

哈尔滨工业大学
HARBIN INSTITUTE OF TECHNOLOGY

数据库系统（下）：管理与技术 国家精品

申请认证证书

评价课程

战德臣、张丽杰、周丽娜、史建森



- 公告
- 评分标准
- 课件
- 测验与作业
- 考试
- 讨论区

课程分享   



 微信提醒课程进度

 扫码下载 APP

帮助中心

课件 > 第4讲（总第20讲）数... > 第20讲模拟练习题

向老师提问

1

已知内存共有8块，若要排序有70块的数据集，应如何组织，才能使磁盘读写次数最少。下列方案中磁盘读写次数最少的方案是_____。

A. 方案I：(1)以8块为一个单位划分子集合，每个子集合进行内排序并存储，形成9个已排序子集合（其中包含一个仅有6块的子集合）；(2)接着在9个子集合中选择3个子集合（其中包含仅有6块的子集合），进行一个三路归并，形成一个已排序子集合；(3)再将剩余5个子集合与刚才归并后形成的子集合，进行一个七路归并，形成最终的已排序集合。这个方案的磁盘读写次数最少。

B. 方案II：(1)以7块为一个单位划分子集合，每个子集合进行内排序并存储，形成10个已排序子集合；(2)接着在10个子集合中任选5个子集合进行一个五路归并，形成一个已排序子集合；(3)再将剩余5个子集合与刚才归并后形成的子集合，进行一个六路归并，形成最终的已排序集合。这个方案的磁盘读写次数最少。

C. 方案III：(1)以8块为一个单位划分子集合，每个子集合进行内排序并存储，形成9个已排序子集合（其中包含一个仅有6块的子集合）；(2)接着在9个子集合中任选七个子集合进行一个七路归并，形成一个已排序子集合；(3)再将剩余2个子集合与刚才归并后形成的子集合，进行一个三路归并，形成最终的已排序集合。这个方案的磁盘读写次数最少。

D. 方案IV：(1)以8块为一个单位划分子集合，每个子集合进行内排序并存储，形成9个已排序子集合；(2)接着在9个子集合中任选5个子集合进行一个五路归并，形成一个已排序子集合；(3)再将剩余4个子集合与刚才归并后形成的子集合，进行一个五路归并，形成最终的已排序集合。这个方案的磁盘读写次数最少。

A

2

已知内存共有100块，若要排序有10000块的数据集，则下列说法正确的是_____。

A. 该数据集可以在两趟内实现排序，磁盘读写次数为40000次。

B. 该数据集不能在两趟内实现排序，磁盘读写次数为40400次。

C. 该数据集可以在两趟内实现排序，磁盘读写次数为40400次。

D. 该数据集不能在两趟内实现排序，磁盘读写次数为60000次。

B

3

已知内存共有8块，若要排序有100块的数据集，则给定多路归并算法如下：(1)以8块为一个单位划分子集合，每个子集合进行内排序并存储，形成13个已排序子集合(含一个仅有4块的子集合)；(2)接着在13个子集合中任选7个子集合（包含仅有4块的子集合）进行一个七路归并，形成一个已排序子集合；(3)再将剩余6个子集合与刚才归并后形成的子集合，进行一个七路归并，形成最终的已排序集合。问：这个方案的磁盘读写次数是_____。

A. 200

B. 400

C. 504

D. 512

C

4 关于基于排序的两趟算法，下列说法不正确的是_____。

- A. 基于排序的两趟算法的第一趟都是划分子表并排序。每一个子表应都能装入内存，并进行排序，然后再存回磁盘。
- B. 基于排序的两趟算法的第二趟是进行归并，在归并的过程中可以边排序边去重复，归并完成即去重复操作完成。
- C. 基于排序的两趟算法的第二趟是进行归并，在归并的过程中可以一边排序一边进行分组并进行聚集计算，归并完成即分组聚集计算操作完成。
- D. 基于排序的两趟算法的第二趟是进行归并，先归并处理第一个关系的元组，边归并边去重复，归并处理完成后，再归并处理第二个关系的元组，边归并边去重复，归并处理完成后即完成两个关系的集合并操作。

D

5 已知关系R和S。关系占用的磁盘块数 $B(R)=1000$ ， $B(S)=1000$ ，已知可用内存页数 $M=40$ 。采用基于排序的算法，下列说法正确的是_____。

- A. 用一趟算法即可实现R和S的集合并操作。
- B. 用两趟算法即可实现R和S的集合并操作。
- C. 用一趟算法即可实现R和S的包的并操作。
- D. 必须用两趟算法才能实现R和S的包的并操作。

C

6 已知关系R和S。关系占用的磁盘块数 $B(R)=1000$ ， $B(S)=500$ ，已知可用内存页数 $M=50$ 。采用基于排序的算法，下列说法正确的是_____。

- A. 用一趟算法即可实现R和S的集合并操作。
- B. 用两趟算法才能实现R和S的集合并操作。
- C. 用一趟算法即可实现R和S的集合交操作。
- D. 用两趟算法才能实现R和S的包的并操作。

B

7 关于基于散列的两趟算法，下列说法不正确的是_____。

- A. 基于散列的两趟算法的第一趟是散列子表。用某一个散列函数，将具有相同散列值的元组散列到相同的子表中并存回磁盘。
- B. 基于散列的两趟算法的第二趟是用与第一趟相同的散列函数，将子表再散列到内存的不同内存块中，在具有相同散列值的所有内存块中去重复，即是在整个关系上去重复。所有子表处理完成，去重复操作即告完成。
- C. 基于散列的两趟算法的第二趟是用与第一趟不同的散列函数，将子表再散列到内存的不同内存块中，在具有相同散列值的所有内存块中去重复，即是在整个关系上去重复。所有子表处理完成，去重复操作即告完成。
- D. 两次散列函数的选择是不同的，第一趟是在大范围上进行散列，将一个大数据集散列成若干个具有相同散列值的散列子表，第二趟是在小范围上进行散列，将具有某相同散列值的散列子表(大范围上散列值相等)散列到内存的某一块或几块(小范围上散列值相等)。

B

- 8 基于散列的两趟算法和基于排序的两趟算法，其中第一趟都是划分子表，都要求子表的存储块数要小于可用内存数，以便子表可以一次性装入内存进行处理。关于划分子表，下列说法正确的是_____。
- A. 基于排序的算法总是可以均匀地划分子表(即每个子表的大小都一样，除最后一块外)；基于散列的算法总是可以均匀地划分子表(即每个子表的大小都一样，除最后一块外)。
 - B. 基于排序的算法不能保证总是均匀地划分子表；基于散列的算法总是可以均匀地划分子表(即每个子表的大小都一样，除最后一块外)。
 - C. 基于排序的算法总是可以均匀地划分子表(即每个子表的大小都一样，除最后一块外)；基于散列的算法不能保证总是均匀地划分子表。
 - D. 基于排序的算法不能保证总是均匀地划分子表；基于散列的算法不能保证总是均匀地划分子表。

C

- 9 关于R与S的并、交、差运算的基于散列的两趟算法，其中第一趟都是划分子表，都要求子表的存储块数要小于可用内存块数，以便子表可以一次性装入内存进行处理。关于划分子表，下列说法正确的是_____。
- A. 必须用相同的散列函数将R和S分别散列成若干个子表。
 - B. 必须用不同的散列函数将R和S分别散列成若干个子表。
 - C. 可以用相同的散列函数，也可以用不同的散列函数，将R和S分别散列成若干子表。
 - D. 划分子表根本不用散列函数。

A

- 10 关于基于散列的两趟算法和基于排序的两趟算法的基本思想，下列说法正确的是_____。
- A. 排序算法是先划分子表，独立处理子表（第一趟），然后再对各子表进行关联性处理（第二趟）；散列算法是先划分子表，独立处理子表（第一趟），然后再对各子表进行关联性处理（第二趟）；
 - B. 排序算法是先划分子表，独立处理子表（第一趟），然后再对各子表进行关联性处理（第二趟）；散列算法是先从关联性角度处理，形成子表（第一趟），然后再独立处理每一个子表（第二趟）。
 - C. 排序算法是先从关联性角度处理，形成子表（第一趟），然后再独立处理每一个子表（第二趟）。散列算法是先划分子表，独立处理子表（第一趟），然后再对各子表进行关联性处理（第二趟）；
 - D. 排序算法是先从关联性角度处理，形成子表（第一趟），然后再独立处理每一个子表（第二趟）；散列算法是先从关联性角度处理，形成子表（第一趟），然后再独立处理每一个子表（第二趟）。

B

- 11 关于连接运算R (JOIN on R.A=S.B) S的基于散列的两趟算法，下列说法不正确的是_____。
- A. 必须以相同的散列函数分别散列R和S，形成若干个散列子表。
 - B. 散列过程中，R必须以A属性值作为散列函数的键值，S必须以B属性值作为散列函数的键值。
 - C. 散列过程中，R必须以A和B属性值作为散列函数的键值，S也必须以A和B属性值作为散列函数的键值。

- D. 第二趟处理中，须将R的子表再完整地散列到内存的若干块中，然后再一块一块处理S对应子表的每一块，以便快速决定可以连接的元组。

C

12 关于基于散列的两趟算法，下列说法正确的是_____。

- A. 第一趟散列和第二趟散列的目的都是提高数据处理的速度。
- B. 第一趟散列的目的是提高数据处理的速度，而第二趟散列的目的是使数据子集具有某一种特性(如具有相同的散列值)。
- C. 第一趟散列的目的是使数据子集具有某一种特性(如具有相同的散列值)，而第二趟散列的目的是提高数据处理的速度。
- D. 第一趟散列和第二趟散列的目的都是使数据子集具有某一种特性。

C

重做