



**ДЕРЖАВНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ**
Факультет інформаційних технологій
Кафедра комп'ютерних наук та інформаційних систем

СИЛАБУС (SYLLABUS)
**Дисципліна «Технології створення розподілених
баз даних та знань/Technologies for creating distributed databases and
knowledge»**

ІНФОРМАЦІЯ ПРО ВИКЛАДАЧА

Викладач	Самойленко Ганна Тимофіївна
Науковий ступінь	Кандидат фізико-математичних наук
Вчене звання	Доцент
Посада	Доцент кафедри комп'ютерних наук та інформаційних систем
Адреса кафедри	м.Київ, вул. Кіото 19, каб. Б-507, Б-526
E-mail	compdep@knute.edu.ua
Консультації	Відповідно до графіку індивідуальних консультацій на сайті кафедри

ПОЛІТИКА АКАДЕМІЧНОЇ ДОБРОЧЕСНОСТІ

<https://knute.edu.ua/file/MzEyMQ==/c12a9f74e87d9154696ca0f761da2e5c.pdf>

Дотримання академічної доброчесності передбачає:

- самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей);
- посилання на джерела інформації у разі використання не авторських ідей, розробок, тверджень, відомостей і т.п.;
- дотримання норм законодавства про авторське право і суміжні права;
- надання достовірної інформації про результати власної наукової діяльності, використанні методики досліджень і джерела інформації.

Порушенням академічної доброчесності вважається:

- академічний плагіат – оприлюднення (частково або повністю) наукових (творчих) результатів, отриманих іншими особами, як результатів власного дослідження (творчості) та/або відтворення опублікованих текстів (оприлюднених творів мистецтва) інших авторів без зазначення авторства;
- самоплагіат – оприлюднення (частково або повністю) власних раніше опублікованих наукових результатів як нових наукових результатів;
- фабрикація – вигадкування даних чи фактів, що використовуються в наукових дослідженнях;
- фальсифікація – свідомо зміна чи модифікація вже наявних даних, що стосуються наукових досліджень.

За порушення академічної доброчесності здобувачі освіти можуть бути притягнені до академічної відповідальності:

- повторне проходження оцінювання (модульний контроль, іспит, залік тощо);
- повторне проходження відповідного освітнього компонента освітньо-професійної програми;
- відрахування з Університету;

- позбавлення наданих університетом пільг;
- відмова у присудженні відповідного ступеня вищої освіти;

ПОЛІТИКА ЩОДО ВІДВІДУВАННЯ ЗАНЯТЬ

- відвідування занять є обов'язковим;
- Студент, який пропустив практичне заняття, самостійно вивчає матеріал (при виникненні питань може звертатися за консультацією згідно розкладу консультацій викладачів оприлюдненого на сайті кафедри) за наведеними джерелами, виконує завдання і здає його викладачу.
- за об'єктивних причин (наприклад, хвороба, міжнародне стажування та ін.) навчання може відбутись в он-лайн формі за погодженням із викладачем дисципліни.

ПОЛОЖЕННЯ ПРО АПЕЛЯЦІЮ РЕЗУЛЬТАТІВ ЕКЗАМЕНІВ У ДТЕУ

<https://knute.edu.ua/file/MjkwNQ==/cf2f392763bdb0447eed3c254854ec5.pdf>

ВРЕГУЛЮВАННЯ КОНФЛІКТНИХ СИТУАЦІЙ

Учасники освітнього процесу повинні дотримуватися принципів гідності, взаємоповаги, толерантності, доброчесності. Адміністрація ДТЕУ забезпечує попередження, запобігання, своєчасне виявлення та врегулювання конфліктних ситуацій, пов'язаних із цькуванням, дискримінацією, сексуальними домаганнями (див. Положення про врегулювання конфліктних ситуацій ДТЕУ (<https://knute.edu.ua/file/MjkwMjQ=/b91ca19cb0c629d8b9938ba46ccc41f5.pdf>)).

УЗГОДЖЕННЯ ЗМІСТУ ДИСЦИПЛІНИ З ЦІЛЯМИ СТАЛОГО РОЗВИТКУ (ЦСР)

Зміст дисципліни «Технології створення розподілених баз даних та знань» узгоджується з кількома ключовими Цілями сталого розвитку (ЦСР). Дисципліна вивчає, як технології створення розподілених баз даних та знань можуть бути використані для аналізу та управління економічними процесами, забезпечуючи сталий розвиток. Студенти досліджуватимуть, як розподілені бази даних і знань застосовувати для зберігання і обробки даних для побудови сталої економіки. Наступні Цілі сталого розвитку (ЦСР) мають особливе відношення до тем, що розглядаються навчальною дисципліною:

1. **ЦСР 4: Забезпечення всеохоплюючої і справедливої якісної освіти та заохочення можливості навчання впродовж усього життя для всіх.** У рамках дисципліни студенти вивчатимуть як розподілені бази даних можуть бути використані для створення розподілених освітніх платформ та баз даних для зберігання освітніх ресурсів, сприяючи інклюзивності та можливостям навчання протягом усього життя.
2. **ЦСР 9: Створення стійкої інфраструктури, сприяння всеохоплюючій і сталій індустріалізації та інноваціям.** Особлива увага приділена вивченню розподілених інформаційних систем для виробничих процесів, безпеці розподілених баз даних та знань, розвитку інфраструктурних об'єктів, що забезпечує впровадження і використання сучасних розподілених інформаційних технологій як основи процесів індустріалізації.

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назва дисципліни / тип дисципліни	Технології створення розподілених баз даних та знань / обов'язкова
Навчальний рік	2024/2025
Факультет	Факультет інформаційних технологій
Курс	I
Семестр	II
Освітній ступінь	магістр
Галузь знань	12 «Інформаційні технології»

Спеціальність	122 «Комп'ютерні науки»
Освітня програма	Комп'ютерні науки
Загальна характеристика	Кількість годин –180 Кількість кредитів – 6 Співвідношення аудиторних годин і годин самостійної роботи - 80/100 Мова викладання – українська Форма викладання – очна
Підсумковий контроль	Екзамен
Програмне забезпечення	SQL, Google Cloud
Обладнання	Проектор, комп'ютерна техніка із встановленим програмним забезпеченням та доступом до мережі Інтернет.
Необхідні попередні дисципліни	«Алгоритмізація та програмування», «Організація баз даних та знань», «Проектування інформаційних систем»
Методика вивчення	Методика вивчення дисципліни полягає у набутті студентами знань теоретичного і практично-прикладного характеру під час лекцій, лабораторних занять, самостійної роботи та вивчення першоджерел і навчально-методичної літератури.
Мета і завдання	Метою вивчення дисципліни «Технології створення розподілених баз даних та знань» є надання поглиблених знань та практичних навичок щодо теорії та практики побудови й використання баз даних. Завданням вивчення дисципліни «Технології створення розподілених баз даних та знань» є засвоєння методів створення розподілених баз даних та знань, технологій їх проектування, наповнення та підтримання в робочому стані.
Місце дисципліни в освітньо-професійній програмі	
Загальні компетентності	ЗК 2 Здатність до абстрактного мислення, аналізу та синтезу. ЗК 5 Здатність вчитися й оволодівати сучасними знаннями.
Фахові компетентності	СК 07 Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень. СК 09 Здатність розробляти та адмініструвати бази даних та знань. СК 011 Здатність ініціювати, планувати та реалізовувати процеси розробки інформаційних та комп'ютерних систем та програмного забезпечення, включно з його розробкою, аналізом, тестуванням, системною інтеграцією, впровадженням і супроводом.
Програмні результати навчання	РН 2 Мати спеціалізовані уміння/навички розв'язання проблем комп'ютерних наук, необхідні для проведення досліджень та/або провадження інноваційної діяльності з метою розвитку нових знань та процедур. РН 12 Проектувати та супроводжувати бази даних та знань.
	Ця дисципліна вивчає, як технології створення розподілених баз даних та знань можуть бути використані для аналізу та управління економічними процесами, забезпечуючи сталий розвиток. Студенти досліджуватимуть, як розподілені бази даних і знань застосовувати для зберігання і обробки даних для побудови сталої економіки. Наступні Цілі сталого розвитку (ЦСР) мають особливе відношення до тем, що розглядаються навчальною дисципліною: ЦСР 4: Забезпечення всеохоплюючої і справедливої якісної освіти та заохочення можливості навчання впродовж усього життя для всіх У рамках дисципліни студенти вивчатимуть, як розподілені бази даних

	<p>можуть бути використані для створення розподілених освітніх платформ та баз даних для зберігання освітніх ресурсів, сприяючи інклюзивності та можливостям навчання протягом усього життя.</p> <p>ЦСР 9: Промисловість, інновації та інфраструктура</p> <p>Студенти досліджуватимуть, як розподілені бази даних і знань можуть оптимізувати процеси в промисловості та інфраструктурі. Особлива увага буде приділена розподіленим інформаційним системам для виробничих процесів та розвитку інфраструктурних об'єктів, які відповідають викликам сучасного світу.</p>
--	--

ТЕМАТИКА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Тема 1. Розподілені бази даних.

Основні принципи, покладені в основу теорії баз даних. Поняття розподіленої бази даних (Distributed DataBase - DDB). Основна задача систем управління розподіленими базами даних. Система управління розподіленими базами даних. Однорідні та неоднорідні розподілені бази даних. Принцип К.Дж. Дейта. Властивості розподілених баз даних згідно К.Дж. Дейту: Локальна автономія. Незалежність від центрального вузла. Безперервні операції. Прозорість розміщення. Прозора фрагментація Прозорість тиражування. Обробка розподілених запитів. Незалежність від обладнання. Незалежність від операційних систем. Прозорість мережі. Незалежність від баз даних. Переваги і недоліки розподілених СУБД (РСУБД).

Тема 2. Архітектура програмно-технічних засобів розподілених СУБД.

Властивості архітектури. Розподіленість. Неоднорідність. Автономність. Два типи РСУБД залежно від типу програмного забезпечення: однорідні та неоднорідні. Різновиди архітектури. Клієнт-серверна архітектура. Архітектура з багатьма незалежними серверами. Архітектура із взаємодіючими серверами. Архітектура однорангової мережі. Розподілене зберігання даних. Два основні механізми розподіленого зберігання даних - фрагментація; реплікація. Поділ розподілених архітектур баз даних по типам: системи недублюючого розбиття (при великому об'ємі даних, що часто змінюються), системи частикового дублювання (при невеликому об'ємі даних, що часто змінюються), системи повного дублювання при невеликому об'ємі даних, що рідко змінюються).

Тема 3. Основні проблеми створення розподілених баз даних

Основні проблеми створення розподілених баз даних: Фрагментація даних і розподілення по комп'ютерам. Створення глобального каталогу з інформацією про кожний фрагмент БД та його місце в мережі. (Каталог може бути на одному вузлі або бути розподіленим). Організація обробки запитів (синхронізація кількох запитів до одних і тих же даних, виключення аномалій видалення і оновлення одних і тих же даних, що знаходяться на різних вузлах, оптимізація послідовності кроків по обробці запитів і т.д.).

Тема 4. Процеси побудови розподілених баз даних.

Принципи побудови розподілених баз даних. Мінімізація інтенсивності обміну даними. Оптимальне розміщення серверних та клієнтських додатків у мережі. Декомпозиція даних на сегменти, що використовуються часто и рідко (для правильного настроювання реплікації – розміщення даних, що часто використовуються на АРМ кінцевих користувачів). Збереження копій даних і виконання дій по підтримці цілісності розподіленої ІС. Паралельні процеси. Методи побудови розподілених баз даних «зверху вниз» і «знизу нагору». Розподілені обчислення. Двомірне подання даних кінцевому користувачеві. Багатомірне подання при описі структур даних. Розклад транзакцій Транзакції, властивості транзакцій.

Тема 5. Обробка розподілених запитів. Засоби SQL для розподілених систем.

Загальна характеристика мови. Реляційна алгебра як основа SQL. Поділення мови SQL на чотири частини: оператори визначення даних (англ. Data Definition Language, DDL) оператори маніпуляції даними (англ. Data Manipulation Language, DML) оператори визначення доступу до даних (англ. Data Control Language, DCL) оператори управління транзакціями (англ. Transaction Control Language, TCL). Засоби пошуку даних. Засоби маніпулювання даними. Операції над схемою бази даних. Віртуальні таблиці та індекси. Транзакції. Обробка розподілених запитів. Переваги та недоліки SQL.

Тема 6. Розподіл даних.

Централізоване розміщення даних. Роздільне (фрагментоване) розміщення даних. Розміщення з повною реплікацією. Розміщення з вибірковою реплікацією. Горизонтальна фрагментація. Вертикальна фрагментація. Змішана фрагментація. Методи поділу відносин на фрагменти й розподіл фрагментів по вузлах. Оптимальний розподіл фрагментів за вузлами мережі. Додаткова інформація, яка стосується розподілу фрагментів за вузлами мережі.

Тема 7. Реплікація.

Визначення. Переваги. Механізми реплікації. Сервери: видавець, дистриб'ютор і передплатник. Два методи відновлення даних передплатників: реплікація за запитом та примусова реплікація. Моделі реплікації. Реплікація моментальних знімків. Реплікація транзакцій. Топологія реплікацій. Функції служби реплікації. Схеми володіння даними. Збереження цілісності транзакцій. Моментальні знімки таблиць. Тригери баз даних. Виявлення та розв'язання конфліктів.

Тема 8. Проектування розподілених реляційних баз даних.

Критерії побудови розподілених баз даних: Всебічний аналіз інформаційних потреб предметної області з виявленням об'ємів x даних, що зберігаються, їх складності, достовірності, взаємозв'язку. Моделювання мережевого трафіку при роботі розподілених баз даних з різними моделями реплікації даних. Кластеризація елементів даних і програм їх обробки з метою - добитися максимальної автономності та слабкої пов'язаності кластерів. Прив'язка кластерів даних до ймовірносних користувачів чи АРМ. Підтримка еталонної копії даних і обмеження реплікаційного механізму. Розробка і реалізація правил приведення локальних і центральної бази даних в несуперечливий стан. Спеціалізовані програмні модулі для створення розподілених баз даних від фірм-розробників програмного забезпечення ГС. Модуль ArcSDE (Spatial Database Engine) фірми ESRI, який дозволяє зовнішнім користувачам працювати з просторовими базами даних. Модуль ArcGIS, що дозволяє працювати із зовнішніми базами даних. Модуль SDE, який забезпечує сучасні функції створення розподілених баз даних, підтримки різних моделей просторових даних).

Тема 9. Організація баз знань

Загальна характеристика баз знань. Базові поняття. Продукційна модель знань. Семантична модель знань. Моделі подання знань. Інтелектуальний аналіз даних. Склад і структура баз знань. Засоби побудови та опрацювання баз знань. Поняття сховищ даних та їх місце в системах оброблення даних. Поняття віртуального, корпоративного, глобального сховища даних та кіоска (вітрини) даних. Поняття залежного та незалежного кіоска (вітрини) даних. Архітектура сховищ даних. Технологія ETL (extract, transform and load) та характеристика її основних фаз. Віртуалізація сховищ. Характеристика основних ознак сховищ даних: предметна орієнтація, інтегрованість, підтримка хронології, незмінність і мінімальна надлишковість. Вимоги до вибору моделей побудови сховищ даних. Характеристика просторової моделі побудови сховища даних і її різновиди моделі „зірка” та „сніжинка”. Характеристики основних елементів багатовимірного представлення даних

Тема 10. Технологія проектування сховищ даних

Оперативна OLTP та аналітична OLAP обробка даних. OLTP (OnLine Transaction Processing): використання, вимоги, переваги і недоліки. OLAP (OnLine Analytical Processing): дія OLAP, правила для систем OLAP. Концепція сховища даних. Основні характеристики сховищ даних. Модель та 12 правил Кодда, яким повинен задовольняти програмний продукт класу OLAP. Тест FASMI. Властивості інформаційних сховищ. Основні елементи і операції OLAP. Типи OLAP. Багатомірний OLAP (MOLAP). Реляційний OLAP (ROLAP) Гібридні OLAP (HOLAP) Основні компоненти інформаційного сховища. Робота з гіперкубом. Реалізація сховищ і вітрин даних. Проблеми інтеграції даних. Неоднорідність програмного середовища. Розподілений характер організації. Підвищення вимог до безпеки даних. Необхідність наявності багаторівневих довідників метаданих. Наявні програмні реалізації. Архітектура OLAP-додатків. Реалізація серверів. Основи побудови банків даних БнД. Підходи до проектування сховищ даних. Автоматизація проектування сховищ даних.

Тема 11. Забезпечення безпеки РСУБД.

Поняття адміністрування баз та сховищ даних. Основні функції адміністратора БД та сховищ даних. Поняття інформаційної безпеки баз і сховищ даних. Типові загрози безпеки та їх запобігання. Захист даних – функція адміністратора. Рівні захисту. Привілеї та їх визначення. Прозорість розподілу. Прозорість фрагментації. Прозорість розташування. Прозорість реплікації. Прозорість локального відображення. Прозорість іменування. Прозорість транзакцій. Прозорість паралельності. Прозорість відмов. Прозорість виконання. Прозорість використання СУБД. Засоби забезпечення конфіденційності баз даних. Ідентифікація та перевірка справжності користувачів. Засоби ідентифікації і аутентифікації об'єктів баз даних. Обліковий запис. Режими аутентифікації. Організація взаємодії СУБД і базової ОС. Управління ключами безпеки. Введення в мережевий обмін даних. Дискреційна модель управління доступом. Основні категорії користувачів. Використання схем для забезпечення безпеки. Поділ користувачів і схем. Види привілеїв. Рольова модель розмежування доступу. Концепція і реалізація механізму ролей. Мандатне управління доступом. Мітки конфіденційності. Рівні конфіденційності об'єктів і рівні довіри суб'єктів доступу. Примусовий контроль доступу.

Тема 12. Засоби контролю цілісності інформації

Загрози безпеці розподілених систем. Розподілені транзакції. Методи розподілу даних. Технологія тиражування даних. Тиражування даних (Data Replication - DR). Реплікатор - спеціальний модуль СУБД - сервер тиражування даних. Переваги та недоліки DR-технології. Цілісність даних. Необхідність відновлення. Транзакції й відновлення. Функції відновлення: механізм резервного копіювання, файл журналу. Створення контрольних вузлів. Методи відновлення. Відновлення з використанням відкладеного відновлення. Відновлення з використанням негайного відновлення. Метод тіншового сторінкового обміну. Засоби забезпечення цілісності даних. Засоби контролю цілісності інформації. Засоби забезпечення доступності баз даних. Аудит системи безпеки бази даних.

Тема 13. Безпека розподілених інформаційних систем

Основні поняття безпеки розподілених інформаційних систем. Криптографічний захист інформації. Захист інформації в каналах зв'язку. Інформаційна безпека в мережах. Захист від несанкціонованого доступу. Принципи організації безпеки комп'ютерних мереж. Методи та засоби забезпечення вимог політики безпеки комп'ютерної мережі. Захист приватної мережі. Етапи проектування комплексної системи захисту розподіленої інформаційної системи.

Тема 14. Інформаційна безпека підприємств і організацій

Модель багаторівневого захисту розподілених інформаційних систем підприємств і організацій. Засоби забезпечення безпеки інформаційних систем. Безпека інформації на об'єктах. Канали витоку інформації. Системи інформаційної безпеки.

**Перелік навчальних робіт студентів та оцінки їх у балах з дисципліни
«Технології створення розподілених баз даних та знань»**

Види робіт	К-сть балів
Лабораторне заняття №1. Тема: «Розподілені бази даних»	2
Лабораторне заняття №2. Тема: «Архітектура програмно-технічних засобів розподілених СУБД»	2
Лабораторне заняття №3. Тема: «Основні проблеми створення розподілених баз даних»	2
Лабораторне заняття №4. Тема: «Процеси побудови розподілених баз даних»	4
Лабораторне заняття №5. Тема: «Обробка розподілених запитів. Засоби SQL для розподілених систем»	4
Лабораторне заняття №6. Тема: «Розподіл даних»	4
Лабораторне заняття №7. Тема: «Реплікація»	4
Лабораторне заняття №8. Тема: «Проектування розподілених реляційних баз даних»	4
Лабораторне заняття №9. Тема: «Організація баз знань»	4
Лабораторне заняття №10. Тема: «Технологія проектування сховищ даних»	4
Лабораторне заняття №11. Тема: «Забезпечення безпеки РСУБД»	4
Лабораторне заняття №12. Тема: «Засоби контролю цілісності інформації»	4
Лабораторне заняття №13. Тема: «Безпека розподілених інформаційних систем»	4
Лабораторне заняття №14. Тема: «Інформаційна безпека підприємств і організацій»	4
Модульний контроль	20
Виконання індивідуального завдання (СР)	30
Разом: Аудиторна робота	70
Самостійна робота (СР)	30
Всього:	100

КОНТРОЛЬ ТА КРИТЕРІЇ ОЦІНЮВАННЯ ЗНАНЬ СТУДЕНТІВ

При вивченні дисципліни використовуються наступні форми контролю знань студентів: поточний; модульний; підсумковий.

Поточний контроль передбачає перевірку теоретичних питань, самостійної роботи, практичних робіт та усне опитування по кожній практичній роботі. По даному виду контролю оцінювання знань здійснюється у відповідності до бального розподілу наведеного в попередній таблиці.

Модульний контроль передбачає виконання модульної контрольної роботи. Всі завдання оцінюються в 20 балів. Перше завдання (теоретичне) – 4 бали, друге завдання (практичне) – 8 балів, третє завдання (практичне) – 8 балів.

Формою підсумкового контролю є екзамен. Екзаменаційна оцінка (100 балів) є результатом виконання двох теоретичних питань (2 x 20 балів = 40 балів) та практичного завдання (60 балів).

Результуюча оцінка з дисципліни визначається як середня від балів набраних протягом семестру та отриманих на іспиті.

СПИСОК РЕКОМЕНДОВАНИХ ДЖЕРЕЛ

1. Ярцев В.П. Розподілені бази даних: навчальний посібник. К. ДУТ. 2018. 97с.
2. Берко А.Ю., Верес О.М., Пасічник В.В. Системи баз даних та знань. Книга 1. Організація баз даних та знань: навч. посібник. Львів : «Магнолія-2006». 2021. 440 с.
3. Берко А.Ю., Верес О.М., Пасічник В.В. Системи баз даних та знань. Книга 2. Системи управління базами даних та знань: навч. посібник. Львів : «Магнолія-2006». 2021. 584 с.