

3

교육전문 전자도서관 저장구조 설계에 관한 연구 - SAN방식을 중심으로 -

김 용

한국통신 연구개발본부 멀티미디어연구소 전자지불연구실 전임연구원
(yongkim@kt.co.kr)

목차

1. 서 론
2. 관련연구
 2. 1 저장장치의 부착형태
 2. 2 네트워크 기반 저장구조(Network Attached Storage)
 2. 3 SAN의 개요
 2. 4 SAN 아키텍처의 특징
3. 교육전문 전자도서관 저장구조의 요구사항
 3. 1 교육용 디지털콘텐츠의 특징
 3. 2 교육용 전자도서관 저장구조 요구사항
4. 실험 및 결과
 4. 1 실험 환경
 4. 2 실험 결과 및 분석
5. 결 론

1. 서론

컴퓨팅파워와 인터넷사용의 폭발적인 증가는 자연스럽게 디지털데이터의 증가를 가져오고 있다. 이러한 디지털데이터의 주요한 형식은 불과 얼마전까지만 하여도 주로 텍스트에 기반을 두고 있었으나 근래의 경우에는 음성/동영상/이미지와 같은 대용량의 멀티미디어요소를 포함하는 멀티미디어콘텐츠의 증가가 큰 특징을 이루고 있다. 기존의 텍스트위주의 정보에서 이미지, 음성과 동영상을 포함하고 있는 멀티미디어콘텐츠의 증가는 자연스럽게 대용량의 저장 장치를 요구하게 되었다. 이러한 저장 장치는 서버와 직접 연결되어서 이용자의 요구가 있는 경우 서버는 요구되는 데이터를 직접 연결된 하드디스크로부터 불러온다. 하지만 이용자의 요구가 매우 빈번한 시스템, 예를 들어 웹에 기반을 두고 있는 포털서비스시스템이나 전자도서관과 같은 경우 서버측과 이용자사이의 처리속도에 비하여 서버와 스토리지 사이의 트래픽의 증가로 인하여 신속성이 요구되는 웹기반시스템에 있어서 많은 문제점을 제공하게 되었다. 또한 데이터베이스, ERP, 전자우편 등의 시스템을 운영하고 있는 관리자는 특정 서버가 다운되는 영향을 최소한으로 줄이기 위해, 복수의 처리를 묶어 1대의 머신으로 운영하는 것을 피하고 싶어하는데, 그러자면 데이터는 복수의 서버로 분산시킬 수밖에 없다. 그런데, 여기에서 문제가 발생할 수 있다. 데이터가 서버마다 분산된 채 증가하게 되면, 그에 따른 유지/보수에 시간과 일손이 많이 들기 때문이다. 그리고 서버마다 다른 제조업체의 하드웨어를 구입한다면, 스토리지를 증설하는데 있어서 상당한 복잡성과 어려움을 가져올 수 있다. 또한, 백업 문제의 심각성도 늘어날 것이다. 각각의 스토리지 내에 저장된 데이터를 각각 백업하는 일과 백업할 때에 발생하는 트래픽이 망에 미치는 영향은 무시할 수 없게 된다. 특히 대용량의 멀티미디어 콘텐츠를 다루고 있는 포털사이트나 전자도서관, 데이터센터와 같은 대용량 시스템의 경우에 있어서 기존의 SCSI방식이나 NAS방식으로 서버와 스토리지를 구성한다면 위에서 언급한 문제점은 더욱 심각하여 질 수 있다. 예를 들어 현재 한국통신에서 구축하고 있는 교육전문 포털사이트의 경우 스토리지의 용량이 약 45TB의 규모로서 이러한 스토리지를 각 개별적인 서버에 SCSI방식으로 구성한다면 그 서버와 스토리지사이의 트래픽의 증가로 인하여 시스템의 성능은 급격하게 낮아질 수 있으며 또한 하나의 서버에 하나의 스토리지장치를 연결하므로써 저장장치의 효율적인 활용이 어렵다. 따라서 기존의 소용량의 데이터를 처리하는데 적용되었던 방식에서 현재 요구되고 있는 대규모의 멀티미디어콘텐츠를 저장하고 이를 처리할 수 있는 새로운 방식의 서버와 스토리지를 구성할 수 있는 구성방식이 개발되어야 할 것이다.

2. 관련 연구

본 장에서는 현재 이용되어지고 있는 저장장치를 구성하는 방식에 대하여 알아보고, 각 개별 방식

들인 버스부착형(bus-attached), NAS와 SAN방식에 대하여 알아보며, 특히 SAN의 정의, 원리 및 SAN을 구성하기 위한 구비조건에 대해서 설명하고, 현재 일반적인 전자도서관과 같은 웹을 기반으로 하고 있는 시스템에서 일반적으로 사용되고 있는 NAS와 SAN과의 차이점에 대하여 비교하여 설명한다.

2. 1 저장장치의 부착 형태

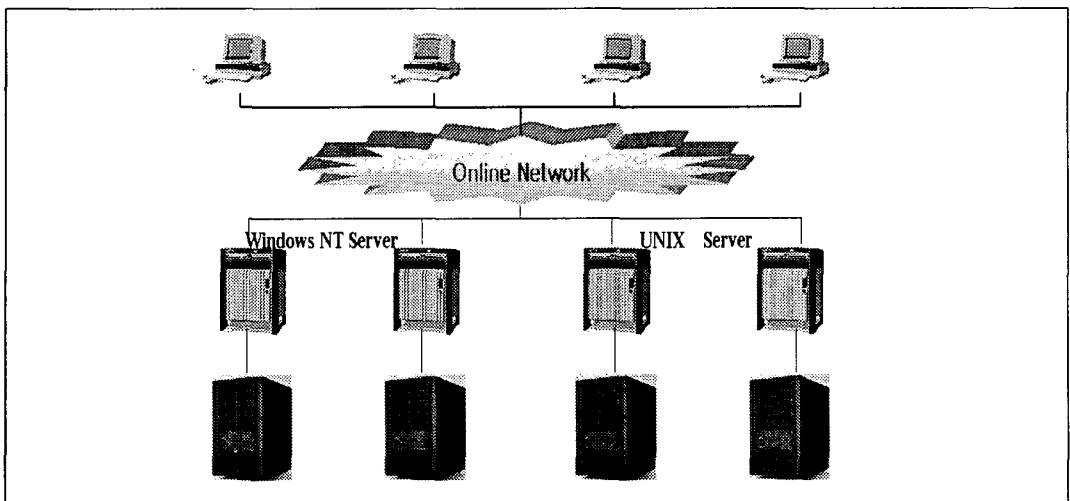
스토리지는 세가지 방법 중 하나로 서버에 부착될 수 있다. 현재 서버 스토리지 접속의 98%는 SCSI나 IDE의 형태를 통한 버스-부착형(bus-attached)이다. 버스-부착형 스토리지는 서버를 통해 작동되며 가용성과 성능은 서버의 용량과 로딩에 제한된다. <그림 2-1>은 전통적인 버스 부착형 스토리지 방식을 보여주고 있다.

이러한 저장구조중에서 현재 많은 이슈가 되고 있는 NAS와 SAN방식에 대하여 좀더 자세히 알아보면 다음과 같다.

2. 2 네트워크기반의 저장구조(NAS: Network Attached Storage)

NAS는 이더넷에 연결되어 파일 서비스를 제공하기 위해 고안된 운영체제와 파일시스템을 가진 서버를 가리킨다. NAS는 NFS 프로토콜을 사용해 데이터를 주고받으며, 이들 프로토콜들은 이더넷이나 인터넷에서 사용되어지는 IP 프로토콜상에서 운영된다. NAS는 이기종간의 파일 공유를 지원하며, 클라이언트의 업무 중단 없이도 손쉽게 설치할 수 있다는 것이 장점이다.

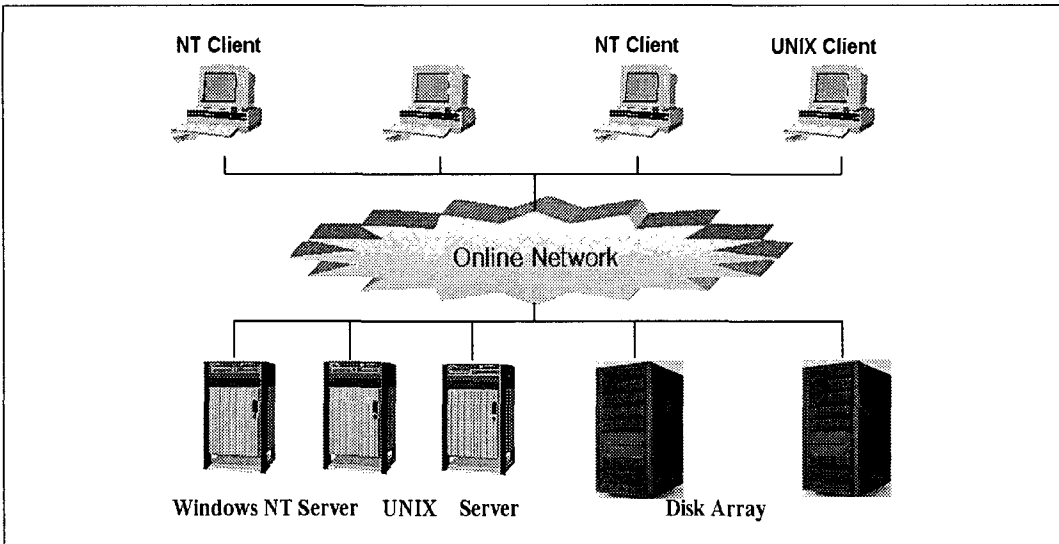
이러한 장점들에 반해 단점도 많다. NAS는 데이터베이스 처리에 문제가 있다. NAS는 파일프로



<그림 2-1> SCSI 방식의 저장장치 구조

토콜을 사용하여 파일 서버의 역할을 하지만 데이터베이스 액세스 프로토콜은 사용하지 않기 때문이다. 또 데이터 백업 및 복구가 매우 어렵다. 일반적인 백업 에이전트를 NAS에 적용하는 것은 불가능하기 때문이다. 게다가 NAS에서는 일반 네트워크, 즉 이더넷으로 연결되기 때문에 대용량 데이터를 읽을 때 부하발생이 크다.

일반적인 NAS의 구성방식을 아래의 <그림 2-2>와 같이 구성될 수 있다.



<그림 2-2> NAS 방식의 스토리지 구성

2. 3 SAN(Storage Area Network)의 개요

현재 기업 규모의 시스템에서는 데이터 웨어하우스의 구축과 ERP 시스템 등의 도입으로 대용량 스토리지에 대한 요구가 계속 높아지고 있다. 높은 신뢰성과 성능, 내장애성, 그리고 통합된 관리와 고속 백업이라는 요구에 대한 솔루션으로 등장한 것이 바로 SAN(Storage Area Network)이다. SAN은 분산 네트워킹에서 주류가 되고 있으며, 조만간 스토리지 부착 및 공유에 대한 일반적인 방법이 될 수 있을 것이라고 많은 이들이 내다보고 있다.

SAN이 등장한 배경에는 디지털콘텐츠 제작의 용이성과 함께 이를 처리하고 저장하는 시스템에 따라 데이터가 폭발적으로 증가하고 있다는 점을 들 수 있다.

2. 4 SAN 아키텍처의 특징

SAN은 망에서 데이터를 이동시키는 거의 모든 어플리케이션들의 성능을 향상시킬 수 있다. 기

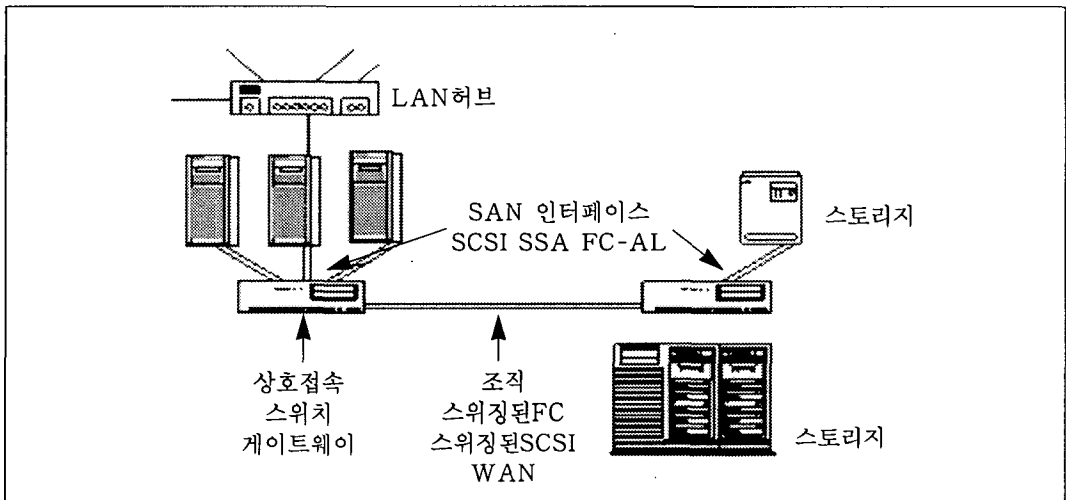
존의 서브넷과 같이, SAN은 기본 망에 대한 부하를 주지않고 특별한 기능을 위한 대역폭을 추가하는 방법으로 LAN과 WAN을 보조할 수 있다. 또한, SAN은 데이터웨어하우스와 같은 고성능 솔루션을 가능하게 하며, 네트워킹 환경에 포괄적으로 적용이 가능하다. <그림 2-3>은 일반적인 SAN 구성을 보여주고 있다.

NAS와 SAN은 여러 부분에서 차이를 보여 주고 있는데 이를 간단히 비교하여 보면 <표 2-1>과 같다.

3. 교육전문 전자도서관 저장구조 요구사항

3.1 교육용 디지털콘텐츠의 특징

기존의 전자도서관에 저장되어지는 디지털데이터는 대부분이 텍스트 기반의 문서들이 주류를 이루고 있다. 따라서 이를 저장하고, 이용자에게 제공하는데 있어서 저장매체나 저장구조의 중요성은 다른 여타의 요소들에 비하여 상대적으로 많은 관심이 되지는 않았다고 할 수 있다. 물론 빠른 검



<그림 2-3> SAN의 일반적 구성도

<표 2-1> SAN과 NAS방식의 특징비교

	SAN	NAS
Inerface	SCSI & Fibre	Network Protocol (IP protocol)
접근방식	물리적인 드라이브	네트워 드라이브
속도	SCSI & Fibre	Network 속도(Fast Ethernet)

색을 위한 효과적인 화일시스템에 대한 많은 연구가 있었으나 실질적으로 데이터가 저장되어 있는 H/W 분야나 아키텍처분야에는 상대적으로 많은 연구가 없었다. 그러나 본 연구에서 대상으로 하고 있는 교육전문 전자도서관에서는 일반적인 전자도서관이 저장하고 있는 데이터와는 형식이나 용량에 있어서 많은 차이점을 보여주고 있다. 예를 들어 교육전문 전자도서관의 콘텐츠는 주로 교육을 위한 교안자료이며 일부가 학습을 지원하는 학습자료가 대부분이라고 할 수 있다. 따라서 현재 국내의 원격교육서비스를 제공하고 있는 6개의 온라인 교육사이트에서 제공되고 있는 대표적인 교육용콘텐츠의 형식, 저작도구, 화일의 용량을 조사하여 본 결과 아래의 <표 3-1>과 같은 결과를 얻을 수 있었다. <표 3-1>에서 볼 수 있는 것과 같이 대부분의 교육용 콘텐츠는 일반적인 문서와는 많은 차이점을 보여주고 있다. 교육용콘텐츠의 특징을 살펴보면 다음과 같다.

1. 파일용량에 있어서 일반 디지털문서와는 달리 대용량이다.
2. 텍스트 기반의 형식이 아닌 주로 이미지, 오디오, 또는 비디오데이터 등을 포함하는 멀티미디어 형식을 주로 띄고 있으며 내용이나 질에 있어서 좋은 평가를 받는 콘텐츠의 경우에 있어서는 텍스트 위주의 형식보다는 주로 오디오와 비디오형식을 포함하는 동영상으로 제공되어지고 있다.
3. 일반 자료들이 콘텐츠의 내용을 열람하는데 있어서 웹브라우저를 이용하여 내용을 브라우징 하거나 다운로드 방식을 이용하고 있으나 교육용 자료의 경우 다운로드 방식과 함께 스트리밍 방식을 취하고 있다.
4. 저작도구에 따라 파일의 형식이 달라진다.

<표 3-1>에서 보는 바와 같이 기존의 전자도서관에 저장되어지는 데이터의 형식과 용량에 있어

<표 3-1> 교육용콘텐츠의 예

사이트	강의명	형식	파일크기	방식	내용
e-campus	Oracle 8 DB administration	Active tutor	1.915 Mb	다운로드	Audio
	Java Script		6.57 Mb	다운로드	Audio
Credu	비즈니스교양골프		3.663 Mb	다운로드	Audio/Video
Educast	Say it right		1.532 Mb	다운로드	Audio/Video
	Thinking in English		2.084 Mb	다운로드	Audio/Video
정보통신사이버대학	네트워크프로그래밍	Rm(real Media)	10 Mb	스트리밍	Audio/Video
메디캠프	간경변	Active tutor	2.109 Mb	다운로드	Audio
	간염처방		2.643 Mb	다운로드	Audio/Video
4Csoft	회계원리 강의	Active tutor	7,429 Mb	스트리밍	Audio/Video
	Unborn Child		4,455 Mb	스트리밍	Audio/Video

서 교육용 콘텐츠는 많은 차이점을 보여주고 있다. 따라서 이러한 콘텐츠의 특성을 고려하지 않고 기존의 저장구조방식을 따른다면 많은 문제점을 내포할 수 있다. 특히 위에서도 언급한 것처럼 교육용 콘텐츠는 일반 문서와는 달리 파일용량에 있어서 많은 차이를 보여주고 있으며 이러한 추세는 교육적인 효과를 고려하여 콘텐츠가 저작되는 추세로 볼 때 개별 콘텐츠의 파일용량은 더욱 더 커질 것이라고 예상할 수 있다. 또한 대부분의 교육용 콘텐츠를 제공하는 방식이 단순히 브라우저를 통한 제공이 아닌 다운로드방식이나 스트리밍방식을 취하고 있기 때문에 서버와 저장장치사이의 지속적인 I/O가 발생하므로써 많은 트래픽이 발생하게 된다. 따라서 효과적인 서비스를 위해서는 네트워크의 성능도 많은 영향을 줄 수 있으나 무엇보다도 서버와 저장매체간의 트래픽을 효율적으로 분산 할 수 있는 저장구조로서 설계되어야 할 것이다. 한편 관리적인 측면에 있어서 장치의 확장성에 대한 경제성이나 전체 저장매체를 관리하기 위한 관리적측면을 고려하여 볼 때 교육용 전자도서관을 위한 저장구조에 있어서는 이러한 새로운 요구사항에 적합한 구조를 가져야 할 것이다.

3. 2 교육용 전자도서관 저장구조 요구사항

위에서 살펴본 바와 같이 교육용콘텐츠의 일반적인 특징과 또한 이를 이용자에게 제공하는 방식을 고려하여 볼 때 기존의 전자도서관 저장구조를 기반으로 교육용콘텐츠를 저장하고 검색, 제공하는기 위해서는 많은 문제점을 야기 할 수 있다.

따라서 교육용콘텐츠를 저장해야 하는 교육전문 전자도서관의 경우 보다 효율적인 서비스를 제공하기 위해서는 저장구조에 있어서 기존의 전자도서관과는 다른 구조를 가져야 한다. 이를 위해 요구되어지는 사항을 정리하여 보면 다음과 같다.

1. 개별 교육용콘텐츠의 용량이 매우 크다고 할 수 있으며 또한 전체 콘텐츠의 용량에 있어서도 매우 많은 저장장치를 필요로 한다.
2. 교육용콘텐츠의 특성으로서 하나의 콘텐츠를 동시에 접속하는 동시접속자수가 매우 많다는 것이다.
3. 플랫폼에 독립적이어야 한다.
4. 확장성이 용이하여야 한다.
5. 디지털콘텐츠의 규모가 매우 크기 때문에 이를 저장하기 위한 저장장치의 규모도 매우 크다고 할 수 있다.

이러한 교육용디지털도서관의 저장구조를 위하여 요구되어지는 요구사항을 고려하여 본 논문에서는 교육용디지털도서관의 저장구조로서 서버와는 독립적으로 저장매체들로서 일련의 네트워크를 구성하여 위에서 요구되는 요구사항을 만족할 수 있는 구조를 알아보고자 한다.

4. 실험 및 결과

4. 1 실험환경

본 실험에서는 실제 SAN방식과 NAS방식의 차이를 비교하기 위하여 현재 교육포탈서비스를 제공하고 있는 한국통신의 포탈사이트인 HANMIR(<http://www.hanmir.com>)에서 이를 실험하였다.

SAN환경을 구축하는데 있어서 사용된 실험 장비는 <표 4-1>과 같다. 성능 측정을 위해서는 현재 가장 일반적으로 많이 사용되고 있는 공개 벤치마크테스트프로그램인 bonnie를 활용하였으며 이 프로그램을 통하여 BMT에서 일반적인 기준으로 삼고 있는 character(1byte)와 block(512kbyte)단위로 sequential input(read), sequential output(write), random seek를 측정하였으며, 또한 CPU가 데이터를 읽고 쓰기 위하여 점유되어지는 점유율의 차이를 측정하였으며 이러한 CPU의 점유율은 해당 데이터를 읽고 쓰는 동안 다른 operation을 수행할 수 있는 능력을 보여주는 것으로서 서버와 저장장치의 성능을 결정하는데 있어서 중요한 기준요소가 될 수 있다. 일반적으로 전자도서관에서 이용되어지는 데이터의 크기는 매우 다양하며 특히 멀티미디어데이터의 경우 그 크기가 매우 다양하다. 따라서 이를 고려하여 실험에 사용되어지는 데이터의 크기를 1M, 4M, 8M, 16M, 32M, 64M, 100M 단위로 구분하여 적용하여 보았다. 각각의 RAID 구성방식은 RAID 0과 1을 적용하였다.

4. 2 실험결과 및 분석

<그림 4-2>에서는 이용자의 요청을 받은 서버에서 요청되어진 파일을 하드디스크로부터 읽어드리고 또한 데이터를 하드디스크에 쓰는데 걸리는 시간과 함께 데이터 입, 출력시간 등에 대하여 SAN방식과 NAS방식과의 처리시간 등을 비교하여 보여 주고 있다. 전체적으로 SAN방식은 NAS방식에 비하여 보다 좋은 성능을 보여주고 있으며 특히 블록단위로 데이터를 읽고 쓰는데 있어서 훨씬 우수한 성능을 보여주고 있다. <표 4-2>에서 보면 SAN방식과 NAS방식에서의 CPU 점유율을 측정한 결과를 보여주고 있다. 여기에서 보면 SAN방식이 전체적인 CPU점유율에서 보다 나은 성능을 보여주고 있으며 특히 Sequential output의 블록단위 CPU점유율에서 SAN이 월등히 나은 성능을 보여주고 있으며 Sequential input의 블록단위 CPU점유율에서 보면 비록 NAS방식이 CPU점

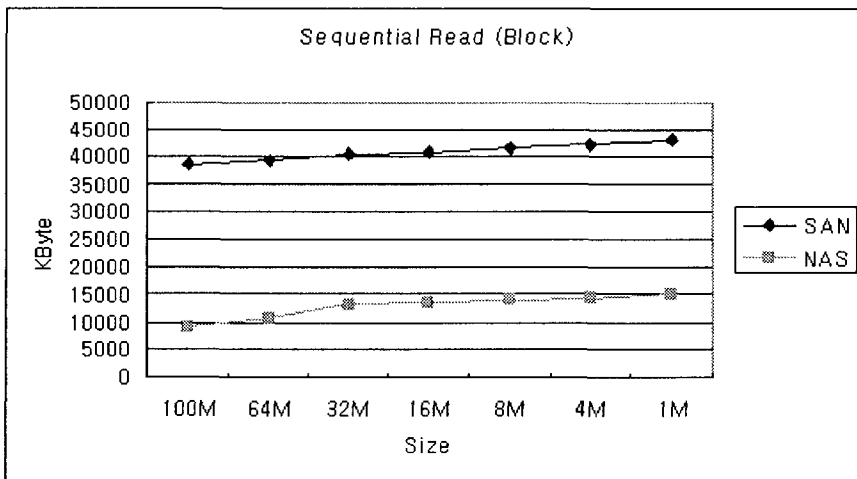
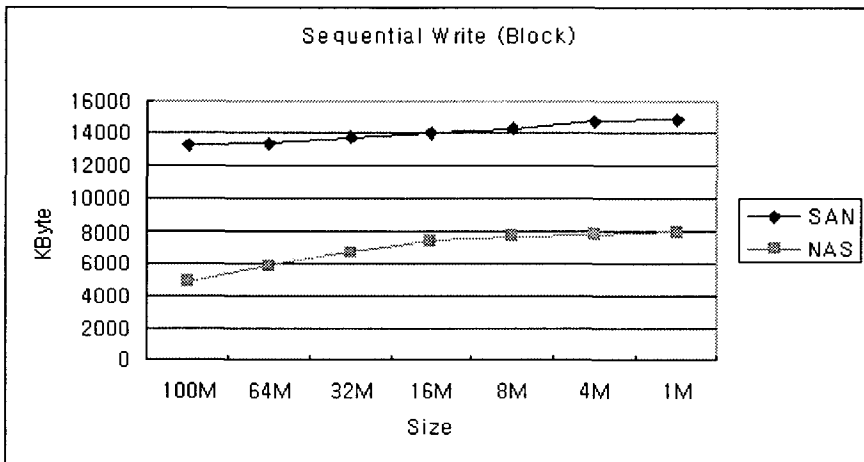
<표 4-1> 사용되어진 실험 장비

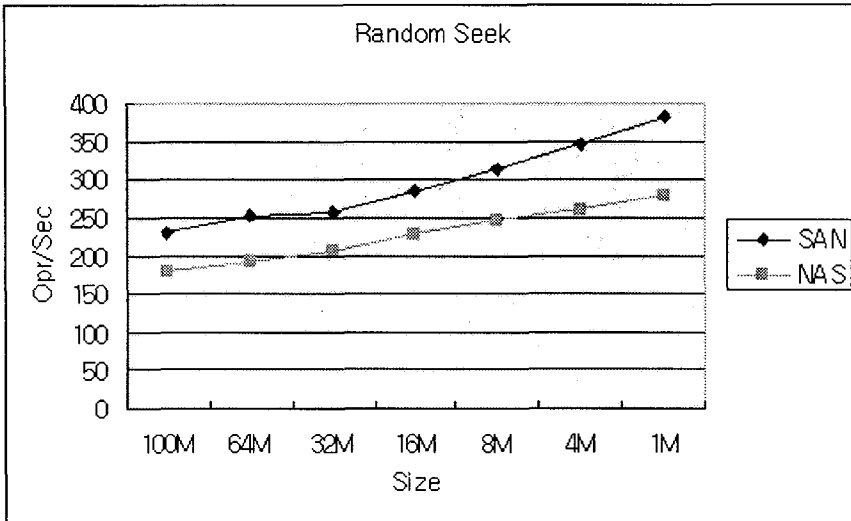
Test Machine	Sun E3500 4CPU 400MHZ 2GB
SAN	SUN T3 250GB(RAID 0+1)
NAS	NetAppliance F760 120GB(RAID 0+1)

유율이 SAN보다 낮지만 해당 데이터를 쓰는 경우에는 약 3배의 성능을 보여주고 있다. 따라서 이러한 결과들을 전체적으로 종합하여 보면 SAN방식이 NAS방식에 비하여 대용량의 데이터를 처리 하는데 있어서 전체적으로 보다 나은 성능을 보여주고 있다.

5. 결 론

컴퓨터기술의 발전과 함께, 디지털저작물 저장을 위한 저장도구의 발전은 이전의 단순한 텍스트 기반의 저작물을 저장하는 것과는 달리 효율적인 데이터의 저장과 이용자의 요구에 따라 요구되어 지는 데이터를 보다 신속하게 제공하기 위한 데이터의 저장구조에 있어서 새로운 요구사항을 필요





〈그림 4-2〉 SAN과 NAS방식에서의 데이터 처리 실험결과

〈표 4-2〉 SAN과 NAS에서의 CPU 점유율

		Sequential Output						Sequential Input				Random	
		Per Char		Block		Rewrite		Per Char		Block		Seeks	
	MB	K/sec	%CPU	K/sec	%CPU	K/sec	%CPU	K/sec	%CPU	K/sec	%CPU	K/sec	%CPU
NAS	100	6851	97.9	37660	75.4	8225	25.2	4486	96.3	19334	45.7	232.5	17.7
SAN	100	4002	19.0	24007	4.9	3764	12.6	14640	97.2	53946	93.2	635.8	18.0

로 하고 있다. 따라서 본 실험 이러한 요구사항에 따라 요구되어지는 SAN 과 NAS방식의 효율적인 방식을 직접 실험을 통하여 측정하였으며 이러한 결과를 분석한 분석결과는 위의 4장에서 언급한 것과 같다. 이러한 결과는 향후 대용량의 멀티미디어데이터를 저장하고 이를 이용자에게 제공하기 위한 대용량의 전자도서관의 경우에 있어서 다른 여타의 요소시스템들 즉, 검색엔진, 웹서버 등도 중요하다고 할 수 있으나 무엇보다도 실제 데이터를 저장하고 이를 읽고 쓰기 위하여 서버와 밀접하게 연동하고 있는 저장장치의 중요성은 매우 높다고 할 수 있을 것이다. 따라서 본 실험 결과를 통하여 알 수 있듯이 기존의 텍스트 기반의 데이터와 같은 소규모의 데이터를 저장하는데 있어서 저장장치의 중요성은 그다지 높다고 할 수 없으며 오히려 다른 여타 시스템의 중요성이 매우 높다고 할 수 있으나 교육용 콘텐츠와 같은 멀티미디어데이터를 저장하기 위한 교육전문 전자도서관과 같은 대용량의 저장장치를 필요로 하는 시스템의 경우에 있어서 저장장치의 중요성은 매우 높다고 할 수 있다. 이러한 관점에서 본다면 위의 4장에서 분석되어진 것과 같이 데이터를 읽고 쓰며 또한 이러한 일련의 작업을 수행하는 동안 다른 오퍼레이션을 수행하는데 있어서 새로운 저장구조

로서 제안 되어지고 있는 SAN은 아마도 가장 적절한 저장구조가 될 수 있을 것이다. 따라서 향후 교육용 전자도서관 뿐만이 아닌 대용량의 데이터를 저장하고 이를 처리 하는 전자도서관의 경우 SAN방식을 활용하여 저장구조를 설계하는 것은 매우 의미 있고 효과적이라고 할 수 있을 것이다.

참 고 문 헌

Dave Fetters, Building a Storage Area Network

<http://www.networkcomputing.com/1109/1109ws1.html>, 2000

Designing storage area networks: a practical reference for implementing fibre channel SANs, Clark, Tom, Reading, Messachusetts: Addison Wesley Longman, Inc., 1999

Gang Ma, Khaleel, A.; Reddy, A.L.N., "Performance evaluation of storage systems based on network-attached disks", Parallel and Distributed Systems, IEEE Transactions on , vol. 11(9), pp 956-968, 2000

Guha, A., "The evolution to network storage architectures for multimedia applications", Multimedia Computing and Systems, 1999. IEEE International Conference on , vol. 1, pp 68-73, 1999

Heath, J.R., Yakutis, P.J., "High speed storage area networks using a fibre channel arbitrated loop interconnect" IEEE Network , vol. 14 (2) , pp 51-56, 2000

Storage Area Network (SAN),

<http://backofficesystems.com/tips/network/san.htm#background>