



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Магнитогорский государственный технический университет им. Г.И. Носова»



УТВЕРЖДАЮ
Директор ИЭиАС
В.Р. Храмшин

10.02.2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

ОСНОВЫ ДАТА ИНЖИНИРИНГА

Направление подготовки (специальность)

10.05.03 Информационная безопасность автоматизированных систем

Направленность (профиль/специализация) программы

10.05.03 специализация N 8 "Разработка автоматизированных систем в защищенном исполнении"

Уровень высшего образования - специалитет

Форма обучения
очная

Институт/ факультет	Институт энергетики и автоматизированных систем
Кафедра	Информатики и информационной безопасности
Курс	2
Семестр	4

Магнитогорск
2023 год

Рабочая программа составлена на основе ФГОС ВО - специалитет по специальности 10.05.03 Информационная безопасность автоматизированных систем (приказ Минобрнауки России от 26.11.2020 г. № 1457)

Рабочая программа рассмотрена и одобрена на заседании кафедры Информатики и информационной безопасности
09.02.2023, протокол № 5

Зав. кафедрой  И.И. Баранкова


Рабочая программа одобрена методической комиссией ИЭиАС
10.02.2023 г. протокол № 7

Председатель  В.Р. Храмшин

Рабочая программа составлена:

доцент кафедры ИиИБ, канд. техн. наук  У.В. Кузьмина

Рецензент:

Проректор по цифровизации, канд. техн. наук  К.А. Рубан

Лист актуализации рабочей программы

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2024 - 2025 учебном году на заседании кафедры Информатики и информационной безопасности

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ И.И. Баранкова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2025 - 2026 учебном году на заседании кафедры Информатики и информационной безопасности

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ И.И. Баранкова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2026 - 2027 учебном году на заседании кафедры Информатики и информационной безопасности

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ И.И. Баранкова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2027 - 2028 учебном году на заседании кафедры Информатики и информационной безопасности

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ И.И. Баранкова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2028 - 2029 учебном году на заседании кафедры Информатики и информационной безопасности

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ И.И. Баранкова

Рабочая программа пересмотрена, обсуждена и одобрена для реализации в 2029 - 2030 учебном году на заседании кафедры Информатики и информационной безопасности

Протокол от _____ 20__ г. № ____
Зав. кафедрой _____ И.И. Баранкова

1 Цели освоения дисциплины (модуля)

Целью дисциплины «Основы Data Engineering» является знакомство обучающихся с современными технологиями создания и обслуживания Big Data; с алгоритмами и методами для их обработки, изучение систем и инструментальных средств для хранения и навыками обработки больших массивов данных в соответствии с требованиями ФГОС ВО для специальности 10.05.03 «Информационная безопасность автоматизированных систем».

2 Место дисциплины (модуля) в структуре образовательной программы

Дисциплина Основы Data инжиниринга входит в обязательную часть учебного плана образовательной программы.

Для изучения дисциплины необходимы знания (умения, владения), сформированные в результате изучения дисциплин/ практик:

Информатика

Теория информации

Основы безопасности цифрового общества

Знания (умения, владения), полученные при изучении данной дисциплины будут необходимы для изучения дисциплин/практик:

Безопасность систем баз данных

Информационные технологии. Базы данных

Защита электронного документооборота

3 Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля) и планируемые результаты обучения

В результате освоения дисциплины (модуля) «Основы Data инжиниринга» обучающийся должен обладать следующими компетенциями:

Код индикатора	Индикатор достижения компетенции
ОПК-1	Способен оценивать роль информации, информационных технологий и информационной безопасности в современном обществе, их значение для обеспечения объективных потребностей личности, общества и государства;
ОПК-1.1	Оценивает роль информации в современном обществе
ОПК-1.2	Владеет современными информационными технологиями
ОПК-1.3	Применяет средства обеспечения информационной безопасности

4. Структура, объём и содержание дисциплины (модуля)

Общая трудоемкость дисциплины составляет 3 зачетных единиц 108 акад. часов, в том числе:

- контактная работа – 69,8 акад. часов;
- аудиторная – 68 акад. часов;
- внеаудиторная – 1,8 акад. часов;
- самостоятельная работа – 38,2 акад. часов;
- в форме практической подготовки – 0 акад. час;

Форма аттестации - зачет

Раздел/ тема дисциплины	Семестр	Аудиторная контактная работа (в акад. часах)			Самостоятельная работа студента	Вид самостоятельной работы	Форма текущего контроля успеваемости и промежуточной аттестации	Код компетенции
		Лек.	лаб. зан.	практ. зан.				
1. Основные понятия курса "основы Data Engineering". Технологии обработки Big Data. Отличие от традиционной технологии.								
1.1 Понятие Big Data. Технологии хранения и обработки Big Data. Три определяющих характеристики - основные принципы работы VVV (volume, velocity, variety) с Big Data.	4	4	4		4,2	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС. Подготовка реферата.	реферат	ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу		4	4		4,2			
2. Методы сбора и обработки данных из сети Internet								
2.1 Основы клиент-серверного взаимодействия. Парсинг API Парсинг HTML. BeautifulSoup, MongoDB Обзор систем управления базами данных MongoDB и SQLite	4	4	6		6	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС. Подготовка к тестированию	Тестирование	ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу		4	6		6			
3. Системы управления Big data.								

3.1 Современные технологии обслуживания Big Data. Техники и методики анализа, применимые к Big data. Технология MapReduce.	4	4	6		6	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС. Подготовка к тестированию	Тестирование	ОПК-1.1, ОПК-1.2, ОПК-1.3
3.2 Программное обеспечение, применяемое для структурирования и анализа Big Data. Консолидация и хранение данных в едином пространстве и единой нормализованной модели		4			4	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС. Подготовка к тестированию	Тестирование	ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу		8	6		10			
4. Инструменты для аналитики Big Data. Системы анализа Big Data								
4.1 Культура сбора и источники данных Предобработка и визуализация данных Улучшение качества работы с данными	4	4	6		4	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС. Подготовка к тестированию	Тестирование	ОПК-1.1, ОПК-1.2, ОПК-1.3
4.2 Использование Hadoop MapReduce – платформы программирования и выполнения распределённых MapReduce-вычислений, с использованием большого количества компьютеров (узлов, nodes), образующих кластер.		4	4		2	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС. Подготовка к защите лабораторной работы	Защита лабораторной работы	ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу		8	10		6			
5. Машинное обучение. Построение и анализ модели. Подготовка данных.								

5.1 Подходы к организации поSQL данных. Кластеризация. Оптимизация запросов.	4	4	6		4	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС. Подготовка к тестированию	Тестирование	ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу		4	6		4			
6. Роль языков программирования Python и R в аналитике больших данных.								
6.1 Сравнительный анализ языков Python и R в применении к Data Science - анализ статистики использования в Google Trends. R — язык программирования для статистической обработки данных и работы с графикой. Преимущества и недостатки Python и R.	4	4			6	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС	Реферат	ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу		4			6			
7. Хранилища данных различных типов. ЦОДы. Витрины данных								
7.1 Хранилища данных (Data Warehouse). Типы хранилищ. ЦОДы. Защищенность, интеграция, агрегация. Разделение наборов данных, используемых для оперативной обработки, и для решения аналитических задач.	4	2	2		2	Самостоятельное изучение учебной и научно литературы, работа с материалами образовательного портала и ЭБС.	Защита лабораторной работы	ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу		2	2		2			
8. зачет								
8.1 Подготовка в зачету	4							ОПК-1.1, ОПК-1.2, ОПК-1.3
Итого по разделу								
Итого за семестр		34	34		38,2		зачёт	
Итого по дисциплине		34	34		38,2		зачет	

5 Образовательные технологии

Для реализации предусмотренных видов учебной работы в качестве образовательных технологий в преподавании дисциплины «Основы Data Engineering» используются традиционная и модульно-компетентностная технологии.

Реализация компетентностного подхода предусматривает использование в учебном процессе активных и интерактивных форм проведения занятий в сочетании с внеаудиторной работой с целью формирования и развития профессиональных навыков обучающихся.

При проведении учебных занятий преподаватель обеспечивает развитие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств посредством проведения интерактивных лекций, групповых дискуссий, ролевых игр, тренингов, анализа ситуаций, учета особенностей профессиональной деятельности выпускников и потребностей работодателей.

Формы учебных занятий с использованием традиционных технологий:

- обзорные лекции – для рассмотрения общих вопросов Информатики и информационных технологий, для систематизации и закрепления знаний;
- информационные – для ознакомления с техническими средствами реализации информационных процессов, со стандартами организации сетей, основными приемами защиты информации, и другой справочной информацией;
- лекции-визуализации – для наглядного представления способов решения алгоритмических и функциональных задач, визуализации результатов решения задач;
- семинар;
- практическое занятие, посвященное освоению конкретных умений и навыков по предложенному алгоритму.

6 Учебно-методическое обеспечение самостоятельной работы обучающихся

Представлено в приложении 1.

7 Оценочные средства для проведения промежуточной аттестации

Представлены в приложении 2.

8 Учебно-методическое и информационное обеспечение дисциплины (модуля)

а) Основная литература:

1. Дадян, Э. Г. Методы, модели, средства хранения и обработки данных : учебник / Э.Г. Дадян, Ю.А. Зеленков. — Москва : Вузовский учебник : ИНФРА-М, 2022. — 168 с. - ISBN 978-5-9558-0490-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1834412> (дата обращения: 02.05.2023). – Режим доступа: по подписке.

2. Внуков, А. А. Защита информации : учебное пособие для вузов / А. А. Внуков. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 161 с. — (Высшее образование). — ISBN 978-5-534-07248-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/512268> (дата обращения: 02.05.2023).

3. Черемухин А.Д. Большие данные : учебное пособие / Черемухин А.Д.. — Москва : Ай Пи Ар Медиа, 2023. — 782 с. — ISBN 978-5-4497-2138-9. — Текст : электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/129721.html> (дата обращения: 02.05.2023). — Режим доступа: для авторизир. пользователей

4. Анализ данных : учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва : Издательство Юрайт, 2023. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511020> (дата обращения:

02.05.2023).

б) Дополнительная литература:

1. Мартишин С. А., Симонов В. Л., Храпченко М.В. Базы данных: Работа с распределенными базами данных и файловыми системами на примере MongoDB и HDFS с использованием Node.js, Express.js, Apache Spark и Scala. / Мартишин С. А., Симонов В. Л., Храпченко М.В. М.: НИЦ ИНФРА-М, 2023 - 235 с. Текст : электронный. - URL: <https://znanium.com/catalog/document?id=416159> (дата обращения: 20.04.2023). – Режим доступа: по подписке

2. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/520544> (дата обращения: 02.05.2023).

3. Костюк А.И. Организация облачных и GRID-вычислений : учебное пособие / Костюк А.И.. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2018. — 121 с. — ISBN 978-5-9275-2879-0. — Текст : электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/87734.html> (дата обращения: 02.05.2023). — Режим доступа: для авторизир. пользователей

4. Баранкова, И. И. Разработка БД в MS SQL Server с использованием SSMS : учебное пособие / И. И. Баранкова, У. В. Михайлова, Г. И. Лукьянов ; МГТУ. - Магнитогорск : МГТУ, 2018. - 1 электрон. опт. диск (CD-ROM). - Загл. с титул. экрана. - URL: <https://magtu.informsystema.ru/uploader/fileUpload?name=3473.pdf&show=dcatalogues/1/1514290/3473.pdf&view=true> (дата обращения: 04.10.2019). - Макрообъект. - Текст : электронный. - ISBN 978-5-9967-1207-6. - Сведения доступны также на CD-ROM.

5. Маркин, А. В. Программирование на SQL в 2 ч. Часть 1 : учебник и практикум для вузов / А. В. Маркин. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 429 с. — (Высшее образование). — ISBN 978-5-534-15817-5. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/509818> (дата обращения: 1.05.2023).

6. Маркин, А. В. Программирование на SQL в 2 ч. Часть 2 : учебник и практикум для вузов / А. В. Маркин. — 3-е изд., испр. и доп. — Москва : Издательство Юрайт, 2023. — 385 с. — (Высшее образование). — ISBN 978-5-534-15818-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/509819> (дата обращения: 1.05.2023).

в) Методические указания:

г) Программное обеспечение и Интернет-ресурсы:

Программное обеспечение

Наименование ПО	№ договора	Срок действия лицензии
LibreOffice	свободно распространяемое ПО	бессрочно
Браузер Mozilla Firefox	свободно распространяемое ПО	бессрочно

Браузер	свободно	бессрочно
FAR Manager	свободно	бессрочно
MS Office 2007	№ 135 от 17.09.2007	бессрочно

Профессиональные базы данных и информационные справочные системы

Название курса	Ссылка
Электронные ресурсы библиотеки МГТУ им. Г.И. Носова	https://magtu.informsystema.ru/Marc.html?locale=ru
Архив научных журналов «Национальный электронно-информационный	https://archive.neicon.ru/xmlui/
Международная реферативная и полнотекстовая справочная база данных научных изданий «Springer	https://www.nature.com/siteindex
Международная база полнотекстовых журналов Springer Journals	http://link.springer.com/
Российская Государственная библиотека. Каталоги	https://www.rsl.ru/ru/4readers/catalogues/
Университетская информационная система РОССИЯ	https://uisrussia.msu.ru
Федеральное государственное бюджетное учреждение «Федеральный институт промышленной	URL: http://www1.fips.ru/
Информационная система - Единое окно доступа к информационным	URL: http://window.edu.ru/
Поисковая система Академия Google (Google Scholar)	URL: https://scholar.google.ru/
Национальная информационно-аналитическая система – Российский индекс научного	URL: https://elibrary.ru/project_risc.asp
Электронная база периодических изданий East View Information Services,	https://dlib.eastview.com/

9 Материально-техническое обеспечение дисциплины (модуля)

Материально-техническое обеспечение дисциплины включает:

Учебные аудитории для проведения занятий лекционного типа

Мультимедийные средства хранения, передачи и представления информации.

Учебные аудитории для проведения практических занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации

Мультимедийные средства хранения, передачи и представления информации.

Помещения для самостоятельной работы обучающихся

Персональные компьютеры с выходом в Интернет и с доступом в электронную информационно-образовательную среду университета

Помещения для хранения и профилактического обслуживания учебного оборудования

Шкафы для хранения учебно-методической документации, учебного оборудования и учебно-наглядных пособий.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ВНЕАУДИТОРНЫХ САМОСТОЯТЕЛЬНЫХ РАБОТ

Общие положения

Настоящие методические указания предназначены для организации внеаудиторной самостоятельной работы студентов и оказания помощи в самостоятельном изучении теоретического и реализации компетенций обучаемых.

Данные методические указания не являются учебным пособием, поэтому перед началом выполнения самостоятельного задания следует изучить соответствующие разделы лекционных занятий, материалов образовательного портала, разделов основной и дополнительной литературы, представленных в пункте 8. «Учебно-методическое и информационное обеспечение дисциплины (модуля)» данной РПД.

Цели и задачи самостоятельной работы

Цель самостоятельной работы – содействие оптимальному усвоению учебного материала обучающимися, развитие их познавательной активности, готовности и потребности в самообразовании.

Задачи самостоятельной работы:

- повышение исходного уровня владения информационными технологиями;
- углубление и систематизация знаний;
- постановка и решение стандартных задач профессиональной деятельности;
- развитие работы с различной по объему и виду информацией, учебной и научной литературой;
- практическое применение знаний, умений;
- самостоятельно использование стандартных программных средств сбора, обработки, хранения и защиты информации
- развитие навыков организации самостоятельного учебного труда и контроля за его эффективностью.

Виды внеаудиторной самостоятельной работы и формы контроля и время на выполнение каждого вида самостоятельной работы указаны в пункте 4. «Структура и содержание дисциплины» данной РПД.

Порядок выполнения

При выполнении текущей внеаудиторной самостоятельной работы обучающемуся следует придерживаться следующего порядка действий:

- 1) внимательно изучить соответствующие теоретические разделы дисциплины, пользуясь материалами (лекционными, презентационными, аудио-визуальными): а) предоставляемыми преподавателем на лекционных занятиях;
б) предоставляемыми преподавателем в рамках электронных образовательных курсов;
с) содержащимися в учебниках и учебных пособиях ЭБС (электронно-библиотечных систем), электронных каталогов университета и интернет-ресурсов.
- 2) Подробно разобрать типовые примеры решения задач, рассмотренные в рамках аудиторной контактной работы с преподавателем.
- 3) Применить полученные теоретические знания и практические навыки к решению индивидуальных заданий, к прохождению компьютерных тестирований.
- 4) При необходимости, сформировать перечень вопросов, вызвавших затруднения в процессе самостоятельной работы. Обсудить возникшие вопросы со студентами группы, в рамках

командно-проектной работы, и с преподавателем, в рамках консультационной помощи, реализованной либо в контактной форме, либо средствами информационно-образовательной среды ВУЗа.

Критерии оценки внеаудиторных самостоятельных работ

Качество выполнения внеаудиторной самостоятельной работы обучающихся оценивается посредством текущего контроля самостоятельной работы обучающихся с использованием балльно-рейтинговой системы.

В качестве форм текущего контроля по дисциплине используются: индивидуальные задания, аудиторные контрольные работы, компьютерное тестирование.

Максимальное количество баллов обучающийся получает, если:

- выполняет индивидуальные задания в соответствии со всеми заявленными требованиями;
- дает правильные формулировки, точные определения, понятия терминов;
- может обосновать рациональность решения текущей задачи.;
- обстоятельно с достаточной полнотой излагает соответствующую теоретический раздел;
- правильно отвечает на дополнительные вопросы преподавателя, имеющие целью выяснить степень понимания им данного материала.

50~85% от максимального количества баллов обучающийся получает, если:

- неполно (не менее 70% от полного), но правильно выполнено задание;
- при изложении были допущены 1-2 несущественные ошибки, которые он исправляет после замечания преподавателя;
- дает правильные формулировки, точные определения, понятия терминов;
- может обосновать свой ответ, привести необходимые примеры;
- правильно отвечает на дополнительные вопросы преподавателя, имеющие целью выяснить степень понимания им данного материала.

36~50% от максимального количества баллов обучающийся получает, если:

- неполно (не менее 50% от полного), но правильно изложено задание;
- при изложении была допущена 1 существенная ошибка;
- знает и понимает основные положения данной темы, но допускает неточности в формулировке понятий;
- излагает выполнение задания недостаточно логично и последовательно;
- затрудняется при ответах на вопросы преподавателя.

35% и менее от максимального количества баллов обучающийся получает, если:

- неполно (менее 50% от полного) изложено задание;
- при изложении были допущены существенные ошибки. В "0" баллов преподаватель вправе оценить выполненное обучающимся задание, если оно не удовлетворяет требованиям, установленным преподавателем к данному виду работы или не было представлено для проверки.

Сумма полученных баллов по всем видам заданий внеаудиторной самостоятельной работы составляет рейтинговый показатель обучающегося. Рейтинговый показатель обучающегося влияет на выставление итоговой оценки по результатам изучения дисциплины.

Показатели и критерии оценивания полученных знаний представлены в пункте 7.6) «Оценочные средства для проведения промежуточной аттестации» данной РПД.

7. Оценочные средства для проведения промежуточной аттестации

Код индикатора	Индикатор достижения компетенции	Оценочные средства
<p>ОПК-1 Способен оценивать роль информации, информационных технологий и информационной безопасности в современном обществе, их значение для обеспечения объективных потребностей личности, общества и государства;</p>		
<p>ОПК-1.1</p>	<p>Оценивает роль информации в современном обществе</p>	<p>Перечень теоретических вопросов</p> <ol style="list-style-type: none"> 1. Значение BigData в развитии современного общества 2. Основные направления развития технологий BigData. 3. Современные аппаратные решения для обработки BigData. 4. Основные методы поиска BigData из различных источников. 5. Основные понятия о способах хранения информационных массивов. 6. Отличительная особенность NoSQL. 7. Назовите основные пять характеристик, присущие BigData. 8. Какие задачи может решать анализ BigData? 9. Назовите принципы разработки параллельных алгоритмов. 10. Свойства больших данных и ограничения RDBMS 11. Какие существуют базовые принципы обработки BigData? 12. Укажите, какой национальный стандарт описывает архитектуру bigdata? <p>Задание. Каждый месяц в сети Facebook выкладывается в открытый доступ 30 млрд новых источников информации. Определите, какой необходим объем для хранения полученных данных за год. Опишите</p>

		возможные модели и методы для хранения полученных массивов данных большого объема.
ОПК-1.2	Владеет современными информационными технологиями	<p>Задание. Провести автоматизированный процесс сбора данных с сайтов по определенным критериям для сбора контента в заданной предметной области. Организовать парсинг с помощью бесплатных сервисов и записать данные в формате CSV. Сервисы, с помощью которых можно осуществлять сбор данных:</p> <ul style="list-style-type: none"> • Import.io • Webhose.io • Dexi.io (ранее CloudScrape) • Scrapinghub • VisualScaper • ParseHub • Spinn3r • Xmldatafeed • Диггернаут Catalogloader□ и др. <p>При автоматизированном сборе информации необходимо соблюдать действующее законодательство РФ, в котором установлены следующие ограничения, имеющие отношение к сети интернет: не допускается нарушение авторских и смежных прав; не допускается неправомерный доступ к охраняемой законом компьютерной информации; не допускается сбор сведений, составляющих коммерческую тайну, незаконным способом; не допускается использование гражданских прав в целях ограничения конкуренции.</p> <p>Задание. Имеется достаточно большой набор данных (файл в формате CSV, см. задание 1), который необходимо загрузить в ЭТ и преобразовать в табличное представление.</p> <p>Задание. Загрузить большие объемы данных различных форматов (CSV и JSON) в JupyterNotebook. Объединить данные из нескольких таблиц. Произвести сортировку и вычисления с помощью оконных функций (для</p>

		<p>вычисления диапазона значений использовать SUM). Примечание: потоки данных большого объема могут генерировать датчики, установленные на авиадвигателе - около 10 Тб за полчаса. Такие же потоки характерны для буровых установок и нефтеперерабатывающих комплексов и т.п.).</p> <p>Задание. Имеются данные в таблице Excel. Преобразуем эти данные в удобный для компьютеров формат с помощью Python-библиотеки Pandas.</p> <p>Загрузите файл с данными себе на компьютер. В Google Colab откройте файл trips_data.xlsx.</p> <p>Выведите первые строки таблицы, используя команду trips_data.head().</p> <p>Выведите последние строки таблицы.</p> <p>Отобразите статистическую информацию по выбранному файлу с данными. Постройте диаграмму по столбцу с заработной платой.</p> <p>Задание. Постройте гистограмму по возрастам клиентов. Людей какого возраста больше всего среди клиентов компании?</p> <p>Задание. Выведите статистику по тому, какое количество людей предпочитает какой вид транспорта. Постройте гистограмму по этим данным.</p> <p>Задание. Создайте и обучите модель данных. На вход модели подаются данные о клиенте, программа выдает наиболее вероятный город направления.</p>
ОПК-1.3	Применяет средства обеспечения информационной безопасности	<p>Задание. Представьте витрину данных, которая будет описывать источники данных, используемые учреждением или организацией. Составьте реестр исходных данных с описанием обменных форматов и передачу данных</p> <p>Задание. Опишите процесс анализа IP- и MAC-адресов с помощью модуля cyberpandas</p>

б) Порядок проведения промежуточной аттестации, показатели и критерии оценивания:

Промежуточная аттестация по дисциплине «Основы Data инжиниринга» включает компьютерное тестирование, позволяющее оценить уровень усвоения обучающимися знаний, и практическое задание, выявляющие степень сформированности умений и владений; проводится в форме зачета.

Показатели и критерии оценивания зачета:

- на оценку «зачтено» – обучающийся должен набрать не менее 50% баллов при прохождении компьютерного тестирования, показав знание теоретического материала на уровне воспроизведения и объяснения информации, а также выполнить лабораторную работу, продемонстрировав умения и навыки решения стандартных задач.
- на оценку «не зачтено» – обучающийся не демонстрирует знание теоретического материала на уровне воспроизведения и объяснения информации, набрав на компьютерном тестировании менее 50% баллов, а также не может выполнить лабораторную работу и не может показать интеллектуальные навыки решения простых задач.

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ

Рекомендации направлены на оказание методической помощи студентам при выполнении лабораторных занятий.

Лабораторное занятие – это занятие, проводимое под руководством преподавателя в учебной аудитории (компьютерном классе университета или учебной специализированной лаборатории университета), направленное на углубление научно-теоретических знаний и получение лабораторных навыков решения типовых и прикладных задач.

Целью лабораторных занятий является формирование и отработка лабораторных умений и навыков, необходимых в последующей деятельности обучающихся.

Основными задачами лабораторных занятий являются:

- углубление уровня освоения общекультурных и профессиональных компетенций;
- обобщение, систематизация, углубление, закрепление полученных лабораторных знаний по конкретным темам дисциплин различных циклов;
- приобретение студентами умений и навыков использования современных теоретических знаний в решении конкретных прикладных задач;
- развитие профессионального мышления, профессиональной и познавательной мотивации.

Перечень тем лабораторных работ определяется рабочей программой дисциплины. План лабораторных занятий отвечает общей направленности лекционного курса и соотнесен с ним в последовательности тем.

Структура лабораторного занятия включает следующие компоненты: вступительная часть; ответы на вопросы обучающихся; практическая часть; заключительное слово преподавателя. Во вступительной части объявляется тема текущей лабораторной работы, ставятся ее цели и задачи, проводится инструктаж по технике безопасности выполнения работы, проверяется исходный уровень готовности студентов к лабораторной работе (выполнение тестов, контрольные вопросы и т.п.), выдается порядок и условия выполнения лабораторной работы.

На лабораторном занятии преподаватель может использовать разнообразные образовательные технологии (методы ИТ, работа в команде, case-study, проблемное обучение, учебные дискуссии и т.п.) по своему выбору для достижения качественного уровня обучения.

Правила по технике безопасности для обучающихся при проведении лабораторных работ

Общие правила:

1. Лабораторные работы проводятся под наблюдением преподавателя. К выполнению лабораторных работ студенты допускаются только после прослушивания инструктажа по технике безопасности, правилам поведения, противопожарным мерам в компьютерном классе и специализированных лабораториях.

2. Обучаемый должен строго выполнять правила техники безопасности и санитарно-гигиенические нормы при работе в компьютерных классах и специализированных лабораториях университета.

Порядок выполнения лабораторных работ

При подготовке к выполнению лабораторных работ студент должен повторить теоретический материал, необходимый для выполнения заданий по текущей теме.

Лабораторная работа выполняется каждым студентом самостоятельно, согласно индивидуальному заданию.

Студенты, пропустившие занятия, выполняют лабораторные работы во внеурочное время.

После выполнения каждой лабораторной работы студент демонстрирует результат выполнения преподавателю в виде отчета по лабораторной работе и отвечает на вопросы. Преподаватель оценивает работу в соответствии с заданными критериями оценки лабораторных работ.

Правила оформления результатов и оценивания лабораторной работы

Результаты выполненной лабораторной работы оформляются в соответствии с требованиями к выполнению конкретной работы.

Лабораторная работа считается выполненной, если студент выполнил ее в полном объеме, без существенных ошибок и сделал вывод по полученным результатам..

Для оценивания работы прилагаются следующие критерии.

Оценка «зачтено» – работа выполнена в полном объеме и без существенных ошибок, могут иметься 2-3 несущественные ошибки, исправленных самостоятельно по требованию преподавателя. Студент отвечает на поставленные вопросы, может объяснить основные положения по теме работы и интерпретировать полученные результаты.

Оценка «незачтено» – допущены существенные ошибки в ходе работы, которые студент не может исправить даже по требованию преподавателя, или работа не выполнена в полном объеме.