

Данные. Модели данных. Ранние подходы к организации данных



Включить звук



Включить видео



Участники 19



Чат



Демонстрация экрана

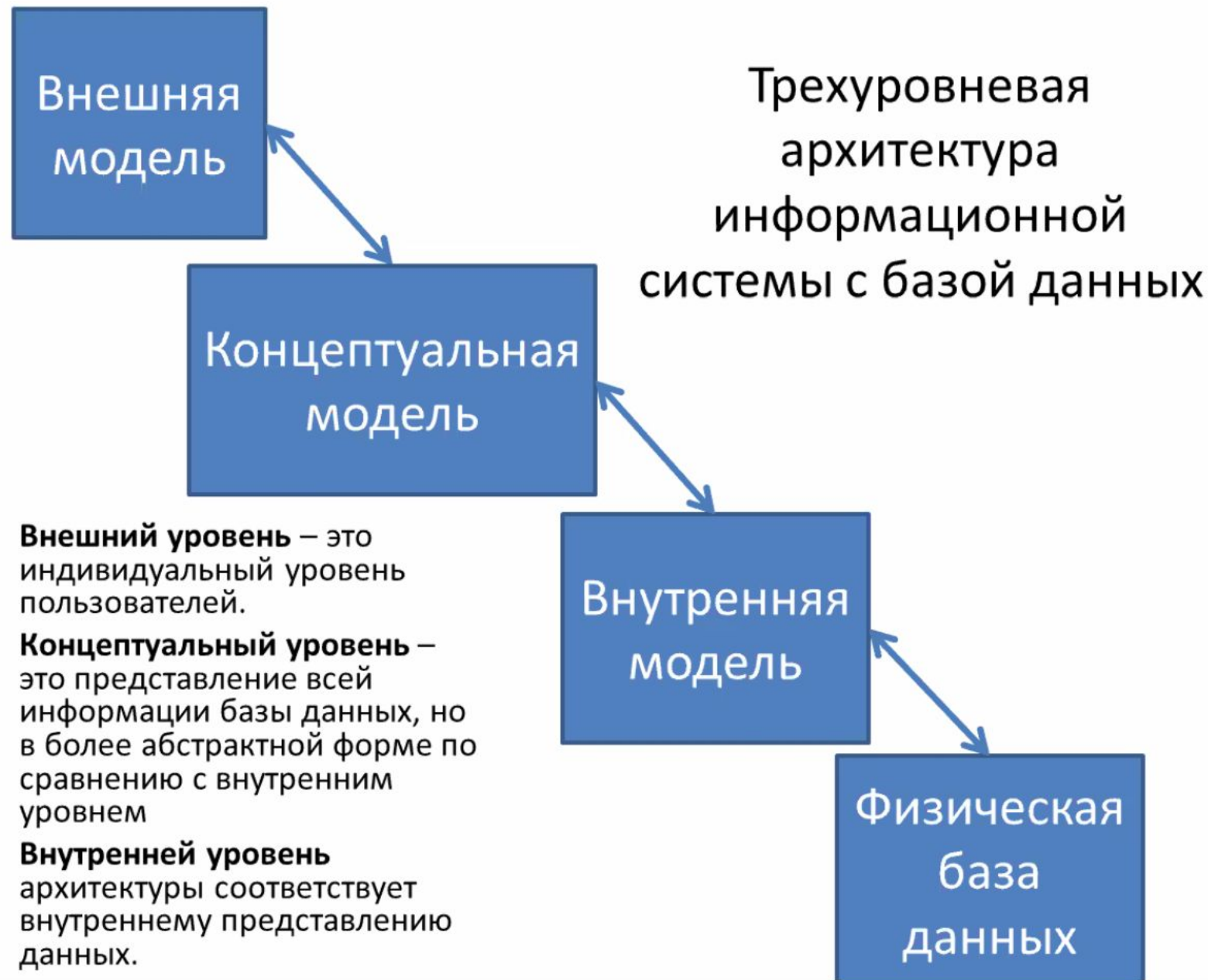


Запись

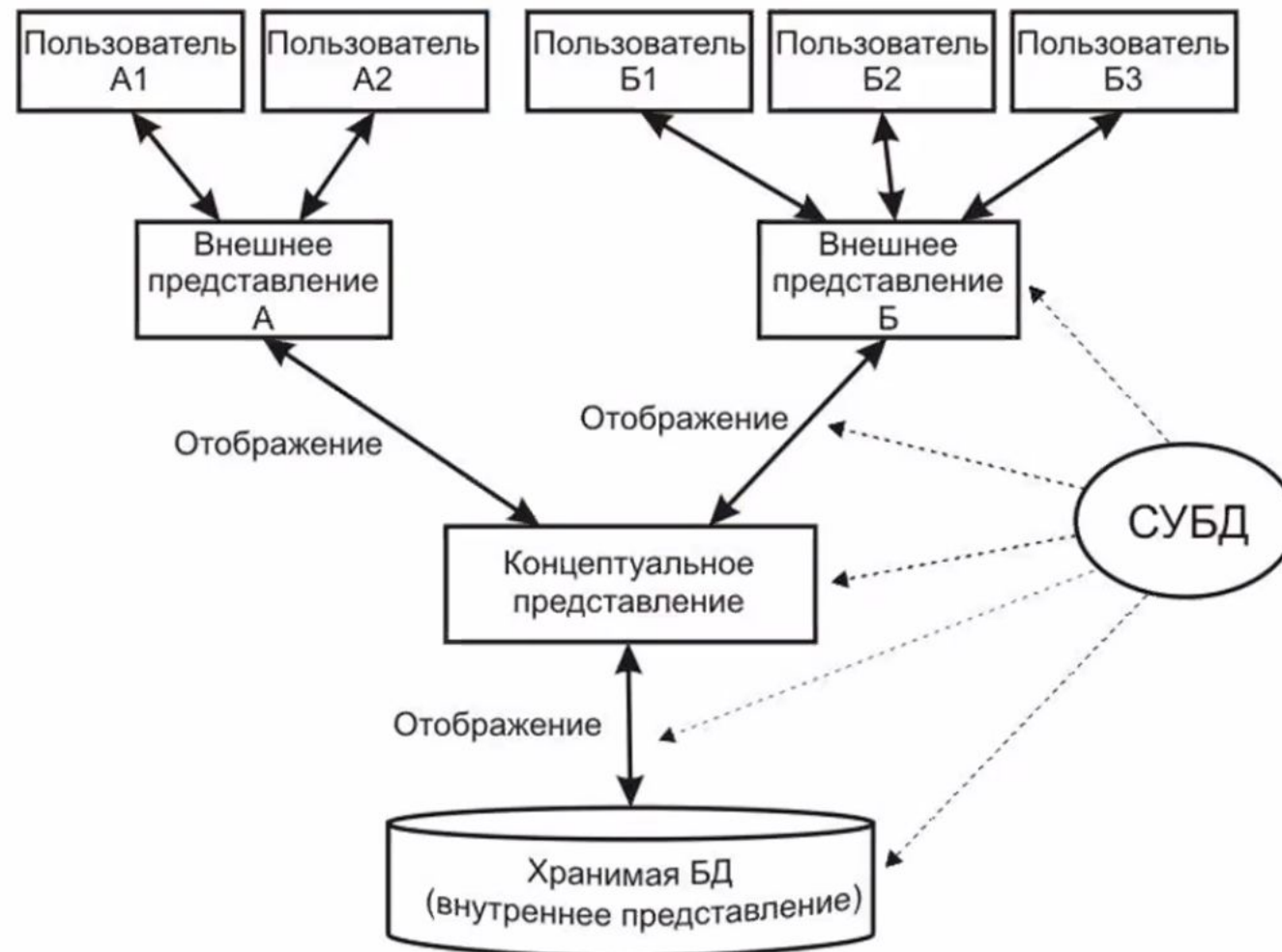
Выйти

Следует учитывать следующие обстоятельства.

- Во-первых, каждому отдельному пользователю в большинстве случаев реально требуется доступ к небольшой, вполне определенной части данных, хранимых в базе данных.
- Во-вторых, необходимое пользователю логическое представление требуемой ему части данных может отличаться от логического представления, принятого в модели данных (единственное общее представление не может одинаково успешно удовлетворить требованиям разных пользователей).
- В-третьих, необходимо обеспечивать дифференцированное управление уровнями доступа пользователей к подмножествам данных, защиту данных от несанкционированного доступа к ним.



Трехуровневая архитектура информационной системы с базой данных



Даталогический аспект

- Для даталогического аспекта характерно рассмотрение конкретной формы представления данных в памяти информационной компьютерной системы.
- При этом выбираются модели и методы представления и интерпретации данных, формулируются правила, обеспечивающие их смысловую интерпретацию, т. е. учитывающие семантическое (смысловое) содержание данных.

Дореляционные системы

- основанные на инвертированных списках,
- иерархические базы данных
- сетевые базы данных.

Инвертированные списки

- Структура базы данных, организованная с помощью инвертированных списков, в определенной мере похожа на реляционную - данные хранятся в виде таблиц, но, в отличие от реляционной БД, хранимые таблицы данных и пути доступа к ним видны и доступны пользователю.
- При этом строки (записи) таблиц упорядочены системой в некоторой физической последовательности.
- Для каждой таблицы можно определить произвольное число ключей поиска, для которых строятся индексы. Эти индексы автоматически поддерживаются системой, но явно видны пользователям.

Код_студ	Имя_студ	Факультет	Курс
0043	Иванов	Физический	1
2004	Петрова	Химический	2
5162	Сидоров	Физический	2
0007	Орлов	Химический	4
0634	Смирнов	Физический	3
0228	Попова	Исторический	4
1735	Кузнецов	Физический	1

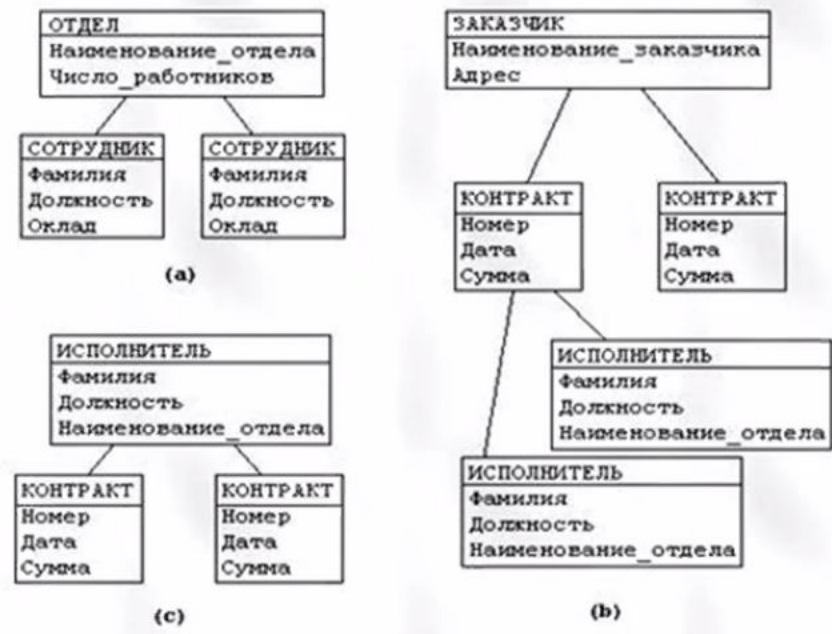


Манипулирование данными

- Типичный набор операторов над адресуемыми записями:
 - LOCATE FIRST - найти первую запись таблицы T в физическом порядке; возвращает адрес записи;
 - LOCATE FIRST WITH SEARCH KEY EQUAL - найти первую запись таблицы T с заданным значением ключа поиска K; возвращает адрес записи;
 - LOCATE NEXT - найти первую запись, следующую за записью с заданным адресом в заданном пути доступа; возвращает адрес записи;
 - LOCATE NEXT WITH SEARCH KEY EQUAL - найти следующую запись таблицы T с заданным значением K;
 - RETRIVE - выбрать запись с указанным адресом;
 - UPDATE - обновить запись с указанным адресом;
 - DELETE - удалить запись с указанным адресом;
 - STORE - включить запись в указанную таблицу; операция генерирует адрес записи.
- Ограничения целостности.
 - Общие правила определения целостности БД отсутствуют.
 - В некоторых системах поддерживаются ограничения уникальности значений некоторых полей, но в основном все возлагается на прикладную программу.
- Представляет интерес сравнение свойств таблиц в СУБД на инвертированных списках со свойствами таблиц-отношений в реляционной модели.
- В современных реляционных СУБД доступ к данным, аналогичный используемому в системах, основанных на инвертированных списках, используется при построении индексных структур и от пользователя скрыт.

Иерархические базы данных

- Структуры данных.
 - Используются упорядоченные структуры данных, организованные в виде «деревьев»
 - Древоподобная структура состоит из одной корневой записи и упорядоченного набора подчиненных записей-потомков (ветвей дерева).



- Для корневой записи обязательно формирование значения ключа.
- Каждая запись-потомок имеет ссылку (указатель) на соответствующую родительскую запись,
- образуя иерархию подчинения от самого нижнего уровня записей – «листьев дерева» – до верхнего уровня, образуемого корневой записью.

Манипулирование данными

- Операторы манипулирования данными осуществляют навигацию по иерархическим структурам, используя ссылки (указатели) для перехода от записей одного уровня к другому.
- Операции над данными:
 - ДОБАВИТЬ в базу данных новую запись;
 - ИЗМЕНИТЬ значение данных предварительно извлеченной записи (ключевые данные не должны подвергаться изменениям);
 - УДАЛИТЬ некоторую запись и все подчиненные ей записи;
 - ИЗВЛЕЧЬ (допускается задание условий выборки): извлечь корневую запись по ключевому значению, допускается также последовательный просмотр корневых записей извлечь следующую запись (следующая запись извлекается в порядке левостороннего обхода дерева).
- Ограничения целостности.
 - Из ограничений целостности автоматически поддерживается только целостность ссылок между записями-предками и записями-потомками
 - Основное правило – никакой «потомок» не может существовать без своего «родителя».



Включить звук



Включить видео



Участники 18



Чат



Демонстрация экрана

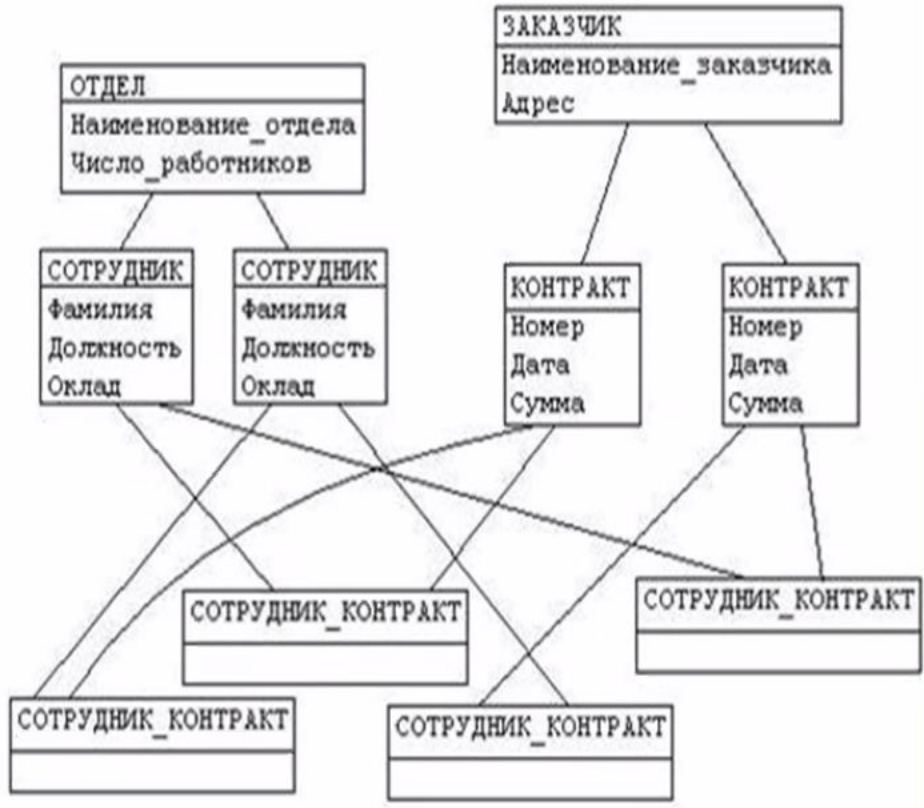


Запись

Выйти

Сетевые базы данных

- Структуры данных.
 - Термины потомок и предок в отношении записей данных в иерархических и сетевых системах отражают тот факт, что в этих системах записи данных связываются между собой с помощью ссылок на адрес, а не по значению данных.
 - К основным понятиям сетевых БД относят:
 - уровень,
 - элемент (узел),
 - связь.



Манипулирование данными

- Операторы манипулирования данными осуществляют низкоуровневую навигацию по сетевой структуре, используя ссылки для перехода от одной записи к другой.
- В сетевой структуре каждый элемент может быть связан с любым другим элементом, поскольку логика процедуры выборки данных зависит от физической организации этих данных, то эта модель не является полностью независимой от приложения.
- Операции над данными:
 - ДОБАВИТЬ - внести запись в БД;
 - ВКЛЮЧИТЬ В ГРУППОВОЕ ОТНОШЕНИЕ - связать существующую подчиненную запись с записью-владельцем;
 - ПЕРЕКЛЮЧИТЬ - связать существующую подчиненную запись с другой записью-владельцем в том же групповом отношении;
 - ОБНОВИТЬ - изменить значение элементов предварительно извлеченной записи;
 - ИЗВЛЕЧЬ - извлечь записи последовательно по значению ключа;
 - УДАЛИТЬ - убрать из БД запись.
- Ограничения целостности
 - в общем виде не поддерживаются.
 - обеспечивается только поддержание целостности по ссылкам (владелец отношения – член отношения).

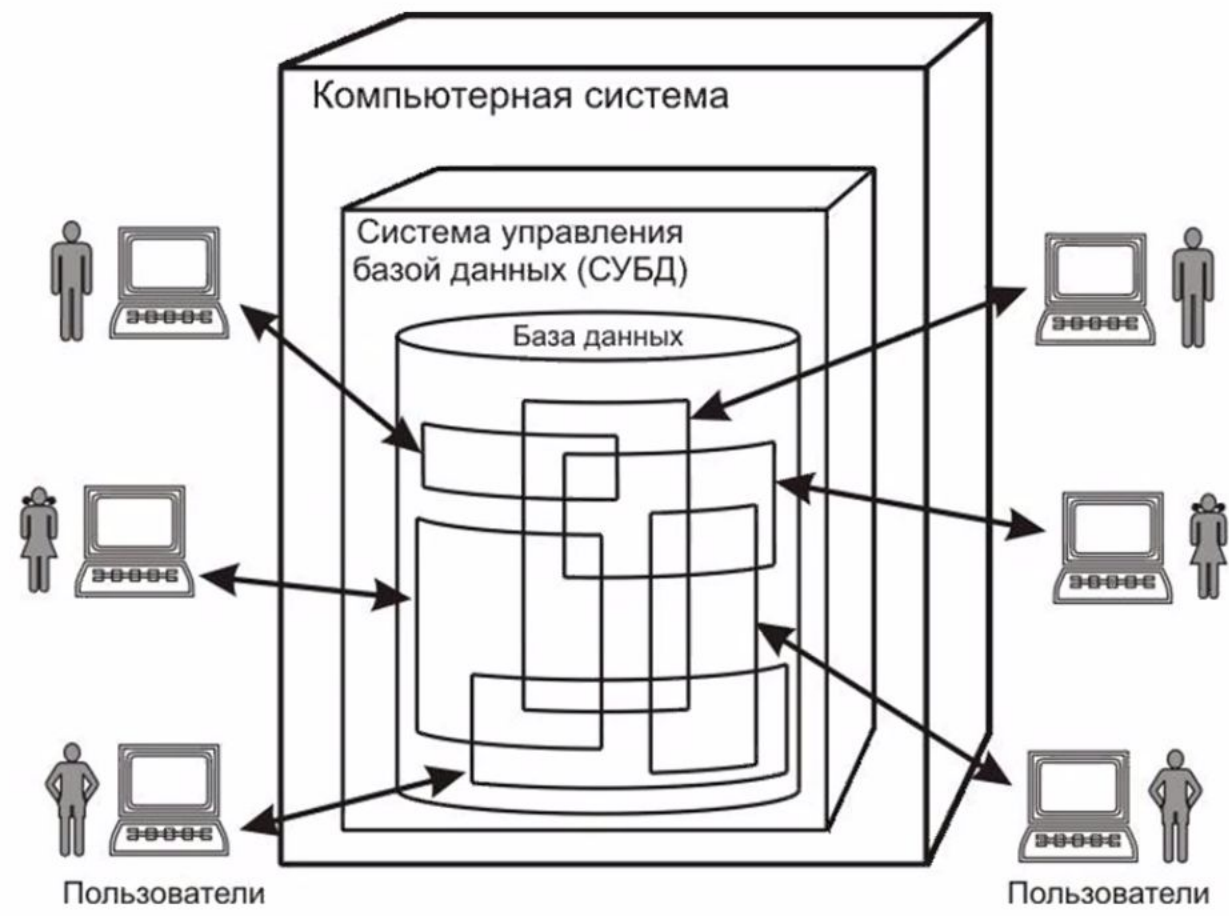
Выводы

- **Общим для всех ранних дореляционных систем является то, что в их основе отсутствуют какие-либо абстрактные модели данных, подкрепленные соответствующим математическим аппаратом.**
 - Понятие модели данных применительно к системам с базами данных стало использоваться именно с появлением реляционного подхода, который во многом родился в результате анализа особенностей, достоинств и в большей степени недостатков существовавших систем управления данными.
- **В ранних системах доступ к данным в базе данных осуществлялся непосредственно на уровне записей.**
 - Навигация по базе данных, поиск, выборка и запись данных в этих системах осуществлялись пользователем с использованием обычных процедурных языков программирования, расширенных функциями работы с СУБД, требуя от пользователя при работе с данными понимания низкоуровневых особенностей хранения данных и связей между ними.
 - Возможность интерактивного доступа к данным реализовывалась ограничено, только путем создания соответствующих прикладных программных средств.
 - Ранние системы имели достаточно слабые средства поддержания целостности данных.
- **Низкоуровневые средства навигации и манипулирования данными позволяли, тем не менее, в ранних системах обеспечивать высокую эффективность реализации этих функций прикладными программами, экономное использование памяти вычислительной системы.**
 - Это можно отнести к достоинствам таких систем, особенно учитывая уровень вычислительных систем того времени (1960-70-е гг.).
 - Одним из следствий низкого уровня языковых средств, используемых в этих моделях для работы с данными, являлась сложность создания прикладных программных систем, необходимость знания при этом особенностей физического представления и хранения данных.
 - Это приводило к большей сложности или даже невозможности выполнения комплекса требований к информационным системам с базами данных.

Основные понятия систем с базами данных

- Информационные компьютерные системы с базами данных - системы информационных, математических, программных, языковых, организационных и технических средств, предназначенных для централизованного накопления и коллективного, многоаспектного использования данных для получения необходимой информации.
- Системы с базами данных включают в себя комплекс методов и средств для поддержания динамической информационной модели предметной области с целью обеспечения информационных запросов пользователей.
- Сам массив данных, хранимый в информационной компьютерной системе, называется базой данных.

Схема информационной системы с базой данных



Пользователи информационной системы

- Главным назначением информационных систем с базами данных является обслуживание информационных запросов пользователей.
- Для этих систем, по сравнению с системами, решающими расчетно-вычислительные задачи, характерно существенное расширение круга их пользователей.

По степени постоянства общения с системой

- постоянные это пользователи или эпизодические, зависит возможность учета типов информационных запросов, которые могут от них последовать.



Включить звук



Включить видео



Участники 17



Чат



Демонстрация экрана



Запись

Выйти

По предоставляемым им возможностям доступа к данным

- информационная система с базами данных должна иметь средства для управления уровнями доступа пользователей к данным, для обеспечения защиты от несанкционированного доступа к данным.
 - например, доступ только на чтение данных, доступ с возможностью их изменения данных, запрет доступа.



по форме представления ими запросов к базе данных

- Выделяют две категории:
 - пользователи-люди и пользователи-задачи.
 - Пользователи-задачи – это соответствующие программные средства, которые обращаются к базе данных с регламентированными, то есть в большой степени заранее определенными, запросами, формируемыми работающей с базой данных программой.

Внешние пользователи

- прикладные программисты
- пользователи – не программисты
 - представляют самый многочисленный круг пользователей, для которых, в отличие от пользователей-программистов, характерным является существенно более низкий уровень специальной профессиональной подготовки в области информационных технологий, программирования и систем с базами данных.

Требования к информационным системам с базами данных

1. Система должна обеспечивать возможность эффективного хранения и модификации больших объемов многоаспектной, разнородной и сложно организованной информации.
2. Система должна обеспечивать заданный уровень достоверности хранимой информации, ее непротиворечивость (обеспечивать целостность базы данных).
3. Система должна удовлетворять выявленным и вновь возникающим запросам внешних пользователей на получение информации.
4. Система должна обеспечивать возможность поиска и выборки информации по произвольной группе признаков.
5. При обработке информационных запросов пользователей система должна удовлетворять заданным требованиям по производительности.
6. Информация должна соответствовать запросам пользователей не только по содержанию, но и по форме ее представления, и система должна обеспечивать возможность выдачи информации в соответствующей форме.
7. Должны быть обеспечены простота и удобство обращения внешних пользователей за информацией.
8. Должна быть обеспечена возможность одновременного доступа к общему информационному хранилищу большого числа внешних пользователей различных категорий.
9. Система должна обеспечивать доступ к хранимым данным только тем пользователям, которые обладают соответствующими полномочиями, должна обладать средствами управления уровнем доступа к информации различных категорий пользователей.
10. Наконец, система должна иметь возможность ее реорганизации и расширения при изменении границ предметной области.

Преимущества и проблемы

- Устраняется дублирование и избыточность хранимых данных, которые чрезвычайно затрудняют обеспечение непротиворечивости данных, их целостности.
- Появляется возможность многоаспектного использования общих данных различными задачами и пользователями, обеспечивает возможность реализации новых программных приложений для работы с уже имеющимися данными.
- В общем случае массивы данных, используемые различными пользователями, могут пересекаться
 - проблема обеспечения такого режима, чтобы одновременно работающие с данными пользователи не создавали друг другу каких-либо помех как связанных с содержанием данных, так и с временем доступа к данным.

Централизованное управление общей базой данных

- дает возможность обеспечения комплексной оптимизации и сбалансированности предъявляемых к системе многочисленных, порой противоречивых требований.
 - Появляется возможность выбора таких структур хранения данных, которые обеспечат наилучшее решение задач и обслуживание пользователей в целом.
 - Появляется возможность стандартизации структур хранения данных, методов работы с ними.
 - Стандартизация в представлении данных упрощает эксплуатацию информационной системы, обмен данными между различными системами и базами данных, в том числе и созданными разными производителями.
 - Появляется возможность эффективного управления уровнями доступа пользователей к данным в соответствии с уровнями их компетенции, в том числе предотвращение несанкционированного доступа к данным.
 - Облегчается возможность системного контроля и обеспечения целостности данных, то есть их корректности, точности и непротиворечивости.

Важные принципы

- Важными принципами, отличающими системы с базами данных от других программных систем, являются:
 - интеграция (объединение) данных,
 - централизованное управление данными
 - обеспечение взаимной независимости данных и использующих их прикладных программ.
- В информационных системах с базами данных сами данные становятся одним из важнейших ресурсов системы.
- И с этой точки зрения информационные системы с базами данных можно рассматривать как средство
 - централизованного управления ресурсами в интересах задач, решаемых всеми пользователями системы

Основные компоненты ИС с базами данных

- Аппаратные средства компьютерной системы.
 - Процессор (процессоры) и основная оперативная память.
 - Системы долговременного хранения информации. Обычно это устройства с магнитными дисками с соответствующими контроллерами (устройствами управления) и каналами ввода-вывода.

Определения

- **База данных (БД)** – это совокупность данных, организованная по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования данными.
- **Система управления базой данных (СУБД)** – это специальный комплекс программ и языковых средств, предназначенный для создания, ведения и использования баз данных, посредством которого реализуется централизованное управление базой данных и обеспечивается доступ к данным.
- В СУБД имеются трансляторы и/или интерпретаторы с языка описания данных и с языка манипулирования данными.
- **Язык описания данных** – это язык высокого уровня, предназначенный для задания схемы базы данных. С его помощью описываются типы данных, подлежащих хранению в базе или выборке из нее, структура данных и их связи между собой. Это язык декларативного (описательного) типа, а не процедурного.
- **Язык манипулирования данными** (или язык запросов к базе данных) представлен набором команд, реализующих операции манипулирования данными. Это операции выборки данных и так называемые операции запоминания, а именно, операции ввода данных, их изменения и удаления.
 - Фактическим стандартом языка для описания данных и манипулирования ими в современных СУБД является язык SQL (Structured Query Language – язык структурированных запросов).
 - В СУБД, как правило, также имеются специальные средства для разграничения доступа к данным пользователей БД, обеспечения защиты данных от аппаратных и программных сбоев, средства контроля достоверности и непротиворечивости данных, средства автоматического мониторинга функционирования системы и др.
- **Словарь (справочник) базы данных** представляет собой специализированную подсистему СУБД, предназначенную для централизованного хранения единообразной информации обо всех хранимых в БД данных, используемой СУБД для доступа к данным. Словарь содержит «данные о данных» (метаданные) в том числе информацию:
 - об объектах, их свойствах и отношениях между объектами для данной предметной области;
 - о данных, хранимых в БД (их наименование, смысловое описание, тип, структура, связи с другими данными);
 - о возможных значениях и форматах представления данных;
 - о кодах защиты и уровнях доступа к данным со стороны пользователей.
- **Схема базы данных** – это описание базы данных в контексте конкретной модели данных

Администратор базы данных

- (АБД) – это лицо или группа специалистов, знакомых с теорией построения информационных систем с базами данных и со спецификой предметной области данной информационной системы.
- Администратор базы данных осуществляет централизованное управление базой данных посредством конкретной СУБД.

Реляционная модель данных.

Основные понятия.

Структуры данных

Операции над данными

История.

- В основе реляционных систем лежит реляционная модель данных
- Принципы реляционной модели были заложены в 1969–1970 гг. американским ученым Е. Ф. Коддом (E. F. Codd)

Три компоненты модели данных

- структуры данных,
- операции, которые можно выполнять над данными, и
- ограничения, связанные с обеспечением целостности данных.
 - Основной структурой данных в реляционной модели являются отношения,
 - называемые – таблицы, для более простого понимания.
 - От термина отношение (по-английски relation) и произошло название модели – реляционная

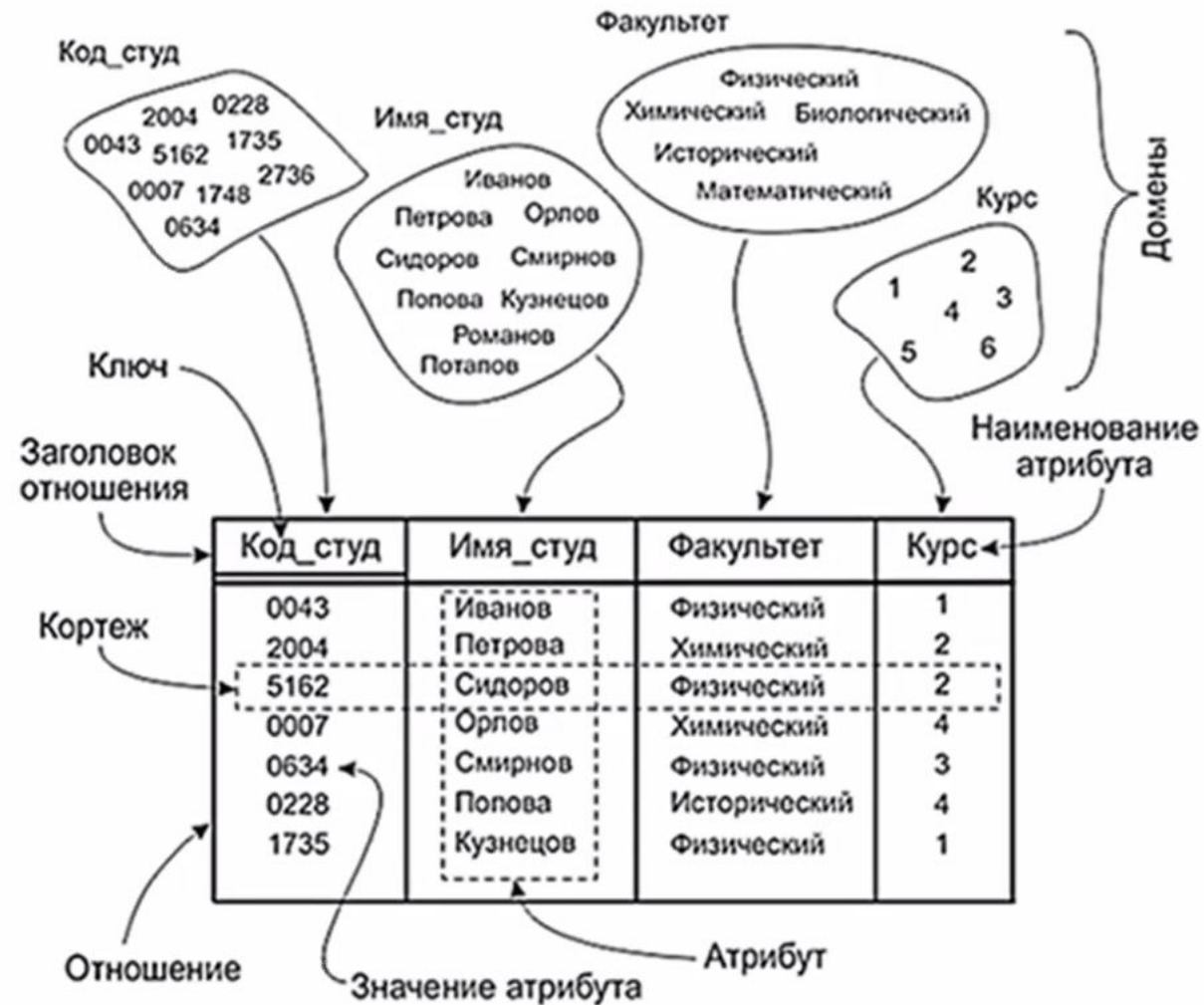
Понятия реляционной модели

- Отношение это таблица, состоящая из строк и столбцов и имеющая вверху строку, называемую заголовком отношения.
- Строки таблицы-отношения называются кортежами (tuple), а столбцы атрибутами (attribute).
- Количество кортежей в отношении называется кардинальным числом отношения, а количество атрибутов называется степенью отношения.
- Каждый атрибут в отношении имеет наименование, которое указывается в заголовочной части отношения.

Понятия реляционной модели

- Ключ отношения – это атрибут или набор атрибутов отношения такие, что в любой момент времени в отношении не существует строк, для которых значение или комбинация значений ключевых атрибутов являются одинаковыми.
 - Ключ, таким образом, является уникальным идентификатором кортежей отношения).
- Домен отношения – это множество значений, из которого могут браться значения конкретного атрибута.
 - Значения атрибута, которые отсутствуют в множестве, задаваемом доменом, являются недопустимыми.

Отношение и его компоненты



Отношение

- Отношение-переменная – это именованный объект, значение которого может изменяться со временем.
- Значение отношения – значение отношение-переменной в любой момент времени.