

МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ УНИВЕРСИТЕТ ТРАНСПОРТА»
(РУТ (МИИТ))



Рабочая программа дисциплины (модуля),
как компонент образовательной программы
высшего образования - программы бакалавриата
по направлению подготовки
09.03.02 Информационные системы и технологии,
утвержденной первым проректором РУТ (МИИТ)
Тимониным В.С.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Информационные хранилища и аналитические системы

Направление подготовки: 09.03.02 Информационные системы и технологии

Направленность (профиль): Информационные системы и технологии на транспорте

Форма обучения: Очная

Рабочая программа дисциплины (модуля) в виде электронного документа выгружена из единой корпоративной информационной системы управления университетом и соответствует оригиналу

Простая электронная подпись, выданная РУТ (МИИТ)
ID подписи: 5665
Подписал: заведующий кафедрой Нутович Вероника Евгеньевна
Дата: 08.04.2024

1. Общие сведения о дисциплине (модуле).

Целью освоения учебной дисциплины (модуля) «Информационные хранилища и аналитические системы» является получение студентами знаний о принципах построения информационных хранилищ и информационно-аналитических систем (ИАС) на основе хранилищ данных, этапах построения хранилищ данных, моделях анализа и представления данных в ИАС, а также об использовании информационных хранилищ и ИАС на железнодорожном транспорте.

В результате изучения дисциплины студенты должны получить необходимые знания об основах теории хранилищ данных, технологии OLAP и интеллектуального анализа данных, о примерах аналитических систем, построенных на изученных принципах и используемых на транспорте. Кроме того, студенты должны овладеть базовыми навыками манипуляции с наборами данных с помощью языка Python и библиотеки Pandas, а также знаниями по построению информационных хранилищ для целей поддержки принятия решений и обеспечения функционирования аналитических систем.

Дисциплина предназначена для получения знаний и решения следующих профессиональных задач (в соответствии с видами деятельности):

Проектно-конструкторская деятельность:

- предпроектное обследование (инжиниринг) объекта проектирования, системный анализ предметной области, их взаимосвязей;
- техническое проектирование (реинжиниринг);
- рабочее проектирование;
- выбор исходных данных для проектирования.

Научно-исследовательская деятельность:

- сбор, анализ научно-технической информации отечественного и зарубежного опыта по тематике исследования.

2. Планируемые результаты обучения по дисциплине (модулю).

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

ОПК-2 - Способен понимать принципы работы современных информационных технологий и программных средств, в том числе отечественного производства, и использовать их при решении задач профессиональной деятельности;

ПК-6 - Способен разрабатывать структурные компоненты баз данных как составной части информационной системы, включая развертывание, сопровождение, оптимизацию функционирования.;

ПК-12 - Способен проектировать системы поддержки принятия решений при управлении транспортным комплексом.

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

Уметь:

-применять современные разработки и тенденции в области проектирования ПО в профессиональной деятельности;

-принимать обоснованный выбор типа хранилища и его архитектуры;

-проектировать витрины данных и области хранилища данных, оптимизировать их структуру с учётом особенностей решаемой задачи.

Знать:

-роль данных и аналитики в бизнес-процессах предприятия;

-назначение и отличия баз данных и информационных хранилищ;

-классические и современные подходы к проектированию хранилищ данных;

-приёмы проектирования моделей данных, обеспечивающие различные свойства информационных хранилищ

Владеть:

-навыками работы с инструментами интеграции и управления данными;

-навыками построения ETL-процессов.

3. Объем дисциплины (модуля).

3.1. Общая трудоемкость дисциплины (модуля).

Общая трудоемкость дисциплины (модуля) составляет 5 з.е. (180 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 1
Контактная работа при проведении учебных занятий (всего):	70	70
В том числе:		

Занятия лекционного типа	40	40
Занятия семинарского типа	30	30

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 110 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	Эволюция архитектуры аналитических систем. Рассматриваемые вопросы: - эволюция архитектур аналитических систем; - единая база данных; - ETL; - витрины данных; - OLTP и OLAP.
2	Хранилище данных как класс информационных систем. Рассматриваемые вопросы: - роль хранилища данных в бизнес-процессах предприятия; - определение хранилища данных/ и его назначение; - отличия от базы данных; - нормализация и денормализация схемы данных.
3	Классические подходы к проектированию хранилища. Многомерность данных. Рассматриваемые вопросы: - многомерное представление данных; - способы представления многомерных данных; - классические модели данных «Звезда» и «Снежинка»; - тест FASMI; - базовые операции OLAP; - «Созвездия».

№ п/п	Тематика лекционных занятий / краткое содержание
4	<p>Архитектурные подходы к проектированию хранилища.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - корпоративная информационная фабрика, подход к проектированию хранилища данных по Биллу Инмону; - хранилище данных с архитектурой шины, подход к проектированию хранилища данных по Ральфу Кимбаллу, ELT.
5	<p>Классификация источников данных для хранилища и способы интеграции.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - плоские файлы различных форматов, переключники, репозитории хранения; - реляционные СУБД; - On-line сообщения и менеджеры очередей; - web-сервисы; - неструктурированные источники информации; - витрины данных хранилища; - ручной ввод данных.
6	<p>Области хранилища данных. Структура ETL-процессов.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - задачи ETL-процессов; - состояние данных на разных этапах ETL; - инструменты построения и выполнения ETL-процессов.
7	<p>Современные тенденции в области проектирования хранилищ данных. Гибкие методологии построения детального слоя данных.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - подход к организации хранилища в виде озера данных (Data Lake); - проектирование хранилища данных с применением подхода Data Vault; - проектирование хранилища данных с якорной моделью (Anchor Modeling).
8	<p>Некоторые шаблоны проектирования моделей данных хранилища.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - модель Entity-Attribute-Value; - иерархия и классификаторы; - типы медленно меняющихся измерений; - бизнес-история и техническая история данных; - оптимизация доступа к быстро растущим измерениям.
9	<p>Способы обеспечения необходимого уровня производительности аналитических систем.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - Требования, предъявляемые к аналитическим системам и системам обеспечения данными; - In-memoгу-вычисления, особенности работы In-memory систем; - Обзор разновидностей СУБД.
10	<p>Принципы функционирования технологий обеспечения необходимого уровня производительности аналитических систем.</p> <p>Рассматриваемые вопросы:</p> <ul style="list-style-type: none"> - колоночные СУБД; - распределённые вычисления в Apache Hadoop и Apache Spark.

№ п/п	Тематика лекционных занятий / краткое содержание
11	Интеллектуальный анализ данных в процессе поддержки принятия решений предприятия. Рассматриваемые вопросы: - подразделения, работающие с системами-источниками; - поддержка работы информационных хранилищ; - отделы анализа данных; - лица, принимающие решения.
12	Обеспечение данными аналитических систем группы корпоративного управления. Рассматриваемые вопросы: - примеры систем группы управления на транспорте; - аналитические задачи систем управления рисками; - данные, необходимые для функционирования системы управления рисками.
13	Обеспечение данными аналитических систем группы управления производственным процессом. Рассматриваемые вопросы: - примеры систем группы управления производственными процессами на транспорте; - аналитические задачи систем оперативного управления; - данные, необходимые для функционирования систем оперативного управления.
14	Обеспечение данными аналитических систем группы управления закупками. Рассматриваемые вопросы: - примеры систем группы управления взаимодействием с поставщиками на транспорте; - аналитические задачи систем управления взаимодействием с поставщиками; - данные, необходимые для функционирования системы управления взаимодействием с поставщиками.
15	Обеспечение данными аналитических систем группы управления продажами. Рассматриваемые вопросы: - примеры систем группы управления продажами на транспорте; - аналитические задачи систем управления продажами и маркетинга; - данные, необходимые для функционирования системы управления продажами.
16	Обеспечение качества данных, поставляемых в аналитические системы. Рассматриваемые вопросы: - причины снижения качества данных; - организационные меры повышения качества данных; - программные методы повышения качества данных; - системы класса Master Data Management, «золотая запись».
17	Этические и моральные аспекты создания и эксплуатации аналитических систем. Рассматриваемые вопросы: - искусственный интеллект, заслуживающий доверия; - кодекс этики в сфере искусственного интеллекта в России; - этические рекомендации экспертной группы альянса искусственного интеллекта ЕС.

4.2. Занятия семинарского типа.

Лабораторные работы

№ п/п	Наименование лабораторных работ / краткое содержание
1	Импорт и интеграция данных. В результате выполнения задания лабораторной работы студент получает навыки проектирования витрин данных по классическим схемам «звезда» и «снежинка», работы с технологиями обработки

№ п/п	Наименование лабораторных работ / краткое содержание
	данных, импорта данных из файлов различного формата средствами языка Python и библиотеки Pandas.
2	Технология построения аналитических отчётов. В результате выполнения задания лабораторной работы студент получает навыки работы с технологиями аналитической обработки данных средствами языка Python и библиотеки Pandas, построению списочных, статистических и графических отчётов и их экспорта в различных форматах.

4.3. Самостоятельная работа обучающихся.

№ п/п	Вид самостоятельной работы
1	Подготовка к выполнению лабораторных работ [1],[3]. Изучение файлов источников данных в лабораторной работе №1 по индивидуальному варианту. Составление технической характеристики файлов с исходными данными.
2	Изучение различных видов источников данных, и особенностей работы с каждым из них.
3	Распределение по областям информационного хранилища наборов данных из лабораторной работы (по индивидуальному варианту).
4	Подготовка отчёта по лабораторной работе №1.
5	Ознакомление с инструментом проектирования хранилища в Anchor –модели.
6	Самостоятельное изучение возможностей low-code инструмента Loginom Community Edition (https://loginom.ru/platform/quick-start).
7	Самостоятельное изучение тем «Колоночные базы данных», «Базы данных NoSQL» [5] , самостоятельное изучение программных продуктов колоночных и NoSQL СУБД по произвольным материалам.
8	Подготовка отчёта по лабораторной работе №2.
9	Самостоятельное изучение видеоматериала «Шаблоны проектирования моделей данных в хранилище».
10	Изучение разделов, посвященным анализу рисков, годовых отчётов отечественных компаний.
11	Практическое упражнение «Анализу рисков учебного процесса».
12	Изучение документа "Кодекс этики в сфере ИИ", разработанный российским Альянсом в сфере искусственного интеллекта.
13	Подготовка к промежуточной аттестации.
14	Подготовка к текущему контролю.

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

№ п/п	Библиографическое описание	Место доступа
----------	----------------------------	---------------

1	Бережной, А. Н. Сохранение данных: теория и практика / А. Н. Бережной. — Москва : ДМК Пресс, 2016. — 317 с. — ISBN 978-5-97060-185-3.	URL: https://e.lanbook.com/book/82823
2	Цехановский, В. В. Управление данными : учебник / В. В. Цехановский, В. Д. Чертовской. — Санкт-Петербург : Лань, 2021. — 432 с. — ISBN 978-5-8114-1853-4.	URL: https://e.lanbook.com/book/168835
3	Макшанов, А. В. Системы поддержки принятия решений : учебное пособие / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. — Санкт-Петербург : Лань, 2020. — 108 с. — ISBN 978-5-8114-5344-3.	URL: https://e.lanbook.com/book/147135
4	Григорьев, Ю. А. Реляционные базы данных и системы NoSQL : учебное пособие / Ю. А. Григорьев, А. Д. Плутенко, О. Ю. Плужникова. — Благовещенск : АмГУ, 2018. — 424 с. — ISBN 978-5-93493-308-2.	URL: https://e.lanbook.com/book/156492
5	Нечитайло, Н.М. Программирование на языке Python. Часть 1. / Н.М. Нечитайло Учебное пособие, 320с. - М.: РУТ (МИИТ), 2021.	URL: https://miitasu.ru/
6	Кодекс этики в сфере искусственного интеллекта. — Москва: Альянс в сфере искусственного интеллекта, 2019 — 18 с.	URL: https://ethics.a-ai.ru/ (Дата обращения 26.05.2023 г.)
7	Ethics guidelines for trustworthy ai. — Brussels: High-Level Expert Group on Artificial Intelligence. European Commission, 2019.	URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (Дата обращения 26.05.2023 г.)

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

Официальная документация Pandas (URL: <https://pandas.pydata.org/pandas-docs/stable/reference/frame.html>).

Электронно-библиотечная система издательства Лань (URL: <https://e.lanbook.com>).

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

Google CoLab, Блокноты Jupyter или локально установленная среда разработки на языке Python – свободно распространяемое программное обеспечение.

Anchor Modeler <https://www.anchor modeling.com/modeler/latest/>.

Loginom Community Edition (<https://loginom.ru/platform/quick-start>) – свободно распространяемое программное обеспечение для некоммерческого использования.

Microsoft Office, включающий MS Word, MS Excel, MS Access, MS Visio – лицензионное программное обеспечение.

Интернет-браузер - свободно распространяемое программное обеспечение.

Улучшенный текстовый редактор с возможностью управления кодировками, отображением непечатаемых символов, подсветкой синтаксиса и разметки (например: Notepad++, EmEditor, PSPad, Atom) - свободно распространяемое или лицензионное программное обеспечение.

При организации обучения по дисциплине (модулю) с применением электронного обучения и дистанционных образовательных технологий необходим доступ каждого студента к информационным ресурсам – библиотечному фонду Университета, сетевым ресурсам и информационно-телекоммуникационной сети «Интернет».

В случае проведения занятий с применением электронного обучения и дистанционных образовательных технологий может потребоваться наличие следующего программного обеспечения (или их аналогов): ОС Windows, Microsoft Office, Интернет-браузер, Microsoft Teams и т.д.

В образовательном процессе, при проведении занятий с применением электронного обучения и дистанционных образовательных технологий, могут применяться следующие средства коммуникаций: ЭИОС РУТ(МИИТ), Microsoft Teams, электронная почта, скайп, Zoom, Telegram и т.п.

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.

В случае проведения занятий с применением электронного обучения и дистанционных образовательных технологий необходимо наличие компьютерной техники, для организации коллективных и индивидуальных форм общения педагогических работников со студентами, посредством используемых средств коммуникации.

9. Форма промежуточной аттестации:

Экзамен в 8 семестре.

10. Оценочные материалы.

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

Авторы:

доцент кафедры «Цифровые
технологии управления
транспортными процессами»

П.О. Козьяков

Согласовано:

Заведующий кафедрой ЦТУТП

В.Е. Нутович

Председатель учебно-методической
комиссии

Н.А. Андриянова