміст навчальної дисципліни

Модуль 1.

Змістовний модуль 1. Базові поняття та основні технології обробки великих даних

Тема 1. Базові поняття великих даних.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 годин.*

- *Лабораторна робота: «Програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD) з використанням фреймворка Spark».*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*

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, слабко структуровані та метадані).

- *Обсяг самостійної роботи здобувачів: 11 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Загальні принципи аналізу даних за допомогою Spark. Особливості роботи зі стійкими розподіленими наборами даних (RDD).

Тема 2. Перехід до великих даних, питання планування та бізнес-аналітика.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 годин.*
- *Лабораторна робота: «Створення та робота з DataFrames з використанням Spark».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*

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. 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 та. бізнес-аналітика для великих даних.

- *Обсяг самостійної роботи здобувачів: 11 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Особливості роботи с парами «ключ-значення». Пакет для роботи зі структурованими даними Spark SQL.

Тема 3. Концепції зберігання та обробки великих даних.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 годин.*

- *Лабораторна робота: «Робота з графами з використанням Spark GraphX».*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*

Кластери. Файлові системи та розподілені файлові системи. NoSQL. Шардінг. Реплікація. Шардінг і реплікація. Теорема CAP. ACID. BASE. Паралельна обробка даних. Розподілена обробка даних. Hadoop. Обробка робочих завдань (пакетна обробка). Транзакційна обробка. Кластер. Обробка у пакетному режимі. 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.

- *Обсяг самостійної роботи здобувачів: 11 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Бібліотека для обробки графів та виконання паралельних обчислень Spark GraphX. Компонент для роботи з потоковими даними Spark Streaming.

Тема 4. Технології зберігання великих даних та основні методи їх аналізу.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*

- *Обсяг аудиторного навантаження: 8 годин.*

- *Лабораторна робота: «Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: статистичні обчислення і перевірка гіпотез».*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*

Дискові пристрої зберігання, бази даних NoSQL. Системи зберігання у оперативній пам'яті (In-memory Data Grid, наскрізне читання, наскрізний запис, відкладений запис, оновлення, In-memory Databases). 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.

- *Обсяг самостійної роботи здобувачів: 10 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Бібліотека Spark Mlib для реалізації машинного навчання.

Модульний контроль 1

- *Форма занять: виконання модульної роботи у вигляді тестування на комп'ютері в аудиторії (за рішенням лектора допускається проведення у дистанційній формі).*
- *Обсяг аудиторного навантаження: 1 година*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленими тестовими завданнями.*
- *Обсяг самостійної роботи здобувачів – за необхідністю.*

Підготовка до модульного контролю.

Змістовний модуль 2. Вирішення питань кібербезпеки під час роботи з великими даними

Тема 5. Основи забезпечення безпеки та конфіденційності великих даних.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 годин.*
- *Лабораторна робота: «Робота зі вбудованими функціями модуля Spark SQL».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*

Вплив безпеки і конфіденційності на характеристики великих даних. Хмарні обчислення й великі дані. Рівні безпеки та конфіденційності великих даних. Інтернет речей, кіберфізичні системи та великі дані. Мобільні пристрої та великі дані. Системний комунікатор для великих даних. Приклади забезпечення безпеки та конфіденційності великих даних в різних галузях: продажі та маркетинг, охорона здоров'я, захист комп'ютерних мереж, урядові організації, освіта, транспорт.

- *Обсяг самостійної роботи здобувачів: 11 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Рекомендації щодо забезпечення трьох рівнів безпеки та конфіденційності еталонної архітектури великих даних від National Institute of Standards and Technology (NIST), USA.

Тема 6. Таксономія аспектів безпеки та конфіденційності великих даних. Рівень безпеки та конфіденційності в еталонній архітектурі великих даних.

Аспекти конфіденційності великих даних. Аспекти походження великих даних. Забезпечення працездатності систем на основі великих даних. Аспекти безпеки та конфіденційності великих даних у сфері державної та соціальної

політики, взаємодії між організаціями. Реєстрація пристроїв та застосунків, що використовують великі дані. Ідентифікація та управління доступом. Управління великими даними. Управління інфраструктурою, що використовує великі дані. Криптографічні технології перетворення великих даних. Фази управління безпекою та конфіденційністю.

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 годин.*
- *Лабораторна робота: «Робота зі Spark Machine Learning Library: Transformers and Estimators».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*
- *Обсяг самостійної роботи здобувачів: 11 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Існуючі стандарти, що стосуються безпеки та конфіденційності великих даних.

Тема 7. Безпека комунікаційних послуг на основі великих даних

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 годин.*
- *Лабораторна робота: «Робота зі Spark Machine Learning Library: Supervised Learning (навчання з учителем)».*
- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*

Характеристики комунікаційних послуг на основі великих даних і категорії даних. Життєвий цикл даних комунікаційних послуг на основі великих даних: збір, передача, зберігання, використання даних, обмін даними, знищення даних. Уразливості безпеки протягом життєвого циклу даних комунікаційних послуг на основі великих даних. Вказівки щодо забезпечення безпеки при управлінні життєвим циклом даних для комунікаційних послуг на основі великих даних. Загрози та виклики великих даних як послуги. Високорівневі концепції великих даних як послуги. Заходи безпеки великих даних як послуги.

- *Обсяг самостійної роботи здобувачів: 11 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Загрози та проблеми безпеки даних моніторингу хмарних обчислень.

Тема 8. Проблеми безпеки хмарних обчислень

- *Форма занять: лекція, лабораторна робота, самостійна робота.*
- *Обсяг аудиторного навантаження: 8 годин.*

- *Лабораторна робота: «Робота зі Spark Machine Learning Library: Recommendation Engines».*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленим фреймворком Apache Spark .*

Загрози безпеці для споживачів і постачальників хмарних обчислень. Проблеми безпеки хмарних обчислень. Можливості забезпечення безпеки хмарних обчислень. Загрози та проблеми безпеки мережі як послуги в середовищі хмарних обчислень. Загрози та проблеми безпеки застосунків мережі як послуги в середовищі хмарних обчислень.

- *Обсяг самостійної роботи здобувачів: 10 годин.*

Опрацювання матеріалу лекцій. Формування питань до викладача. Вимоги до безпеки даних моніторингу хмарних обчислень.

Модульний контроль 2

- *Форма занять: виконання модульної роботи у вигляді тестування на комп'ютері в аудиторії (за рішенням лектора допускається проведення у дистанційній формі).*

- *Обсяг аудиторного навантаження: 1 година*

- *Обов'язкові предмети та засоби (обладнання, устаткування, матеріали, інструменти): комп'ютер зі встановленими тестовими завданнями.*

- *Обсяг самостійної роботи здобувачів – за необхідністю.*

Підготовка до модульного контролю.

5. Індивідуальні завдання

Не передбачено навчальним планом

6. Методи навчання

Словесні, наочні, практичні.

7. Методи контролю

Поточний контроль (теоретичне опитування й розв'язання практичних завдань), модульний контроль (тестування за розділами курсу) та підсумковий (семестровий) контроль (іспит).

8. Критерії оцінювання та розподіл балів, які отримують здобувачі

Складові навчальної роботи	Бали за одне заняття (завдання)	Кількість занять (завдань)	Сумарна кількість балів
Змістовний модуль 1			
Робота на лекціях	0…2	8	0…16
Виконання і захист лабораторних робіт	0...4	4	0…16
Модульний контроль	0…18	1	0…18
Змістовний модуль 2			
Робота на лекціях	0…2	8	0…16
Виконання і захист лабораторних робіт	0...4	4	0…16
Модульний контроль	0…18	1	0…18
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 – 4 бали;

– виконані всі завдання лабораторної роботи, мають місце незначні відхилення від оформлення відповідно до вимог, надано правильні відповіді на всі запитання – 3 бали;

– виконано 75-99% завдань лабораторної роботи, мають місце незначні відхилення від оформлення відповідно до вимог, надано правильні відповіді не на всі запитання – 2 бали;

– виконано 50-74% 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 18 балів) виносяться всі пройдені за контрольований період теми, які включаються у варіанти тестових завдань, що містять по 18 питань. 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.

Прийнята шкала оцінювання

Сума балів за всі види навчальної діяльності	Оцінка для екзамену, *курсового проекту* (роботи), практики
90-100	відмінно
75-89	добре
60-74	задовільно
01-59	незадовільно з можливістю повторного складання

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 програмувати операції зі стійкими розподіленими наборами даних та DataFrames з використанням Spark, вміти забезпечувати безпеку та конфіденційність великих даних для заданої архітектури.

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, аналізу, забезпечення безпеки та конфіденційності 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. Вміти удосконалювати існуючий рівень безпеки та конфіденційності архітектури великих даних при зміні вимог. Вміти забезпечувати безпеку великих даних при управлінні їх життєвим циклом для комунікаційних послуг.

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, а також для забезпечення безпеки та конфіденційності 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. Вміти удосконалювати існуючий рівень безпеки та конфіденційності архітектури великих даних при зміні вимог. Вміти забезпечувати безпеку великих даних при управлінні їх життєвим циклом для комунікаційних послуг. Розуміти загрози та вирішувати проблеми безпеки хмарних обчислень. Безпомилково виконувати та захищати всі лабораторні роботи в обумовлений викладачем строк з докладним обґрунтуванням рішень та заходів, які запропоновано у роботах.

9. Політика навчального курсу

Відпрацювання пропущених занять відбувається відповідно до розкладу консультацій, за попереднім погодженням з викладачем. Питання, що стосуються академічної доброчесності, розглядає викладач або за процедурою, визначеною у Положенні про академічну доброчесність.

10. Методичне забезпечення та інформаційні ресурси

Дистанційний курс в системі дистанційного навчання Mentor, розташований за адресою: .

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.
3. Mastering Big Data Analytics. URL:https://www.mygreatlearning.com/academy/learn-for-free/courses/mastering-big-data-analytics.

11. 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.