

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«Великі дані (Big Data) та хмарні технології (Cloud technology)»



Ступінь освіти	бакалавр
Спеціальність	122 Комп'ютерні науки
Освітня програма	Комп'ютерні науки
Загальний обсяг	5 кредитів ЄКТС (150 годин)
Тривалість викладання	8 семестр (15 чверть)
Заняття:	
лекції:	2 години
лабораторні заняття:	3 години
Мова викладання	українська

Сторінка курсу в СДО НТУ «ДП»: <https://do.nmu.org.ua/enrol/index.php?id=5121>.

Кафедра, що викладає Програмного забезпечення комп'ютерних систем



Викладач:

Іванченко Олег Васильович
доктор технічних наук, доцент, професор
кафедри програмного забезпечення комп'ютерних систем

Персональна сторінка

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

e-mail: ivanchenko.o.v@nmu.one



Викладач:

Кабак Леонід Віталійович
Доцент, к.т.н.

Персональна сторінка

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

E-mail: kabak.l.v@nmu.one

1. Анотація до курсу

Основними завданнями викладання навчальної дисципліни «Великі дані (Big Data) та хмарні технології (Cloud technology)» є: отримання знань по основним принципам побудови систем управління великими даними, котрі

використовуються у різноманітних областях виробництва, науки і техніки і також в сучасних автоматизованих інформаційних системах. Виявлення складу та вивчення можливостей програмного забезпечення сучасних систем на прикладі Google Big Data, яка використовується для збереження та обробки великих даних в інформаційних системах. Розглянуто різні типи сучасних систем Big Data. Значна увага приділена вивченню системи адміністрування Big Data.

2. Мета та завдання курсу

Мета дисципліни – формування стійких знань та засвоєння базових понять сучасної теорії організації баз великих даних (Big Data). Вивчення загальних принципів та методів застосування сучасної теорії великих даних (Big Data) для проектування та розробки автоматизованих інформаційно-пошукових систем.

Завдання курсу:

Завданнями вивчення дисципліни є:

- освоєння здобувачами принципів аналізу даних у сучасних BigData 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 сформовано на основі ПРН освітньо-професійної програми «Комп'ютерні науки» першого (бакалаврського) рівня вищої освіти (ПР10, ПР12):

- володіти навичками адміністрування клієнт-серверних систем;
- вміти здійснювати попередню класифікацію Big Data, створювати сховища та вітрини даних на хмарних сервісах;
- вміти будувати і досліджувати моделі інформаційних процесів у Big Data із застосуванням хмарних сервісів;
- володіти навичками застосовувати інформаційні технології обробки, зберігання та передачі даних, використання програмних інструментів підтримки багатовимірного та кластерного аналізу;
- 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.

4. Структура курсу

Види та тематика навчальних занять	Внесок в загальну оцінку, %
ЛЕКЦІЇ	40
<p>Тема 1. Визначення Cloud Technogy. Системи Big Date. Основні поняття та архітектура. Уведення у систему Hadoop Горизонтальна масштабованість. Відмовостійкість. Локальність даних. Технології і тенденції роботи з Big Data. Методи і техніка аналізу великих даних. Зберігання та аналіз даних. Порівняння з іншими системами. Розподілені обчислення. Apache Hadoop і екосистема Hadoop. Імена конфігурацій Hadoop. MapReduce API. Сумісність Hadoop з операційними системами.</p>	
<p>Тема 2. Адміністрування Hadoop. Дискові структури даних. Безпечний режим Журнали аудиту. Інструменти. Моніторинг. Ведення журналів. Метрики. Супровід. . Стандартні адміністративні процедури. Включення і виключення вузлів. Оновлення. Установка і запуск Pig. Установка Hive. Знайомство з HBase.</p>	
<p>Тема 3. Архітектура HDFS. Структура HDFS Основні концепції HDFS. Блоки. Вузли імен та вузли даних. NameNode, Secondary NameNode, Datanode. Інтерфейс командного рядка. Основні операції файлової системи. Файлові системи Hadoop. Інтерфейси. Інтерфейс Java. Читання даних Hadoop по URL-адресою. Читання даних з використанням Filesystem API. Запис даних. Отримання інформації від файлової системи Видалення даних. Потік даних. Читання файлу. Запис в файли. Модель цілісності. Переміщення даних: Flume і Sqoop.</p>	
<p>Тема 4. HBASE Особливості HBase. Історія. Коли потрібно використовувати Hbase. Реалізація. Регіони. Клієнти.</p>	
<p>Тема 5. Hive Оболонка Hive. Адміністрування Hive. Клієнти Hive. Метахосховище . Порівняння з традиційними базами даних. Двійкові формати зберігання: SequenceFile . файли даних Avro та RCFile. Імпортування даних. Зовнішні з'єднання. Підзапити. Уявлення. Уявлення Написання функції користувача.</p>	
<p>Тема 6. Введення / вивід в Hadoop Цілісність даних. Цілісність даних в HDFS.LocalFileSystem.ChecksumFileSystem. Стиснення. Кодеки. Стиснення і розбиття вхідних даних. Використання стиснення в MapReduce. Серіалізація. Інтерфейс Writable. Типи даних і схеми Avro. Файлові структури даних.</p>	
<p>Тема 7. YARN Основні визначення. Ключові компоненти. Протоколи взаємодії YARN. Обробка збоїв. Використання YARN.</p>	
<p><i>Тестова контрольна робота №1 (за темами 1-7).</i></p>	20
<p>Тема 8 ZooKeeper Архітектура. Особливості ZooKeeper Підтримувані операції. Zookeeper та Java</p>	
<p>Тема 9. Scoop Встановлення та запуск Sqoop .Конектори Sqoop. Імпортування даних. Текстові та двійкові формати. Експортування даних в кластері. Упаковка завдання. Запуск завдання.</p>	

Види та тематика навчальних занять	Внесок в загальну оцінку, %
Тема 10 Архітектурні моделі Big Data 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.	
Тема 11. MapReduce - модель програмування. Потік даних. Комбінують функції. Визначення комбінуює функції. Виконання розподіленого завдання MapReduce.	
Тема 12. Створення кластера Hadoop Устаткування кластера. Мережева топологія. Налагодження та установка кластера. Установка Java. Створення користувача Hadoop. Установка Hadoop. Тестування установки. Конфігурація SSH. Конфігурація Hadoop. Управління конфігурацією. Налаштування оточення. Важливі властивості демонів Hadoop. CP18-4Адреси та порти демонів Hadoop. Інші властивості HadoopСтворення облікових записів користувачів Конфігурація YARN. Важливі властивості демонів YARN. Адреси та порти демонів YARN Безпека. Kerberos і Hadoop. Маркери делегування. Інші поліпшення в області безпеки.Тестування кластера Hadoop.	
<i>Тестова контрольна робота №2 (за темами 8-12).</i>	20
ЛАБОРАТОРНІ ЗАНЯТТЯ	60
Лабораторна робота 1 Основи мови Scala	
<i>Звіт з роботи № 1 та захист лабораторної роботи.</i>	5
Лабораторна робота 2 Керуючі структури і функції мови Scala	
<i>Звіт з роботи № 2 та захист лабораторної роботи.</i>	5
Лабораторна робота 3 Робота з масивами	
<i>Звіт з роботи № 3 та захист лабораторної роботи.</i>	5
Лабораторна робота 4 Масиви що асоціюються та кортежі	
<i>Звіт з роботи № 4 та захист лабораторної роботи.</i>	5
Лабораторна робота 5 Класи	
<i>Звіт з роботи № 5 та захист лабораторної роботи.</i>	5
Лабораторна робота 6 Підзапити	
<i>Звіт з роботи № 6 та захист лабораторної роботи.</i>	5
Лабораторна робота 7 Об'єкти	
<i>Звіт з роботи № 7 та захист лабораторної роботи.</i>	5
Лабораторна робота 8 Пакети та імпортування	
<i>Звіт з роботи № 8 та захист лабораторної роботи.</i>	5
Лабораторна робота 9 Наслідування	
<i>Звіт з роботи № 9 та захист лабораторної роботи.</i>	5
Лабораторна робота 10 Файли та регулярні вирази	
<i>Звіт з роботи № 10 та захист лабораторної роботи.</i>	5

Види та тематика навчальних занять	Внесок в загальну оцінку, %
Лабораторна робота 11 Змінення таблиць та обмежень	
<i>Звіт з роботи № 11 та захист лабораторної роботи.</i>	5
Лабораторна робота 12 Створення трейтів	
<i>Звіт з роботи № 12 та захист лабораторної роботи.</i>	5
Разом	100

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, Oracle Virtual Box, Oracle XE, Microsoft Visual Studio Community.

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

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

Лабораторні роботи (12 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). У сумі за лабораторну частину курсу при поточному оцінюванні отримується **максимум 60 балів**.

Отримані бали за теоретичну частину та лабораторні роботи додаються і є поточною успішністю 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
40	60	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 (екзамен)** під час сесії.

Екзамен 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 2 завдань з практичної частини, кожне з запитань оцінюється максимум у 30 балів (разом 60 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/MCfh5kv>

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.

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. 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>
6. Byte Size Infographic: Visualising data // Електронний ресурс. Режим доступу: <https://www.redcentricplc.com/resources/infographics/byte-size/>
7. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>
8. Apache Hadoop // Електронний ресурс. Режим доступу: <http://hadoop.apache.org/>
9. HDFS // Електронний ресурс. Режим доступу: <https://www.ibm.com/analytics/hadoop/hdfs>
10. MapReduce Tutorial // Електронний ресурс. Режим доступу: https://hadoop.apache.org/docs/r1.2.1/mapred_tutorial.html
11. Big Data: the 3 VS explained // Електронний ресурс. Режим доступу: <https://bigdataidn.com/intelligence/big-data-the-3-vs-explained/>
12. Computing // Електронний ресурс. Режим доступу: <https://home.cern/science/computing>
13. Open Knowledge Foundation // Електронний ресурс. Режим доступу: <https://okfn.org>
14. Garminder // Електронний ресурс. Режим доступу: <https://www.garminder.org>
15. Портал відкритих даних України // Електронний ресурс. Режим доступу: <https://data.gov.ua>

Додаткова:

1. Cay S. Horstmann / Scala for the Impatient, 2nd Edition // University of Connecticut Електронний ресурс. Режим доступу: <https://www.pearson.com/en-us/subject-catalog/p/scala-for-the-impatient/P200000000333/9780134540566>
2. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>
3. Jupyter Notebook // Електронний ресурс. Режим доступу: <https://jupyter.org/>