

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
“Засоби роботи з великими даними”,
що викладається в межах ОПП
“Інженерія програмного забезпечення”
першого (бакалаврського) рівня вищої освіти для здобувачів з спеціальності 121 – Інженерія програмного забезпечення

Львів 2023 р.

Назва дисципліни	Засоби роботи з великими даними
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 121 – Інженерія програмного забезпечення
Викладачі дисципліни	Павлишенко Б.М., докт. техн. наук, професор Мисюк Ірина Володимирівна
Контактна інформація викладачів	bohdan.pavlyshenko@lnu.edu.ua iryna.mysiuk@lnu.edu.ua https://electronics.lnu.edu.ua/employee/mysiuk-i-v
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/лабораторних занять (за попередньою домовленістю): ауд. 201, корпус факультету електроніки та комп'ютерних технологій, вул. Драгоманова 50, м. Львів
Сторінка курсу	https://electronics.lnu.edu.ua/course/web-tekhnohii-ipz
Коротка анотація дисципліни	Дисципліна “Засоби роботи з великими даними” є дисципліною з циклу професійної та практичної підготовки спеціальності 121 Програмна інженерія для освітньої програми «Інженерія програмного забезпечення», яка викладається в 6 семестрі в обсязі 4 кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Інформація про дисципліну	Курс розроблено для ознайомлення студентів з основними технологіями, обов'язкові для того, щоб розв'язувати типові задачі при роботі з великими даними.
Мета та цілі дисципліни	<i>Мета:</i> одержання студентами знань про сучасні технології великих даних, які використовуються для їх обробки. <i>Цілі:</i> забезпечити знайомство студентів з загальною теорією використання засобів та технологій для роботи з великими даними і сформуванню навички їх практичного створення; навчити студента базових інструментів Apache Hadoop, Pig, Apache Spark ELK, EFK, HDFS, RabbitMQ, Kafka, ZeroMQ
Література для вивчення дисципліни	Основна література: 1. Документація Apache Hadoop [Електронний ресурс] // Apache Hadoop. – 2021. – Режим доступу до ресурсу: https://hadoop.apache.org/docs/stable/ . 2. Документація Apache Spark [Електронний ресурс] // Apache Spark. – 2019. – Режим доступу до ресурсу: https://spark.apache.org/docs/latest/ . 3. Документація HBase [Електронний ресурс] // HBase. – 2019. – Режим доступу до ресурсу: https://hbase.apache.org/book.html . 4. RabbitMq [Електронний ресурс] // RabbitMq. – 2020. – Режим доступу до ресурсу: https://www.rabbitmq.com/documentation.html .

	<p>5. 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.</p> <p>6. Michael Armbrust. Makeing Apache Spark better with Delta Lake: Databricks, 2020. - 399 p.</p>
Обсяг курсу	64 годин аудиторних занять. З них 32 години лекцій, 32 годин лабораторних робіт.
Очікувані результати навчання	<p>У результаті вивчення даного курсу студент буде:</p> <p>знати: основні поняття, визначення і проблеми курсу; вимоги до постановки основних задач та інструментів роботи з великими даними; призначення й особливості застосування основних складових механізмів роботи з великими даними;</p> <p>вміти: володіти базовими знаннями Apache Hadoop, Pig, Apache Spark, ELK, EFK, HDFS, RabbitMQ, Kafka, ZeroMQ.</p> <p>Після вивчення даного курсу «Засоби роботи з великими даними» здобувачі набудуть таких Загальних та Фахових компетентностей та Програмних результатів навчання:</p> <p>ЗК1. Здатність до абстрактного мислення, аналізу та синтезу. ЗК2. Здатність застосовувати знання у практичних ситуаціях. ЗК6. Здатність до пошуку, оброблення та аналізу інформації з різних джерел. 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. ФК19. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних. ФК28. Володіння методами сучасних веб-технологій, хмарних технологій, великих даних та засобами розробки веб-застосувань. ПРН18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних. ПРН26. Знати та вміти застосовувати засоби інженерії програмного забезпечення для реалізації проєктів з використанням технологій науки про дані та штучного інтелекту.</p>
Ключові слова	Великі дані, big data, Apache Spark, Kafka, Map reduce
Формат курсу	Очний

Теми	<p>Тема 1. Великі дані. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Парадигма Map-Reduce.</p> <p>Тема 2. Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування.</p> <p>Тема 3. Загальний розгляд технологій Hadoop. Файлова система HDFS. Його Архітектура.</p> <p>Тема 4. Розгляд системи Map reduce. Розгляд платформи Apache Pig. Apache Hive. Hive QL. Архітектура MRv1. Планувальник YARN.</p> <p>Тема 5. Розгляд бази даних Hbase. Її модель даних. Життєвий цикл. Архітектура.</p> <p>Тема 6. MongoDB та Neo4j. Архітектура баз даних та їхнє використання у Великих даних.</p> <p>Тема 7. Розгляд бази даних Cassandra. Її принцип роботи. Використання. Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.</p> <p>Тема 8. Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.</p> <p>Тема 9. ELK стек. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи.</p> <p>Тема 10. EFK стек. Elasticsearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.</p> <p>Тема 11. RabbitMQ. Опис. Взаємозв'язок. Використання.</p> <p>Тема 12. Apache Kafka і RabbitMQ. Взаємодія сервісів через меседж брокери.</p>
Підсумковий контроль, форма	Залік
Пререквізити	<p>Для вивчення даного курсу студентам потрібні базові знання з курсів:</p> <ul style="list-style-type: none"> - основи програмування; - алгоритмізація і програмування; - об'єктно-орієнтоване програмування; - бази даних; - нереляційні та розподілені бази даних.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	<p>Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація);</p> <p>дедуктивні методи на основі узагальнень;</p> <p>евристичні методи (проблемна лекція);</p> <p>інтерактивні методи (дискусія).</p>
Необхідне обладнання	<p>Мультимедійне обладнання, комп'ютерний клас, Moodle, Databricks, Hadoop, Apache Spark, PySpark, Apache Kafka, Python, Pig, Apache Spark ELK, EFK, HDFS, RabbitMQ, Kafka, ZeroMQ</p>
Критерії оцінювання (окремо для кожного виду нав-	<p>Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням:</p> <ul style="list-style-type: none"> • Змістовий модуль 1: 25% семестрової оцінки за виконані лабораторні роботи. • Змістовий модуль 2: 25% семестрової оцінки за виконані лабораторні

<p>чальної діяльності)</p>	<p>роботи.</p> <ul style="list-style-type: none"> • залік: 50% семестрової оцінки, максимальна кількість балів 50. <p>Підсумкова максимальна кількість балів 100.</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.</p> <p>Відвідання занять є важливою складовою навчання. Очікується, що всі студенти відвідають усі лекції та лабораторні зайняття 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 підсумкового тестування. При цьому обов'язково враховуються присутність на заняттях та активність студента під час практичного заняття; недопустимість пропусків та запізньєнь 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>Жодні форми порушення академічної доброчесності не толеруються.</p> <p>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.</p>
<p>Питання до заліку чи екзамену.</p>	<p>Великі дані. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Парадигма Map-Reduce.</p> <p>Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування.</p> <p>Загальний розгляд технології Hadoop. Файлова система HDFS. Його Архітектура.</p> <p>Розгляд системи Map reduce. Розгляд платформи Apache Pig.</p> <p>Apache Hive. Hive QL. Архітектура MRv1. Планувальник YARN.</p> <p>Розгляд бази даних Hbase. Її модель даних. Життєвий цикл. Архітектура. MongoDB та Neo4j. Архітектура баз даних та їхнє використання у Великих даних.</p> <p>Розгляд бази даних Cassandra. Її принцип роботи. Використання.</p>

	<p>Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.</p> <p>Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.</p> <p>ELK стек. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи.</p> <p>EFK стек. Elasticsearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.</p> <p>RabbitMQ. Опис. Взаємозв'язок. Використання.</p> <p>Apache Kafka і RabbitMQ. Взаємодія сервісів через меседж брокери.</p>
Опитування	Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.

**Схема курсу “Засоби роботи з великими даними”
для студентів спеціальності 121 – Інженерія програмного забезпечення**

Тиж.	Тема, план, короткі тези	Форма діяльності (заняття)* *лекція, самостійна, дискусія, групова робота)	Література.** * Ресурси в інтернеті	Завдання, год	Термін виконання
1	Вступ. Великі дані. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Парадигма Map-Reduce.	Лекція	[1], [2], [3], [4], [5] Сайт курсу	2	кінець поточного тижня
1	Лаб. 1. Ознайомлення, встановлення та розгляд технологій великих даних	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
2	Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування.	Лекція	[1] Сайт курсу	2	кінець поточного тижня
2	Лаб. 2. Розгляд технології Hadoop. Реалізація простої обробки даних.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
3	Загальний розгляд технології Hadoop. Файлова система HDFS. Його Архітектура.	Лекція	[1] Сайт курсу	2	кінець поточного тижня
3	Лаб. 3. Використання Apache Pig в Apache Hadoop.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
4	Розгляд системи Map reduce. Розгляд платформи Apache Pig.	Лекція	[1] Сайт курсу	2	кінець поточного тижня
4	Лаб. 4. Використання Apache Hive в Apache Hadoop.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
5	Apache Hive. Hive QL. Архітектура MRv1. Планувальник YARN.	Лекція	[1] Сайт курсу	2	кінець поточного тижня
5	Лаб. 5. Розгляд бази даних Hbase. Її модель даних. Життєвий цикл. Архітектура.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
6	MongoDB та Neo4j. Архітектура баз даних та їхнє використання у Великих даних.	Лекція	[3] Сайт курсу	2	кінець поточного тижня
6	Лаб. 6. Розробка проекту з вивченими технологіями..	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
7	Розгляд бази даних Cassandra. Її принцип роботи. Використання	Лекція	[3], Сайт курсу	2	кінець поточного тижня

7	Лаб. 7. Розгляд технології Apache Spark. Реалізація простої обробки даних.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
8	Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.	Лекція	[3], [6] Сайт курсу	2	кінець поточного тижня
8	Лаб. 8 Налаштування легування з використанням ELK	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
9	Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.	Лекція	[2], Сайт курсу	2	Кінець поточного тижня
9	Лаб. 8. Налаштування легування з використанням EFK	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
10	ELK стек. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи	Лекція	[2], Сайт курсу	2	кінець поточного тижня
10	Лаб. 9. Розробка проекту з вивченими технологіями.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
11	EFK стек. Elasticsearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.	Лекція	[5], Сайт курсу	2	кінець поточного тижня
11	Лаб.10. Використання Apache Kafka для передавання даних між сервісами	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
12	RabbitMQ. Опис. Взаємозв'язок. Використання.	Лекція	[5], Сайт курсу	2	кінець поточного тижня
12	Лаб. 11. Використання RabbitMQ для передавання даних між сервісами.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня
13	Apache Kafka і RabbitMQ. Взаємодія сервісів через меседж брокери.	Лекція	[4], [5], Сайт курсу	2	кінець поточного тижня
13	Лаб. 12. Розробка проекту з вивченими технологіями.	Лабораторна робота	Сайт курсу	2	кінець поточного тижня