



## Силабус курсу Інструментальні засоби обробки даних

Освітньо-професійна програма: «Комп'ютерні науки»  
Ступінь вищої освіти - бакалавр  
Спеціальність 122 «Комп'ютерні науки»

Рік навчання: III, Семестр: VI

Кредитів: 4 Мова викладання: українська

### Керівник курсу

ППП

викладач Владислав Пойдич

Контактна інформація

vladyslav.poidych@wunu.edu.ua

### Опис дисципліни

Мета дисципліни «Інструментальні засоби обробки даних» полягає в підготовці кваліфікованих фахівців, здатних ефективно використовувати сучасні інструменти та технології для збору, обробки, аналізу та візуалізації даних різних типів і розмірів. Студенти отримають компетенції, необхідні для роботи з великими масивами даних, їх оптимізації та трансформації з метою підтримки прийняття обґрунтованих рішень.

Завдання навчальної дисципліни спрямоване на вивчення фундаментальні концепції та принципів, які лежать в основі обробки і аналізу даних, з оглядом на сучасні методи та підходи, розвиток практичних навичок в області використання сучасних інструментів мов програмування платформ і технологій для ефективною обробки даних, вивчення технологій роботи з великими даними та робота з сучасними методами та засобами аналітики даних.

### Структура курсу

Години (лек./лаб.)	Тема	Результати навчання	Завдання
2/-	Тема 1. Вступ в аналіз і обробку даних.	Розуміти поняття даних і концепцію роботи над ними. Проектування структури.	Питання
2/4	Тема 2. Запити за даними, фільтрування.	Розуміти та використовувати запити направлені на отримання даних і фільтрування даних.	Питання, лабораторна робота
4/4	Тема 3. Об'єднання таблиць: теорія та практика. Особливості типів даних у системах баз даних.	Знати типи об'єднань таблиць та задачі які вони вирішують. Володіти базовими типами даних.	Питання, лабораторна робота
2/2	Тема 4. Маніпуляція даними, групування елементів.	Знати техніки маніпуляції даними, групування елементів та сортування.	Питання, лабораторна робота
6/6	Тема 5. Транзакції та підзапити. Ефективність дотримання цілісності бази даних.	Знати типи транзакцій, розуміти як вони виконуються і які проблеми вирішують, розуміти де потрібно використовувати кожен з типів.	Питання, лабораторна робота
4/4	Тема 6. Обмеження для забезпечення цілісності та консистентності	Знати які обмеження використовуються для забезпечення цілісності і консистентності даних.	Питання, лабораторна робота

	збережених даних.		
2/-	Тема 7. Бази даних та сховища даних.	Розуміти різницю між базою даних та сховищем даних.	Питання
2/-	Тема 8. Розподілені платформи.	Знати розподілені платформи та сфери застосування.	Питання, лабораторна робота
4/4	Тема 9. Інструментальні засоби.	Знати інструментальні засоби які використовуються для обробки даних.	Питання, лабораторна робота
4/4	Тема 10. Платформи обробки даних.	Знати платформи які надають можливості роботи з великими даними.	Питання, лабораторна робота
4/4	Тема 11. Методи запуску циклічних задач по оброці даних Apache Airflow.	Розуміти концепцію Airflow та циклічних задач, виконання таких задач у власному середовищі.	Питання, лабораторна робота
2/4	Тема 12. Методи перетворення даних у процесі ETL та ELT, DBT.	Розуміти концепцію DBT та процесу перетворення. Виконання таких процесів у власному середовищі.	Питання, лабораторна робота
4/4	Тема 13. Засоби візуалізації чистих даних: PowerBI, Looker.	Розуміти коцепцію PowerBI та процесу візуалізації. Виконання таких процесів у власному середовищі.	Питання, лабораторна робота
4/4	Тема 14. PowerBI запити, підключення до зовнішніх джерел даних, перетворення та особливості агрегування даних.	Розуміти концепцію звязків у PowerBI та для чого вони використовуються. Розуміти роботу з зовнішніми джерелами даних.	Питання, лабораторна робота

### Літературні джерела

1. Data Engineering with dbt: A practical guide to building a cloud-based, pragmatic, and dependable data platform with SQL 1st Edition, Kindle Edition, Roberto Zagni, 2023.
2. Data Pipelines with Apache Airflow, Bas P. Harenslak, 2021.
3. Microsoft Power BI Dashboards Step by Step, Errin O'Connor, 2018.
4. PostgreSQL: Up and Running: A Practical Guide to the Advanced Open Source Database, Regina Obe, Leo Hsu, 2017.
5. Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter, 3rd Edition, Wes McKinney, 2022.
6. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. - Springer, 2020. – 298 p.
7. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. - Springer, 2019. – 105 p.
8. Bühlmann P., Drineas P., Kane M., van der Laan M. (eds.) Handbook of Big Data. -Taylor and Francis, 2016. – 456 p.
9. Oussous A., Benjelloun F.-Z., Lahcen A.A. and Belfkih S. Big Data technologies: A survey. // Journal of King Saud University. Computer and Information Sciences. - 2018. - Vol. 30, Issue 4. - P. 431–448.
10. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. – 110 p.
11. Raheem N. Big Data: A Tutorial-Based Approach. - Taylor & Francis Group LLC, CRC Press, 2019. – 203 p.

12. Комар М.П. «Методологічні основи інформаційної технології інтелектуального аналізу та обробки великих даних». Дисертація на здобуття наукового ступеня доктора технічних наук, 2021. – 363 с.
13. Костюченко А.О. Основи програмування мовою Python: навчальний посібник. - Ч.: ФОП Баликіна С.М., 2020. - 180 с.

### Політика оцінювання

**Політика щодо дедлайнів та перескладання:** Перескладання модулів відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

**Політика щодо академічної доброчесності:** Списування під час контрольних робіт та екзаменів заборонені (в т.ч. із використанням мобільних пристроїв). Мобільні пристрої дозволяється використовувати лише під час он-лайн тестування.

**Політика щодо відвідування:** Відвідування занять є обов'язковим компонентом оцінювання. За об'єктивних причин (наприклад, хвороба, міжнародне стажування) навчання може відбуватись в он-лайн формі за погодженням із керівником курсу.

### Оцінювання

Модуль 1		Модуль 2		Модуль 3	Модуль 4	Модуль 5
10%	10%	10%	10%	5%	15%	40%
Поточне оцінювання виставляється як середнє арифметичне за лабораторні роботи та опитування на заняттях.	Модульний контроль проводиться за допомогою тестового контролю по пройдених темах у системі Moodle 20 тестів по 2 бали  Дві задачі по 15 балів  Теоретичне питання 30 балів	Поточне оцінювання виставляється як середнє арифметичне за лабораторні роботи та опитування на заняттях.	Модульний контроль проводиться за допомогою тестового контролю по пройдених темах у системі Moodle  10 тестів по 2 бали  Дві задачі по 20 балів  Два Теоретичних питання по 20 балів	Оцінювання тренінгу проводиться на основі виконання завдання комплексного проекту тренінгу.	Поточне оцінювання виставляється як середнє арифметичне з оцінок отриманих під час самостійної роботи.	Екзамен містить 50 білетів. Структура білета відповідає наступному:  Два теоретичних питання по 20 балів.  Дві задачі по 30 балів.

### Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90-100	відмінно	A (відмінно)
85-89	добре	B (дуже добре)
75-84		C (добре)
65-74	задовільно	D (задовільно)
60-64		E (достатньо)
35-59	незадовільно	FX (незадовільно з можливістю повторного складання)
1-34		F (незадовільно з обов'язковим повторним курсом)