

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Кафедра (циклова комісія) _____ системного проектування _____

“ЗАТВЕРДЖУЮ”

Декан факультету _____

доц. Юрій ФУРГАЛА

“ ”

2022 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

ЗАСОБИ РОБОТИ З ВЕЛИКИМИ ДАНИМИ.

(шифр і назва навчальної дисципліни)

спеціальність _____ 121 – Інженерія програмного забезпечення _____

(шифр і назва спеціальності)

спеціалізація _____

(назва спеціалізації)

факультет _____ електроніки та комп'ютерних технологій _____

(назва інституту, факультету, відділення)

2022 – 2023 навчальний рік

Робоча програма “Засоби роботи з великими даними” для студентів
(назва навчальної дисципліни)
галузі знань “12 – Інформаційні технології”
за спеціальністю “121 Інженерія програмного забезпечення”

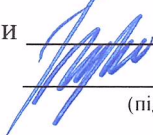
Розробники:

Ірина МИСЮК, асистент кафедри системного проектування)

(вказати авторів, їхні посади, наукові ступені та вчені звання)

Робочу програму схвалено на засіданні кафедри (циклової комісії) _____
системного проектування

Протокол від “ 30 ” 08 2022 року № 1

Завідувач кафедри _____ системного проектування
 (підпис) (Роман ШУВАР)
(прізвище та ініціали)

Ухвалено Вченою радою факультету електроніки та комп’ютерних технологій

Протокол від “ 31 ” 08 2022 року № 28/22

© Мисюк І., 2022 рік
© Шувар Р., 2022 рік
© ЛНУ ім. І. Франка, 2022 рік

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		денна форма навчання	заочна форма навчання
Кількість кредитів 4	Галузь знань <u>12 Інформаційні технології</u> (шифр і назва)	Вибірковий (за вибором)	
Модулів – <i>немає</i>	Спеціальність: <u>121 Інженерія програмного забезпечення</u>	Рік підготовки	
Змістових модулів – 2		1-й	
Індивідуальне науково-дослідне завдання <u>немає</u> (назва)		Семестр	
Загальна кількість годин – 120		2-й	
Тижневих годин для денної форми навчання: аудиторних – 4 самостійної роботи студента – 3,5	Освітній ступінь <u>бакалавр</u>	Лекції	
		32 год.	
		Практичні, семінарські	
		<i>немає</i>	
		Лабораторні	
		32 год.	
		Самостійна робота	
		56 год.	
		Індивідуальні завдання:	
		<i>немає</i>	
Вид контролю:			
<i>залік</i>			

Примітка.

Співвідношення кількості годин аудиторних занять до самостійної і індивідуальної роботи становить

для денної форми навчання – 1,14

для заочної форми навчання – немає

2. Мета та завдання навчальної дисципліни

Мета: одержання студентами знань про сучасні технології великих даних, які використовуються для їх обробки.

Цілі: забезпечити знайомство студентів з загальною теорією використання засобів та технологій для роботи з великими даними і сформувати навички їх практичного створення; навчити студента базових інструментів Apache Hadoop, Pig, Apache Spark, ELK, EFK, HDFS, RabbitMQ, Kafka, ZeroMQ.

У результаті вивчення навчальної дисципліни студент повинен **знати**: основні поняття, визначення і проблеми курсу; вимоги до постановки основних задач та інструментів роботи з великими даними; призначення й особливості застосування основних складових механізмів роботи з великими даними;

вміти: володіти базовими знаннями Apache Hadoop, Pig, Apache Spark, ELK, EFK, HDFS, RabbitMQ, Kafka, ZeroMQ.

Після вивчення даного курсу «Засоби роботи з великими даними» здобувачі набудуть таких Загальних та Фахових компетентностей та Програмних результатів навчання:

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

ЗК6. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ФК17. Здатність брати участь у проектуванні програмного забезпечення, включаючи проведення моделювання (формальний опис) його структури, поведінки та процесів функціонування.

ФК18. Здатність розробляти архітектури, модулі та компоненти програмних систем.

ФК22. Володіння знаннями про інформаційні моделі даних, здатність створювати програмне забезпечення для зберігання, видобування та опрацювання даних.

ФК28. Володіння методами сучасних веб-технологій, хмарних технологій, великих даних та засобами розробки веб-застосувань.

ПРН18. Знати та вміти застосовувати інформаційні технології обробки, зберігання та передачі даних.

ПРН26. Знати та вміти застосовувати засоби інженерії програмного забезпечення для реалізації проектів з використанням технологій науки про дані та штучного інтелекту.

3. Програма навчальної дисципліни

Тема 1. *Великі дані*. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Парадигма Map-Reduce.

Тема 2. *Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування*.

Тема 3. *Загальний розгляд технології Hadoop*. Файлова система HDFS. Його Архітектура.

Тема 4. *Розгляд системи Map reduce*. Розгляд платформи Apache Pig.

Apache Hive. Hive QL. Архітектра MRv1. Планувальник YARN.

Тема 5. *Розгляд бази даних Hbase*. Її модель даних. Життєвий цикл. Архітектура.

Тема 6. *MongoDB та Neo4j*. Архітектура баз даних та їхнє використання у Великих даних.

Тема 7. *Розгляд бази даних Cassandra*. Її принцип роботи. Використання Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.

Тема 8. *Apache Spark*. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.

Тема 9. *ELK стек*. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи.

Тема 10. *EFK стек*. Elasticsearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.

Тема 11. *RabbitMQ*. Опис. Взаємозв'язок. Використання.

Тема 12. *Apache Kafka і RabbitMQ*. Взаємодія сервісів через меседж брокери.

4. Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	денна форма						заочна форма					
	усьо го	у тому числі					усьог о	у тому числі				
		л	п	лаб	інд	с. р.		л	п	лаб	інд	с. р.
1	2	3	4	5	6	7	8	9	10	11	12	13
Тема 1. Великі дані. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Парадигма Map-Reduce.		2		2		3,5						
Тема 2. Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування.		2		2		3,5						
Тема 3. Загальний розгляд технології Hadoop. Файлова система HDFS. Його Архітектура.		2		2		3,5						
Тема 4. Розгляд системи Map reduce. Розгляд платформи Apache Pig.		2		2		3,5						
Тема 4. Apache Hive. Hive QL. Архітектура MRv1. Планувальник YARN.		2		2		3,5						
Тема 5. Розгляд бази даних Hbase. Її модель даних. Життєвий цикл. Архітектура.		2		2		3,5						
Тема 6. MongoDB та Neo4j. Архітектура баз даних та їхнє використання у Великих даних.		2		2		3,5						
Тема 7. Розгляд бази даних Cassandra. Її принцип роботи. Використання.		2		2		3,5						
Тема 8. Apache Spark. Розгляд.		2		2		3,5						

<i>Архітектура. Порівняння технологією Hadoop. Принцип роботи.</i>											
<i>Тема 8. Apache Spark. Розгляд. Архітектура. Порівняння технологією Hadoop. Принцип роботи.</i>		2	2		3,5						
<i>Тема 9. ELK стек. ElasticSearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи.</i>		2	2		3,5						
<i>Тема 10. EFK стек. ElasticSearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.</i>		2	2		3,5						
<i>Тема 11. RabbitMQ. Опис. Взаємозв'язок. Використання.</i>		2	2		3,5						
<i>Тема 12. Apache Kafka і RabbitMQ. Взаємодія сервісів через меседж брокери.</i>		2	2		3,5						
<i>Тема 12. Продовження Apache Kafka і RabbitMQ. Взаємодія сервісів через меседж брокери.</i>		2	2		3,5						
Усього годин		32	32		56						

5. Теми семінарських занять

6. Теми практичних занять

7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	<i>Лаб. 1. Ознайомлення, встановлення та розгляд технологій великих даних.</i>	2
2	<i>Лаб. 2. Розгляд технології Hadoop. Реалізація простої</i>	2

	<i>оброки даних.</i>	
3	<i>Лаб. 3. Використання Apache Pig в Apache Hadoop.</i>	2
4	<i>Лаб. 4. Використання Apache Hive в Apache Hadoop.</i>	2
5	<i>Лаб. 5. Розгляд бази даних Hbase. Її модель даних. Життєвий цикл. Архітектура.</i>	2
6	<i>Лаб. 6. Розробка проекту з вивченими технологіями..</i>	2
7	<i>Лаб. 7. Розгляд технології Apache Spark. Реалізація простої оброки даних.</i>	4
8	<i>Лаб. 8 Налаштування легування з використанням ELK</i>	4
9	<i>Лаб. 9. Розробка проекту з вивченими технологіями.</i>	2
10	<i>Лаб.10. Використання Apache Kafka для передавання даних між сервісами</i>	2
11	<i>Лаб. 11. Використання RabbitMQ для передавання даних між сервісами.</i>	2
12	<i>Лаб. 12. Розробка проекту з вивченими технологіями.</i>	2
13	<i>Підсумкове заняття</i>	4
	Разом	32

8. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	<i>Великі дані. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Парадигма Map-Reduce.</i>	3,5
2	<i>Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування.</i>	3,5
3	<i>Загальний розгляд технології Hadoop. Файлова система HDFS. Його Архітектура.</i>	3,5
4	<i>Розгляд системи Map reduce. Розгляд платформи Apache Pig.</i>	3,5
5	<i>Apache Hive. Hive QL. Архітектра MRv1. Планувальник YARN.</i>	3,5
6	<i>Розгляд бази даних Hbase. Її модель даних. Життєвий цикл. Архітектура.</i>	3,5
7	<i>MongoDB та Neo4j. Архітектура баз даних та їхнє використання у Великих даних.</i>	3,5
8	<i>Розгляд бази даних Cassandra. Її принцип роботи. Використання Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи</i>	7
9	<i>Apache Spark. Розгляд. Архітектура. Порівняння з технологією Hadoop. Принцип роботи.</i>	3,5
10	<i>ELK стек. Elasticsearch. Logstash. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи.</i>	3,5
11	<i>EFK стек. Elasticsearch. Fluentd. Kibana. Опис. Взаємозв'язок. Використання. Принцип роботи. Різниця між ELK і EFK.</i>	3,5
12	<i>RabbitMQ. Опис. Взаємозв'язок. Використання Apache Kafka і RabbitMQ. Взаємодія сервісів через меседж брокери</i>	3,5
13	<i>Тема 1. Великі дані. Загальне використання. Принцип роботи. Проблеми Великих даних. Спільні характеристики. Парадигма Map-Reduce.</i>	3,5
14	<i>Тема 2. Розгляд технологій Hadoop, Spark, Kafka. Variety, Velocity, Volume та їхнє застосування.</i>	7
	Разом	56

9. Індивідуальні завдання

10. Методи навчання

Інформаційні методи (лекція, бесіда, ілюстрація, демонстрація); дедуктивні методи на основі узагальнень; евристичні методи (проблемна лекція); інтерактивні методи (дискусія).

11. Методи контролю

Поточний контроль здійснюється шляхом проведення усного опитування та написання письмових звітів по виконаних лабораторних роботах. У кінці курсу проводиться екзамен.

12. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота																Підсумковий тест (екзамен)	Сума
Змістовий модуль 1								Змістовий модуль 2								50	100
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16		
2	2	3	3	3	4	4	4	4	3	3	3	3	3	3	3		

T1, T2 ... T12 – теми змістових модулів.

Шкала оцінювання: національна та ЄКТС

Оцінка ЄКТС	Сума балів за всі види навчальної діяльності	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
A	90 – 100	відмінно	зараховано
B	81-89	добре	
C	71-80		
D	61-70		
E	51-60	задовільно	
FX	21-50	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
F	0-20	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

* кількість балів для оцінок «незадовільно» (FX і F) визначається Вченими радами факультетів (педагогічними радами коледжів).

13. Методичне забезпечення

14. Рекомендована література

- 1) Документація Apache Hadoop [Електронний ресурс] // Apache Hadoop. – 2021. – Режим доступу до ресурсу: <https://hadoop.apache.org/docs/stable/>.
 - 2) Документація Apache Spark [Електронний ресурс] // Apache Spark. – 2019. – Режим доступу до ресурсу: <https://spark.apache.org/docs/latest/>.
 - 3) Документація HBase [Електронний ресурс] // HBase. – 2019. – Режим доступу до ресурсу: <https://hbase.apache.org/book.html>.
 - 4) RabbitMQ [Електронний ресурс] // RabbitMQ. – 2020. – Режим доступу до ресурсу: <https://www.rabbitmq.com/documentation.html>.
- Ifeyinwa A. A. Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications / A. A. Ifeyinwa, H. N. Friday. – Nigeria: Abakaliki, 2019. – 30 с.

15. Інформаційні ресурси

1. Internet – джерела.
2. Наукова бібліотека Львівського національного університету імені Івана Франка (<https://www.lnulibrary.lviv.ua/to-users-2/paid-services/internet/>).
3. Львівська національна наукова бібліотека України імені Василя Стефаника (<https://www.lsl.lviv.ua/index.php/uk/elektronni-resursy1/>).