



Проектування високонавантажених систем Робоча програма навчальної дисципліни (Силабус)

1. 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>Математичні методи криптографічного захисту інформації</i>
Статус дисципліни	<i>Вибіркова</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>4 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредити 120 годин Лекції – 36 годин Лабораторні роботи – 18 годин Самостійна робота – 66 годин</i>
Семестровий контроль/ контрольні заходи	<i>Залік, Модульна контрольна робота</i>
Розклад занять	<i>Лекції 2 години раз на тиждень, Лабораторні роботи 1 година щотижня</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	Лектор: доц., к.т.н., Родіонов Андрій Миколайович
Розміщення курсу	https://drive.google.com/drive/folders/0B6VW3ge7g2HXUnpsaEU5U1pkUDQ?resourcekey=0-qk0NCrkzMB2p6rhgyHO4Uw&usp=share_link

2. Програма навчальної дисципліни

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

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

У сучасному світі кількість даних, що генеруються, передаються, обробляються, накопичуються та аналізуються стрімко зростає. Обсяг та інтенсивність даних суттєво впливає на архітектуру інформаційних систем та засоби які застосовуються, починаючи від систем зберігання і закінчуючи підходами до передачі та аналізу. Дана навчальна дисципліна дає розуміння проблематики високонавантажених систем та підходам до їх проектування.

Метою навчальної дисципліни є оволодіння проблематикою, архітектурними рішеннями та програмними засобами при побудові високонавантажених систем, у розрізі зберігання та доступності до даних.

Предметом дисципліни є високонавантажені системи.

Програмні результати навчання є наступними:

- **Знання:**
 - Обмеження традиційних підходів на основі реляційних баз даних
 - Переваги та недоліки NoSQL баз даних для зберігання та обробки даних
 - Типи та моделі NoSQL баз даних
 - Забезпечення високої доступності та відмовостійкості при зберіганні та обробці даних
 - CAP- теорема
 - Розподілені запити на основі Map/Reduce та сучасні підходи
- **Уміння:**
 - Визначати необхідний тип архітектури
 - Вибирати відповідне сховище (базу даних) для зберігання та обробки даних
 - Забезпечувати масштабування та відмовостійкості системи
- **Досвід:**
 - Робота з різними типами NoSQL баз даних
 - Конфігурація реплікації та шардінгу
 - Створення розподілених запитів

Одержані знання та уміння посилюють такі компетентності та результати навчання:

Здатність до самонавчання, пошуку, оброблення та інтелектуального аналізу інформації з різних джерел, вміння виявляти, ставити та вирішувати проблеми.

Здатність генерувати нові ідеї та нестандартні підходи до їх реалізації, адаптуватись та діяти в нових ситуаціях, виявляти ініціативу, інноваційність та підприємливість.

Навички та вміння міжособистісної комунікації, здатність представляти і доносити знання й ідеї та працювати в команді

Здатність діяти у відповідності до норм інтелектуальної власності

Здатність формалізувати та розв'язувати складні задачі й проблеми, які потребують оновлення й інтеграції знань, часто в умовах неповної, неточної чи недостатньої інформації та суперечливих вимог.

Здатність формалізувати, будувати та використовувати у практичній діяльності моделі та методи інтелектуального аналізу даних.

Здатність використовувати та впроваджувати існуючі механізми, протоколи та системи криптографічного захисту інформації

Використовувати та адаптувати математичні теорії та моделі для забезпечення теоретичного підґрунтя розв'язання наукових та практичних задач

Ситуативно й професійно спілкуватись та аналізувати науково-технічну інформацію однією з іноземних мов, організовувати багатосторонню (у тому числі міжкультурну) комунікацію та управляти нею.

Виявляти ініціативу та підприємливість, організовувати власну професійну діяльність, розробляти інноваційні підприємницькі проекти та створювати компанії для їх реалізації

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Необхідними умовами для проходження дисципліни є:

- базовий рівень володіння англійською мовою не нижче А2
- Бази даних - ACID властивості, транзакції
- Операційні системи - робота з командним рядком, запуск сервісів та редагування конфігурацій, робота з Docker
- Комп'ютерні мережі - мережева взаємодія, особливості протоколів, конфігурування мережевих сервісів, HTTP-протокол
- Програмування - володіння мовою програмування високого рівня, архітектура Веб-додатків, формат JSON, взаємодія з БД

3. Зміст навчальної дисципліни

Розділ 1 - Особливості високонавантажених систем та недоліки реляційних БД

1. Databases mechanics
2. Властивості ACID
3. Рівні ізоляції транзакцій
4. Оптимістичне та песимістичне блокування

Розділ 2 - NoSQL бази даних

1. Недоліки RDBMS. Нормалізована та де-нормалізована модель даних
2. Особливості NoSQL баз даних
3. Моделювання даних
4. Типи NoSQL баз даних

Розділ 3 - Масштабованість та відмовостійкість

1. Масштабованість та відмовостійкість в NoSQL баз даних
2. Проблема цілісності даних
3. CAP-теорема

Розділ 4 - Розподілені запити

1. Проблема “локальності даних”
2. Підхід Map/Reduce
3. Підхід Apache Spark

4. Навчальні матеріали та ресурси

- NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence. Pramod Sadalage, Martin Fowler. Addison-Wesley Professional; 1st edition (August 8, 2012). - 192 pages
- Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. Martin Kleppmann. O'Reilly Media; 1st edition (April 18, 2017) - 616 pages
- Database Internals: A Deep Dive into How Distributed Data Systems Work. Alex Petrov. O'Reilly Media; 1st edition (October 22, 2019). - 376 pages
- Запис лекцій

<https://www.youtube.com/playlist?list=PLj3AfeFCdPVz9NdfYZtkxwqOZTbLwaEF>

3. Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Лекційні заняття за розділами

- Лекційні заняття

№ з/п	Назва лекції та перелік основних питань
1	Розділ 1 - Особливості високонавантажених систем та недоліки реляційних БД
2	Лекція 1. Вступ <ul style="list-style-type: none">- Особливості високонавантажених систем- Сфери застосування та використання- Проблеми при побудові високонавантажених систем
2	Лекція 2. ACID. Рівні ізоляції транзакцій <ul style="list-style-type: none">- Реляційна модель даних- Недоліки нормальних форм- Властивості ACID- Рівні ізоляції транзакцій
3	Лекція 3. Оптимістичне та песимістичне блокування <ul style="list-style-type: none">- Механізми реалізації ізоляції транзакцій- Multiversion concurrency control- Проблема 'lost update'- Оптимістичне та песимістичне блокування
4	Розділ 2 - NoSQL бази даних
5	Лекція 4. NoSQL бази даних <ul style="list-style-type: none">- Недоліки RDBMS- Нормалізована та де-нормалізована модель даних- Властивості NoSQL
6	Лекція 5. Типи NoSQL бази даних <ul style="list-style-type: none">- Типи NoSQL бази даних- Модель Ключ-значення- Документо-орієнтована модель- Графова модель- Колонкова модель
7	Лекція 6. Колонкова модель <ul style="list-style-type: none">- Колонкова модель даних на прикладі Apache Cassandra- Primary key, partition та clustering keys- Блюм-фільтр- Запити у Cassandra- Обмеження Cassandra
8	Лекція 7. Polyglot Persistence <ul style="list-style-type: none">- Критерії до обрання типу БД

	<ul style="list-style-type: none"> - Переваги, недоліки, обмеження
9	Розділ 3 - Масштабованість та відмовостійкість
10	<p>Лекція 8. Реплікація та шардінг даних</p> <ul style="list-style-type: none"> - Підходи до масштабування - шардінг - Види шардінгу - Підходи до відмовостійкості - реплікація - Види реплікації
11	<p>Лекція 9. Проблема цілісності даних</p> <ul style="list-style-type: none"> - Реплікація та ‘цілісність в кінцевому рахунку’ - Моделі цілісності: eventual consistency, monotonic reads, read your writes, strong consistency - Гарантії цілісності
12	<p>Лекція 10. CAP-теорема</p> <ul style="list-style-type: none"> - Цілісність, доступність та відмовостійкість - CAP-теорема - Системи типу CA, AP, CP - Підходи до обрання систем CA, AP, CP - BASE
13	Розділ 4 - Розподілені запити
14	<p>Лекція 11. Основи Map/Reduce</p> <ul style="list-style-type: none"> - Розподілені запити - ‘Локальність даних’ - Підхід Map/Reduce - Фази Map/Reduce
15	<p>Лекція 12. Особливості Map/Reduce</p> <ul style="list-style-type: none"> - Додаткові фази Map/Reduce - Вимоги до функцій Map, Reduce, Combine - Основні алгоритми Map/Reduce: Word count, Inverted index, Top N
16	<p>Лекція 13. Альтернативи Map/Reduce</p> <ul style="list-style-type: none"> - Недоліки підходу Map/Reduce - Моделювання даних як потоку - Datastream processing model - Проміжні та термінальні функції Datastream processing model
17	<p>Лекція 14. Apache Spark</p> <ul style="list-style-type: none"> - Модель обробки та запитів даних на основі Apache Spark - Переваги Apache Spark - Функції Apache Spark: filter, map, flatMap, distinct, reduce, groupBy, collect
18	<p>Лекція 15. Висновки по курсу</p> <ul style="list-style-type: none"> - Огляд та узагальнення основних моментів

Завданням лабораторних робіт є отримати практичний досвід з використання та налаштування найбільш популярних NoSQL БД.

Теми основних лабораторних робіт:

1. Лабораторна 1. Тестування швидкості вставки значень в БД
2. Лабораторна 2. Основи роботи з MongoDB
3. Лабораторна 3. Основи роботи з Neo4j
4. Лабораторна 4. Основи роботи з Apache Cassandra
5. Лабораторна 5. Налаштування реплікації в MongoDB
6. Лабораторна 6. Налаштування реплікації в Apache Cassandra

Теми додаткових лабораторних робіт:

7. Лабораторна 7. Написання розподілених запитів з використанням Apache Spark
8. Лабораторна 8. Map/Reduce

6. Самостійна робота студента

До самостійної роботи відноситься встановлення та налаштування інфраструктури для виконання перелічених лабораторних робіт, а також виконання пунктів завдання перелічених в описі лабораторних робіт.

4. Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Політика навчальної дисципліни повністю визначається нормативними документами НТУУ «КПІ ім.Ігоря Сікорського».

- Відвідування лекцій не є обов'язковим, але у цьому разі вимагає самостійного опрацювання матеріалу лекції та перегляд їх у запису
- Для зарахування лабораторних робіт необхідно їх захищати особисто, перед цим надіславши протокол, і при необхідності виконати додаткові завдання, на основі завдань з опису лабораторної роботи
- Заохочувальні бали даються за вихід за межі завдання наведеного в описі до лабораторної роботи
- Також заохочувальні бали даються за виконання додаткових робіт
- У випадку порушення дедлайнів по здачі лабораторних роботи, обов'язковою для виконання стає додаткова лабораторна робота, і за кількістю балів вони рахуються як одна
- У разі виявлення несамостійного виконання роботи, робота не зараховується і призначається додаткова - штрафна робота

Академічна доброчесність

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Норми етичної поведінки

Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі

Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Процедура оскарження результатів контрольних заходів

Студенти мають можливість поставити будь-яке питання, яке стосується процедури проведення та/або оцінювання контрольних заходів, та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами.

Студенти мають право оскаржити результати контрольних заходів, але обов'язково аргументовано, пояснивши, з яким критерієм не погоджуються відповідно до оціночного листа та/або зауважень.

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

Основними видами контролю, за які надаються рейтингові бали, є лабораторні роботи та МКР

N	Контрольний захід	%	Ваговий бал	Кіл-ть	Всього
1.	Лабораторні роботи	80	10	8	80
2.	МКР	20	20	1	20
	Всього				100

Система рейтингових (вагових) балів та критерії оцінювання

Модульна контрольна робота

Ваговий бал – 20.

Проводиться у вигляді тесту, де оцінюється відповідь на кожне окреме питання. Сумма правильних відповідей складає загальний бал.

- правильна відповідь на запитання у тесті - 1
- неправильна відповідь на запитання у тесті - 0

Лабораторні роботи

Передбачається 5 обов'язкових завдань та 3 додаткових завдання.

Ваговий бал за завдання – 10

Максимальна кількість балів за лабораторні роботи 8 x 10 = 80 балів

Здача перших 5ти обов'язкових завдань є умовою допуску до Заліку.

У разі наявності штрафних лабораторних робіт їх здача також є умовою допуску до Заліку.

Штрафні та заохочувальні бали за:

- активність на заняттях та систематична самостійна робота протягом семестру+ 1... + 5
- участь на олімпіадах, а також ВНО та наукових конференціях, модернізація лабораторних робіт, виконання завдань із удосконалення методичних та дидактичних матеріалів з дисципліни.....+ 1...+ 5
- здача додаткових лабораторних робіт +10
- за затримку здачі лабораторних робіт даються додаткові (штрафні) лабораторні роботи у максимальній кількості 3 штуки

Залік - 30 балів

Поточний контроль

МКР, який проводиться наприкінці курсу за теоретичними матеріалами курсу

Семестровий контроль

Семестровим контролем є залік

Розрахунок шкали (R) рейтингу:

Сума вагових балів контрольних заходів складає $R = 20+80 = 100$ балів;

Складова залікового контролю шкали дорівнює $R_3 = 30$ балів:

Необхідною умовою допуску до заліку є зарахування 5-ти обов'язкових лабораторних робіт

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

<i>Кількість балів</i>	<i>Оцінка</i>
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)

Перелік питань які виносяться на семестровий контроль:

1. ACID - properties of database transactions
2. Transaction isolation levels
3. Pessimistic vs optimistic locking. Lost update problem
4. SQL vs NoSQL. RDBMS problems. Normalized vs “Denormalized” data model
5. NoSQL properties (shemaless, aggregate orientation, transactions, ...)
6. Types of NoSQL databases
7. Column-family store
8. Cassandra data model: primary key, partition and clustering keys
9. Cassandra Queries. Cassandra limitations
10. Distribution Models. Consistency problem
11. Distribution Models. Sharding
12. Distribution Models. Replication
13. Distributed systems. Consistency models: eventual consistency, monotonic reads, read your writes, strong consistency
14. Distributed systems. Consistency guarantee
15. Distributed systems. MongoDB and Cassandra parameters for consistency guarantee
16. Distributed systems. CAP theorem. BASE.
17. CAP theorem with SQL and NoSQL DBs
18. Polyglot Persistence
19. MapReduce. Data locality. Phases

20. MapReduce. Standard algorithm (Word count, Inverted index, Top N)
21. MapReduce. Map/Reduce/Combine functions requirements
22. MapReduce alternatives
23. Datastream processing model: intermediate and terminal operations, data processing parallelisation
24. Apache Spark
25. Apache Spark RDD transformations and actions: filter, map, flatMap, distinct, reduce, groupBy, collect

Робочу програму навчальної дисципліни (силабус):

Складено доц., к.т.н, Родіонов Андрій Миколайович, andrey.rodionov@gmail.com

Ухвалено кафедрою інформаційної безпеки (протокол No 5/2022 від 22.06.2022)

Погоджено Методичною комісією НН ФТІ (протокол № 6 від 30.06.2022)