



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	Комп'ютерної інженерії та інформаційних систем
Статус навчальної дисципліни	ОК-25 ОПІ «Комп'ютерні мережі та Інтернет»
Форма навчання	Денна

Викладач

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



Старший викладач кафедри Комп'ютерної інженерії та інформаційних систем, кандидат технічних наук

Загальна інформація про дисципліну

Анотація до дисципліни	Програма вивчення навчальної дисципліни «Технології BigData» складена відповідно до освітньо-професійної програми підготовки бакалаврів напряму (спеціальності) 123 Комп'ютерна інженерія. Предметом вивчення навчальної дисципліни є технології, в основі яких лежать принципи отримання, перетворення, розподіленого зберігання та обробки, а також аналізу Великих Даних.
-------------------------------	--

Мета дисципліни	Формування у студентів базових умінь та компетенцій щодо методів та інформаційних технологій збирання, зберігання та обробки даних, розмір яких перевищує можливості звичайних програмних аналітичних платформ та баз даних по аналізу, зберіганню і управлінню інформацією.
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.
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. Приклади можливостей 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.	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.
Тема 3.	Принципи онлайн-обробки транзакцій (OLTP) та аналітичної онлайн-обробки даних (OLAP). Багатовимірний гіперкуб даних, зріз гіперкубу. Представлення гіперкубу за допомогою схем «Зірка» та «Сніжинка».

Тема 4.	Поняття Сховища Даних (Data Warehouse, DWH). Базова тришарова архітектура DWH (джерел даних, сховища даних, користувачів). Вітрини даних. Процеси ETL та ELT. Додавання прошарків проміжних даних та вітрин даних до базової архітектури DWH. Традиційні підходи до організації сховища даних (Ральфа Кімбала, Біла Інмона). Поняття Озера Даних (Data Lake). Особливості озер даних у порівнянні з DWH. Data lakehouse - гібридне рішення для зберігання даних.
Тема 5.	Обчислювальна парадигма MapReduce. Переваги та недоліки моделі MapReduce. Проект Apache Hadoop - відкрита реалізація MapReduce. Основні компоненти Hadoop. HDFS - розподілена файлова система Hadoop. Архітектура кластерної системи Hadoop 1.0. Версія Hadoop MapReduce 2.0, менеджер ресурсів YARN. Фреймворк Apache Spark для розподіленої обробки даних. Компоненти Apache Spark. Apache Hive.
Тема 6.	Основи аналізу Великих Даних. Три класи задач аналізу: інформаційно-пошуковий, оперативно-аналітичний, інтелектуальний (Data Mining). Методологія дослідження даних CRISP-DM. Основні задачі Data Mining: описові (кластеризація, пошук асоціативних правил), прогнозуючі (класифікація, регресія).
Тема 7.	Методи вирішення задач Data Mining. Ієрархічний англомеративний та метод k-середніх для вирішення задач кластеризації. Методи класифікації: Наївний Байєсовський алгоритм, метод k-найближчих сусідів, метод опорних векторів, дерева рішень. Метод найменших квадратів вирішення задачі лінійної регресії. Алгоритм Apriori для пошуку асоціативних правил.

Список рекомендованих джерел

1. Albert Y. Zomaya, Sherif Sakr (editors). Handbook of Big Data Technologies. Springer, 2017. 890 p.
2. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020, 298 p. Режим доступу: <https://vdoc.pub/download/big-data-conceptual-analysis-and-applications-40crds1qhta0>
3. Thomas Erl, Wajid Khattak, and Paul Buhler. Big Data Fundamentals: Concepts, Drives & Techniques. Prentice Hall, 2016. 235 p.
4. Ghavami P. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. CreateSpace Independent Publishing Platform, 2016. 204 p.
5. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. 110 p.
6. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019. 105 p.

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		Н а р а х у в а н н я б а л і в	Бали нараховуються таким чином: <i>Оцінювання знань здобувачів вищої освіти здійснюється за 100-бальною шкалою</i> і становить: за поточну успішність (участь у практичних заняттях, виконання практичних завдань та контрольних робіт) – до 90 балів, за результати екзамену – до 10 балів.
		для іспиту	для заліку		
90-100	A	Відмінно	зараховано		
82-89	B	Добре			
74-81	C				
64-73	D	Задовільно			
60-63	E				
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.

Інші вимоги: Загальна оцінка з дисципліни – максимум 100 балів. У випадку отримання менше 60 балів, здобувач обов'язково здійснює перескладання для ліквідації академічної заборгованості.