

ПРОЕКТНАЯ РАБОТА
по модулю «Базы данных и SQL»
курса «Машинное обучение и анализ данных для преподавания в
организациях высшего и среднего профессионального образования»

Исходные данные:

1. Описание предметной области для моделирования:

Разработку базы данных проведём на примере процесса обучения студентов в колледже. В колледже протекает достаточно много различных процессов, по которым нужно вести учет данных. Мы рассмотрим только основной процесс – это процесс учета успеваемости студентов.

Назовем рассматриваемую предметную область «Колледж».

Студенты в течение определенного срока (например, 4 или 3 года) изучают дисциплины в соответствии с учебным планом выбранной специальности.

Изучение каждой дисциплины имеет две стадии: приобретение знаний и контроль усвоения знаний.

Учебная часть ведет учет изучаемых дисциплин и результатов сдачи экзаменов, зачетов, курсовых работ и прочих видов контроля.

Дисциплины ведут преподаватели и мастера колледжа. Один преподаватель / мастер, как правило, ведет несколько дисциплин, а одну дисциплину могут вести несколько преподавателей.

Качество обучения характеризуется оценками, которые студенты получают во время экзаменационной сессии.

Каждый студент изучает много дисциплин и поэтому имеет много оценок.

Необходимо вести учет полученных оценок.

Введём следующие ограничения предметной области:

- срок обучения в институте по всем специальностям одинаковый – 4 года (8 семестров);

- будем учитывать дату создания группы, но не будем учитывать дату поступления в колледж конкретного студента. В реальной ситуации студент

может начать обучение в одной группе, уйти в академический отпуск и затем продолжить учиться в группе другого года набора;

- год окончания колледжа студентов приравнивается к году выпуска всей группы;

- один студент учится только в одной группе.

2. Проектирование концептуальной, логической и физической модели базы данных.

В предметной области выделим следующие сущности: цикловые комиссии, преподаватели, группы, студенты, дисциплины, оценки.

Сущности предметной области «КОЛЛЕДЖ»

Сущность	Описание
Цикловая комиссия	Структурная единица колледжа, обозначающая совокупность преподавателей одной преподаваемой предметной области. Характеризуется обозначением (код цикловой комиссии), которое может быть буквенно-символьным; полным наименованием комиссии
Преподаватель	Сведения о преподавателях и мастерах колледжа: фамилия, имя, отчество, пол, дата рождения, дата приема на работу
Дисциплина	Сведения о названиях дисциплин, изучаемых в колледже по разным специальностям, в разное время
Студент	Сведения о студентах: фамилия, имя, отчество, пол, дата рождения, группа, серия паспорта, номер паспорта, адрес регистрации: страна, город, улица, дом, квартира
Группа	Структурная единица колледжа, обозначающая совокупность студентов, обучающихся по одной специальности в один и тот же период. Характеризуется обозначением (номер группы), которое может быть буквенно-символьным; годом образования, который совпадает с годом набора студентов группы на первый курс; годом окончания обучения, который на 4 года больше года образования. Может содержать также сведения о специальности, получаемой студентом в колледже
Оценка	Сведения об оценках, полученных студентами во время обучения: сведения о студенте, дата сдачи, семестр, дисциплина, вид контроля (экзамен, зачет и пр.), оценка, сведения о преподавателе, поставившем оценку

Обучающиеся в колледже студенты объединены в группы. В одной группе учится много студентов. Один студент учится только в одной группе,

поэтому связь между сущностями «Группа» ↔ «Студент» – «один-ко-многим».

Преподаватели (мастера) колледжа распределены по цикловым комиссиям. В одной цикловой комиссии несколько преподавателей. Один преподаватель прикреплен только к одной цикловой комиссии, поэтому связь между сущностями «Цикловая комиссия» ↔ «Преподаватель» – «один-ко-многим».

Сущность «Студент» имеет так называемые составные атрибуты — «ФИО», «Адрес» и «Паспорт». Составные атрибуты затрудняют обработку данных. Для упрощения обработки удобнее разбить поле «ФИО» на отдельные поля («Фамилия», «Имя», «Отчество»); поле «Паспорт» – на поля «Серия» и «Номер».

Приведение к 1НФ в данном случае вызовет появление в таблице новых столбцов, а данные в отдельных полях опять могут дублироваться, потому что много людей живет в одном городе или на одной улице. То есть, в результате приведения к 1НФ может появиться избыточность данных и, как следствие, аномалии добавления, удаления и модификации. Избыточность данных говорит о том, что реквизиты данного отношения, возможно, не находятся в полной функциональной зависимости.

Описательные реквизиты информационного объекта логически связаны с общим для них ключом – эта связь называется функциональной зависимостью реквизитов.

Так, в сущности «Студент» атрибуты «Фамилия», «Имя», «Отчество» и пр. находятся в функциональной зависимости от ключевого реквизита «Код студента (Номер личного дела)».

Функциональная зависимость реквизитов означает, что для каждого экземпляра сущности значению его ключевого атрибута соответствует только одно значение описательного атрибута.

Такое определение функциональной зависимости позволяет при анализе всех взаимосвязей реквизитов предметной области выделить самостоятельные сущности.

В случае наличия в таблице составного ключа вводится понятие функционально полной зависимости.

Функционально полная зависимость неключевых атрибутов заключается в том, что каждый неключевой атрибут функционально зависит от ключа, но не находится в функциональной зависимости ни от какой части составного ключа. Таблица находится во второй нормальной форме, если она находится в 1НФ, и каждый неключевой атрибут функционально полно зависит от составного ключа.

Таким образом, если отношение имеет простой ключ и находится в 1НФ, то требования 2НФ удовлетворяются автоматически. А в случае составного ключа надо проверить наличие полной функциональной зависимости. Сущность «Оценки студента» находится в первой нормальной форме и имеет составной ключ: «Код студента» + «Дисциплина». Это отношение не находится во второй нормальной форме, так как атрибуты «Фамилия», «Имя», «Отчество» зависят только от номера личного дела и не зависят от дисциплины. Отсутствие полной функциональной зависимости приводит также к появлению избыточности данных и аномалиям работы с данными.

Приведение таблицы (отношения) к 2НФ осуществляется по следующим правилам:

- атрибуты, зависящие от части составного ключа, и сама эта часть выносятся в отдельную (новую) таблицу и исключаются из исходной;
- ключом новой таблицы становится часть ключа исходной, от которой имеет место функционально неполная зависимость;
- между новой и исходной таблицами устанавливается связь типа 1:M.

Приведение к 2НФ позволяет выявить в модели упущенные сущности и связи и, кроме того, устранить избыточность данных в исходной модели.

Для нормализации сущностей «Студент» и «Преподаватель» введём дополнительные сущности «Сведения о преподавателях» и «Сведения о студентах».

Реквизиты «Дисциплина» и «Оценка» можно выделить в отдельную таблицу «Успеваемость», а остальные реквизиты оставить в исходной таблице.

Кроме нормализации отношений, при построении информационно-логической модели необходимо устранить связи типа «многие-ко-многим

Между сущностями «Преподаватель» и «Дисциплина» имеется связь типа «многие-ко-многим» – один преподаватель ведет много дисциплин, одну дисциплину ведут многие преподаватели.

Устранение таких связей выполняется по следующим правилам:

- создается новая (так называемая вспомогательная) таблица;
- между исходными и новой таблицами устанавливаются связи типа 1:M;
- из модели удаляется связь M:M.

В общем случае вспомогательные таблицы могут не иметь собственных атрибутов; если это так, то определенные для них связи становятся ключевыми.

Обучающиеся в колледже студенты объединены в группы. В одной группе учится много студентов. Один студент учится только в одной группе, поэтому связь между сущностями «Группа» ↔ «Студент» – «один-ко-многим».

Преподаватели (мастера) колледжа распределены по цикловым комиссиям. В одной цикловой комиссии несколько преподавателей. Один преподаватель прикреплен только к одной цикловой комиссии, поэтому связь

между сущностями «Цикловая комиссия» ↔ «Преподаватель» – «один-ко-многим».

Один преподаватель может вести несколько дисциплин. Одну дисциплину могут вести несколько преподавателей. Поэтому связь между сущностями «Преподаватель» ↔ «Дисциплина» – «многие-ко-многим». Для установления этой связи введем вспомогательную (промежуточную) сущность «Преподаватель-Дисциплина», в которую включим ключи исходных сущностей. Таким образом, получим следующие связи: между сущностями «Преподаватель» ↔ «Преподаватель-Дисциплина» – «один-ко-многим», между сущностями «Дисциплина» ↔ «Преподаватель-Дисциплина» – «один-ко-многим».

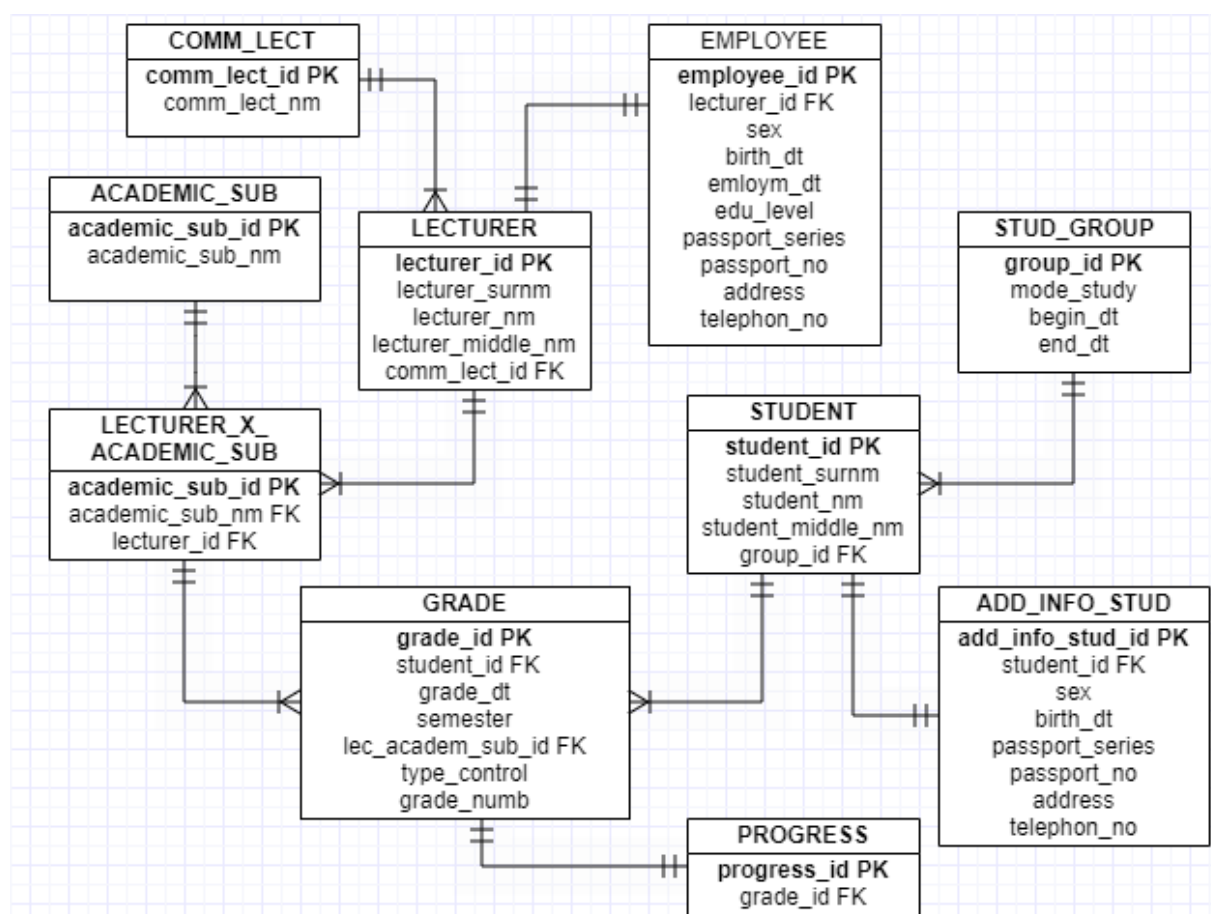
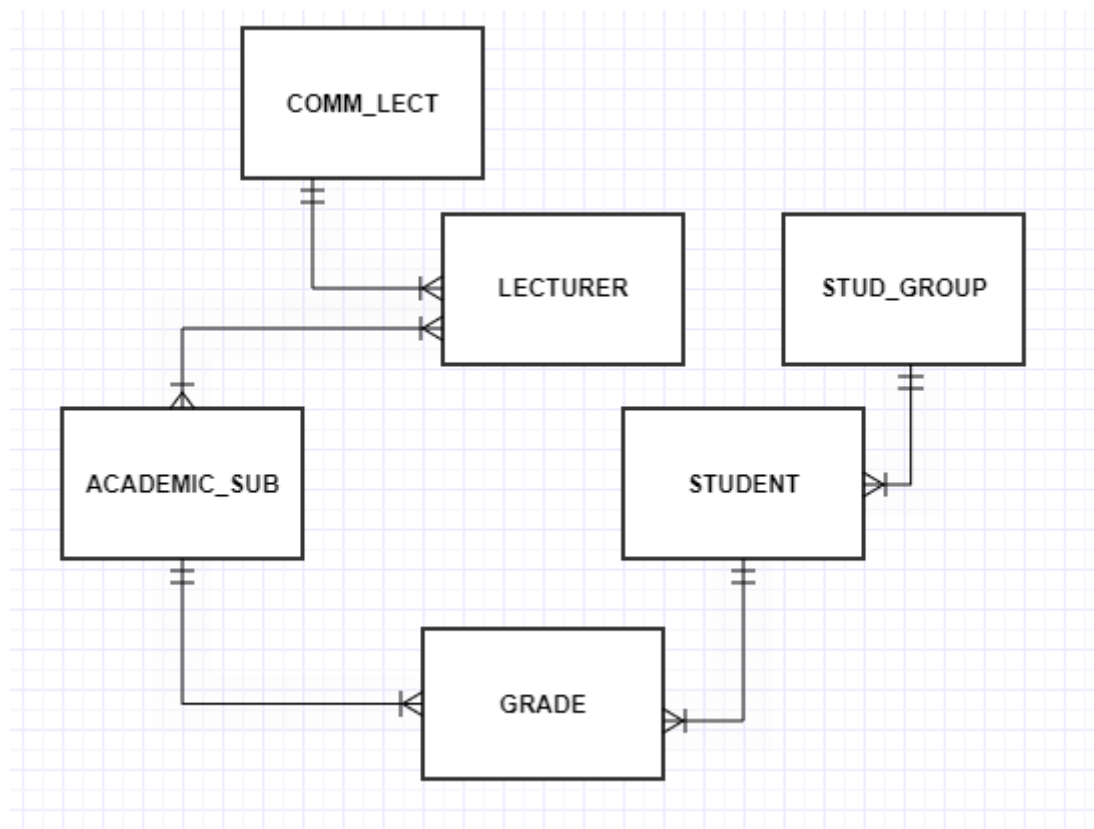
Определим связи сущности «Оценка». Оценку получает студент по изученной дисциплине. Но дисциплину преподает конкретный преподаватель. Сведения о дисциплинах, которые ведет каждый преподаватель, учитываются в таблице «Преподаватель-Дисциплина». Поэтому сущность «Оценка» имеет связи с сущностями «Студент», «Преподаватель-Дисциплина».

Один студент получает много оценок. Но каждая полученная оценка относится только к одному студенту, поэтому связь между сущностями «Студент» ↔ «Оценка» – «один-ко-многим».

Каждый студент имеет оценки по разным учебным дисциплинам, добавим дополнительно сущность «Успеваемость». Связь между сущностями «Студент» ↔ «Успеваемость» – «один-к-одному».

Один преподаватель по дисциплине выставляет много оценок. В то же время каждая конкретная оценка относится к одному конкретному экземпляру сущности «Преподаватель-Дисциплина». Поэтому имеет место связь между сущностями «Преподаватель-Дисциплина» ↔ «Оценка» – «один-ко-многим».

На основании выявленных связей получаем концептуальную и логическую модели:



Выделим атрибуты сущностей предметной области «КОЛЛЕДЖ» и спроектируем физическую модель базы данных:

Название сущности	Название атрибута		Описание атрибута	Тип данных	ОГРАНИЧЕНИЕ
Цикловая комиссия COMM_LECT	Код комиссии	comm_lect_id	Код обозначения комиссии (до 4 букв)	VARCHAR(4)	NOT NULL PK
	Полное название	comm_lect_nm	Наименование цикловой комиссии	VARCHAR(80)	
Преподаватель LECTURER	Код преподавателя	lecturer_id	Код преподавателя в виде персонального номера (6 цифр)	VARCHAR(6)	NOT NULL PK
	Фамилия	lecturer_surnm	Фамилия	VARCHAR(20)	
	Имя	lecturer_nm	Имя	VARCHAR(20)	
	Отчество	lecturer_middle_nm	Отчество	VARCHAR(20)	
	Код комиссии	comm_lect_id	Код обозначения комиссии (до 4 букв)	VARCHAR(4)	FK Связь с сущностью «Цикловая комиссия»
Сотрудник EMPLOYEE	Код сотрудника	employee_id	Код сотрудника в виде номера личного дела (6 цифр, символ «/» и 4 цифры года приёма на работу)	VARCHAR(11)	NOT NULL PK
	Код преподавателя	lecturer_id	Код преподавателя в виде персонального номера (6 цифр)	VARCHAR(6)	FK Связь с сущностью «Преподаватель»
	Пол	sex	Наименование пола («мужской» или «женский»)	VARCHAR(10)	
	Дата рождения	birth_dt	Дата рождения	DATE	
	Дата приема	emloym_dt	Дата приема на работу	DATE	
	Образование	edu_level	Уровень образования (высшее, неполное высшее, среднее профессиональное, начальное профессиональное и пр.)	VARCHAR(50)	
	Серия паспорта	passport_series	Серия паспорта	VARCHAR(4)	
	Номер паспорта	passport_no	Номер паспорта	VARCHAR(6)	

Название сущности	Название атрибута		Описание атрибута	Тип данных	ОГРАНИЧЕНИЕ
	Адрес	address	Адрес места проживания студента в виде списка: страна, регион, населённый пункт, улица, дом, квартира	VARCHAR(100)	
	Номер телефона	telephon_no	Номер телефона	VARCHAR(20)	
Дисциплина ACADEMIC_SUB	Код дисциплины	academic_sub_id	Код обозначения дисциплины (аббревиатура до 4 букв, символ «.» и двузначное число)	VARCHAR(7)	NOT NULL PK
	Название	academic_sub_nm	Наименование учебной дисциплины/ междисциплинарного курса / профессионального модуля согласно учебному плану	VARCHAR(50)	
Преподаватель – Дисциплина LECTURER_X_ ACADEM_SUB	Код ПД	lect_x_academ_sub_id	Код связи преподавателя и дисциплины	INT	NOT NULL PK
	Код дисциплины	academic_sub_id	Код обозначения дисциплины (аббревиатура до 4 букв, символ «.» и двузначное число)	VARCHAR(7)	FK Связь с сущностью «Дисциплина»
	Код преподавателя	lecturer_id	Код преподавателя в виде номера личного дела (6 цифр, символ «/» и 4 цифры года приёма на работу)	VARCHAR(11)	FK Связь с сущностью «Преподаватель»
Студент STUDENT	Код студента	student_id	Код студента в виде номера личного дела (6 цифр, символ «/» и 4 цифры года поступления)	VARCHAR(11)	NOT NULL PK
	Фамилия	student_surnm	Фамилия	VARCHAR(20)	
	Имя	student_nm	Имя	VARCHAR(20)	
	Отчество	student_middle_nm	Отчество	VARCHAR(20)	
	Код группы	group_id	Номер группы (4 цифры)	INT	FK Связь с сущностью «Группа»

Название сущности	Название атрибута		Описание атрибута	Тип данных	ОГРАНИЧЕНИЕ
Сведения о студенте ADD_INFO_STUD	Код сведений	add_info_stud_id	Код записи персональных сведений о студенте в виде шестизначного числа	INT	NOT NULL PK
	Код студента	student_id	Код студента в виде номера личного дела (6 цифр, символ «/» и 4 цифры года поступления)	VARCHAR(11)	FK Связь с сущностью «Студент»
	Пол	sex	Пол	VARCHAR(10)	
	Дата рождения	birth_dt	Дата рождения	DATE	
	Серия паспорта	passport_series	Серия паспорта в виде четырёхзначного числа	INT	
	Номер паспорта	passport_no	Серия паспорта в виде четырёхзначного числа	INT	
	Адрес	address	Адрес места проживания студента в виде списка: страна, регион, населённый пункт, улица, дом, квартира	VARCHAR(100)	
	Номер телефона	telephon_no	Номер телефона	VARCHAR(20)	
Группа STUD_GROUP	Код группы	group_id	Обозначение группы (аббревиатура специальности и номер)	VARCHAR(10)	NOT NULL PK
	Форма обучения	mode_study	Список: очная, очно-заочная, заочная	VARCHAR(20)	
	Дата образования	begin_dt	Дата создания группы приказом о зачислении студентов и распределении по группам	DATE	
	Дата окончания	end_dt	Дата окончания обучения студентов группы в виде четырёхзначного числа от даты окончания срока обучения	DATE	
Оценка GRADE	Код оценки	grade_id	Код записи выставленной оценки	INT	NOT NULL PK

Название сущности	Название атрибута		Описание атрибута	Тип данных	ОГРАНИЧЕНИЕ
	Код студента	student_id	Код студента в виде номера личного дела (6 цифр, символ «/» и 4 цифры года поступления)	VARCHAR(11)	FK Связь с сущностью «Студент»
	Дата	grade_dt	Дата получения оценки	DATE	
	Семестр	semester	Номер семестра обучения в виде целого числа от 1 до 8	SMALLINT	
	Код ПД	lect_x_academ_sub_id	Код связи преподавателя и дисциплины	INT	FK Связь с сущностью «Преподаватель-дисциплина»
	Вид контроля	type_control	Вид контроля (экзамен, зачёт, курсовая работа, отчёт)	VARCHAR(20)	
	Оценка	grade_numb	Результат оценивания в виде балла: 0, 2, 3, 4, 5	SMALLINT	
Успеваемость PROGRESS	Код записи	progress_id	Код записи в ведомости успеваемости	INT	NOT NULL PK
	Код оценки	grade_id	Код записи выставленной оценки	INT	FK Связь с сущностью «Оценка»