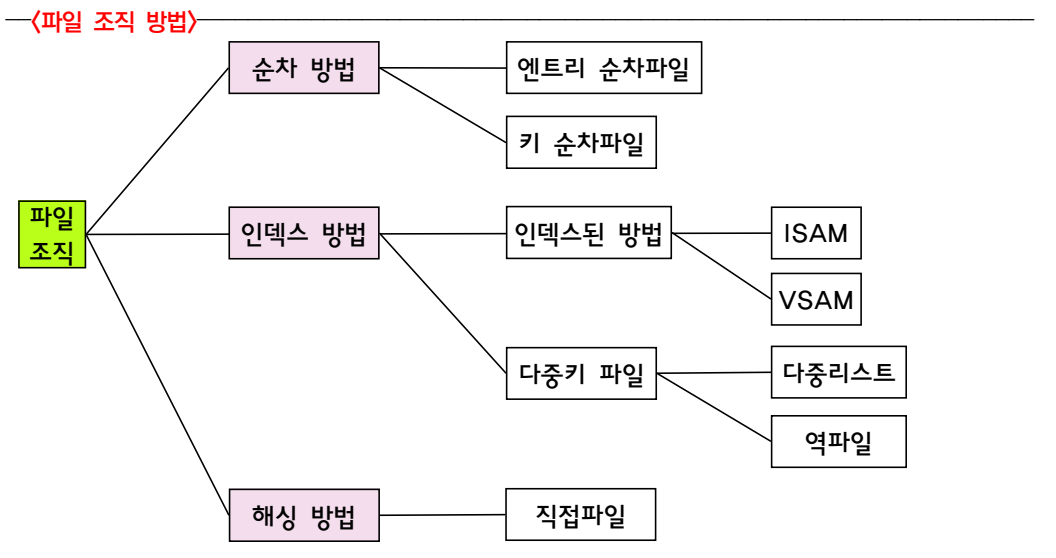


제7장 파일조직과 인덱스

궁극적으로 데이터베이스는 시스템 내부에서 파일(file)로 구현된다.



- 파일(file)은 디스크에 저장되는 단위이다.
- 파일(file)은 같은 타입(type)의 레코드 집합으로 구성된다.
- 파일조직은 파일의 레코드들을 물리적 저장장치에 저장시키기 위한 배치 방법이다.
- 데이터베이스시스템에서 파일조직은 데이터베이스의 물리적 저장 방법이 된다.

1. 순차파일

순차 방법(sequential method)은

레코드들의 물리적 순서와 논리적 순서가 같도록 순차적으로 저장하는 방법이다.

엔트리 순차파일	<ul style="list-style-type: none"> • 엔트리 순차파일(entry-sequence file)은 비순서파일이다. • 엔트리 순차파일은 레코드가 시스템에 입력되는 순서로 저장된다. • 엔트리 순차파일은 힙(heap) 또는 파일(pile)이라 한다. • 일반적으로, 새로 입력되는 레코드는 파일 끝에 저장된다. • 만약, 파일 중간에 빈 공간이 있으면 저장된다. • 레코드 검색은 모든 레코드들을 순차적으로 접근해야 한다. • 힙 파일은 좋은 성능 유지를 위해서는 주기적으로 재조직할 필요가 있다.
키 순차파일	<ul style="list-style-type: none"> • 키 순차파일(key-sequential file)은 레코드 키 순서로 저장된 파일이다. • 키 순차파일은 엔트리 순차파일이 키 순서로 가공된 구조이다. • 일반적으로, 순차파일이라고 하면 키 순차파일을 의미한다.

- 순차파일에서 레코드 접근은 레코드가 저장된 물리적 순서에 의한다.
- 순차파일에서 중간에 있는 특정 레코드를 접근하려면, 이전의 모든 레코드를 접근해야 한다.
- 해서, 순차파일은 레코드들을 일괄처리하는 응용에 많이 사용된다.
- 다음은 엔트리 순차파일(entry-sequence file)과 키 순차파일(key-sequential file) 모습이다.

학생			
학번	이름	학과	학년
3	강감찬	건축	4
6	이세영	토목	4
2	이순신	토목	3
5	홍하은	전산	3
1	홍재연	전산	4
4	김유신	전산	2

〈엔트리 순차파일〉

학생			
학번	이름	학과	학년
1	홍재연	전산	4
2	이순신	토목	3
3	강감찬	건축	4
4	김유신	전산	2
5	홍하은	전산	3
6	이세영	토목	4

〈키 순차파일〉



엔트리 순차파일은 각 학생의 자료가 시스템에 도착되는 순서로 저장된다.

해서, 학번 순서로 저장되지 않고 자료가 시스템에 도착되는 순서로 저장되어 있다.

기출문제 분석

1. 데이터베이스 저장구조에 대한 설명으로 옳지 않은 것은? [2014년 국가 7급]

- ① 힙 파일(heap file)은 파일에서 빈 공간이 있다면 레코드를 어느 위치에나 저장할 수 있다.
- ② 순차파일(sequential file)은 검색 키 값의 순서에 따라 정렬된 형태로 레코드들을 저장한다.
- ③ 해싱파일(hashing file)은 각 레코드의 애트리뷰트에 해싱함수를 적용한 후 레코드들을 정렬하여 저장한다.
- ④ 다중테이블 클러스터링 파일(multitable clustering file)은 서로 다른 릴레이션들의 관련 레코드들을 같은 블록에 저장한다.

☞ **데이터베이스 저장구조**

- 해싱 파일(hashing file)은 각 레코드의 애트리뷰트에 해싱함수를 적용한 후 레코드들을 정렬하여 저장한다.(×)
 - 해싱 파일은 해시함수를 적용한 후 계산된 값에 의거하여 레코드들을 랜덤하게 저장
 - 해싱 파일은 레코드들을 정렬하여 저장하지 않는다.
- 순차파일(sequential file)은 키 값의 순서에 따라 정렬된 형태로 레코드들을 저장한다.

// 힙 파일(heap file) - 비순서 파일

- 힙 파일은 가장 단순한 비순서 파일 조직이다.
 - 레코드들이 입력된 순서대로 시스템에 저장된다.
 - 힙 파일은 파일(pile) 또는 엔트리 순서 파일(entry-sequence file)을 의미한다.
- 일반적으로, 새로 입력되는 레코드는 파일 끝에 저장된다.
 - 만약, 파일 중간에 빈 공간이 있으면 저장된다.
- 레코드 검색은 모든 레코드들을 순차적으로 접근해야 한다.
- 힙 파일은 좋은 성능 유지를 위해서는 주기적으로 재조직할 필요가 있다.

// 다중테이블 클러스터링 파일(multitable clustering file)

- 다중테이블 클러스터링은 단위 클러스터에 두 개 이상의 테이블을 같이 저장하는 것이다.
- 그리고, 하나의 블록에는 여러 개의 단위 클러스터가 저장될 수 있다.
- 다중테이블 클러스터링은 조인하는 속도를 향상시키고자 할 때 주로 활용된다.

2. 테이블의 튜플 데이터를 파일 내에 저장하기 위한 구조는 힙(heap) 파일구조, 순차(sequential) 파일구조, 해시(hash) 파일구조로 구분될 수 있다. 일반적으로 테이블에 가장 빈번하게 이루어지는 연산에 의해 사용될 파일구조가 결정된다. 다음에 주어진 각각의 연산에 대하여 가장 효율적인 파일구조를 바르게 연결한 것은? [2017년 국가 7급]

ㄱ. 속성 값의 범위 구간을 주고 이에 해당하는 튜플 검색(range search)

ㄴ. 빈번한 튜플 삽입

ㄷ. 특정 속성 값과 일치하는 튜플 검색

- | | | |
|------|----|----|
| ㉠ | ㉡ | ㉢ |
| ① 순차 | 해시 | 힙 |
| ② 순차 | 힙 | 해시 |
| ③ 힙 | 순차 | 해시 |
| ④ 해시 | 힙 | 순차 |

☞ 효율적인 파일구조

ㄱ. 속성 값의 범위 구간을 주고 이에 해당하는 튜플 검색(range search) - 순차

→ 이유는, 속성 값의 범위 구간을 주고 순차 탐색하므로

→ 순차구조는 키를 기준으로 데이터가 정렬되어 있다.(범위 검색에 적절)

ㄴ. 빈번한 튜플 삽입 - 힙

→ 이유는, 파일 안에 빈 공간이 있으면 튜플은 어디든지 삽입 가능하다.

→ 참고로, 힙 파일구조와 자료구조 힙은 같은 개념이 아니다.

ㄷ. 특정 속성 값과 일치하는 튜플 검색 - 해시

→ 이유는, 해시구조에서 특정 값을 탐색하는 시간복잡도는 O(1)이다.

엔트리 순차파일	<ul style="list-style-type: none">• 엔트리 순차파일(entry-sequence file)은 비순서파일이다.• 엔트리 순차파일은 레코드가 시스템에 입력되는 순서로 저장된다.• 엔트리 순차파일은 힙(heap) 또는 파일(pile)이라 한다.• 일반적으로, 새로 입력되는 레코드는 파일 끝에 저장된다.• 만약, 파일 중간에 빈 공간이 있으면 저장된다.• 레코드 검색은 모든 레코드들을 순차적으로 접근해야 한다.• 힙 파일은 좋은 성능 유지를 위해서는 주기적으로 재조직할 필요가 있다.
키 순차파일	<ul style="list-style-type: none">• 키 순차파일(key-sequential file)은 레코드 키 순서로 저장된 파일이다.• 키 순차파일은 엔트리 순차파일이 키 순서로 가공된 구조이다.• 일반적으로, 순차파일이라고 하면 키 순차파일을 의미한다.