



КУРСОВА РОБОТА З ОБРОБЛЕННЯ НАДВЕЛИКИХ МАСИВІВ ДАНИХ

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>другий (магістерський)</i>
Галузь знань	<i>Інформаційні технології</i>
Спеціальність	<i>121 Інженерія програмного забезпечення</i>
Освітня програма	<i>Інженерія програмного забезпечення інформаційних систем</i>
Статус дисципліни	<i>основна</i>
Форма навчання	<i>заочна</i>
Рік підготовки, семестр	<i>1-й курс, осінній семестр</i>
Обсяг дисципліни	<i>1 кредит, 30 годин</i>
Семестровий контроль/ контрольні заходи	<i>Залік (захист курсової роботи)</i>
Розклад занять	<i>Згідно розкладу на осінній семестр поточного навчального року https://schedule.kpi.ua/</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	доцент кафедри ІПІ Олійник Ю.О., yurii.oliinyk-fiot@iit.kpi.ua
Розміщення курсу	https://classroom.google.com/ Код курсу за запрошеним посиланням

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

Мета вивчення дисципліни – набуття ключових фахових компетентностей, теоретичних знань і практичних навичок аналізу, аргументування, прийняття рішень при розв'язанні задач та практичних проблем оброблення та аналізу надвеликих масивів даних.

Предметом вивчення дисципліни є технології, моделі, архітектура розподілених сховищ даних; процеси обробки та аналізу надвеликих масивів даних.

Завдання вивчення дисципліни: – оволодіння основними поняттями обробки та аналізу надвеликих масивів даних; – ознайомлення з новітніми підходами створення розподілених сховищ даних; – набуття практичних навичок обробки та аналізу надвеликих масивів даних для вирішення задач підтримки прийняття рішень.

Навчальна дисципліна покликана допомогти студенту отримати:

- вивчення сучасних концепцій та підходів до оброблення надвеликих масивів даних та створення сховищ даних;
- уміння вільно орієнтуватися на сучасному світі розподілених сховищ даних; проектувати та створювати розподілені сховища даних, застосовувати сучасні методи та технології обробки та аналізу надвеликих масивів даних.
- здатність використовувати можливості сучасних засобів та технологій обробки потоків даних.

КОМПЕТЕНТНОСТІ

Спеціальні (фахові, предметні) компетентності

- ФК02 - Здатність розробляти і реалізовувати наукові та/або прикладні проекти у сфері інженерії програмного забезпечення.
- ФК13 - Здатність до аналізу, проектування та розробки нових та використання існуючих систем зберігання та обробки надвеликих масивів даних.

ПРОГРАМНІ РЕЗУЛЬТАТИ НАВЧАННЯ

- ПРН17 - Збирати, аналізувати, оцінювати необхідну для розв'язання наукових і прикладних задач інформацію, використовуючи науково-технічну літературу, бази даних та інші джерела.
- ПРН20 - Планувати і виконувати наукові дослідження в сфері інженерії програмного забезпечення, обирати методики та інструменти, аналізувати результати, обґрунтовувати висновки.
- ПРН23 – Розробляти, реалізувати та застосовувати різні методи інтелектуального аналізу даних до Big Data, формулювати алгоритми обробки в парадигмі Map Reduce, обирати відповідну технологію зберігання і оброблення надвеликих даних, використовувати сучасні високонавантажені системи зберігання та оброблення великих даних.

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

Для успішного засвоєння дисципліни студент повинен володіти базовими знаннями з:

- Баз даних.
- Аналізу даних.
- Проектування інформаційних систем.
- Теорії алгоритмів.

Знання, отримані студентами при вивченні дисципліни, використовуються у таких дисциплінах:

- ПО 02 – Науково-дослідна практика;
- ПО 03 – Виконання магістерської дисертації

3. Зміст навчальної дисципліни

Основні типові етапи виконання курсової роботи складається з етапів, наведених в таблиці 1.

Таблиця 1. Етапи написання курсової роботи

Найменування	Результат	Години
Підготовчий етап	Узгоджене завдання на курсову роботу керівником та виконавцем	1
Виконавчий етап. <ul style="list-style-type: none">• постановка задачі;• розробка процесів підготовки даних, ETL процесів, проектування структури БД;• опис методів обробки даних;• дослідження ефективності методів обробки даних;	Розділи курсової роботи	26
Підготовка пояснювальної записки та доповіді	Пояснювальна записка та доповідь	3
Публічний захист курсової роботи		

Етапи виконання можуть відрізнятися в залежності від теми.

Тема курсової роботи повинна корелюватися з задачами магістерської дисертації. Задачі повинні вирішені за допомогою технологій BigData.

Приклади тем:

- Сегментація вподобань користувачів на основі технології Apache Spark.
- Класифікація наукових текстів за кодом УДК.
- Розпізнавання ієрогліфів японської мови в потоці даних.

4. Навчальні матеріали та ресурси

Основна література

1. Дистанційний курс на платформі Google Classroom.
2. Олещенко, Л. М. Технології оброблення великих даних. Конспект лекцій [Електронний ресурс] : навчальний посібник для студентів спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем») / Л. М. Олещенко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 227 с.. (доступ за посиланням <https://ela.kpi.ua/handle/123456789/42206>)
3. Ланде, Д. В. Оброблення надвеликих масивів даних (Big Data) [Електронний ресурс] : навчальний посібник для використання у навчальному процесі з підготовки фахівців другого (магістерського) рівня вищої освіти зі спеціальності 122 «Комп'ютерні науки» / Д. В. Ланде, І. Ю. Субач, А. Я. Гладун ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 6,95 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2021. – 168 с. (доступ за посиланням <https://ela.kpi.ua/handle/123456789/46129>)

Додаткова література

1. Пасічник В.В., Шаховська Н.Б. Сховища даних: Навчальний посібник. - Львів: “Магнолія 2006”, 2008.-496 С. ISBN 978-966-2025-18-7.
2. Damji, Jules S., et al. Learning Spark. O'Reilly Media, 2020
3. Saha, B., Shah, H., Seth, S. et al. Apache Tez: A Unifying Framework for Modeling and Building Data Processing Applications // Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (SIGMOD'15). Melbourne, Australia, 2015. P. 1357–1369. DOI: 10.1145/2723372.2742790
4. Graph Databases for Beginners [Електронний ресурс] / Neo4j, Inc. – Neo4j – Режим доступу: https://neo4j.com/wp-content/themes/neo4jweb/assets/images/Graph_Databases_for_Beginners.pdf
5. Van Bruggen, R. Learning Neo4j [Електронний ресурс]. – Birmingham : Packt Publishing, 2014. – 195 с. – Режим доступу: https://go.neo4j.com/rs/710-RRC-335/images/LearningNeo4j_eBook.pdf
6. Apache Software Foundation. Apache Hive Documentation [Електронний ресурс]. – Режим доступу: <https://hive.apache.org/docs/latest/>

Додаткові матеріали до курсу «Оброблення надвеликих масивів даних» доступні для завантаження здобувачам, які зареєстровані на курс на платформі classroom.google.com/

Інформаційні ресурси

- На персональному комп'ютері або в хмарному середовищі мають бути встановлені: Apache Hadoop, Apache Hive, Apache Spark версії 3 та вище, які розповсюджуються по безкоштовній ліцензії.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Для опанування навчальної дисципліни застосовуються наступні методи:

- Пояснювально-ілюстративного метод. Послідовна та логічно ув'язана подача матеріалу надає уявлення та знання у його логічної цілісності.
- Метод проблемного викладу надає уяву та методи отримання нових знань та фактів з використанням вже відомих фактів та тверджень.
- Репродуктивний метод, завдяки якому студенти закріплюють вивчений теоретичний матеріал та навчаються використовувати його в конкретних задачах.
- Проблемний метод, при застосуванні якого студенти залучаються до обговорення та вирішення задач, пов'язаних з новітніми інформаційними технологіями аналітичної обробки інформації.
- Самостійна робота з можливістю особистих консультацій з викладачем.

6. Самостійна робота студента

Підготовка курсової роботи відбувається в межах самостійної роботи згідно плану етапів підготовки.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Форми організації освітнього процесу, види навчальних занять і оцінювання результатів навчання регламентуються Положенням про організацію освітнього процесу в Національному технічному університеті України «Київському політехнічному інституті імені Ігоря Сікорського».

Політика виставлення оцінок: кожна оцінка виставляється відповідно до розроблених викладачем та заздалегідь оголошених студентам критеріїв, а також мотивується в індивідуальному порядку на вимогу студента.

Дотримання академічної доброчесності. Обов'язковою умовою виконання завдань з освітньої компоненти є дотримання політики та принципів академічної доброчесності (<https://kpi.ua/academic-integrity>), які, у тому числі, викладено у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>), Положенні про систему запобігання академічному плагіату в КПІ ім. Ігоря Сікорського (<https://osvita.kpi.ua/node/47>). У разі виявлення дублювання робіт, плагіату роботи здобувачі отримують нульовий рейтинг.

Політика використання штучного інтелекту. Використання штучного інтелекту (далі, ШІ) регламентується «Політикою використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського» (<https://osvita.kpi.ua/node/1225>). Усі навчальні завдання з дисципліни мають бути результатом власної оригінальної роботи здобувача.

Норми академічної етики: дисциплінованість; дотримання субординації; чесність; відповідальність; робота в аудиторії з відключеними мобільними телефонами. Повага один до одного дає можливість ефективніше досягати поставлених командних результатів. Якщо ви використовуєте свій ноутбук чи телефон для аудіо чи відеозапису, необхідно заздалегідь отримати дозвіл викладача.

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

Поточний контроль: перевірка дотримання графіку виконання календарного плану курсової роботи.

Семестровий контроль: залік.

Керівник здійснює контроль за ходом виконання студентом курсової роботи, надає йому необхідну консультативну допомогу. Протягом семестру студент демонструє викладачу поточні результати роботи над проектом. Після перевірки роботи викладач призначає день, час і місце захисту.

Напередодні захисту студенту необхідно повторити теоретичний матеріал, що стосується роботи, та переглянути безпосередньо її зміст. Захист КР проводиться у формі співбесіди зі з'ясуванням усіх питань, що виникли у керівника під час перевірки роботи.

За результатами захисту, у відповідності до критеріїв оцінювання викладачі виставляють студенту оцінку.

На оцінку за КР впливають:

- якість розробленого програмного забезпечення;
- якість розробленої програмної документації;
- компетентність та загальна ерудиція студента при відповідях на запитання під час захисту;
- ступінь виконання графіку підготовки курсової роботи.

Якщо студент подав на захист не самостійно виконану роботу, про що свідчить його некомпетентність у рішеннях та матеріалах роботи, КР до захисту не допускається, що супроводжується записом "не допущений" у екзаменаційній відомості. Такий самий запис робиться у випадку, якщо КР не завершена на час захисту. В цих випадках запис "не допущений" еквівалентний отриманню оцінки "незадовільно".

Система рейтингових балів.

Рейтингова оцінка з курсової роботи має дві складові: виконання курсової роботи та її захист.

Перша (стартова) складова характеризує роботу студента з курсового проектування та її результат - якість пояснювальної записки та розробленого програмного забезпечення; друга складова характеризує якість захисту студентом курсової роботи.

Розмір шкали першої складової (r1) дорівнює 70 балів, а другої складової (r2)- 30 балів.

Загальна рейтингова оцінка $R = r1 + r2$

Стартова складова виконання курсової роботи (r1):

- ступінь розкриття теоретичних аспектів теми та коректність використання понятійного апарату – до 5 балів;
- повнота та коректність предметної області та задачі обробки даних – до 10 балів;
- якість написання та оформлення програмного коду та моделі бази даних та ETL процесів – до 20 балів;
- складність обраних методів обробки – до 20 балів;
- якість оформлення пояснювальної записки з урахуванням виконання вимог нормативних документів – до 15 балів;

Складова захисту курсової роботи (r2):

- ступінь володіння теоретичним матеріалом – до 5 балів;
- ступінь володіння кодом програми в цілому – до 15 балів;

- вміння внести зміни у програмний код – до 10 балів.

Узагальнені критерії оцінювання захисту (складова r2) курсової роботи та деталізовані бальні шкали наведено в наступній таблиці

Параметри оцінювання	Діапазон балів	Критерії оцінювання за бальною шкалою
Ступінь володіння теоретичним матеріалом	0-5	0 – студент не дав відповіді на теоретичні питання
		1-4 – відповідь, на одне чи два теоретичних питання
		5 - відповідь, на усі теоретичні питання
Ступінь володіння кодом програми в цілому	0-15	0 – студент не володіє кодом програми
		1-3 – студент володіє лише власним кодом
		4-14 – студент частково володіє кодом своїх колег по курсовій роботі
		15 – студент повністю володіє кодом своїх колег по курсовій роботі
Вміння внести зміни у програмний код	0-10	0 – студент не може виконати додаткові завдання, пов'язані із внесенням змін у програмний код
		1-9 – студент частково може виконати додаткові завдання, пов'язані із внесенням змін у програмний код
		10 – студент грамотно виконує додаткові завдання, пов'язані із внесенням змін у програмний код

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

R = r1+ r2	Національна оцінка
95... 100	відмінно
85 ... 94	добре
75 ... 84	
65 ... 74	задовільно
60 ... 64	
Менше 60	незадовільно
Курсовий проект не допущено до захисту	не допущено

Робочу програму навчальної дисципліни (силабус):

Складено доцент кафедри ІІІ, Олійник Ю.О.

Ухвалено кафедрою ІІІ (протокол №16 від 29.05.2024 р.)

Погоджено Методичною комісією факультету (протокол № 10 від 21.06.2024 р.)