

# НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Кафедра комп'ютерних наук

**«ЗАТВЕРДЖУЮ»**  
Декан факультету інформаційних технологій  
Олена ГЛАЗУНОВА  
«12» серпня 2023 р.

**«СХВАЛЕНО»**  
на засіданні кафедри комп'ютерних наук  
Протокол № 12 від «01» 06 2023 р.

Завідувач кафедри  
Белла ГОЛУБ

**«РОЗГЛЯНУТО»**  
Гарант ОП Інформаційні управляючі системи і технології  
Гарант ОП  
Олег ГУСТЕРА

## РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

### ТЕХНОЛОГІЇ BIG DATA

**Спеціальність** – 122 «Комп'ютерні науки»

**Освітня програма** – «Інформаційні управляючі системи і технології»

**Факультет** інформаційних технологій

**Розробник:** професор кафедри комп'ютерних наук, д.т.н., проф. Хиленко В. В.

Київ 2023

## Опис навчальної дисципліни Технології Big Data

| Галузь знань, спеціальність, освітній ступінь                        |                                                |
|----------------------------------------------------------------------|------------------------------------------------|
| Освітній ступінь                                                     | Магістр                                        |
| Спеціальність                                                        | 122 «Комп'ютерні науки»                        |
| Освітня програма                                                     | «Інформаційні управляючі системи і технології» |
| Характеристика навчальної дисципліни                                 |                                                |
|                                                                      |                                                |
| Вид                                                                  | Обов'язкова                                    |
| Загальна кількість годин                                             | 120                                            |
| Кількість кредитів ECTS                                              | 4                                              |
| Кількість змістових модулів                                          | 2                                              |
| Курсовий проект (робота) (за наявності)                              | -                                              |
| Форма контролю                                                       | Іспит                                          |
| Показники навчальної дисципліни для денної та заочної форми навчання |                                                |
|                                                                      |                                                |
|                                                                      | денна форма навчання                           |
| Курс (рік підготовки)                                                | 1                                              |
| Семестр                                                              | 1                                              |
| Лекційні заняття                                                     | 30                                             |
| Практичні, семінарські заняття                                       |                                                |
| Лабораторні заняття                                                  | 30                                             |
| Самостійна робота                                                    | 60                                             |
| Індивідуальні завдання                                               |                                                |
| Кількість тижневих аудиторних годин для денної форми навчання        | 12                                             |

## 1. Мета, завдання та компетентності навчальної дисципліни

**Мета:** Навчити студента використовувати відомі методи роботи з великими об'ємами даних при вирішенні професійних (виробничих) задач. Ознайомити студента з відповідними існуючими програмними продуктами. Надати базові знання щодо математичного апарату (математичних методів та алгоритмів) які використовуються для обробки великих об'ємів даних, методів аналітичної обробки великих даних та імітаційного моделювання. Навчити студента формулювати задачі створення баз даних, баз знань, сховищ даних для розв'язання аналітичних задач (системного аналізу / Data Analysis, імітаційного моделювання) з урахуванням особливостей предметної області.

У результаті вивчення навчальної дисципліни студент повинен

### **Знати:**

- принципи та методи побудови інформаційних моделей;
- принципи та методи побудови сховищ даних та баз знань;
- принципи і підходи до організації процедур імітаційного моделювання,
- особливості математичних методів, математичного і алгоритмічного забезпечення яке може використовуватись для роботи з великими об'ємами даних;
- методології аналізу результатів імітаційного моделювання;
- методологію та методи оцінки достовірності отриманих результатів;
- методи побудови та дослідження математичних моделей складних динамічних систем.

### **Уміти:**

- формулювати задачу обробки великих інформаційних масивів відповідно до задачі, що вирішується;
- формувати бази знань (сховища даних) з урахування специфіки предметної галузі досліджень;
- розробляти структуру інформаційної моделі об'єкта дослідження;
- будувати інформаційну модель (модель інформаційних потоків) об'єкту або процесу, який вивчається;
- виконати реструктуризацію моделі або, при необхідності, її декомпозицію;
- оцінити рівень достовірності отриманих результатів;
- модифікувати побудовані інформаційні моделі у відповідності до вимог і цілей замовника досліджень з метою створення спектру моделей;
- використовувати існуючі на ринку інструментальні та програмно-алгоритмічні засоби, що можуть бути використані для вирішення задач обробки великих інформаційних масивів.

**Інтегральна компетентність (ІК):**

Здатність розв'язувати задачі в галузі професійної та/або дослідницько-інноваційної діяльності у сфері комп'ютерних наук, що передбачає як вільне володіння наявними знаннями, так і спроможність їх застосування у професійній практиці.

**Загальні компетентності (ЗК):**

ЗК3. Здатність спілкуватися державною мовою як усно, так і письмово.

**Фахові (спеціальні) компетентності (СК):**

СК1. Розуміння теоретичних засад комп'ютерних наук для об'єктивного оцінювання можливостей використання обчислювальної техніки в певних процесах людської діяльності і визначення перспективних інформаційних технологій, у тому числі, у природоохоронній галуззі.

СК6. Здатність збирати і аналізувати дані (включно з великими), насамперед, пов'язаних з природоохоронною галуззю, для забезпечення якості прийняття рішень.

СК8. Здатність застосовувати існуючі і розробляти нові алгоритми

СК11. Здатність розробляти та адмініструвати бази даних та знань, володіти сучасними теоріями та моделями даних та знань, методами їх інтерактивної та автоматизованої розробки, технологіями обробки та візуалізації.

**Програмні результати навчання (РН):**

РН8. Розробляти та забезпечувати заходи з моніторингу, оптимізації, технічного обслуговування, виявлення відмов тощо.

**2. Програма та структура навчальної дисципліни для:  
повного терміну денної форми навчання.**

**ЗМІСТОВНИЙ МОДУЛЬ 1. ФОРМУЛЮВАННЯ І ФОРМАЛІЗАЦІЯ  
ЗАДАЧ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ**

**ТЕМА 1 ВСТУП ДО ДИСЦИПЛІНИ. СУЧАСНІ ТЕНДЕНЦІЇ В РОЗВИТКУ СИСТЕМ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ.**

Цілі і задачі курсу «Біг Дейта»; зв'язок з іншими навчальними курсами та базовою системою знань, яка необхідна для вирішення задач, що потребують використання великих обсягів даних (задачі: системного аналізу / «Data analysis, імітаційного моделювання, машинного навчання, штучного інтелекту, аналітичних задач). Бази знань та сховища даних. Загальні поняття та відомості. Сучасні тенденції в розвитку систем обробки великих обсягів даних. Системи розподіленої обробки інформації. Технології блокчейн.

## ТЕМА 2. ВИКОРИСТАННЯ СУЧАСНИХ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ ДЛЯ ФОРМУВАННЯ ІНФОРМАЦІЙНОЇ БАЗИ (СПЕКТРУ) МОДЕЛЕЙ. МАСШТАБУВАННЯ БД

Апаратні засоби для обробки великих обсягів даних, суперкомп'ютери та багатопроцесорні обчислювальні комплекси. Використання хмарних технологій, використання існуючих і формування індивідуальних вузькопрофільних баз знань, використання існуючих і формування сховищ даних для створення спектру моделей імітаційного моделювання екологічних систем, перспективні програмні продукти провідних софтверних компаній та їх використання для моделювання предметної області.

## ТЕМА 3. ПАРСІНГ. ІМІТАЦІЙНИЙ ЕКСПЕРИМЕНТ І ПОБУДОВА ІМІТАЦІЙНИХ СИСТЕМ

Формування вимог до системи, розробка структури системи, врахування факторів невизначеності, формування вимог до результатів імітаційних експериментів; класифікація вимог; стратегії вибору рішення; аналіз очікуваних результатів; аналіз невідповідності; аналіз невизначеності цілей; оптимізація процесу вибору рішення при побудові імітаційної системи.

## ТЕМА 4. СТИСНЕННЯ ДАНИХ. МАТЕМАТИЧНА ФОРМАЛІЗАЦІЯ ЗАДАЧ В ІМІТАЦІЙНОМУ МОДЕЛЮВАННІ

Алгоритми стиснення даних. Повнота моделі; ясність (недвозначність, можливість використання бібліотечних програмних продуктів); коректність і несуперечність; масштабованість; економічна доцільність обраного типу моделі відповідно до вимог замовника; відповідність до вагових коефіцієнтів замовника.

## ТЕМА 5. АНАЛІЗ ДОСТОВІРНОСТІ ОТРИМАНИХ ДАНИХ

Аналіз завдань замовника до можливої похибки імітаційного моделювання, аналіз база даних щодо об'єкту дослідження, аналіз відповідності обраного класу математичних моделей з точки зору забезпечення ефективності імітаційного моделювання, аналіз структурних похибок, аналіз похибок обчислень, аналіз невідповідності.

## ТЕМА 6. РЕСТРУКТУРІЗАЦІЯ МОДЕЛЕЙ. ДЕКОМПОЗИЦІЯ

Методи корегування моделей у відповідності до цілі досліджень, існуюче математичне забезпечення для розв'язання окресленого кола задач, методи декомпозиції, реструктуризація моделей відповідно до вимог та цілей замовника, аналіз проблемної області; аналіз ступеня деталізації математичних моделей, що використовуються.

## **ЗМІСТОВНИЙ МОДУЛЬ 2. ПРОГРАМНО-АЛГОРИТМІЧНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ**

**ТЕМА 7. МАТЕМАТИЧНИЙ АПАРАТ ДЛЯ ВИРІШЕННЯ ЗАДАЧ  
ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ. ОБГРУНТУВАННЯ ДОЦІЛЬНОСТІ  
ВИКОРИСТАННЯ АЛГОРИТМІВ СТИСНЕННЯ ДАНИХ**

Методи системного аналізу, типові задачі системного аналізу; невизначеність цілей, математичне моделювання, випадкові процеси, теорія хаосу, теорія катастроф, математичні методи дослідження операцій.

**ТЕМА 8. ПОБУДОВА СИСТЕМИ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ.  
ВИБІР СТРУКТУРИ**

Оптимізація вибору структури системи відповідно до цілей модельного експерименту, визначення критеріїв відповідності і прийнятності, визначення «суттєвих» параметрів, функціональність системи, діаграма потоків даних, діаграма станів, аналіз впливу факторів невизначеності.

**ТЕМА 9. СУЧАСНІ ПРОГРАМНО-АЛГОРИТМІЧНІ ЗАСОБИ ДЛЯ  
РОБОТИ З ВЕЛИКИМИ ДАНИМИ**

Програмні оболонки, вибір інтерфейсу та мови програмування при створенні програмних продуктів BIG DATA, Apache Hadoop, MapReduce, використання бібліотек та стандартних (ринкових) програмних продуктів, потоки даних.

**ТЕМА 10. ОСОБЛИВОСТІ ЕКОЛОГІЧНИХ СИСТЕМ ЯК ОБ'ЄКТІВ  
ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ BIG  
DATA**

Особливості типових математичних моделей екологічних систем; типові проблемні ситуації процесу імітаційного моделювання екологічних систем; які типи моделей ефективно використовувати у відповідності до цілей імітаційного експерименту, діаграма варіантів використання; діаграма дій; альтернативні мови моделювання.

**ТЕМА 11. МОДИФІКАЦІЯ ІНФОРМАЦІЙНИХ ТА МАТЕМАТИЧНИХ  
МОДЕЛЕЙ.**

Аналіз чутливості моделей до зміни окремих параметрів та коефіцієнтів, оцінка витрат часу при створенні різних груп моделей і оцінка об'єму досліджень (інтенсивність використання) різних моделей відповідно до градації частоти вирішення практичних задач, що вирішуються для екологічних систем, побудова спектру моделей для забезпечення можливості вирішення різнопланових завдань замовника.

|                               |                 |
|-------------------------------|-----------------|
| Назви змістових модулів і тем | Кількість годин |
|-------------------------------|-----------------|

|                                                                                                                                                  | денна форма |              |   |           |         |           | Заочна форма |              |    |         |         |          |
|--------------------------------------------------------------------------------------------------------------------------------------------------|-------------|--------------|---|-----------|---------|-----------|--------------|--------------|----|---------|---------|----------|
|                                                                                                                                                  | усть<br>ого | у тому числі |   |           |         |           | усть<br>ого  | у тому числі |    |         |         |          |
|                                                                                                                                                  |             | л            | п | ла<br>б   | ін<br>д | с.<br>р.  |              | л            | п  | ла<br>б | ін<br>д | с.р<br>. |
| 1                                                                                                                                                | 2           | 3            | 4 | 5         | 6       | 7         | 8            | 9            | 10 | 11      | 12      | 13       |
| <b>Змістовий модуль 1. ФОРМУЛЮВАННЯ І ФОРМАЛІЗАЦІЯ ЗАДАЧ ІМІТАЦІЙНОГО МОДЕЛЮВАННЯ</b>                                                            |             |              |   |           |         |           |              |              |    |         |         |          |
| Тема 1. Вступ до дисципліни. Сучасні тенденції в розвитку систем обробки великих обсягів даних                                                   | 7           | 2            |   |           |         | 5         |              |              |    |         |         |          |
| Тема 2. Використання сучасних інформаційно-комунікаційних технологій для формування інформаційної бази (спектру) моделей. Масштабування БД       | 11          | 4            |   | 2         |         | 5         |              |              |    |         |         |          |
| Тема 3. Парсинг. Імітаційний експеримент і побудова імітаційних систем                                                                           | 11          | 4            |   | 2         |         | 5         |              |              |    |         |         |          |
| Тема 4. Стиснення даних. Математична формалізація задач в імітаційному моделюванні                                                               | 9           | 2            |   | 2         |         | 5         |              |              |    |         |         |          |
| Тема 5. Аналіз достовірності отриманих даних                                                                                                     | 9           | 2            |   | 2         |         | 5         |              |              |    |         |         |          |
| Тема 6. Реструктуризація моделей. Декомпозиція                                                                                                   | 13          | 4            |   | 4         |         | 5         |              |              |    |         |         |          |
| <b>Разом за змістовим модулем 1</b>                                                                                                              | <b>60</b>   | <b>18</b>    |   | <b>12</b> |         | <b>30</b> |              |              |    |         |         |          |
| <b>Змістовий модуль 2. ПРОГРАМНО-АЛГОРИТМІЧНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ОБРОБКИ ВЕЛИКИХ ОБСЯГІВ ДАНИХ</b>                                 |             |              |   |           |         |           |              |              |    |         |         |          |
| Тема 7. Математичний апарат для вирішення задач обробки великих обсягів даних. Обґрунтування доцільності використання алгоритмів стиснення даних | 11          | 2            |   | 4         |         | 5         |              |              |    |         |         |          |
| Тема 8. Побудова системи імітаційного моделювання. Вибір структури                                                                               | 9           | 2            |   | 2         |         | 5         |              |              |    |         |         |          |
| Тема 9. Сучасні програмно-алгоритмічні засоби для роботи з великими даними                                                                       | 11          | 2            |   | 4         |         | 5         |              |              |    |         |         |          |
| Тема 10. Особливості екологічних систем як об'єктів імітаційного моделювання з використанням технологій Big Data                                 | 11          | 2            |   | 4         |         | 5         |              |              |    |         |         |          |
| Тема 11. Модифікація інформаційних та математичних моделей                                                                                       | 18          | 4            |   | 4         |         | 10        |              |              |    |         |         |          |

|                                     |            |                |  |           |  |                |  |  |  |  |  |  |
|-------------------------------------|------------|----------------|--|-----------|--|----------------|--|--|--|--|--|--|
| <i>Разом за змістовим модулем 2</i> | <b>60</b>  | <b>1<br/>2</b> |  | <b>18</b> |  | <b>3<br/>0</b> |  |  |  |  |  |  |
| <b>Усього годин</b>                 | <b>120</b> | <b>3<br/>0</b> |  | <b>30</b> |  | <b>6<br/>0</b> |  |  |  |  |  |  |

### 3. Теми семінарських занять

Не передбачено програмою

### 4. Теми практичних занять

Не передбачено програмою

### 5. Теми лабораторних занять

| № з/п | Назва теми                                                                                                               | Кількість годин |
|-------|--------------------------------------------------------------------------------------------------------------------------|-----------------|
| 1.    | Вибір предмета дослідження при імітаційному моделюванні якого потрібно опрацьовувати великі обсяги даних                 | 2               |
| 2.    | Аналіз достовірності отриманих результатів, хмарні обчислення                                                            | 2               |
| 3.    | Побудова інформаційної моделі, бази даних, бази знань                                                                    | 2               |
| 4.    | Математична та програмно-алгоритмічна формалізація задачі                                                                | 2               |
| 5.    | Аналіз структурних похибок, аналіз похибок обчислень, аналіз невідповідності                                             | 4               |
| 6.    | Методи побудови та застосування альтернативних моделей, аналіз можливості застосування штучного інтелекту                | 4               |
| 7.    | Діаграма потоків даних, діаграма станів                                                                                  | 2               |
| 8.    | Вибір програмних оболонок, бібліотек та обчислювальних методів для комплексного вирішення задач імітаційного моделювання | 4               |
| 9.    | Діаграма варіантів використання; діаграма дій                                                                            | 4               |
| 10.   | Оцінка витрат часу при створенні різних груп моделей                                                                     | 4               |
| Разом |                                                                                                                          | 30              |

### 6. Самостійна робота

| № з/п | Назва теми                                                         | Кількість годин |
|-------|--------------------------------------------------------------------|-----------------|
| 1.    | Сучасні тенденції в розвитку систем обробки великих обсягів даних. | 5               |



|       |                                                                                                                                               |    |
|-------|-----------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.    | Перспективні програмні продукти провідних софтверних компаній та їх використання для моделювання предметної області.                          | 5  |
| 3.    | Вибір екологічних систем як об'єкта імітаційного моделювання і обґрунтування вибраної структури бази даних (сховища даних)                    | 5  |
| 4.    | Математична формалізація задач імітаційного моделювання і побудова інформаційних моделей для конкретних класів екологічних систем.            | 5  |
| 5.    | Аналіз вибору математичного і програмно-алгоритмічного забезпечення для розв'язання сформованих задач                                         | 5  |
| 6.    | Реструктуризація задач імітаційного моделювання                                                                                               | 5  |
| 7.    | Обґрунтування доцільності використання методів на прикладах конкретних задач.                                                                 | 5  |
| 8.    | Обґрунтування доцільності використання методів і розподілених технологій обробки даних та паралельних обчислень на прикладах конкретних задач | 5  |
| 9.    | Альтернативні мови моделювання                                                                                                                | 5  |
| 10.   | Типові проблемні ситуації процесу імітаційного моделювання екологічних систем                                                                 | 5  |
| 11.   | Побудова спектру моделей для забезпечення можливості вирішення різнопланових завдань замовника                                                | 10 |
| Разом |                                                                                                                                               | 60 |

## 7. Методи навчання.

При викладанні навчальної дисципліни використовуються такі методи навчання:

М1. Лекція (проблемна, інтерактивна).

М2. Лабораторна робота.

М3. Проблемне навчання.

М5. Онлайн навчання

М8. Дослідницький метод.

## 8. Форми контролю.

При викладанні навчальної дисципліни використовуються такі методи контролю:

МК1. Тестування.

МК4. Методи усного контролю.

МК5. Екзамен.

## 9. Розподіл балів, які отримують студенти.

Оцінювання знань студента відбувається за 100-бальною шкалою і переводиться в національні оцінки згідно з табл. 1 «Положення про екзамени та заліки у НУБіП України» наказ про уведення в дію від 26.04.2023 р. протокол № 8)

| Рейтинг студента,<br>бали | Оцінка національна<br>за результати складання |               |
|---------------------------|-----------------------------------------------|---------------|
|                           | екзаменів                                     | заліків       |
| 90-100                    | Відмінно                                      | Зараховано    |
| 74-89                     | Добре                                         |               |
| 60-73                     | Задовільно                                    |               |
| 0-59                      | Незадовільно                                  | Не зараховано |

Для визначення рейтингу студента (слухача) із засвоєння дисципліни  $R_{\text{дис}}$  (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи  $R_{\text{нр}}$  (до 70 балів):  $R_{\text{дис}} = R_{\text{нр}} + R_{\text{ат}}$ .

#### 10. Методичне забезпечення

1. Опорні конспекти з курсу. Тестові завдання для поточного та підсумкового контролю. Електронний навчальний курс. (знаходиться в процесі розробки).
2. ПОЛОЖЕННЯ про екзамени та заліки у Національному університеті біоресурсів і природокористування України, затверджене Вченою радою НУБіП України № 8 від «26» квітня 2023 р.

#### 11. Рекомендовані джерела інформації

##### Базові

1. Zgurovsky M.Z. Big Data: Conceptual Analysis and Applications. / M.Z. Zgurovsky, Y.P. Zaychenko // Springer, 2020, 298 p.
2. Akerkar R. Models of Computation for Big Data / R. Akerkar // Springer, 2018, 110 p.
3. Ghavami P. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics / P. Ghavami // CreateSpace Independent Publishing Platform, 2016, 204 p.
4. Feeney K. Engineering Agile Big-Data Systems / K. Feeney, J. Davies, J. Welch, S. Hellmann, C. Dirschl, A. Koller, P. Francois, A. Marciniak // River Publishers, 2018, 436 p.
5. Олійник А. О. Інтелектуальний аналіз даних : навчальний посібник. — Запоріжжя : ЗНТУ, 2012. — 278 с.

6. Ситник В.Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): Навч. пос. — К:КНЕУ, 2007 — 376 с.
7. Rajkumar Buyya. Big Data. Principles and Paradigms. — Elsevier, 2016. — 496p.

### **Додаткові**

1. Kozhevnikov A. V. Estimation of the population density spatial distribution using clutter model / A. V. Kozhevnikov, Ye. A. Krivosheyev // Науковий вісник НГУ – Дніпропетровськ: НГУ, 2011, №2, с. 31 – 36.
2. Thomas Erl. Big Data Fundamentals. Concepts, Drivers & Techniques. — Prentice Hall, 2016. — 235p.
3. Robert Slane. Big Data Essentials. — 2018. — 356p.
4. Phil Simon. Too Big to Ignore: The Business Case for Big Data. — Wiley, 2019. — 257p.

### **Інформаційні ресурси**

1. Microsoft Solutions Framework Core Whitepapers <https://www.microsoft.com/en-us/download/details.aspx?id=3214>
2. Big Data Fundamentals courses. <https://cognitiveclass.ai/learn/big-data>
3. UCI Machine Learning Repository. <https://archive.ics.uci.edu>
4. Introduction to R. Online Course. <https://www.datacamp.com/courses/free-introduction-to-r>