

주문형 비디오 서버에서 효율적인 데이터 배치를 위한 고려 요소에 관한 연구

○안 유 정 원 유 헌
홍익대학교 컴퓨터공학과

A Study on Points to be Considered for Efficient Placement of Data in VOD Server

You-Jung Ahn Yoo-Hun Won
Department of Computer Engineering, Hong-Ik University

요 약

주문형 비디오 서버의 중요한 목표인, 동시에 보다 많은 클라이언트들의 요구를 만족시키기 위해서는 저장 매체 상에 비디오 데이터를 효율적으로 배치하는 것이 중요하다. 이와 같은 주문형 비디오 서버를 위한 배치 정책으로서 디스크 그룹화를 통한 배치정책이 소개되었고 실험을 통해 기존의 배치정책들 보다 검색 성능이 향상됨을 보였다[1]. 이 정책에서는 유사한 성능의 디스크들을 한 그룹으로 묶고 각 그룹마다 독립된 디스크 배열과 같이 취급하며 각 그룹의 성능을 고려하여 가장 적절한 비디오 데이터를 스트라이핑 시킨다. 이 정책에서 고려되어야 할 중요한 요소는 디스크들을 그룹화하는 과정에 있어서 각 그룹의 크기와 성능의 유사도를 어떻게 정할 것인가의 문제이다. 본 연구에서는 위와 같은 고려 요소들이 검색 성능의 향상에 미치는 영향에 대해 실험을 통해 밝히고자 한다.

1. 서 론

주문형 비디오(Video On Demand) 서버는 비디오 데이터를 저장 관리하는 시스템으로서, 비디오 데이터를 영상 데이터베이스에 저장하고 클라이언트의 요청이 있을 때 빠르게 검색하여 고속으로 원하는 부분을 제공하는 기능을 한다. 특히 주문형 비디오 시스템에서 비디오에 대한 사용자의 요구는 집중되는 경향이 있으며 인기 있는 비디오의 경우에는 여러 클라이언트가 동일한 비디오의 여러 다른 부분을 동시에 요구할 가능성이 높다. 따라서 동시에 보다 많은 클라이언트들의 요구를 만족시키기 위해서는 비디오 데이터들을 저장 매체 상에 효율적으로 배치시키는 것이 중요하다.

이와 같은 주문형 비디오 서버를 위한 배치 정책으로서 디스크 그룹화를 통한 배치 정책이 제안되었다[1]. 이 정책에서는 주문형 비디오 서버 내의 대형의 디스크 배열을 유사한 성능의 디스크들로 그룹화하고 각 그룹을 독립된 디스크 배열과 같이 취급하며 각 그룹의 성능에 가장 적합한 비디오 데이터를 스트라이핑시킨다. 이 정책을 주문형 비디오 서버에서 사용할 경우 임의의 시간에 특정 비디오에 대한 요구가 집중될 때 이로부터 발생하는 작업 부하를 여러 저장 매체에 분산시킬 수 있으며 나아가서는 오류 발생 시 신뢰성도 제공할 수 있다.

이 정책에서 특히 고려되어야 할 중요한 요소는 디스크들을 그룹화하는 과정에 있어서 각 그룹의 최적의 크

기와 한 그룹 내에 속하는 디스크들의 성능의 유사도를 어떻게 결정할 것인가이다. 또한 일부 디스크 배열에서는 이 크기와 유사도 사이에 상충하는 경우도 있으므로 이 값을 정하는 일은 쉽지 않다. 따라서 본 연구에서는 실험을 통해 다양한 디스크 배열 환경을 구성하고 각 환경마다 이 두 요소들의 관계를 밝혀보고자 한다.

본 논문의 구성을 살펴보면, 2장에서는 주문형 비디오 서버에서의 데이터 배치 정책들과 그들의 문제들에 대해 살펴보고 3장에서는 앞에서 언급한 고려 요소들이 검색 성능에 미치는 영향을 실험을 통해 측정하고 4장에서 결론을 맺는다.

2. 주문형 비디오 서버에서의 배치 정책

주문형 비디오 서버의 저장 시스템은 보통 수십 개 이상의 디스크들이 소프트웨어적으로 구성된 대형 디스크 배열의 형태로 존재한다.

기존 파일 시스템에서의 데이터 배치 방식은 임의의 데이터를 한 디스크 상에 순차적으로 배치시키는 형태이다. 그러나 멀티미디어 데이터는 대용량이며 또한 높은 전송 대역폭을 요구하므로 이와 같은 배치 방법으로는 정상적인 서비스 요구를 충족시킬 수 없다. 따라서 멀티미디어 서버에서는 데이터를 디스크 배열 내의 여러 디스크들에 같은 크기의 블록 단위로 분산 배치시키는 uniform striping 방식이 사용되어 왔다. 그러나 보다 빠르고 큰 용량의 디스크들이 등장함에 따라 저장

시스템 내에는 다양한 성능의 디스크들이 공존하게 되었으므로, uniform striping만을 사용할 경우 저장 시스템에서 전체 시스템의 검색 성능은 성능이 가장 낮은 디스크에 의해 제한될 수밖에 없다[2].

따라서 이러한 문제를 개선하기 위해, 디스크 배열 내의 디스크들을 유사한 성능별로 그룹화하여 그룹 내에서 uniform striping 방식으로 데이터를 배치하는 디스크 그룹화에 따른 배치 정책이 제안되었다[1]. 이 방식은 각 디스크들의 입출력 대역폭 성능을 충분히 이용할 수 있으므로 기존의 uniform striping 정책만을 사용할 때 보다 결과적으로 검색 성능이 향상된다. 그러나 다양한 저장 시스템들의 환경을 고려해볼때 검색 성능을 최대화할 수 있도록 디스크들을 그룹화하기는 쉽지 않다.

디스크 그룹화 작업 시 가장 고려해야 할 점은 각 디스크 그룹의 크기와 한 그룹으로 묶여질 디스크들의 성능 유사도이다. 이 중 디스크 그룹의 크기에 대해서는 이미 연구된 바가 있으며 그 결과는 그림 1과 같다[4]. 이 연구에서는 한 종류로 구성된 디스크 배열을 대상으로 디스크 배열의 크기에 따라 최적의 디스크 그룹 크기를 보이고 있다.

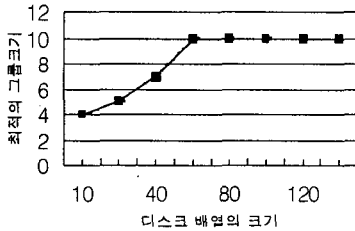


그림 1 디스크 배열의 크기에 따른 최적의 그룹 크기

디스크 그룹화 과정에서 또한 가지 고려 사항은 같은 그룹 내의 디스크들 성능의 유사도이다. 즉 다양한 성능의 디스크들이 존재할 때 검색 성능을 높이기 위해서는 한 그룹 내의 디스크들의 성능 차이를 얼마나 허용할 것인가의 문제이다. 그러나 그룹화 작업을 하다 보면 그룹의 크기와 성능의 유사도 사이에 서로 상충되는 경우도 있다. 따라서 본 논문에서는 실험을 통하여 이 문제에 접근하고자 한다.

3. 실험

3.1 실험 방법 및 환경

실험에 사용될 비디오 데이터는 MPEG-II로 압축된 영화 "terminator II"이다. 이 비디오 클립의 총 프레임 수는 52075개이고 4.6 Mbps의 전송 대역폭을 요구한다. 또한 비디오 데이터에 대한 요청 간격은 10초로 가정하고 스트라이핑 단위 크기는 100 KB로 한다.

실험에서는 다양한 저장 시스템의 환경을 설정하기 위해 표1에서 소개하고 있는 세 종류의 디스크들을 사용하며, 이 세 종류의 디스크들을 사용하여 크기가 12, 24, 36, 48, 60인 디스크 배열을 각각 구성하는데 자세한 내용은 표2와 같다.

표1. 실험에 이용된 디스크 종류

디스크 모델명	저장능력 (GB)	평균 탐색 시간 (ms)	평균 회전율 (ms)	전송률 (KB/ms)
ST31200W	1.001	12	7.5	2.33-4.17
Hawk 2XL	2.149	10	5.54	5.0-8.75
Ultrastar XP	4.51	8.5	4.17	9.59-12.58

표2. 실험에 사용된 디스크 배열

(S: ST31200W, H: Hawk 2XL, U: Ultrastar XP)

디스크 배열의 크기	디스크들의 구성
12	S 4개, H 4개, U 4개
24	S 8개, H 8개, U 8개
36	S 9개, H 18개, U 9개
48	S 21개, H 18개, U 9개
60	S 24개, H 24개, U 12개

표2에서와 같이 본 논문에서는 다양한 크기의 디스크 배열에 대해서 뿐만 아니라 디스크 배열 내에서의 성능별 구성비를 다양하게하여 실험하였다. 표1의 세 가지 디스크들인 ST31200W(이하 S), Hawk 2XL(이하 H) 그리고 Ultrastar XP(이하 U)의 성능을 비교해보면, S<H<U의 순인데, 표2에서 보면 디스크 배열의 크기가 12 또는 24일 경우 S:H:U의 구성비를 1:1:1로 동일하게 하였으며 크기가 36일때는 1:2:1로 48일때는 7:6:3 그리고 디스크 배열의 크기가 60일때는 2:2:1로 다양하게 구성하므로써, 다양한 디스크 배열의 환경에서 최적의 검색 성능을 내기 위해서는 어떻게 디스크 그룹을 구성해야 하는지를 측정하였다.

3.2 실험 결과

먼저 디스크 배열의 크기가 12일때 디스크들을 그룹화하여 검색 성능을 측정한다. 디스크 배열의 크기가 12일 때 [4]에서 제시한 최적의 그룹 크기를 살펴보면 그룹의 적정 크기는 4인데 표2에서 S, H, U 모두 4개씩이므로 간단히 세 그룹으로 그룹화된다.

디스크 배열의 크기가 24일 때 최적의 그룹 크기는 약 5 정도인데 현재 디스크들의 구성을 보면 S, H, U 모두 8개씩이므로 이들에 대해 다양하게 그룹화 한 결과는 표3과 같다. 표에서 검색 성능이란 초당 검색 가능한 비디오 스트림의 수를 수치화 한 것이다. 이 실험에서 가장 성능이 우수한 경우는 case 3으로서 이 경우 그룹 내에서 디스크들간의 성능 유사도는 100%이며 그룹의 크기는 최적의 크기에 80%에 해당한다. 그룹의 크기가 검색 성능에 미치는 영향을 보기 위해 case 2와 5를 비교해 본다. case 2와 5는 그룹화 정도가 거의 유사한데 그룹 2만 디스크 S와 H로된 혼합 디스크 그룹이며 이때 S:H의 구성비도 1:1로 같다. 그러나 차이점은 case 2의 경우 디스크 그룹의 크기가 6, 4, 6, 8로 최적의 크기에 각각 120%, 80%, 120%, 160%로써 다소 차이가 있는 편이지만 case 5의 경우는 case 2와 디스크 구성은 유사하나 각 그룹의 크기가 5, 6, 5, 4, 4로 최적의 크기의 80~120% 범위으로써 더 근접하며 그 결과 case 5가 검색 성능이 더 좋다. 다음으로 그룹 내에서 디스크들간의 유사도가 검색 성능에 미치는 영향을 보기 위한 예가 case 1과 2이다. 이 두 case 모두 각 디스크 그룹의 크

기는 8, 6, 4, 6으로서 같으나 각 case에서 혼합 그룹의 구성에 차이가 있다. case 1의 그룹 3은 디스크 H와 U의 혼합 그룹이고 case 2의 그룹 2는 S와 H의 혼합 그룹이다. 이때 디스크 S:H의 유사도는 56.85 : 100이고 H:U의 유사도는 64.92 : 100이므로 H와 U의 성능이 더 유사하다. 실험의 성능 비교에서도 보다 유사한 디스크들로 그룹화를 한 case 1의 검색 성능이 더 높게 나타나고 있다. 이와 거의 유사한 경우가 case 6과 7의 비교이다. S, H보다는 유사도가 더 높은 H, U 그룹에 더 많은 수의 디스크들을 배정시킨 case 7이 더 우수한 성능을 보인다.

표3. 디스크 배열의 크기=24일 경우의 디스크 그룹화 예 (8S: ST31200W 8개를 의미함)

구성 case 번호	각 case별 그룹 구성						case별 검색 성능
	그룹1	그룹2	그룹3	그룹4	그룹5	그룹6	
case 1	8S	6H	2H, 2U	6U			66.9476
case 2	6S	2S, 2H	6H	8U			58.8916
case 3	4S	4S	4H	4H	4U	4U	70.0093
case 4	4S	4S	5H	3H, 3U	5U		65.4475
case 5	5S	3S, 3H	5H	4U	4U		66.3596
case 6	5S	3S, 2H	5H	1H, 3U	5U		63.0246
case 7	5S	3S, 1H	5H	2H, 3U	5U		64.2358
...							...

이와 같은 방식으로 디스크 배열의 크기가 36일때와 48일때 디스크 그룹화 과정의 예 중 중요한 부분만을 표현한 것이 각각 표 4와 5이며, 모든 실험 결과를 종합하여 최적의 디스크 그룹을 구성한 결과가 표 6이다.

표4. 디스크 배열의 크기=36일 경우의 디스크 그룹화 예

구성 case 번호	각 case별 그룹 구성						case별 검색 성능
	그룹1	그룹2	그룹3	그룹4	그룹5	그룹6	
case 1	6S	3S, 5H	6H	6H	1H, 3U	6U	93.424
case 2	6S	3S, 4H	6H	6H	2H, 3U	6U	94.638
case 3	6S	3S, 3H	6H	6H	3H, 3U	6U	95.852
case 4	6S	3S, 2H	6H	6H	4H, 3U	6U	97.066
case 5	6S	3S, 1H	6H	6H	5H, 3U	6U	98.28
...							...

표5. 디스크 배열의 크기=48일 경우의 디스크 그룹화 예

구성 case 번호	각 case별 그룹 구성							case별 검색 성능
	그룹1	그룹2	그룹3	그룹4	그룹5	그룹6	그룹7	
case 1	8S	8S	5S, 2H	8H	8H	9U		120.829
case 2	7S	7S	7S	7H	7H	4H, 2U	7U	114.594
case 3	7S	7S	7S	9H	9H	9U		123.26
...								...

표6. 그룹화된 디스크 배열의 구성

디스크 배열의 크기	그룹의 최적크기	디스크들의 구성					
		그룹1	그룹2	그룹3	그룹4	그룹5	그룹6
12	4	4S	4H	4U			
24	5	4S	4S	4H	4H	4U	4U
36	6	6S	3S, 1H	6H	6H	5H, 3U	6U
48	8	7S	7S	7S	9H	9H	9U
60	10	12S	12S	12H	12H	12U	

이 실험의 결과 디스크 그룹의 크기가 최적의 크기의 80~120% 범위에서 좋은 검색 성능을 보이며 그룹 크기가 같을 때는 성능간의 유사도가 높을수록 전체 검색 성능도 높다. 만일 성능이 상이한 둘 이상의 디스크들로

구성된 혼합 그룹이 여럿일 경우에는 유사도가 더 높은 그룹의 디스크 수가 더 많은 것이 높은 성능을 보인다.

표6에서 보인 디스크 그룹들을 통해 디스크 그룹화를 통한 배치 정책을 사용한 경우와 uniform striping을 사용한 경우의 검색 성능을 측정된 결과가 그림 2이다.

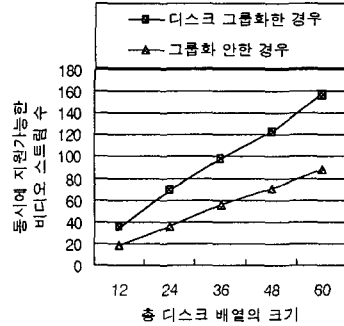


그림 2. 배치 정책에 따른 검색 성능의 비교

4. 결 론

본 연구에서는 주문형 비디오 서버에 적합한 배치 정책인 디스크 그룹화를 통한 배치 정책에서 디스크 그룹화 작업 시 그룹의 크기와 성능의 유사도가 전체 저장 시스템의 검색 성능에 어떤 영향을 미치는지를 실험을 통해 측정하였다. 그 결과, 디스크 그룹의 크기는 [4]에서 제시한 최적의 크기의 80~120%일때 상대적으로 좋은 성능을 나타냈고 디스크 그룹의 크기가 같을 때에는 그룹 내에서 디스크 성능 간의 유사도가 높을수록 전체 검색 성능도 높았다. 그리고 만일 성능이 상이한 둘 이상의 디스크들로 구성된 혼합 그룹이 여럿일 경우에는 유사도가 더 높은 그룹을 더 많은 수의 디스크들로 구성하는 것이 전체 검색 성능을 높일 수 있음을 알 수 있었다.

5. 참 고 문 헌

- [1] 안유정, 원유현, "클러스터드 주문형 비디오 서버에서 비디오 스트림의 검색 효율을 높이는 배치 정책", 통산학회 논문지 제24권 제9호, 1999.
- [2] Yuewei Wang, David H. C. Du, "Weighted Striping in Multimedia Servers", IEEE multimedia systems p102 - 109, June, 1997
- [3] Renu Tewari, Daniel M. Dias, Rajat Mukherjee, Harrick M. Vin, "High Availability in Clustered Multimedia Servers", Proceedings of the USENIX 1996 Annual Technical Conference, Jan. 1996
- [4] Prashant J. Shenoy, Harrick M. Vin, "Efficient Striping Techniques for Multimedia File Servers", Univ. of Texas at Austin, Technical Report CS-TR-96-27, 1997
- [5] Peter M. Chen, Edward K. Lee, David A. Peterson, "RAID: High-Performance, Reliable Secondary Storage", ACM Computing Surveys, Vol 26, No. 2, pp. 152-158, June 1994.