

8.2 解:

(1) 文件创建、文件删除、记录插入、记录删除、扫描。

(2) 索引是磁盘中组织数据记录的一种数据结构，用于优化某类数据检索的操作；搜索码是存储在索引中用来检索的字段的集合；索引使得我们可以有效地检索满足索引的搜索码字段上的搜索条件上的那些记录。

(3) ①数据项 k^* 是一个真正的数据记录；

②数据项是一个对，其中rid是搜索码值为k的数据记录的记录id；

③数据项是一个对，其中rid-list是搜索码值为k的数据记录的记录id的列表。

(4) 主索引的字段集合包含主码，次索引的字段集合不包含主码；索引中的重复数据项是指具有相同索引搜索码的数据项；主索引的字段集合包含主码，主码能够惟一区别字段，所以主码不可以重复。

(5) 聚簇索引中数据记录的顺序与索引的数据项顺序相同或类似，非聚簇索引数据记录的顺序与索引的数据项顺序不相同；如果一个索引将数据记录作为数据项，它的数据记录的顺序与数据项顺序相同，一定是聚簇索引。

(6) 一个文件上最多有一个聚簇索引，为一个文件是否创建一个聚簇索引取决于具体问题，维护一个聚簇索引的代价一般大于一个非聚簇索引。

(7) 索引数据项的第二种与第三种形式适合于次索引；第一种形式由于其数据项是真正的数据记录，是聚簇索引，而在实践中，数据更新时维护文件有序性的代价很高所以第一种形式不适合于次索引。

8.5 解:

哈希索引中，文件中的记录被分组存放在不同的桶中，其中一个桶由一个主页构成，或者是由一个主页和多个页构成的链组成。一个记录属于哪一个桶可以由一个哈希函数用于搜索码来决定。B+树索引中，记录的组织方式是使用树形的数据结构，数据项按照搜索码值进行排列，并且维护一个层次化的搜索数据结构，以便将搜索定向到数据项所属的页面。

关于等值搜索，哈希索引相较于B+树索引代价更低。如果我们检索文件中的一条记录，哈希索引的平均成本约为1.2个I/O，而B+树索引的平均成本约为3-4个I/O。关于范围搜索，B+树索引优于哈希索引。在范围搜索中，哈希索引需要扫描整个文件记录，B+树索引当检索到满足条件的第一条记录后即可顺序检索数据页。

8.6 解:

文件类型	扫描	等值搜索	范围搜索	插入	删除
堆文件	BD	0.5BD	BD	2D	Search+...
排序文件	BD	$D\log_2 B$	$D\log_2 B + \#$ matching pages	Search+BD	Search+E...
聚簇文件	1.5BD	$D\log_F 1.5B$	$D\log_F 1.5B$ + # matching pages	Search+D	Search+...
非聚簇树文件	$BDR + 0.15BD$	$D + D\log_F 0.15B$	$D\log_F 0.15B$ + # matching records	$3D + D\log_F 0.15B$	Search+2...
非聚簇哈希文件	$BDR + 0.125BD$	2D	BD	4D	Search+2...