

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Кафедра програмної інженерії та інформаційних технологій управління
(назва)

«ЗАТВЕРДЖУЮ»

Голова науково-методичної комісії _____
(назва комісії)

_____ (підпис) _____ (ініціали та прізвище)

« _____ » _____ 20 _____ року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Вступ до Big Data

(назва навчальної дисципліни)

рівень вищої освіти другий (магістерський)
перший (бакалаврський) / другий (магістерський)

галузь знань 12 Інформаційні технології
(шифр і назва)

спеціальність 126 Інформаційні системи та технології
(шифр і назва)

вид дисципліни професійна підготовка (вибіркова)
(загальна підготовка / професійна підготовка)

форма навчання денна
(денна / заочна)

Харків – 2017 рік

ЛИСТ ЗАТВЕРДЖЕННЯ

Робоча програма з навчальної дисципліни

Вступ до Big Data

(назва дисципліни)

Розробники:

доцент, к.т.н., доц.

(посада, науковий ступінь та вчене звання)

(підпис)

Шматко О.В.

(ініціали та прізвище)

(посада, науковий ступінь та вчене звання)

(підпис)

(ініціали та прізвище)

Робоча програма розглянута та затверджена на засіданні кафедри

програмної інженерії та інформаційних технологій управління

(назва кафедри)

Протокол від « 31 » _____ 2017 року № 1

Завідувач кафедри _____

(назва кафедри)

(підпис)

Годлевський М.Д.

(ініціали та прізвище)

ЛИСТ ПЕРЕЗАТВЕРДЖЕННЯ РОБОЧОЇ НАВЧАЛЬНОЇ ПРОГРАМИ

Дата засідання кафедри – розробника РПНД	Номер протоколу	Підпис завідувача кафедри	Підпис голови НМК (для дисциплін загальної підготовки та дисциплін професійної підготовки за спеціальністю) або завідувача випускової кафедри (для дисциплін професійної підготовки зі спеціалізації, якщо РПНД розроблена не випусковою кафедрою)

МЕТА, КОМПЕТЕНТНОСТІ, РЕЗУЛЬТАТИ НАВЧАННЯ ТА СТРУКТУРНО-ЛОГІЧНА СХЕМА ВИВЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Мета навчальної дисципліни “Вступ до Big Data” є формування компетентностей щодо аналізу великих даних як інструменту для вирішення дослідницькі питання. Курс знайомить з базовими поняттями великих даних і висвітлює питання аналізу великих даних та пов'язаних з ними технічні, концептуальні та етичні проблеми. Здобувачі вищої освіти набувають практичні навички з розробки та представлення концепції великих даних для конкретних предметних областей реального світу. Курс містить практичні вправи для ознайомлення студентів з форматом великих даних, практичний досвід обробки та аналізу великих, складних структур даних.

Компетентності:

Загальні компетентності:

- Здатність застосовувати знання у практичних ситуаціях.
- Знання та розуміння предметної області та розуміння професійної діяльності.
- Здатність вчитися і оволодівати сучасними знаннями.
- Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

Спеціальні (фахові) компетентності:

- здатність формування моделі представлення великих даних, визначення основних методів видобування та аналізу даних.

Нормативний зміст підготовки здобувачів вищої освіти, сформульований у термінах результатів навчання

Класифікація компетентностей за НРК	Знання	Уміння	Комунікація	Автономія та відповідальність
Вміння розкрити наукову суть проблем у професійній галузі Здатність вирішувати проблеми в професійній галузі на основі аналізу та синтезу	Знання наукових і методологічних основ створення та застосування інформаційних технологій та інформаційних систем для автоматизованої переробки інформації та управління	Уміння застосовувати теоретичні знання у практичних ситуаціях у науковій діяльності.	Спілкування в діалоговому режимі з широкою науковою спільнотою та громадськістю в певній галузі наукової та/або професійної діяльності	Ініціювання інноваційних комплексних проєктів, лідерство та повна автономність під час їх реалізації Соціальна відповідальність за результати прийняття стратегічних рішень. Здатність саморозвиватися і самовдосконалюватися протягом життя, відповідальність

Класифікація компетентностей за НРК	Знання	Уміння	Комунікація	Автономія та відповідальність
				ь за навчання інших

Структурно-логічна схема вивчення навчальної дисципліни

Попередні дисципліни:	Наступні дисципліни:
Бази даних та сховища даних	Преддипломна практика
	Дипломне проектування

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

(розподіл навчального часу за семестрами та видами навчальних занять)

Семестр	Загальний обсяг (годин) / кредитів ECTS	З них		За видами аудиторних занять (годин)			Індивідуальні завдання студентів (КП, КР, РГ, Р, РЕ)	Поточний контроль	Семестровий контроль	
		Аудиторні заняття (годин)	Самостійна робота (годин)	Лекції	Лабораторні заняття	Практичні заняття, семінари			Контрольні роботи (кількість робіт)	Залік
1	2	3	4	5	6	7	8	9	10	11
1	120 /4	60	60	0	0	60		1	3	

Співвідношення кількості годин аудиторних занять до загального обсягу складає 50% (%):

СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

№ з/п.	Види навчальних занять (Л, ЛЗ, ПЗ, СР)	Кількість годин	Номер семестру (якщо дисципліна викладається у декількох семестрах). Назви змістових модулів. Найменування тем та питань кожного заняття. Завдання на самостійну роботу.	Рекомендована література (базова, допоміжна)
1	2	3	4	5

№ з/п.	Види навчальних занять (Л, ЛЗ, ПЗ, СР)	Кількість годин	Номер семестру (якщо дисципліна викладається у декількох семестрах). Назви змістових модулів. Найменування тем та питань кожного заняття. Завдання на самостійну роботу.	Рекомендована література (базова, допоміжна)
Змістовий модуль 1. Вступ до Big Data				
1	ПЗ	10	Тема 1. Основні положення концепції Big Data Що такі великі дані і чому ви майте? Що таке аналітика великих даних? Як отримати великі дані від традиційних джерел? .	[1-3]
	СР	10		
2	ПЗ	10	Тема 2. BIG DATA, екосистема APACHE HADOOP Основні концепції Hadoop®; Встановлення та налаштування: локально і в хмарі; HDFS: Архітектура, реплікація, читання і запис даних; Команди HDFS;	[1-3]
	СР	10		
3	ПЗ	10	Тема 3. MapReduce в Hadoop 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.	[4-5]
	СР	10		
Змістовий модуль 2. Основи аналізу Big Data				
4	ПЗ	10	Тема 4. Вступ в Pig і Hive 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 (середства діагностики, угруповання, Inner і Outer bag, FOREACH, функція TOKENIZE, оператор FLATTEN, WordCount, Inner і Outer Join). Hive (архітектура, інтерфейс, концепція, створення таблиці, завантаження даних, виконання запиту, Inner і Outer Join, WordCount).	[5-8]
	СР	10		
5	ПЗ	10	Тема 5. Apache Mahout Що таке Apache Mahout. Реалізовані алгоритми. Класифікація (Naive Bayes, k-Means). Рекомендації (коллаборативна фільтрація, Item-based, алгоритм Slop One, Apache.teste, Item-based з Hadoop, Mahout зі Spark, co-occurrence recommenders).	[8-10]
	СР	10		
6	ПЗ	10	Тема 6. NoSQL, HBase, Cassandra Масштабування вгору. Масштабування RDBMS (master / slave, sharding). Що таке NoSQL. Dynamo і BigTable. Теорія CAP. Модель консистентності. Eventual Consistency. Типи NoSQL. Key / Value.	[8-10]

№ з/п.	Види навчальних занять (Л, ЛЗ, ПЗ, СР)	Кількість годин	Номер семестру (якщо дисципліна викладається у декількох семестрах). Назви змістових модулів. Найменування тем та питань кожного заняття. Завдання на самостійну роботу.	Рекомендована література (базова, допоміжна)
	СР	10	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.	
Разом (годин)		120		

САМОСТІЙНА РОБОТА

№ з/п	Назва видів самостійної роботи	Кількість годин
1	Опрацювання практичного матеріалу	15
2	Підготовка до практичних(лабораторних, семінарських) занять	15
3	Самостійне вивчення тем та питань, які не викладаються на практичних заняттях	15
4	Виконання індивідуального завдання:	10
5	Інші види самостійної роботи	5
	Разом	60

ІНДИВІДУАЛЬНІ ЗАВДАННЯ

Розрахункова робота

(вид індивідуального завдання)

Розрахункова робота пов'язана з аналізом великих даних. Проект можна виконувати самостійно або в парах. Аналіз даних виконується у відповідності до наборів даних.

- 1. Статистика бідності.** Джерело: Дані Світового банку. Опис: Для країн, які мають активну програму моніторингу бідності, Світовий банк у співпраці з національними установами, іншими агенціями розвитку та громадянським суспільством регулярно проводить аналітичну роботу для оцінки ступеня та причин бідності та нерівності та вивчення впливу зростання та державна політика, а також переглянути дані обстежень домогосподарств та методи вимірювання. Дані включають заходи з бідності та нерівності, отримані з аналітичних звітів, національних програм моніторингу бідності, а також з групи досліджень Світового банку з розвитку, яка з 1990 року виробляє порівняно міжнародні та глобальні оцінки та рівень бідності.
- 2. Скарги споживачів.** Посилання для завантаження: [Consumer_Complaints.xls](#). Джерело: База даних споживчих скарг. Опис: База даних скарг, що надійшла до Бюро фінансового захисту споживачів, щодо фінансових продуктів та послуг.
- 3. Індекс споживчих цін США.** Посилання для завантаження: [historicalcpi.xls](#). Джерело: Міністерство сільського господарства США з питань економічних досліджень. Опис: Індекс споживчих цін (ІСЦ) на продукти є складовою частиною ІСЦ. ІСЦ вимірює середню зміну в часі цін, сплачених міськими споживачами за репрезентативний ринковий кошик споживчих товарів та послуг. У той час як споживчий індекс цін вимірює зміни цін на всі споживчі товари та послуги, включаючи продовольчі товари, ІСЦ для продуктів харчування вимірює зміни лише в роздрібних цінах на продукти харчування.
- 4. Показники жінок та чоловіків.** Посилання для завантаження:
 - а. Правовий вік для шлюбу.

- b. Шлюби
- c. Пільги по вагітності та пологах
- d. Неповна зайнятість
- e. Педагогічний колектив
- f. Жіночі законодавці та менеджери

Джерело: Відділ статистики ООН (UNSD)

Опис: Показники жінок та чоловіків містять найновіші статистичні дані та показники щодо жінок та чоловіків у шести конкретних галузях, що викликають занепокоєння: населення, жінки та чоловіки в сім'ях, охорона здоров'я, освіта, робота та прийняття політичних рішень. Статистика та показники стосуються останнього року, за який доступні дані, розбиті за статтю. Дані були зібрані як з офіційних національних джерел, так і з міжнародних джерел.

5. Стартапи: фінансування та придбання. Посилання для завантаження:

Дані про стартап-компанії, інвестиції та придбання (папка з блискавками, що включає багато файлів CSV). Джерело: Crunchbase. Опис: Дані Crunchbase містять інформацію про велику кількість стартапів, включаючи, хто в них інвестував і скільки. Дані включають компанії по всьому світу, які зібрали гроші, інвестори (індивідуальні та інституційні), які інвестували в ці компанії, раунди фінансування інвестицій та записи про всі придбання цих стартапів. Також міститься інша інформація про компанії (наприклад, категорія, місцезнаходження).

6. Показники злочинності та соціально-економічні показники.

Посилання для завантаження: Злочини - 2001 рік по теперішній час (Чикаго) (Прес "Експорт") Дані перепису - вибрані з 2008 - 2012 роки. Опис: Злочини - зафіксовано випадки злочинів (крім вбивств) у місті Чикаго з 2001 року по сьогодні, мінус останніх сім днів.

7. Оцінки малого доходу та бідності - Файли у каталозі даних містять

оцінки бідності та доходів за 2013 рік. Є один файл даних для кожного штату та США, з даними для всіх статистичних даних за 2013 рік. Крім того, є один файл, який містить дані для США та кожного штату та

округу. Дані перепису - вибір шести соціально-економічних показників, що мають значення для здоров'я населення, та індексу труднощів. (Приклад області для дослідження: Чи існує взаємозв'язок між кількістю злочинів у місті Чикаго та соціально-економічними показниками, такими як середній дохід, бідність та освіта?)

8. **Нью-Йорк.** Посилання для завантаження: Відкриті дані Нью-Йорка. Результати інспекції ресторану в Нью-Йорку. Джерело: data.nyu.gov. Опис: Відкриті дані Нью-Йорка містять дані про найрізноманітніші аспекти Нью-Йорка (наприклад, освіта, безпека, відпочинок та багато іншого). Результати інспекції ресторанів у Нью-Йорку фіксують інформацію про інспекції, порушення, оцінки та судові рішення у Нью-Йорку.
9. **Веллмарт.** Посилання для завантаження: [магазини.csv](#), [особливості.csv](#), [поїзд.csv](#). Джерело: Kaggle Competition. Опис: Історичні дані про продажі для 45 магазинів Walmart в різних регіонах. (Приклад області для вивчення: вплив на продаж погоди, температури, витрати палива, сезон відпусток та інші фактори.)
10. **Світове здоров'я.** Посилання для завантаження: Показники - ми рекомендуємо будь-який із наборів даних у розділі Здоров'я. Джерело: Дані Світового банку. Опис: Дані про здоров'я в усьому світі, що охоплюють такі фактори, як рівень народжуваності, ВІЛ, імунізація, населення, тривалість життя, народжуваність, рівень смертності та багато іншого.

МЕТОДИ НАВЧАННЯ

При викладанні навчальної дисципліни для активізації навчального процесу передбачено застосування сучасних навчальних технологій, таких, як: проблемні лекції; робота в малих групах; семінари-дискусії; кейс-метод; ділові ігри.

Проблемні лекції спрямовані на розвиток логічного мислення студентів. Коло питань теми лекції обмежується двома-трьома ключовими моментами, увага студентів концентрується на матеріалі, що не знайшов широкого відображення в підручниках, використовується досвід закордонних навчальних закладів з роздаванням студентам під час лекцій друкованого матеріалу та виділенням головних висновків з питань, що розглядаються. При викладанні лекційного матеріалу студентам пропонуються питання для самостійного розмірковування. При цьому лектор задає запитання, які спонукають студента шукати розв'язання проблемної ситуації. Така система примушує студентів сконцентруватися і почати активно мислити в пошуках правильної відповіді.

На початку проведення проблемної лекції необхідно чітко сформулювати проблему, яку необхідно вирішити студентам. При викладанні лекційного матеріалу слід уникати прямої відповіді на поставлені запитання, а висвітлювати лекційний матеріал таким чином, щоб отриману інформацію студент міг використовувати при розв'язанні проблеми.

Міні-лекції передбачають викладання навчального матеріалу за короткий проміжок часу й характеризуються значною ємністю, складністю логічних побудов, образів, доказів та узагальнень. Міні-лекції проводяться, як правило, як частина заняття-дослідження. На початку проведення міні-лекції за вказаними темами лектор акцентує увагу студентів на необхідності представити викладений лекційний матеріал у так званому структурно-логічному вигляді. На розгляд виносяться питання, які зафіксовані у плані лекцій, але викладаються вони стисло. Лекційне заняття, проведене у такий спосіб, пробуджує у студента активність та увагу при сприйнятті матеріалу, а також спрямовує його на використання системного підходу при відтворенні інформації, яку він одержав від викладача. Проблемні лекції та міні-лекції доцільно поєднувати з такою формою активізації навчального процесу, як робота в малих групах.

Робота в малих групах дає змогу структурувати лекційні або лабораторні заняття за формою і змістом, створює можливості для участі кожного студента в роботі за темою заняття, забезпечує формування особистісних якостей та досвіду соціального спілкування. Після висвітлення проблеми (при використанні проблемних лекцій) або стислого викладання матеріалу (при використанні міні-лекцій) студентам пропонується об'єднуватися у групи по 5-6 осіб та презентувати наприкінці заняття своє бачення та сприйняття матеріалу.

Презентації – виступи перед аудиторією, що використовуються для представлення певних досягнень, результатів роботи групи, звіту про виконання індивідуальних завдань. Однією з позитивних рис презентації та її переваг при використанні в навчальному процесі є обмін досвідом, який здобули студенти при роботі у певній малій групі.

Лабораторні заняття (з елементами семінарської дискусії) дозволяють формувати у студентів навички особистого експериментального дослідження фізичних процесів що відбуваються під час роботи компонентів операційної системи, проводити аналіз умов її функціонування, а також розробляти нові елементи та системні компоненти відповідно до вимог, що пред'являються до них, узагальнювати отримані результати, формулювати висновки та думки, вести подальший обмін думками та поглядами з іншими учасниками щодо отриманих результатів досліджень з даної теми, а також розвивають творче мислення, допомагають формувати погляди і переконання, вчать об'єктивно оцінювати результати і пропозиції опонентів, критично підходити до власних результатів та поглядів.

Ділові та рольові ігри – форма активізації студентів, за якої вони задіяні в процесі інсценізації певної виробничої ситуації у ролі безпосередніх учасників подій.

Кейс-метод – метод аналізу конкретних ситуацій, який дає змогу наблизити процес навчання до реальної практичної діяльності спеціалістів і передбачає розгляд виробничих, управлінських та інших ситуацій, складних конфліктних випадків, проблемних ситуацій, інцидентів у процесі вивчення навчального матеріалу.

Розподіл форм та методів активізації процесу навчання за темами навчальної дисципліни

Тема	Практичне застосування навчальних технологій
Тема 2. BIG DATA, екосистема APACHE HADOOP	<i>Лекція проблемного характеру “How is big data different from traditional data sources?”</i>
Тема 4. Вступ в Pig і Hive	<i>Міні-лекція “ Difference between Pig and Hive”</i>
Тема 6. NoSQL, HBase, Cassandra	<i>Кейс-метод “Implementation of HBase at Facebook”</i>

МЕТОДИ КОНТРОЛЮ

Система оцінювання знань, вмінь та навичок студентів передбачає виставлення оцінок за усіма формами проведення занять. Перевірка та оцінювання знань студентів може проводитись у таких формах:

1. Оцінювання роботи студентів у процесі практичних занять.
2. Проведення проміжного контролю.
3. Проведення модульного контролю.

Загальна модульна оцінка складається з поточної оцінки, яку студент отримує під час практичних занять та оцінки за виконання модульної контрольної роботи.

Загальна оцінка з дисципліни визначається як середнє арифметичне модульних оцінок та оцінки яка отримана під час проведення заліку.

Порядок поточного оцінювання знань студентів

Поточне оцінювання здійснюється під час проведення лабораторних занять і має на меті перевірку рівня підготовленості студента до виконання конкретної роботи. Об'єктами поточного контролю є:

- 1) активність та результативність роботи студента протягом семестру над вивченням програмного матеріалу дисципліни; відвідування занять;
- 2) виконання проміжного контролю;
- 3) виконання модульного контрольного завдання.

Контроль систематичного виконання самостійної роботи та активності на лабораторних заняттях

Оцінювання проводиться за 5-бальною шкалою за такими критеріями:

- 1) розуміння, ступінь засвоєння теорії та методології проблем, що розглядаються;
- 2) ступінь засвоєння матеріалу дисципліни;
- 3) ознайомлення з рекомендованою літературою, а також із сучасною літературою з питань, що розглядаються;
- 4) уміння поєднувати теорію з практикою при розгляді виробничих ситуацій, розв'язанні задач, проведенні розрахунків при виконанні завдань, винесених для самостійного опрацювання, та завдань, винесених на розгляд в аудиторії;
- 5) логіка, структура, стиль викладу матеріалу в письмових роботах і при виступах в аудиторії, вміння обґрунтовувати свою позицію, здійснювати узагальнення інформації та робити висновки.

Оцінка "відмінно" ставиться за умови відповідності виконаного завдання студента або його усної відповіді до всіх п'яти зазначених критеріїв.

Відсутність тієї чи іншої складової знижує оцінку на відповідну кількість балів.

При оцінюванні практичних завдань увага приділяється також їх якості та самостійності, своєчасності здачі виконаних завдань викладачу (згідно з графіком навчального процесу). Якщо якась із вимог не буде виконана, то оцінка буде знижена.

Проміжний модульний контроль

Проміжний модульний контроль рівня знань передбачає виявлення опанування студентом матеріалу лекційного модуля та вміння застосовувати його для вирішення практичної ситуації і проводиться у вигляді контрольної роботи за темами 1-го або 2-го модулю.

Проведення модульного контролю

Модульний контроль здійснюється та оцінюється за допомогою проведення контрольної роботи за всіма темами дисципліни.

Підсумковий/семестровий контроль проводиться у формі семестрового екзамену. Семестрові екзамени – форма оцінки підсумкового засвоєння студентами теоретичного та практичного матеріалу з окремої навчальної дисципліни, що проводиться як контрольний захід.

Підсумкова оцінка з дисципліни розраховується як середня з кількох складових, що враховує оцінки кожного виду контролю (дві оцінки за результатами поточного модульного контролю, оцінку за курсовий проект і оцінку за семестрову контрольну роботу).

РОЗПОДІЛ БАЛІВ, ЯКІ ОТРИМУЮТЬ СТУДЕНТИ, ТА ШКАЛА ОЦІНЮВАННЯ ЗНАНЬ ТА УМІНЬ (НАЦІОНАЛЬНА ТА ECTS)

Таблиця 1 – Розподіл балів

	Поточний контроль			Семестровий контроль	Всього за семестр
	КР	ПЗ	ІНДЗ		
Підсумкові бали	75			25	100
Макс. проміжні бали	12	5	26		
Кільк. од. обліку у семестрі	2	5	1		
Макс. проміжних балів, всього	24	25	26		100
Коеф.. перерахунку	1				
Макс. кільк. підсумкових балів	25	25	25	25	100

Таблиця 2 – Шкала оцінювання знань та умінь: національна та ЄКТС

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою
90 ... 100	A	відмінно
82 ... 89	B	добре
74 ... 81	C	
64 ... 73	D	задовільно
60 ... 63	E	
35 ... 59	FX	незадовільно з можливістю повторного складання
0 ... 34	F	незадовільно з обов'язковим повторним вивченням дисципліни

НАВЧАЛЬНО-МЕТОДИЧНЕ ЗАБЕЗПЕЧЕННЯ НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

1. Комплект слайдів презентацій з матеріалами лекцій.
2. Методичні вказівки до практичних занять з дисципліни “Вступ до BigData” – Режим доступу: <https://trello.com/b/4w4jmR8S/introduction-to-bigdata>

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. Akerkar R. Models of Computation for Big Data Cham: Springer International Publishing, 2018. — 110 p.
2. Ghavami Peter. Big Data Governance: Modern Data Management Principles for Hadoop, NoSQL & Big Data Analytics. CreateSpace Independent Publishing Platform, 2016. — 204 p
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.
4. Raheem N. Big Data: A Tutorial-Based Approach. Taylor & Francis Group LLC, CRC Press, 2019. — 203 p
5. Радченко И.А., Николаев И.Н. Технологии и инфраструктура Big Data. Учебное пособие. — СПб.: Университет ИТМО, 2018. — 52 с.

Допоміжна література

6. Волкова Светлана (ред.). Просто BIG DATA. СПб.: Страта, 2019. — 148 с.
7. Вайгенд Андреас. BIG DATA. Вся технология в одной книге. М.: Эксмо, 2018. — 384 с
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. Бн Анналин, Су Кеннет. Теоретический минимум по Big Data. Всё что нужно знать о больших данных. СПб.: Питер, 2019. — 208 с.

ІНФОРМАЦІЙНІ РЕСУРСИ В ІНТЕРНЕТІ

11. <https://www.coursera.org/learn/big-data-introduction>
12. <https://www.edx.org/course/introduction-to-big-data-4>