



Технології BigData

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>12 Інформаційні технології</i>
Спеціальність	<i>123 Комп'ютерна інженерія</i>
Освітня програма	<i>Комп'ютерні системи та мережі</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>очна (денна)</i>
Рік підготовки, семестр	<i>1 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредити</i>
Семестровий контроль/ контрольні заходи	<i>Залік</i>
Розклад занять	<i>Лекцій 36 годи Лабораторні 18 годин Самостійна робота 66 годин</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	<i>Лектор: Таран Владислав Ігорович, taran@comsys.kpi.ua Лабораторні: Таран Владислав Ігорович, taran@comsys.kpi.ua</i>
Розміщення курсу	<i>https://cloud.comsys.kpi.ua/s/weFE5o9BwpjEs76</i>

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:

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

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.

ПРН7 Вирішувати задачі аналізу та синтезу комп'ютерних систем та мереж.

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 Тему 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. Поток роботи у Apache
 Oozie
 Тема 6.3. Особливості налаштування диспетчера
 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. Потoki 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.

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)

PCO з дисципліни, семестровий контроль з якої передбачений у формі заліку, для очної форми навчання розробляється за типом PCO-1 і включає оцінювання заходів поточного контролю з дисципліни впродовж семестру.

Рейтингова оцінка здобувача складається з балів, отриманих здобувачем за результатами заходів поточного контролю.

Види поточного контролю з навчальної дисципліни "Технології BigData" включають:

Лабораторні роботи:

Заплановано 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.

Співбесіда (за бажанням):

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	10-20
Виконання та захист лабораторної роботи № 2	10-20
Виконання та захист лабораторної роботи № 3	10-20
Виконання та захист лабораторної роботи № 4	10-20
Виконання та захист лабораторної роботи № 5	10-20
Rп	50-100
Співбесіда (за бажанням студента для підвищення рейтингу)	Rд 20

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 – 20, мінімальна – 10. Бали нараховуються за:

- Теоретична складова – 10 (5) балів,
- Практична складова – 10 (5) балів

Максимальний можливий бал за лабораторну роботу – 20 балів.

Максимальна кількість балів за всі лабораторні роботи $20 \times 5 = 100$ балів.

Розрахунок розміру шкали (R) рейтингу.

Сума вагових балів контрольних заходів протягом семестру становить:

$$R = R_{п},$$

де $R_{п}$ – семестровий рейтинг студента (лабораторні роботи).

Розмір рейтингової шкали для навчальної дисципліни становить:

$$RD = R_{п} = 100 \text{ балів.}$$

Умови допуску до семестрової атестації з кредитного модуля :

- виконання усіх лабораторних робіт.

При невиконанні згаданих вимог студент до заліку не допускається.

Студенти, які виконали всі умови допуску до семестрової атестації з кредитного модуля та мають рейтингову оцінку (RD) не менше 60 балів отримують відповідну позитивну оцінку (залік) без додаткових випробувань.

Зі здобувачами, які виконали всі умови допуску до заліку та мають рейтингову оцінку менше 60 балів, а також з тими здобувачами, хто бажає підвищити свою рейтингову оцінку, на останньому за розкладом занятті з дисципліни в семестрі викладач проводить семестровий контроль у вигляді співбесіди ($R_{д}$). Максимальне значення $R_{д} = 20$.

В цьому випадку

$$RD = RD + R_{д}$$

Таблиця 1 відповідності рейтингових балів оцінкам за університетською шкали

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)