

데이터 비만도를 개선한 데이터 모델링에 관한 연구

이혜경*, 김희완**

용인송담대학교 컴퓨터게임과^{*}, 삼육대학교 컴퓨터학부^{**}

A Study on the Data Modeling decreasing the Data Obesity

Rhee Hye Kyung^{*}, Hee Wan Kim^{**}

Dept. of Computer Game, Yongin Songdam College^{*}

Division of Computer Engineering, Shamyook University^{**}

요 약 현업 데이터베이스 응답 속도가 점차 느려지는 원인을 찾기 위하여 데이터 비만도가 거기에 어떤 영향을 초래할 수 있는지에 대해 연구했다. 게임정보시스템을 대상으로 게임 데이터 인프라가 어느 정도 잘 갖추어져 있는지 분석함으로써 연구를 실행했다. 데이터 인프라 수준을 측정하는 다양한 방법이 있지만 본 논문에서는 정보시스템 데이터 설계 결과물을 놓고 실물 평가를 실시하였다. 데이터 모델링의 여러 단계 중 실존하는 게임정보시스템의 논리데이터모델 설계 산출물을 기준으로 새로 모델링 한 개체관계 모형도 비교 분석하였다. 데이터 인프라의 주요 지표가 되는 데이터 비만도가 게임정보시스템 전체 평균 60%에 달하여 기준치인 15%를 무려 45% 상회하는 것으로 나타났다. 본 논문에서 데이터모델링 절차를 수행한 후의 데이터 중복률은 41%로 나타나서 기존 모델의 64%에 비하여 23%의 개선효과를 가져왔다.

주제어 : 데이터 비만도, 게임정보시스템, 개체관계모델, 속성식별

Abstract In this paper, we studied how the data obesity can affect in which the response speed of database gradually slows down. Our research is performed by analyzing how the game data infrastructure is well-formed. Although there are a variety of ways to evaluate to measure the level of infrastructure, we performed with real information system. We analyzed data obesity by comparing the entity-relationship models between the products of real game information system and newly modeled databases. We could find data obesity is over 60% among overall average of game information system. It shows that 45% higher than standard obesity which is 15%. In this paper, data redundancy rate after performing the procedure of the data modeling was 41% resulting in an improvement of 23% compared to 64% of an existing model.

Key Words : data obesity, game information system, entity-relationship model, attributes identifying

1. 서론

1.1 배경

이 시대를 정보화 시대라 일컫지만 정작 데이터의 시

대라는 의미를 지니기 보다는 여전히 전산화 프로그램 시대에서 아직도 벗어나지 못하고 있는 것이 현실이다. 데이터 마인드 부족현상은 데이터 마인드에 충실하지 못하여 데이터의 응답시간이 예전보다 점점 느려지게 만드

Received 11 September 2013, Revised 1 October 2013
Accepted 20 November 2013
Corresponding Author: Hee Wan Kim(Shamyook University)
Email: hwkim@syu.ac.kr

ISSN: 1738-1916

© The Society of Digital Policy & Management. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

는 주요 원인이 되고 있다. 진정한 의미의 데이터 마인드란 정보화에 있어서 어떤 것이 데이터화 될 것이고 어떤 것이 프로그램 코드화 될 것인지를 적확히 구분하는 데 있다. 이러한 현상은 데이터의 불일치[1, 2, 3], 또는 단순히 편의상 데이터가 중복되었다든지, 데이터 홍수 현상[4], 또는 데이터 과적현상[5]때문이라 할 수 있다. 데이터 홍수현상은 최근 주목을 받아오고 있으나 데이터 불일치 문제는 아주 긴박하고 기본적인 문제로 간주되고 있다.

인체 비만처럼 데이터 비만이 데이터 응답속도에 어떤 영향을 미치는 지에 대한 연구로 다른 과거 연구로는 이전에 분산되어 관리되어 오던 회사 내의 데이터들을 단일 통합된 데이터 시스템인 엔터프라이즈 데이터 모델(EDM)로 교체함으로써 데이터의 품질을 높일 수 있다고 하였다[6]. 데이터는 중요한 기업의 자산이므로, 품질은 매우 중요하며, 중복 데이터 제거가 데이터 품질 유지에 대한 기본 요인 중 하나이며 EDM 데이터 품질에 대한 필수적이라 하였다.

데이터베이스 성능향상의 이유로 데이터의 중복을 허용하는 방식의 설계가 이루어지고 있는 점을 연구한 논문에서는 데이터의 중복은 국내기업들의 평균 비율로 65%나 될 정도로 위험수위에 육박하고 있으며 데이터 중복을 허용하는 방식의 설계는 데이터의 무결성을 낮추고 그 결과 데이터 품질을 저하시킨다는 점을 지적하였다. 중복을 포함한 데이터모델과 정규화를 통해 중복을 제거한 데이터 모델의 성능 비교를 위한 실험 결과 응답 시간은 중복 관계를 제거한 모델이 중복을 허용한 모델보다 10% 안팎으로 향상된 것으로 나옴으로써, 중복을 제거한 모델이 성능 및 데이터 품질에서 우수함을 증명하였다[7].

데이터 품질 관리를 위한 개발 방법론에 대한 연구[8]에서는 데이터베이스 설계 과정에서 세심하게 데이터를 수집하며 데이터베이스를 설계하는 과정을 소개하고 있지만 데이터 중복에 대한 중요성은 잠깐 동안 언급만 했을 뿐 어떻게 설계를 해야 중복을 피하는 품질 좋은 데이터베이스를 설계하는지는 연구하지 않았다.

위 논문들에서 본 바와 같이 데이터 중복에 대한 중요성은 인식은 하였지만 정확한 학문적 연구를 통해 데이터 비만도에 대한 산출은 하지 못하였다. 데이터 비만도를 정확히 산출해 내기 위해서는 현업 데이터베이스 확

보가 필수적이거나, 이러한 실제 데이터베이스를 들여다 보기는 매우 힘든 상황이기 때문에 학문적 연구를 위한 데이터 비만도 측정은 시도하지 못하였다.

1.2 동기

데이터의 비만도에 연구를 하게 된 동기로는 과연 산업에서 데이터 모델링을 할 때 이러한 균형감있는 설계를 하고 있는지에 대한 의구심을 갖게 되었기 때문이다. 왜냐하면 데이터 모델링은 완전히 설계자의 역량에 따라 데이터베이스의 품질이 좌우되는 것이 현실이다. 설계를 담당하고 있는 실무자들이 과연 어떤 지침서를 보고 설계에 대한 학습을 하는지 출판되어 있는 관련 교재들을 조사해본 결과, 데이터 중복으로 인한 비만도가 데이터베이스의 품질에 막대한 영향을 미치고 있는데도 불구하고 데이터 비만도에 대한 언급이 부족하였다. 단순히 중복이 좋지 않기 때문에 중복을 피하라는 지침만 있을 뿐 어떻게 설계해야 중복이 줄어들 수 있는지는 보여주지 않았다[9, 10].

데이터 모델링이란 기업의 '전사적 데이터 지도'[11, 12]를 만드는 일련의 방법론을 말한다. 많은 양의 데이터를 효과적으로 운용하기 위해 데이터 중복을 최대한 제거하고 속도와 질을 모두 만족시키는 데이터 지도만이 기업의 데이터 무결성을 보장하는 동시에 정보시스템의 응답 속도 및 서비스 품질을 향상시킬 수 있다. 불가피한 중복으로 인한 평균 데이터 중복률은 15%가 적정하다고 본다.

데이터 모델에서 데이터의 유형은 개체(entity)와 개체 간의 연결고리 관계를 형성해주는 행위(relationship)로 반드시 양분되어야 한다. 그러나 현실적으로 컴퓨터 게임 정보시스템의 논리데이터모델에서는 어떤 것이 관계인지를 분간할 수 없게끔 설계되어 있는 실정이다. 이렇게 관계를 설계도면 상에서 명시적으로 구분해낼 수 없는 경우에는 설계의 균형성이 파괴될 수 있다. 균형성이란 개체와 관계간의 균형 감각을 가리키는 것으로서 균형성의 기준은 관계의 총 수가 개체의 총 수의 차이가 10% 범위 내를 넘지 않아야 함을 원칙으로 하고 있다. 이러한 균형 파괴 설계가 나타나는 원인은 주로 설계를 관계(행위) 중심으로 하지 않고 개체 중심으로 하는 잘못된 관행에 기인한다.

데이터베이스 설계 구조의 균형이 파괴되면 데이터

질의에 대해 응답하는 경우 응답속도가 항상 일정 시간(예: 3초)내에 불가능해지며 설계상 편향성(skewness)으로 인해 어떤 질의에 대해서는 3초, 또 어떤 질의에 대해서는 30초, 경우에 따라서는 300초를 나타내는 등으로 응답속도 편차가 커지게 된다. 현대의 기술 수준을 기준으로 할 때 질의에 대한 응답속도는 어떤 질의이든 간에 막론하고 항상 3초 이내를 만족해야 컴퓨터 게임 정보시스템이 질적으로 문제 없다고 볼 수 있다. 현행 컴퓨터 게임 정보시스템의 경우에는 대부분의 질의에 대해 3초를 벗어나는 경우가 다반사로 발생할 수 밖에 없을 것으로 예상된다. 실제 현장에서 데이터베이스 비만도가 높은 이유는 데이터베이스 설계에서 방사형 설계가 그 원인으로 추측이 된다. 본 논문에서는 이를 정방향 구조로 다시 설계하여 데이터 중복률을 비교해 봄으로써 보다 정량적인 차이점을 도출하고자 하였다.

1.3 목적

우선 현장 데이터 비만도를 측정하기 위해 실제 게임 데이터베이스를 사례로, 기존의 데이터베이스에 대한 평균 데이터 중복률을 조사했다. 또한 데이터베이스의 설계에 따른 비만도가 어떻게 달라지는지 살펴보기 위해 새로 균형감있는 설계를 통해 데이터베이스를 설계하였으며, 기존 데이터베이스와 개선된 데이터베이스의 평균 데이터 비만도를 비교 분석해 보았다.

사례로 든 데이터베이스는 실제 모바일 게임 데이터베이스로서 규모가 그리 크지는 않지만, 규모가 큰 데이터베이스와 비교해 볼 때 본 논문에서 주장하고자 하는 데이터 비만도 측정 면에서는 전혀 차이가 없다고 본다. 왜냐하면 사례로 든 데이터베이스는 방사형 구조로서 규모가 큰 대부분의 데이터베이스들도 마찬가지로 구조일 것으로 예상하기 때문이다. 새로 설계한 데이터베이스는 정방향 구조를 가진 데이터베이스로서 데이터베이스의 규모가 확장하여도 동일한 구조를 가지게 될 것이다.

본 논문의 연구를 통해 ‘데이터 체지방율’ 혹은 ‘데이터 비만도’에 해당하는 데이터 중복률[13]을 컴퓨터 게임 정보시스템 사상 최초로 산정해 봤다는 말은 컴퓨터 게임 정보시스템이 과거와 현재에 얼마나 정보시스템 개선을 위한 목표 의식이 부재하고 정교함이 갖추어지지 않은 채 계획성 없이 기획 운영되어 왔는지 알 수 있는 것이다.

데이터 지도에 지도의 일부분이 조각 혹은 파편 형태로 고립된 상태가 존재해서는 지도로서의 자격이 상실된다. 그러나, 현행 컴퓨터 게임 정보시스템의 경우, 이런 고립된 파편 형태가 빈번히 발생하여 설계의 품질이 격하되어 있는 상태이다. 이러한 편향 설계의 폐단은 데이터 비만도의 과도함으로 드러나게 된다[14]. 데이터가 전혀 쓸데없이 불필요하게 중복되어 있는 정도를 나타내는 데이터 비만도는 데이터 중복율을 계산해 봄으로써 알 수 있다. 컴퓨터 게임 정보시스템의 정확한 평균 데이터 비만도 산정 결과는 아래 제4장에서 밝혔다. 데이터 비만도가 만약 과다하게 나타나는 경우에는 방사형 및 고립형 설계를 지양하고 사각정방향으로 데이터 경로가 어느 경우에도 선명하게 가시화되게끔 현행 시스템을 개선하는 것이 요구된다.

2. 현행 시스템의 데이터의 중복 과다 체증 현상

2.1 평균 비만도 최적 기준치

본 논문에서 예를 든 컴퓨터 게임 정보시스템의 평균 데이터 중복율(즉, 데이터 비만도)은 평균 64%로 집계됐다. 이는 과도 비만 상태로서 2000년 이후 IT 기술 수준으로는 평균 15%를 상회하면 잘못 설계된 것이라는 점에서 상태의 심각성을 상세 분석하는 것이 중요하다. 결론적으로는 무려 50%를 초과하는 비만도를 지니고 있는 현행 컴퓨터 게임 정보시스템에 대해서 개선 전략을 준비해야 할 것으로 판단된다.

현대 데이터베이스 기술이 최적으로 적용된 경우에는 데이터(이후 본 논문에서 데이터라 함은 실질 데이터를 지칭하는 속성 차원을 의미함)가 설정 중복되는 경우가 발생하더라도 데이터 중복율이 최대 15%를 상회한다면 정보시스템의 품질에 문제가 존재하는 것이다. 데이터 중복율(혹은 데이터 비만도)은 인체로 말하면 체지방율에 해당하는 것으로서 예를 들어 데이터 중복율이 50%라고 한다면 이는 인체의 체지방율이 바로 50%에 달한다는 말과 완전히 동일한 의미이다.

2.2 현행 컴퓨터 게임 정보시스템 모형도

현행 컴퓨터 게임 정보시스템(가칭: Combat DB)은 다

음과 같은 개체와 관계에 대한 속성으로 설계되어 있다.

USER(id, account_type, user_name, passwd, invite_accept, gift_accept, high_score, record_time, best_score, level, exp, join_date, attendance_continuance, latest_connect_time, latest_play_time, phone_number)

HAVE(id, item_code, item_count)

PERIOD(id, item_code, start_date, end_date)

//아이템 중 특수한 한 아이템(초코 하루 무제한)에 대한 사용 유효기간(24시간)

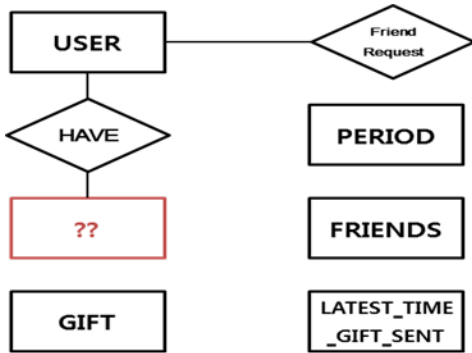
FRIENDS(id, friend_id)

LATEST_TIME_GIFT_SENT(sender_id, receiver_id, time_sent)

FRIENDREQUEST(id, friend_id)

GIFT(gift_id, sender_id, receiver_id, time_sent, item_code, item_count, message)

위의 릴레이션들을 바탕으로 작성한 ER-D는 [Fig. 1]과 같다.



[Fig. 1] ER-D of Combat DB

[Fig. 1]에서 보는 바와 같이 현행 Combat DB 설계의 문제점은 대략 두 가지로 나타난다. 우선 Combat Data Modeling 방식은 제1정규형(1NF)에 해당된다. 기본 키에 대한 개념이 전혀 설정되어 있지 않고, 종류별로 필요한 데이터를 수집해 놓은 상태에 불과하다. 두 번째 문제점은 속성(Attribute)으로 잡혀야 하는 것들이 현재 테이블로 설정되어 있는 점이다. 예를 들어 GIFT테이블은 속

성으로 잡혀야 하나 테이블로 잡혀있는 것을 볼 수 있다.

3. Combat DB(New_Combat DB) 적법 Data Modeling

이미 구현되어 실용화하고 있는 Combat DB의 문제점을 개선하여 데이터 중복률을 줄이기 위해서는 새로운 Data Modeling이 필요하다. 우선 데이터 모델링을 하기 위해서는 제일 먼저 업무 분석을 위한 요구형성 과정을 실시해야 한다. 새로운 Combat DB(가칭: New_Combat DB)의 Data Modeling의 요구형성을 통한 업무분석은 다음과 같다.

[New_Combat DB 데이터 모델링 업무설계도]

- ① User는 자신이 아닌 다른 User에게 친구 요청을 할 수 있다.
 - 친구 요청을 보낸 user는 자신이 요청을 보낸 user와 자동적으로 친구 관계를 맺게 된다.
 - 친구 요청을 받은 user는 친구 요청을 받을지 받지 않을지 친구 수락여부를 결정한다.
 - 친구 수락된 user는 서로 친구 관계를 맺는다.
- ② User는 친구 관계가 맺어진(상대방 수락여부에 관계없이) 자신의 친구에게 선물(item 종류 중에 choco item)을 보낼 수 있다.
 - 한 친구에게 한번 선물(item 종류 중에 choco item)을 보내면 일정 시간을 기다리고 다시 보낼 수 있다.
- ③ User는 선물(item 종류 중에 choco item)을 받을 수 있다.
 - 친구 관계가 아닌(자신에게 친구요청을 한) user로부터 선물(item 종류 중에 choco item)을 받을 수 있다.
 - 서로 친구 관계가 맺어진 자신의 친구로부터 선물(item 종류 중에 choco item)을 받을 수 있다.
 - 선물을 받은 user는 선물 수락여부를 결정한다.
 - 선물을 받은 user는 수락한 선물을 사용할 수 있다.
- ④ 선물을 받은 user는 수락하지 않은 선물을 자신의

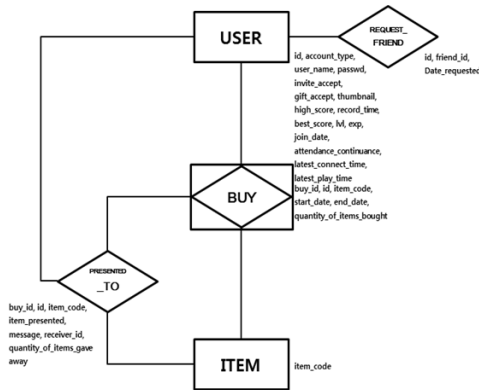
item 보관함(현재 자신의 모든 item 보유현황을 보여줌)에 저장한다.

- User 자신이 직접 구입한 'choco 하루 무제한' item 을 사용 시작 후 24시간 동안만 사용 가능함

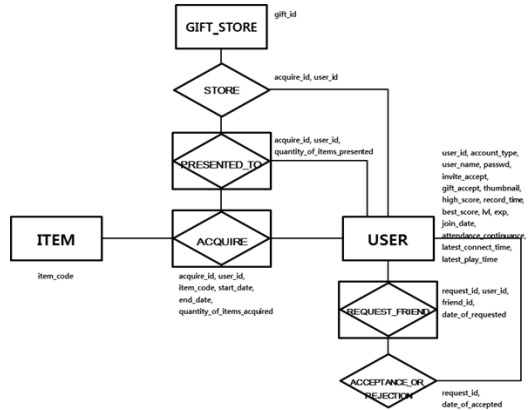
위와 같이 요구형성과정을 통해 업무설계도가 작성되면 다음 단계로 데이터 자체 즉, 행위를 강조한 ERD(Entity-Relationship Diagram)를 설계한다. ERD는 현행 개체 중심 설계(8개의 테이블(Entity))를 DB 설계 기본에 입각한 행위(relationship) 중심으로 설계할 경우 단 3개의 테이블(Entity)로 설계가 가능하게 된다. 즉, 행위를 도출하기 위한 노력을 통해 행위 위주의 설계도를 작성해야만 한다.

행위(relationship) 중심으로 설계할 경우 Entity는 USER, ITEM, GIFT_STORE 세 개로 구성할 수 있으며, Relationship은 REQUEST_FRIEND, ACCEPTANCE_OR_REJECTION, ACQUIRE, PRESENTED_TO, STORE 로 설계할 수 있다. 이와 같이 행위를 중심으로 설계하게 되면 제1정규형을 분해하여 각 필드 간의 상호 관계를 좀 더 세부적으로 규정함으로써 제3정규형(3NF)을 구성할 수 있게 된다.

위의 업무분석을 통한 1차 개선된 Combat Data Modeling의 ER-D는 [Fig. 2]와 같다. [Fig. 3]은 1차 개선된 ERD를 개선한 최종적으로 구현한 ER-D이다.



[Fig. 2] New_Combat DB Data Modeling



[Fig. 3] Final Improved New_Combat DB Data Modeling

4. Data Modeling의 데이터 중복률 개선 방법

Data Modeling을 전혀 고려하지 않고 설계된 Combat DB와 Data Modeling 설계 절차를 거쳐 설계된 New_Combat DB의 데이터 중복률을 비교하면 다음과 같다.

Combat DB의 데이터 중복률(data redundancy ratio)

$$\frac{\text{중복된 속성들의 총 빈도 수}}{\text{DB Schema에 포함된 총 속성 종류의 갯수}} = \frac{18}{28} = 64\%$$

New_Combat DB의 데이터 중복률

$$\frac{\text{중복된 속성들의 총 빈도 수}}{\text{DB Schema에 포함된 총 속성 종류의 갯수}} = \frac{12}{29} = 41\%$$

2.2절의 현행 컴퓨터게임 정보시스템 모형에서 7개의 엔터티에서 총 속성의 수는 28개이며, 그 중에서 18개의 속성들이 외래키나 id로 중복되는 속성으로 중복률은 64%가 된다. 반면에, [Fig. 3]의 개선된 최종 Data Modeling에서의 총 속성의 수는 29개이며, 중복된 속성의 수는 12개로서 중복률은 41%가 되어 두 DB간의 중복률이 23%차이가 남을 알 수 있었다. 즉, 새로 설계된 DB의 중복률이 현저히 감소되었다. Database normalization, 즉

1NF에서 3NF로 변환함으로써 데이터 중복을 방지하고 저장 가용공간을 최적화할 수 있었다. 또한 적절한 외래 키를 사용함으로써 데이터 중복뿐만 아니라 destructive anomalies 의 가능성을 최소화 할 수 있었다.

4.1 최적 중복율은 15% 선

현대의 기술을 기준으로 할 때 IT 기술 자체의 한계선인 중복율 15%를 상회한다는 이야기 자체는 기술적으로는 주로 primary key 속성을 부주의하게 foreign key 속성으로 빈번히 여러 군데에서 방만하게 중복되게끔 허용했다는 뜻이다. 물론 이외에도 중복이 유발되는 경우는 존재한다. 이는 현대의 데이터 설계의 규범과 기초를 무시한 편법 설계로서 반드시 개선해야 할 대상이다. 인체 기준으로 해도 인체 체지방율이 30%를 넘는다는 것은 고도 비만 상태의 심각한 지경으로 누구나 상식적으로 평소 들어서 알 내용이다. 게임정보시스템의 데이터 중복 체지방율이 현재 평균 60%로 나왔다는 본 논문의 연구 결과는 게임정보시스템이 심각한 수준의 동맥경화 상태에 처해 있다는 단적인 증거라 할 수 있다.

중복율이 최적 중복율 15%를 기준으로 삼을 때 심지어 45% 만큼이나 과다하게 상회 초과하고 있다는 점이 시사하는 바는 게임정보시스템이 현재 전체 보유하고 있는 데이터 중 45%를 과감히 도려내더라도 게임정보시스템은 전혀 문제 없이 오히려 효율적으로 작동할 것이라는 점이다. 뿐만 아니라 불필요 부분 제거 작업을 수행함으로써 신체의 경우처럼 몸이 가벼워져 질의에 대한 응답속도 또한 제거한 만큼에 자동 비례하여 훨씬 더 신속하게 나오게 되므로 보다 질 높은 게임운영을 위한 중복 데이터 제거 작업 계획은 조기에 마련될수록 바람직하다고 본다.

4.2 데이터 인프라 개선 방향

현행 컴퓨터 게임 정보시스템에는 무엇이 데이터인지 그리고 무엇이 프로그램인지 구분이 애매 모호하게 설계되어 있는 현실이다. 데이터 인프라에는 철칙상 Raw Data가 아니면 나타나면 안되게 되어 있으나 그렇지 않은 가공 데이터(예: 누계성 데이터 혹은 집계성 데이터)들까지 데이터 모델에 포함되어 있어 데이터 인프라 형성 제작에 대한 기본 이해의 근간이 잘 안되어 있는 상태

이다. 차체에 데이터 인프라에서는 Raw Data만 존재하도록 개선하고 일체의 집계성 가공 데이터는 모두 제거하여 프로그램에서 자동 산출하게끔 구분해야 한다.

이런 근원적 문제점을 개선하지 않고는 집계성 데이터가 계속 데이터 모델상에 존재할 경우 데이터 일관성이 시스템에 의해 자동 유지 보장될 수 없고, 집계 데이터마다 수작업에 가까운 작업으로 Raw Data와의 일관성을 매일 혹은 경우에 따라서는 매시간 사람이 일일이 점검하는 수밖에 없는 문제점을 벗어날 길이 막연하다. 사람은 전문가라 하더라도 실수할 수 있으며 한번의 실수가 전장 데이터의 경우 치명적인 패배를 자초할 수 있는 바 일관성 유지 문제는 반드시 시스템이 자동으로 처리하도록 차체에 대책을 마련해야 할 것이다. 일관성 체크를 어떤 경우에도 시스템 자동으로 처리하지 않을 경우 그로 인한 시스템 응답 속도의 저하 현상은 심각해질 수 있다.

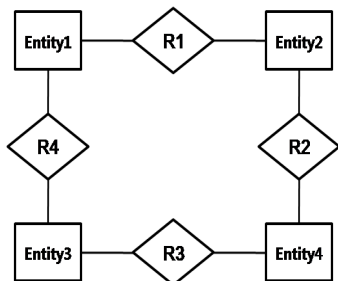
개체와 관계 간에는 데이터 연결 경로가 한번도 끊어짐이 없이 선명히 이어져야 상시 양방향 접근이 가능함. ‘개체-행위-개체-행위-개체-행위-개체...’ 식으로 개체간 데이터 경로가 행위라는 이음새로 끊김 없이 연결되어 어떤 데이터(개체 혹은 행위) 둘 간에도 방향성을 갖고 서로 찾아갈 수 있는 길(경로)이 존재해야 한다. 이런 경로가 존재하는 설계도면을 ‘데이터 지도(Data Map)’라고 부르며, 한쪽 방향으로 무한정 뻗어 나가지만 하고 결코 돌아올 수 없는 경로가 전혀 부재한 맵은 데이터맵이라고 부를 수 없는 것이다.

데이터 지도 제작의 첫걸음은 행위를 중심으로 행위를 야기하는 행위주체와 행위의 대상이 되는 행위객체를 명확히 규명하고 이를 인지 식별할 수 있게 도식화하는데 있다. 그러나 현행 컴퓨터 게임 정보시스템들은 대부분 ‘주체(개체)-행위-객체(개체)’의 설계 사상을 완전히 무시하고 천편일률적으로 ‘개체-행위(진정한 행위가 아닌 유사 행위)-유사행위-유사행위...-유사행위’라는 방식의 편법적 설계를 적용하고 있어 데이터 설계의 철학이 전반적으로 결여된 상태이다. 진정한 행위란 ‘개체(A)-행위(B)-개체(C)’형태의 행위의 주체(A 혹은 C)와 행위의 대상(C 혹은 A)이 명확히 드러나게 묘사된 B를 가리키나 현행 시스템 일체에서는 행위의 대상이 거의 불명확한 상태이다. 따라서 이러한 행위는 진정한 행위의 자격이 없으며 편법적 유사행위로 판단되며, 혹은 병

존할 수 없는 여러 행위의 결합으로 판단된다.

일부 컴퓨터 게임 정보시스템에서 극히 일부분에 사각경로 형태의 설계가 보이는 것 같으나, 이는 자세히 보면 ‘개체(A)-개체-개체-----개체(A)’ 형태의 정도를 벗어난 설계로서 개체와 개체가 무슨 근거 행위를 기반으로 상호 관계가 맺어지는지 납득할 수 없는 형태이다. 그러므로 현행 컴퓨터 게임 정보시스템 일체에서 논리 데이터모델을 ‘개체-행위-개체-행위-개체’의 철저한 행위 중심의 모델링 규범을 일단 준수하도록 개선해야만 할 것이다.

이런 과도 비만을 유발한 설계 방식을 지양하는 것에 더하여 상시 3초내 응답속도 보장을 위해서는 특히 방식형(skewed) 설계를 탈피하는 형태로 논리데이터모델을 개선하는 것이 필요하다. 이에 더하여 고립형 설계까지 탈피하면 설계 결과는 자동적으로 정방형 설계 혹은 사각경로형 설계의 형태를 지니게 된다. 이런 사각정방형 설계가 바로 응답시간 3초를 원천적으로 상시 보장하는 열쇠가 된다. 보편적으로 DB ERD 설계 시 [Fig. 4]의 예와 같이 path에 cycle이 존재하도록 설계함으로써 maximum path를 제한할 수 있다.



[Fig. 4] A Cycled ER-D

4.3 데이터 확장성

Data Modeling을 할 때는 반드시 데이터의 증가로 인한 데이터 확장성(Scalability)을 고려해야만 한다. 본 논문에서 제안한 New_Combat DB인 경우 행위 중심으로 설계함으로써 DB의 확장 가능 및 변화에 대한 대응이 강력해질 수 있었다. 반면 Combat DB의 설계 방식은 편향(skewed) 설계라고 할 수 있다. 즉, 균형이 깨진 설계, 다시 말해 뻗어 나가기만 하며 사각 경로(cycle)가 부재한 설계라 할 수 있다. 이러한 구조로 설계된 DB는 확장성

면에서 매우 불안정적이다.

5. 결론

작은 오차가 예상치 않은 고도의 비만을 유발할 수 있으며 고도 비만은 연이어 데이터 일관성 유지의 어려움을 유발할 수 있다. 더 나아가서는 데이터 상호운용성 저하로 이어지고, 또 더 나아가서는 다가오는 빅 데이터 시대에 대처하기 힘들어지는 부정적 연쇄 과급효과가 나타날 수 있음을 간과해서는 아니 될 것이다.

불필요한 중복을 불가피한 중복으로 오도 및 둔감시키는 잘못된 관행의 현행 시스템 개발 방식을 지양하고 바람직한 방향으로 진화해 나가기 위해서는 최적의 표준형 설계 지향해야만 한다. 특히 현행 시스템들의 데이터 인프라 상 상호운용성이 미흡한 이면에는 기술적으로 데이터 설계의 종착 목표가 정해져 있지 않기 때문인 점을 간과해서는 아니 될 것이다. 현행 시스템들은 전부 데이터 설계 원칙 상 가장 초보적 수준인 제1차 정규형 수준에도 훨씬 못 미치는 수준의 설계 결과물임을 인지해야 할 것이다. 적어도 제3차 정규형 설계가 자동으로 보장되도록 하는 데이터 인프라 설계 방법이 이미 기준에 나와 있고 다른 현장에서 이를 잘 활용 적용한 사례가 존재하는 상황에서 1차정규형에도 미달되는 수준의 설계를 계속 가져간다면 여간 심각한 문제가 아닐 수 없다. 데이터 중복율이 게임정보시스템의 전반적 품질 제고에 있어서도 중요하므로 데이터 인프라 설계 시 반드시 데이터 설계 원칙을 준수해야 할 것이다.

REFERENCES

[1] C. W. Fisher, B. R. Kingma, Criticality of data quality as exemplified in two disasters, Information Systems, 39, pp.109-116, 2010.
 [2] [DB] Elements of a good data model, <http://blog.daum.net/fmddn/1787002>, 2012.12.17.
 [3] C. B. Cinzia Cappiello, C. Francalanci, A. Maurino, Methodologies for data quality assessment and improvement, ACM Computing Surveys 41(3), p.

- 52, 2009.
- [4] D. Katz, M. Bommaroti, J. Zelner, The data deluge, The Economist, Mar 1, 2010
- [5] T. Shanker, M. Richtel, Data overload can be deadly, The New York Times, Jan 16, 2011.
- [6] Noreen Kendle, The Enterprise Data Model, The Data Administration Newsletter, Jul 1, 2005.
- [7] Min Kyu Lee, Data Performance Cases in relation to the Removal of Repetitive Data Connections, Graduate School of Information Science and Technology of Soongsil University , 2010.
- [8] Richard Y. Wang, Henry B. Kon, Stuart E. Madnick, Data quality requirements analysis and modeling, [Proceedings. IEEE Ninth International Conference on Data Engineering](#), pp.670 - 677, 1993.
- [9] Yong Rak Choi, Data Modelling Practices, Munwoondang, 2010.
- [10] Ki Won Jung, Data Modelling Practices, Brain Korea, 2004.
- [11] Practical Project of Data Model Normalization / De-normalization, <http://blog.naver.com/jooyong3/40035951092>, Mar 29, 2007.
- [12] I. Davies, P. Green, M. Rosemann, M. Indulska, S. Galo, How do practitioners use conceptual modeling in practice?, Data and Knowledge Engineering, 58, pp.358-380, 2006.
- [13] H. Rhee, Corporate data obesity: 50 percent redundant, Journal of Computer Science and Technology, 10(5), pp.7-11, 2010.
- [14] H. Rhee, A rare example of pitfall in corporate data modeling practices, Journal of Computer Science and Technology, 10(5), pp.7-11, 2011.

이혜경(Rhee, Hae Kyung)



- 1979년 2월 : 숭실대학교 전자계산학과 졸업
- 1985년 4월 : University of Illinois (Urbana-Champaign) 전산학과(공학석사)
- 2000년 2월 : 성균관대학교 정보공학과(공학박사)
- 1988년 3월 ~ 1989년 2월 : 국립천안공업전문대학 전자계산과 전임강사
- 1992년 3월 ~ 2001년 8월 : 경인여자대학 멀티미디어정보전산학부 조교수
- 2001년 9월 ~ 현재 : 용인송담대학교 컴퓨터게임과 부교수
- 관심분야 : 데이터베이스, 데이터 모델링, 데이터 보안, 동시성제어
- E-Mail : leehk@ysc.ac.kr

김희완(Kim, Hee Wan)



- 1995년 8월 : 성균관대학교 정보공학과(공학석사)
- 2002년 2월 : 성균관대학교 컴퓨터공학과(공학박사)
- 1996년 5월 : 정보관리기술사 취득
- 1997년 1월 : 정보시스템 수석감리원 자격 취득
- 2001년 3월 ~ 현재 : 삼육대학교 컴퓨터학부 교수
- 관심분야 : 정보시스템 감리, 프로젝트 관리, 데이터베이스, 소프트웨어 공학
- E-Mail : hwkim@syu.ac.kr