

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Львівський національний університет імені Івана Франка
Механіко-математичний факультет
Кафедра математичної статистики і диференціальних рівнянь

Затверджено

На засіданні
кафедри математичної статистики і
диференціальних рівнянь
механіко-математичного факультету
Львівського національного університету
імені Івана Франка
(протокол № 1 від 22.06.2023 р.)



Завідувач кафедри:

A handwritten signature in black ink, appearing to be "Олег Бугрій".

Олег БУГРІЙ

Силабус з навчальної дисципліни
“ Методи і техніки аналізу великих даних ”,
що викладається в межах ОПШ
“ Статистичний аналіз даних ”
другого (магістерського) рівня вищої освіти для здобувачів з
спеціальності 112 - Статистика

Львів 2023 р.

Назва дисципліни	Методи і техніки аналізу великих даних
Адреса викладання дисципліни	Головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Факультет та кафедра, за якою закріплена дисципліна	Механіко-математичний факультет Кафедра математичної статистики і диференціальних рівнянь
Галузь знань, шифр та назва спеціальності	11 - Математика та статистика 112 – Статистика
Викладачі дисципліни	Бугрій Олег Миколайович, доктор фізико-математичних наук, професор, завідувач кафедри математичної статистики і диференціальних рівнянь
Контактна інформація викладачів	oleh.buhrii@lnu.edu.ua , http://new.mmf.lnu.edu.ua/employee/buhrii_o_m ; Головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Консультації з питань навчання по дисципліні відбуваються	Консультації в день проведення лекцій/практичних занять (за попередньою домовленістю): ауд. 267, головний корпус ЛНУ ім. І. Франка, вул. Університетська 1, м. Львів, 79000
Сторінка курсу	https://new.mmf.lnu.edu.ua/course/mtavd-112-mag
Інформація про дисципліну	Дисципліна “Методи і техніки аналізу великих даних” є вибірковою дисципліною з спеціальності 112-Статистика для освітньої програми “Статистичний аналіз даних”, яка викладається в 1-му семестрі в обсязі 6-ох кредитів (за Європейською Кредитно-Трансферною Системою ECTS).
Коротка анотація дисципліни	Курс розроблено для ознайомлення студентів з основними технологіями для роботи з великими даними та їх використанню для аналізу даних.
Мета та цілі дисципліни	<i>Мета:</i> надати студентам знання про сучасні технології великих даних, які використовуються для обробки даних <i>Цілі:</i> забезпечити знайомство студентів з загальною теорією використання засобів та технологій для роботи з великими даними і сформувані навички їх практичного створення; навчити студента базових інструментів Apache Hadoop, Pig, Apache Spark ELK, EFK, HDFS, RabbitMQ, Kafka, ZeroMQ
Література для вивчення дисципліни	1) Документація Apache Hadoop [Електронний ресурс] // Apache Hadoop. – 2023. – Режим доступу до ресурсу: https://hadoop.apache.org/docs/stable/ . 2) Документація Apache Spark [Електронний ресурс] // Apache Spark. – 2023. – Режим доступу до ресурсу: https://spark.apache.org/docs/latest/ . 3) Документація HBase [Електронний ресурс] // HBase. – 2023. – Режим доступу до ресурсу: https://hbase.apache.org/book.html . 4) RabbitMQ [Електронний ресурс] // RabbitMQ. – 2023. – Режим доступу до ресурсу: https://www.rabbitmq.com/documentation.html . 5) Ifeyinwa A., Friday H. Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. Nigeria: Abakaliki, 2019. 6) Marz N., Warren J. Big Data: Principles and best practices of scalable realtime data systems, Manning Publ. Co., NY, 2015. 7) Cielen D., Meysman A., Ali M. Introducing Data Science: Big Data, Machine

	learning. Manning Publ. Co., NY, 2016.
Обсяг курсу	Загальний обсяг: 180 годин. Аудиторних занять: 64 год., з них 32 год. лекційних та 32 год. практичних занять. Самостійної роботи: 116 год.
Очікувані результати навчання	У результаті вивчення даного курсу студент повинен: 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; вміти: володіти базовими знаннями Apache Hadoop, Pig, Apache Spark, ELK, EFK, HDFS, RabbitMQ, Kafka, ZeroMQ.
Ключові слова	Великі дані, big data, Apache Spark, Kafka, Map reduce.
Формат курсу	Очний
Теми	Див. Схема курсу
Підсумковий контроль, форма	Залік у кінці семестру
Пререквізити	Для вивчення даного курсу студентам потрібні базові знання з: - основи програмування; - алгоритмізація і програмування; - об'єктно-орієнтоване програмування; - нереляційні та розподілені бази даних.
Навчальні методи та техніки, які будуть використовуватися під час викладання курсу	Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація); дедуктивні методи на основі узагальнень; евристичні методи (проблемна лекція); інтерактивні методи (дискусія)
Необхідне обладнання	Комп'ютер із необхідним програмним забезпеченням, доступ до Internet мережі.
Критерії оцінювання (окремо для кожного виду навчальної діяльності)	Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням: • Змістовий модуль 1: 7% семестрової оцінки за активну роботу на заняттях, 12% семестрової оцінки за самостійну роботу і виконання практичних завдань, 5% семестрової оцінки за колоквиум, максимальна кількість балів 24. • Змістовий модуль 2: 7% семестрової оцінки за активну роботу на заняттях, 14% семестрової оцінки за самостійну роботу і виконання практичних завдань, 5% семестрової оцінки за колоквиум, максимальна кількість балів 26. • підсумковий тест: 50% семестрової оцінки, максимальна кількість балів 50. Підсумкова максимальна кількість балів 100. 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.

	<p>чесності 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 практичні 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 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 і т. ін.</p> <p>Жодні форми порушення академічної доброчесності не толеруються.</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 їх 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.</p> <p>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>
<p>Опитування</p>	<p>Анкету-оцінку з метою оцінювання якості курсу буде надано по завершенню курсу.</p>

**Схема курсу “ Методи і техніки аналізу великих даних ”
для студентів спеціальності 112 – Статистика**

Тижні	Лекційний курс		Практичні заняття		К-сть год СР	Літе- ратура
	Назва теми	К-сть год	Назва теми	К-сть год		
1	2	3	4	5	6	7
1	Вступ. Великі дані. Загальне використання. Принцип роботи.	2	Принципи побудови баз даних	2	7	[1]-[7], Сайт курсу
2	Проблеми великих даних. Спільні характеристики. Парадигма Map-Reduce.	2	Встановлення та розгляд технологій великих даних	2	7	[1]-[7], Сайт курсу
3	Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування.	2	Розгляд технології Hadoop. Реалізація простої оброки даних.	2	7	[1], Сайт курсу
4	Загальний розгляд технології Hadoop. Файлова система HDFS та її архітектура.	2	Використання Apache Pig в Apache Hadoop.	2	7	[1], Сайт курсу
5	Розгляд системи Map reduce. Розгляд платформи Apache Pig.	2	Використання Apache Hive в Apache Hadoop.	2	7	[1], Сайт курсу
6	Apache Hive. Hive QL. Архітектра MRv1. Планувальник YARN.	2	Розгляд бази даних Hbase. Її модель даних. Життєвий цикл. Архітектура.	2	7	[1], Сайт курсу
7	MongoDB та Neo4j. Архітектура баз даних та їх використання у великих даних.	2	Розробка проекту з вивченими технологіями.	2	7	[3], Сайт курсу
8	Колоквіум № 1	2	Підсумкове заняття ЗМ 1	2	7	[1]-[7], Сайт курсу
9	База даних Cassandra. Принцип її роботи та використання.	2	Розгляд технології Apache Spark. Реалізація простої оброки даних	2	7	[3], Сайт курсу
10	Apache Spark. Розгляд архітектури.	2	Налаштування легування з використанням ELK	2	7	[3], Сайт курсу
11	Apache Spark. Порівняння з технологією Hadoop. Принцип роботи.	2	Налаштування легування з використанням EFK	2	7	[2], Сайт курсу
12	ELK стек. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи.	2	Розробка проекту з вивченими технологіями.	2	7	[2], Сайт курсу
13	EFK стек. Elasticsearch. Fluentd. Kibana. Опис. Взаємозв'язок. Викорис-	2	Використання Apache Kafka для передавання даних між сервісами	2	7	[5], Сайт курсу

	тання. Принцип роботи. Різниця між ELK і EFK.					
14	RabbitMQ. Опис. Взаємозв'язок. Використання.	2	Використання RabbitMQ для передавання даних між сервісами.	2	7	[5], Сайт курсу
15	Apache Kafka і RabbitMQ. Взаємодія сервісів через меседж брокери.	2	Розробка проекту з вивченими технологіями.	2	7	[4], [5], Сайт курсу
16	Колоквіум № 2	2	Підсумкове заняття	2	11	[1]-[7], Сайт курсу
	Разом	32		32	66	
	Викладач: Бугрій О.М.		Викладач: Бугрій О.М.			