

КЛАСИЧНИЙ ПРИВАТНИЙ УНІВЕРСИТЕТ
КАФЕДРА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ

СИЛАБУС

навчальної дисципліни
«АНАЛІТИЧНІ СХОВИЩА ДАНИХ ТА
ХМАРНІ ТЕХНОЛОГІЇ»

КОНТАКТНА ІНФОРМАЦІЯ ТА ТЕХНІЧНОЇ ДОПОМОГА
(включаючи електронну пошту, робочий час / місцезнаходження тощо).

Викладач (-і)	Бречко Наталія Анатоліївна
Контактний тел.	+38(061) 228-07-85 (внутр. 224)
E-mail:	brechko.na@gmail.com
Сторінка курсу на сайті підтримки навчальних програм КПУ	http://www.zhu.edu.ua/cpu_edu/course/view.php?id=5247
Консультації	<i>Очні консультації:</i> за графіком консультацій викладача, а. 311, головний корпус КПУ <i>Консультації off-line:</i> шляхом повідомлення на сторінці навчальної дисципліни сайту підтримки навчальних програм КПУ http://www.zhu.edu.ua/cpu_edu/message

АНОТАЦІЯ

Навчальна дисципліна «Аналітичні сховища даних та хмарні технології» є нормативною для студентів першого (бакалаврського) рівня вищої освіти галузі знань 11 Математика та статистика, спеціальності 112 Статистика, освітньої програми Комп'ютерна статистика та аналіз даних. Згідно з навчальним планом денної форми навчання вивчення дисципліни заплановано на 8 семестр (4 курс).

Метою дисципліни є отримання студентами базових знань та практичних навичок проектування структури даних для аналізу, використання хмарних сервісів для обробки Big Data та побудови звітності для ефективного вирішення завдань у сфері Data Science та Big Data. Для засвоєння програми дисципліни необхідні знання, вміння та компетенції з дисциплін "Інформатика та теорія алгоритмів", "Бази даних та SQL", "Java для Data Science".

Освітній процес з дисципліни здійснюється за такими формами: навчальні заняття; самостійна робота; контрольні заходи. Видами навчальних занять згідно з навчальним планом є: лекції; лабораторні заняття, а також консультації.

Лабораторні заняття передбачають: отримання студентами базових знань і практичних навичок використання технологій та проектування структури даних для аналізу, використання

хмарних сервісів для обробки Big Data для ефективного вирішення задач у сфері Data Science та Big Data.

Самостійна робота студентів полягає у засвоєнні вивченого навчального матеріалу в час, вільний від обов'язкових навчальних занять, без участі викладача.

Консультації призначені для роз'яснення студентам теоретичних або практичних питань.

Засвоєння навчального матеріалу перевіряється за допомогою поточного контролю, який здійснюється на лабораторних заняттях у формі усних відповідей, виконання робіт та підсумкових тематичних тестових завдань.

Підсумковий (семестровий) контроль після завершення 8 семестру здійснюється у формі іспиту.

ФОРМАТ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Загальна кількість годин – 120 год., у т. ч. 48 годин аудиторних занять і 72 години самостійної роботи студента. Кількість кредитів ECTS – 4.

Всього кредитів	Всього годин	Аудиторних годин	У тому числі			Сам. робота
			Лекц.	Лабор.	Семін. (практ.)	
4	120	48	20	28	-	72

ОЗНАКИ ДИСЦИПЛІНИ

Навчальний рік	Курс (рік навчання)	Семестр	Цикл підготовки	Нормативна/вибіркова
2025/2026	4	8	професійна	нормативна

МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета навчальної дисципліни

Метою дисципліни є отримання студентами базових знань та практичних навичок проектування структури даних для аналізу, використання хмарних сервісів для обробки Big Data та побудови звітності для ефективного вирішення завдань у сфері Data Science та Big Data.

Завдання навчальної дисципліни

- формування чіткого уявлення про відмінності між операційними (OLTP) та аналітичними (OLAP) системами.;
- проектування схем даних (зокрема, схеми «Зірка» та «Сніжинка»), які спрощують агрегацію та побудову часових рядів, необхідних для статистичних прогнозів;
- надання знань про сучасні хмарні платформи (AWS, GCP, Azure), моделі обслуговування (IaaS, PaaS, SaaS) та принципи тарифікації хмарних ресурсів;
- використання SQL не просто для вибірки рядків, а як інструмент первинного статистичного аналізу;
- розвинути здатність виявляти аномалії, пропущені значення та помилки в даних на етапі їх потрапляння в сховище, щоб гарантувати валідність подальших статистичних висновків.

ЗАПЛАНОВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ

У результаті вивчення дисципліни студенти повинні знати:

- ✧ різницю між транзакційними системами (OLTP) та аналітичними системами (OLAP), а також їх призначення у статистичному аналізі;
- ✧ принципи побудови багатовимірних моделей даних (Multidimensional Data Modeling), особливості схем «Зірка» (Star schema) та «Сніжинка» (Snowflake schema);
- ✧ поняття нормалізації та денормалізації даних;
- ✧ різницю між традиційним сховищем (Data Warehouse), озером даних (Data Lake) та сучасною концепцією Data Lakehouse;
- ✧ сутність та сфери застосування хмарних моделей обслуговування: IaaS (інфраструктура), PaaS (платформа), SaaS (програмне забезпечення);
- ✧ поняття вертикального та горизонтального масштабування обчислювальних ресурсів;
- ✧ критерії якості даних (точність, повнота, узгодженість, актуальність) та методи їх перевірки перед статистичним моделюванням;
- ✧ правила побудови ефективних дашбордів, вибір правильного типу візуалізації залежно від типу змінних (категоріальні, кількісні) та статистичної задачі.

Після вивчення дисципліни студенти повинні вміти:

- ✧ проектувати схеми даних «Зірка» та «Сніжинка» на основі бізнес-вимог, визначати таблиці фактів та вимірів для конкретних статистичних задач;
- ✧ обґрунтовано вибирати технологічний стек (тип бази даних, хмарного провайдера) залежно від обсягу даних та бюджету проекту;
- ✧ створювати та налаштовувати облікові записи у хмарних консолях (AWS/GCP/Azure), керувати правами доступу користувачів (IAM);
- ✧ писати складні SQL-запити з використанням віконних функцій (OVER, PARTITION BY, RANK, LEAD/LAG) для розрахунку динаміки, ковзних середніх та накопичувальних підсумків;
- ✧ налаштовувати хмарні сховища (наприклад, S3 бакети), завантажувати в них файли великих обсягів та керувати життєвим циклом об'єктів;
- ✧ виконувати запити до даних у форматах JSON/Array безпосередньо в базі даних (наприклад, розгортати масиви в рядки);
- ✧ налаштовувати автоматизовані процеси завантаження даних із зовнішніх джерел (CSV, API) у сховище даних;
- ✧ програмно виявляти та обробляти пропущені значення, дублікати та аномалії засобами SQL або Python (в рамках хмарних функцій) перед початком статистичного аналізу;
- ✧ розробляти зрозумілі та інформативні дашборди, налаштовувати фільтри, drill-down (деталізацію) та перехресну фільтрацію.

Відповідно до освітньо-професійної програми підготовки магістра галузі знань 11 Математика та статистика, спеціальності 112 Статистика, освітньої програми Комп'ютерна статистика та аналіз даних сприяє формуванню компетентностей та програмних результатів навчання :

Інтегральна компетентність:

Здатність розв'язувати складні спеціалізовані математичні та статистичні задачі, що характеризується комплексністю і невизначеністю умов і передбачає застосування теоретико-ймовірнісних і статистичних методів, комп'ютерних технологій.

Загальні компетентності:

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК 2. Здатність застосовувати знання у практичних ситуаціях.

ЗК 6. Навички використання інформаційних і комунікаційних технологій.

ЗК 7. Здатність вчитися і оволодівати сучасними знаннями.

ЗК 8. Здатність до пошуку, обробки та аналізу інформації з різних джерел.

Професійні (фахові) компетенції

СК 4. Здатність до математичного формулювання задач та вибору методів їх розв'язання.

СК 7. Здатність робити якісні висновки з кількісних даних.

СК 8. Уміння працювати з інформаційними базами даних.

СК 11. Здатність використання обчислювальної техніки, спеціалізованих мов програмування та програмних засобів для розв'язання задач і здобуття додаткової інформації.

СК 14. Здатність до аналізу основ і властивостей статистичних алгоритмів та розуміння переваг і обмежень тих чи інших підходів, у тому числі до оцінки їх обґрунтованості й ефективності.

Програмні результати навчання:

РН 12. Вміти збирати та обробляти дані, застосовувати статистичні процедури для аналізу даних за допомогою обчислювальної техніки та програмних засобів.

РН 13. Вміти моделювати реалізації випадкових величин і процесів та використовувати результати моделювання для верифікації й аналізування ефективності статистичних процедур.

РН 14. Володіти сучасними інформаційними технологіями для створення презентацій, роботи з базами даних, пошуку інформації та обміну нею.

РН 15. Володіти математичними та статистичними методами аналізу, прогнозування та оцінки параметрів математичних моделей, статистичними методами інтерпретації та обробки числових даних.

РН 16. Вміти використовувати в практичній діяльності спеціалізоване статистичне програмне забезпечення.

ПЛАН КУРСУ

Назва змістових модулів та тем	Лекц.	Лабор.	Завдання для самостійної роботи
Тема № 1. “Концепції сховищ даних та багатовимірне моделювання”.	2	4	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: ▲ вивчити теоретичний матеріал залікового модуля; ▲ виконати та захистити індивідуальний варіант самостійної роботи; ▲ здійснити перевірку отриманих знань шляхом комп'ютерного тестування.
Тема №2. “Основи хмарних обчислень та інфраструктура”.	4	2	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань:

Назва змістових модулів та тем	Лекц.	Лабор.	Завдання для самостійної роботи
			✦ вивчити теоретичний матеріал залікового модуля; ✦ виконати та захистити індивідуальний варіант самостійної роботи; ✦ здійснити перевірку отриманих знань шляхом комп'ютерного тестування.
Тема №3. “Розширений SQL для статистичного аналізу”.	4	4	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: ✦ вивчити теоретичний матеріал залікового модуля; ✦ виконати та захистити індивідуальний варіант самостійної роботи; ✦ здійснити перевірку отриманих знань шляхом комп'ютерного тестування.
Тема №4. “Архітектура сучасних хмарних DWH (Columnar & MPP)”	4	6	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: ✦ вивчити теоретичний матеріал залікового модуля; ✦ виконати та захистити індивідуальний варіант самостійної роботи; ✦ здійснити перевірку отриманих знань шляхом комп'ютерного тестування.
Тема №5. “Інтеграція даних: стратегії ETL/ELT та якість даних”.	2	6	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: ✦ вивчити теоретичний матеріал залікового модуля; ✦ виконати та захистити індивідуальний варіант самостійної роботи; ✦ здійснити перевірку отриманих знань шляхом комп'ютерного тестування.
Тема №6. “Візуалізація та презентація даних”	4	6	1. Опрацювання лекційного матеріалу, 2. Самостійне опрацювання теоретичних питань: ✦ вивчити теоретичний матеріал залікового модуля; ✦ виконати та захистити індивідуальний варіант самостійної роботи; ✦ здійснити перевірку отриманих знань шляхом комп'ютерного тестування.

ФОРМИ КОНТРОЛЮ ТА КРИТЕРІЇ ОЦІНЮВАННЯ

У процесі вивчення навчальної дисципліни «Аналітичні сховища даних та хмарні технології» використовуються наступні види контролю:

1. Поточний контроль – здійснюється протягом семестру шляхом опитування на семінарських (практичних) заняттях, перевірки виконання тестових завдань, виконання практичних робіт, модульних контрольних робіт тощо. За змістом він включає перевірку ступеню засвоєння студентом навчального матеріалу, який охоплюється темою лекційного та семінарського заняття, уміння самостійно опрацьовувати навчально-методичну літературу, здатність осмислювати зміст теми, уміння публічно та письмово представити певний матеріал, а також виконання завдань самостійної роботи.

2. Підсумковий семестровий контроль – здійснюється у формі іспиту відповідно до графіка освітнього процесу.

Для оцінювання студентів використовується система накопичування балів. Згідно з «Положенням про організацію освітнього процесу в КПУ» підсумкова оцінка з дисципліни

виставляється за 100-бальною шкалою з наступним переведенням у національну шкалу та шкалу ECTS.

Бали нараховуються за виконання практичних завдань, лабораторних робіт, контрольних (модульних) завдань, тестів.

Результати поточного модульного контролю здобувачів вищої освіти є складовими елементами підсумкової оцінки з дисципліни.

Оцінка рівня роботи студента протягом семестру під час навчальних занять та самостійної роботи здійснюється у межах 60 балів. Вага екзамену у підсумковій оцінці складає 40 балів.

РОЗПОДІЛ БАЛІВ ПОТОЧНОГО ТА ПІДСУМКОВОГО (СЕМЕСТРОВОГО) ОЦІНЮВАННЯ

Поточне оцінювання під час навчальних занять та самостійної роботи						Екзамен	Сума
Змістовий модуль 1			Змістовий модуль 2				
ПМ1	ПМ2	ПМ3	ПМ4	ПМ5	ПМ6	40	100
10	10	10	10	10	10		

Шкала оцінювання: 100-бальна, національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка за шкалою ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90-100	A	відмінно	зараховано
82-89	B	добре	
75-81	C		
67-74	D	задовільно	
60-66	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

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Kimball, R., & Ross, M. (2013). *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (3rd ed.). Wiley. - 600 p
2. Tanimura, C. (2021). *SQL for Data Analysis: Advanced Techniques for Transforming Data into Insights*. O'Reilly Media. - 360 p.
3. Берко, А. Ю., & Верес, О. М. (2013). *Організація сховищ даних та знань: підручник*. Львів: Видавництво Львівської політехніки. -340 с.
4. Knafllic, C. N. (2015). *Storytelling with Data: A Data Visualization Guide for Business Professionals*. Wiley. - 288 p.
5. Reis, J., & Housley, M. (2022). *Fundamentals of Data Engineering*. O'Reilly Media. - 436 p.

Додаткові матеріали та ресурси

1. Kleppmann, M. (2017). *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*. O'Reilly Media. - 616 p.
2. Inmon, W. H. (2005). *Building the Data Warehouse* (4th ed.). Wiley. - 543 p.
3. Lakshmanan, V., & Tigani, J. (2019). *Google BigQuery: The Definitive Guide*. O'Reilly Media. - 480 p.
4. Molinaro, A. (2020). *SQL Cookbook: Query Solutions and Techniques for Database Developers* (2nd ed.). O'Reilly Media. - 554 p.
5. Wexler, S., Shaffer, J., & Cotgreave, A. (2017). *The Big Book of Dashboards: Visualizing Your Data Using Real-World Business Scenarios*. Wiley. - 448 p.
6. Глибовець, М. М. (2005). *Основи баз даних*. Київ: КМ-Академія. - 270 с

12. ІНФОРМАЦІЙНІ РЕСУРСИ В ІНТЕРНЕТІ

1. Документація Google Cloud BigQuery. URL:
<https://cloud.google.com/bigquery/docs>.
2. Документація Snowflake. URL: <https://docs.snowflake.com>.
3. Kimball Group Archive (статті з моделювання даних). URL:
<https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/>
4. Kaggle Datasets (репозиторій наборів даних для лабораторних). URL:
<https://www.kaggle.com/datasets>