

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 28.08, 2023 р.)

Завідувач кафедри:



Роман ШУВАР

Силабус з навчальної дисципліни
“Нереляційні та розподілені бази даних”,
що викладається в межах ОПП
“ Високопродуктивний комп'ютинг ”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 121 – Інженерія програмного забезпечення

Львів 2023 р.

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	12 Інформаційні технології 121 Інженерія програмного забезпечення (ВПК)
Викладачі дисципліни	Ляшкевич Василь Яремович, канд. тех. наук, доцент
Контактна інформація	vasyl.lyashkevych@lnu.edu.ua , https://electronics.lnu.edu.ua/employee/liashkevych-v-ya
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=317
Інформація про дисципліну	Дисципліна “Нереляційні та розподілені бази даних” є вибірковою дисципліною з циклу професійної та практичної підготовки спеціальності 121 Інженерія програмного забезпечення для освітньої програми «Високопродуктивний комп'ютинг», яка викладається в 6 семестрі в обсязі 4,5 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Навчальна дисципліна забезпечує оволодіння базових понять нереляційних баз даних, сховищ даних, проектування розподіленого середовища баз та сховищ даних, управлінні ресурсами віддалених розподілених систем, методах завантаження, трансформації та зберігання даних, оцінці якості ресурсів 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 нереляційних баз даних, питання їх масштабування, проектування розподілених систем і паралельних обчислень на основі сховищ даних, формування системи теоретичних знань і набуття практичних умінь та навичок щодо застосування, налагодження й адміністрування систем на базі технологій розподілених сховищ даних та проектування

	<p>відповідних надійних та економічно привабливих систем для збереження великих об'ємів даних. З цією метою у дисципліні представлено як огляд базових понять та інструментів розподілений 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.</p> <p>Цілями дисципліни 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"> - нереляційні бази даних та їх властивості і характеристики, засоби проектування та обслуговування нереляційних баз та сховищ даних; - основні принципи побудови розподілених файлових сховищ даних та загальні властивості масштабування баз даних.
<p>Література для вивчення дисципліни</p>	<p>Основна література:</p> <ol style="list-style-type: none"> 1. Dan McCreary, Ann Kelly. Making Sense of NoSQL: Manning, 2014. - 314 p. 2. Gaurav Vaish. Getting Startged with NoSQL: Packt Publishing, 2013. - 142 p. 3. Michael Armbrust. Makeing Apache Spark better with Delta Lake: Databricks, 2020. - 399 p. 4. Gerardus Blokdyk. Databricks A complete Guide, 2021. - 205 p. - [Режим доступу]: https://www.everand.com/book/487839900/Databricks-A-Complete-Guide-2021-Edition 5. Tom White. Hadoop: The definitive Guide: O'Reilly, 2015. - 805 p. 6. Michael Collier, Robin Shashan. Fundamentals of Azure. Second Edition. Microsoft Azure Essentials: Microsoft, 2016. - 263p. 7. Michael Crump, Chris Pietschmann, Vahe Minasyan. The Developer's Guide to Azure. Microsoft Press, A division of Microsoft Corporation One Microsoft Way, Redmond, Washington 98052-6399. 8. Kai Hwang, Min Chen. Big-Data Analytics for Cloud, IoT and Cognitive Computing: Wiley, 2017. - 428 p. 9. Designing Distributed System. - [Режим доступу]: https://azure.microsoft.com/mediahandler/files/resourcefiles/designing-distributed-systems/Designing_Distributed_Systems.pdf 10. Kristina Chodorow. Scaling MongoDB: O'Reilly, 2011. - 58 p. 11. Query-By-Example (QBE). - [Електронний ресурс]. - Режим доступу: https://link.springer.com/chapter/10.1007/978-3-642-58763-4_10 12. Google file system. - [Електронний ресурс]. - Режим доступу: https://docplayer.net/10419940-The-google-file-system.html 13. Google. Cloud Bigtable. - [Електронний ресурс]. - Режим доступу: https://cloud.google.com/bigtable 14. Rik Van Bruggen. Learning Neo4j: Packt Publishing, 2014. - 222 p. 15. 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 16. Alex Holmes. Hadoop in Practice: Manning Publications, 2012. - 537 p. - Режим доступу: https://ia600201.us.archive.org/7/items/HadoopInPractice/Hadoop%20in%20Practice.pdf 17. Neha Narkhede. Kafka: The Definitive Guide: O'Reilly, 2017. - 322 p. - [Електронний ресурс]. - Режим доступу: https://book.huihoo.com/pdf/confluent-kafka-definitive-guide-complete.pdf 18. Bas Harenslak, Julian de Rooter. Data Pipelines with Apache Airflow: Manning Publications, 2021. - 482 p. - [Електронний ресурс]. - Режим доступу: https://biconsult.ru/files/Data_warehouse/Bas_P_Harenslak%2C_Julian_Rutger_de_Rooter_Data_Pipelines_with_Apache.pdf 19. Apache HBase Team. Apache HBase™ Reference Guide. - [Електронний ресурс]. - Режим доступу: https://hbase.apache.org/apache_hbase_reference_guide.pdf 20. Informatica PowerCenter Designer Guide 10.4.0: Informatica, 2019. - 286. - [Електронний ресурс]. - Режим доступу: https://docs.informatica.com/content/dam/source/GUID-B54ED1F4-60B8-4F11-8E22-48C4BECE109A/27/en/PC_1040_DesignerGuide_en.pdf 21. Joshua N.Milligan. Learning Tableau 2019. Third Edition: Packt Publications, 2019. - 808 p. - [Електронний ресурс]. - Режим доступу:

	<p>http://projanco.com/Library/Learning%20Tableau%202019%20Tools%20for%20Business%20Intelligence.%20data%20prep.%20and%20visual%20analytics.pdf</p> <p>22. Nagios. - [Електронний ресурс]. - Режим доступу: https://www.tutorialspoint.com/nagios/nagios_tutorial.pdf</p> <p>23. Icinga2open source monitoring. - [Електронний ресурс]. - Режим доступу: https://docplayer.net/11107242-Icinga2-open-source-monitoring.html</p>
Обсяг курсу	Кількість кредитів ЄКТС: 4,5 (135 год), з них: 64 годин аудиторних занять (лекції: 32 год, лабораторні: 32 год.) та 71 год. самостійної роботи.
Очікувані результати навчання	<p>Після вивчення даного курсу здобувачі набудуть таких Загальних та Фахових компетентностей та Програмних результатів навчання:</p> <p>ЗК01. Здатність до абстрактного мислення, аналізу та синтезу.</p> <p>ЗК02. Здатність застосовувати знання у практичних ситуаціях.</p> <p>ЗК03. Здатність спілкуватися державною мовою як усно, так і письмово.</p> <p>ЗК04. Здатність спілкуватися іноземною мовою як усно, так і письмово.</p> <p>ЗК05. Здатність вчитися і оволодівати сучасними знаннями.</p> <p>ЗК06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.</p> <p>ЗК07. Здатність працювати в команді.</p> <p>ЗК09. Прагнення до збереження навколишнього середовища.</p> <p>ФК20. Здатність застосовувати фундаментальні і міждисциплінарні знання для успішного розв'язання завдань інженерії програмного забезпечення.</p> <p>ФК22. Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя.</p> <p>ПРН02. Знати кодекс професійної етики, розуміти соціальну значимість та культурні аспекти інженерії програмного забезпечення і дотримуватись їх в професійній діяльності.</p> <p>ПРН10. Проводити передпроектне обстеження предметної області, системний аналіз об'єкта проектування.</p> <p>ПРН13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.</p> <p>ПРН14. Застосовувати на практиці інструментальні програмні засоби доменного аналізу, проектування, тестування, візуалізації, вимірювань та документування програмного забезпечення.</p> <p>ПРН16. Мати навички командної розробки, погодження, оформлення і випуску всіх видів програмної документації.</p> <p>ПРН22. Знати та вміти застосовувати методи та засоби управління проектами.</p> <p>ПРН23. Вміти документувати та презентувати результати розробки програмного забезпечення.</p> <p>ПРН28. Зберігати та примножувати цивілізаційні цінності і досягнення суспільства, діяти соціально відповідально та</p>

	свідомо, зберігати навколишнє середовище, знати правила ведення здорового способу життя і надання першої медичної допомоги.
Ключові слова	Бази даних, нереляційні бази даних, розподілені системи, розподілені бази даних, кластери даних, сховища даних, великі дані, конвеєри обробки даних, якість ресурсів 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: “Вибрані розділи сучасних мов та засобів програмування” та “Бази даних”.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Лекції, презентація, виконання лабораторних робіт, обговорення, консультації для кращого розуміння тем, самостійна робота, робота в команді
Необхідне обладнання	Мультимедійне обладнання, комп’ютерний клас, програми та сервіси MS Teams, Moodle, Databricks, Hadoop, Apache Spark, PySpark, Airflow, Microsoft Azure, Azure Cosmos, Apache Kafka, Python
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.</p> <p>Бали нараховуються за такими видами робіт з наступним співвідношенням:</p> <ul style="list-style-type: none"> • контрольні заміри: 70% семестрової оцінки, максимальна кількість балів 70: <ul style="list-style-type: none"> - лабораторні роботи: 68,5% оцінки контрольного заміру; максимальна кількість балів 48 (16 лабораторних робіт по 3 бали за кожену роботу). - теоретичний матеріал: 31,5% оцінки контрольного заміру; максимальна кількість балів 22 (2 тести по 11 балів за кожний тест, де 1 тест складається із 22 запитань по 0.5 бала за одне питання). • залік: 30% семестрової оцінки, максимум 30 балів. <p>Оцінки за лабораторні заняття розподіляються наступним чином:</p> <p>3 (від 0 до 2-х балів за виконання разом в сумі від 0 до 1-го балу за тестування/опитування) – студент в повному обсязі володіє навчальним матеріалом, має повне розуміння</p>

розглянутої теми, надає правильні відповіді на запитання по темі, код програми функціонує відповідно до завдання;

2 (від 0 до 1-х балів за виконання разом в сумі від 0 до 1-го балу за тестування/опитування) – студент не досить добре розуміє розглянутий матеріал та написаний ним код програми, вагається та надає неточні/не конкретні відповіді на запитання по темі, код програми функціонує неточно, або з помірними недоліками;

1 (від 0 до 1-го балів за виконання разом в сумі від 0 до 1-го балу за тестування/опитування) - студент погано розуміє розглянутий матеріал та написаний ним код програми, код програми не функціонує належним чином;

0 (робота не виконана, тест не пройдено) - студент зовсім не засвоїв розглянутий матеріал, робота не виконана чи не відповідає темі чи не функціонує взагалі й тест не пройдено.

Оцінювання залікових питань:

10 балів - розглянута тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно, які містять аналіз і систематизацію, аргументовані висновки. Засвідчено глибоке володіння матеріалом. Наведені приклади коду повністю робочі та відповідають темі. Можуть бути присутні несуттєві помилки та невідповідності;

8 балів - відтворюється значна частина розглянутої теми. Виявлено знання і розуміння основних положень навчальної дисципліни, проте присутні неточності та/або невідповідності основній темі. Наведені приклади коду частково робочі, проте в загальному відповідають темі;

5 балів - відстежується загальне розуміння розглянутої теми. Виявлені множинні неточності та невідповідності, пояснення наведеного коду відсутні, код функціонує із значними неточностями (або відсутні приклади запуску коду на виконання взагалі);

3 бали – студент погано розуміє розглянуту тему. Виявлені суттєві неточності та невідповідності. Наведені приклади коду з суттєвими недоліками, або не відповідають темі;

Менше 3 балів – студент взагалі не розуміє розглянуту тему. Тему не розкрито, кількість викладеного матеріалу не відповідає загальним нормам обраного виду роботи. Наведений код не робочий, або відсутній як такий.

Академічна доброчесність: Очікується, що лабораторні та контрольні роботи студентів будуть їх оригінальними дослідженнями чи міркуваннями. Відсутність посилань на використані джерела, фабрикавання 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.

Відвідання занять є важливою складовою навчання.

	<p>Очікується, що всі студенти відвідають усі лекції і лабораторні заняття курсу. Студенти мають інформувати викладача про неможливість відвідати заняття.</p> <p>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> <p>Критерії оцінювання результатів неформальної освіти:</p> <p>Нарахування балів відбувається за написання студентом тез доповідей на конференціях, наукових статей, участь у діяльності наукових гуртків, участь у наукових семінарах та круглих столах, конкурсах, участь у заходах неформальної освіти за отримання сертифікатів про проходження навчання на різних освітніх платформах (Coursera, Prometheus тощо), курсах на провідних ІТ компаніях за тематикою навчальної дисципліни. Кількість балів визначається відсотком покриття результатів відповідної активності до вимог результатів навчання з навчальної дисципліни.</p>
<p>Питання до контрольних робіт</p>	<p>Перелік питань та завдань для проведення підсумкової оцінки знань певних тем до контрольних робіт:</p> <ol style="list-style-type: none"> 1. Концепція розподілених систем (<i>Приклади типових розподілених систем. Компоненти розподілених систем. Інтернет та інтранет. Використання пристроїв у розподілених системах. Обмін ресурсів в WWW. Веб-сервери та веб-браузери. HTTP/HTTPS протоколи. Програмні та апаратні сервісні рівні. Клуєнт-серверні моделі. Веб-прохі сервери.</i>) 2. Веб-застосунки (<i>Веб-застосунки. Мобільні агенти. Комп'ютерні мережі. Синхронні та асинхронні розподілені системи. Події в системі. Порядок подій в реальному часі. Канали зв'язку. Питання захисту.</i>) 3. Поняття нереляційних баз даних (<i>Поняття NoSQL. Модель даних. Типи моделей даних. Логічна модель. Створення фізичної моделі даних. Фізична модель даних для</i>

реляційних баз даних. Інструменти для моделювання даних.)

4. MongoDB (Гнучке моделювання даних за допомогою MongoDB Atlas. Приклад моделювання бази даних. Запити і агрегування. Детальніше про мову запитів в MongoDB.)
5. MongoDB SQL (Розуміння зіставлення MongoDB та SQL. MongoDB SQL: термінологія. MongoDB SQL: Виконувани файли бази даних. MongoDB SQL: команди. Приклад CRUD функцій в MongoDB. Побудова запитів в MongoDB.)
6. Графові бази даних (Що таке є мова запитів GraphQL. Переваги GraphQL. GraphQL в порівнянні REST. Схеми GraphQL. Схема визначення. Вирішувачі (Resolvers). Запит даних. Повернення даних. Apollo GraphQL. Приклад з NodeJS та Express. Приклад запиту для React client.)
7. Резидентна система управління базами даних Redis (Розподілене сховище пар ключ-значення. Конфігурації Redis. Типи даних Redis. Команди з використанням ключа. Redis server. Redis on python. Backup. Безпека Redis. Redis benchmark. Транзакції Redis. Redis pipelining. Redis на docker.)
8. Графова база даних Neo4j (Graph Databases. Приклади графових баз даних. Порівняння графових та реляційних баз даних. Графова база даних Neo4j. Neo4j браузер. Схема Neo4j. Структура зашифрованого запиту. Написання зашифрованих запитів.)
9. Робота з Neo4j (Схема Neo4j. Структура зашифрованого запиту. Написання зашифрованих запитів. Створення та запит до вершин. Встановлення взаємозв'язку між вершинами. Використання транзакційного зашифрованого HTTP end-point. Використовувані протоколи.)
10. Графові моделі Orient db (Огляд Orient db. Мульти-модель систем керування базами даних (СКДБ). Моделі документів. Графові моделі. Orient db синтаксис. Типи транзакцій. ETL (Extract, Transform, Load). Можливості Orient db. Використання Orient db.)
11. База даних HBase (Кластерна архітектура HBase. Відмінності архітектури HBase від інших розподілених файлових систем. Особливості побудови архітектури HBase. Hbase на Python. Операції HBase. Функції операцій HBase. Огляд переваг операцій HBase над іншими файловими системами.)
12. Особливості розроблення веб-додатків і веб-сервісів із застосуванням технологій розподілених сховищ даних (Поняття веб-додатку та веб-сервісу. Розподілені бази даних. Сховища даних. Розподілені технології. Особливості розподілених сховищ даних)

13. Концепція Hadoop (Знайомство з Hadoop. Історія Hadoop. Компоненти Hadoop. Вузли та демони Hadoop. Архітектура Hadoop. Hadoop характеристики. oogle File System (GFS). HDFS. Концепція та архітектура HDFS. Відмінності архітектури HDFS від інших розподілених файлових систем.)
14. Особливості HDFS (Особливості побудви архітектури HDFS. Операції HDFS. Функції операцій HDFS. Огляд переваг операцій HDFS над іншими файловими системами. Типи запитів. Мовна підтримка. MapReduce. Властивості MapReduce. Трекер задач. Hive. HiveQL. Hadoop Fea.)
15. Різновид реляційних мов запиту (Реляційні мови запитів. Запит як приклад (Query-by-Example). Опис запитів мовою QBE. Вибірка даних з умовою. Базові оператори мови SQL та особливості їх запису.)
16. Формування SQL запитів (Формування запитів мовою SQL. Вибірка рядків конструкцією WHERE. Сортування результатів (конструкція ORDER BY). Вкладені запити (підзапити). Запити для кількох зв'язків. Умовний ящик. Microsoft Access. Datalog.)
17. Знайомство з Apache Spark (Виклики та рішення. Що таке Apache Spark? Модель Spark. Потужний стек – гнучка розробка. Компоненти Apache Spark)
18. Spark SQL (Spark SQL. Інтерфейс програмування. Модель даних. Операції DataFrame. Запити рідних наборів даних. Функції, визначені користувачем. Оптимізація та виконання плану. Логічний план. Фізичний план. Фізичний план з предикатом Pushdown і Column Pruning. Генерація коду. Розширення. Spark MLlib конвеєр.)
19. Поняття хмарної платформи Snowflake (Поняття платформи даних. Переваги хмарних платформ даних. Використання платформ даних для бізнесу. Традиційні архітектури. Сучасна архітектура даних з Snowflake. Порівняльна характеристика Snowflake в порівнянні з іншими платформами.)
20. Хмарні сервіси з Snowflake (Архітектура Snowflake. Хмарні сервіси. Збільшення існуючих озер даних. Низька затримка. Транзакція перетворення, що масштабується. Безпечний доступ до даних. Інтеграція з Snowflake. Snowflake з Tableau. Робота з Snowflake.)
21. Оркестрування потоків операцій в Airflow (Поняття Apache Airflow. Основні відомості та призначення. Поняття прямих ациклічних графів (DAG). Робочий процес. Airflow веб-сервер. Запуск DAG. Оператори Airflow. Налаштування Airflow операторів. Airflow сенсори. Проектне навчання. Використання Python для програмування DAG. Панель керування Airflow.)

	<p>22. Оцінка якості даних (<i>Якість даних. Поняття контролю якості даних. Якість великих даних. Якість даних у потоках даних. Якість даних в хмарі. Профілювання даних. Статистика профілювання даних у хмарі. Правила специфікацій. Словники. Очищення даних. Верифікація даних. Процес якості даних. Компоненти для аналізу даних. Використання Python.</i>)</p> <p>23. Оцінка якості даних в Informatica (<i>Компоненти Informatica. Процес оцінки якості даних. Компоненти для аналізу даних. Використання Python. Використання агрегаторів та сортування. Табличні процесори.</i>)</p> <p>24. Робота з Tableau (<i>Поняття Tableau. Візуалізація даних. Тип діаграми та потоки інформаційної панелі. Попередня уважна обробка. Заголовок і підказка. Наступні кроки та додаткові ресурси. Вибір правильного графіку. Порівняння методів графічного представлення інформації. Розподіли. Взаємозв'язок. Панель інструментів. Шарування. Тестування. Попередня уважна обробка даних. Кольорові гами. Вибір стилю та кольору. Заголовок. Контекст. Tooltip.</i>)</p> <p>25. Масштабування даних (<i>Перша зустріч науки з великими даними. Поняття машинного навчання. Дані і знання. Великі масштаби даних. Надвеликі розміри моделей. Класичні алгоритми машинного навчання. Питання масштабування. Стратегія паралелізму. Використання MapReduce.</i>)</p>
<p>Опитування</p>	<p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p>

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література. Ресурси в Інтернеті	Завдання, год	Термін виконання, тиж.
1	Поняття конвеєрів даних. Поняття паралельної обчислювальної системи. Види паралелізму: багатопроцесорна обробка, конвейерна обробка, векторна обробка, векторно-конвейерна обробка. Типи паралелізму: Паралелізм на рівні бітів. Паралелізм на рівні інструкцій. Паралелізм даних. Паралелізм завдань. Поняття багатопоточності.	лекція	1 - 2	2	кінець поточного тижня
	Лаб 1. Реєстрація та знайомство з Databricks	лаб. робота	3-4	2	кінець поточного тижня
	Способи організації паралельної обчислювальної системи	сам. робота	1-2	4,4375	кінець поточного тижня
2	Принципи побудови розподіленої системи. Приклади типових розподілених систем. Компоненти розподілених систем. Інтернет та інтранет. Використання пристроїв у розподілених ситемах. Обмін ресурсів в WWW. Веб-сервери та веб-браузери. HTTP/HTTPS протоколи. Програмні та апаратні сервісні рівні. Клуєнт-серверні моделі. Веб-прохі сервери. Веб-застосунки. Мобільні агенти. Комп'ютерні мережі. Синхронні та асинхронні розподілені системи. Події в системі. Порядок подій в реальному часі. Канали зв'язку. Питання захисту.	лекція	1-5	2	кінець поточного тижня
	Лаб 2. Визначення ресурсів даних та послідовності їх опрацювання	лаб. робота	3-4, 6-7	2	кінець поточного тижня
	Програмна побудова розподілених систем.	сам. робота	8-9	4,4375	кінець поточного тижня
3	Основи нереляційних баз даних. Важко змінити таблиці та відносини. Різноманітність даних. Властивості ACID: Атомарність. Послідовність.	лекція	1-2, 8-9	2	кінець поточного тижня

	Ізоляція. Довговічність. Теорема CAP. Основи NoSQL. Огляд NoSQL. Характеристики баз даних NoSQL. Категорії бази даних NoSQL. Проблеми під час переходу від RDBMS до баз даних NoSQL.				
	Лаб 3. Реалізація конвеєру даних на Databriks	лаб. робота	3-4, 6-7	2	кінець поточного тижня
	Обливості проектування паралельних методів	сам. робота	1-2, 9	4,4375	кінець поточного тижня
4	Робота з MongoDB. Модель даних. Типи моделей даних. Логічна модель. Створення фізичної моделі даних. Фізична модель даних для реляційних баз даних. Інструменти для моделювання даних. Гнучке моделювання даних за допомогою MongoDB Atlas. Приклад моделювання бази даних. Запити і агрегування. Детальніше про мову запитів в MongoDB. Розуміння зіставлення MongoDB та SQL. MongoDB SQL: термінологія. MongoDB SQL: Виконувати файли бази даних. MongoDB SQL: команди. Приклад CRUD функцій в MongoDB. Побудова запитів в MongoDB.	лекція	1-2, 10	2	кінець поточного тижня
	Лаб 4. Робота з MongoDB	лаб. робота	6-7	2	кінець поточного тижня
	Методи синхронізації процесів та даних в паралельних програмах	сам. робота	1-2, 8-9	4,4375	кінець поточного тижня
5	Запити GraphQL. Що таке є мова запитів GraphQL. Переваги GraphQL. GraphQL в порівнянні REST. Схеми GraphQL. Схема визначення. Вирішувачі (Resolvers). Запит даних. Повернення даних. Apollo GraphQL. Приклад з NodeJS та Express. Приклад запиту для React client.	лекція	6, 10-11	2	кінець поточного тижня
	Лаб 5. Робота із запитамі GraphQL	лаб. робота	1-3, 5	2	кінець поточного тижня

	Особливості побудови архітектури GFS	сам. робота	12-13	4,4375	кінець поточного тижня
6	Резидентна система управління базами даних Redis. Розподілене сховище пар ключ-значення. Конфігурації Redis. Типи даних Redis. Команди з використанням ключа. Redis server. Redis on python. Backup. Безпека Redis. Redis benchmark. Транзакції Redis. Redis pipelining. Redis на docker.	лекція	1-2	2	кінець поточного тижня
	Лаб 4. Робота з Redis	лаб. робота	1-2, 4	2	кінець поточного тижня
	Розподілені файлові сховища даних	сам. робота	4-6, 8-10	4,4375	кінець поточного тижня
7	Робота з Neo4j. Graph Databases. Приклади графових баз даних. Порівняння графових та реляційних баз даних. Графова база даних Neo4j. Neo4j браузер. Схема Neo4j. Структура зашифрованого запиту. Написання зашифрованих запитів. Створення та запит до вершин. Встановлення взаємозв'язку між вершинами. Використання транзакційного зашифрованого HTTP end-point. Використовувані протоколи.	лекція	8, 9, 14	2	кінець поточного тижня
	Лаб 7. Побудова конвеєру машинного навчання на Spark	лаб. робота	3, 5	2	кінець поточного тижня
	Особливості мови запитів QBE	сам. робота	11	4,4375	кінець поточного тижня
8	Бази даних Hbase та Orient db. Особливості розроблення веб-додатків і веб-сервісів із застосуванням технологій розподілених сховищ даних. Кластерна архітектура HBase. Відмінності архітектури HBase від інших розподілених файлових систем. Особливості побудови архітектури HBase. Hbase на Python. Операції HBase. Функції операцій HBase. Огляд переваг операцій HBase над іншими файловими	лекція	3, 5, 8	2	кінець поточного тижня

	системами. Огляд Orient db. Мульти-модель систем керування базами даних (СКДБ). Моделі документів. Графові моделі. Orient db синтаксис. Типи транзакцій. ETL (Extract, Transform, Load). Можливості Orient db. Використання Orient db.				
	Лаб 8. Основи структурованої потокової подачі даних	лаб. робота	8-9, 14	2	кінець поточного тижня
	Особливості мови запитів SQL	сам. робота	15	4,4375	кінець поточного тижня
9	Концепція та компоненти Hadoop. Знайомство з Hadoop. Історія Hadoop. Компоненти Hadoop. Вузли та демони Hadoop. Архітектура Hadoop. Hadoop характеристики. oogle File System (GFS). HDFS. Концепція та архітектура HDFS. Відмінності архітектури HDFS від інших розподілених файлових систем. Особливості побудви архітектури HDFS. Операції HDFS. Функції операцій HDFS. Огляд переваг операцій HDFS над іншими файловими системами. Типи запитів. Мовна підтримка. MapReduce. Властивості MapReduce. Трекер задач. Hive. HiveQL. Hadoop Fea.	лекція	5, 16	2	кінець поточного тижня
	Лаб 9. Структурована потокова подача даних з Apache Kafka	лаб. робота	1-2, 8-9, 17	2	кінець поточного тижня
	Методи управління паралельним доступом до баз даних	сам. робота	9-10, 13	4,4375	кінець поточного тижня
10	Різновид реляційних мов запити. Реляційні мови запитів. Запит як приклад (Query-by-Example). 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). Запити для кількох зв'язків. Умовний ящик. Microsoft Access. Datalog.	лекція	5, 11, 12	2	кінець поточного тижня

	Лаб 10. Робота із дельта-форматом даних	лаб. робота	3, 9-10, 13	2	кінець поточного тижня
	Особливості розподіленими системами баз даних	сам. робота	9-11, 13-15	4,4375	кінець поточного тижня
11	Опрацювання даних із Spark SQL. Виклики та рішення. Що таке Apache Spark? Модель Spark. Потужний стек – гнучка розробка. Spark SQL. Інтерфейс програмування. Модель даних. Операції DataFrame. Запити рідних наборів даних. Функції, визначені користувачем. Оптимізація та виконання плану. Логічний план. Фізичний план. Фізичний план з предикатом Pushdown і Column Pruning. Генерація коду. Розширення. Spark MLlib конвеєр.	лекція	9, 11, 15	2	кінець поточного тижня
	Лаб 11. Знайомство з Azure Data Services	лаб. робота	6-7	2	кінець поточного тижня
	Резертування та відновлення у розподілених базах даних	сам. робота	1-2, 9	4,4375	кінець поточного тижня
12	Хмарна платформа даних Snowflake. Поняття платформи даних. Переваги хмарних платформ даних. Використання платформ даних для бізнесу. Традиційні архітектури. Сучасна архітектура даних з Snowflake. Порівняльна характеристика Snowflake в порівнянні з іншими платформами. Архітектура Snowflake. Хмарні сервіси. Збільшення існуючих озер даних. Низька затримка. Транзакція перетворення, що масштабується. Безпечний доступ до даних. Інтеграція з Snowflake. Snowflake з Tableau. Робота з Snowflake.	лекція	6-7	2	кінець поточного тижня
	Лаб 12. Налаштування та використання Azure Data Services	лаб. робота	6-7	2	кінець поточного тижня
	Особливості розподілених файлових систем	сам. робота	9-10	4,4375	кінець поточного тижня

13	<p>Оркестрування потоків операцій в Airflow. Поняття Apache Airflow. Основні відомості та призначення. Поняття прямих ациклічних графів (DAG). Робочий процес. Airflow веб-сервер. Запуск DAG. Оператори Airflow. Налаштування Airflow операторів. Airflow сенсори. Проектне навчання. Використання Python для програмування DAG. Панель керування Airflow.</p>	лекція	14, 18	2	кінець поточного тижня
	<p>Лаб 13. Проектування Delta Live Tables (DLT)</p>	лаб. робота	3	2	кінець поточного тижня
	<p>Особливості розподіленої файлової системи HDFS</p>	сам. робота	19	4,4375	кінець поточного тижня
14	<p>Призначення та обробка даних в Informatica. Якість даних. Поняття контролю якості даних. Якість великих даних. Якість даних у потоках даних. Якість даних в хмарі. Профілювання даних. Статистика профілювання даних у хмарі. Правила специфікацій. Словники. Очищення даних. Верифікація даних. Процес якості даних. Компоненти для аналізу даних. Використання Python.</p>	лекція	9, 20	2	кінець поточного тижня
	<p>Лаб 14. Побудова Delta Live Tables (DLT)</p>	лаб. робота	3	2	кінець поточного тижня
	<p>Особливості розподіленої файлової системи HBase</p>	сам. робота	19	4,4375	кінець поточного тижня
15	<p>Знайомство із Tableau. Вступ. Візуалізація даних. Тип діаграми та потоки інформаційної панелі. Попередня уважна обробка. Заголовок і підказка. Наступні кроки та додаткові ресурси. Вибір правильного графіку. Порівняння методів графічного представлення інформації. Розподіли. Взаємозв'язок. Панель інструментів. Шарування. Тестування. Попередня уважна обробка даних. Кольорові гами. Вибір стилю та кольору. Заголовок. Контекст. Tooltip.</p>	лекція	13, 21	2	кінець поточного тижня

	Лаб 15. Проектування SQL Warehouse на Microsoft Azure	лаб. робота	3-4, 6, 15, 18	2	кінець поточного тижня
	Операції та функції операцій Google BigTable	сам. робота	13	4,4375	кінець поточного тижня
16	Розподілені бази даних для машинного навчання на великих даних. Перша зустріч науки з великими даними. Поняття машинного навчання. Дані і знання. Великі масштаби даних. Надвеликі розміри моделей. Класичні алгоритми машинного навчання. Питання масштабування. Стратегія паралелізму. Використання MapReduce. Традиційна обробка даних. Паралелізм на рівні даних та моделей. Внутрішні властивості програм машинного навчання. GraphLab: Паралельне моделювання через графіки. Огляд Petuum. Складність розпаралелювання алгоритмів. Паралелізація з урахуванням структури (SAP). Динамічний планувальник з урахуванням структури (STRADS). Обчислювальна ефективність параметрів моделі.	лекція	1-4, 8-9, 16-17	2	кінець поточного тижня
	Лаб 16. Побудова SQL Warehouse на Azure	лаб. робота	3-4, 6-7, 15, 18	2	кінець поточного тижня
	Моніторинг стану розподілених обчислювальних систем	сам. робота	22-23	4,4375	кінець поточного тижня