



Силабус освітнього компонента

Програма навчальної дисципліни



Методи та технології роботи з великими даними

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
Магістр

Тип дисципліни
Профільна, Вибіркова

Семестр
1

Мова викладання
Українська

Викладачі, розробники



Любчик Леонід Михайлович

leonid.liubchuk@khpі.edu.ua

Доктор технічних наук, професор.

Кількість наукових та навчальних публікацій понад 200.

Провідний лектор з дисциплін: «Теорія керування», «Некоректні задачі обробки даних», «Прогнозний аналіз».

[Детальніше про викладача на сайті кафедри](#)



Ямковий Клим Сергійович

klym.yamkovyi@cs.khpі.edu.ua

асистент

Кількість наукових та навчальних публікацій 10.

Загальна інформація

Анотація

Вивчаються функціональні можливості основних компонентів інфраструктури та інструменти Великих Даних. Особлива увага приділяється розумінню та використанню екосистеми Apache Hadoop як основної платформи для Великих Даних, її основних функціональних компонентів MapReduce, Spark, HBase, Hive, Pig та підтримуваних мов програмування Pig Latin та Hive. Курс також дає інформацію щодо проблеми безпеки та відповідності галузевим вимогам по управлінню даних, включаючи питання, пов'язані із Загальним Регламентом Захисту Даних ЄС (GDPR – General Data Protection Regulation).

Мета та цілі дисципліни

Метою вивчення дисципліни є вивчення основних понять Великих Даних та пов'язаних з ними технологій, набуття знань та навичок вибору та оцінювання інфраструктурних сервісів Великих Даних від основних постачальників хмарних послуг (Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP) та інші), для вирішення основних задач управління та аналізу даних

підприємств, набуття навичок для вибору, та розгортання кластеру Hadoop або Spark на одній із хмарних платформ (Azure HDInsight, Amazon EMR), або інші), програмування завдань з використанням однієї із скриптових мов програмування HiveQL, Pig Latin.

Формат занять

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

Компетентності

ЗК 4. Здатність виявляти, ставити та вирішувати проблеми у професійній діяльності.

ЗК 7. Здатність працювати з інформацією: знаходити і використовувати інформацію з різних джерел, потрібну для розв'язання професійних завдань.

ЗК 10. Здатність здійснювати професійну наукову та проєктно-виробничу діяльність у міжнародному середовищі.

СК 5. Здатність до проведення математичного і комп'ютерного моделювання та обчислювального експерименту, збору, візуалізації, аналізу та обробки отриманих даних, розв'язання формалізованих задач за допомогою спеціалізованих програмних засобів.

СК 7. Здатність до пошуку, вивчення та аналізу науково-технічної інформації, вітчизняного і закордонного досвіду, пов'язаного із застосуванням математичних методів для дослідження процесів та систем.

СК 10. Здатність обирати, розробляти, досліджувати та застосовувати математичні моделі і методи для інтелектуального аналізу даних в умовах невизначеності.

СК 13. Здатність до розробки та експлуатації спеціалізованих програмних засобів обробки великих масивів даних на основі інформаційних технологій розподілених і хмарних обчислень.

СК 14. Здатність до використання сучасних інформаційних технологій інтелектуального аналізу даних, прогнозування, прийняття рішень, інформаційного пошуку і видобування знань..

Результати навчання

РН 4. Уміти поєднувати методи математичного і комп'ютерного моделювання з неформальними процедурами експертного аналізу для пошуку оптимальних рішень.

РН 7. Уміти застосовувати сучасні технології програмування та розроблення програмного забезпечення, програмної реалізації чисельних і символічних алгоритмів.

РН 8. Уміти застосовувати у практичній роботі спеціалізовані програмні продукти і програмні системи комп'ютерної математики, аналізу великих даних тощо.

РН 11. Демонструвати навички професійного спілкування, усної та письмової комунікації українською мовою і принаймні ще однією з європейських мов.

РН 14. Уміти застосовувати наявні існуючі і розробляти нові алгоритми та програмні засоби для статистичного й інтелектуального аналізу невизначених даних.

РН 16. Уміти застосовувати сучасні інформаційні технології та програмне забезпечення для обробки великих масивів даних на основі розподілених і хмарних сервісів.

Обсяг дисципліни

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

Передумови вивчення дисципліни (пререквізити)

Бакалаврський рівень підготовки.

Особливості дисципліни, методи та технології навчання

Інтерактивні лекції з презентаціями. Проблемні лекції, робота в малих групах, командна робота. Лабораторні роботи виконуються на хмарній платформі та кластері Hadoop (AWS або Azure).

Програма навчальної дисципліни

Теми лекційних занять

- Тема 1. Основи хмарних технологій.
- Тема 2. Моделі хмарних сервісів, хмарні ресурси, функціонування хмарних служб.
- Тема 3. Еталонна архітектура Великих Даних та приклади використання.
- Тема 4. Основи Великих Даних.
- Тема 5. Хмарні платформи для великих даних. Огляд та порівняння.
- Тема 6. Хмарні реляційні бази даних AWS RDS, AWS Aurora.
- Тема 7. Базова концепція оркестрації та AirFlow.
- Тема 8. Компоненти AirFlow.
- Тема 9. Сучасні великомасштабні бази даних AWS Aurora, Azure CosmosDB, Google Spanner.
- Тема 10. Основи Secondary Index: SOLR.
- Тема 11. Основи Elasticsearch/Opensearch.
- Тема 12. Основи повнотекстового пошуку.
- Тема 13. Основи векторного пошуку.
- Тема 14. Архітектура великих даних підприємства та управління великими даними.
- Тема 15. Проблеми безпеки Великих Даних, захист даних. Контроль доступу та управління ідентичністю.
- Тема 16. Семінар. Огляд курсу та обговорення досягнутих результатів.

Теми практичних занять

Немає

Теми лабораторних робіт

- Лабораторне заняття 1. Основні провайдери хмарних послуг AWS, Microsoft Azure, Google Cloud Plt.
- Лабораторне заняття 2. Робота з хмарою Amazon Web Services(AWS).
- Лабораторне заняття 3. Розгортання та доступ до екземплярів EC2, S3, VM.
- Лабораторне заняття 4. Конфігурація клієнта SSH і доступ до VM.
- Лабораторне заняття 5. Установка та налаштування Airflow.
- Лабораторне заняття 6. Створення ETL pipeline в Airflow.
- Лабораторне заняття 7. Виконання простого завдання з MapReduce.
- Лабораторне заняття 8. Установка та налаштування автономного кластера Hadoop.
- Лабораторне заняття 9. Робота з кластером Hadoop.
- Лабораторне заняття 10. Ознайомлення з інтерфейсом Hue, завантаження даних і файлів.
- Лабораторне заняття 11. Створення ETL pipeline в Hue.
- Лабораторне заняття 12. Робота з SQL, базові команди.
- Лабораторне заняття 13. Ознайомлення з хмарними сервісами AWS RDS для створення власної реляційної бази даних.
- Лабораторне заняття 14. Робота з AWS RDS Maria DB.
- Лабораторне заняття 15. Робота з AWS RDS PostgreSQL.
- Лабораторне заняття 16. Робота з навчальним онлайн кластером Databrick.
- Лабораторне заняття 17. Робота зі Spark SQL та DataFrame API .
- Лабораторне заняття 18. Ознайомлення з Amazon OpenSearch Service.
- Лабораторне заняття 19. Робота з повнотекстовим пошуком в OpenSearch.
- Лабораторне заняття 20. Робота з векторним пошуком в OpenSearch.
- Лабораторне заняття 21. Управління проектами щодо Великих Даних.
- Лабораторне заняття 22. Управління життєвим циклом даних та DataOps.
- Лабораторне заняття 23. Визначення корпоративної інфраструктури Великих Даних та сервісів.
- Лабораторне заняття 24. Вибір інфраструктурних послуг та компонентів.

Самостійна робота

- Віртуальний гібридний/ динамічний хмарний центр обробки даних
- Хмарний аутсорсинг ІТ-інфраструктури підприємства.
- Інфраструктура Великих Даних та її компоненти.

Великі Дані в промисловості та концепція Industry 4.0.
Знайомство з хмарними платформами: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP).
Віртуальний гібридний/ динамічний хмарний центр обробки даних
Хмарний аутсорсинг ІТ-інфраструктури підприємства.
Інфраструктура Великих Даних та її компоненти.
Великі Дані в промисловості та концепція Industry 4.0.
Знайомство з хмарними платформами: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP).

Література та навчальні матеріали

Основна література

1. Л. М. Олещенко. Технології оброблення великих даних. Конспект лекцій. Київ КПІ ім. Ігоря Сікорського 2021. – 227 с.
https://ela.kpi.ua/bitstream/123456789/42206/1/%D0%9AonspLekts_Tekhnolohii-obroblennia-velykykh-danykh_%D0%9Eleshchenko.pdf
2. Balusamy, Balamurugan, Seifedine Kadry, and Amir H. Gandomi. Big Data: Concepts, Technology, and Architecture. John Wiley & Sons, 2021.
https://media.wiley.com/product_data/excerpt/21/11197018/1119701821-11.pdf
3. Rajni Bhalla. Introduction to Big Data. Lovely Professional University. 2020.
https://www.lpude.in/SLMs/Master%20of%20Computer%20Applications/Sem_2/DECAP456_INTRODUCTION_TO_BIG_DATA.pdf
4. Introduction to Big Data Analytics. Publisher: The Palm Publications. 2022. – [Електронні ресурси].
https://www.researchgate.net/publication/360410918_INTRODUCTION_TO_BIG_DATA_ANALYTICS

Інтернет-ресурси

5. Apache Hadoop. Режим доступу: <https://hadoop.apache.org>
6. MapReduce Tutorial. Режим доступу: https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
7. Apache Spark. Режим доступу: <https://spark.apache.org/>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Бали студента з дисципліни нараховуються за наступним співвідношенням: контрольні роботи: 20% семестрової оцінки; самостійна робота: 15% семестрової оцінки; колоквіум: 25% семестрової оцінки; іспит: 40% семестрової оцінки.

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 АХІЗЕР

Дата погодження, підпис
31.08.2023 р.



Гарант ОП
Леонід ЛЮБЧИК