



Силабус освітнього компонента Програма навчальної дисципліни



Інфраструктура та менеджмент великих даних

Шифр та назва спеціальності
113 – Прикладна математика

Інститут
ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма
Інтелектуальний аналіз даних

Кафедра
Комп'ютерна математика і аналіз даних

Рівень освіти
Бакалавр

Тип дисципліни
Спеціальна (фахова), Вибіркова

Семестр
7

Мова викладання
Українська

Викладачі, розробники



Костюк Ольга Василівна

Olha.Kostiuk@khi.edu.ua

Кандидат технічних наук, доцент кафедри комп'ютерної математики і аналізу даних НТУ «ХПІ»

Автор наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Нечіткі моделі та методи», «Теорія прийняття рішень», «Вступ до спеціальності та інженерної діяльності»

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна спрямована на оволодіння теоретичними основами основних принципів, підходів та напрямів технологій та інфраструктури великих даних. Розглянуто підходи та визначення, надано огляд системи великих даних та розкрито тему систем управління великими даними і їх практичне застосування.

Мета та цілі дисципліни

Метою вивчення дисципліни є набуття необхідних компетентностей для прикладного застосування основ управління великомасштабними базами даних, зокрема: здатність обговорити і вирішити проблеми, які виникають у зв'язку з великими даними; здатність порівняти та співставити різні програмні та інфраструктурні архітектури, які можна застосувати для управління великими даними, і вибрати відповідну архітектуру для даної проблеми; здатність порівнювати різні підходи до реалізації систем великих даних; здатність розробити та створити велику масштабовану базу даних і використовувати її для відповідей на складні запити;

здатність виявлення та усунення вузьких місць у продуктивності архітектури та впровадження середовища даних.

Формат занять

Лекції, лабораторні роботи, розрахункові завдання, самостійна робота, консультації. Підсумковий контроль – іспит.

Компетентності

- ЗК 1. Здатність учитися й оволодівати сучасними знаннями.
- ЗК 2. Здатність застосовувати знання у практичних ситуаціях.
- ЗК 5. Здатність проведення досліджень на відповідному рівні.
- ЗК 7. Здатність до пошуку, оброблення й аналізу інформації з різноманітних джерел.
- ЗК 8. Знання і розуміння предметної області та розуміння професійної діяльності.
- ЗК 9. Здатність спілкуватися з представниками інших професійних груп різного рівня (з експертами з інших галузей знань/видів економічної діяльності)
- ЗК 10. Навички у використанні інформаційних і комунікаційних технологій.
- СК 5. Здатність розробляти алгоритми та структури даних, програмні засоби та програмну документацію.
- СК 6. Здатність проектувати бази даних, інформаційні системи та ресурси.
- СК 7. Здатність розв'язувати професійні задачі за допомогою комп'ютерної техніки, комп'ютерних мереж та Інтернету, в середовищі сучасних операційних систем, із використанням стандартних офісних додатків.
- СК 8. Здатність експлуатувати та обслуговувати програмне забезпечення автоматизованих та інформаційних систем різного призначення.
- СК 21. Здатність до розробки та експлуатації програмних засобів обробки великих масивів даних на основі інформаційних технологій розподілених і хмарних обчислень.

Результати навчання

- РН 1. Демонструвати знання й розуміння основних концепцій, принципів, теорій прикладної математики і використовувати їх на практиці.
- РН 7. Вміти проводити практичні дослідження та знаходити розв'язок некоректних задач.
- РН 11. Вміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символічних алгоритмів.
- РН 14. Виявляти здатність до самонавчання та продовження професійного розвитку.
- РН 25. Вміти застосовувати сучасні інформаційні технології та програмне забезпечення для обробки великих масивів даних на основі розподілених і хмарних сервісів.

Обсяг дисципліни

Загальний обсяг дисципліни 120 год.: лекції – 30 год., лабораторні роботи – 30 год., самостійна робота – 60 год.

Передумови вивчення дисципліни (пререквізити)

«Теорія і проектування баз даних», «Алгоритмічні мови (за вибором)», «Алгоритмізація та програмування», «Дискретні структури і структури даних».

Особливості дисципліни, методи та технології навчання

Особливістю викладання є використання елементів проектної роботи.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Введення у великі дані

- визначення великих даних;
- історія та важливість великих даних.

Тема 2. Архітектура великих даних

- огляд основних компонентів;
- принципи проектування.

Тема 3. Технології обробки великих даних

- Hadoop;
- Spark;
- інші фреймворки.

Тема 4. Зберігання та управління даними

- бази даних NoSQL;
- системи управління базами даних;
- сховища даних.

Тема 5. Хмарні платформи для великих даних

- AWS;
- Google Cloud;
- Azure та інші.

Тема 6. Аналіз великих даних

- інструменти та методи для отримання цінної інформації з великих обсягів даних.

Тема 7. Візуалізація великих даних

- інструменти та техніки для представлення великих даних у зрозумілій формі.

Тема 8. Безпека та конфіденційність у великих даних

- проблеми та рішення.

Тема 9. Машинне навчання та великі дані

- використання великих даних для побудови моделей машинного навчання.

Тема 10. Інтеграція та обробка поточкових даних

- робота з реальними потоками даних.

Тема 11. Управління проектами великих даних

- планування;
- керування;
- моніторинг проектів великих даних.

Тема 12. Застосування великих даних у різних галузях

- приклади використання великих даних у охороні здоров'я;
- приклади використання великих даних у фінансах;
- приклади використання великих даних у роздрібній торгівлі та інших галузях.

Тема 13. Майбутнє великих даних

- тренди та прогнози розвитку.

Тема 14. Етика та великі дані

- обговорення етичних питань, пов'язаних із використанням великих даних.

Тема 15. Аналіз конкретних прикладів успішного використання великих даних

- приклади успішного використання великих даних у бізнесі;
- приклади успішного використання великих даних у науці.

Теми практичних занять

Практичні заняття не передбачені навчальним планом.

Теми лабораторних робіт

Лабораторна робота 1. Установка та налаштування Hadoop: Практичне знайомство з основним фреймворком для роботи з великими даними

Лабораторна робота 2. Робота з HDFS: Вивчення основних операцій із файловою системою Hadoop

Лабораторна робота 3. MapReduce: Розробка та тестування простих завдань MapReduce

Лабораторна робота 4. Встановлення та налаштування Spark: Знайомство з іншим популярним фреймворком для роботи з великими даними

Лабораторна робота 5. Робота з RDD у Spark: Створення та трансформація RDD, основного абстрактного типу даних у Spark

Лабораторна робота 6. Spark SQL та DataFrames: Використання Spark для роботи з даними у форматі SQL та DataFrame

Лабораторна робота 7. Встановлення та використання NoSQL баз даних: Робота з такими системами як MongoDB або Cassandra
Лабораторна робота 8. Хмарні послуги для великих даних: Практичне знайомство з хмарними платформами, такими як AWS або Google Cloud
Лабораторна робота 9. Візуалізація великих даних: Створення інтерактивних візуалізацій великих даних.
Лабораторна робота 10. Машинне навчання з використанням Spark MLlib: Розробка та тестування моделей машинного навчання на великих даних
Лабораторна робота 11. Обробка поточкових даних за допомогою Kafka: Робота з реальними потоками даних
Лабораторна робота 12. Управління проектами великих даних із використанням Jira: Планування та відстеження прогресу проекту
Лабораторна робота 13. Великі дані у різних галузях: Розробка рішень для конкретних галузей з урахуванням великих даних
Лабораторна робота 14. Безпека великих даних: Розробка та тестування заходів безпеки для захисту великих даних
Лабораторна робота 15. Розробка та аналіз рішень для реальних бізнес-проблем на основі великих даних

Самостійна робота

Парадигма MapReduce.
Платформа для розподілених обчислень Hadoop.
Застосунок Apache Pig.
Застосунок Apache Hive.
Диспетчер робіт Apache.
Створення та налаштування власного потоку робіт для кластеру Hadoop.

Література та навчальні матеріали

Основна література

1. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019. – 105 p.
2. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020. – 298 p.
3. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. – 110 p.
4. Raheem N. Big Data: A Tutorial-Based Approach. Taylor & Francis Group LLC, CRC Press, 2019. – 203 p.

Додаткова література

5. Davy Cielen, Arno D. B. Meysman, and Mohamed Ali. Introducing Data Science. Big data, machine learning, and more, using Python tools.
<https://www.manning.com/books/introducing-data-science>
6. Технології Big Data. Лабораторний практикум. Навч. посібник для здобувачів ступеня магістр за спеціальністю 123 «Комп'ютерні системи та мережі» / Таран В. К.: КПІ ім. Ігоря Сікорського, 2022. – 26 с.
<https://comsys.kpi.ua/metodichni-vkazannya-po-disciplinam>
7. Peter Ghavami. Big Data Management. Data Governance Principles for Big Data Analytics.
<https://doi.org/10.1515/9783110664065>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

100% підсумкової оцінки складаються з результатів оцінювання у вигляді іспиту (40%) та поточного оцінювання (60%).

Іспит: письмове завдання (2 теоретичних питання і дві задачі) та усна доповідь.

Поточне оцінювання: оцінки за лабораторні роботи та розрахункове завдання.

Шкала оцінювання

Сума балів	Національна оцінка	ECTS
90–100	Відмінно	A
82–89	Добре	B
75–81	Добре	C
64–74	Задовільно	D
60–63	Задовільно	E
35–59	Незадовільно (потрібне додаткове вивчення)	FX
1–34	Незадовільно (потрібне повторне вивчення)	F

Норми академічної етики і політика курсу

Студент повинен дотримуватися «Кодексу етики академічних взаємовідносин та доброчесності НТУ «ХПІ»: виявляти дисциплінованість, вихованість, доброзичливість, чесність, відповідальність. Конфліктні ситуації повинні відкрито обговорюватися в навчальних групах з викладачем, а при неможливості вирішення конфлікту – доводитися до відома співробітників дирекції інституту. Нормативно-правове забезпечення впровадження принципів академічної доброчесності НТУ «ХПІ» розміщено на сайті: <http://blogs.kpi.kharkov.ua/v2/nv/akademichna-dobrochesnist/>

Погодження

Силабус погоджено

Дата погодження, підпис
31.08.2023 р.

Завідувач кафедри
Олена АХІЄЗЕР

Дата погодження, підпис
31.08.2023 р.

Гарант ОП
Олена АХІЄЗЕР