

제 7 장

- 7.1 어느 유통 회사에서는 전산화 된 판매 시스템을 가지고 있어서, 매일 고객에 대한 판매 기록을 입력하고 월말에는 두 개의 보고서를 출력한다. 하나는 각 고객에 대한 고객 번호, 이름, 주소, 지불할 금액을 출력하는 것이고, 다른 하나는 고객별 그 달 총 판매액을 출력하는 것이다. 이 판매 시스템에는 두 개의 화일이 있는데 하나는 고객 마스터 화일이고, 다른 하나는 일일 판매액을 기록하는 판매 화일이다.

(1) 각 화일의 구조를 설계하라.

답) 매일 고객의 판매기록 입력 작업시에는 고객에 대한 직접 접근이 가능해야 하며, 월말의 고객별 판매액 출력시에는 전체 고객에 대한 순차 접근이 행해진다. 따라서 고객 마스터 화일은 인덱스된 순차 화일의 구조로 구성하는 것이 효율적이다. 또, 월말에 판매 화일의 판매기록을 차례로 읽어 각 고객에 대한 판매액을 계산하므로 판매화일은 순차화일이 적당하다.

(2) 각 화일의 레코드 키는 무엇인가?

답) 고객 마스터 화일은 고객 번호를 레코드 키로 사용해야 하며, 판매 화일은 순차적으로 증가되는 판매번호를 레코드 키로 사용하는 것이 적당하다.

- 7.2 어느 소매상에서 재고 관리 시스템을 가지고 있다. 매일 매일 직원이 터미널을 통하여 그 날 판매된 각 상품의 수를 입력하면 상품화일의 각 상품재고량이 수정된다. 만일 특정 상품의 재고량이 지정된 양 이하로 줄어들면 상품을 주문하게 된다. 재고 관리 시스템의 출력은 다음과 같다.

- ① 직원이 그날 판매한 상품 정보를 입력한 후의 현재 재고량
- ② 각 상품의 일일 재고량에 대한 요약 보고서
- ③ 도매상의 이름, 주소, 날짜, 주문량, 상품 번호, 단위 가격 등을 포함한 주문서

(1) 상품 화일은 어떤 구조의 화일이어야 하는가?

답) 상품 화일은 특정 상품의 재고량을 알기 위해 그 상품 레코드에 대한 직접 접근이 가능해야 하며, 각 상품의 일일 재고량의 요약 보고서를 위해서 상품에 대한 순차 접근이 가능해야 한다. 따라서 인덱스된 순차 화일이 상품화일로 적절하다.

(2) 화일에서 레코드의 키는 무엇이어야 하는가?

답) 상품 화일의 레코드의 키로는 상품 번호가 되어야 한다.

7.3 어느 종합 병원에서 환자 기록 시스템을 가지고 있다. 각 간호원실에는 터미널이 있어서 환자가 입원이 되면 환자 개인 정보와 병력(medical history)이 포함된 레코드를 생성한다. 환자를 진료하는 의사는 진단 결과 및 처방을 기록한다. 간호원은 의사가 기록한 정보에 따라 투약하고 간호를 한다. 또한 환자의 체온, 혈압 등 각종 통계를 수정하여 나중에 담당 의사가 간단히 참조할 수 있도록 한다.

(1) 환자 화일은 어떤 구조로 구성하여야 하는가?

답) 환자 화일은 특정 환자에 대한 직접 접근이 수행될 수 있어야 하나, 환자들에 대한 순차 접근은 거의 수행되지 않는다. 따라서, 환자 화일은 직접 화일이 적당하다.

(2) 화일의 각 레코드 키는 무엇인가?

답) 환자 화일의 레코드 키로는 환자 번호가 되어야 한다.

7.4 인덱스된 순차 화일의 순차적 접근 과정, 특정 레코드의 직접 접근 과정, 그리고 하나의 레코드를 삽입하는 과정과 삭제하는 과정을 설명하여라.

답) 인덱스된 순차 화일은 크게 정적 인덱스 방법과 동적 인덱스 방법으로 나뉘고 그에 따라 위의 과정들이 달라진다.

정적 인덱스 방법은 기억 장소의 물리적 구조를 이용하여 화일을 구성하므로 화일의 순차적 탐색의 경우 기본 데이터 화일 구역의 실린더와 트랙을 순차적으로 탐색하면 된다. 이때 오버플로 구역을 고려하여야 한다. 특정 레코드 접근시 마스터 인덱스, 실린더 인덱스 및 트랙 인덱스를 통하여 기본 데이터 화일에 있는 레코드를 접근한다. 삽입은 기본적으로 해당 위치에 비어 있는 공간에 삽입하여야 하나 만약 비어있는 공간이 없는 경우

오버플로 구역을 이용한다. 오버플로 구역은 오버플로된 키와 구역내의 체인을 사용하여 관리한다. 삭제는 물리적으로 실제 삭제하는 방법과 단순히 삭제될 레코드에 '삭제' 표시만 하는 방법이 있다. 첫째 방법은 기본 데이터 구역일 경우에 엔트리 관리를 필요로 하고 오버플로 구역일 경우 체인의 관리등을 필요로 하고, 둘째 방법은 이후에 쓰레기 수집(garbage collection)을 행해야 할 필요가 있게 된다.

동적 인덱스 방법에서 데이터 레코드에 대한 순차 접근은 데이터 블록의 체인을 이용하여 데이터 블록에 대한 순차적 접근을 통하여 수행된다. 특정 레코드의 검색은 다단계로 구성된 인덱스를 이용하여 수행된다. 데이터 레코드의 삽입 및 삭제는 데이터 화일의 해당 위치에 수행되면 되지만, 이 경우 만일 삽입시에 삽입할 레코드를 저장할 공간이 없거나, 삭제 후 일정 수의 레코드를 유지하지 못할 경우 데이터 블록을 분할하거나 합병하고 이에 맞춰 인덱스 블록을 조정한다.

- 7.6 블록 크기가 4000바이트, 데이터 레코드 길이가 200바이트, 키값의 크기가 10바이트이고, 포인터 크기가 4바이트일 때, 100만 개의 데이터 레코드를 인덱스된 순차 화일로 구성하려 한다. 20퍼센트의 자유 공간을 고려하면 필요한 인덱스 단계수는 얼마인가?

답) 20퍼센트의 자유 공간을 고려하면 총 125만개($100만개 * 5/4$)의 데이터 레코드를 위한 공간이 필요하다. 데이터의 블로킹 인수는 20이고 125만개의 데이터 레코드를 위해서는 62500개의 인덱스 엔트리가 있어야 한다.

또 하나의 인덱스 엔트리 크기는 $14(10+4)$ 바이트이고, 블록의 크기가 4000바이트이므로 인덱스 블록의 분기율은 285가 된다. 따라서 62500개의 데이터 블록에 대한 첫번째 단계의 인덱스 엔트리를 위해 $219(=62500/285)$ 개의 인덱스 블록이 필요하고, 두번째 단계의 인덱스는 1개의 블록만 있으면 된다. 따라서 총 2단계의 인덱스로 구성할 수 있다.

- 7.7 순차 화일과 인덱스된 순차 화일의 공통점은 무엇인가? 인덱스된 순차 화일이 데이터를 접근하고 갱신하기 위해 가지고 있는 기능 가운데 순차 화일이 갖지 않은 것은 무엇인가 설명해 보라.

답) 순차 화일과 인덱스된 순차 화일의 공통점은 화일을 키값 순서대로 각 레코드들을 순차적으로 접근하는 것이 가능한 것이다. 그러나 인덱스된 순차 화일은 순차 화일이 가지고 있지 않은 기능인 특정 레코드에 대한 직접 접근이 가능하다는 장점이 있다.

7.8 인덱스된 순차 화일에서 레코드의 삽입이 왜 문제가 되는가를 설명하라. 또 이 삽입을 처리하는 두 가지 방법을 열거하고 설명해 보라.

답) 순차 데이터 화일에서 레코드의 삽입시 레코드의 순서를 유지하고 인덱스를 갱신해야 하며, 만약 오버플로를 처리할 수 있어야 한다. 이를 처리하는 데는 정적 인덱스 방법과 동적 인덱스 방법이 있다.

정적 인덱스 방법에서는 데이터 화일의 레코드가 삽입, 삭제시 인덱스의 내용은 변하지만 구조 자체는 변경되지 않는다. 만일 데이터 화일에서 새로 삽입될 레코드를 저장할 공간이 없으면 오버플로 구역(overflow area)을 사용한다. 동적 인덱스 방법은 인덱스나 데이터 화일은 블록으로 구성되고, 각 블록에는 나중에 레코드가 삽입될 것을 감안하여 빈 공간을 미리 준비해 두는 방법이다. 만약 어떤 블록이 가득차게 되면 동적으로 분열(split)되고 또 일정 수의 레코드를 유지하지 못하는 블록은 합병(merge)된다.

7.9 인덱스된 순차 화일의 인덱스 구조로 B-트리보다 B⁺-트리가 더 적절하다는 이유를 설명하라.

답) B⁺-트리는 B-트리에 없는 순차 세트를 가지고 있기 때문에 인덱스된 순차 화일의 기능에 필요한 키값에 의한 레코드들의 순차 접근이 가능하게 된다.

7.11 어떤 경우에 -트리 대신에 B-트리를 사용하는 것이 유리한가?

답) B-트리에서 노드의 분할, 합병 또는 재분배해야하는 연산때문에 오버헤드를 줄일때

7.12 어떤 경우에 B-트리 대신에 -트리를 사용하는 것이 유리한가?

답) 순차접근이 필요할때