



# 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,<br>семестр                        | <i>1 курс, осінній семестр</i>  |
| Обсяг дисципліни                                  | <i>4 кредитів</i>   |
| Семестровий<br>контроль/<br>контрольні заходи     | <i>Залік</i>  |
| Розклад занять                                    | <i>Лекцій 18 (36 годин) Лабораторних 9 (18 годин)</i>   |
| Мова викладання                                   | <i>Українська</i>   |
| Інформація про<br>керівника курсу /<br>викладачів | Лектор: <i>ст. викладач, доктор філософії, Таран Владислав Ігорович, taran@comsys.kpi.ua</i><br>Лабораторні: <i>ст. викладач, доктор філософії, Таран Владислав Ігорович, taran@comsys.kpi.ua</i> |
| Розміщення курсу                                  | <a href="https://cloud.comsys.kpi.ua/s/weFE5o9BwpjEs76">https://cloud.comsys.kpi.ua/s/weFE5o9BwpjEs76</a>   |

### 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.

Дисципліна “Технології BigData” забезпечує наступні програмні компетентності і програмні результати освітньо-наукової програми (ОНП): ФК2, ФК11, ФК18, ПРН14, ПРН21.

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 “Технології 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".

## **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. Потіки 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.

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 |
| <b>Розділ 1. Вступ до Big Data</b><br>Тема 1.1. Основні поняття Big Data<br>Тема 1.2. Характеристика Big Data<br>Тема 1.3. Приклади задач Big Data<br>Тема 1.4. Сучасні технології Big Data | 10              | 4            |                  | 6   |
| <b>Розділ 2. Парадигма MapReduce</b><br>Тема 2.1. MapReduce у розподілених обчисленнях  | 24              | 8            | 4                | 12  |

|  |     |    |    |    |
|--|-----|----|----|----|
| Тема 2.2. Аспекти роботи MapReduce<br>Тема 2.3. Основні етапи MapReduce<br>Тема 2.4. Приклади застосування   |     |    |    |    |
| <b>Розділ 3. Платформа для розподілених обчислень Hadoop</b><br>Тема 3.1. Характеристики кластеру Hadoop<br>Тема 3.2. Розподілена файлова система – HDFS<br>Тема 3.3. Менеджер ресурсів кластеру – YARN<br>Тема 3.4. Засоби моніторингу роботи кластеру Hadoop<br>Тема 3.5. Запуск обчислювальних задач на кластері Hadoop<br>Тема 3.6. Особливості інсталяція та налаштування кластеру Hadoop | 26  | 8  | 6  | 12 |
| <b>Розділ 4. Застосунок Apache Pig</b><br>Тема 4.1. Основні поняття Apache Pig<br>Тема 4.2. Порівняння з MapReduce<br>Тема 4.3. Мова Pig Latin<br>Тема 4.4. Написання скриптів для кластеру Hadoop   | 20  | 6  | 2  | 12 |
| <b>Розділ 5. Застосунок Apache Hive</b><br>Тема 5.1. Основні поняття Apache Hive<br>Тема 5.2. Порівняння між Apache Pig та MapReduce<br>Тема 5.3. Спосіб організації даних у Apache Hive<br>Тема 5.4. Мова запитів HiveQL<br>Тема 5.5. Написання запитів для кластеру Hadoop   | 18  | 4  | 2  | 12 |
| <b>Розділ 6. Диспетчер робіт Apache Oozie</b><br>Тема 6.1. Основні поняття Apache Oozie<br>Тема 6.2. Потоки робіт у Apache Oozie<br>Тема 6.3. Особливості налаштування диспетчеру<br>Тема 6.4. Засоби моніторингу робіт<br>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)

Поточний контроль: [опитування за темою заняття](#)

Календарний контроль: проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Семестровий контроль: [залік](#)

Умови допуску до семестрового контролю: [зарахування усіх лабораторних робіт](#)

Оцінювання окремих видів виконаної студентом навчальної роботи – лабораторних робіт здійснюється в балах за 100 бальною шкалою. Оскільки кредитний модуль має семестровий контроль у вигляді заліку, рейтинг студента з кредитного модуля складається з середнього балу за всі виконані лабораторні роботи.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

| <i>Кількість балів</i>    | <i>Оцінка</i> |
|---------------------------|---------------|
| 100-95                    | Відмінно      |
| 94-85                     | Дуже добре    |
| 84-75                     | Добре         |
| 74-65                     | Задовільно    |
| 64-60                     | Достатньо     |
| Менше 60                  | Незадовільно  |
| Не виконані умови допуску | 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):**

**Складено** ст. викладач кафедри обчислювальної техніки, доктор філософії, Таран В. І.

**Ухвалено** кафедрою обчислювальної техніки (протокол № 10 від 25.05.2023)

**Погоджено** методичною комісією факультету інформатики та обчислювальної техніки (протокол № 11 від 30.06.2023)