



# БАЗИ ДАНИХ

## Робоча програма навчальної дисципліни (Силабус)

### Реквізити навчальної дисципліни

|   |  |
|---|--|
| Рівень вищої освіти                         | <i>Перший (бакалаврський)</i>  |
| Галузь знань                                | <i>11 Математика і статистика</i>  |
| Спеціальність                               | <i>113 Прикладна математика</i>  |
| Освітня програма                            | <i>Наука про дані та математичне моделювання</i>   |
| Статус дисципліни                           | <i>Нормативна</i>  |
| Форма навчання                              | <i>очна(денна)</i>   |
| Рік підготовки, семестр                     | <i>3 курс, весняний семестр</i>  |
| Обсяг дисципліни                            | <i>4 кредити ЄКТС / 120 годин (36 год. - лекції, 18 год. - лабораторні роботи, 66 год. - СРС)</i>  |
| Семестровий контроль/ контрольні заходи     | <i>Екзамен, МКР та лабораторні роботи</i>  |
| Розклад занять                              | <i>Лекції – 1 раз на тиждень (18 лекцій)<br/>Лабораторні заняття – 1 раз на 2 тижні (9 занять)</i>   |
| Мова викладання                             | <i>Українська</i>  |
| Інформація про керівника курсу / викладачів | Лектор: Дмитренко Олександра Анатоліївна<br>Лабораторні: Жук Іван Сергійович   |
| Розміщення курсу                            | Google classroom.<br><a href="https://classroom.google.com/c/NjYyMjE4Mjg4OTAw?cjc=2i64sdu">https://classroom.google.com/c/NjYyMjE4Mjg4OTAw?cjc=2i64sdu</a> |

### Програма навчальної дисципліни

#### 1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна «Бази даних» посідає одне з чільних місць в підготовці фахівців з прикладної математики й відноситься до професійно-орієнтованих дисциплін. Це визначається тим, що для широкого класу автоматизованих систем бази даних є системоутворюючим фактором.

**Метою** навчальної дисципліни є формування у студентів здатностей:

- розуміти основні сучасні інформаційні технології обробки даних;
- розуміти місце і роль баз даних в інформаційних системах обробки інформації, методи проектування і управління базами даних;
- розуміти основні моделі структур даних (списки, ієрархії, відношення, мережеві структури);
- розуміти класифікацію систем управління базами даних (СУБД), як реляційних так і нереляційних (NoSQL);
- знати та вміти застосовувати на практиці транзакції в реляційних СУБД;
- розуміти обмеження та переваги нереляційних СУБД;
- розуміти основні шаблони та підходи роботи з базами даних;

- розуміти що таке розподілені БД;
- розбиратись, яку БД вибрати в якій ситуації;

## 2. Пререквізити та постреквізити дисципліни

Дисципліна базується на знаннях і навичках, які студенти одержали при вивченні дисципліни "Вступ до баз даних та інформаційних систем", "Програмування на мові Python".

Дисципліна забезпечує засвоєння студентами дисципліни «Інформаційні системи».

## 3. Зміст навчальної дисципліни

### Розділ 1. Робота з реляційними СУБД

1. Робота з запитами.
2. Оптимізація SQL-запитів. Індекси.
3. Робота з транзакціями та контроль цілісності даних в реляційних СУБД.
4. Принципи побудови БД - бінарні дерева.

### Розділ 2. Основи розгортання інформаційних систем

1. Багаторівнева архітектура програмного застосунку. Моделювання реляційних баз даних. Концептуальна, логічна та фізична ERD.
2. Міграція/рефакторинг БД.
3. Шаблони роботи з базами даних.

### Розділ 3. Робота з нереляційними СУБД

1. Види нереляційних СУБД. Базові принципи Роботи з ними.
2. CAP/PACELC теореми
3. Робота з Cassandra та MongoDB
4. Шардінг та партішинінг. Реплікація
5. Розподілені системи. Великі дані. Map-Reduce
6. Робота з Elasticsearch. Векторні бази даних

## 4. Навчальні матеріали та ресурси

### Базові джерела:

- Матеріали курсу, що публікуються викладачами у слак каналі курсу
- Інформація по SQL запитам - <https://www.geeksforgeeks.org/sql-tutorial/?ref=lbp>
- Alex Petrov. "Database Internals. A Deep Dive into How Distributed Data Systems Work". — O'Reilly, 2019. — 350 с.
- Принципи The Twelve-Factor App (<https://12factor.net/uk/>)
- Буй Д.Б., Сільвейструк Л.М. Формалізація моделі "Сутність-зв'язок" — Київ: Видавничо поліграфічний центр "Київський університет", 2011. — 175 с. (<http://csc.knu.ua/uk/library/books/bui-silveistruk-33.pdf>)
- <https://db-engines.com/en/ranking>
- <https://jepson.io/analyses>

### Додаткові джерела:

- *Designing Data-Intensive Applications*, by Martin Kleppmann, O'Reilly Media, Inc., 2017 (<https://dataintensive.net/>)
- *Patterns of enterprise application architecture*, by Martin Fowler, with Dave Rice, Matthew Foemmel, Edward Hieatt, Robert Mee, and Randy Stafford, Addison-Wesley,

2003

- *Документація по PostgreSQL* (<https://www.postgresql.org/docs/current/index.html>)
- *Документація по Flyway* (<https://flywaydb.org/documentation/>)
- *Документація по SQLAlchemy* (<https://docs.sqlalchemy.org/en/14/orm/tutorial.html>)
- *Evolutionary Database Design, by Pramod Sadalage, Martin Fowler* (<https://martinfowler.com/articles/evodb.html>)
- *The SQLAlchemy Session - In Depth* (<https://www.youtube.com/watch?v=uAtaKr5HOdA>)
- *Consistency Tradeoffs in Modern Distributed Database System Design, by Daniel Abadi* (<http://www.cs.umd.edu/~abadi/papers/abadi-pacelc.pdf>)

## Навчальний контент

### 5. Методика опанування навчальної дисципліни (освітнього компонента)

#### Лекції

##### Розділ 1. Робота з реляційними СУБД

*Лекція 1. Узагальнення SQL (SQL, DDL, DML). Повторення основних понять. Поняття ETL*

*Поняття CRUD. Операції над даними - перетин JOIN, підзапити та групування даних, використання having, groupBy, orderBy. Агрегатні функції SQL: Count(), Sum(), Avg(), Min(), Max().*

*Лекція 2. Оптимізація SQL-запитів. Загальні правила. Індeksi, їх види. Розріз даних (Data Slicing) для оптимізації запитів та аналізу даних.*

*Лекція 3. Транзакції та їх властивості. ACID принципи. Рівні ізоляції транзакцій. Поняття Deadlock.*

*Лекція 4. Принципи побудови бази даних. Бінарні дерева.*

##### Розділ 2. Основи розгортання інформаційних систем

*Лекція 5. Багаторівнева архітектура. Принципи. Проблеми, які вона вирішує.*

*Моделювання реляційних баз даних. Концептуальна, логічна та фізична ERD*

*Лекція 6. Міграція/рефакторинг БД. Принципи роботи Flyway та Liquibase.*

*Принципи роботи Flask та SQLAlchemy.*

*Лекція 7. Шаблони роботи з базами даних. Transaction Script. Domain model. Service Layer. Table Data Gateway. Row Data Gateway. ORM шаблони роботи з базами даних. Active Record. Session.*

##### Розділ 3. Робота з нереляційними СУБД: NoSQL та Big Data.

*Лекція 8. Загальні принципи роботи нереляційних СУБД, їх необхідність.*

*Лекція 9. CAP теорема. Класифікація нереляційних СУБД.*

*Лекція 10. Класифікація нереляційних СУБД. Ключ-значення СУБД.*

*Документ-орієнтовані СУБД. Графові СУБД. Column-Family СУБД. Time-series СУБД.*

*Лекція 11. Колоночні СУБД. PACELC теорема. Компроміси при виборі технології бази даних.*

Лекція 12. Робота з розподіленими базами даних на прикладі MongoDB.

Лекція 13. Повнотекстовий пошук. Реалізація в PostgreSQL. Реалізація в ElasticSearch.

Лекція 14. Робота з ElasticSearch, як з основною базою даних для системи.

Лекція 15. Розподілені системи. Необхідність, спосіб розподілення баз даних.

Лекція 16. BigData бази даних. Принципи роботи Spark. Принципи роботи HBase.

Лекція 17. Векторні бази даних. Принцип їх роботи. Бази даних для штучного інтелекту.

Лекція 18. Модульна контрольна робота.

### Лабораторні роботи

| № з/п | Тема лабораторної роботи   |
|-------|--|
| 1     | Lost-update, In-place update, row-level locking, optimistic concurrency control в Postgres |
| 2     | Реалізація своєї бази даних на базі структури бінарного дерева B-Tree                      |
| 3     | Міграція БД з використанням liquibase або flyway та мови Java або Python                   |
| 4     | Робота з MongoDB (Document Store), під'єднання її до програмного проекту                   |
| 5     | Робота з базовими функціями на вибір: Neo4j (graph) чи Cassandra (wide column store)       |

### 6. Самостійна робота студента

Студент повинен завчасно готуватись до лекцій та практичних занять. Перед лекціями потрібно повторити вивчений раніше теоретичний матеріал, підготувати питання щодо нової теми для обговорення на лекції. Перед практичними заняттями потрібно повторити відповідний теоретичний матеріал.

| № з/п | Вид самостійної роботи            | Кількість годин СРС |
|-------|-----------------------------------|---------------------|
| 1.    | Підготовка до лабораторних занять | 35                  |
| 2.    | Підготовка до лекцій              | 5                   |
| 3.    | Опрацювання лекційного матеріалу  | 10                  |
| 4.    | Підготовка до МКР                 | 6                   |
| 5.    | Підготовка до екзамену            | 10                  |
|       | <b>Загалом</b>                    | <b>66</b>           |

### До самостійного опрацювання виносяться наступні теми

| № з/п | Тема  | СРС, год |
|-------|---|----------|
| 1     | Вивчення синтаксису мови запитів MongoDB<br>( <a href="https://docs.mongodb.com/manual/reference/sql-comparison/">https://docs.mongodb.com/manual/reference/sql-comparison/</a> )   | 2        |
| 2     | Вивчення синтаксису мови запитів Elasticsearch<br>( <a href="https://www.elastic.co/guide/en/elasticsearch/reference/current/getting-started.html">https://www.elastic.co/guide/en/elasticsearch/reference/current/getting-started.html</a> ) | 2        |
| 3     | Вивчення синтаксису мови запитів GraphQL -<br><a href="https://graphql.org/learn/">https://graphql.org/learn/</a>   | 2        |
| 4     | Вивчення особливостей роботи з SpringData   | 2        |
| 5     | Вивчення особливостей роботи з Hibernate  | 2        |

### Додаткова практична робота:

Додаткову практичну роботу можна виконати щоб заробити додаткові бали (максимум 10).

| № з/п | Тема   |
|-------|--|
| 1     | Робота з розподіленням БД на вибір: MongoDB, Cassandra, Hazelcast. Також можна спробувати роботу з Elasticsearch чи S3Bucket (на Amazon) |

### Додаткові ресурси для практики

| № з/п | Ресурси   |
|-------|---|
|       | [1]<br>Організація баз даних NoSQL [Електронний ресурс] : практикум / Н.В. Ситник, І.С. Зінов'єва. — К. КНЕУ, 2022. — 167, [1] с. , ISBN 978-966-926-406-0.<br>– Режим доступу:<br><a href="https://www.researchgate.net/publication/361775185_ORGANIZACIA_BAZ_DANIH_NoSQL_PRAKTIKUM">https://www.researchgate.net/publication/361775185_ORGANIZACIA_BAZ_DANIH_NoSQL_PRAKTIKUM</a>  |
| 1     | Практика SQL запитів -<br><a href="https://www.datacamp.com/blog/5-sql-challenges-to-practice-your-skills?utm_source=google&amp;utm_medium=paid_search&amp;utm_campaignid=19589720818&amp;utm_adgroupid=152984010654&amp;utm_device=c&amp;utm_keyword=&amp;utm_matchtype=&amp;utm_network=g&amp;utm_adposition=&amp;utm_creative=684592138976&amp;utm_targetid=dsa-2222697809758&amp;utm_loc_interest_ms=&amp;utm_loc_physical_ms=9111792&amp;utm_content=DSA~blog~SQL&amp;utm_campaign=230119_1-sea~dsa~tofu_2-b2c_3-eu_4-prc_5-na_6-na_7-le_8-pdsh-go_9-na_10-na_11-na&amp;gad">https://www.datacamp.com/blog/5-sql-challenges-to-practice-your-skills?utm_source=google&amp;utm_medium=paid_search&amp;utm_campaignid=19589720818&amp;utm_adgroupid=152984010654&amp;utm_device=c&amp;utm_keyword=&amp;utm_matchtype=&amp;utm_network=g&amp;utm_adposition=&amp;utm_creative=684592138976&amp;utm_targetid=dsa-2222697809758&amp;utm_loc_interest_ms=&amp;utm_loc_physical_ms=9111792&amp;utm_content=DSA~blog~SQL&amp;utm_campaign=230119_1-sea~dsa~tofu_2-b2c_3-eu_4-prc_5-na_6-na_7-le_8-pdsh-go_9-na_10-na_11-na&amp;gad</a> |

|   |  |
|---|--|
|   | <a href="https://source=1&amp;gclid=CjwKCAiAq4KuBhA6EiwArMAw1F0JhLXiI5zAzefeWqwNulU0tylUXjzjXo5bkwCB4j1iOHi0G_LnyBoCWE0QAvD_BwE">source=1&amp;gclid=CjwKCAiAq4KuBhA6EiwArMAw1F0JhLXiI5zAzefeWqwNulU0tylUXjzjXo5bkwCB4j1iOHi0G_LnyBoCWE0QAvD_BwE</a>  |
| 2 | BigQuery Studio - <a href="https://console.cloud.google.com/">https://console.cloud.google.com/</a> , how to use Analytics Dataset - <a href="https://support.google.com/analytics/answer/7586738?hl=en#zippy=%2Cin-this-article">https://support.google.com/analytics/answer/7586738?hl=en#zippy=%2Cin-this-article</a>   |
| 3 | Google Analytics Dataset - web analytics data from a data warehouse that stores web analytics from Google for a commercial website - <a href="https://console.cloud.google.com/marketplace/product/obfuscated-ga360-data/obfuscated-ga360-data?project=lexical-script-761">https://console.cloud.google.com/marketplace/product/obfuscated-ga360-data/obfuscated-ga360-data?project=lexical-script-761</a> |
| 4 | БД ЗНО 2022: <a href="https://osvita.ua/news/data/87870/">https://osvita.ua/news/data/87870/</a>   |

## Політика та контроль

### 7. Політика навчальної дисципліни (освітнього компонента)

- **Відвідування лекцій та лабораторних занять.** Відсутність на лекціях та лабораторних заняттях без поважної причини не допускається.
- **Правила поведінки на заняттях.** На лекційних та лабораторних заняттях студенти мають вимкнути мобільні телефони або увімкнути їх на беззвучний режим. Забороняється користуватися телефонами та іншими засобами зв'язку для пошуку інформації на гугл-диску викладача, в Інтернеті, отримання відповідей за допомогою чату GPT тощо.
- **Правила захисту лабораторних робіт.** Усі лабораторні роботи оформляються і здаються студентами у вигляді демонстрації коду, прийнятної для читання. Захист лабораторних робіт відбувається у вигляді презентації студентом результатів виконаної роботи, показування демо, а також написання коду у випадку запиту викладача на уроці. Також студенту можуть бути поставлені запитання щодо змісту його роботи. Формат звіту передбачений тільки для деяких робіт, або для всіх у випадку вимушеної асинхронної роботи студента.
- **Правила призначення заохочувальних та штрафних балів.**  
Заохочувальні бали нараховуються за:
  - активність студента на лекціях та лабораторних заняттях;
  - проходження онлайн курсів, погоджених з викладачем.
 Максимальна загальна кількість бонусних балів може складати не більше, ніж 10% від максимального балу за семестр (відповідно до РСО).  
Штрафні бали призначаються за несвоєчасне виконання лабораторних робіт (відповідно до РСО), недотримання принципів академічної доброчесності та плагіат.
- **Політика щодо академічної доброчесності.** Згідно з Кодексом честі студента КПІ, при виконанні лабораторних чи контрольних робіт **забороняється** користуватися чужими виконаними лабораторними чи контрольними роботами та/або їх фрагментами. У випадку виявлення плагіату лабораторна робота може бути оцінена від 0 до 1/3 сумарної кількості запланованих за неї балів. Під час виконання контрольних робіт забороняється користуватись будь-якими сторонніми джерелами.

### 8. Види контролю та рейтингова система оцінювання результатів навчання (РСО)

Протягом семестру відбуваються наступні види контролю

*Поточний контроль: МКР, лабораторні роботи*

*Семестровий контроль: екзамен*

*Умови допуску до семестрового контролю: зараховані щонайменше 3 лабораторних роботи та семестровий рейтинг не менший від 50% від максимальної кількості балів, яку може отримати студент до даного календарного контролю.*

*Умови допуску до семестрового контролю: виконання всіх частин МКР та семестровий рейтинг не менший від 25 балів.*

*Рейтинг студента з кредитного модуля складається з:*

- 1) балів за модульну контрольну роботу (МКР),*
- 2) балів за лабораторні роботи;*
- 3) балів за відповідь на екзамені.*

*Також, можна отримати додаткові бали за*

- 1) за відповіді на лекціях - до 7 балів;*
- 2) за виконання додаткової лабораторної роботи - до 10 балів.*

*Критерії оцінювання якості виконання лабораторної роботи (звіту):*

*5 балів – робота виконана якісно, в повному обсязі;*

*3-4 бали – робота виконана не в повному обсязі, або містить незначні помилки;*

*0-2 бали – робота містить суттєві прогалини та/або значні помилки.*

*Критерії оцінювання захисту лабораторної роботи:*

*4-5 балів – презентація та доповідь повністю розкриває контекст завдання, якісно виконані, відповіді повні, усі висновки добре аргументовані;*

*1-3 бали – презентація та доповідь мають прогалини, у відповідях є помилки;*

*0 балів – презентація отриманих результатів відсутня.*

**Максимальна кількість балів за відповідь на екзамені:**

*Завдання на МКР складається з 14 тестових питань (8 питань з однією правильною відповіддю та 6 питань з декількома правильними відповідями). Кожне питання з однією правильною відповіддю оцінюється в 1 бали, кожне питання з декількома правильними відповідями оцінюється в 4 бали. 2 питання буде оцінено в 5 балів, та вони полягатимуть у самостійному написанні відповіді.*

*Критерії оцінювання кожного тестового запитання з однією правильною відповіддю:*

*2 бали – відповідь правильна;*

*0 балів – немає відповіді або відповідь невірна.*

*Критерії оцінювання кожного тестового запитання з декількома правильними відповідями:*

*4 бали – обрано всі правильні відповіді та жодної невірної;*

*2 бали – обрано щонайменше 50% від всіх правильних відповідей;*

*0 балів – немає відповіді або всі відповіді невірні.*

*Критерії оцінювання кожної вільної відповіді:*

*5 балів – відповідь повністю вірна;*

*2,5 бали – написано близько 50% від правильної відповіді; можливі й інші бали з 0 до 5 відповідно до пропорції правильності відповіді.*

*0 балів – немає відповіді або все невірно.*

*Максимальна кількість балів за екзамен:*

*$R = 2 \text{ бали} \times 8 \text{ тестових питань з однією правильною відповіддю} + 4 \text{ бали} \times 6 \text{ питань з декількома правильними відповідями} + 5 \text{ балів} \times 2 \text{ питання} = 15 \text{ балів.}$*

20 балів × 1 теоретичне питання + 10 балів × 3 задачі = **50 балів**

**Розрахунок шкали (R) рейтингу:**

Сума вагових балів контрольних заходів протягом семестру складає:

$$R_c = 30 + 20 = 50 \text{ балів.}$$

Із цієї суми за семестр студент може набрати певну кількість балів  $r_c$ .  
Екзаменаційна складова ( $R_E$ ) шкали дорівнює 50 балам, тобто складає 50% від  $R$ . За відповідь на екзамені студент може набрати певну кількість балів  $r_E$ .

Рейтингова шкала з дисципліни складає

$$R = R_c + R_E = 100 \text{ балів.}$$

Неодмінною умовою допуску до екзамену є виконання всіх складових модульної контрольної роботи, а також набрані  $r_c = 25$  балів.

Якщо студент не виконав умов допуску до екзамену, то він не може отримати оцінку в першу відомість, а натомість отримує додаткову можливість виконати ці умови (повторно писати складові МКР доти, доки балів для допуску не стане достатньо), після чого складає екзамен за наведеними раніше правилами, а його остаточну рейтингову оцінку  $R_D$  розраховують за формулою:  $R_D = r_c + r_E$ .

**Атестації**

На першій атестації (8-й тиждень) студент отримує «зараховано», якщо його поточний рейтинг складає не менше 12,5 балів (50 % від максимальної кількості балів, яку може отримати студент до першої атестації).

На другій атестації (14-й тиждень) студент отримує «зараховано», якщо його поточний рейтинг складає не менше 25 балів (50 % від максимальної кількості балів, яку може отримати студент до другої атестації).

**Семестровий контроль:** екзамен. Умови допуску до семестрового контролю: зарахування усіх лабораторних робіт та семестровий рейтинг рівний або більше 25 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

| Кількість балів           | Оцінка       |
|---------------------------|--------------|
| 100-95                    | Відмінно     |
| 94-85                     | Дуже добре   |
| 84-75                     | Добре        |
| 74-65                     | Задовільно   |
| 64-60                     | Достатньо    |
| Менше 60                  | Незадовільно |
| Не виконані умови допуску | Не допущено  |

**9. Додаткова інформація з дисципліни (освітнього компонента)**

- Курсом передбачена можливість зарахування сертифікатів проходження дистанційних чи онлайн курсів за відповідною тематикою у випадку неможливості чи недоцільності складання передбачених лабораторних та виключно після домовленості з викладачем.

**Робочу програму навчальної дисципліни (силабус):**

Складено асистентом Дмитренко Олександром Анатоліївною.

Ухвалено кафедрою прикладної математики (протокол № 11 від 17.05.2023 )

Погоджено Методичною комісією факультету прикладної математики (протокол № 10 від 26.05.2023 )