

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Факультет електроніки та комп'ютерних технологій
Кафедра системного проектування

Затверджено

на засіданні
кафедри системного проектування
факультету електроніки та комп'ютерних
технологій
Львівського національного університету
імені Івана Франка
(протокол № 1 від 30.08.2022 р.)

Завідувач кафедри,



Роман ШУВАР

Силабус з навчальної дисципліни
«Нереляційні та розподілені бази даних»,
що викладається в межах ОПП «Інженерія програмного забезпечення»
першого (бакалаврського) рівня вищої
освіти для здобувачів з спеціальності
121 «Інженерія програмного забезпечення»

Львів 2022

Назва дисципліни	Нереляційні та розподілені бази даних
Адреса викладання дисципліни	м. Львів, вул. Драгоманова, 50
Факультет та кафедра, за якою закріплена дисципліна	Факультет електроніки та комп'ютерних технологій, кафедра системного проектування
Галузь знань, шифр та назва спеціальності	121 Інженерія програмного забезпечення
Викладачі дисципліни	Ляшкевич Марія Юріївна, асистент кафедри системного проектування
Контактна інформація	maria.liashkevych@lnu.edu.ua , https://electronics.lnu.edu.ua/employee/liashkevych-m-yu
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекційних занять (за попередньою домовленістю). Також можливі он-лайн консультації через MS Teams або систему електронного навчання Moodle. Для погодження часу онлайн консультацій слід писати на електронну пошту викладача.
Сторінка дисципліни	https://moodle.elct.lnu.edu.ua/course/view.php?id=256
Інформація про дисципліну	Дисципліна “ Нереляційні та розподілені бази даних ” є дисципліною з циклу професійної та практичної підготовки спеціальності 121 Програмна інженерія для освітньої програми «Інженерія програмного забезпечення», яка викладається в 6 семестрі в обсязі 4 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено таким чином, щоб оволодіти базовими поняттями, пов'язаними з організацією та використанням розподілених технологій, управлінні ресурсами віддалених розподілених систем, баз та сховищ даних, використанням технологій розподілених обчислень. З цією метою у дисципліні представлено як огляд базових понять та інструментів розподілених технологій, так і засобів, які потрібні для вирішення типових завдань збереження та опрацювання даних таких як розподілені бази та сховища даних, розроблення програм та програмних інтерфейсів для розподілених додатків і баз даних.
Мета та цілі дисципліни	Метою вивчення дисципліни “ Нереляційні та розподілені бази даних ” є надання поглиблених знань та практичних навичок щодо теорії та практики побудови й використання розподілених систем і паралельних обчислень, формування

	<p>системи теоретичних знань і набуття практичних умінь та 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.</p> <p>Цілями дисципліни “ Нереляційні та розподілені бази даних ” є засвоєння методів створення розподілених систем та технології їх проектування, наповнення даними та підтримання в робочому стані, вивчення методів і засобів паралельних обчислень та їх застосувань на практиці, а також сформувані знання про:</p> <ul style="list-style-type: none"> - особливості та характерні ознаки звичайного хостингу веб-ресурсів, оренди віртуальних приватних машин та систем хмарних обчислень; - програмні рішення для серверних систем віртуалізації та комплексні рішення, що здатні сформувані приватне хмарне середовище підприємства чи корпорації; - основні принципи побудови розподілених файлових сховищ даних та загальні властивості масштабування баз даних; - концепції комп'ютерної реалізації моделей предмету дослідження на основі алгоритмічного, структурного, об'єктно-зорієнтованого, компонентного, аспектно-орієнтованого, сервіс-орієнтованого, мультиагентного та інших сучасних підходів, використовувати концепції паралельної обробки інформації; - задачі автоматичного проектування систем управління, створення та випробування автоматичних систем управління.
<p>Література для вивчення дисципліни</p>	<p>Основна література:</p> <ol style="list-style-type: none"> 1. Павленко Л. А. Корпоративні інформаційні системи: Навчальний посібник./ Л. А. Павленко - Харків: ВД "ІНЖЕК", 2005. – 260 2. Катренко А.В., Системний аналіз об'єктів та процесів комп'ютеризації : Навчальний посібник./А.В. Катренко - Львів: "Новий світ-2000".-2003.-424с. 3. Google file system. - [Електронний ресурс]. - Режим доступу: https://docplayer.net/10419940-The-google-file-system.html 4. SAS, NAS, SAN Past, present, and future. - [Електронний ресурс]. - Режим доступу: http://www.imexresearch.com/pdfs/sasnassan.pdf 5. MySQL Cluster Manager 8.0.31 User Manual. - [Електронний ресурс]. - Режим доступу: https://downloads.mysql.com/docs/mysql-cluster-manager-1.4-en.a4.pdf 6. Query-By-Example (QBE). - [Електронний ресурс]. - Режим доступу: https://link.springer.com/chapter/10.1007/978-3-642-58763-4_10 7. Kristina Chodorow. Scaling MongoDB: O'Reilly, 2011. - 58 p. 8. Alex Holmes. Hadoop in Practice: Manning Publications, 2012. - 537 p. - Режим доступу: https://ia600201.us.archive.org/7/items/HadoopInPractice/Hadoop%20in%20Practice.pdf 9. Apache HBase Team. Apache HBase™ Reference Guide. - [Електронний ресурс]. - Режим доступу: https://hbase.apache.org/apache_hbase_reference_guide.pdf 10. Google. Cloud Bigtable. - [Електронний ресурс]. - Режим доступу: https://cloud.google.com/bigtable 11. Nagios. - [Електронний ресурс]. - Режим доступу: https://www.tutorialspoint.com/nagios/nagios_tutorial.pdf 12. Icinga2 open source monitoring. - [Електронний ресурс]. - Режим доступу: https://docplayer.net/11107242-Icinga2-open-source-monitoring.html
<p>Обсяг курсу</p>	<p>64 годин аудиторних занять. З них 32 години лекцій, 32 годин</p>

	лабораторних робіт.
Очікувані результати навчання	<p>Після завершення цього курсу студент буде:</p> <ul style="list-style-type: none"> - мати здатність до абстрактного мислення, аналізу та синтезу; - мати здатність застосовувати знання у практичних ситуаціях; - мати здатність вчитися і оволодівати сучасними знаннями; - мати здатність до пошуку, оброблення та аналізу інформації з різних джерел; - мати здатність розробляти архітектури, модулі та компоненти програмних систем; - мати здатність використовувати для розробки програмного забезпечення перспективні технології, зокрема, системи штучного інтелекту, тощо; - володіння методами, серверними технологіями та інструментальними засобами проектування веб-застосунків; - мати здатність здійснювати розробку програмного забезпечення використовуючи різні парадигми програмування і розподілених баз даних; - знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення; - знати мови запитів QBE та SQL; - уміти керувати транзакціями у розподілених базах даних та у сховищах великих даних; - знати концепцію проектування розподілених баз даних та уміти проектувати інформаційні системи на основі розподілених баз даних.
Ключові слова	Розподілені системи, розподілені та паралельні обчислення, розподілені бази даних, кластери даних, бази даних, великі дані, сховища даних.
Формат курсу	Очний. Проведення лекцій, практичних робіт та консультації для кращого розуміння тем.
Теми	Див. СХЕМА КУРСУ
Підсумковий контроль, форма	Залік в кінці семестру
Пререквізити	Для вивчення курсу студенти потребують базових знань з дисциплін: "Комп'ютерні мережі" та "Бази даних".
Навчальні методи та Презентація, лекції, лабораторні роботи, обговорення, дискусія. техніки, які будуть використовуватися під час викладання курсу	Презентація, лекції, лабораторні роботи, обговорення, дискусія.

Необхідне обладнання	Мультимедіа, платформа Moodle, комп'ютерне програмне забезпечення
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться упродовж семестру за 100-бальною шкалою, де враховано бали за два контрольні заміри по 35 балів за кожний модуль та 30 балів за складання заліку. Бали нараховуються за видами робіт з співвідношенням:</p> <ul style="list-style-type: none"> • контрольні заміри (2 модулі): 70% семестрової оцінки; максимальна кількість балів 70, а саме: <ul style="list-style-type: none"> - лабораторні роботи: 60% оцінки контрольного заміру; максимальна кількість балів 42 (8 лабораторних робіт). - лекційний матеріал: 40% оцінки контрольного заміру; максимальна кількість балів 28 (2 тести по 14 балів кожний). • залік: 30% семестрової оцінки, максимально 30 балів. <p>Оцінки за лабораторні заняття розподіляються наступним чином: виконання лабораторних завдань – 60 %, відповіді на запитання викладача по темі заняття – 40 %.</p> <p>Академічна доброчесність: Очікується, що лабораторні та контрольні роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в роботі студента є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції і лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття. Студенти зобов'язані дотримуватися усіх термінів визначених для виконання усіх видів робіт, передбачених курсом.</p> <p>Література. Уся література, яку студенти не зможуть знайти самостійно, буде надана викладачем виключно в освітніх цілях без права її передачі третім особам. Студенти заохочуються до використання також й іншої літератури та джерел, яких немає серед рекомендованих.</p> <p>Політика виставлення балів. Враховуються бали набрані на поточному тестуванні, самостійній роботі та бали підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізнь на заняття; користування мобільним телефоном, планшетом чи іншими мобільними пристроями під час заняття в цілях не пов'язаних з навчанням; списування та плагіат; несвоєчасне виконання поставленого завдання. Жодні форми порушення академічної доброчесності не толеруються.</p>
Питання до контрольних робіт	Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт розміщені на веб-

	сторінці курсу (система електронного навчання Moodle)
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література. Ресурси в Інтернеті	Завдання (лаб. робота), год	Термін виконання, тиж.
1	Паралельні обчислювальні системи. Поняття паралельної обчислювальної системи. Паралельна обробка. Види паралелізму: багатопроцесорна обробка, конвейерна обробка, векторна обробка, векторно-конвейерна обробка. Паралельні обчислювальні системи. Типи паралелізму: Паралелізм на рівні бітів. Паралелізм на рівні інструкцій. Паралелізм даних Паралелізм завдань (багатопоточність).	Лекція	1 - 5	Планування завантаження багатопроцесорного конвеєра, 2 год.	2
2	Принципи побудови розподіленої системи. Апаратна побудова розподілених систем. Програмна побудова розподілених систем. Види паралельної взаємодії. Класифікація архітектур паралельних обчислювальних систем: Класифікація Флінна. Класифікація Фенга. Класифікація Хокні. Класифікація Скїллікорна. Поняття та склад проміжного програмного забезпечення грід-систем. Промислові грід-системи, засоби доступу та запуск завдань.	Лекція	1 - 5	Планування конфліктів по даних у багатопотокових конвеєрах даних, 4 год	3
3	Принципи розробки та етапи проектування паралельних	Лекція	1 - 5		4

	<p>методів. Аналіз обчислювальних систем і декомпозиція їх на частини (підзадачі), які можуть бути реалізовані незалежно одна від одної. Виділення для сформованого набору під задач інформаційних взаємодій, що відбуваються в процесі вирішення поставленої задачі. Визначення обчислювальної системи, необхідної для вирішення задачі. Розподілення набору підзадач між процесорами системи.</p>				
4	<p>Моделювання паралельних програм. Представлення паралельного методу на стадії проектування графом «підзадачі – повідомлення». Модель для опису паралельної програми на стадії виконання у вигляді графа «процеси – канали». Особливості застосування моделі «процеси – канали». Синхронізація процесів та даних в паралельних програмах. Паралелізм задач та паралелізм даних. Виділення інформаційних залежностей. Принципи організації оброблення даних в розподілених системах. Концепція машин потоків даних.</p>	Лекція	1 - 5	Планування конвеєрів даних для систем паралельної обробки даних, 4 год.	5
5	<p>Архітектура GFS. Відмінності архітектури GFS від інших розподілених файлових систем. Особливості побудови архітектури GFS. Операції GFS Функції операцій GFS. Огляд переваг операцій GFS над іншими файловими системами.</p>	Лекція	6		6
6	<p>Розподілені файлові сховища даних на базі технологій SAN та NAS. Масштабування систем</p>	Лекція	7, 8	Організація MySQL кластеру засобами AWS і	7

	збереження даних на базі реляційних СУБД. Кластери на базі MySQL та PostgreSQL.			DataBricks, 4 год.	
7	Мови запитів QBE та SQL. Опис запитів мовою QBE. Вибірка даних з умовою. Базові оператори мови SQL та особливості їх запису. Формування запитів мовою SQL. Вибірка рядків конструкцією WHERE. Сортування результатів (конструкція ORDER BY). Вкладені запити (підзапити).	Лекція	8, 9		8
8	Управління транзакціями. Поняття та властивості транзакції. Порядок виконання операції транзакції. Проблеми управління паралельним доступом. Впорядкування та відновлення транзакцій. Методи управління паралельним доступом. Методи блокувань. Методи обробки взаємоблокувань. Методи управління паралельним доступом з використанням часових відміток. Оптимістичні методи впорядкування транзакцій. Механізми відновлення бази даних. Функції СКБД по відновленню бази даних. Журнал транзакцій. Створення контрольних точок. Метод відкладеного оновлення. Метод поточного оновлення. Метод тіньового сторінкового обміну.	Лекція	3, 5, 8	Маштабування MongoDB на AWS, 4 год.	9
9	Технології нереляційних СУБД у розподілених сховищах даних. Особливості систем MongoDB, CouchDB та Redis. Маштабування MongoDB.	Лекція	10		10
10	Концепція розподілених баз даних. Основні поняття. Відмінності між розподіленими	Лекція	5, 11, 12	Проектування розподіленої бази даних на основі	11

	<p>системами баз даних, засобами розподіленої обробки даних та паралельними системами баз даних. Класифікація розподілених баз даних. Переваги та недоліки системи керування розподіленими базами даних. Принципи створення розподілених БД. Проектування розподіленої БД. Фрагментація даних. Реплікація даних. Розміщення даних. Методологія проектування розподілених БД.</p>			заданих вимог, 4 год.	
11	<p>Управління паралельним доступом у розподільному середовищі. Протоколи блокування для управління паралельним виконанням у розподіленій базі даних. Усунення взаємних блокувань у розподіленому середовищі. Особливості використання часових відміток для управління паралельним виконанням у розподіленій базі даних. Відновлення розподілених баз даних. Особливості відновлення розподілених баз даних. Протокол двофазної фіксації транзакцій (2PC). Неблокуючий протокол трифазної фіксації транзакцій (3PC). Протоколи відновлення. Протоколи аварійного завершення.</p>	Лекція	11, 12		12
12	<p>Застосування та особливості проектування рішень на базі розподілених сховищ даних. Поняття про великі дані. Архітектура HDFS. Відмінності архітектури HDFS від інших розподілених файлових систем. Особливості побудви архітектури HDFS. Операції HDFS. Функції операцій HDFS. Огляд переваг операцій HDFS</p>	Лекція	12	Робота з MapReduce. Створення та старт роботи, 4 год.	13

	над іншими файловими системами.				
13	Особливості розроблення веб-додатків і веб-сервісів із застосуванням технологій розподілених сховищ даних. Архітектура HBase. Відмінності архітектури HBase від інших розподілених файлових систем. Особливості побудови архітектури HBase. Операції HBase. Функції операцій HBase. Огляд переваг операцій HBase над іншими файловими системами.	Лекція	11		14
14	Застосування технологій розподілених сховищ даних у хмарних обчисленнях. Місце та значення технологій розподілених сховищ даних під час оброблення великих масивів даних. Перспективи розвитку систем та технологій розподілених сховищ даних.	Лекція	12	Робота з Google BigTable, 6 год.	15
15	Архітектура Google BigTable. Відмінності архітектури Google BigTable від інших розподілених файлових систем. Операції та функції операцій Google BigTable. Переваги операцій Google BigTable над іншими файловими системами.	Лекція	9		16
16	Моніторинг стану розподілених обчислювальних систем і сховищ даних на базі відкритих програмних засобів Nagios та Icinga.	Лекція	11, 12		16