

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ЗАХІДНОУКРАЇНСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

ЗАТВЕРДЖУЮ

**В.о. декана факультету комп'ютерних
інформаційних технологій**



Ігор ЯКИМЕНКО

2023 р.

ЗАТВЕРДЖУЮ

**В.о. проректора з науково-педагогічної
роботи**



Віктор ОСТРОВЕРХОВ

2023 р.

РОБОЧА ПРОГРАМА
з дисципліни

«Обробка інформації та аналітика даних»

Ступінь вищої освіти – магістр

галузь знань – 12 Інформаційні технології

спеціальність – 122 Комп'ютерні науки

Освітньо-професійна програма – «Управління проєктами»

Кафедра інформаційно-обчислювальних систем і управління

Форма навчання	Курс	Семестр	Лекції (год.)	Практичні (год.)	ІРС (год.)	Тренінг (год.)	СРС (год.)	Разом (год.)	Іспит (сем.)
Денна	1	1	30	15	5	4	96	150	1

31.08.2023
[Signature]

Тернопіль – ЗУНУ
2023

Робоча програма складена на основі освітньо-професійної програми «Управління проєктами» підготовки магістрів галузі знань 12 «Інформаційні технології» спеціальності 122 «Комп'ютерні науки», затвердженої Вченою радою ЗУНУ (протокол № 10 від 23.06.2023 р.).

Робочу програму склал професор кафедри інформаційно-обчислювальних систем і управління, д.т.н., професор Мирослав КОМАР

Робоча програма затверджена на засіданні кафедри інформаційно-обчислювальних систем і управління, протокол № 1 від 29 серпня 2023 р.

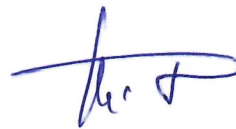
Завідувач кафедри
д.т.н., професор



Мирослав КОМАР

Розглянуто та схвалено групою забезпечення спеціальності „Комп'ютерні науки”, протокол № 1 від 29 серпня 2023 р.

Голова групи
забезпечення спеціальності,
д.т.н., професор



Мирослав КОМАР

Гарант освітньо-професійної
програми «Управління проєктами»



Михайло ДОМБРОВСЬКИЙ

**СТРУКТУРА РОБОЧОЇ ПРОГРАМИ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«ОБРОБКА ІНФОРМАЦІЇ ТА АНАЛІТИКА ДАНИХ»**

**1. ОПИС ДИСЦИПЛІНИ
«ОБРОБКА ІНФОРМАЦІЇ ТА АНАЛІТИКА ДАНИХ»**

Дисципліна «Обробка інформації та аналітика даних»	Галузь знань, спеціальність, СВО	Характеристика навчальної дисципліни
Кількість кредитів ECTS – 5	Галузь знань – 12 «Інформаційні технології»	Статус дисципліни – обов'язкова Мова навчання – українська
Кількість залікових модулів – 4	Спеціальність – 122 «Комп'ютерні науки»	Рік підготовки: 1 Семестр: 1
Кількість змістових модулів – 2	Освітньо-професійна програма: Управління проектами	Лекції: 30 год. Практичні заняття: 15 год.
Загальна кількість годин – 150	Ступінь вищої освіти – магістр	Самостійна робота: 100 год., у т.ч. тренінг – 4 год. Індивідуальна робота: 5 год.
Тижневих годин – 10, з них аудиторних – 3 год.		Вид підсумкового контролю – екзамен

2. МЕТА Й ЗАВДАННЯ ДИСЦИПЛІНИ «ОБРОБКА ІНФОРМАЦІЇ ТА АНАЛІТИКА ДАНИХ»

2.1. Мета дисципліни

Мета дисципліни «Обробка інформації та аналітика даних» – сформувати у студентів фундаментальні знання з теорії та практики в області розробки і використання систем обробки і аналізу великих масивів даних. Курс знайомить з базовими поняттями великих даних і висвітлює питання аналізу великих даних та пов'язаних з ними технічних, концептуальних та етичних проблем. Здобувачі вищої освіти набувають практичні навички з розробки та представлення концепції великих даних для конкретних предметних областей реального світу. Курс містить практичні завдання для ознайомлення студентів з форматом великих даних, практичний досвід обробки та аналізу великих, складних структур даних.

2.2. Завдання дисципліни:

- сформувати у слухачів сукупність теоретичних знань і практичних навичок щодо технологій обробки інформації та аналітики даних;
- навчити використовувати технології та інструменти для роботи з великими даними;
- навчити застосовувати сучасні методи та засоби аналітики даних;
- навчити реалізовувати програми для аналітики великих даних.

2.3. Найменування та опис компетентностей, формування яких забезпечує вивчення дисципліни:

СК01. Усвідомлення теоретичних засад комп'ютерних наук.

СК02. Здатність формалізувати предметну область певного проєкту у вигляді відповідної інформаційної моделі.

СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.

СК04. Здатність збирати та аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень.

СК06. Здатність застосовувати існуючі та розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук.

СК09. Здатність розробляти та адмініструвати бази даних та знань.

2.4. Передумови для вивчення дисципліни

Засвоєння знань за програмою вступного фахового випробування по спеціальності.

2.5. Результати навчання:

РН6. Розробляти концептуальну модель інформаційної або комп'ютерної системи.

PH7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.

PH8. Розробляти математичні моделі та методи аналізу даних (включно з великим).

PH9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).

PH11. Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.

PH12. Проєктувати та супроводжувати бази даних та знань.

PH19. Аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій.

3. ПРОГРАМА ДИСЦИПЛІНИ «ОБРОБКА ІНФОРМАЦІЇ ТА АНАЛІТИКА ДАНИХ»

ЗМІСТОВИЙ МОДУЛЬ 1. Вступ в обробку інформації та аналітику даних.

Тема 1. Вступ в аналіз і обробку великих даних.

Актуальність та зростання даних в сучасному світі. Поняття та характеристики великих даних. Приклади великих даних.

Тема 2. Типи даних та їх обробка.

Відкриті та приватні дані. Структуровані та неструктуровані дані. Хмарні та туманні обчислення. Статичні та динамічні дані. Інфраструктура великих даних. Розподілені дані та їх обробка.

Тема 3. Інструменти для аналізу даних.

Вимоги до інструментів аналізу даних. Аналітика великих даних. Життєвий цикл аналізу даних. Формати та засоби обробки даних.

Тема 4. Інженерія великих даних.

Архітектурні моделі інженерії великих даних. Центри обробки даних та хмарні обчислення. Технології віртуалізації. Шари абстракції. Гіпервізори. Контейнерна технологія виконання програмного коду на сервері. Інжиніринг даних.

ЗМІСТОВИЙ МОДУЛЬ 2. Технології та інструменти роботи з великими даними.

Тема 5. Бази даних та сховища даних.

Особливості баз даних NoSQL. Механізми для доступу до інформації в NoSQL базах даних. Порівняння баз даних SQL та NoSQL. Організація та принципи функціонування MongoDB. Організація та принципи функціонування PostgreSQL. Дослідження швидкодії роботи різних баз даних.

Тема 6. Розподілена платформа Hadoop.

Масштабованість за допомогою великих даних. Зберігання та оброблення даних в розподілених файлових системах. Розподілені бази даних. Розподілена файлова система Hadoop (HDFS). MapReduce.

Тема 7. Розподілена платформа Kafka.

Проблема прийому даних. Характеристики Kafka. Компоненти API Kafka. Методи передачі повідомлень.

Тема 8. Система управління розподіленими базами даних з відкритим кодом NoSQL Cassandra.

Опис Apache Cassandra. Переваги Cassandra. Cassandra File System. Модель даних. Типи даних.

Тема 9. Розподілена платформа Apache Spark.

Проблеми роботи з великими даними. Технологія Spark. Відмінності Spark та MapReduce. Засоби R для роботи з великими даними.

Тема 10. Архітектури обробки великих даних Lambda та Карра.

Lambda-архітектура. Переваги і недоліки Lambda-архітектур. Карра-архітектура. Переваги і недоліки Карра-архітектури.

Тема 11. Апаратно-програмні комплекси для обробки великих даних.

Комплекс інструментів Oracle Exalytics. Апаратно-програмний комплекс SAP HANA. Апаратні рішення Netezza, Teradata та Exadata. Апаратні рішення DAS.

4. СТРУКТУРА ЗАЛІКОВОГО КРЕДИТУ ДИСЦИПЛІНИ «ОБРОБКА ІНФОРМАЦІЇ ТА АНАЛІТИКА ДАНИХ»

	Кількість годин				
	Лек.	Прак.	СРС	ІР	Контрольні заходи
ЗМІСТОВИЙ МОДУЛЬ 1. Вступ в обробку інформації та аналітику даних					
Тема 1. Вступ в аналіз і обробку великих даних.	2	1	8	1	Опитування під час заняття
Тема 2. Типи даних та їх обробка.	2	1	6		Опитування під час заняття
Тема 3. Інструменти для аналізу даних.	2	1	6		Опитування під час заняття
Тема 4. Інженерія великих даних.	4	1	10		Опитування під час заняття
ЗМІСТОВИЙ МОДУЛЬ 2. Технології та інструменти роботи з великими даними					
Тема 4. Бази даних та сховища даних.	4	2	10	1	Опитування під час заняття
Тема 6. Розподілена платформа Hadoop.	4	2	12	1	Опитування під час заняття
Тема 7. Розподілена платформа Kafka.	2	2	10	1	Опитування під час заняття
Тема 8. Система управління розподіленими базами даних з відкритим кодом NoSQL Cassandra.	2	2	8	1	Опитування під час заняття
Тема 9. Розподілена платформа Apache Spark.	4	1	8		Опитування під час заняття
Тема 10. Архітектури обробки великих даних Lambda та Карра.	2	1	8		Опитування під час заняття
Тема 11. Апаратно-програмні комплекси для обробки великих даних.	2	1	10		Опитування під час заняття
Тренінг	-	-	4	-	Оцінювання виконання завдань під час тренінгу
Всього	30	15	100	5	

5. ТЕМАТИКА ПРАКТИЧНИХ ЗАНЯТЬ

Практичне заняття №1.

Тема: Методи аналізу та обробки даних.

Мета: Освоїти методи аналізу даних та визначити переваги їх використання.

Питання для обговорення:

1. Розподілене зберігання даних.
2. Розподілена обробка, з урахуванням локальності даних.
3. Збір, обробка та структурування даних.
4. Агрегація і аналіз даних.
5. Аналітика даних в реальному часі.

Література: 1-7, 9, 11, 12–14.

Практичне заняття №2.

Тема: Алгоритми кластеризації великих даних.

Мета: Освоїти алгоритми кластеризації великих даних та визначити особливості їх застосування.

Питання для обговорення:

1. Алгоритм k-середніх.
2. Алгоритм HCM (Hard C - Means).
3. Алгоритм нечіткої кластеризації Fuzzy C-means.
4. Алгоритм FRiS-Cluster.
5. Алгоритм FRiS-Stolp.

Література: 1-6.

Практичне заняття №3.

Тема: Розподілена платформа Apache Hadoop.

Мета: Освоїти розподілену платформу Apache Hadoop та визначити особливості її застосування.

Питання для обговорення:

1. Загальний огляд Apache Hadoop.
2. Hadoop Distributed File System (HDFS).
3. Hadoop YARN.

Література: 1–7.

Практичне заняття №4.

Тема: Розподілена платформа Apache Spark.

Мета: Освоїти розподілену платформу Apache Spark та визначити особливості її застосування.

Питання для обговорення:

1. Загальний огляд Apache Spark.
2. Spark Core.
3. Spark MLib.
4. Бібліотека GraphX.

Література: 1–7.

Практичне заняття №5.

Тема: Налаштування та запуск кластера Hadoop, YARN та Spark.

Мета: Навчитися налаштовувати та запускати віртуальний кластер із трьома вузлами Hadoop, YARN та Spark.

Питання для обговорення:

1. Створення віртуальні машини.
2. Встановлення та налаштування Hadoop.
3. Встановлення та налаштування YARN.
4. Встановлення та налаштування Spark.

Література: 1–7.

Практичне заняття №6.

Тема: Реалізація методів аналізу та обробки даних.

Мета: Навчитися створювати програмне забезпечення для вирішення задач аналізу та обробки даних.

Питання для обговорення:

1. Атрибути та методи класів Python.
2. Встановлення канонічної імплементації MCL.
3. Генерування даних мовою програмування Python.
4. Дослідження методу кластеризації даних.
5. Дослідження методів на основі глибоких неронних мереж.
6. Дослідження когнітивного аналізу даних.

Література: 15, 16.

6. КОМПЛЕКСНЕ ПРАКТИЧНЕ ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

Комплексні практичні індивідуальні завдання з дисципліни «Обробка інформації та аналітика даних» виконуються самостійно кожним студентом і охоплюють усі основні теми дисципліни. КПЗ оформляється у відповідності з встановленими вимогами. Виконання КПЗ є одним із обов'язкових складових модулів залікового кредиту. КПЗ пов'язане з аналізом великих даних. Аналіз даних виконується у відповідності до наборів даних офіційного сайту Державної служби статистики України (ukrstat.gov.ua) (розділ «Статистична інформація»):

Варіанти КПЗ з дисципліни «Обробка інформації та аналітика даних»:

1. Будівництво – ukrstat.gov.ua/operativ/menu/menu_u/bud.htm.
2. Внутрішня торгівля –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/spr.htm.
3. Державні фінанси, податки та публічний сектор –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/ekon/publ_u.htm.
4. Діяльність підприємств –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/sze.htm.

5. Доходи та умови життя –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/virdg.htm.
6. Енергетика – http://www.ukrstat.gov.ua/operativ/menu/menu_u/energ.htm.
7. Зовнішньоекономічна діяльність –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/zed.htm.
8. Інформаційне суспільство –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/zv.htm.
9. Капітальні інвестиції –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/ioz.htm.
10. Комплексна статистика –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/mp.htm.
11. Культура – http://www.ukrstat.gov.ua/operativ/menu/menu_u/cult.htm.
12. Макроекономічна статистика –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/tda.htm.
13. Навколишнє середовище –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/ns.htm.
14. Населені пункти та житло –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/if.htm.
15. Наука, технології та інновації –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/ni.htm.
16. Національні рахунки –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/nac_r.htm.
17. Освіта – http://www.ukrstat.gov.ua/operativ/menu/menu_u/osv.htm.
18. Основні засоби – http://www.ukrstat.gov.ua/operativ/menu/menu_u/voz.htm.
19. Охорона здоров'я –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/oz.htm.
20. Правосуддя та злочинність –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/ppr.htm.
21. Промисловість – http://www.ukrstat.gov.ua/operativ/menu/menu_u/prom.htm.
22. Реєстр статистичних одиниць –
<http://www.ukrstat.gov.ua/operativ/operativ2013/kap/kap.htm>.
23. Ринок праці –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/dem/r_pr.htm.
24. Сільське, лісове та рибне господарство –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/cg.htm.
25. Соціальний захист –
http://www.ukrstat.gov.ua/operativ/menu/menu_u/sz.htm.
26. Транспорт – http://www.ukrstat.gov.ua/operativ/menu/menu_u/tr.htm.
27. Туризм – http://www.ukrstat.gov.ua/operativ/menu/menu_u/tur.htm.
28. Ціни – http://www.ukrstat.gov.ua/operativ/menu/menu_u/cit.htm.

7. САМОСТІЙНА РОБОТА

№ з/п	Тематика	К-сть годин
1.	Основні характеристики великих даних	2
2.	Роль великих даних в різних галузях	4
3.	Консолідація даних	6
4.	Візуалізація даних, Graphi	6
5.	Основні конструкції мови R, консолідація даних, візуалізація	6
6.	HDFS – основи організації	4
7.	Архітектура Hadoop	4
8.	Виконання Map/Reduce	4
9.	Виконання програм в Hadoop	4
10.	Основи YARN	4
11.	Аналітика поточкових даних в платформі Storm	4
12.	Архітектура Apache Spark	4
13.	Організація даних в Apache Spark	4
14.	Обробка даних в GraphX	6
15.	Алгоритми класифікації	6
16.	Алгоритми кластеризації	6
17.	Методи машинного навчання для обробки великих даних	8
18.	Методи та засоби візуалізації великих даних	8
19.	Застосування технологій великих даних для задач управління в реальному часі.	6
	Разом	96

8. ТРЕНІНГ З ДИСЦИПЛІНИ

Порядок проведення:

1. Вступна частина: ознайомлення студентів з темою тренінгового заняття і видача завдання.
2. Практична частина: виконання завдань студентами згідно з індивідуальним завданням; оформлення короткого звіту.
3. Підведення підсумків: обговорення результатів виконаних завдань. Завдання. Провести аналіз та візуалізацію великих даних відповідно до вибраної предметної області.

9. ЗАСОБИ ОЦІНЮВАННЯ ТА МЕТОДИ ДЕМОНСТРУВАННЯ РЕЗУЛЬТАТІВ НАВЧАННЯ

У навчальному процесі застосовуються: лекції, в тому числі з використанням мультимедійного проектора та інших ТЗН, практичні заняття, консультації, індивідуальна та самостійна робота студента.

В процесі вивчення дисципліни «Обробка інформації та аналітика даних» використовуються наступні методи оцінювання навчальної роботи студента:

- поточне опитування;
- оцінювання практичних завдань;
- ректорська контрольна робота;
- оцінювання виконання КППЗ;
- оцінювання виконання завдань під час тренінгу;
- екзамен.

10. КРИТЕРІЙ, ФОРМИ ПОТОЧНОГО ТА ПІДСУМКОВОГО КОНТРОЛЮ

Підсумковий бал (за 100-бальною шкалою) з дисципліни «Обробка інформації та аналітика даних» визначається як середньозважена величина, в залежності від питомої ваги кожної складової залікового кредиту:

Заліковий модуль 1	Заліковий модуль 2	Заліковий модуль 3	Екзамен
20%	20%	20%	40%
1. Поточне опитування (6 тем по 5 балів) – 30 балів 2. Практичні завдання – 70 балів	1. Поточне опитування (5 тем по 5 балів) – 25 балів 2. Практичні завдання – 35 балів 3. Ректорська контрольна робота - 40 балів	1. Написання та захист КППЗ – 80 балів 2. Виконання завдань під час тренінгу – 20 балів	1. Тестові завдання (25 тестів по 2 бали) – 50 балів 2. Завдання 1 – 25 балів 3. Завдання 2 – 25 балів

Шкала оцінювання:

За шкалою Університету	За національною шкалою	За шкалою ECTS
90-100	відмінно	A (відмінно)
85-89	добре	B (дуже добре)
75-84		C (добре)
65-74	задовільно	D (задовільно)
60-64		E (достатньо)
35-59	незадовільно	FX (незадовільно з можливістю повторного
1-34		F (незадовільно з обов'язковим повторним

10. ІНСТРУМЕНТИ, ОБЛАДНАННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ, ВИКОРИСТАННЯ ЯКИХ ПЕРЕДБАЧАЄ НАВЧАЛЬНА ДИСЦИПЛІНА

№	Найменування	Номер теми
1.	Мультимедійне обладнання	1-11
2.	Комп'ютери з доступом до мережі Інтернет, обчислювальний кластер, суперкомп'ютер	1-11
3.	Системи розподілених обчислень, прикладні інструментальні засоби	6-11

РЕКОМЕНДОВАНІ ДЖЕРЕЛА ІНФОРМАЦІЇ

Основна література:

1. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. - Springer, 2020. – 298 p.
2. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. - Springer, 2019. – 105 p.
3. Bühlmann P., Drineas P., Kane M., van der Laan M. (eds.) Handbook of Big Data. -Taylor and Francis, 2016. – 456 p.
4. Oussous A., Benjelloun F.-Z., Lahcen A.A. and Belfkih S. Big Data technologies: A survey. // Journal of King Saud University. Computer and Information Sciences. - 2018. - Vol. 30, Issue 4. - P. 431–448.
5. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. – 110 p.
6. Raheem N. Big Data: A Tutorial-Based Approach. - Taylor & Francis Group LLC, CRC Press, 2019. – 203 p.
7. Ghavami Peter. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. - CreateSpace Independent Publishing Platform, 2016. – 204 p.

8. IoT Fundamentals: Big Data & Analytics // Електронний ресурс. Режим доступу: <https://www.netacad.com/courses/iot/big-data-analytics>.
9. Комар М.П. «Методологічні основи інформаційної технології інтелектуального аналізу та обробки великих даних». Дисертація на здобуття наукового ступеня доктора технічних наук, 2021. – 363 с.
10. Wang C., Shakhovska N., Sachenko A., Komar M. A New Approach for Missing Data Imputation in Big Data Interface. Information Technology and Control. 2020. Vol. 49. No 4. Pp. 541-555.
11. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем»)/ Л.М. Олещенко; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 5,55 Мбайт). – Київ: КПІ ім. Ігоря Сікорського, 2021. – 227 с.
12. Koval V., Komar M. Deep Learning For IoT. Internet of Things for Industry and Human Application. - Kharkiv, 2019. -P. 268-302.
13. Шаховська Н.Б. Організація просторів даних у складних інформаційних систем. Дисертація на здобуття наукового ступеня доктора технічних наук, 2012. - 414 с.
14. Болюбаш Ю. Я. Методи та засоби опрацювання інформаційних ресурсів великих даних в системах територіального управління. Дисертація на здобуття наукового ступеня кандидата технічних наук, 2017. - 178 с.
15. Костюченко А.О. Основи програмування мовою Python: навчальний посібник. - Ч.: ФОП Баликіна С.М., 2020. - 180 с.
16. Основи програмування : методичні вказівки до виконання комп'ютерних практикумів з дисципліни «Основи програмування». Основи програмування мовою Python / Уклад.: А.В. Яковенко. – К.: НТУУ «КПІ ім. І. Сікорського», 2018. – 113 с.

Додаткова

17. Tshepo Chris Nokeri. Data Science Solutions with Python. 1st Ed. – Apress, 2022. – 119 p.
18. Wiktorski T. Data-intensive Systems Principles and Fundamentals using Hadoop and Spark / Tomasz Wiktorski. – Springer Nature Switzerland AG. : Springer International Publishing, 2019. – 524 p.
19. О'Ніл К. BIG DATA. Зброя математичного знищення (MIM) / Кейт О'Ніл. – Київ: Форс, 2019. – 336 с.
20. Feeney K. et al. Engineering Agile Big-Data Systems. – River Publishers, 2018. – 436 p.
21. Bill Chambers and Matei Zaharia. Spark: The Definitive Guide Big Data Processing Made Simple. - O'Reilly Media, Inc., 2018. – 600 с.
22. Rajat Mehta. Big Data Analytics with Java. – Packt Publishing, 2017. – 418 p.

23. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. – 110 p.
24. Ghavami Peter. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. – CreateSpace Independent Publishing Platform, 2016. – 204 p.
25. Raheem N. Big Data: A Tutorial-Based Approach. – Taylor & Francis Group LLC, CRC Press, 2019. – 203 p.
26. A Guide to the Business Analysis Body of Knowledge (BABOK Guide). – 3rd Edition. – IIBA. – 2015. – 502 p.
27. Ferrari A. Introducing Microsoft Power BI / Alberto Ferrari and Marco Russo // Microsoft Press, 2016. – 189 p.
28. Collier Michael S. Microsoft Azure Essentials: Fundamentals of Azure, Second Edition / Michael S. Collier and Robin E. Shahan // Microsoft Press, 2016. – 246 p.
29. Barnes J. Microsoft Azure Essentials: Azure Machine Learning / Jeff Barnes // Microsoft Press, 2015. – 237 p.
30. Акіменко В. В. Прикладні задачі інтелектуального аналізу даних (Data Mining) / В. Акіменко. – Київ : КНУ ім. Тараса Шевченка, 2018 р. – 152 с.

Інформаційні ресурси

31. Система дистанційного навчання MOODLE [Електронний ресурс]. – Режим доступу: <http://www.moodle.tneu.edu.ua>.
32. What is Big Data [Електронний ресурс]. – Режим доступу: <http://datascience.berkeley.edu/what-is-big-data/>
33. Шаховська Н. Б., Болюбаш Ю. Я. Модель великих даних “сутність-характеристика” [Електронний ресурс]. – Режим доступу: http://ena.lp.edu.ua:8080/bitstream/ntb/29775/1/20_186-196.pdf
34. Shaw J. Why “Big Data” Is a Big Deal [Електронний ресурс]. – Режим доступу: <http://harvardmag.com/pdf/2014/03-pdfs/0314-HarvardMag.pdf>.
35. Schutt P. What is Big Data? [Електронний ресурс]. – Режим доступу: <https://blogs.oracle.com/bigdata/big-data-andanalytic-top-10-trends-for-2014/>
36. Відкритий посібник з відкритих даних [Електронний ресурс]. – Режим доступу: <https://socialdata.org.ua/manual/>
37. Big Data Та Блокчейн – Прорив В Області Аналізу Даних [Електронний ресурс]. – Режим доступу: <https://busines.in.ua/big-data-ta-blokchejn-proryv-v-oblasti-analizu-danyh/>
38. Weka Machine learning software to solve data mining problems [Електронний ресурс]. – Режим доступу: https://sourceforge.net/projects/weka/?source=typ_redirect
39. Books Ngram Viewer [Електронний ресурс]. – Режим доступу: <https://books.google.com/ngrams>
40. Мова програмування R [Електронний ресурс]. – Режим доступу: <https://cran.r-project.org>
41. Середовище для розробки програм на R – R Studio [Електронний ресурс]. – Режим доступу: <http://www.r-studio.com>

42. Manyika James and others. Big data: The next frontier for innovation, competition, and productivity [Электронный ресурс]. – Режим доступа: <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>
43. Guest post from Paul Burkhardt, a Research Developer at SRA International, Inc. A profile of Apache Hadoop MapReduce computing efficiency (part 2) by Jon Zuanich resource [Электронный ресурс]. – Режим доступа: <http://blog.cloudera.com/blog/2010/12/a-profile-of-hadoop-mapreduce-computing-efficiency-continued/>
44. IBM Analytics [Электронный ресурс]. – Режим доступа: <http://www.ibm.com/analytics/us/en/technology/hadoop/hadoop-trials.html>
45. IBM Cloud [Электронный ресурс]. – Режим доступа: https://www.ibm.com/cloud-computing/bluemix/?lnk=hp_trials_uauk
46. IBM Bluemix Promo Code - 6 Month Trial [Электронный ресурс]. – Режим доступа: <https://ibm.onthehub.com/WebStore/OfferingDetails.aspx?o=bb3528b7-2b63-e611-9420-b8ca3a5db7a1>
47. Hadoop: Built for big data, insights, and innovation [Электронный ресурс]. – Режим доступа: <http://www.ibm.com/analytics/us/en/technology/hadoop/>
48. IBM BigInsights [Электронный ресурс]. – Режим доступа: <http://www.ibm.com/analytics/us/en/technology/biginsights/>
49. It's (not) elementary: How Watson works [Электронный ресурс]. – Режим доступа: <https://www.pcworld.com/article/3128401/its-not-elementary-howwatson-works.html>.
50. A Computer Called Watson [Электронный ресурс]. – Режим доступа: <https://www.ibm.com/ibm/history/ibm100/us/en/icons/watson/>
51. IBM Watson supercomputer [Электронный ресурс]. – Режим доступа: <https://searchenterpriseai.techtarget.com/definition/IBM-Watson-supercomputer>.
52. Training helps users get more out of Watson Developer Cloud [Электронный ресурс]. – Режим доступа: <https://www.ibm.com/blogs/cloudcomputing/2017/06/06/training-helps-users-get-watson-developer-cloud>.