



本节主题:

索引文件

索引文件

索引文件组织方法

- 数据文件/主文件：用于存储数据，以记录为单位
- 索引表：另外建立的一张表，用于指明逻辑记录和物理记录之间的一一对应关系。
- 索引表和主文件一起构成的文件称作索引文件。

索引表

	关键字	物理地址
[1000]	1	100
[1010]	3	300
[1020]	4	500
[1030]	5	200
[1040]	8	400

	学号	姓名	性别	民族	班级
[100]	1	李明	男	汉族	99101
[200]	5	王平	男	汉族	99101
[300]	3	张萍	女	满族	99102
[400]	8	陈强	男	汉族	99101
[500]	4	马伟	男	回族	99102

主文件

索引表

- 索引表中的每一条记录称作索引项
- 一般索引项都是由主关键字和该关键字所在记录的物理地址组成的。
- 索引表必须按主关键字有序
- 主文件本身则可以按主关键字有序或无序
 - 索引顺序文件
 - 索引非顺序文件

索引表的建立和更新

建立

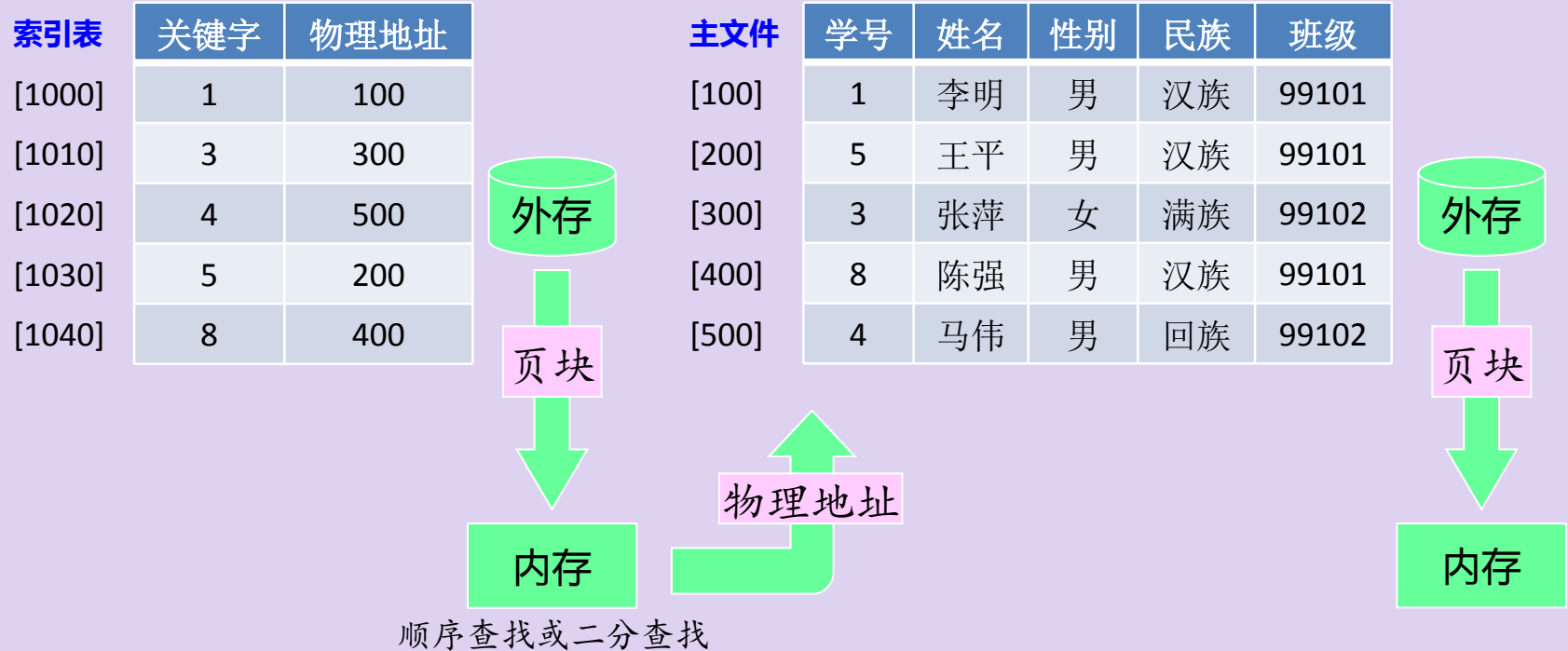
- 先按输入记录的先后次序建立主文件和索引表（索引表中关键字是无序的）
- 待全部记录输入完毕后再对索引表进行排序

	学号	姓名	其他		关键字	物理地址		关键字	物理地址		
[100]	1	李明		➡	[1000]	1	100	➡	[1000]	1	100
[200]	5	王平			[1010]	5	200		[1010]	3	300
[300]	3	张萍			[1020]	3	300		[1020]	4	500
[400]	8	陈强			[1030]	8	400		[1030]	5	200
[500]	4	马伟			[1040]	4	500		[1040]	8	400

更新

- 插入：“记录”插入在主文件的末尾，而相应的“索引项”必须插入在索引的合适位置上；
- 删除：需删除索引表中相应的索引项，在主文件中做标记

检索过程



顺序查找或二分查找

多级索引

适用：当数据量特别大，以致于索引表在一个页块中放不下时

查找表	关键字	物理地址
[200000]	9011278	100000
[200010]	9112827	100900

查找表：对索引表建立的索引。

索引表1	关键字	物理地址
[100000]	9010101	100
[100010]	9010103	300
[100020]	9010104	500
[100030]	9010105	200
[100040]	9010108	400
[100850]	9011278	8600

索引表2	关键字	物理地址
[100900]	9110011	8700
[100010]		

主文件	学号	姓名	其他
[100]	9010101	李明	
[200]	9010105	王平	
[300]	9010103	张萍	
[400]	9010108	陈强	
[500]	9010104	马伟	
[8600]	9011278	秦奋	
[8700]	9110011	施恩	
...			

稠密索引和稀疏索引

- 对于索引非顺序文件，由于主文件中记录是无序的，则必须为每个记录建立一个索引项，这样建立的索引表称为**稠密索引**。
- 对于索引顺序文件，由于主文件中记录按关键字有序，则可对一组记录建立一个索引项，例如，让文件中每个页块对应一个索引项，这种索引表称之为**稀疏索引**。

稀疏索引表

	关键字	物理地址
[200000]	9011278	100
[200010]	9112827	10000

主文件

	学号	姓名	其他
[100]	9010101	李明	
[200]	9010103	张萍	
[300]	9010104	马伟	
[400]	9010105	王平	
[500]	9010108	陈强	
[8600]	9011278	秦奋	
[10000]	9110011	施恩	
...			

索引文件的检索方式

❏ 直接存取

📁 数据文件独立，但单独操作可能会破坏索引，慎用

❏ 按关键字存取

📁 先查索引

📁 然后根据索引中指针所指索取记录；

索引非顺序文件

☺ 适合随机存取

☹ 不适合按关键字的顺序存取：会引起磁头的频繁移动

索引顺序文件

☺ 既适合随机存取

☺ 又适合按关键字的顺序存取

☹ 保持顺序不容易