

## 화일색인에 의한 대용량공간자료봉사방법

박 정 호

망지리정보체계개발에서 대용량자료봉사는 매우 중요한 문제의 하나로 제기된다. 일반적으로 망지리정보체계들에서는 기존자료기지봉사프로그램들을 리용하여 자료봉사를 진행하고있다. 그러나 이것은 수십TB이상 되는 초대용량공간자료봉사에는 적합하지 않은 것으로 되고있다.

론문에서는 이러한 문제를 해결하기 위하여 화일색인방법에 의한 초대용량공간자료봉사방법에 대하여 서술한다.

### 1. 문 제 설 정

망환경에서의 자료봉사를 제공하는 기존프로그램들은 수없이 많이 출현하여 보급리용되고있다. 그중에서 제일 많이 리용되는 대표적인 망자료기지봉사프로그램들을 보면 Oracle, MySQL, Post grid SQL, Microsoft SQL Server 등이다.[1, 2]

최근 인터넷봉사기에서 가장 많이 리용되는 자료기지는 Oracle로 알려져있다. Oracle은 분산망환경과 Grid망환경에까지 대응하도록 설계되고 개발되고있으며 현재까지 개발된 망자료기지봉사프로그램들중에서 가장 큰 자료관리능력을 지원하고있다.

그러나 이러한 기존프로그램들을 리용하여 대용량공간자료봉사를 실현하는데는 일정한 제한성이 있다. 그것은 첫째로, 자료용량이 수십TB이상인 경우 이것을 Oracle자료기지로 넘기는것은 시간과 비용의 측면에서 비효율적이다.

자료용량이 수십TB이상인 공간자료를 Oracle자료기지로 넘기자면 우선 원본자료의 용량을 초과하는 물리적기억매체가 있어야 한다. 또한 수억개이상의 레코드를 만들어 Oracle자료기지의 표안에 추가해야 한다. 이것은 하루에 100만개의 레코드를 생성한다고 해도 100일 즉 3달이라는 시간이 필요된다는것을 의미한다.

둘째로, 자료기지의 대피(Backup)와 복구(Restore)가 매우 어렵다.

Oracle에 들어가있는 수십TB이상의 공간자료를 여벌복사하자면 또 그만 한 용량의 물리적기억매체가 있어야 하며 그 시간도 매우 오래다. 그리고 자료기지회복에도 적지 않은 시간이 요구된다. 실험에 의하면 15TB의 지리공간자료를 Oracle을 리용하여 대피하고 복구하는데는 10일이상의 시간이 걸렸다. 일반적으로 자료기지를 대피하고 복구하는 동안에는 자료기지봉사를 할수 없는것으로 되어있다. 결국은 10일동안 자료봉사를 진행할수 없는것으로 된다. 결과적으로 Oracle과 같은 기존자료기지봉사프로그램들을 리용하여 초대용량공간자료봉사를 실현하자면 비용이 많이 들고 자료의 관리와 유지보수가 매우 어려운 부족점을 가진다. 이러한 문제로 하여 현실에서 초대용량공간자료봉사를 기존자료기지봉사프로그램을 리용하여 실현하지 못하고있다.

## 2. 화일색인에 의한 대용량공간자료봉사방법

우리는 기존자료기지봉사프로그램을 리용하는데서 제기되는 제한성을 자료화일 그 자체를 리용하여 공간자료를 봉사하는 방법을 서술하였다. 즉 화일들을 구조화하고 색인화하여 검색의 신속성을 보장하면서도 자료관리와 복구, 대피를 손쉽게 진행하자는 것이다. 화일색인이란 공간자료화일에 들어있는 공간자료단위들의 색인정보 다시말하여 레코드들의 식별번호(ID)와 그 레코드가 위치하고있는 화일에서의 자료위치(Byte Offset)의 모임을 말한다.

매 자료화일들에는 색인정보를 가지고있는 색인화일들을 대응시켜 실지 자료를 읽을 때에는 색인화일만을 먼저 읽어 해당하는 레코드의 위치를 얻은 다음 자료화일에서 그 레코드마당들을 읽어 자료를 봉사한다. 화일구조화 및 화일색인구축방법은 다음과 같다.

화일들을 구조화한다.

일정한 ID영역에 들어가는 자료들을 ID값에 따르는 순서대로 배열정돈하여 하나의 화일로 묶고 그 화일의 이름을 ID영역과 관련되게 만든다. 만일 이렇게 묶어지는 화일들이 많은 경우 화일검색속도가 떨어질 수 있으므로 ID묶음들을 또다시 몇개의 등록부로 갈라 등록부들을 만들고 그 등록부들에 해당하는 화일들을 만들어준다.

화일색인을 구축한다.

매 자료화일들의 ID들을 순서대로 배열정돈하여 그 정돈된 순서대로 화일에서의 자료위치(Byte Offset)만을 4B씩 쓰기한다. 경우에 따라 추가적인 정보를 함께 쓰기할 수 있다. 화일색인작성에서 중요한것은 매 레코드들의 색인크기가 같아야 한다는 것이다.

화일색인으로부터의 레코드검색방법은 다음과 같다.

우선 검색ID로부터 해당하는 등록부이름과 화일이름을 계산한다. 다음 색인화일을 열고 검색 ID에서 색인화일의 첫 ID를 덜어 색인화일에서의 색인정보위치를 계산한다. 그리고 색인정보에 있는 자료위치를 리용하여 자료화일로부터 레코드자료를 읽는다. 만일 색인정보에 추가적인 정보가 있는 경우 이것도 검색조건확인에 리용할 수 있다.

우와 같은 화일색인을 리용한 대용량공간자료봉사방법은 다음의 우점을 가진다.

첫째로, Oracle과 같은 기존봉사프로그램이 없이도 간단한 봉사프로그램을 자체로 개발하여 봉사를 진행할 수 있다.

둘째로, 자료보안을 보다 강화할 수 있다.

기존프로그램들을 리용하는 경우에는 자료보안의 믿음성을 그 프로그램에 맡겨야 한다. 그러나 자체로 봉사프로그램을 개발하여 리용하는 경우 보안을 보다 더 철저히 할 수 있다.

셋째로, 봉사기구축과 자료유지보수를 위한 비용을 줄일 수 있다.

화일색인을 리용하여 봉사를 구축하는 경우 자료화일 그자체가 자료기지로 되므로 추가적인 물리적기억매체가 없어도 되며 그 대피와 복구도 단순한 화일복사로 진행되므로 매우 신속하고 봉사를 끄지 않고도 진행할 수 있는 우점을 가진다.

넷째로, 자료검색속도를 기존프로그램들이상으로 보장한다.

ID에 의한 자료검색속도는 구조화된 화일들을 리용하여 진행하므로 기존프로그램들에서의 2진검색속도보다 더 빠르다.

이상과 같이 화일색인을 리용하면 특별한 기존자료기지봉사프로그램이 없이도 초대용량공간자료봉사를 보다 손쉽고 안전하며 편리하게 실현할수 있다.

우리는 화일색인에 의한 공간자료봉사를 초대용량화상자료를 리용하는 파노라마통합지리정보체계에 응용하였다.

7지역을 3~10m 간격으로 순회하면서 수집한 파노라마화상자료의 레코드개수는 1억 개이상이고 그 용량은 15TB에 달하였다. 이러한 자료를 1개 화일에 2 000개씩 ID순서로 묶어 50개의 등록부에 구조화하였다. ID번호는 1부터 1억까지이므로 검색하려는 ID로부터 등록부번호와 화일번호를 계산한 다음 해당한 레코드를 읽어 봉사하는 방법으로 파노라마공간자료봉사를 실현하였다.

### 맺 는 말

화일색인화방법에 대한 연구를 보다 심화시키면 ID에 의한 검색뿐만아니라 다른 속성 정보에 의한 효율적인 검색도 실현할수 있다.

### 참 고 문 헌

- [1] Computers & Geosciences, 31, 221, 2005.
- [2] Computers & Geosciences, 31, 199, 2005.

주체104(2015)년 6월 5일 원고접수

## Mass Storage Spatial Data Service by File Indexing

*Pak Jong Ho*

The spatial data service by file indexing is more effective than the ones by traditional data service programs such as Oracle, MSSQL, MySQL, etc which have some disadvantages in database building, backup and restoring.

The more works on file indexing would make it possible to support searching function by extra criteria as well as record ID.

Key words: file indexing, spatial data