

Программа учебной дисциплины «Теория баз данных»

Утверждена
Академическим советом ООП
Протокол № 2.03-09/2706-01
от «27» июня 2018г.

Автор	Бекларян Армен Леонович
Число кредитов	6
Контактная работа (час.)	56
Самостоятельная работа (час.)	172
Курс	1
Формат изучения дисциплины	без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целью освоения учебной дисциплины является формирование знаний, умений и навыков проектирования баз данных (БД), хранилищ данных (ХД), администрирования баз и хранилищ данных, разработки приложений (внешних интерфейсов) БД и ХД, использования средств многомерного анализа данных класса OLAP и др.

Задачи курса:

- изучить теорию реляционных баз данных;
- получить представление о реализации промышленных баз данных;
- овладеть навыками проектирования баз данных с использованием CASE-инструментов;
- выработать компетенции в области выполнения запросов к базам данных и проектирования бизнес-логики баз данных;
- хранилища данных (ХД) – причины возникновения;
- архитектура ХД;
- основные принципы проектирования ХД;
- проектирование и разработка процесса наполнения ХД;
- физическая модель ХД;
- использование ХД в процессе анализа данных.

В результате освоения дисциплины студент должен:

знать:

- основные понятия теории баз данных и хранилищ данных;
- подходы к построению БД и ХД и сферы их применимости;
- особенности реляционной модели БД и многомерной модели ХД;
- понятие предметной области и способы ее описания;
- теорию реляционных БД: основные понятия, реляционная алгебра и реляционное исчисление, теория нормализации;
- теорию многомерных ХД: схемы типа «звезда», «снежинка», метод Data Vault Modeling, ETL-процедуры, ROLAP и др.
- методологии ER-моделирования, проектирование БД и ХД на основе ER-моделей;
- классификацию и способы задания ограничений целостности;

- языки описания и манипулирования данными разных классов;
- особенности работы в распределенной многопользовательской среде;
- методы организации доступа к данным;
- архитектуру программных средств, работающих с БД и ХД.

уметь:

- описывать предметную область, выявлять сущности и связи между ними;
- проектировать структуру БД и ХД вручную и с использованием выбранного CASE-средства;
- использовать инструментальные средства СУБД при реализации модели БД и ХД;
- разрабатывать бизнес-логику работы с БД и ХД;
- строить запросы, используя различные языковые средства;
- использовать БД и ХД при построении отчетов и разработке приложений.
- разрабатывать WEB-приложения к БД и ХД.

владеть:

- навыками проектирования структуры БД и ХД;
- навыками формирования SQL-запросов к БД и MDX-запросов к ХД;
- навыками анализа данных.

Изучение дисциплины «Теория баз данных» базируется на следующих дисциплинах:

- Проектирование информационных систем;
- Системный анализ и проектирование.

Для освоения учебной дисциплины студенты должны знать концептуальные основы архитектуры предприятия, основные классы информационных систем управления бизнесом, лучшие практики и современные стандарты в сфере информационных технологий.

Также студенты должны владеть методами проектирования информационных систем, уметь систематизировать и обобщать информацию, разрабатывать конкретные предложения по результатам исследований, готовить справочно-аналитические материалы для принятия управленческих решений в сфере информационных технологий.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

- Банковская ИТ-инфраструктура;
- Современные методы принятия решения: алгоритмы обработки больших данных.

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1. Введение в курс

Структура курса, контрольные мероприятия, оценка за работу в семестре. Понятия данные, информация, управление данными. Аспекты изучения информации. Понятие банка данных (БНД) Предпосылки создания БНД. Преимущества и недостатки БНД. Компоненты БНД. База данных (БД) – ядро БНД. Программные средства БНД. Языковые средства современных СУБД. Классификация языковых средств. Языки четвертого поколения и их свойства. Технические средства БНД. Организационно-методические средства. Понятие СУБД. Функции СУБД. Выбор СУБД. Классификация БНД. Иерархические, сетевые БД, БД на инвертированных списках. OLAP и OLTP-системы. Хранилища данных. Смешанные БД. XML-базы данных. Тенденции развития БНД. Пользователи БНД.

Тема 2. Основы теории реляционных баз данных

Отличительные особенности реляционных моделей. Основные понятия: отношение, кортеж, атрибут, ключ (простой, составной; первичный, альтернативный). Домен. Связывание таблиц. Внешний ключ. Функциональные зависимости. Многозначные зависимости. Операции реляционной алгебры. Трёхзначная логика (3VL). Теория нормализации. Понятие 1НФ, 2НФ, 3НФ, 4НФ, 5НФ. Алгоритм нормализации. Достоинства и недостатки подхода.

Тема 3. Проектирование баз данных – логическое и физическое моделирование

Общие сведения о даталогическом моделировании. Особенности даталогических моделей. Влияние особенностей модели на проектные решения. Факторы, влияющие на проектирование БД. Критерии оценки проекта БД. Проектирование структуры базы на основе использования ER-моделей. Влияние нотации ER-модели на сферу ее использования и алгоритм проектирования. Алгоритм проектирования (алгоритм перехода от базовой ER-модели к структуре реляционной базы). ERwin Data Modeler. Уточнение терминологии. Интерфейс системы. Создание логических и физических моделей. Методология проектирования в среде ERwin Data Modeler. Методология построения физических моделей. Типы данных. Физическое хранение данных (файлы, файловые группы, страницы, экстенды, индексы, секции). Влияние нормализации/денормализации.

Тема 4. Целостность данных

Понятие целостности и ограничения целостности. Классификация ограничений целостности. Причины, вызывающие нарушение ограничений целостности. Описание ограничений целостности в CASE-средствах. Способы задания ограничений целостности в современных СУБД: процедурный и декларативный способы задания ограничений целостности. Ссылочная целостность (целостность связи).

Тема 5. Язык SQL

Введение в SQL. Стандарты. Диалекты. Составные части языка SQL (DDL, DML, DCL). Расширения языка. Визуальные построители запросов. Структура языка и синтаксис основных операторов SQL. Запросы и подзапросы. Управление выполнением сценария. Использование SQL для выборки данных из таблиц. Использование булевых и реляционных операторов создания более сложных предикатов. Использование специальных операторов в «условиях». Суммирование данных с помощью функций агрегирования. Форматирование результатов запросов. Использование множества таблиц в одном запросе. Вложенные запросы. Связанные подзапросы. Использование операторов EXIST, ANY, ALL, SOME. Использование предложения UNION. Ввод, удаление, изменение значений полей. Использование подзапросов с командами обновления. Создание таблиц, команда CREATE. Табличные и скалярные функции. Представления. Работа с метаданными на основе системных представлений. Динамическое формирование запросов. Планы выполнения запросов.

Тема 6. Триггеры и хранимые процедуры

Разработка триггеров БД. Условия выполнения триггера (триггерные события). Тело триггера. Использование триггеров для проверки корректности введенных данных и выполнения сложных ограничений целостности данных. Использование триггеров для обновления данных в смежных таблицах. Разработка хранимых процедур БД. Общее представление о языках разработки хранимых процедур (TSQL и PL/SQL).

Тема 7. Вывод информации из баз данных

Возможности генераторов отчетов современных СУБД. Задание формы и состава документа. Введение вычисляемых полей. Получение документов, включающих несколько степеней итогов. Получение документов на основе нескольких связанных файлов. Графическое оформление документа. Вывод документа на печать, на дисплей и в файл. Использование генераторов форм для получения выходных документов.

Тема 8. Архитектура ХД

Типовые архитектуры хранилищ данных, рассматривается глобальное хранилище данных, централизованное хранилище данных, распределенное хранилище данных, витрины данных, взаимосвязанные киоски данных, независимые киоски данных, корпоративная информационная фабрика, хранилище данных с архитектурой шины данных.

Тема 9. Основные принципы проектирования ХД

Инструменты (CASE-средства) для проектирования ХД. Методы проектирования Центрального хранилища данных. Метод многомерного моделирования: схема «Звезда», схема «Снежинка», преимущества и недостатки. Подход к построению ХД – Data Vault. Моделирование временных данных. Жизненный цикл разработки хранилища данных.

Тема 10. Распределенные БД

Понятие распределенных БД. Классификация БД. Дополнительные критерии оценки СУРБД. Централизованные и распределенные системы. Работа с базами данных в режимах «файл-сервер» и «клиент-сервер». Двух- и трехуровневые системы клиент-сервер. Особенности работы с базами данных в многопользовательском режиме. Обеспечение целостности БД в распределенных БД. Уровни изолированности в ANSI SQL. Механизм блокировок. Технологии тиражирования. Проектирование распределенных баз данных. Сетевые возможности современных СУБД. Особенности создания баз данных, функционирующих в локальных сетях. Особенности программирования обработки баз данных при многопользовательском режиме доступа. Проектирование приложений типа «клиент-сервер». Клиентское и серверное ПО. Особенности работы в гетерогенной среде. Стандарты интерфейсов. Проблемы, возникающие при параллельном доступе, и пути их решения.

Тема 11. Вычислительная парадигма MapReduce и концепция NoSQL

Вычисления некоторых наборов распределенных задач с использованием большого количества компьютеров, образующих кластер. Шаги Map и Reduce. Предварительная обработка входных данных и свёртка данных. Концепция параллелизма. Шаблоны доступа к данным, хеш-таблица, деревья, таксономия NoSQL, колоночные СУБД, bigtable.

Тема 12. Экосистема Apache

Разработка и выполнение распределённых программ, расширение вычислительных мощностей посредством добавления в кластер дополнительных узлов, технология Hadoop, распределённая файловая система HDFS (Hadoop Distributed File System), интеграция с NoSQL и MapReduce. Движки: MapReduce, Spark, Tez.

Тема 13. SQL-подобный язык запросов: Hive

Распределённая версионированная нереляционная СУБД HBase. Трёхмерные таблицы (строковый ключ, штамп времени и квалифицированное имя колонки). Консистентность за счёт write-ahead-log (WAL). Партиция и бакет.

III. ОЦЕНИВАНИЕ

Формами текущего контроля являются две контрольные работы. Каждая из форм текущего контроля оценивается по 10-балльной шкале. Общая оценка за текущий контроль (по 10-балльной шкале) рассчитывается по формуле:

$$O_{\text{текущий}} = 0,5 \cdot O_{\text{кр1}} + 0,5 \cdot O_{\text{кр2}},$$

где $O_{\text{кр1}}$ – оценка за первую контрольную работу;

$O_{\text{кр2}}$ – оценка за вторую контрольную работу.

При определении накопленной оценки (по 10-балльной шкале) самостоятельная вне-аудиторная работа не оцениваются. Поэтому накопленная оценка формируется из оценки за текущий контроль и оценки за работу на аудиторных занятиях, и рассчитывается по формуле:

$$O_{\text{накопленная}} = 0,8 \cdot O_{\text{текущий}} + 0,2 \cdot O_{\text{ауд}} + 0,0 \cdot O_{\text{сам.работа}},$$

где $O_{\text{текущий}}$ – оценка за текущий контроль;

$O_{\text{ауд}}$ – оценка за аудиторную работу;

$O_{\text{сам.работа}}$ – оценка за самостоятельную работу.

Оценка за аудиторную работу выставляется на основе пропорции посещаемости студента к общему числу проведенных занятий, исходя из максимума в 10 баллов.

Результирующая оценка (выставляется в диплом) формируется на основе итоговой оценки за экзамен (по 10-балльной шкале) и накопленной оценки. Результирующая оценка рассчитывается по формуле:

$$O_{\text{результ}} = 0,4 \cdot O_{\text{экз}} + 0,6 \cdot O_{\text{накопленная}},$$

где $O_{\text{экз}}$ – оценка за итоговый контроль (экзамен);

$O_{\text{накопленная}}$ – накопленная оценка.

При формировании результирующей оценки на основе весовых коэффициентов применяется арифметическое округление до целого числа. В случае точного равенства дробной части пяти десятым округление применяется в большую сторону.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

4.1 Содержание заданий контроля

Первая контрольная работа проверяет освоение материала по разделу «Базы данных». Каждый вариант содержит 4 задания. В первом задании требуется построить дата-логическую схему по описанию предметной области при этом ставится требование соблюдения ЗНФ. В рамках второй задачи требуется заполнение данными готовой базы данных. Третье задание требует выделения функциональных связей и построения дата-логической схемы в ЗНФ по данной 1НФ. Четвертое задание представляет из себя написание SQL запросов к готовой базе данных.

Вторая контрольная работа предполагает создание OLAP куба по заданному хранилищу и проведения анализа данных указанного типа (временные ряды, линейная регрессия, кластеризация и пр.).

Экзаменационная работа представляет из себя тест из, примерно, 25 вопросов закрытого типа.

4.2 Пример вопросов экзамена

1. Построить ER-модель для проектирования БД кафедры, содержащую информацию, необходимую для закрепления тем курсового проектирования. Исходными данными является:

- Выписка из учебного плана с указанием, какие КП (по каким предметам) на каких потоках и в каком семестре должны быть выполнены в данном учебном году.
- Какие преподаватели могут руководить КП по данному предмету.

- Плановое число КП по каждой дисциплине на каждом потоке для каждого преподавателя.
- Численность студентов в каждой студенческой группе, в том числе платных и бесплатных.
- Темы КП по каждой дисциплине.

2. В базе данных содержатся данные о нагрузке преподавателей:

NAGR (KODPREP, DATA, VIDNAGR, CHAS)

и справочники:

PREP (KODPREP, FIO, KODKAF),

KAF (KODKAF, NAIMKAF),

где:

KODPREP – код преподавателя

DATA – дата

VIDNAGR – вид нагрузки

CHAS – количество часов

KODKAF – код кафедры

NAIMKAF – наименование кафедры.

Выдать справку о нагрузке по кафедрам за январь 2010 г. с детализацией по видам нагрузки, содержащую поля NAIMKAF, VIDNAGR и суммарный объем нагрузки (SQL).

4.3 Вопросы для оценки качества освоения дисциплины

1. Дайте определение реляционной База Данных? Роль системы управления базами данными (СУБД) в организации.
2. Архитектура современных СУБД. В чем суть логического и инфологического моделирования?
3. Этапы проектирования СУБД. Каковы принципы построения ER-модели?
4. Нормализация БД. В чем особенности первой, второй и третьей нормальной формы?
5. Какие задачи по проектированию БД могут решаться с помощью инструментальных средств EgWin/BrWin.
6. Каково назначение языков DDL и DML. Как создать таблицу БД используя DDL.
7. Какие существуют операторы языка SQL для чтения и записи данных?
8. Какие существуют операторы языка SQL для обновления (редактирования) и удаления данных?
9. Какие существуют операторы языка SQL предоставления прав доступа к отдельным объектам БД различным пользователям?
10. В чем отличия архитектуры клиент-сервер от архитектуры файл-сервер? Каковы особенности работы БД в многопользовательском режиме?
11. Перечислите основные инструментальные средства инструментальные средства СУБД MS SQL Server?

12. Каково предназначение инструментального средства инструментального средства служб Reporting Services?
13. Каково назначение оператора BULK INSERT в Transact SQL?
14. Каково назначение функций-указателей в Transact SQL?
15. Каковы основные причины внедрения ХД в организации? Чем ХД отличается от транзакционной БД?
16. Опишите основные особенности ROLAP, HOLAP и MOLAP.
17. Опишите схему типа «звезда» и «снежинка».
18. Особенности метода DataVault.
19. Какие типы инфо-кубов поддерживаются в современных ХД.
20. Зачем используются агрегаты в ХД.
21. Расскажите об основных функциональных возможностях языка MDX.
22. Опишите принципы формирования OLAP-отчетности.
23. Опишите процедуру создания аналитического хранилища данных в среде MS SQL Server (SSAS).
24. Что такое ETL-процедура? Опишите возможную схему загрузки данных в ХД из плоских файлов.

V. РЕСУРСЫ

5.1 Основная литература

1. Гасанов Э.Э. Интеллектуальные системы. Теория хранения и поиска информации: учебник для бакалавриата и магистратуры / Э.Э. Гасанов, В.Б. Кудрявцев. – 2-е изд., испр. и доп. – М.: Издательство Юрайт, 2017. – 289 с. Режим доступа: www.biblio-online.ru/book/2771E75A-5B2D-4E2D-BD2B-B13DFB2916EB.
2. Гордеев С.И. Организация баз данных в 2 ч. Часть 1: учебник для вузов / С.И. Гордеев, В.Н. Волошина. – 2-е изд., испр. и доп. – М.: Издательство Юрайт, 2017. – 311 с. Режим доступа: www.biblio-online.ru/book/6006DA14-2EF3-4C35-B618-D9E87AB9173D.
3. Илюшечкин В.М. Основы использования и проектирования баз данных: учебник для академического бакалавриата / В. М. Илюшечкин. – М.: Издательство Юрайт, 2017. – 213 с. Режим доступа: www.biblio-online.ru/book/A8B884B0-CFEB-4279-98AF-DD2412E7700E.
4. Маркин А.В. Программирование на sql в 2 ч. Часть 1: учебник и практикум для бакалавриата и магистратуры / А.В. Маркин. – М.: Издательство Юрайт, 2017. – 362 с. Режим доступа: www.biblio-online.ru/book/95F7BDBD-B425-48BD-9DB2-744D039AFCAE.
5. Парфенов Ю.П. Постреляционные хранилища данных: учеб. пособие для вузов / Ю.П. Парфенов; под науч. ред. Н.В. Папуловской. – М.: Издательство Юрайт, 2017. – 121 с. доступа: www.biblio-online.ru/book/12E9FB88-0247-486A-AA58-6E4506281707.

5.2 Дополнительная литература

1. Калятин В.О. Право интеллектуальной собственности. Правовое регулирование баз данных: учеб. пособие для бакалавриата и магистратуры / В.О. Калятин. – М.: Издательство Юрайт, 2018. – 186 с. Режим доступа: www.biblio-online.ru/book/64BFB04C-1C8F-4034-B435-7D39ACB5BDA7.
2. Кудрявцев В.Б. Интеллектуальные системы: учебник и практикум для бакалавриата и магистратуры / В.Б. Кудрявцев, Э.Э. Гасанов, А.С. Подколзин. – 2-е изд., испр. и доп.

– М.: Издательство Юрайт, 2017. – 219 с. Режим доступа: www.biblio-online.ru/book/1DAA117E-A40C-4F22-B6EA-642C255D29CB.

3. Стасышин В.М. Базы данных: технологии доступа: учеб. пособие для академического бакалавриата / В.М. Стасышин, Т.Л. Стасышина. – 2-е изд., испр. и доп. – М.: Издательство Юрайт, 2017. – 178 с. Режим доступа: www.biblio-online.ru/book/20A6CE69-114D-4B78-A7CD-FA7D64D33E71.
4. Стружкин Н.П. Базы данных: проектирование. Практикум: учеб. пособие для академического бакалавриата / Н.П. Стружкин, В.В. Годин. – М.: Издательство Юрайт, 2017. – 291 с. Режим доступа: www.biblio-online.ru/book/61C51717-0D09-4A82-8DC8-42BAD8347E3D.

5.3 Программное обеспечение

№ п/п	Наименование	Условия доступа
1.	Microsoft Windows 7 Professional или более новая версия	Из внутренней сети университета (договор)
2.	Microsoft Office Professional Plus 2013 или более новая версия	Из внутренней сети университета (договор)
3.	Microsoft SQL Server 2014 Enterprise Edition или более новая версия	Из внутренней сети университета (договор)
4.	Microsoft Power BI	Свободно распространяемое ПО
5.	JDK 8	Свободно распространяемое ПО
6.	Notepad++	Свободно распространяемое ПО
7.	R 3.1.2 или более новая версия	Из внутренней сети университета (договор)
8.	RStudio	Из внутренней сети университета (договор)
9.	Anaconda 3 x64	Из внутренней сети университета (договор)
10.	Faronics Insight	Из внутренней сети университета (договор)
11.	SQL Server Data Tools 2015 или более новая версия	Из внутренней сети университета (договор)

5.4 Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№ п/п	Наименование	Условия доступа
	<i>Профессиональные базы данных, информационно-справочные системы</i>	
1.	Электронно-библиотечная система Юрайт	URL: https://biblio-online.ru/
	<i>Интернет-ресурсы (электронные образовательные ресурсы)</i>	

1.	Веб-сервис для хостинга IT-проектов и их совместной разработки	URL: https://github.com
----	--	--

5.5 Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, анти-вирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ, а также с установленным требуемым программным обеспечением, в количестве одна единица на каждого слушателя дисциплины.