


СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Шифр та назва спеціальності:	Дана дисципліна передбачає базову підготовку в галузі інформаційних технологій і може бути рекомендована для студентів, які навчаються за галузями знань: 12 – Інформаційні технології, 11 – Математика та статистика
Назва освітньої програми:	Основи Data Engineering
Рівень освіти:	бакалаврський

ВИКЛАДАЧ

	Коротач Ігор
	Позиція в компанії Senior data engineer

ЗАГАЛЬНА ІНФОРМАЦІЯ ПРО ДИСЦИПЛІНУ

Анотація	Дисципліна «Основи Data Engineering» передбачає ознайомлення та вирішення основних задач сучасного data engineering. Формує загальне розуміння проектування, аналізу, обробки даних за допомогою таких засобів як високопродуктивних рушіїв обробки даних, сховищ даних, мікроконтролерів, хмарних сервісів. В курсі передбачено вивчення аналітичних інструментів, необхідних під час курсу за допомогою практичних робіт та оформлення звітів
Мета	Метою кредитного модуля «Основи Data Engineering» є формування у студентів здатностей: <ul style="list-style-type: none">- до розуміння основних принципів Data Engineering;- проектування систем повного циклу Data Engineering;- практичних навичок використання основних інструментів області;- створення звітів на основі здобутих під час практичних робіт навичок
Формат	Лекції, практичні заняття, індивідуальні завдання, екзамен

Обсяг	Загальний обсяг дисципліни 180 год.: лекції – 26 год., практичні заняття 26 год., самостійна робота 128 год. (рекомендований загальний обсяг = 180 год., - маються на увазі академічні години = 45 хв).
Пререквізити	Володіння основами баз даних, основами мови програмування Python, теоретичні навички роботи з Docker, навички роботи з віддаленими комп'ютерами на основі ОС Linux
Вимоги викладача	

СТРУКТУРА ДИСЦИПЛІНИ

Лекція	Тема лекції	Завдання на самостійну практичну роботу	Тема та завдання на самостійну роботу
Лекція 1	Введення у Data Engineering. Основні задачі області, ознайомлення з інструментарієм. Огляд роботи у контексті загального рішення (взаємодія з Data Science, DevOps, Big Data). Огляд найпопулярніших рішень.		
Лекція 2	Концепція хмарних рішень. Огляд основних Cloud Providers (AWS, GCP, Azure). Ознайомлення з Cloud Services, які можуть використовуватись (AWS IoT Core, AWS EKS, AWS EC2 та їх аналоги на інших хмарних платформах)	Завдання 1	Створити акаунт на одному з AWS, GCP, Azure. Ознайомитись з сервісами та інтерфейсом користувача. Створення та налаштування серверу за допомогою AWS EC2, або аналогу іншого Cloud Provider
Лекція 3	Введення до parallel computing. Введення до обробки даних за допомогою Spark	Завдання 2	Налаштування Spark. Обробка датасету за допомогою PySpark
Лекція 4	Ознайомлення з Task processing та Batch	Завдання 3	Оформлення звіту за темою “Task processing та

	processing. Вивчення концепту MapReduce, DAG у Big Data системах. Інструменти Airflow, Luigi		DAG”
Лекція 5	Введення до сховищ даних, що використовуються у Data Engineering. Data Lake та Data Warehouse. Приклади застосування у різних системах	Завдання 4	Створення та налаштування серверу СУБД (Postgresql) у контейнері Docker на сервері створеному під час лабораторної роботи 1.
Лекція 6	Введення до мікроконтролерів. Подібності та різниці у роботі порівняно до персональних комп’ютерів. Приклади мікроконтролерів, їх плюси та мінуси	Завдання 5	Налаштування Raspberry Pi 3, приклади програмування під Raspberry Pi 3
Лекція 7	Програмування для мікроконтролерів. Збір та обробка даних за мікроконтролерів	Завдання 6	Розробка застосування для збору інформації з датчику температури, вологості
Лекція 8	Введення до роботи з чергами повідомлень. Обробкою даних у реальному часі	Завдання 7	Звіт з порівнянням різних черг повідомлень
Лекція 9	Ознайомлення з протоколом MQTT, чергою повідомлень Kafka	Завдання 8	Налаштування MQTT брокеру на віддаленому сервері створеному під час лабораторної роботи 1.
		Завдання 9	Налаштування мікроконтролеру на роботу за допомогою MQTT брокеру створеному під час лабораторної роботи 7
Лекція 10	Проектування Data Engineering системи для подальшої праці у взаємодії з Data Science методами обробки та аналізу	Завдання 10	Створення схеми СУБД для збереження даних, отриманих з мікроконтролеру для їх подальшої обробки.
		Завдання 11	Створення Design Document для розробленого під час курсу рішення
Лекція 11	Методи представлення та візуалізації даних.	Завдання 12	Обробка даних збережених у роботі 10 за допомогою PySpark
Лекція 12	Огляд рішень для оформлення та	Завдання 13	Візуалізація даних оброблених у роботі 12 за

	візуалізації даних		допомогою Matplotlib
Лекція 13	Перспективи розвитку сфери Data Engineering, сучасні тенденції		
Додаткові питання до самостійної роботи	Самостійна робота передбачає: опрацювання лекційного матеріалу (30), підготовка до практичних занять (20), самостійне вивчення тем та питань, які не викладаються на лекційних заняттях (50).		
	1. Порівняння Cloud Provider'ів. Навести приклади сервісів для Data Engineering		
	2. Високорівнева архітектура Spark. Навести приклади моделей паралельного обчислення		
	3. Опишіть популярні алгоритми Big Data окрім MapReduce		
	4. Порівняння застосувань SQL та NoSQL СУБД для створення Data Lake та Data Warehouse		
	5. Застосування мікроконтролерів для вирішення завдання збору даних. Приклади мікроконтролерів.		
	6. Підходи до написання програм під мікроконтролери. Пам'ять, бистродія, модулі розширення		
	7. Черги повідомлень. Архітектура, відмінності		
	8. Протокол MQTT. Особливості вибору для роботи з малопотужними пристроями		
	9. Огляд архітектури застосування на базі AWS IoT Core, або аналогу іншого Cloud Provider		
	10. Обробка даних. Відказостійкі системи.		
	11. Design Document, основні положення		
	12. Фільтрація зіпсованих даних		
13. Інструменти візуалізації даних, приклади			
ЛІТЕРАТУРА ТА НАВЧАЛЬНІ МАТЕРІАЛИ			
Основна	1. Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems		
Додаткова	1. Data Engineering with Python: Work with Massive Datasets to Design Data Models and Automate Data Pipelines Using Python		
ПЕРЕЛІК ОБЛАДНАННЯ			
Лабораторні роботи мають проводитись у комп'ютерному залі. Припустимо застосування студентами власних ноутбуків. Необхідно мати мікроконтролери на базі Raspberry Pi 3 з датчиком ... (TBD, має мати не тільки C, а й Python бібліотеку для роботи)			

СИСТЕМА ОЦІНЮВАННЯ

Розподіл балів для оцінювання успішності студента	Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою
	90-100	A	відмінно
	82-89	B	добре
	74-81	C	
	64-73	D	задовільно
	60-63	E	
	35-59	FX	незадовільно з можливістю повторного складання
	0-34	F	незадовільно з обов'язковим повторним вивченням дисципліни

Нарахування балів

Бали нараховуються за наступним співвідношенням:

- Практичні роботи разом до 60 балів
- Теоретична підготовка, яка складається на екзамені, або може бути достроково захищена під час теоретичного не обов'язкового колоквиуму (КМ), які складаються з 3х питань по одному з кожного змістовного модуля.

Загальна оцінка за теоретичну підготовку 40 балів

Правила нарахування балів та підсумкової атестації:
 Рейтингове оцінювання формується з результатів накопичення балів отриманих студентом під час виконання практичної, теоретичної та самостійної роботи

$$П(60) + КМ (40) = Екз (100)$$

Виконання практичної роботи є допуском до екзамену.
 Студент має право відмовитись від оцінок отриманих за теоретичну підготовку та скласти екзамен.

НОРМИ АКАДЕМІЧНОЇ ЕТИКИ

При підготовці звітів ПР, або інших робіт, дослівне запозичення з тексту інших робіт, підручників, конспекту, методичних вказівок, інших джерел не допускається. Загальні положення, висновки осмислюються і викладаються в тексті реферату самостійно.