

# Модели данных

## А3. Физические модели данных



Московский государственный технический университет  
имени Н.Э. Баумана

**Факультет ИБМ**

сен 2024 года

Москва

Артемьев Валерий Иванович © 2024

# Курс «Моделирование данных»

## 3. Физические модели данных

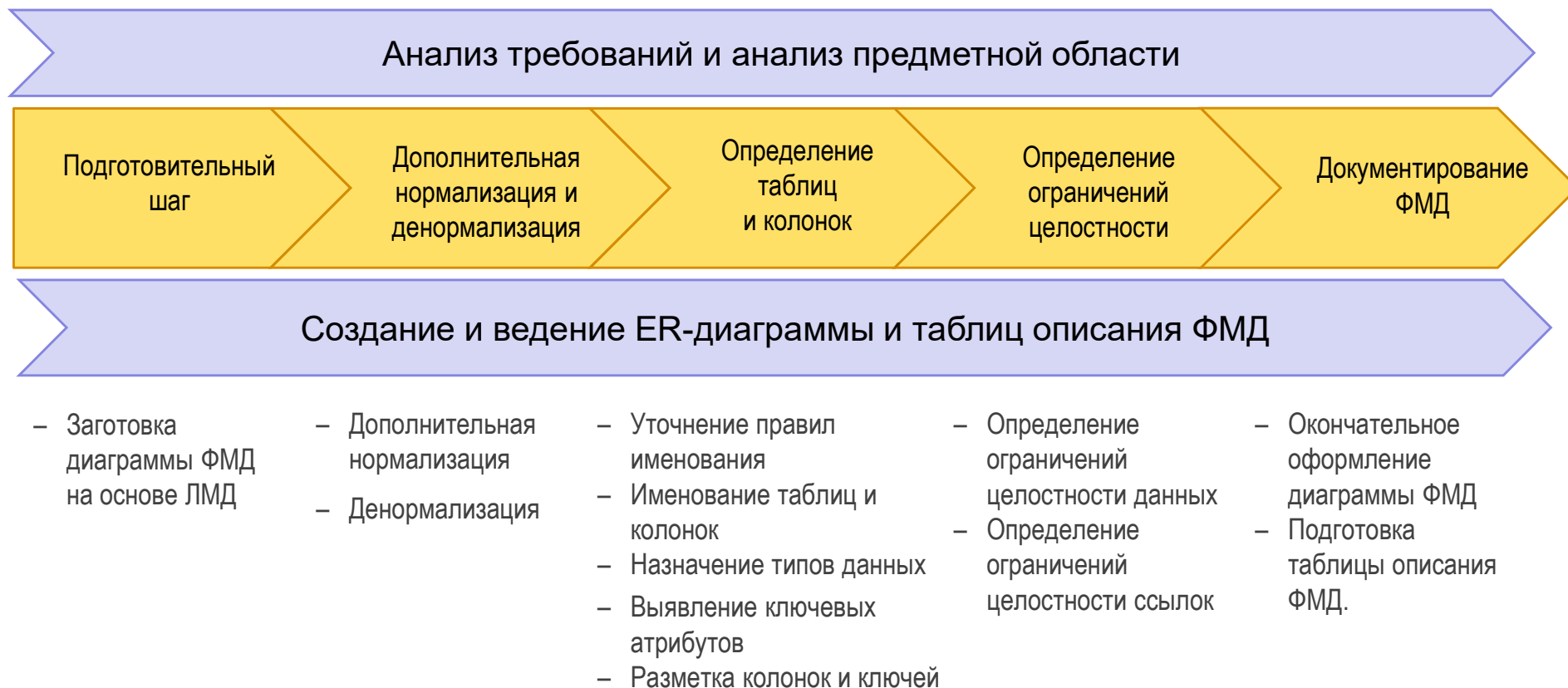
- Определение физической модели данных
  - Шаги создания физической модели данных
  - Подготовительный шаг
- 
- Именованное таблиц и колонок ФМД
  - Типы данных СУБД
  - Необязательные значения NULL
  - Выбор имён для таблиц, колонок примера
  - Определение типов данных БД примера
  - Целостность данных и бизнес-правила
  - Определение процедур контроля целостности данных для примера
- 
- Первичные и внешние ключи (РК и FK)
  - Объявление ключевых атрибутов примера
  - Целостность ссылок
  - Определение процедур контроля целостности ссылок для примера
- 
- Нотации физической модели данных
  - Создание диаграммы «сущность-связь» для ФМД примера
  - Оформление таблицы описания физической модели данных примера
  - Описание данных – в комментариях к таблицам и атрибутам

# Физическая модель данных

*Детальная модель данных отражает необходимые потребности бизнеса, ограничения выбранной абстракции и конкретной реализации в виде БД.*

- *Выполняется дополнительная нормализация или денормализация баз данных.*
- *Сущности данных представляются в виде таблиц и представлений базы данных.*
- *Атрибуты реализуются в виде колонок таблицы БД.*
- *Идентификаторы и ссылки преобразуются в первичные (primary key) и внешние ключи (foreign key) БД.*
- *Использование типов данных SQL.*
- *На основе физической модели БД формируется DDL-скрипты для генерации объектов БД .*

# Шаги создания физической модели данных



# Заготовка ER-диаграммы ФМД на основе ЛМД

Скопировать ER-диаграмму ЛМД:

- Открыть диаграмму ЛМД.
- Сохранить её под новым именем для создания ФМД.

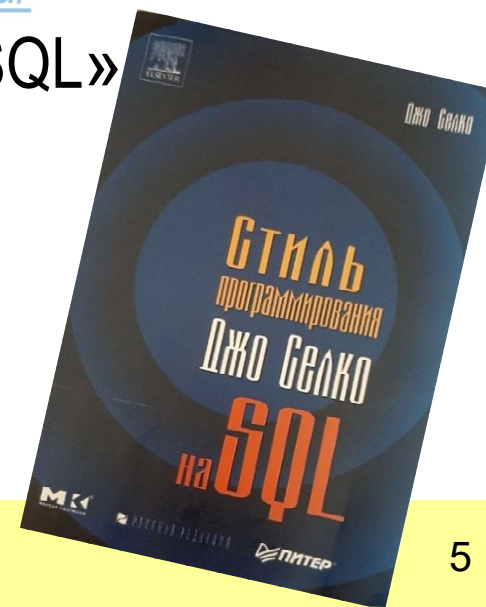
В зависимости от возможности графических редакторов и различий блоков ER-диаграмм ЛМД и ФМД:

- В PowerPoint можно использовать блоки ЛМД для дальнейшего редактирования.
- В других графических редакторах в худшем случае потребуется пересоздать блоки сущностей, если нотация блоков ФМД сильно отличается от нотации ЛМД.

# Именованние таблиц и колонок ФМД

- Именуйте таблицы и колонки ФМД на основе имён сущностей и атрибутов ЛМД
- Сохранять выбранный порядок слов в именах
- Русские сокращения применять, но не переводить
- Формат имени змейка (snake\_case): **creator\_role**
- Есть повод поговорить о правилах именования
- Правила именования <https://www.sqlstyle.guide/ru/>
- Книга «Стиль программирования Джо Селко на SQL»
- Стандарт именования элементов данных ГОСТ Р ИСО/МЭК 11179-1-2010

Правила именования элементов данных имеют более широкую область применения в ИТ.



# Стандарт именования данных ГОСТ Р ИСО/МЭК 11179-1-2010

## Семантика имён

- Имена таблиц и колонок имеют следующие составляющие:
  - Термы классов объектов (сущностей)
  - Термы свойств объектов
  - Термы представления
  - Уточняющие термы
- Должны быть ровно один терм классов, один терм свойств и может быть один терм представления

Пример: **Общая Сумма Стоимости за Бюджетный Период**

## Синтаксис имён

- Терм класса должен занимать 1-ое место в имени
- Уточняющие термы должны предшествовать уточняемым термам
- Терм свойства должен занимать следующую позицию
- Терм представления занимает последнее место

Пример: **Стоимость\_БюджетныйПериод\_Общая\_Сумма**  
**Cost\_BudgetPeriod\_Total\_Amount**

# Преобразования имён ЛМД в имена ФМД

- Варианты правил именования
  - *по-английски*, нужно переводить и вести словарь
  - *транслитерация* с настройкой <http://translit-online.ru/>
  - *русский язык* поддерживается в некоторых БД
- Преобразования имён атрибутов (сущностей) в имена колонок (таблиц)

Кто знает,  
что такое  
транслитерация?

Формат имени	Имя атрибута (сущности)	Транслитерация		Перевод	Имя в кавычках
		Прямой порядок слов	Обратный порядок слов	Обратный порядок слов	
Прямой порядок слов	имя_издателя	imja_izdatelja			
Обратный порядок слов	издатель_имя		izdatel_imja	publisher_name	
Имя в кавычках	“имя издателя”				“имя издателя”

- Стандарт транслитерации ГОСТ 7.79-2000



# Основные классы типов данных

*Тип данных – именованная категория множества значений и характеристик элемента данных, а также набор допустимых операций над ними.*

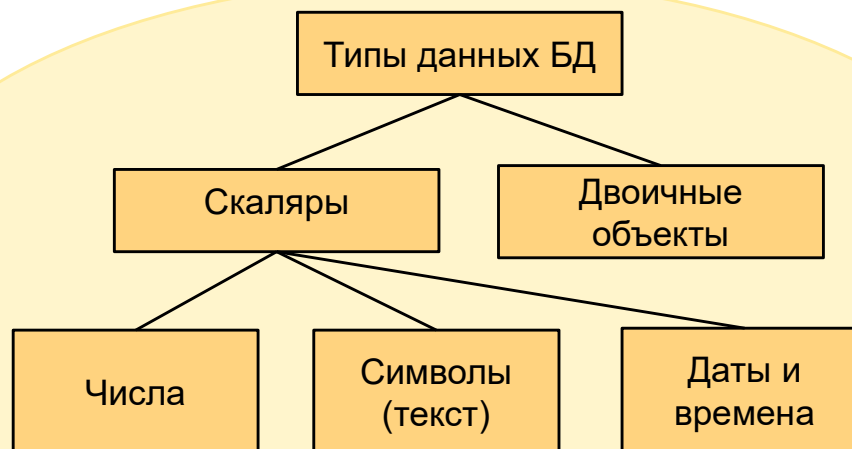


*Какие типы данных вы знаете  
или сталкивались на практике  
в программировании?*

**Выбор правильного типа данных уменьшает объём памяти и повышает производительность обработки данных**

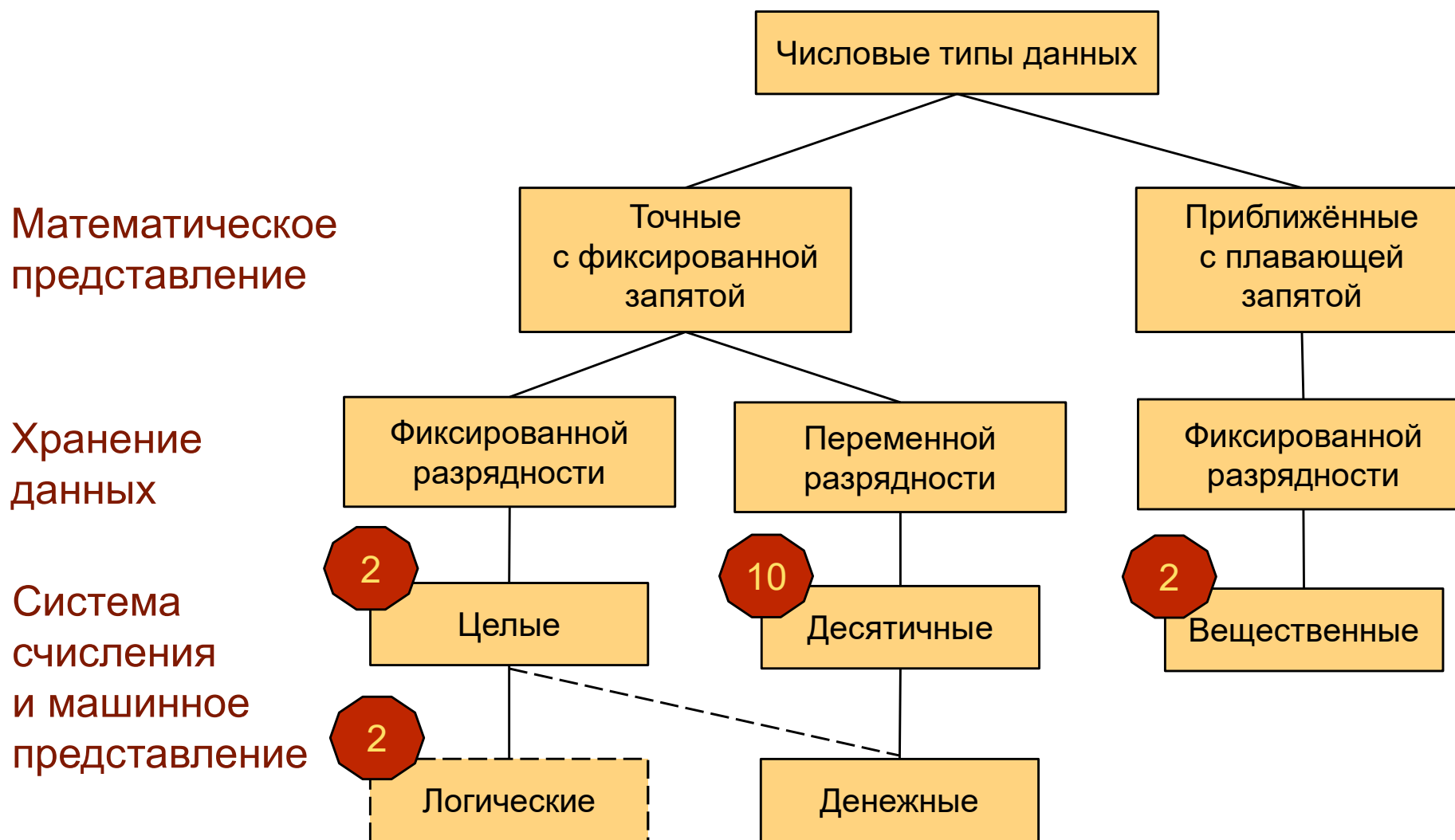
# Основные классы типов данных

*Тип данных – именованная категория множества значений и характеристик элемента данных, а также набор допустимых операций над ними.*



ФМД содержит *типы данных в реляционных БД*  
ЛМД могут содержать *математические или прикладные типы данных*.

# Классификация числовых типов данных



Разные СУБД MySQL, PostgreSQL, Oracle, SQL Server имеют свои особенности и не все типы данных поддерживаются.

# Числовые и логические типы данных

Тип данных	PostgreSQL	MySQL	Oracle	SQL Server	Байты	Макс. дес. разрядов
<b>SMALLINT</b>	Y	Y	Y	Y	2	4
<b>INTEGER</b>	Y	Y	Y	Y	4	9
<b>INT</b>		Y	Y	Y	4	9
<b>BIGINT</b>	Y	Y		Y	8	19
<b>DECIMAL NUMERIC</b>	Y <sup>1)</sup>	Y	Y <sup>2)</sup>	Y	до 17	38
<b>MONEY</b>	Y			Y	до 9 б	19
<b>REAL</b>	Y	Y	Y	Y	4	7
<b>FLOAT</b>		Y	Y	Y	4 / 8	7 / 15
<b>DOUBLE PRECISION</b>	Y	Y	Y	Y	8	15
<b>BIT</b>	Y	Y		Y	1	
<b>BOOLEAN</b>	Y			Y	1	

1) Поддерживает до 1000 разрядов

2) Тип NUMBER

# Символьные типы данных

	Тип данных	PostgreSQL	MySQL	Oracle	SQL Server
Фиксированной длины	CHAR	до 10 млн символов	до 255 б	до 2000 б	до 8000 б
	NCHAR			до 2000 б	до 8000 б
	CHARACTER	до 10 млн символов		до 2000 б	до 8000 б
Переменной длины	VARCHAR	до 10 млн символов	до 64 Кб	до 4000 б	до 8000 б
	NVARCHAR			до 4000 б	до 8000 б
	CHARACTER VARYING	до 10 млн символов		до 4000 б	до 8000 б
	TEXT (CLOB)	неограниченная длина	до 64 Кб	до 4 Гб	до 2 Гб
	MEDIUMTEXT		до 16 Мб		
	LONGTEXT		до 4 Гб		
	NTEXT (NCLOB)			до 4 Гб	до 2 Гб

Название типов, начинающихся с N, для шрифтов Unicode.

# Типы данных даты и времени

Тип данных	PostgreSQL	MySQL	Oracle	SQL Server
DATE	Y	Y	Y <sup>1)</sup>	Y
TIME	Y	Y		Y
TIMESTAMP	Y	Y	Y	
DATETIME		Y		Y
INTERVAL	Y	Y		Y

<sup>1)</sup> Дата и время

# Типы двоичных объектов данных

Тип данных	PostgreSQL	MySQL	Oracle	SQL Server
BLOB		До 64 Кб	До 4 Гб	
MEDIUMBLOB		До 16 Мб		
LOB		До 4 Гб		
BINARY				До 8000 б
VARBINARY		До 64 Кб		До 8000 б
RAW			До 2000 б	
BYTEA	До 1 Гб			
BFILE			До 4 Гб	

Для хранения изображений, мультимедиа и сырых двоичных данных

# Домены данных

## ■ Что такое домен данных?

Именованная область допустимых значений  
однотипных атрибутов – прикладной тип данных,  
задаёт ограничения для значений:

- Тип данных
- Формат
- Диапазон значений
- Список допустимых значений
- Условие

## ■ Примеры доменов

- Тип\_ИНН\_юрлиц или Цифры(10)
- Деньги(20,4)
- Список: “малый”; “средний”; “большой”



# Объявление необязательных значений

Отсутствие значения, неизвестное, неопределённое или недопустимое значение в этом контексте в базах данных помечается признаком **NULL**.

Такие значения специальным образом обрабатываются.

При описании ФМД мы должны явно указать обязательность значения - **NOT NULL**, или необязательность значения - **NULL**

На ER-диаграмме ФМД можно не указывать везде **NOT NULL**, но надо задать признак по умолчанию.

# Объявление ключей

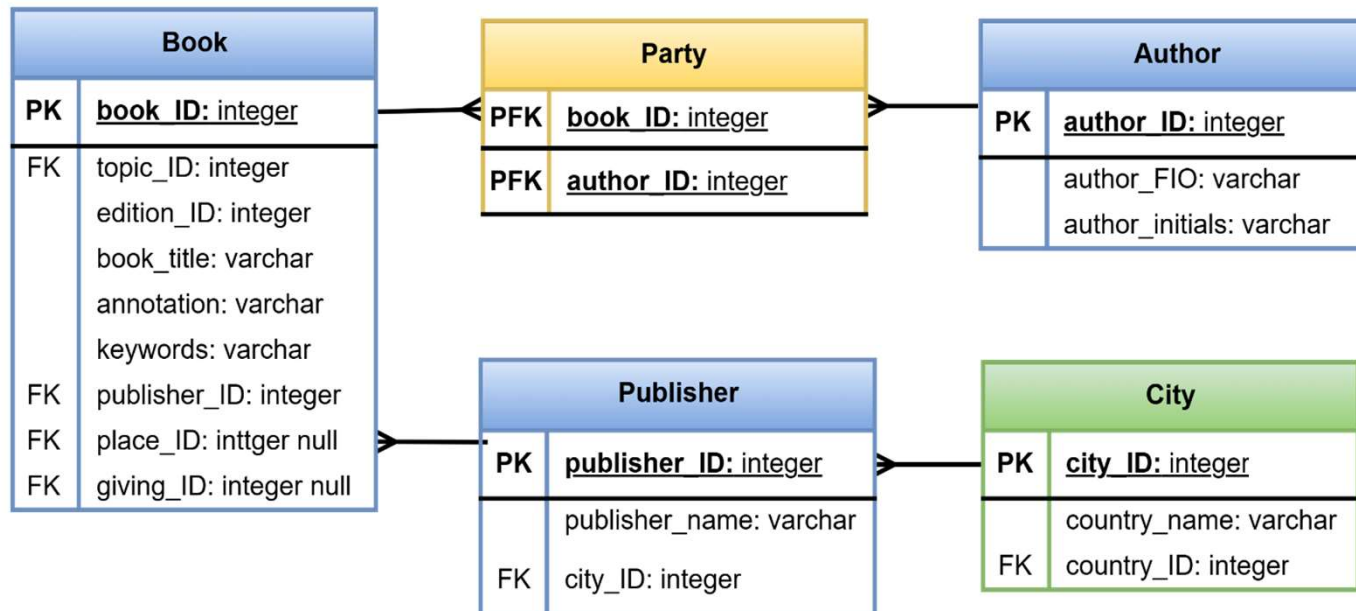
Первичный ключ PK  
PRIMARY KEY

Альтернативные ключи  
UN  
UNIQUE

Внешние ключи  
FK  
FOREIGN KEY

Процедуры проверки уникальности и целостности ссылок

# Нотация ER-диаграммы физической модели данных



## Нотация Р. Баркера

- Имя таблицы/представления существительное латиницей в единственном или во множественном числе на полочке или подчёркнуто.
- В блоке задаются имена колонок латиницей без пробелов, без повторения или с повтором имени сущности.
- Следует указывать (PK) для идентификатора и (FK) для ссылки.
- Указывать (null) для необязательных атрибутов.
- Указывать типы данных для колонки через «:».

# Состав таблицы описания физической модели данных

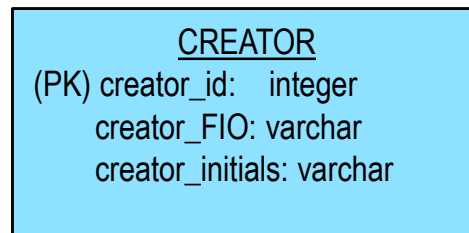
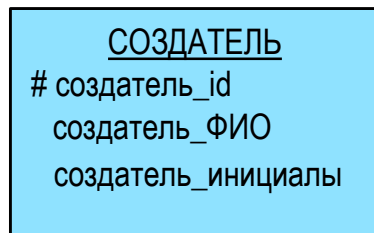
- Имя базы данных (имя предметной области)
- Вид набора данных: таблица/ представление
- Имя таблицы (сущности)
- Имя колонки (атрибута)
- Описание колонки (атрибута)
- Обязательность значения: NULL или NOT NULL
- Имя домена данных и его описание
- Тип данных и *длина/ разрядность*
- Процедуры целостности данных и ссылок (допустимые ограничения типа, правила целостности, первичные и внешние ключи)
- [Примеры значения данных]

# Пример: Домашняя библиотека

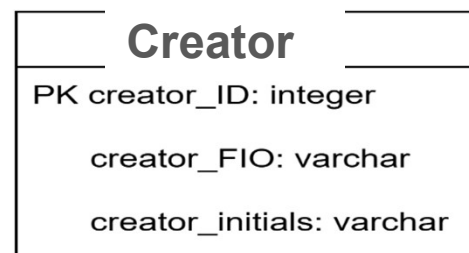
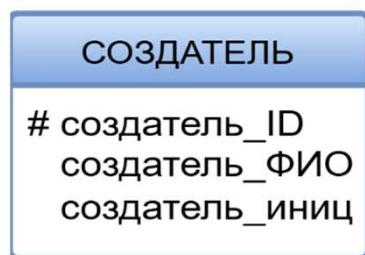
## Создать каталог домашней библиотеки

- **Атрибуты поиска книги:**
  - Тематика, вид издания
  - Название издания и/ или ключевые слова
  - Фамилию, имя (инициалы) автора, редактора, составителя, переводчика или художника
  - Название и место издательства
  - Год выпуска.
- **Результаты поиска:**
  - Хранимые атрибуты книги (включая аннотацию, описание приложения и сведения об оригинале перевода)
  - Место хранения издания (шкаф, полка)
  - Текущий держатель (читатель) издания, его контакты
- **В библиотеке каждая книга присутствует в одном экземпляре**
- **Расширения видов изданий:** электронные книги и аудиокниги, хранимые на компьютерах, ноутбуках, планшетах и смартфонах
- **Расширение мест хранения:** дача, офис

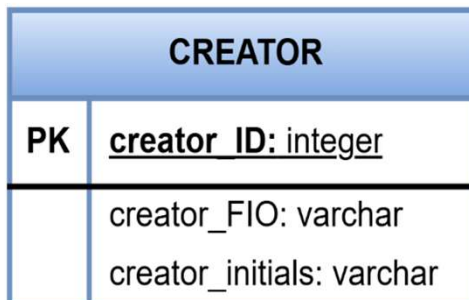
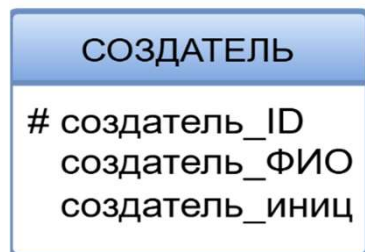
# Преобразование блоков ER-диаграммы



PowerPoint  
Замена и редактирование  
названий атрибутов

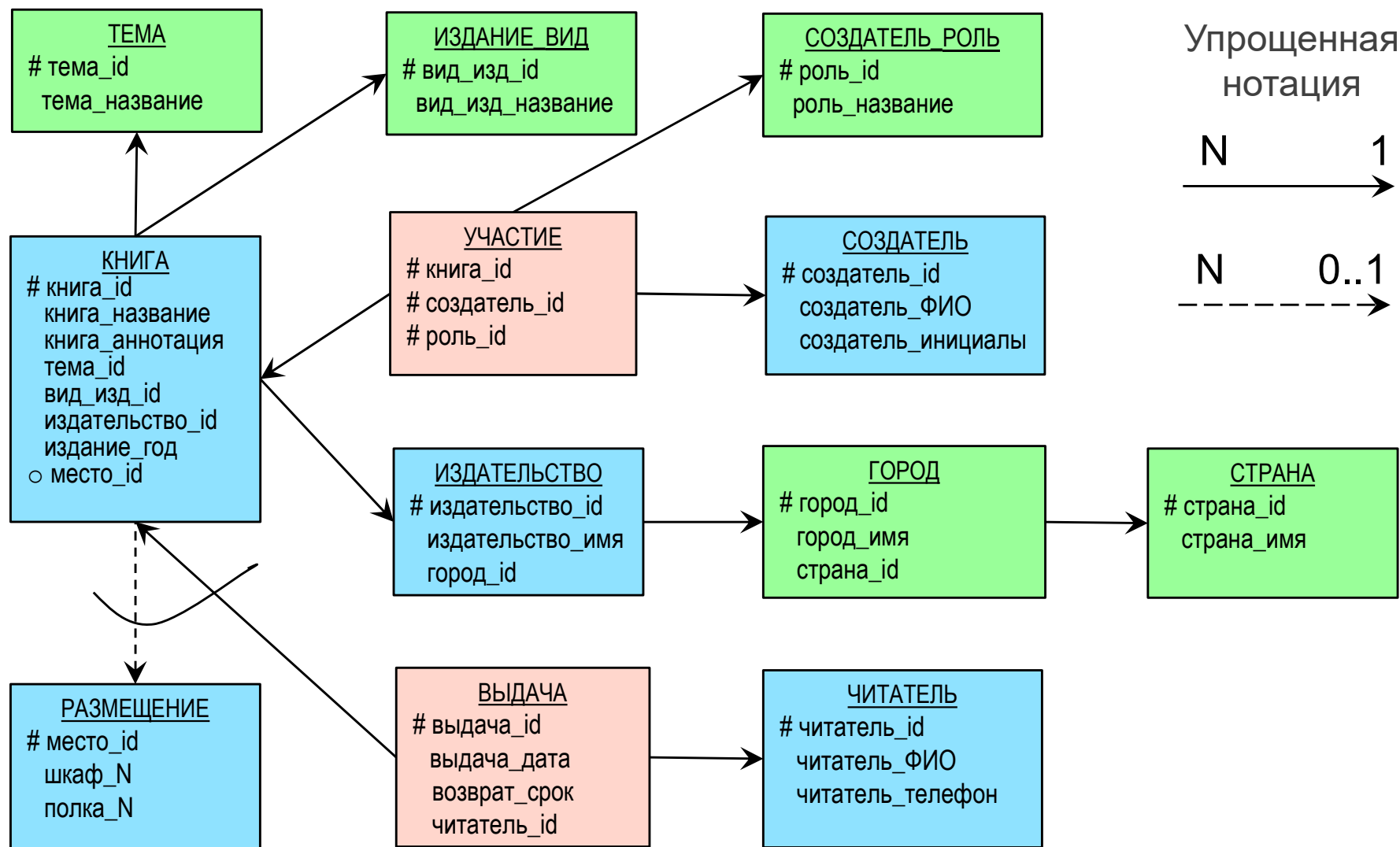


Draw.io  
Замена и редактирование  
названий атрибутов

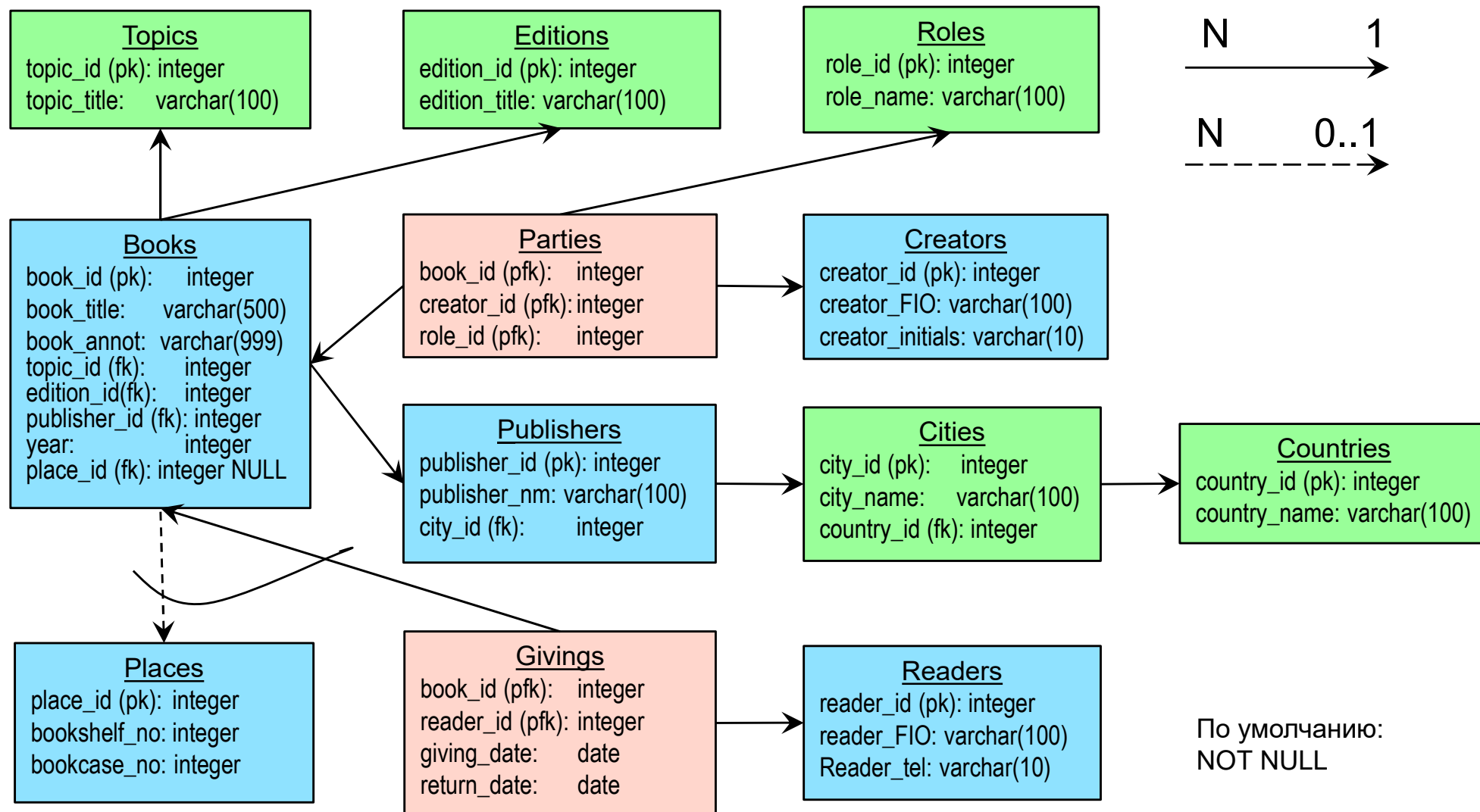


Draw.io  
Замена блоков,  
ввод и копирование  
названий атрибутов

# Логическая модель данных «Домашняя библиотека»

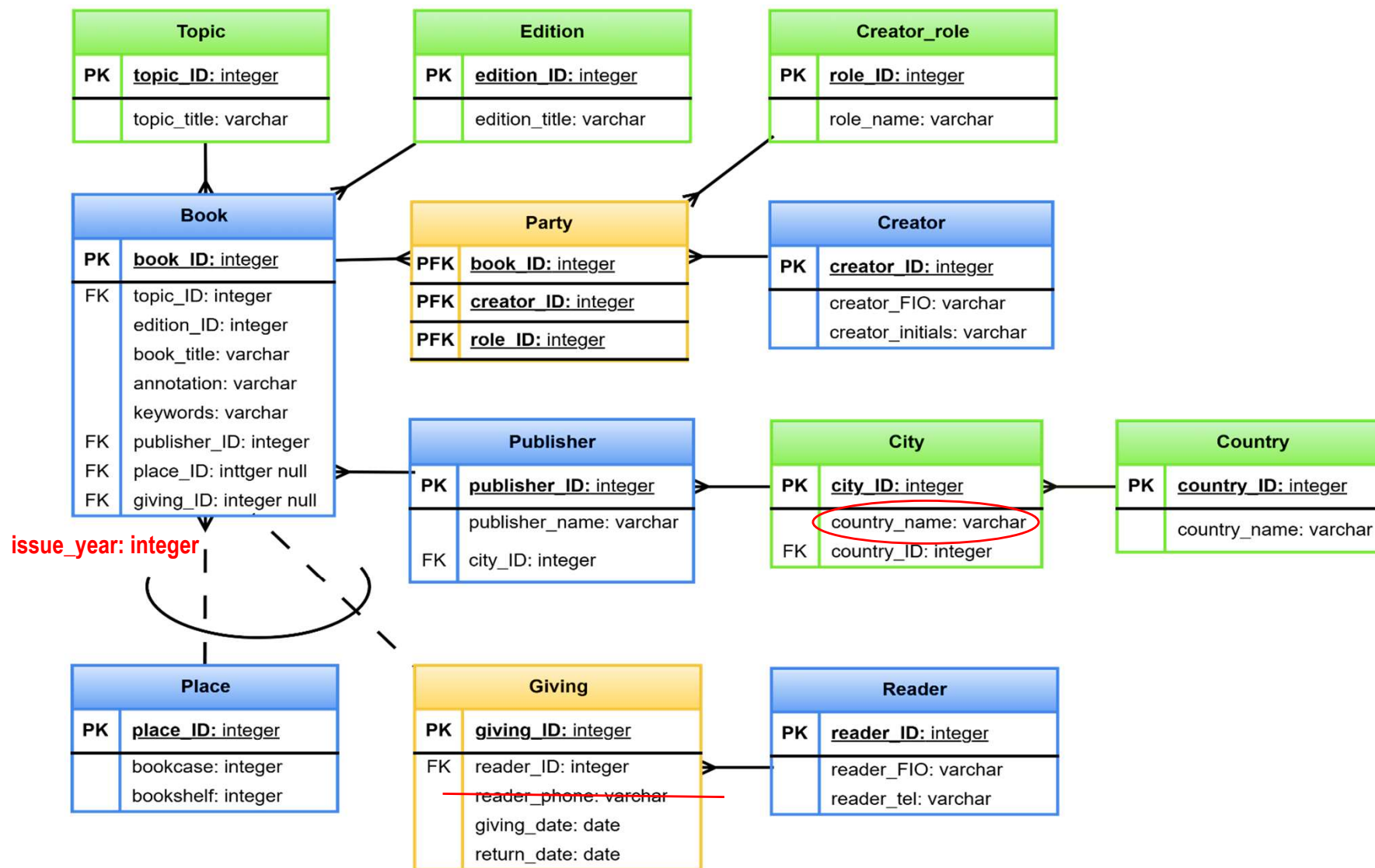


# Физическая модель данных «Домашняя библиотека» (PowerPoint)





# Физическая модель данных «Домашняя библиотека» (draw.io)



# Описание физической модели данных «Домашняя библиотека»

Описание атрибута	Имя таблицы	ключ	Имя колонки	Тип данных	Null / Not null	Длина / разрядность	Область допустимых значений	Бизнес-правила
Идентификатор Реестра КНИГА	books	PK	book_ID	Integer	Not null		Уникальное значение ≤10000	
Название книги	books		book_title	Varchar	Not null	200		
Аннотация книги	books		annotation	Varchar	Not null	1000		
Ссылка на позицию Классификатора ТЕМА (тема книги)	books	FK	topic_ID	Integer	Not null		Соответствует значениям topics.topic_ID	
Ссылка на позицию Классификатора ИЗДАНИЕ_ВИД (вид издания)	books	FK	edition_ID	Integer	Not null		Соответствует значениям editions.edition_ID	
Ссылка на позицию Реестра ИЗДАТЕЛЬСТВО	books	FK	publisher_ID	Integer	Not null		Соответствует значениям publishers.publisher_ID	
Год издания книги	books		year	Integer	Not null	4 цифры		year ≤ текущий год
Ссылка на позицию Реестра РАЗМЕЩЕНИЕ (места хранения книг). Если ссылка пустая, то должна быть непустая ссылка на читателя.	books	FK	place_ID	Integer	Null		Соответствует значениям places.place_ID	Должно быть задано place_ID или giving.book_ID
Идентификатор Классификатора ТЕМА	topics	PK	topic_ID	Integer	Not null			
Название темы книги	topics		topic_name	Varchar	Not null			
Идентификатор Классификатора ИЗДАНИЕ_ВИД	editions		edition_ID	Integer	Not null			
Название вида издания	editions		edition_name	Varchar	Not null			
Идентификатор Реестра СОЗДАТЕЛЬ	creators	PK	creator_ID	Integer	Not null			
Фамилия, имя и отчество или инициалы создателя (редактора, автора, переводчика, художника и т.п.)	creators		creator_FIO	Varchar	Not null			
Идентификатор Классификатора СОЗДАТЕЛЬ_РОЛЬ	creator_roles	PK	role_ID	Integer	Not null			
Название роли создателя: редактора, автора, переводчика, художника и т.п.	creator_roles		role_name	Varchar	Not null			
Часть составного ключа участия и ссылка на книгу	parties	PFK	book_ID	Integer	Not null			
Часть составного ключа участия и ссылка на создателя	parties	PFK	creator_ID	Integer	Not null			
Часть составного ключа участия и ссылка на роль создателя	parties	PFK	role_ID	Integer	Not null			
Идентификатор Реестра ИЗДАТЕЛЬСТВО	publishers	PK	publisher_ID	Integer	Not null			
Наименование издательства	publishers		publisher_NM	Varchar	Not null			

# В качестве резюме

- Физическая модель данных строится на основе ЛМД для реализации базы данных в *среде конкретной СУБД*.
- *Денормализация – компромисс между объёмом БД и временем выполнения запросов.*
- Вместо естественных ключей рекомендуется использовать *суррогатные ключи фиксированной длины с автоинкрементом или глобальные уникальные идентификаторы GUID*.
- *ER-диаграммы ФМД строятся не так часто, как для ЛМД, из-за потери обозримости при большом объёме информации. Чаще используется табличное описание ФМД.*
- Помимо физического моделирования данных *существует проектирование БД: создание представлений, индексов, планирование пространства и распределение данных.*

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

**Спасибо за внимание!**

**Валерий Иванович Артемьев**

**Департамент данных, проектов и процессов  
Банк России**

**Тел.: +7(495) 753-96-25**

**e-mail: [avi@cbr.ru](mailto:avi@cbr.ru)**