

2020-02 CSE4095(파일처리)

Chapter3 homework

서강대학교 수학과 박세원 (20161229)

1. 어떤 디스크 드라이브는 512 바이트 크기의 섹터를 사용한다. 만약 한 프로그램이 128 바이트 크기의 레코드를 디스크로 쓸 것을 요청한다면, 파일 관리기는 레코드를 기록하기 전에 디스크로부터 한 섹터를 읽어야 한다. 그러한 불필요한 디스크 읽기의 횟수를 줄이기 위해 무엇을 할 수 있겠는가? 그 이유는 무엇인가?

디스크 드라이브가 512 byte 크기의 섹터를 사용하므로 디스크에 데이터를 쓸 때 512 byte 크기 만큼 써야한다. 따라서 128 byte 만큼 쓰려고 해도 나머지 384 byte에 대한 정보가 필요하기 때문에 디스크로부터 한 섹터를 읽어야 한다. 만약 파일에 순차적으로 접근한다면, 출력을 4개의 레코드로 묶어서 512 byte를 한번에 쓰는 방법을 이용해 이러한 불필요한 읽기를 줄일 수 있다.

2. 초기 UNIX 시스템에서, 최소한의 블록 크기는 하나의 클러스터의 크기로 512바이트였다. 그러나 4.BSD에서는 1,024바이트로 블록의 크기는 증가되었고, 작업량은 두 배 이상이 되었다. 이러한 것이 어떻게 발생할 수 있는지 설명하라.

파일의 크기가 512 바이트보다 클 경우, 최소 블록 크기가 1024일 때보다 512일 때가 블록의 수가 더 많아지고 탐색 시간도 늘어나게 된다. 또 최소 블록 크기가 작으면 파일이 여러 디스크에 나누어 저장되는 경우가 많아진다. 따라서 최소 블록의 크기를 1024로 할 경우 이러한 탐색 시간이 반 이상 줄어들어 작업량이 2배 이상 늘어날 수 있다.

3. IBM 3350 디스크 드라이브는 블록 주소화 방법을 사용한다. 다음의 두 서브 블록 구성이 사용 가능하다.

계수-데이터: 185 바이트에 해당하는, 계수 서브블록과 블록간 갭을 위해 사용되는 특별한 공간.

계수-키-데이터: 267 바이트와 키 크기의 합에 해당하는, 계수와 키 서브 블록과 갭을 위해 사용되는 특별한 공간.

IBM 3350은 트랙당 19,069의 바이트를 사용할 수 있고, 실린더마다 30개의 트랙이 존재하며, 드라이브마다 555개의 실린더가 있다. 350,000개의 80-바이트 레코드를 가지고 있는 파일을 3350 드라이브에 저장하고자 한다. 다음의 질문에 대한 답을 하라. 특별한 언급이 없으면, 블록킹 계수는 10이고, 계수-데이터 서브블록이 사용된다.

- a) 각 트랙마다 몇 개의 블록이 저장될 수 있는가? 몇 개의 레코드가 저장될 수 있는가?

한 블록 당 $80\text{byte} \times 10 = 800\text{byte}$ 의 레코드를 저장할 수 있고, 185 byte의 nondata overhead 공간이 필요하므로 985byte의 공간이 필요하다. 따라서 트랙당 $19069/985 = 19.36 = 19$ 개의 블록이 저장될 수 있다.

블록킹 계수가 10이므로 트랙당 $19 \times 10 = 190$ 개의 레코드가 저장될 수 있다.

b) 만약에 계수-키-데이터 서브블럭이 사용되고, 키 크기가 13 바이트라면 한 트랙당 몇 개의 블록이 저장될 수 있는가?

한 블록 당 800byte의 데이터 공간과 $267 + 13 = 280$ byte의 nondat overhead 공간이 필요하므로 총 1080byte의 공간이 필요하다. 따라서 트랙당 $19069/1080 = 17.66 = 17$ 개의 블록이 저장될 수 있다.

c) 계수-데이터 서브블럭을 가정할 때 저장 효율을 위한 효과적인 블록 크기를 보여주는 그래프를 그려 보아라. 저장 효율을 사용해서 최선과 최악의 블록킹 계수를 예측하는데 이 그래프를 사용하라.

각 블록킹 계수당 다음과 같은 과정으로 저장 효율을 구한다.

트랙당 레코드 개수 = $(19069/(\text{블록킹 계수} \times 80 + 185)) \times (\text{블록킹 계수})$

총 레코드 데이터의 byte = $80 \times (\text{트랙당 레코드 개수})$

저장 효율 = $(\text{총 레코드 데이터의 byte})/19069$

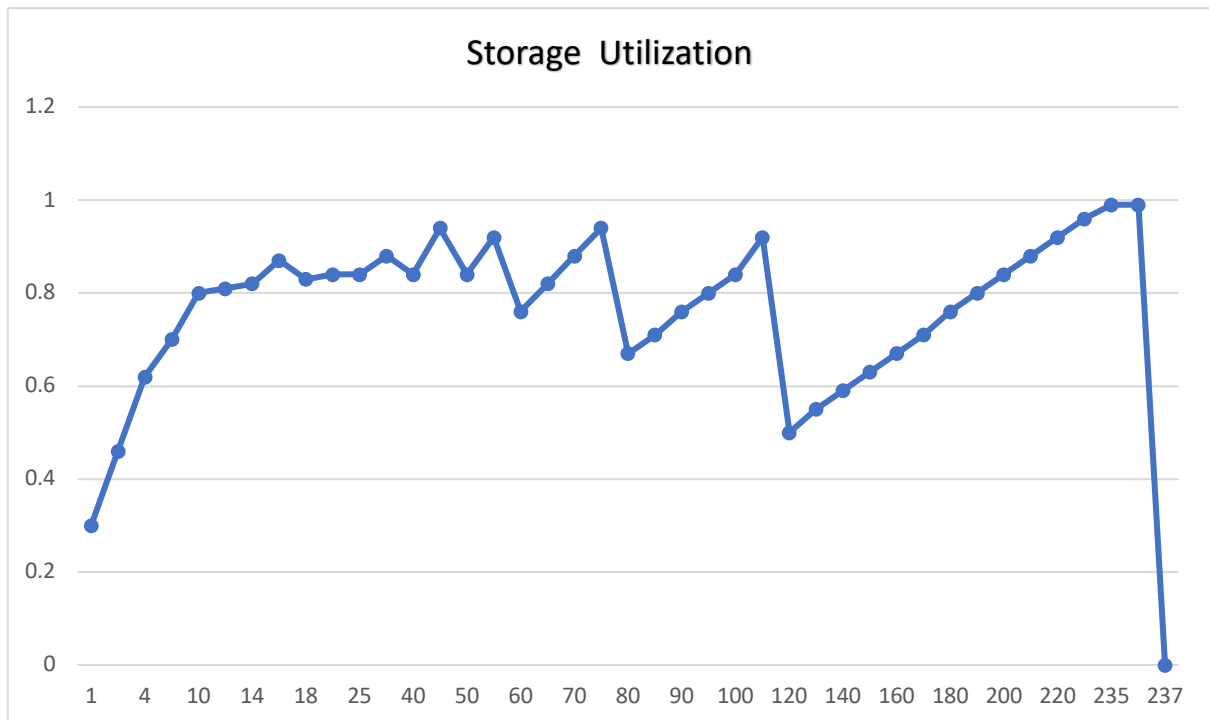


Table. Storage Utilization (x축: 블록킹 계수, y축: 저장 효율)

d) 파일이 항상 순차적으로 접근된다고 가정할 때, 이전의 문제에서 최선과 최악의 블록킹 계수를 각각 예측하라. 저장 효율과 처리시간을 이용하여 그 예측의 정당성을 밝혀라.

파일을 순차적으로 접근한다면 블록의 크기가 클수록 탐색 시간이 줄어들어 유리하다. 따라서 저장 효율이 가장 낮은 경우인 블록킹 계수가 1일 때가 최악이고, 저장 효율이 가장 높은 경우인 블록킹 계수가 236인 경우가 최선이다.

- e) 파일을 저장하기 위해서 얼마나 많은 실린더가 요구되는가? (블록킹 계수 10과 계수-데이터 포맷) 내부 트랙 단편화로 인해 사용되지 않은 많은 공간은 얼마나 되는가?

한 트랙당 190개의 레코드가 저장될 수 있으므로 한 실린더당 $190 \times 30 = 5700$ 개의 레코드가 저장될 수 있다. 따라서 350000개의 레코드를 저장하기 위해선 $350000/5700 = 61.4 = 62$ 개의 실린더가 필요하다.

한 트랙당 총 $19 \times 985 = 18715$ byte가 사용되므로 $19069 - 18715 = 354$ byte가 사용되지 않고 있다. 62개의 실린더를 사용하므로 총 $62 \times 30 \times 354 = 658440$ byte가 사용되지 않는다.

- f) 만약에 파일이 연속적인 실린더에 저장되고, 디스크 드라이브를 사용하는 다른 프로세서의 방해가 없다면, 파일에 임의적으로 접근할 때 평균 탐색시간은 12msec이다. 이것을 사용해서 한 레코드가 임의적으로 접근될 때 필요한 평균 시간을 계산하라.

하드 디스크의 경우 평균적으로 회전 지연은 6msec 정도이다.

전송 시간 = $(985/19069) \times 12 \text{ msec} = 0.62 \text{ msec}$

따라서 평균 접근 시간은 $12 + 6 + 0.62 = 18.62 \text{ msec}$

- g) 레코드의 임의 접근에 대한 탐색 시간이 어떻게 블록 크기에 영향을 받는지를 설명하라. 서로 다른 블록 크기가 사용될 때, 저장 효율과 탐색 사이의 상쇄효과에 대해서 설명하라. 설명을 위해 서로 다른 블록 크기를 갖는 테이블을 작성하라.

블록의 크기가 커지면 전송될 바이트의 수가 커지므로 전송 시간은 늘어나지만 사용하는 실린더의 개수가 줄어들어 탐색 시간이 짧아지고 회전 지연은 변하지 않는다.

blocking factor	cylinder	transfer time(비율)
1	106	0.04
2	107	0.05
10	62	0.04
180	65	0.76
235	50	0.98

Table. 블록 크기에 따른 실린더 개수와 전송 시간의 차이

4. 다루는 1,000,000개의 레코드로 된 우편 목록 파일에 대해서 고려해보자. 그 파일은 0.3-인치 블록간 갭을 가지는 6,250-bpi의 2,400-feet 릴 테이프에 백업될 것이다. 테이프의 속도는 초당 200인치이다.

- a) 오직 하나의 테이프가 블록킹 계수가 50인 경우 그 파일의 백업을 위해 사용되어진다는 것을 보여라.

레코드의 크기가 100byte라고 가정하면

데이터 블록의 물리적 길이는 $(50 \times 100)/6250 = 0.8 \text{ inch}$

데이터 블록의 수는 $1,000,000/50 = 20,000$

따라서 필요한 총 테이프의 길이는 $20,000 \times (0.8 + 0.3) = 22,000 \text{ inch} = 1833.333 \text{ feet}$ 이므로 하나의 테이프가 충분하다.

b) 블록킹 계수가 50이라면, 얼마나 많은 여분의 레코드가 2,400-feet 테이프에 대해 필요로 되는지는가?

$2400 - 1833.3 = 566.7 \text{ feet} = 6800 \text{ inch}$ 의 공간이 남으므로

$6800 / 1.1 = 6181.8 = 6181$ 개의 블록을 사용할 수 있고 따라서 $6181 * 50 = 309050$ 개의 레코드를 저장 가능하다.

c) 블록킹 계수가 50이 사용될 때, 유효 레코딩 밀도는 얼마인가?

$(\text{블럭당 바이트 수}) / (\text{블록을 저장하기 위해 요구되는 인치 수}) = 100 * 50 / 1.1 = 4545.45 \text{ inch}$

d) 블록킹 계수가 최대 유효 레코딩 밀도를 갖기 위해서는 얼마나 커야만 하는가? 블록킹 계수를 증가시키는 것으로 인해 어떤 문제가 발생하게 되는가?

하나의 갭만 존재할 때, 즉 블록킹 계수가 1,000,000일 때 최대 유효 레코딩 밀도를 갖게 된다. 이러한 블록을 포함하기에 충분히 큰 I/O 버퍼가 할당되어야 하고, 블록당 I/O 전송 시간이 증가하므로 전체 데이터가 전송 완료될 때 까지 시간이 많이 소요된다는 문제가 발생한다.

e) 그 테이프에 대해 앞의 파일을 저장하기 위해서 요구되는 최소 블록킹 계수는 얼마인가?

블록킹 계수가 x 일 때, 필요한 총 테이프의 길이는

$S = (1,000,000/x) * (0.3 + 100x/6250) \text{ inch}$ 이고, s 가 $2400 \text{ feet} = 28800 \text{ inch}$ 보다 작거나 같아야 하므로 x 는 24보다 크거나 같아야 한다. 따라서 최소 블록킹 계수는 24이다.

f) 블록킹 계수가 50일 때, 하나의 블록을 판독하는데 걸리는 시간은 갭까지 포함해서 얼마인가? 유효 전송율은 얼마인가? 완전히 파일을 판독하기 위해서는 얼마의 시간이 걸리는가?

한 블록의 길이는 1.1 inch 이므로, $1.1/200 = 0.0055 \text{ sec}$ 가 소요된다.

유효 전송율 = (유효 레코딩 밀도) * (테이프 속도) = $4545.45 * 200 = 909,090 \text{ bytes}$

전체 파일을 위해 필요한 블록의 개수는 20,000개 이므로 $20,000 * 0.0055 = 110 \text{ sec}$ 가 전체 파일 판독을 위해 소요된다.

g) 테이프의 역방향으로 판독하는 것이 불가능하다고 할 때, 앞의 파일에서 하나의 레코드에 대한 이진 탐색을 수행하는데 걸리는 시간은 얼마인가?

이진 탐색을 수행할 때 좌측 인덱스의 파일로 접근해야 하는 경우가 생기는데, 역방향 판독이 불가능하다면 테이프를 처음부터 다시 읽어야 한다. 즉 우측 인덱스의 파일로 접근하는 경우와 좌측 인덱스의 파일로 접근하는 경우의 시간 차이가 상당히 때문에 정확한 시간을 측정하기 어렵다.