



ТЕХНОТРЕК

Занятие №1

Архитектура и проектирование Баз Данных

Дина Сафина, Павел Лысак

Mail.Ru Games



Цель курса



1. Отвечать на вопросы по БД на собеседовании
2. Писать SQL-запросы любой сложности
3. Грамотно проектировать базы данных
4. Обосновывать выбор инструментов для хранения данных

Структура курса



1. SQL
2. Проектирование
3. Архитектура

10 лекций + защита проекта



Всё выкладывается на github.

Оценивается:

1. Проработанность области деятельности
2. Логическая и физическая схема базы данных
3. Правильность подсчёта основных метрик
4. Оптимальность построения запросов

Баллы за курс



За что начисляются баллы:

- 75 – домашнее задание
- 15 – защита проекта
- 10 – доклад

Баллы	Оценка
0 – 50	2
51 – 69	3
70 – 84	4
85 – 199	5



<https://t.me/joinchat/A9Gm9hJXOPr0-gyyqEBjng>

План первой лекции



1. Что такое Базы Данных
2. Реляционные Базы Данных
3. Реляционная алгебра
4. Ключи
5. Немного про MySQL
6. Типы данных в MySQL
7. Домашняя работа

Где используются базы данных?



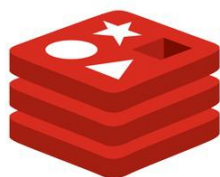
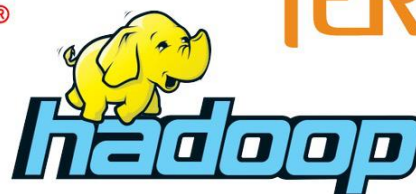
Самые популярные Базы Данных



PostgreSQL

ORACLE®

TERADATA



redis



mongoDB®



ClickHouse



- **База данных** — совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных
- **Система управления базами данных (СУБД)** — совокупность программных и лингвистических средств общего или специального назначения, обеспечивающих управление созданием и использованием баз данных

Чем СУБД лучше файликов на диске



- + Совместный доступ
 - + Непротиворечивость
 - + Целостность данных
 - + Масштабируемость
 - + Транзакционность
 - + Скорость обработки
 - + Готовые способы манипуляции данными
- Нужно знать SQL
 - Имеет смысл только с какого-то объёма информации



- Иерархическая
- Сетевая
- Объектно-ориентированная
- **Реляционная**

Реляционная модель данных



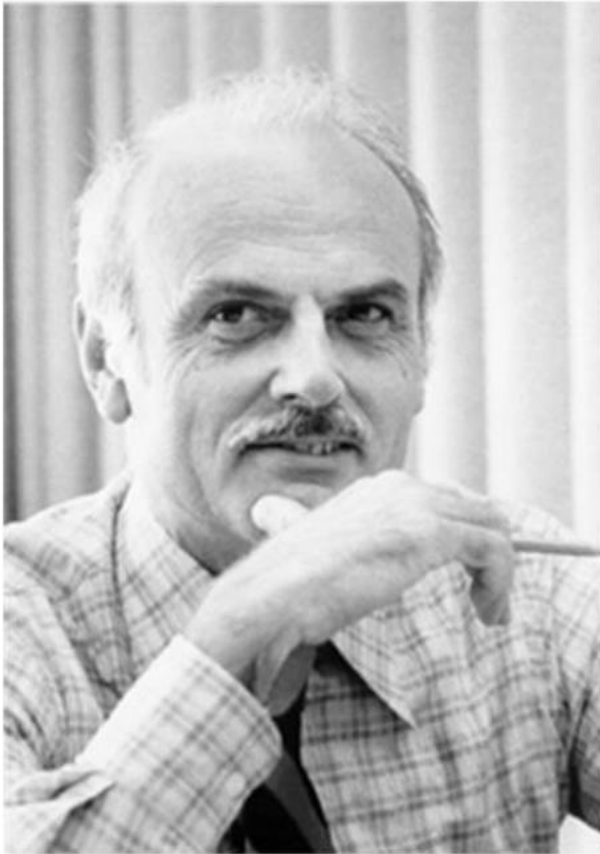
- **Реляционной моделью** называется база данных, в которой все данные, доступные пользователю, организованы в виде таблиц, а все операции над данными сводятся к операциям над этими таблицами
- Сформулирована в 1970 Эдгаром Коддом
- Три аспекта:
 1. Структурный
 2. Манипуляционный
 3. Целостный



- **Structured Query Language** — декларативный язык программирования, применяемый для создания, модификации и управления данными в реляционной базе данных, управляемой соответствующей СУБД



Операции реляционной алгебры



- Выборка
- Проекция
- Объединение
- Пересечение
- Разность
- Произведение
- Деление
- Соединение



Персоны

Имя	Возраст	Вес
Harry	34	80
Donald	29	70
Helena	54	54
Peter	34	80

σ Возраст ≥ 34 (Персоны)

Имя	Возраст	Вес
Harry	34	80
Helena	54	54
Peter	34	80



Персоны

Имя	Возраст	Вес
Harry	34	80
Donald	29	70
Helena	54	54
Peter	34	80

πВозраст, Вес (Персоны)

Возраст	Вес
34	80
29	70
54	54
34	80

Объединение



Персоны

Имя	Возраст	Вес
Harry	34	80
Donald	29	70
Helena	54	54
Peter	34	80

Персонажи

Имя	Возраст	Вес
Daffy	24	19
Donald	29	70
Scrooge	81	27

Имя	Возраст	Вес
Harry	34	80
Donald	29	70
Helena	54	54
Peter	34	80
Duffy	24	19
Scrooge	81	27

Пересечение



Персоны

Имя	Возраст	Вес
Harry	34	80
Donald	29	70
Helena	54	54
Peter	34	80

Персонажи

Имя	Возраст	Вес
Daffy	24	19
Donald	29	70
Scrooge	81	27

Имя	Возраст	Вес
Donald	29	70

Разность



Персоны

Имя	Возраст	Вес
Harry	34	80
Donald	29	70
Helena	54	54
Peter	34	80

Персонажи

Имя	Возраст	Вес
Daffy	24	19
Donald	29	70
Scrooge	81	27

Имя	Возраст	Вес
Harry	34	80
Helena	54	54
Peter	34	80

Произведение



Мультфильмы

Код	Название
0	Gravity Falls
1	Duck Tales
2	Over the Garden Wall

Каналы

Код	Название
0	Disney
1	Cartoon Network

Мультфильмы.Код	Мультфильмы.Название	Каналы.Код	Каналы.Название
0	Gravity Falls	0	Disney
0	Gravity Falls	1	Cartoon Network
1	Duck Tales	0	Disney
1	Duck Tales	1	Cartoon Network
2	Over the Garden Wall	0	Disney
2	Over the Garden Wall	1	Cartoon Network

Деление



Мультфильмы

Код	Название	Перевод
0	Gravity Falls	Сыендук
0	Gravity Falls	LostFilm
0	Gravity Falls	Кубик в Кубе
1	Duck Tales	LostFilm
1	Duck Tales	Кубик в Кубе
2	Over the Garden Wall	Сыендук
2	Over the Garden Wall	LostFilm

Перевод

Название
Сыендук
LostFilm

Мультфильмы.Код	Мультфильмы.Название
0	Gravity Falls
2	Over the Garden Wall

Соединение



Мультфильмы

Код	Название	Каналы.Код
0	Gravity Falls	0
1	Duck Tales	0
2	Over the Garden Wall	1

Каналы

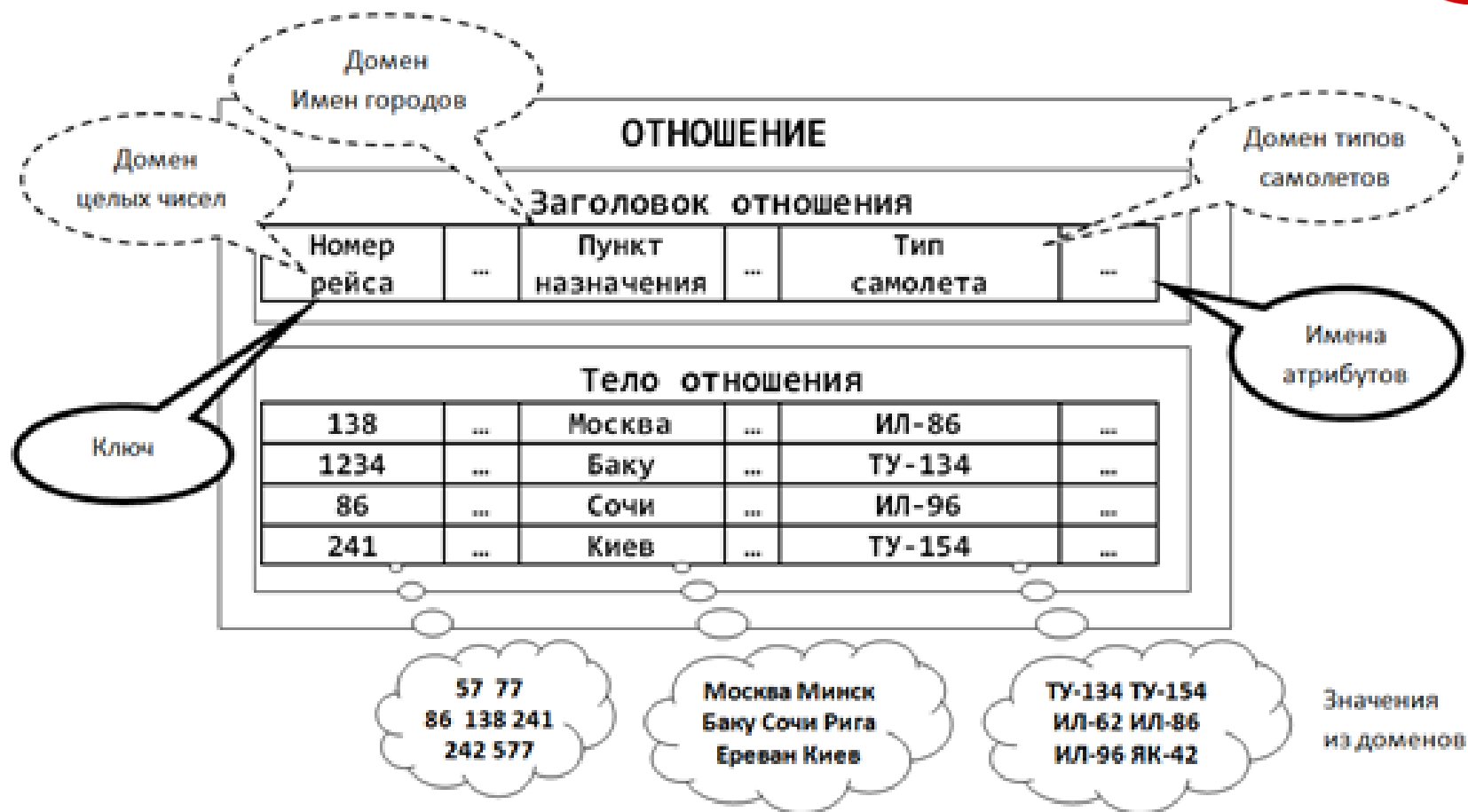
Код	Название
0	Disney
1	Cartoon Network

Мультфильмы.Код	Мультфильмы.Название	Каналы.Код	Каналы.Название
0	Gravity Falls	0	Disney
1	Duck Tales	0	Disney
2	Over the Garden Wall	1	Cartoon Network



- **Атрибут** – свойство сущности в предметной области
- **Кортеж** – множество именованных атрибутов, которое содержит одно вхождение каждого имени атрибута, принадлежащего схеме отношения
- **Домен** – допустимое множество значений
- **Отношение** – множество кортежей (неупорядоченное)
- **Целостность** – соответствие имеющейся в БД информации её внутренней логике, структуре и всем явно заданным правилам

Сущности Базы Данных



Первичный ключ



Потенциальный ключ – уникальность + минимальность

Первичный ключ – один из потенциальных

Альтернативный ключ – все остальные

Суррогатный ключ – искусственно введённый идентификатор

Внешний ключ – использование первичного ключа одного отношения в другом

Суррогатный ключ



- | | |
|--|---|
| <ul style="list-style-type: none">+ Неизменность+ Гарантированная уникальность+ Гибкость+ Эффективность (длина ссылок)+ Упрощение разработки | <ul style="list-style-type: none">- Уязвимость- Неинформативность- Неоптимальность (одновременное обслуживание уникальности естественного и суррогатного ключа) |
|--|---|

Почему MySQL?



345 systems in ranking, September 2018

Rank			DBMS	Database Model	Score		
Sep 2018	Aug 2018	Sep 2017			Sep 2018	Aug 2018	Sep 2017
1.	1.	1.	Oracle +	Relational DBMS	1309.12	-2.91	-49.97
2.	2.	2.	MySQL +	Relational DBMS	1180.48	-26.33	-132.13
3.	3.	3.	Microsoft SQL Server +	Relational DBMS	1051.28	-21.37	-161.26
4.	4.	4.	PostgreSQL +	Relational DBMS	406.43	-11.07	+34.07
5.	5.	5.	MongoDB +	Document store	358.79	+7.81	+26.06
6.	6.	6.	DB2 +	Relational DBMS	181.06	-0.78	-17.28
7.	↑ 8.	↑ 10.	Elasticsearch +	Search engine	142.61	+4.49	+22.61
8.	↓ 7.	↑ 9.	Redis +	Key-value store	140.94	+2.37	+20.54
9.	9.	↓ 7.	Microsoft Access	Relational DBMS	133.39	+4.30	+4.58
10.	10.	↓ 8.	Cassandra +	Wide column store	119.55	-0.02	-6.65

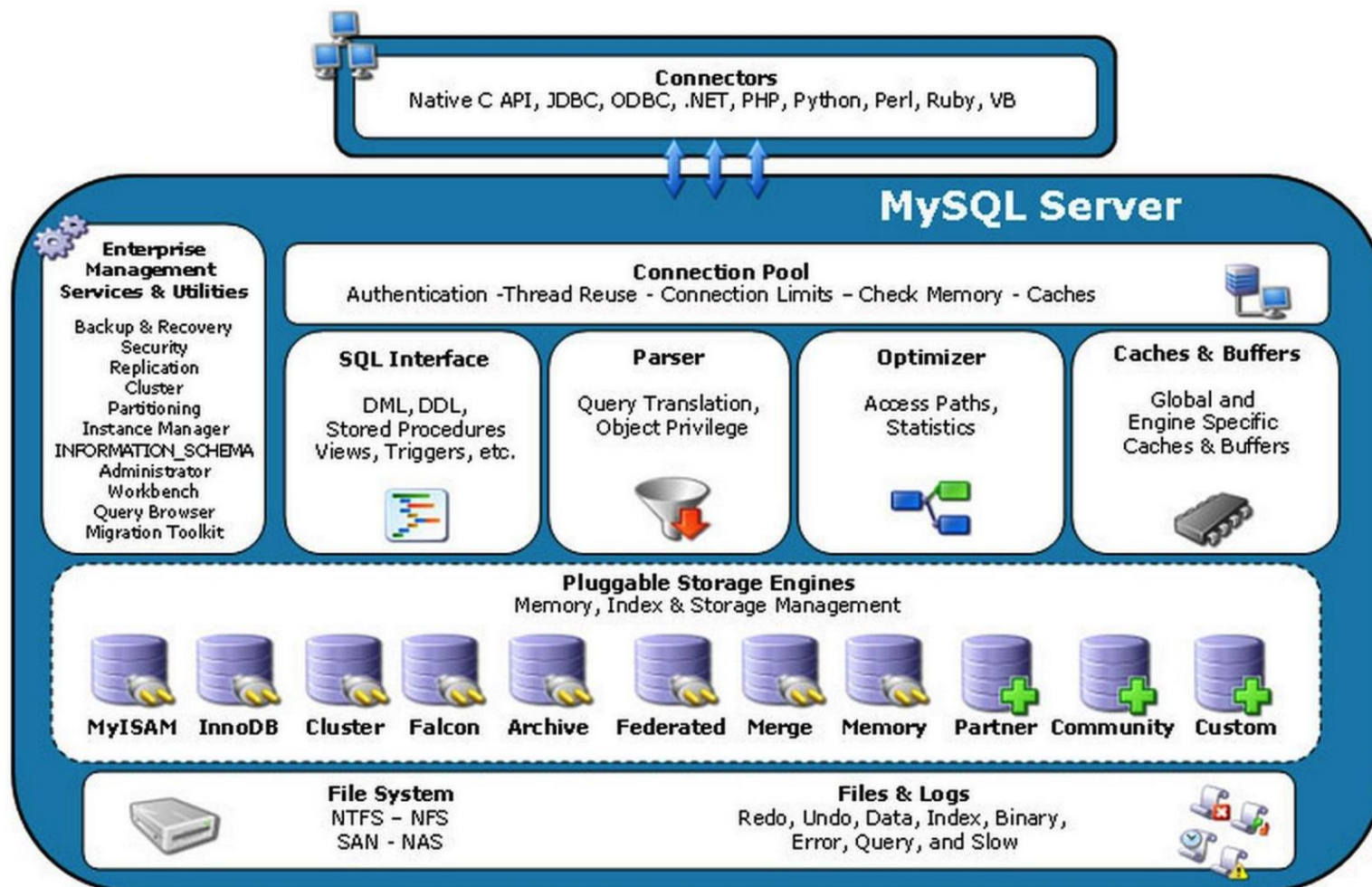
<https://db-engines.com>

Немного о MySQL



- Реляционная СУБД
 - Кроссплатформенная
 - Несколько разных движков
 - Двойное лицензирование (GPL + платное)
-
- Появилась в 1995г
 - В 2008г Sun Microsystems MySQL AB
 - В 2009г от MySQL отпочковалась MariaDB
 - В 2010г Oracle купил Sun Microsystems

Архитектура MySQL



Движок MyISAM



- До MySQL 5.5 была движком по умолчанию
- Полнотекстовый поиск
- Отсутствие транзакций
- Отсутствие внешних ключей
- Блокировка на уровне таблицы
- Заброшен разработчиками

Движок InnoDB



- Основной движок MySQL
- Поддерживает транзакции
- Поддерживает внешние ключи
- Блокировки на уровне строк
- Ближе всего к поддержке спецификации SQL-92
- С версии MySQL 5.6.4 поддерживает полнотекстовый поиск

Типы данных в MySQL



1. Числовые
2. С плавающей точкой
3. Бинарные
4. Строковые
5. Календарные
6. Географические
7. JSON

Числовые типы данных



Тип поля	Объём памяти	Диапазон значений
TINYINT [(M)]	1 байт	От -128 до 127 От 0 до 255
SMALLINT [(M)]	2 байта	От -32 768 до 32 767 От 0 до 65 535
MEDIUM INT [(M)]	3 байта	От -8 388 608 до 8 388 608 От 0 до 16 777 215
INT [(M)]	4 байта	От -2 147 683 648 до 2 147 683 648 От 0 до 4 294 967 295
BIGINT [(M)]	8 байт	От -2^{63} до $2^{63} - 1$ От 0 до 2^{64}

Числа с плавающей точкой



Тип поля	Объём памяти	Диапазон значений
DECIMAL [(M [, D])], DEC [(M [, D])], NUMERIC [(M [, D])]	M + 2 байта	Повышенная точность; зависимость параметров
FLOAT [(M, D)]	4 байта	От $-3.402823466E + 38$ до $1.175494351E - 38$ От $1.175494351E - 38$ до $3.402823466E + 38$
DOUBLE [(M, D)], REAL [(M, D)], DOUBLE PRECISION [(M, D)]	8 байт	От $-1.7976931348623157E + 308$ до $-2.2250738585072014E - 308$ От $2.2250738585072014E - 308$ до $1.7976931348623157E + 308$

Бинарные типы



Тип поля	Объём памяти	Диапазон значений
BIT [(M)]	$(M + 7) / 8$ байт	От 1 до 64 бита, в зависимости от M
BOOLEAN [(M)]	1 байт	0 или 1

Строковые типы



Тип поля	Объём памяти	Диапазон значений
CHAR (M)	M	M символов
VARCHAR (M)	L + 1	M символов
TINYBLOB, TINYTEXT	L + 1	$2^8 - 1$ символов
BLOB, TEXT	L + 2	$2^{16} - 1$ символов
MEDIUMBLOB, MEDIUMTEXT	L + 3	$2^{24} - 1$ символов
LOB, LONGTEXT	L + 4	$2^{32} - 1$ символов
ENUM ('value1', 'value2',...)	1 или 2 байта	65 536 элементов
SET ('value1', 'value2',...)	1, 2, 3, 4 или 8	64 элемента

Календарные типы



Тип поля	Объём памяти	Диапазон значений
DATE	3 байта	От «1000-01-01» до «9999-12-31»
TIME	3 байта	От «-828:59:59» до «828:59:59»
DATETIME	8 байт	От «1000-01-01 00:00:00» до «9999-12-31 59:59:59»
TIMESTAMP	4 байта	От «1970-01-01 00:00:00» до «2038-12-31 59:59:59»
YEAR [(2 / 4)]	1 байт	От 1901 до 2115 для YEAR(4) От 1970 до 2069 для YEAR(2)

Географические типы



Тип поля	Смысл
GEOMETRY GEOMETRYCOLLECTION	Геометрии любого типа
POINT MULTIPOINT	Точка в пространстве
LINESTRING MULTILINESTRING	Набор точек, кривая
POLYGON MULTIPOLYGON	Замкнутый набор точек, граница

JSON



- JSON_TYPE()
- JSON_ARRAY()
- JSON_OBJECT()
- JSON_VALID()
- JSON_EXTRACT()
- JSON_REPLACE()
- JSON_INSERT()
- JSON_REMOVE()
- ...

NULL



Запрос	Результат
1 = NULL	NULL
1 != NULL	NULL
1 > NULL	NULL
1 < NULL	NULL
1 + NULL	NULL
1 IS NULL	0
1 IS NOT NULL	1
NULL = NULL	NULL
NULL != NULL	NULL
NULL IS NULL	1
NULL IS NOT NULL	0

Примеры тем проектов



- Управление городским транспортом
- Ломбард
- Издательство манги
- Продвинутая Мафия
- Рекомендательная музыкальная система
- Поисковик авиабилетов
- Спортивный трекер

Домашнее задание №1



- Установить MySQL Server
- Установить MySQL WorkBench
- Завести репозиторий на GitHub'e

- Выбрать тему проекта
- Перечислить сущности

Срок сдачи

26 сентября 2018