

**СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«ВЕЛИКІ ДАНІ (BIG DATA) ТА ХМАРНІ ТЕХНОЛОГІЇ
(CLOUD TECHNOLOGY)»**



Ступінь освіти	магістр
Спеціальність	121 Інженерія програмного забезпечення
Освітня програма	Інженерія програмного забезпечення
Тривалість викладання	осінній семестр (1, 2 чверті)
Кількість кредитів	4 кредити ЄКТС (120 годин)
Заняття:	
лекції:	2 год/тиждень
лабораторні заняття:	2 год/тиждень
Мова викладання	українська

Сторінка курсу в СДО НТУ «ДП»: <https://do.nmu.org.ua/course/view.php?id=4539>

Кафедра, що викладає

Програмного забезпечення комп'ютерних систем



Викладач:

Іванченко Олег Васильович
Професор, д.т.н.

Персональна сторінка

<https://pzks.nmu.org.ua/ua/teachers/ivanchenkoov.php>

E-mail:

ivanchenko.o.v@nmu.one

1. Анотація до курсу

Основними завданнями викладання навчальної дисципліни “Великі дані (Big Data) та хмарні технології (Cloud technology)” є: отримання знань щодо основних принципів побудови систем управління великими даними, котрі використовуються у різноманітних галузях виробництва, науки і техніки. Виявлення складу та вивчення можливостей програмного забезпечення сучасних систем на прикладі Google Big Data, які використовуються для збереження та обробки великих даних в інформаційних системах за допомогою хмарних технологій. У рамках курсу викладено матеріали щодо характеристик сучасних систем Big Data, розвитку систем обробки та збереження інформації. Розглянуто різні типи сучасних систем Big Data, а

також використання обробки хмарними технологіями з відкритим кодом та сервісами компанії Amazon Web Services. Значна увага приділена вивченню системи аналізу даних в Big Data. Висвітлені загальні питання аналізу даних у сучасних системах Big Data в промисловості, а також впровадження хмарних технологій в них.

2. Мета та завдання курсу

Мета дисципліни – формування стійких знань та засвоєння базових понять сучасної теорії організації баз великих даних (Big Data) та їх обробки за допомогою хмарних технологій. Вивчення загальних принципів та методів застосування сучасної теорії великих даних (Big Data) для проєктування та розробки програмного забезпечення, а також використання обробки великих даних за допомогою сучасних хмарних технологій.

Завданнями дисципліни є:

- освоєння студентами принципів аналізу даних у сучасних Big Data за допомогою хмарних технологій;
- вивчення структур організації сучасних Big Data;
- вивчення особливостей роботи з Big Data та їх обробки в хмарних сховищах.

3. Результати навчання

Дисциплінарні результати навчання:

- будувати і досліджувати моделі інформаційних процесів у Big Data із застосуванням хмарних сервісів;
- здійснювати попередню класифікацію Big Data;
- розробляти і оцінювати стратегії проєктування програмних засобів за допомогою хмарних технологій;
- аналізувати, оцінювати і застосовувати на системному рівні Big Data та хмарні технології;
- аналізувати та обирати оптимальні рішення щодо використання технологій хмарних обчислень у відповідності з вимогами готовності, гнучкості та масштабованості сервіс-орієнтованих ресурсів, застосувати їх у власних дослідженнях.

Дисциплінарні результати навчання сформовано на основі ПРН освітньо-професійної програми «Інженерія програмного забезпечення» другого (магістерського) рівня вищої освіти (PH03, PH04, PH06, PH07, PH18).

4. Структура курсу

Види та тематика навчальних занять	Внесок в загальну оцінку, %
ЛЕКЦІЇ	44
<p>Тема 1. Визначення Cloud Technology. Передумови і історія виникнення. Типи хмарних технологій – приватні, публічні і гібридні. Етапи розвитку, від IaaS до SaaS. Основні провайдери публічних хмарних сервісів. Основні властивості хмарних сервісів.</p>	

Види та тематика навчальних занять	Внесок в загальну оцінку, %
<p>Тема 2. Архітектурні моделі Big Data. Технології віртуалізації. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері SaaS, PaaS і IaaS.</p> <p>Архітектурні моделі інженерії Big Data. Центри обробки даних та хмарні обчислення. Технології віртуалізації. Шари абстракції. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері. Інжиніринг даних.</p>	
<p>Тема 3. Lambda та Карра архітектури оброблення великих даних.</p> <p>Lambda архітектура. Переваги і недоліки Lambda -архітектури. Карра - архітектура. Переваги і недоліки Карра-архітектури</p>	
<p>Тема 4. Основні інфраструктурні сервіси на прикладі AWS.</p> <p>VPC, EC2, ECS, Load Balancers, API Gateway, RDS, Dynamo DB, S3, Lambda functions, SNS, SQS, CloudFront, etc.</p>	
<p>Тема 5. Додаткові сервіси AWS для роботи з BigData та ML.</p> <p>Amazon EMR, Amazon Glue, Amazon StepFunctions, Amazon Textract, Amazon Lex, Amazon Polly, Amazon Rekognition.</p>	
<p>Тема 6. Стохастичні методи оцінювання готовності хмарних систем і технологій.</p> <p>Метод побудови логіко-ймовірнісних моделей готовності хмарних систем і технологій. Метод побудови дерева відмов хмарної системи. Метод побудови структурних схем надійності хмарних систем. Метод побудови структурних схем безпеки хмарних систем і технологій. Метод оцінювання готовності та часу простою хмарних систем.</p>	
<p>Тема 7. Марковські моделі готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів</p> <p>Характеристика марковського випадкового процесу. Дискретний однорідний марковський ланцюг. Неперервний марковський ланцюг. Застосування систем диференціальних рівнянь Колмогорова-Чепмена в завданнях з марковського моделювання. Розв'язування задач марковського моделювання з застосуванням перетворювання Лапласу. Розв'язування систем лінійних арифметичних рівнянь. Характеристика стаціонарних та нестаціонарних режимів застосування хмарних систем з сервіс-орієнтованим розподіленням ресурсів. Параметризація марковських моделей готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів. Цикломатична складність графів станів марковських моделей готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів. Прикладні аспекти застосування апарату марковського моделювання в задачах щодо оцінювання готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів.</p>	
<p>Тема 8. Напівмарковські моделі готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів.</p> <p>Характеристика напівмарковського випадкового процесу. Завдання щодо побудови напівмарковської моделі. Стаціонарне розв'язування завдань на основі застосування напівмарковських процесів. Розв'язування завдань з напівмарковського моделювання з використанням вкладених марковських ланцюгів. Розв'язування завдань щодо підтримання хмарних систем з сервіс-орієнтованим</p>	

Види та тематика навчальних занять	Внесок в загальну оцінку, %
розподіленням ресурсів в готовності до застосування на основі використання апарату напівмарковського моделювання. Параметризація напівмарковських моделей готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів. Прикладні аспекти застосування апарату напівмарковського моделювання в задачах щодо оцінювання готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів.	
<i>Тестова контрольна робота №1 (за темами 1-8).</i>	22
Тема 9. Вступ до аналізу даних за допомогою Spark. Застосування Apache Spark. Уніфікований стек DrivenData. Spark SQL. Spark Streaming. Відкриті дані. MLlib. Диспетчери кластерів. Дані в спокої та дані в русі. Мета використання та особливості Spark. Коротка історія розвитку Spark.	
Тема 10. Програмування операцій з RDD. Основи RDD. Створення RDD. Операції з RDD: Перетворення. Дії. Відстрочені обчислення. Передача функцій у Spark. Часто використовувані перетворення Збереження (кешування).	
Тема 11. Робота з парами ключ/значення. Створення наборів пар. Створення RDD. Перетворення наборів пар. Агрегування. Угруповання даних. З'єднання. Сортування. Управління розподілом даних.	
Тема 12. Завантаження та збереження даних. Формати файлів. Текстові файли. JSON. Файли у форматі CSV. Об'єктні файли. Формати Hadoop для введення та виведення. Стиснення файлів. Управління розподілом даних. Файлові системи. Структуровані дані та Spark SQL.	
Тема 13. Додаткові можливості Spark. Акумулятори. Широкомовні змінні. Робота з розділами окремо. Взаємодія із зовнішніми програмами. Числові операції над наборами RDD тестовими даними. Локальний запуск завдання. Тестування керуючої програми. Запуск в кластері. Упаковка завдання. Запуск завдання.	
Тема 14. Робота з кластером Архітектура середовища Spark. Час виконання. Розгортання програм за допомогою spark submit. Пакування програмного коду та залежностей. Планування додатків у додатках Spark. Диспетчери кластерів.	
<i>Тестова контрольна робота №2 (за темами 9-14)</i>	22
ЛАБОРАТОРНІ ЗАНЯТТЯ	56
Робота №1 Тема: <u>ОСНОВНІ ПОНЯТТЯ MAPREDUCE</u> Мета та основне завдання роботи: ознайомитися із парадигмою програмування MapReduce, отримати навички написання та запуску простої програми, яка слідує парадигмі MapReduce. Спробувати реалізувати етапи, що знаходяться між Map та Reduce (наприклад, Shuffle та Sorting) без використання фреймворку.	
<i>Звіт з виконання та захист лабораторної роботи №1</i>	8

Види та тематика навчальних занять	Внесок в загальну оцінку, %
<p align="center">Робота №2</p> <p align="center">Тема: <u>ОСНОВНІ ПОНЯТТЯ HADOOP</u></p> <p>Мета та основне завдання роботи: ознайомитися із програмним забезпеченням для розподілених обчислень Hadoop, що включає в себе розподілену файлову систему Hadoop – HDFS, менеджер ресурсів кластеру Hadoop – YARN та засоби Hadoop для моніторингу кластеру. Навчитися розробляти та запускати програму MapReduce на кластері Hadoop.</p>	
<p><i>Звіт з виконання та захист лабораторної роботи №2</i></p>	8
<p align="center">Робота №3</p> <p align="center">Тема: <u>ОСНОВНІ ПОНЯТТЯ APACHE PIG</u></p> <p>Мета та основне завдання роботи: ознайомитися із програмним забезпеченням високого рівня для виконання розподілених обчислень на кластері Hadoop. Ознайомитися із мовою програмування високого рівня Pig Latin. Навчитися розробляти та запускати програму/скрипт Pig Latin на кластері Hadoop.</p>	
<p><i>Звіт з виконання та захист лабораторної роботи №3</i></p>	8
<p align="center">Робота №4</p> <p align="center">Тема: <u>APACHE HIVE</u></p> <p>Мета та основне завдання роботи: ознайомитися із програмним забезпеченням високого рівня для виконання розподілених обчислень на кластері Hadoop. Ознайомитися із засобами Apache Hive та мовою HiveQL. Навчитися розробляти та виконувати запити HiveQL на кластері Hadoop.</p>	
<p><i>Звіт з виконання та захист лабораторної роботи №4</i></p>	8
<p align="center">Робота №5</p> <p align="center">Тема: <u>ДИСПЕТЧЕР APACHE OOZIE</u></p> <p>Мета та основне завдання роботи: ознайомитися із диспетчером робіт для кластеру Hadoop. Ознайомитися із принципами роботи диспетчера Apache Oozie. Навчитися створювати простий потік завдань для кластеру Hadoop, використовуючи диспетчер Apache Oozie.</p>	
<p><i>Звіт з виконання та захист лабораторної роботи №5</i></p>	8
<p align="center">Робота №6</p> <p align="center">Тема: <u>Програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD) з використанням Spark</u></p> <p>Мета та основне завдання роботи: вивчити яким чином створюються RDD набори даних; порядок виконання функцій filter, collect, first для RDD; яким чином реалізуються функції map та flatmap; яким чином реалізуються функції join та union; функції агрегування Spark; яким чином здійснюється робота з парами ключ / значення.</p>	
<p><i>Звіт з виконання та захист лабораторної роботи №6</i></p>	8
<p align="center">Робота №7</p> <p align="center">Тема: <u>Створення та робота з DataFrames з використанням Spark</u></p> <p>Мета та основне завдання роботи: вивчити яким чином створюються DataFrames як ключові абстракції Spark; порядок створення DataFrames з файлів JSON; яким чином створюються DataFrames з RDD і</p>	

Види та тематика навчальних занять	Внесок в загальну оцінку, %
реалізуються операції з ними; яким чином застосовуються DataFrame для аналізу даних; порядок створення реляційних баз даних з застосуванням наборів DataFrames; яким чином створюються кластери об'єднаної інформації DataFrames.	
<i>Звіт з виконання та захист лабораторної роботи №7</i>	8
Загальна кількість балів	100

5. Технічне обладнання та/або програмне забезпечення*

Активованій акаунт університетської пошти (student.i.p.@nmu.one) на Офіс365. Використовуються лабораторії кафедри програмного забезпечення комп'ютерних систем (комп'ютерне та мультимедійне обладнання). Дистанційна платформа Moodle, MS Office 365, Microsoft Teams, система керування і застосунки для великих даних (Big Data) Hadoop, Microsoft Azure, Python.

6. Система оцінювання та вимоги

6.1. Навчальні досягнення здобувачів вищої освіти за результатами вивчення курсу оцінюватимуться за шкалою, що наведена нижче:

Рейтингова шкала	Інституційна шкала
90 – 100	відмінно
74-89	добре
60-73	задовільно
0-59	незадовільно

Загальні критерії досягнення результатів навчання відповідають описам 7-го кваліфікаційного рівня НРК.

6.2. Здобувачі вищої освіти можуть отримати **підсумкову оцінку** з навчальної дисципліни **на підставі поточного оцінювання знань** за умови, якщо набрана кількість балів з поточного тестування та виконання і захисту лабораторних робіт складатиме не менше 60 балів.

Теоретична частина оцінюється за результатами здачі двох контрольних тестових робіт, кожна з яких містить тестові закриті запитання з однією вірною відповіддю. Загалом за дві контрольні тестові роботи отримується **максимум 44 бали**, тобто **44%** від оцінки за дисципліну.

Лабораторні роботи (сім робіт – у вигляді індивідуального завдання з кожної, розподіл % див. в таблиці розділу 4) виконуються у письмовому вигляді (звіт з кожної роботи оцінюється в межах 8 балів) та загалом враховуються, як **56% (максимум 56 балів)**. Лабораторні роботи захищаються у вигляді опитування (1 запитання -1 бал (максимум 4 бали) та 4 бали за правильність виконання індивідуального завдання). У сумі за практичну частину курсу при поточному оцінюванні отримується **максимум 56 балів**.

Отримані бали за теоретичну частину та лабораторні роботи додаються і є підсумковою оцінкою за вивчення навчальної дисципліни. Максимально за поточною успішністю здобувач вищої освіти може набрати 100 балів.

Максимальне оцінювання поточного контролю в балах:

Теоретична частина	Практична частина	Разом
44	56	100

6.3. Критерії оцінювання підсумкової роботи. У випадку якщо здобувач вищої освіти за поточною успішністю отримав менше 60 балів та/або прагне поліпшити оцінку проводиться **підсумкове оцінювання (іспит)** під час сесії..

Іспит проводиться у вигляді комплексної контрольної роботи, яка включає запитання з теоретичної та практичної частини курсу. Білет складається з **40 тестових завдань** з чотирма варіантами відповідей, одна правильна відповідь оцінюється в 1 бал (**разом 40 балів**) та **15 тестових завдань** з практичної частини, кожне з запитань оцінюється максимум у 4 бали (**разом 60 балів**), причому:

- 4 бали – відповідність еталону;
- 3 бали – відповідність еталону з незначними помилками;
- 2 бали – часткова відповідність еталону, питання повністю не розкрито;
- 1 бал – невідповідність еталону, але відповідність темі запитання;
- 0 балів – відповідь не наведена або не відноситься до теми запитання.

Отримані бали за відповіді на запитання з теоретичної та практичної частини курсу додаються і є підсумковою оцінкою за вивчення навчальної дисципліни. Максимально за підсумковою роботою здобувач вищої освіти може набрати 100 балів.

7. Політика курсу

7.1. Політика щодо академічної доброчесності. Академічна доброчесність здобувачів вищої освіти є важливою умовою для опанування результатами навчання за дисципліною і отримання задовільної оцінки з поточного та підсумкового контролів. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). Політика щодо академічної доброчесності регламентується положенням "Положення про систему запобігання та виявлення плагіату у Національному технічному університеті "Дніпровська політехніка" (<https://bit.ly/3ExtVKY>).

У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно. При цьому викладач залишає за собою право змінити тему завдання.

7.2. Комунікаційна політика. Здобувачі вищої освіти повинні мати активовану корпоративну університетську пошту.

Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту.

7.3. Політика щодо перескладання. Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання

підсумкового оцінювання відбувається із дозволу деканату за наявності поважних причин (наприклад, лікарняний).

7.4 Політика щодо оскарження оцінювання. Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку.

7.5. Відвідування занять. Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, участь в університетських заходах, академічна мобільність, які необхідно підтверджувати документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту.

7.6. Опитування. Наприкінці вивчення курсу та перед початком сесії здобувача вищої освіти буде запропоновано анонімно заповнити електронні анкети (Microsoft Forms Office 365), які буде розіслано на ваші університетські поштові скриньки. Заповнення анкет є важливою складовою вашої навчальної активності, що дозволить оцінити дієвість застосованих методів викладання та врахувати ваші пропозиції стосовно покращення змісту навчальної дисципліни.

8 Рекомендовані джерела інформації

1. Олещенко Л.М. Технології оброблення великих даних. Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с. [Електронний ресурс] (istu.edu.ua).

2. Тарн В.І., Гордієнко Ю. Г., Стіренко С.Г. / Технології BIG DATA Практикум // Київ: КПІ ім. Ігоря Сікорського, 2022. – 56 с.

3. Зінченко О.В., Іщеряков С.М., Прокопов С.В., Серих С.О., Василенко В.В. Хмарні технології. – Навчальний посібник. – К: ФОП Гуляєва В.М., 2020. –

4. Хмарні технології в освіті. Навчально-методичний посібник для студентів фізико-математичного факультету. – Житомир: вид-во ЖДУ, 2016. – 72 с.

5. Byte Size Infographic: Visualising data // Електронний ресурс. Режим доступу: <https://www.redcentricplc.com/resources/infographics/byte-size/>

6. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>

7. Kaggle// Електронний ресурс. Режим доступу: <https://www.kaggle.com/>

8. DrivenData // Електронний ресурс. Режим доступу: <https://www.drivendata.org/>

9. Big Data: the 3 VS explained // Електронний ресурс. Режим доступу: <https://bigdataldn.com/intelligence/big-data-the-3-vs-explained/>

10. Computing // Електронний ресурс. Режим доступу: <https://home.cern/science/computing>

Додаткова:

11. Python – Object Oriented // Електронний ресурс. Режим доступу: https://www.tutorialspoint.com/python/python_classes_objects.htm

12. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>

13. Jupyter Notebook // Електронний ресурс. Режим доступу: <https://jupyter.org/>

14. Іванченко О. В. Методологічні основи та інформаційна технологія забезпечення готовності хмарних систем критичних інфраструктур. –Кваліфікаційна наукова праця на правах рукопису. Електронний ресурс. Режим доступу: <http://dissertations.karazin.ua/tech/files/Ivan-01/dis-Ivan.pdf>