



СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ХМАРНІ СХОВИЩА І ТЕХНОЛОГІЇ BIG DATA

Галузь знань	12 Інформаційні технології
Шифр та назва спеціальності	126 Інформаційні системи та технології
Назва освітньо-професійної програми	Інформаційні системи в економіці та бізнесі
Рівень вищої освіти	Перший (бакалаврський)
Факультет	Інформаційних технологій та кібербезпеки
Кафедра	Комп'ютерної інженерії та інформаційних систем
Статус навчальної дисципліни	ОК-25 ОПП «Інформаційні системи в економіці та бізнесі»
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	<p>Програма вивчення навчальної дисципліни «Хмарні сховища і технології Big Data» складена відповідно до освітньо-професійної програми підготовки бакалаврів напряму (спеціальності) 126 Інформаційні системи та технології.</p> <p>Предметом вивчення навчальної дисципліни є технології, в основі яких лежать принципи отримання, перетворення, розподіленого зберігання та обробки, а також аналізу великих обсягів даних. Навчання спрямоване на отримання студентами навичок практичного використання типових технологій та засобів аналітики даних, таких як MapReduce, Hadoop, додаткових бібліотек Python (NumPy, Pandas, Matplotlib).</p> <p>Міждисциплінарні зв'язки: забезпечується вивченням дисциплін – «Мережі та системи інфокомунікацій», «Технології IP-мереж», «Системи керування базами даних», «Технології програмування на мові Python».</p>
-------------------------------	---

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-1. Здатність до абстрактного мислення, аналізу та синтезу. ЗК-3. Здатність до розуміння предметної області та професійної діяльності. СК-2. Здатність застосовувати стандарти в області інформаційних систем та технологій при розробці функціональних профілів, побудові та інтеграції систем, продуктів, сервісів і елементів інфраструктури організації. СК-6. Здатність використовувати сучасні інформаційні системи та технології (виробничі, підтримки прийняття рішень, інтелектуального аналізу даних та інші), методики й техніки кібербезпеки під час виконання функціональних завдань та обов'язків. СК-10. Здатність вибору, проектування, розгортання, інтегрування, управління, адміністрування та супроводження інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації. СК-12. Здатність управляти та користуватися сучасними інформаційно-комунікаційними системами та технологіями (у тому числі такими, що базуються на використанні Інтернет). СК-15. Здатність використовувати технології віртуалізації та хмарних обчислень під час проектування, розробки, реконфігурації та реконструкції інформаційних систем.
Результати навчання	ПРН-7. Обґрунтовувати вибір технічної структури та розробляти відповідне програмне забезпечення, що входить до складу інформаційних систем та технологій. ПРН-13. Знати сучасні методи і технології проектування баз даних та знань та вміти застосовувати їх на практиці. ПРН-15. Знати методи захисту інформації, моделі безпеки інформаційних систем, використовувати ці знання при створенні безпечних інформаційних систем. ПРН-17. Застосовувати інформаційні технології та засоби для створення ІТ інфраструктури та її компонентів, вміти здійснювати їх технічне обслуговування.
Обсяг дисципліни	Загальний обсяг дисципліни: 5 кредитів ЄКТС (150 годин). Для денної форми навчання: лекції –24 години, практичні заняття –16 годин, лабораторні заняття –16 годин, самостійна робота – 94 години.
Форма підсумкового контролю	Екзамен
Терміни викладання дисципліни	Дисципліна викладається у 1-му семестрі 4 курсу (1–18 тижні)

Програма дисципліни

Тема 1.	<i>Вступ до Big Data.</i> Передумови виникнення великих даних. Сфери застосування та переваги використання для бізнесу. Основні положення концепції Big Data. Джерела надходження великих даних. Порівняння «традиційних» даних та «великих» даних. Ступінь структурованості даних.
----------------	---

V-характеристики великих даних.

Реляційні бази даних, мова структурованих запитів SQL. ACID - набір вимог до транзакційної системи, що забезпечує її надійну та передбачувану роботу (Atomicity - атомарність, Consistency - узгодженість, Isolation - ізольованість, Durability - надійність). Проблеми реляційних баз даних. Нереляційні бази даних, концепція NoSQL (NOSQL). Переваги та недоліки NOSQL підходу у порівнянні з реляційними базами даних. Використання розподіленої архітектури у NOSQL базах даних. Теорема Брюера, або CAP-теорема (consistency - узгодженість, availability - доступність, та partition tolerance - стійкість до розділення). Основні види нереляційних баз даних (NOSQL БД) - документоорієнтовані, бази даних «ключ-значення», графові, колонкові; приклади систем управління (СУБД) нереляційними базами даних. Бази даних NewSQL. Принципи онлайн-обробки транзакцій (OLTP) та аналітичної онлайн-обробки даних (OLAP).

Тема 2. *Технології зберігання і обробки великих даних.*

Сховище даних (Data Warehouse, DWH). Відмінності сховищ даних від баз даних. Вітрина даних. Компоненти сховища даних: інструменти інтеграції даних, центральна БД, метадані, інструменти доступу. Розповсюджені архітектури DWH: Ральфа Кімбала, Біла Інмона, EAV-архітектура. Процес ETL (Extract - витягти, Transform - перетворити, Load - завантажити). Переваги та недоліки сховищ даних.

Озеро даних (Data Lake); відмінність озер даних від баз даних та сховищ даних. Переваги та недоліки зберігання даних в Data Lake. Data lakehouse - гібридне рішення для зберігання даних.

Розподілене зберігання та обробка великих даних. Обчислювальна парадигма MapReduce, фази відображення (Map) і згортки (Reduce); приклади використання. Компоненти MapReduce. Переваги та недоліки моделі MapReduce. Еволюція систем розподілених обчислень Hadoop. Проект Apache Hadoop - відкрита реалізація MapReduce. Основні компоненти Hadoop. HDFS - розподілена файлова система Hadoop. Архітектура кластерної системи Hadoop. Завантаження даних в Hadoop за принципом «схема-на-читання».

Хмарні сервіси, локальне та хмарне зберігання великих даних. Приватне, публічне та гібридне облако. Особливості використання, переваги та недоліки хмарного сховища. Структура хмарних систем зберігання даних. Способи зберігання даних у хмарі: блоковий, об'єктний, файловий. Популярні хмарні сервіси для роботи з Big Data: Amazon Web Services, Microsoft Azure Data Lake, Google BigQuery та інші.

Інформаційна безпека у сфері Big Data. Специфіка захисту великих даних. Загрози та вразливості, підходи щодо забезпечення безпеки Big Data. Світовий ринок засобів захисту великих даних.

Тема 3. *Основи аналізу великих даних.*

Три класи задач аналізу: інформаційно-пошуковий, оперативно-аналітичний, інтелектуальний (Data Mining). Життєвий цикл аналітики даних, методологія CRISP-DM (Cross-Industry Standard Process for Data Mining). Основні етапи CRISP-DM: розуміння цілей і вимог проекту, розуміння даних, підготовка та попередня обробка даних, моделювання, оцінка, розгортання.

Основні задачі Data Mining:

- описові (пошук частих закономірностей, асоціацій, аналіз та виявлення відхилень, кластеризація)
- прогнозуючі (класифікація, регресія)

Алгоритми кластеризації. Регресійний аналіз. Алгоритми пошуку асоціативних правил. Статистичний аналіз. Застосування машинного навчання у аналітиці.

Програмні інструменти Big Data. Популярні аналітичні платформи: Power BI, Tableau, Information is Beautiful, Storytelling with Data, Python Graph Gallery. Аналіз даних засобами мови Python. Вбудовані структури даних Python: списки, кортежі, словники, множини.

Бібліотека NumPy, багатовимірні масиви даних, векторизація обчислень. Бібліотека Pandas, серії та датафрейми, групування і агрегація даних, об'єднання датафреймів. Попередня перевірка набору даних, методи info() та describe(), описова статистика. Візуалізація даних та результатів аналізу. Бібліотеки Matplotlib та Seaborn.

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. Ghavami P. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. CreateSpace Independent Publishing Platform, 2016, 204 p.
4. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. 110 p.
5. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019. 105 p.
6. Thomas Erl, Wajid Khattak, and Paul Buhler. Big Data Fundamentals: Concepts, Drives & Techniques. Prentice Hall, 2016. 235 p.
7. Sh. Sakr, A.Y.Zomaya. Encyclopaedia of Big Data Technologies. Springer, 2019. 1853 p.
8. Підручник з великих даних для початківців: <https://uk.myservername.com/big-data-tutorial-beginners-what-is-big-data>.
9. Сергєєв-Горчинський О. О., Іщенко Г. В. Інтелектуальний аналіз даних. Комп'ютерний практикум. КПІ ім. Ігоря Сікорського, 2018. https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf.
10. Data Science With Python Tutorial : <https://www.geeksforgeeks.org/data-science-tutorial/>.
11. Beginner's Guide to Python : <https://wiki.python.org/moin/BeginnersGuide>.
12. NumPy tutorials : <https://numpy.org/numpy-tutorials/index.html>.
13. Pandas User Guide : https://pandas.pydata.org/docs/user_guide/index.html.
14. Matplotlib Tutorials : <https://matplotlib.org/stable/tutorials/index#introductory>.
15. Seaborn User guide and tutorial : <https://seaborn.pydata.org/tutorial.html>.

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:
		для іспиту	для заліку		
90-100	A	Відмінно	зараховано		<p>Оцінювання знань здобувачів вищої освіти здійснюється за 100-бальною шкалою і становить: за поточну успішність і самостійну роботу – до 80 балів, за результати заліку – до 20 балів.</p> <p>При оформленні документів за екзаменаційну сесію використовується таблиця відповідності оцінювання знань здобувачів вищої освіти за різними системами.</p>
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 підсумкового заліку допускаються здобувачі, які виконали практичні та лабораторні завдання. Здобувач, який не з'явився на залік 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.