



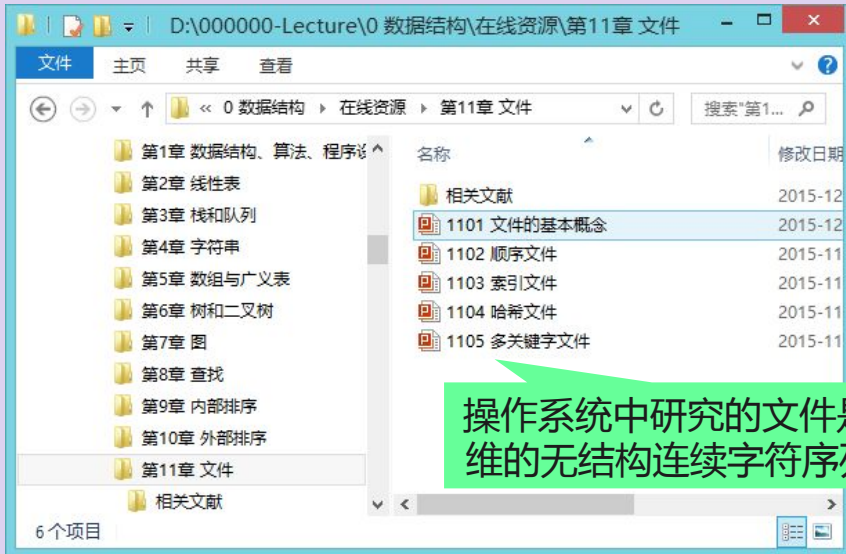
本节主题：

文件的基本概念

# 文件

- 文件是性质相同的记录的集合。
- 文件的数据量通常很大，它被放置在外存上。
- 区分
  - 数据库意义上的文件
  - 操作系统意义上的文件

学号	姓名	性别	民族	班级
1	李明	男	汉族	99101
5	王平	男	汉族	99101
3	张萍	女	满族	99102
8	陈强	男	汉族	99101
4	马伟	男	回族	99102



数据库中所研究的文件是带有结构的记录集合



# 相关术语

❏ **记录**是文件中存取的基本单位

❏ **数据项**是文件可使用的最小单位

❏ 数据项有时也称为**字段**。

❏ 关键字项/关键字

❏ 其值能惟一标识一个记录的数据项或数据项的组合称为**主关键字项**

❏ 其他不能惟一标识一个记录的数据项则称为**次关键字项**

❏ 主关键字项的值称为**主关键字**。

❏ 次关键字项的值称为**次关键字**。

❏ 约定：不严格区分**关键字项**和**关键字**

❏ 在不易混淆时,将主(或次)关键字项简称为主(或次)关键字

❏ 假定主关键字项只含一个数据项

## 关键字

☺ 从唯一标识记录的角度

☺ 从查找角度

学号	姓名	性别	民族	班级
1	李明	男	汉族	99101
5	王平	男	汉族	99101
3	张萍	女	满族	99102
8	陈强	男	汉族	99101
4	马伟	男	回族	99102

学号	课号	成绩
1	c01	89
5	c01	78
1	c02	91
3	c03	63
3	c01	72

# 文件分类

## □ 按记录中关键字的多少

### 📁 单关键字文件

- 📄 若文件中的记录只有一个惟一标识记录的主关键字，则称其为**单关键字文件**

### 📁 多关键字文件

- 📄 若文件中的记录除了含有一个主关键字外，还含有若干个次关键字，则称为**多关键字文件**。

## □ 按记录的长度是否确定

### 📁 定长文件

- 📄 若文件中记录含有的信息长度相同，则称这类记录为**定长记录**，由这种定长记录组成的文件称做**定长文件**

### 📁 不定长文件

- 📄 若文件中记录含有的信息长度不等，则称做**不定长文件**

学号	姓名	性别	民族	班级	简历
1	李明	男	汉族	99101	...
5	王平	男	汉族	99101	...
3	张萍	女	满族	99102	...
8	陈强	男	汉族	99101	...
4	马伟	男	回族	99102	...

# 文件的结构

## 逻辑结构

在用户和程序员面前呈现的记录的形式，是用户对数据的表示方式

学号	姓名	性别	民族	班级
1	李明	男	汉族	99101
5	王平	男	汉族	99101
3	张萍	女	满族	99102
8	陈强	男	汉族	99101
4	马伟	男	回族	99102

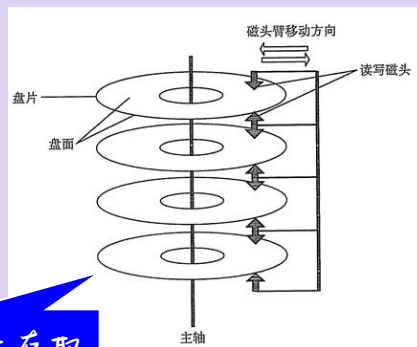
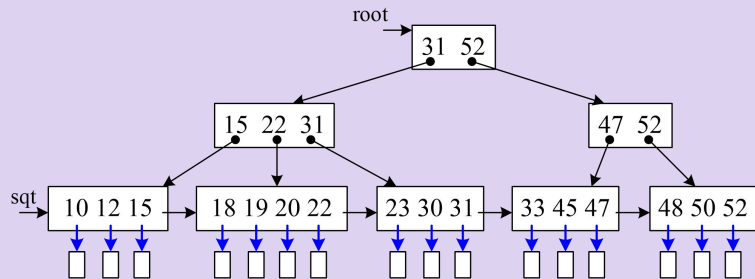
着眼用户使用方便。

研究数据本身的关联，确定要实施的操作。

以“块”为单位存取

## 物理结构

数据在物理存储器上存储的方式，是数据的物理表示和组织。



考虑提高存储空间的利用率，以及减少存取记录的时间。

根据不同需要，以及设备的特征，选择合适的结构。

# 文件的逻辑结构及操作

## □ 文件可看成是一种线性结构

📁 一个文件的各个记录按照某种次序排列，记录之间就自然地形成了一种线性关系。

📄 可以按记录中关键字的大小为序

📄 也可以按各个记录存入该文件的时间先后为序

## 📁 线性结构的特点

📄 文件中每个记录最多只有一个后继记录和一个前驱记录

📄 文件的第一个记录只有后继没有前驱

📄 文件的最后一个记录只有前驱而没有后继

## □ 文件上的操作

### 📁 检索/查找

📄 按记录的逻辑号/按关键字

### 📁 维护

📄 插入/删除/修改

📄 文件的再组织/备份/恢复/安全保护

学号	姓名	性别	民族	班级
1	李明	男	汉族	99101
5	王平	男	汉族	99101
3	张萍	女	满族	99102
8	陈强	男	汉族	99101
4	马伟	男	回族	99102

# 文件的物理/存储结构

- 文件的存储结构是指文件在外存上的组织方式。
- 采用不同的组织方式就得到不同的存储结构。

## 基本的存储结构

顺序组织  
索引组织  
哈希组织  
链组织

基本存储结构的结合

取决于对文件中记录的使用方式和频繁程度、存取要求、外存的性质和容量。

## 文件组织方式

顺序文件  
索引文件  
哈希文件  
多关键字文件

接下来的话题,在各种存储结构下,基本操作的实现.....