

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ**

"ЗАТВЕРДЖУЮ"
Проректор з навчально-методичної роботи
Каріна ПЕМАШКАЛО




Технології роботи з BIG DATA

робоча програма навчальної дисципліни

Галузь знань *12 Інформаційні технології*
Спеціальність *126 Інформаційні системи та технології*
Освітній рівень *перший (бакалаврський)*
Освітня програма *Інформаційні системи та технології*

Статус дисципліни *вибіркова*
Мова викладання, навчання та оцінювання *українська*

Завідувач кафедри
інформатики та комп'ютерної техніки  Сергій УДОВЕНКО

Харків
2021

ЗАТВЕРДЖЕНО

на засіданні кафедри інформатики та комп'ютерної техніки
Протокол №1 від 27 серпня 2021 р.

Розробник:

Кобзев І. В., к.т.н., доц., доцент кафедри інформатики та комп'ютерної техніки

**Лист оновлення та перезатвердження
робочої програми навчальної дисципліни**

Навчальний рік	Дата засідання кафедри - розробника РПНД	Номер протоколу	Підпис завідувача кафедри

Анотація навчальної дисципліни

Навчальна дисципліна "Технології роботи з BIG DATA" є вибірковою навчальною дисципліною та вивчається згідно з навчальним планом підготовки студентів за спеціальністю 126 "Інформаційні системи та технології" першого (бакалаврського) рівня усіх форм навчання. Програму навчальної дисципліни розроблено у відповідності до вимог галузевого стандарту вищої освіти на базі освітньо-професійної програми підготовки бакалавра.

Метою викладання навчальної дисципліни є формування у майбутніх фахівців теоретичних і практичних знань з основ застосування методів інтелектуального аналізу даних для розв'язання спеціалізованих задач попередньої обробки, перетворення, класифікації даних та дослідження процесів у сфері використання інформаційних технологій.

Навчальна дисципліна "Технології роботи з BIG DATA" надає студентам теоретичні знання та практичні навички вибору та застосування математичних моделей та методів опрацювання експериментальних даних та інших спеціально не підготованих великих даних (Big Data), як основи задач математичного моделювання процесів у різних галузях людської діяльності. Завданням навчальної дисципліни "Технології роботи з BIG DATA" є формування у студентів теоретичних знань про основні поняття та методи аналізу і пошуку закономірностей у великих наборах даних та їх застосування при розв'язанні конкретних задач, розвиток математичного та логічного мислення студентів, навичок самостійної роботи над літературою.

Програма навчальної дисципліни передбачає навчання у формі лекцій, лабораторних занять та самостійної роботи студентів. Лекції, лабораторні заняття, індивідуальна робота та консультації проводяться з застосуванням персональних комп'ютерів, локальної мережі та мережі Інтернет у комп'ютерних класах (або з використанням дистанційних технологій навчання в режимі он-лайн). Всі види занять забезпечуються необхідними електронними методичними матеріалами.

З метою підвищення ефективності вивчення навчальної дисципліни студенти мають змогу користуватись системою дистанційного навчання ХНЕУ ім. С. Кузнеця.

Характеристика навчальної дисципліни

Курс	4
Семестр	8
Кількість кредитів ECTS	5
Форма підсумкового контролю	екзамен

Структурно-логічна схема вивчення дисципліни

Пререквізити	Постреквізити
Бази даних, Сучасні інформаційні технології	Дипломний проект

Компетентності та результати навчання за дисципліною

Компетентності	Результати навчання
Здатність аналізувати об'єкт проектування або функціонування та його предметну область	Здатність збирати, формалізувати, систематизувати і аналізувати потреби та вимоги до комп'ютерної системи, що розробляється, експлуатується чи супроводжується. Здатність формалізувати предметну область певного проекту як складну систему з визначенням ключових елементів та зв'язків між ними, мети та критеріїв оцінки її функціонування у вигляді відповідної інформаційної моделі
Здатність проектувати, розробляти та використовувати засоби реалізації інформаційних систем, технологій та	Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття рішень. Здатність застосовувати існуючі і розробляти

інфокомунікацій (методичні, інформаційні, алгоритмічні, технічні, програмні та інші)	нові алгоритми розв'язування задач у галузі комп'ютерних наук: алгоритми розв'язання обчислювальних та логічних задач, алгоритми паралельних та розподілених обчислень, алгоритми аналітичної обробки й інтелектуального аналізу великих даних з оцінкою їх ефективності та складності. 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
Здатність вибору, проектування, розгортання, інтегрування, управління, адміністрування та супроводжування інформаційних систем, технологій та інфокомунікацій, сервісів та інфраструктури організації	Здатність оцінювати якість ІТ-проектів, комп'ютерних і програмних систем різного призначення, володіти методологіями, методами і технологіями забезпечення та вдосконалення якості ІТ-проектів, комп'ютерних та програмних систем на основі міжнародних стандартів оцінки якості програмного забезпечення інформаційних систем, моделей оцінки зрілості процесів розробки інформаційних та програмних систем
Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень	Здатність ініціювати та планувати процеси розробки комп'ютерних систем та програмного забезпечення, включно з його розробкою, аналізом, тестуванням, системною інтеграцією, впровадженням і супроводом. Здатність проводити обчислювальні експерименти, порівнювати результати експериментальних даних і отриманих рішень

Програма навчальної дисципліни

Змістовий модуль 1. Основи Big Data

Тема 1. Головні поняття, функції і задачі великих даних

Основні поняття та визначення. Історія розвитку. Джерела великих даних. Застосування в економіці, бізнесі, сільському господарстві, промисловості. Приклади використання. Великі дані в наукових сферах. Особливості застосування. Вимоги до професії аналітика великих даних.

Тема 2. Методи аналізу великих даних

Основні етапи життєвого циклу. Збір, консолідація і очищення даних. Побудова моделей, роль машинного навчання.

Тема 3. Технології роботи з великими даними

Збір та консолідація даних, "аналітична пісочниця", "озеро даних", пакетна аналітика, аналітика реального часу, гібридна аналітика, робота з СКБД.

Тема 4. Розподілена обробка великих даних та метод Map-Reduce

Збереження великих даних. Технології: in-memory (Oracle Exadata, SAP HANA), NoSQL, Hadoop/MapReduce. Алгоритм MapReduce. Apache Hadoop, HDFS, HBase, YARN, Hive, Pig, Storm як система потокової обробки.

Змістовий модуль 2. Технології та інструменти роботи з Big Data

Тема 5. Google Cloud Platform

Огляд сервісів. Засоби для розробників. Основні компоненти платформи.

Тема 6. Основи BigQuery, Google Cloud Storage

Огляд сервісів. Засоби для розробників. Основні компоненти платформи.

Тема 7. Інструменти аналітичної обробки даних у сховищі великих даних

Apache Spark – інфраструктура кластерних обчислень. Компоненти Spark – MapReduce, HDFS, YARN, Hive, Pig, Zookeeper, Flume та інші. Алгоритм Apache Hive. Мова Pig Latin (Apache Pig). Планування завдань Apache Oozie.

Тема 8. Основні типи віртуалізації

Програмна та апаратна віртуалізація. Контейнерна віртуалізація. Огляд платформ віртуалізації.

Перелік лабораторних занять, а також питань та завдань до самостійної роботи наведено у таблиці "Рейтинг-план навчальної дисципліни".

Методи навчання та викладання

Методи навчання – взаємодія між викладачем і студентом, під час якої відбувається передача та засвоєння знань, умінь і навичок від викладача до студента, а також самостійна та індивідуальна робота студента

Дисципліна "Технології роботи з BIG DATA" передбачає використання таких методів навчання та викладання:

– під час викладання навчального матеріалу: словесні (проблемні лекції за темами 1-8, пояснення, дискусії та інструктаж – під час проведення лабораторних робіт за темами 1-8); наочні (ілюстрування – під час проведення лабораторних робіт за темою 3, 4 та 7, демонстрування – проблемна лекція за темами 3 та 4, самостійне спостереження – виконання лабораторних робіт за темами 5 та 6); практичні (під час виконання лабораторних робіт за темами 2-8);

– методи передачі та сприймання навчальної інформації: індуктивні, дедуктивні та аналітичні (під час виконання індивідуальних завдань лабораторних робіт за темами 2-8);

– за організаційним характером навчання: методи контролю та самоконтролю у навчанні (теми 1-8).

У разі здобуття освіти за дистанційною формою або з використанням дистанційних технологій навчання лекційні заняття проводяться в режимі он-лайн (відео-конференції ZOOM та GoogleMeet).

Порядок оцінювання результатів навчання

ХНЕУ ім. С. Кузнеця використовує накопичувальну (100-бальну) систему оцінювання.

Оцінювання здійснюється за такими видами контролю:

– поточний контроль, що здійснюється протягом семестру під час проведення лекційних, лабораторних занять та контрольних робіт і оцінюється сумою набраних балів (максимальна сума – 60 балів; мінімальна сума, що дозволяє студенту отримати допуск до екзамену – 35 балів);

– підсумковий/семестровий контроль здійснюється у формі екзамену.

Поточний контроль включає оцінювання студентів під час: лекційних занять – активна робота на парі (1 бал за кожне заняття) за умови участі студента в обговоренні питань лекції (максимальна кількість балів за активну роботу на лекціях – 8 балів); лабораторних занять – активна робота на парі (1 бал за кожне заняття) за умови виконання студентом завдань з 8 лабораторних робіт, що згруповані у 8 занять, об'єднаних спільними темами. Захист кожної подвійної лабораторної роботи оцінюється у 5 або 6 балів. Оцінка за лабораторну роботу отримується студентом за наявності звіту з лабораторної роботи, виконаних завдань лабораторної роботи, розгорнутої відповіді на запитання та виконання контрольних прикладів. Максимальна кількість балів за активну роботу на лабораторних заняттях та захист лабораторних робіт складає 44 бали.

Модульний контроль проводиться з урахуванням поточного контролю за відповідний змістовий модуль у вигляді контрольної роботи.

Контрольні роботи виконуються на комп'ютері з застосуванням системи дистанційного навчання. Кожна контрольна робота містить два або три практичні завдання (задачі) та оцінюється у: 2 бали (контрольна робота №1) та 2 бали (контрольна робота №2). Оцінка за контрольну роботу знижується при відсутності виконаного завдання, припущенні помилок у

розрахунках, неповному виконанні завдань. Максимальна кількість за виконання та захист контрольних робіт складає 4 бали.

Тестування знань виконується на комп'ютері з застосуванням системи дистанційного навчання. Кожен тест містить 10-15 завдань з варіантами вибору вірних відповідей та оцінюється у 2 бали. Максимальна кількість за тестів складає 4 бали.

Самостійна робота студента включає ознайомлення із літературними джерелами за відповідними темами, підготовку до: виконання і захисту лабораторних та контрольних робіт, а також до складання іспиту.

Підсумковий контроль:

Екзамен виконується на комп'ютері з застосуванням системи дистанційного навчання. Екзамен передбачає необхідність виконання чотирьох завдань (задач), кожне з яких оцінюється у 10 балів. Оцінка за екзамен знижується при відсутності виконаного завдання, припущенні помилок у розрахунках, неповному виконанні завдань. Максимальна оцінка за екзамен складає 40 балів.

Підсумкова оцінка з навчальної дисципліни розраховується з урахуванням балів, отриманих під час поточного та підсумкового контролю за накопичувальною системою. Виставлення підсумкової оцінки здійснюється за шкалою, наведено в таблиці "Шкала оцінювання: національна та ЄКТС" (табл. 1).

Форми оцінювання та розподіл балів наведено у таблиці "Рейтинг-план навчальної дисципліни" (табл. 2).

Таблиця 1

Шкала оцінювання: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ЄКТС	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
82 – 89	B	добре	
74 – 81	C		
64 – 73	D	задовільно	
60 – 63	E		
35 – 59	FX	незадовільно	не зараховано

Таблиця 2

Рейтинг-план навчальної дисципліни

	Форми та види навчання		Форми оцінювання	Мак бал
Тема 1	<i>Аудиторна робота</i>			
	Лекція	Лекція 1. Головні поняття, функції і задачі великих даних	Активна робота на парі	1
	<i>Самостійна робота</i>			
	Питання та завдання до самостійного опрацювання	Ознайомлення із додатковою інформацією щодо задач великих даних	-	-
Тема 2	<i>Аудиторна робота</i>			
	Лекція	Лекція 2. Методи аналізу великих даних	Активна робота на парі	1

Тема 2	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Пошук, підбір та огляд літературних джерел за заданою тематикою. Ознайомлення із особливостями моделювання та аналізу Big Data. Визначення завдання на лабораторний практикум та знайомство з предметною областю		
Тема 3	Аудиторна робота			
	Лекція	Лекція 3. Технології роботи з великими даними	Активна робота на парі	1
	Лабораторне заняття	Лабораторна робота 1. Робота з даними в Data Studio	Активна робота на парі. Лабораторні роботи (захист)	5
Тема 3	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Ознайомлення із додатковою інформацією щодо аналітики великих даних в галузях промисловості. Підготовка до захисту лабораторної роботи № 1	-	-
Тема 4	Аудиторна робота			
	Лекція	Лекція 4. Розподілена обробка великих даних та метод Map-Reduce	Активна робота на парі	1
	Лабораторне заняття	Лабораторна робота 2. Основи Hadoop	Активна робота на парі. Лабораторні роботи (захист)	6
	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Ознайомлення із додатковою інформацією щодо технологій обчислень . Підготовка до захисту лабораторної роботи № 2	-	-
Тема 5	Аудиторна робота			
	с	Лабораторна робота 3. Виконання MapReduce-задач засобами Hadoop-Streaming	Активна робота на парі. Лабораторні роботи (захист)	5
	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Підготовка до захисту лабораторної роботи № 3. Контрольна робота №1 (теми 1-4)	Контрольна робота 1	2
Тема 6	Аудиторна робота			
	Лекція	Лекція 5. Google Cloud Platform	Активна робота на парі	1
	Лабораторне заняття	Лабораторна робота 4. Основи роботи з Google Cloud Platform	Активна робота на парі. Лабораторні роботи (захист)	6

Тема 6	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Підготовка до захисту лабораторної роботи № 4. Виконання тестових завдань (теми 1-4)	Тести для поточної роботи (Тема 1-4)	2
Аудиторна робота				
Тема 7		Лекція 6. Основи BigQuery, Google Cloud Storage	Активна робота на парі	1
	Лабораторне заняття	Лабораторна робота 5. Аналіз наборів даних з використанням BigQuery та Google Cloud Storage	Активна робота на парі. Лабораторні роботи (захист)	6
	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Ознайомитися з хмарними платформами аналітики. Підготовка до захисту лабораторної роботи № 5		
Аудиторна робота				
Тема 8	Лекція	Лекція 7. Інструменти аналітичної обробки даних у сховищі великих даних	Активна робота на парі	1
	Лабораторне заняття	Лабораторна робота 6. Організація запитів до даних в BigQuery	Активна робота на парі. Лабораторні роботи (захист)	6
	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Ознайомитися с запитами до баз даних. Підготовка до захисту лабораторної роботи № 6		
Аудиторна робота				
Тема 9	Лекція	Лекція 8. Основні типи віртуалізації	Активна робота на парі	1
	Лабораторне заняття	Лабораторна робота 7. Основи роботи в Cloud AI Platform	Активна робота на парі. Лабораторні роботи (захист)	5
	Лабораторне заняття	Лабораторна робота 8. Основи роботи в Cloud AI Platform	Активна робота на парі. Лабораторні роботи (захист)	5
	Самостійна робота			
	Питання та завдання до самостійного опрацювання	Ознайомитися с запитами до баз даних. Підготовка до захисту лабораторної роботи № 7. Контрольна робота №2 (теми 5-9)	Контрольна робота 2	2
	Питання та завдання до самостійного опрацювання	Підготовка до захисту лабораторної роботи № 8. Виконання тестових завдань (теми 5-9)	Тести для поточної роботи (Тема 5-9)	2
	Екзамен			

Рекомендована література

Основна

1. Wiktorski T. Data-intensive Systems Principles and Fundamentals using Hadoop and Spark / Tomasz Wiktorski. – Springer Nature Switzerland AG. : Springer International Publishing, 2019. – 524 p.
2. О'Ніл К. BIG DATA. Зброя математичного знищення (МІМ) / Кейт О'Ніл. – Київ: Форс, 2019. – 336 с.
3. Джеймс У. Большие данные: принципы и практика построения масштабируемых систем обработки данных в реальном времени / У. Джеймс, Н. Марц. – Киев: Manning Publications, 2018. – 368 с
4. Акіменко В. В. Прикладні задачі інтелектуального аналізу даних (Data Mining) / В. Акіменко. – Київ : КНУ ім. Тараса Шевченка, 2018 р. – 152 с.

Додаткова

5. Гілдер Д. Життя після Google. Занепад великих даних і становлення блокчейн-економіки / Джордж Гілдер. – Київ: Форс, 2021. – 320 с.

Інформаційні ресурси

6. Технології роботи з BIG DATA. Сайт персональних навчальних систем ХНЕУ ім. С. Кузнеця [Електронний ресурс]. – Режим доступу : <https://pns.hneu.edu.ua/course/view.php?id=8358>