



## 1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины является формирование у обучающихся компетенций в соответствии с требованиями образовательного стандарта и в ознакомлении студентов с методами извлечения, трансформации и загрузки (ETL) больших объемов данных, а также в развитии навыков работы с современными инструментами и технологиями для эффективной обработки и анализа данных.

Задачи дисциплины включают изучение методов и инструментов для извлечения, трансформации и загрузки данных из различных источников в хранилища данных. Студенты также научатся проектировать и оптимизировать ETL-процессы для обеспечения высокой производительности и надежности обработки данных. Кроме того, курс предполагает освоение практических навыков работы с большими данными, включая их анализ, визуализацию и применение в бизнес-решениях.

## 2. Планируемые результаты обучения по дисциплине (модулю).

Перечень формируемых результатов освоения образовательной программы (компетенций) в результате обучения по дисциплине (модулю):

**ПК-2** - Способен осуществлять разработку и внедрение специального программного обеспечения цифровой обработки сигналов, цифрового программного управления на языках высокого и низкого уровней.

Обучение по дисциплине (модулю) предполагает, что по его результатам обучающийся будет:

### **Знать:**

- основные методы и принципы ETL-процессов и работы с большими данными.

### **Уметь:**

- применять современные инструменты для извлечения, трансформации и загрузки данных, а также оптимизировать эти процессы для повышения их эффективности.

### **Владеть:**

- навыками анализа и визуализации данных, что позволит им принимать обоснованные решения на основе полученной информации.

## 3. Объем дисциплины (модуля).

### 3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 4 з.е. (144 академических часа(ов)).

3.2. Объем дисциплины (модуля) в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Тип учебных занятий	Количество часов	
	Всего	Семестр №3
Контактная работа при проведении учебных занятий (всего):	32	32
В том числе:		
Занятия лекционного типа	16	16
Занятия семинарского типа	16	16

3.3. Объем дисциплины (модуля) в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации составляет 112 академических часа (ов).

3.4. При обучении по индивидуальному учебному плану, в том числе при ускоренном обучении, объем дисциплины (модуля) может быть реализован полностью в форме самостоятельной работы обучающихся, а также в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении промежуточной аттестации.

## 4. Содержание дисциплины (модуля).

### 4.1. Занятия лекционного типа.

№ п/п	Тематика лекционных занятий / краткое содержание
1	Введение в большие данные и их экосистему Рассматриваемые вопросы: Определение и характеристики больших данных Архитектуры и компоненты экосистемы больших данных Применение больших данных в бизнесе и науке
2	Глубокое понимание ETL-процессов. Рассматриваемые вопросы:

№ п/п	Тематика лекционных занятий / краткое содержание
	<p>Подробное изучение этапов ETL: извлечение, трансформация, загрузка</p> <p>Жизненный цикл ETL-процессов и их проектирование</p> <p>Сравнение ETL и ELT: когда и как использовать</p>
3	<p><b>Современные инструменты и технологии для извлечения данных.</b></p> <p>Рассматриваемые вопросы:</p> <p>Обзор ETL-инструментов (Apache NiFi, Talend, Informatica, Apache Kafka)</p> <p>Подключение к различным источникам данных (реляционные и нереляционные БД, API, потоковые данные)</p> <p>Практические примеры извлечения данных в реальных сценариях</p>
4	<p><b>Трансформация данных: методы и технологии</b></p> <p>Рассматриваемые вопросы:</p> <p>Подходы к трансформации данных: правила, агрегация, обогащение</p> <p>Стандартизация, очистка и валидация данных</p> <p>Использование SQL и других языков для трансформации данных.</p>
5	<p><b>Загрузка данных и работа с хранилищами</b></p> <p>Рассматриваемые вопросы:</p> <p>Типы хранилищ данных (OLAP, OLTP, Data Lakes, Data Warehouses)</p> <p>Оптимизация процессов загрузки данных: инкрементальная и полная загрузка</p> <p>Архитектурные решения для эффективного хранения и обработки данных</p>
6	<p><b>Мониторинг, управление и оптимизация ETL-процессов</b></p> <p>Рассматриваемые вопросы:</p> <p>Значение мониторинга и управления ETL-процессами</p> <p>Инструменты и методы для мониторинга (Apache Airflow, Luigi, Grafana)</p> <p>Обработка ошибок, восстановление после сбоев и обеспечение надежности</p>
7	<p><b>Анализ данных и визуализация для принятия решений</b></p> <p>Рассматриваемые вопросы:</p> <p>Методы анализа больших данных: статистические и машинные методы</p> <p>Инструменты визуализации данных (Tableau, Power BI, D3.js)</p> <p>Применение аналитики для выявления инсайтов и поддержки бизнес-решений</p>
8	<p><b>Будущее ETL и большие данные: тренды и инновации</b></p> <p>Рассматриваемые вопросы:</p> <p>Новые технологии в области ETL и больших данных (например, Serverless, Streaming ETL)</p> <p>Влияние искусственного интеллекта и машинного обучения на ETL-процессы</p> <p>Этические аспекты и безопасность данных в эпоху больших данных</p>

## 4.2. Занятия семинарского типа.

### Практические занятия

№ п/п	Тематика практических занятий/краткое содержание
1	<p><b>Практика извлечения данных</b></p> <p>Рассматриваемые вопросы:</p> <p>Настройка среды для работы с ETL-инструментами (например, Apache NiFi или Talend)</p> <p>Извлечение данных из различных источников (реляционные БД, REST API, CSV-файлы)</p> <p>Создание простого ETL-процесса для извлечения данных</p>
2	<p><b>Трансформация данных на практике.</b></p> <p>Рассматриваемые вопросы:</p> <p>Применение методов очистки и стандартизации данных</p>

№ п/п	Тематика практических занятий/краткое содержание
	Использование SQL для трансформации данных (JOIN, GROUP BY, CASE) Создание бизнес-правил для трансформации данных в зависимости от сценариев
3	Практические кейсы и проекты Рассматриваемые вопросы: - реализация ETL-процессов на практике; - решение типичных задач в индустрии.

#### 4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Самостоятельное изучение и конспектирование отдельных тем учебной литературы, связанных с разделами дисциплины
2	Работа с лекционным материалом
3	Подготовка к практическим занятиям
4	Подготовка к промежуточной аттестации.
5	Подготовка к текущему контролю.

5. Перечень изданий, которые рекомендуется использовать при освоении дисциплины (модуля).

№ п/п	Библиографическое описание	Место доступа
1	Баламирзоев, А. Г. Технологии и инфраструктура Big Data : учебно-методическое пособие / А. Г. Баламирзоев. — Махачкала : ДГПУ, 2024. — 51 с.	<a href="https://e.lanbook.com/book/442649">https://e.lanbook.com/book/442649</a>
2	Макшанов, А. В. Большие данные. Big Data / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкаръ. — 4-е изд., стер. — Санкт-Петербург : Лань, 2024. — 188 с. — ISBN 978-5-507-47346-5.	<a href="https://e.lanbook.com/book/362318">https://e.lanbook.com/book/362318</a>
3	Лебедев, А. С. Методы Big Data : учебно-методическое пособие / А. С. Лебедев, Ш. Г. Магомедов. — Москва : РТУ МИРЭА, 2021. — 91 с.	<a href="https://e.lanbook.com/book/182452">https://e.lanbook.com/book/182452</a>

6. Перечень современных профессиональных баз данных и информационных справочных систем, которые могут использоваться при освоении дисциплины (модуля).

Информационный портал Научная электронная библиотека eLIBRARY.RU ([www.elibrary.ru](http://www.elibrary.ru));

Единая коллекция цифровых образовательных ресурсов (<http://window.edu.ru>);

Научно-техническая библиотека РУТ (МИИТ) (<http://library.miit.ru>);  
Поисковые системы «Яндекс» для доступа к тематическим информационным ресурсам;  
Электронно-библиотечная система издательства «Лань» – <http://e.lanbook.com/>;  
Электронно-библиотечная система [ibooks.ru](http://ibooks.ru) – <http://ibooks.ru/>;  
Электронно-библиотечная система «УМЦ» – <http://www.umcздt.ru/>;  
Электронно-библиотечная система «BOOK.ru» – <http://www.book.ru/>;  
Электронно-библиотечная система «ZNANIUM.COM» – <http://www.znanium.com/>

7. Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства, необходимого для освоения дисциплины (модуля).

1. Операционная система windows microsoft office 2003 и выше;
2. Браузер Internet Explorer 8.0 и выше с установленным Adobe Flash player версии 10.3 и выше;
3. Adobe acrobat.

8. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю).

Учебные аудитории для проведения учебных занятий, оснащенные компьютерной техникой и наборами демонстрационного оборудования.

9. Форма промежуточной аттестации:

Зачет в 3 семестре.

10. Оценочные материалы.

Оценочные материалы, применяемые при проведении промежуточной аттестации, разрабатываются в соответствии с локальным нормативным актом РУТ (МИИТ).

Авторы:

старший преподаватель кафедры  
«Системы управления транспортной  
инфраструктурой»

А.Н. Малых

Согласовано:

Директор

Б.В. Игольников

Руководитель образовательной  
программы

А.С. Киселёва

Председатель учебно-методической  
комиссии

Д.В. Паринов