

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ**  
**ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

**ЗАТВЕРДЖЕНО**

на засіданні кафедри  
кібербезпеки та інформаційних технологій  
Протокол № 2 від 29.08.2024 р.

**ПОГОДЖЕНО**

Проректор з навчально-методичної роботи

Каріна ЦЕМАШКАЛО



**РОЗПОДІЛЕНІ СХОВИЩА ДАНИХ**

**робоча програма навчальної дисципліни (РПНД)**

Галузь знань  
Спеціальність  
Освітній рівень  
Освітня програма

**12 "Інформаційні технології"**  
**122 "Комп'ютерні науки"**  
**другий (магістерський)**  
**" Комп'ютерні науки "**

Статус дисципліни  
Мова викладання, навчання та оцінювання

**обов'язкова**  
**українська**

Розробник(и):  
доктор технічних наук,  
професор

підписано КЕП

Володимир АЛЕКСІЄВ

Завідувач кафедри  
кібербезпеки та  
інформаційних технологій

Ольга СТАРКОВА

Гарант програми  
доктор технічних наук,  
професор

підписано КЕП

Сергій МІНУХІН

**Харків**  
**2024**

## ВСТУП

Розподілені сховища даних є основою для побудови ефективних інформаційних систем від рівня конфігурації серверу центру оброблення даних (ЦОД) до створення рішень масштабу хмарних обчислень (Cloud Computing). Принципи побудови, методи та технології створення, розгортання та підтримки розподілених сховищ даних є основою для масштабування сучасних веб-рішень, веб-додатків та веб-сервісів. Зараз існує багато програмних систем і сервісів для побудови великих за ємністю та доступних у мережі Інтернет сховищ даних, які надають можливості підвищити безпеку інформаційних систем, покращити їх надійність, а також надають засоби для масштабування проектних рішень. Знання та компетентності щодо вибору архітектури побудови сучасних засобів для збереження даних та впровадження промислових систем на їх основі є актуальними та необхідними у підготовці фахівців з комп'ютерних наук.

Метою навчальної дисципліни «Розподілені сховища даних» є формування системи теоретичних знань і набуття практичних умінь і навичок щодо застосування, налагодження й адміністрування систем на базі технологій розподілених сховищ даних та проектування відповідних надійних та економічно привабливих систем для збереження великих обсягів даних.

Завданнями навчальної дисципліни є: огляд існуючих рішень рівня окремого серверу, побудова мережевих сховищ даних (SAN та NAS систем), кластерних файлових та об'єктних сховищ даних, а також визначення ролі застосування децентралізованих систем у галузевих рішеннях. Поруч з цим, дисципліна розглядає особливості побудови та масштабування реляційних баз даних та NoSQL-рішень.

Предметом навчальної дисципліни є програмні засоби побудови розподілених сховищ даних.

Об'єкт навчальної дисципліни є процеси розгортання, адміністрування та впровадження у виробництво рішень для збереження даних та організація захищеного доступу до них.

Результати навчання та компетентності, які формує навчальна дисципліна визначено в табл. 1.

Таблиця 1

Результати навчання та компетентності, які формує навчальна дисципліна

<b>Результати навчання</b>	<b>Компетентності, якими повинен оволодіти здобувач вищої освіти</b>
РН4.	ЗК07, СК09.
РН6.	ЗК02, ЗК05, ЗК07, СК01, СК02, СК05, СК09.
РН7.	ЗК01, ЗК03, СК01.
РН9.	ЗК02.
РН10.	ЗК07, СК09.

Завершення табл. 1

Результати навчання	Компетентності, якими повинен оволодіти здобувач вищої освіти
РН12.	ЗК01, ЗК02, ЗК03, ЗК05, ЗК07, СК02, СК04, СК05, СК06, СК07, СК08, СК09, СК11, СК12.
РН13.	ЗК02, СК05.
РН17.	СК05.
РН18.	СК09.
РН20.	ЗК07, СК05, СК09.

де, РН4. Управляти робочими процесами у сфері інформаційних технологій, які є складними, непередбачуваними та потребують нових стратегічних підходів.

РН6. Розробляти концептуальну модель інформаційної або комп'ютерної системи.

РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

РН10. Проєктувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.

РН12. Проєктувати та супроводжувати бази даних та знань.

РН13. Оцінювати та забезпечувати якість інформаційних та комп'ютерних систем різного призначення.

РН17. Виявляти та усувати проблемні ситуації в процесі експлуатації програмного забезпечення, формулювати завдання для його модифікації або реінжинірингу.

РН18. Збирати, формалізувати, систематизувати і аналізувати потреби та вимоги до інформаційної або комп'ютерної системи, що розробляється, експлуатується чи супроводжується.

РН20. Розробляти алгоритми та компоненти програмного забезпечення комп'ютерних інформаційних систем для надпродуктивних систем оброблення великих даних (включно з розподіленими та паралельними обчисленнями) та сервісів хмарних платформ.

ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК05. Здатність вчитися й оволодівати сучасними знаннями.

ЗК07. Здатність генерувати нові ідеї (креативність).

СК01. Усвідомлення теоретичних засад комп'ютерних наук.

СК02. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.

СК04. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень.

СК05. Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.

СК06. Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

СК07. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

СК08. Здатність розробляти і реалізовувати проєкти зі створення програмного

забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом.

СК09. Здатність розробляти та адмініструвати бази даних та знань.

СК11. 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.

СК12. Здатність розробляти, застосовувати та інтегрувати технології оброблення та аналізу даних в надпродуктивних системах та хмарних платформах для забезпечення ефективного використання обчислювальних ресурсів комп'ютерних систем.

## **ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

### **Зміст навчальної дисципліни**

**Змістовий модуль 1. Розподілені файлові системи та бази даних.**

**Тема 1. Технології розподілених сховищ даних для рішення завдань обробки великих обсягів даних.**

**1.1. Класифікація технологій створення розподілених баз і сховищ даних.** Визначення основних напрямків вивчення дисципліни.

**1.2. Обробка великих даних (Big Data) та технології побудови озер даних (Data Lake).**

Визначення проблеми обробки великих обсягів даних. Рішення щодо побудови гетерогенних сховищ даних на базі хмарних обчислень (Cloud Computing).

**1.3. Сучасний центр обробки даних. Технології віртуалізації.**

Сучасні технології віртуалізації рівня робочої станції (Desktop) та серверу (Server). Основи побудови засобів хмарних обчислень. Приватна хмара. Шар збереження даних.

**1.4. Засоби серверної віртуалізації.**

Технологія живої міграції. Відмінності систем віртуалізації на базі гіпервізора та рівня застосування контейнерів.

**Тема 2. Розподілені файлові сховища даних на базі технологій SAN та NAS. Хмарні теки.**

**2.1. Побудова сховища даних рівня окремого серверу.**

Особливості файлових систем. Технологія LVM. Застосування накопичувачів, які поєднано у RAID-масив.

**2.2. Побудова мережевих сховищ даних.**

Застосування SAN та NAS рішень на прикладі технологій TrueNAS.

**2.3. Застосування рішень для побудови хмарної теки.**

Практичний досвід роботи з ownCloud та Nextcloud системами.

### **Тема 3. Об'єктні сховища даних. Кластерні файлові системи.**

#### **3.1. Кластерні файлові системи.**

Побудова кластеру для збереження даних на прикладі концепції GoogleFS.  
Алгоритм MapReduce.

#### **3.2. Розподілене сховище даних Ceph.**

Забезпечення зберігання даних засобами кластеру Ceph.

### **Тема 4. Масштабування систем збереження даних та створення баз знань на основі реляційних СКБД.**

#### **4.1. Особливості технології OLAP (online analytical processing).**

Застосування реляційних баз даних як платформи для побудови баз знань для систем прийняття рішень.

#### **4.2. Реплікація бази даних.**

Рішення завдань реплікації даних на прикладі MySQL.

#### **4.3. Побудова кластеру реляційної бази даних.**

Засоби побудови кластеру PostgreSQL. Рішення Galera Cluster для MariaDB або MySQL.

### **Тема 5. Технології нереляційних баз даних.**

#### **5.1. Особливості бази даних типу «ключ-значення».**

Розподілене сховище, на прикладі технології Redis, для рішення завдань побудови веб-застосунків, що масштабуються.

#### **5.2. База даних MongoDB.**

Особливості застосування та масштабування рішень, що побудовані на основі нереляційної бази даних.

#### **5.3. Масштабовані та надійні сховища даних на базі технологій Hadoop та Cassandra.**

Побудова рішень для збереження та обробки великих даних.

**Змістовий модуль 2.** Застосування та особливості проектування рішень на базі розподілених сховищ даних.

### **Тема 6. Моніторинг стану розподілених обчислювальних систем і сховищ даних на базі відкритих програмних засобів.**

#### **6.1. Моніторинг засобами Nagios, Zabbix.**

Особливості використання загальних мультіагентних систем моніторингу.

#### **6.1. Моніторинг Prometheus та візуалізація даних Grafana.**

Особливості використання систем моніторингу, які побудовано за модульним принципом.

## **Тема 7. Особливості розроблення веб-додатків і веб-сервісів із застосуванням технологій розподілених сховищ даних.**

### **7.1. Теорема CAP (теорема Брюера).**

Поєднання узгодженості даних, доступності та стійкості до розділення у розподілених системах.

### **7.2. Особливості масштабування веб-додатків.**

Розуміння ролі сховищ даних у рішеннях завдань побудови веб-застосунків. Особливості архітектури сучасних веб-застосунків, наприклад, застосування мікросервісів, використання ланцюжків повідомлень тощо.

### **7.3. Централізовані та децентралізовані системи.**

Особливості побудови децентралізованих систем збереження даних. Застосування торрент-мереж. Технологія блокчейн.

## **Тема 8. Застосування технологій розподілених сховищ даних у хмарних обчисленнях.**

Особливості реалізації парадигми «сховище даних як сервіс» у хмарних рішеннях: Amazon Web Services, Microsoft Azure, Google Cloud та Oracle Cloud Infrastructure.

## **Тема 9. Перспективи розвитку систем та технологій розподілених сховищ даних.**

Огляд сучасних тенденцій та перспектив розвитку технологій та комплексних рішень щодо побудови й застосування розподілених сховищ даних.

Перелік лабораторних занять / завдань за навчальною дисципліною наведено в табл. 2.

Таблиця 2

### **Перелік лабораторних занять / завдань**

Назва теми та / або завдання	Зміст
Тема 2. Завдання 1.	Розгорнути у середовищі віртуалізації систему TrueNAS або аналог. Відкрити доступ за протоколом FTP для збереження та завантаження файлів.
Тема 2. Завдання 2.	Дослідити роботу засобів ownCloud або Nextcloud.
Тема 3. Завдання 3.	Розгорнути у середовищі віртуалізації кластер (три вузла) Serf. Дослідити роботу відповідного рішення.
Тема 4. Завдання 4.	У середовищі віртуалізації дослідити особливості налаштування реплікації MySQL (реляційна база даних).
Тема 4. Завдання 5.	У середовищі віртуалізації дослідити особливості розгортання кластеру реляційної бази даних за технологією Galera Cluster або аналог.
Тема 5. Завдання 6.	Розгорнути три вузла MongoDB та на її основі дослідити алгоритм MapReduce (рішення NoSQL).

3. Перелік самостійної роботи за навчальною дисципліною наведено в табл.

Таблиця 3

### Перелік самостійної роботи

Назва теми та / або завдання	Зміст
Тема 1. Завдання 1.	За матеріалами Інтернет розглянути архітектурні рішення побудови Data Lake.
Тема 2. Завдання 2.	Розгорнути віртуальну машину на базі сучасного дистрибутиву, наприклад, Ubuntu Server. Визначити можливість застосування технології LVM. Розгорнути веб-сервер Apache та надати користувачам можливість завантаження файлів веб-застосунку за технологією FTP. Виконати дослідження безпеки відповідного рішення.
Тема 5. Завдання 3.	Дослідити особливості розгортання та застосування рішень Redis або аналог.
Тема 5. Завдання 4.	За матеріалами Інтернет порівняти особливості технологій Hadoop та Cassandra.
Тема 6. Завдання 5.	Розгорнути (за наявності техн. можливості) та дослідити засоби моніторингу Zabbix.
Тема 6. Завдання 6.	Розгорнути (за наявності техн. можливості) і виконати дослідження засобів моніторингу Prometheus та візуалізації даних Grafana.
Тема 7. Завдання 7.	За матеріалами Інтернет розглянути архітектурні рішення побудови децентралізованих системи збереження даних.
Тема 8. Завдання 8.	За матеріалами Інтернет виконати дослідження та порівняти реалізацію сервісів, які застосовують розподілені сховища даних, у рішеннях хмарних обчислень.
Тема 9. Завдання 9.	За матеріалами Інтернет дослідити тенденції розвитку технологій розподілених сховищ даних. Написання есе.

Кількість годин лекційних, лабораторних занять та годин самостійної роботи наведено в робочому плані (технологічній карті) з навчальної дисципліни.

### МЕТОДИ НАВЧАННЯ

У процесі викладання навчальної дисципліни для набуття визначених результатів навчання, активізації освітнього процесу передбачено застосування таких методів навчання, як:

Словесні (лекція (Тема 2, 3, 4, 5, 6, 8), проблемна лекція (Тема 1, 7, 9)).

Наочні (демонстрація (Тема 2, 3, 5, 6)).

Практичні (лабораторна робота (Тема 2, 3, 4, 5), есе (Тема 9), кейс-метод (Тема 8)).

## ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

Університет використовує 100 бальну накопичувальну систему оцінювання результатів навчання здобувачів вищої освіти.

**Поточний контроль** здійснюється під час проведення лекційних та лабораторних занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретної роботи і оцінюється сумою набраних балів: – для дисциплін з формою семестрового контролю залік: максимальна сума – 100 балів; мінімальна сума – 60 балів.

**Підсумковий контроль** включає семестровий контроль та атестацію здобувача вищої освіти.

**Семестровий контроль** проводиться у формах, диференційованого заліку або заліку.

**Підсумкова оцінка за навчальною дисципліною** визначається:

– для дисциплін з формою семестрового контролю залік – сумуванням всіх балів, отриманих під час поточного контролю.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: Індивідуальні навчально-дослідні завдання (20 балів), письмова контрольна робота (10 балів), есе (10 балів), виконання лабораторного практикуму (60 балів).

Семестровий контроль: Залік.

Більш детальну інформацію щодо системи оцінювання наведено в робочому плані (технологічній карті) з навчальної дисципліни.

## РЕКОМЕНДОВАНА ЛІТЕРАТУРА

### Основна

1. Бази даних : лабораторний практикум для студентів галузі знань 12 "Інформаційні технології" першого (бакалаврського) рівня [Електронний ресурс] / укл. В. В. Федько, В. П. Бурдаєв; Харківський національний економічний університет ім. С. Кузнеця. - Х.: ХНЕУ ім. С. Кузнеця, 2019. - 229с. [Електронний ресурс]. – Режим доступу: <http://repository.hneu.edu.ua/handle/123456789/21526>

2. Олещенко, Л. М. Технології оброблення великих даних. Конспект лекцій [Електронний ресурс] : навчальний посібник для студентів спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем») / Л. М. Олещенко ; КПІ ім. Ігоря Сікорського.– Київ : КПІ ім. Ігоря Сікорського, 2021. – 227 с. – Режим доступу: <https://ela.kpi.ua/handle/123456789/42206>

3. Навчальний посібник з дисципліни “Технології Big Data” для студентів спеціальності 123 - “Комп’ютерна інженерія” / Таран В.І., Гордієнко Ю.Г.,



Стіренко С.Г. - Київ: КПІ, 2022. - 56 с. [Електронний ресурс]. – Режим доступу: <https://comsys.kpi.ua/tecnologi-yi-big-data>

4. Гордієнко Ю.Г., Таран В.І. ХМАРНІ ОБЧИСЛЕННЯ. Конспект лекцій. Навчальний посібник для здобувачів ступеня магістра. [Електронне мережне навчальне видання] КПІ, 2022. - 1224 с. – Режим доступу: <https://comsys.kpi.ua/katalog/files/konspekt-lekciy-1.pdf>

### Додаткова

5. Технології роботи з Big Data. Методичні рекомендації до лабораторних робіт для студентів спеціальності 126 "Інформаційні системи та технології" освітньої програми "Інформаційні системи та технології" першого (бакалаврського) рівня [Електронний ресурс] / уклад. І. В. Кобзев, Н. М. Кобзева, О. В. Тесленко; ХНЕУ ім. С. Кузнеця. – Харків: ХНЕУ ім. С. Кузнеця, 2023. — 82 с. [Електронний ресурс]. – Режим доступу: <http://repository.hneu.edu.ua/handle/123456789/30046>

6. Розподілені сховища даних: методичні рекомендації до викон. лабораторних робіт для студ. спец. 122 "Комп'ютерні науки та інформаційні технології" другого (магістерського) рівня [Електронний ресурс] / укл. В.О. Алексієв; ХНЕУ ім. С. Кузнеця. - Х. : ХНЕУ ім. С. Кузнеця, 2017. - 65 с. [Електронний ресурс]. – Режим доступу: <http://www.repository.hneu.edu.ua/jspui/handle/123456789/17914>

7. Горденко Ю., Таран В. ХМАРНІ ОБЧИСЛЕННЯ. Лабораторний практикум. Навчальний посібник для здобувачів ступеня магістра [Електронне мережне навчальне видання] КПІ, 2021. - 38 с. – Режим доступу: <https://comsys.kpi.ua/hmarni-obchislennya>

8. Каштан В.Ю., Іванов Д.В. Конспект лекцій з дисципліни “Бази даних в інформаційних системах”. Для студентів галузі знань 12 “Інформаційні технології” спеціальності 126 “Інформаційні системи та технології”. – Д.: НТУ «ДП», 2021. – 58 с. [Електронний ресурс]. – Режим доступу: [https://it.nmu.org.ua/ua/scientific\\_method\\_materials/textbooks.php](https://it.nmu.org.ua/ua/scientific_method_materials/textbooks.php)

9. Trent R. Hein, Evi Nemeth, Garth Snyder, Ben Whaley, Dan Mackin. UNIX and Linux System Administration Handbook, 5th Edition / Addison-Wesley Professional, 2017. – 1232 p.

10. Initial Server Setup with Ubuntu 20.04 [Электронный ресурс] / Brian Boucheron. DigitalOcean, 2021. – Режим доступа : <https://www.digitalocean.com/community/tutorials/initial-server-setup-with-ubuntu-20-04>.

11. Jamon Camisso. Sysadmin eBook: Making Servers Work – DigitalOcean, 2020. – 281 p.

12. M. van Steen and A.S. Tanenbaum, Distributed Systems, 4th ed., distributed-systems.net, 2023. – 685 p. [Electronic resource]. – Access mode: <https://www.distributed-systems.net/index.php/books/ds4/>.

13. Кравченко П. Блокчейн і децентралізовані системи : навч. посібник для студ. закладів вищ. освіти : в 3 частинах. Ч. 1 / П. Кравченко, Б. Скрябін, О. Дубініна. – Харків : ПРОМАРТ, 2019. – 452 с.

14. Кравченко П. Блокчейн і децентралізовані системи: навч. посібник для студ. закладів вищ. освіти: в 3 частинах. Ч. 2 / П. Кравченко, Б. Скрябін, О. Курбатов, О. Дубініна. - Харків, 2019. – 412 с.

15. Kravchenko, P. Blockchain and decentralized systems : in three volumes. V.3 / P. Kravchenko, B. Skriabin, O. Kurbatov, O. Dubinina. – Kharkiv : 2020. 298 p.

16. Martovytskyi V. Technology for monitoring the functioning state of distributed computer systems / V. Martovytskyi, Y. Koltun, D. Holubnychyi et al. // Системи управління, навігації та зв'язку : зб. наук. пр. – 2022. – Вип. 1 (67). – С. 75-80. Режим доступа :

<http://www.repository.hneu.edu.ua/handle/123456789/27369>.

17. Minukhin S. Performance study of the DTU model for relational databases on the Azure platform / S. Minukhin // Сучасний стан наукових досліджень та технологій в промисловості. – 2022. – № 1 (19). – С. 27-39. Режим доступа : <http://repository.hneu.edu.ua/handle/123456789/27739>.

### **Інформаційні ресурси**

18. TrueNAS Documentation Hub [Electronic resource]. –Access mode: <https://www.truenas.com/docs/core/13.0/gettingstarted/>

19. Nextcloud Server Administration Guide [Electronic resource]. –Access mode: [https://docs.nextcloud.com/server/latest/admin\\_manual/](https://docs.nextcloud.com/server/latest/admin_manual/)

20. The Ceph Documentation [Electronic resource]. –Access mode: <https://docs.ceph.com/en/latest/>

21. MySQL Documentation [Electronic resource]. –Access mode: <https://dev.mysql.com/doc/>

22. MariaDB Server documentation [Electronic resource]. –Access mode: <https://mariadb.org/documentation/>

23. What is MariaDB Galera Cluster? [Electronic resource]. –Access mode: <https://mariadb.com/kb/en/what-is-mariadb-galera-cluster/>

24. MongoDB Documentation [Electronic resource]. –Access mode: <https://www.mongodb.com/docs/>

25. Zabbix Documentation [Electronic resource]. –Access mode: <https://www.zabbix.com/manuals>

26. What is Prometheus? [Electronic resource]. –Access mode: <https://prometheus.io/docs/introduction/overview/>

27. Apache Cassandra documentation [Electronic resource]. –Access mode: <https://cassandra.apache.org/doc/latest/>

28. Apache Hadoop Documentation [Electronic resource]. –Access mode: <https://hadoop.apache.org/docs/current/>

29. Сайт персональних навчальних систем ХНЕУ ім. С. Кузнеця за дисципліною " Розподілені сховища даних" [Electronic resource]. –Access mode: <https://pns.hneu.edu.ua/course/view.php?id=11473>.