

Національний університет біоресурсів і природокористування України

Кафедра комп'ютерних наук

«ЗАТВЕРДЖУЮ»

Декан факультету інформаційних технологій

_____ О. Г. Глазунова

« _____ » _____ 20 ____ р.

«СХВАЛЕНО»

на засіданні кафедри комп'ютерних наук

Протокол № _____ від « ____ » _____ 20 ____ р.

Завідувач кафедри

_____ Б. Л. Голуб

«РОЗГЛЯНУТО»

Гарант ОП «Програмне забезпечення
інформаційних систем»

_____ доцент, к.т.н. Голуб Б.Л.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Організація сховищ даних

Спеціальність – 121 «Інженерія програмного забезпечення»

Освітня програма «Програмне забезпечення інформаційних систем»

Факультет інформаційних технологій

Розробник: завідувач кафедри комп'ютерних наук , доцент, к.т.н Голуб Б. Л.

Київ – 2021 р.

1 ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Організація сховищ даних»

Галузь знань, спеціальність, освітній ступінь	
Галузь знань	12 Інформаційні технології
Спеціальність	121 «Інженерія програмного забезпечення»
Освітня програма	«Програмне забезпечення інформаційних систем»
Освітній ступінь	Магістр
Характеристика навчальної дисципліни	
Вид	вибіркова
Загальна кількість годин	120
Кількість кредитів ECTS	4
Кількість змістових модулів	2
Форма контролю	іспит
Показники навчальної дисципліни для денної форми навчання	
	денна форма навчання
Рік підготовки	1
Семестр	1
Лекційні заняття	30 год.
Лабораторні заняття	30 год.
Курсова робота	КП
Самостійної роботи студента	60 год.
Кількість тижневих годин для денної форми навчання: аудиторних самостійної роботи студента –	4 год.

2 МЕТА ТА ЗАВДАННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Метою викладання дисципліни “ Організація сховищ даних ” є ознайомити магістрів з існуючими методичними підходами і технологічними засобами розробки сховищ даних, вивчити методики побудови та супроводження таких систем.

У курсі розглядаються принципи побудови систем, орієнтованих на аналіз даних, різні моделі даних, які використовуються для побудови сховищ даних. Також розглянуті питання побудови систем на основі сховищ даних, доставка даних в сховищі, технологія інтелектуального аналізу даних та інші питання.

Як інструментарій у курсі вивчається СУБД MS SQL Server. Саме у середовищі цього продукту студенти навчаються розроблювати інформаційно-аналітичні системи, інтегрувати дані із різних джерел у сховище даних.

Оволодіння такими знаннями дозволить реалізовувати задачі автоматизації обробки інформації, автоматизації керування об'єктами, в тому числі, сільськогосподарськими, за допомогою комп'ютерної техніки, та, найголовніше, створювати системи підтримки прийняття рішень. Такі знання майбутній спеціаліст зможе застосовувати як при подальшому навчанні, так і після отримання вищої освіти у своїй професійній діяльності.

Викладання дисципліни ґрунтується на знаннях по таких напрямках як:

- організація баз даних та знань;
- системи і технології управління базами даних;
- інформаційно-управляючі системи та технології в АПК.

Матеріал, який викладається у цій дисципліні, використовується магістрами для написання дипломної роботи та при вивченні дисциплін у магістратурі, перш за все, дисципліни «Проектування інформаційно-управляючих та інтелектуальних систем».

Набуття компетентностей

1. Відповідно ОПП ПЗІС

ЗК04. Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами інших галузей знань/видів економічної діяльності), у тому числі, з експертами природоохоронної галузі.

ЗК05. Здатність генерувати нові ідеї (креативність).

СК01. Здатність аналізувати предметні області, формувати, класифікувати вимоги до програмного забезпечення, насамперед, пов'язаних з природоохоронною галуззю.

СК02. Здатність розробляти і реалізовувати наукові та/або прикладні проекти у сфері інженерії програмного забезпечення.

СК04. Здатність розвивати і реалізовувати нові конкурентоспроможні ідеї в інженерії програмного забезпечення.

СК07. Здатність критично осмислювати проблеми у галузі інформаційних технологій та на межі галузей знань, інтегрувати відповідні знання та розв'язувати складні задачі у широких або мультидисциплінарних контекстах.

СК08. Здатність розробляти і координувати процеси, етапи та ітерації

життєвого циклу програмного забезпечення на основі застосування сучасних моделей, методів та технологій розроблення програмного забезпечення.

РН01. Знати і застосовувати сучасні професійні стандарти і інші нормативно-правові документи з інженерії програмного забезпечення.

РН03. Будувати і досліджувати моделі інформаційних процесів у прикладній області, насамперед, пов'язаної з природоохоронною галуззю.

РН04. Виявляти інформаційні потреби і класифікувати дані для проєктування програмного забезпечення.

РН07. Аналізувати, оцінювати і застосовувати на системному рівні сучасні програмні та апаратні платформи для розв'язання складних задач інженерії програмного забезпечення.

РН10. Модифікувати існуючі та розробляти нові алгоритмічні рішення детального проєктування програмного забезпечення.

РН12. Приймати ефективні організаційно-управлінські рішення в умовах невизначеності та зміни вимог, порівнювати альтернативи, оцінювати ризики.

РН17. Збирати, аналізувати, оцінювати необхідну для розв'язання наукових і прикладних задач інформацію, використовуючи науково-технічну літературу, бази даних та інші джерела.

2. Відповідно ОПП ІУСТ та КЕЕМ

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях, насамперед, пов'язаних з природоохоронною галуззю.

ЗК3. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК6. Здатність бути критичним і самокритичним.

ЗК7. Здатність генерувати нові ідеї (креативність).

СК2. Здатність комунікувати з представниками різних галузей знань, насамперед, природоохоронної галуззі, та сфер діяльності з метою з'ясування їх потреб в автоматизації обробки інформації.

СК3. Здатність збирати, формалізувати, систематизувати і аналізувати потреби та вимоги до комп'ютерної системи, що розробляється, експлуатується чи супроводжується.

СК5. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області певного проєкту в процесі його реалізації і супроводження.

СК6. Здатність збирати і аналізувати дані (включно з великими), насамперед, пов'язаних з природоохоронною галуззю, для забезпечення якості прийняття рішень.

СК8. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук: алгоритми розв'язання обчислювальних та логічних задач, алгоритми паралельних та розподілених обчислень, алгоритми аналітичної обробки й інтелектуального аналізу великих даних з оцінкою їх ефективності та складності.

СК10. Здатність використовувати програмні інструменти для організації командної роботи над проєктом.

СК11. Здатність розробляти та адмініструвати бази даних та знань, володіти сучасними теоріями та моделями даних та знань, методами їх інтерактивної та автоматизованої розробки, технологіями обробки та візуалізації.

СК13. Здатність ініціювати та планувати процеси розробки комп'ютерних систем та програмного забезпечення, включно з його розробкою, аналізом, тестуванням, системною інтеграцією, впровадженням і супроводом.

РН1. Ідентифікувати поняття, алгоритми та структури даних необхідні для опису, розробки або дослідження предметної області, що пов'язана, насамперед, з природоохоронною галуззю; забезпечити декомпозицію поставленої задачі з метою застосування відомих методів і технологій для її вирішення.

РН3. Аналізувати проміжні результати розробки або дослідження з метою з'ясування їх відповідності вимогам; розробляти тести та використовувати засоби верифікації, щоб переконатися у якості прийнятих рішень.

РН4. Аналізувати предметну область, насамперед, пов'язану з природоохоронною галуззю, розробки або дослідження, використовуючи наявну документацію, консультації з стейкхолдерами; розробляти документацію, що фіксує як функціональні, так і нефункціональні вимоги до розробки чи дослідження.

РН8. Розробляти та забезпечувати заходи з моніторингу, оптимізації, технічного обслуговування, виявлення відмов тощо.

3 ПРОГРАМА ТА СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Змістовий модуль 1. Проблеми представлення атомарних та багатовимірних структур даних

Тема лекційного заняття 1. Реляційна база даних. Правила Кодда (3 години лекцій)

Необхідною умовою роботи з СУБД є знання реляційної моделі БД. Це найбільш популярна модель зберігання даних. Технічна стаття «Реляційна модель даних для великих поділюваних банків даних» доктора Є.Ф. Кодда, представлена в 1970р., є родоначальницею сучасної теорії реляційних БД. Доктор Коддом визначив 13 правил реляційної моделі (які називають 12 правилами Кодда).

Тема лекційного заняття 2. Управління багатомірним представленням даних. Ідеологія і концепція. Правила Кодда (4 години лекцій)

До теперішнього часу в багатьох організаціях накопичені значні обсяги даних, на основі яких є можливість вирішення різноманітних аналітичних і управлінських завдань. Проблеми збереження та обробки аналітичної інформації стають все більш актуальними і привертають увагу фахівців і фірм, які працюють в області інформаційних технологій, що призвело до формування повноцінного ринку технологій бізнес-аналізу.

Тема лекційного заняття 3. Поняття сховища даних (4 години лекцій)

СД - предметно-орієнтований, інтегрований, рідко змінний, що підтримує хронологію, набір даних, організований для цілей підтримки прийняття рішень. Предметна орієнтація означає, що СД інтегрують інформацію, яка відображає різні точки зору на предметну область. Інтеграція припускає, що дані, що зберігаються в СД, приводяться до єдиного формату. Підтримка хронології означає, що всі дані в СД відповідають послідовним інтервалам часу.

Тема лекційного заняття 4. Архітектура OLAP-системи (4 години лекцій)

OLAP (Online Analytical Processing) - технологія оперативної аналітичної обробки даних, що використовує методи і засоби для збору, зберігання та аналізу багатовимірних даних з метою підтримки процесів прийняття рішень.

Змістовий модуль 2. Побудова сховищ даних

Тема лекційного заняття 5. Проблематика побудови сховищ даних (3 години лекцій)

Інтеграція різнорідних даних. Ефективне зберігання і обробка великих обсягів даних. Організація багаторівневих довідників метаданих. Забезпечення інформаційної безпеки СД.

Тема лекційного заняття 6. Використання технології «data flow» для отримання даних у сховище (3 години лекцій)

Скорочення витрат на проектування та розробку СД може бути досягнуто шляхом створення вітрин даних (ВД). ВД - це спрощений варіант СД, який

містить лише тематично об'єднані дані.

Тема лекційного заняття 7. Створення куба в SQL Server (3 години лекцій)

Методом SQL Server Business Intelligence Development Studio створити СД для предметної області «Абітурієнт».

Тема лекційного заняття 8. Проекти служб BI (3 години лекцій)

Служби BI: SSAS, SSTS, SSRS.

Тема лекційного заняття 9. Служба SSAS: обчислення (3 години лекцій)

Обчислення є багатовимірним виразом чи сценарієм, який використовується для визначення обчислюваного елемента, іменованого набору або призначення із зазначенням області в кубі в службах SSAS. Обчислення дозволяють додавати об'єкти, визначені не даними куба, а виразами, які можуть посилалися на інші частини куба, інші куби, або навіть дані поза бази даних служб SSAS. Обчислення дозволяють розширити можливості куба, збільшуючи гнучкість і продуктивність додатків бізнес-аналітики.

Основне призначення OLAP-систем - підтримка аналітичної діяльності, довільних запитів користувачів - аналітиків. Мета OLAP-аналізу - перевірка гіпотез.

У процесі аналізу даних часто виникає необхідність побудови залежностей між різними параметрами, число яких може бути значним. Можливість аналізу залежностей між різними параметрами припускає можливість подання даних у вигляді багатовимірної моделі - гіперкуба, або OLAP-куба.

СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин													
	денна форма							Заочна форма						
	тижні	усього	у тому числі					усього	у тому числі					
			л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.	
1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Змістовий модуль 1. Проблеми представлення атомарних та багатовимірних структур даних														
Тема 1 Введення в сховища даних	1-3	28	7		6		15							
Тема 2 Формати даних OLAP	4-7	31	8		8		15							
Разом за змістовим модулем 1	7	59	15		14		30							
Змістовий модуль 2. Побудова сховищ даних														
Тема 3 Розгортання служб аналізу	8-11	30	7		8		15							
Тема 4 Обчислення	12-15	31	8		8		15							
Разом за змістовим модулем 2	8	61	15		16		30							
Усього	15	120	30		30		60							

6 ТЕМИ ЛАБОРАТОРНИХ ЗАНЯТЬ

№ з/п	Назва теми	Кількість годин
1	Збір вимог щодо сховища даних	4 години
2	Побудова сховища даних	8 години
3	Розгортання кубу	4 години
4	Заповнення даними побудованого кубу	4 години
5	Обчислення KPI	4 години
6	Побудова звітів	6 години

7 КОНТРОЛЬНІ ПИТАННЯ, КОМПЛЕКТИ ТЕСТІВ ДЛЯ ВИЗНАЧЕННЯ РІВНЯ ЗАСВОЄННЯ ЗНАТЬ СТУДЕНТАМИ

САМОСТІЙНА РОБОТА

Для кращого засвоєння матеріалу студентам пропонується самостійно опанувати наведеними нижче темами і, як результат освоєння відповідного матеріалу, підготувати реферати на ці теми.

1. Об'єктно-орієнтована методологія розробки систем. Принципи об'єктно-орієнтованого підходу. Складові об'єктно-орієнтованої методології: об'єктно-орієнтований аналіз, об'єктно-орієнтоване проектування, об'єктно-орієнтоване програмування (10 годин).

2. Класи та об'єкти. Відношення між класами та об'єктами. Якість класів та об'єктів (5 годин).

3. Задача класифікації. Три основні підходи рішення задачі класифікації (5 годин).

4. Система позначень об'єктно-орієнтованого проектування (5 годин).

5. Уніфікована мова візуального моделювання UML. Діаграми функцій (варіантів виконання), послідовностей, взаємодії, класів, станів, компонентів, розміщення (10 годин).

6. Процес об'єктно-орієнтованого проектування. Переваги і недоліки об'єктно-орієнтованого підходу (10 годин).

7. CASE-системи - інструментальні засоби розробки систем (9 годин).

8. CASE-системи як засоби автоматизації розробки систем (9 годин).

9. Класифікація CASE-систем. Методи специфікації в CASE-системах (9 годин).

10. Об'єктно-орієнтоване CASE-засіб Rational Rose (9 годин).

11. Концепція відкритих систем. Цілі і завдання розвитку концепції відкритих систем. Архітектура концепцій і методів відкритих систем. Напрямки розвитку та моделі концепції відкритих систем (9 годин).

ПРИКЛАД ЕКЗАМЕНАЦІЙНОГО БІЛЕТУ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ ТА ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ			
ОС <u>магістр</u> Спеціальність «Комп'ютерні науки»	Кафедра комп'ютерних наук	ЕКЗАМЕНАЦІЙНИЙ БІЛЕТ № 4 з дисципліни <u>ОСД</u>	ЗАТВЕРДЖУЮ Зав. кафедри _____ Голуб Б.Л. 20 ____ р.
Експериментальне запитання (максимальна оцінка 10 балів за відповідь на кожне запитання)			
1. Детально поясніть, що представлено на цьому рисунку?			
2. Дайте визначення поняттю «KPI», та зазначте, які кроки необхідно виконати для його побудови.			

Тестові завдання

1. Розставити у відповідності до елементів OLAP-системи їхні визначення:

A. Факт – це	1) атомарна структура куба, що відповідає повному набору конкретних значень вимірів
B. Клітина – це	2) угруповання об'єктів одного виміру в об'єкти більш високого рівня
C. Ієрархія – це	3) безліч об'єктів одного або декількох типів, які забезпечують інформаційний контекст числового показника.
D. Вимір – це	4) числова величина яка розташовується в осередках гіперкуба.

2. До особливостей ROLAP-схеми типу "сніжинка" відносяться такі моменти (більше однієї правильної відповіді):

1	одна центральна таблиця фактів, що є ненормалізованою.
2	декілька ненормалізованих таблиць вимірів, які містять описову інформацію.
3	обчислені дані можуть зберігатися окремо від вихідних.
4	декілька нормалізованих таблиць вимірів, які містять описову інформацію.

3. Яким командам мови SQL відповідають наведені пояснення?

1. create	A. Створення нового рядка таблиці
2. drop	B. Видалення одного рядка таблиці
3. delete	C. Створення нового об'єкта у базі даних
4. insert	D. Видалення об'єкта базиданих

4. Скільки вимірів має таблиця реляційної бази даних?

(у бланку відповідей впишіть вірну відповідь одним словом)

5. Ієрархія вимірів OLAP- системи, в якій кількість рівнів визначено структурою незмінно, та кожна гілка ієрархічного дерева містить об'єкти кожного з рівнів, має назву:

1	збалансованої.
2	незбалансованої.
3	не рівної.

4	рівної.
6. Чи допустимо використання гіперкубу без часового виміру?	
<i>(у бланку відповідей впишіть вірну відповідь одним словом)</i>	
7. Утворюється переміщенням детальних даних, кількість звернень до яких знизилася:	
1	вхідний потік.
2	потік узагальнення.
3	архівний потік.
4	потік метаданих.
8. Факти, пов'язані з подіями або станом об'єкта, мають таку назву:	
1	Snapshot facts.
2	Transaction facts.
3	Line-item facts.
4	Event or state facts.
9. Які дані називаються агрегованими?	
<i>(у бланку відповідей впишіть вірну відповідь)</i>	
10. Чи входить до складу кубу сховище даних відповідно концепції C1F?	
<i>(у бланку відповідей впишіть вірну відповідь)</i>	

8 МЕТОДИ НАВЧАННЯ

M1. Лекція (проблемна, інтерактивна). M2. Лабораторна робота.
M3. Проблемне навчання. M4. Проектне навчання (індивідуальне). M8.
Дослідницький метод.

9 ФОРМИ КОНТРОЛЮ

МК1. Тестування. МК4. Методи усного контролю. МК5. Екзамен. МК7. Звіт.

10 РОЗПОДІЛ БАЛІВ, ЯКІ ОТРИМУЮТЬ СТУДЕНТИ

Оцінювання знань студента відбувається за 100-бальною шкалою і переводиться в національні оцінки згідно з табл. 1 «Положення про екзамени та заліки у НУБіП України» (наказ про уведення в дію від 27.12.2019 р. № 1371).

Поточний контроль		Рейтинг з навчальної роботи $R_{\text{НР}}$	Рейтинг з додаткової роботи $R_{\text{ДР}}$	Рейтинг штрафний $R_{\text{ШТР}}$	Підсумкова атестація (екзамен чи залік)	Загальна кількість балів
Змістовий модуль 1 0-100	Змістовий модуль 2 0-100	0-70	0-20	0-5	0-30	0-100

Шкала оцінювання

Рейтинг здобувача вищої освіти, бали	Оцінка національна за результати складання екзаменів заліків	
	екзаменів	заліків
	90-100 74-89	відмінно добре
		зараховано

60-73	задовільно	
0-59	незадовільно	не зараховано

Для визначення рейтингу студента із засвоєння дисципліни $R_{\text{дис}}$ (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи $R_{\text{нр}}$ (до 70 балів): $R_{\text{дис}} = R_{\text{нр}} + R_{\text{ат}}$.

11

МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ

1. Голуб Б. Л., Ящук Д. Ю. Організація сховищ даних: Навчальний посібник до вивчення дисципліни «Організація сховищданих». – Київ, 2018. – 150 с.

2. 1. Методичні вказівки до виконання курсового проекту з дисципліни «Організація сховища даних» для студентів, що навчаються за спеціальністю 121 «Інженерія програмного забезпечення», 122 «Комп'ютерні науки» ОС «Магістр» / Голуб Б.Л., Ящук Д.Ю. – Редакційно-видавничий відділ НУБіП України, 2018. – 20 с.

12

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. Перминов Г.И. Хранилища данных. Интернет-университет информационных технологий - ИНТУИТ.ру, 2009.

2. Чубукова И.А. Data Mining. БИНОМ. Лаборатория знаний, Интернет-университет информационных технологий - ИНТУИТ.ру, 2008.

Додаткова

3. Тельнов Ю.Ф. Интеллектуальные информационные системы в экономике: Учеб. Пособие. –М.:СИНТЕГ, 1999.

4. Туманов В.Е. Проектирование хранилищ данных для систем бизнес-аналитики. Интернет-университет информационных технологий - ИНТУИТ.ру, БИНОМ. Лаборатория знаний, 615 стр. 2010 г.

5. Козлов В.А. Открытые информационные системы. –М.:Финансы и статистика, 1999.

6. Липаев В.В. Системное проектирование сложных программных средств для информационных систем. –М.:Синтег, 1999.

13

ІНФОРМАЦІЙНІ РЕСУРСИ

1. ЕНК по дисципліні знаходиться за електронною адресою: <http://it.nubip.edu.ua/course/view.php?id=109>