



Силабус курсу Обробка інформації та аналітика даних

Освітньо-професійна програма: «Комп'ютерні науки»
Ступінь вищої освіти - магістр
Спеціальність 122 «Комп'ютерні науки»

Рік навчання: I, Семестр: I

Кредитів: 5 Мова викладання: українська

Керівник курсу

ППП

д.т.н., професор Комар Мирослав Петрович

Контактна інформація

mko@wunu.edu.ua, +380985865009

Опис дисципліни

Дисципліна «Обробка інформації та аналітика даних» спрямована на ознайомлення студентів з фундаментальними знаннями з теорії та практики в області розробки і використання систем обробки і аналізу великих масивів даних. Дисципліна знайомить з базовими поняттями великих даних і висвітлює питання аналізу великих даних та пов'язаних з ними технічних, концептуальних та етичних проблем. Здобувачі вищої освіти набувають практичні навички з розробки та представлення концепції великих даних для конкретних предметних областей реального світу. Дисципліна містить практичні завдання для ознайомлення студентів з форматом великих даних, практичний досвід обробки та аналізу великих, складних структур даних.

Завдання дисципліни: сформулювати у слухачів сукупність теоретичних знань і практичних навичок щодо технологій обробки інформації та аналітики даних; навчити використовувати технології та інструменти для роботи з великими даними; навчити застосовувати сучасні методи та засоби аналітики даних; навчити реалізовувати програми для аналітики великих даних.

Структура курсу

Години (лек./пр.)	Тема	Результати навчання	Завдання
2/1	Тема 1. Вступ в аналіз і обробку великих даних.	Розуміти актуальність даних в сучасному світі. Розуміти поняття великих даних. Знати характеристики великих даних. Знати та вміти навести приклади застосування великих даних.	Питання, практична робота
2/1	Тема 2. Типи даних та їх обробка.	Розуміти поняття відкриті та приватні дані, структуровані та неструктуровані дані. Вміти застосовувати хмарні та туманні обчислення для обробки великих даних. Розуміти поняття статичні та динамічні дані. Розуміти та вміти застосовувати інфраструктуру великих даних. Розуміти поняття розподілених даних та їх обробки.	Питання, практична робота
2/1	Тема 3. Інструменти для аналізу даних.	Знати вимоги до інструментів аналізу даних. Вміти використовувати аналітику великих даних. Знати життєвий цикл аналізу даних. Знати формати та засоби обробки даних.	Питання, практична робота

4/1	Тема 4. Інженерія великих даних.	Знати архітектурні моделі інженерії великих даних. Вміти використовувати центри обробки даних та хмарні обчислення. Розуміти та вміти застосовувати технології віртуалізації, шари абстракції, гіпервізори. Розуміти та вміти застосовувати контейнерну технологію виконання програмного коду на сервері. Розуміти поняття та вміти застосовувати інжиніринг даних.	Питання, практична робота
4/2	Тема 5. Бази даних та сховища даних.	Розуміти особливості та вміти застосовувати бази даних NoSQL. Знати принципи організації та функціонування і вміти застосовувати MongoDB, PostgreSQL. Вміти проводити дослідження швидкодії роботи різних баз даних.	Питання, практична робота
4/2	Тема 6. Розподілена платформа Hadoop.	Знати та вміти використовувати масштабованість за допомогою великих даних. Вміти зберігати та обробляти дані в розподілених файлових системах. Розуміти та вміти застосовувати розподілені бази даних. Знати та вміти використовувати розподілену файлову систему Hadoop (HDFS) та MapReduce.	Питання, практична робота
2/2	Тема 7. Розподілена платформа Kafka.	Розуміти проблему прийому даних. Знати характеристики Kafka. Знати та вміти використовувати компоненти API Kafka. Вміти використовувати методи передачі повідомлень.	Питання, практична робота
2/1	Тема 8. Система управління розподіленими базами даних з відкритим кодом NoSQL Cassandra.	Знати особливості Apache Cassandra. Знати та вміти використовувати переваги Cassandra для обробки великих даних. Розуміти та вміти застосовувати Cassandra File System, модель даних, типи даних.	Питання, практична робота
4/1	Тема 9. Розподілена платформа Apache Spark.	Розуміти проблеми роботи з великими даними. Знати та вміти використовувати технологію Spark. Розуміти відмінності Spark та MapReduce. Знати та вміти використовувати засоби R для роботи з великими даними.	Питання, практична робота
2/1	Тема 10. Архітектури обробки великих даних Lambda та Карра.	Знати та вміти використовувати Lambda-архітектуру. Розуміти переваги і недоліки Lambda-архітектур. Знати та вміти використовувати Карра- архітектуру. Розуміти переваги і недоліки Карра-архітектури.	Питання, практична робота
2/1	Тема 11. Апаратно-програмні комплекси для обробки великих даних.	Знати та вміти використовувати комплекс інструментів Oracle Exalytics для аналітики даних. Знати та вміти використовувати апаратно-програмний комплекс SAP HANA. Вміти застосовувати апаратні рішення Netezza, Teradata та Exadata, апаратні рішення DAS.	Питання, практична робота

Літературні джерела

1. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. - Springer, 2020. – 298 p.
2. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. - Springer, 2019. – 105 p.
3. Bühlmann P., Drineas P., Kane M., van der Laan M. (eds.) Handbook of Big Data. -Taylor and Francis, 2016. – 456 p.
4. Oussous A., Benjelloun F.-Z., Lahcen A.A. and Belfkih S. Big Data technologies: A survey. // Journal of King Saud University. Computer and Information Sciences. - 2018. - Vol. 30, Issue 4. - P. 431–448.
5. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. – 110 p.
6. Raheem N. Big Data: A Tutorial-Based Approach. - Taylor & Francis Group LLC, CRC Press, 2019. – 203 p.
7. Ghavami Peter. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. - CreateSpace Independent Publishing Platform, 2016. – 204 p.
8. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>.
9. Комар М.П. «Методологічні основи інформаційної технології інтелектуального аналізу та обробки великих даних». Дисертація на здобуття наукового ступеня доктора технічних наук, 2021. – 363 с.
10. Wang C., Shakhovska N., Sachenko A., Komar M. A New Approach for Missing Data Imputation in Big Data Interface. Information Technology and Control. 2020. Vol. 49. No 4. Pp. 541-555.
11. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. / Л.М. Олещенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с.
12. Koval V., Komar M. Deep Learning For IoT. Internet of Things for Industry and Human Application. - Kharkiv, 2019. -P. 268-302.
13. Шаховська Н.Б. Організація просторів даних у складних інформаційних систем. Дисертація на здобуття наукового ступеня доктора технічних наук, 2012. - 414 с.
14. Болюбаш Ю. Я. Методи та засоби опрацювання інформаційних ресурсів великих даних в системах територіального управління. Дисертація на здобуття наукового ступеня кандидата технічних наук, 2017. - 178 с.
15. Костюченко А.О. Основи програмування мовою Python: навчальний посібник. - Ч.: ФОП Баликіна С.М., 2020. - 180 с.
16. Основи програмування : методичні вказівки до виконання комп'ютерних практикумів з дисципліни «Основи програмування». Основи програмування мовою Python / Уклад.: А.В. Яковенко. – К.: НТУУ «КПІ ім. І. Сікорського», 2018. – 113 с.

Політика оцінювання

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3	Екзамен
20%	20%	20%	40%
1. Поточне опитування (6 тем по 5 балів) – 30 балів 2. Практичні завдання – 70 балів	1. Поточне опитування (5 тем по 5 балів) – 25 балів 2. Практичні завдання – 35 балів 3. Ректорська контрольна робота - 40 балів	1. Написання та захист КПЗ – 80 балів 2. Виконання завдань під час тренінгу – 20 балів	1. Тестові завдання (25 тестів по 2 бали) – 50 балів 2. Завдання 1 – 25 балів 3. Завдання 2 – 25 балів

Шкала оцінювання:

За шкалою ЗУНУ	За національною шкалою	За шкалою ECTS
90-100	відмінно	A (відмінно)
85-89	добре	B (дуже добре)
75-84		C (добре)
65-74	задовільно	D (задовільно)
60-64		E (достатньо)
35-59	незадовільно	FX (незадовільно з можливістю повторного складання)
1-34		F (незадовільно з обов'язковим повторним курсом)