

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

Затверджено

На засіданні кафедри СП

факультету електроніки та комп'ютерних
технологій

Львівського національного університету
імені Івана Франка

(протокол № 1 від 31 серпня 2021 р.)

Завідувач кафедри _____

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

Львів 2021

Назва дисципліни	Розподілені технології та 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 (ВПК)
Викладачі дисципліни	Ляшкевич Василь Яремович, 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=256
Інформація про дисципліну	Дисципліна “Розподілені технології та бази даних” є дисципліною з циклу професійної та практичної підготовки спеціальності 121 Програмна інженерія 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	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 “Розподілених технологій та баз даних” є надання поглиблених знань та практичних навичок щодо теорії та практики побудови й використання розподілених систем і паралельних обчислень, формування системи теоретичних знань і набуття практичних умінь та

	<p>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 “Розподілені технології та бази даних” є засвоєння методів створення розподілених систем та технологій їх проектування, наповнення даними та підтримання в робочому стані, вивчення методів і засобів паралельних обчислень та їх застосувань на практиці, а також сформулювати заняття 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 сформулювати 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. Michael Armbrust. Making 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: Willey, 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

	<p>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</p> <p>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</p> <p>19. Apache HBase Team. Apache HBase™ Reference Guide. - [Електронний ресурс]. - Режим доступу: https://hbase.apache.org/apache_hbase_reference_guide.pdf</p> <p>20. Informatica PowerCenter Designer Guide 10.4.0: Informatica, 2019. - 286. - [Електронний ресурс]. - Режим доступу: https://docs.informatica.com/content/dam/source/GUID-B/GUID-B54ED1F4-60B8-4F11-8E22-48C4BECE109A/27/en/PC_1040_DesignerGuide_en.pdf</p> <p>21. Joshua N.Milligan. Learning Tableau 2019. Third Edition: Packt Publications, 2019. - 808 p. - [Електронний ресурс]. - Режим доступу: 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. Icinga2 open source monitoring. - [Електронний ресурс]. - Режим доступу: https://docplayer.net/11107242-Icinga2-open-source-monitoring.html</p>
<p>Обсяг курсу</p>	<p>Кількість кредитів ЄКТС: 4,5 (135 год), з них: 64 годин аудиторних занять (лекції: 32 год, лабораторні: 32 год.) та 71 год. самостійної роботи.</p>
<p>Очікувані результати навчання</p>	<p>Після завершення цього курсу студент буде:</p> <ul style="list-style-type: none"> - ЗК1. Здатність до абстрактного мислення, аналізу та синтезу. - ЗК2. Здатність застосовувати знання у практичних ситуаціях. - ЗК4. Здатність спілкуватися іноземною мовою як усно, так і письмово. - ЗК05. Здатність вчитися і оволодівати сучасними знаннями. - ФК13. Здатність ідентифікувати, класифікувати та формулювати вимоги до програмного забезпечення. - ФК14. Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування. - ФК15. Здатність розробляти архітектури, модулі та компоненти програмних систем. - ФК22. Здатність накопичувати, обробляти та систематизувати професійні знання щодо створення і супроводження програмного забезпечення та визнання важливості навчання протягом всього життя. - ФК24. Здатність здійснювати процес інтеграції системи, застосовувати стандарти і процедури управління змінами для підтримки цілісності, загальної функціональності і надійності програмного забезпечення. - ФК25. Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення. - ФК27. Здатність розробляти високопродуктивні програмні комплекси для вирішення завдань наук про дані, систем штучного інтелекту, вбудованих та інших інноваційних систем.

	<ul style="list-style-type: none"> - ФК28. Володіння методами розроблення систем підвищеної продуктивності, серверними та розподіленими технологіями, інструментальними засобами проектування та розробки веб-застосовань і нових технологій. - ФК29. Здатність здійснювати розробку програмного забезпечення використовуючи різні методології та засоби програмування з метою забезпечення їх високої надійності та продуктивності в роботі. - ПРН01. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки. - ПРН09. Знати та вміти використовувати методи та засоби збору, формулювання та аналізу вимог до програмного забезпечення. - ПРН10. Проводити передпроектне обстеження предметної області, системний аналіз об'єкта проектування. - ПРН13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних знань. - ПРН14. Застосовувати на практиці інструментальні програми засоби доменного аналізу, проектування, тестування, візуалізації, вимірювань та документування програмного забезпечення. - ПРН18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних. - ПРН21. Знати, аналізувати, вибирати, кваліфіковано застосовувати засоби забезпечення інформаційної безпеки (в тому числі кібербезпеки) і цілісності даних відповідно до розв'язуваних прикладних завдань та створюваних програмних систем. - ПРН26. Знати засоби інтеграції, розгортання та підтримки спеціалізованих програмних компонентів, розроблених на основі інноваційних технологій для вирішення завдань високопродуктивних обчислень. - ПРН27. Знати основи інженерії даних і конструювання конвеєрів даних та вміти обирати оптимальні алгоритми і технології розробки інноваційних рішень, зокрема для вирішення задач наук про дані та вбудованих систем.
Ключові слова	Розподілені системи, розподілені та паралельні обчислення, розподілені бази даних, кластери даних, бази даних, великі дані, сховища даних.
Формат курсу	Лекції, презентація, виконання лабораторних робіт, обговорення, консультації для кращого розуміння тем, самостійна робота, робота в команді
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	Презентація, лекції, лабораторні роботи, обговорення, дискусія.
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 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 (11 лабораторних робіт). - теоретичний 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>Бали за лабораторними роботами розподіляються так:</p> <ul style="list-style-type: none"> - лаб. робота 1 оцінюється в 2 бали, де від 0 до 1 балу за виконання та від 0 до 1 балу за тестування чи опитування. - лаб. робота 2 оцінюється в 5 балів, де від 0 до 4 балів за виконання та від 0 до 1 балу за тестування чи опитування. - лаб. робота 3 оцінюється в 3 бали, де від 0 до 2 балів за виконання та від 0 до 1 балу за тестування чи опитування. - лаб. робота 4 оцінюється в 3 бали, де від 0 до 2 балів за виконання та від 0 до 1 балу за тестування чи опитування. - лаб. робота 5 оцінюється в 3 бали, де від 0 до 2 балів за виконання та від 0 до 1 балу за тестування чи опитування. - лаб. робота 6 оцінюється в 5 балів, де від 0 до 4 балів за виконання та від 0 до 1 балу за тестування чи опитування. - лаб. робота 7 оцінюється в 3 бали, де від 0 до 2 балів за

виконання та від 0 до 1 балу за тестування чи опитування.

- лаб. робота 8 оцінюється в 5 балів, де від 0 до 4 балів за виконання та від 0 до 1 балу за тестування чи опитування.

- лаб. робота 9 оцінюється в 5 балів, де від 0 до 4 балів за виконання та від 0 до 1 балу за тестування чи опитування.

- лаб. робота 10 оцінюється в 3 бали, де від 0 до 2 балів за виконання та від 0 до 1 балу за тестування чи опитування.

- лаб. робота 11 оцінюється в 5 балів, де від 0 до 4 балів за виконання та від 0 до 1 балу за тестування чи опитування.

При тестування, вага питання має 0.2 бала (5 питань для одного лабораторної роботи). Критерії виконання та оцінювання більш детально розписані у завданнях для лабораторних робіт.

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, фабрикування джерел, списування, втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності.

	<p>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.</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>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. Паралельні обчислювальні системи (<i>Поняття паралельної обчислювальної системи. Паралельна обробка. Види паралелізму: багатопроцесорна обробка, конвейєрна обробка, векторна обробка, векторно-конвейєрна обробка.</i>) 2. Поняття паралелізму (<i>Типи паралелізму: Паралелізм на рівні бітів. Паралелізм на рівні інструкцій. Паралелізм даних Паралелізм завдань. Поняття багатопоточності.</i>) 3. Концепція розподілених систем (<i>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.</i>) 4. 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>) 5. Принципи розробки та етапи проектування паралельних методів (<i>Аналіз обчислювальних систем і декомпозиція їх на частини (підзадачі), які можуть бути реалізовані</i>

незалежно одна від одної. Типи розбиття. Планування комунікацій. Планування обчислень. Виділення для сформованого набору під задач інформаційних взаємодій, що відбуваються в процесі вирішення задачі)

6. Поняття паралельних програм (Стадії розроблення паралельних програм. Визначення обчислювальної системи, необхідної для вирішення задачі. Розподілення набору підзадач між процесорами системи.)
7. Поняття нереляційних баз даних (Поняття NoSQL. Модель даних. Типи моделей даних. Логічна модель. Створення фізичної моделі даних. Фізична модель даних для реляційних баз даних. Інструменти для моделювання даних.)
8. MongoDB (Гнучке моделювання даних за допомогою MongoDB Atlas. Приклад моделювання бази даних. Запити і агрегування. Детальніше про мову запитів в MongoDB.)
9. MongoDB SQL (Розуміння зіставлення MongoDB та SQL. MongoDB SQL: термінологія. MongoDB SQL: Виконувані файли бази даних. MongoDB SQL: команди. Приклад CRUD функцій в MongoDB. Побудова запитів в MongoDB.)
10. Графові бази даних (Що таке є мова запитів GraphQL. Переваги GraphQL. GraphQL в порівнянні REST. Схеми GraphQL. Схеми визначення. Вирішувачі (Resolvers). Запит даних. Повернення даних. Apollo GraphQL. Приклад з NodeJS та Express. Приклад запиту для React client.)
11. 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.)
12. Графова база даних Neo4j (Graph Databases. Приклади графових баз даних. Порівняння графових та реляційних баз даних. Графова база даних Neo4j. Neo4j браузер. Схеми Neo4j. Структура зашифрованого запиту. Написання зашифрованих запитів.)
13. Робота з Neo4j (Схеми 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.)
14. Графові моделі Orient db (Огляд Orient db. Мульти-модель систем керування базами даних (СКДБ). Моделі документів. Графові моделі. Orient db синтаксис.

	<p><i>Типи транзакцій. ETL (Extract, Transform, Load). Можливості Orient db. Використання Orient db.)</i></p> <p>15. 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.)</p> <p>16. 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)</p> <p>17. 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.)</p> <p>18. Особливості HDFS (Особливості побудови 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.)</p> <p>19. 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.)</p> <p>20. 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.)</p> <p>21. Знайомство з Apache Spark (Виклики та рішення. Що таке Apache Spark? Модель Spark. Потужний стек – гнучка розробка. Компоненти Apache Spark)</p> <p>22. 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 MLib конвеєр.)</p> <p>23. 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.)</p>
--	--

	<p>24. Хмарні сервіси з Snowflake (<i>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.</i>)</p> <p>25. Оркестрування потоків операцій в Airflow (<i>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.</i>)</p> <p>26. 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>27. 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>28. 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>29. 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>30. Поняття паралелізації в конвеєрах обробки даних (<i>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.</i>)</p>
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано

	по завершению курсу.
--	----------------------

СХЕМА КУРСУ

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)	Література. Ресурси в Інтернеті	Завдання, 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. 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. Організація конвеєру даних на Databricks	лаб. робота	3-4, 6-7	4	кінець 3-го тижня
	Програмна побудова розподілених систем.	сам. робота	8-9	4,4375	кінець поточного тижня
3	Принципи розробки та етапи проектування паралельних методів.	лекція	1-2, 8-9	2	кінець поточного тижня

	<p>Аналіз обчислювальних систем і декомпозиція їх на частини (підзадачі), які можуть бути реалізовані незалежно одна від одної. Типи розбиття.</p> <p>Планування комунікацій. Планування обчислень. Виділення для сформованого набору під задач інформаційних взаємодій, що відбуваються в процесі вирішення поставленої задачі. Стадії розроблення паралельних програм.</p> <p>Визначення обчислювальної системи, необхідної для вирішення задачі.</p> <p>Розподілення набору підзадач між процесорами системи.</p>				
	Обливості проектування паралельних методів	сам. робота	1-2, 9	4,4375	кінець поточного тижня
4	<p>Робота з MongoDB.</p> <p>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.</p> <p>Детальніше про мову запитів в MongoDB. Розуміння зіставлення MongoDB та SQL. MongoDB SQL: термінологія. MongoDB SQL: Виконувані файли бази даних. MongoDB SQL: команди. Приклад CRUD функцій в MongoDB. Побудова запитів в MongoDB.</p>	лекція	1-2, 10	2	кінець поточного тижня
	Лаб 3. Створення сховища даних на Microsoft Azure	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	<p>Графова база даних GraphQL.</p> <p>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.</p>	лекція	6, 10-11	2	кінець поточного тижня

	Лаб 4. Знайомство з Pyspark	лаб. робота	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	7, 8	2	кінець поточного тижня
	Лаб 5. Побудова конвеєру машинного навчання на Spark	лаб. робота	3, 5	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 6. Основи структурованої потокової подачі даних	лаб. робота	8-9, 14	4	кінець 8-го 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	лекція	3, 5, 8	2	кінець поточного тижня

	<p>HBase. Hbase на Python. Операції HBase. Функції операцій HBase. Огляд переваг операцій HBase над іншими файловими системами. Огляд Orient db.</p> <p>Мульти-модель систем керування базами даних (СКДБ). Моделі документів.</p> <p>Графові моделі. Orient db синтаксис.</p> <p>Типи транзакцій. ETL (Extract, Transform, Load). Можливості Orient db.</p> <p>Використання Orient db.</p>				
	Особливості мови запитів SQL	сам. робота	15	4,4375	кінець ПОТОЧНОГО ТИЖНЯ
9	<p>Концепція та компоненти Hadoop.</p> <p>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.</p>	лекція	5, 16	2	кінець ПОТОЧНОГО ТИЖНЯ
	Лаб 7. Структурована потокова подача даних з Apache Kafka	лаб. робота	1-2, 8-9, 17	2	кінець 10-го ТИЖНЯ
	Методи управління паралельним доступом до баз даних	сам. робота	9-10, 13	4,4375	кінець ПОТОЧНОГО ТИЖНЯ
10	<p>Різновид реляційних мов запити.</p> <p>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.</p>	лекція	5, 11, 12	2	кінець ПОТОЧНОГО ТИЖНЯ

	Лаб 8. Робота із дельта-форматом даних	лаб. робота	3, 9-10, 13	4	кінець 11-го 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
	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 9. Облаштування за допомогою Azure Data Services	лаб. робота	6-7	4	кінець 13-го тижня
	Особливості розподілених файлових систем	сам. робота	9-10	4,4375	кінець поточного тижня
13	Оркестрування потоків операцій в Airflow. Поняття Apache Airflow. Основні відомості та призначення. Поняття	лекція	14, 18	2	кінець поточного тижня

	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.				
	Особливості розподіленої файлової системи HDFS	сам. робота	19	4,4375	кінець поточного тижня
14	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.	лекція	9, 20	2	кінець поточного тижня
	Лаб 10. Побудова Delta Live Tables (DLT)	лаб. робота	3	2	кінець поточного тижня
	Особливості розподіленої файлової системи HBase	сам. робота	19	4,4375	кінець поточного тижня
15	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.	лекція	13, 21	2	кінець поточного тижня
	Лаб 11. Побудова SQL Warehouse	лаб. робота	3-4, 6, 15, 18	4	кінець 16-го тижня
	Операції та функції операцій Google BigTable	сам. робота	13	4,4375	кінець поточного тижня

16	<p>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 Retuum. Складність розпаралелювання алгоритмів. Паралелізація з урахуванням структури (SAP). Динамічний планувальник з урахуванням структури (STRADS). Обчислювальна ефективність параметрів моделі.</p>	лекція	1-4, 8-9, 16-17	2	кінець поточного тижня
	Моніторинг стану розподілених обчислювальних систем	сам. робота	22-23	4,4375	кінець поточного тижня