

Лекция 1

Основные понятия теории баз данных


БАЗЫ ДАННЫХ И СУБД

С самого начала развития вычислительной техники (далее – ВТ) образовались **два основных направления** ее использования:

- ? применение **ВТ для сложных численных расчетов;**
- ? применение **ВТ для хранения и обработки информации.**

Первоначально ВТ использовалась только для сложных численных расчетов. Именно потребностью обеспечить большие объемы расчетов, требовавшихся в новых отраслях науки и техники (ядерной энергетике, ракетостроении и др.), и было вызвано появление ЭВМ.

Это направление дало толчок развитию численных методов и языков программирования, ориентированных на удобную запись численных алгоритмов.




БАЗЫ ДАННЫХ И СУБД

Второе направление возникло, когда в компьютерах появилось достаточно оперативной и внешней памяти (магнитные ленты, магнитные барабаны и, главное, магнитные диски). Оно дало развитие информационным системам (далее – ИС).

Классическими примерами ИС являются банковские системы, системы резервирования билетов, библиотечные системы и т.п.

Появление магнитных дисков, способных хранить большие объемы информации, потребовало создания систем управления данными во внешней памяти. Так появились **системы управления файлами** (далее – **СУФ**), или **файловые системы**.

С точки зрения прикладной программы **файл** – это именованная область внешней памяти, в которую можно записывать и из которой можно считывать данные.



БАЗЫ ДАННЫХ И СУБД

Почему файловых систем недостаточно для эффективной обработки информации?

В файловой системе хранятся следующие виды файлов:

- ? **текстовые** (документы, программы); обрабатываются текстовыми редакторами; их структура очень проста – последовательность строк или байтов (символов);
- ? **объектные и исполняемые** файлы, которые также имеют простую структуру.

Таким образом, файловые системы обеспечивают хранение слабоструктурированной информации, оставляя дальнейшую структуризацию прикладным программам.

Сами файлы часто изменяются, удаляются, создаются заново.



ИНФОРМАЦИОННЫЕ СИСТЕМЫ

Одним из важнейших условий обеспечения эффективного функционирования любой организации является наличие развитой информационной системы.

Информационная система представляет собой систему, реализующую автоматизированный сбор, обработку и манипулирование данными и включающая технические средства обработки данных, программное обеспечение и обслуживающий персонал.



БАНК ДАННЫХ

Современной формой информационных систем являются **банки данных**, которые включают в свой состав вычислительную систему, одну или несколько баз данных (БД), систему управления базами данных (СУБД) и набор прикладных программ (ПП).

Основными функциями банков данных являются:

- ? хранение данных и их защита;
- ? изменение (обновление, добавление и удаление) хранимых данных;
- ? поиск и отбор данных по запросам пользователей;
- ? обработка данных и вывод результатов.



БАНК ДАННЫХ

База данных обеспечивает хранение информации и представляет собой поименованную совокупность данных, организованных по определенным правилам, включающим общие принципы описания, хранения и манипулирования данными.

Система управления базами данных представляет собой пакет прикладных программ и совокупность языковых средств, предназначенных для создания, сопровождения и использования баз данных.

Прикладные программы (приложения) в составе банков данных служат для обработки данных, вычислений и формирования выходных документов по заданной форме.

Для работы с базой данных во многих случаях можно обойтись только средствами СУБД, скажем, создавая запросы и отчеты. Приложения разрабатывают главным образом в случаях, когда требуется обеспечить удобство работы с БД

▶ неквалифицированным пользователям или интерфейс СУБД не устраивает пользователя.

Понятие БАЗЫ ДАННЫХ

Неформально, **база данных (БД)** – это большое по объему хранилище данных, в которое можно добавлять и из которого можно извлекать данные.

Более строгое определение БД в «*Мартин Дж. Организация баз данных в вычислительных системах. М: Мир, 1978*».

База данных – это совокупность взаимосвязанных хранящихся вместе данных при наличии такой **минимальной избыточности**, которая допускает их эффективное использование для одного или нескольких приложений (задач);

данные запоминаются так, чтобы они были **независимы от программ**, использующих эти данные;

для добавления новых и модификации существующих данных, а также для поиска данных применяется **общий управляемый способ**.



СУБД

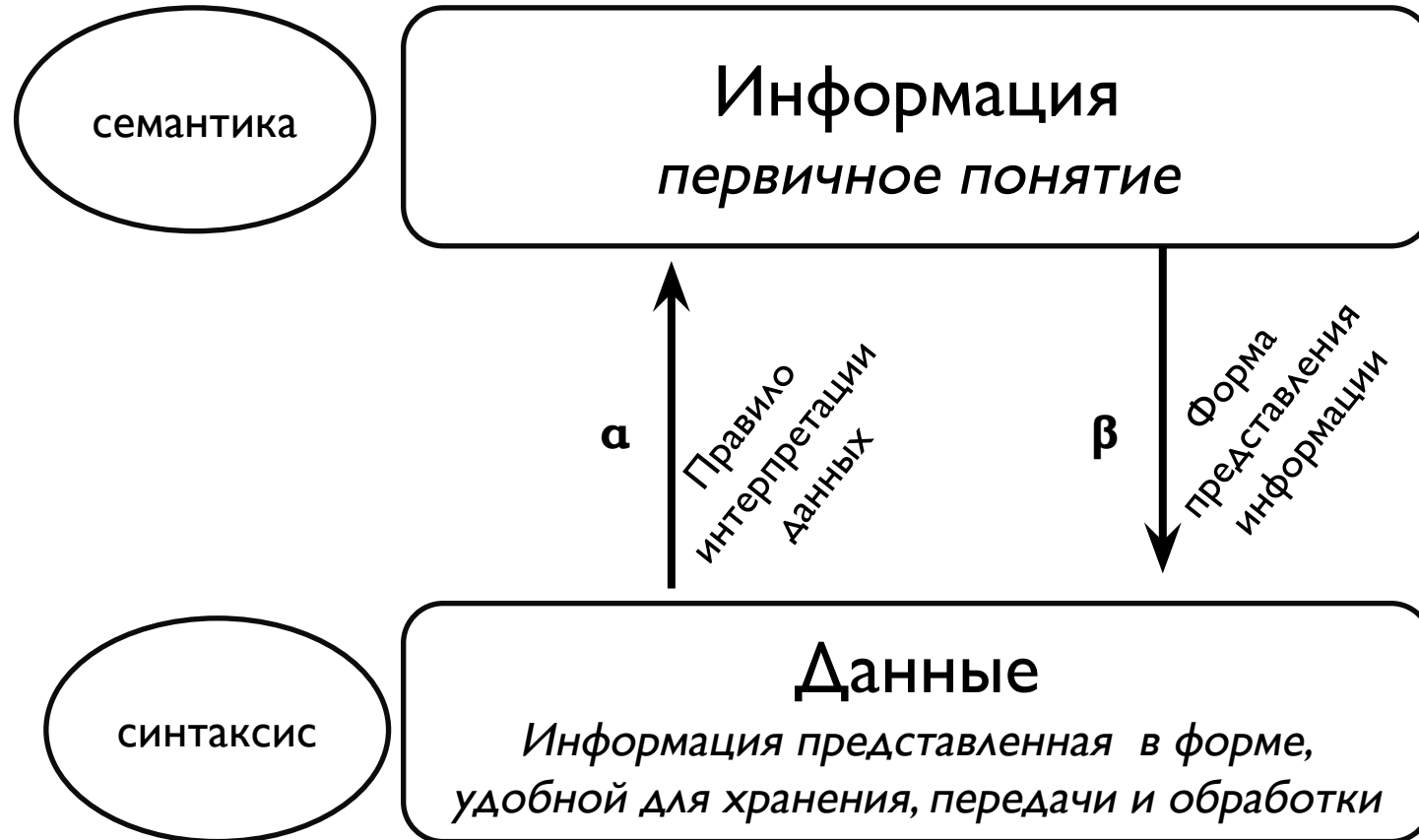
Система управления данными является **СУБД**, если она

- ? поддерживает логически согласованный набор файлов;
- ? предоставляет язык манипулирования данными;
- ? обеспечивает восстановление информации после сбоев;
- ? обеспечивает параллельную работу нескольких пользователей.

Примеры СУБД: VisualFoxPro, Paradox, Clipper, dBase, Access, Oracle, MySQL, MS SQL Server, Informix

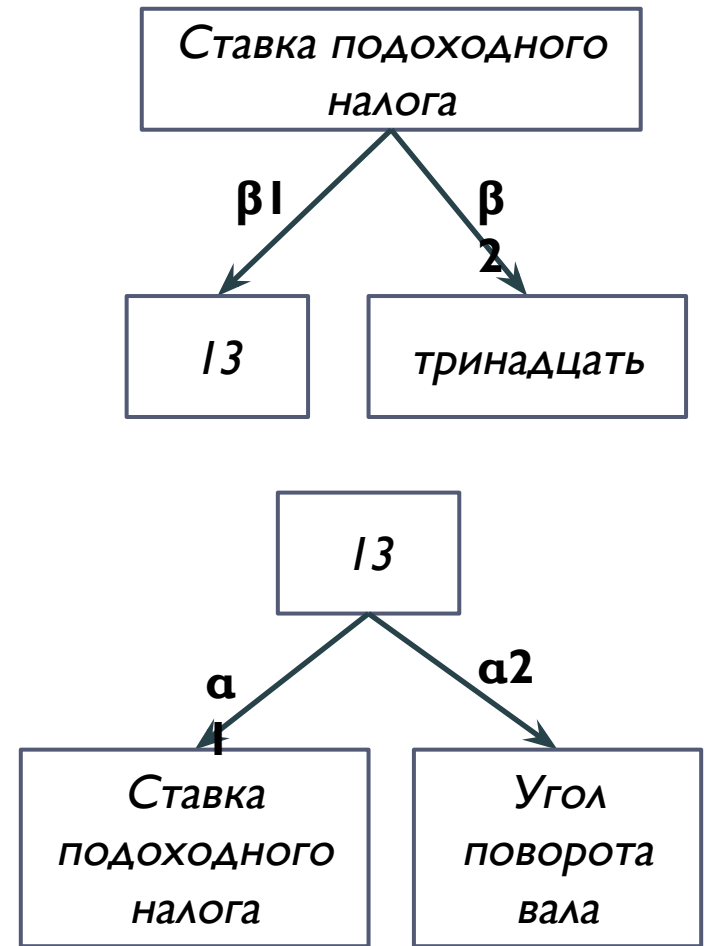
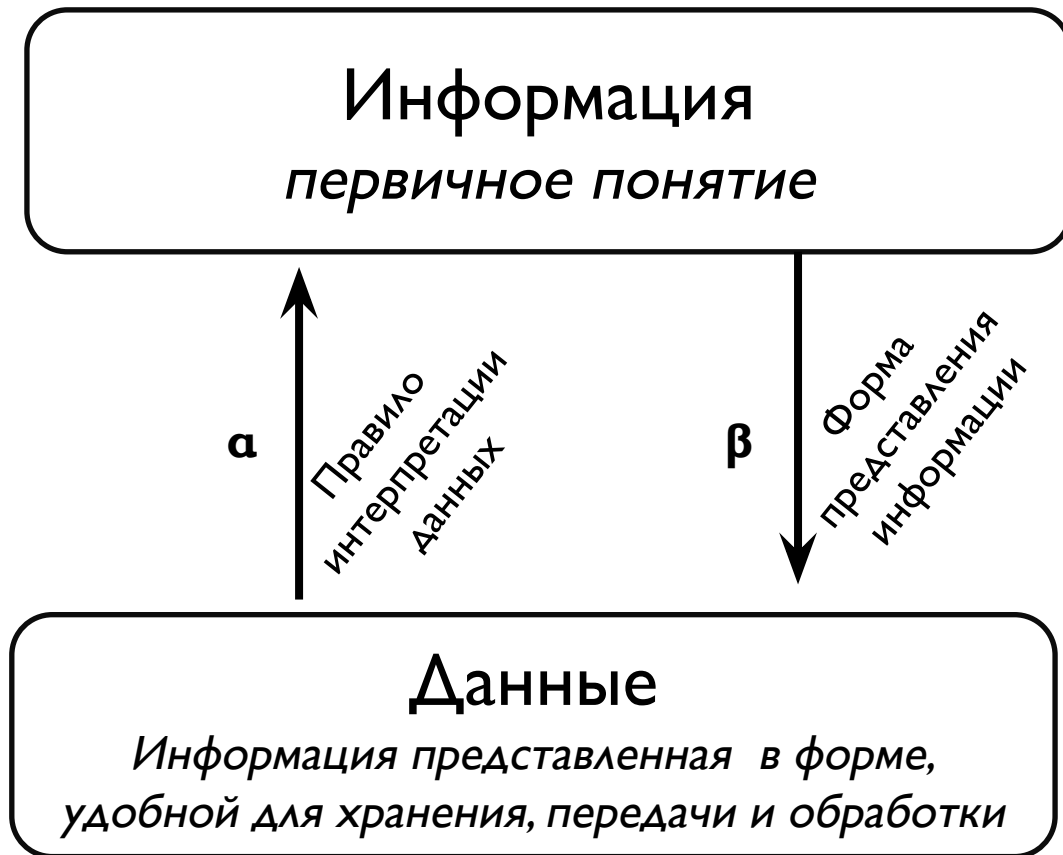


Информация и данные

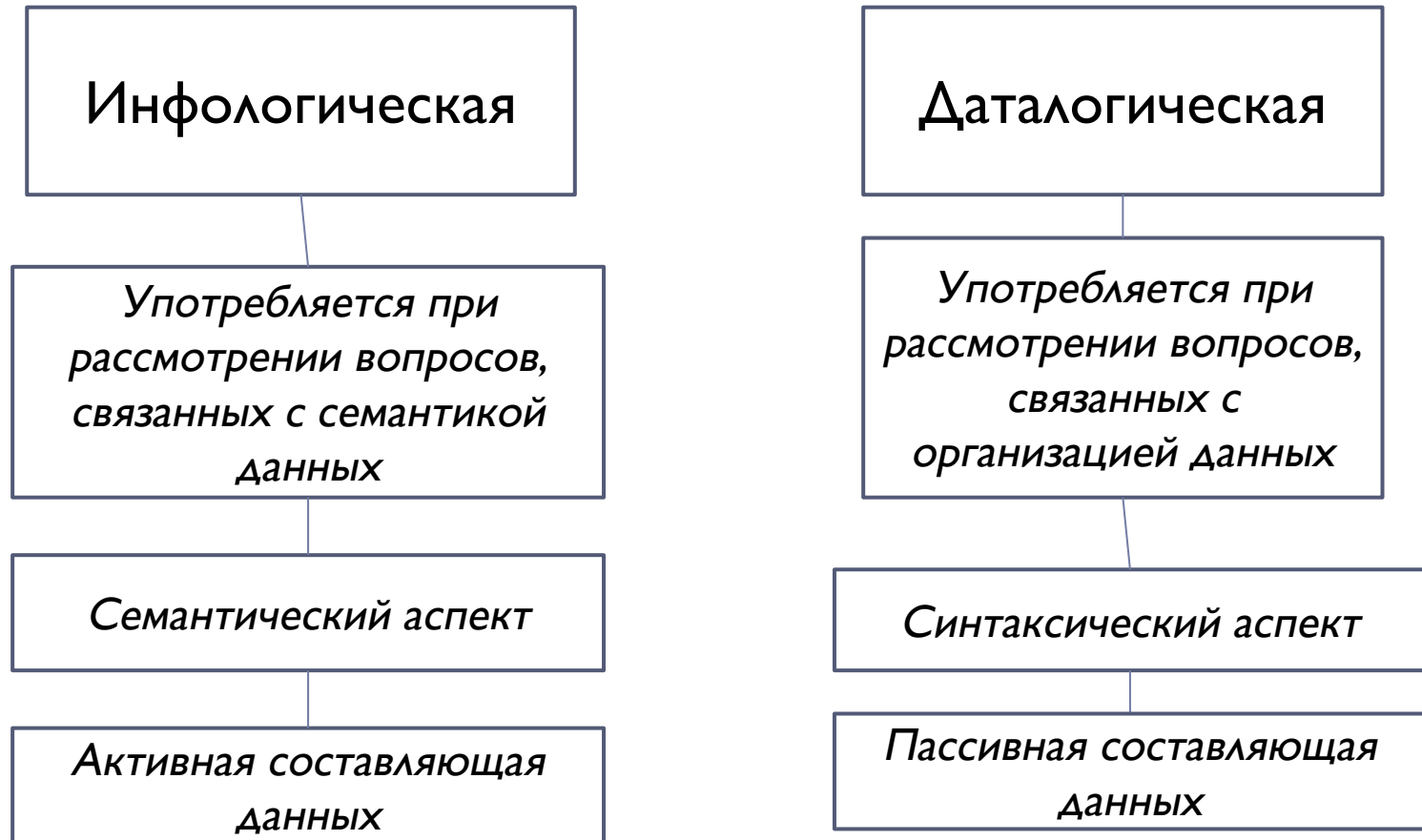


Одни и те же данные могут быть различным образом истолкованы. Одну и ту же информацию можно представить (записать) многообразными способами.

Информация и данные

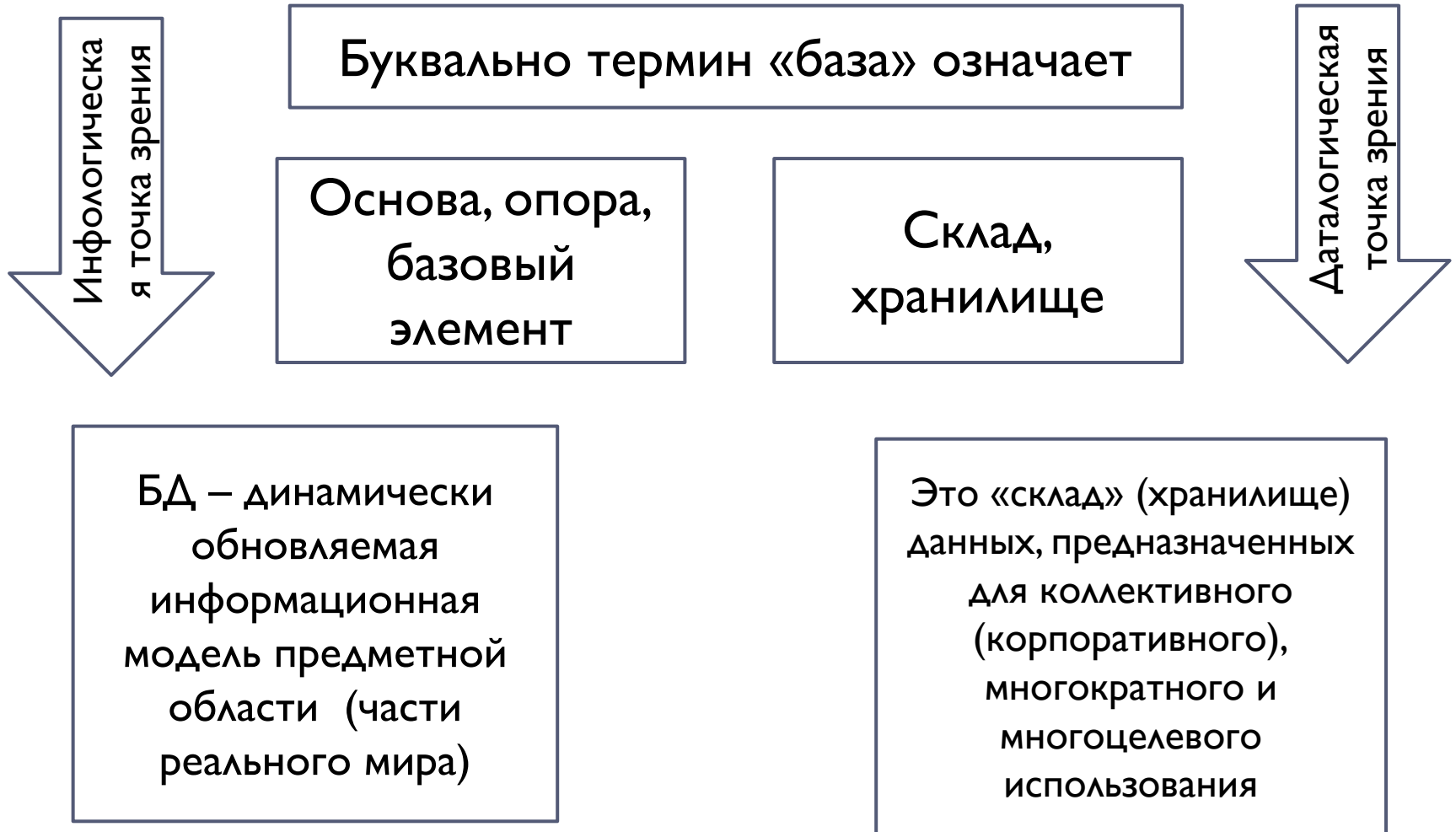


Две точки зрения на данные, используемые в теории баз данных



Сами данные пассивны, активность им придают прикладные программы – ввода, вывода и обработки данных. Только программы обеспечивают интерпретацию и связность данных.

Две точки зрения на базу данных



1: БД как информационная модель

БД – динамически обновляемая информационная модель предметной области (части реального мира)



Объектом моделирования являются информационные процессы, протекающие в предметной области. Всякое моделирование предполагает абстрагирование и отражение самого существенного (основного).

В базе данных подлежат хранению только опорные данные, из которых можно получить другие.

Здесь везде база данных опора, основа.

Проектирование базы данных - моделирование предметной области – создание семантических и синтаксических моделей разной степени формализации: от словесного описания предметной области до определения структуры базы данных на формальном языке описания данных
СУБД

2: БД как склад данных

Это «склад» (хранилище) данных, предназначенных для коллективного (корпоративного), многократного и многоцелевого использования

Если «склад» данных используется многими пользователями и для многих целей, то он должен иметь единую организацию и единое управление, которые удовлетворяют ряду **требований**

Требования к организации

БД:

- **НЕ**избыточность
- **НЕ**противоречивость (целостность)
- **НЕ**зависимость от приложений

Требования к управлению

БД:

- Защита данных
- Эффективность доступа для каждого пользователя

НЕизбыточность и НЕпротиворечивость

База данных избыточна, если удаление какого либо элемента данных ведет к потере информации о предметной области

База данных непротиворечива (целостна), если все хранящиеся в ней данные удовлетворяют определенным условиям

Избыточность базы данных проявляется в возникновении дублирования (повторения) данных.

Различают избыточное и избыточное дублирование данных.

При избыточном дублировании повторяются не просто данные, а информация о предметной области

Условия, которым должны удовлетворять хранимые в базе данные, называют **ограничениями целостности** .

Всякое ограничение целостности задает некоторые логико-семантическое свойство данных и определяется спецификой предметной области. Это семантическая составляющая данных.



Демонстрационный пример

База данных «Список рабочих завода»

Таб. номер	ФИО	Год поступления	должность	разряд	Номер цеха	Название цеха
1	2	3	4	5	6	7
100	Иванов И.И.	2004	Токарь	5	20	Сборочный
101	Петров П.П.	1990	Токарь	5	20	Сборочный
105	Егоров Е.Е.	2007	Токарь	7	20	Сборочный
170	Иванов Н.Н.	2000	Кузнец	7	30	Литейный
200	Жуков Р.Р.	1995	Кузнец	4	30	Кузнечный

Ограничения целостности:

- Каждый рабочий имеет уникальный табельный номер
- Номер цеха однозначно определяет название цеха
- Номер цеха двузначное положительное число
- Допустимые значения для каждого столбца таблицы



Демонстрационный пример

База данных «Список рабочих завода»

Таб. номер	ФИО	Год поступления	должность	разряд	Номер цеха
1	2	3	4	5	6
100	Иванов И.И.	2004	Токарь	5	20
101	Петров П.П.	1990	Токарь	5	20
105	Егоров Е.Е.	2007	Токарь	7	20
170	Иванов Н.Н.	2000	Кузнец	7	30
200	Жуков Р.Р.	1995	Кузнец	4	30

Номер цеха	Название цеха
20	Сборочный
30	Литейный

► Для устранения избыточности базу данных следует представить в виде двух таблиц

НЕизбыточность

Однако на практике во многих базах данных избыточность присутствует. В этом случае говорят об **управляемой** или **минимальной избыточности**. Она нужна для

- уменьшения времени доступа к данным или упрощения способов адресации;
- обеспечения возможности восстановления данных при их случайной потере.



Независимость БД от приложений

Программы, с помощью которых пользователи работают с базой данных, называются приложениями. В общем случае с одной базой данных могут работать множество различных приложений.

Например, если база данных моделирует некоторое предприятие, то для работы с ней может быть создано приложение, которое обслуживает подсистему учета кадров, другое приложение может быть посвящено работе подсистемы расчета заработной платы сотрудников, третье приложение работает как подсистемы складского учета, четвертое приложение посвящено планированию производственного процесса.

Под **независимостью данных от приложений** и наоборот, приложений от данных, подразумевается, что изменение одних не приводит к изменению других.

При рассмотрении приложений, работающих с одной базой данных, предполагается, что они могут работать параллельно и независимо друг от друга, и именно СУБД призвана обеспечить работу множества приложений с единой базой данных таким образом, чтобы каждое из них выполнялось корректно, но учитывало все изменения в базе данных, вносимые другими приложениями.