

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ДВНЗ «ПРИКАРПАТСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ВАСИЛЯ СТЕФАНІКА»**

Факультет/інститут економічний

Кафедра економічної кібернетики

**СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

**Технології та інфраструктура Big Data**

Освітня програма Економіка

Спеціальність 051 Економіка

Галузь знань 05 Соціальні та поведінкові науки

Затверджено на засіданні кафедри  
Протокол № 1 від 29 серпня 2019 р.

## **ЗМІСТ**

1. Загальна інформація
2. Анотація до курсу
3. Мета та цілі курсу
4. Результати навчання (компетентності)
5. Організація навчання курсу
6. Система оцінювання курсу
7. Політика курсу
8. Рекомендована література

<b>1. Загальна інформація</b>	
<b>Назва дисципліни</b>	Технології та інфраструктура Big Data
<b>Викладач (-і)</b>	к. ф.м. н., доц. Дмитришин М.І.
<b>Контактний телефон викладача</b>	+38(096)5346498
<b>Е-mail викладача</b>	<a href="mailto:marian.dmytryshyn@pnu.edu.ua">marian.dmytryshyn@pnu.edu.ua</a>
<b>Формат дисципліни</b>	Очний
<b>Обсяг дисципліни</b>	6 кредитів ECTS
<b>Посилання на сайт дистанційного навчання</b>	<a href="http://www.d-learn.pu.if.ua/index.php?">http://www.d-learn.pu.if.ua/index.php?</a>
<b>Консультації</b>	Очні консультації: згідно розкладу консультацій
<b>2. Анотація до курсу</b>	
Курс «Технології та інфраструктура Big Data» присвячений базовим поняттям великих даних і питанням аналізу великих даних та пов'язаних з ними технічних, концептуальних та етичних проблем. В рамках курсу передбачається ознайомлення студентів з форматом великих даних, практикою обробки та аналізу великих, складних структур даних.	
<b>3. Мета та цілі курсу</b>	
Метою вивчення дисципліни є ознайомлення з технологіями аналізу великих даних як інструменту для розв'язання прикладних проблем. Цілі курсу передбачають формування здатностей до пошуку, оброблення та аналізу інформації з різних джерел, формування моделей представлення великих даних, визначення основних методів отримання та аналізу даних.	
<b>4. Результати навчання (компетентності)</b>	
<b>Результати навчання:</b>	
5. Застосовувати аналітичний та методичний інструментарій для обґрунтування пропозицій та прийняття управлінських рішень різними економічними агентами (індивідуумами, домогосподарствами, підприємствами та органами державної влади).	
7. Пояснювати моделі соціально-економічних явищ з погляду фундаментальних принципів і знань на основі розуміння основних напрямів розвитку економічної науки.	
13. Ідентифікувати джерела та розуміти методологію визначення і методи отримання соціально-економічних даних, збирати та аналізувати необхідну інформацію, розраховувати економічні та соціальні показники.	
19. Використовувати інформаційні та комунікаційні технології для вирішення соціально-економічних завдань, підготовки та представлення аналітичних звітів.	
21. Вміти абстрактно мислити, застосовувати аналіз та синтез для виявлення ключових характеристик економічних систем різного рівня, а також особливостей поведінки їх суб'єктів.	
<b>Компетентності:</b>	
ІК - Здатність розв'язувати складні спеціалізовані задачі та практичні проблеми в економічній сфері, які характеризуються комплексністю та невизначеністю умов, що передбачає застосування теорій та методів економічної науки.	
ЗК3. Здатність до абстрактного мислення, аналізу та синтезу.	
ЗК4. Здатність застосовувати знання у практичних ситуаціях.	
ЗК7. Навички використання інформаційних і комунікаційних технологій.	
ЗК8. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.	
СК7. Здатність застосовувати комп'ютерні технології та програмне забезпечення з обробки даних для вирішення економічних завдань, аналізу інформації та підготовки аналітичних звітів.	
СК10. Здатність використовувати сучасні джерела економічної, соціальної, управлінської, облікової інформації для складання службових документів та аналітичних звітів.	
<b>5. Організація навчання курсу</b>	
Обсяг курсу - 180 год.	
Вид заняття	Загальна кількість годин
Лекції	24
Практичні	36

Самостійна робота		120			
Ознаки курсу					
Семестр	Спеціальність	Курс (рік навчання)	Обов'язковий / вибірковий		
8	051 Економіка	IV	Цикл професійної підготовки Вибіркові дисципліни		
Тематика курсу					
Тема, план	Форма заняття	Літерату ра	Завдання, год	Вага оцінки	Термін виконання
<b>Тема 1. Основні положення концепції Big Data</b> Що такі великі дані і чому ви майте? Що таке аналітика великих даних? Як отримати великі дані від традиційних джерел?	Лекція, практичне заняття	[1-10]	Опрацювати лекційний матеріал, підготуватися до практичного заняття	0,16	До наступного заняття за розкладом
<b>Тема 2. BIG DATA, екосистема APACHE HADOOP</b> Основні концепції Hadoop®; Встановлення та налаштування: локально і в хмарі; HDFS: Архітектура, реплікація, читання і запис даних; Команди HDFS	Лекція, практичне заняття	[1-10]	Опрацювати лекційний матеріал, підготуватися до практичного заняття	0,17	До наступного заняття за розкладом
<b>Тема 3. MapReduce в Hadoop</b> Робочий процес MapReduce. Hadoop MapReduce і HDFS. Виконання MapReduce. Архітектура і робота першої версії MapReduce. Hadoop API (типи, класи). WordCount (Congigure Job, Mapper, Reducer). Reducer як Combiner. Типи даних в Hadoop. InputSplit, InputFormat, OutputFormat. Shuffle і Sort в Hadoop. Запуск і налагодження завдань. Hadoop Streaming. Streaming в MapReduce.	Лекція, практичне заняття	[1-10]	Опрацювати лекційний матеріал, підготуватися до практичного заняття, розв'язати задачі Контрольна робота	0,17	До наступного заняття за розкладом
<b>Тема 4. Вступ в Pig і Hive</b> Що таке Pig, для чого використовується. Pig і MapReduce. Основні можливості. Компоненти. Режими виконання. Запуск Pig. Pig Latin. Операції DUMP і STORE. Великий обсяг даних. Команда LOAD. Типи даних для схеми. Pig Latin (засоби діагностики, групування, Inner і Outer bag, FOREACH, функція TOKENIZE, оператор FLATTEN, WordCount, Inner і Outer Join). Hive (архітектура,	Лекція, практичне заняття	[1-10]	Опрацювати лекційний матеріал, Пройти тестування до теми	0,17	До наступного заняття за розкладом

інтерфейс, концепція, створення таблиці, завантаження даних, виконання запиту, Inner і Outer Join, WordCount).					
<b>Тема 5. Apache Mahout</b> Що таке Apache Mahout. Реалізовані алгоритми. Класифікація (Naive Bayes, k-Means). Рекомендації (колаборативна фільтрація, Item-based, алгоритм Slop One, Apache.teste, Item-based з Hadoop, Mahout зі Spark, co-occurrence recommenders).	Лекція, практичне заняття	[1-10]	Опрацювати лекційний матеріал, підготуватися до практичного заняття	0,16	До наступного заняття за розкладом
<b>Тема 6. NoSQL, HBase, Cassandra</b> Масштабування вгору. Масштабування RDBMS (master / slave, sharding). Що таке NoSQL. Dynamo і BigTable. Теорія CAP. Модель консистентності. Eventual Consistency. Типи NoSQL. Key / Value. Schema-Less. Що таке HBase, коли його потрібно і коли не потрібно використовувати. Модель даних HBase. Column Family. Timestamp. Осередки. Архітектура і компоненти HBase. Розподіл ключів в RegionServer. Зберігання даних в HBase. Master і Zookeeper. Доступ до HBase. Column Family як одиниця зберігання. Запит даних з HBase. Що таке Cassandra. Типовий NoSQL API. Модель даних. Cassandra і консистентність.	Лекція, практичне заняття	[1-10]	Опрацювати лекційний матеріал, підготуватися до практичного заняття Контрольна робота	0,17	Згідно розкладу
<b>6. Система оцінювання курсу</b>					
Загальна система оцінювання курсу	<b>100 бальна – 50 балів</b> протягом семестру та <b>50 балів</b> за екзамен; <b>“відмінно”</b> – студент демонструє повні і глибокі знання навчального матеріалу, достовірний рівень розвитку умінь та навичок, правильне й обґрунтоване формулювання практичних висновків, наводить повний обґрунтований розв’язок прикладів та задач, аналізує причинно-наслідкові зв’язки; вільно володіє науковими термінами; <b>“добре”</b> – студент демонструє повні знання навчального матеріалу, але допускає незначні пропуски фактичного матеріалу, вміє застосувати його до розв’язання конкретних прикладів та задач, у деяких випадках нечітко формулює загалом правильні відповіді, допускає окремі несуттєві помилки та неточності в розв’язках;				

	<p><b>“задовільно”</b> – студент володіє більшою частиною фактичного матеріалу, але викладає його не досить послідовно і логічно, допускає істотні пропуски у відповідях, не завжди вміє правильно застосувати набуті знання до розв’язання конкретних прикладів та задач, нечітко, а інколи й невірно формулює основні твердження та причинно-наслідкові зв’язки;</p> <p><b>“незадовільно”</b> – студент не володіє достатнім рівнем необхідних знань, умінь, навичок, науковими термінами.</p>
Вимоги до письмової роботи	Відповідно до навчального плану, студент виконує одну контрольну роботу. Головна її мета – перевірка самостійної роботи студентів в процесі навчання, виявлення ступеня засвоєння ними теоретичних положень курсу. При розв’язанні задач студент має детально вказувати, яким саме був хід його роздумів, якими формулами він користувався.
Семінарські заняття	Практичне заняття проводиться з метою формування у студентів умінь і навичок з предмету, вирішення сформульованих завдань, їх перевірка та оцінювання. За метою і структурою практичні заняття є ланцюжком, який пов’язує теоретичне навчання і навчальну практику з дисципліни, а також передбачає попередній контроль знань студентів. Оцінка за практичне заняття враховується при виставленні підсумкової оцінки з дисципліни.
Умови допуску до підсумкового контролю	<ul style="list-style-type: none"> <li>– оцінка за поточне тестування (10 балів);</li> <li>– оцінка за відповіді на всі основні та додаткові запитання під час аудиторних занять (15 балів);</li> <li>– оцінка за контрольну роботу (15 балів);</li> <li>– оцінка за самостійну роботу (10 балів).</li> </ul>
<b>7. Політика курсу</b>	
<p>- самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей);</p> <p>- посилання на джерела інформації у разі використання ідей, розробок, тверджень, відомостей;</p> <p>- надання достовірної інформації про результати власної навчальної (наукової, творчої) діяльності, використані методики досліджень і джерела інформації.</p> <p>Засвоєння пропущеної теми лекції з поважної причини перевіряється під час складання підсумкового контролю. Пропуск лекції з неповажної причини відпрацьовується студентом відповідно до вимог кафедри, що встановлені на засіданні кафедри (співбесіда, реферат тощо).</p> <p>Пропущені практичні, семінарські та лабораторні заняття, незалежно від причини пропуску, студент відпрацьовує згідно з графіком консультацій. Поточні „2”, отримані студентом під час засвоєння відповідної теми на практичному, семінарському та лабораторному занятті перескладаються викладачеві, який веде заняття до складання підсумкового контролю з обов'язковою відміткою у журналі обліку роботи академічних груп.</p>	
<b>8. Рекомендована література</b>	
<ol style="list-style-type: none"> <li>1. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. — 110 p.</li> <li>2. Ghavami Peter. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL &amp; Big Data Analytics. CreateSpace Independent Publishing Platform, 2016. — 204 p</li> <li>3. Feeney K. et al. Engineering Agile Big-Data Systems //Kevin Feeney, Jim Davies, James Welch, Sebastian Hellmann, Christian Dirschl, Andreas Koller, Pieter Francois, Arkadiusz Marciniak. — River Publishers, 2018. — 436 p.</li> <li>4. Raheem N. Big Data: A Tutorial-Based Approach. Taylor &amp; Francis Group LLC, CRC Press, 2019. — 203 p</li> <li>5. Радченко И.А., Николаев И.Н. Технологии и инфраструктура Big Data. Учебное пособие. — СПб.: Университет ИТМО, 2018. — 52 с.</li> <li>6. Волкова Светлана (ред.). Просто BIG DATA. СПб.: Страта, 2019. — 148 с.</li> <li>7. Вайгенд Андреас. BIG DATA. Вся технология в одной книге. М.: Эксмо, 2018. — 384 с</li> </ol>	

8. Zgurovsky M.Z., Zaychenko Y.P. Big Data: Conceptual Analysis and Applications. Springer, 2020. — 298 p.
9. Wiktorski Tomasz. Data-intensive Systems: Principles and Fundamentals using Hadoop and Spark. Springer, 2019. — 105 p
10. Ын Анналин, Су Кеннет. Теоретический минимум по Big Data. Всё что нужно знать о больших данных. СПб.: Питер, 2019. — 208 с.

**Викладач**

**Дмитришин М.І.**