Основы технологий баз данных

Б. А. Новиков
Санкт-Петербургский государственный университет, JetBrains Research

PROFESSIONAL Postgres

Об этом курсе

PROFESSIONAL Postgres

Содержание курса: где это можно прочитать?

Основное содержание курса изложено в этой книге:

Б. А. Новиков, Е. А. Горшкова, Графеева Н. Г.
Основы технологий баз данных, часть 1.
Postgres Professional, 2018.
https://postgrespro.ru/education/books/dbtech

Лектор оставляет за собой право излагать материал с дополнениями, сокращениями и в измененном порядке.

Для освоения этого курса

★Необходимы

- Навыки работы с компьютером
- Знания основных определений теории множеств
- Знание базовых определений теории графов
- Знакомство с архитектурами компьютеров и вычислительных сетей

★Желательны

- Знакомство с каким-либо языком программирования
- Знакомство с алгоритмами и структурами данных

Что можно и что невозможно узнать в результате освоения этого курса

+ Теоретические модели баз данных

+ Назначение и свойства языков запросов

+ Принципы организации взаимодействия баз данных и приложений

+ Основные возможности СУБД PostgreSQL

- Практические навыки проектирования баз данных

- Навыки программирования запросов

- Детальная информация о конкретных системах управления базами данных

Postgres PROFESSIONAL

Кто-нибудь еще об этом писал? (рекомендуемая литература)

- Date C. Introduction to Database Systems. (имеются русские переводы)
- Г. Гарсиа-Молина, Дж. Ульман, Дж. Уидом. Системы баз данных. Полный курс. М.: Вильямс, 2004.
- J. Ullman. Principles of database and knowledge base systems. Vol. 1. Academic Press, 1988—1989.
- С. Д. Кузнецов. Основы баз данных. М.: Бином, 2007.
- Б. А. Новиков, Г. Р. Домбровская: Настройка приложений баз данных. СПб.: БХВ, 2006.

Postgres PROFESSIONAL

К сожалению, писали многие (НЕ рекомендуемая литература)

- Базы данных для «чайников»
- Базы данных для домохозяек
- Случайные источники в Интернет
- Руководства типа hands-on
- Рекламные материалы по коммерческим продуктам

Postgres PROFESSIONAL

О лекторе

Борис Арсенович Новиков

регалии (степени, звания и т. п.) есть, но это не так важно

профессор кафедры информационно-аналитических систем Санкт-Петербургского государственного университета

Руководитель лаборатории MLIM в JetBrains Research

Немного больше можно найти здесь:
https://research.jetbrains.org/groups/mach_learning

Введение

PROFESSIONAL Postgres

4000 лет до нашей эры

Возрождение

Luca Pachioli

+

Правила ведения бухгалтерских записе

Леонардо

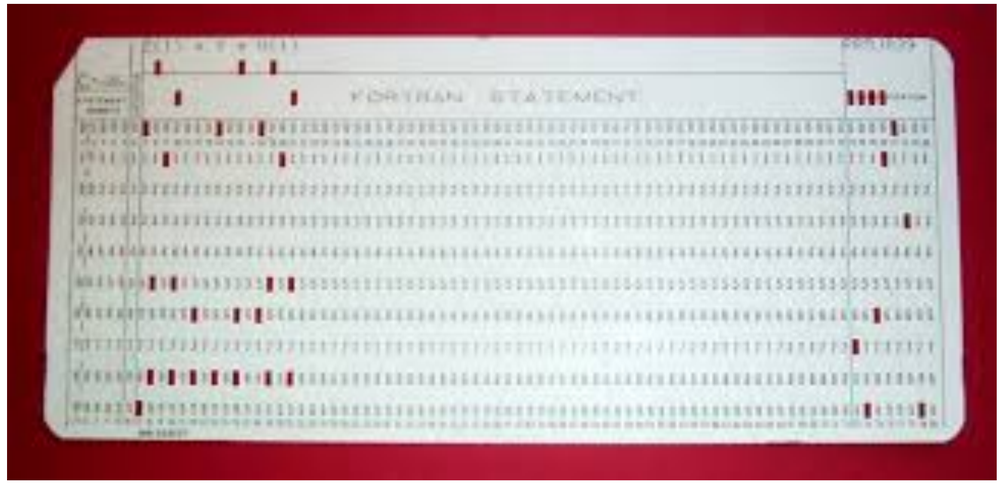
Summa de
Arithmetica geo
metria. Proportioni : et proportionalita:

Continentia de tutta lopera:

De numeri e misure in tutti modi occurrenti.
Proportioni e ꝓportionalita a notitia del 5.° de Euclide e de tutti li altri soi libri.
Chiauiouero euidentie numero. 13. per le quantita continue ꝓportionali del 6.° e 7.° de Euclide extratte.
Tutte le parti de lalgorismo: cioe relevare partire multiplicare sumare: e sotrarre con tutte sue ꝓpe in fani e rotti e radici e progressioni.
De la regola mercantesca vsitta del. 3. e soi fondamenti cō casi exemplari p cl m° §.5. guadagni: perdite: transportationi: e investite.
Partire multiplicare: sumare: e sotrar de le ꝓportioni: e de tutte sorti radici.
De le tre regole del Catayn vsitta positione: e sua origine.
Euidentie generali: ouer conclusioni numero. 66. absolutere ogni caso che per regole ordinarie non si potesse:
Tutte sorte binomii e recisi: e altre linee irrationali del decimo de Euclide.
Tutte regole de Algebra vsitte de la cosa e loro fabriche e fondamenti.
Spagnie in tutti modi: e loro partire.
Socide de bestiami: e loro partire.
Fitti: pensioni: ottimi: liuelli: logagioni: e godimenti.
Baratti in tutti modi semplici: composite col tempo.
Cambi reali: secchi: fittitij: e vsimati: ouer communi.
Meriti semplici e a capo danno: e altri termini
Resti: saldi: sconti: de tēpo e denari: e de recare a vn di piu partite.
Oriargenti: oloro affinare: e caratare.
Molti casi e ragioni straordinarie: varie e diuerse a tutte occurrēte: cōmo nella seguente tauola appare ordinatamente de tutte.
Ordine a saper tener ogni cōtoe scriptura nel quaderno in vinegia.
Tarifa de tutte vsance e costumi mercantesci in tutto el mondo.
Pratica e theorica de geometria: e de li cinque corpi regulari: e altri ꝓpedenti
E molte altre cose de grandissimi piacere e frutto: comuno difficilmente per la seguente tauola appare.

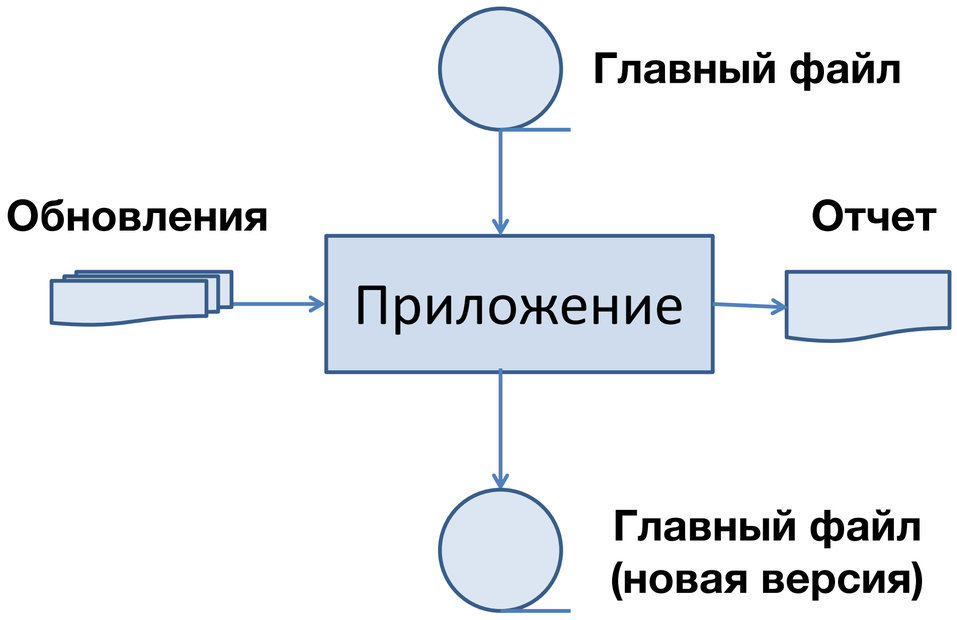
Fil. Dr GUSTAF MESSTRÖM
BIBLIOTEK
STOCKHOLMS HÖGSKOLA

Перепись населения США (1899)



Все было хорошо…

- Основной носитель данных — магнитные ленты
- Строго последовательная обработка
- Простая логика работы приложения
- Пакетное обновление (ежесуточно)
- Предыдущая версия сохраняется
- Простое повторение при ошибках и сбоях



Все хорошо, но…

- Время отклика системы на любой запрос составляет примерно сутки, независимо от сложности запроса
- Полная перезапись всех данных, независимо от количества обновлений или запросов на чтение

Устройства прямого доступа

+ Возможность независимой обработки отдельных блоков данных
+ Малое время доступа
+ Возможность частичного обновления
+ Структуры данных потенциально любой сложности

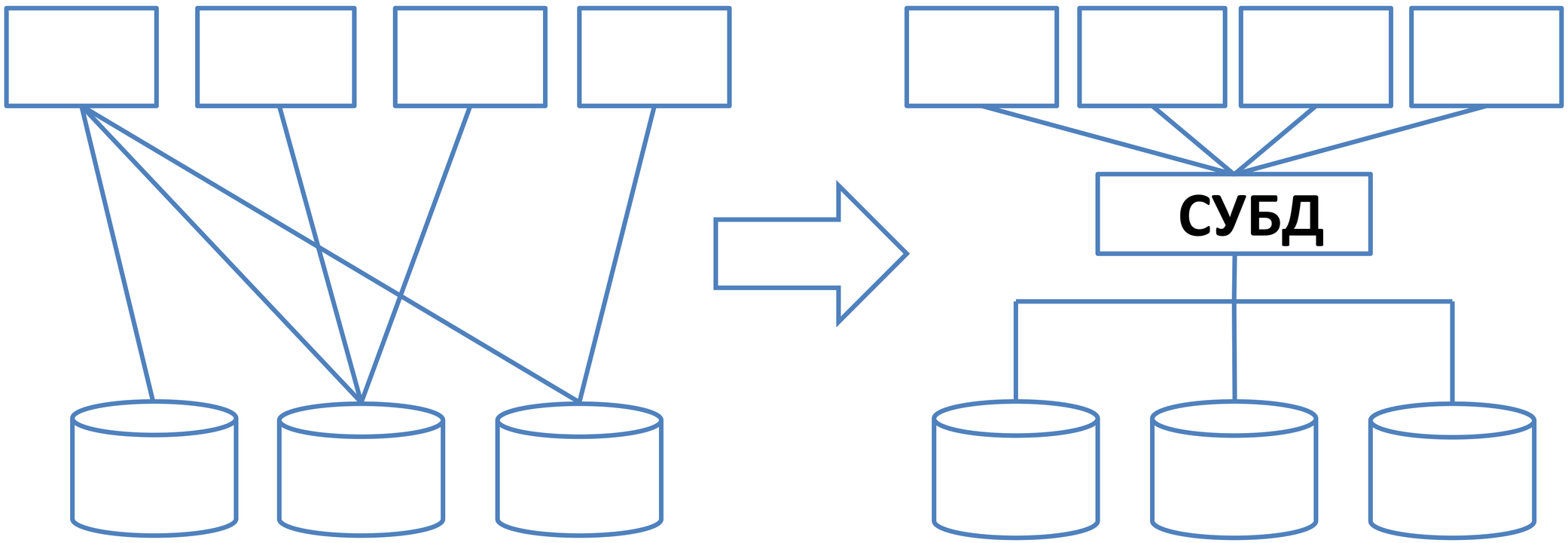
- Сложность программирования приложения
- Необходимость резервного копирования
- Сложность одновременного доступа к разделяемым данным

Postgres PROFESSIONAL

Предпосылки появления СУБД

- Появление устройств прямого доступа
- Выборочное обновление
- Сложные структуры данных
- Высокая стоимость данных в важнейших областях применения
- Дублирование средств управления данными в каждом приложении
- Высокая сложность создания приложений
- Чрезмерная сложность организации доступа к общим (разделяемым между приложениями) данным

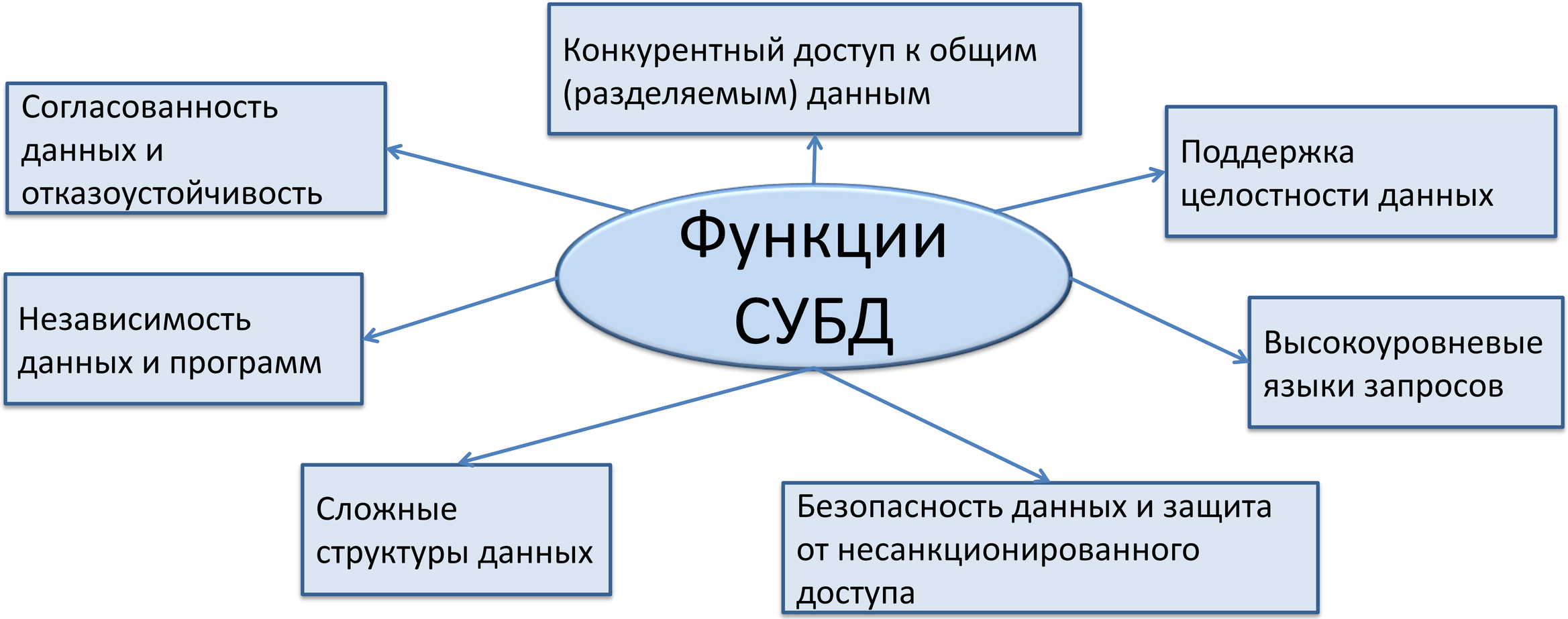
Централизация управления данными



Требования к системам управления базами данных (1971)

- Наличие средств описания данных
- Независимость данных и программ
- Высокоуровневый язык запросов
- Оперативный доступ к данным
- Совместное использование данных разными приложениями
- Целостность
- Согласованность
- Надежность
- Безопасность

Postgres PROFESSIONAL

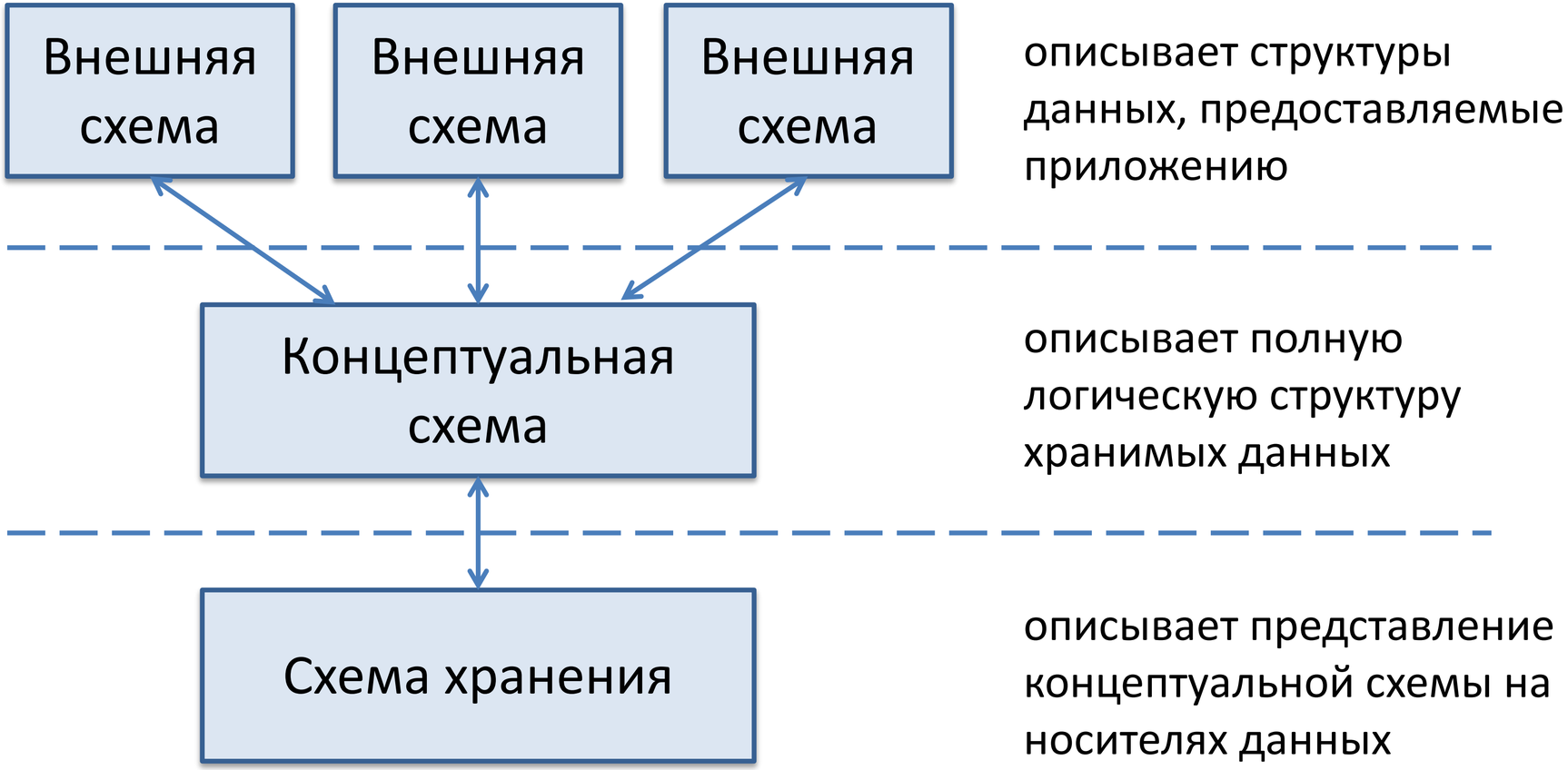

Функции СУБД
Конкурентный доступ к общим (разделяемым) данным
Согласованность данных и отказоустойчивость
Поддержка целостности данных
Независимость данных и программ
Высокоуровневые языки запросов
Сложные структуры данных
Безопасность данных и защита от несанкционированного доступа

Независимость данных и программ

- Возможность внесения изменений в структуру данных без изменения функциональности приложений
- Обеспечивается языком описания данных (Data Definition Language, DDL)
- Наличие логического описания данных — один из основных критериев, отличающих базы данных от файловых систем
- В настоящее время значение независимости данных и программ зачастую недооценивается
- **Само по себе описание данных не обеспечивает независимость**

Postgres PROFESSIONAL

Трехуровневая структура описания данных (1975)



Высокоуровневые языки запросов

- Описание структуры данных обеспечивает возможность задания операций логического уровня без создания приложения
- Современные языки запросов можно и целесообразно применять как декларативные:
  - Запрос задает требования к результату (что нужно получить), но не алгоритм его получения
  - Декларативные языки позволяют получить высокую эффективность

Postgres PROFESSIONAL

Целостность (Integrity)

- Средства поддержки целостности гарантируют соблюдение ограничений на значения и структуру данных
- СУБД предотвращает нарушение ограничений целостности
- Ограничения целостности необходимо учитывать при разработке приложений
- Наличие ограничений целостности позволяет исключить проверку их выполнения в приложении

Согласованность (Consistency)

- Корректность данных может нарушаться при одновременной работе приложений, каждое из которых работает корректно при отсутствии помех
- Средства поддержки согласованности позволяют обеспечить корректность одновременного выполнения нескольких приложений, работающих с общими данными
- Какие именно состояния базы данных являются согласованными, зависит от семантики приложений

Надежность хранения (Reliability)

- Ценность данных может быть очень высока и значительно превышать ценность оборудования или программного кода
- СУБД может гарантировать сохранность данных при отказах оборудования или системы
- Современные СУБД позволяют добиться очень высокой надежности, однако высоконадежные конфигурации оказываются весьма дорогостоящими
- Значение средств обеспечения надежности часто недооценивается

Postgres PROFESSIONAL

Безопасность (Security)

- Защита от несанкционированного доступа к данным
  - Работать с базой данных могут только зарегистрированные пользователи
- Разграничение доступа
  - Пользователь может работать только с той частью данных, к которой ему предоставлен доступ
- Ограничение полномочий
  - Пользователь может выполнять только разрешенные ему операции над данными
- Во многих современных системах средства безопасности СУБД не используются

Postgres PROFESSIONAL

Производительность

- Основные метрики производительности:
  - Пропускная способность
  - Время отклика
  - Доступность
  - Масштабируемость?
- Как измерить производительность?

Postgres PROFESSIONAL

Уточним термины

База данных (БД)	Совокупность хранимых данных вместе с описаниями
Система управления базами данных (СУБД)	Комплекс программ, реализующих функции управления базами данных
Сервер баз данных (кластер)	СУБД, установленная на вычислительной системе в состоянии выполнения
Экземпляр (Instance) базы данных	База данных, доступ к которой в данный момент обеспечивается сервером базы данных

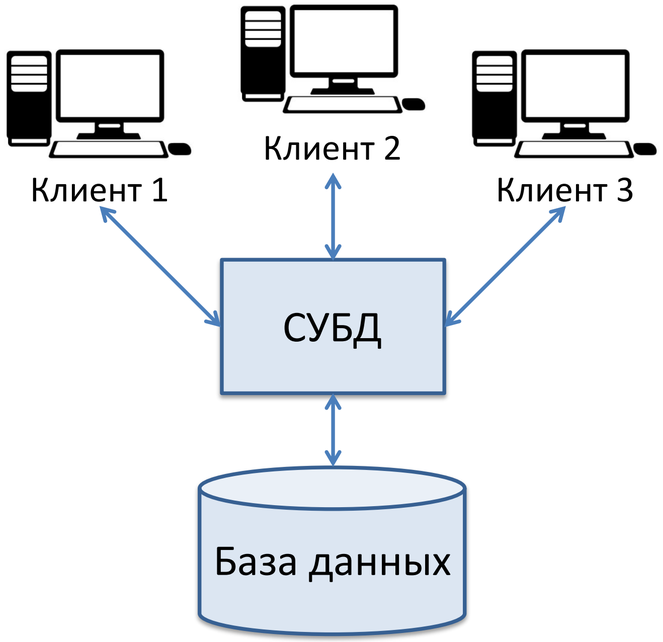
Postgres PROFESSIONAL

Конфигурации серверов баз даных

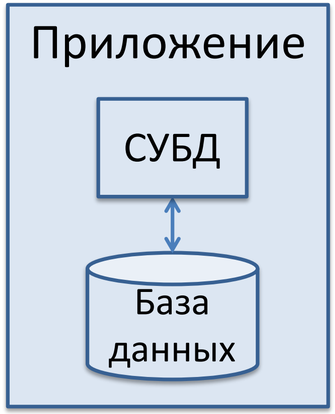
- Централизованные, на одном сервере
- Централизованные, параллельные
- Распределенные

Взаимодействие с приложениями

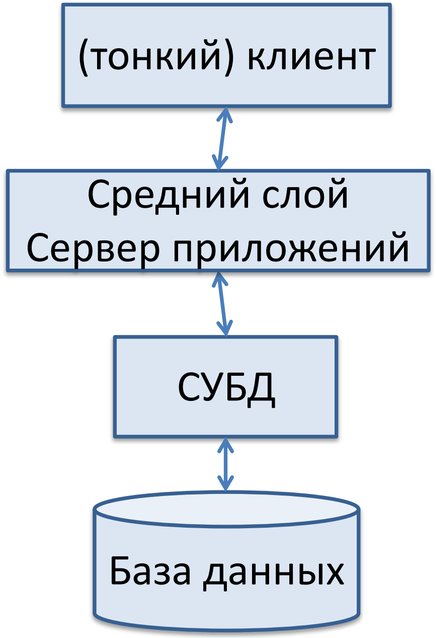
Клиент-сервер

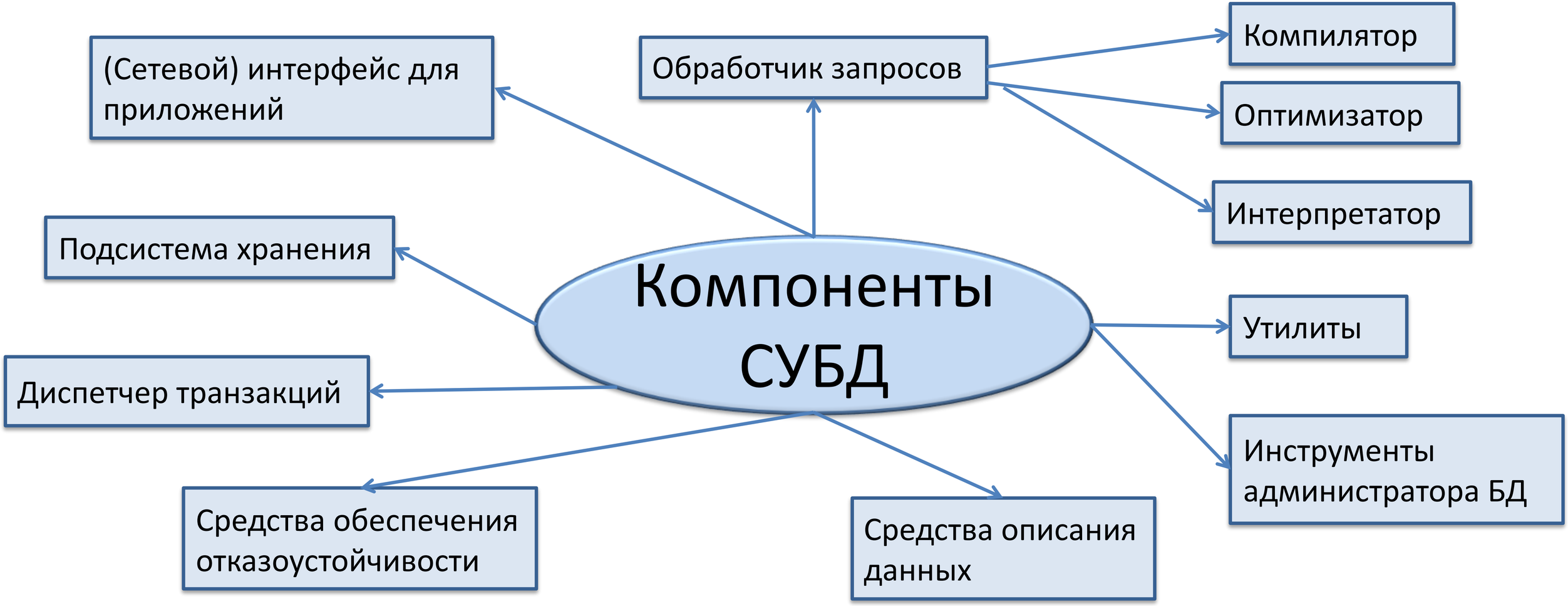


Встроенная СУБД



Сервер приложений




Компоненты СУБД
(Сетевой) интерфейс для приложений
Обработчик запросов
Компилятор
Оптимизатор
Интерпретатор
Подсистема хранения
Утилиты
Диспетчер транзакций
Инструменты администратора БД
Средства обеспечения отказоустойчивости
Средства описания данных

Демонстрационная база данных PostgreSQL

https://postgrespro.ru/education/demodb

Базы данных в настоящее время

- Популярность в учебных заведениях и в промышленности
- Базы данных и «Большие данные»
- Проектирование системы по требованиям или эмпирический анализ?