

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
“Аналітика великих даних у хмарних технологіях”,
що викладається в межах ОПП “Комп'ютерні науки”
першого (бакалаврського) рівня вищої освіти для здобувачів з
спеціальності 122 – Комп'ютерні науки

Львів 2023 р.

Назва дисципліни	Аналітика великих даних у хмарних технологіях
Адреса викладання дисципліни	Корпус факультету електроніки та комп'ютерних технологій, Львівський національний університет імені Івана Франка, вул. Драгоманова 50, м. Львів, 79005, вул. Ген. Тарнавського 107, м. Львів, 79011
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 122 – Комп'ютерні науки
Викладачі дисципліни	Стахіра Роман Йосипович, кандидат фізико-математичних наук, доцент;
Контактна інформація викладачів	roman.stakhirar@lnu.edu.ua , https://electronics.lnu.edu.ua/employee/stahira-roman-josypovych
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/лабораторних занять (за попередньою домовленістю): ауд. 305, корпус факультету електроніки та комп'ютерних технологій, вул. Драгоманова 50, м. Львів
Сторінка курсу	https://moodle.elct.lnu.edu.ua/course/view.php?id=321
Інформація про дисципліну	Дисципліна “ Аналітика великих даних у хмарних технологіях ” є вибірковою дисципліною зі спеціальності 122 – Комп'ютерні науки для освітньої програми “Комп'ютерні науки”, яка викладається в 7-му семестрі в обсязі 5-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Навчальну дисципліну розроблено на основі аналізу актуальних проблем при роботі із великими даними у різних сферах їх застосування, що регулярно проводиться на відповідних конференціях чи семінарах. Серед найголовніших питань, що розглядаються в дисципліні, є питання базових понять великих даних, організації та використання розподілених технологій для опрацювання великих даних, управлінні ресурсами віддалених розподілених систем та сховищ великих та надвеликих даних, побудові конвеєрів для опрацювання великих даних, вирішенні питань аналітики та інших задач в галузі науки про дані та систем штучного інтелекту.
Мета та цілі дисципліни	Метою вивчення дисципліни “ Аналітика великих даних у хмарних технологіях ” є надання поглиблених знань та практичних навичок щодо роботи з великими даними, побудови й використання розподілених систем для побудови конвеєрів опрацювання великих даних, формування системи теоретичних знань і набуття практичних умінь та навичок щодо застосування, налагодження й адміністрування систем на базі технологій великих даних та розподілених баз та сховищ даних, проектуванні надійних та економічно доцільних систем для збереження великих об'ємів

	<p>даних.</p> <p>Цілями дисципліни є отримання практичних навичок та засвоєння методів опрацювання великих даних, створення кластерів даних та технології їх проектування для вирішення задач наук про дані на основі великих даних, наповнення даними та підтримання в робочому стані систем з даними.</p>
<p>Література для вивчення дисципліни</p>	<ol style="list-style-type: none"> 1. Rajkumar Buууа, Rodrigo N. Calheiros, Amir Vahid Dastjerdi. Big Data: Principles and Paradigms: Elsevier, 2016. - 496 p. - Режим доступу: https://dphoto.lecturer.pens.ac.id/lecture_notes/internet_of_things/Big%20Data%20Principles%20and%20Paradigms.pdf 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. 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. Документація Apache Hadoop [Електронний ресурс] // Apache Hadoop. – 2021. – Режим доступу до ресурсу: https://hadoop.apache.org/docs/stable/. 7. Документація Apache Spark [Електронний ресурс] // Apache Spark. – 2019. – Режим доступу до ресурсу: https://spark.apache.org/docs/latest/. 8. Документація HBase [Електронний ресурс] // HBase. – 2019. – Режим доступу до ресурсу: https://hbase.apache.org/book.html. 9. RabbitMq [Електронний ресурс] // RabbitMq. – 2020. – Режим доступу до ресурсу: https://www.rabbitmq.com/documentation.html. 10. Ifeyinwa A. A. Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications / A. A. Ifeyinwa, H. N. Friday. – Nigeria: Abakaliki, 2019. – 30 с. 11. Donald Miner, Adam Sbook. MapReduce Design Patterns: O'Reilly, 2013. - 251 p. - Режим доступу: http://vargas-solar.com/bigdata-fest/wp-content/uploads/sites/33/2014/11/MapReduce-Design-Patterns-V413HAV.pdf 12. 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. 13. Kai Hwang, Min Chen. Big-Data Analytics for Cloud, IoT and Cognitive Computing: Willey, 2017. - 428 p. 14. Designing Distributed System. - [Режим доступу]: https://azure.microsoft.com/mediahandler/files/resourcefiles/designing-distributed-systems/Designing_Distributed_Systems.pdf 15. Kristina Chodorow. Scaling MongoDB: O'Reilly, 2011. - 58 p. 16. Joyce Kay Avila. Snowflake The Definitive Guide: O'Reilly, 2022. - 468 p. - [Електронний ресурс]. - Режим доступу: https://www.snowflake.com/wp-content/uploads/2022/03/Snowflake-The-Definitive-Guide-1.pdf 17. Google. Cloud Bigtable. - [Електронний ресурс]. - Режим доступу: https://cloud.google.com/bigtable 18. Rik Van Bruggen. Learning Neo4j: Packt Publishing, 2014. - 222 p. 19. 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 20. Alex Holmes. Hadoop in Practice: Manning Publications, 2012. - 537 p. - Режим доступу: https://ia600201.us.archive.org/7/items/HadoopInPractice/Hadoop%20in%20Practice.pdf 21. 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 22. 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 23. Apache HBase Team. Apache HBase™ Reference Guide. - [Електронний ресурс]. - Режим доступу: https://hbase.apache.org/apache_hbase_reference_guide.pdf

	24. 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
Обсяг курсу	Загальний обсяг: 150 годин. Аудиторних занять: 96 год., з них 48 год. лекційних та 48 годин лабораторних робіт. Самостійної роботи: 54 год.
Очікувані результати навчання	<p>Після вивчення даного курсу здобувачі будуть знати і вміти:</p> <p>Здатність до абстрактного мислення, аналізу та синтезу.</p> <p>Здатність застосовувати знання у практичних ситуаціях.</p> <p>Здатність вчитися і оволодівати сучасними знаннями.</p> <p>Здатність до пошуку, оброблення та аналізу інформації з різних джерел.</p> <p>Здатність розробляти архітектури, модулі та компоненти програмних систем.</p> <p>Здатність до алгоритмічного та логічного мислення.</p> <p>Здатність розробляти високопродуктивні програмні комплекси для вирішення задач наук про дані, систем штучного інтелекту, вбудованих та інших інноваційних систем.</p> <p>Здатність здійснювати розробку програмних продуктів використовуючи різні методології і технології із забезпеченням їх інноваційності та високої продуктивності.</p> <p>Знати і застосовувати відповідні математичні поняття, методи доменного, системного і об'єктно-орієнтованого аналізу та математичного моделювання для розробки програмного забезпечення.</p> <p>Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення.</p> <p>Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.</p> <p>Знати засоби інтеграції, розгортання та підтримки спеціалізованих програмних компонентів, розроблених на основі інноваційних технологій для вирішення завдань високопродуктивних технологій.</p>
Ключові слова	Хмарні обчислення, великі дані (Big Data), data warehouse, data lake, AWS, GCP, Azure, Hadoop, Spark, MapReduce, Hive, Pig.
Формат курсу	Очний
Теми	Див. Схема курсу
Підсумковий контроль, форма	Залік у кінці семестру
Пререквізити	<p>Для вивчення даного курсу студентам потрібні базові знання з курсів:</p> <ul style="list-style-type: none"> - вища математика; - дискретна математика; - основи програмування; - алгоритми і структури даних; - об'єктно-орієнтоване програмування.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Інформаційні методи (лекції, презентації, лабораторні роботи, написання рефератів, виконання індивідуальних завдань, робота у групі, командна робота, обговорення, консультації для поглибленого розуміння тем, бесіда, ілюстрація, демонстрація), дедуктивні методи на основі узагальнень, евристичні методи (проблемна лекція), інтерактивні методи (дискусія).

Необхідне обладнання	Мультимедійне обладнання, комп'ютерний клас, програми та сервіси MS Teams, Moodle, Хмарна платформа AWS (CloudFormation, S3, AWS Cloud9, S3 Data Lake, Elasticsearch, IAM, CloudWatch, Kinesis Data Firehose, DWH, Amazon Athena, Visual ETL)
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	<p>Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням:</p> <ul style="list-style-type: none"> • Змістовий модуль 1: до 25 балів за виконання модульного завдання – написання есе за обраною тематикою з переліку. • Змістовий модуль 2: до 25 балів за виконання модульного завдання – написання есе за обраною тематикою з переліку. • Виконання лабораторних робіт: до 50 балів. <p>Підсумкова максимальна кількість балів 100.</p> <p>Змістовий модуль – самостійна робота студента оформлена у вигляді есе – робота друкованим текстом, рекомендованим обсягом до 10 сторінок (шрифт Times New Roman, 14). Есе включає в себе детальний розгляд обраної індивідуальної теми, приведення прикладів та лістингів коду програм, огляду технологій, літератури. Код програм повинен бути обов'язково прокоментований та пояснений, необхідно також продемонструвати його роботу у разі, якщо в якості прикладу наводяться не окремі елементи технології, а суцільна програма. Есе також повинно містити посилання на літературні джерела/інтернет ресурси, що були використані під час його написання. Фінальна версія есе здається студентом у електронному форматі .pdf викладачу для оцінки. Теми для змістових модулів див. у розділі Питання до модульного контролю.</p> <p>Академічна доброчесність: Очікується, що роботи студентів будуть оригінальними дослідженнями чи міркуваннями. Списування та втручання в роботу інших студентів становлять, але не обмежують, приклади можливої академічної недоброчесності. Виявлення ознак академічної недоброчесності в написанні завдань є підставою для її незарахування викладачем, незалежно від масштабів плагіату чи обману. Жодні форми порушення академічної доброчесності не толеруються.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та лабораторні заняття курсу. Студенти повинні інформувати викладача про неможливість відвідати заняття. У будь-якому випадку студенти зобов'язані дотримуватися термінів визначених для виконання всіх 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 при поточному контролі та бали за виконання лабораторних робіт. При цьому обов'язково враховуються присутність на заняттях та активність студента під час 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 і т. ін.</p> <p>Оцінювання лабораторних робіт (10 лабораторних робіт, максимальна</p>

кількість балів: 50) .

Бали оцінювання лабораторних робіт нараховуються за наступним співвідношенням:

5 – студент в повному обсязі володіє навчальним матеріалом, має повне розуміння розглянутої теми, надає правильні відповіді на запитання по темі, код програми функціонує відповідно до завдання;

4 – студент достатньо розуміє розглянутий матеріал та принципи написаного ним коду програми, присутні неточності та незначні помилки у відповідях на запитання по темі, код програми функціонує відповідно до завдання (або з несуттєвими недоліками);

3 – 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;

2 – студент погано розуміє розглянутий матеріал та написаний ним код програми, студент в більшості надає помилкові відповіді на питання по темі, код програми функціонує з суттєвими недоліками;

1 - студент погано розуміє розглянутий матеріал та написаний ним код програми, код програми не функціонує належним чином;

0 - студент зовсім не засвоїв розглянутий матеріал, написаний ним код програми не відповідає темі/не функціонує взагалі.

Оцінювання змістових модулів (2 змістових модулі, 25 балів за кожний)
— за результатами написаних студентом есе, тестів, програм, тощо.

Бали оцінювання змістових модулів нараховуються за наступним співвідношенням:

25-20 - розглянута тема відтворюється в повному обсязі, правильно, обґрунтовано, логічно, містить аналіз і систематизацію, аргументовані висновки. Засвідчено глибоке володіння матеріалом. Наведені приклади коду повністю робочі та відповідають темі. Можуть бути присутні несуттєві помилки та невідповідності;

20-15 - 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;

15-10 - відстежується загальне розуміння розглянутої теми. Виявлені множинні неточності та невідповідності, пояснення наведеного коду відсутні, код функціонує зі значними неточностями (або відсутні приклади запуску коду на виконання взагалі);

10-5 – студент погано розуміє розглянуту тему. Виявлені суттєві неточності та невідповідності. Наведені приклади коду з суттєвими недоліками, або не відповідають темі;

5 – 0 – 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.

Критерії оцінювання результатів неформальної освіти:

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.
Питання до модульного контролю	<ol style="list-style-type: none"> 1. Вступ до великих даних (Великі дані. Атрибути великих даних. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Джерела великих даних. Загальні поняття для накопичення та опрацювання великих даних.) 2. Парадигма MapReduce та доступ до даних (<i>Історія MapReduce. Розгляд системи MapReduce. Розгляд платформи Apache Pig. Apache Hive. Hive QL. Архітектура MRv1. Планувальник YARN.</i>) 3. Мова запитів GraphQL (<i>Що таке є мова запитів GraphQL. Переваги GraphQL. GraphQL в порівнянні REST. Схеми GraphQL. Схеми визначення. Вирішувачі (Resolvers). Запит даних. Повернення даних. Apollo GraphQL.</i>) 4. Шаблони у MapReduce (<i>Поняття шаблону. Питання проектування шаблонів MapReduce. Шаблон фільтрування. Фільтр Bloom. Шаблони об'єднання. Мета шаблони.</i>) 5. Шаблони сумаризації у MapReduce (<i>Шаблон сумаризації. Числова сумаризація. Сумаризація з інвертованих індексом. Сумаризація обрахунків.</i>) 6. Шаблони організації даних у MapReduce (<i>Шаблон організації даних. Структурований шаблон. Ієрархічний шаблон. Шаблон роздріблення та зв'язування.</i>) 7. Концепція розподілених систем (<i>Приклади типових розподілених систем. Компоненти розподілених систем. Інтернет та інтранет. Використання пристроїв у розподілених системах.</i>) 8. Протоколи обміну даними в розподілених системах (<i>Обмін ресурсів в WWW. Веб-сервери та веб-браузери. HTTP/HTTPS протоколи. Програмні та апаратні сервісні рівні. Клієнт-серверні моделі. Веб-прохі сервери.</i>) 9. 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>) 10. MongoDB (<i>Гнучке моделювання даних за допомогою MongoDB Atlas. Приклад моделювання бази даних. Запити і агрегування. Детальніше про мову запитів в MongoDB.</i>) 11. MongoDB SQL (<i>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.</i>) 12. Резидентна система управління базами даних Redis (<i>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.</i>) 13. Графова база даних Neo4j (<i>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.</i>)

	<p>14. Робота з Neo4j (<i>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.</i>)</p> <p>15. База даних Hbase (<i>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.</i>)</p> <p>16. Особливості розроблення веб-додатків і веб-сервісів із застосуванням технологій розподілених сховищ даних (<i>Поняття веб-додатку та веб-сервісу. Розподілені бази даних. Сховища даних. Розподілені технології. Особливості розподілених сховищ даних</i>)</p> <p>17. Концепція Hadoop (<i>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.</i>)</p> <p>18. Особливості HDFS (<i>Особливості побудви архітектури HDFS. Операції HDFS. Функції операцій HDFS. Огляд переваг операцій HDFS над іншими файловими системами. Типи запитів. Мовна підтримка. Hadoop Fea.</i>)</p> <p>19. Формування SQL запитів (<i>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.</i>)</p> <p>20. Знайомство з Apache Spark (<i>Виклики та рішення. Що таке Apache Spark? Модель Spark. Потужний стек – гнучка розробка. Компоненти Apache Spark</i>)</p> <p>21. Spark SQL (<i>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.</i>)</p> <p>22. Поняття хмарної платформи Snowflake (<i>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.</i>)</p> <p>23. Поняття стеку ELK (<i>ELK стек. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.</i>)</p> <p>24. Поняття стеку EFK (<i>EFK стек. Elasticsearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.</i>)</p> <p>25. 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>26. 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>27. Організація конвеєрів великих даних на RabbitMQ (<i>Поняття RabbitMQ. Опис. Взаємозв'язок. Використання. Протокол AMQ. Масштабування RabbitMQ. Інтеграція з базами даних.</i>)</p> <p>28. Організація конвеєрів великих даних на Kafka (<i>Поняття технології Apache Kafka. Поняття брокера. Поняття споживача. Взаємодія сервісів через меседж брокери. Порівняння Kafka і RabbitMQ.</i>)</p> <p>29. 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.</i>)</p>
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, год	Термін виконання, тиж.
1	Вступ до великих даних. Великі дані. Атрибути великих даних. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Джерела великих даних. Загальні поняття для накопичення та опрацювання великих даних.	лекція	1 - 10	3	кінець поточного тижня
	Реєстрація та знайомство з Databricks	лаб. робота	3-4	3	кінець поточного тижня
	Природа та особливості великих даних	сам. робота	1-2, 10	3	кінець поточного тижня
2	Поняття розподілених систем. 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.	лекція	1-9	3	кінець поточного тижня

	Поняття керування великими даними на Databriks	лаб. робота	3-4	3	кінець поточного тижня
	Проблеми побудови розподілених систем для роботи з великими даними	сам. робота	5, 8-10	3	кінець поточного тижня
3	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.	лекція	1-2, 8-16	3	кінець поточного тижня
	Організація конвеєру на Databriks для роботи із великими даними	лаб. робота	3-4, 6-7	3	кінець поточного тижня
	Обливості проектування кластерів для опрацювання великих даних	сам. робота	1-2, 9	3	кінець поточного тижня
4	Парадигма MapReduce та доступ до даних. Історія MapReduce. Розгляд системи MapReduce. Розгляд платформи Apache Pig. Apache Hive. Hive QL. Архітектра MRv1. Планувальник YARN. Мовна підтримка. MapReduce. Властивості MapReduce. Трекер задач. Hive. HiveQL.	лекція	1-2, 10	3	кінець поточного тижня

	Знайомство з Pyspark	лаб. робота	3, 7	3	кінець поточного тижня
	Методи синхронізації процесів та даних в розподілених програмах	сам. робота	1-2, 8-9	3	кінець поточного тижня
5	Шаблони MapReduce. Поняття шаблону. Питання проектування шаблонів MapReduce. Шаблон сумаризації. Числова сумаризація. Сумаризація з інвертованих індексом. Сумаризація обрахунків. Шаблон фільтрування. Фільтр Bloom. Шаблон організації даних. Структурований шаблон. Ієрархічний шаблон. Шаблон розділення та зв'язування. Шаблони об'єднання. Мета шаблони.	лекція	1, 11	3	кінець поточного тижня
	Побудова конвеєру машинного навчання на Spark	лаб. робота	3, 7	3	кінець поточного тижня
	Особливості побудови архітектури системи на основі технології MapReduce	сам. робота	11	3	кінець поточного тижня
6	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.	лекція	6, 10-15	3	кінець поточного тижня

	MongoDB SQL: термінологія. MongoDB SQL: Виконувані файли бази даних. MongoDB SQL: команди. Приклад CRUD функцій в MongoDB. Побудова запитів в MongoDB.				
	Основи структурованої потокової подачі даних	лаб. робота	8-9, 14	3	кінець поточного тижня
	Розподілені файлові сховища даних на основі MongoDB	сам. робота	8-10, 15-17	3	кінець поточного тижня
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.	лекція	7-8	3	кінець поточного тижня
	Структурована потокова подача даних з Apache Spark	лаб. робота	3, 7	3	кінець поточного тижня
	Поняття кешування в системах з великими даними	сам. робота	11	3	кінець поточного тижня
8	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	лекція	8, 9, 18	3	кінець поточного тижня

	запитів. Створення та запит до вершин. Встановлення взаємозв'язку між вершинами. Використання транзакційного зашифрованого HTTP end-point. Використовувані протоколи.				
	Поняття сучасних форматів для зберігання великих даних	лаб. робота	3, 9-10, 13	3	кінець поточного тижня
	Особливості мови запитів SQL для великих даних	сам. робота	20-25	3	кінець поточного тижня
9	Бази даних Hbase. 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 системами.	лекція	8, 24	3	кінець поточного тижня
	Робота із дельта-форматом даних	лаб. робота	3, 9-10, 13	3	кінець поточного тижня
	Методи управління паралельним доступом до сховищ з великими даними	сам. робота	9-10, 13	3	кінець поточного тижня
10	Стеки великих даних. ELK стек. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок.	лекція	1, 6-8	3	кінець поточного тижня

	Використання. Принцип роботи. EFK стек. ElasticSearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.				
	Побудова Delta Live Tables (DLT)	лаб. робота	3	3	кінець поточного тижня
	Поняття IoT систем	сам. робота	13	4	кінець поточного тижня
11	Робота з 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	3, 7	3	кінець поточного тижня
	Проектування конвеєру даних для міграції даних з NoSQL у SQL	лаб. робота	3-4, 6, 15, 18	3	кінець поточного тижня
	Резервування та відновлення у розподілених базах даних	сам. робота	1-2, 9	3	кінець поточного тижня
12	Хмарна платформа Snowflake. Поняття платформи даних. Переваги хмарних платформ даних. Використання платформ даних для	лекція	16	3	кінець поточного тижня

	бізнесу. Традиційні архітектури. Сучасна архітектура даних 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.				
	Програмування конвєсру для міграції даних з NoSQL у SQL	лаб. робота	3-4, 6, 15, 18	3	кінець поточного тижня
	Особливості розподілених файлових систем Snowflake	сам. робота	9-10	4	кінець поточного тижня
13	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.	лекція	22	3	кінець поточного тижня
	Проектування конвєсру даних для міграції даних з SQL у NoSQL сховище даних	лаб. робота	3-4, 6, 15, 18	3	кінець поточного тижня
	Використання Airflow у конвєсрах обробки великих даних	сам. робота	18	4	кінець поточного тижня

14	Організація конвеєрів великих даних. Поняття RabbitMQ. Опис. Взаємозв'язок. Використання. Протокол AMQ. Масштабування RabbitMQ. Інтеграція з базами даних. Поняття технології Apache Kafka. Взаємодія сервісів через меседж брокери. Порівняння Kafka і RabbitMQ.	лекція	1, 9, 20-22	3	кінець поточного тижня
	Програмування конвеєру для міграції даних з SQL у NoSQL	лаб. робота	1-4, 6, 15, 18	3	кінець поточного тижня
	Особливості роботи з RabbitMQ	сам. робота	19	4	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.	лекція	17, 24	3	кінець поточного тижня
	Проектування Data Warehouse для великих даних	лаб. робота	1-4, 6, 15, 18	3	кінець поточного тижня
	Операції та функції операцій Google BigTable	сам. робота	13	4	кінець поточного тижня

16	Машинне 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.	лекція	1-4, 8-9, 16-17	3	кінець поточного тижня
	Проектування Data Lake для великих даних	лаб. робота	1, 3-4, 6, 15, 18	3	кінець поточного тижня
	Побудова сховищ даних для великих даних	сам. робота	1, 14, 19-23	4	кінець поточного тижня