

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



Ступінь освіти	магістр
Спеціальність	122 Комп'ютерні науки
Освітня програма	Комп'ютерні науки
Загальний обсяг	4 кредити ЄКТС (120 годин)
Тривалість викладання	1-й семестр (1,2 чверть)
Заняття:	
лекції:	2 години/тиждень
лабораторні заняття:	2 години/тиждень(1чв) 1 година/тиждень(2чв)
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:



Мартиненко Андрій Анатолійович  
PhD, доцент каф. ПЗКС

**Персональна сторінка**

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

**E-mail:** martynenko.a.a@nmu.one

## 1. Анотація до курсу

Основними завданнями викладання навчальної дисципліни “Великі дані (Big Data) та хмарні технології (Cloud technology)” є: отримання знань та умінь з основних принципів побудови систем управління великими даними, котрі використовуються у різноманітних областях виробництва, науки і техніки а також в сучасних автоматизованих інформаційних системах. Виявлення складу та вивчення можливостей програмного забезпечення сучасних систем, які використовуються для збереження та обробки великих даних в інформаційних системах. В рамках курсу розглянуто різні типи сучасних систем Big Data. Значна увага приділена вивченню системи аналізу даних в Big Data. Висвітлені загальні питання аналізу даних в сучасних систем Big Data в різних галузях.

## 2. Мета та завдання курсу

**Мета дисципліни** формування компетентностей щодо роботи з великими даними (BigData) та хмарними технологіями.

### Завдання курсу:

Завданнями вивчення дисципліни є:

- опанування студентами принципів аналізу даних в сучасних BigData;
- вивчення структур організації сучасних BigData;
- вивчення особливостей роботи з BigData;
- вивчення структур організації та використання сучасних хмарних технологій.

## 3. Результати навчання:

Дисциплінарні результати навчання:

- Розробляти та використовувати математичні моделі та методи аналізу великих даних у спеціалізованих середовищах та технологіях.
- Розробляти та використовувати алгоритмічне та програмне забезпечення для аналізу великих даних у спеціалізованих середовищах та технологіях.
- Проєктувати та супроводжувати та використовувати бази даних та знань для обробки та аналізу великих даних.
- Виявляти та усувати проблемні ситуації в процесі експлуатації програмного забезпечення, формулювати завдання для його модифікації або реінжинірингу при обробці великих даних у спеціалізованих середовищах та технологіях.
- Аналізувати та обирати оптимальні рішення щодо використання технологій хмарних обчислень при обробці великих даних у спеціалізованих середовищах та технологіях, застосувати їх у власних дослідженнях у сфері комп'ютерних наук.

Дисциплінарні результати навчання сформовано на основі ПРН освітньо-професійної програми «Комп'ютерні науки» другого (магістерського) рівня вищої освіти (PH08, PH09, PH12, PH17, PH20).

## 4. Структура курсу.

Види та тематика навчальних занять	Внесок в загальну оцінку, %
<b>ЛЕКЦІЇ</b>	<b>40</b>
<b>Тема 1. Визначення Cloud Technology.</b> 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.	
<b>Тема 2. Основні інфраструктурні сервіси на прикладі AWS.</b> VPC, EC2, ECS, Load Balancers, API Gateway, RDS, Dynamo DB, S3, Lambda functions, SNS, SQS, CloudFront, etc.	
<b>Тема 3. Додаткові сервіси AWS для роботи з BigData та ML.</b> Amazon EMR, Amazon Glue, Amazon StepFunctions, Amazon Textract, Amazon Lex, Amazon Polly, Amazon Rekognition.	

<b>Види та тематика навчальних занять</b>	<b>Внесок в загальну оцінку, %</b>
<p><b>Тема 4. Стохастичні методи оцінювання готовності хмарних систем і технологій</b></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><b>Тема 5. Марковські моделі готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів</b></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><b>Тема 6. Напівмарковські моделі готовності хмарних систем з сервіс-орієнтованим розподіленням ресурсів</b></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 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.</p>	
<p><b>Тема 7. Вступ до аналізу даних за допомогою Spark</b></p> <p>Визначення Apache Spark. Уніфікований стек DrivenData. Spark SQL. Spark Streaming. Відкриті дані. MLlib. Диспетчери кластерів. Дані в спокої та дані в русі.</p>	
<p><i>Контрольна робота №1 (теми 1-7)</i></p>	<p>20</p>

<b>Види та тематика навчальних занять</b>	<b>Внесок в загальну оцінку, %</b>
<b>Тема 8. Програмування операцій з RDD</b> Основи RDD Створення RDD Операції з RDD: Перетворення Дії Відстрочені обчислення Передача функцій у Spark. Часто використовувані перетворення Збереження (кешування).	
<b>Тема 9. Робота з парами ключ/значення</b> Створення наборів пар. Створення RDD. Перетворення наборів пар. Агрегування. Угрупування даних. З'єднання, Сортування, Управління розподілом даних.	
<b>Тема 10. Знавантаження та збереження даних</b> 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.	
<b>Тема 11. Додаткові можливості Spark</b> 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.	
<b>Тема 12. Робота з кластером</b> Архітектура середовища Spark час виконання розгортання програм за допомогою spark submit. пакування програмного коду та залежностей. Планування додатків у додатках Spark. Диспетчери кластерів.	
<b>Тема 13 Архітектурні моделі Big Data. Технології віртуалізації. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері. SaaS, PaaS і IaaS</b> 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.	
<b>Тема 14 Lambda та Карра архітектури оброблення великих даних.</b> Lambda архітектура. Переваги і недоліки Lambda -архітектури. Карра - архітектура. Переваги і недоліки Карра-архітектури.	
<i>Контрольна робота №2 (теми 8-14)</i>	20
<b>ЛАБОРАТОРНІ ЗАНЯТТЯ</b>	<b>60</b>
1. Основні поняття та можливості MapReduce	6
2. Основні поняття та можливості Hadoop	6
3. Основні поняття та можливості Apache Pig	6
4. Основні поняття та можливості Apache Hive	6
5. Основні поняття та можливості диспетчеру Apache Oozie	6
6. Програмування операцій зі стійкими розподіленими наборами даних (Resilient Distributed Datasets, RDD) з використанням Spark	6
7. Створення та робота з DataFrames з використанням Spark	6
8. Робота з графами з використанням інструментарію Spark	6

Види та тематика навчальних занять	Внесок в загальну оцінку, %
GraphX	
9. Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: статистичні обчислення і перевірка гіпотез	6
10. Реалізація алгоритмів машинного навчання з використанням Spark MLlib і пакета spark.ml: вилучення (TF-IDF, Word2Vec, CountVectorizer і CountVectorizerModel) і перетворення (Tokenizer, StopWordsRemover, n-gram, Binarizer, StringIndexer, IndexToString)	6
<b>РАЗОМ</b>	<b>100</b>

### 5. Технічне обладнання та/або програмне забезпечення.

Використовуються лабораторії кафедри програмного забезпечення комп'ютерних систем (комп'ютерне та мультимедійне обладнання). Дистанційна платформа Moodle, MS Office 365, Microsoft Teams. Система BigData Hortonworks Sandbox HDP 3.0.

### 6. Система оцінювання та вимоги

**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

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 (максимальна кількість – 20 балів за кожною тестовою роботою). Загалом за дві контрольні тестові роботи отримується **максимум 40 балів**, тобто 40% від оцінки за дисципліну.

**Лабораторні роботи** (10 робіт – у вигляді індивідуального завдання з кожної, розподіл % див. в таблиці розділу 4) виконуються у письмовому вигляді (звіт з кожної роботи оцінюється в межах балів, представлених в таблиці розділу 4, загалом лабораторні враховуються як 60% (**максимум 60 балів**). При несвоєчасному здаванні роботи оцінка знижується вдвічі. 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.

**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**.

**Екзамен** проводиться у вигляді комплексної контрольної роботи, яка включає запитання з теоретичної та практичної частини курсу. Білет складається з **30 тестових завдань** з чотирма варіантами відповідей, одна правильна відповідь оцінюється в 2 бали (**разом 60 балів**) та **2 завдань** з практичної частини, кожне з запитань оцінюється максимум у 20 балів (**разом 40 балів**), причому:

- 20 балів – відповідність еталону;
- 15 балів – відповідність еталону з незначними помилками;
- 10 балів – часткова відповідність еталону, питання повністю не розкрито;
- 5 балів – невідповідність еталону, але відповідність темі запитання;
- 0 балів – відповідь не наведена або не відноситься до теми запитання.

Отримані бали за тестові завдання та завдання з практичної частини 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://cutt.ly/MCfh5ky>

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.

За об'єктивних причин (наприклад, міжнародна мобільність) навчання може відбуватись в он-лайн формі за погодженням з керівником курсу.

#### **7.6. Участь в анкетуванні**

Наприкінці вивчення курсу та перед початком сесії здобувачам вищої освіти буде запропоновано анонімно заповнити електронні анкети (Microsoft Forms Office 365), які буде розіслано на їхні університетські поштові скриньки. Заповнення анкет є важливою складовою їхньої навчальної активності, що дозволить оцінити дієвість застосованих методів викладання та врахувати надані 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] Олещенко Л. М. Технології оброблення великих даних: конспект лекцій | КЕС МНТУ (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., Сєрих С.О., Василенко В.В. Хмарні технології. – Навчальний посібник. – К: ФОП Гуляєва В.М., 2020.

4. Іванченко О. В. Методологічні основи та інформаційна технологія забезпечення готовності хмарних систем критичних інфраструктур. – Кваліфікаційна наукова праця на правах рукопису. 2021р. <http://dissertations.karazin.ua/tech/files/Ivan-01/dis-Ivan.pdf>

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>

11. Open Knowledge Foundation // Електронний ресурс. Режим доступу: <https://okfn.org>

12. Garminder // Електронний ресурс. Режим доступу: <https://www.garminder.org>

13. Портал відкритих даних України // Електронний ресурс. Режим доступу: <https://data.gov.ua>

#### **Додаткова:**

14. Python – Object Oriented // Електронний ресурс. Режим доступу: [https://www.tutorialspoint.com/python/python\\_classes\\_objects.htm](https://www.tutorialspoint.com/python/python_classes_objects.htm)

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

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