

**Приватний вищий навчальний заклад  
«Буковинський університет»**

**Факультет інформаційних технологій та економіки**

**ЗАТВЕРДЖУЮ**

**Перший проректор**

 Світлана ПЕЛИПЧУК

30 серпня 2024р.


**РОБОЧА ПРОГРАМА  
ОБОВ'ЯЗКОВОЇ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ  
«ТЕХНОЛОГІЇ BIG DATA»**

**для студентів другого (магістерського) рівня  
спеціальності 122 «Комп'ютерні науки»  
галузі знань 12 «Інформаційні технології»**

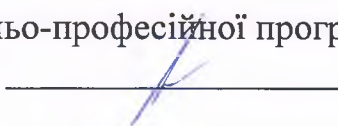
Чернівці

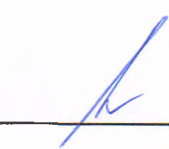
2024

Робочу програму обов'язкової навчальної дисципліни «Технології Big Data» розроблено на основі освітньо-професійної програми «Комп'ютерні науки», навчальних планів підготовки здобувачів вищої освіти другого (магістерського) рівня за спеціальністю 122 «Комп'ютерні науки» та відповідних нормативних документів.

Робочу програму розробила:  
кандидат технічних наук, доцент  О. І. Артеменко

Обговорено та схвалено на засіданні випускової кафедри освітньо-професійної програми «Комп'ютерні науки» - кафедри комп'ютерних систем і технологій протокол № 1 від 29 серпня 2024 року.

Гарант освітньо-професійної програми «Комп'ютерні науки»  
к.т.н., доцент  О. І. Артеменко

Завідувач кафедри  
комп'ютерних систем  
і технологій, к.т.н., доцент  О. І. Артеменко

Робочу програму обговорено та узгоджено на засіданні науково-методичної ради факультету інформаційних технологій та економіки протокол № 1 від 29 серпня 2024 року.

Голова  
науково-методичної ради факультету  Т. В. Штерма

## 1. Анотація дисципліни (призначення навчальної дисципліни).

Дисципліна «Технології Big Data» формує у здобувачів необхідний обсяг теоретичних і практичних знань про великі дані Big Data, методи та засоби їхнього аналізу та обробки, основних принципів побудови систем управління великими даними. Виявлення складу та вивчення можливостей програмного забезпечення сучасних систем, які використовуються для збереження та обробки великих даних в інформаційних системах.

Після вивчення дисципліни студенти повинні:

### **Знати:**

- основні поняття та визначення в теорії Big Data;
- моделі та методи аналізу великих даних;
- сучасні програмні засоби для проектування і розробки систем аналізу великих даних;
- методологію та методи оцінки достовірності отриманих результатів.

### **Вміти:**

- 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 даних.

## 2. Мета та завдання вивчення дисципліни:

**Мета вивчення дисципліни:** формування у слухачів знань, вмінь та навичок аналізу даних, застосування відповідних інформаційних технологій для обробки даних.

### **Завдання:**

- сформулювати знання та отримати практичні навички для використання інформаційних технологій аналізу великих даних при вирішенні задач аналізу даних, розпізнавання образів, пошуку знань та ін;
- огляд методів, програмних продуктів та різних інструментальних засобів, що використовуються для побудови інтелектуальних систем аналізу великих даних.

## 3. Пререквізити.

Знання за програмою єдиного фахового вступного випробування за спеціальністю.

## 4. Компетентності та результати навчання

### **Інтегральна компетентність**

Здатність розв'язувати задачі дослідницького та/або інноваційного характеру у сфері комп'ютерних наук.

### **Загальні компетентності (ЗК)**

ЗК02. Здатність застосовувати знання у практичних ситуаціях.

ЗК07. Здатність генерувати нові ідеї (креативність).

### **Спеціальні (фахові) компетентності:**

СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК04. Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проектних рішень.

СК07. Здатність розробляти програмне забезпечення відповідно до сформульованих вимог з урахуванням наявних ресурсів та обмежень.

СК09. Здатність розробляти та адмініструвати бази даних та знань.

СК12. Здатність застосовувати методи, моделі та алгоритми для виявлення прихованих корисних закономірностей у великих наборах даних щодо вирішення задач прогнозування, керування, класифікації та інтелектуального аналізу.

**Отримуються наступні програмні результати навчання:**

PH8. Розробляти математичні моделі та методи аналізу даних (включно з великим).

PH9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

PH12. Проектувати та супроводжувати бази даних та знань.

PH17. Виявляти та усувати проблемні ситуації в процесі експлуатації програмного забезпечення, формулювати завдання для його модифікації або реінжинірингу.

PH20. Виявляти, ідентифікувати, класифікувати закономірності в великих наборах даних та ефективно їх використовувати в міжгалузевих задачах.

**5. Опис навчальної дисципліни****5.1 Загальна інформація**

Форма навчання	Рік підготовки	Семестр	Кількість		Кількість годин					Вид підсумкового контролю
			кредитів	годин	лекції	практичні	семінарські	лабораторні	самостійна робота	
Денна	1	1	5	150	20	30	-	-	100	екзамен

**5.2 Структура навчальної дисципліни****Теми лекційних занять навчальної дисципліни**

Назви тем	Кількість годин
	денна форма
<b>Тема 1.</b> Вступ до Big Data	2
<b>Тема 2.</b> Типові задачі Big Data та сучасні інструменти їх вирішення	2
<b>Тема 3.</b> HADOOP. Архітектура HADOOP. Робота з HDFS.	2
<b>Тема 4.</b> Парадигма map-reduce.	2
<b>Тема 5.</b> Модель обчислень. Map, Shuffle та Reduce фази. Опис job, task, розбиття операцій на задачі.	2
<b>Тема 6.</b> Життєвий цикл проєкту з аналізу Big Data.	2
<b>Тема 7.</b> Основні технології та інструменти для аналізу великих даних.	2
<b>Тема 8.</b> Аналіз і візуалізація великих даних	2
<b>Тема 9.</b> Базові методи аналізу даних з використанням R.	2
<b>Тема 10.</b> Наукові проблеми аналізу великих даних у складних соціально-економічних системах для досягнення цілей сталого розвитку	2
<b>Усього годин</b>	<b>20</b>

**Теми практичних занять навчальної дисципліни**

Назви тем	Кількість годин
	денна форма
<b>Тема 1.</b> Призначення та функції платформи Apache Hadoop	2
<b>Тема 2.</b> Призначення та функції файлової системи HDFS	2
<b>Тема 3.</b> Адміністрування середовища Apache Hadoop	4
<b>Тема 4.</b> Розроблення застосувань MapReduce в Apache Hadoop.	4
<b>Тема 5.</b> Екосистема Apache Hadoop: функціональні можливості та використання СУБД HBase.	4
<b>Тема 6.</b> Екосистема Apache Hadoop: функціональні можливості	4

та використання Apache Pig.	
<b>Тема 7.</b> Екосистема Apache Hadoop: функціональні можливості та використання Apache Hive.	4
<b>Тема 8.</b> Аналітика та візуалізація великих даних засобами Microsoft Power BI Desktop.	6
<b>Разом</b>	<b>30</b>

#### Теми самостійної роботи студента

Назва теми	денна форма
<b>Тема 1.</b> Менеджер ресурсів кластеру – YARN	20
<b>Тема 2.</b> Застосунок Apache Hive.	20
<b>Тема 3.</b> Застосунок Apache Pig	20
<b>Тема 4.</b> Диспетчер робіт Apache Oozie.	20
<b>Тема 5.</b> Приклади задач аналізу великих даних у складних соціально-економічних системах для досягнення цілей сталого розвитку.	20
<b>Разом</b>	<b>100</b>

\* ІНДЗ – до змістового модуля, або в цілому до навчальної дисципліни визначається викладачем, з урахуванням специфіки дисципліни.

#### 6. Система контролю та оцінювання

##### Освітні технології, методи навчання і викладання навчальної дисципліни

Методи навчання та викладання: лекції, практичні заняття, електронне навчання з використанням системи дистанційного навчання, тестування, усне опитування, виконання завдань самостійної роботи студента.

Об'єктом поточного контролю знань студентів є:

- систематичність, якість та своєчасність виконання і захисту практичних робіт;
- систематичність та своєчасність виконання завдань самостійної роботи студента;
- якість виконання модульних контрольних робіт.

Дисципліна «Технології Big Data» складається з двох модулів, оцінювання яких здійснюється за результатами 8 практичних робіт, двох модульних контрольних робіт та самостійної роботи студента.

Підсумковий контроль знань проводиться у формі екзамена.

На іспит виносяться вузлові питання програми та типові задачі, що потребують творчої відповіді та вміння синтезувати отримані знання і застосувати їх при вирішенні практичних задач. Максимально можлива оцінка за іспит 30 балів.

#### Критерії оцінювання результатів навчання з навчальної дисципліни

##### Шкала оцінювання: національна та ECTS

СУМА БАЛІВ	ОЦІНКА ECTS	ОЦІНКА ЗА НАЦІОНАЛЬНОЮ ШКАЛОЮ	
		екзамен	залік
90-100	A	відмінно	відмінно
82-89	B	добре	добре
74-81	C		
64-73	D	задовільно	задовільно
60-63	E		
35-59	FX	незадовільно	незадовільно
1-34	F		

#### Визначення балів за знання та вміння студентів

**A (90-100) «відмінно»** виставляється за високий рівень знань навчального матеріалу (допускаються деякі неточності), що міститься в основних і додаткових рекомендованих джерелах, вміння аналізувати явища, які вивчаються, у їхньому взаємозв'язку і розвитку, чітко, лаконічно, логічно, послідовно відповідати на поставлені запитання, вміння застосовувати теоретичні положення під

час розв'язання практичних задач, практичні завдання виконані в повному обсязі та своєчасно, оформлення робіт/звітів повністю відповідає вимогам;

**В 82-89 «дуже добре»** виставляється за знання навчального матеріалу *вище від середнього рівня*, включаючи розрахунки та алгоритми, аргументовані відповіді на поставлені запитання (*можлива невелика кількість неточностей*), вміння застосовувати теоретичні положення під час розв'язання практичних задач, практичні завдання виконані в повному обсязі та своєчасно, але можуть містити *невелику кількість неточностей*, оформлення робіт/звітів в цілому відповідає вимогам;

**С 74-81 «добре»** виставляється за *загалом правильне* розуміння навчального матеріалу, включаючи розрахунки та алгоритми, аргументовані відповіді на поставлені запитання, які, однак, містять певні неістотні недоліки, вміння в цілому застосовувати теоретичні положення під час розв'язання практичних задач, практичні завдання виконані в не повному обсязі та/або без дотримання термінів здачі робіт, також можуть містити *невелику кількість неточностей*, оформлення робіт/звітів в цілому відповідає вимогам;

**Д 63-73 «задовільно»** виставляється за посередні знання навчального матеріалу, мало аргументовані відповіді, слабе застосування теоретичних положень під час розв'язання практичних задач, практичні завдання виконані в не повному обсязі та без дотримання термінів здачі робіт, містять *неточності*, оформлення робіт/звітів не відповідає вимогам;

**Е 60-63 «достатньо»** виставляється за слабкі знання навчального матеріалу компонента, неточні або мало аргументовані відповіді, з порушенням послідовності викладення, за слабе застосування теоретичних положень під час розв'язання практичних задач, практичні завдання виконані в не повному обсязі та без дотримання термінів здачі робіт, містять помилки, оформлення робіт/звітів не відповідає вимогам;

**ФХ 35-59 «незадовільно» (з можливістю повторного складання)** виставляється за незнання значної частини навчального матеріалу, істотні помилки у відповідях на запитання, невміння застосувати теоретичні положення під час розв'язання практичних задач, практичні завдання виконані в не повному обсязі та без дотримання термінів здачі робіт, містять значні помилки, оформлення робіт/звітів не відповідає вимогам;

**Ф 1-34 «незадовільно» з обов'язковим повторним вивченням дисципліни)** виставляється за незнання значної частини навчального матеріалу, істотні помилки у відповідях на запитання, невміння орієнтуватися під час розв'язання практичних задач, незнання основних фундаментальних положень, практичні завдання не виконані.

#### Розподіл балів з навчальної дисципліни

Поточний контроль		Іспит	Загальна кількість балів
Модуль I	Модуль II		
35	35	30	100

#### Політика академічної доброчесності

Студент повинен ознайомитися і дотримуватися Положення про забезпечення академічної доброчесності у ПВНЗ «Буковинський університет» (<http://bukuniver.edu.ua/wp-content/themes/bukuniver/docs/polozhennia-pro-zabezpechennia-akademichnoi-dobrochesnosti.pdf>). Він повинен виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем.

Дотримання академічної доброчесності здобувачами освіти передбачає: самостійне виконання навчальних завдань, завдань підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей); посилення на джерела інформації у разі використання ідей, розробок, тверджень, відомостей; дотримання норм законодавства про авторське право і суміжні права; надання достовірної інформації про результати власної навчальної (наукової, творчої) діяльності, використані методики досліджень і джерела інформації. Списування (копіювання тексту) під час виконання письмових робіт та екзаменів заборонені. Користування мобільними пристроями допускається лише з дозволу викладача під час онлайн-тестування та підготовки практичних завдань. Самостійні роботи у вигляді рефератів, доповідей, презентацій повинні мати коректні текстові посилання на використані інформаційні джерела.

#### 7. Література

1. Гороховатський В.О., Творошенко І.С. Методи інтелектуального аналізу та оброблення даних: навч. посібник. – Харків: ХНУРЕ, 2021. – 92 с.

2. Майборода Р. Комп'ютерна статистика. Київ : ВПЦ «Київський університет», 2019. 589 с. URL: <http://probability.univ.kiev.ua/userfiles/mre/cscolor.pdf>
3. Майборода Р. Багатомірна дескриптивна статистика. Методичні рекомендації по курсу. Київ : КНУ ім. Тараса Шевченка, 2020. 35 с. URL: <https://probability.knu.ua/userfiles/mre/descrtasks.pdf>
4. Zgurovsky M.Z. Big Data: Conceptual Analysis and Applications. / M.Z. Zgurovsky, Y.P. Zaychenko // Springer, 2020, 298 p.
5. Rajkumar Buyya. Big Data. Principles and Paradigms. — Elsevier, 2016. — 496p.

### **Інформаційні ресурси в інтернеті**

1. TextMining with R. Share and Discover Knowledge on SlideShare. URL: <http://www.slideshare.net/whitish/textmining-with-r> (date of access: 1.09.2024).
2. R: The R Project for Statistical Computing. R: The R Project for Statistical Computing. URL: <https://www.r-project.org> (date of access: 1.09.2024).
3. Package "timeSeries". The Comprehensive R Archive Network. URL: <https://cran.r-project.org/web/packages/timeSeries/timeSeries.pdf> (date of access: 1.09.2024).
4. Create Elegant Data Visualisations Using the Grammar of Graphics. Create Elegant Data Visualisations Using the Grammar of Graphics • ggplot2. URL: <https://ggplot2.tidyverse.org/> (date of access: 1.09.2024).