

# Dev Camp

Введение в базы данных  
Анализ данных

# План

1. Основные понятия
2. Реляционные базы данных.
3. Этапы проектирования баз данных.
4. Анализ данных. Пример.

# 1. Основные понятия

## Информация и данные

**Информация** - данные, которым придается некий смысл (интерпретация) в конкретной ситуации в рамках некоторой системы понятий

**Данные** – информация, представленная в определенной форме, пригодной для последующей обработки, хранения и передачи.

**Основные преобразования в ИС:**



# 1. Основные понятия

## Информация и данные

a) Data entry screen

[illegible]

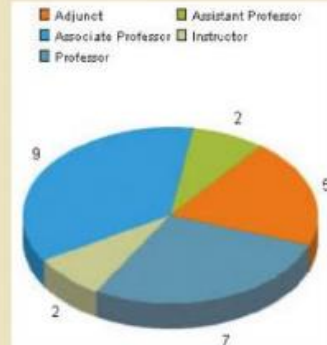
b) Raw data

Player	Team	Position	Height	Weight	Age	Experience	College	Drafted	Round	Picks	Notes
1. J. Rodriguez	Atlanta Braves	Pitcher	6'2"	185	24	1	Georgia Tech	1998	1	1	1998-99
2. Mike Johnson	Los Angeles Angels	Outfielder	6'0"	175	23	1	UCLA	1998	2	2	1998-99
3. David Lee	San Francisco Giants	Infielder	5'10"	165	22	1	Stanford	1998	3	3	1998-99
4. Jason Miller	St. Louis Cardinals	Pitcher	6'4"	200	25	2	Missouri State	1998	4	4	1998-99
5. Chris Davis	Baltimore Orioles	Outfielder	6'1"	180	21	1	Wake Forest	1998	5	5	1998-99
6. Ryan Howard	Philadelphia Phillies	Infielder	6'3"	210	23	1	Temple	1998	6	6	1998-99
7. Tim Lincecum	Seattle Mariners	Pitcher	5'11"	160	22	1	UCLA	1998	7	7	1998-99
8. Alex Rodriguez	New York Yankees	Infielder	6'0"	175	21	1	Baylor	1998	8	8	1998-99
9. J. D. Drew	Boston Red Sox	Outfielder	6'0"	175	23	1	Stanford	1998	9	9	1998-99
10. Matt Kemp	Los Angeles Dodgers	Outfielder	6'1"	185	22	1	Florida State	1998	10	10	1998-99
11. Adam Lind	San Diego Padres	Infielder	6'0"	180	23	1	Arizona State	1998	11	11	1998-99
12. Ryan Braun	Milwaukee Brewers	Infielder	6'0"	180	22	1	UCLA	1998	12	12	1998-99
13. Prince Fielder	Detroit Tigers	Infielder	6'5"	215	23	1	Nebraska	1998	13	13	1998-99
14. Justin Verlander	Detroit Tigers	Pitcher	6'5"	215	23	1	Florida State	1998	14	14	1998-99
15. Jacob deGrom	New York Mets	Pitcher	6'4"	230	24	1	UT Austin	1998	15	15	1998-99
16. Clayton Kershaw	Los Angeles Dodgers	Pitcher	6'4"	200	24	1	University of Michigan	1998	16	16	1998-99
17. Max Scherzer	Washington Nationals	Pitcher	6'3"	200	25	1	University of Texas at Arlington	1998	17	17	1998-99
18. Bryce Harper	Washington Nationals	Outfielder	6'3"	200	22	1	University of California, Berkeley	1998	18	18	1998-99
19. Freddie Freeman	Los Angeles Dodgers	Infielder	6'0"	185	23	1	University of California, Berkeley	1998	19	19	1998-99
20. Mookie Betts	Boston Red Sox	Outfielder	5'9"	165	23	1	University of California, Berkeley	1998	20	20	1998-99
21. Gleyber Torres	New York Yankees	Infielder	6'0"	180	22	1	University of California, Berkeley	1998	21	21	1998-99
22. Anthony Rendon	Washington Nationals	Infielder	6'3"	200	23	1	University of California, Berkeley	1998	22	22	1998-99
23. Juan Soto	Washington Nationals	Outfielder	6'0"	180	22	1	University of California, Berkeley	1998	23	23	1998-99
24. Vladimir Guerrero Jr.	Toronto Blue Jays	Infielder	6'0"	180	22	1	University of California, Berkeley	1998	24	24	1998-99
25. Shohei Ohtani	Los Angeles Angels	Outfielder/Pitcher	6'2"	180	23	1	University of California, Berkeley	1998	25	25	1998-99
26. Jacob deGrom	New York Mets	Pitcher	6'4"	230	24	1	UT Austin	1998	15	15	1998-99
27. Clayton Kershaw	Los Angeles Dodgers	Pitcher	6'4"	200	24	1	University of Michigan	1998	16	16	1998-99
28. Max Scherzer	Washington Nationals	Pitcher	6'3"	200	25	1	University of Texas at Arlington	1998	17	17	1998-99
29. Bryce Harper	Washington Nationals	Outfielder	6'3"	200	22	1	University of California, Berkeley	1998	18	18	1998-99
30. Freddie Freeman	Los Angeles Dodgers	Infielder	6'0"	185	23	1	University of California, Berkeley	1998	19	19	1998-99
31. Mookie Betts	Boston Red Sox	Outfielder	5'9"	165	23	1	University of California, Berkeley	1998	20	20	1998-99
32. Gleyber Torres	New York Yankees	Infielder	6'0"	180	22	1	University of California, Berkeley	1998	21	21	1998-99
33. Anthony Rendon	Washington Nationals	Infielder	6'3"	200	23	1	University of California, Berkeley	1998	22	22	1998-99
34. Juan Soto	Washington Nationals	Outfielder	6'0"	180	22	1	University of California, Berkeley	1998	23	23	1998-99
35. Vladimir Guerrero Jr.	Toronto Blue Jays	Infielder	6'0"	180	22	1	University of California, Berkeley	1998	24	24	1998-99

c) Information in summary format

Rank	COUNT	%INFS	TOT/COL	%COL. TOT.	%COL. FAC.
Adjunct	5	26.00%	23	21.74%	3.27%
Assistant Professor	2	8.00%	26	7.14%	1.31%
Associate Professor	9	36.00%	37	24.32%	5.88%
Instructor	2	8.00%	18	11.11%	1.31%
Professor	7	28.00%	47	14.89%	4.58%

d) Information in graphical format



# 1. Основные понятия

## Информационная система

**Информационная система** - организационная совокупность ресурсов, реализующая

✓ сбор,  
✓ обработка,  
✓ хранение,  
✓ поиск,  
✓ выдача,  
✓ передача  
информации

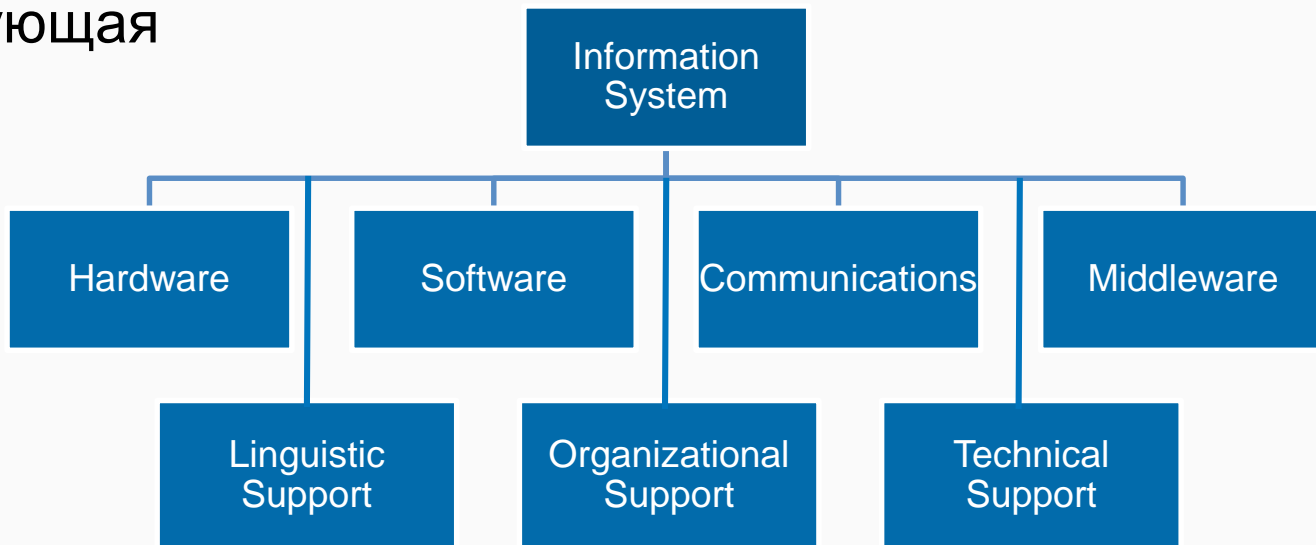


**Цель:**  
удовлетворить  
потребность группы  
пользователей в  
информации

# 1. Основные понятия

## Информационная система

**Информационная система** - организационная совокупность ресурсов, реализующая



# 1. Основные понятия

## Классификация информационных систем

- ❑ **OLTP-системы** - Online Transaction Processing -
  - ❑ системы оперативной обработки транзакций
- ❑ **OLAP-системы** - Online Analysis Processing -
  - ❑ системы делового анализа/хранилища данных

# 1. Основные понятия

## Классификация информационных систем

Характеристика ИС	Содержимое БД	Типичный размер таблиц	Данные, с которыми работает приложение	Интенсивность обращения к БД	Тип доступа
OLTP-система	Текущие данные	Тысячи строк	Отдельные строки	Много транзакций в секунду	Выборка, вставка, обновление
OLAP-система	Накопленные данные	Миллионы строк	Группа строк	Одна транзакция в час	Выборка



# 1. Основные понятия

## Базы данных

**База данных** – структурированный набор данных

**База данных** – упорядоченный набор **логически взаимосвязанных данных**, которые совместно используются, и предназначены для удовлетворения информационных потребностей пользователей.

**База данных** – унифицированная совокупность **хранимых** и воспроизводимых данных, используемых в рамках организации

**База данных** – совокупность взаимосвязанных данных, которые **совместно используются** несколькими приложениями и хранятся с **минимальной регулируемой избыточностью**

# 1. Основные понятия Базы данных

## Data Definition Language

- Описывает логическую модель
- Описывает способ хранения данных
- Описывает способ доступа к данным

## Data Manipulation Language

- Обработка элементов данных
- Структуры для создания приложений

# 1. Основные понятия

## Системы управления базами данных

**Система управления базами данных** – набор программных средств, обеспечивающих создание и обслуживание БД и выполнение операций над данными БД.

### **Функции:**

- ✓ описание и манипулирование данными,
- ✓ поддержка логической модели данных,
- ✓ взаимодействия логической и физической структур данных,
- ✓ актуальность, защита и целостность данных

# 1. Основные понятия

## Классическая трехуровневая архитектура БД

Представление КМ с  
учетом интересов  
пользователей/  
приложений

Интегрированное  
независимое описание  
предметной области

Представление КМ с  
учетом способов хранения  
и доступа к данным



# 1. Основные понятия

## Системы управления базами данных

**Концептуальный уровень** - интегрированное описание предметной области, для которой разрабатывается БД, **независимо от ее восприятия отдельными пользователями и способов реализации** в компьютерной системе.

**Внешний уровень** – определяет уровень доступа, предоставляет пользователю/приложению лишь те данные, которые ему нужны и в нужном виде.

**Внутренний уровень** – описывает способ хранения данных и методы доступа к ним.

# 1. Основные понятия

## Модель данных

### Концепция модели данных

- ❑ При проектировании базы данных основное внимание уделяется тому, как **структура БД будет использоваться для хранения данных конечных пользователей и управления ими**.
- ❑ **Модель данных** - это относительно **простое представление**, обычно графическое, более **сложных** реальных **структур** данных.
- ❑ Основная **функция** модели - помочь вам **понять сложности** реальной среды.
- ❑ Моделирование данных - **итеративный процесс**

# 1. Основные понятия

## Модель данных

### Модель



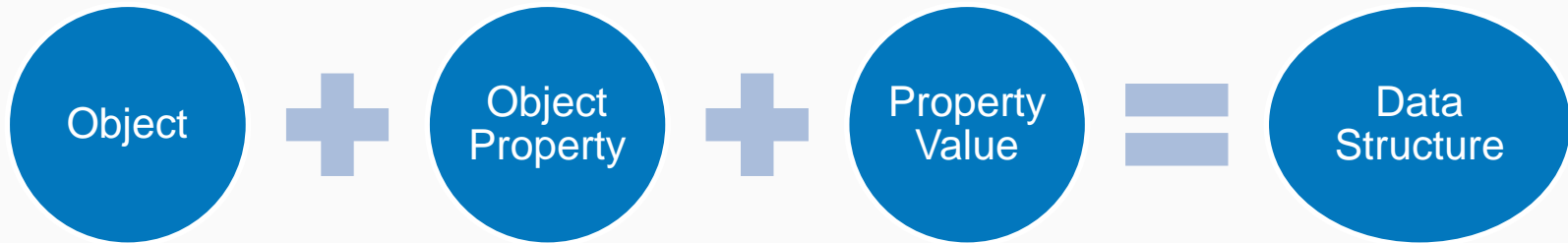
### Критерии выбора:

- ☐ объем информации,
- ☐ задачи,
- ☐ техническое и программное обеспечение,
- ☐ средства манипулирования данными,
- ☐ целостность и защита,
- ☐ критерии качества и т.п.

# 1. Основные понятия

## Модель данных

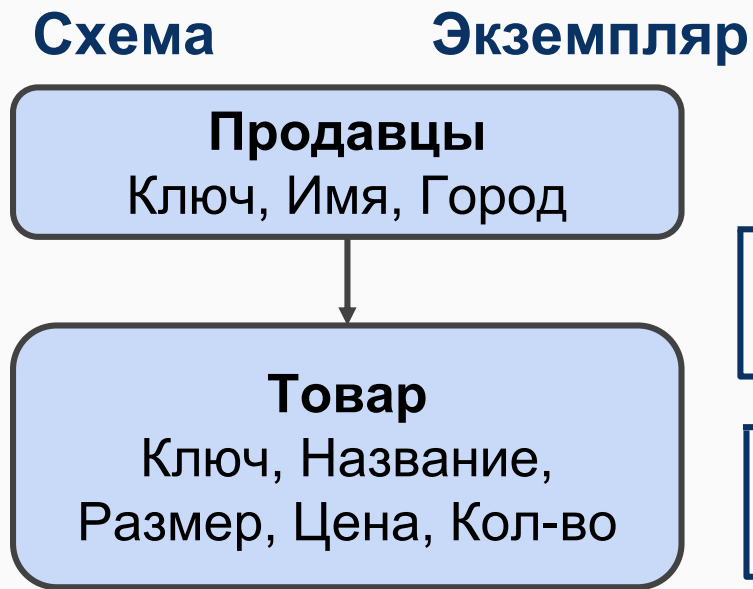
### Общая структура данных





# 1. Основные понятия.

## Иерархическая модель данных



Ключ	Имя	Город	Товар
101	Иванов	Сумы	
102	Петров	Харьков	

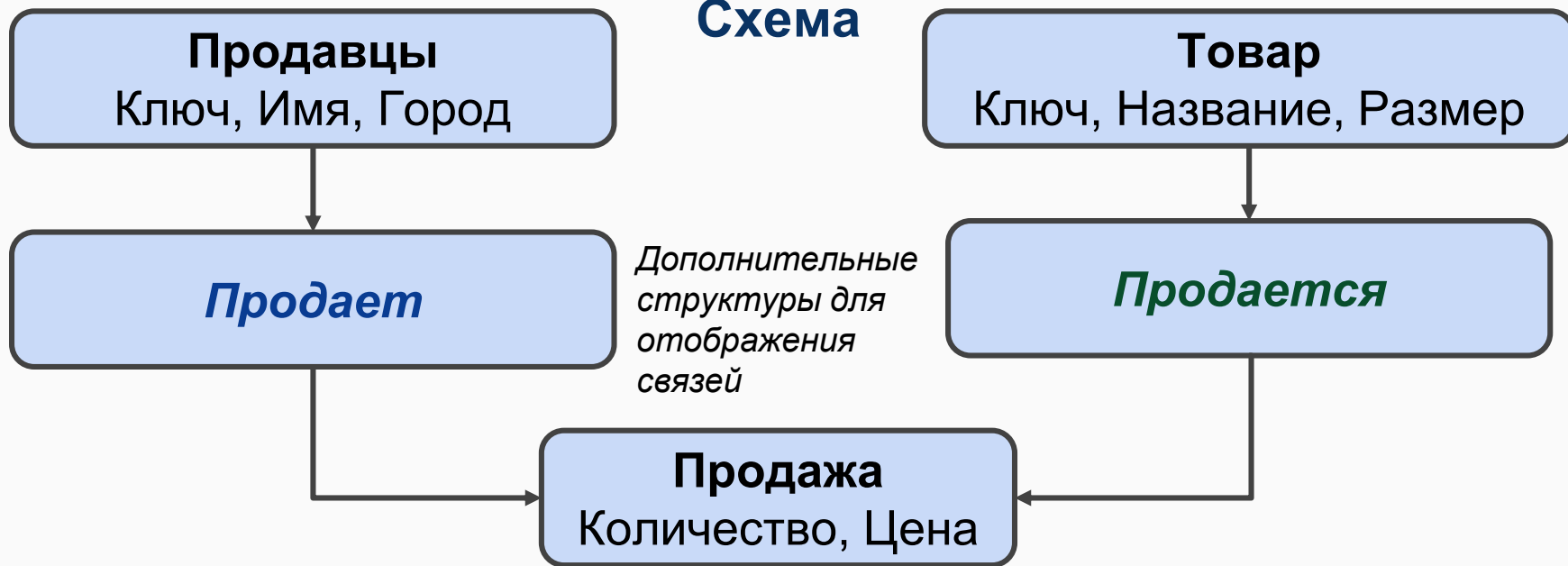
Ключ	Название	Размер	Цена	Кол-во
15	Чайник	1,5	320,00	10

Ключ	Название	Размер	Цена	Кол-во
15	Чайник	1,5	325,00	14

# 1. Основные понятия.

## Сетевая модель данных

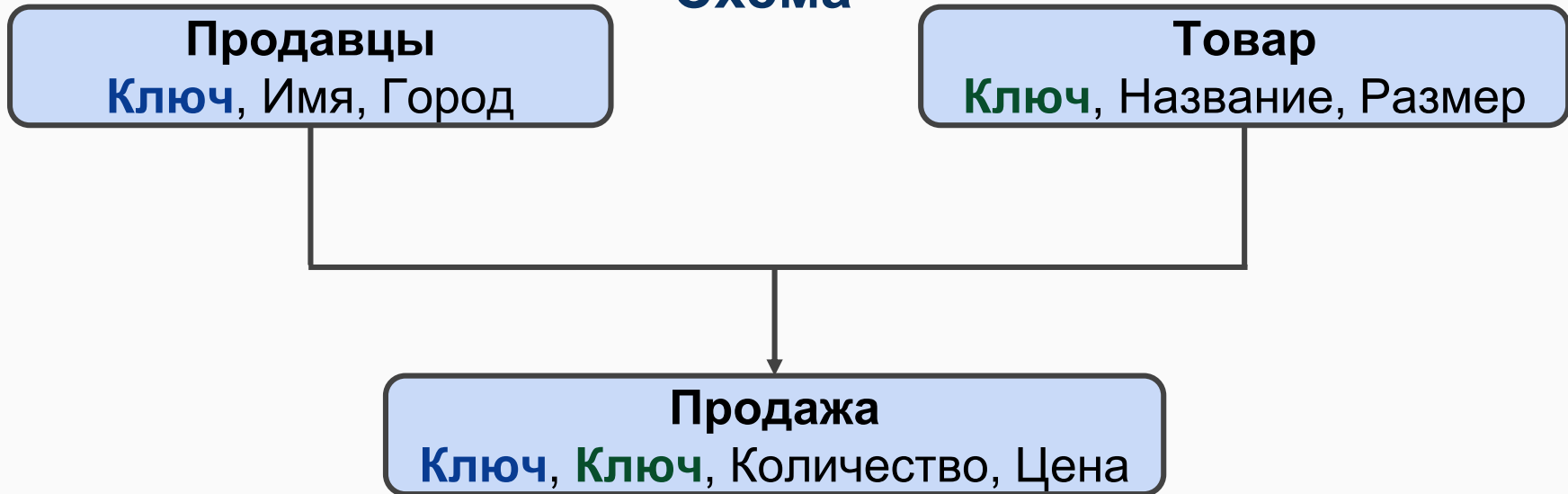
### Схема



# 1. Основные понятия.

## Реляционная модель данных

### Схема



# 1. Основные понятия.

## Объектно-ориентированная модель данных

### Реляционная модель

Ключ	Имя	Город
101	Иванов	Сумы
102	Петров	Харьков

КлючП	Телефон
101	0503025645
102	0661457878
102	0971452565

### Объектно-ориентированная модель

Продавец
<pre>{   "Ключ": 102,   "Имя": "Петров",   "Город": "Харьков",   "Телефон": [ "0661457878",                "0971452565"             ] }</pre>

## 2. Реляционная модель данных

### Реляционная модель данных

- Организованная совокупность информации, которая сохраняется во взаимосвязанных двухмерных таблицах

### Таблица

- поименованное двухмерное изображение реляционного отношения, состоящего из одного или более поименованных столбцов и нуля или более строк

## 2. Реляционная модель данных

### Элементы модели

Пример табличного представления реляционного отношения  
**Факультет**

#F	Name	Head	Building
1	EIIT	Oleksii Drozdenko	Electrotechnical building
2	TESET	Alexandr Gusak	Laboratory building
3	IFSK	Olena Sushkova	Main building

## 2. Реляционная модель данных

### Элементы модели

Атрибуты	#F	Name	Head	Building
	1	EIIT	Serhiy Protsenko	Electrotechnical building
	2	TESET	Alexandr Gusak	Laboratory building
Запись	3	IFSK	Olena Sushkova	Main building

**Атрибут – поименованный столбец отношения**  
Атрибуты могут быть приведены в любом порядке

## 2. Реляционная модель данных

### Элементы модели

#F	Name	Head	Building
1	EIIT	Oleksii Drozdenko	Electrotechnical building
2	TESET	Alexandr Gusak	Laboratory building
3	IFSK	Olena Sushkova	Main building

**Домен - набор допустимых значений для одного или нескольких атрибутов**

Каждый атрибут в отношении определяется в домене



## 2. Реляционная модель данных

### Элементы модели

#F	Name	Head	Building
1	EIIT	Oleksii Drozdenko	Electrotechnical building
2	TESET	Alexandr Gusak	Laboratory building
3	IFSK	Olena Sushkova	Main building

Степень

**Степень отношения - количество атрибутов**

## 2. Реляционная модель данных

### Элементы модели

#F	Name	Head	Building
1	EIIT	Oleksii Drozdenko	Electrotechnical building
2	TESET	Alexandr Gusak	Laboratory building
3	IFSK	Olena Sushkova	Main building

Мощность

**Кардинальность (мощность) отношения -  
количество кортежей**

## 2. Реляционная модель данных

### Элементы модели

#### Ключ

#F	Name	Head	Building
1	EIIT	Oleksii Drozdenko	Electrotechnical building
2	TESET	Alexandr Gusak	Laboratory building
3	IFSK	Olena Sushkova	Main building

**Ключ – уникальный атрибут для сущности в одном или нескольких столбцах, который отличает один экземпляр сущности (кортеж) от всех других экземпляров.**

## 2. Реляционная модель данных

### Элементы модели



## 2. Реляционная модель данных

### Элементы модели

#F	Name	Head	Building
1	EIIT	Oleksii Drozdenko	Electrotechnical building
2	TESET	Alexandr Gusak	Laboratory building
3	IFSK	Olena Sushkova	Main building

#D	#F	Name	Head	Building
1	1	CS	Anatoliy Dovbysh	Central building
2	1	ECT	Anatoliy Opanasyk	Electrotechnical building
3	3	GPh	Iryna Kobyakova	Electrotechnical building

**Внешний ключ  
(Foreign Key) –  
атрибут  
моделирования  
связи**

## 2. Реляционная модель данных

### Соответствие терминологии


Formal Terminology	Informal Terminology
Relation/Отношение	Table/Таблица
Attribute/Атрибут	Column/Колонка, столбец
Tuple/Кортеж	Row, record/Строка
Cardinality/Мощность	Row numbers/Количество строк
Degree/Степень	Column numbers/Количество колонок, столбцов
Domain/Домен	Values domain/Область значений
Key/Ключ	Identificatory/Идентификатор

## 2. Реляционная модель данных

### Связи между таблицами

#### Характеристики связи

**Связь** – определяет отношения между таблицами реляционной модели

Сила	<ul style="list-style-type: none"><li>Идентифицирующая —</li><li>Неидентифицирующая - - -</li></ul>
Мощность	<ul style="list-style-type: none"><li>1:1  </li><li>1:M</li><li>M:N </li></ul>
Участие сущности	<ul style="list-style-type: none"><li>Обязательная  </li><li>Факультативная ○</li></ul>
Степень	<ul style="list-style-type: none"><li>Число ассоциируемых сущностей</li></ul>

# 3. Этапы проектирования БД

## Методология проектирования

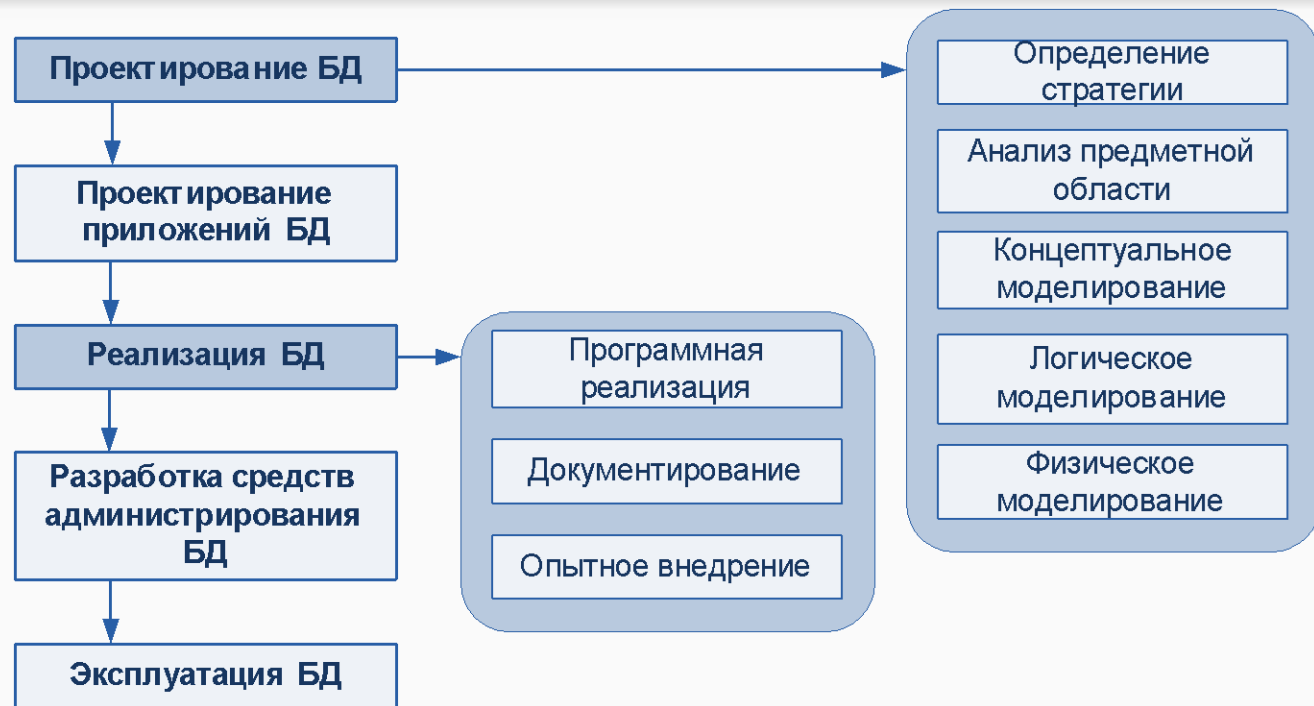
совокупность принципов, методов, инструментов и средств, применяемых для последовательной разработки базы данных

- ❑ Процесс проектирования
- ❑ Методика оценивания альтернативных решений
- ❑ Информационные требования
- ❑ Средства описания результатов проектирования



# 3. Этапы проектирования БД

Жизненный  
цикл системы  
баз данных



### 3. Этапы проектирования БД

## Критерии оценивания модели данных

#### Количественные

- Время выполнения запроса
- Стоимость операций манипулирования данными
- Затраты памяти

#### Качественные

- Гибкость • Адаптивность
- Совместимость • Восприимчивость

# 3. Этапы проектирования БД

## Анализ ПО

- Анализ задач
- Анализ данных

## Концептуальное моделирование

- Описание предметной области в терминах формализованного языка

## Логическое моделирование

- Разработка структуры БД без привязки к конкретной СУБД

## Физическое моделирование

- Описание структуры БД в терминах конкретной СУБД

## 4. Анализ данных. Пример.

### **Этапы анализа данных**

1. Проанализировать функции, которые будет выполнять информационная система с базой данных
2. Проанализировать сущности моделируемой предметной области (все существительные в описании связей)
3. Описать атрибуты (свойства) выделенных сущностей предметной области
4. Сформулировать связи между сущностями предметной области

# 4. Анализ данных. Пример.

Разработка базы данных для системы учета нагрузки сотрудников ИТ компании

## 1. Функции системы:

- Вести учет нагрузки сотрудников компании
- Вести учет проектов
- Вести историю клиентов
- Хранить информацию о сотрудниках компании.

## 2. Определение сущностей (объекты):

- Учет - основная функция системы, не учитываем
- Нагрузка +
- Сотрудники +
- Компания – локальный заказ, не храним инфу
- Проекты +
- История - связка между клиентом и проектами, связь
- Клиенты +

# 4. Анализ данных. Пример.

## 2. Определение атрибутов (характеристики) сущностей:

- Нагрузка: ключ, сотрудник, проект, задачи, объем часов, сроки
- Сотрудники: ключ, ФИО, специализация, навыки, моб телефон, дата приема, дата увольнения, отдел, ставка (график работы), образование?
- Проекты: ключ, название, дата старта, дата завершения, менеджер, стек технологий, стоимость, тип разработки, клиент
- Клиенты: ключ, ФИО, моб телефон, тип клиента (ЮЛ-ФЛ), страна, конт.лицо, моб.конт.лицо

# 4. Анализ данных. Пример.

## 4. Описание связей:

- Нагрузка - Сотрудники:

Одна нагрузка **должна** быть выполнена **одним или более** сотрудниками

Один сотрудник **может** быть задействован во **многих** задачах проектов

- Нагрузка- Проект:

Одна нагрузка **должна** быть выполнена для **одного** проекта

Один проект **может** быть выполнен во **многих** нагрузках

- Проекты - Клиенты:

Один проект **должен** быть связанный с **одним** клиентом

Один клиент **может** иметь **много** проектов

# Домашнее задание

**Задание:** Разработать базу данных интернет-магазина компьютерной техники и комплектующих

Основная **цель разработки** интернет-магазина: организовать продажи и отслеживать историю работы с клиентами (пример compservise, rosetka, eldorado).

**Задание 1:** Выполнить анализ предметной области (данных). Выделить не менее 4-ех функций

**Оформление:** текстовый документ