



한국정보과학회

KOREAN INSTITUTE OF INFORMATION SCIENTISTS AND ENGINEERS

<http://www.kiise.or.kr>



한국정보과학회

제40회 정기총회 및 추계학술발표회

2013년 11월 15일(금)~16일(토)

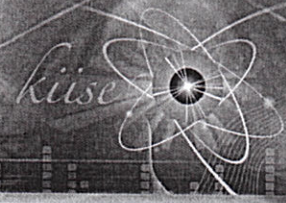
제주대학교

<http://www.kiise.or.kr/conference02/>



2013 추계학술발표회 논문집

2013년 11월 15일~16일, 제주대학교



48. Intelligent SSD에서 이질 스케줄링 알고리즘 적용을 통한 프로그램 성능 향상 방안	조성우 · 조용연 · 배덕호 · 김상욱 · 오현옥	292
49. 과학기술 연관관계 추출을 위한 테스트 컬렉션 구축	황미영 · 정창후 · 이승우 · 김순영 · 최성필 · 정한민	294
50. 인터넷 게시판 분류에 따른 사용자 연결망 특성 분석	탁해성 · 조환규	296
51. 국가 현안 주제 선정 및 키워드 추출 방법론	홍진성 · 최희석 · 한희준 · 김재수 · 유은지 · 임소라 · 김남규	299
52. 고차원 멀티미디어 데이터의 인덱싱을 위한 Locality Sensitive Hashing	문병문 · 홍지원 · 정교성 · 김상욱	302
53. 국방 M&S 가상 시험 연동을 위한 효율적인 DB설계	최성우 · 이예나 · 최미선 · 김영국	304
54. SAM 파일을 활용한 유사 프로파일을 가지는 리드 생성기	이상민 · 탁해성 · 박기정 · 조환규 · 이도훈	307
55. 의료환경에서 데이터 객체의 민감도를 고려한 상황인식 접근제어 기법	이용우 · 정강수 · 박 석	310
56. 지시적 분석을 위한 연구자 히스토리 추출 방법	김장원 · 황명권 · 송사광 · 김진형 · 정도현 · 정한민	313
57. Classification of Discretization Schemes	Rahman Ali · Muhammad Hameed Saddiqi · Donghai Guan · Sungyoung Lee	315
▶ 학부생논문		
58. OLTP환경에서 SSD의 성능 분석	정승훈 · 이재성 · 강윤석 · 조용연 · 배덕호 · 김상욱 · 강주영 · 차재혁	318
59. 온라인 뉴스 서비스에서 뉴스 추천 방법의 정확도 평가	이정민 · 박 정 · 황원석 · 김상욱	320
60. 추천 방법의 정확도 향상을 위한 신뢰 네트워크 기반 대치 방법	김지곤 · 황원석 · 이수우 · 김상욱	322
61. 데이터 마이닝을 이용한 1, 2세대 전자식 디젤 엔진 성능향상 연구	안성식 · 이재영 · 안대현 · 이준환	324
62. [우수논문] 칼만 필터를 이용한 무선 센서 기반의 에너지 효율적 실시간 수질 모니터링 시스템	민세희 · 오연주 · 박 석	327
63. 빅데이터화를 위한 관계 데이터 모델의 빅데이터 테이블화	양효석 · 황준순 · 강건희 · 권하은 · 박보경 · 김동호 · 박용범 · 김영철	330

모바일응용및시스템

1. 탭 패턴 입력을 이용한 스마트폰 UX 설계 및 구현	김영빈 · 곽문상 · 김유희	333
2. [우수논문] 터치 기반 환경에서의 그래프를 이용한 데이터관리 스케치 인터페이스 시스템	박범준 · 정한길 · 전재용 · 최윤철	336
3. 가변적인 측정 간격을 사용하는 효율적인 이동 물체 파악 기법	최중현 · 윤태섭 · 조은선	339
4. 스마트폰 기반 행위인지 기법을 활용한 운동량 측정 및 행동 모니터링 시스템	이원주 · 이호성 · 이승룡	342
5. 스마트 기기를 이용한 악보인식에서 프리뷰 분석을 통한 악보 기율기 보정 기법의 설계 및 구현	김성룡 · 김택희 · 김미선 · 이보람 · 김근정 · 김계영 · 이상준	345
6. [우수논문] 결정 트리과 베이지안 네트워크를 이용한 모바일 상황공유 시스템	전기춘 · 조성배	347
7. XmlPullParser를 이용한 안드로이드 기반 무선 IVEF 프로토콜의 성능개선	김경환 · 김주영 · 이병길 · 박남제	350

빅데이터화를 위한 관계 데이터 모델의 빅데이터 테이블화

양효석⁰ 황준순 강건희 권하은 박보경 김동호 박용범* 김영철
홍익대학교 소프트웨어공학연구소
단국대학교*

{yhs⁰, Hwang, Kwon, Kang, bkpark, bob}@selab.hongik.ac.kr, ybpark*@dankook.ac.kr

Big Data Tablization with Relational Data Model for generating Big Data

Hyeo Seok Yang⁰, Jun Sun Hwang, Geon-Hee Kang, Ha En Kwon, Bo Kyung Park, Dong Ho Kim,
Y. B. Park*, R. Y. Kim

SELab, Dept. of. CIC, Hongik University

*Dept. of. Computer Science, Dankook University

요 약

현재의 데이터베이스시스템에서 빅데이터를 다루기 어려우며, 방대한 양의 데이터를 연산할 때 관계형 데이터베이스 모델(Relational Database Model)은 테이블 간의 Join 연산으로 인해 데이터 처리 시간이 오래 걸리는 단점이 있다. 본 논문에서는 기존 관계데이터모델의 빅데이터 테이블화 하고자 한다. 기존 관계데이터모델의 함수적 접근 패스(Functional Access Path) 메카니즘[1]적용하여 빅 데이터 테이블화 방안을 제안한다. 이는 기존 DB 정보의 재사용에 초점을 두고자 한다.

1. 서 론

빅데이터란 데이터의 생성 양, 주기, 형식 등이 기존 데이터에 비해 크기 때문에 수집, 저장 등이 어려운 방대한 데이터를 말한다. 최근 소셜 네트워크 서비스, 클라우드 컴퓨팅 등장과 스마트폰의 발달로 많은 양의 데이터가 생산되거나 소비되고 있다. 따라서 빅데이터 분석 작업은 IT 산업뿐만 아닌 모든 산업의 이익 창출 효과가 나타나고 있다. 또한 사회현상 및 소비자의 니즈와 시장 변화에 관한 새로운 시각이나 법칙을 발견할 가능성이 커졌다[1].

실제 기업들은 빅데이터를 활용하여 자사의 이익을 높이는데 사용하고 있다. 하루 440만 명의 방문자, 900만개의 상품을 주문하는 아마존은 1억 2,000만 명의 고객정보와 고객 상품 탐색 및 구매 내역 분석을 위해서 위와 같은 방대한 양의 데이터를 저장 및 처리하고 있다. 이는 단순 분석에서 그치지 않으며 사용자 정보처리를 통해 제안되는 '추천' 시스템은 아마존 전체 매출의 30%를 차지한다[1,2,3,4].

현재 빅데이터의 연구는 '저장되어진 빅데이터를 어떻게 하면 각 산업 혹은 기업에 이익이 될 수 있는가'에 대한 분석 혹은 관리, 분석된 빅데이터의 비전화, 빅 데이터 아키텍처, 하둡 분산파일/에코 시스템, 빅 데이터 Map reduce 등에 대한 연구들이 많이 진행되고 있다. 하지만 빅데이터의 구축방법 및 기존 관계 데이터베이스의 빅데이터 변환을 위한 관계 데이터베이스의 재사용에 대한 관점이 부족하다.

본 논문에서는 방대한 양의 데이터를 저장 및 처리했던 관계형 데이터베이스 모델의 재사용을 통해 기존 데이터베이스의 빅데이터 테이블화를 제안한다. 논문의 구성은 다음과 같다. 2장에서는 관련연구로써 기존 관계형 데이터베이스 테이블들의 Functional Access Path에 대해서 설명한다. 3장에서는 Functional Access Path로 나누어진 테이블들의 빅데이터 테이블화에 대해서 제안한다. 4장은 결론에 대해서 언급한다.

2. 관련연구

* 본 연구는 미래창조과학부 및 한국산업기술평가관리원의 산업원천기술개발사업[10035708, 고신뢰 자율제어 SW를 위한 CPS(Cyber-Physical Systems) 핵심 기술 개발]과 2013년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2013R1A1A2011601)

2.1 Dynamo

Dynamo는 AWS(Amazon Web Service)의 NoSQL 서비스이다. Key-Value 형태로 대용량의 데이터를 저장할 수 있으며, 고속의 데이터 접근을 제공한다. RDBMS처럼 테이블이라는 개념을 가지고 있으며 테이블은 테이블명과 각각의 ROW로 구성된다. 테이블은 Unique한 Primary Key를 가지고 있으며 이를 key라 정의한다. RDBMS와 다른점은 각 ROW는 똑같은 Columns을 갖는 것이 아니라, 각 row 마다 다른 Column을 가질 수 있는 것이다.

2.3 NoSQL의 컬럼-패밀리 저장소

컬럼-패밀리 데이터베이스는 칼럼을 칼럼 패밀리로 구조화한다. 특정 칼럼 패밀리에 대한 데이터는 보통 함께 접근된다는 가정하에, 각 칼럼은 칼럼 패밀리 하나의 일부가 되어야 하고, 칼럼은 접근 단위로 동작한다[5].

3. 본 론

포함 의존성(Inclusion dependency) 개념은 여러 의미적 구조(Constructs)들 예를들면, 일반성, 집합성, 참조 무결성)을 모델링에 유용한 역할을 한다. 또한 이 의존성 개념은 Complex object를 유추가 가능하다[6].

정의 1: A,B 가 릴레이션 일 때, A는 B에 종속(subordinate)한다. 조건: 임의의 X(속성)에 대해, X가 A내 속성이며, 임의의 Y(속성)에 대해, Y가 B내 속성이고, $X < Y$ 이다.

정의 2: 종속 의존성 그래프 (A Subordinate dependency graph)는 방향성 그래프 $G(V,E)$ 이다. 조건:V는 하나의 릴레이션, E는 두 릴레이션 간의 종속 의존성을 나타냄.

이 두 정의를 통해, 기존 DB에서 릴레이션 데이터 모델(E-R) 내에 각각의 릴레이션들간의 의존성을 통해 연관성을 찾을 수 있다. 즉 관련된 정보는 근접화(locality)해 놓는 것이다. 이는 빅 데이터 테이블화에서 관련된 정보를 묶어 저장하자는 것이다.

이 개념은 원래 데이터 모델을 내포 릴레이션 데이터 모델(Nested Relational model)화에 있었다. 우리는 이를 관련된 릴레이션을 묶는 방법으로 변경하였다.

본 논문에서는 데이터베이스진흥원 2013년 DA설계공모전 요구사항의 일부를 사용하여,

“많은 대학생과 직장인들이 취업이나 경력관리, 직무수행능력 향상, 자격증 취득 등을 목적으로 수업을 하는 IT전문교육센터의 운영시스템 개선이다. 전국에 위치한 교육센터들이 개별적으로 운영되어 전체적으로 유사한 엔터티와 속성들의 중복이 심각하며, 데이터 분산과 이중관리에 따른 불편함이 존재했다. 이로 인한 데이터 정합성 및 공유성의 저하뿐만 아니라 교육센터를 운영하는

데 있어 시스템 사용상에서 서로 상이한 데이터로 인한 혼란과 불편 때문에 교육센터 운영관리에 어려움을 겪고 있다.....”등의 요구사항을 통해 데이터베이스를 설계 했다.

이러한 문제점 해결을 위한 관계형 데이터베이스 설계를 진행하고, 다음의 그림 1은 요구사항에 맞는 데이터 아키텍처를 구축하고 데이터 품질 향상을 위해 진행한 설계이다.

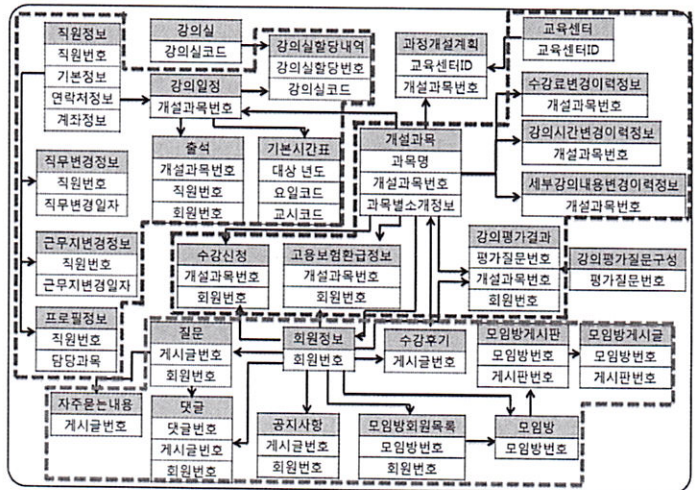


그림 1 교육관리 시스템 구축을 위한 데이터 모델링

그림 1은 간결화한 교육관리 시스템 구축을 위한 데이터 모델링이다. 총 30개의 테이블과 36개의 릴레이션으로 구성되어 있다.

본 논문에서는 그림 1과 같은 요구사항을 바탕으로 구축한 ‘교육관리 시스템 구축을 위한 데이터 모델’을 활용하여 관계형 데이터베이스의 빅데이터화를 제안을 위한 예제로 사용한다. 이런 관계형 데이터베이스는 위와 같이 복잡한 형태의 구조를 가지고 있다. 이러한 구조에서 빅데이터를 처리하기에는 무리가 있다. 또한 이런 관계형 데이터베이스는 데이터의 지속성, 동시성 제어, 통합 메커니즘 제공 등 많은 장점이 있지만 빅데이터 저장 및 처리에 있어 클러스터에서 실행되는 엄청난 양의 데이터를 지원하기에 효율적으로 동작 및 설계되어 있지 않으며, 각 테이블간의 릴레이션으로 연결 및 세분화되어있어 테이블간 연산에서 많은 시간 소비로 인해 비용 증감으로 문제점이 있다.

현 시점에서 빅데이터의 저장 및 처리를 위해 관계형 데이터베이스를 완전히 배제되어 있고, 새로운 저장구조와 인덱싱으로 구축하여 효율성을 높이려 노력 중이다. 이 논문에서는 기존에 구축된 관계형 데이터베이스를 재사용 관점에서 빅데이터 저장 및 처리를 효율적으로 할 수 있도록 빅데이터화 하는 것이 목표이다.

기존 NoSQL 데이터베이스 중 구글의 빅테이블은 SQL 을 사용하지 않으며 모든 데이터를 테이블에 저장하여 사용하는 것이다. 하지만 우리는 Functional Access Path[6]를 이용하여 관계형 데이터베이스에서 비슷한 유형의 테이블들을 묶어 빅테이블화 하였다. 이는 빅데이터의 분석을 수행할 때 짧은 시간에 효율적인 분석 효과를 얻을 수 있을 것이다

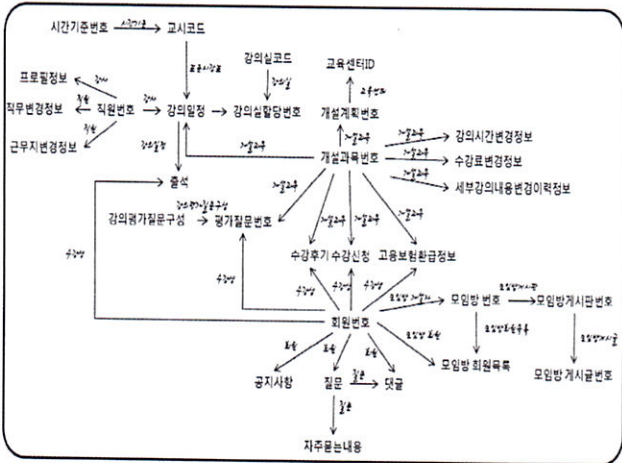


그림 2 릴레이션 데이터 상에서의 functional access paths 위의 functional access 기반으로 릴레이션간의 연관 관계를 연결할 수 있으며, Flat complex object [6] 메카니즘 기반으로, 여러 개의 complex 덩어리화가 가능하다. 그림 2는 릴레이션들 상에서의 functional access를 보여 준다. 이를 통해서, 교육관리 시스템 구축을 위한 데이터 모델링의 전체 설계 모델은 Functional Access Path를 이용하여 3가지 큰 모델덩어리로 나뉜다. 3가지 complex 모델링 덩어리는 직원정보, 개설과목, 회원정보를 중심으로 구성되어 있다.

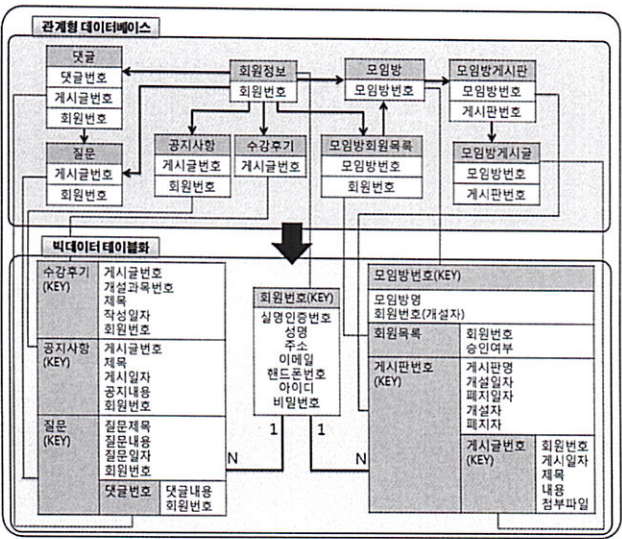


그림 3 회원정보 테이블의 빅데이터 테이블화

그림 3은 '교육관리 시스템 구축을 위한 데이터 모델링'에서 회원 정보 부분의 해당하는 테이블들을 Functional Access Path를 적용하여 나눈 테이블이다. 우리는 회원 정보 테이블의 릴레이션이라 할 수 있는 '회원 정보'테이블을 기준으로 '수강후기', '공지사항', '질문', '댓글', '모임방', '모임방 게시판', '모임방 게시글', '모임방회원목록'으로 구성 했다.

회원정보 모델에서 회원정보 테이블은 주변 테이블들의 키를 가지고 있는 상위테이블이다. 그러므로 회원정보 테이블을 그대로 이용하고 나머지 테이블들을 빅테이블화 하여 설계하였다.

4. 결 론

기존의 관계형 데이터베이스는 빅데이터의 이슈를 다루기 어려우며, 많은 양의 데이터를 연산하는 빅데이터 처리에 있어 테이블 간의 Join 연산으로 시간이 오래 걸리는 단점을 가지고 있다. 기존의 관계형 데이터베이스 모델을 재사용 관점에서 빅데이터 테이블화를 제안한다. 이를 해결하기 위하여 관계형 데이터베이스 모델에서 functional access path 메카니즘 기반으로 빅데이터 덩어리화하여, 빅 데이터 테이블화(big data table) 방안을 제안한다. 관련 된 정보는 근접화(locality)하여야 하는 메카니즘으로, 관계형 데이터 모델링내의 테이블간의 함수적 접근 경로(functional access path) 연결까지 하나의 빅테이블화하는 방식이다. 아직은 빅 데이터 덩어리를 빅 데이터 테이블화에 대한 규칙을 집중하고 있다.

5. 참고 문헌

[1] 이미영, 최완, "빅데이터 처리 및 저장관리 기술동향 및 전망", 한국정보통신학회지 13권 1호 pp.33-39, 2013
 [2] 이충희, 허정, 오효정, 김현진, 류범모, 김현기, "소셜 빅데이터이슈 탐지 및 예측분석 기술 동향", 전자통신동향분석 제 28권 제1호 pp.62-71, 2013.02
 [3] 이진형, "데이터 빅뱅, 빅 데이터(BIG DATA)의 동향 빅 데이터와 신 가치창출 재인용", 방송통신전파저널 통권47호, 2012
 [4] 한소월, 이민수, "소셜 정보 추천 기법을 위한 빅 데이터 모델", 정보과학회논문지 : 데이터베이스 제 39권 제 6호, 2012
 [5] 프라노드 사달게이, 마틴 파울러, "NoSQL 빅데이터 세상으로 떠나는 간결한 안내서", 인사이드, 2013
 [6] Carlson, C. R., and Arora, A.K., The updatability of Retional views based on Functional Dependency, Proc. IEEE COMPSAC pp.415-420, Chicago, Nov. 1989