

Модели данных

(с) Пыхалов А.В, ЮГИНФО ЮФУ, 2011

Модель данных

- Модель данных состоит из
 - ЯОД,
 - ЯМД,
 - средств описания целостности данных.
- МД : $M=(G,O)$
 - G - множество правил порождения схем
 - O – множество операций (ЯМД) (функции на множестве состояний БД)

База данных

- БД – (D, C, d)
 - D – схема БД,
 - d – экземпляр БД,
 - C – ограничения целостности на D .

Трехуровневая архитектура БД ANSI/SPARC

- Внешний уровень – уровень пользовательского представления
- Концептуальный уровень – отражает данные “как они есть”, полное содержание БД (уровень, на котором работает АБД)
- Внутренний уровень – уровень хранения данных (табличные пространства, файлы, организация индексов, таблиц и т.д.)

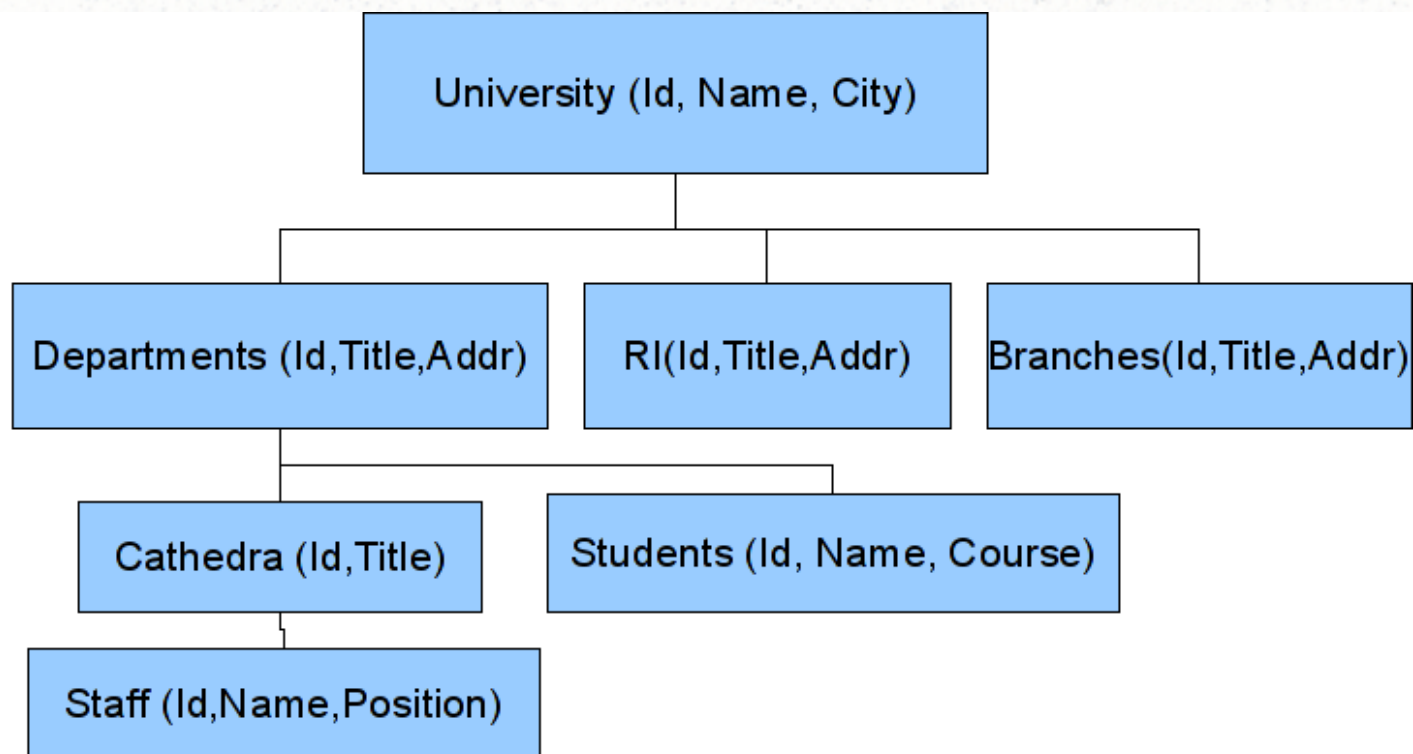
Основные модели данных

- Иерархическая модель
- Сетевая модель
- Семантические сети
- ER-модель
- Реляционная модель

Иерархическая модель

- Объекты представлены записями (атрибуты записи определены на определенном домене)
- Записи связаны отношениями родитель-потомок (один родитель может иметь множество потомков, но не наоборот)
- Примеры:
 - СУБД IBM IMS (появилась в 1966)
 - Реестр Windows или БД Gconf
 - LDAP каталоги (например, AD)

Пример иерархической СД



Пример запроса к иерархической СД

GET UNIQUE

UNIVERSITY (NAME='SFEDU')

DEPARTMENTS (TITLE='MMCS')

CATHEDRA (TITLE='VV IKT')

STAFF

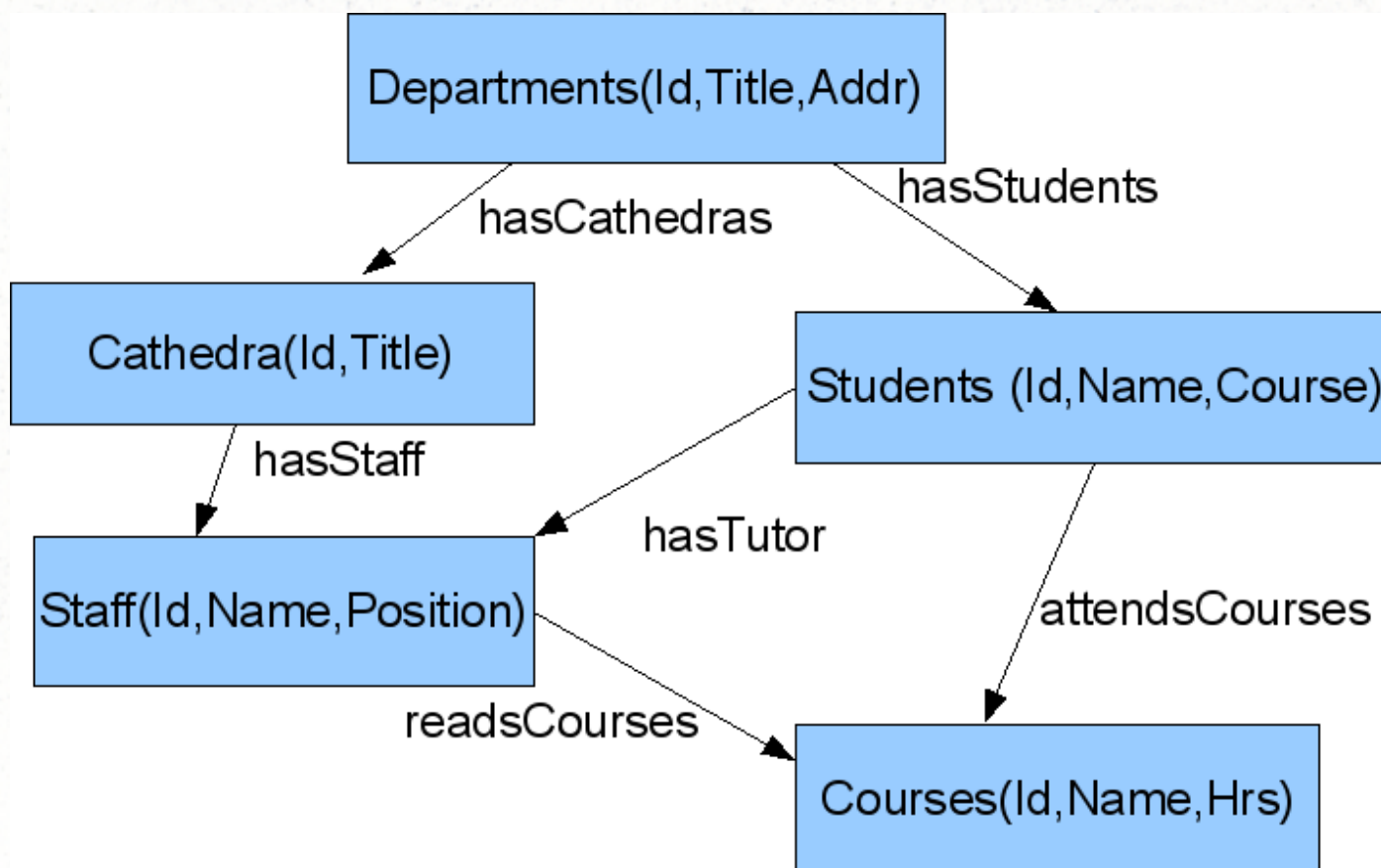
NS GET NEXT STAFF

GOTO NS

Сетевая модель (CODASYL)

- Объекты представлены записями (атрибуты записи определены на определенном домене)
- Записи связаны произвольными именованными отношениями
- Примеры
 - СУБД CA IDMS (появилась в 60-е)
 - Oracle CODASYL DBMS (досталась от DEC)

Пример сетевой СД



Пример запроса к сетевой СД

```
FIND CATHEDRA RECORD WHERE TITLE='VV IKT'  
NO: OBTAIN NEXT STAFF RECORD OF hasStaff  
GOTO NO
```

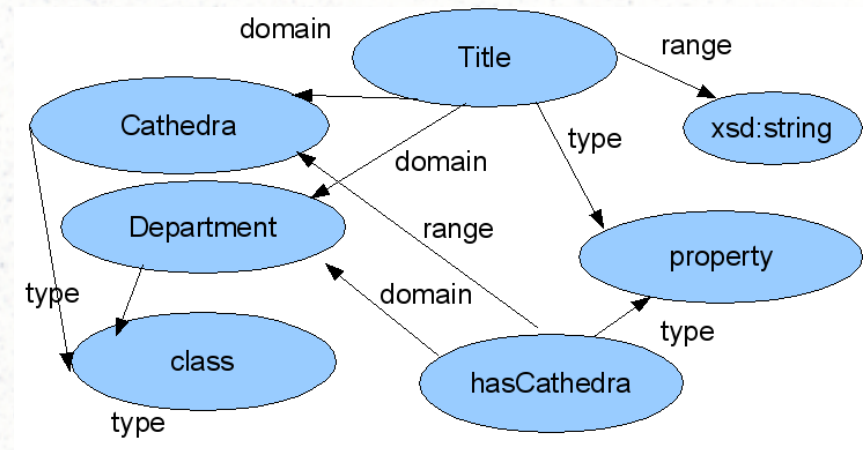
Семантические сети

- Графовая модель $G(V,E)$:
 - V – вершины, представляющие объекты
 - E – связи между объектами
 - Дуги обычно имеют имена
- Представление в исчислении предикатов
 - $e(v1,v2)$
- Примеры:
 - Semantic Web (RDF и др.)
 - WordNet

Пример СД семантической сети

```
rdfs:Class(Department)
rdfs:Class(Cathedra)
rdf:Property(Title)
rdf:Property(hasCathedras)
```

```
rdfs:domain(Title,
  Departments)
rdfs:domain(Title, Cathedra)
rdfs:range(Title, xsd:string)
rdfs:domain(hasCathedras,
  Departments)
rdfs:range(hasCathedras,
  Cathedra)
```



Пример семантической сети

```
rdf:type(MMCS,Department)
rdf:type(VV_IKT,Cathedra)
hasCathedras(MMCS,
  VV_IKT)
title(VV_IKT,"BB ИКТ")
```

Пример запроса

PREFIX univ:

http://sfedu.ru/univOnt#

SELECT ?title

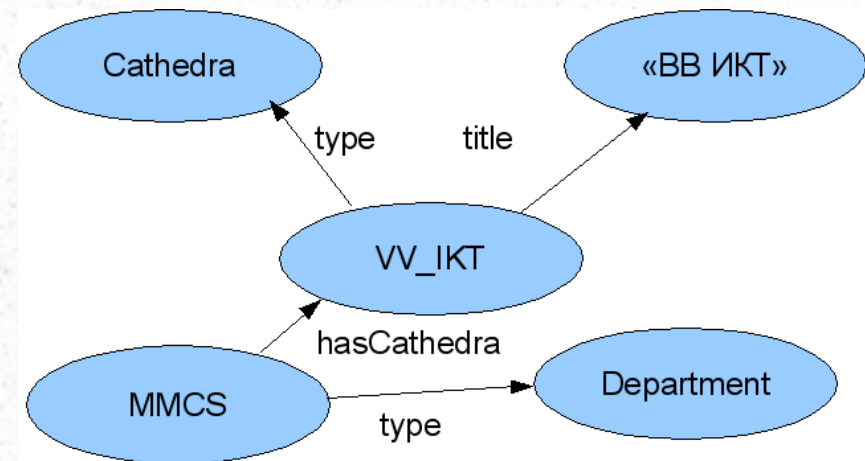
WHERE {

?x rdf:type Department

?x univ:hasCathedra ?y

?y univ:title ?title

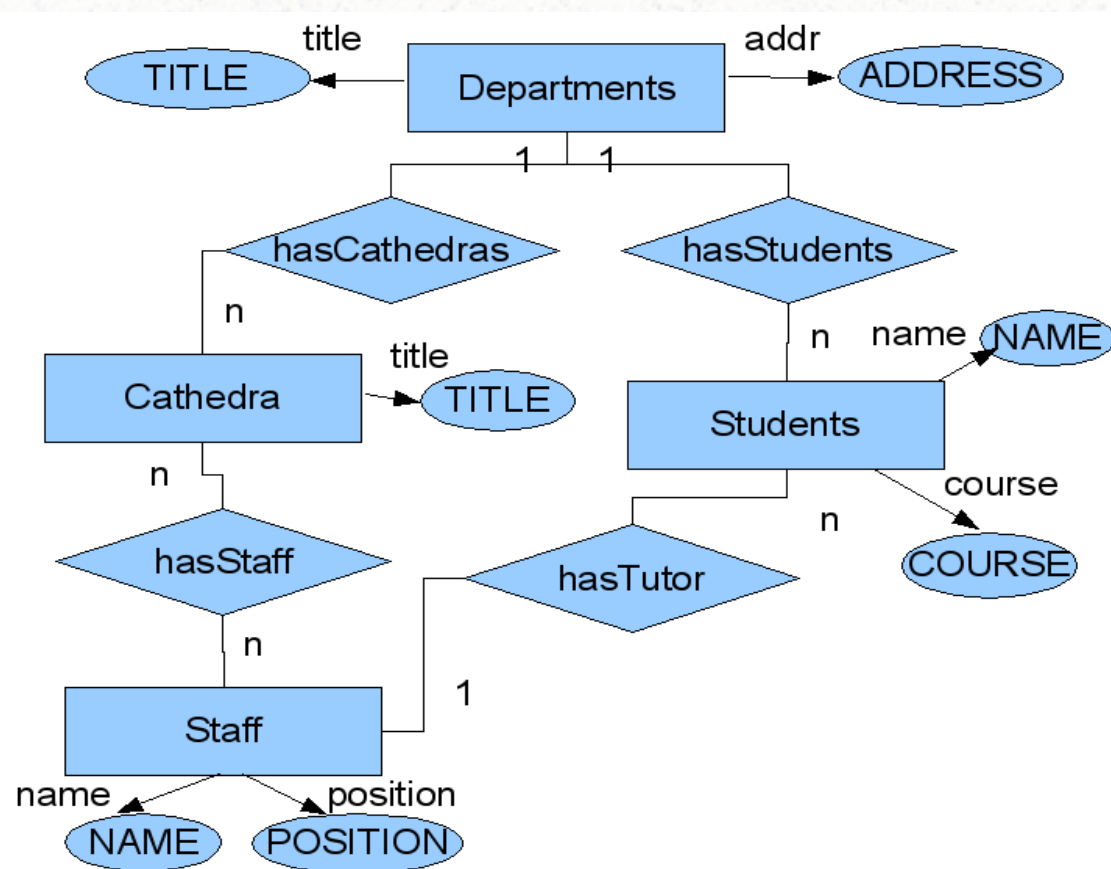
}



ER-модель (модель сущность-связь)

- ER-модель является *средством моделирования БД*
- Основные понятия — тип сущности и тип связи
- Связи могут быть n-арными, иметь тип многие-ко-многим, быть рекурсивными
- Атрибут — отображение между множеством сущностей *и* связей и множеством значений(доменом)
- Атрибуты могут быть многозначными

Пример СД ER-модели



Пример запроса к СД ER-модели

OUTPUT Staff.name

```
SELECT Departments.title='мехмат' /  
Cathedra.title='ВВ ИКТ'
```

Реляционная модель данных

- Данные воспринимаются как таблицы (отношения)
 - схема отношения определяет его структуру
 - других типов данных не существует
 - отношения
 - неупорядочены
 - не содержат дублирующихся строк (не являются мульти-множествами)
 - соответствуют таблице истинности предикатов и следовательно не содержат NULL-значений (первоначально)

Формальное определение РА

- Пусть дано конечное множество именованных атрибутов T , и для каждого A из T задан $\text{dom}(A)$ – домен атрибута A
- Схема отношения – последовательность атрибутов из T
- Кортеж t схемы $R = (A_1, \dots, A_n)$ - элемент множества $\text{dom}(A_1) \cup \{null\} \times \dots \times \text{dom}(A_n) \cup \{null\}$
- Отношение r – конечное множество кортежей схемы R

Операции РА

- Выборка $\sigma_{\text{условие}} e$
- Проекция $\pi_{\text{ИмяАтрибута}, \dots, \text{ИмяАтрибута}} e$
- Соединение (естественное) $e_1 \text{ JOIN } e_2$
- Произведение $e_1 \times e_2$
- Объединение $e_1 \cup e_2$
- Пересечение $e_1 \cap e_2$
- Разность $e_1 - e_2$
- Переименование $e_1 \text{ RENAME } a_1 \text{ AS } a_2$

Операции РА при отказе от имен атрибутов

- Выборка $\sigma_{\text{условие}} e$
- Проекция $\pi_{\text{НомерАтрибута}, \dots, \text{НомерАтрибута}} e$
- Произведение $e_1 \times e_2$
- Объединение $e_1 \cup e_2$
- Пересечение $e_1 \cap e_2$
- Разность $e_1 - e_2$

Таблица истинности операций

- При любой операции сравнения с null, кроме is_null, результат – unknown (в SQL – NULL)

AND	T	F	U
T	T	F	U
F	F	F	F
U	U	F	U
OR	T	F	U
T	T	T	T
F	T	F	U
U	T	U	U
NOT	F	T	U
IS NULL	F	F	T

Пример СД реляционной МД

```
DEPARTMENTS(Id DID# PK, Title DTITLE,  
  Addr ADDRESS);  
CATHEDRA(Id CID# PK, Title CTITLE,  
  DPT DID# REFERENCES DEPARTMENTS(Id));  
STAFF(Id SFID# PK, Name NAME, Position POSITION,  
  CTH CID# REFERENCES CATHEDRA(Id));  
COURSES(Id CRID# PK, Name CNAME#, Hrs HOURS,  
  PROF SFID#REFERENCES STAFF(Id));  
STUDENTS(Id STID# PK, Name NAME,  
  Course COURSE_NUMBER,  
  DPT DID# REFERENCES DEPARTMENTS(Id),  
  PROF SFID# REFERENCES STAFF(Id));  
STUD_COURSES(  
  Sid STID# REFERENCES STUDENTS(Id),  
  Cid CRID# REFERENCES COURSES(Id)  
  PK (Sid,Cid));
```

Пример запроса к реляционной СД

$\pi_{Name} \sigma_{DTITLE=' мехмат ' \wedge CTITLE=' ВВ ИКТ '}$ (DEPARTMENTS
RENAME id AS DID, TITLE AS DTITLE)
JOIN (CATHEDRA RENAME ID AS CID, TITLE AS CTITLE)
JOIN (STAFF RENAME ID AS SID)
 $\pi_8 \sigma_{2=' мехмат ' \wedge 5=' ВВ ИКТ ' \wedge 6=1 \wedge 10=4}$ DEPARTMENTS \times
 \times CATHEDRA \times STAFF

Нормализация БД. Первая нормальная форма

- 1 НФ – отношение находится в 1 НФ если каждый атрибут отношения атомарен.

Иванов Иван Иванович	212-34-56 309-32-25
Семенов Семен Семенович	312-32-55

Сотрудник (ФИО, Телефон)

Иванов	Иван	Иванович	212-34-56
Иванов	Иван	Иванович	309-32-25
Семенов	Семен	Семенович	312-32-55

Сотрудник (Фамилия, Имя,
Отчество, Телефон)

Вторая нормальная форма

- Отношение находится во 2 НФ, если оно находится в 1НФ и в нем нет неключевых атрибутов, зависящих от части составного ключа.

Приведение ко 2 НФ. Исходная БД

Гришин	Кладовщик	20000	Нет
Васильев	Программист	40000	Есть
Васильев	Кладовщик	25000	Нет

Сотрудник (Фамилия, Должность, Зарплата, Наличие Компьютера)

Потенциальный ключ (Фамилия, Должность)

ФЗ: Должность => Наличие Компьютера

Приведение ко 2 НФ. Преобразованная БД

Гришин	Кладовщик	20000
Васильев	Программист	40000
Васильев	Кладовщик	25000

Сотрудник (Фамилия, Должность, Зарплата)

Потенциальный ключ (Фамилия, Должность)

Кладовщик	Нет
Программист	Есть

Наличие компьютера (Должность, Наличие компьютера)

Третья нормальная форма

- Отношение находится в 3НФ, если оно находится в 2НФ и в каждый неключевой атрибут находится в нетранзитивной зависимости от каждого ключа.

Сотрудник (Фамилия, Отдел, Телефон)

Потенциальный ключ (Фамилия)

ФЗ: Отдел => Телефон

Гришин	Бухгалтерия	11-22-33
Васильев	Бухгалтерия	11-22-33
Васильев	Снабжение	44-55-66

Приведение к ЗНФ. Преобразованная БД

Гришин	Бухгалтерия
Васильев	Бухгалтерия
Васильев	Снабжение

Сотрудник (Фамилия, Отдел)

Бухгалтерия	11-22-33
Снабжение	44-55-66

Отделы(Отдел, Телефон)

Литература

- К.Дж. Дэйт. Введение в системы баз данных //Седьмое издание, Москва, Вильямс, 2001
- Д.Цикритзис, Ф.Лоховски. Модели данных // Москва, “Финансы и статистика”, 1985
- W3C. RDF Primer // <http://www.w3.org/TR/rdf-primer/>, 2004

P.S. Задание на семестр

- Warehouses (w_id,address,...)
- Products(p_id,title,price,photo,...)
- PW (w_id,p_id,quantity,...)
- Users (u_id,...)
- Managers (...) - связаны с Warehouses
- Cart(u_id, p_id, w_id, quantity)
- Управляющий может управлять товарами на складе
- Администратор может управлять управляющими
- Пользователи могут регистрироваться и заказывать товары

P.S. Требования к приложению

- Возможна одновременная работа множества пользователей, менеджеров, администратора
- Если пользователь отложил товары в корзину, их не может заказать другой пользователь
- Если пользователь не подтвердил заказ в течении N часов, содержимое его корзины возвращается на склады
- Работа с данными – на PL/SQL, интерфейс – как удобно (JSP, PHP, Python, etc...)