



Силабус освітнього компонента

Програма навчальної дисципліни



Технології BigData

Шифр та назва спеціальності

121 – Інженерія програмного забезпечення
122 – Комп'ютерні науки

Інститут

ННІ Комп'ютерних наук та інформаційних технологій

Освітня програма

Інженерія програмного забезпечення
Комп'ютерні науки та інтелектуальні системи

Кафедра

Програмна інженерія та інтелектуальні технології управління (321)

Рівень освіти

Бакалавр

Тип дисципліни

Спеціальна (фахова), Вибіркова

Семестр

5

Мова викладання

Українська, англійська

Викладачі, розробники



Бурдаєв Володимир Петрович

volodymyr.burdaev@khpi.edu.ua

К.ф.-м.н., с.н.с., доцент кафедри програмної інженерії та інтелектуальних технологій управління

Google

Scholar: <https://scholar.google.com/citations?user=&user=RX9JedIAAAA>

ORCID: <https://orcid.org/0000-0001-9848-9059>

Scopus: <https://www.scopus.com/authid/detail.uri?authorId=6507982230>,
<https://www.scopus.com/authid/detail.uri?authorId=57224197566>.

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Завданням дисципліни є засвоєння студентами необхідного рівня знань використання технологій великих даних: зберігання, обробка, аналіз, візуалізація та застосування. Студенти вивчатимуть основи сучасних обчислювальних платформ для великих даних, а також віддалені сервісні функції, що базуються на організації та застосуванні технологій хмарних обчислень

Мета та цілі дисципліни

Формування у студентів системи теоретичних знань та набуття практичних умінь та навичок щодо використання технологій великих даних, основних моделей, надання послуг хмарних обчислень, розробок веб-додатків для проведення наукових досліджень у хмарному середовищі.

Формат занять

Лекції, лабораторні роботи, самостійна робота, консультації. Підсумковий контроль – залік.

Компетентності

121 - Інженерія програмного забезпечення

K05. Здатність вчитися і оволодівати сучасними знаннями.

K06. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

K19. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.

K24. Здатність здійснювати процес інтеграції системи, застосовувати стандарти і процедури управління змінами для підтримки цілісності, загальної функціональності і надійності програмного забезпечення.

K25. Здатність обґрунтовано обирати та освоювати інструментарій з розробки та супроводження програмного забезпечення.

122 - Комп'ютерні науки та інтелектуальні системи

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.

ЗК6. Здатність вчитися й оволодівати сучасними знаннями.

ЗК9. Здатність працювати в команді.

СК8. Здатність проектувати та розробляти програмне забезпечення із застосуванням різних парадигм програмування: узагальненого, об'єктно-орієнтованого, функціонального, логічного, з відповідними моделями, методами й алгоритмами обчислень, структурами даних і механізмами управління.

СК9. Здатність реалізувати багаторівневу обчислювальну модель на основі архітектури клієнт-сервер, включаючи бази даних, знань і сховища даних, виконувати розподілену обробку великих наборів даних на кластерах стандартних серверів для забезпечення обчислювальних потреб користувачів, у тому числі на хмарних сервісах.

Результати навчання

121 - Інженерія програмного забезпечення

ПР01. Аналізувати, цілеспрямовано шукати і вибирати необхідні для вирішення професійних завдань інформаційно-довідникові ресурси і знання з урахуванням сучасних досягнень науки і техніки.

ПР07. Знати і застосовувати на практиці фундаментальні концепції, парадигми і основні принципи функціонування мовних, інструментальних і обчислювальних засобів інженерії програмного забезпечення.

ПР13. Знати і застосовувати методи розробки алгоритмів, конструювання програмного забезпечення та структур даних і знань.

ПР14. Застосовувати на практиці інструментальні програмні засоби доменного аналізу, проектування, тестування, візуалізації, вимірювань та документування програмного забезпечення.

ПР15. Мотивовано обирати мови програмування та технології розробки для розв'язання завдань створення і супроводження програмного забезпечення.

ПР18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.

122 - Комп'ютерні науки та інтелектуальні системи

ПР9. Розробляти програмні моделі предметних середовищ, вибирати парадигму програмування з позицій зручності та якості застосування для реалізації методів та алгоритмів розв'язання задач в галузі комп'ютерних наук.

ПР10. Використовувати інструментальні засоби розробки клієнт-серверних застосувань, проектувати концептуальні, логічні та фізичні моделі баз даних, розробляти та оптимізувати запити до них, створювати розподілені бази даних, сховища та вітрини даних, бази знань, у тому числі на хмарних сервісах, із застосуванням мов веб-програмування.

Обсяг дисципліни

Загальний обсяг дисципліни 150 год. (5 кредити ECTS): лекції – 32 год., лабораторні роботи – 32 год., самостійна робота – 86 год.

Передумови вивчення дисципліни (пререквізити)

Об'єктно-орієнтоване програмування

Основи веб-розробки

Особливості дисципліни, методи та технології навчання

Методи викладання та навчання:

інтерактивні лекції з презентаціями, дискусії, лабораторні заняття, командна робота, кейс-метод, метод зворотного зв'язку з боку студентів, проблемне навчання.

Форми оцінювання:

письмові індивідуальні завдання до лабораторних робіт (CAS), оцінювання знань на лабораторних заняттях (CAS), експрес-опитування (CAS), онлайн-тести (CAS), підсумковий/семестровий контроль у формі семестрового екзамену, відповідно до графіку навчального процесу (FAS).

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Джерела великих даних. Визначення Big Data

Інтернет Речей та зростання даних. Визначення великих даних. Приклади великих даних у реальному світі. Відкриті дані. Приватність даних. Структуровані та неструктуровані дані. Хмарні обчислення. Інфраструктура великих даних. Розподілені дані та їх обробка.

Тема 2. Онтологія для великих даних

Людський мозок та онтологія. Властивості онтологій. Переваги онтологій. Компоненти онтологій. Роль онтологій у великих даних. Вирівнювання онтологій. Цілі онтологій у великих даних. Проблеми з онтологією у великих даних. RDF – універсальний формат даних. Використання OWL, мови веб-онтологій.

Тема 3. Вивчення великих даних

Контрольоване та неконтрольоване машинне навчання. Модель програмування Spark. Бібліотека Spark MLlib. Регресійний аналіз. Лінійна регресія. Метод найменших квадратів. Кластеризація даних. Алгоритм K-середніх. Системи рекомендацій на основі контенту.

Тема 4. Технології віртуалізації

Базові складові хмарних обчислень. Поняття віртуалізації комп'ютерних систем та мереж. Огляд систем віртуалізації мереж, комп'ютерних ресурсів, додатків та сховищ даних. Визначення віртуалізації рівня додатків та операційних систем. Серверна віртуалізація. Поняття віртуалізації операційних систем. Ознайомлення із поняттями та технологіями перетворення серверного рішення до віртуальної машини, міграції віртуальних машин та «живої міграції» даних.

Тема 5. Стеки хмарних платформ

Класифікація систем хмарних обчислень. Визначення систем: IaaS, PaaS, SaaS. IaaS – інфраструктура як сервіс. PaaS – платформа як сервіс. SaaS – програмне забезпечення як сервіс. Поняття бізнес-моделі надання програмного забезпечення у оренду. Огляд основних провайдерів хмарних обчислень.

Тема 6. Хмарна платформа Microsoft Azure

Особливості платформи. Історичні відомості про впровадження платформи. Основні складові платформи. Технології, що підтримуються хмарою Microsoft Azure. Приклади застосування.

Тема 7. Відкриті дані, їх формати та засоби оброблення

Можливості інструментів аналізу даних за допомогою Python Pandas. Роль Python в аналізі даних. Традиційна аналітика великих даних та аналітика нового покоління. Життєвий цикл аналізу даних. Відкриті дані, їх формати та засоби обробки Веб-скрепінг. Витягування, перетворення та завантаження даних.

Тема 8. Форматування даних про час та дату, читання та запис файлів в Python

Взаємодія із зовнішніми додатками. Форматування даних про час та дату у Python. Читання та запис файлів в Python. Взаємодія із зовнішніми додатками.

Тема 9. Програмування Python та основні операції SQL

Робота Python з SQL. Python та реляційні бази даних трьох типів: SQLite, MySQL та PostgreSQL. Використання Jupyter Notebook для обробки даних. Ефективна навігація в блокноті. Особливості Python для практичної науки про дані. Словники. Класи. Візуалізації в блокноті Jupyter Notebook.

Тема 10. Процедура імпорту даних із файлів у Pandas

Імпорт даних з мережі Інтернет. Засоби для кореляційного аналізу в Pandas. Статистичні підходи до аналітики великих даних. Використання Pandas. Імпорт даних з файлів. Імпорт даних з мережі Інтернет. Описова статистика в Pandas. Засоби для кореляційного аналізу в Pandas.

Тема 11. Перетворення типів даних та маніпулювання данафреймами у Python

Обробка відсутніх даних. Перетворення типів даних. Маніпулювання датафреймами. Методи та типи аналізу машинного навчання. Регресійний аналіз. Типи регресійного аналізу. Застосування регресійного аналізу.

Тема 12. Помилки в аналізі даних та прогнозній аналітиці

Оцінка помилок регресії засобами Python. Призначення бібліотеки scikit-learn. Помилки в аналізі даних та прогнозній аналітиці. Оцінка помилок регресії засобами Python. Призначення бібліотеки scikit-learn.

Тема 13. Алгоритми класифікації даних

Застосування та проблеми класифікацій. Модель класифікатора дерева рішень. Проблеми класифікації. Алгоритми класифікації. Візуалізація класифікацій. Застосування та валідація класифікацій.

Тема 14. Типи візуалізації даних

Візуалізація аномалій. Модуль Pypplot. Інструмент Plotly. Візуалізація аномалій.

Тема 15. Платформа Apache Spark

Проблема обчислювальної функції. Технологія Spark. Порівняння Spark та MapReduce. Resilient Distributed Dataset (RDD). DataFrame. DataSet. Бібліотека машинного навчання – Mllib. ML Pipelines.

Тема 16. Великі дані та системи штучного інтелекту

Піраміда результатів. Що людський мозок робить найкраще. Сенсорне введення. Зберігання. Обчислювальна потужність. Низьке енергоспоживання. Що електронний мозок робить найкраще. Обробка повним перебором. Великі дані. Еволюція від тупих до розумних машин. Інтелект. Типи інтелекту. Класифікація розвідувальних завдань. Платформи великих даних. Пакетна обробка. Обробка в реальному часі. Інтелектуальні програми з великими даними. Области штучного інтелекту.

Теми практичних занять

Практичні заняття в рамках дисципліни не передбачені.

Теми лабораторних робіт

Тема 1. Розгортання веб-додатків ASP.NET в службі додатків Azure за допомогою Visual Studio

Тема 2. Системи контролю версій (Git)

Тема 3. Розгортання бази даних у Microsoft Azure. Створення інфологічної моделі та запитів

Тема 4. Міграція БД SQL Server в Microsoft Azure за допомогою Data Migration Assistant

Тема 5. Обробка великих даних з Apache Spark

Тема 6. Обробка великих даних з Apache Spark (DataFrame API та Spark SQL)

Тема 7. Побудова прогностичних моделей в Microsoft Azure ML Studio (модель регресії)

Тема 8. Побудова прогностичних моделей в Microsoft Azure ML Studio (модель класифікації, модель кластеризації)

Самостійна робота

Індивідуальні завдання навчальним планом не передбачені.

Студентам рекомендуються додаткові матеріали (відео, статті) для самостійного вивчення та опрацювання.

Література та навчальні матеріали

Основна література

1. J. Perrin, Spark in action (2nd ed.). (Covers Apache Spark 3 with examples in Java, Python, and Scala), Manning, 2020, 576 p.
2. J. Damji, B. Wenig, T. Das, D. Lee, Learning spark (2nd ed.) O'Reilly Media Inc., 2020, 672 p.
3. S. Nudurupati, Essential PySpark for scalable data analytics: A beginner's guide to harnessing the power and ease of PySpark 3 Packt Publishing, 2021, 308 p.

4. K. Ramcharan, K. Sundar, S. Alla, Applied data science using PySpark: Learn the end-to-end predictive model-building cycle Apress, 2020, 138 p.
5. M. Lathkar, Python Data Persistence: With SQL and NoSQL Databases, BPB Publications, 2019, 316 p.

Додаткова література

1. B. Murray, Big Data for Beginners: Book 1 - An Introduction to the Data Collection, Storage, Data Cleaning and Preprocessing, Independently published, 2023, 57 p.
2. Saswat Sarangi, Pankaj Sharma, Big Data: A Beginner's Introduction, Routledge India, 2019, 138 p.
3. B. Murray. Big Data for Beginners: Book 2 - An Introduction to the Data Analysis, Visualization, Integration, Interoperability, Governance and Ethics, Independently published, 2023, 60 p.
4. Thomas H. Davenport, Nitin Mittal, All-in On AI: How Smart Companies Win Big with Artificial Intelligence, Harvard Business Review Press , 2023, 224 p.
5. A. J. Gutman, J. Goldmeier, Becoming a Data Head: How to Think, Speak, and Understand Data Science, Statistics, and Machine Learning, Wiley, 2021, 272 p.

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

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

40% залік

60% поточне оцінювання:

Контрольна робота №1 (10%)

Контрольна робота №2 (10%)

Лабораторні роботи (40%)

Лабораторна робота №1 (5%)

Лабораторна робота №2 (5%)

Лабораторна робота №3 (5%)

Лабораторна робота №4 (5%)

Лабораторна робота №5 (5%)

Лабораторна робота №6 (5%)

Лабораторна робота №7 (5%)

Лабораторна робота №8 (5%)

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

08.06.2023

Завідувач кафедри
Ігор ГАМАЮН

08.06.2023

Гарант ОП
Андрій КОПП
Юлія ЛІТВИНОВА

