



Технології BigData

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти

Другий (магістерський)

Галузь знань

12 Інформаційні технології

Спеціальність

123 Комп'ютерна інженерія

Освітня програма

Комп'ютерні системи та мережі

Статус дисципліни

Нормативна

Форма навчання

очна (денна)

Рік підготовки, семестр

1 курс, осінній семестр

Обсяг дисципліни

4 кредити

Семестровий контроль/
контрольні заходи

Залік

Розклад занять

*Лекцій 36 годи Лабораторні 18 годин Самостійна
робота 66 годин*

Мова викладання

Українська

Інформація про керівника курсу / викладачів

*Лектор: Таран Владислав Ігорович, taran@comsys.kpi.ua
Лабораторні: Таран Владислав Ігорович,
taran@comsys.kpi.ua*

Розміщення курсу

<https://cloud.comsys.kpi.ua/s/weFE5o9BwpjEs76>

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна "Технології BigData" спрямована на вивчення підходів, методів і механізмів функціонування та використання інфраструктури для розподілених обчислень на базі кластеру Hadoop та парадигми MapReduce. Необхідність в використанні нових підходів обумовлена тим, що сучасні підходи до вирішення складних завдань, які потребують обробки надзвичайно великого обсягу даних, потребують використання великої кількості обчислювальних ресурсів. Вивчення даної дисципліни майбутніми фахівцями дозволить їм набути важливих компетенцій в плані розвитку існуючих і використанню нових підходів для організації розподілених обчислень.

Метою вивчення дисципліни "Технології BigData" є підготовка фахівців, здатних розв'язувати комплексні проблеми у сфері обробки "великих даних" та використовувати сучасні засоби для організації обчислень в розподілених системах.

Предметом дисципліни є:

- підходи та методи обробки "великих даних" в розподілених системах;
- механізми організації обчислень в розподілених системах.

Згідно з вимогами ОНП здобувачі після засвоєння дисципліни "Технології BigData" мають продемонструвати такі програмні результати навчання:

- Здатність використовувати сучасні інфраструктури для розподілених обчислень;
- Здатність створювати та використовувати програмне забезпечення для розподілених обчислень;
- Прогнозувати вплив і ефект застосовуваних методів, технічних засобів і технологій BigData;
- Розв'язувати складні задачі і проблеми, що виникають у професійній діяльності;
- Самостійно обирати та безпечно застосувати відповідні технічні та програмні засоби для використання в комп'ютерній інженерії.

За результатами вивчення навчальної дисципліни "Технології BigData" мають бути отримані такі знання:

- Мати передові концептуальні та методологічні знання у сфері обробки та аналізу великих обсягів даних;
- Мати методологічні знання в плані застосування сучасних підходів та засобів для організації обчислень у розподілених обчислювальних системах.

Уміння, які мають бути отримані у рамках вивчення навчальної дисципліни "Технологія BigData":

- Вміти ефективно здійснювати пошук та критичний аналіз організації розподілених обчислювальних систем;
- Вміти розв'язувати задачі налаштування основних компонентів розподілених обчислювальних систем;
- Вміти розробляти програмне забезпечення для обробки великих даних у розподілених обчислювальних системах;
- Вміти застосовувати технології організації обчислень у розподілених системах;
- Вміти застосовувати технології організації складних послідовностей обчислень у розподілених системах.

Здобувачі наукового ступеня також мають бути здатні:

- Застосовувати прикладні бібліотеки та програмні системи, які використовуються при розробці основних компонентів розподілених обчислювальних систем.
- Володіти методами та технологіями програмування з використанням прикладних бібліотек та програмних систем, призначених для інтеграції основних компонентів розподілених обчислювальних систем.

Таке поєднання загальних та спеціальних компетентностей, теоретичних та практичних знань, умінь та здатностей сприяє підвищенню науково-практичного рівня здобувачів наукового ступеня магістра задля здійснення ними ефективних наукових досліджень.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Необхідні дисципліни: "Програмування", "Архітектура комп'ютерних систем", "Комп'ютерні системи", "Структури даних та алгоритми", "Дослідження і проектування комп'ютерних систем", "Алгоритми та методи обчислень", "Дискретна математика".

Дисципліна "Технології BigData" забезпечує наступні програмні компетентності і програмні результати навчання ОНП/ОПП:

ЗК2 Здатність до абстрактного мислення, аналізу і синтезу; ЗК3 Здатність проводити дослідження на відповідному рівні; ЗК7 Здатність приймати обґрунтовані рішення;

ФК1 Здатність до визначення технічних характеристик, конструктивних особливостей, застосування і експлуатації програмних, програмно-технічних засобів, комп'ютерних систем та мереж різного призначення;

ФК2 Здатність розробляти алгоритмічне та програмне забезпечення, компоненти комп'ютерних систем та мереж, Інтернет додатків, кіберфізичних систем з використанням сучасних методів і мов програмування, а також засобів і систем автоматизації проектування;

ФК7 Здатність досліджувати, розробляти та обирати технології створення великих і надвеликих систем;

ФК11 Здатність обирати ефективні методи розв'язування складних задач комп'ютерної інженерії, критично оцінювати отримані результати та аргументувати прийняті рішення;

ФК15 Здатність розробляти системи обробки великих обсягів даних;

ПРН1 Застосовувати загальні підходи пізнання, методи математики, природничих та інженерних наук до розв'язання складних задач комп'ютерної інженерії.

ПРН6 Аналізувати проблематику, ідентифікувати та формулювати конкретні проблеми, що потребують вирішення, обирати ефективні методи їх вирішення.

ПРН8 Застосовувати знання технічних характеристик, конструктивних особливостей, призначення і правил експлуатації програмно-технічних засобів комп'ютерних систем та мереж для вирішення складних задач комп'ютерної інженерії та дотичних проблем;

ПРН9 Розробляти програмне забезпечення для вбудованих і розподілених застосувань, мобільних і гібридних систем;

ПРН15 Здійснювати дослідження та проектування високопродуктивних комп'ютерних систем.

3. Зміст навчальної дисципліни

Перелік основних тем, що входять до програми вивчення дисципліни "Технології BigData":

Розділ 1. Вступ до Big Data
Тема 1.1. Основні поняття Big Data
Тема 1.2. Характеристика Big Data
Тема 1.3. Приклади

задач Big Data

Тема 1.4. Сучасні технології Big Data

Розділ 2. Парадигма MapReduce

Тема 2.1. MapReduce у розподілених обчисленнях

Тема 2.2. Аспекти роботи MapReduce

Тема 2.3. Основні етапи MapReduce

Тема 2.4. Приклади застосування

Розділ 3. Платформа для розподілених обчислень Hadoop

Тема 3.1. Характеристики кластеру Hadoop

Тема 3.2. Розподілена файлова система –

HDFS
Тема 3.3. Менеджер ресурсів кластеру
– YARN

Тема 3.4. Засоби моніторингу роботи кластеру Hadoop

Тема 3.5. Запуск обчислювальних задач на кластері Hadoop

Тема 3.6. Особливості інсталяція та налаштування кластеру Hadoop

Розділ 4. Застосунок Apache Pig

Тема 4.1. Основні поняття Apache

Pig
Тема 4.2. Порівняння з

MapReduce
Тема 4.3. Мова Pig

Latin

Тема 4.4. Написання скриптів для кластеру Hadoop

Розділ 5. Застосунок Apache Hive

Тема 5.1. Основні поняття Apache Hive

Тема 5.2. Порівняння між Apache Pig та

MapReduce
Тема 5.3. Спосіб організації даних у

Apache Hive
Тема 5.4. Мова запитів HiveQL

Тема 5.5. Написання запитів для кластеру Hadoop

Розділ 6. Диспетчер робіт Apache

Oozie
Тема 6.1. Основні поняття Apache

Oozie
Тема 6.2. Потоки робіт у Apache

Oozie

Тема 6.3. Особливості налаштування диспетчеру

Тема 6.4. Засоби моніторингу робіт

Тема 6.3. Створення та налаштування власного потоку робіт для
кластеру Hadoop

4. Навчальні матеріали та ресурси

Базова:

1. Alex Holmes. Hadoop in Practice. – Manning Publications, 2012;
2. Tom White. Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale. – O'Reilly Media, 2010;
3. Donald Miner, Adam Shook. MapReduce Design Patterns. – O'Reilly Media, 2012.
4. Технології Big Data. Лабораторний практикум. Навч. посібник для здобувачів ступеня магістр за спеціальністю 123 «Комп'ютерні системи та мережі» / Таран В. К.: КПІ ім. Ігоря Сікорського, 2022. – 26 с. [Електронний ресурс], <https://comsys.kpi.ua/metodichni-vkazannya-po-disciplinam>

Додаткова:

1. Apache Hadoop (<https://hadoop.apache.org/docs/r2.10.1/>)

2. Apache Pig (<http://pig.apache.org/docs/r0.17.0/>)
3. Apache Hive (<https://cwiki.apache.org/confluence/display/Hive/LanguageManual>)
4. Apache Oozie (<https://oozie.apache.org/docs/4.3.1/index.html>)

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Розділ 1. Вступ до Big Data Тема 1.1. Основні поняття Big Data Тема 1.2. Характеристика Big Data Тема 1.3. Приклади задач Big Data Тема 1.4. Сучасні технології Big Data	10	4		6
Розділ 2. Парадигма MapReduce Тема 2.1. MapReduce у розподілених обчисленнях Тема 2.2. Аспекти роботи MapReduce Тема 2.3. Основні етапи MapReduce Тема 2.4. Приклади застосування	24	8	4	12
Розділ 3. Платформа для розподілених обчислень Hadoop Тема 3.1. Характеристики кластеру Hadoop Тема 3.2. Розподілена файлова система – HDFS Тема 3.3. Менеджер ресурсів кластеру – YARN Тема 3.4. Засоби моніторингу роботи кластеру Hadoop Тема 3.5. Запуск обчислювальних задач на кластері Hadoop Тема 3.6. Особливості інсталяція та налаштування кластеру Hadoop	26	8	6	12
Розділ 4. Застосунок Apache Pig Тема 4.1. Основні поняття Apache Pig Тема 4.2. Порівняння з MapReduce Тема 4.3. Мова Pig Latin Тема 4.4. Написання скриптів для кластеру Hadoop	20	6	2	12
Розділ 5. Застосунок Apache Hive Тема 5.1. Основні поняття Apache Hive Тема 5.2. Порівняння між Apache Pig та MapReduce Тема 5.3. Спосіб організації даних у Apache Hive Тема 5.4. Мова запитів HiveQL Тема 5.5. Написання запитів для кластеру Hadoop	18	4	2	12

Розділ 6. Диспетчер робіт Apache Oozie Тема 6.1. Основні поняття Apache Oozie Тема 6.2. Потоки робіт у Apache Oozie Тема 6.3. Особливості налаштування диспетчеру Тема 6.4. Засоби моніторингу робіт Тема 6.3. Створення та налаштування власного потоку робіт для кластеру Hadoop	22	6	4	12
Всього в семестрі	120	36	18	66

Лабораторні заняття:

Метою проведення лабораторних занять є набуття студентами необхідних практичних навичок роботи із сучасними технологіями для організації розподілених обчислень.

- Лабораторна робота №1: Основні поняття MapReduce;
- Лабораторна робота №2: Основні поняття Hadoop;
- Лабораторна робота №3: Основні поняття Apache Pig;
- Лабораторна робота №4: Apache Hive;
- Лабораторна робота №5: Apache Oozie.

6. Самостійна робота студента

- підготовка до лекційних занять по вивченню попереднього лекційного матеріалу;
- підготовка до лабораторних робіт з вивченням теорії лабораторного заняття з усною відповіддю на наведені питання розділу;
- оформлення результатів лабораторної роботи у вигляді протоколу.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Під час занять з навчальної дисципліни студенти повинні дотримуватись певних дисциплінарних правил:

- забороняється запізнюватись на заняття;
 - не допускаються сторонні розмови або інший шум, що заважає проведенню занять;
 - не допускається користування мобільними телефонами та іншими технічними засобами без дозволу викладача.
- Лабораторні роботи здаються особисто з попередньою перевіркою теоретичних знань, які необхідні для виконання лабораторної роботи. Перевірка практичних результатів включає перевірку коду та виконання тестових завдань. Роботи, які здаються із порушенням зазначених термінів без поважних причин, оцінюються на нижчу оцінку.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Види контролю з навчальної дисципліни «Системи штучного інтелекту» включають:

Лабораторні роботи:

Заплановано самостійне виконання 5 лабораторних робіт. Теми лабораторних робіт узгоджені у часі та за змістом з темами лекцій. Виконання лабораторних робіт у повному обсязі дозволяє набути практичних навичок використання систем штучного інтелекту.

Поточний контроль:

Передбачено проведення опитування

Залік: проводиться у вигляді співбесіди зі студентом для об'єктивного визначення рівня знань, умінь та практичних навичок, отриманих за семестр

Семестровий рейтинг студента складається з балів, які він отримує за види робіт відповідно до таблиці 1.

Таблиця 1

Оцінювання окремих видів навчальної роботи студента (у балах)

Вид навчальної роботи	Всього за видом роботи
Виконання та захист лабораторної роботи № 1	12
Виконання та захист лабораторної роботи № 2	12
Виконання та захист лабораторної роботи № 3	12
Виконання та захист лабораторної роботи № 4	12
Виконання та захист лабораторної роботи № 5	12
Rп	60
Залік (Rз)	40
Усього за семестр (R = Rп+ Rз)	100

Індивідуальний поточний рейтинг студента (**Rп**) складається з балів, які він отримує за виконання лабораторних робіт. Протягом семестру студенти виконують 5 лабораторних робіт. Максимальна кількість балів за кожну лабораторну роботу – 12. Бали нараховуються за:

- Теоретична складова – 6 балів,
- Практична складова – 6 балів

Максимальний можливий бал за лабораторну роботу – 12 балів.

Максимальна кількість балів за всі лабораторні роботи $12 \times 5 = 60$ балів.

Розрахунок розміру шкали (R) рейтингу.

Сума вагових балів контрольних заходів протягом семестру становить:

$$R = Rп + Rз ,$$

де **Rп** – семестровий рейтинг студента (лабораторні роботи).

Rз – залік.

Розмір рейтингової шкали для навчальної дисципліни становить:

$$R = Rп + Rз = 60 + 40 = 100 \text{ балів.}$$

Необхідною умовою допуску студента до заліку є його індивідуальний семестровий рейтинг (**Rп**), не менший, ніж 59 балів, та відсутність заборгованості з лабораторних робіт. При невиконанні згаданих вимог студент до заліку не допускається.

Таблиця відповідності рейтингових балів оцінкам за університетською шкали

Табл.1 Відповідність рейтингових балів оцінкам за університетською шкалою

Кількість балів	Оцінка ECTS
95...100	Відмінно
85...94	Дуже добре
75...84	Добре
65...74	Задовільно

60...64	Достатньо
$R_c < 60$	Незадовільно
$R_c \leq 50$ або не виконані інші умови допуску до екзамену	Не допущено

9. Додаткова інформація з дисципліни (освітнього компонента)

теоретичні та практичні питання, що виносяться під час захисту лабораторних робіт та семестрового контролю, відповідають переліку основних тем, що входять до програми вивчення дисципліни "Технології BigData".

Робочу програму навчальної дисципліни (силабус) :

Склав асистент кафедри обчислювальної техніки, Таран В. І.

Ухвалено кафедрою обчислювальної техніки (протокол № 10 від 25.05.2022)

Погоджено методичною комісією факультету (протокол № 10 від 09.06.2022)