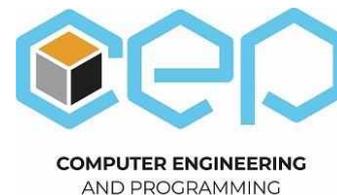




Силабус освітнього компонента Програма навчальної дисципліни



Паралельні та хмарні обчислювальні системи

Шифр та назва спеціальності
123 – Комп'ютерна інженерія

Інститут
ННІ комп'ютерних наук та інформаційних технологій

Освітня програма
Сучасне програмування, мобільні пристрої та комп'ютерні ігри (інноваційний кампус/
Прикладна комп'ютерна інженерія)

Кафедра
Комп'ютерна інженерія та програмування
(326)

Рівень освіти
Бакалавр

Тип дисципліни
Вільного вибору студента профільної
підготовки

Семестр
7

Мова викладання
Українська

Викладачі, розробники



Любченко Наталія Юріївна

Nataliia.Liubchenko@khpri.edu.ua

Кандидат технічних наук, доцент, доцент кафедри комп'ютерної інженерії та програмування

Досвід роботи – 20 років. Автор понад 100 наукових та навчально-методичних праць. Провідний лектор з дисциплін: «Основи програмування», «Об'єктно-орієнтоване програмування», «Технології та засоби великих даних», «Паралельні та хмарні обчислювальні системи», «Основи розподілених та паралельних обчислень».

[Детальніше про викладача на сайті кафедри](#)

Загальна інформація

Анотація

Дисципліна забезпечує теоретичну і практичну підготовку в області паралельних та розподілених обчислень, оволодіння концепціями сучасного програмування в рамках парадигм паралельного та розподіленого програмування. Основу вивчення складають підходи до програмування в багатопотокових системах, розподілених системах, системах паралельних розрахунків, розглядаються проблеми сумісної роботи процесів паралельної програми та їх синхронізації.

Мета та цілі дисципліни

Вивчення теоретичних основ та практичних аспектів використання паралельних обчислювальних систем для вирішення складних прикладних задач з великим об'ємом обчислень.

Формат занять

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

Компетентності

ФК 7. Здатність використовувати та впроваджувати нові технології, включаючи технології розумних, мобільних, зелених і безпечних обчислень, брати участь в модернізації та реконструкції комп'ютерних систем та мереж, різноманітних вбудованих і розподілених додатків, зокрема, з метою підвищення їх ефективності.

ФК 8. Готовність брати участь у роботах з впровадження комп'ютерних систем та мереж, введення їх до експлуатації на об'єктах різного призначення.

Результати навчання

ПРН 6. Вміти застосовувати знання для ідентифікації, формулювання і розв'язування технічних задач спеціальності, використовуючи методи, що є найбільш придатними для досягнення поставлених цілей.

ПРН16. Вміти оцінювати отримані результати та аргументовано захищати прийняті рішення..

Обсяг дисципліни

Загальний обсяг дисципліни 120 год. (4 кредитів ECTS): лекції – 16 год., практичні заняття – 32 год., самостійна робота – 72 год.

Передумови вивчення дисципліни (пререквізити)

Для успішного проходження курсу необхідно мати знання та практичні навички з наступних дисциплін: «Об'єктно-орієнтоване програмування», «Системне програмування», «Веб-програмування», «Організація та проектування баз даних».

Особливості дисципліни, методи та технології навчання

На лекційних заняттях викладання матеріалу здійснюється в усній формі із записом основних положень лекції у конспект. Для демонстрації презентацій застосовується медіа проектор та комп'ютер. На практичних заняттях студенти виконують індивідуальні завдання на комп'ютерах у середовищі IntelliJ IDEA, Eclipse для роботи з проектами Scala, за допомогою інструментів Oracle JDK, Scala Build Tool (SBT), Scala IDE, Apache Spark, Apache Hadoop.

Програма навчальної дисципліни

Теми лекційних занять

Тема 1. Введення в розподілені та паралельні обчислення.

Поняття про паралельні та розподілені обчислення. Послідовні обчислення. Паралельні обчислення. Засоби для здійснення паралельних обчислень.

Тема 2. Введення в розподілені розрахунки. Модель, спеціалізації розподілених розрахунків.

Поняття розподілених обчислень та розподіленої системи. Цілі побудови розподілених систем. Вимоги до розподілених систем.

Тема 3. Розуміння великих даних.

Концепції та термінологія. Характеристики великих даних. Проблеми використання великих даних. Визначення еталонної архітектури. Обґрунтування використання архітектурних патернів. Огляд інструментів для роботи з великими даними та їх аналізу.

Тема 4. Концепції зберігання та обробки великих даних.

Кластери. Файлові системи та розподілені файлові системи. База даних NoSQL. Шардінг (Sharding). Реплікація. Поєднання шардінга і реплікації. Теорема CAP. Принцип проектування бази даних ACID. Принцип BASE. Паралельна обробка даних. Розподілена обробка даних. Обробка робочих завдань. Кластер. Обробка в пакетному режимі. Обробка в режимі реального часу.

Тема 5. Розподілена пакетна і потокова обробка неструктурованих і слабоструктурованих даних.

Фреймворк Apache Spark. Особливості, обмеження, архітектура і принцип роботи компонентів Apache Spark. Робота з даними в Spark – RDD, DataFrame, DataSet. Операції Spark RDD. Spark SQL DataFrame.

Тема 6. Розробка додатків обробки даних, що виконуються в розподіленому обчислювальному середовищі.

Apache Hadoop. Екосистема Hadoop. Архітектурна концепція Hadoop. Демони Hadoop. Методи розгортання Hadoop. Розподілена файлова система Hadoop. Архітектура HDFS. Реплікації даних. Файлові операції HDFS – операція запису, операція читання, операція видалення.

Тема 7. Управління ресурсами кластерів і планування завдань.

YARN – планувальник ресурсів Hadoop. Архітектура YARN. Ключові компоненти Hadoop YARN.

Тема 8. Модель програмування для паралельної і розподіленої обробки великих наборів даних.

Apache Hadoop MapReduce. Стадії MapReduce. Завдання Map и Reduce. Приклади MapReduce

Теми практичних занять

Практичне заняття №1.

Знайомство з основними сервісами Apache Hadoop. Установка програмного забезпечення. Середовище розробки INTELLIJ IDEA.

Практичне заняття №2.

Програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD).

Практичне заняття №3.

RDD. Агрегування в Spark.

Практичне заняття №4.

Створення та робота з DataFrames з використанням Spark.

Практичне заняття №5.

Робота з графами з використанням Spark GraphX.

Практичне заняття №6.

Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml.

Практичне заняття №7.

Реалізація алгоритмів машинного навчання: статистичні обчислення і перевірка гіпотез.

Практичне заняття №8.

Установка Hadoop.

Практичне заняття №9.

Робота з розподіленою файловою системою Hadoop (HDFS). Рішення задачі запису даних в HDFS. Особливості управління ресурсами в режимі Hadoop кластера з YARN.

Практичне заняття №10.

Тестування Hadoop и Spark.

Практичне заняття №11-12.

Мова програмування Scala. Ознайомлення з синтаксисом і основними можливостями мови Scala. Введення в аналіз даних за допомогою Scala і Spark.

Практичне заняття №13-14.

Модель розподілених обчислень MapReduce.

Практичне заняття №15.

Рішення задач MapReduce в Spark.

Практичне заняття №16.

Рішення задач MapReduce в Hadoop.

Теми лабораторних робіт

Лабораторні роботи в рамках дисципліни не передбачені.

Самостійна робота

Опрацювання лекційного матеріалу. Підготовка до практичних занять та модульних контролів. Виконання, оформлення звітних матеріалів. Самостійне вивчення тем та питань, які не викладаються на лекційних заняттях: Система Hive, Система Impala, Система Mahout, Файлова система MapR, Технологія паралельного програмування OpenMP.

Література та навчальні матеріали

ОСНОВНА ЛІТЕРАТУРА

1. Минайленко Р.М. Паралельні та розподілені обчислення : навч. посіб. — Кропивницький : Видавець Лисенко В. Ф., 2021. — 153 с.

<https://dspace.kntu.kr.ua/server/api/core/bitstreams/396e02d2-725b-47b5-a1c0-ae07a9bec326/content>

2. Олещенко Л.М. Технології оброблення великих даних: навч. посіб. / Л.М. Олещенко; КПІ ім. Ігоря Сікорського. –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](https://ela.kpi.ua/bitstream/123456789/42206/1/%D0%9AonspLekts%20Tekhnolohii-obroblennia-velykykh-danykh%20%D0%9Eleshchenko.pdf)

3. Jules S. Damji. Learning Spark. Lightning-Fast Data Analytics. Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee / Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. - 2020. - 399 p.

<https://pages.databricks.com/rs/094-YMS-629/images/LearningSpark2.0.pdf>

4. Spark by Robert D. Schneider and Jeff Karmiol / Published by John Wiley & Sons, Inc. 111 River St. Hoboken, NJ 07030-5774. – 2019. – 80 p. <https://www.ibm.com/downloads/cas/WEB4XBOR>

5. Таран В. І. Технології Big Data: навч. посіб. / В. І. Таран, Ю. Г. Гордієнко, С. Г. Стіренко. – Київ КПІ ім. Ігоря Сікорського, 2022. – 56 с.

https://ela.kpi.ua/bitstream/123456789/50014/1/BigData_praktykum.pdf

6. Dean Wampler Programming Scala / Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472. – 2021. – 556 p. <https://books-library.net/files/books-library.net-10211445Tf3F5.pdf>

ДОДАТКОВА ЛІТЕРАТУРА

7. Любченко Н. Ю. Основи мови програмування SCALA [Електронний ресурс] : навч.-метод. посібник / Н. Ю. Любченко, А. О. Подорожняк, О. П. Черних ; Нац. техн. ун-т "Харків. політехн. ін-т". – Електрон. текст. дані. – Харків, 2023. – 148 с. <https://repository.kpi.kharkov.ua/items/3f5ee53b-258b-472a-ba35-3fdc412bfbed>

8. Проект Apache Hadoop [Електронний ресурс]. – Режим доступу: <https://hadoop.apache.org/>

9. Unified engine for large-scale data analytics [Електронний ресурс]. – Режим доступу: <https://spark.apache.org/>

10. Корочкін О.В. Паралельні та розподілені обчислення. Вибрані розділи: Навч. посібник. / О.В.Корочкін, Русанова О.В. – Київ : КПІ ім. Ігоря Сікорського, 2020. – 123 с

<https://ela.kpi.ua/bitstream/123456789/48224/1/Paralelni%20.pdf>

11. Big Data [Електронний ресурс]. Режим доступу: <https://www.it.ua/knowledge-base/technology-innovation/big-data-bolshie-dannye>

Система оцінювання

Критерії оцінювання успішності студента та розподіл балів

Оцінювання проводиться за 100-бальною шкалою. Бали нараховуються за наступним співвідношенням:
модульні контролі - 40 балів; практичні заняття - 30 балів; іспит - 30 балів.

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
22.04.2024



Завідувач кафедри
Олександр ЗАКОВОРОТНИЙ

Дата погодження, підпис
22.04.2024



Гарант ОП
Олександр ЗАКОВОРОТНИЙ