

Модели данных

А2. Логические модели данных



Московский государственный университет
имени Н.Э. Баумана

Факультет ИБМ

Июль 2024 года

Москва

Артемьев Валерий Иванович © 2024

Курс «Моделирование данных»

2. Логические модели данных

- Определение логической модели данных
- Этапы создания логической модели данных
- Именование таблиц и атрибутов ЛМД
- Виды атрибутов
- Типы и домены данных
- Необязательные значения
- Ключевые атрибуты
- Нотация логической модели данных
- Нормализация реляционных моделей данных
- 1-ая нормальная форма
- 2-ая нормальная форма
- 3-ья нормальная форма
- Прочие нормальные формы
- Логическая модель данных «Домашняя библиотека» (карточка книги)
- Создание диаграммы «сущность-связь»
- Проверка выполнения правил нормализации
- Оформление таблицы описания логической модели данных
- Пример заполнения таблиц на Excel
- Вопросы для самопроверки



Логическая модель данных

Детальная модель данных отражает необходимые потребности бизнеса и ограничения реляционной абстракции в терминах предметной области, не зависит от конкретной реализации в виде базы данных.

- Не содержит многосторонних и многозначных связей
- Детализирует содержание, добавляя все атрибуты для сущностей
- Детализирует структуру данных, определяя ключевые атрибуты связи
- Рекомендуемая документация по информационному обеспечению
- Средство коммуникации заказчиков и разработчиков приложений
- Позволяет избежать ошибок по содержанию и структуре БД
- Ещё не отражает специфику конкретной системы управления БД

Шаги создания логической модели данных



- | | | | | |
|---|--|--|---|--|
| – Заготовка диаграммы ЛМД на основе КМД | – Анализ КМД для выявления сложных связей
– Преобразование n-арных связей в бинарные
– Декомпозиция связей «многие ко многим». | – Анализ предметной области и требований
– Выбор правил именования
– Определение и именование атрибутов
– Детализация диаграммы
– Выявление ключевых атрибутов
– Назначение идентификаторов и ссылок связей
– Выявление необязательных атрибутов
– Разметка атрибутов на диаграмме. | – Проверка соответствия ЛМД нормальным формам:
1NF
2NF
3NF
– Устранение аномалий ЛМД. | – Окончательное оформление диаграммы ЛМД
– Подготовка таблицы описания ЛМД. |
|---|--|--|---|--|

Подготовительный этап

Вариант 1 (ранее создана КМД)

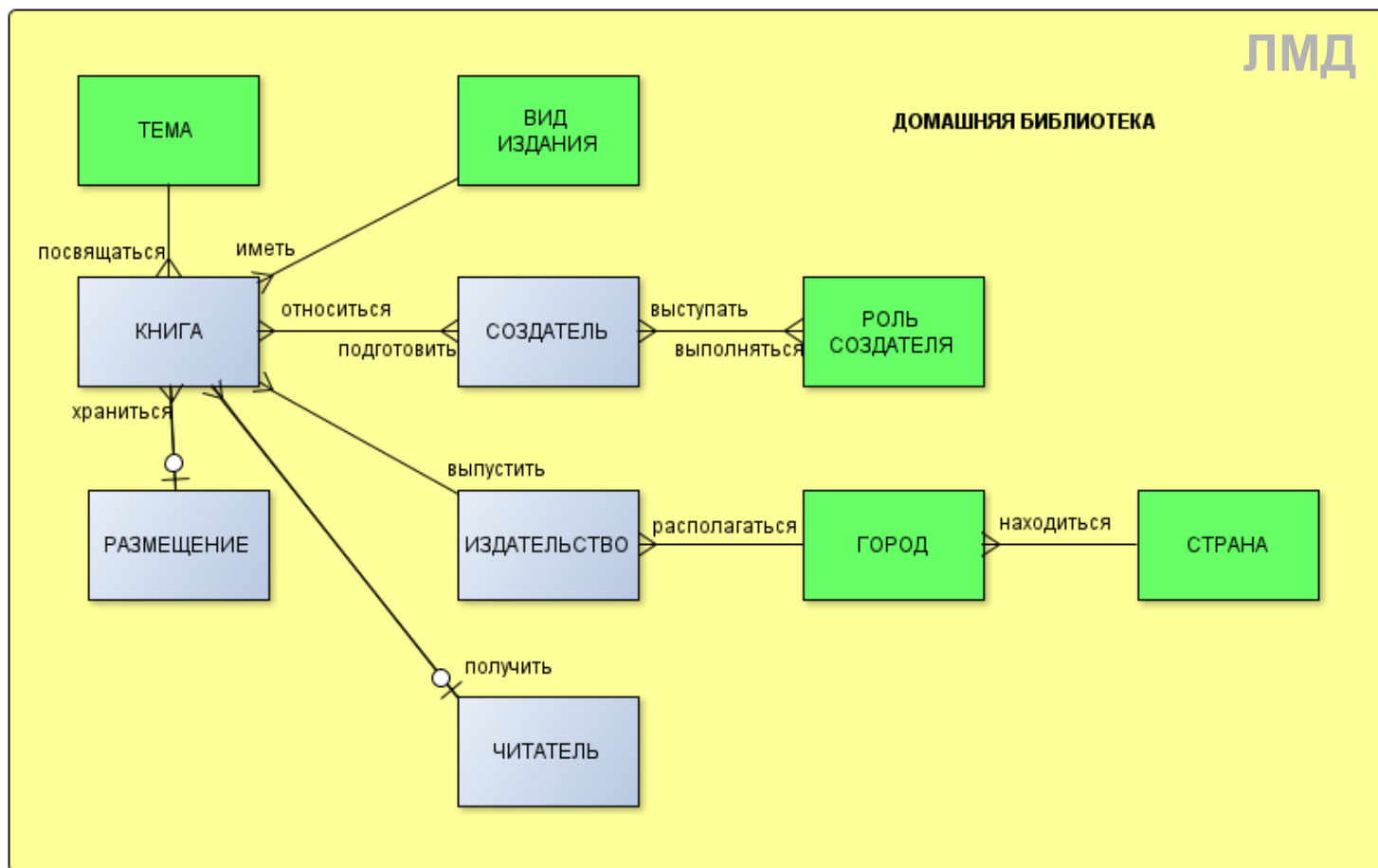
- Заготовка диаграммы ЛМД на основе КМД:
 - открыть ER-диаграмму КМД в редакторе диаграмм
 - сохранить диаграмму с именем, выбранным для ЛМД.

Вариант 2 (ранее не была создана КМД)

- Выполнить действия:
 - анализ требований и предметной области
 - нормализация и выделение бизнес-сущностей и связей
 - создать ER-диаграмму и таблицы описания КМД.

Пример. Подготовительный этап

Заготовка диаграммы ЛМД на основе КМД

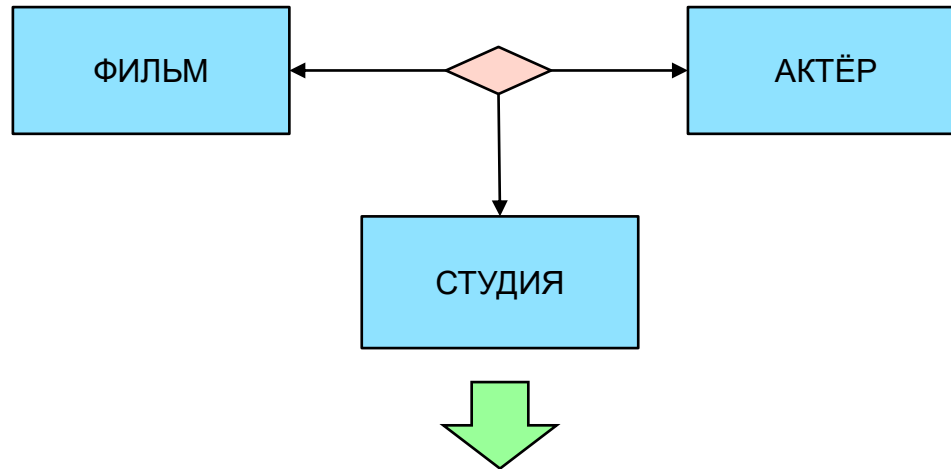


Преобразование неподдерживаемых типов связей

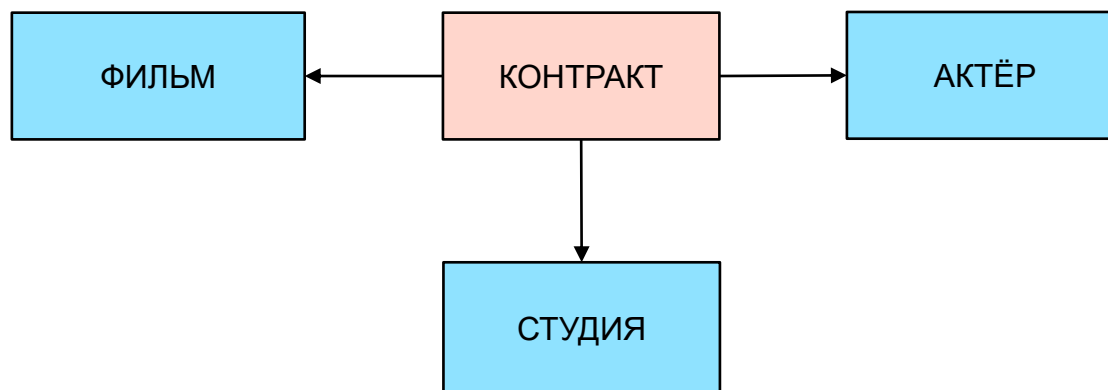
- Анализ КМД для выявления сложных связей:
 - многосторонние связи (n-арки)
 - связи «многие ко многим»
- Преобразование n-арных связей в бинарные:
 - добавить новую сущность для представления многосторонних связей
 - заменить каждую связь n-арки на бинарную связь того же типа с новой сущностью
- Декомпозиция многозначных связей:
 - добавить новую сущность вместо связи «многие ко многим»
 - установить связи «многие к одному» новой сущности с двумя исходными сущностями

Преобразование многосторонних связей в бинарные

Многосторонняя связь до преобразования модели данных

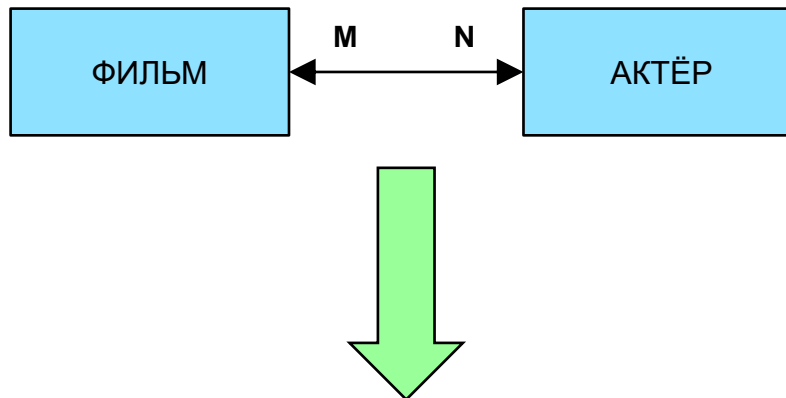


Добавление сущности для представления многосторонних связей

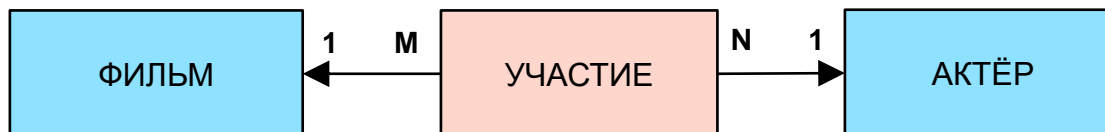


Декомпозиция многозначных связей

Многозначная связь до преобразования модели данных



Добавление сущности для представления многозначной связи

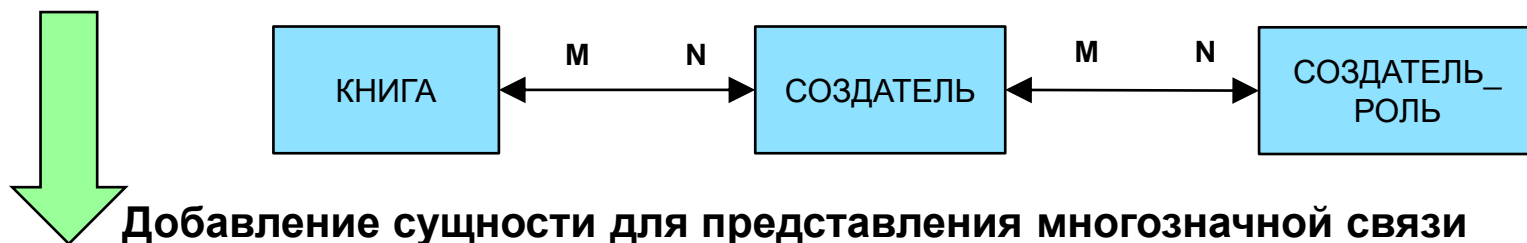


Пример. Преобразование неподдерживаемых типов связей

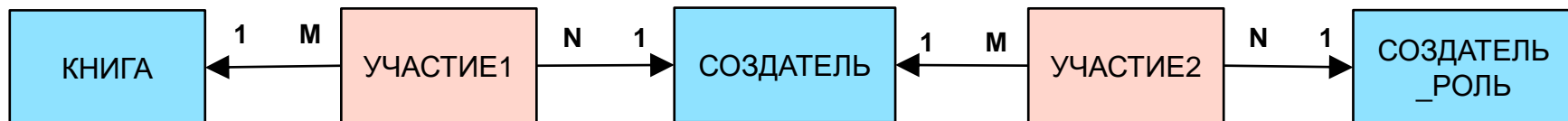
- Анализ КМД показал:
 - нет многосторонних связей, все связи бинарные
 - наличие многозначных связей:
 - КНИГА – СОЗДАТЕЛЬ
 - КНИГА – РОЛЬ_СОЗДАТЕЛЯ
- Декомпозируем связи типа «многие ко многим»:
 - добавим новую сущность УЧАСТИЕ для многозначной связи
 - установить связи «многие к одному» новой сущности с сущностями КНИГА и СОЗДАТЕЛЬ
 - связь КНИГА – РОЛЬ_СОЗДАТЕЛЯ перенесём к сущности УЧАСТИЕ

Пример. Декомпозиция многозначных связей

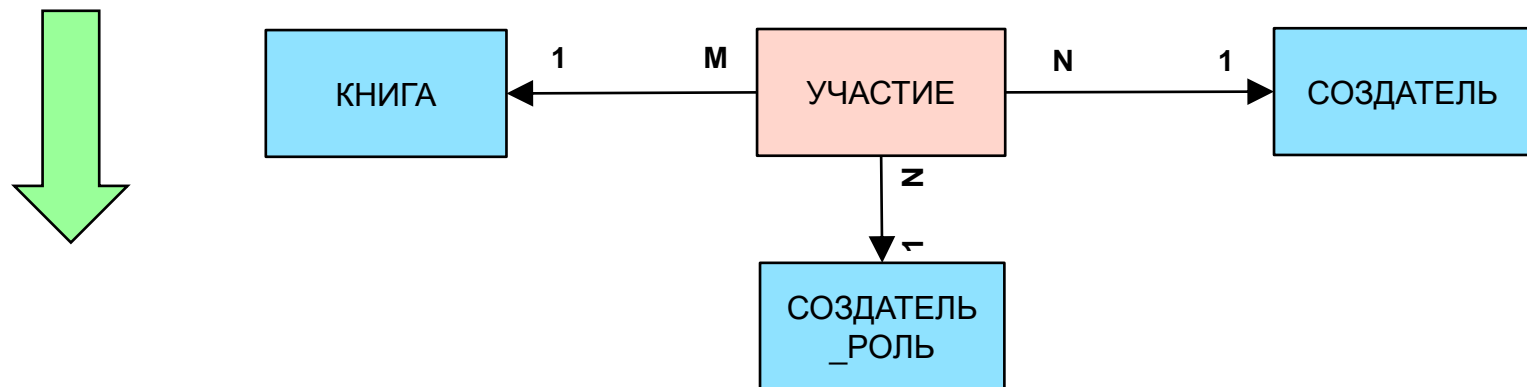
Многозначная связь до преобразования модели данных



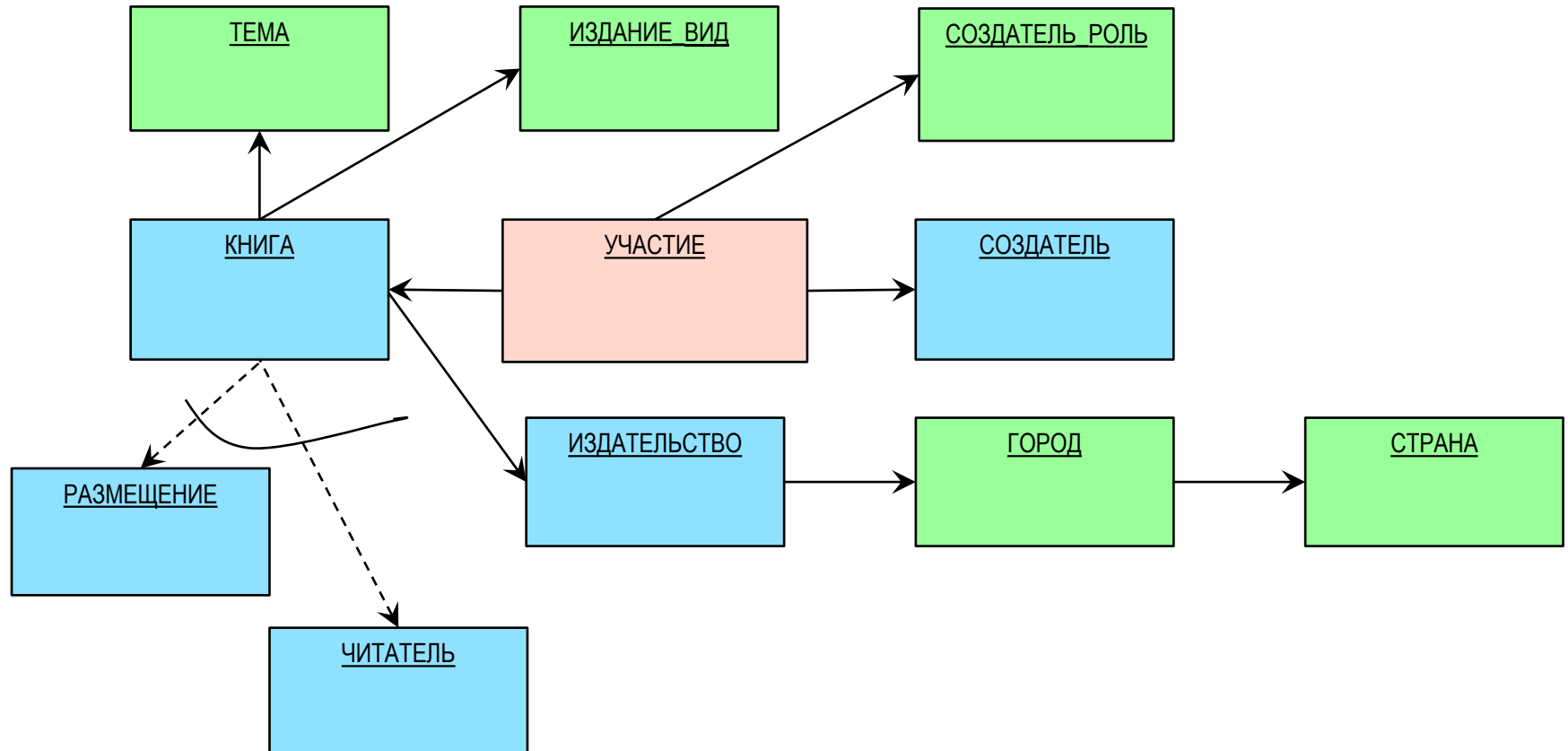
Добавление сущности для представления многозначной связи



Слияние промежуточных сущностей для представления многозначной связи



Пример. Диаграмма ЛМД после преобразования сложных связей



Определение атрибутов и ключей

- Анализ предметной области и требований
- Выбор правил именования
- Определение и именование атрибутов
- Детализация диаграммы
- Выявление ключевых атрибутов (ключей)
- Назначение и именование идентификаторов и ссылок
- Выявление необязательных атрибутов
- Разметка атрибутов на диаграмме.

Что такое атрибут?

Атрибут – свойство сущности данных, его значения служат для идентификации, описания и измерения экземпляров сущности, а также определяют его связь с экземплярами других сущностей.

- Измерения, количества:
 - количество (сколько?)
 - размер (насколько широко? и т.п.)
 - денежная сумма (сколько стоит?)
 - фактор (какая доля?)
 - момент времени (когда?)
 - периодичность (как часто?)
 - длительность (как долго?)
 - место (где? куда? откуда?)
- Категория, признак
 - коды, мнемокоды (какой код вида/ типа?)
 - наименования категорий/ классов (как называется вид/ тип?)
 - логические значения (да/ нет?)
- Описания — пояснения и примечания (что собой представляет?)
- Идентификатор — уникальный код, имя, название, наименование (кто? что?)
- Ссылка — указание идентификатора

Виды атрибутов

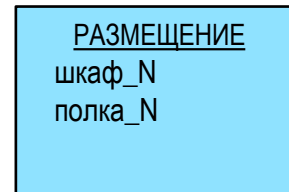
- Простые и составные атрибуты
 - год издания книги: *2013*
 - фамилия, имя и отчество автора книги: *Эмиль Золя*
 - адрес издательства: *Москва, ул. Соколиной горы, дом 5, корп. 3*
- Однозначные и многозначные атрибуты
 - номер телефона сотрудника: *678-23-41*
 - список номеров телефонов: *678-23-41, 678-25-12*
- Обязательные атрибуты
 - должны иметь значения
- Необязательные атрибуты
 - могут не иметь значения
- Ключевые атрибуты
 - идентификатор книги
 - ссылка на идентификатор книги
 - комбинация атрибутов для идентификации.

Именованние атрибутов ЛМД

- Имена пишутся по-русски
- Имена сущностей и атрибутов не равны их описанию
- Лаконичные или подробные имена:
 - атрибут название с полным именем КНИГА.название
 - атрибут название_книги (КНИГА.название_книги)
- Формат имени
 - ✓ змейка (snake_case): год_издания
 - ✗ верблюд (camelCase): годИздания
- Порядок слов в имени
 - прямой порядок: год_издания
 - обратный порядок: издание_год
- Сокращение слов в составном имени
 - усечение слов, но хотя бы одно должно быть полное
- Применение сокращений: только общепринятые
- Правила именования <https://www.sqlstyle.guide/ru/>
- Частные случаи именования ключевых атрибутов (см. ниже)

Пример. Определение и именование атрибутов

1. В качестве кандидатов в атрибуты определены выделенные в задании термины
2. Выбираем полное именование атрибутов «змейкой» строчными буквами, с обратным порядком слов
3. В блоке сущности вместо полочки для названия сущности можно просто подчёркивать название
4. Название каждого атрибута занимает отдельную строку в блоке сущности



Пример. Анализ требований

Создать каталог домашней библиотеки

■ Атрибуты поиска книги:

- Тема, вид издания
- Название издания и/ или ключевые слова
- Фамилия, имя (инициалы) автора, редактора, составителя, переводчика или художника
- Название издательства и местонахождение
- Год выпуска.

Выделены кандидаты
в атрибуты сущностей

■ Результаты поиска:

- Хранимые атрибуты книги (включая аннотацию, сведения об оригинале перевода)
- Место хранения издания (шкаф, полка)
- Текущий держатель (читатель) издания, его контакты

■ В библиотеке каждая книга присутствует в *одном экземпляре*

Пример. Анализ предметной области

УДК 004.65
ББК 32.973.26-018.1
К89

С.Д. Кузнецов

Базы данных: языки и модели. Учебник — М.: ООО «Бином-Пресс», 2008 г. — 720 с.: ил.

Книга является учебником по современным моделям и языкам баз данных и частично основана на материалах курса, читающегося автором более десяти лет на факультете вычислительной математики и кибернетики Московского государственного университета. Предназначается для студентов старших курсов и аспирантов классических и технических университетов, специализирующихся в области информационных технологий, также для преподавателей, читающих различные курсы, связанные с тематикой баз данных. Книга может использоваться и для самообразования специалистами, желающими лучше узнать современное состояние технологии баз данных. Книга опирается на материал многочисленных классических и недавно изданных монографий, статей и текстов стандартов.

Основные главы книги посвящены реляционной модели данных (в современной

© Кузнецов С.Д., 2008

© Издательство Бином,

ISBN 978-5-9518-0132-6

Учебное издание

Кузнецов Сергей Дмитриевич

Базы данных: языки и модели

Оформление обложки *И.Ю. Буровой*
Компьютерная верстка *В.В. Краснова*

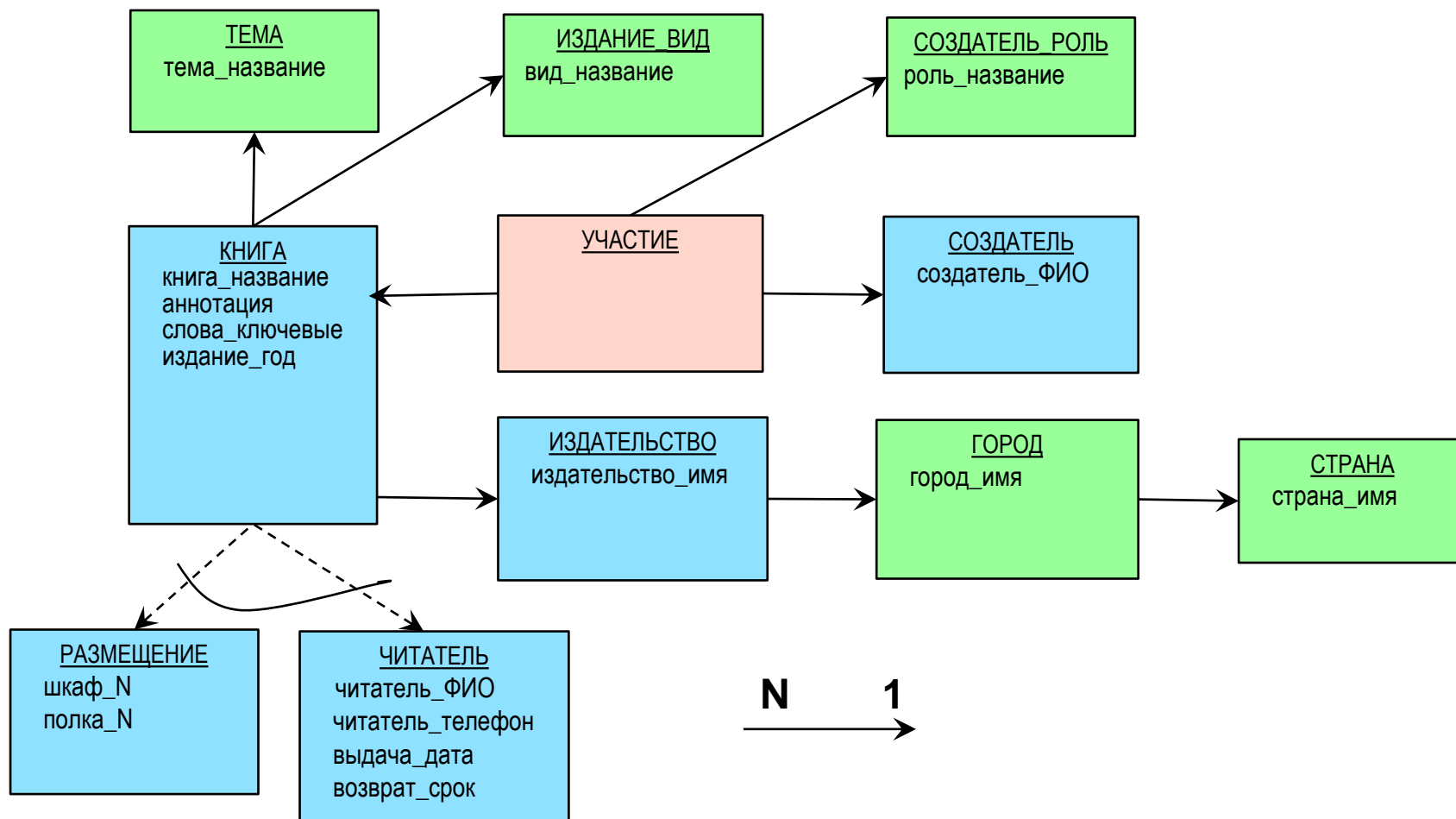
Подписано в печать 27.04.2007. Формат 70×100/16. Усл. печ. л. 58,5.

Гарнитура Петербург. Бумага газетная. Печать офсетная.

Тираж 3000 экз. Заказ 6275

Издательство «Бином-Пресс», 2008 г.
141077, Королев, Московской обл., ул. 50 лет ВЛКСМ, 4-Г

Пример. Детализация диаграммы ЛМД



Что такое ключевые атрибуты?

Атрибуты, используемые для обеспечения уникальной идентификации экземпляров сущности и связывания экземпляров сущностей между собой.



Идентификатор – адрес дома

Идентификаторы и ссылки
Простой и составной ключи
Естественные и суррогатные ключи
Первичный ключ (идентификатор)
Альтернативные ключи
Внешний ключ (ссылка)



Адресная ссылка на дом

Именованние ключевых атрибутов (идентификаторов и ссылок)

- **Вариант 1** (простые имена идентификаторов, ссылки отличаются)
 - Идентификатор любой сущности получает простое имя: **ID**
 - Такие атрибуты различают по полному имени в запросах: **ТЕМА.ID**
 - Имя ссылки на такой идентификатор включает имя сущности, на которую она ссылается: **тема_ID**
- **Вариант 2** (имена одинаковые, обнаружение связей)
 - В имя атрибута идентификатора включают имя сущности: **тема_ID**
 - Ссылка на такой идентификатор имеет такое же имя: **тема_ID**
 - Такие атрибуты различают по полному имени в запросах:
ТЕМА.тема_ID – идентификатор темы; **КНИГА.тема_ID** – ссылка на тему
- **Вариант 3** (имена идентификаторов и ссылок различимы по суффиксам)
 - Имя идентификатора включает имя сущности: **тема_ID**
 - Имя ссылки имеет вид: **тема_REF**
 - В запросах можно не использовать полные имена ключевых атрибутов

Пример. Выявление ключевых атрибутов

- Какие комбинации атрибутов могут уникально идентифицировать экземпляры сущностей?
 - КНИГА: книга_название, год_издания (составной ключ)
 - ЧИТАТЕЛЬ: читатель_телефон (простой ключ)
- Не так много естественных уникальных ключей
- Поэтому используем для всех бизнес-сущностей искусственные – простые суррогатные идентификаторы (ключи)
- Выбираем одинаковые имена идентификаторов и ссылок на них: книга_ID; тема_ID; издание_вид_ID; издательство_ID; создатель_ID
- Для сущности-моста УЧАСТИЕ, связывающей 3 сущности, в качестве составного ключа возьмём 3 ссылки, составляющих содержание этой сущности: книга_ID; создатель_ID; создатель_роль_ID

Необязательные значения

Специальное значение атрибута или визуальное обозначение, указывающее на отсутствие значимых данных у атрибута по разным причинам.

- значение отсутствует
 - значение неизвестно
 - значение неопределенно
 - значение недопустимо в этом контексте и т.п.
- Свойство обязательности значения атрибута говорит о необходимости его присутствия
- В базах данных имеется необязательное значение Null
- Для строк часто используется пустая строка
- Для чисел в языках программирования применяется Nan
- Визуально оформляется по разному: “–” или “X”

Что может означать необязательное значение?

- Смысл необязательного значения сильно зависит от контекста, нужно различать разные варианты отсутствия данных.

Пусть имеем сущности:

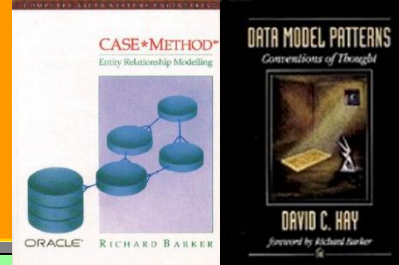
ПЕРСОНА (персона_ID, персона_Имя, подразд_ID)

ПОДРАЗДЕЛЕНИЕ (подразд_ID, подразд_Имя)

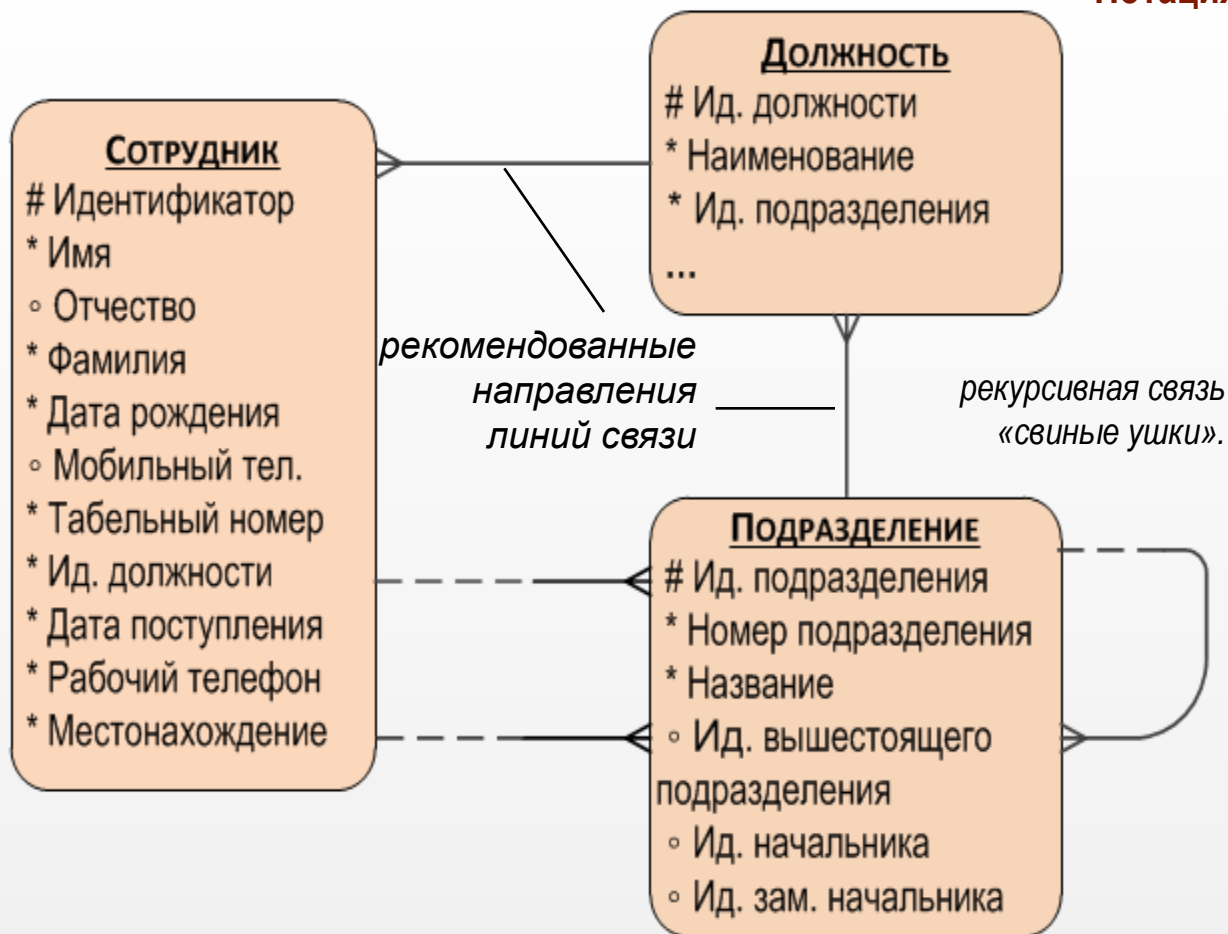
Что может означать необязательное значение в ПЕРСОНА.подразд_ID?

- **Неизвестно** – нет сведений, из какого подразделения персона
 - **Неверно** – неправильная ссылка на подразделение (ошибка)
 - **Не существует ещё или уже** – сотрудник принят, но подразделение ещё не назначено, или он уже уволен, или реорганизация
 - **Не имеет смысла** – персона не работает в этой организации
 - **Недоступно** – недостаточно прав узнать группу.
- Может быть результатом множества операций
 - Можно обойтись без них, усложняя модель данных.

Нотация ER-диаграммы логической модели данных



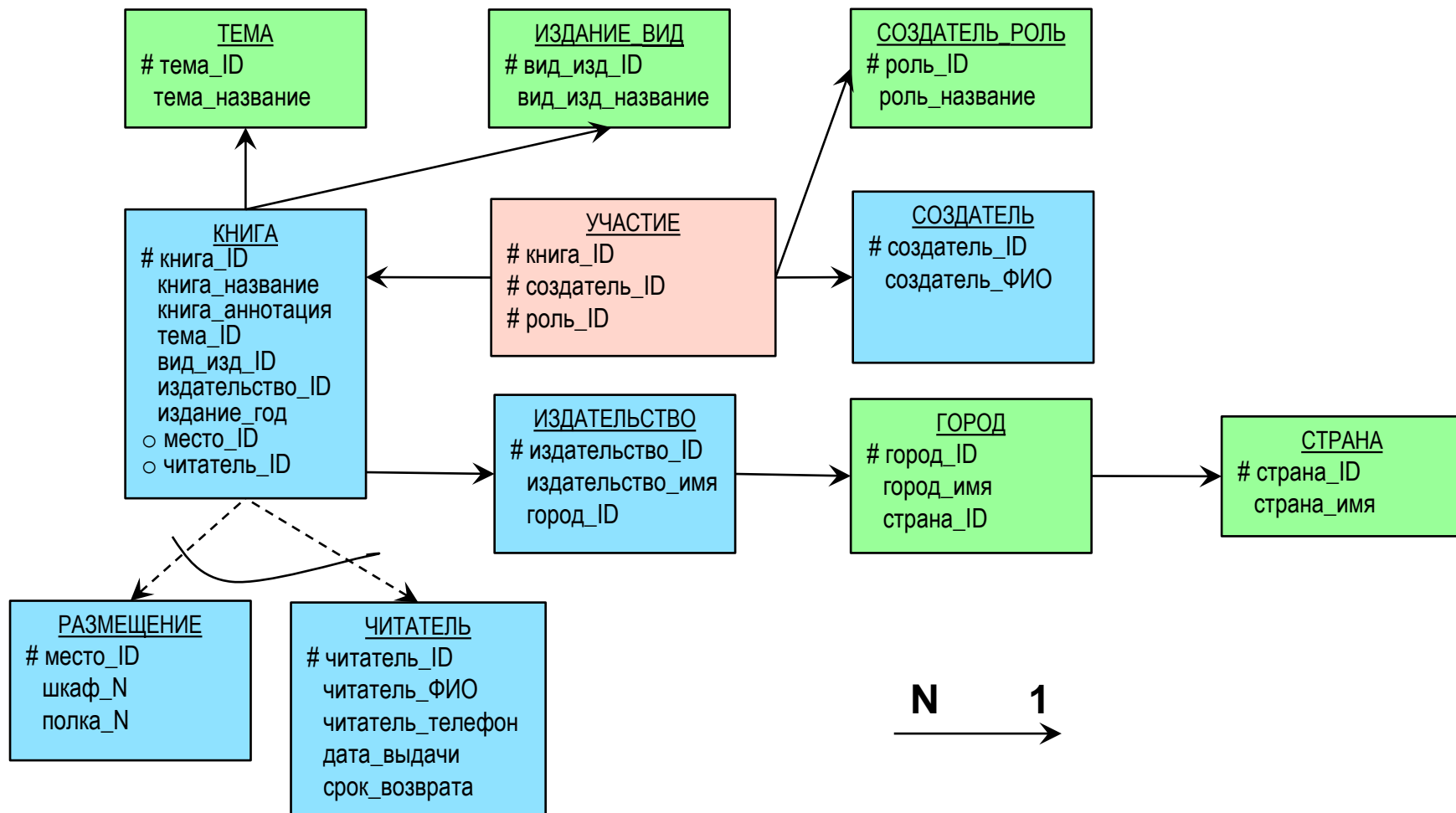
Нотация Р. Баркера



- Данные разной детальности по разным сущностям
- Не допустимы связи «многие ко многим».
- Атрибут должен иметь одно значение.
- Имя сущности в единственном числе по-русски на полочке или подчёркнуто.
- В блоке сущности задаются атрибуты, перед ними знаки:
 - # -- идентификатор
 - * -- обязательный атрибут
 - ° -- необязательный атрибут
- Допустимо не размечать атрибуты.
- Не рекомендуется указывать типы / домены атрибутов.
- Именованние атрибутов: без повторения или с повтором имени сущности.

Для простоты предлагается отмечать только необязательные атрибуты.

Пример. Логическая модель данных «Домашняя библиотека»



Табличное описание логической модели данных

- Имя сущности данных
- Имя атрибута
- Описание атрибута
- Тип ключевого атрибута
(идентификатор/ссылка)
- Обязательность значения (да/нет)
- [Примеры значения атрибута]

Пример. Таблица описания логической модели данных

Имя сущности	Имя атрибута	Описание атрибута	Тип ключевого атрибута	Обязательное значение
КНИГА	книга_ID	Идентификатор Реестра КНИГА	идентификатор	да
КНИГА	книга_название	Название книги		да
КНИГА	книга_аннотация	Аннотация книги		да
КНИГА	тема_ID	Ссылка на позицию Классификатора ТЕМА (тема книги)	ссылка	да
КНИГА	издание_вид_ID	Ссылка на позицию Классификатора ИЗДАНИЕ_ВИД (вид издания)	ссылка	да
КНИГА	издательство_ID	Ссылка на позицию Реестра ИЗДАТЕЛЬСТВО	ссылка	да
КНИГА	издание_год	Год издания книги		да
КНИГА	размещение_ID	Ссылка на позицию Реестра РАЗМЕЩЕНИЕ (места хранения книг). Если ссылка пустая, то должна быть непустая ссылка на читателя.	ссылка	нет
КНИГА	читатель_ID	Ссылка на позицию Реестра ЧИТАТЕЛЬ. Если ссылка пустая, то должна быть непустая ссылка на размещение книги.	ссылка	нет
ТЕМА	тема_ID	Идентификатор Классификатора ТЕМА	идентификатор	да
ТЕМА	тема_название	Название темы книги		да
ИЗДАНИЕ_ВИД	вид_изд_ID	Идентификатор Классификатора ИЗДАНИЕ_ВИД		да
ИЗДАНИЕ_ВИД	вид_изд_название	Название вида издания		да
СОЗДАТЕЛЬ	создатель_ID	Идентификатор Реестра СОЗДАТЕЛЬ	идентификатор	да
СОЗДАТЕЛЬ	создатель_ФИО	Фамилия, имя и отчество или инициалы создателя (редактора, автора, переводчика, художника и т.п.)		да
СОЗДАТЕЛЬ_РОЛЬ	роль_ID	Идентификатор Классификатора СОЗДАТЕЛЬ_РОЛЬ	идентификатор	да
СОЗДАТЕЛЬ_РОЛЬ	роль_название	Название роли создателя: редактора, автора, переводчика, художника и т.п.		да
УЧАСТИЕ	книга_ID	Часть составного ключа участия и ссылка на книгу	часть ключа, ссылка	да
УЧАСТИЕ	создатель_ID	Часть составного ключа участия и ссылка на создателя	часть ключа, ссылка	да
УЧАСТИЕ	роль_ID	Часть составного ключа участия и ссылка на роль создателя	часть ключа, ссылка	да
ИЗДАТЕЛЬСТВО	издательство_ID	Идентификатор Реестра ИЗДАТЕЛЬСТВО	идентификатор	да
ИЗДАТЕЛЬСТВО	издательство_имя	Наименование издательства		да
ИЗДАТЕЛЬСТВО	город_ID	Ссылка на позицию Классификатора ГОРОД	ссылка	да
ГОРОД	город_ID	Идентификатор Классификатора ГОРОД	идентификатор	да
ГОРОД	город_имя	Наименование города		да
ГОРОД	страна_ID	Ссылка на позицию Классификатора СТРАНА	ссылка	да
СТРАНА	страна_ID	Идентификатор Классификатора СТРАНА	идентификатор	да
СТРАНА	страна_имя	Наименование страны		да
РАЗМЕЩЕНИЕ	размещение_ID	Идентификатор Реестра РАЗМЕЩЕНИЕ	идентификатор	да
РАЗМЕЩЕНИЕ	шкаф_N	Номер шкафа (стеллажа) размещения книги		да
РАЗМЕЩЕНИЕ	полка_N	Номер полки размещения книги		да
ЧИТАТЕЛЬ	читатель_ID	Идентификатор Реестра ЧИТАТЕЛЬ	идентификатор	да
ЧИТАТЕЛЬ	читатель_ФИО	Фамилия, имя и отчество или инициалы читателя		да
ЧИТАТЕЛЬ	читатель_телефон	Номер телефона читателя		да
ЧИТАТЕЛЬ	дата_выдачи	Дата выдачи книги читателю		да
ЧИТАТЕЛЬ	срок_возврата	Информация о сроке возврата		

**Терпения и удачи всем, кто связан
с моделированием данных**

Спасибо за внимание!

Валерий Иванович Артемьев

МГТУ имени Н.Э. Баумана, кафедра ИУ-5

Банк России

Департамент данных, проектов и процессов

Тел.: +7(495) 753-96-25

e-mail: viart@bmstu.ru