

데이터 모델링을 위한 엔터티의 표현방법과 활용방안

이화식*, 김현산**, 양해술**

*(주)엔코아컨설팅.

**호서대학교 벤처전문대학원

kimsan1@hotmail.com

hsiee@en-core.com, hsyang@office.hoseo.ac.kr

Expression method and application plan of the entity for a data modeling

Hwa-Sik Lee*, Hyun-San Kim**, Hae-Sool Yang**

*EN-CORE Consulting,

**Dept. management Information, hoseo univercity

요 약

엔터티는 데이터 모델링에서 가장 중요한 요소라고 할 수 있다. 엔터티를 우리말로 '실체(實體)'라고 번역하고 있다. 말 그대로 '실질적으로 육체를 가지고 있는 것'을 의미한다. 만약 이 실체가 모호하다면 그 다음 과정이 아무리 완벽하다고 하더라도 모델링은 전체가 모호한 모습일 수밖에 없다. 이 논문에서는 엔터티를 보다 쉽게 도출하고 전략적인 판단을 해 가는 구체적인 단계와 객관적인 판단의 근거를 매우 상세하게 설명할 것이다. 엔터티는 매우 명확하게 표현해야 한다. 이처럼 데이터 모델링에서는 실체 집합인 '엔터티'를 정확하게 도출할 수 있는 것이 무엇보다도 중요하다. 잘못된 실체를 가지고 아무리 집을 지어본들 그것은 단지 사상누각에 불과 하다. 본 연구에서는 데이터베이스 엔터티 각 각의 표현 방법에 대해서 연구함으로써, 데이터 모델을 보다 정밀하게 표현할 수 있고, 남들이 모델을 좀 더 정확하게 이해하는데 기여할 수 있다.

1. 서론

데이터 모델링의 근본은 엔터티를 명확하게 정의하는 것에서부터 비롯된다는 것을 확신한다. 실체가 모호한 것보다 아무리 릴레이션십을 맺고 속성을 채워 본들 모두가 사상누각일 수밖에 없다는 것은 절대 변할 수 없는 진리라고 믿기 때문이다. 어쩌면 너무 어렵고 복잡하다고 생각하는 사람도 있을 것이다. 그렇지만 사고(思考)의 세계는 명령어를 취급하는 세계와는 분명히 다르다. 사고의 세계는 어차피 고도의 판단력으로 결정된다. 판단력이란 확신을 가질 수 있는 관련 요소를 얼마나 완벽하게 확보했느냐에 따라 크게 좌우된다. 이러한 이유로 이 논문은 판단 기준을 최대한 구체적으로 제시하려 노력했고, 실전에서 일어나는 현실적인 사례를 최대한 다양하게 소개하기 위해 노력 했다.

2. 유형별 엔터티 표현방법

지금까지 우리가 정의해 왔던 엔터티들을 목적이

나 용도, 혹은 뚜렷한 특징들을 나타내기 위해 형태별로 구분해 둘 필요가 있다.

세상(데이터 모델)에는 그 주체가 되는 인간(엔터티)이 존재한다. 인간에게 존재하는 이러한 유형들이 엔터티에서도 거의 유사하게 존재한다. 살아 있는 정상적인 사람들을 일반(normal) 엔터티, 마치 식물 인간처럼 앞으로 다른 것으로 대체를 하거나 통합시키기 위해 아직은 없애지 않고 남겨 두었지만 삭제한 것이나 다름없는 의미의 제거(drop) 엔터티, 외계에서 온 우주인과 유사한 외부(external) 엔터티, 만화의 주인공과 비슷한 가상(pseudo) 엔터티, 손오공의 분신술로 만든 복제(clone) 엔터티, 역할을 대신하기 위해 존재하는 대체(代替, substitute) 엔터티, 임신 중이기는 하지만 곧 세상에 태어날 준비를 하고 있는 의미인 추가(additional) 엔터티로 분류했다.

이와 같은 여러 유형의 엔터티를 적절하게 분류해서 지정해 두는 것은 여러 가지의 의미를 가진다. 가령, 데이터 모델을 보다 정밀하게 표현할 수 있고,

본 연구는 대학 IT 연구센터 육성.지원사업의 연구결과로 수행되었음.

남들이 모델을 좀 더 정확하게 이해하는데 기여할 수 있으며, 모델링 진행과정에서 작업 대상을 식별하거나 진행관리를 할 수 있다는 부가적인 장점도 있으므로 많이 활용할 필요가 있다.

위와 같은 여러 유형의 엔터티를 적절하게 분류해서 지정해 두는 것은 여러 가지의 의미를 가진다. 가령, 데이터 모델을 보다 정밀하게 표현할 수 있고, 남들이 모델을 좀 더 정확하게 이해하는데 기여할 수 있으며, 모델링 진행과정에서 작업 대상을 식별하거나 진행관리를 할 수 있다는 부가적인 장점도 있으므로 많이 활용할 필요가 있다.

2.1 제거(drop) 엔터티

제거 엔터티를 남겨두는 이유는 마치 본인이 개인용 컴퓨터에서 파일을 삭제시키면 원래의 디렉토리에서는 삭제되지만 '휴지통'에 들어가 있음으로써 이를 복원하거나 내용을 참조해 보거나 영구히 삭제시킬 수 있는 기회를 제공해 주려고 배려를 하는 것과 같은 이유이다. 이미 나름대로의 확실한 역할을 하고 있던 어떤 존재가 갑자기 혼자만 함부로 사라질 수는 없기 때문에 설혹 제거를 하겠다고 하더라도 잠정적인 삭제 상태로 만들어 두고, 주변 관계를 말끔하게 정리한 후에야 원하는 대로 사라지게 하여야 한다. 엔터티를 제거 상태로 표시하게 되는 또 하나의 활용 형태는 제거 엔터티란 의미는 앞으로 없애 버릴 엔터티를 표시해 두겠다는 것이다. 이 말은 곧 ERP패키지의 데이터 모델을 커스터마이징할 때나 리버스 모델링을 하여 만들어진 현행(as-is) 데이터 모델에서 새로운 목표(to-be) 데이터 모델을 결정해갈 때 앞으로는 사용하지 않고 버리겠다는 의사표시로도 활용할 수 있다는 것을 말한다.

2.2 외부(external) 엔터티

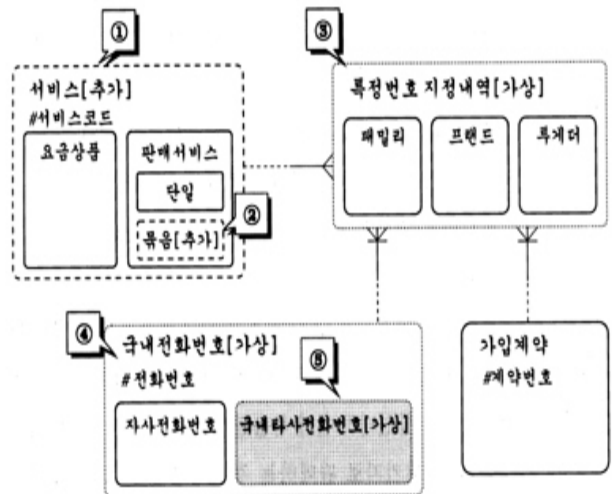
외부 엔터티는 자기 집합을 제외한 나머지 의미를 가지고 있다. 즉, 외부라는 집합을 정확히 정의하기 위해서는 먼저 내부 집합이 구체적으로 정의되어야 함을 뜻한다. 사실 일견 당연해 보이는 이 개념은 보는 사람의 시각에 따라 다를 수가 있다. 가령, 회사 전체 시스템을 내부로 보고 타사에 있는 개념적인 집합들을 외부로 볼 수 있는 방법이 있다. 회사 내에서도 OLTP, DW, 혹은 계정계, 정보계라는 이름으로도 시스템을 구분해 볼 수 있다. 아니면 H/W 환경이나 DBMS에 따라 별도의 시스템으로 구분하는 방법도 있을 것이다. 또한 협의로 본다면 같은 OLTP내에서도 '인사, 회계, ...'와 같이 서브시스템 별로 나눈 각각을 내부로 생각할 수도 있다.

그러나 외부 엔터티란 우리 중의 누구도 관리하거나 책임질 필요가 없으며, 우리가 관리하는 어떤 ERD에서도 소유주가 존재하지 않는 그야말로 자신이 아닌 제3자가 관리하는 엔터티란 의미가 숨어 있다. 이 외부 엔터티는 우리가 소유하고 있는 엔터티가 아니므로 나중에 구현 단계에 가서라도 직접 릴레이 선섭을 가지고 액세스 할 수 없는 엔터티라는 의미도 같이 포함되어 있다.

2.3 가상(pseudo) 엔터티

가상 엔터티란 말 그대로 상상 속에서 존재한다는 뜻이다. 다시 말해서 개념적으로만 존재하는 엔터티를 말한다. 하나의 엔터티에는 여러개의 슈퍼타입과 여러개의 서브타입이 존재-즉, 다양한 논리적인 집합이 존재 하고 있다. 이러한 논리적인 집합들은 나중에 가서 물리적인 테이블이 될 자격은 가지고 있지만 반드시 물리적인 테이블이 되어야 할 필요는 없다. 가상 집합을 활용했을 때 커다란 의미를 가질 수 있는 경우에 대해 연구해 보았다. 첫 번째로 기존의 시스템을 리버스 모델링하면서 논리화 시킬 때 적용하는 형태이다. 이 방법은 현재 물리적으로 존재하고 있는 테이블이 원래 논리적으로는 어떤 엔터티에 어떤 형태로 소속되었어야 하는가를 표현할 때 사용 하는 것으로써 리버스 모델링에서 반드시 필요한 개념이다. 두 번째 활용 형태는 어떤 행위 엔터티의 본질 식별자를 정확하게 밝히고자 할 때 부모중에서 원래는 별도의 엔터티로 존재해야 했지만 어떤 이유로 인해 실존 엔터티를 가질 수 없는 경우 그 엔터티를 표현하기 위한 방법이다. 세 번째로 엔터티가 아니라 서브타입에 대해서도 동일하게 가상의 집합을 표현하는 것이다. 위의 세 가지 활용 형태를 좀더 구체적으로 살펴보기 위해 활용형태를 보

(그림1) 가상엔터티의 세가지 활용형태



졌다. 위 그림에 있는 번호들을 보면 ①,②는쇄선이고 ③,④,⑤는점선으로 되어있다. ①번은 쇠선이므로 이 엔터티는 지금 새롭게 정의하려고 시도하는 과정에 있다는 것을 의미 하고있다. ②에 있는 '묶음' 서브타입을 쇠선으로 표시한 것은 지금까지는 전혀 존재하지 않았던 새로운 개념의 집합을 추가해 보겠다는 의미로 사용된 것이다. ③번 엔터티는 어떤 가입계약이 보유한 서비스 중에 대응되는 몇 개의 상대전화번호를 지정하면 그들과 통화할 때는 통화료의 일정부분을 할인해 주겠다는 할인대상정보를 담고 있는 엔터티이다. 이렇게 모인 국내의 모든 전화번호가 있는 집합은 실존 할 수 없기 때문에 점선으로 표시 할 수밖에 없다.그림에 있는 ③은 ①번과 유사해 보이기도 하지만 이를 점선으로 표현한 것에는 미묘한 차이가 있기 때문이다. ③은 현재 존재하지 않는 집합이다. 지금은 유사한 목적을 위해 생성된 것처럼 보이는 것들을 단지 한 곳에 보아서 표현해 두려는 단순한 목적이었다면 가상의 엔터티로 보는 것이 좋겠다는 것을 보여 주려는 것일 뿐이다. 물론 쇠선으로 표현하여 현존하지 않지만 앞으로는 존재해야 할 것이라고 표현 돼는 것도 당연히 가능하다.

2.4 대체(substitute) 엔터티

대체 엔터티란 야구에서 사용하는 '대타'나 '대주자', 혹은 조립공정을 가진 생산라인에서 주로 사용하는 '대체 자재'라는 개념과 매우 유사하다. 이들이 가지고 있는 공통적인 특성은 동일한 하나의 역할을 위해서 한 가지 이상의 실체가 존재할 수 있다는 것이다.

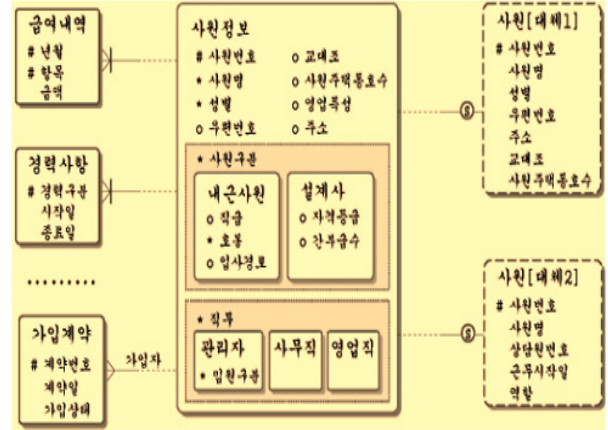
2.4.1 대체 엔터티 - 대안수립용

가령 A~I까지 9개의 엔터티로 구성된 데이터 모델이 있다고 가정해 보자. 그런데 A엔터티 대신에 A'로 하는 것도 생각해 볼 수 있고, B엔터티를 B'로 하는 경우에 대해서도 충분히 검토할 가치가 있다고 가정하자. 대신 생각을 단순하게 하기 위하여 나머지 7개의 엔터티는 복수안이 없다고 가정한다. 그러나 이처럼 단 2개의 엔터티가 복수안을 가지고 있다고 하더라도 우리는 총 4벌의 ERD를 추가로 그려야 각각의 경우에 대비할 수 있다.

엔터티가 단 9개뿐인 ERD 에서 2개 엔터티만 복수안을 가지고 있어도 이럴진대 수백 개의 엔터티를 가지고 있으면서 수십 개의 엔터티가 여러 개의 대안을 가지고 있다면 도대체 몇 벌의 ERD가 필요하겠는가?

이와 같은 경우 대표되는 9개의 엔터티를 선정하고, A' 및 B'는 대안으로서 관리하면 한벌의 ERD만 가지고서 각각의 경우에 대처 할 수 있을 것이다.

2.4.2 대체 엔터티의 형태조정



(그림2) 엔터티 형태 일부 조정 시

어떤 논리적 모델링을 가지고 다음 단계인 데이터베이스 설계 단계에서 분산 시스템으로 설계하여야 한다고 가정해 보았다. 데이터베이스 설계 단계에서는 시스템의 물리적인 요소들을 감안해야 하고 해당 노드의 특성도 반영되어야 하므로 그림 우측처럼 각 노드별로 가져 갈 속성과 개체집합의 범위가 서로 다르게 결정되기로 했다고 가정했다.

이때 우리가 취할 수 있는 방법은 크게 두 가지가 있다.

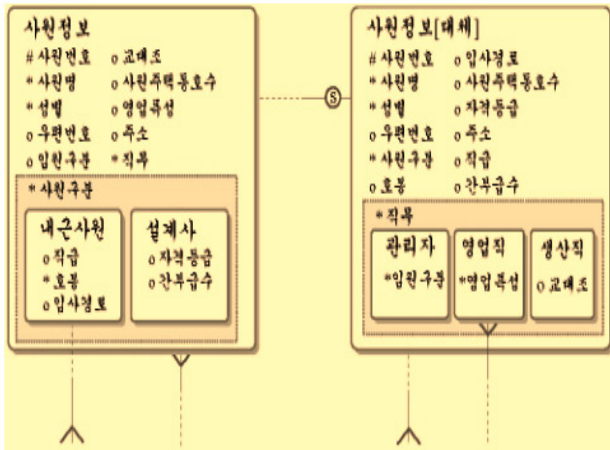
첫 번째 방법은 데이터 모델에서는 하나의 엔터티만 정의하고 데이터베이스 설계 단계에서 노드별로 서로 다르게 테이블을 정의하는 것이다.

두 번째 방법은 지금 설명하고자 하는 대체 엔터티를 활용하는 방법이다. 데이터 모델링 단계에서 미리 각 노드별 특성을 감안하여 각 노드별로 대체 엔터티를 매달아 두는 방법이다. 이 대체 엔터티는 원래의 엔터티와 동일한 집합일 수도 있지만 자기만 필요한 속성(가로)과 개체(세로)들로 구성할 수가 있다.

마찬가지로 방법으로 우측 하단에 있는 대체 엔터티는 요즈음 고객관리를 위해 많이 활용하고 있는 콜센터나 고객지원센터 시스템에서 발생할 수 있는 예제이다. 대부분의 회사에는 이처럼 이미 특수한 목적을 위해 존재하는 시스템이 있는데 대개의 경우 기존과는 조금 다른 아키텍처로 구축되어 있는 경우가 많다. 물론 새로운 통합 데이터 모델을 사용하도록 시스템을 변경할 수도 있겠지만 만약 특별한 제

한요소가 있다면 어쩔 수 없이 그들만의 고유한 엔터티를 생성할 수밖에 없을 것이다. 이러한 경우에도 별도의 ERD를 작성할 필요 없이 대체 엔터티를 활용함으로써 쉽게 해결할 수 있다

2.4.3 대체엔터티-서브타입의 상세표현



(그림3) 서브타입의 상세표현

이 활용 형태는 데이터 모델을 좀 더 명확하게 표현하기 위해 여러 각도에서 바라 본 모습을 각기 다른 엔터티로 그려 볼 때 활용하는 방법이다.

물론 우리는 서브타입세트를 이용하여 하나의 엔터티 내에서도 여러 차원에서 바라 본 데이터 모델을 그려낼 수가 있다. 사실 서브타입세트를 이용하여 여러 각도에서 바라 본 집합의 형태를 표현하는 경우는 다음과 같은 상황일 때 가장 적절하다. 즉, 가장 대표적인 서브타입세트로 전체를 표현했지만 그 중에서 일부분에 대해서는 좀 더 상세하게 보강하여 표현하고 싶을 때 이러한 부분만 강조한 서브타입세트를 그리고자 하는 경우이다.

이와는 다르게 만약 모든 유형의 서브타입세트마다 전체 집합을 나타내려 한다면 불가능하지는 않지만 그림이 너무 복잡해지게 된다. 이러한 경우 대체 엔터티를 사용하면 매우 쉽고 단순한 모양으로 표현할 수 있다. 다음 그림은 이런 경우에 대체 엔터티를 사용한 사례를 보여주고 있다.

그림에는 나타나 있듯이 서브타입세트에 따라 데이터 모델을 바라 본 모습이 서로 다른 것을 발견할 수 있다. 각 속성들의 위치 - 즉, 슈퍼타입이나 서브타입에 대한 소속 - 가 서로 달라질 수 있으며, 속성에 붙어 있는 선택사양은 물론 릴레이션쉽 또한 들어오고 나가는 위치가 서로 다르게 표현된다는 것을 알 수 있다.

2.5 추가(additional) 엔터티

추가 엔터티를 활용하는 경우는 크게 세 가지로 나눌 수 있다. 첫 번째는 리버스 모델링에서 현재 (as-is) 시스템에는 없지만 리버스를 하는 과정에, 혹은 논리화를 하는 과정에서 앞으로 있어야 할 것으로 예상되는 엔터티가 나타났을 때 이를 일단 정의해 두고 나중에 최종적으로 확정하겠다고 할 때 적용하는 경우이다. 두 번째의 활용 형태는 이미 구축되어 운용되고 있는 시스템의 데이터 모델에 업무의 변화로 인해 데이터 모델의 변경을 시도하는 경우에 활용하는 경우이다. 운용 중인 시스템의 데이터 모델이라면 이미 거기에 있는 엔터티는 확정되어 있는 것들이다. 확정되어 있는 수많은 엔터티들 사이에 새롭게 정의 - 즉, 추가 할 엔터티가 별도의 표시를 가지지 않는다면 뒤섞여 버려서 찾기가 어려워질 것이다. 이 엔터티는 아직 완전히 확정이 되지 않았을 뿐이지 결코 가상의 집합이라고는 할 수 없으므로 가상엔터티와는 별도로 관리해야 한다.

활용 형태의 세 번째는 ERP패키지를 도입하여 우리에게 맞도록 데이터 모델을 커스텀마이징하는 경우에 적용하는 경우이다. 즉, 패키지에 원래부터 있던 데이터 모델을 기존의 것으로 보고 우리 업무에 맞도록 새롭게 추가한 엔터티들을 추가 엔터티로 표시하는 방법이다. 적용방법은 다르지 않지만 패키지를 도입할 때는 충분히 의미가 있을 것이다.

3. 결론

우리는 이 와같이 엔터티의 여러표현 방법들을 연구해 보았다. 즉 서론에서 언급했듯이 데이터 모든 근원이 엔터티로 정의되어질 수 있다는 것을 각각의 유형별로 살펴본 것이다. 어떠한 시스템의 데이터모델링은 확실한 엔터티의 영역으로 수립관리할 필요가 있다는 것이 입증된 것이다. 그 것으로부터 우리는 구체적인 시스템 분석 및 설계를 위한 데이터 모델링의 기본사상을 구축할 수 있고 보다 전략적으로 비즈니스 문제해결 능력을 갖추게 될 것이다.

본 논문에서는 엔터티의 표현방법에 대해서 연구해 보았지만 향후 연구 과제는 엔터티 정의서작성 방법에 대해서 연구를 진행할 예정이다.

참고문헌

[1]Data Architecture Solution 이화식 지음
 [2]대용량데이터베이스솔루션 이화식,조광원 지음
 [3]엔터프라이즈애플리케이션아키텍처패턴 마틴파올러
 [4]CRM을 위한 데이터 마이닝 알렉스버슨,스테판 스미스 지음
 [5]데이터베이스 시스템 SILBERSCHATZ