

Разработано НИЯУ МИФИ

**Методические рекомендации для учителей по подготовке обучающихся
к прохождению теоретического этапа Московского конкурса
межпредметных навыков и знаний «Интеллектуальный мегаполис.
Потенциал»
в номинации «Академический класс»
по направлению «Информационно-технологическое»
на основе разбора демонстрационного варианта**

Москва, 2022 г.

Содержание

Раздел 1. Введение	3
Раздел 2. Методические рекомендации по решению задач по разделу “Введение в базы данных”	5
Раздел 3. Методические рекомендации по решению задач по разделу “Понятие реляционной модели данных”	12
Раздел 4. Методические рекомендации по решению задач по разделу “SQL- запросы”	25
Раздел 5. Методические рекомендации по решению задач по разделу “ Анализ и первичная обработка данных”	33

Раздел 1. Введение

Методические рекомендации для учителей по подготовке обучающихся к прохождению теоретического этапа Московского конкурса межпредметных навыков и знаний «Интеллектуальный мегаполис. Потенциал» в номинации «Академический класс» по направлению «Информационно-технологическое» составлены НИЯУ МИФИ на основе разбора демонстрационного варианта и включают четыре основных раздела с рекомендациями по решению заданий по разделам «введение в базы данных», «понятие реляционной модели данных», «SQL-запросы» и «анализ и первичная обработка данных».

Демонстрационный вариант теоретического этапа по медико-инженерному направлению в 2022 году состоит из 13 заданий различного уровня сложности. Обобщённый план конкурсных материалов для проведения теоретического этапа Конкурса представлен ниже в таблице 1.1.

Таблица 1.1.

№ задания	Уровень сложности	Темы элективного курса	Контролируемые требования к проверяемым умениям	Балл
1.	базовый	Введение в базы данных	Владеть основами баз данных	3
2.	повышенный	Введение в базы данных	Владеть основами баз данных	6
3.	базовый	Понятие реляционной модели данных	Владеть основными понятиями реляционной модели данных	3
4.	повышенный	Понятие реляционной модели данных	Владеть основными понятиями реляционной модели данных	6
5.	повышенный	Понятие реляционной модели данных	Владеть основными понятиями реляционной модели данных	6

6.	базовый	Понятие реляционной модели данных	Владеть основными понятиями реляционной модели данных. Составлять SQL-запросы	3
7.	базовый	SQL-запросы	Составлять SQL-запросы	3
8.	повышенный	SQL-запросы	Составлять SQL-запросы	6
9.	повышенный	SQL-запросы	Составлять SQL-запросы	6
10.	базовый	SQL-запросы	Составлять SQL-запросы	3
11.	повышенный	Сбор и первичная обработка данных	Проводить сбор и первичную обработку данных	6
12.	базовый	Анализ данных	Проводить анализ данных	3
13.	повышенный	Анализ данных	Проводить анализ данных	6
Сумма баллов:				60

Раздел 2. Методические рекомендации по решению задач по разделу “Введение в базы данных”

Задания по разделу “Введение в базы данных” представлены двумя заданиями: одно задание базового уровня сложности, одно задание повышенного уровня сложности.

Задание 1. (3 балла)

Контролируемые требования к проверяемым умениям:

Владение основами баз данных

Уровень сложности – базовый

Текст задания демоварианта

Закончите фразу: «База данных — это»:

- 1) это организованная на машинном носителе совокупность взаимосвязанных данных, которая содержит сведения о различных сущностях одной предметной области.
- 2) совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных.
- 3) система, предназначенная для сбора, передачи, обработки, хранения и выдачи информации потребителям.
- 4) верный ответ отсутствует.

Для выполнения задания необходимо

Первое задание в тесте относится к теме “Введение в базы данных”. Поэтому для его выполнения необходимо знать основные базовые определения баз данных. В частности, определения базы данных, СУБД, реляционной базы данных и др.

При этом на практике часто путают понятия базы данных и СУБД, хотя это совершенно разные понятия.

Если база данных — это набор данных, то СУБД (система управления базами данных) — это специальное программное обеспечение, разработанное для управления этими базами данных. Существует множество СУБД, например: MS Access, PostgreSQL, SQLite, Oracle Database и MySQL.

В литературе предлагается множество определений понятия «база данных», отражающих скорее субъективное мнение тех или иных авторов, однако общепризнанная единая формулировка понятия базы данных отсутствует.

Для примера приведем несколько различных определений. Первое из госта по Организации данных в системах обработки данных. Второе это стандарт ISO - международная организация по стандартизации, которая как раз отвечает за стандарт языка SQL. Третье из учебника по базам данных.

База данных — совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных.

База данных — совокупность данных, организованных в соответствии с концептуальной структурой, описывающей характеристики этих данных и взаимоотношения между ними, которая поддерживает одну или более областей применения.

База данных — это организованная на машинном носителе совокупность взаимосвязанных данных, которая содержит сведения о различных сущностях одной предметной области.

Безусловно в чем-то эти определения схожи, а именно в том, что бд — это совокупность данных, которая организована в соответствии с некими правилами.

Для верного ответа на первый вопрос теста также необходимо помнить классификацию СУБД по основным признакам: по модели данных, по способу доступа к Базе данных и по степени распределенности.

Также необходимо помнить основные функции СУБД:

- Управление данными во внешней памяти;
- Управление буферами оперативной памяти;
- Управление транзакциями (РСУБД);
- Журнализация;
- Поддержка языков баз данных.

Решение задания

Давайте поочередно разберем все предлагаемые варианты ответа.

Первый вариант совпадает с одним из определений приведенных ранее и соответственно является правильным.

Второй вариант “совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных” является определением СУБД (системы управления базами данных).

Проще говоря СУБД — это комплекс программ, необходимых для того, чтобы сначала создать базу данных, затем манипулировать данными (удалять, добавлять, изменять, искать) Этот программный комплекс обеспечивает безопасность, надежность и целостность данных, а также предоставляет средства для администрирования.

Третий вариант ответа “система, предназначенная для сбора, передачи, обработки, хранения и выдачи информации потребителям” является определением информационной системы.

По своей сути любая БД представляет собой скрупулезно построенное хранилище структурированных данных. Само по себе это хранилище ни для

кого особого интереса не представляет до тех пор, пока оно не интегрируется в состав более сложной надсистемы, называемой информационной системой.

То есть база данных это один из основных компонентов информационных систем, которая предназначена для своевременного обеспечения надлежащих людей надлежащей информацией.

А благодаря БД данные представляются и интерпретируются в удобном для анализа виде.

Четвертый вариант ответа “верный ответ отсутствует” является неверным, так ранее уже был найден ответ, являющийся верным.

Ответ: 1

Задание 2. (6 баллов)

Контролируемые требования к проверяемым умениям:

Владение основами баз данных

Уровень сложности – повышенный

Текст задания демоварианта

Даталогический уровень проектирования базы данных описывает ...

- 1) Предметную область без привязки к виду данных.
- 2) Предметную область с привязкой к виду базы данных или даже конкретной СУБД.
- 3) Технические аспекты реализации базы данных под управлением конкретной СУБД.
- 4) Верный ответ отсутствует.

Для выполнения задания необходимо

Второе задание в тесте относится к теме “Введение в базы данных”. Поэтому для его выполнения необходимо знать этапы проектирования базы данных. Вообще говоря, проектирование базы данных разделяется на три этапа:

- Концептуальный (или еще его называют инфологическим);
- Логический (его еще называют даталогическим);
- Физический.

Каждый из этих этапов выполняет свои функции и решает свои задачи.

Первый этап проектирования заключается в представлении предметной области в математических (как правило теоретико-множественных) терминах в результате чего формируется концептуальная (или концептуально информационная) модель предметной области. Такая модель фиксирует понятийный базис будущей системы.

На сегодняшний момент существует несколько эффективных способов или языков представления концептуальной модели. Многие из них в той или иной степени базируются на модели, разработанной Питером Ченом, которую еще называют ER-модель или модель “сущность-связь”.

После концептуального проектирования выполняется логическое (или даталогическое) проектирование — создание схемы базы данных на основе конкретной модели данных, например, реляционной модели данных.

Основной целью данного уровня проектирования является детализация инфологической модели и превращение ее в схему, на которой ранее выявленные сущности, атрибуты и связи оформляются согласно правилам моделирования для выбранной модели данных (то есть для выбранного вида базы данных).

Так, например, для реляционной модели данных даталогическая модель — набор схем отношений, обычно с указанием первичных ключей, а также «связей» между отношениями, представляющих собой внешние ключи.

На этапе логического проектирования учитывается специфика конкретной модели данных, но может не учитываться специфика конкретной СУБД.

По завершении фазы логического проектирования утверждается финальная версия ТЗ на БД. В техническом задании закрепляются требования к проекту, утверждается логическая модель БД, описываются необходимые отчеты итд.

Переход к следующей фазе физического проектирования БД производится после выбора СУБД, именно она определяет особенности будущего программного продукта.

Замечание. Ряд специалистов (и источников) придерживается точки зрения, что выбор СУБД надо проводить не после логического проектирования, а ранее - после концептуального. Главное, что, выбор СУБД должен быть осуществлен до начала этапа физического проектирования.

Физическое проектирование — это уточнение решения с учетом имеющихся в наличие технологий, возможностей реализации и требуемой производительности.

В ходе физического проектирования выполняются следующие действия:

- создание отношений (таблиц) и связей между ними;
- создание необходимых индексов;
- разработка представлений;
- реализация бизнес-логики (в первую очередь с помощью триггеров и хранимых процедур);
- определяются характеристики транзакций;
- внедряются механизмы защиты (как минимум предусматривается наличие разных пользователей и/ или групп пользователей с различными уровнями доступа к данным).

Решение задания

Первый вариант ответа относится к концептуальному (инфологическому) уровню проектирования.

Второй вариант ответа относится к логическому (датологическому) уровню проектирования и поэтому является верным.

Третий вариант ответа относится к физическому уровню проектированию.

Четвертый вариант ответа “верный ответ отсутствует” является неверным, так ранее уже был найден ответ, являющийся верным.

Ответ: 2

Раздел 3. Методические рекомендации по решению задач по разделу “Понятие реляционной модели данных”

Задания по предмету “Понятие реляционной модели данных” представлены четырьмя заданиями: два задания базового уровня сложности, два задания повышенного уровня сложности.

Задание 3. (3 балла)

Контролируемые требования к проверяемым умениям:

Владение основными понятиями реляционной модели данных

Уровень сложности – базовый

Текст задания демоварианта

Выберите бинарные операции реляционной алгебры. Возможны несколько корректных вариантов ответа.

- 1) Проекция.
- 2) Удвоение.
- 3) Деление.
- 4) Объединение.
- 5) Все ответы верны.

Для выполнения задания необходимо

Необходимо знать правила составления запросов на реляционной алгебре, все операции реляционной алгебры и их особенности.

Реляционная алгебра — это процедурный способ записи запроса, который базируется на теории множеств. Последовательность операций в выражениях реляционной алгебры влияет на стратегию вычисления запроса, хотя существуют способы эквивалентного преобразования алгебраических выражений.

Реляционная алгебра — это замкнутая система операций над отношениями в реляционной модели данных.

Первоначальный набор из восьми операций был сформирован Э.Коддом в 1970х годах. Некоторые операции можно выразить через другие, то есть это множество избыточное.

Результатом каждой операции реляционной алгебры над отношениями является также отношение, данное свойство алгебры называется замкнутостью.

Запрос на реляционной алгебре — это некая последовательность операций над исходными отношениями.

Ответом или результатом запроса является производное отношение, которое определяется алгебраическим выражением над исходными отношениями.

Множество операций реляционной алгебры подразделяется на две группы: теоретико-множественные операции и специальные. В некоторых источниках выделяется третья группа, состоящая из одной только операции декартова произведения.

В первую группу входят операции объединения, пересечения и разности.

Это бинарные операции, которые могут выполняться только над структурно идентичными отношениями. Схемы таких отношений называют соответственными.

То есть у отношений должна быть одна и та же степень и соответствующие атрибуты определены на одних и тех же доменах.

В группу специальных операций входят операции проекции, ограничения, соединения и деления.

Решение задания

Поочередно разберем все предлагаемые варианты ответа.

Операция проекции – это унарная операция, которая определяет список атрибутов, подлежащих выборке. Результатом операции является множество кортежей, в котором будут присутствовать только атрибуты, перечисленные в списке. Операция проекции дает "вертикальный срез" отношения, в котором удалены все возникшие при таком срезе дубликаты кортежей.

Удвоение - такая операция в реляционной алгебре отсутствует

Деление отношений – бинарная операция. Реляционное деление достаточно нетривиально описать. Пусть схемы исходных отношений $R_1(A_1, \dots, A_n, B_1, \dots, B_m)$ и $R_2(B_1, \dots, B_m)$, причем атрибуты B_1, \dots, B_m – общие для двух отношений. Делением отношения R_1 на R_2 называется отношение, схема которого имеет вид – $R_{рез}(A_1, \dots, A_n)$, а его тело содержит множество кортежей $\langle a_1, \dots, a_n \rangle$ таких, что для каждого из кортежей отношения R_2 , $\langle b_1, \dots, b_m \rangle \in R_2$, в R_1 найдется кортеж $\langle a_1, \dots, a_n, b_1, \dots, b_m \rangle$.

Объединение — это бинарная операция. Объединением двух отношений называется отношение, содержащее множество кортежей, принадлежащих либо первому, либо второму отношению, либо обоим отношениям сразу.

Ответ: 3 и 4

Задание 4. (6 баллов)

Контролируемые требования к проверяемым умениям:

Владение основными понятиями реляционной модели данных

Уровень сложности – повышенный

Текст задания демоварианта

Выберите все неверные утверждения из предложенных ниже:

1) В реляционной модели данных порядок кортежей в отношении не определен.

- 2) В реляционной теории считается, что в отношении не может быть двух идентичных элементов (кортежей).
- 3) Мощность отношения – это количество атрибутов в отношении.
- 4) Домен можно рассматривать как множество, каждым элементом которого является кортеж.
- 5) Атрибут — именованное свойство сущности (отношения).

Краткая теоретическая справка

Реляционная модель — математическая теория, описывающая структуры данных, логику контроля целостности данных и правила управления данными.

Упрощённо: модель для описания реляционных баз данных.

Принципы реляционной модели были сформулированы в 1969—1970 годах Э. Ф. Коддом (E. F. Codd).

На момент изобретения именно эта модель предоставила полный комплекс решений по описанию базы данных, контролю целостности данных и управлению данными — в отличие от множества других альтернативных подходов, позволяющих реализовать лишь часть задач.

Термин «реляционный» означает, что теория основана на математическом понятии отношение (relation). В качестве неформального синонима термину «отношение» часто встречается слово таблица. Необходимо помнить, что «таблица» есть понятие нестрогое и неформальное и часто означает не «отношение» как абстрактное понятие, а визуальное представление отношения на бумаге или экране.

В контексте описания структур данных реляционной моделью вводится понятие нормализованного отношения. Более того, нормализованное отношение является единственной базовой структурой, которой оперирует эта модель

В контексте целостности данных речь идет о целостности сущностей и внешних ключей.

И, наконец, в контексте правил управления данными речь идет о реляционной алгебре и реляционном счислении.

Давайте поговорим о фундаментальных понятиях реляционной модели данных, которые необходимо знать для верного ответа на четвертый вопрос теста.

Отношение (relation) — множество сущностей, обладающих одинаковым набором атрибутов. В контексте реляционных баз данных отношение состоит из заголовка (схемы) и тела (набора кортежей).

Важно разделять понятия «отношение» (relation) и «связь» (relationship) несмотря на то, что очень часто в русскоязычной (переводной) литературе их путают.

Тогда для упрощения запоминания можно сформулировать более простое определение отношения.

Отношение — множество кортежей (записей, строк таблицы), обладающих одинаковым набором атрибутов (свойств, полей, столбцов таблицы).

Перейдем к рассмотрению компонентов отношения.

Тип данных (data type) — набор объектов данных определенной структуры и набор допустимых операций, в любой из которых такие объекты могут выступать операндами.

Упрощенно, на примере: числа, строки, даты и т.д.

В общем случае понятие «тип данных» здесь эквивалентно аналогичному понятию в языках программирования (например, «строки» или «числа» трактуются в базах данных и языках программирования почти идентично).

Все значения, которыми оперирует база данных, являются типизированными (т.е. в любой момент времени известен их тип). Для

большинства типов данных существуют операции их преобразования (например, строка «12.34» может быть преобразована в дробное число 12.34).

На типы данных могут налагаться дополнительные ограничения, что приводит к появлению понятия «домен данных».

Домен данных (attribute domain) — набор всех возможных значений атрибута отношения.

Упрощенно, на примере: номер телефона, фамилия, название улицы и т.д.

Суть этого понятия проще всего пояснить на примере. Допустим, в некотором отношении есть атрибут «Код товара», который представлен строкой. И также известно, что все коды товара формируются по правилу «три буквы английского алфавита в верхнем регистре, четыре цифры, две буквы английского алфавита в нижнем регистре» (например, «ABC1234yz»). Очевидно, что строка «понедельник» не является кодом товара, т.е. не входит в соответствующий домен, хоть и относится к тому же типу данных (к строкам).

Понятие домена важно в силу того, что данные считаются сравнимыми только в случае принадлежности к одному домену, т.е. бессмысленно сравнивать код товара и название дня недели (хотя технически такую операцию позволит вам выполнить любая СУБД).

Атрибут (attribute) — именованное свойство сущности (отношения).

Упрощенно: столбец (колонка) таблицы.

Кортеж (tuple) — часть отношения, представляющая собой уникальную взаимосвязанную комбинацию значений, каждое из которых соответствует своему атрибуту.

Упрощенно: строка (запись) таблицы.

Мощность отношения (кардинальность) — количество кортежей в отношении.

Степень отношения (арность) — количество атрибутов в отношении.

Основные свойства отношения.

В реляционной теории считается, что все элементы множества попарно различимы, соответственно в отношении не может быть двух идентичных элементов (кортежей)

Порядок кортежей* в отношении не определен.

Порядок атрибутов в схеме отношения не определен.

*В принятой трактовке реляционной модели кортеж — это множество, в отличие от математического понятия кортежа как последовательности.

Важно помнить, что термины «таблица», «столбец», «строка» являются лишь упрощенными и более удобными в повседневной речи аналогами соответствующих понятий реляционной теории. Но если подходить к вопросу строго математически, такое упрощение будет неверным.

Решение задания

Верными утверждения являются варианты ответа под номерами 1, 2 и 5. Соответственно варианты ответа под номерами 3 и 4 являются неверными утверждениями.

Ответ: 3 и 4

Задание 5. (6 баллов)

Контролируемые требования к проверяемым умениям:

Владение основными понятиями реляционной модели данных

Уровень сложности – повышенный

Текст задания демоварианта

В какой нормальной форме находится приведенная ниже таблица:

ID	Автор	Название	Издательство	ДР автора	Тел. Изд
1	Иван Дмитриев	Сборник рассказов	Книжный пресс	11.12.1980	111-111

2	Ольга Петрова	Сборник стихов	Очепятка	18.02.1958	222- 222
3	Василий Сумкин	Новый роман	Книжный пресс	01.02.1949	111- 111
4	Мария Иванова	Пьеса о главном	МКС	23.08.1997	333- 333
5	Иван Дмитриев	Романтичный роман	ОМСК	11.12.1980	444- 444

- 1) В третьей нормальной форме.
- 2) Во второй нормальной форме.
- 3) В первой нормальной форме.
- 4) Правильный ответ отсутствует.

Ответ: 3.

Краткая теоретическая справка

Нормализация — это пошаговый обратимый процесс разложения исходных отношений на более мелкие и простые.

Нормальная форма — свойство отношения в реляционной модели данных, характеризующее его с точки зрения избыточности, потенциально приводящей к логически ошибочным результатам выборки, изменения или удаления данных. Нормальная форма определяется как совокупность требований, которым должно удовлетворять отношение.

Процесс преобразования отношений базы данных к виду, отвечающему нормальным формам, называется нормализацией. она предназначена для приведения структуры БД к виду, обеспечивающему минимальную логическую избыточность, и не имеет целью уменьшение или увеличение производительности работы или же уменьшение или увеличение физического объема базы данных.

Как отмечает К. Дейт, общее назначение процесса нормализации заключается в следующем:

- исключение некоторых типов избыточности;
- устранение некоторых аномалий обновления;
- разработка проекта базы данных, который является достаточно «качественным» представлением реального мира, интуитивно понятен и может служить хорошей основой для последующего расширения;
- упрощение процедуры применения необходимых ограничений целостности.

Устранение избыточности производится, как правило, за счёт декомпозиции отношений таким образом, чтобы в каждом отношении хранились только первичные факты (то есть факты, не выводимые из других хранимых фактов).

Аномалией называется такая ситуация в таблице БД, которая приводит к противоречию в БД либо существенно усложняет обработку БД. Причиной является излишнее дублирование данных в таблице, которое вызывается наличием функциональных зависимостей от не ключевых атрибутов.

Аномалии операций с данными — некорректное выполнение операций с данными или возникновение побочных эффектов операций с данными, ставшее результатом нарушения требования адекватности базы данных предметной области.

Упрощенно: выполняя правильную (с точки зрения пользователя) операцию правильным SQL-запросом, мы получаем ошибочный результат или нежелательный побочный эффект.

Традиционно выделяют следующие три вида аномалий операций с данными:

- Аномалия вставки (или аномалии добавления).
- Аномалия удаления
- Аномалия обновления (или аномалия модификации)

Важно понимать, что аномалии операций с данными ничего общего не имеют со сбоями в работе СУБД или ошибками в SQL-запросе. При возникновении аномалии СУБД работает совершенно корректно, и выполняемый SQL-запрос тоже корректен и соответствует поставленной цели.

В этом и кроется опасность аномалий: всё работает, верно, а результаты получаются неправильными.

В теории нормализации схема отношения определяется тройкой: $\langle R, U, F \rangle$

- R – имя схемы
- U – множество атрибутов
- F – множество нетривиальных функциональных зависимостей (ФЗ) между атрибутами

ФЗ: значению из области определения одного атрибута соответствует ровно одно значение из области определения другого.

Говорят, что A_n функционально полно зависит от $\{A_1, \dots, A_m\}$, если из этого множества нельзя выбросить ни одного атрибута, не разрушая ФЗ.

Левая часть ФЗ называется детерминативом.

В отношении всегда присутствует функциональная зависимость – это зависимость вторичных(неключевых) атрибутов от ключа отношения.

Понятие возможного (потенциального) ключа: Любой детерминатив, от которого функционально зависят все атрибуты схемы.

Если атрибут A_i входит в какой-нибудь возможный ключ, то он входит в подмножество (U_0) основных атрибутов, если нет - в подмножество (U_n) неосновных атрибутов. То есть $U = U_0 \cup U_n$

Каждой нормальной форме соответствует определенный набор ограничений, и отношение находится в некоторой нормальной форме, если удовлетворяет свойственному ей набору ограничений.

Основные свойства нормальных форм состоят в следующем:

1) каждая следующая нормальная форма в некотором смысле лучше предыдущей нормальной формы;

2) при переходе к следующей нормальной форме свойства предыдущих нормальных форм сохраняются.

Отношение находится в 1НФ, если все его атрибуты являются простыми (атомарными) те не списками). Другими словами, элемент является атомарным, если его нельзя разделить на части, которые могут использоваться в таблице независимо друг от друга. Не должно быть повторений строк в таблице. Первая нормальная форма не накладывает никаких ограничений на множество функциональных зависимостей F .

Отношение находится в 2НФ, если оно находится в 1НФ и каждый не основной атрибут функционально полно зависит от его потенциального ключа.

Говорят, что A_n функционально полно зависит от $\{A_1, \dots, A_m\}$, если из этого множества нельзя выделить меньшее подмножество атрибутов, от которого также можно вывести данную функциональную зависимость.

Отношение находится в 3НФ, если оно находится в 2НФ и нет функциональных зависимостей между неосновными атрибутами. (или по-другому каждый неключевой атрибут нетранзитивно зависит от первичного ключа.)

Отношение находится в нормальной форме Бойса-Кодда (BCNF), если оно находится в 3НФ и любой детерминатив является ключом. (отсутствуют зависимости атрибутов ключа от вторичных (не входящих в ключ) атрибутов)

Решение задания

Приведенное отношение не содержит атрибутов-списков и не имеет повторений строк в таблицы, поэтому можно сделать вывод, что оно находится в первой нормальной форме.

Если мы предполагаем, что книга в нашей таблице может выпускаться только одним издательством, тогда комбинация атрибутов Автор и Название является потенциальным ключом. Если мы предполагаем, что книга в нашей таблице может выпускаться разными издательствами, тогда комбинация атрибутов Автор, Название и Издательство является потенциальным ключом. В любом из этих случаев в отношении есть атрибуты, которые функционально полно зависят только от части ключа: день рождения автора зависит только от автора, а телефон издательства только от издательства.

Ответ: 3

Задание 6. (3 балла)

Контролируемые требования к проверяемым умениям:

Владение основными понятиями реляционной модели данных

Уровень сложности – базовый

Текст задания демоварианта

Укажите строковый тип данных, который позволяет хранить строки фиксированной длины.

- 1) varchar(size)
- 2) char(size)
- 3) text

4) tinytext(size)

Решение задания

- 1) varchar(size) - тип данных для хранения строк переменной длины.
- 2) char(size) -тип данных для хранения строк фиксированной длины.
- 3) text - тип данных в некоторых СУБД, в котором можно хранить строки произвольной длины.
- 4) tinytext(size) – данный тип данных не существует

Ответ: 2

Раздел 4. Методические рекомендации по решению задач по разделу “SQL-запросы”

Задания по предмету “SQL-запросы” представлены четырьмя заданиями: два задания базового уровня сложности, два задания повышенного уровня сложности.

Задание 7. (3 балла)

Контролируемые требования к проверяемым умениям:

Умение составлять SQL-запросы

Уровень сложности – базовый

Текст задания демоварианта

Укажите верные особенности языка SQL. Возможны несколько корректных вариантов ответа.

- 1) SQL поддерживает неопределенные значения (NULL) в отличие от реляционной теории.
- 2) Несмотря на наличие международного стандарта, многие разработчики СУБД вносят изменения в язык SQL.
- 3) Функции и названия объектов в SQL чувствительны к регистру.
- 4) Язык SQL чувствителен к переносу строк.
- 5) SQL – это процедурный язык программирования, предназначенный для описания, изменения и извлечения данных, хранимых в реляционных базах данных.

Краткая теоретическая справка

SQL – это непроцедурный язык программирования, применяемый для создания, модификации и управления данными в реляционной базе данных, управляемой соответствующей системой управления базами данных. Он декларирует, что нужно сделать, а исполнение возлагается на СУБД.

Особенности SQL

Декларативность

С помощью SQL программист описывает только то, какие данные нужно извлечь или модифицировать. То, каким образом это сделать, решает СУБД непосредственно при обработке SQL-запроса.

Несоответствие реляционной модели данных

Создатели реляционной модели данных Эдгар Кодд, Кристофер Дейт и их сторонники указывают на то, что SQL не является истинно реляционным языком. В частности, они указывают на следующие дефекты SQL с точки зрения реляционной теории:

- допущение строк-дубликатов в таблицах и результатах выборок, что в рамках реляционной модели данных невозможно и недопустимо;
- поддержка неопределённых значений (NULL), создающую фактически многозначную логику;
- значимость порядка столбцов, возможность ссылок на столбцы по номерам (в реляционной модели столбцы должны быть равноправны);
- допущение столбцов без имени, дублирующихся имен столбцов.

Сложность

Хотя SQL и задумывался как средство работы конечного пользователя, в конце концов он стал настолько сложным, что превратился в инструмент программиста.

Отступления от стандартов

Несмотря на наличие международного стандарта ANSI SQL-92, многие разработчики СУБД вносят изменения в язык SQL, применяемый в разрабатываемой СУБД, тем самым отступая от стандарта.

Диалекты SQL

Каждая из реализаций языка SQL в конкретной СУБД называется диалектом. Функции, которые добавляются к стандарту языка разработчиками коммерческих реализаций принято называть расширениями. Например, в

стандарте языка SQL определены конкретные типы данных, которые могут храниться в базах данных.

Разделы языка SQL

Выделяют три раздела языка SQL:

- язык определения данных (DDL – Data Definition Language);
- язык манипулирования данными (DML – Data Manipulation Language);
- язык управления данными (DCL – Data Control Language);
- В некоторых источниках выделяется еще 4ое подмножество - TCL (Transaction Control Language) - язык управления транзакциями.

Синтаксис SQL

- Функции и названия объектов нечувствительны к регистру: SELECT = sELeCt;
- Однако согласно общепринятому стилю программирования, операторы (и другие зарезервированные слова) в SQL обычно рекомендуется писать прописными буквами;
 - Однако при поиске по текстовым полям регистр учитывается;
 - SQL не чувствителен к переносу строк;
 - Отсутствуют обязательные символы, завершающие строки;
 - Каждую транзакцию принято завершать точкой с запятой, но при выполнении отдельных команд их употребление не обязательно.

Решение задания

В тексте находим правильные утверждения: варианты ответа под номерами 1 и 2.

Ответ: 1 и 2

Задание 8. (6 баллов)

Контролируемые требования к проверяемым умениям:

Умение составлять SQL-запросы

Уровень сложности – повышенный

Текст задания демоварианта

Выберите запрос, который посчитает количество студентов (таблица students(student_id, name, group)) в каждой группе и выведет только те группы, в которых более десяти студентов?

- 1) SELECT count(student_id), group
FROM students
WHERE count(student_id) > 10
GROUP BY group;
- 2) SELECT count(student_id), group, name
FROM students
GROUP BY group
HAVING count(student_id) > 10;
- 3) SELECT count(student_id), group
FROM students
HAVING count(student_id) > 10;
- 4) SELECT count(student_id), group
FROM students
GROUP BY group
HAVING count(student_id) > 10;

Краткая теоретическая справка

Решение задания

- 1) SELECT count(student_id), group
FROM students
WHERE count(student_id) > 10 – (использование агрегационных функций в условии WHERE недопустимо)
GROUP BY group;

- 2) SELECT count(student_id), group, name – (использование name недопустимо, так как это не агрегационная функция и по данному атрибуту не проводится группировка)
FROM students
GROUP BY group
HAVING count(student_id) > 10;
- 3) SELECT count(student_id), group - (в запросе отсутствует группировка)
FROM students
HAVING count(student_id) > 10;
- 4) SELECT count(student_id), group - все верно
FROM students
GROUP BY group
HAVING count(student_id) > 10; - **верный запрос**

Ответ: 4

Задание 9. (6 баллов)

Контролируемые требования к проверяемым умениям:

Умение составлять SQL-запросы

Уровень сложности – повышенный

Текст задания демоварианта

Выберите верные утверждения о псевдозначении NULL на языке SQL.

Возможны несколько корректных вариантов ответа.

- 1) NULL <> 1
- 2) NULL <> NULL
- 3) NULL IS NOT NULL
- 4) NULL IS NULL

Краткая теоретическая справка

NULL в Системах управления базами данных (СУБД) — специальное значение (псевдозначение), которое может быть записано в поле таблицы базы данных (БД). NULL соответствует понятию «пустое поле», то есть «поле, не содержащее никакого значения». Введено для того, чтобы различать в полях БД пустые (визуально не отображаемые) значения (например, строку нулевой длины) и отсутствующие значения (когда в поле не записано вообще никакого значения, даже пустого).

NULL означает отсутствие, неизвестность информации. Значение NULL не является значением в полном смысле слова.

Поэтому NULL не равно ни логическому значению FALSE, ни пустой строке, ни нулю ни другому нулю (NULL). При сравнении NULL с любым значением будет получен результат NULL, а не FALSE и не 0. Более того, NULL не равно NULL!

Поэтому для того, чтобы узнать, что значение NULL или не NULL используется конструкция IS NULL, IS NOT NULL. То есть через больше, меньше и равно NULL сравнивать не имеет смысла.

В реляционных СУБД у всех полей, если явно не сказано обратное, может быть значение NULL.

Поскольку NULL не является, в общем смысле, значением, использование его в арифметических, строковых, логических и других операциях, строго говоря, некорректно. Тем не менее, большинство СУБД поддерживают такие операции, но вводят для них специальные правила:

- NULL может присваиваться переменным и записываться в поля, независимо от объявленного типа данных этих переменных (полей);
- NULL может передаваться в процедуры и функции как легальное значение параметра. Результаты выполнения такой процедуры или функции определяются операциями, выполняемыми с параметрами внутри неё;
- Любая операция сравнения с NULL (даже операция «NULL = NULL»), даёт в результате значение «неизвестность» (UNKNOWN). Окончательный результат при этом зависит от полного логического

выражения в соответствии с таблицей истинности логических операций. Если сравнение с NULL есть вся логическая операция целиком (а не её часть), то результат ее аналогичен FALSE (выражение вида IF <что-то> = NULL THEN <действие1> ELSE <действие2> END IF всегда будет приводить к выполнению действия2);

- Агрегатные и аналитические функции (используемые в SQL в качестве операций на множествах и списках), как правило, игнорируют значения NULL в пользу допустимых значений остальных элементов множества. Например, для функции AVG, предназначенной для нахождения среднего арифметического значения какого-либо выражения, вычисленного для каждой строки из группы, результат получается таким же, как если бы строки, содержащие для этого выражения значение NULL, вообще не содержались бы в группе;

- Существует специальная системная функция или операция (обычно expr IS [NOT] NULL), возвращающая логическое значение «истина» (TRUE), если expr является (не является) NULL и FALSE в противном случае;

Решение задания

В тексте находим правильное утверждение: выражение во варианте ответа под номером 4 имеет логическое значение «истина».

Ответ: 4

Задание 10. (3 балла)

Контролируемые требования к проверяемым умениям:

Умение составлять SQL-запросы

Уровень сложности – базовый

Текст задания демоварианта

Какое действие выполняет SQL команда «DELETE FROM users WHERE users.id > 200» ?

- 1) Удаляет таблицу из базы данных;
- 2) Удаляет базу данных;

- 3) Удаляет из указанной таблицы строки, удовлетворяющие заданному условию;
- 4) Удаляет из указанной таблицы все строки.

Краткая теоретическая справка

DELETE — в языке SQL — команда удаления записей из таблицы. Критерий отбора записей для удаления определяется выражением WHERE. В случае, если критерий отбора не определён, выполняется удаление всех записей.

Решение задания

В тексте находим правильный ответ: «Удаляет из указанной таблицы строки, удовлетворяющие заданному условию».

Ответ: 3

Раздел 5. Методические рекомендации по решению задач по разделу “Анализ и первичная обработка данных”

Задания по разделу “Анализ и первичная обработка данных” представлены четырьмя заданиями: одно задание базового уровня сложности, два задания повышенного уровня сложности.

Задание 11. (6 баллов)

Контролируемые требования к проверяемым умениям:

Умение проводить сбор, первичную обработку и анализ данных

Уровень сложности – повышенный

Текст задания демоварианта

Григорий проводит опрос жителей центрального округа Москвы о благоустройстве детских площадок. Однако далеко не все жители готовы сотрудничать. В первый день Григорий обратился к 200 жителям, из которых успешно опросить у него получилось лишь 130. Второй день был еще более печальным. Из 300 человек согласилось пройти опрос лишь 120. Можно ли при $\alpha=0,01$ утверждать, что первый день у Григория прошел более эффективно?

- 1) Можно;
- 2) Нельзя;
- 3) Недостаточно данных для принятия решения.

Решение задания

Пусть p_1 - процент жителей, опрошенных в первый день, p_2 - процент жителей, опрошенных во второй день. Введем нулевую гипотезу: $H_0 : p_1 = p_2$ при конкурирующей гипотезе $H_1 : p_1 > p_2$. Вычислим наблюдаемое значение критерия по формуле:

$$U_{\text{набл}} = \frac{\frac{m_1}{n_1} - \frac{m_2}{n_2}}{\sqrt{\frac{m_1+m_2}{n_1+n_2} \left(1 - \frac{m_1+m_2}{n_1+n_2}\right) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

где $m_1 = 130$, $m_2 = 120$, $n_1 = 200$, где $n_2 = 300$

Значит, $U_{\text{набл}} = 5,48$.

Критическая точка $U_{\text{крит}}$ находится из условия $\Phi(U_{\text{крит}}) = 1 - 2a = 1 - 2 \cdot 0,01 = 0,98$.

Значит, $U_{\text{крит}} = 2,33$. Так как $U_{\text{набл}} = 5,48 > U_{\text{крит}} = 2,33$, нулевую гипотезу следует отвергнуть на данном уровне значимости, и можно считать, что первый день прошел более эффективно.

Ответ: 1

Задание 12. (3 балла)

Контролируемые требования к проверяемым умениям:

Умение проводить сбор, первичную обработку и анализ данных

Уровень сложности – базовый

Текст задания демоварианта

Медиана набора чисел – это ...

- 1) число, которое находится в середине этого набора, если его упорядочить по возрастанию.
- 2) число, которое встречается наиболее часто в данном наборе.
- 3) наименьшее число из набора.
- 4) число, равное сумме всех чисел набора, делённой на их количество.

Краткая теоретическая справка

Медиана ряда чисел – это число, которое стоит строго посередине ряда нечётного количества чисел, упорядоченного от наименьшего к наибольшему. Для четного количества чисел в ряду медианой является половина суммы двух чисел, которые стоят посередине ряда, упорядоченного по возрастанию.

Решение задания

В тексте находим правильный ответ: число, которое находится в середине этого набора, если его упорядочить по возрастанию

Ответ: 1

Задание 13. (6 баллов)

Контролируемые требования к проверяемым умениям:

Умение проводить сбор, первичную обработку и анализ данных

Уровень сложности – повышенный

Текст задания демоварианта

Дан набор чисел: 30 33 39 29 34. Необходимо вычислить дисперсию данного набора чисел.

- 1) 33
- 2) 10
- 3) 12,4
- 4) 0

Краткая теоретическая справка

Термин «дисперсия» относится к степени дисперсии точек данных набора данных от его среднего значения, которое вычисляется как среднее квадрата отклонения каждой точки данных от среднего значения совокупности. Формула для дисперсии может быть получена путем суммирования квадрата отклонения каждой точки данных и последующего деления результата на общее количество точек данных в наборе данных.

Решение задания

Расчет дисперсии производят по формуле:

$$s^2 = \frac{\sum_{i=1}^n (X - \bar{X})^2}{n}$$

, где

s^2 – выборочная дисперсия, рассчитанная по данным наблюдений,

X – отдельные значения,

N – количество значений,

\bar{X} – среднее арифметическое по выборке.

Формула для определения среднего арифметического наблюдений:

$$\bar{X} = (X_1 + X_2 + \dots + X_n) / n$$

$$\bar{X} = (30 + 33 + 39 + 29 + 34) / 5 = 33$$

$$\begin{aligned} s^2 &= ((30 - 33)^2 + (33 - 33)^2 + (39 - 33)^2 + (29 - 33)^2 + (34 - 33)^2) / 5 = \\ &= ((-3)^2 + (0)^2 + (6)^2 + (-4)^2 + (1)^2) / 5 = \\ &= (9 + 0 + 36 + 16 + 1) / 5 = 12,4 \end{aligned}$$

Ответ: 3