

NEC 시스템의 I/O 성능 향상을 위한 파일 시스템 최적화 구성

이영주, 성진우, 장지훈, 이상동, 김종권
한국과학기술정보연구원
e-mail:yjlee@kisti.re.kr

File System Optimization Implement for I/O Efficient Performance of NEC System

Young-Joo Lee, Jin-Woo Sung, Ji-Hoon Jang, Sang-Dong Lee,
Joong-Kwon Kim
Korea Institute of Science Technology Information

요 약

최근 인터넷과 같은 개방된 네트워크의 발달로 데이터의 양이 기하급수적으로 증가하고 있으며 데이터의 단위 크기도 커지고 있다. NEC 시스템에서 이러한 크고 많은 데이터들을 효율적으로 관리하고 컴퓨터의 성능을 최대로 활용하기 위하여 파일 시스템을 설계하고 구현하여 그 성능을 시험하고 분석하였다. 전체 컴퓨터의 성능을 결정하는 요소는 CPU와 메모리, 디스크, 그리고 이들을 연결하는 네트워크 등 여러 가지가 있지만 이러한 각각의 부분적인 시스템 중에서 사용자가 수행할 프로그램 특성에 맞게 설계하여 사용할 수 있는 부분은 파일 시스템이다. 파일 시스템은 전체 시스템의 특성과 사용자의 프로그램 용도를 고려하여 알맞게 설계되면 시스템의 성능을 더 향상시킬 수 있다. 본 논문은 NEC 시스템에서의 파일 시스템을 시스템의 특성과 용도에 맞게 설계하고 테스트하여 그 성능 결과를 분석하였다.

1. 서론

최근 인터넷의 발달로 인하여 네트워크 환경에서 데이터가 기하급수적으로 증가하고 있으며, 데이터의 단위 크기도 커지고 있다. 이러한 파일시스템을 관리하고 다루기 위한 기술도 발달되고 있으며 이러한 파일 시스템의 관리가 컴퓨터의 처리 능력과 연계되어 시스템의 성능에 큰 영향을 준다.

파일 시스템은 사용자가 컴퓨터를 이용하여 작업을 할 때 프로그램이나 데이터 또는 시스템 관련 라이브러리 등을 저장하는 장소이다. 컴퓨터의 전체 성능을 좌우하는 것은 CPU나 메모리, 디스크의 성능이 중요하지만 이 중에서 사용자의 사용 환경에 맞게 재설계가 가능한 부분은 파일 시스템이다. 파일 시스템은 작업의 실행에 영향을 준다. 프로그램 실행에 있어서 처리 시간의 대부분은 데이터의 입출력에 소요됨으로 시스템의 특성과 디스크 성능을 고려하여 파일시스템의 구성을 최적화 하면 시스템 전체의 성능을 향상할 수가 있다.

본 논문에서는 NEC에서의 공유 파일 시스템인 GFS와 파일의 용도에 알맞은 클러스터 크기를 사용하여 파일 시스템을 설계하고 테스트하였다.

2. 관련 연구

2.1 NFS(Network File System)

NFS는 썬마이크로 시스템즈에서 개발한 네트워크 상에서 파일시스템을 공유하도록 설계된 파일시스템의 한종류로서 다른 호스트에 있는 파일시스템의 일부를 자신의 디렉토리인 것처럼 사용할 수 있도록 해주는 방식이다. 즉 네트워크상에 디스크를 한서버에 집중관리하고 나머지 시스템들은 그것을 공유하여 사용할 수 있게 해주는 것이다. 지금까지 일반적으로 가장 많이 사용되어왔지만 전송속도와 보안상의 문제로 인해 NFS 서버의 사용을 점점 제한해 가는 추세이다. NFS 설정은 파일시스템을 공유하는 서버의 설정과 서버의 파일시스템을 마운트해가는 클라이언트 설정으로 나누어진다.

2.2 GPFS(General Parallel File System)

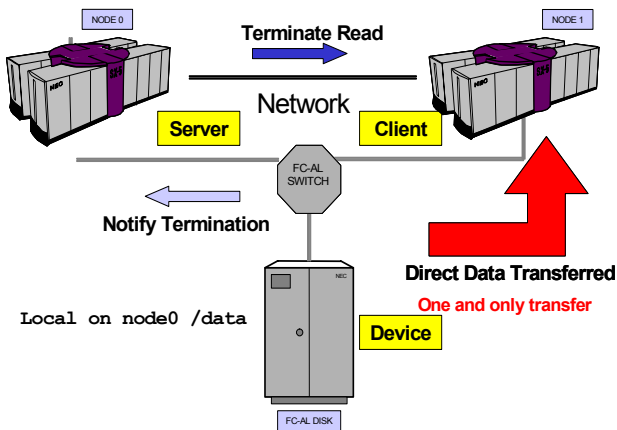
GPFS는 여러노드에 걸쳐 있는 다수의 디스크를 사용하여 파일 시스템을 만들고 여러 노드의 사용자들이 하나의 시스템을 동시에 읽고 쓸 수 있는 병렬 파일 시스템이다. GPFS는 SMP 시스템에서 빠르게 병렬 프로그램을 수행하기 위하여 개발되었다.

GPFS는 멀티미디어 서비스를 위한 고성능 파일 시스템을 구현하기 위해 개발한 만큼 여러노드에서 고속으로 동시에 read/write을 하기 위한 구조를 가지고 있다. GPFS는 노드의 메인메모리 중 일부를 캐시로 사용하여 전체적인 I/O 성능을 향상시킬 수 있으며, 각 노드마다 4MB에서 최대 512MB까지 캐시를 사용할 수가 있다.

2.3 GFS(Global File System)

GFS는 기존의 유닉스 파일 시스템을 개선하여 SAN 환경에서 UNIX 파일시스템으로 사용하도록 개발되었다. GFS는 커다란 스토리지 용량과 대역폭을 요구하는 멀티미디어와 과학용 프로그램들을 주요 타깃으로 삼았다. 다른 파일 시스템과 다른 점은 캐시데이터를 호스트의 메인메모리가 아닌 스토리지 디바이스에 위치하는 것이다. 그리고 캐시 일관성을 유지하기 위한 락 매커니즘은 스토리지 디바이스 컨트롤러에 의해 관리되어 진다.

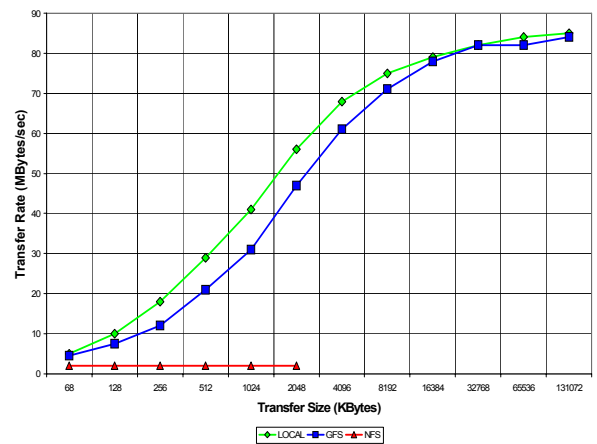
(그림 1)은 클라이언트에서 서버에 요청한 정보를 다시 클라이언트에서 받을 때까지의 절차를 나타내고 있다. 클라이언트에서 서버에 있는 어떤 파일을 액세스 하려면 먼저 클라이언트에서 서버에 해당 정보를 요청하고 서버는 다시 서버에 연결된 저장장치에 요청하여 해당 정보를 서버를 통하여 클라이언트에 해당 정보를 전송하게 된다.



(그림 1) GFS 파일 전송 과정

파일의 구성은 NFS와 같지만 액세스 절차에서 한 가지 차이가 있다. 클라이언트에서 서버에 있는 어떤 파일을 액세스 하려고 할 때, 클라이언트에서 서버에 해당 정보를 요청하고 서버는 다시 서버에 연결된 저장장치에 요청하는 과정은 같으나 저장 장치의 해당 정보를 서버를 통하지 않고 직접 클라이언트에 해당 정보를 전송하게 되는 것이 다르다. 이와 같이 한단계가 줄어들기 때문에 그만큼 전송 속도가 빠르게 이루어질 수 있다.

(그림 2)에서와 보는바와 같이 GFS의 성능이 로컬 파일시스템의 성능과 거의 비슷하고 NFS의 성능에 비하면 매우 우수한 것을 알 수 있다.

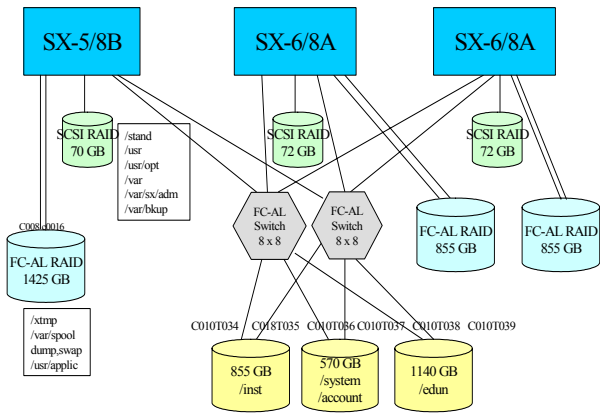


(그림 2) GFS와 로컬 파일시스템의 전송 속도

3. 파일시스템 설계

3.1 NEC 시스템 구성 환경

NEC 시스템은 (그림3)과 같이 SX-5 1노드, SX-6 2노드의 모두 3노드로 구성되어 있으며, 이들 3노드의 로긴 서버는 SX-6a이다. 두 종류의 시스템 중에서 로긴 노드와 서버를 SX-6로 정한 것은 SX-6 시스템이 SX-5에 비하여 CPU 클럭 속도가 빠르고 스칼라 성능이 SX-5 시스템에 비하여 좋기 때문이다. NEC의 SX-5와 SX-6의 사용자의 홈디렉토리는 SX-6a에 로컬로 연결하고 SX-5와 SX-6b는 GFS를 통하여 공유한다. 사용자의 작업 공간을 위하여 각각의 노드에 스크래치 디스크 /xtmp를 로컬로 연결하고 세노드에서 모두 공유할 수 있는 /ytmp는 SX-6a에 로컬로 연결하였다. 홈 및 스크래치 디스크는 모두 듀얼의 FC-AL로 연결하였다.

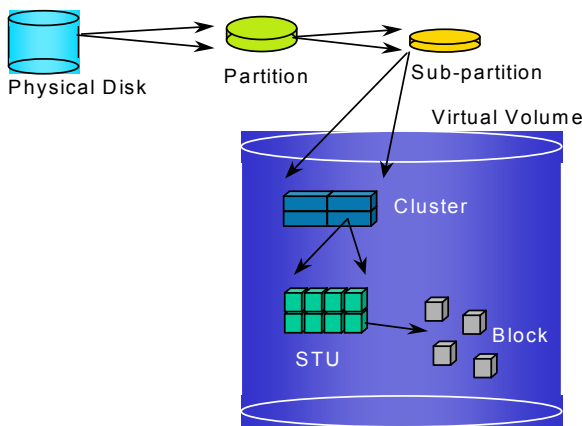


(그림 3) NEC의 파일시스템 구성

3.2 NEC 공유 파일시스템 구성

NEC 파일 시스템 구성에 있어서 성능에 관련되는 부분은 크게 두가지로 나눌 수 있다. 하나는 클러스터에 연결된 파일 시스템을 공유하는 방식과 다른 하나는 각각의 디스크 클러스터 크기에 따른 구성이다.

NEC 공유파일 시스템은 GFS를 사용한다. GFS는 클러스터 노드들이 다수의 스토리지를 공유하는 SAN 환경에서 데이터를 일관성으로 유지하기 위한 락킹 메커니즘을 제공하는 분산 시스템이다. GFS는 입, 출력 요구가 일어나고 있는 동안에 노드가 가지고 있는 버퍼 크기만큼 데이터를 캐시하고 있다가 요구가 끝나면 데이터가 스토리지 디바이스에 쓰게 된다.

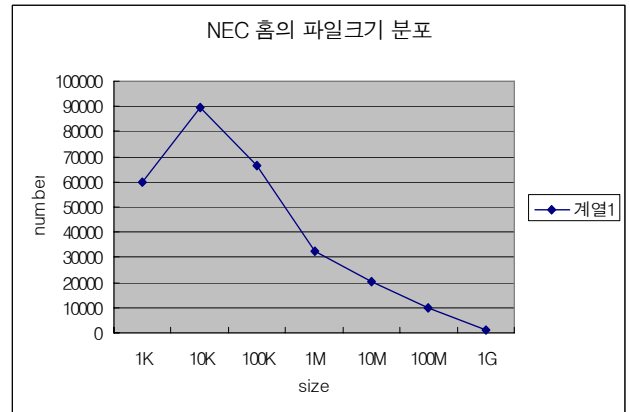


(그림 4) NEC 파일 시스템 연계

3.3 NEC 파일시스템 구성

파일 시스템의 종류는 크게 시스템 파일과 일반 사용자의 파일의 두가지로 나눌 수 있다. 시스템 파일은 I/O가 적기 때문에 클러스터의 크기를 4K로 하여 파일 저장의 낭비를 줄였으며 홈과 스크래치

디스크는 각각의 용도에 따라서 사용자의 파일의 크기를 분석하여 클러스터의 크기를 다르게 하였다.



(그림 5) NEC 홈디렉토리의 파일 분포

NEC에서 제공하는 파일의 클러스터 크기는 최소 4KB에서 최대 16MB이다. 클러스터는 I/O의 단위로써 디스크에 1byte를 기록하려면 1클러스터가 필요하게 된다. 따라서 클러스터의 크기가 크면 디스크의 낭비가 많으나 I/O의 성능은 증가하며 이와 반대로 클러스터의 크기가 작으면 디스크의 낭비는 적지만 I/O의 성능이 저하된다. NEC 파일 시스템의 구성은 (그림 5)의 통계를 바탕으로 <표 1>와 같은 디스크의 용도와 클러스터 크기를 결정하였다.

<표 1> 공유파일 시스템 구성

File System	Phase1(GB)	Cluster	
System /	4	4K	
Home	/edun	278	128K
	/inst	278	128K
	/system	48	128K
Scratch /xtmp	557	4M	
System (FC-AL)	/var/spool/nqs/restart	126	128K
	/usr/applic	10	128K
	dump, swap	128	4M
	/tmp	4	128K
	/work	4	4M
	/archive	210	128K
	/account	20	128K
	/export	425	128K
	/PCS_file	557	128K
Total	2,092		

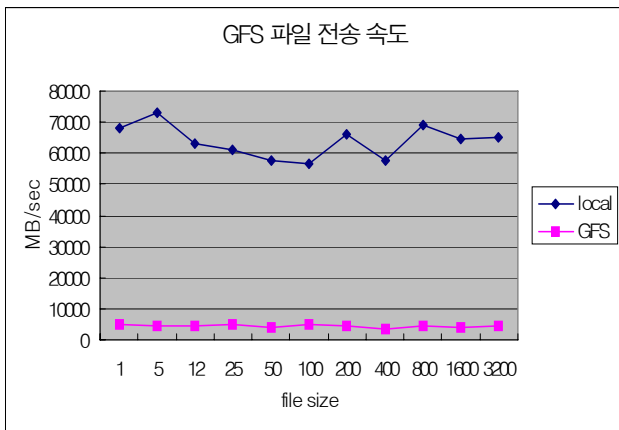
4. 파일 시스템 성능 시험

4.1 GFS와 로컬 파일 시스템의 전송 속도 비교

NEC 디스크 구조는 디스크당 73GB, 10,000rpm, 버퍼는 256MB의 RAID3으로 구성되었다. 성능 테스트 시의 시스템 환경은 다른 프로그램들이 수행되고 있는 가운데 진행하였다. 이는 성능 테스트를 위하여 시스템 전체를 정지시키기가 어려운 환경뿐 아니라 실제로 사용 중에 I/O에 대한 성능을 분석하기 위해서이다. I/O 성능을 위한 측정 도구 프로그램은 bonnie를 사용하였다. 테스트 파일의 크기는 1MB 이상의 파일을 테스트 하였다. 이는 파일의 크기가 디스크의 I/O 버퍼의 크기보다 큰 파일을 사용하여 버퍼를 사용하지 않고 파일의 전송 속도를 측정하였다.

전송 테스트는 SX-6a에서 홈 및 스크래치를 파일 시스템을 사용하였으며 bonnie 프로그램의 측정 결과 중에서 output의 block 단위의 속도의 결과를 참조하였다.

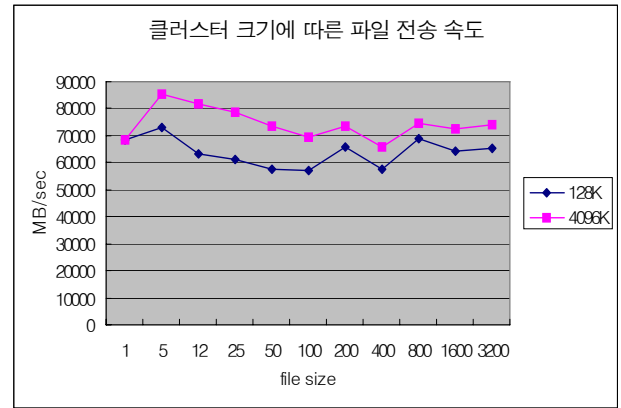
(그림 6)은 GFS와 로컬 파일시스템과의 전송 속도를 보여주고 있다. GFS의 전송 속도가 로컬 시스템에 비하여 약 7배정도의 속도 향상을 나타내고 있다.



(그림 6) GFS와 로컬 파일시스템의 전송속도비교

4.2 클러스터 크기에 따른 파일 전송 속도 비교

(그림 7)은 클러스터 크기에 따른 파일의 I/O 전송 속도를 보여주고 있다. 전송 테스트는 (그림 6)에서 시행한 환경과 동일하며 클러스터의 크기가 다른 파일 시스템을 사용하여 클러스터에 따른 전송속도를 비교하였다. (그림 7)에서와 같이 파일 클러스터의 크기가 128K와 4096K인 두 디스크 간에 전송 속도는 평균 17%의 차이를 나타냈다.



(그림 7) 클러스터 크기에 따른 전송 속도 비교

5. 결론

NEC에서 제공한 GFS의 성능 자료에 따르면 GFS의 성능이 로컬 파일시스템과 거의 동등하게 보여주고 있지만 실제로 시스템을 가동하면서 사용하는 환경에서 테스트한 결과는 이론 성능과의 차이가 큰 것을 알 수 있었다. 이것은 GFS 시스템이 NFS에 비하여 성능이 우수한 것은 사실이지만 사용자 환경에서의 체감 속도와는 많은 차이가 있다는 것을 알 수 있었다.

디스크의 용도에 따라 클러스터 크기가 다른 파일 시스템은 비교적 좋은 성능을 나타냈다.

향후에는 GFS 파일 시스템의 클러스터 노드를 증가시키면서 파일 전송 속도를 측정하여 노드 수가 많아질수록 파일 시스템의 부하가 어떻게 변화하는지 비교 연구를 하고자 한다.

참고문헌

- [1] NEC, "File System Administrator's Guide." NEC Corporation
- [2] NEC System User's Guide 2002 KISTI
- [3] IBM "GPFS on AIX Clusters: High Performance File System Administration Simplified" IBM red book
- [4] 김주호, SAN 기반 클러스터 파일 시스템 GFS의 입, 출력 성능 분석, 과학기술연구논문집, 2002
- [5] 김선일 SAN 환경하에서 파일 공유 시스템의 설계 및 구현, 2001