

SYLLABUS / РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ
ДИСЦИПЛІНИ

Course of study (code) / Назва дисципліни (шифр)	Великі дані IT122ВД
Academic year / Навчальний рік - Семестр	2021/2022 8 семестр
Course of study / Назва спеціальності	122 Комп'ютерні науки
Educational program / Освітня програма Education - ECTS / Рівень – Кредити Status / Статус Learning language / Мова навчання	«Комп'ютерні науки» Перший (бакалаврський) рівень 3 ECTS Обов'язкова Українська
Author / Укладач	Белінський Андрій Олександрович Державний університет економіки і технологій e-mail: krivogame@gmail.com . https://orcid.org/0000-0002-2821-2895 моб. +380984341701
Консультації	Четвер, 12.30-13.50

A. OBJECTIVE OF THE SUBJECT / МЕТА ТА ЗАВДАННЯ ДИСЦИПЛІНИ

Мета дисципліни “Великі дані” полягає в тому, щоб надати студентам можливість отримувати інформацію та приймати рішення на основі даних, використовуючи великі обсяги даних, які зазвичай є складними, різноманітними та створюються з високою швидкістю.

Завдання дисципліни “Великі дані” є наступними:

1. Розуміння фундаментальних концепцій і принципів великих даних, включаючи збір, зберігання, обробку, аналіз і візуалізацію даних.
2. Ознайомлення з різними типами даних, які зазвичай зустрічаються у великих даних, такими як структуровані, напівструктуровані та неструктуровані дані.
3. Навчитися працювати з великими даними за допомогою відповідних інструментів і технологій, включаючи сховища даних, бази даних Hadoop, NoSQL і інструменти візуалізації даних.
4. Набуття навичок аналізу даних і статистичного моделювання для виявлення закономірностей, тенденцій і розуміння великих даних.
5. Розвиток досвіду в алгоритмах і техніках машинного навчання для побудови прогнозних моделей і отримання цінної інформації з великих даних.
6. Навчитися застосовувати етичні та правові принципи під час роботи з великими даними, включаючи питання, пов'язані з конфіденційністю даних, безпекою та правом власності.
7. Розуміння бізнес-застосувань Big Data та їх впливу на різні галузі, включаючи охорону здоров'я, фінанси, маркетинг і транспорт.

Загалом, мета цього курсу полягає в тому, щоб надати студентам всебічне розуміння дисципліни “Великі дані” і навички, необхідні для застосування концепцій, інструментів і методів Big Data до проблем реального світу.

B. SUBJECT PROGRAM / ПРОГРАМА ДИСЦИПЛІНИ

Тема 1. Вступ до великих даних.

Що таке Hadoop
Історія Hadoop
Екосистема Hadoop
Дистрибутиви Hadoop

Тема 2. Розподілена файлова система HDFS

Архітектура HDFS
Shell-команди HDFS
Java API HDFS

Тема 3. Вступ до MapReduce

Парадигма MapReduce
Фреймворк MapReduce
Java API MapReduce



Hadoop Streaming

Тема 4. Розв'язування задач за допомогою MapReduce

Алгоритми на MapReduce

Реляційні функції

Розрахунок TF-IDF

Тема 5. Алгоритми на графах у MapReduce

Графи в MapReduce

Пошук найкоротшого шляху

PageRank

Проблеми MR-алгоритмів на графах

Тема 6. Pig та Hive

Pig

Основні оператори PigLatin

Hive

Pig vs Hive

Тема 7. NoSQL бази даних: HBase та Cassandra

Способи зберігання даних

NoSQL

Вступ до HBase

Архітектура HBase

Cassandra

Тема 8. Spark

Основні поняття Spark

Оператори Spark

Фреймворк Spark

Тема 9. YARN. MapReduce 2.0

Поняття YARN

Основні компоненти YARN

Концепт MapReduce 2.0

C. LIST OF COMPETENCIES AND STUDIES TARGETED RESULTS / ПЕРЕЛІК КОМПЕТЕНТНОСТЕЙ ТА ПРОГРАМНИХ РЕЗУЛЬТАТІВ НАВЧАННЯ

Загальні компетентності
(ЗК)

- ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.
- ЗК2. Здатність застосовувати знання у практичних ситуаціях.
- ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.
- ЗК4. Здатність спілкуватися державною мовою як усно, так і письмово.
- ЗК6. Здатність вчитися й оволодівати сучасними знаннями.
- ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- ЗК9. Здатність працювати в команді.
- ЗК 12. Здатність оцінювати та забезпечувати якість виконуваних робіт.
- ЗК13. Здатність діяти на основі етичних міркувань.
- ЗК 14. Здатність реалізувати свої права і обов'язки як члена суспільства, усвідомлювати цінності громадянського (вільного демократичного) суспільства та необхідність його сталого розвитку, верховенства права, прав і свобод людини і громадянина в Україні.
- ЗК15. Здатність зберігати та примножувати моральні, культурні, наукові цінності і досягнення суспільства на основі розуміння історії та закономірностей розвитку предметної області, її місця у загальній системі знань про природу і суспільство та у розвитку суспільства, техніки і технологій, використовувати різні види та форми рухової активності для активного відпочинку та ведення здорового способу життя.
- СК4. Здатність використовувати сучасні методи математичного моделювання об'єктів, процесів і явищ, розробляти моделі й алгоритми чисельного розв'язування задач математичного моделювання, враховувати похибки наближеного чисельного розв'язування професійних задач.

Спеціальні (фахові)
компетентності (СК)



SYLLABUS / РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ
ДИСЦИПЛІНИ

Програмні результати
навчання(ПР)

СК5. Здатність здійснювати формалізований опис задач дослідження операцій в організаційно-технічних і соціально-економічних системах різного призначення, визначати їх оптимальні розв'язки, будувати моделі оптимального управління з урахуванням змін економічної ситуації, оптимізувати процеси управління в системах різного призначення та рівня ієрархії.

СК7. Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів.

СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу в процесі розв'язування прикладних задач.

СК14. Здатність застосовувати методи та засоби забезпечення інформаційної безпеки, розробляти й експлуатувати спеціальне програмне забезпечення захисту інформаційних ресурсів об'єктів критичної інформаційної інфраструктури.

СК16. Здатність реалізовувати високопродуктивні обчислення на основі хмарних сервісів і технологій, паралельних і розподілених обчислень при розробці й експлуатації розподілених систем паралельної обробки інформації.

ПР1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук

ПР3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.

ПР4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

ПР12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірного аналізу даних на основі технологій Data Mining, Text Mining, Web Mining.

ПР16. Розуміти концепцію інформаційної безпеки, принципи безпечного проектування програмного забезпечення, забезпечувати безпеку комп'ютерних мереж в умовах неповноти та невизначеності вихідних даних.

D. SEMESTER PLAN / СЕМЕСТРОВИЙ ПЛАН

Тиждень Дата	Тема, план/короткі тези	Форма діяльності (заняття), години, формат	Завдання для СРС (література, ресурси в інтернеті, презентація, відеокурси)
За розкладом	Тема 1. Вступ до великих даних. Що таке Hadoop Історія Hadoop Екосистема Hadoop Дистрибутиви Hadoop	лекція (1 год), практичні заняття (2 год.)	Опрацювання літератури: основна 1 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 2. Розподілена файлова система HDFS Архітектура HDFS Shell-команди HDFS Java API HDFS	лекція (1 год), практичні заняття (2 год.)	Опрацювання літератури: основна 1 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 3. Вступ до MapReduce Парадигма MapReduce	лекція (1 год),	Опрацювання літератури:

SYLLABUS / РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ
ДИСЦИПЛІНИ

	Фреймворк MapReduce Java API MapReduce Hadoop Streaming	практичні заняття (2 год.)	основна 1, 2, 4 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 4. Розв'язування задач за допомогою MapReduce Алгоритми на MapReduce Реляційні функції Розрахунок TF-IDF	лекція (1 год), практичні заняття (2 год.)	Опрацювання літератури: основна 1, 2, 4 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 5. Алгоритми на графах у MapReduce Графи в MapReduce Пошук найкоротшого шляху PageRank Проблеми MR-алгоритмів на графах	лекція (1 год), практичні заняття (2 год.)	Опрацювання літератури: основна 1, 2, 4 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 6. Pig та Hive Pig Основні оператори PigLatin Hive Pig vs Hive	лекція (1 год), практичні заняття (2 год.)	Опрацювання літератури: основна 5, 6 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 7. NoSQL бази даних: HBase та Cassandra Способи зберігання даних NoSQL Вступ до HBase Архітектура HBase Cassandra	лекція (1 год) практичні заняття (3 год.)	Опрацювання літератури: основна 3 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 8. Spark Основні поняття Spark Оператори Spark Фреймворк Spark	лекція (2 год) практичні заняття (3 год.)	Опрацювання літератури: основна 1-6 додаткова 1-4 Інтернет-ресурси: 1-4
	Тема 9. YARN. MapReduce 2.0 Поняття YARN Основні компоненти YARN Концепт MapReduce 2.0	лекція (1 год) практичні заняття (2 год.)	Опрацювання літератури: основна 1-6 додаткова 1-4 Інтернет-ресурси: 1-4

E. BASIC LITERATURE (OBLIGATORY TEXTBOOKS) / ОСНОВНА ЛІТЕРАТУРА (ОБОВ'ЯЗКОВІ ПІДРУЧНИКИ)

1. Hadoop: The Definitive Guide. Tom White. O'Reilly Media; 4rd Edition
2. MapReduce Design Patterns. Donald Miner, Adam Shook. O'Reilly Media
3. HBase: The Definitive Guide. Lars George. O'Reilly Media; 1 edition
4. Data-Intensive Text Processing with MapReduce. Jimmy Lin and Chris Dyer (April, 2010)
5. Programming Pig. Alan Gates. O'Reilly Media; 1st Edition
6. Programming Hive. Edward Capriolo, Dean Wampler, Jason Rutherglen. O'Reilly Media; 1 edition

F. COMPLEMENTARY LITERATURE / ДОДАТКОВА ЛІТЕРАТУРА

1. "Big Data Technologies: A Survey." Journal of King Saud University - Computer and Information Sciences 30, no. 4 (2018): 431–48. <https://doi.org/10.1016/j.jksuci.2017.06.001>.
2. Ahonen, Pertti. "Institutionalizing Big Data Methods in Social and Political Research , Institutionalizing Big Data Methods in Social and Political Research." Big Data & Society 2, no. 2 (2015): 2053951715591224.
3. "Beyond the Hype: Big Data Concepts, Methods, and Analytics." International Journal of Information Management 35, no. 2 (2015): 137–44. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>.
4. Hashem, Ibrahim Abaker Targio, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. "The Rise of 'Big Data' on Cloud Computing: Review and Open Research Issues." Information Systems 47 (2015): 98–115. <https://doi.org/10.1016/j.is.2014.07.006>.

Інформаційні ресурси:

1. Introduction to Big Data. URL: <https://www.coursera.org/learn/big-data-introduction?specialization=big-data>
2. Big Data Modeling and Management Systems. URL: <https://www.coursera.org/learn/big-data-management?specialization=big-data>

3. Big Data Integration and Processing. URL: <https://www.coursera.org/learn/big-data-integration-processing?specialization=big-data>
4. Machine Learning With Big Data. URL: <https://www.coursera.org/learn/big-data-machine-learning?specialization=big-data>

G. THE MOST IMPORTANT PUBLICATIONS OF THE AUTHOR(S) CONCERNING PROPOSED CLASSES / ОСНОВНІ ПУБЛІКАЦІЇ АВТОРА, ЩО ПОВ'ЯЗАНІ З ТЕМАТИКОЮ ЗАПЛАНОВАНИХ ЗАНЯТЬ

1. Bielinskyi, A.O., Serdyuk, O.A., Semerikov, S.O., Soloviev, V.N. Econophysics of cryptocurrency crashes: A systematic review (2021) CEUR Workshop Proceedings, 3048, pp. 31-133.
2. Bielinskyi, A.O., Hushko, S.V., Matviychuk, A.V., Serdyuk, O.A., Semerikov, S.O., Soloviev, V.N. Irreversibility of financial time series: A case of crisis (2021) CEUR Workshop Proceedings, 3048, pp. 134-150.
3. Soloviev, V., Bielinskyi, A., Solovieva, V. Entropy analysis of crisis phenomena for DJIA index (2019) CEUR Workshop Proceedings, 2393, pp. 434-449.
4. Soloviev, V.N., Belinskiy, A. Complex Systems Theory and Crashes of Cryptocurrency Market (2019) Communications in Computer and Information Science, 1007, pp. 276-297.
5. Bielinskyi, A.O., Soloviev, V.N. Complex network precursors of crashes and critical events in the cryptocurrency market (2018) CEUR Workshop Proceedings, 2292, pp. 37-45.
6. Soloviev, V., Belinskij, A. Methods of nonlinear dynamics and the construction of cryptocurrency crisis phenomena precursors (2018) CEUR Workshop Proceedings, 2104, pp. 116-127.
7. Bielinskyi, A.O., Khvostina, I., Mamanazarov, A., Matviychuk, A., Semerikov, S., Serdyuk, O., Solovieva, V., Soloviev, V.N. Predictors of oil shocks. Econophysical approach in environmental science (2021) IOP Conference Series: Earth and Environmental Science, 628 (1), № 012019.
8. Посилання на сертифікат за завершення курсу «Neural Networks and Deep Learning» [Electronic resource] - Access mode: <https://coursera.org/share/0395ea9b644d52a73c99f7a55d16430c>
9. Посилання на сертифікат за завершення курсу «Text Retrieval and Search Engine» [Electronic resource] - Access mode: <https://coursera.org/share/0c428f0f1b573d58982e7a7ab56dbcb9>
10. Посилання на сертифікат за завершення курсу «Text Mining and Analytics» [Electronic resource] - Access mode: <https://coursera.org/share/3015183244ac0daf4321bc9b3a7887f0>
11. Посилання на сертифікат за завершення курсу «Pattern Discovery in Data Mining» [Electronic resource] - Access mode: <https://coursera.org/share/9cea0fae980858dbe319a7adae626a9e>
12. Посилання на сертифікат за завершення курсу «Cluster Analysis in Data Mining» [Electronic resource] - Access mode: <https://coursera.org/share/83cf3ed7c80223d4d59c5cf24abc4f8a>

H. PREREQUISITE AND POSTREQUISITE / ПРЕРЕКВІЗИТИ ТА ПОСТРЕКВІЗИТИ

Навчальна дисципліна базується на нормативних та за вибором навчальних дисциплінах, зокрема дисципліні «Програмування на Python», і є необхідною для опанування подальших навчальних дисциплін, виконання дипломної роботи.

Знання англійської мови навіть на початковому рівні вітається.

I. SCOPE AND TYPE / КІЛЬКІСТЬ ВІДВЕДЕНИХ ГОДИН ТА ФОРМА ПРОВЕДЕННЯ ЗАНЯТЬ

Форма навчання	Денна	Заочна
Лекції	10	-
Практичні (лабораторні)	20	-
Самостійна робота студента (СРС)	50	-
Індивідуально-консультативна робота (ІКР)	10	-
Курсова робота	-	-
Разом годин	90	-

J. CURRENT AND FINAL EVALUATION / ПОТОЧНЕ ТА ПІДСУМКОВЕ ОЦІНЮВАННЯ

Складові поточного контролю результатів освітньої діяльності здобувачів вищої освіти	Форма підсумкового контролю	
	залік	
	за формами навчання	
	Денна	Заочна
Поточний контроль		

**SYLLABUS / РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ
ДИСЦИПЛІНИ**

Знання, уміння та навички, продемонстровані на аудиторних заняттях	20	-
Виконання модульних контрольних робіт	30	-
Виконання і захист завдань самостійної роботи	50	-
Підсумковий контроль		
	-	-
Усього	100	-

Шкала балів	Оцінка за 4-бальною шкалою	Шкала ECTS
90 – 100	Відмінно	A
80 – 89	Добре	B
70 – 79		C
66 – 69	Задовільно	D
60 – 65		E
21 – 59	незадовільно з можливістю повторного складання екзамену (заліку)	FX
0 – 20	незадовільно з можливістю вивчення дисципліни за індивідуальним графіком у формі додаткової індивідуально-консультативної роботи	F

K. CODE OF CONDUCT OF THE COURSE / КОДЕКС ПОВЕДІНКИ ПІД ЧАС ВИВЧЕННЯ КУРСУ

обов'язків:

- ❖ не запізнюватися на заняття;
- ❖ не пропускати заняття (як лекційні, так і практичні), в разі хвороби мати довідку або її ксерокопію;
- ❖ самостійно опрацювати весь лекційний матеріал та ресурси для самостійної роботи;
- ❖ при підготовці есе, есе-рефлексії на статтю, що запропонована вам для читання (див. семестровий план), оцінюється якість та оригінальність наведених вами аргументів. Есе повинно бути надіслано до 16:00 у попередній день перед семінаром. Усі повинні обов'язково підготувати есе, а його відсутність з будь-яких причин не може бути виправданням.
- ❖ конструктивно підтримувати зворотній зв'язок з викладачем на всіх етапах проходження курсу (особливо під час виконання індивідуальних проектів/курсового проекту);
- ❖ своєчасно і самостійно виконувати всі передбачені програмою лабораторні та практичні завдання;
- ❖ брати очну участь у контрольних заходах

L. METHODS OF CONDUCTING / МЕТОДИ НАВЧАННЯ

Для формувань уміння та навичок застосовуються такі методи навчання:

- вербальні/словесні (лекція, пояснення, розповідь, бесіда, інструктаж);
- наочні (спостереження, ілюстрація, демонстрація);
- практичні (різні види вправлення, виконання графічних робіт, проведення експерименту, практики);
- пояснювально-ілюстративний або інформаційно-рецептивний, який передбачає пред'явлення готової інформації викладачем та її засвоєння студентами;
- репродуктивний, в основу якого покладено виконання різного роду завдань за зразком;
- метод проблемного викладу.
- частково-пошуковий або евристичний.
- Дослідницький.

M. TOOLS, EQUIPMENT AND SOFTWARE / ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

мультимедійний клас з ПК, цифровий проектор, комп'ютерний клас або власні ноутбуки

Zoom – хмарна платформа для відео і аудіо конференцій та вебінарів.

Telegram – програма месенджер.

viber – програма для відео та голосового зв'язку.

ZELIS - система призначена для тестування знань студентів в двох режимах: автоматизований контроль знань та тестування по бланкам.

**N. STUDENT RESOURCES, MOOC PLATFORMS / ЦИФРОВІ РЕСУРСИ ДЛЯ СТУДЕНТІВ ТА ВІДКРИТІ
ДИСТАНЦІЙНІ ОНЛАЙН КУРСИ**

Студентам пропонується доступ до навчальних матеріалів дисципліни - moodle.kneu.dp.ua:

[Coursera](#) – безкоштовні онлайн-курси з різних дисциплін, у разі успішного закінчення яких користувач отримує сертифікат про проходження курсу.

[EdX](#) – онлайн-курси від закладів вищої освіти.

[Prometheus](#) – український громадський проєкт масових відкритих онлайн-курсів.



**SYLLABUS / РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ
ДИСЦИПЛІНИ**

O. FEEDBACK/ ЗВОРОТНІЙ ЗВ'ЯЗОК

Електронні листи є найкращим способом зв'язатися з керівником курсу, і, будь ласка, додайте шифр групи в темі листа. Якщо ви надішлете мені електронне повідомлення, надайте мені, принаймні, 24 години, щоб відповісти. Якщо ви не отримуєте відповідь, відправте листа повторно.

P. ACADEMIC HONESTY/ АКАДЕМІЧНА ДОБРОЧЕСНІСТЬ

Державний університет економіки і технологій очікує від студентів розуміння та підтримання високих стандартів академічної чесності. Приклади академічної не доброчесності включають такі: плагіат, зловживання інформацією із застарілих джерел мережі. Очікується, що вся робота, виконана відповідно до вимог курсу, є власною роботою студента. Під час підготовки роботи, яка відповідає вимогам курсу, студенти повинні відрізняти власні ідеї від інформації, отриманої з інших джерел. Без попереднього письмового схвалення викладачем, студенти можуть не подавати один і той же звіт двічі.

Обов'язково вказати на положення про доброчесність й зробити гіперпосилання на сайт Університету (Положення про академічну доброчесність у Державному університеті економіки і технологій. Затверджено Вченою радою Державного університету економіки і технологій, Протокол № 5 від 25 листопада 2021 р.) https://www.duet.edu.ua/uploads/normbase/243/pol_AD.pdf

APPROVED / ЗАТВЕРДЖЕНО

Рішенням кафедри «Економіки та цифрового бізнесу» Державного університету економіки і технологій - протокол № 1 від 05 жовтня 2022 року

Укладач

Андрій БЕЛІНСЬКИЙ.

ЗАТВЕРДЖЕНО:

Кафедрою економіки та цифрового бізнесу
Протокол № 1 від 05 жовтня 2022 року
В.о. завідувача кафедри

Вікторія СОЛОВЙОВА

Науково-методичною радою Державного університету
економіки і технологій
Протокол № 4 від 30 листопада 2022 року
Голова науково-методичної ради

Валентин ОРЛОВ