КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

ННІ «Інститут геології»

Кафедра ***геоінформатики***

«ЗАТВЕРДЖУЮ»

Заступник директора інституту
з навчальної роботи
_____
«____»_____2026 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

АНАЛІЗ ВЕЛИКИХ МАСИВІВ ДАНИХ І БАЗИ ДАНИХ (BIG DATA & DATABASE)

(повна назва навчальної дисципліни)

для студентів

галузь знань	**Е Природничі науки, математика та статистика**
спеціальність	**Е4 Науки про Землю**
освітній ступінь	**Магістр**
освітня-наукова програма	**Геоінформатика**
блок дисциплін	**Дистанційні зондування (Remote Sensing)**
вид дисципліни	**Вибіркова**

Форма навчання	**денна**
Навчальний рік	**2026/2027**
Семестр	**3**
Кількість кредитів ECTS	**4**
Мова викладання, навчання та оцінювання	**українська**
Форма заключного контролю	**залік**

Викладачі: *Зацерковний Віталій Іванович, доктор технічних наук, професор зав. кафедри геоінформатики; Гудак Василь Михайлович, асистент кафедри геоінформатики*

Пролонговано: на 20__/20__ н.р. _____(_____) «__»___ 20__р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____(_____) «__»___ 20__р.
(підпис, ПІБ, дата)

на 20__/20__ н.р. _____(_____) «__»___ 20__р.
(підпис, ПІБ, дата)

© Зацерковний В.І., Гудак В.М., 2026 рік

КИЇВ – 2026

Розробники: ***Зацерковний В. І.,*** *доктор технічних наук, професор, зав.кафедри геоінформатики;* ***Гудак В.М.,*** *асистент кафедри геоінформатики*

Затверджено на засіданні кафедри геоінформатики
Протокол *№ 1 від «31» серпня 2026 р.*

зав. кафедри геоінформатики
_____ (Віталій ЗАЦЕРКОВНИЙ)
(підпис)

Схвалено науково - методичною комісією ***ННІ «Інститут геології»***

Протокол *№ 1* від *«31» серпня 2026 року*

Голова науково-методичної комісії _____ (Всеволод ДЕМИДОВ)
(підпис)

Мета дисципліни полягає у формуванні у студентів теоретичних знань і практичних навичок із проєктування, створення та використання баз даних, а також роботи з великими масивами даних (Big Data). Дисципліна передбачає вивчення сучасних технологій зберігання, організації, інтеграції, розподіленої обробки й аналізу структурованих, напівструктурованих і неструктурованих даних; опанування принципів роботи з реляційними та нереляційними базами даних, сховищами даних, озерами даних і хмарними платформами; набуття навичок створення аналітичних рішень для підтримки прийняття рішень.

Вимоги до вибору навчальної дисципліни:

- базове розуміння структури даних, інформаційних систем і принципів організації цифрової інформації;
- знання основ програмування та алгоритмізації;
- навички роботи з операційними системами та командним рядком;
- базові знання мови Python для опрацювання й аналізу даних;
- розуміння принципів реляційної моделі даних і побудови SQL-запитів;
- навички роботи з електронними таблицями, файлами CSV, JSON та структурованими наборами даних.

Анотація навчальної дисципліни/референс**:**

Навчальна дисципліна спрямована на вивчення принципів організації баз даних і сучасних технологій роботи з великими масивами даних (Big Data), методів їх зберігання, інтеграції, обробки та аналізу. Розглядаються реляційні й нереляційні моделі даних, системи керування базами даних, SQL-запити, типи та джерела великих даних, а також принципи побудови сховищ даних, озер даних, архітектур Lakehouse і розподілених обчислювальних систем.
У межах дисципліни вивчаються архітектури сховищ даних, хмарних озер даних і Lakehouse; технології NoSQL; ETL/ELT-процеси; розподілена пакетна та потокова обробка; Apache Spark, PySpark і сучасні хмарні платформи даних. Особлива увага приділяється практичному створенню баз даних, інтеграції різнорідних джерел, виконанню аналітичних запитів та застосуванню сучасних інструментів для опрацювання великих масивів даних і підтримки прийняття рішень.
Академічна доброчесність. Виконання всіх навчальних і контрольних завдань здійснюється з дотриманням Хартії академічної доброчесності та чинних положень Київського національного університету імені Тараса Шевченка. Запозичення, використання даних, ідей, текстів, зображень, цифрових матеріалів та інших результатів мають бути належно атрибутовані. Порушення академічної доброчесності розглядаються відповідно до університетських процедур.
Опанування програмного матеріалу дисципліни, поточний та підсумковий контроль здійснюються на принципах академічної доброчесності відповідно до ст. 42 Закону України «Про освіту» та рекомендацій Міністерства освіти і науки України щодо запобігання академічному плагіату (Лист МОН № 1/9-565 від 15.08.2018 р.), а також Положення про використання штучного інтелекту при реалізації освітніх програм у Київському національному університеті імені Тараса Шевченка https://senate.knu.ua/?p=3466
Використання штучного інтелекту. Для освітнього компонента загалом встановлюється рівень 2 використання генеративного штучного інтелекту згідно «Положення про використання штучного інтелекту при реалізації освітніх програм у Київському національному університеті імені Тараса Шевченка».
Для окремих видів оцінюваних робіт у п.п. 7.1 Форми оцінювання студентів робочої програми

визначені спеціальні обмеження, зокрема повна заборона використання генеративного ШІ під час контрольних заходів, що виконуються під наглядом викладача.
Визнання результатів неформальної та/або інформальної освіти. Результати навчання, здобуті в неформальній та/або інформальній освіті, можуть бути подані до визнання у встановленому Університетом порядку за умови їх відповідності результатам навчання цієї дисципліни та наявності підтвердних матеріалів.

Завдання:

- ознайомити студентів із концепцією Big Data, характеристиками великих даних і сферами їх застосування;
- сформувати знання про принципи проєктування, створення та використання баз даних;
- розглянути реляційні, нереляційні та розподілені моделі даних;
- сформувати навички створення таблиць, зв'язків, ключів, обмежень цілісності та запитів до баз даних;
- вивчити типи, формати та джерела структурованих, напівструктурованих і неструктурованих даних;
- розглянути архітектури баз даних, сховищ даних, хмарних Data Lake і Lakehouse;
- сформувати навички роботи з розподіленими та хмарними системами зберігання даних;
- опанувати принципи роботи з реляційними та NoSQL-базами даних;
- сформувати навички побудови ETL/ELT-процесів для завантаження, очищення, перетворення та інтеграції даних;
- вивчити принципи розподіленої пакетної та потокової обробки великих масивів даних;
- засвоїти методи опрацювання й аналізу даних засобами Apache Spark і PySpark;
- ознайомити студентів із сучасними форматами та технологіями Lakehouse, зокрема Parquet, Delta Lake та Apache Iceberg;
- сформувати навички управління якістю, метаданими, доступом і захистом даних;
- навчити студентів створювати аналітичні рішення для підтримки прийняття управлінських рішень.

Результати навчання:

Результат навчання (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність)		*Форма/Методи викладання і навчання*	*Форма/Методи оцінювання*	*Відсоток у підсумковій оцінці з дисципліни*
Код	Результат навчання			
1.1	*Основні поняття Big Data, характеристики великих даних і сфери їх застосування*	*лекція, практичне заняття*	*Письмова робота*	*до 5%*
1.2	*Структуровані, напівструктуровані й неструктуровані дані, їх формати та джерела*	*лекція, практичне заняття*	*Письмова робота*	*до 5%*
1.3	*Принципи організації, проєктування та функціонування реляційних, нереляційних і розподілених баз даних*	*лекція, практичне заняття*	*Письмова робота*	*до 10%*
1.4	*Принципи розподіленої, пакетної та потокової обробки великих масивів даних*	*лекція, практичне заняття*	*Письмова робота*	*до 10%*
1.5	*Основи архітектури сховищ даних, хмарних Data Lake і Lakehouse, а також особливості їх застосування*	*лекція, практичне заняття*	*Письмова робота*	*до 10%*

1.6	*Архітектуру та можливості NoSQL, Apache Spark, PySpark, Delta Lake й сучасних хмарних платформ даних*	*лекція, практичне заняття*	*Письмова робота*	*до 10%*
2.1	*Проєктувати та створювати реляційні бази даних, визначати таблиці, ключі, зв'язки й обмеження цілісності*	*практичне заняття, самостійне навчання*	*Письмова робота, програмне представлення*	*до 10%*
2.2	*Завантажувати, очищувати, перетворювати та інтегрувати дані у базах даних, хмарних Data Lake і Lakehouse-середовищах*	*практичне заняття, самостійне навчання*	*Письмова робота, програмне представлення*	*до 10%*
2.3	*Розробляти програми пакетної та потокової обробки й аналізу великих даних із використанням Spark та PySpark*	*практичне заняття, самостійне навчання*	*Письмова робота, програмне представлення*	*до 10%*
3.1	*Представляти результати аналізу великих даних у формі звітів, візуалізацій, інформаційних панелей та аналітичних висновків*	*практичне заняття*	*Програмне представлення*	*до 5%*
3.2	*Розробляти технічну документацію до баз даних, потоків даних та аналітичних рішень, описувати їх архітектуру й результати роботи*	*практичне заняття, самостійне навчання*	*Письмова робота*	*До 10%*
4.1	*Самостійно обирати технології баз даних і Big Data, оцінювати якість, безпеку та достовірність даних і нести відповідальність за отримані результати*	*--/-*	*--/-*	*до 5%*

Структура курсу: *лекційні і практичні заняття.*

Схема формування оцінки:

Форми оцінювання студентів

1. Семестрове оцінювання:

1) Контрольна робота з основ Big Data, типів даних, реляційних і нереляційних баз даних та архітектур зберігання даних– 10 балів (рубіжна оцінка 6 балів). *Під час підготовки письмової роботи використання ШІ заборонено.*

2) Контрольна робота з технологій інтеграції, розподіленої обробки та аналізу великих даних (ETL/ELT, Cloud Data Lake, Lakehouse, Spark, PySpark) – 10 балів (рубіжна оцінка 6 балів). *Під час підготовки письмової роботи використання ШІ заборонено.*

3) Оцінка за роботу на практичних заняттях – 60 балів (рубіжна оцінка 24 бали). *Дозволяється використовувати ШІ при підготовці практичних робіт як допоміжного інструменту за умови збереження авторського внеску в підготовку програмного коду практичної роботи та розрахункових задач. Дозволено використовувати ШІ для довідкових функцій під час підготовки практичної роботи та як інструмент перевірки орфографії, технічного оформлення роботи. Декларування факту та способу використання ШІ є обов'язковим. Під час захисту практичних робіт використання ШІ заборонено.*

2. Підсумкове оцінювання *у формі заліку*: максимальна оцінка 20 балів, рубіжна оцінка 12 балів Під час заліку студент виконує тестове або практичне завдання на перевірку знань з основ Big Data, баз даних, технологій інтеграції та розподіленої обробки даних, Cloud Data Lake, Lakehouse, Spark і PySpark. *Під час підготовки письмової підсумкової роботи використання ШІ заборонено.*

Результати навчальної діяльності студентів оцінюються за 100 бальною шкалою.

Залік виставляється за результатами роботи студента впродовж усього семестру, *як сума (проста або зважена) балів за систематичну роботу впродовж семестру.*

	Семестрова кількість балів	**ПКР** (підсумкова контрольна робота) *чи/або* **залік**	**Підсумкова оцінка**
Мінімум	*36*	*24*	*60*
Максимум	***60***	***40***	***100***

*Студент не допускається до **підсумкового оцінювання у формі заліку**, якщо під час семестру набрав менше 40 балів.*

Організація оцінювання: *Контроль здійснюється за модульно-рейтинговою системою та передбачає: виконання 7 практичних робіт (де студенти мають продемонструвати якість засвоєних знань та вирішити поставлені задачі без обмеження інструментарію та техніки вирішення проблеми) та проведення 2 письмових модульних контрольних робіт. Підсумкове оцінювання проводиться у формі заліку.*

Шкала відповідності

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ТЕМАТИЧНИЙ ПЛАН ЛЕКЦІЙ І ПРАКТИЧНИХ ЗАНЯТЬ

№ п/п	Назва теми	Кількість годин		
		лекції	**практичні**	**самостійна робота**
***Розділ1** Основи Big Data, бази даних та сучасні архітектури зберігання даних*				
1	**Вступ**. **Тема 1 Основні поняття Big Data, типи та джерела даних.** Характеристики великих даних, структуровані, напівструктуровані та неструктуровані дані, сфери застосування Big Data.	2	8	20
2	**Тема 2. Реляційні та нереляційні бази даних.** Проєктування баз даних, реляційна модель, SQL, NoSQL-бази даних та особливості їх застосування для великих масивів даних.	4		18
3	**Тема 3. Сховища даних, Cloud Data Lake та Lakehouse.** Архітектури сховищ і хмарних озер даних, Lakehouse-підхід, формати Parquet, Delta Lake та Apache Iceberg.	2		1
	Контрольна робота 1			2

Розділ 2 *Інтеграція, обробка та аналіз великих масивів даних*				
4	**Тема 4. Інтеграція та розподілена обробка даних.** ETL/ELT-процеси, очищення й перетворення даних, пакетна та потокова обробка, розподілені обчислення.	4	6	18
5	**Тема 5. Apache Spark для обробки великих даних**. Архітектура Spark, розподілені структури даних, Spark SQL, пакетна та потокова обробка.	6		
6	**Тема 6. PySpark та аналітика великих даних.** Завантаження, підготовка, аналіз і агрегування даних засобами PySpark. Побудова аналітичних рішень і представлення результатів.	6		20
	Контрольна робота 2			2
	ВСЬОГО	**24**	**14**	**81**

Загальний обсяг *120* *год.,* у тому числі:
Лекцій – ***24*** *год.*
Практичні заняття - ***14*** *год.*
Консультації - ***1*** *год.*
Самостійна робота **– 81** *год.*

Теми для самостійного навчання:

1. Методи виявлення пропущених значень, дублікатів, аномалій і помилок у наборах даних.
2. Індексування й оптимізація продуктивності баз даних під час роботи з великими обсягами інформації.
3. Транзакції, узгодженість і масштабування у розподілених базах даних.
4. Метадані, каталоги даних і простежуваність походження даних у корпоративних системах.
5. Архітектура Medallion та організація рівнів Bronze, Silver і Gold у хмарному середовищі.
6. Оркестрація потоків даних, планування завдань і моніторинг процесів обробки.
7. Потокове опрацювання даних у реальному часі та використання брокерів повідомлень.
8. Контроль доступу, конфіденційність, резервне копіювання та захист великих даних.
9. Відтворюваність аналітичних процесів, документування моделей і контроль версій даних.
10. Оптимізація витрат на зберігання й обробку даних у хмарних середовищах.
11. Візуалізація великих масивів даних та побудова інформаційних панелей для підтримки прийняття рішень.

РЕКОМЕНДОВАНІ ДЖЕРЕЛА:

Основні:

Таран В. І., Гордієнко Ю. Г., Стіренко С. Г. (2022). Технології Big Data. Практикум : навч. посіб. Київ : КПІ ім. Ігоря Сікорського. 56 с. URL: https://ela.kpi.ua/handle/123456789/50014

Олещенко Л. М. (2021). Технології оброблення великих даних. Конспект лекцій : навч. посіб. Київ : КПІ ім. Ігоря Сікорського. 227 с. URL: https://ela.kpi.ua/handle/123456789/42206

Reis J., Housley M. (2022). Fundamentals of Data Engineering: Plan and Build Robust Data Systems. Sebastopol, CA : O'Reilly Media. 450 p.

Kleppmann M., Riccomini C. (2026). Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. 2nd ed. Sebastopol, CA : O'Reilly Media. 672 p.

Damji J. S., Wenig B., Das T., Lee D. (2020). Learning Spark: Lightning-Fast Data Analytics. 2nd ed. Sebastopol, CA : O'Reilly Media. 397 p.

Lee D., Wentling T., Haines S., Babu P. (2024). Delta Lake: The Definitive Guide. Sebastopol, CA : O'Reilly Media. 382 p.

Shapira G., Palino T., Sivaram R., Petty K. (2021). Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale. 2nd ed. Sebastopol, CA : O'Reilly Media. 485 p.

Додаткові:

Зацерковний В.І., Бурачек В.Г., Железняк О.О., Терещенко А.О. (2014). Геоінформаційні системи і

Зацерковний В. І., Бурачек В. Г., Железняк О. О., Терещенко А. О. (2014). Геоінформаційні системи і бази даних : монографія. Кн. 1. Ніжин : НДУ ім. М. Гоголя.

Марченко О. О., Россада Т. В. (2017). Актуальні проблеми Data Mining : навч. посіб. Київ.

Berman J. J. (2018). Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information. 2nd ed. Cambridge, MA : Academic Press.

Pal A., Pal S. K. (Eds.). (2017). Pattern Recognition and Big Data. Singapore : World Scientific Publishing.

Chambers B., Zaharia M. (2018). Spark: The Definitive Guide: Big Data Processing Made Simple. Sebastopol, CA : O'Reilly Media.

Apache Software Foundation. (б. р.). Apache Hadoop Documentation. URL: https://hadoop.apache.org/docs/current/

Apache Software Foundation. (б. р.). Apache Spark Documentation. URL: https://spark.apache.org/documentation.html

Apache Software Foundation. (б. р.). Apache Kafka Documentation. URL: https://kafka.apache.org/documentation/

PostgreSQL Global Development Group. (б. р.). PostgreSQL Documentation. URL: https://www.postgresql.org/docs/current/

MongoDB, Inc. (б. р.). MongoDB Documentation. URL: https://www.mongodb.com/docs/

Apache Software Foundation. (б. р.). Apache Iceberg Documentation. URL: https://iceberg.apache.org/docs/latest/

Delta Lake Project. (б. р.). Delta Lake Documentation. URL: https://docs.delta.io/