清华大学数据库技术与应用

SQL I

授课教师：计算机系王健楠

授课学期：2026年（春季）

课程大纲

- 数据模型概述
- 关系模型基础
- 使用 SQL 定义关系模式

课程大纲

- **数据模型概述**
- 关系模型基础
- 使用 SQL 定义关系模式

数据库与数据库系统

清華大學 Tsinghua University

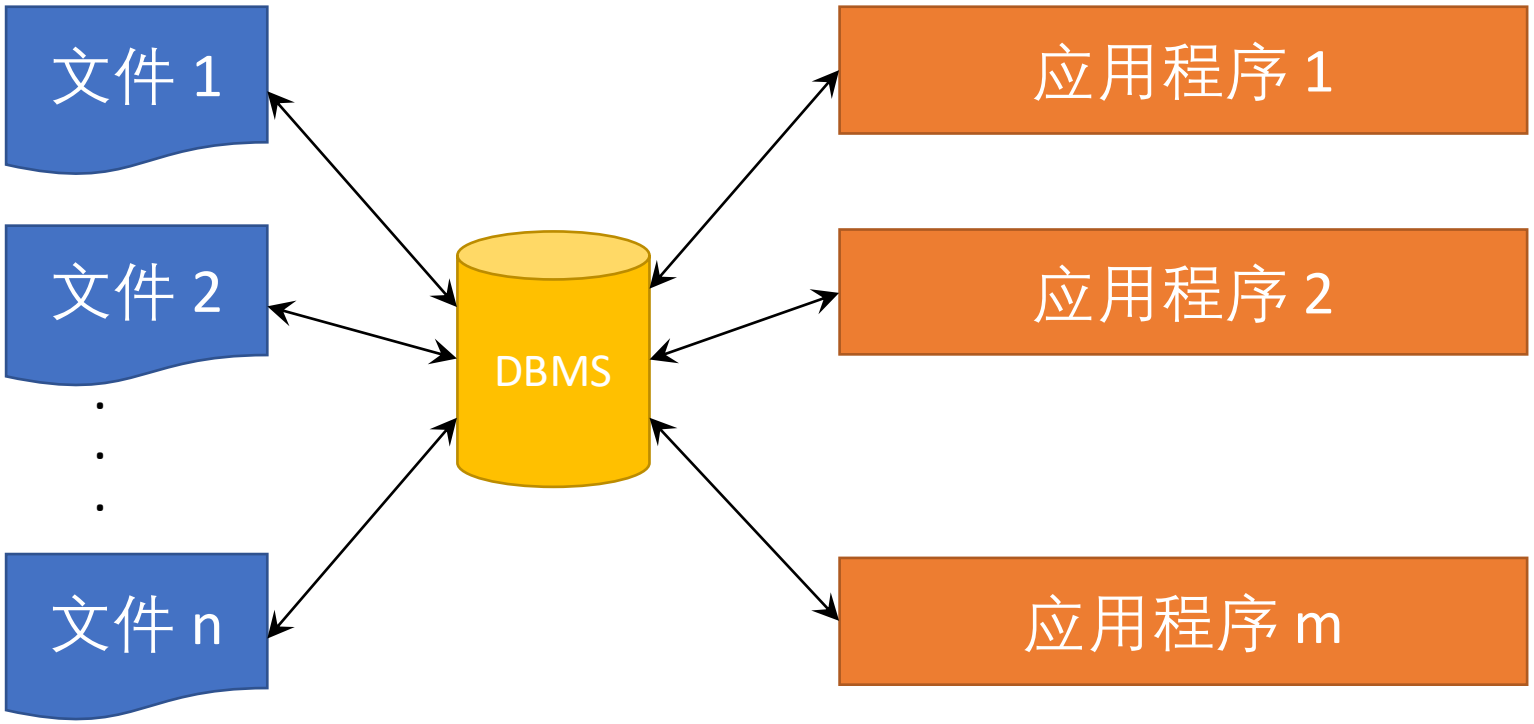
什么是数据库?

- 存储相关数据的文件的集合

什么是数据库管理系统（DBMS）？

- 用于存储和管理数据库的软件

使用DBMS存储数据



数据模型

数据模型

- 用于描述数据的数学形式化表示（或概念化方式）

描述通常包含三个部分：

- 数据的结构
- 数据的操作
- 数据的约束

数据的结构

清華大學 Tsinghua University

模式（例如：表名、属性名）

- 描述数据的概念结构

不同于数据结构（例如：链表、数组）

- 数据结构可视为一种物理数据模型

数据的操作

查询语言（例如 SQL）

- 描述可以对数据执行哪些操作

两类操作

- 检索信息的操作
- 修改数据库的操作

不同于编程语言（例如 C、Java）

- 仅支持一组有限的操作
- 便于进行查询优化

数据的约束

清華大學 Tsinghua University

约束（例如 age > 0、student# 唯一）

- 描述对数据取值的限制

不同类型的约束

- 域约束（Domain constraints）
- 完整性约束（Integrity constraints）

为什么重要?

- 确保数据的正确性

常见的数据模型

- 关系数据模型
- 键值数据模型
- 半结构化数据模型（例如 JSON、XML）

关系模型简介

Id	Name	Age	GPA
1000	Mike	21	3.8
1001	Bill	19	3.4
1002	Alice	20	3.6

数据的结构

- 表结构

查询语言

- SQL

数据的约束

- 例如：id 唯一，age > 10，name 非空（NOT NULL）

键值模型简介

Key	Value
1000	(Mike, 21, 3.8)
1001	(Bill, 19, 3.4)
1002	(Alice, 20, 3.6)

数据的结构

- (Key, Value) 键值对
- 键为整数或字符串，值可以是任意对象

查询语言

- `get(key)`, `put(key, value)`

数据的约束

- 例如：键唯一，值非空（NOT NULL）

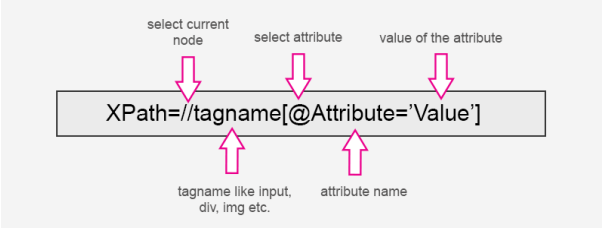
半结构化模型简介

数据的结构

- 树形结构

查询语言

- XPath



约束

- 例如：<Age> 必须为整数；每个 <Student> 中都嵌套有一个 <Name> 元素

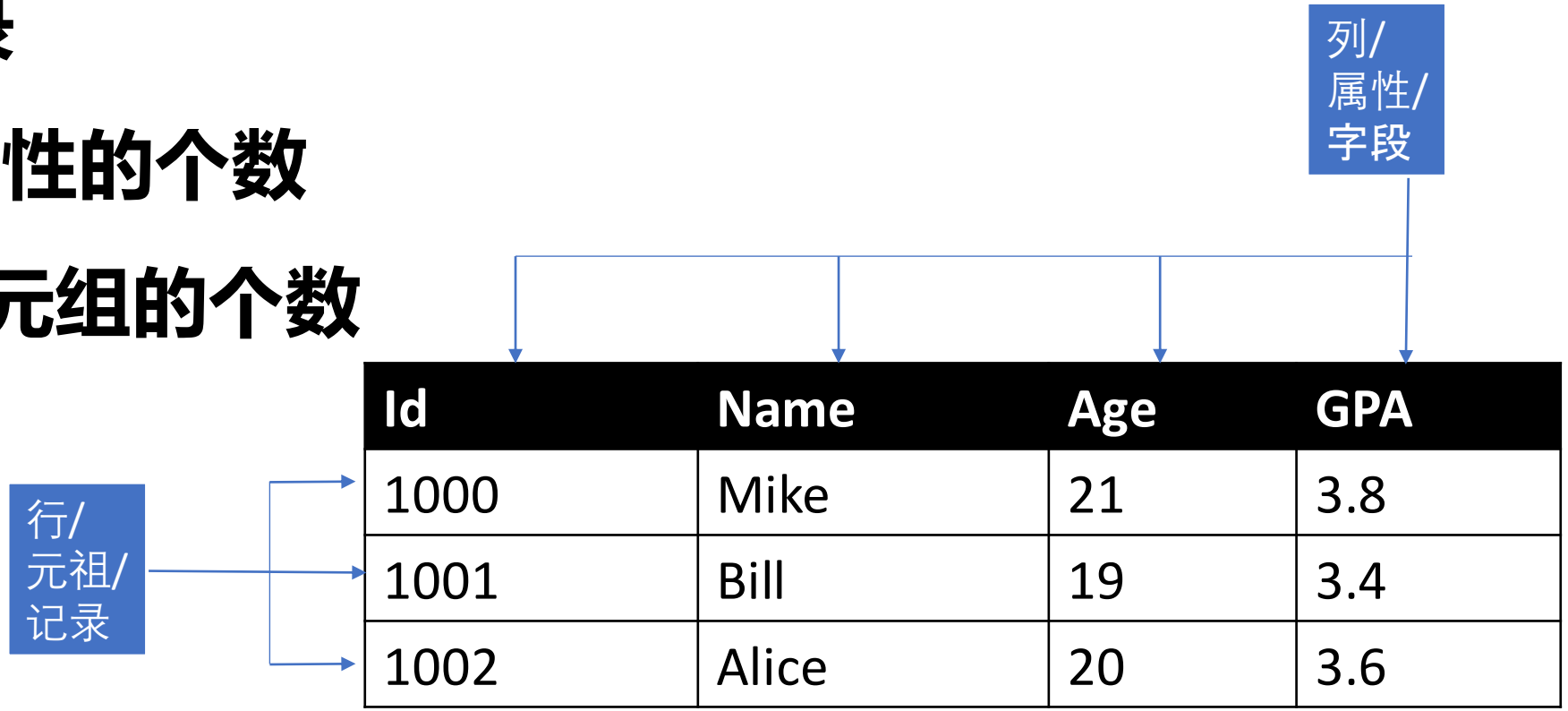
```
<Students>
  <Student id=1000>
    <Name>Mike</Name>
    <Age>20</Age>
    <GPA>3.8</GPA>
  </Student>
  <Student id=1001>
    <Name>Bill</Name>
    <Age>19</Age>
    <GPA>3.4</GPA>
  </Student>
  <Student id=1002>
    <Name>Alice</Name>
    <Age>21</Age>
    <GPA>3.6</GPA>
  </Student>
</Students>
```

课程大纲

- 数据模型概述
- **关系模型基础**
- 使用 SQL 定义关系模式

术语

- 关系（Relations）/ 表
- 列/ 属性/ 字段
- 行/ 元组/ 记录
- 关系的度 = 属性的个数
- 关系的基数= 元组的个数



Id	Name	Age	GPA
1000	Mike	21	3.8
1001	Bill	19	3.4
1002	Alice	20	3.6

模式（Schema）

关系模式

- 关系的名称 + 关系的属性集合

Student(id, sname, age, gpa)

数据库模式

- 数据库中所有关系模式的集合

Student (sid, sname, age, gpa)

Take (sid, cid)

Course (cid, cname, credit)

域（Domain）

清華大學 Tsinghua University

每个属性都有一个域（数据类型）

示例

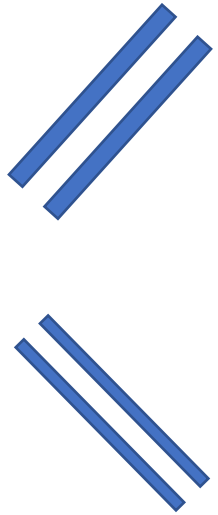
- 文本：CHAR(20), VARCHAR(50), TEXT
- 整数：INT, SMALLINT
- 实数：DOUBLE, FLOAT
- 少量依赖具体数据库厂商的类型

```
Student(id:INT, sname:VARCHAR(50), age:INT, gpa:FLOAT)
```

关系的等价表示

顺序无关紧要！

Id	Name	Age	GPA
1000	Mike	21	3.8
1001	Bill	19	3.4
1002	Alice	20	3.6



Id	Name	GPA	Age
1000	Mike	3.8	21
1001	Bill	3.4	19
1002	Alice	3.6	20

Id	Name	Age	GPA
1000	Mike	21	3.8
1002	Alice	20	3.6
1001	Bill	19	3.4

练习-1：术语

清華大學 Tsinghua University

Accounts

AcctNo	Type	Balance
12345	savings	12000
23456	checking	1000
34567	savings	25

Customers

fname	lname	idNo	account
Robbie	Banks	901-222	12345
Lena	Hand	805-333	12345
Lena	Hand	805-333	23456

1. 列举另外两种“行”的说法
2. 列举另外两种“列”的说法
3. 列举另一种“表”的说法

练习-2：术语

清華大學 Tsinghua University

Accounts

AcctNo	Type	Balance
12345	savings	12000
23456	checking	1000
34567	savings	25

Customers

fname	lname	idNo	account
Robbie	Banks	901-222	12345
Lena	Hand	805-333	12345
Lena	Hand	805-333	23456

4. 指出每个关系表的属性
5. 指出每个关系表的元组
6. 指出每个关系表的度
7. 指出每个关系表的基数

练习-3：术语

清華大學 Tsinghua University

Accounts

AcctNo	Type	Balance
12345	savings	12000
23456	checking	1000
34567	savings	25

Customers

fname	lname	idNo	account
Robbie	Banks	901-222	12345
Lena	Hand	805-333	12345
Lena	Hand	805-333	23456

8. 指出每个关系的模式
9. 指出数据库模式
10. 为每个属性指定合适的域

练习-4：术语

Customers

fname	lname	idNo	account
Robbie	Banks	901-222	12345
Lena	Hand	805-333	12345
Lena	Hand	805-333	23456

11.给出该关系表的另一种等价表示

12.该关系表有多少种不同的表示方式?

键（Key）

清華大學 Tsinghua University

键 = 能够唯一标识一条记录的一个（或多个）属性

AcctNo	Type	Balance
12345	savings	12000
23456	checking	1000
34567	savings	25

键（Key）

清華大學 Tsinghua University

键 = 能够唯一标识一条记录的一个（或多个）属性

键	非键	非键
AcctNo	Type	Balance
12345	savings	12000
23456	checking	1000
34567	savings	25

多属性键

清華大學 Tsinghua University

多属性键 = 能够唯一标识一条记录的多个属性的组合

Key = fname, lname

fname	lname	age	salary
Robbie	Banks	20	10k
Alice	Banks	30	8k
Alice	Smith	25	12k

多个键

键

另一个键

SIN	fname	lname	age	salary
123-456-789	Robbie	Banks	20	10k
222-111-709	Alice	Banks	30	8k
345-498-712	Alice	Smith	25	12k

- 可以选择一个键作为主键（primary key，例如 SSN）

外键（Foreign Key）

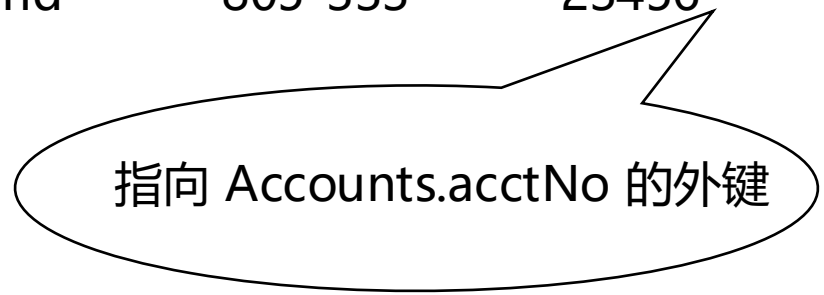
属性（或多个属性）的取值是另一关系中某条记录的键

Accounts

acctNo	type	balance
12345	savings	12000
23456	checking	1000
34567	savings	25

Customers

fname	lname	idNo	account
Robbie	Banks	901-222	12345
Lena	Hand	805-333	12345
Lena	Hand	805-333	23456



讨论

清華大學 Tsinghua University

表是无序的

- 它们是集合（set）或多重集合（multiset, bag）

表并不规定它们在磁盘上如何实现与存储

- 这被称为物理数据独立性

课程大纲

- 数据模型概述
- 关系模型基础
- **使用 SQL 定义关系模式**

为什么要学习SQL?

清華大學 Tsinghua University

2000年代：SQL的低谷

- 关系型数据库已经消亡了吗?

2010年代：SQL又回来了

- Pig, Hive, Impala
- SparkSQL

NoSQL

- Non SQL
- Not-Only-SQL
- Not-Yet-SQL

SQL简介

清華大學 Tsinghua University

SQL 表示结构化查询语言（Structured Query Language）

读音：读作“S Q L”或者 “sequel”

普适性：受到所有主流商业数据库系统的支持

标准化：标准不断演进，持续引入大量新特性

语言类型：声明式语言

SQL 语言的两大核心组成

清華大學 Tsinghua University

数据定义语言（DDL）

- 定义关系模式 (*Schema*)
- 对数据表及其字段属性进行创建 、 修改和删除

数据操纵语言 (DML)

- 插入、删除或修改表中的行
- 查询单表或多表数据

创建表

清華大學 Tsinghua University

使用 CREATE TABLE 语句创建表

- 指定表名、字段名及其域（数据类型）

```
CREATE TABLE Customer (
    sin             CHAR(11),
    firstName       CHAR(20),
    lastName        CHAR(20),
    age             INTEGER,
    income          REAL)
```

问题：SQL 区分大小写吗?

- 答：SQL 的关键字（例如 create 和 table）不区分大小写。
- 命名对象（表、列等）可能区分大小写。

删除表

使用 DROP TABLE 语句删除表

- 不仅会删除所有记录，还会删除表的模式

```
DROP TABLE Customer
```

修改表

清華大學 Tsinghua University

使用 ALTER TABLE 语句可以为表添加或删除列

- ADD 用于添加列
- DROP 用于删除列

```
ALTER TABLE Customer
ADD height INTEGER
```

```
ALTER TABLE Customer
DROP height
```

插入记录

清華大學 Tsinghua University

使用 INSERT 语句向已有表中插入记录

- 列名列表是可选的
- 若省略，则值必须按与列相同的顺序给出

```
INSERT INTO Customer(sin, firstName, lastName, age, income)
VALUES ('111', 'Sam', 'Spade', 23, 65234)
```

删除记录

清華大學 Tsinghua University

使用 DELETE 语句删除记录

- WHERE 子句指定要删除的记录

```
DELETE
FROM Customer
WHERE sin = '111'
```

- 注意：下面这条 SQL 查询会删除表中所有记录

```
DELETE
FROM Customer
```

修改记录

使用 UPDATE 语句修改表中的一条或多条记录

- 注意：WHERE 子句会在 SET 子句之前被求值
- 与 DELETE 类似，WHERE 子句指定需要更新哪些记录

```
UPDATE Customer
SET age = 37
WHERE sin = '111'
```