

# Технологии и разработка СУБД

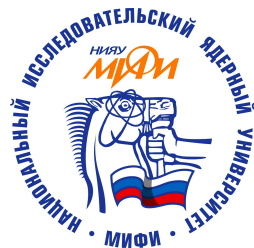
Анастасия Лубенникова  
Александр Алексеев

# Лекция 1

- Знакомство
- Обзор предметной области
- Ответы на вопросы
- Если позволит время: демонстрация

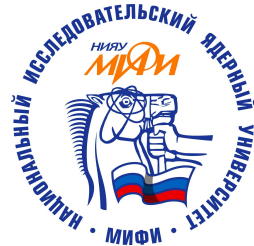
# О нас. Настя

- 2014 - стажировка в GSoC
- 2015 - закончила бакалавриат НИЯУ МИФИ
- 2014-2015 - работа в Cybertec
- с 2015 - работа в Postgres Professional
- 2016 - курс Hacking PostgreSQL
- 2017 - ментор в GSoC



# О нас. Саша

- Закончил НИЯУ МИФИ в 2010, специализация “Прикладная математика и информатика”
- Работаю программистом с 2007
- Контрибьютор в PostgreSQL с 2015
- Интересы: FOSS, функциональное программирование, электроника (микроконтроллеры, FPGA, SDR, ...), распределенные системы, блогинг, подкастинг
- <https://eax.me/> & <https://devzen.ru/>



# О вас

- уровень английского
- C / C++ / Go
- git
- алгоритмы и структуры данных
  - hash, деревья, сортировки, оценка сложности
- основы ОС
- базы данных
  - реляционная модель
  - SQL / NoSQL
  - применение на практике

# О курсе

- О чем: *практика* разработки СУБД
- 13 лекций по средам с 16:20 до 18:00, ауд 248 (в ГУМ 2)
- Короткие тесты по предыдущим лекциям
- Практические задания
- Активные обсуждения на семинарах
- Экзамен в конце

Группа в Telegram (вместе с ВШЭ): <https://t.me/dbmsdev>

# О базах данных и СУБД

База данных - структурированный организованный набор данных.

СУБД (Система Управления Базами Данных) - программное обеспечение, предназначенное для организации работы с БД.

Транзакция - группа последовательных операций, которая представляет собой логическую единицу работы с данными.

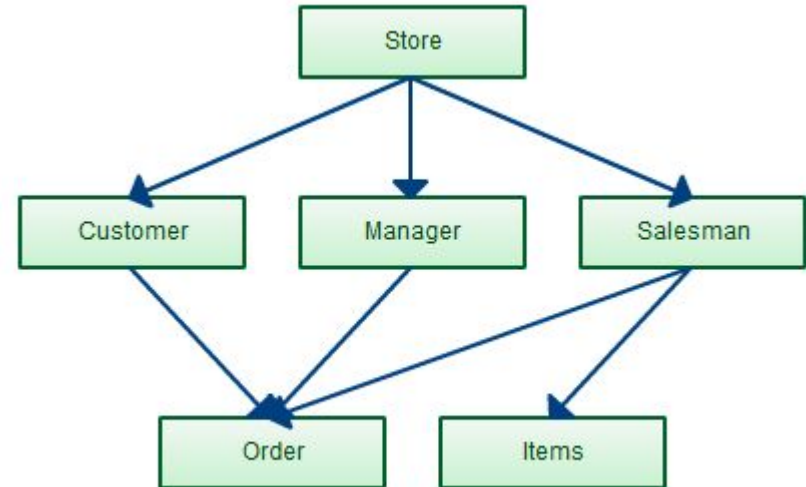
# О технологиях СУБД

- Красивые теоретические идеи
  - реляционная алгебра
  - контроль конкурентного доступа
  - распределенные системы
- Требования из реального мира
  - высокая надежность и доступность данных (High availability)
  - безопасность доступа к данным
  - требования к производительности
  - множество компромиссов (trade-off)
- Работа со всеми уровнями абстракции
  - От разбора грамматики до принципов работы дисков и CPU



# История СУБД. 1960-е.

- Иерархические и Сетевые СУБД
- IBM Information Management System (IMS) разрабатывалась для космической программы “Аполлон”
- CODASYL
- Слишком сложны в использовании



## История СУБД. 1970-е.

“Будущие пользователи больших банков данных должны быть освобождены от необходимости знать организацию данных в машине (внутреннее представление)”

Э.Кодд, 1970г.

# Реляционная модель

- Relation (Table)
- Attribute (Column, Field)
- Tuple (Row, Record)

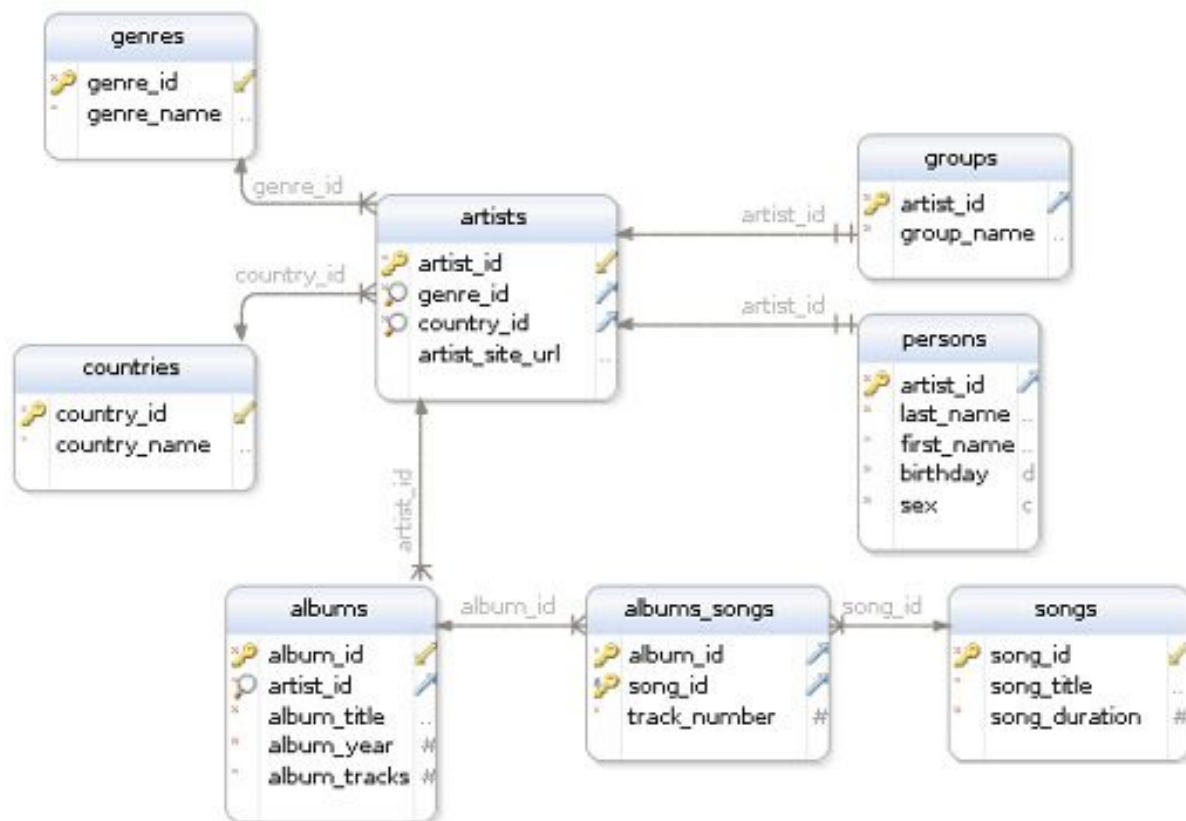
Activity Code	Activity Name
23	Patching
24	Overlay
25	Crack Sealing

Key = 24

Activity Code	Date	Route No.
24	01/12/01	I-95
24	02/08/01	I-66

Date	Activity Code	Route No.
01/12/01	24	I-95
01/15/01	23	I-495
02/08/01	24	I-66

# Реляционная модель. Ключи.



# SQL (Structured Query Language)

- Декларативный язык
  - “что” нужно сделать, а не “как”
- Примеры других декларативных языков
  - Haskell, Prolog

# SQL. DDL (Data Definition Language)

```
CREATE TABLE users (id SERIAL PRIMARY KEY, login CHAR(64),  
                    password CHAR(64));
```

```
ALTER TABLE users ADD COLUMN birthday date;
```

```
ALTER TABLE users DROP COLUMN birthday;
```

```
ALTER TABLE users RENAME COLUMN password TO pass;
```

```
DROP TABLE users;
```

# SQL. DML (Data Manipulation Language)

**INSERT INTO users (login, pass) values ('alex', 'qwerty');**

**SELECT \* FROM users;**

**UPDATE users SET pass = 'secret' WHERE id = 1;**

**DELETE FROM users WHERE id = 1;**

# OLTP & OLAP

- OLTP (Online Transaction Processing)
  - много маленьких транзакций, которые должны выполняться быстро
- OLAP (Online Analytical Processing)
  - сложные и долгие аналитические запросы



# ACID

- Atomicity
- Consistency
- Isolation
- Durability

# Классификация СУБД

- Relational: PostgreSQL, MySQL, MS SQL Server, Oracle, DB2, SQLite
- Column: Vertica, BigTable, ClickHouse, Druid, Pinot
- Document: MongoDB, CouchDB, RethinkDB
- Key-value: Redis, Riak, Couchbase
- Graph: Neo4J
- Spatial: PostGIS, TerraLib, ...
- Full Text Search: ElasticSearch, Solr, Sphinx
- Queue: RabbitMQ, Kafka, ...
- MapReduce & Co: Hadoop, Spark, ...
- ...

# PostgreSQL

- Open source object-relational database system
- Свободная лицензия BSD
- Больше 20ти лет работы над проектом (старше Linux!)
- Сотни разработчиков по всему миру
- Написан на C
- Кроссплатформенный (Windows, Linux, MacOS, Solaris, \*BSD, AIX, ...)
- Высокое качество кода\*
- Отличная документация\*

# Домашнее задание №1

- Настроить Linux-окружение для разработки (можно виртуалку)
- Освежить знания C и Git
- Собрать PostgreSQL из исходников и установить
- Спроектировать схему базы данных “телефонная книга”

# Рекомендуемые материалы

- **Database System Implementation** by Hector Garcia-Molina, Jeffrey Ullman
- **Readings in Database Systems, Fifth Edition** by Peter Bailis, Joseph M. Hellerstein, Michael Stonebraker (<http://www.redbook.io>)
- **Distributed Systems for Fun and Profit** by Mikito Takada (<http://book.mixu.net/distsys/>)
- **Database System Concepts** by Avi Silberschatz, Henry F. Korth, S. Sudarshan (<http://db-book.com/>)
- **Начало работы с PostgreSQL** <https://eax.me/postgresql-install/>
- **PostgreSQL: сборка из исходников и настройка под Linux** <https://eax.me/postgresql-build/>
- **Становимся контрибьютером в PostgreSQL** (<https://habr.ru/p/308442/> + далее по ссылкам)

## Вопросы и ответы.

- [a.lubennikova@postgrespro.ru](mailto:a.lubennikova@postgrespro.ru)
- [a.alekseev@postgrespro.ru](mailto:a.alekseev@postgrespro.ru)
- Telegram: <https://t.me/dbmsdev>