

TPC-H 데이터모델을 이용한 SQL 교육 도구

박인혜, 김지은, 전민아, 심재희, 강현정, 박우창
덕성여자대학교 컴퓨터학과

e-mail: {in073548, jieun1103, minah88, shimjaehee, hyunjeung,
ucpark}@duksung.ac.kr

SQL Learning Tool Using TPC-H model

Inhye Pack, Jieun Kim, Minah Jeon, Jaehye Shim, Hyunjeong Kang,
Uchang Park

*School of Information and Media Duksung Women's University

요 약

본 연구에서는 SQL를 배우고자 하는 개발자들에게 SQL 문법을 학습할 수 있는 교육용 Tool을 개발한다. 개발자가 예제와 설명을 통하여 SQL 문법을 배우고 ER-Diagram을 보면서 논리적인 DB의 개념을 이용하여 쉽게 학습할 수 있다. 예제는 초급과 중급으로 나누어져 있어 사용자의 수준에 맞는 학습이 선택가능하다. TPC-H 데이터는 DSS 환경에서 사용되는 표준 데이터 모델로 Database Generator를 통해 생성하며 본 연구에서 사용자가 데이터량의 조정이 가능하도록 구성하였다.

1. 서론

SQL 언어는 데이터베이스 프로그래머에게 중요한 언어이다. SQL 학습 시 다양한 데이터 모델로 SQL의 기능들을 학습하게 되는데 표준화된 데이터 모델을 사용하면 실무에서 적응력을 높일 수 있다. 본 연구는 이러한 점을 감안하여 TPC-H[1] 모델을 SQL 교육에 활용하는 것을 목표로 했다. 또 DBMS를 학습하는 프로그래머들이 기존 프로그램과 달리 시각화된 프로그램을 통해 보다 쉬운 학습이 가능하도록 한다.

기존의 TOAD, SQLTools[2], Orange[3]같은 데이터베이스 SQL 인터페이스들은 데이터베이스를 생성, 관리하고 SQL 작성/수행에 특화되어 있다. 본 연구에서는 DB 프로그램을 처음 접하는 사용자들이 책의 설명으로 이해하기 힘든 쿼리들을 표준화된 데이터 모델과 실행 가능한 설명과 함께 제공된 예제를 통해 학습 효율을 높일 수 있다.

2절에서는 TPC-H의 데이터 생성을, 3절에서는 구현된 시스템에 대한 설명을 한다. 마지막으로 4절에서는 결론과 향후 더 연구해야할 점들을 논한다.

2. TPC-H의 데이터 생성

TPC(Transaction Processing Performance Council)는 비영리 기관으로서 트랜잭션 프로세싱 및 TPC-C, TPC-H, TPC-W 벤치마크와 같은 데이터베이스 성능 벤치마크를 정의하고, 이러한 벤치마크 결과를 기초로 한 객관적인 성능 자료를 배포하기 위해 설립되었다. TPC-H는 TPC-D 모델을 개선한 것으로 DSS 환경에서 대량의 데이터를 Ad-hoc 쿼리를 이용하여 시험하는 벤치마크 모델이다. TPC-H는 Database Generator와 Ad-hoc Query Generator로 구성되어있는데, Database Generator를 통해

표준화된 데이터를 생성하여 샘플 데이터로 이용하였다.

TPC-H 데이터 생성을 위하여 사용된 스크립트의 일부는 다음과 같다. 지정된 데이터가 들어있는 리스트와 랜덤 함수 등을 이용하여 만들어진 스크립트에 따라 데이터를 생성한다. 아래 그림 1은 customer 데이터를 생성하는 예를 보여준다.

```
long
mk_cust(DSS_HUGE n_cust, customer_t * c)
{
    DSS_HUGE i;
    static int binit = 0;
    static char szFormat[100];

    if (!binit)
    {
        sprintf(szFormat, C_NAME_FMT, 9, HUGE_FORMAT + 1);
        binit = 1;
    }
    c->custkey = n_cust;
    sprintf(c->name, szFormat, C_NAME_TAG, n_cust);
    V_STR(C_ADDR_LEN, C_ADDR_SD, c->address);
    c->alen = (int)strlen(c->address);
    RANDOM(i, 0, (nations.count - 1), C_NTRG_SD);
    c->nation_code = i;
    gen_phone(i, c->phone, (long) C_PHNE_SD);
    RANDOM(c->acctbal, C_ABAL_MIN, C_ABAL_MAX, C_ABAL_SD);
    pick_str(&c_mseg_set, C_MSEG_SD, c->mktsegment);
    TEXT(C_CMNT_LEN, C_CMNT_SD, c->comment);
    c->clen = (int)strlen(c->comment);

    return (0);
}
```

그림 1 : customer 데이터 생성 예

TPC-H 데이터모델에 의하여 생성되는 데이터는 대용량이지만 본 연구에서는 다양한 크기의 데이터를 생성할 수 있도록 scale factor 값을 두고 scale factor에 따라 소규모 및 대용량 데이터를 사용자의 선택에 따라 생성하도록 하였다. Scale factor는 benchmark에 사용된 데이터베이스의 크기를 TPC-H 산출 공식에 따라 표시한 형태이다. 예를 들어 사용된 Database의 크기가 1GB 이면 SF=1, 3TB이면 SF=3000이다. 아래 그림 2는 사용자가 scale factor를 지정하는 과정을 보여준다.



그림 2 : TPC-H에서 scale factor 지정 화면

3. 도구의 기능

개발된 도구의 화면은 아래의 그림 3과 같다. 화면은 메뉴, SQL 입력창, SQL 예제 리스트창, SQL 예제 설명창, SQL 결과창, ER 다이어그램 창,으로 구성된다. 도구는 SQL 인터페이스인 DBedit을 기반으로 만들어져 있다[4]. DBedit은 SQL 문을 입력하고 실행하는 기능만 있는 단순한 오픈소스코드이다.

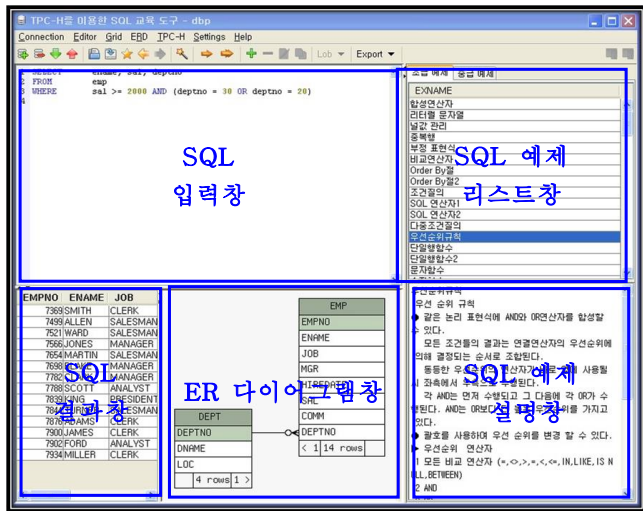


그림 3 : SQL 교육도구 실행 창

SQL 입력창은 SQL 예제 리스트창에서 선택된 SQL문이 표시되는 창이다. 사용자는 예제를 수정하거나 직접 예제를 입력할 수 있다.

SQL 결과창은 입력한 SQL문의 실행 결과를 보여준다.

SQL 예제 리스트창과 SQL 예제 설명창의 예제는 초급과 중급으로 나누어 제공하였다[5][6][7]. 초급은 기본적인 SELECT 문과 SQL 내장함수를 이용하여 SQL 쿼리 방법에 대해 실습할 수 있게 하였고, DML의 사용법에 대해 초점을 두고 다루었다. 중급에서는 심화된 SQL문과 JOIN 함수에 대한 기초설명부터 여러 JOIN함수를 다룰 수 있는 설명을 제공한다. 따라서 사용자들의 수준에 맞는 단계를 선택하여 SQL를 공부할 수 있다.

ER 다이어그램창은 엔티티간의 관계를 시각적으로 보여주는 창이다. DA#, ER-Win등의 Data Modelling Tool에서는 사용자가 직접 연결하는 과정과 만드는 과정이 있다. 하지만 복잡한 과정은 제외하고 기존에 있는 엔티티 간의 관계를 표현할 수 있어 기존 Data Modelling Tool보다 프로그램의 복잡도를 낮추었다. 또한 데이터무결성에 위배되는 DML 사용을 방지할 수 있다. 예를 들어 UPDATE시 데이터 Cascade 등의 동작들이 있다.

ER 다이어그램창을 이용하여 자신이 조작하고 있는 데이터의 구조 및 관계를 정확하게 이해할 수 있다.

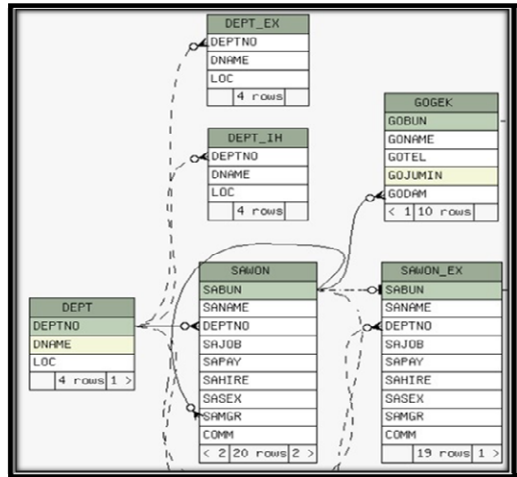


그림 4 : ER 다이어그램창의 결과

ER 다이어그램 생성은 오픈소스코드인 SchemaSpy[8]를 기반으로 Graphviz Tool[9] 도구를 사용하여 파일로 본 프로그램에 삽입했다. 사용자들은 새로운 Diagram 프로그램을 동작하지 않더라도 하나의 프로그램 인터페이스에서 시각적으로 ER 다이어그램까지 볼 수 있다.

4. 결론

본 프로그램은 SQL를 배우고자 하는 사용자들에게 샘플 데이터 모델, 예제, 실행인터페이스를 종합적인 기능을 제공하는 교육용 Tool이다. 예제와 설명을 통하여 SQL를 상세하고 쉽게 따라할 수 있으며 특히 ER 다이어그램창의 시각적 효과를 통해 논리적인 DB의 개념을 이해하면서 실습이 가능하다. DSS 응용과 비슷한 TPC-H 데이터 모델과 scale factor로 조절된 데이터 량을 이용하여 Database에 대한 실습을 가능하게 해준다. 단계적 선택학습을 통한 맞춤 교육을 제공하는 것이 본 프로그램의 장점이다.

참고문헌

- [1] <http://www.tpc.org/>, Transaction Processing Performance Council
- [2] www.wikipedia.org, Comparison of database tools.
- [3] <http://www.warevalley.com/koMain.asp>, Orange.
- [4] <http://sourceforge.net/projects/dbedit/>, DBEdit.
- [5] 한용수, The Guide for SQL Professional SQL 전문가 가이드 2010 Edition. KDB 한국 데이터베이스 진흥원, 2010.
- [6] http://www.koug.net/xe/oracle_lecture, 오라클 강좌.
- [7] <http://dblab.changwon.ac.kr/>, 창원대학교 데이터베이스 연구실.
- [8] <http://schemaspy.sourceforge.net/>, Schema Spy.
- [9] <http://www.graphviz.org/>, Graphviz Tool.