

국토대상자료기지구축에서 MongoDB의 응용

곽 남 일

경애하는 김정은동지께서는 다음과 같이 말씀하시였다.

《국토관리와 환경보호부문의 과학연구기관들에서는 강성국가건설의 요구에 맞게 과학 연구사업을 힘있게 벌려 국토관리와 환경보호사업에서 나서는 과학기술적문제들을 제때에 높은 수준에서 원만히 풀어나가야 합니다.》

국토대상자료기지를 합리적으로 구축하는것은 국토건설과 계획작성의 과학화, 정보화를 실현하는데서 선차적인 공정이다.

지난 시기에는 국토대상자료기지를 주로 관계형자료기지에 기초하여 실현하기 위한 연구가 진행되였다.

그러나 국토대상자료와 같이 자료의 구조가 복잡하고 자료용량이 큰 경우 관계형자료기지의 실현에서는 일련의 제한성이 나타났으며 이 문제를 해결하기 위한 방도의 하나로 NoSQL자료기지가 널리 리용되고있다.[2, 3]

본문에서는 NoSQL자료기지의 하나인 MongoDB를 리용하여 국토대상자료기지를 구축하는데서 나서는 방법론적문제들에 대하여 서술하였다.

1. 국토대상자료기지의 설계요구

국토자료의 특성은 다음과 같다.[1]

첫째로, 국토자료는 시공간성을 가진다.

국토자료는 고정불변한것이 아니라 국토요소들의 변화와 함께 시계열적으로 변화된다. 또한 모든 국토건설대상들은 자기의 고유한 지리적위치를 가지고 공간에 분포되어있으므로 지리공간자료를 중요한 속성자료로 포함하고있다.

둘째로, 국토자료는 련관성을 가진다.

국토대상들은 독립적으로 존재하는것이 아니라 서로 일정한 련관관계를 맺고있으며 유기적으로 결합되어있다.

셋째로, 국토자료는 다양성과 다차원성을 가진다.

국토자료는 수자, 문건, 도면 등 형식이 다양하고 공간적으로 점, 선, 면 등의 형태로 다차원적인 특성을 가진다.

넷째로, 국토자료는 광범하고 방대한 특성을 가진다.

국토자료는 대상들이 광범하고 자료의 종류가 많으며 시공간적으로 부단히 변화되는 시계열자료인것으로 하여 자료의 수집과 처리에서 품이 많이 들고 량적으로 방대하다.

국토대상자료기지설계에서 나서는 요구는 다음과 같다.[1]

첫째로, 국토자료의 특성을 옳게 반영하여야 한다.

둘째로, 모든 국토요소자료들을 충분히 관리하고 신속정확하게 보장하며 그것에 대한 분석을 원만히 진행할수 있어야 한다.

셋째로, 다른 부문의 자료기지 및 정보체계들과의 자료교환(리용, 봉사)을 진행할수 있어야 한다.

2. MongoDB에 기초한 국토대상자료기지구축

1) 국토대상자료기지의 자료모형설계

국토대상자료기지는 국토관리사업의 특성으로부터 속성자료, 기준자료, 공간자료, 메타자료로 이루어진다.

속성자료는 필수적으로 입력하여야 할 필수항목과 대상의 특성에 따라 입력하여야 할 항목으로 나누어 구성되며 공간자료는 대상의 지리적위치와 도형형태(점, 선, 면 등), 지도현시를 위한 각종 부호정보 등을 반영하여 구성된다.

기준자료는 국토대상들에 대한 실태조사와 계획작성에서 나서는 기준지표들에 대한 자료로 구성되며 메타자료는 자료의 수집과 보관, 처리, 갱신, 관리, 용도, 첨부화일정보 등에 대한 자료로 구성된다.

2) 속성지표체계의 구성

국토대상자료기지를 구축하는데서 중요한것은 속성지표체계를 국토계획작성의 요구와 자료의 특성에 맞게 합리적으로 구성하는것이다.

우선 국토대상들을 그 특성에 따라 분류하고 부문분류체계를 작성하였다.(표 1)

여기서는 647개의 국토대상들에 대하여 1차분류에서는 혁명사적부문, 공업부문, 농업부문, 산림부문, 수역부문, 교통운수부문, 기술시설부문, 주민지 및 사회문화부문, 지하자원부문, 환경보호부문 등 10개로, 2차분류에서는 55개로 부문분류를 진행하여 분류체계를 확립하였으며 이에 기초하여 국토대상의 부문분류collection을 설계하였다.

표 1. 국토대상의 부문분류표 실례

번호	1차분류	2차분류	3차분류(국토대상)
			림소반
		산림토지	산불막이선
			⋮
1	산림		산림경영소
		산림시설	채종림사업소
			⋮
⋮	⋮	⋮	⋮

표 1에서 보는바와 같이 부문분류는 나무자료구조로 체계화되며 국토대상의 특성에 따라 분류나무의 높이와 차수를 변경시킬수 있다.

MongoDB는 문서형(document-oriented)자료기지로써 매물형문서(embedded document)와 배열(Array)을 리용하여 하나의 레코드로도 복잡한 계층구조를 표현할수 있는 우점을 가지고있다.

MongoDB를 리용하여 국토대상분류체계를 하나의 collection에 반영할수 있으며 마당은 표 2와 같이 구성한다.

표 2. 국토대상분류 collection의 마당

번호	마당이름	의미	자료형
1	_id	분류마디의 유일식별번호	int
2	info	분류마디의 명칭	string

표계속

번호	마당이름	의미	자료형
3	code_ancestors	웃준위분류마디의 프로그램정보	array
4	info_ancestors	웃준위분류마디의 명칭	string
5	level	분류준위	int
6	cfCode	같은 분류준위에서 유일프로그램정보	int
7	order	같은 분류준위에서 현시순서	int
8	isEnabled	리용상태정보	boolean
9	isLeaf	잎마디인가를 판정	boolean
10	isVisible	현시상태정보	boolean
11	note	비고	string
12	features	소속된 대상들의 정보	array

표 2에서 features마당은 해당 분류마디에 소속되어있는 국토대상들의 정보를 반영하며 features_id마당과 order마당을 가진 매물형문서(embedded document)의 배열로 이루어진다.

속성지표관리를 위한 collection은 국토대상들이 가질수 있는 속성지표마당들을 표준화하고 유일적으로 관리하기 위한데 목적을 두고 설계하였다.

속성지표관리를 위한 collection의 마당은 표 3과 같이 설계하였다.

표 3. 속성지표관리를 위한 collection의 마당

번호	마당이름	의미	자료형
1	_id	속성지표의 유일식별번호	int
2	info	속성지표의 명칭	string
3	data_type	자료형	string
4	data_length	자료길이	Int
5	data_precision	자료의 소수부길이	Int
6	unit_id	계량단위식별번호	Int
7	min_value	최소지표값	Int
8	max_value	최대지표값	Int
9	isEnum	속성자료유무상태	boolean
10	type	속성분류번호	Int
11	note	비고	string
12	isEnabled	리용상태정보	boolean
13	isVisible	현시상태정보	boolean
14	info_en	영문표기명칭	string
15	values	표준화된 입력값정보	array

표 3에서 보는바와 같이 values는 속성값으로 입력할수 있는 표준화된 값정보로서 매물형문서들의 배열로 구성된다.

그 아래준위마당은 표 4와 같다.

표 4. values마당의 아래준위마당

번호	마당이름	의미	자료형
1	value	속성값	Int
2	info	속성값정보	String
3	isEnabled	리용상태정보	Boolean
4	isVisible	현시상태정보	Boolean
5	note	비고	String

우와 같이 국토대상의 부문분류 collection과 속성지표관리collection을 설계한데 기초하여 국토대상별로 속성마당들을 관리하기 위한 collection에 대한 설계를 진행하였다.(표 5)

표 5. 국토대상의 속성마당관리collection

번호	마당이름	의미	자료형
1	_id	대상의 유일식별번호	Int
2	info	대상의 명칭	String
3	info_en	대상의 영문명칭	String
4	isPoint	점대상인가	boolean
5	isLine	선대상인가	boolean
6	isPolygon	면대상인가	boolean
7	isOverlap	중첩관계	boolean
8	doLanduse	토지리용상태	boolean
9	isEnabled	리용상태	boolean
10	isVisible	현시상태	boolean
11	symbol_info	기호정보	String
12	note	비고	String
13	fields	속성 항목	Array
14	geometry	공간속성	Array

표 5에서 보는바와 같이 fields마당은 국토대상이 가지고있는 속성 항목들에 대한 정보를 포함하고있는데 그 아래준위마당으로서는 field_id(속성마당의 id)와 order(현시순서)가 있다.

geometry마당은 국토대상의 지리적위치와 공간적인 도형형태를 반영하는 마당으로서 그 아래준위마당으로서 type(공간도형형태)마당과 coordinates(지리자리표)마당을 가진다.

type마당에서는 GeoJSON형식으로 대상도형의 형태를 지적하는데 여기에는 “Point”, “LineString”, “Polygon”, “MultiPoint”, “MultiLineString”, “MultiPolygon”, “GeometryCollection” 등을 설정할수 있다.

coordinates마당은 대상의 지리적위치를 반영하며 경위도자리표값들의 배열로서 표시된다. 이와 같이 국토대상들에 대한 자료기지를 설계한데 기초하여 국토대상자료기지를 구축하였다.

MongoDB에 기초한 국토대상자료기지는 종전에 리용하던 관계형자료기지와는 달리 매물형문서를 리용하여 표의 개수를 최대한 줄이고 빈값자료에 의한 기억용량의 낭비를 감소시켰으며 자료의 검색, 삭제, 갱신조작 등 자료의 처리속도를 현저히 높일수 있게 되었다.

분석실험은 255 000개의 림소반자료가 포함되어있는 전국의 림소반대상자료기지를 가지고 관계형자료기지(PostgreSQL)로 구축했을 때와 MongoDB로 구축했을 때의 처리속도를 비교하는 방식으로 진행하였다. Intel Core i5 2.5GHz 4GB의 성능을 가진 컴퓨터에서 PostgreSQL 9.6과 MongoDB 4.0을 리용하였는데 비교결과는 표 6과 같다.

표 6. 자료처리속도 비교

No.	지표	PostgreSQL	MongoDB
1	속성자료검색	3.512s	0.624s
2	자료삭제	0.925s	0.392s
3	자료갱신	0.003s	0.002s
4	공간분석	0.86s	0.27s

표 6에서 보는바와 같이 1번지표는 255 000개의 림소반자료에 대하여 관리단위, 주수종, 용도, 면적 등 항목별로 집계를 내는데 걸리는 시간으로서 MongoDB가 관계형자료기
지보다 약 5.6배정도, 삭제조작에서는 약 2.4배, 공간분석(완충구역분석)에서는 약 3.18배
정도 빠르다는것을 알수 있다.

맺 는 말

NoSQL자료기지의 하나인 MongoDB에 기초하여 국토대상자료기지를 구축하는데서 나
서는 방법론적문제를 해결함으로써 대용량국토대상자료를 보다 합리적으로 구축하고 관리
할수 있게 되었으며 자료의 용량을 줄이고 자료처리성능을 현저히 높일수 있게 되었다.

참 고 문 헌

- [1] 리영성; 국토정보처리, 김일성종합대학출판사, 1~43, 주체104(2015).
- [2] 윤주환; 컴퓨터와 프로그래밍기술, 1, 34, 주체107(2018).
- [3] 刘亚 等; 基于MongoDB的地名信息管理, 测绘通报, 10, 117, 2014.

주체110(2021)년 4월 5일 원고접수

Application of MongoDB in Construction of Land Object Database

Kwak Nam Il

We studied the methodological problems for construction of land object database using MongoDB which is one of the NoSQL database.

Keywords: land object database, non-relational database, NoSQL, MongoDB