

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

"ЗАТВЕРДЖУЮ"

Заступник керівника

(професор з науково-педагогічної роботи)



М. В. Афанасьєв
М. В. Афанасьєв

РОЗПОДІЛЕНІ СХОВИЩА ДАНИХ

робоча програма навчальної дисципліни

Галузь знань **12 "ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ"**
Спеціальність **122 "КОМП'ЮТЕРНІ НАУКИ"**
Освітній рівень **другий (магістерський)**
Освітня програма **"КОМП'ЮТЕРНІ НАУКИ"**

Вид дисципліни
Мова викладання, навчання та оцінювання

**вибіркова
українська**

Завідувач кафедри кібербезпеки
та інформаційних технологій

Євсєєв С.П.

Харків
ХНЕУ ім. С. Кузнеця
2019

ЗАТВЕРДЖЕНО
на засіданні кафедри Кібербезпеки та інформаційних технологій
Протокол № 1 від 26.08.2019 р.

Розробники:
Алексієв В. О., д. т. н.,
проф. кафедри кібербезпеки та інформаційних технологій

**Лист оновлення та перезатвердження
робочої програми навчальної дисципліни**

Навчальний рік	Дата засідання кафедри – розробника РПНД	Номер протоколу	Підпис завідувача кафедри

1. Вступ

Анотація навчальної дисципліни:

Подано тематичний план навчальної дисципліни й її змістовність за модулями та темами, вміщено плани лекцій і лабораторних занять, матеріал щодо закріплення знань (завдання для самостійної роботи, контрольні запитання), методичні рекомендації та оцінювання знань студентів.

Розподілені сховища даних і розподілені бази даних є основою побудови ефективних інформаційних систем від рівня центру оброблення даних до створення рішень масштабу глобальних грид-систем і систем хмарних обчислень. Принципи побудови, методи та технології створення, розгортання та підтримки розподілених сховищ даних є основою для масштабування сучасних веб-рішень, веб-додатків та веб-сервісів.

Мета навчальної дисципліни:

Метою викладання навчальної дисципліни 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 об'ємів даних.

Курс	1 М	
Семестр	1	
Кількість кредитів ECTS	5	
Аудиторні навчальні заняття	лекції	12
	семінарські, практичні	–
	лабораторні	28
Самостійна робота		110
Форма підсумкового контролю	екзамен	

Структурно-логічна схема вивчення навчальної дисципліни:

Попередні дисципліни	Наступні дисципліни
Операційні системи	Хмарні обчислення
Комп'ютерні мережі	Технології обробки даних в інформаційно- комунікаційних системах
Організація баз даних	Дипломна робота

2. Компетентності та результати навчання за дисципліною:

Компетентності	Результати навчання
Здатність до проведення аналізу завдань щодо застосування розподілених сховищ даних	Знати термінологію та сучасний стан розвитку розподілених сховищ даних
Розрізняти технології SAN і NAS для рішення завдань побудови розподілених файлових сховищ даних	Використовувати основні принципи побудови розподілених файлових сховищ даних
Визначати ефективні напрями масштабування систем збереження даних на базі реляційних СУБД	Володіти основами проектування розподілених баз даних

Компетентності	Результати навчання
Використовувати технології нереляційних СУБД	Володіти основами проектування розподілених баз даних
Вміти розгортати та налаштовувати системи моніторингу стану вузлів грид-систем, розподілених обчислювальних систем і сховищ даних	Знати методи розгортати та налаштування систем моніторингу стану розподілених обчислювальних систем і сховищ даних
Вміти оцінювати складність завдань щодо розгортання розподілених сховищ даних	Вміти аналізувати та обирати оптимальні рішення під час використання програмно-апаратних засобів для побудови розподілених сховищ даних
Орієнтуватися у напрямках рішень завдань щодо оброблення великих даних та застосування для цього розподілених сховищ даних	Знати основи комп'ютерних наук щодо галузі застосування у рішенні завдань обробки великих даних
Визначати оптимальні напрями застосування масштабованих та надійних сховища даних та вміти розгортати та адмініструвати рішення на базі розподілених сховищ даних	Розв'язувати проблеми масштабованості, проектування та експлуатації систем розподілених сховищ даних

3. Програма навчальної дисципліни

Змістовий модуль 1. Розподілені файлові системи. Технології розподілених баз даних.

Тема 1. Вступ. Основні терміни, визначення та класифікація технологій створення розподілених баз і сховищ даних.

Основні терміни та визначення. Огляд особливостей та технологій, що застосовуються у розподілених інформаційних системах та застосування розподілених сховищ даних для рішення сучасних бізнес-завдань підприємствами та під час проведення наукових досліджень. Визначення концепції *Big Data*.

Класифікація технологій створення розподілених баз і сховищ даних. Тенденції розвитку сучасних центрів оброблення даних. Основи масштабування веб-додатків.

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

Надійність збереження даних і масштабування файлових сховищ даних. Дискові масиви RAID. Менеджер логічних томів LVM. Мережева система зберігання даних NAS. Протоколи NFS, SMB, FTP та технологія доступу до даних *WebDAV*. Мережа зберігання даних SAN. Особливості застосування інтерфейсу *iSCSI*.

Особливості застосування технологій розподілених файлових систем на прикладі систем *GlusterFS* та *Ceph*.

Тема 3. Масштабування систем збереження даних на базі реляційних СУБД. Кластери на базі СУБД *MySQL* та *PostgreSQL*.

Технології реплікації. Особливості розгортання та застосування СУБД, які реалізують модель реляційних баз даних. Надійність збереження даних і доступність рішень на базі технологій реляційних баз даних.

Кластерні рішення. Масштабування SQL-рішень на основі сегментування даних. Масштабування рішень на базі реляційних баз даних.

Тема 4. Технології нереляційних СУБД у розподілених сховищах даних. Особливості систем *MongoDB*, *CouchDB* та *Redis*.

Особливості розробки інформаційних систем на базі *NoSQL*-рішень. Технологічні рішення на прикладі СУБД *MongoDB* *CouchDB* та *Redis*. Порівняння та оцінювання сучасних рішень на базі концепції *NoSQL*.

Технології реплікації. Масштабування *NoSQL*-рішень на основі сегментування даних. Особливості розгортання та підтримки рішень на базі розподіленої бази даних, яка застосовує нереляційну модель даних.

Тема 5. Моніторинг стану розподілених обчислювальних систем і сховищ даних на базі відкритих програмних засобів *Nagios*, *Icinga* та *Zabbix*.

Особливості рішення завдань моніторингу стану вузлів та сервісів у рішеннях на базі хмарних обчислень. Моніторинг стану розподілених систем та сховищ даних на прикладі систем *Nagios*, *Icinga* та *Zabbix*.

Застосування систем моніторингу у час упровадження та підтримки розподілених баз і сховищ даних.

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

Тема 6. Особливості розроблення веб-додатків і веб-сервісів із застосуванням технологій розподілених сховищ даних.

Масштабування рішень. Основні ідеї та технологічні рішення застосування розподілених сховищ даних для створення архітектури інформаційних систем для підтримки сучасних веб-сервісів та розподілених веб-додатків. Особливості застосування й масштабування систем на базі SQL та *NoSQL*-рішень. Особливості рішень у ситуації надвеликих навантажень на веб-ресурс. Особливості та недоліки технології розподілених систем кешу даних в оперативній пам'яті (*memcached*).

Застосування технологій хмарних обчислень. Особливості автоматизації отримання замовником ресурсів хмарних систем. Автоматизація процесів розгортання та підтримки розподілених баз та сховищ даних. Системи *Puppet*, *Chef* та *Ansible*. Технології віртуалізації. Контейнери *Docker*. Особливості технологій *OpenStack*.

Тема 7. Застосування технологій розподілених сховищ даних у ґрид-системах та у хмарних обчисленнях.

Особливості супроводження та підтримки грид-систем. Розгортання та супроводження розподілених систем на прикладі кластеру баз даних на базі PostgreSQL.

Тема 8. Місце та значення технологій розподілених сховищ даних під час оброблення великих масивів даних.

Програмна модель *Map Reduce*. Особливості розподіленого паралельного оброблення великих масивів даних із використанням кластерів. Переваги та недоліки застосування моделі *Map Reduce*.

Проект *Apache Hadoop* та технології *Apache Spark*. Платформа *Apache Hadoop* для організації розподіленого оброблення великих об'ємів даних. Технології *Apache Spark* для розроблення високопродуктивних розподілених систем для вирішення завдань оброблення великих даних і машинного навчання.

Тема 9. Масштабовані та надійні сховища даних на базі технологій *Apache HBase* та *Apache Cassandra*.

Технологія моделі великих табличних даних (*Big Table*), що реалізовані у системі *HBase*. Розподілена файлова система *Hadoop Distributed Filesystem* (HDFS) для забезпечення надійності збереження даних. Особливості архітектури *HBase*.

Технологія моделі великих табличних даних (*Big Table*), що реалізовані у системі *Cassandra*. Особливості архітектури *Apache Cassandra*.

Тема 10. Перспективи розвитку систем та технологій розподілених сховищ даних.

Аналіз платформ сучасних веб-порталів, сервісів пошуку та платформ соціальних мереж й загальних рішень веб-сервісів. Особливості застосування РСД у завданнях масштабування веб-рішень та підтримки безвідмовної роботи інформаційних систем в умовах значного навантаження на веб-сервіси та веб-ресурси.

Застосування технологій розподілених сховищ даних у задачах розроблення сервісів хмарних обчислень. Розподілені сховища даних як сервіс, що надається у рамках хмарних обчислень. Особливості супроводження та адміністрування систем РСД як складових технологічної платформи хмарних обчислень.

Лабораторний практикум

Лабораторна робота 1. "Організація сховищ даних типу SAN та NAS на основі дистрибутиву FreeNAS".

Лабораторна робота 2. "Налаштування розподіленої файлової системи Serf в операційній системі Linux на базі віртуальних машин".

Лабораторна робота 3. "Розгортання СУБД MySQL, налаштування механізму реплікації даних із застосуванням технологій віртуалізації та налаштування кластера Percona XtraDB Cluster на базі технологій віртуалізації".

Лабораторна робота 4. "Розгортання та робота із розподіленою базою даних на основі СУБД MongoDB у середовищі віртуальних машин".

4. Порядок оцінювання результатів навчання

Система оцінювання сформованих компетентностей у студентів враховує види занять, які згідно з програмою навчальної дисципліни передбачають лекційні, лабораторні заняття, а також виконання самостійної роботи. Оцінювання сформованих компетентностей у студентів здійснюється за накопичувальною 100-бальною системою. Відповідно до Тимчасового положення “Про порядок оцінювання результатів навчання студентів за накопичувальною бально-рейтинговою системою” ХНЕУ ім. С. Кузнеця, контрольні заходи включають:

поточний контроль, що здійснюється протягом семестру під час проведення лекційних, лабораторних занять і оцінюється сумою набраних балів (максимальна сума – 60 балів; мінімальна сума, що дозволяє студенту скласти іспит, – 35 балів);

модульний контроль, що проводиться у формі колоквиуму як проміжний міні-екзамен з ініціативи викладача з урахуванням поточного контролю за відповідний змістовий модуль і має на меті *інтегровану* оцінку результатів навчання студента після вивчення матеріалу з логічно завершеної частини дисципліни – змістового модуля;

підсумковий/семестровий контроль, що проводиться у формі семестрового екзамену, відповідно до графіку навчального процесу.

Порядок проведення поточного оцінювання знань студентів. Оцінювання знань студента під час семінарських, практичних і лабораторних занять та виконання індивідуальних завдань проводиться за такими критеріями:

розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються; ступінь засвоєння фактичного матеріалу навчальної дисципліни; ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються; вміння поєднувати теорію з практикою при розгляді виробничих ситуацій, розв'язанні задач, проведенні розрахунків у процесі виконання індивідуальних завдань та завдань, винесених на розгляд в аудиторії; логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки; арифметична правильність виконання індивідуального та комплексного розрахункового завдання; здатність проводити критичну та незалежну оцінку певних проблемних питань; вміння пояснювати альтернативні погляди та наявність власної точки зору, позиції на певне проблемне питання; застосування аналітичних підходів; якість і чіткість викладення міркувань; логіка, структуризація та обґрунтованість висновків щодо конкретної проблеми; самостійність виконання роботи; грамотність подачі матеріалу; використання методів порівняння, узагальнення понять та явищ; оформлення роботи.

Загальними критеріями, за якими здійснюється оцінювання позааудиторної самостійної роботи студентів, є: глибина і міцність знань, рівень мислення, вміння систематизувати знання за окремими темами, вміння робити обґрунтовані висновки, володіння категорійним апаратом, навички і прийоми виконання практичних завдань, вміння знаходити необхідну інформацію, здійснювати її систематизацію та обробку, самореалізація на лабораторних заняттях.

Підсумковий контроль знань та компетентностей студентів з навчальної дисципліни здійснюється на підставі проведення семестрового екзамену, завданням якого є перевірка розуміння студентом програмного матеріалу в цілому, логіки та взаємозв'язків між окремими розділами, здатності творчого використання накопичених знань, вміння формулювати своє ставлення до певної проблеми навчальної дисципліни тощо.

Екзаменаційний білет охоплює програму дисципліни і передбачає визначення рівня знань та ступеня опанування студентами компетентностей.

Кожен екзаменаційний білет складається із 3 практичних ситуацій (один стереотипний, один діагностичних та одне евристичне завдання), які передбачають вирішення типових професійних завдань фахівця на робочому місці та дозволяють діагностувати рівень теоретичної підготовки студента і рівень його компетентності з навчальної дисципліни.

Результат семестрового екзамену оцінюється в балах (максимальна кількість – 40 балів, мінімальна кількість, що зараховується, – 25 балів) і проставляється у відповідній графі екзаменаційної "Відомості обліку успішності".

Студента слід **вважати атестованим**, якщо сума балів, одержаних за результатами підсумкової/семестрової перевірки успішності, дорівнює або перевищує 60. Мінімумально можлива кількість балів за поточний і модульний контроль упродовж семестру – 35 та мінімумально можлива кількість балів, набраних на екзамені, – 25.

Підсумкова оцінка з навчальної дисципліни розраховується з урахуванням балів, отриманих під час екзамену, та балів, отриманих під час поточного контролю за накопичувальною системою. Сумарний результат у балах за семестр складає: "60 і більше балів – зараховано", "59 і менше балів – не зараховано" та заноситься у залікову "Відомість обліку успішності" навчальної дисципліни.

Розподіл балів за тижнями

(вказати засоби оцінювання згідно з технологічною картою)

Теми змістовних модулів		Лекційні заняття	Виконання лабораторних робіт	Захист лабораторних робіт	Поточні КР	Екзамен	Усього
Змістовий модуль 1	Тема 1	1 тиждень	0,5				0,5
		2 тиждень		0,5			0,5
	Тема 2-3	3 тиждень	0,5	0,5			1
		4 тиждень		0,5			0,5
	Тема 4-5	5 тиждень	0,5	0,5	10		11
		6 тиждень		0,5			0,5
	Тема 6-7	7 тиждень	0,5	0,5			1
		8 тиждень		0,5		5	5,5
	Тема 8-9	9 тиждень	0,5	0,5	10		11

Теми змістовних модулів		Лекційні заняття	Виконання лабораторних робіт	Захист лабораторних робіт	Поточні КР	Екзамен	Усього
Змістовий модуль 2	10 тиждень		0,5				0,5
	Тема 10	11 тиждень	0,5	0,5			1
		12 тиждень					0
		13 тиждень		0,5	10		10,5
		14 тиждень		0,5		5	5,5
		15 тиждень		0,5			0,5
		16 тиждень		0,5	10		10,5
		17 тиждень					40
Усього		3	7	40	10	40	100

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
82 – 89	B	добре	
74 – 81	C		
64 – 73	D	задовільно	
60 – 63	E		
35 – 59	FX	незадовільно	не зараховано
1 – 34	F		

5. Рекомендована література

5.1 Основна

1. Дейт К. Дж. Введение в системы баз данных / К. Дж. Дейт ; пер. с англ. – М. : ИД "Вильямс", 2005. – 1328 с.
2. Редмонд Э. Семь баз данных за семь недель. Введение в современные базы данных и идеологию NoSQL / Э. Редмонд, Дж. Р. Уилсон. ; пер. с англ. А. А. Слинкин. – М. : ДМК Пресс, 2013. – 384 с.
3. Тарасов С. В. СУБД для программиста. Базы данных изнутри / С. В. Тарасов. – М. : СОЛОН-Пресс, 2015. – 320 с.
4. Уайт Т. Nadoor: Подробное руководство / Т. Уайт. – СПб. : Питер, 2013. – 672 с.
5. MySQL. Оптимизация производительности / Б. Шварц, П. Зайцев, В. Ткаченко и др. – 2-е изд. – СПб. : Символ-Плюс, 2010. – 832 с.
6. Unix и Linux. Руководство системного администратора / [Э. Немет, Г. Снайдер, Т. Хейн и др.] – М. : ИД "Вильямс", 2012. – 1312 с.
7. Бьюли А. Изучаем SQL / А. Бьюли ; пер. с англ. – СПб. : Символ-Плюс, 2007. – 312 с.
8. Бэнкер К. MongoDB в действии / К. Бэнкер. ; пер. с англ. А. А. Слинкина – М. : ДМК Пресс, 2012. – 394 с.
9. Грофф Дж. Р. SQL: полное руководство / Дж. Р. Грофф, П. Н. Вайн-берг, Э. Дж. Оппель ; пер. с англ. – 3-е изд. – М. : ИД "Вильямс", 2015. – 960 с.
10. Уорсли Дж. PostgreSQL. Для профессионалов / Дж. Уорсли, Дж. Дрейк. – СПб. : Питер, 2003. – 496 с.
11. Фаулер М. NoSQL: новая методология разработки нереляционных баз данных / М. Фаулер, П. Дж. Садаладж. ; пер. с англ. – М. : ИД "Вильямс", 2013. – 192 с.
12. Monitoring with Ganglia / M. Massie, B. Li, B. Nicholes, V. Vuksan. – O'Reilly Media, Inc. – Sebastopol, CA. – 2012. – 256 p.
13. Wojciech K. Learning Nagios 4 / K. Wojciech. – Packt Publishing. – Birmingham. – 2014. – 400 p.

5.2 Інформаційні ресурси

14. Васильев А. Ю. Работа с PostgreSQL: настройка и масштабирование [Электронный ресурс] / А. Ю. Васильев. – Режим доступа : <http://postgresql.leopard.in.ua/>.
15. Гудвилл Дж. Memcached и Grails. Часть 1: Установка и применение memcached [Электронный ресурс] / Дж. Гудвилл. – IBM developerWorks. – Режим доступа : <http://www.ibm.com/developerworks/ru/library/j-memcached1/>.

16. Как устроена apache cassandra [Электронный ресурс] / Хабра-хабр. – Режим доступа : <http://habrahabr.ru/post/155115/>.

17. Климонтович В. Apache Hadoop (ADD-2010) [Электронный ресурс] / В. Климонтович.– Режим доступа : [http://lib.custis.ru/Apache_Hadoop_\(Владимир_Климонтович_на_ADD-2010\)](http://lib.custis.ru/Apache_Hadoop_(Владимир_Климонтович_на_ADD-2010)).

18. Лекции Техносферы. 2 семестр. Методы распределенной обработки больших объемов данных в Hadoop [Электронный ресурс] / Блог компании Mail.Ru Group. – Режим доступа : <http://habrahabr.ru/company/mailru/blog/258045/>.

19. Лекции Технопарка. 3 семестр. Проектирование высоконагруженных систем [Электронный ресурс] / Блог компании Mail.Ru Group. – Режим доступа : <http://habrahabr.ru/company/mailru/blog/254843/>.

20. Перера С. Погружение в СУБД Apache Cassandra [Электронный ресурс] / С. Перера. – IBM developerWorks. – Режим доступа : <https://www.ibm.com/developerworks/ru/library/os-apache-cassandra/>.

21. Распределенные базы и хранилища данных : Электронный учебник / Н. Аносова, О. Бородин, Е. Гаврилов и др. – НОУ "ИНТУИТ" [Электронный ресурс]. – Режим доступа : <http://www.intuit.ru/studies/courses/1145/214/info>.

22. Распределенные файловые системы. Технологии хранения и обработки больших объемов данных / Computer Science Center [Электронный ресурс]. – Режим доступа : <https://compscicenter.ru/courses/big-data/2015-spring/classes/1117/>.

23. Распределенная файловая система Ceph FS за 15 минут / Хабра-хабр [Электронный ресурс]. – Режим доступа : <http://habrahabr.ru/post/179823/>.

24. Руководство пользователя Zabbix [Электронный ресурс]. – Режим доступа : <http://www.zabbix.com/ru/documentation.php>.

25. Тестируем распределение контента в GlusterFS / Хабрахабр [Электронный ресурс]. – Режим доступа : <http://habrahabr.ru/post/251931/>.

26. Установка и настройка распределенной файловой системы GlusterFS в ОС CentOS 7.0 / Блог OS CONFIG [Электронный ресурс]. – Режим доступа : http://osc.dondub.com/articles/2015/04/article_852.

27. Depardon B. Analysis of Six Distributed File Systems / B. Depardon, G. Le Mahec, C. Séguin. – Research Report. – 2013. – 44 p. [Electronic re-source]. – Access mode : <https://hal.archives-ouvertes.fr/hal-00789086>.

28. FreeNAS User Guide [Electronic resource]. – Access mode : <http://doc.freenas.org/>.

29. Memcached и PHP ликбез [Электронный ресурс] / Хабрахабр. – Режим доступа : <http://habrahabr.ru/post/108274/>.

30. Алексієв В. О. Застосування GRID-технології у транспортному ВНЗ : навч.-метод. посіб. / В. О. Алексієв. – Х. : ХНАДУ, 2008. – 208 с.

31. Методы и модели планирования ресурсов в GRID-системах : монография / В. С. Пономаренко, С. В. Листровой, С. В. Минухин и др. ; Хар. нац. экон. ун-т. – Х. : ИД "ИНЖЭК", 2008. – 407 с.

32. Методи та моделі розроблення комп'ютерних систем і мереж : монографія / В. С. Пономаренко, С. В. Мінухін, С. В. Кавун та ін. – Х. : Вид. ХНЕУ, 2008. – 315 с.