



## СИЛАБУС

навчальної дисципліни «Аналітика великих даних»

### 1. Загальна інформація

Назва дисципліни	Аналітика великих даних
Статус дисципліни	Нормативна
Рівень вищої освіти, форма навчання	другий(магістерський), дення форма
Освітньо-професійна програма	Комп'ютерні науки
Спеціальність	122 Комп'ютерні науки
Рік навчання, семестр	2-й рік (3 семестр)
Мова викладання	українська
Викладач	Придатко Олександр Володимирович, к.т.н., доцент, начальник кафедри інформаційних технологій та систем електронних комунікацій
E-mail	<a href="mailto:o_prydatko@ukr.net">o_prydatko@ukr.net</a>
Сторінка курсу в ВУ	<a href="https://virt.ldubgd.edu.ua/course/view.php?id=3348">https://virt.ldubgd.edu.ua/course/view.php?id=3348</a>
Консультації	Згідно розкладу консультацій кафедри інформаційних технологій та систем електронних комунікацій

### 2. Анотація до курсу

Освітня програма підготовки магістра зі спеціальності «Комп'ютерні науки» передбачає оволодіння здобувачами освіти низки фахових компетентностей щодо побудови ефективних систем аналітики даних (у тому числі великих обсягів) для забезпечення процесів підтримки прийняття управлінських рішень. Оволодіння означеними та іншими компетентностями, в рамках освітньої програми, передбачено шляхом вивчення курсів «Прикладна математика та статистика для Data Sciencs», «Python для Data Science», «Візуалізація даних», «Бази та сховища даних» тощо. Сам курс «Аналітика великих даних» є узагальнюючим та орієнтований на закріплення здобувачами освіти теоретичних знань та практичних навиків щодо побудови та використання ефективних систем аналітики даних із використанням багатокритеріальних SQL запитів, середовища Power BI, мови формул DAX та існуючих принципів побудови систем великих даних.

Предметом вивчення навчальної дисципліни є сучасні технології збору, зберігання, оброблення та аналітики великих даних. Інструментальними засобами для оволодіння предметом вивчення навчальної дисципліни є мова структурованих запитів SQL, системи управління нереляційними базами даних NoSQL, середовище Apache Spark, середовище Power BI, платформа Databricks.



### 3. Мета і завдання курсу

**3.1. Метою** вивчення дисципліни «Аналітика великих даних» є отримання знань, вмінь та навичок в області роботи з даними, оброблення великих обсягів даних для встановлення закономірностей між ними, аномалій, а також формування ефективних рішень на основі отриманих результатів аналізу даних. В рамках курсу здобувачі освіти вивчають сучасні інструменти та технології аналізу великих Data-сетів, а також побудови систем збору, зберігання та оброблення великих масивів даних.

#### 3.2. Завдання:

- удосконалення практичних навиків щодо формування багатокритеріальних SQL запитів для отримання необхідних даних з Data-сетів;
- удосконалення практичних навиків щодо роботи з реляційними базами даних;
- освоєння інструментальних засобів аналітики великих даних: Power BI; Tableau тощо;
- оволодіння базовими навиками роботи з Apache Spark та платформою Databricks;
- оволодіння основними принципами побудови систем зберігання та оброблення великих даних.

#### 3.3. Компетентності:

*Загальні компетентності:*

- ЗК1 Здатність до абстрактного мислення, аналізу та синтезу.

*Спеціальні (фахові) компетентності:*

- СК4 Здатність збирати і аналізувати дані (включно з великими), для забезпечення якості прийняття проєктних рішень.
- СК6 Здатність застосовувати існуючі і розробляти нові алгоритми розв'язування задач у галузі комп'ютерних наук (алгоритми розв'язання обчислювальних та логічних задач, алгоритми паралельних та розподілених обчислень, алгоритми аналітичної обробки й інтелектуального аналізу великих даних тощо).
- СКО1 Здатність до проєктування, розробки, впровадження і супроводу програмного забезпечення інтелектуальних систем збору та аналітики даних, а також інформаційних, комп'ютерних та програмних систем безпеко-орієнтованого спрямування.
- СКО2 Здатність застосовувати технології моделювання та прогнозування процесів для одержання нових знань і встановлення ймовірнісних зв'язків між даними, а також математичних методів, програмних засобів і технологій інтелектуального аналізу та обробки даних (включно з великими) для підвищення їх інформативності та значущості задля підтримки прийняття управлінських рішень.

#### 3.4. Програмні результати навчання:

- РН1 Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.
- РН8 Розробляти математичні моделі та методи аналізу даних (включно з великим).
- РН9 Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).
- РН11 Створювати нові алгоритми розв'язування задач у сфері комп'ютерних наук, оцінювати їх ефективність та обмеження на їх застосування.
- РНО3 Застосовувати методи і засоби роботи з даними (збір, обробка, зберігання), а також володіти програмним забезпеченням аналізу великих масивів даних для розробки безпеко-орієнтованих систем.



#### 4. Формат і обсяг курсу

<b>Формат курсу</b>	Засвоєння теоретичного матеріалу курсу передбачає відвідування 8 лекційних занять, здачу тестових завдань на базі електронного освітнього середовища. Засвоєння практичної частини курсу зосереджено в рамках 8 лабораторних занять, під час яких здобувачі освіти виконують 4 лабораторних роботи. Виконані роботи необхідно завантажувати до відповідної категорії електронного освітнього середовища для їх подальшого захисту на оцінку.
<b>Обсяг дисципліни:</b>	3,0 кредити / 90 академічних годин, з яких: лекцій 16 годин, лабораторних 16 годин, самостійної роботи 58 годин.
<b>Форми навчання</b>	лекції, лабораторні заняття, консультації, самостійна робота.

#### 5. Тематика та зміст курсу

Назви змістових модулів і тем	Кількість годин				
	денна форма				
	усього	у тому числі			
л		п	лаб	с.р.	
1	2	3	4	5	6
<b>Тема 1.</b> Вступ до аналітики даних.	11	2	-	2	7
<b>Тема 2.</b> Бази даних NoSQL.	11	2	-	2	7
<b>Тема 3.</b> ETL та ELT процеси.	11	2	-	2	7
<b>Тема 4.</b> Машинне навчання в аналізі великих даних.	22	4	-	4	14
<b>Тема 5.</b> Візуалізація великих даних.	12	2	-	2	8
<b>Тема 6.</b> Безпека і конфіденційність при роботі з великими даними.	11	2	-	2	7
<b>Тема 7.</b> Програмні платформи для роботи з Big Data.	12	2	-	2	8
<b>Разом</b>	<b>90</b>	<b>16</b>	<b>-</b>	<b>16</b>	<b>58</b>

#### 6. Інформаційний обсяг навчальної дисципліни

##### Тема 1. Вступ до аналітики даних.

Поняття великих даних. Джерела великих даних. Структуровані та неструктуровані дані. Основні ознаки великих даних (3V). Сфери та приклади застосування великих даних. Основні етапи роботи з великими даними. Збір, підготовка, зберігання, аналіз, візуалізація.

##### Тема 2. Бази даних NoSQL

Порівняння реляційних і нереляційних баз даних. Стовпчикові, графові, документоорієнтовані бази даних. Бази даних «ключ – значення». Найпоширеніші СУБД. MongoDB, Cassandra, Redis, DynamoDB, Couchbase, Neo4j, HBase, Firebase Realtime Database, Amazon SimpleDB, OrientDB.

##### Тема 3. ETL та ELT процеси.

Extract, Transform, Load та Extract, Load, Transform концепції. Поняття пайплайну. Основні процеси: екстракція (вилучення даних із різних джерел), трансформація (очищення, агрегування та перетворення даних у потрібний формат), завантаження (збереження підготовлених даних у цільову систему).

**Тема 4. Машинне навчання в аналізі великих даних.** Задачі машинного навчання. Архітектури штучних нейронних мереж. Обробка природної мови. Трансформери. Лематизація, стемінг, мішок слів. Механізм уваги, енкодери та декодери.

##### Тема 5. Візуалізація великих даних.

Поняття дашборду. Key performance indicators. Інструменти створення дашбордів. PowerBI. Створення візуальних елементів. Таблиця та Матриця. Форматування. Зміна методу агрегації



даних. Відсоткові розрахунки. Картки та багаторядкові картки. Гістограми та діаграми: звичайні, нормовані, з накопиченням. Лінійний графік та діаграма з областями. Діаграми секторна, кільцева, деревоподібна. Комбіновані діаграми. Візуалізації карт. Візуальний елемент Фільтр. Водоспад, воронка, точкова та стрічкова діаграми.

**Тема 6. Безпека і конфіденційність при роботі з великими даними.** Загрози безпеці великих даних. Принципи забезпечення конфіденційності. Шифрування даних. Управління доступом. Безпечне зберігання даних. Анонімізація даних. Забезпечення відповідності нормативним вимогам. Аудит і моніторинг. Захист від зовнішніх атак. Управління ризиками при роботі з даними.

#### **Тема 7. Програмні платформи для роботи з Big Data.**

Історична ретроспектива розвитку розподілених обчислень. Парадигма Map Reduce, Hadoop, Apache Spark, Databricks.

### **7. Завдання для самостійного опрацювання**

1. З метою закріплення отриманих практичних навиків, здобувачі освіти виконують лабораторні роботи, завдання на які отримують в під час або наприкінці лабораторного заняття. Зміст та варіанти виконання завдань (методичні вказівки) відображені на платформі електронного освітнього середовища «Віртуальний університет». Звіт про виконання завдань (лабораторних робіт) здається для подальшого оцінювання, залежно від теми, двома способами:

– розв'язки (результати виконання) лабораторних завдань заносяться до тестової системи електронного освітнього середовища «Віртуальний університет» з метою автоматизованого оцінювання отриманих результатів;

– звіти лабораторних робіт завантажуються до відповідної категорії електронного освітнього середовища для подальшої перевірки викладачем та їх захисту на оцінку (на наступному лабораторному занятті).

### **8. Методи навчання**

Основні форми організації навчання: лекції; лабораторні заняття; індивідуальні практичні завдання, консультації.

Методи організації та здійснення навчально-пізнавальної діяльності:

- лекції – словесні та пояснювально-ілюстративний (наочний) метод (демонстрація, ілюстрація);
- лабораторні роботи та індивідуальні практичні завдання – дослідницький метод (метод спостереження), репродуктивний метод (виконання завдання за визначеним алгоритмом), частково-пошуковий метод навчання (певні елементи матеріалу відомі, решту здобувачі освіти отримують самостійно виконуючи завдання, розв'язуючи задачі тощо);
- консультації – словесний та дискусійний методи.

### **9. Технічне й програмне забезпечення /обладнання**

Комп'ютери на базі процесорів Intel Pentium Gold G5400 або Intel Core i5, компоненти програмного забезпечення MS Office 365 (Teams, PowerPoint, Word, Excel), мова структурованих запитів SQL, система управління базами даних MySQL, MySQL Server, середовище MySQL Workbench, середовище Power BI, середовище Tableau, навчальні/тренувальні Data-сети, інтегроване середовище розробки Eclipse або IntelliJ IDEA, комплект розробника Java (JDK), платформа Databricks, електронне освітнє середовище «Віртуальний університет» (на базі платформи Moodle).



## 10. Критерії оцінювання

Оцінювання результатів навчання здобувачів вищої освіти здійснюється відповідно до «Положення про організацію освітнього процесу у ЛДУ БЖД» <https://cutt.ly/OWRAkEh> та «Положення про порядок та критерії оцінювання результатів навчання здобувачів вищої освіти у ЛДУ БЖД» <https://cutt.ly/iWRAWF3>.

<b>Поточний контроль</b>	
<p>Поточний контроль проводиться у формі виконання тестових завдань на базі платформи електронного освітнього середовища «Віртуальний університет» та виконання лабораторних робіт з подальшим їх захистом на оцінку (викладачеві або через автоматизовану систему оцінювання). Оцінювання результатів поточного контролю здійснюється за національною (чотирибальною) шкалою. Результати поточного контролю (поточна успішність) враховуються викладачем при визначенні допуску до підсумкового контролю та виставленні підсумкової оцінки за екзамен.</p>	
<b>Вид робіт</b>	<b>Формат проведення та критерії оцінювання</b>
Тестові завдання	Курсом передбачено проходження 6 тестових завдань за матеріалами лекцій. Оцінювання здійснюється за національною (чотирибальною) шкалою, відповідно до Додатку Б «Положення про порядок та критерії оцінювання результатів навчання здобувачів вищої освіти у ЛДУ БЖД». За успішне виконання тестових завдань сумарно можна отримати до <b>60 балів</b> .
Лабораторні роботи (індивідуальні практичні завдання)	Курсом передбачено виконання та захист 4 лабораторних робіт. За успішне виконання та захист лабораторних робіт виводиться середня оцінка за національною шкалою та шляхом пропорційного співвідношення переводиться у 100-бальну шкалу. Здобувачі освіти мають можливість одержати за виконання та захист усіх лабораторних робіт сумарно до <b>40 балів</b> . Критерії оцінювання наведені в електронному курсі «Віртуального університету».

За виконання завдань та тестувань здобувач може отримати сумарно до **100 балів**

<b>Підсумковий контроль</b>
<p>Підсумковий контроль проводиться у формі екзамену. Допуск до підсумкового контролю здійснюється за умови виконання здобувачем освіти лабораторних робіт і тестових завдань та одержання понад <b>60</b> зі <b>100</b> можливих балів за результатами проходження курсу на базі електронного освітнього середовища «Віртуальний університет».</p> <p><b>Екзамен (максимально 70 балів)</b> складається із теоретичної та практичної частини і проводиться на базі електронного освітнього середовища «Віртуальний університет».</p> <p>На теоретичну частину (максимально 40 балів) виноситься 40 випадково згенерованих тестових питань закритої форми з вибором однієї правильної відповіді за матеріалами курсу. Вага кожного питання тесту = 1 балу.</p> <p>Практична частина (максимально 30 балів) передбачає розв'язок практичного завдання, яке оцінюється у 30 балів.</p> <p><i>Критерії оцінювання практичного завдання:</i></p> <ul style="list-style-type: none"><li>- 30 балів – вірно розв'язання запропонованої практичної задачі у повній мірі;</li><li>- 25-29 балів – вірно розв'язання запропонованої практичної задачі у повній мірі, проте присутні незначні помилки;</li><li>- 15-24 балів – розв'язання запропонованої практичної задачі не в повній мірі відповідає завданню, або розв'язане частково, проте спостерігається вірний хід розв'язання завдання;</li></ul>





- 10-14 балів – наведене розв'язання запропонованої практичної задачі містить неточності, не враховує усі особливості реалізації або частково не відповідає змісту завдання, проте містить вірний хід розв'язання поставленого завдання;
- 1-9 балів – розв'язання запропонованої практичної задачі не вірне або відсутнє, однак спостерігається вірно обраний напрям вирішення завдання;
- 0 балів – завдання не виконане або розв'язок не відповідає поставленому завданню.

Підсумкова семестрова оцінка обчислюється як сума балів поточного (з коефіцієнтом 0,3 та округлення до цілого числа) та підсумкового контролю за 100-бальною шкалою і переводяться в національну (чотирибальну) шкалу (“відмінно”, “добре”, “задовільно”, “незадовільно”).

Підсумкові оцінки вносяться до екзаменаційної відомості в національній, 100-бальній шкалі та шкалі ЄКТС відповідно до співвідношень, поданих у наступній таблиці.

### Шкала оцінювання результатів навчання здобувачів вищої освіти

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, диференційованого заліку, курсового проекту (роботи), практики	для заліку
91 – 100	<b>A</b>	відмінно	зараховано
81-90	<b>B</b>	добре	
71-80	<b>C</b>		
61-70	<b>D</b>		
51-60	<b>E</b>	задовільно	не зараховано
36-50	<b>FX</b>	незадовільно	
0-35	<b>F</b>		

## 11. Політика курсу

Виконання навчальних завдань і робота на курсі має відповідати вимогам «Кодексу академічної доброчесності та корпоративної культури ЛДУ БЖД» <https://cutt.ly/5WTlc6y>.

*Академічні очікування від здобувачів* – своєчасне виконання тестових завдань, передбачених силабусом дисципліни; обов'язкове відвідування лабораторних занять і виконання ідивідуальних практичних завдань (лабораторних робіт).

*Політика щодо термінів виконання завдань та ліквідації академічної заборгованості:* терміни виконання завдань вказуються у електронному курсі «Віртуального університету». Після завершення терміну прийому завдань, система блокує можливість їх завантаження для подальшої оцінки викладачем, окрім випадків пов'язаних із поважними причинами, про що здобувач особисто повідомляє викладача. Відпрацювання академічної заборгованості з дисципліни можливо до дня проведення підсумкового контролю (відповідно до розкладу).

Недопущені до підсумкового контролю здобувачі освіти здійснюють перездачу в терміни, відведені для усунення академічної заборгованості у два етапи:

- заборгованість із поточного контролю;
- заборгованість із підсумкового контролю.

Ліквідація заборгованості поточного контролю відбувається шляхом проходження тестових завдань та виконання лабораторних робіт згідно із тематичним планом курсу. Ліквідація заборгованості з підсумкового контролю організовується в форматі перездачі екзамену.

*Дотримання принципів академічної доброчесності:* роботи (завдання) виконуються здобувачами самостійно, ідеї та ініціативи інших авторів використовуються лише при належно оформленому цитуванні.

*Поведінка в аудиторії* – неприпустимо запізнення та користування телефоном на заняттях, за винятком виконання громіздких обчислень та використанні додаткових програм в освітніх цілях; повага до думки інших колег; дотримання норм культури мовлення та ін.



## 12. Рекомендована література

### 12.1. Основна:

1. *Analyzing Data With Power BI and Power Pivot for Excel* / Альберто Феррарі, Марко Руссо. - Print2print, 2020. - 288 с.
2. *The Definitive Guide to DAX* / Альберто Феррарі, Марко Руссо. - Print2print, 2021. – 776 с.
3. *Expert Data Modeling with Power BI* / Сохейл Бахшію - Print2print, 2020. – 490 с.
4. *Big Data: Principles and best practices of scalable realtime data systems* / Nathan Marz, James Warren. – Manning, 2015. – 328 p.
5. *Big Data Fundamentals: Concepts, Drivers & Techniques* / Thomas Erl, Wajid Khattak, Paul Buhler. – Pearson, 2016. – 240 p.
6. *MySQL Cookbook: Solutions for Database Developers and Administrators 3rd Edition* / Paul DuBois. – O`Reilly, 2014. – 866 p.

### 12.2. Додаткова:

1. *SQL для початківців, 9-е видання* / А. Тейлор. – Київ: ДІАЛЕКТИКА, 2020. – 544 с.
2. *Системи баз даних та знань. Книга 1. Організація баз даних та знань* : [навч. посіб.] / А. Ю. Берко, О. М. Верес, В. В. Пасічник. – Львів: «Магнолія-2006», 2019. – 584 с.
3. *Системи баз даних та знань. Книга 2. Системи управління базами даних та знань* : [навч. посіб.] / А. Ю. Берко, О. М. Верес, В. В. Пасічник. – Львів: «Магнолія-2006», 2019. – 584 с.
4. *Консолідовані інформаційні ресурси баз даних та знань*: Навч. посібник. / І.П. Жежнич. – Львів: Вид-во НУ “Львівська політехніка”, 2010. – 212 с.
5. *Пономаренко В. С. Інструментальні засоби розробки та підтримки баз даних розподілених інформаційних систем* / В. С. Пономаренко, Л. А. Павленко. – Х.: Вид. ХДЕУ, 2001. – 132 с.
6. *Придатко О. В. Інтеграція новаційного методу мобільного навчання в освітні проекти підготовки розробників програмного забезпечення* / О. В. Придатко, В. В. Придатко, Ю. О. Борзов, В. Є. Дзень // Вісник ЛДУБЖД: Зб. наук. праць. Львів: ЛДУ БЖД, 2018. – №18. – С.70-80.
7. *Martyn Ye. Software for Shelter’s Fire Safety and Comfort Levels Evaluation* / Martyn Ye., Smotr O., Burak N., Prydatko O., Malets I. // *Communications in Computer and Information Science*, Springer, Cham. – Vol. 1158, 2020. pp. 457-469 [https://doi.org/10.1007/978-3-030-61656-4\\_31](https://doi.org/10.1007/978-3-030-61656-4_31)
8. *Придатко О. В. Адаптивна інформаційно-довідкова система "UniBell" як складова частина проєкту "Smart-університет"* / О. В. Придатко, Н. Є. Бурак, В. Є. Дзень, М. С. Кунинець // Науковий вісник НЛТУ України : Зб. наук. праць. Львів: НЛТУ, 2020. - т. 30, № 5 – С. 113–121.
9. *Prydatko O. Informational System of Project Management in the Areas of Regional Security Systems' Development* / O. Prydatko, O. Smotr, Yu. Borzov, I. Solotvinskyi, O. Didyk // 2018 IEEE Second Conference on Data Stream Mining & Processing. Lviv, 2018. – №2 – 187-192.

### 12.3. Інформаційні ресурси:

1. *Віртуальний університет ЛДУ БЖД* [Електронний ресурс]. — Доступний з <https://virt.ldubgd.edu.ua/>
2. *Основи аналітики даних на базі платформи Prometheus* [Електронний ресурс]. — Доступний з <https://prometheus.org.ua/prometheus-plus/data-analytics-basics/>




Львівський державний університет безпеки життєдіяльності  
Навчально-науковий інститут цивільного захисту

Розглянуто на засіданні кафедри інформаційних технологій та систем електронних комунікацій  
протокол від «30» серпня 2023 року № 1


**РОЗРОБНИК**

Начальник кафедри інформаційних технологій та систем електронних комунікацій  
кандидат технічних наук, доцент

  
Олександр ПРИДАТКО  
« 30 » серпня 2023 р.


**ЗАТВЕРДЖЕНО**

Начальник кафедри інформаційних технологій та систем електронних комунікацій  
кандидат технічних наук, доцент

  
Олександр ПРИДАТКО  
« 30 » серпня 2023 р.


**ПОГОДЖЕНО**

Гарант освітньої програми «Комп'ютерні науки»  
другого (магістерського) рівня вищої освіти  
кандидат технічних наук, доцент

  
Назарій БУРАК  
« 30 » серпня 2023 р.

**ПОГОДЖЕНО**

Заступник начальника навчально-наукового інституту цивільного захисту  
кандидат фізико-математичних наук, доцент

  
Ольга МЕНЬШИКОВА  
« 30 » серпня 2023 р.

Дата актуалізації*					
Підпис					
Ім'я, прізвище завідувача кафедри					