Основы проектирования баз данных

# Основы. Что такое Базы Данных?

(Первые досье собирали в средние века. В 60-х годах началось развитие сохранение информации на электронные носители.)

База данных – инструмент для сбора и структурирования информации.

Информационные системы (ИС) – системы для обработки документов, а точнее содержащейся в ней информации.

1. Для обеспечения их работы нужны сравнительно низкие вычислительные мощности.
2. Данные, которые они используют, имеют сложную структуру.
3. Предметная область является частью реального мира подлежащая изучению с целью организации управления и автоматизации.
4. Метаданные – данные независимые от программ пользователя
5. Система Управления Базой Данных (СУБД) – важнейший компонент информационной системы

Основные функции СУБД:

1. Управление данными во внешней памяти
2. Управление данными в ОЗУ
3. Логирование изменений
4. Восстановление после сбоев

Классификация БД:

1. По форме представления информации
2. По типу используемой модели данных: иерархические, сетевые, реляционные(табличные), объектные, функциональные.
3. По типологии хранения данных: локальные и удаленные

-----------------------------------------------------------------------------------------------------------------------------------

Базы данных – набор сведений, хранящихся некоторым упорядоченным образом.

СУБД – совокупность языковых и программных средств, которая осуществляет доступ к данным, позволяет их создавать, изменять и удалять, обеспечивает безопасность данных и т.д.

Международные стандарты:

База данных – совокупность данных, хранимых в соответствии со схемой данных, манипулирование которыми выполняют в соответствии с правилами средств моделирования данных.

База данных – совокупность данных, организованных в соответствии с концептуальной структурой, описывающей характеристики этих данных и взаимоотношения между ними, такое собрание данных, которое поддерживает одну или более областей применения.

Классификация по модели:

1. Иерархическая
2. Объектная и объектно-ориентированная
3. Объектно-реляционная
4. Реляционная
5. Сетевая
6. Функциональная

Классификация по содержимому:

1. Географическая
2. Историческая
3. Научная
4. Мультимедийная
5. Клиентская

Классификация по степени распределенности:

1. Централизованная или сосредоточенная - БД полностью на одном компьютере.
2. Распределенная - БД, Составные части которой размещаются в разных узлах сети.
3. Неоднородная - БД, фрагменты которой размещаются в более одной СУБД.

Другие виды БД:

1. Пространственная - БД, в которой поддерживаются пространственные свойства сущностей предметной области. (Геоинформатика)
2. Временная или темпоральная – БД, в которой поддерживается какой-либо аспект времени, определяемого пользователя. (Физика)
3. Пространственно-временной – И пространство и время.

Циклическая – один и тот же элемент перезаписывается по использовании некоторого объема.

Простейшая работа БД

SQL - запрос

Система управления базой данных

Данные

По характеру использования СУБД делят:

1. Однопользовательские – один человек/компьютер/система
2. Многопользовательские – одна БД, но пользователей много

Структурирование – набор соглашений о способах представления данных.

Реляционная структура – все данные представлены в виде простых таблиц, разбитых на строки и столбцы, на пересечении которых расположены данные.

# Интерфейс

Интерфейс – способ отображения информации и ведения диалога между пользователем и компьютером.

Понятие надежности системы:

1. Сохранность информации, не зависящая от любых сбоев
2. Безотказность работы системы в любых условиях
3. Обеспечение защиты от несанкционированного доступа

Вычислительные задачи – имеют сложные алгоритмы по обработке данных по мере необходимости использования данных простой структурой. Как правило эти данные имеют сложные алгоритмы, и содержатся в отдельном файле. Файлы данных имеют простую структуру в виде одномерных или двумерных массивов.

К информационным задачам относят:

1. Большие объемы обрабатываемой информации
2. Данные сложной структуры
3. Сравнительно простые алгоритмы обработки

Объект – нечто существующее и различимое, обладающее набором свойств.

Отличие одного объекта от другого определяется конкретными значениями свойств. Объекты бывают материальными и идеальными. Материальным объектом относят предметы материального мира: автомобили, мебель. К абстрактным объектам относят только такие, которые существуют в представлении или памяти человека. К таким объектам можно отнести спектакль, сюжет книги.

Параметр – конкретное значение свойства объектов.

Сущность – отображение объекта в памяти человека или компьютера. Сущность всегда меньше объекта.

Атрибут – это конкретное значение свойства сущности.

В одной БД бумага - это атрибут, а в другой это параметр.

Ключевыми элементами данных (ключом) - называется такой атрибут, по значению которого можно определить значения для других не ключевых атрибутов.

Связь – функциональная зависимость между сущностями. Если между некоторыми сущностями существует связь, то атрибуты из одной сущности ссылаются на атрибуты другой сущности. Связи описываются основными характеристиками.

Типы связи:

1. Идентифицирующие - когда дочерняя сущность однозначно определяется через её связь с родительской сущностью
2. Родительская сущность
3. Дочерняя (зависимая) сущность

Мощность связи – отношение количества родительских сущностей к соответствующему количеству дочерних сущностей.

# Этапы жизненного цикла ПО

1. Анализ – любая работа начинается с анализа БД. Это определение потребности клиента и специфики предметной области.   
   2. Проектирование – что из чего состоит, кто с кем работает. Это логическая разработка спецификации программного продукта.  
   3. Реализация (по заданному макету) – само программирование. Это физическая реализация логических спецификаций.  
   4. Тестирование. Этот выявление некорректной работы ПО.  
   5. Внедрение. Это передача программного продукта заказчику и введение его в эксплуатацию.

Работа с предметной областью

Анализ предметной области – это совокупность методик и правил, направленных на максимальное получение информации о предприятии, требованиях клиента, специфике организации и области применения.

Цели:

1. Определение потребностей клиента
2. Определение функций программного продукта.
3. Определение бизнесс-процессов, протекающих на предприятии.
4. Проведение анализа, составление специфик существующего программного продукта.
5. Определение техно-экономических показателей.
6. Определение методик разработки.
7. Определение технологии разработки.

Для анализа предметной области необходимо выделять отделы предприятия в подсистему, которые могут выполнять более узкие функции. В теории БД отделы называются подсистемами. Это часть предметной области, направленных на выполнение конкретного круга задач.   
Пример: отдел кадров, бухгалтерия. Важно понимать:  
Подсистемы работают как единое целое и могут взаимодействовать друг с другом ……. Множества информационных объектов (персонал, заказ, должность, оплата и т.д.) позволяют связывать системы между собой. Информационный объект – часть подсистемы, которая хранит или передает информацию о том или ином объекте.

Структура вопросов для проведения анализа.

1. Полное наименование организации.
2. Род деятельности.
3. Наличие филиалов.
4. Круг потребителей.
5. Выпуск продукции.
6. Нормативные документы(внутренние и внешние).
7. Структура предприятия.
8. Штатная структура предприятия.

Пример: МПТ. ФБГОУ ВО РЭУ им. Г.В.Плеханова МПТ. Обучение абитуриентов по специальностям в IT сфере в соотв с ФГосСПО. Два филиала. 1. Нежинская 7. 2. Нахимовский проспект. Абитуриенты, поступающие на базе 9 и 11 классов. Готовый знающий студент, защитивший диплом. По окончании обучения студенту присваивается СПспециальность в рамках одного из выбранных направлений. Документы: зачетка, студенческий, журнал, билеты, расписание, аттестат. Внешние документы: закон об образовании, лицензия, аккредитация, ГОС, договор с охраной, с буфетом.

# Анализ предметной области с помощью методологии SADT, стандарт IDEEF 0

Методология – набор спецификаций, с помощью которых можно организовать структуру.

Задачи анализа:

1. Фиксация требований в установленной форме
2. Получение полноценной картины о разрабатываемом программном продукте
3. Фиксирование и утверждение специфик программного продукта с заказчиком
4. Выявление неточностей у существующих программных обеспечений

Сбор информации о предметной области

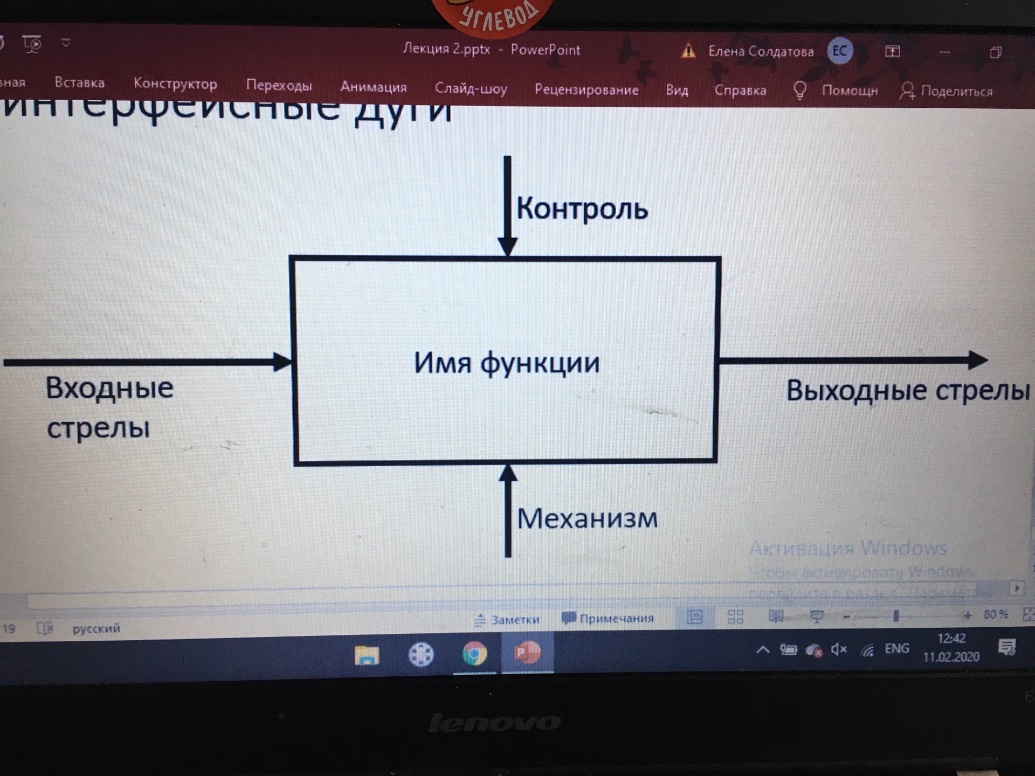
Анализ, как и начало любого проекта – это сбор и обработка информации о спецификах предметной области

1. Устное – общение с заказчиком
2. Документационное – внутренние документы, описывающие предприятие
3. Техническая документация – ранее разработанные на программный продукт документы
4. Интернет

На основе методологии SADT была разработана методология IDEF 0 (Icam Defenition) – которая является основной частью программы ICAM (интегрированная компьютеризация производства)

Направлена на выполнение основных задач:

1. Пояснения функциям предметной области
2. Определение объектов предметной области
3. Определение механизмов выполнения той или иной функции



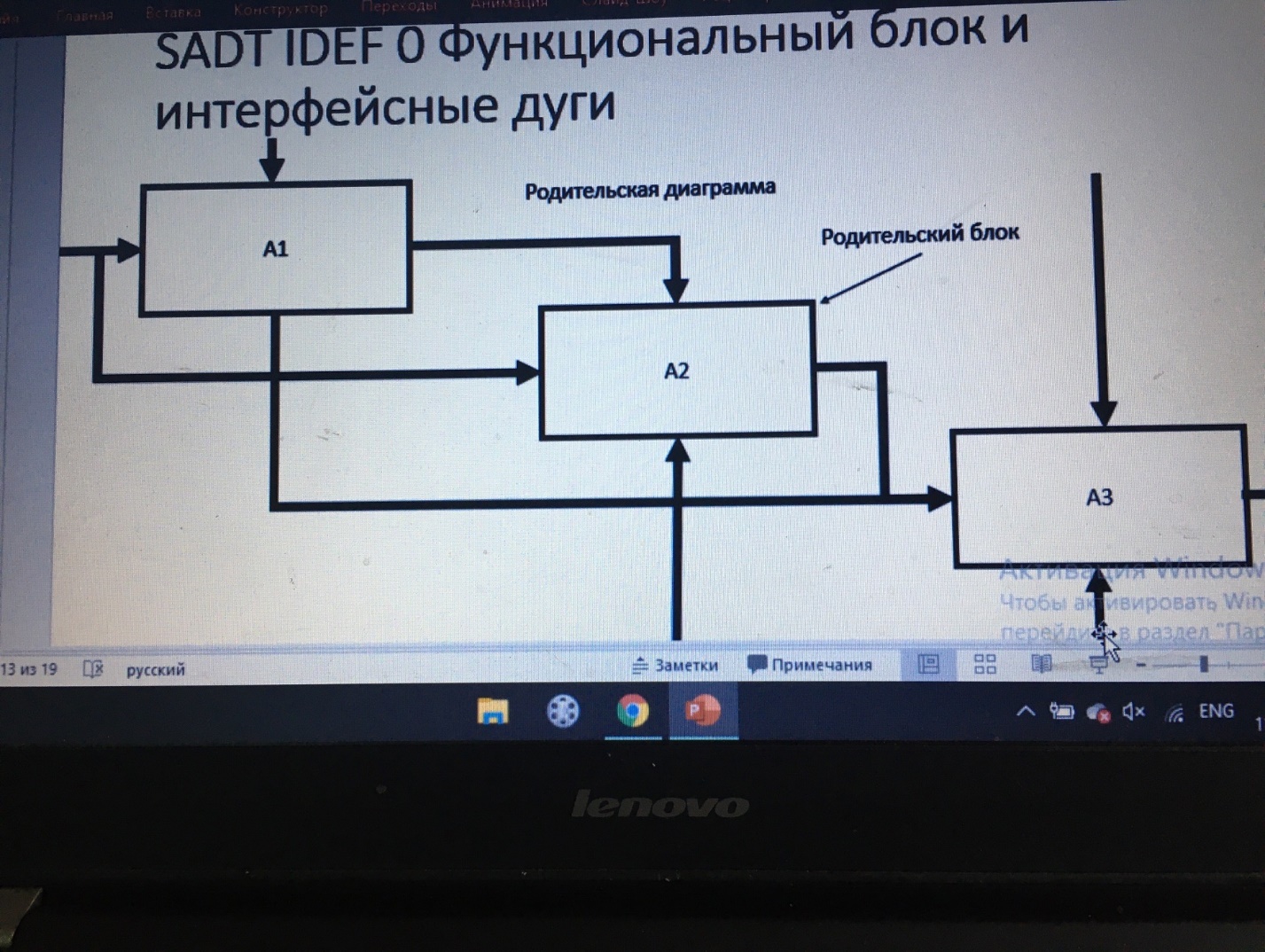
Блок имени функции – определяет процесс, действие, функцию которая выполняется в предметной области, подсистеме или процессе

Входные стрелы – данные об объектах, людях, событиях, которые подлежат обработке процессом и дальнейшем преобразованием в результат

Контроль – объекты стандартов, правил, нормативных документов предметной области, обеспечивающие выполнение процессов/функций

Механизм – объекты за счет чего выполняются процессы, штатные или структурные единицы предприятия, некое оборудование или физические механизмы

Выходные стрелы – объект, который является продуктом входных стрел, за счет выполнения функции некими механизмами на основе контроля



Декомпозиция функции – достижение полноты картины действий предметной области и подсистем зависит от более глубокого анализа

Процесс углубления в каждый родительский процесс именуется декомпозицией

Пример декомпозиции:

1. Учебный процесс
   1. Приветствие
   2. Перекличка
   3. Объяснение материала
   4. Опрос
   5. Контроль качества
   6. Выдача задания

Дуги/стрелы не могут присутствовать на нижних уровнях декомпозиции и при этом отсутствовать на родительском уровне.

# СУБД

Свойства данных, которые подлежат хранению в базе

Свойства данных являются правилами проектирования, реализация, сопровождение.

К основным свойствам относят:

1. Не избыточность
2. Адекватность
3. Достоверность
4. Целостность
5. Не противоречивость

Не избыточность – заключается в удалении излишней информации за счет упразднения повторяющихся полей, групп полей и данных.

Адекватность – заключается в хранении в каждом столбце соответствующей информации, которая может не относиться к предметной области.

Достоверность – соответствие данных предметной области.

Непротиворечивость – данные об одном и том же объекте в разных частях БД имеют одинаковое значение атрибута.

Целостность – заключается в полноценном и сохранном хранении и выводе информации в БД. Обеспечивается с двух основных сторон:

1. Со стороны приложения БД.
2. Со стороны сервера.

Целостность данных со стороны приложения:

* Обеспечение надежного ввода информации за счет проверки на пустые поля ввода.
* Маски ввода (работа с полем).
* Шифрование и дешифрование данных.
* Разграничение прав доступа.

Целостность данных со стороны сервера:

* Мониторинг операций в БД.
* Разграничение внутренней политики безопасности (роли и группы ролей).
* Проверка на отказоустойчивость работы сетевого оборудования.
* Обеспечение внутренних защищенных каналов связи сервера с клиентами (приложениями БД).
* Обеспечение проверки перед /после добавлением/ изменением/ удалением данных за счет триггера.

Архитектура БД

Виды архитектур БД.

Архитектура БД (Система БД) – это логическое расположение компонентов системы БД (приложение, СУБД и БД) в рамках разрабатываемой базы и предметной области.

Виды:

1. Локальная – все находится на одном ПК.

Достоинства:   
простота  
используется для частных задач  
простота в сопровождении  
  
Недостатки:  
в случае поломки ПК данные будут утеряны  
невозможность обмена данными  
минимальная сохранность информации  
трудность восстановления

1. Централизованная – компоненты системы БД расположены следующим образом: на ПК пользователя - приложение, на сервере – СУБД и БД.

П и БД

СУБД и БД

П и БД

Достоинства:  
многопользовательский режим  
возможность сохранно хранить информацию  
  
Недостатки:  
невозможность использования на крупных предприятиях  
недостатки сети  
сложность в администрировании  
более серьезная по масштабам разработка, требуется группа разработчиков

1. Распределенная – все компоненты системы расположены на разных взаимосвязанных элементах.

БД3

БД2

БД1

П

СУБД

П

П

Достоинства:  
глобальное решение корпоративных задач  
возможность повысить целостность данных  
достоинства как у централизованной архитектуры  
  
Недостатки:  
проблемы с сетью  
требуется команда разработчиков и администраторов

Виды информационных систем по степени распределенности:

* Файл-серверная архитектура

Предназначена для небольших организаций в 10-15 одновременных подключений, пользователь работает со всей БД целиком.

СУБД + П



БД

СУБД + П



Достоинства:  
возможность разработки 1-2 разработчиками  
простота в администрировании  
  
Недостатки:  
перегрузка каналов связи  
светофорный режим  
невозможность обращения двух и более пользователей к одной ячейке памяти, пока она находится под редактированием другого пользователя

* Клиент-серверная архитектура

-двухзвенная архитектура  
-трехзвенная архитектура  
-многозвенная архитектура

Visio – установить

Повторить все лекции, письменный опрос по сегодняшней лекции!

# Категория пользователей БД

Частые роли при эксплуатации БД:

Конечный пользователь - категория лиц, непосредственной задачей которого является актуализация данных.

Функции:

Манипулирование данными

Вывод информации в отчётном и статическом виде

Системный аналитик и программист

Задачи:

Собирать и обрабатывать информацию о потребностях клиента/заказчика

Разрабатывать технические спецификации

Переработка требований заказчика

Подготавливать проектную спецификацию

Обслуживающий персонал

Задачи:

Поддерживать в рабочем состоянии телекоммуникационную технику и сетевое оборудование

Отвечать на запросы конечных пользователях о неполадках

Администратор БД

Задачи:

Ведение БД

Поддержание структуры БД в работоспособном состоянии

Структурированные и неструктурированные данные. Учёт и автоматический учёт.

Неструктурированными данными называют данные, например, записанные в текстовом файле.

Структурные данные - это данные, которые были сгруппированы по определённым параметрам.

Учёт - это установление наличия, кол-во чего-либо кого-либо и регистрации с занесением в список.

Виды учёта:

Оперативный учёт - формирует оперативную информацию для текущего управления и контроля за текущими процессами и фактами.

Статистический учёт - представляет систему сбора и обобщения информации о состоянии массовых процессах и явлений, происходящих в отрасли, стране и т.д..

Бухгалтерский учёт - представляет собой систему наблюдения, сбора, регистрации и группировки данных, и итогового обобщении информации.

Финансовый учёт - представляет собой информацию о имущественном и финансовом состоянии.

Управленческий учёт - представляет собой систему формирования информации менеджерами для оперативного управления предпринимательской деятельностью.

Налоговый учёт - используется для получения данных для необходимых для правильного исчисления налогов.

Автоматизированный учёт - учёт, который выполняется с использованием вычислительной техники.

Построение межтабличных связей

Межтабличная связь организуется с целью оптимального взаимодействия между всеми сущностями, атрибутами и данными в базе.

Виды ключей

Первичный ключ (Primary Key) - это атрибут по значению, которого однозначно определяется запись в таблице БД.

Внешний ключ (Foreign Key) - данный ключ позволяет ссылаться на записи родительской таблицы, находится в дочерней таблице и название атрибута берётся от названия родительской таблицы.

Суррогатный ключ – выполняет роль первичного ключа, но не является параметром объекта

Составной ключ – выполняет роль составного ключа, в том случае, когда невозможно определить запись по одному атрибуту, берется совокупность 2-3 атрибутов

1 к 1 - Одна запись таблицы ссылается на одну запись в другой таблице. Данная связь используется, когда необходимо разделить одну сущность, состоящую из 50-60 атрибутов, на более мелкие сущности по 5-10 атрибутов в каждой.

Супер-сущности – вид связи, 1 к 1, когда невозможно структурировать данные в одну таблицу.  
Общие атрибуты сущности определяются в общую таблицу (Супер-сущность), а остальные уникальные данные выносят в отдельные уникальные сущности.

1 ко многим – одна запись одной таблицы ссылается на множество записей другой таблицы

Много ко многим - много записей одной таблицы ссылаются на много записей другой таблицы

Виды таблиц по степени работы:

|  |  |  |
| --- | --- | --- |
|  | Частота манипуляций | Паузы между манипулированием |
| Справочная | Низкая | Максимальная |
| Операционная (системная) | Высокая | Минимальная |

Виды таблиц по приоритетам:  
Родительская таблица – со стороны единички  
Дочерняя таблица со стороны М

# Алгоритм подготовки предметной области к нормализации

1. Анализ предметной области (См. 1 практическую работу)
2. Анализ предметной области методология SADT (См. 2 практическую работу)
3. Описание технической области (См практическую)
4. Описание объектов предметной области, параметров, ключевых элементов данных
   1. Описание объектов. В данном разделе дать описание информационным объектам. В соответствии с 1 практической работой. Группировать объекты по категориям – материальный или абстрактный объект. Название объекта – пояснение о его назначения.
   2. Описание параметров. На основе описания объектов выделить основные параметры и описать их. Название параметра – описать его назначение.
   3. Описание сущностей. На основе описания объектов, преобразовать объекты в конкретные сущности и дать пояснение к сущностям. Объект1 + Объект2 = Сущность1 – описание сущности;
   4. Описание атрибутов. На основе описания сущностей и описания параметров к каждой сущности определить оптимальное количество параметров и их описание. Атрибут– описание атрибута.
   5. Выявление ключевого элемента данных. У каждой из сущности из списка атрибутов определить ключевые элементы данных.

Информационные объекты

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Наименование подсистемы | Информационные объекты | Параметры | Сущность | Атрибуты | Ключевое значение данных |
| Отдел техподдержки | Работники, техника, должность, вопрос, ответ, пользователь, отдел, доступ, справочная информация, жалобы | | | | |
|  | Работники | ФИО, дата рождения, должность, пол, вес, возраст, национальность, паспортные данные, прописка, семейное положение | Сотрудники, личное дело сотрудника | Сотрудники, ФИО, дата рождения, пол, паспортные данные, личное дело сотрудника, данные об образовании | Сотрудники:  Серия и номер паспорта  Личное дело: Номер личного дела |
|  | Техника | Тип техники, название, марка, инвентарный номер, | Техника, тип техники | Техника:  Номер, название, марка, тип техники | Техника:  Номер, название, марка, тип техники |
|  | Должность | Наименование, оклад, требования ЗП, график работы | Должность | Наименование, оклад, номер | Номер |
|  | Вопрос | Тема вопроса, пользователь, письмо, тип письма, дата отправки, наличие графических объектов, статус | Вопрос:  Тип вопроса | Вопрос:  Тема вопроса, дата отправки, текст письма, наличие графических объектов, номер сотрудника | Номер |
|  | Ответ | Номер вопроса, номер сотрудника, дата и время |  |  |  |
|  | Пользователь | ФИ, дата рождения, почта, телефон, пароль, графический контент, карьера, военная служба, город |  |  | ФИ, дата рождения, почта, телефон, пароль, графический контент, карьера, военная служба, город |