

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт экономики и менеджмента

УТВЕРЖДАЮ:
Директор Института
экономики и менеджмента
_____ Е.В. Нехода

« 20 » _____ 04 20 23 г.

Рабочая программа дисциплины

Управление данными

по направлению подготовки

38.04.01 Экономика

Направленность (профиль) подготовки:
«Экономика»

Форма обучения
Очная

Квалификация
Магистр

Год приема
2023

СОГЛАСОВАНО:
Руководитель ОП
_____ Н.А. Скрыльникова

Председатель УМК
_____ М.В. Герман

Томск – 2023

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

– ПК-2 – Способен разрабатывать стратегии управления изменениями в организации.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ПК-2.2. Определяет основные параметры и ключевые показатели эффективности разрабатываемых стратегических изменений в организации;

2. Задачи освоения дисциплины

– Освоить понятийный аппарат и инструментарий управления данными для принятия решений.

– Научиться применять понятийный аппарат и инструментарий управления данными при решении практических задач профессиональной деятельности.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, является обязательной для изучения. Относится к профессиональному модулю «Бизнес аналитика».

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Семестр 3, экзамен.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины требуются результаты обучения по следующим дисциплинам: «Эконометрика», «Финансово-экономический анализ для бизнес-решений», «Python и R для анализа данных», «Анализ и прогнозирование временных рядов», «Большие данные», «Интеллектуальный анализ данных»

6. Язык реализации

Русский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 6 з.е., 216 часов, из которых:

– лекции: 18 ч.;

– практические занятия: 28 ч.;

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. Основные понятия базы данных.

Понятие модели представления данных. Уровни модели данных (внешний, концептуальный, физический). Уровни модели данных в базе данных: экстенциональный и интенциональный (метаданные). Понятие реляционной модели данных (отношение, схема отношения, запись, поле). Ключи (возможный, первичный, внешний, суррогатный). Представление типов объектов и типов связей в реляционной модели. Понятие нормализации реляционной модели. Постреляционная (расширенная реляционная) модель данных.

Тема 2. Основные особенности хранилища данных (Data Warehouse).

Понятие хранилища данных (Data Warehouse). Назначение хранилища данных, его отличие от базы данных. Источники данных для хранилища. Основные принципы организации хранилища данных. Основные процессы работы с данными в хранилище.

Тема 3. Основные особенности OLAP-технологии.

Понятие OLAP. Основная цель OLAP-средств. Понятия измерения, метки и факта, способы определения фактов. Многомерная модель данных, понятие куба. Поликубическая и гиперкубическая модель данных. Операции, применяемые к кубам (агрегация, детализация, формирования среза, вращение).

Тема 4. Структура хранилища данных.

Состав хранилища данных (таблицы измерений, таблица фактов). Типовые структурные схемы (схема звезды, схема снежинки), их достоинства и недостатки. Многоуровневые структуры данных для OLAP (сбалансированные, несбалансированные и неровные иерархии).

Тема 5. Основные особенности продуктов OLAP.

Требования, предъявляемые к продуктам OLAP (тест FASMI). Особенности OLAP-средств, встроенных в настольные приложения. Особенности сетевых OLAP-систем. Классификация OLAP-средств с точки зрения способа представления данных в хранилище (MOLAP, ROLAP, HOLAP). Достоинства и недостатки средств каждого класса. Физический OLAP и виртуальный OLAP.

Тема 6. Архитектурные концепции хранилища данных.

Виртуальное хранилище данных, его достоинства и недостатки. Аналитическая пирамида. Краткая характеристика элементов архитектуры хранилища данных. Оперативный склад данных. Витрины данных (принципы построения, источники данных, виды витрин, преимущества и недостатки). Метаданные.

Тема 7. Виды архитектур хранилища данных.

SIF –архитектура ("корпоративная информационная фабрика"). Хранилище данных с архитектурой шины BUS. Гибридная многоуровневая архитектура. Слоистая архитектура хранилища данных. Достоинства и недостатки архитектур каждого вида.

Тема 8. Средства управления хранилищем данных.

Типовая структура программной системы управления хранилищем данных (подсистема загрузки данных, подсистема представления данных и обработки запросов, подсистема администрирования хранилища). Технология ETL: назначение, процессы, реализуемые в ETL-системах. Особенности выполнения этапа извлечения данных из источников. Особенности выполнения этапа преобразования данных (перевод значений, внешнее и внутреннее обогащение, очистка и агрегирование данных). Особенности выполнения этапа загрузки данных в хранилище.

Тема 9. Основные особенности Deductor Studio.

Основные понятия Deductor Warehouse: семантический слой, измерение, ссылка, атрибут, процесс. Структура программного окна Deductor Studio; панель управления и область визуализаторов; подключения и сценарии; основные компоненты для работы в системе (мастера подключений, импорта, экспорта, обработки и визуализации), последовательность работ.

Тема 10. Создание хранилища данных.

Понятие проекта хранилища данных. Создание типовой структуры хранилища данных, формирование метаданных, создание сценариев. Модификация структуры хранилища данных (очистка процессов и измерений, удаление измерения и процесса в хранилище, добавление атрибута, факта и измерения в процесс).

Импорт данных. Очистка данных (фильтрация, выявление дубликатов и противоречий, обнаружение и редактирование аномалий, сглаживание) и загрузка импортированных данных в хранилище данных.

Тема 11. Извлечение данных из хранилища данных.

Создание сценариев импорта измерений с формированием таблиц, кубов, срезов. Назначение и виды визуализаторов, их возможности и использование. Создание сценариев импорта из процесса с формированием таблиц и кубов.

Тема 12. Построение OLAP-куба в хранилище данных.

Создание куба в хранилище данных, импорт куба из хранилища данных и построение его среза. Отличие кубов в хранилище данных и в сценарии. Использование кросс-диаграмм и кросс-таблиц.

Тема 13. Использование OLAP-куба для оперативного анализа данных.

Настройка кросс-таблицы и куба, фильтрация в таблице и кубе, преобразование даты и времени, настройка набора данных, разбиение области значений непрерывного числового параметра на интервалы (квантование), замена данных, добавление полей, значения которых вычисляются из значений других полей. Объединение (слияние) данных.

Тема 14. Построение модели и ее использование для прогнозирования.

Проверка сезонности. Подготовка данных для построения модели. Построение пользовательской модели и ее использование для прогнозирования. Построение стандартных моделей (на примере линейной регрессии и нейронной сети) и их использование для прогнозирования.

Тема 15. Выявление закономерностей между связанными событиями и их использование для оперативного анализа данных.

Ассоциативные правила. Параметры, характеризующие ассоциативные правила (поддержка, достоверность, лифт). Настройка параметров поиска ассоциативных правил. Поиск ассоциативных правил. Визуализаторы, используемые для оперативного анализа ассоциативных правил («правила», «популярные наборы», «что-если», «дерево правил»). Интерпретация ассоциативных правил.

Тема 16. Использование ранее созданных моделей для анализа новых данных (скрипты).

Скрипты (определение термина, настройка скрипта). Использование в сценариях ранее созданных ветвей обработки данных. Использование ранее созданных моделей для анализа новых данных и прогнозирования с помощью скрипта.

9. Текущий контроль по дисциплине

Текущий контроль приобретения компетенций осуществляется путем контроля посещаемости, проверки уровня усвоения знаний, с помощью блиц-опросов, выполнения индивидуальных заданий и контрольных работ, и фиксируется в форме контрольной точки не менее одного раза в семестр. Индивидуальные задания и контрольные работы

обучающийся выполняет, основываясь исключительно на знаниях и навыках, полученных во время аудиторных занятий и самостоятельной работы, без использования справочных и других подобных материалов.

Выполнение заданий текущего контроля является обязательным, поскольку демонстрирует уровень освоения дисциплины и позволяет студенту подготовиться к итоговому контролю.

Вклад результатов текущего контроля в итоговой оценке по дисциплине составляет – 60 баллов (60%).

10. Порядок проведения и критерии оценивания промежуточной аттестации

Экзамен в третьем семестре проводится в письменной форме по билетам. Продолжительность экзамена 1 час.

Примерный перечень теоретических вопросов.

1. Основные принципы организации хранилища данных.
2. Основные процессы работы с данными в хранилище.
3. Основные особенности OLAP.
4. Многомерная модель данных, понятие куба.
5. Агрегирование данных, многоуровневые структуры для OLAP.
6. Краткая характеристика элементов архитектуры хранилища данных.
7. Виртуальное хранилище данных, его достоинства и недостатки.
8. Достоинства и недостатки архитектур хранилища данных разного вида.
9. Типовая структура программной системы управления хранилищем данных.
10. Назначение технологии ETL, процессы, реализуемые в ETL-системах.

Критерии выставления баллов за письменный экзамен:

Баллы	Характеристика
40 баллов	Даны полные и развернутые ответы на вопрос.
30 баллов	Даны неполные ответы на вопрос.
20 баллов	Даны фрагментарные ответы на вопрос.
10 баллов	Даны ограниченные ответы на вопрос.

Итоговая оценка по дисциплине складывается из результатов текущего контроля (60%) и результатов промежуточной аттестации (40%) и составляет максимум 100 баллов.

Результаты экзамена определяются оценками «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». Механизм перевода результатов балльно-рейтинговой системы пятибалльную шкалу:

Баллы	Итоговая оценка
85-100 баллов	«отлично»
70-84 балла	«хорошо»
55-69 баллов	«удовлетворительно»
54 балла и менее	«неудовлетворительно»

11. Учебно-методическое обеспечение

а) Электронный учебный курс по дисциплине в электронном университете «Moodle» - <https://moodle.tsu.ru/course/view.php?id=16474>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.

Примерный перечень теоретических вопросов, проверяющих ПК-2.2:

1. Понятие проекта хранилища данных.
2. Семантический слой в хранилище данных.
3. Последовательность выполнения работ с хранилищем данных.
4. Создание типовой структуры хранилища данных.
5. Формирование метаданных.
6. Загрузка данных в хранилище.
7. Создание куба в хранилище данных.
8. Подготовка данных для построения модели.
9. Построение пользовательской модели.
10. Построение модели нейронной сети.
11. Прогнозирование с помощью нейронной сети.
12. Ассоциативные правила.
13. Использование в сценариях ранее созданных ветвей обработки данных.

Примерный перечень практических задач, проверяющих ПК-2.2

Задача № 1. Создание хранилища данных «Кредитование физических лиц».

1. Проектирование структуры хранилища данных.

Исходные данные представлены в трех таблицах:

1) КЛИЕНТЫ(КлиентКод, Имя клиента, Возраст, Пол, Образование, Стаж работы, ОбеспеченностьКод).

Здесь *КлиентКод* – измерение, и *ОбеспеченностьКод* – ссылка, а остальные – атрибуты.

2) ОБЕСПЕЧЕННОСТЬ КЛИЕНТОВ(ОбеспеченностьКод, Месячный доход, Жилая недвижимость в собственности, Стоимость недвижимости, Наличие кредитов, Возврат кредита).

Здесь *ОбеспеченностьКод* – измерение, остальные – атрибуты.

3) КРЕДИТЫ(КредитКод, КлиентКод, Цель кредитования, Срок кредита, Сумма кредита)

Здесь *КредитКод*, *КлиентКод*, *Цель кредитования* – измерения, *Срок кредита* и *Сумма кредита* – факты.

2. Создание структуры хранилища данных.

1) Файл пустого хранилища данных: *credit.gdb*.

2) Имя, метка и описание ХД

Имя: *CREDITS*
 Метка: *Кредитование*
 Описание: *Кредитование физических лиц*

3. Формирование метаданных.

1) Измерения.

Измерение	Имя	Метка	Описание	Тип данных
ОбеспеченностьКод	Provide_ID	ОбеспеченностьКод	Код обеспеченности	Целый
КлиентКод	Client_ID	КлиентКод	Код клиента	Целый
КредитКод	Credit_ID	КредитКод	Код кредита	Целый
Цель кредитования	Aim_ID	Цель кредитования		Строковый

Для них атрибуты:

Измерение	Имя	Метка	Описание	Тип данных
ОбеспеченностьКод	Prov_Income	Доход	Месячный доход	Вещественный
	Prov_Home	Жилье	Жилая недвижимость в собственности	Строковый
	Prov_Cos	СтоимЖилья	Стоимость недвижимости	Вещественный
	Prov_AddCredit	НалКредитов	Наличие кредитов	Строковый
	Prov_RetCredit	ВозвратКредита	Возврат кредита	Целый
КлиентКод	Client_Name	КлиентИмя	Имя клиента	Строковый
	Client_Age	КлиентВозраст	Возраст клиента	Целый
	Client_Sex	КлиентПол	Пол клиента	Строковый
	Client_Edu	Образование	Образование клиента	Строковый
	Client_Work	СтажРаботы	Стаж работы клиента	Вещественный

Для *КлиентКод* – ссылка *ОбеспеченностьКод*.

2) Процесс.

Измерения: *КредитКод*, *КлиентКод*, *Цель кредитования*.

Факты (вид – непрерывный):

Имя	Метка	Описание	Тип Данных
F_Interval	Срок	Срок кредита	Вещественный
F_Sum	Сумма	Сумма кредита	Вещественный

Имя: *Proc_Credit*
 Метка: *Кредитование*
 Описание: *Кредитование физических лиц*

Задача № 2. Наполнение ХД «Кредитование физических лиц» первичной информацией и извлечение из него данных

1. Подготовка исходных данных.

Исходные данные представлены в трех таблицах файла *DataCredit.xls*:

1) КЛИЕНТЫ(КлиентКод, Имя клиента, Возраст, Пол, Образование, Стаж работы, ОбеспеченностьКод).

2) ОБЕСПЕЧЕННОСТЬ КЛИЕНТОВ(ОбеспеченностьКод, Месячный доход, Жилая недвижимость в собственности, Стоимость недвижимости, Наличие кредитов, Возврат кредита).

3) КРЕДИТЫ(КредитКод, КлиентКод, Цель кредитования, Срок кредита, Сумма кредита).

Произвести экспорт исходных данных из формата Excel в текстовый формат (с разделителем табуляции).

2. Импорт данных из текстовых файлов.

Метки сценариев:

Импорт из "Обеспеченность клиентов.txt"

Импорт из "Клиенты.txt"

Импорт из "Кредиты.txt"

3. Загрузка импортированных данных в ХД

Метки сценариев: *Загрузка в "Кредитование"*

4. Импорт измерений с формированием таблицы и сведений.

Метки сценариев:

Кредитование: Клиенты

Кредитование: Обеспеченность клиентов.

5. Импорт из процесса с формированием таблицы.

Измерения: Цель кредитования, Образование, Месячный доход, Стоимость недвижимости.

Факты: Срок кредита (Сумма) и Сумма кредита (Сумма)

Способы отображения: Таблица.

Метка сценария: Импорт из процесса «Кредитование»

Задача № 3. Создание и импорт куба в ХД «Кредитование физических лиц».

1. Создание куба в хранилище данных

Измерения: Цель кредитования, Пол.

Факты: Срок кредита (Среднее) и Сумма кредита (Среднее)

Имя: CubeCredit

Метка: Зависимость кредита от пола и цели

Динамическое обновление: установить

2. Импорт куба из хранилища данных

Способы отображения: *Таблица* и *Куб*.

Размещение измерений в кубе: Колонки – *Пол*, Строки – *Цель кредитования*.

Факты: *Срок кредита* и *Сумма кредита*

Варианты агрегации: *Среднее*

Вариант отображения: *Значение*

Метка сценария: *Импорт куба «Зависимость кредита от пола и цели»*

3. Изменение сценария импорта из процесса *Кредитование* с формированием куба.

Для импорта:

Измерения: *Цель кредитования, Образование, Месячный доход, Стоимость недвижимости.*

Факты: *Срок кредита, Сумма кредита*

Вариант агрегации: *Среднее*
Способы отображения: *Таблица и Куб*.
Для куба:
Назначения полей:
Измерения: *Цель кредитования, Образование, Месячный доход, Стоимость недвижимости*
Факты: *Срок кредита, Сумма кредита*
Размещение измерений в кубе:
Колонки – *Образование*, Строки – *Цель кредитования*.
Настройка фактов *Срок кредита* и *Сумма кредита*:
Вариант агрегации: *Среднее*
Вариант отображения: *Значение*
Метка сценария: *Импорт из процесса «Кредитование»*

в) Методические указания по организации самостоятельной работы студентов.

Самостоятельная работа магистрантов включает в себя:

- самостоятельную подготовку к занятиям по заявленным темам курса в соответствии с приведенным планом (содержание дисциплины) и литературой. Контроль выполнения производится на занятиях в блиц-опросах;
- самостоятельную подготовку к проблемным дискуссиям;
- самостоятельную работу в аудитории при ответах на вопросы, решении задач и разборе деловых ситуаций (кейсов). Контроль выполнения осуществляется сразу же при оценке полученных результатов;
- самостоятельное выполнение индивидуальных аналитических заданий. Контроль выполнения осуществляется в сроки, предусмотренные для сдачи индивидуальных заданий, которые оговариваются с магистрантами;
- самостоятельную подготовку к зачету. Контроль выполнения заключается в предоставлении итоговой оценки по итогам обучения.

Для закрепления теоретического материала предполагается самостоятельное выполнение заданий по каждой пройденной теме, что позволяет обратить внимание на наиболее сложные, ключевые и дискуссионные аспекты изучаемой темы, помочь магистрантам систематизировать и лучше усвоить пройденный материал. При выполнении заданий магистрант должен не просто воспроизводить полученные знания по заданной теме, но и творчески переосмыслить существующие подходы к пониманию тех или иных проблем, явлений, событий; продемонстрировать и убедительно аргументировать собственную позицию.

В ходе выполнения заданий магистрантам необходимо ознакомиться с вопросами, вынесенными на обсуждение, изучить соответствующий материал по бизнес-кейсам, дополнительную литературу.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Архипенков С.Я. Хранилища данных. От концепции до внедрения / С.Я. Архипенков, Д.В. Голубев, О.Б. Максименко, под ред. С. Архипенкова. – М. : Диалог-МИФИ, 2002. – 528 с.
 - Лакшманан В., Тайджани Д. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении. – СПб. : Питер, 2021. – 491 с.
 - Ralph Kimball, Margy Ross. The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling. – John Wiley and Sons, Ltd, 2013. – 600 p.
- Электронная версия публикации представлена в «Ebook Central(ProQuest)»:
<https://ebookcentral-proquest-com.ez.lib.tsu.ru/lib/tomskuniv->

ebooks/detail.action?docID=1313513&query=9781118530801

Ссылка с компьютеров сети ТГУ.

б) дополнительная литература:

- Балдин К.В. Уткин В.Б. Информационные системы в экономике : Учебник для вузов. – М. : Дашков и К, 2013. – 393 с.
- Горбенко А. О. Информационные системы экономики. – М. : БИНОМ, 2010. – 292 с.
- Чистов Д.В. Информационные системы в экономике. Учебное пособие.– М. : ИНФРА-М, 2015. – 234 с.
- Просто о больших данных / Гурвиц Джудит, Ньюджент Алан, Халпер Ферн, Кауфман Марсия. – М. : Эксмо, 2015. – 400 с.

в) журналы (периодические издания):

«Программные продукты и системы», «Информационные технологии в управлении», «Информационные технологии», «Информационные технологии и вычислительные системы», «КомпьютерПресс», «Engineering Economist», «Information Sciences», «Big Data and Advanced Analytics», «Science & Healthcare», «Information and Control Systems» .

г) ресурсы сети Интернет:

- Федоров А., Елманова Н. Введение в OLAP: часть 1. Основы OLAP [Электронный ресурс]. – URL: http://www.olap.ru/basic/OLAP_intro1.asp.
- Федоров А., Елманова Н. Введение в OLAP: часть 2. Хранилища данных [Электронный ресурс]. – URL: http://www.olap.ru/basic/olap_intro2.asp.
- Deductor Studio Academic [Электронный ресурс]. – М. : BaseGroup Labs. – URL: <https://basegroup.ru/deductor/components/studio>.
- Deductor. Руководство аналитика [Электронный ресурс]. – М. : BaseGroup Labs. – URL: <https://basegroup.ru/deductor/manual>.
- Loginom Skills. Станьте аналитиком данных без программирования. Электронные курсы [Электронный ресурс]. – М. : Loginom Company. – URL: <https://loginom.ru/skills>.
- OLAP.RU: Business intelligence - effective data mining & analysis [Электронный ресурс]. – URL: <http://www.olap.ru>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

– Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office OneNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook), Deductor Studio Academic;

– публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения практических занятий, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

15. Информация о разработчиках

Лещинский Борис Семенович, кандидат технических наук, доцент кафедры информационных технологий и бизнес-аналитики Института экономики и менеджмента Национального исследовательского Томского государственного университета.