

Распределенные базы данных

Введение

Список литературы

Основная литература:

- Коннолли Т., Бегг К. Базы данных: проектирование, реализация, сопровождение. Теория и практика, 3-е изд. : Пер. с англ. : Уч. пос. – М.: Изд. дом "Вильямс", 2003. – 1440 с.
 - глава 22. Концепции и разработка распределённых СУБД
 - глава 23. Распределённые СУБД: дополнительные концепции.
 - глава 19. "Управление транзакциями".
 - глава 20. "Обработка запросов", приложение В.
- Кузнецов С.Д. Основы современных баз данных. – Учебник для вузов. (глава 8) = <http://citforum.ru/database/osbd/contents.shtml>
- М. Тамер Оззу, Патрик Валдуриз. Распределенные и параллельные системы баз данных. / Журнал СУБД, №4, 1996. // Новая редакция: Сергей Кузнецов, 2009 г. – http://citforum.ru/database/classics/distr_and_parallel_sdb/
- Энсор Д., Стивенсон Й. Oracle. Проектирование баз данных: Пер. с англ. – Киев: Изд. группа ВНУ, 1999. – 560 с.
- Глеб Ладыженский. Распределенные информационные системы и базы данных. = <http://citforum.ru/database/kbd96/45.shtml>

Дополнительные материалы и материалы для повторения:

- Карпова И.П. Базы данных. – Учебное пособие по курсу "Базы данных". – РИО, МГИЭМ, 2009. – 118 с. – <http://rema44.ru/resurs/students/karpova>
- Грабер Мартин. SQL. (любое издание).
- Изучение основ языка SQL: Метод. указания к лабораторным работам 1-4 по курсу "Базы данных" / МИЭМ; Сост.: И. П. Карпова. М., 2012. – 32 с.

Интернет-источники:

- citforum.ru/database
- rema44.ru -> Студентам -> Карпова И.П. - <http://rema44.ru/resurs/students/karpova/>

Базы данных: термины

Информация – любые сведения о каком-либо событии, объекте или процессе, являющиеся объектом некоторых операций: восприятия, передачи, преобразования, хранения или использования.

Данные – это информация, зафиксированная в некоторой форме, пригодной для последующей обработки, передачи и хранения, например, находящаяся в памяти ЭВМ или подготовленная для ввода в ЭВМ.

Обработка данных – это совокупность задач, осуществляющих преобразование массивов данных. Обработка данных включает в себя ввод данных в ЭВМ, отбор данных по каким-либо критериям, преобразование структуры данных, перемещение данных на внешней памяти ЭВМ, вывод данных, являющихся результатом решения задач, в табличном или в каком-либо ином удобном для пользователя виде.

Система обработки данных (СОД) – это набор аппаратных и программных средств, осуществляющих выполнение задач по управлению данными.

Управление данными – совокупность функций обеспечения требуемого представления данных, их накопления и хранения, обновления, удаления, поиска по заданному критерию и выдачи данных. [ГОСТ 20886-85]

Базы данных: термины

База данных (БД) – это совокупность взаимосвязанных структурированных данных, относящихся к определенной предметной области и организованных так, чтобы обеспечить независимость данных от программ обработки.

Фактически база данных – это модель предметной области (ПрО).

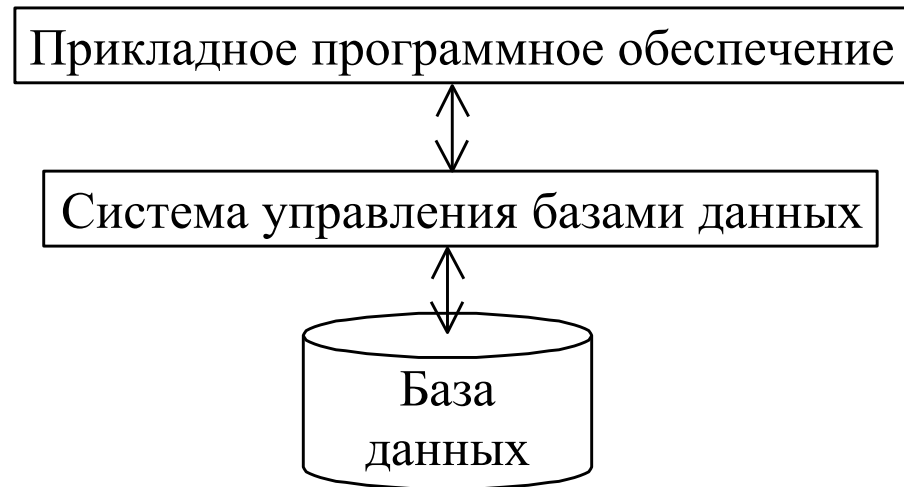
Предметная область (ПрО) – часть реального мира, подлежащая изучению с целью организации управления и, в конечном итоге, автоматизации.

Ведение базы данных – деятельность по обновлению, восстановлению и изменению структуры базы данных с целью обеспечения её целостности, сохранности и эффективности использования [ГОСТ 20886-85].

Система управления базами данных (СУБД) – это совокупность программ и языковых средств, предназначенных для управления данными в базе данных, ведения базы данных и обеспечения взаимодействия её с прикладными программами [ГОСТ 20886-85].

Автоматизированная информационная система (АИС) представляет собой совокупность данных, экономико-математических методов и моделей, технических, программных средств и специалистов, предназначенную для обработки информации и принятия управленческих решений.

Компоненты системы баз данных



Основным принципом организации базы данных является совместное хранение данных и их описаний. Это отличает базу данных от любого другого набора данных, хранящихся в ЭВМ.

Описание базы данных хранится в так называемом **словаре-справочнике данных (ССД) или каталоге данных**. Хранение данных вместе с их описанием и позволяет обеспечивать независимость данных от программ, т.е. интерпретация данных определяется не программой, а описанием данных.

Уровни представления данных архитектуры ANSI/SPARC



Концептуальный уровень: поддерживает единый взгляд на базу данных, общий для всех её приложений и независимый от них и от среды хранения.


Внутренний уровень: схема хранения данных в среде хранения.

Внешний уровень (внешние схемы): предназначены для групп пользователей.

Схема базы данных – это описание базы данных в терминах конкретной модели данных.

Физическая и логическая независимость данных.

Модели данных

- Иерархическая модель данных (ИМД); система IMS (Information Management System, IBM).
 - Сетевая модель данных (СМД); система IDS (Integrated Data Store, GE).
 - Реляционная модель данных (РМД); Oracle, DB2, MySQL, MS SQL Server, MS Access, Firebird, ЛИНТЕР, ...
 - Объектно-реляционная модель данных (ОРМД).
Стандарт SQL-3 (SQL-2003). Oracle (с версии 8.0), DB2, Informix, PostgreSQL, SQL Server 2008 и др.)
 - Объектно-ориентированная модель данных (ООМД). O2, GemStone, Iris и др.
Стандарт ODMG 3.0 (Object Database Management Group).
 - Многомерные базы данных.
 - Поточковые базы данных.
 - ...
- } I поколение
- II поколение
- III поколение
- 

Реляционная модель данных (РМД)

- Домен, D – множество значений, которые может принимать элемент данных.
- Декартово произведение доменов – множество всех возможных комбинаций значений доменов:
 $D_1 \times D_2 \times \dots \times D_n = \{(d_{1i}, d_{2i}, \dots, d_{ni})\}$, где $d_{ki} \in D_k$
- Пример: $D_1 = (1, 2)$, $D_2 = (a, b, c)$.
 $D_1 \times D_2 = \{(1,a), (1,b), (1,c), (2,a), (2,b), (2,c)\}$
- Отношение – подмножество декартова произведения доменов. Пример:

<i>Табельный номер</i>	<i>ФИО сотрудника</i>	<i>Должность</i>	<i>Оклад</i>	<i>Год рождения</i>	<i>Отдел</i>
023	Волкова Елена Павловна	секретарь	26000	1985	2
113	Белов Сергей Юрьевич	инженер	39800	1980	1
101	Рогов Сергей Михайлович	директор	62000	1972	2
056	Панина Анна Алексеевна	инженер-программист	41800	1978	1
...
098	Фролов Юрий Вадимович	начальник отдела	49200	1971	9

РМД. Термины. Свойства отношения

первичный ключ

Отношение, таблица

столбец, поле, атрибут

<i>Табельный номер</i>	<i>ФИО сотрудника</i>	<i>Должность</i>	<i>Оклад</i>	<i>Год рождения</i>	описание (схема отношения)
023	Волкова Елена Павловна	секретарь	26000	1985	
113	Белов Сергей Юрьевич	инженер	39800	1980	строка, запись, кортеж

Отношение обладает двумя основными свойствами:

1. В отношении не должно быть одинаковых кортежей, т.к. это множество.
2. Порядок кортежей в отношении несущественен.

РМД. Ключи отношения

- **Ключ** – атрибут (группа атрибутов), которые позволяют классифицировать кортеж (запись таблицы).
- **Потенциальный ключ (уникальный ключ)** – атрибут (группа атрибутов), которые позволяют идентифицировать кортеж (запись таблицы).
- **Первичный ключ** – обязательный уникальный ключ. Для каждой таблицы может быть определен только один первичный ключ.
- **Вторичный ключ** – любой другой ключ, кроме первичного. Может быть необязательным и неуникальным.
- **Внешний ключ** – служит для организации связей между таблицами.

РМД. Организация связей между таблицами

Связь один-ко-многим: Отделы – Сотрудники

Таблица «Сотрудники»

<i>Табельный номер</i>	<i>ФИО сотрудника</i>	<i>Отдел</i>
023	Волкова Елена Павловна	2
113	Белов Сергей Юрьевич	1
101	Рогов Сергей Михайлович	2
056	Панина Анна Алексеевна	1
...
098	Фролов Юрий Вадимович	9

Таблица «Отделы»

<i>Номер отдела</i>	<i>Название отдела</i>
1	Информационный отдел
2	Администрация
3	Отдел кадров
...	...
9	Проектный отдел

«Номер отдела» - первичный ключ в таблице «Отделы»

«Отдел» – внешний ключ в таблице «Сотрудники»

РМД. Организация связей между таблицами

Связь многие-ко-многим: Проекты – Сотрудники

Таблица «Сотрудники»

<i>ФИО</i>	<i>Номер</i>
Волкова Е.П.	023
Белов С.Ю.	113
Рогов С.М.	101
Панина А.А.	056
Фролов Ю.В.	098
...	...

Таблица «Участие»

<u><i>Участник</i></u>	<i>Роль</i>	<u><i>Проект</i></u>
113	исполнитель	23/Н
101	руководитель	18-К
056	исполнитель	18-К
101	консультант	09/Р
098	руководитель	23/Н
...

Таблица «Проекты»

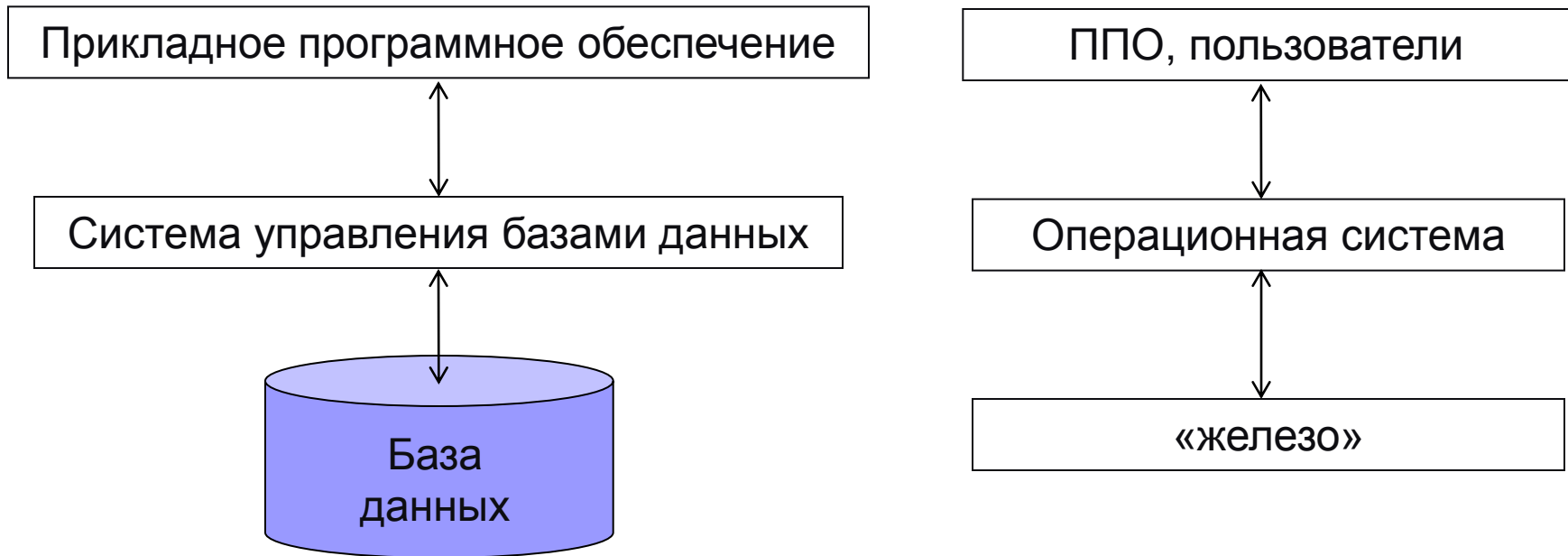
<i>Шифр</i>	<i>Название проекта</i>
23/Н	АИС "Налог"-2
18-К	ИПС "Жители"
09/Р	ГИС "Город"
...	...

В таблице «Участие»:

«Участник» – внешний ключ к таблице «Сотрудники»

«Проект» – внешний ключ к таблице «Проекты»

СУБД. Основные функции



- Обеспечение доступа ППО к базе данных
- Управление базой данных

Классификация СУБД

По степени универсальности СУБД делят на два класса:

- СУБД общего назначения (СУБД ОН);
- специализированные СУБД (СпСУБД).

По модели данных различают **иерархические, сетевые, реляционные, объектно-реляционные и объектно-ориентированные СУБД.**

По методам организации хранения и обработки данных СУБД делят на

- **Централизованные**
- **Распределённые.**

СУБД. Состав и функции

1. Хранение, извлечение и обновление данных.
2. Каталог (словарь-справочник данных, ССД), доступный конечным пользователям.

Обычно в системном каталоге хранятся следующие сведения:

- имена, типы и размеры элементов данных;
- имена связей;
- накладываемые на данные ограничения целостности;
- имена пользователей, которым предоставлено право доступа к данным;
- внешняя, концептуальная и внутренняя схемы и отображения между ними;
- статистические данные, например частота транзакций и счетчики обращений к объектам базы данных.

СУБД. Состав и функции

3. Поддержка транзакций.
4. Служба управления параллельной работой.
5. Службы восстановления.
6. Службы контроля доступа к данным.
7. Службы поддержки целостности данных.
8. Службы поддержки независимости от данных.
9. Вспомогательные службы.

Основные объекты базы данных

- **База данных (DATABASE)** – объект, который находится на самом верхнем уровне физической организации базы данных. База данных состоит из словаря-справочника данных, собственно данных и различных вспомогательных объектов, состав которых зависит от СУБД. База данных может быть создана автоматически при инсталляции СУБД или вручную с помощью команды `CREATE DATABASE`.
- **Пользователь (USER)** – объект, обладающий возможностью создавать и использовать другие объекты БД, а также запрашивать выполнение функций сервера.

К числу таких функций относятся организация сессии, изменение состояния сервера и базы данных, создание других объектов БД, запросы на выполнение операторов SQL и проч. В некоторых СУБД (Oracle, ЛИНТЕР) имя пользователя совпадает с именем схемы. Создается командой `CREATE USER`. Каждый объект БД принадлежит тому пользователю, который его создал, и находится в его схеме. Полное имя объекта БД (кроме базы данных и пользователей) состоит из имени схемы, в которой он создан, и собственно имени объекта, например:

`scott.emp`

Основные объекты базы данных

- **Таблица (TABLE)** является базовой структурой реляционной модели. Как известно, вся информация в базе данных хранится в таблицах. Таблицы состоят из множества поименованных столбцов или атрибутов. Множество значений столбца определено с помощью ограничений целостности (множества допустимых значений). Таблица может быть пустой или состоять из одной или более строк значений атрибутов. Создается командой `CREATE TABLE`.
- **Представление (VIEW)** – это поименованная, динамически поддерживаемая сервером выборка данных из одной или нескольких таблиц. В основе представления лежит оператор `SELECT`, который определяет базовый запрос представления. Данные в представлении не хранятся: сервер формирует представление каждый раз при обращении к нему (это называется *материализация представления*). Используя представления, администратор безопасности может ограничить доступную пользователям часть базы данных только теми данными, которые реально необходимы им для выполнения работы. Создается командой `CREATE VIEW`.

Основные объекты базы данных

- **Индекс (INDEX)** – это объект базы данных, создаваемый для повышения производительности выборки данных. Индекс создается для столбца (столбцов) таблицы и обеспечивает более быстрый доступ к данным этой таблицы за счет упорядочения данных столбца (столбцов) по значению. Создается командой `CREATE INDEX`.

Специфичными для распределенных систем являются такие объекты Oracle как *снимок* и *связь базы данных*.

- **Снимок (SNAPSHOT)** – локальная копия таблицы удаленной базы данных, которая используется либо для тиражирования (копирования) всей или части таблицы, либо для тиражирования результата запроса данных из нескольких таблиц. Создается командой `CREATE SNAPSHOT`.
- **Связь базы данных (DATABASE LINK)** – это объект базы данных, который позволяет обратиться к объектам удаленной базы данных. Имя связи базы данных можно рассматривать как ссылку на параметры механизма доступа к удаленной базе данных (имя узла, протокол и т. п.). Создается командой `CREATE DATABASE LINK`.

Основные объекты базы данных

Для программирования алгоритмов обработки данных, поддержки сложных правил целостности данных использует процедурные объекты (далее состав приведен для СУБД Oracle):

- **Процедура (PROCEDURE)** – это подпрограмма на языке PL/SQL, предназначенная для решения конкретной задачи обработки данных. Создается командой `CREATE PROCEDURE`.
- **Функция (FUNCTION)** – это подпрограмма на языке PL/SQL, предназначенная для решения конкретной задачи и возвращающая конкретное значение. (`CREATE FUNCTION`).
- **Пакет (PACKAGE)** – это поименованный, структурированный набор переменных, процедур и функций, связанных единым функциональным замыслом. Пакет состоит из спецификации и тела пакета. Спецификация содержит описания внешних переменных, констант, типов и подпрограмм, а тело пакета – реализацию подпрограмм и описание внутренних переменных, констант и типов, которые доступны только внутри пакета. Спецификация пакета создается командой `CREATE PACKAGE`, а тело пакета – `CREATE PACKAGE BODY`.
- **Триггер (TRIGGER)** – это хранимая процедура, которая автоматически запускается тогда, когда происходит связанное с триггером событие. Создается командой `CREATE TRIGGER`.