

# Технологии баз данных

Б.А. Новиков

НИУВШЭ в Санкт-Петербурге

[https://www.hse.ru/staff/Boris\\_Novikov](https://www.hse.ru/staff/Boris_Novikov)

**Курс разработан при поддержке компании Postgres Professional**

## Аннотация

В курсе рассматриваются модели, структуры данных и алгоритмы, применяемые в современных промышленных системах управления базами данных, обсуждаются задачи администрирования баз данных и дополнительные возможности СУБД. Значительное внимание уделяется особенностям системы PostgreSQL, в том числе реализации рассматриваемых методов и алгоритмов в этой системе, особенности применения различных структур данных, а также возможности расширения, предоставляемые этой системой. Для успешного освоения материала этого курса необходимо знакомство с основными разделами, обычно включаемыми в начальный курс по системам управления базами данных. В частности, предполагается знакомство с основными моделями данных, основами реляционной теории баз данных (включая реляционную алгебру), основными конструкциями языка SQL и основными понятиями, связанными с применением транзакций. В курсе используются базовые определения теории графов и теории множеств. Для выполнения упражнений необходим доступ к компьютеру. Знакомство с содержанием курса необходимо для разработчиков систем управления базами данных и администраторов данных. Оно полезно для проектировщиков больших прикладных систем, в которых интенсивно применяются базы данных.

## Результаты освоения курса:

В результате освоения курса студенты получают:

- Знание структур данных и алгоритмов, применяемых в промышленных системах управления базами данных;
- Представление о современном научно-исследовательском контексте в области систем управления базами данных;
- Представление о методах и средствах программирования на сервере баз данных и дополнительных возможностях языка SQL и расширений, имеющихся в СУБД PostgreSQL;
- Знание методов расширения, существующих в СУБД PostgreSQL, и методов реализации расширяемости;
- Знание особенностей параллельных и распределенных архитектур систем управления базами данных и применяемыми в таких системах алгоритмами и протоколами.

## Содержание курса

Примерное распределение материала курса по часам лекций:

1. Введение: архитектуры СУБД

2. Структуры хранения
3. Индексные структуры
4. Алгоритмы реляционных операций
5. Оптимизация и выполнение запросов
6. Управление транзакциями: критерии корректности
7. Управление транзакциями: протоколы
8. Надежность: восстановление после отказов
9. Дополнительные средства SQL
10. Проектирование запросов
11. Функции и процедуры
12. PL/pgSQL
13. Расширяемость PostgreSQL
14. Полнотекстовый поиск
15. Администрирование
16. Репликация
17. Выполнение запросов в параллельных и распределенных системах
18. Согласованность в распределенных системах.

## Контрольные инструменты

Для проверки подготовленности студентов и степени освоения материала будут применяться:

- Экспресс-тесты, реализуемые во время лекций с применением средства голосования системы ZOOM;
- Проверка выполнения домашних заданий, отчеты о выполнении домашних заданий должны высылаться на электронную почту;
- Зачет в устной форме на основе ответов на контрольные вопросы (устный зачет применяется в случае неполных или неточных ответов на экспресс-тесты и неполного или неточного выполнения домашних заданий).

# Контрольные вопросы для зачета в устной форме

1. Основные компоненты СУБД
2. Структуры хранения объектов базы данных
3. Одномерные индексы: деревья и хеширование
4. Пространственные индексы
5. Алгоритмы выборки данных
6. Алгоритмы соединения
7. Этапы обработки запросов
8. Задача оптимизации запросов
9. Алгоритмы оптимизации запросов снизу вверх
10. Модели стоимости и целевые функции
11. Транзакции, истории и расписания
12. Семантика Эрбрана и критерии корректности по конечному и по видимому состоянию
13. Конфликты и критерий сериализуемости по конфликтам
14. Классические протоколы управления транзакциями
15. Протоколы на основе меток времени и мгновенных снимков
16. Многоверсионные протоколы управления транзакциями
17. Откаты транзакций и восстановимость
18. Ведение журнала обновлений
19. Алгоритмы восстановления после системных отказов
20. Контрольные точки и точки восстановления
21. Стратегии восстановления носителя базы данных
22. Рекурсия в SQL
23. Применение функциональных и условных индексов в PostgreSQL
24. Оконные функции в SQL
25. Требования к расширяемости и средства расширения в PostgreSQL
26. Пользовательские типы, функции, процедуры и агрегаты в PostgreSQL

27. Обобщенные индексные структуры в PostgreSQL
28. Полнотекстовый и нечеткий поиск в PostgreSQL
29. Средства разграничения доступа
30. Основные функции администратора базы данных
31. Задачи настройки баз данных и приложений
32. Выполнение запросов в параллельных системах управления базами данных
33. Выполнение распределенных запросов
34. Поддержка согласованности в распределенных системах

## Основная и дополнительная литература

Материал курса подготовлен на основе книги [1], которую следует использовать в качестве основной литературы. Дополнительная литература включает [2] и многочисленные исследовательские статьи, процитированные в [1]. Предварительные сведения, необходимые для освоения курса, можно почерпнуть из первой части той же книги, из учебника [3] или из других учебников по системам управления базами данных.

# Литература

- [1] Б. Новиков, Е. Горшкова, Н. Графеева "Основы технологий баз данных". ДМК Пресс, Москва, 03 2020
- [2] Б. Новиков, Г. Добровская "Настройка приложений баз данных"ВНУ, Санкт-Петербург, 2006
- [3] С.Д. Кузнецов "Базы данных: учебник для студ. учреждений высшего проф. образования"Университетский учебник. Прикладная математика и информатика.