

네트워크 장치를 이용하여 디스크 기반 백업을 지원하는 백업 소프트웨어

최길성*, 윤중현**, 복경수**, 유재수**

*동아방송대학 정보통신계열

**충북대학교 정보통신공학과

kschoi@dabc.ac.kr,

{bluette, ksbok}@netdb.chungbuk.ac.kr,

yjs@cbucc.chungbuk.ac.kr

The Backup Software Supporting Disk Based Backup Using Network Devices

Kil Seong Choi*, Jong Hyoun Yun**, Kyoung Soo Bok**,
Jae Soo Yoo**

*Division of Telecommunication, Dong-Ah Broadcasting College

**Department of Computer & Communication Engineering,

Chungbuk National University

요 약

최근 데이터 통신 기술 및 저장 매체의 발전과 더불어 대용량의 데이터를 가공 처리하기 위한 많은 연구들이 진행되고 있다. 이와 함께 NAS와 SAN 등의 보급으로 저장 매체의 대용량화가 가속화되면서 대용량의 데이터를 효과적으로 백업하기 위한 요구들이 증가되고 있다. 본 논문에서는 네트워크 기반의 백업을 위해 제안된 개방형 표준 프로토콜인 NDMP를 이용하여 디스크 기반을 지원하는 백업 소프트웨어를 설계하고 구현한다. 제안하는 백업 소프트웨어는 네트워크에 연결된 다수의 장치들을 관리하고 NDMP 프로토콜을 확장하여 디스크 백업을 지원할 수 있도록 한다.

1. 서론

90년대 중반 이후 산업체 전반에 걸쳐 전산 업무 물론 대부분의 고객 서비스가 통신을 통해 온라인화 되어 가고 있다. 특히, e-비즈니스와 인터넷과 같은 통신 기술의 발달은 급격한 데이터의 증가를 가져왔다. 만일 이러한 상황에서 컴퓨터 장애로 인해 서비스를 제공하지 못하거나 사용자의 실수 또는 재난으로 개인 또는 회사의 중요한 데이터가 손실될 경우 경제적으로나 사회적으로 큰 손실을 입게 된다. 따라서, 각종 데이터 및 파일 시스템을 보호하기 위해서는 데이터를 주기적 또는 비주기적으로 백업하는 것이 필수적이다[1].

기존에 사용되는 DAS(Direct Attached Storage)는 서버에 저장 장치를 직접 연결하는 가장 일반적인 방법으로 직접적인 일대일 SCSI 연결을 사용한다. 본 연구는 산업자원부의 지역혁신 인력양성사업의 연구결과로 수행되었음”

다. DAS 시스템은 서버별로 독립적인 저장장치들을 관리하기 때문에 효율적으로 데이터를 공유하거나 관리하는 것이 어려우며 하나의 서버에 문제가 발생할 경우 데이터에 대한 접근이 불가능하게 된다. 이러한 문제점을 해결하기 위해 NAS(Network Attached Storage)와 SAN(Storage Area Network)과 같은 시스템들이 개발되었다.

NAS와 SAN는 시스템의 통합 및 공유, 중앙집중식 관리, 편리한 백업, 고가용성을 효과적으로 제공할 수 있다. 본 논문에서는 네트워크 기반의 백업을 위해 제안된 개방형 표준 프로토콜인 NDMP[2]를 확장하여 네트워크 환경에서 디스크 기반 백업을 지원하는 백업 소프트웨어를 설계하고 구현한다. 제안하는 백업 소프트웨어는 네트워크에 연결된 다수의 장치를 관리하여 네트워크에 연결된 디스크에 백업을 수행할 수 있도록 한다. 또한, 사용자의 스케줄링에 의한 자동 백업과 다양한 유형의 백업 및 복구

유형을 지원하며 백업 파일들에 대한 생명 주기 관리 기능, 스크립트 처리 기능, 장치 및 미디어 관리 기능을 지원한다. 본 논문의 구성은 다음과 같다. 2장에서는 관련 분야에 대해 살펴보고, 3장에서 제안하는 백업 시스템의 구조를 설명한다. 4장에서는 제안하는 백업 소프트웨어의 특징에 대해 설명한다. 마지막으로 5장에서는 결론 및 향후 연구 계획에 대해서 설명한다.

2. 관련연구

NAS는 TCP/IP를 이용하여 기존의 네트워크를 이용하기 때문에 SAN에 비해 구축 비용이 저렴하고 편리하게 설치할 수 있지만 기존 저장 구조가 특화되어 있다. SAN은 파이버 채널(Fiber Channel)을 이용하여 서버와 저장 장치를 연결한 특별한 네트워크이다. SAN은 클러스터링 된 저장 장치에 의한 뛰어난 가용성과 확장성을 제공하며 데이터 공유 및 관리가 중앙집중식 관리가 가능하다. 또한 파이버 채널을 통해 연결되기 때문에 높은 처리속도를 제공할 수 있으며, 백업 및 복구 시 보안성 향상과 시간 단축 등의 이점을 얻을 수 있다.

NDMP는 NAS 기반의 이기종 간의 백업을 위한 개방형 표준 프로토콜이다. NDMP는 클라이언트에 해당하는 DMA(Data Management Application)와 서버에 해당하는 DSP(Data Service Provider)로 구성된다. DMA는 NDMP의 정책, 사용자 인터페이스 그리고 DSP를 관리, 제어한다. DSP는 실제적인 백업 및 복구 작업을 수행하기 위한 테이프 서비스, 데이터 서비스, 그리고 전송 서비스 등을 제공하고 있다. 데이터 서비스는 주 저장장치에 대한 NDMP 인터페이스를 제공하며, 테이프 서비스는 보조 저장장치에 대한 NDMP 인터페이스를 제공한다. NDMP는 최근 버전 5에 이르러 SAN에 대한 지원을 포함하도록 확장되었다. Veritas사의 NetBackup[3], Legato사의 NetWorker[4], Atempo 사의 Time Navigator[5]등이 NDMP를 지원하고 있다. 이들 제품은 전체 백업, 마지막 전체 백업 이후 변경된 내용들에 대해 백업을 수행하는 차등(Differential) 백업과 점진(Incremental) 백업을 지원한다. 또한 사용자의 요청을 스케줄링하여 다중 백업 및 복구를 가능하게 하고 백업 장치 공유 및 관리 기능을 제공하여 효과적인 백업 작업을 수행할 수 있도록 지원하고 있다.

3. 제안하는 백업 소프트웨어의 구조

3.1 백업 소프트웨어의 구조

네트워크 기반의 백업을 효과적으로 수행하기 위

해 많은 연구들이 진행되고 있으며 이에 따라 여러 상용화된 제품들이 개발되고 사용되고 있다. 그러나 대부분의 상업용 소프트웨어들이 외국에서 개발된 제품으로 국내에서는 이러한 제품을 이용하여 시스템을 구축하고 있는 실정이다. 본 논문에서 NDMP 프로토콜을 확장하여 네트워크 기반 백업을 수행하는 소프트웨어를 개발한다. 본 논문에서 개발한 백업 소프트웨어의 구조는 그림 1과 같다. 개발한 백업 소프트웨어는 크게 백업 서버(Backup Server), 백업 클라이언트(Backup Client) 그리고 미디어 서버(Media Server)로 구성되어 있다.

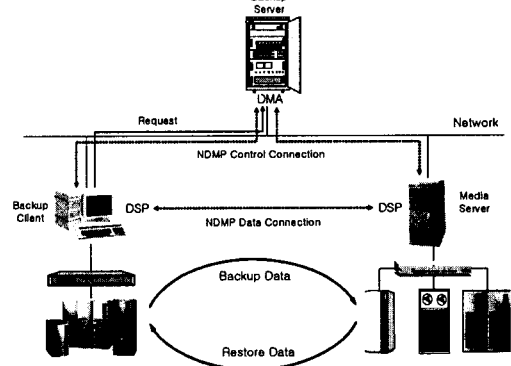


그림 1 제안하는 백업 소프트웨어 구조

백업 클라이언트는 실제 작업을 요청하는 사용자로서 사용자의 로컬 컴퓨터에 저장된 데이터에 대한 작업 요청뿐만 아니라 원격지에 존재하는 특정 저장장치에 저장된 데이터에 대한 작업 요청을 수행할 수 있다. 백업서버는 사용자의 작업 요청을 처리하기 위해 서버 모듈과 함께 NDMP 프로토콜에 의해 DSP들을 제어하고 통신할 수 있는 DMA 모듈을 포함하고 있어야 한다. DMA는 백업 및 복구 요청을 수행하기 위해 제어 신호를 전달하면서 실제 데이터 전송을 수행하는 DSP들의 상태를 제어하는 역할을 수행한다. 미디어 서버는 백업 장치를 논리적 또는 물리적으로 관리하고 백업 및 복구 요청에 따라 데이터 서비스를 제공하는 DSP와 데이터를 송수신하기 위해 테이프 서비스를 제공하는 DSP 모듈을 포함하고 있어야 한다. 이러한 미디어 서버는 별도로 존재할 수도 있지만 백업 서버 내에 함께 존재할 수도 있다. 제안하는 백업 소프트웨어는 테이프 기반의 백업 뿐만 아니라 디스크 기반의 백업을 수행하기 때문에 기존 테이프 서비스를 제공하는 기능을 확장한 DSP 기능을 제공한다.

3.2 백업 처리 과정

사용자는 클라이언트 인터페이스를 통해 서버에 백업 작업을 요청한다. 사용자의 백업 요청 정보는 서버의 백업 관리자에 의해 분석된다. 사용자의 백업 요청은 스케줄 관리자에 의해 작성된 주기에 의해 수행되는 자동 백업과 사용자의 직접적인 요청에 의해 수행되는 수동 백업으로 구분된다. 그림 2는 백업 처리 과정을 나타낸 것이다. 사용자의 백업 요청의 분석 결과에 따라 사용자가 요청한 백업은 클럭 데몬을 통해 스케줄 데몬에게 전달된다. 만약 수동 백업을 요청하는 경우에는 클럭 데몬을 거치지 않고 스케줄 데몬에 전달된다. 백업 관리자는 사용자가 요청한 백업 정보를 바탕으로 백업 작업 파일을 작성한다. 작성된 백업 작업 파일을 파라미터로 하여 스케줄 데몬은 NDMP 프로세스를 호출하고, NDMP 프로토콜에 따라 백업이 수행되도록 한다. NDMP 프로세스가 모든 작업을 종료하면 로그 및 리포트 관리자를 통해 백업 과정에서 발생한 로그 및 백업된 파일들에 대한 히스토리 정보를 기록한다. 기록된 히스토리 정보는 복구를 수행하기 위해 백업된 내용을 검색할 목적으로 사용된다. 또한 사용자가 지정한 백업 파일의 생명 주기에 따라 백업 파일에 대한 생명 주기 정보를 기록한다.

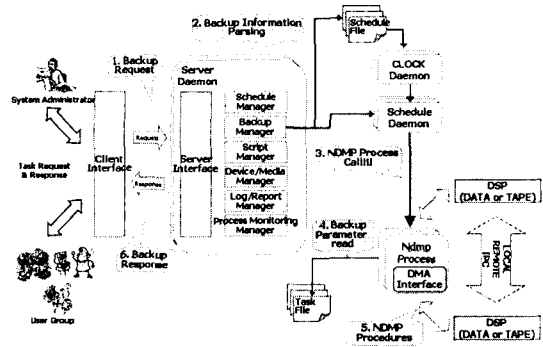


그림 3 백업 처리 과정

백업 작업을 수행하기 위해서 사용자는 백업된 데이터를 저장할 저장 공간을 확보해야 한다. 구현된 백업 소프트웨어에서는 다수의 디스크 장치를 풀(Pool)이라는 하나의 논리적인 공간을 사용하여 각 사용자에게 할당한다. 백업 서버는 장치 및 미디어 관리자를 통해 시스템에 부착된 디스크 백업 장치를 검색하고 사용 가능한 장치 목록을 지정한다.

사용자는 백업의 이전 또는 이후에 수행할 작업을 스크립트 관리자를 통해 작성할 수 있다. 사용자가 작성한 스크립트는 백업을 수행하는 과정에서 백업 서버에 의해 백업 전 또는 후에 처리할 작업을 수행시킨다. 사용자는 백업을 수행하면서 백업된 데이터

에 유지 기간을 설정할 수 있다. 이러한 유지 기간을 생명 주기라 한다. 백업된 데이터는 사용자가 요청한 생명 주기에 의해 관리된다. 만약 사용자가 백업에 대한 생명 주기를 설정하였다면 백업 서버는 생명 주기가 종료되는 작업을 확인하고 해당되는 백업 파일 및 백업에 관련된 히스토리 정보를 모두 삭제한다.

실제적인 백업 작업은 NDMP 프로토콜에 의해 수행되며 백업 서버에는 DSP들을 제어하고 통신할 수 있는 DMA(Data Management Applications)를 포함하고 있다. 백업 서버에서 DMA는 하나의 프로세스에 대응된다.

3.3 복구 처리 과정

클라이언트 인터페이스를 통해 전달된 사용자의 복구 요청에 따라 백업된 내용에 대한 전체 또는 부분 복구를 수행한다. 그림 3은 복구 처리 과정을 나타낸 것이다. 복구할 정보는 서버의 복구 관리자와 인덱스 관리자를 통해 결정한다.

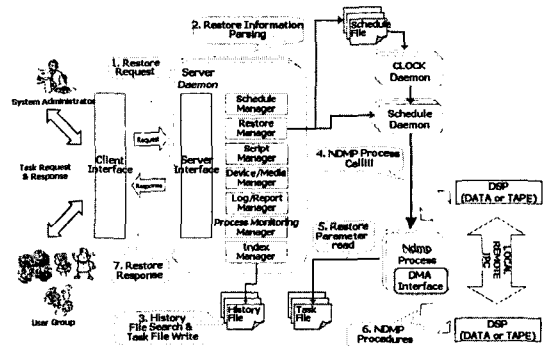


그림 4 복구 처리 과정

인덱스 관리자는 백업 작업 후 기록된 파일 히스토리 정보를 검색한다. 사용자는 인덱스 관리자를 통해 검색된 히스토리 정보를 바탕으로 복구할 파일 목록을 지정할 수 있다. 사용자가 지정한 파일 목록이 서버의 복구 관리자에게 전달되면 복구 관리자는 사용자로부터 전달받은 정보를 이용하여 복구 작업 파일을 작성한다. 작성된 작업 파일을 파라미터로 하여 스케줄 데몬은 NDMP 프로세스를 호출하고 복구 작업을 수행한다. NDMP 프로세스는 백업 파일이 저장된 사용자 풀로부터 백업 파일을 읽어 지정된 복구 위치에 복구 작업을 수행한다. NDMP 프로세스가 모든 작업을 종료하면 로그 및 리포트 관리자를 통해 복구 과정에서 발생한 로그 정보를 기록하고, 사용자에게 작업 결과를 반환한다.

백업 작업과 마찬가지로 복구 작업은 클릭 대몬을 통해 스케줄 관리자가 지정한 주기에 스케줄 대몬을 거쳐 수행되는 예약 복구와 직접 스케줄 대몬을 거쳐 수행되는 즉시 복구로 나눌 수 있다. 또한 백업 작업과 마찬가지로 복구 작업을 수행하는데 필요한 전, 후처리 작업은 스크립트를 사용하여 처리할 수 있다.

4. 제안하는 백업 소프트웨어의 특징

4.1 디스크 기반 백업

기존의 테이프 기반의 백업은 장치의 I/O속도와 순차적인 I/O에 의해 속도가 많이 저하된다. 이에 비해 디스크 기반의 백업은 I/O속도가 향상되며 동시에 하나의 디스크에 백업이 가능하다. 디스크 기반의 백업은 원본 데이터에 대한 백업 장치로 사용되며 테이프에 백업하기 위한 중간 단계로도 사용된다. 이는 백업 속도를 향상 시키고 백업이 동작하지 않는 시간을 이용하여 비교적 백업 수행 시간이 느린 테이프에 백업을 함으로써 보다 효율적인 백업을 가능하게 한다.

4.2 생명주기(Life Cycle) 기능

디스크 기반의 백업의 경우 더 이상 사용되지 않는 백업 이미지들이 디스크에 남아있을 수 있다. 사용되지 않는 오래된 백업 이미지들은 저장 공간을 낭비하여 디스크의 가용성을 떨어트리고 추가적인 저장 공간을 요구한다. 따라서 이들 백업 이미지들마다 적절한 보관 기간을 명시하고 보관 기간이 지난 백업 이미지들에 대해서는 2차 백업 혹은 백업 이미지 삭제와 같은 관리 작업을 수행하는 기능이 필요하다. 개발한 백업 소프트웨어에서는 백업 이미지에 대해 보관 기간을 명시하고, 보관 기능이 지난 백업 이미지를 관리하는 생명 주기 기능을 제공하여 디스크의 저장 공간 낭비를 제거하고 가용성을 높인다.

4.3 장치 및 미디어 관리 기능

백업 시스템에서 SAN 클러스터 내에 존재하는 백업 장치 및 미디어를 관리 기능은 전체 백업용 장치를 효과적으로 사용할 수 있게 하고 백업 과정에서 발생할 수 있는 용량부족으로 인한 에러를 줄일 수 있게 한다. 이를 위해 먼저 각 사용자별 백업용 장치 및 풀(Pool)을 설정할 수 있게 한다. 이것은 사용자는 자신이 관리하는 시스템이 백업될 공간을 미리 확보하고 확보된 저장 공간을 본인이 스스로 관리하게 함으로써 백업 수행에 대한 신뢰성을 높일 수 있게 된다. 또한, 사용자 풀을 백업된 데이터의

크기가 큰 전체백업을 위한 전체백업(Full Backup) 풀과 전체 백업이 아닌 비교적 용량이 작은 백업에 이용하는 non-전체백업(non-Full Backup) 풀로 나누어 관리하도록 함으로써 디스크의 용량을 효과적으로 사용할 수 있도록 한다. 사용자는 본인에게 할당된 백업용 공간을 관리하며 백업을 수행할 수 있도록 하는 기능을 제공하고 있다. 또한 관리자는 디폴트 풀을 설정하여 각 사용자 별 백업 풀/장치의 관리 소홀로 인해 발생하는 문제를 해결할 수 있다. 즉, 지정된 풀의 용량이 부족한 경우에 디폴트 풀로 대체해 백업을 계속 수행할 수 있게 한다.

4.4 Pre/Post 스크립트 기능

Pre/Post 스크립트 기능은 백업 또는 복구를 수행하기 이전 또는 이후에 수행할 작업들을 사용자가 미리 정의하여 작성한 스크립트나 프로그램을 사용하여 작업을 처리하는 기능이다. Pre 스크립트는 백업이나 복구 작업이 수행되기 이전에 수행해야 할 여러 작업들을 처리할 수 있으며, Post 스크립트는 백업이나 복구가 수행된 이후 처리해야 하는 여러 작업들을 처리할 수 있다. Pre/Post 스크립트를 사용함으로써 사용자는 소프트웨어 측면에서 자동적으로 지원하기 힘든 작업 또는 사용자의 의사가 많이 반영되는 여러 작업들을 수행할 수 있다.

5. 결론 및 향후연구

본 논문에서는 네트워크 기반의 백업을 위해 제안된 개방형 표준 프로토콜인 NDMP를 확장하여 디스크 기반을 수행하는 백업 소프트웨어를 설계하고 구현했다. 제안하는 백업 소프트웨어는 사용자의 스케줄링에 의한 자동 백업과 다양한 유형의 백업 및 복구 타입을 지원하며, 백업 파일들에 대한 생명주기 기능, 스크립트 기능, 장치 및 미디어 관리 기능을 지원한다. 향후 연구 방향으로 다양한 시스템 환경에 적합한 백업 소프트웨어를 개발할 예정이다.

참고문헌

- [1] A. L. Chervenak, V. Vellanki, Z. Kurmas and V. Gupta, "Protecting File System, A Survey of Backup Techniques", Pro. Joint NASA and IEEE Mass Storage Conference, 1998
- [2] NDMP, "http://www.ndmp.org"
- [3] NetBackup, "http://www.veritas.com"
- [4] Networker, "http://www.legato.com"
- [5] Time Navigator, "http://www.atempo.com"