

# 데이터 그리드를 이용한 초고전압 투과전자현미경

## 데이터베이스 시스템

안영현<sup>o</sup> 권희석 김윤중  
한국기초과학지원연구원 전자현미경팀  
{yhahn<sup>o</sup>, hskweon, y-jkim}@kbsi.re.kr

### High Voltage Electron Microscopy DataBase System Using Data Grid

Young-Heon Ahn<sup>o</sup>, Hee-Seok Kweon, Youn-Joong Kim  
Electron Microscopy Team, Korea Basic Science Institute

#### 요 약

바이오 및 재료 분야 등 NT(Nano Technology), BT(Bio Technology)에 관련된 연구를 보다 더 수월하게 하기 위해 본원에 설치되어 있는 초고전압 투과전자현미경(High Voltage Electron Microscopy - 이하 HVEM)을 이용한다. HVEM을 통해 획득한 이미지의 정보는 매우 방대하여 하나의 시료를 관찰하는데 있어서 수백 메가 이상의 용량을 차지하고, 연구를 수행하는데 있어서 데이터를 여러 형태로 관찰 분석하기 때문에 수월한 지원을 위해 체계적으로 관리할 수 있는 데이터베이스 시스템이 필요하다. 그러나 일반적인 범용 데이터베이스로는 이러한 대규모의 데이터를 저장할 수 없다. 따라서 본 논문에서는 이러한 용량 데이터를 체계적으로 관리할 수 있도록, 데이터 그리드와 연구 데이터의 정보를 갖는 metadata 테이블을 통해 서로 먼 거리에 있는 연구원들이 데이터를 접근하고 대규모 저장 공간을 갖는 데이터베이스 시스템을 제안한다.

#### 1. 서 론

90년대 중반에 등장한 그리드는 기존의 병렬/분산 컴퓨팅이 가지고 있는 성능과 확장성, 그리고 지리적인 한계를 극복하여, 네트워크로 상호 연동된 컴퓨팅 자원들을 각 자원을 보유하고 있는 기관의 자율성을 침해하지 않으면서 온라인상에서 안전하게 사용할 수 있는 환경을 말한다[1]. 기존의 병렬/분산 시스템 분야는 작은 컴퓨팅 자원을 여러 개 연결하여 보다 큰 컴퓨팅 성능을 제공하기 위한 방법으로 발전해 왔지만, 범용성이나 단일 시스템 이미지 등을 제공하기에는 많은 한계가 있었다. 그리드 컴퓨팅은 병렬/분산 컴퓨팅의 연장선상에서 이러한 한계를 극복하기 위하여 대응하고 발전해 온 기술이다[2].

그리드 컴퓨팅은 주로 과학 포탈, 분산 컴퓨팅, 대용량 데이터 분석 등에 응용되고 있다[2]. 본원에 장비된 HVEM의 경우 NT(Nano Technology) 및 BT(Bio Technology) 연구를 위해 매우 중요한 장비이며 이를 통하여 얻는 데이터도 보통 수백 메가 - 수 기가의 용량을 차지한다. 이러한 데이터의 효과적인 관리를 통해 연구원들이 쉽게 연구를 할 수 있도록 데이터베이스 시스템을 구축할 필요가 있다. 본원은 2003년부터 HVEM 장비의 원격거리 사용자들을 위해 그리드 컴퓨터를 이용한 원격제어시스템을 시범 운영한 바 있으며, 원격 연구원들의 원활한 연구자원을 위해 데이터 그리드를 이용하여 데이터베이스 시스템을 설계하였다.

본 논문의 순서는 다음과 같다.

2장 관련 연구에서는 본원에 설치된 HVEM 및 데이터 그리드 시스템에 대해 설명한다.

3장 데이터베이스 설계에서는 그리드 시스템을 이용한 데이터베이스 설계를 설명한다.

마지막 4장 결론 및 향후계획에서는 본 데이터베이스 구현

방향 및 향후과제에 대해서 설명한다.

#### 2. 관련 연구

전자현미경은 1931년 독일 베를린공대의 에른스트 루스카와 막스 크놀이 광학 현미경과 같은 형태의 현미경에서 가시광선 대신 전자빔을 이용하여 만든 것이 시초다.

투과전자현미경은 높은 진공 상태에서 고속으로 가속되는 전자선을 광원으로 이용하여 이 전자선이 표본을 투과, 형광판이나 필름에 영상이 나타난다. 기본적으로 광학 현미경과 유사하지만 광선 대신 전자빔을 사용하며 유리 렌즈 대신 빔을 집중하기 위해 자기코일을 사용한다. 전자빔이 표본을 통과해야 하므로 표본은 매우 얇아야 한다. 가속전압에 따라 투과력이 달라지는데 보통 가속 전압 200-300 KV의 현미경은 0.2 $\mu$ m의 시료를 투과할 수 있다[3]. 본원에 갖고 있는 전자현미경의 경우 모델명 JEM-ARM1300S, 1250 KV의 가속전압에 의해 얻어지는 고 투과력으로 생물 시료의 경우 3-5 $\mu$ m의 투과할 수 있는 고성능의 투과전자현미경이다.

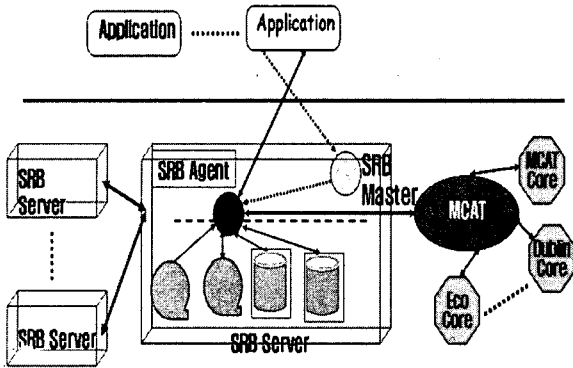
그리드 컴퓨터는 크게 계산 그리드, 데이터 그리드, 액세스 그리드 3가지로 구분할 수 있다.

- 계산 그리드 : 그리드 시스템으로 묶여져 있는 컴퓨팅을 보다 많은 자원을 연결하여 작업 전체의 수행 시간을 줄여준다.
- 데이터 그리드 : 분산된 자료를 통합해 분석할 수 있게 해주는 그리드로 여러 곳에 분산되어 있는 데이터를 하나의 저장 장치처럼 처리할 수 있다.
- 액세스 그리드 : 분산 처리를 필요로 하는 어플리케이션을 위한 그리드이다[4].

데이터 그리드는 서로 떨어져 있는 기관과 기관의 데이터를 공유할 수 있는 편리한 기능을 제공한다. 그리고 현재 보유

하고 있는 시스템에 추가 시스템을 연결하여 클러스터 화가 가능하게 하는 장점이 있다.

SRB는 Storage Resource Broker의 약자로서, 미국 샌디에고 슈퍼컴퓨터센터(San Diego Super Computer Center - 이하 SDSC)에서 개발한 데이터 그리드 미들웨어이다. SRB는 네트워크 상에 존재하는 Heterogeneous Data Resource, 예를 들어 High Performance System(이하 HPSS), 다양한 DBMS(DB2, Oracle, Postgress 등) 그리고 File System에 대해서 일관된 인터페이스를 제공함으로써 데이터에 쉽게 접근하고 복제와 저장에 관해 여러 가지 기능을 제공한다[5].

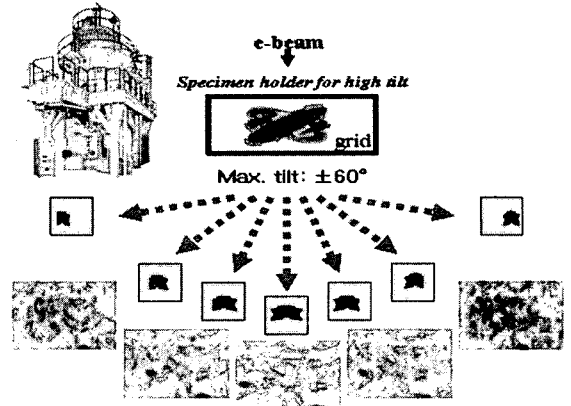


[그림 1] SRB Architecture

[그림 1]은 SRB의 구조이다. SRB는 MCAT(Metadata Catalog)와 SRB 서버로 구성된다. MCAT은 서로 연결되어 있는 SRB 서버의 정보 및 저장된 파일 및 유저에 대한 정보를 갖고 있다. MCAT은 일반적인 DB로 구성되어 있으며, SRB에 저장되어 있는 데이터에 접근하기 전에 먼저 MCAT을 통하여 접근하게 된다. SRB 서버는 SRB 마스터, SRB 에이전트와 데이터 저장소로 구성되어 있다. 마스터는 외부의 명령을 입력받는 역할을 하며, 에이전트는 마스터로부터 받은 명령을 MCAT에게 전달하고, 데이터 저장소로부터 데이터를 입출력하는 역할을 담당한다. 하나의 SRB 서버를 Container라고 부르며, 하나의 MCAT은 여러 개의 SRB 서버를 관리할 수 있으며, 하나의 MCAT이 관리하는 SRB 서버들의 집합을 Zone이라고 한다. Zone과 Zone은 서로 인증을 통해서 데이터를 공유할 수 있다.

### 3. 데이터베이스 설계

HVEM은 X, Y축 각도를 조절하여 관찰하고자 하는 이미지를 얻는다. 이를 통해 얻어지는 데이터는 약 2048 \* 2048이상의 영상 이미지이다. HVEM은 Y축을 고정하고, X축만을 -60° ~ +60°로 약 1도 간격으로 촬영하여 동일한 Y축에서 관찰하는데 이것을 Single Tilting 이미지라고 한다. 이러한 이미지는 2차원 영상으로써 3차원 영상으로 변환할 수 있으며, 자세한 3차원 영상 정보를 관찰하기 위해서 Y축을 변환하여 고정 후 다시 X축의 각도를 움직이며 사진을 촬영한다. Tilting에 관한 정보가 많을수록 관찰하고자 하는 3차원 영상의 정보는 좀 더 자세하게 얻을 수 있다. [그림 2]는 2차원 이미지를 통해 3차원 영상으로 변환하는 과정을 보여준다.

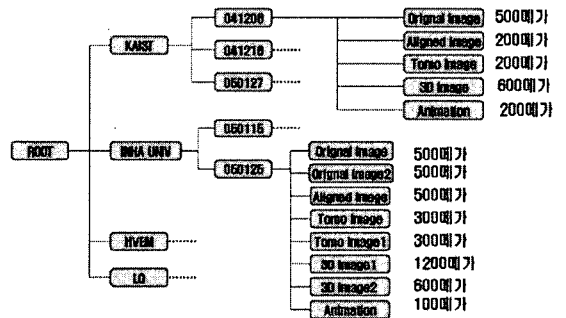


[그림 2] 2차원 영상을 통한 3차원 영상 변환 과정

[그림 3]은 HVEM을 이용하여 획득하고 3D로 변환하는데 필요한 데이터들을 저장한 SRB 구성도이다.

첫 번째 폴더를 보면 HVEM을 이용하여 데이터를 얻은 연구 기관들의 이름이며, 두 번째 폴더의 경우 이 영상을 얻은 날짜 정보이다.

세 번째 폴더의 경우 Original Image라는 이름을 갖는 폴더는 HVEM으로부터 얻은 데이터, Aligned Image의 경우 Original Image를 3차원으로 변환할 수 있도록 정렬과정을 거친 데이터이다. Tomo Image는 Aligned Image로부터 얻은 최초의 3차원 영상 데이터이다. 이 Tomo image의 경우 Aligned image로부터 프로그램을 통해 실사 영상으로 얻어진다. 3D Image의 경우 Tomo image로부터 관찰하고자 하는 부분중 관찰하고자 하는 부분만 추출하는 것이다. 이상의 모든 영상 데이터를 avi 등 애니메이션 형태로 저장하는 수 있는데 이 폴더를 Animation으로 하였다. 하나의 시편 데이터를 통해 여러 개의 Tomo Image 및 3차원으로 변환을 하는데 이에 대한 체계적인 관리가 필요하다.



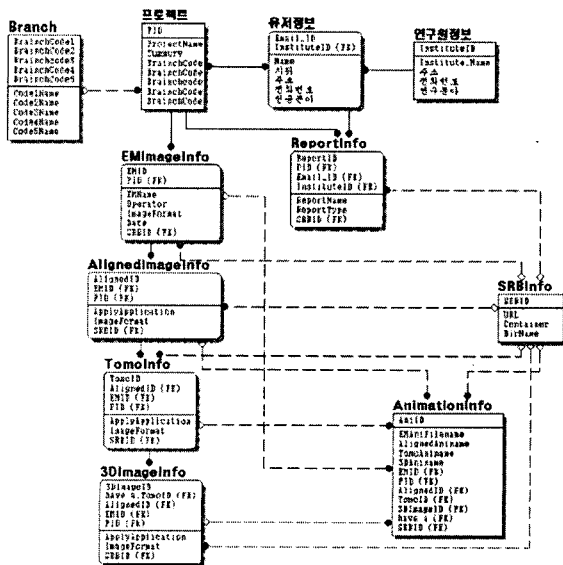
[그림 3] HVEM을 통해 얻은 이미지 저장도 (SRB System)

이러한 데이터는 하나의 연구 단위 형태로 관리할 수 있다. 하나의 연구를 위해서는 여러 개의 시편 데이터를 관찰하고, 여러 개의 3차원 데이터를 얻을 수 있으며 이를 통해서 결과물(논문 또는 보고서)이 생성되고, SRB에 저장될 수 있다.

따라서 위의 실험방법을 만족하도록 요구 조건을 설계해야

한다.

- 1) 사용자가 속한 연구기관과 연구정보를 알 수 있다.
- 2) 하나의 연구에 참여하는 유저와 기관정보를 알 수 있다.
- 3) 현재 연구하는 분야의 정보를 알 수 있다.
- 4) 연구를 위해 HVEM은 여러 개의 시편자료를 촬영할 수 있으며 이에 대한 자료를 쉽게 검색할 수 있어야 한다.
- 5) 시편자료는 어떠한 파일 형식으로 저장되어 있는지 이것을 촬영한 유저는 누구인지 시편의 이름은 무엇인지 알 수 있어야 한다.
- 6) 하나의 시편 이미지를 통해 여러 개의 정렬된 이미지를 찾을 수 있다. Tomo를 위한 응용프로그램이 여러 개가 있기 때문에 이 응용프로그램이 사용하기 위한 Aligned 데이터의 형태도 다르다.
- 7) Aligned 및 3차원 영상화를 위해 사용된 응용프로그램의 정보를 알 수 있다.
- 8) 하나의 Aligned image로 여러 개의 Tomo image로 변환될 수 있다. 이 image는 사실 image 형태로 변환되어 진다. 현재 많이 사용되는 Tomo 응용프로그램의 경우, gtomo[6], imod[7] 등의 프로그램이 있다.
- 9) Tomo image를 통해 유저가 원하는 부분만을 다시 3차원으로 재구성한다. 유저는 필요한 정보만을 추출한다.
- 10) HVEM을 통해 촬영된 이미지, 정렬이미지, Tomo 이미지, 3D 이미지의 경우 애니메이션 형태로 저장할 수 있다. 이에 대한 정보를 알 수 있다.
- 11) 10)에서 언급한 이미지 및 애니메이션은 파일 형태로 데이터 그리드 즉 SRB 시스템에 저장된다. 따라서 이들을 검색할 수 있는 metadata 정보가 필요하다.
- 12) 11)을 통해 저장된 파일정보를 통해 유저는 자신이 원하는 데이터를 검색 및 추가, 삭제, 수정이 가능하다.
- 13) 하나의 연구를 통해 나온 결과물(논문, 연구보고서) 등에 대한 정보를 검색하거나, 추가, 삭제할 수 있다.
- 14) 13)에서 언급된 결과물 또한 데이터그리드를 통해 저장 가능하다.



[그림 4] 데이터그리드-HVEM 데이터베이스 모델

[그림 4]는 ERWin[8]으로 작성한 데이터베이스 모델이다. SRB는 네트워크상으로 연결된 시스템의 저장 장치를 하나의 저장장치처럼 사용할 수 있도록 해준다. 따라서 실험데이터를 저장할 수 있고, 데이터베이스는 이에 대한 저장 위치를 갖고 있어야 한다.

'프로젝트' 테이블은 'Branch' 테이블을 통해 자신이 속한 연구 분야를 알 수 있다. 이 '프로젝트'에 참여한 유저들의 정보를 알 수 있고, 이 '프로젝트'를 하면서 획득한 실험 데이터 및 결과물 등에 대해서도 검색이 가능하다.

실험 데이터 및 결과물은 'SRBInfo' 테이블의 키를 외래키로 갖고 있어서 자신들의 실제 데이터의 위치를 알 수 있다. SRB는 이 위치정보를 통해서 사용자가 원하는 데이터를 인터페이스를 통해서 사용자에게 전달해주게 된다.

#### 4. 결론 및 향후 계획

이상으로 데이터 그리드를 이용한 HVEM을 통해 얻은 데이터베이스 시스템 설계를 해 보았다. 데이터 그리드 이클래스인 SRB는 이 기존 데이터베이스 및 파일 시스템을 갖는 기관과 기관의 데이터를 쉽게 서로 공유할 수 있도록 공통의 인터페이스를 제공하여 준다. 본 논문에서 보듯이 HVEM을 통해 나온 데이터의 용량 및 정보는 상당한 저장 용량이 필요로 하고, 일반적인 데이터베이스로는 이러한 데이터를 저장할 수 없다. 따라서 본 시스템을 이용하면 이러한 문제점을 해결할 수 있다. 특히 데이터 그리드를 사용하여 원거리의 유저들끼리 마치 하나의 저장장치를 사용하도록 할 수 있으며 쉽게 저장 및 검색 등 범용 데이터베이스와 같은 시스템을 구성할 수 있다.

본 논문에서 언급한 시스템을 갖고 데이터베이스를 설계하고 구축하면 이러한 대규모 데이터를 체계적으로 정리하고, 유저들이 원하는 대로 검색이 가능하여 데이터 그리드를 통해 연결된 시스템에 접근하여 언제 어디서든지 원하는 데이터를 빠르고 쉽게 찾아서 BT, NT 연구 활동에 도움을 줄 수 있다.

향후 계획으로는 본 설계를 토대로 실제 구현을 통해 유저들이 데이터베이스를 통해 쉽게 데이터에 접근하는지 알아볼 필요가 있다. 무엇보다도 SRB시스템과 데이터베이스 시스템을 연동하고 웹 어플리케이션 형태로 제작하여 그리드 시스템으로 연결된 연구원들이 웹 페이지를 통해서 데이터베이스를 쉽게 검색 및 저장 등이 가능하도록 이에 대한 구현이 필요하다.

#### 참고문헌

- [1] I. Foster, C.Kesselman, and S. Tuecke, "The Anatomy of the Grid: Enabling Scalable Virtual Organization," International J. Supercomputer Applications, 15(3), 2001.
- [2] 최재영, "웹 서비스를 적용한 그리드 서비스", 정보과학회지, 제22권 제 10호, 2004년 10월, pp54-59.
- [3] 대전매일신문 2003년 4월 21일.
- [4] 이정훈, 오영은, 김정석 공저, "리눅스 클러스터로 만드는 슈퍼 컴퓨터", 영진닷컴.
- [5] 국가그리드기반구축사업보고서(III), 한국과학기술정보연구원, 2004년 1월.
- [6] gtomo, <http://juggler.ucsd.edu/gtomo/>
- [7] imod, <http://bio3d.colorado.edu/imod/>
- [8] ERWin, ER Data base Moudler, <http://ca.com/>