



СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ТЕХНОЛОГІЇ BIGDATA

Галузь знань	12 Інформаційні технології
Шифр та назва спеціальності	123 Комп'ютерна інженерія
Назва освітньо-професійної програми	Комп'ютерні мережі та Інтернет
Рівень вищої освіти	Перший (бакалаврський)
Факультет	Інформаційних технологій та кібербезпеки
Кафедра	Інформаційних та комп'ютерних систем
Статус навчальної дисципліни	ОК-25 ОПП «Комп'ютерні мережі та Інтернет»
Форма навчання	Денна

Викладач

Тихонова Олена Вікторівна
elena.tykhonova@suitt.edu.ua



Старший викладач кафедри Інформаційних та комп'ютерних систем,
кандидат технічних наук

Загальна інформація про дисципліну

Анотація до дисципліни	Дисципліна спрямована на формування у здобувачів теоретичних основ і практичних уявлень про концепцію Big Data, еволюцію систем керування даними та сучасні підходи до зберігання й аналітичної обробки великих обсягів даних. У курсі розглядаються архітектури сховищ даних і Data Lake, підходи ETL/ELT та методології проєктування аналітичних систем. Вивчаються обчислювальні моделі та інструменти екосистеми Hadoop, а також сучасні фреймворки розподіленої обробки даних. Курс охоплює базові методи інтелектуального аналізу даних, включно з кластеризацією, класифікацією, пошуком асоціативних правил та застосуванням нейронних мереж.
Мета дисципліни	Формування у здобувачів системного розуміння основних концепцій, архітектурних рішень та технологічних інструментів, що використовуються для зберігання, обробки та аналізу великих даних.

Компетентності, формуванню яких сприяє дисципліна	ЗК-3. Здатність застосовувати знання у практичних ситуаціях. СК-7. 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. СК-13. Здатність вирішувати проблеми у галузі комп'ютерних та інформаційних технологій, визначати обмеження цих технологій.
Результати навчання	ПРН-2. Мати навички проведення експериментів, збирання даних та моделювання в комп'ютерних системах. ПРН-3. Знати новітні технології в галузі комп'ютерної інженерії. ПРН-15. Вміти виконувати експериментальні дослідження за професійною тематикою. ПРН-20. Усвідомлювати необхідність навчання впродовж усього життя з метою поглиблення набутих та здобуття нових фахових знань, удосконалення креативного мислення.
Обсяг дисципліни	Загальний обсяг дисципліни: 5 кредитів ЄКТС (150 годин). Для денної форми навчання: лекції –28 годин, практичні заняття –14 годин, лабораторні заняття –14 годин, самостійна робота – 94 години. Для заочної форми навчання: лекції – 10 годин, практичні заняття –4 години, лабораторні заняття – 4 години, самостійна робота –132 години.
Форма підсумкового контролю	Екзамен
Терміни викладання дисципліни	Дисципліна викладається у 6-му семестрі

Програма дисципліни

Тема 1.	Основні положення концепції Big Data. Визначення понять великих даних та причини їх появи. Сфери застосування Big Data та переваги їх використання для бізнесу. Порівняння «традиційних» даних та «великих» даних. Джерела надходження великих даних. Описова модель великих даних 3V: Volume (об'єм), Velocity (швидкість), Variety (різноманітність).
Тема 2.	Еволюція систем керування базами даних: від реляційних моделей до Big Data, NoSQL та аналітичних систем OLAP. Проблеми реляційних баз даних у контексті Big Data. Концепція NoSQL (NOSQL). Переваги та недоліки NOSQL підходу у порівнянні з реляційними базами даних. CAP-теорема. Основні види нереляційних баз даних (NOSQL БД) - документоорієнтовані, бази даних «ключ-значення», графові, колонкові. Бази даних NewSQL. Принципи онлайн-обробки транзакцій (OLTP) та аналітичної онлайн-обробки даних (OLAP). Багатовимірна модель даних. Гіперкуб даних, зріз гіперкубу. Представлення гіперкубу за допомогою схем «Зірка» та «Сніжинка».
Тема 3.	Архітектури сховищ даних та сучасні підходи до управління корпоративними даними: від DWH до Data Lakehouse. Поняття Сховища Даних (Data Warehouse, DWH). Базова тришарова архітектура екосистеми DWH (шари джерел даних, сховища даних, користувачів). Вітрини даних. Додавання прошарків проміжних даних та вітрин даних до базової архітектури DWH. Процеси ETL та ELT. Традиційні підходи до організації сховища даних (Ральфа Кімбала, Біла Інмона). Сучасні методології побудови

	сховища даних (Data Vault, Anchor Modelling). Поняття Озера Даних (Data Lake). Особливості озер даних у порівнянні з DWH. Data lakehouse - гібридне рішення для зберігання даних.
Тема 4.	Обчислювальна парадигма MapReduce та екосистема Apache Hadoop: основи обробки великих даних. Обчислювальна парадигма MapReduce, фази відображення (Map) і згортки (Reduce), проміжний етап Shuffle. Переваги та недоліки моделі MapReduce. Актуальність технології MapReduce на поточний момент. Еволюція ролі MapReduce. Проект Apache Hadoop - відкрита реалізація концепції MapReduce. Основні компоненти Hadoop. HDFS - розподілена файлова система Hadoop. Концепція блоків в HDFS. Архітектура кластерної системи Hadoop 1.0.
Тема 5.	Еволюція та сучасні фреймворки екосистеми Hadoop: від MapReduce та YARN до Apache Spark та Hive. Еволюція систем розподілених обчислень Hadoop. Обмеження Hadoop MapReduce v1. Версія Hadoop MapReduce 2.0, менеджер ресурсів YARN. Оновлення третьої версії Hadoop. Фреймворк Apache Spark для розподіленої обробки даних, його особливості. Компоненти Apache Spark. Apache Hive. Сучасне використання Apache Hive та Apache Spark.
Тема 6.	Основи інтелектуального аналізу даних (Data Mining): методологія CRISP-DM та практичні задачі. Інтелектуальний аналіз даних. Три класи задач аналізу: інформаційно-пошуковий, оперативно-аналітичний, інтелектуальний (Data Mining). Життєвий цикл аналітики даних. Методологія дослідження даних CRISP-DM. Основні етапи CRISP-DM: розуміння цілей і вимог проекту, розуміння даних, підготовка та попередня обробка даних, моделювання, оцінка, розгортання. Основні задачі Data Mining (описові, прогнозуючі). Дослідження науковців ДУІТЗ у сфері обробки великих даних, зокрема великих текстових масивів.
Тема 7.	Описові задачі Data Mining. Актуальність задачі пошуку асоціативних правил. Алгоритм Apriori для пошуку асоціативних правил. Ієрархічні та неієрархічні методи кластеризації. Англомеративний алгоритм кластеризації. Метод k-середніх для вирішення задач кластеризації. Правило «ліктя» визначення кількості кластерів для методу k-середніх.
Тема 8.	Прогнозуючі задачі Data Mining. Методи класифікації: наївний Байєсовський алгоритм, метод k-найближчих сусідів. Метрики для обчислення відстані між елементами даних. Нормалізація даних при обчисленні відстані між елементами. Метод опорних векторів та дерева рішень. Показник «забруднення» вузла дерева, інформаційний приріст способу розбиття, вибір найкращого розбиття вузла дерева рішень. Метод найменших квадратів вирішення задачі лінійної регресії.
Тема 9.	Використання штучних нейронних мереж для вирішення задач інтелектуального аналізу даних. Поняття «штучний інтелект», «машинне навчання», «нейронні мережі». Основні компоненти нейронної мережі (вхідний, вихідний, прихований шари; ваги та зміщення; функція активації). Стратегія ініціалізації вагових коефіцієнтів та зміщень. Алгоритми нейронних мереж за принципом навчання: зі вчителем, без вчителя, з підкріпленням. Класифікація даних за допомогою нейронної мережі «зі вчителем». Кластеризація даних. Алгоритм нейронної мережі, що навчається «без вчителя». Автоенкодер. Нейронні мережі «з підкріпленням» в задачах аналізу даних. Процес навчання нейронної мережі; поняття функції втрат, градієнтного спуску та зворотного поширення помилки. Гіперпараметри алгоритмів нейронних мереж. Переваги нейронних мереж для вирішення задач аналізу даних.

Список рекомендованих джерел

1. Junxiu An, Sian Jin. Cloud Computing and Big Data Technology, 2nd edition. World Scientific, 2025. 376 p.
2. Joe Reis, Matt Housley. Fundamentals of Data Engineering. O'Reilly, 2022. 450 p.
3. Brij B Gupta, Mamta. Big Data Management and Analytics. World Scientific, 2024. 288 p.
4. Методичні вказівки до практичних занять з дисципліни «Технології Big Data» / уклад.: В.І. Тіхонов, О.В. Тихонова. Одеса : ДУІТЗ, 2025. 84 с.
5. Методичні рекомендації до лабораторних робіт з дисципліни «Технології Big Data» / уклад.: В.І. Тіхонов, О.В. Тихонова. Одеса : ДУІТЗ, 2025. 84 с.
6. An introduction to big data. URL : <https://opensource.com/resources/big-data> (дата звернення 20.08.24).
7. Big Data Resource. URL : <https://www.sciencedirect.com/topics/computer-science/big-data-resource> (дата звернення 20.08.24).
8. Best Resources to learn Big Data. URL : <https://www.analyticsvidhya.com/resources-big-data/> (дата звернення 20.08.24).

Інформація про консультації

Згідно визначеного розкладу: ауд. 402 або онлайн за посиланням

<https://us04web.zoom.us/j/3185149804?pwd=TmUybHZZYzBRK2dleUQrNVhPaG1wdz09>

Загальна схема оцінювання

Сума балів за всі види навчальної діяльності	Шкала ЄКТС	Оцінка за національною шкалою		Нарахування балів	Бали нараховуються таким чином: <i>Оцінювання знань здобувачів вищої освіти здійснюється за 100-бальною шкалою і становить: за поточну успішність (участь у практичних заняттях, виконання практичних завдань, лабораторних та контрольних робіт) та за результати заліку/екзамену)</i>
		для іспиту	для заліку		
90-100	A	Відмінно	зараховано		
82-89	B	Добре			
74-81	C				
64-73	D				
60-63	E	Задовільно			
35-59	FX	Незадовільно з можливістю повторного складання	Не зараховано з можливістю повторного складання		

0-34	F	Незадовільно з обов'язковим повторним вивченням дисципліни	Не зараховано з обов'язковим повторним вивченням дисципліни		
------	---	--	---	--	--

Політика опанування дисципліни

Відвідування занять: відвідування здобувачами навчальних занять є обов'язковим, запізнення на заняття на 15 хвилин і більше не допускається.

При проведенні занять в онлайн-режимі присутність здобувача зараховується у разі включення ним камери та/або мікрофона.

Умови зарахування пропущених занять: зарахування пропущених практичних/лабораторних занять здійснюється за умови виконання та захисту відповідних завдань. До екзамену допускаються здобувачі, які виконали практичні та лабораторні завдання. Здобувач, який не з'явився на екзамен або не був допущений на момент його проведення, має право повторно його пройти у визначений викладачем термін.

Дотримання принципів академічної доброчесності: Підготовка усіх завдань, письмових робіт і т. ін., що виконуються в межах дисципліни, здійснюється здобувачем вищої освіти самостійно, на засадах академічної доброчесності. У разі порушення здобувачем принципів академічної доброчесності робота оцінюється незадовільно та має бути виконана повторно.