

이론과 실무를 겸비한 최고의 전략서!

RELATIONAL DATA MODELING PREMIUM GUIDE

관계형 데이터 모델링 프리미엄 가이드

김기창 지음



01 데이터 모델링에 대한 상념

- 1.1. 데이터 모델링이 왜 어려운가? | 15
- 1.2. 데이터 모델링의 매력 | 17
- 1.3. 모델링은 상식적이다 | 18
- 1.4. 모델러와 바둑 프로기사 | 19
- 1.5. 좋은 모델은? | 20
- 1.6. 모델링이 왜 필요한가? | 21
- 1.7. 좋은 모델러란? | 23
- 1.8. 모델링 목표 | 25

02 데이터 모델링 기본 개념

- 2.1. 관계형 데이터 모델링 | 29
- 2.2. 무결성 | 32
- 2.3. 데이터베이스 라이프 사이클 | 37
- 2.4. 주제 영역 | 39
- 2.5. 데이터 표준화 | 42
- 2.6. ERD | 46

03 개념 모델 & 논리 모델 & 물리 모델

- 3.1. 개념 모델 | 51
- 3.2. 논리 모델 | 60
- 3.3. 물리 모델 | 65

04 정규화(Normalization)

- 4.1. 정규화란? | 73
- 4.2. 정규화의 목적 | 76
- 4.3. 아노말리 | 78
- 4.4. 함수 종속 | 82
- 4.5. 정규형의 종류 | 86
 - 1정규형 | 87
 - 2정규형 | 96
 - 3정규형 | 98
 - 보이스코드 정규형 | 102
 - 4정규형 | 106
 - 5정규형 | 110
- 4.6. 정규형과 성능 | 114

05 데이터 통합(Generalization)

- 5.1. 데이터 통합 | 119
- 5.2. 데이터 통합의 장단점 | 124
 - 데이터 통합의 장점 | 124
 - 데이터 통합의 단점 | 125
- 5.3. 엔터티 통합 대상 | 128
- 5.4. 데이터 오너십 | 143

06 슈퍼타입(Supertypes)과 서브타입(Subtypes)

- 6.1. 슈퍼타입과 서브타입 정의 | 147
- 6.2. 슈퍼타입과 서브타입의 사용 방법 | 153
- 6.3. 서브타입의 종류 | 158
- 6.4. 서브타입의 물리모델 변환 | 169
 - 서브타입별 엔터티로 분할(타입1) | 173
 - 하나의 엔터티로 통합(타입2) | 177
 - 통합 엔터티와 개별 엔터티 통합(타입3) | 179

07 엔터티(Entity)

- 7.1. 엔터티란? | 189
- 7.2. 자립 엔터티 & 종속 엔터티 | 192
- 7.3. 엔터티 종류 | 198
 - 실체 엔터티 | 200
 - 행위 엔터티 | 202
 - 가공 엔터티 | 204
 - 기준 엔터티 | 206
- 7.4. 엔터티 도출 원칙 | 211
- 7.5. 엔터티명 | 217
- 7.6. 엔터티 검증 | 222

08 식별자(Unique Identifier)

- 8.1. 업무 식별자란? | 225
- 8.2. 식별자와 키 종류 | 231
- 8.3. 주 식별자 선정 원칙 | 236
- 8.4. 주 식별자 선정 절차 | 254
- 8.5. 주 식별자 상속 | 260
- 8.6. 인조 식별자 채번 | 264
- 8.7. 복잡한 주 식별자 | 266
- 8.8. 식별자 검증 | 273

09 속성(Attributes)

- 9.1. 속성이란? | 277
- 9.2. 속성 종류 | 280
 - 기초 속성 | 281
 - 관계 속성 | 282
 - 추출 속성 | 282
 - 시스템 속성 | 284
- 9.3. 도메인 | 285
- 9.4. 복합 속성과 다가 속성 | 288
- 9.5. 속성 명 | 292
- 9.6. 코드 | 294
 - 속성 코드와 식별자 코드 | 296
 - 코드 모델 | 301
- 9.7. 널 | 311
- 9.8. 데이터 타입 | 316
- 9.9. 속성 검증 | 327

10 관계(Relationships)

- 10.1. 관계란? | 329
- 10.2. 관계와 속성 그리고 엔터티 | 336
- 10.3. 관계 구성 요소 | 345
 - 카디널리티 | 346
 - 옵셔널리티 | 349
 - 관계 디그리 | 359
- 10.4. 관계 표현 | 364
 - 관계명 | 364
 - 관계선 | 372
- 10.5. 관계 종류 | 382
 - 일대일(1:1) 관계 | 382
 - 배타 관계 | 390
 - 재귀 관계 | 398
 - 추출 관계 | 409
 - 양방향 관계 | 412
- 10.6. 참조 무결성 | 415
- 10.7. 관계 검증 | 417

11 모델링 방법론 (Modeling Methodology)

- 11.1. 하향식과 상향식 | 421
- 11.2. 리버스 모델링 | 427
- 11.3. 모델링 프로젝트 WBS | 446

12 이력 관리

- 12.1. 이력 데이터란? | 461
- 12.2. 선분 이력 | 466
- 12.3. 이력 엔터티 선정 절차 | 470
- 12.4. 이력 관리 모델 유형 | 472
- 12.5. 이력 엔터티의 주 식별자 | 496
- 12.6. 정정 데이터 | 502

13 비정규화(Denormalization)

- 13.1. 비정규화란? | 507
- 13.2. 비정규화 과정 | 511
- 13.3. 비정규화의 단점 | 515
- 13.4. 비정규화 방법 | 518
 - 엔터티 합체 | 518
 - 엔터티 분해 | 521
 - 요약 엔터티 추가 | 525
 - 중복 속성 채택 | 526
 - 추출 속성 채택 | 529
 - 반복 속성 채택 | 534
 - 중복 데이터 채택 | 535

14 물리적(DBMS) 요소

- 14.1. 테이블 타입 | 537
 - 클러스터링 테이블 | 538
 - IOT | 539
- 14.2. 파티션 | 540
- 14.3. 인덱스 | 542
- 14.4. 뷰 | 545

Bibliography | 550
Index | 551

Prologue

“데이터 프로페셔널(Data Professional)을 위해...”

1만 시간의 법칙이 있다고 한다. 어느 분야에서든 세계적인 전문가가 되기 위해서는 1만 시간의 노력이 필요하다고 한다. 1만 시간보다 적은 시간을 몰두하고 그 분야에서 세계적인 전문가가 된 예는 찾아볼 수 없다고 한다.

1만 시간은 하루에 7~8시간씩 약 5년 동안의 시간이다. 엄청난 시간일 수 있지만, 필자는 그다지 오랜 시간이 아니라고 생각한다. IT 분야에서 일하면서 한 분야에서 1만 시간을 채우는 것은 어렵지 않다. 다만, 몰두해야 하는 시간이기 때문에 쉽게 이를 수 있는 시간은 결코 아니다.

진정한 프로페셔널이 되고자 하는 사람은 1만 시간을 도전해 보았으면 한다. 모델을 고민하고 연구하면서 모델링에 몰두하는 시간이 10년 정도 되면 데이터 모델링 분야에서 세계 최고가 될 수 있다. 단순히 모델링 분야에 참여한 노동의 시간이 아니라 고민하고 연구하는 몰두의 시간이어야 한다.

누군가의 경험을 그대로 답습만 한다면 그 분야의 최고가 될 수 없다. 데이터 모델링 분야도 마찬가지로 응용하고 연구해야 전문가가 될 수 있다.

데이터 모델링 분야는 세계 최고의 전문가가 되는 것을 도전해볼 만한 가치가 있는 분야다. 독자들의 1만 시간으로 가는 길에 이 책이 10시간이라도 차지했으면 한다.

“데이터 모델링은...”

데이터 모델링은 어려운 분야지만 상식적인 분야이기도 하다. 상식적이라는 것은 쉽다는 것을 의미한다. 상식을 벗어나면 한없이 어려워질 수 있다. 기초적인 이론부터 차근차근 습득하면 결국 상식처럼 쉬워지는 것이 모델링이다.

데이터 모델링을 예술에 비유하는 전문가가 많다. 하얀 백지 위에 작도하는 느낌이 비슷하고 무에서 유를 창조한다는 면에서 예술과 비슷하다. 데이터 요건을 고뇌해서 모델을 만들어내는 과정은 창작의 그것과 유사하다.

하지만 모델은 상상력이나 예술성보다는 이론을 기반으로 만든다는 점에서 예술과 약간의 거리가 있다. 머릿속에서는 함수 종속과 SQL이 흐르고 있다. 정규화와 데이터 통합, 서브타입, 이력, 비정규화 등은 이론으로 이루어진다.

데이터 모델링은 매력적인 분야다. 이론을 기반으로 다양하고 종합적인 판단을 통해 이루어진다. 창작의 기쁨과 유사한 희열을 느낄 수도 있다.

기업의 핵심 요소인 데이터를 최전선에서 다루는 사람이 데이터 모델러다. 데이터 요건(Information Requirements)은 더욱 다양해질 것이고 인간의 판단력으로 이루어지는 데이터 모델링은 더욱 필요해질 것이다. 이론과 통찰력을 바탕으로 진행되는 데이터 모델링은 유망하면서 매력적인 분야이다.

“책의 내용은...”

이 책은 필자의 경험을 위주로 작성됐다. 경험과 사고를 통해 얻은 지식을 정리한 책이다. 책의 내용 중에는 이미 상식적으로 알려진 내용도 있을 테지만 통용되지 않는 내용도 있을 수 있다.

논란이 될 수 있음에도 필자의 생각을 그대로 지면에 옮긴 것은 맞고 틀리고를 판별하려는 것보다는, 독자들이 많은 사고를 통해서 스스로 결정할 수 있는 힘을 키웠으면 하는 바람 때문이다. 고민과 연구의 단초를 제공했으면 한다.

이 책에서는 가능한 많은 예제를 사용하려 노력했다. 하지만 예제는 제시된 개념이나 기법, 의도된 결과에 초점을 맞추고 그것을 설명하기 위해 사용한 것이므로 예제에 집착하지 않았으면 한다. 설명하는 것은 개념이나 기법이지 예제를 설명하기 위함은 아니다.

또한 이 책에서 사용된 많은 예제는 실무에서의 다양한 환경과 요건을 대표하지 않는다. 일반적이라면 필

자가 생각하는 최선의 모델 또는 차선의 모델일 뿐 항상 예제 모델처럼 사용해야 하는 것은 아니다.

특별한 언급이 없는 한 현행 데이터베이스가 존재한다는 가정에서 설명했다. 최근 프로젝트는 대부분 현행 데이터베이스가 존재해 그것을 활용하여 수행된다.

“실무와 이론을 겸비한 전략서-관계형 데이터 모델링 프리미엄 가이드”

필자는 이 책의 제목이 촌스럽다고 생각한다. 하지만 이 책을 가장 잘 설명한 것으로 생각한다. 이 책은 실무에서 모델링을 수행하면서 얻은 이론이나 노하우를 기반으로 쓰였다. 실무 경험을 토대로 한 모델링 기법과 노하우는 필자 나름의 확고한 이론적 뒷받침이 돼 있다.

필자는 모델링을 잘하고자 오랜 기간 노력했고, 더 잘하려고 즐기면서 모델을 만들었다. 지금도 항상 즐거운 마음으로 연구한다. 독자들도 데이터 모델링의 묘미를 만끽하면서 모델을 만들 수 있길 기원한다. 알려진 대로 즐기는 사람이 천재도, 노력하는 사람도 뛰어넘을 수 있다.

이 책은 관계형 이론을 기반으로 썼지만 현실적인 면에서도 접근했으므로 모델링 전략서의 역할을 할 수 있으리라 기대한다. 전반적인 책의 내용을 습득하고 지식을 확장하면 데이터 전략가가 될 수 있을 것이다.

책 내용이 이론과 현실 사이에서 갈팡질팡하기도 하지만 필자는 이론을 중요하게 생각하는 사람이다. 독자들도 이론 없이 지은 50층짜리 아파트에서 살고 싶지는 않을 것이다.

이 책에서 이론 이외에 강조하는 것은 실전 모델링이다. 현실을 외면한 이론이 되지 않도록 하는 것이 이 책을 쓰는 목적이다. 실전적인 전략가가 돼야 한다.

책 전반적으로 모델링 일정을 자주 언급한 이유는 실전을 위함이다. 리버스 모델링을 별도로 설명한 이유도 실전 모델링이라 생각하기 때문이다. 일정이 고려되지 않은 모델링은 실전에서 의미가 반감된다.

다양한 환경의 프로젝트 일정을 고려한 모델링 전략을 세울 수 있어야 한다. 이론과 실전을 전달하는 것이 이 책을 쓰는 필자의 의도이다.

“감사를 전합니다”

모델링에 관한 첫 책을 출판해 주신 오픈메이드컨설팅 최영철 사장님과 출판을 총괄해 주신 김기호 전 무님께 감사드립니다

이 책은 약 3년여 동안 쓰였습니다. 시작할 때 이렇게 오래 걸릴 줄 알았다면 아마 엄두가 나지 않아 시작조차 못 했을 것입니다. 긴 시간 동안 가족의 지원이 없었다면 중간에 포기했을 것입니다.

아빠를 언제나 기다리던 어린 겐이와 솔이에게도 인내의 시간이었을 것입니다. 누구보다 필자와 같은 고통의 시간을 보냈을 아내에게 고마움을 전하고 싶습니다.

위즈덤마인드 데이터 모델링 시리즈 - 프리미엄 가이드 1
이론과 실무를 겸비한 전략서

관계형 데이터 모델링 프리미엄 가이드 개정판

저자 김기창 bluepupi@gmail.com

데이터 분야에서 15년 이상 일하고 있으며, 현재는 위즈덤마인드(주)에서 대표 컨설턴트로서 데이터 모델링(Data Modeling)과 DA(Data Architecture) 컨설팅을 하고 있다. 특별히 풍부한 실전을 바탕으로 데이터 모델링을 직접 수행하며, 실무 DA 컨설팅을 하는 것이 강점이다.

[관계형 데이터 모델링 프리미엄 가이드]에 이어 [관계형 데이터 모델링 노트]를 저술했으며, 2007년에는 [전사적 데이터 아키텍처 프레임워크에 대한 개념모델 개발] 논문을 발표했다.

모델러가 기업에게 제공할 최고의 가치는 좋은 모델 제공이라고 생각하고 있다. 소명을 갖고 많은 기업에서 진짜 모델이 운영 되는 것을 꿈꾸고 있으며, 그런 모델을 설계하는 진짜 모델러가 많아질 수 있도록 노력하고 있다.

관계형 데이터 모델링 프리미엄 가이드 개정판 Copyright© 2020 by Ki Chang, Kim

Published by WISDOMMIND, Korea

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage or retrieval system, without the prior written permission of the copyright owner.

초 판 1쇄 발행 | 2014년 1월 2일

개정판 1쇄 발행 | 2020년 8월 10일

저은이 | 김기창

출판사 | 위즈덤마인드

주 소 | 경기도 부천시 부일로 448, 401호

유통사 | 글봄크리에이티브

주 소 | 서울특별시 관악구 신림로 29 8 A동 402호

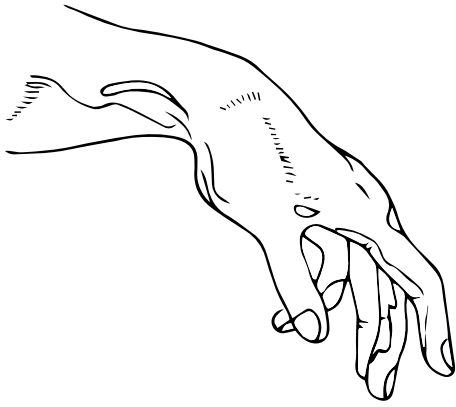
책주문 | 전화: 02)507-2340, 팩스: 02)507-2350, friend@mustree.com

ISBN 979-11-963192-2-9

값 37,000원

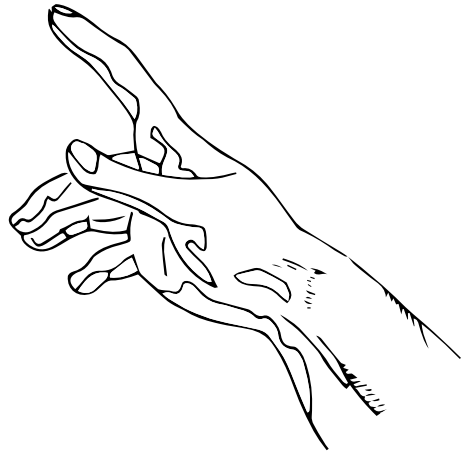
이 책은 저작권의 보호를 받으며 저자의 허락 없이는

어떠한 형태의 번역이나 번안, 유표, 복사, 출판, 게재, 디지털 매체로의 저장 및 전송, 촬영, 녹음도 할 수 없습니다.



네가 자기의 일에 능숙한 사람을 보았느냐
이러한 사람은 왕 앞에 설 것이요
천한 자 앞에 서지 아니하리라

- 잠언 22:29 -



데이터 모델링에 대한 상념

1.1. 데이터 모델링이 왜 어려운가?

데이터 모델링은 어렵다. 책의 첫 문장이 독자의 의욕을 저하시키는 게 아닌가 우려되지만 사실이다. 모델링 이론이 어렵다는 말은 아니다. 이론은 다른 분야의 것과 비교해서 어렵지 않다고 생각한다. 하지만 이론을 알고 있더라도 막상 시작하려면 무엇을 어떻게 해야 할지 막막해진다. 또한 완료하고 나서는 모델링이 제대로 됐는지에 대한 확신이 없다. 시작하기 어렵고 결과를 확신할 수 없으니 어려운 분야임이 틀림없다.

이유는 여러 가지가 있을 것 같다. 우선 모델링 이론 외에 알아야 하는 분야의 폭이 넓다. 모델링은 데이터의 본질을 통찰해야 하는 추상적인 개념에서부터 DBMS(데이터베이스)의 구체적인 특징과 기능까지 알아야 제대로 수행할 수 있다.

데이터의 집합을 정의하는 것은 쉽지 않다. 유사한 집합을 일반화(Generalization)하는 것도 쉬운 작업이 아니다. DBMS의 기본적인 특징을 알아야 하는 것은 기본이고 성능 관점의 구체적인 지식을 습득해야 한다. 개발자보다 SQL 작성 능력이 뛰어나야 하고 PLAN을 분석해 효율적인 구조의 모델을 선택할 수 있어야 한다.

그리고 사실상 정답이 없다는 점이 어렵다. 공식이나 틀에 맞추기보다는 상황에 따라 판단이 달라져서 정답이 없이 느껴진다. 똑같은 요건도 상황에 따라 최적의 모델이 달라질 수 있다. 하지만 분명히 상황에 맞는 정답은 있다. 최소한 모범 답안은 존재한다. 100점짜리 모델과 0점짜리

모델은 존재하지 않을 수 있지만 10점에 가까운 모델과 90점에 가까운 모델은 존재한다.

흔히 데이터 모델을 데이터 저장소와 같이 인식하고 사용한다. 마치 블랙박스과 같은 엔터티에 무엇이 어떻게 들어 있는지는 상관없이 결과만 제대로 나오면 프로젝트는 이상 없이 진행됐다. 하지만 최근에는 데이터 모델에 대한 관심이 높아져서 블랙박스 안에 존재하는 모델을 공개해 더 좋은 모델이 될 수 있도록 노력한다. 분명히 90점에 가까운 모델은 있으며, 어떻게든 그 모델을 지향해야 한다.

또 다른 어려움은 어떤 모델이 90점에 가까운 모델인지 확인할 수가 없다는 것이다. 잘 된 모델을 확인하는 방법에 대한 연구가 절실하다고 생각한다. 프로그램은 원하는 결과가 나오면 확인할 수 있고 튜닝은 원하는 수치나 현재보다 월등한 수치가 나오면 확인할 수 있지만 좋은 모델은 확인하기 쉽지 않다.

어떻게 하면 좋은 모델이라는 것을 확인할 수 있을까? 전문가는 척 보면 알 수도 있지만 결국 객관화를 시키려면 수치화돼야 할 것이란 생각이 든다. 어느 정도까지 정량화가 가능할지에 대한 연구가 필요하지만, 정량화를 해서 수치로 평가해야 좋은 모델이라는 것을 어느 정도 확인할 수 있다. 물론 수치만으로 평가할 수 없는 분야가 모델링이다. 데이터 모델링은 모델러의 판단력과 통찰력에 의존적인 분야다.

필자가 모델링을 수행하면서 가장 어려워하는 부분 중 하나는 구축된 모델을 관련자한테 설득하는 것이다. 사실은 설명하는 것인데 결국 설득으로 이어지는 경우가 많다. 사용자를 설득해야 하고 개발자를 설득해야 한다. 데이터 집합(주제)을 정의한 개념적인 내용을 설명하고 이해시키는 것은 쉽지 않다. 다양한 이해관계가 있는 사람을 설득하는 것 또한 쉽지 않다. 특히 실제로 모델을 사용해서 개발해야 하는 개발자를 설득하는 것은 힘들 때가 많다.

필자도 개발하는 것으로 일을 시작해 개발자의 노고를 알 수 있다. 필자는 아직도 개발자가 가장 순수한 집단이라고 생각한다. 하지만 현행 소스를 그대로 사용할 목적으로 모델 변경을 완강하게 반대하는 주장에는 실망스러운 게 사실이다. 비록 효율적인 모델이 되더라도 이미 익숙해 있는 모델을 바꾸는 것에 대해 반대하는 운영자도 마찬가지다. 이런 운영자를 설득하는 것도 힘이 든다. 주객이 바뀐 것에 절망감을 느끼기도 한다. 이렇게 관련자를 설득할 때는 모델러 역시 객관적이고 구체적인 방법으로 설득해야 한다. 이런 여러 가지 이유 때문에 데이터 모델링은 쉬운 분야가 아니다.

1.2. 데이터 모델링의 매력

데이터 모델링은 어려운 분야지만 대단히 매력적인 분야이기도 하다. 필자가 데이터 모델링을 하면서 느끼는 희열은 무에서 유를 창조한다는 느낌이다. 백지상태에서 시작해 데이터 구조가 생성되는 과정은 마치 창작의 과정과 흡사하다. 글을 쓸 때도 타이핑하는 시간보다 고민해서 쓰고 지우고를 반복하는 시간이 훨씬 오래 걸리는 것과 같이 모델링 과정도 마찬가지다. 겉으로 보기에 박스와 선과 텍스트지만 그 안에 담긴 내용은 창작물이다. 필자는 모델링을 하면서 창작의 기쁨을 얻는다. 필자의 느낌에 동의하는 모델러가 많을 것으로 생각한다.

모델링은 일하는 자체도 매력적이다. 인터뷰와 분석을 반복하면서 모델의 구조가 형성된다. 그리고 모델 구조를 관련자에게 설명하고 논의한다. 이 과정에서 많은 업무 지식을 습득한다. 필자는 업무에는 관심이 없는 편이지만 업무 지식을 습득하는 것은 커다란 장점이다. 이런 과정을 거치면서 많은 사람을 만난다. 새로운 사람을 만나는 것을 꺼리는 사람이 모델러가 되면 약간의 고역이 될 것이다. 최근에는 덜 하지만 예전에는 모델러를 우대해 주는 분위기가 있었다. 모델러로서 의견이 존중받는다면 이것도 일하는 과정 중에 얻게 되는 기분 좋은 일이다.

아직도 그렇지만 예전에는 데이터 모델이나 데이터베이스를 블랙박스 정도로 생각하는 경향이 컸다. 쉽게 얘기해서 아무렇게나 집어넣고 아무렇게나 꺼내 와도 사용만 하면 된다는 식이었다. 데이터에 대한 중요성은 20년 전에도 다 같이 인식하는 것이었지만 실무에서는 이렇게 막 대했던 것이 데이터였다.

하지만 지금은 그렇지 않다. 데이터의 실체를 보려는 노력을 많이 한다. 무엇보다 데이터 품질이 매우 중요해져서 누구라도 데이터에 관심을 가진다. 이는 데이터 표준화에 대한 구축으로 이어졌으며 자연스럽게 데이터 모델링과도 연결된다. 더욱이 DA(Data Architect)에 대한 수요가 급격히 늘어났다. 많은 프로젝트에서 DA를 채용하고 있으며, 이는 앞으로 더욱 늘어날 것이다. EA(Enterprise Architecture)나 DQ(Data Quality), 마스터 데이터, ISP(Information Strategic Planning), PI(Process Innovation) 등과 같은 프로젝트에서도 데이터 전문가(Data Professional)가 필요하다. 데이터 전문가가 되기 위한 가장 중요한 관문이 데이터 모델이다. 모델링을 할 수 있다면 데이터 관련 다양한 분야에서 활약할 수 있다.

최근 시스템 구축 프로젝트에서 데이터 모델의 가치가 더욱 커지고 있다. 데이터 모델이 시스템의 근간이 된다는 점을 인식한 것은 20년이 넘었는데 그동안은 알면서도 데이터 모델의 중요성을 간과했다. 하지만 이제는 알면서 그냥 넘어갈 수 없을 정도로 많은 사람이 데이터 모델의 중요성을 인식하고 있다. 데이터 모델은 시스템에서 가장 중요한 토대이므로 시스템을 좌지우지한

다. 그렇기에 중압감은 있지만 도전할 만한 매력적인 분야다.

이론이라는 근거에 통찰력과 사고력, 경험을 바탕으로 한 데이터 모델이 시스템 구축 프로젝트의 중심을 잡아준다면, 프로젝트가 근간이 흔들리는 일은 없을 것이다. 데이터 모델링은 유망하고 매력적인 분야다.

1.3. 모델링은 상식적이다

데이터 모델링은 상식적이다. 만약에 모델링이 상식적이지 않은 분야였다면 필자는 모델러로서 지금까지 일할 수 없었을 것이다.

이 책에서 설명하는 모델링과 관련된 내용은 크게 두 가지로 나눌 수 있다. 하나는 이론이고 다른 하나는 경험이다. 그중에 모델링 이론은 지극히 상식적이다. 모델링 이론이 다른 분야와 비교해서 어렵지 않다고 느껴지는 것은 상식적이기 때문이다. 하지만 이론은 상식적이지만 실제로 모델링을 하면 어렵게 느껴진다. 상식적으로 알 수 있는 쉬운 이론으로 모델링을 수행해도 어려운 것이 모델링인데 이론을 모른다면 더욱 막막할 것이다.

상식적이라는 판단은 다소 개인적일 수밖에 없다. 필자가 상식적이라고 느껴지는 부분을 다른 사람은 동의하지 않을 수도 있다. 개인마다 기준이 다를 수 있지만, 정규화를 기초로 한 이론은 상식적이라고 생각한다. 데이터 무결성을 최우선 과제로 삼는 데이터 모델링의 이론은 지극히 상식적이다. 데이터를 제대로 관리하는 데 필요하므로 상식적일 수밖에 없다.

반면에 경험을 통해 터득하는 부분이 존재한다. 엔터티를 구성하는 데이터의 정체성에 대한 판단이나 데이터를 통합하는 기준은 개인마다 다를 수 있다. 그리고 많은 경험에 의해 단련된다.

정규화는 처음 경험하는 초보자든 경험이 많은 전문가든 이론을 정확히 알고 있다면 같은 결과가 나온다. 하지만 데이터 통합과 관련해서는 전문가 사이에서도 이견이 존재할 수 있다. 물론 전문가가 구축한 모델은 상세한 면에서 조금 다를 뿐 크게 보면 대단히 유사하다고 생각한다. 이는 경험에 의해 자의적으로 모델링이 이루어지는 것 같지만, 이론이라는 상식을 기반으로 모델링이 이루어지기 때문이다. 자신이 사용해 본 모델만을 고수하면 근거가 없으며 상식적이지 않은 것이다. 경험이 상식이 되지 못하면 고집에 지나지 않는다.

데이터 모델링은 상식적인 이론이 토대가 돼야 하며 상식을 기반으로 한 경험이 축적돼야 좋은 모델을 만들 수 있다. 상식이라는 근거가 없는 모델은 누구도 확신시킬 수 없으며 혼란을 일으킬

뿐이다. 이론을 무시하면 또한 상식적이지 않은 것이다. 이론을 바탕으로 통찰력과 분석력, 경험이 조화를 이룬다면 데이터 모델링은 어렵지 않으며 상식적인 것이 된다.

1.4. 모델러와 바둑 프로기사

필자는 바둑을 좋아한다. 바둑을 두고 있으면 시간 가는 줄을 모른다. 두어 시간 동안 아무 생각 없이 집중할 수 있는 시간을 좋아한다. 사람은 극도로 집중할 수 있는 일에 빠진다.

모델링은 바둑과 유사한 점이 많다. 모델링을 바둑에 비유하는 모델러는 필자만이 아닐 것이다. 네 귀퉁이에서 정석을 바탕으로 한 판의 바둑이 펼쳐진다. 이는 중요 엔터티를 기준으로 주변의 관계 엔터티로 확장해 나가면서 모델링을 수행하는 것과 유사하다. 필자는 모델링 툴의 빈 여백이 바둑판으로 생각될 때가 많다. 첫수를 둘 때의 느낌이 첫 엔터티를 그릴 때의 느낌과 유사하다. 판이 어떻게 짜질지를 기대하면서 첫수를 둔다. 한 수씩 진행하면서 흥미는 증폭된다.

바둑은 정석을 모르면 제대로 두기 어렵다. 모델링도 엔터티와 관계, 속성과 관련된 다양한 이론이 존재하며, 때로는 기계적으로 적용되기도 하는 정석과 같은 이론을 모르고서는 제대로 수행하기 어렵다. 초반 정석을 잘못 선택하거나 빼놓으면 판 전체가 엉망이 되기도 하듯이 처음에 선택한 핵심적인 엔터티를 제대로 정의하지 않으면 모델링이 전반적으로 힘들어진다.

정석, 포석, 행마, 끝내기 등의 중요한 요소를 열심히 공부해도 프로 바둑 기사가 되기 어렵다. 끊임없는 연구와 실전 대국, 자신만의 스타일, 감각, 집중력, 체력 등이 있어야 프로 기사가 될 수 있다. 모델링도 마찬가지로 이론은 쉽게 익힐 수 있지만 직접 모델링을 하기는 쉽지 않다. 엔터티, 속성, 관계, 정규화 등의 이론은 기본이며 다양한 상황에 따른 최선의 모델을 판단할 수 있어야 하며 분석력과 논리력, 종합적인 사고능력, 설득력, 섬세함 등이 있어야 전문 모델러가 될 수 있다.

바둑에서 수 읽기가 중요하듯이 모델링을 수행하면서 데이터가 생성되는 것을 읽어야 한다. 사례 데이터는 머릿속에서도 발생시킬 수 있지만, 확인을 위해서 직접 작성해보는 것이 좋다. 타인을 위해서라도 만들어 보는 것이 좋다.

바둑은 기력이 조금 있더라도 프로 기사가 둔 수에 대해서 명확하게 이해할 수 있다. 상식적이고 자연스러운 위치여서 당연하게 생각된다. 하지만 프로가 둔 수를 스스로 두기란 결코 쉬운 게 아니다. 몇 번은 프로처럼 멋진 수를 둘 수 있지만 항상 그렇게 두지는 못한다. 데이터 모델도 마찬

가지로 전문 모델러가 구축한 모델을 보거나 설명을 들으면 이해가 되고 그런 구조가 맞는 것 같다. 하지만 실제로 그런 모델을 구축하기란 쉽지 않다. 매번 그렇게 하기는 더욱 어렵다

많은 점에서 바둑과 모델링은 닮았다. 모델러와 바둑 기사의 성향도 닮았다고 생각한다. 이론적인 지식은 기본이며 실전 경험과 연구를 통해 더 좋은 모델이 될 수 있도록 노력해야 한다. 프로 바둑 기사는 서문에서 소개한 1만 시간을 채우고도 남을 정도로 바둑에 몰두했을 것이다. 프로 모델러가 돼 더 좋은 데이터 모델을 구축하고 국가 경쟁력을 높이는 데 이바지할 수 있도록 노력해야 한다.

1.5. 좋은 모델은?

필자가 생각하는 좋은 모델은 단순하고 명확한 모델이다. 명확한 모델은 명확한 정의에서 비롯된다. 특히 엔터티에 대한 명확한 정의가 명확한 모델의 출발이다. 애매한 정의는 항상 오용이 따라다니게 된다.

경험상 단순한 모델이 좋은 모델이 될 가능성이 크다. 세상 이치도 마찬가지라 생각한다. 어떤 것을 단순하게 만드는 능력은 참으로 대단하다. 복잡한 것을 단순하게 만드는 것은 위대하며, 단순한 것을 단순하게 만드는 것도 대단하다. 단순한 것을 복잡하게 만들지 말아야 한다. 모르기 때문에 복잡하게 만드는 것이다. 데이터 모델링은 상식적이라 단순하다고 생각한다.

단순하고 명확한 모델은 이론을 기반으로 해서 나온다. 데이터를 보고 데이터의 성격을 한눈에 판단할 수 있는 감각도 이론이 바탕이 되었을 때 생길 수 있다. 이론을 기반으로 많은 경험이 더 해지면 감각적인 판단이 생길 수 있다. 이론을 원칙으로 적절한 판단이 적재적소에 가미된 상식적인 모델이 좋은 모델이다.

그리고 데이터 무결성(Integrity)이 보장되는 모델이 좋은 모델이다. 무결성은 이 책 전반에 걸쳐 강조하는 핵심 요소이다. 무결성이 훼손된 데이터는 아무리 비즈니스 요구 사항을 충족한다 해도, 아무리 빠른 성능이 보장된다 해도 좋은 모델이 될 수 없다(물론 성능만 빠르면 최고인 때도 존재한다). 데이터 모델링의 최우선 목표는 데이터 무결성을 보장하도록 데이터 모델을 구축하는 것이다. 그리고 데이터 무결성을 보장하는 가장 기본적인 방법은 중복을 제거하는 것이다. 중복(추출) 속성과 중복 엔터티를 제거하는 것이 데이터 무결성을 보장하는 핵심이다.

데이터 무결성 다음으로 고려해야 하는 부분이 성능이다. 성능이 좋은 모델은 분명히 존재한다.

대개 원칙을 지켰을 때 성능은 좋아지지만 성능만을 고려해 모델을 구현해야 할 때도 존재한다. 데이터 무결성을 다소 포기해야 하는 상황도 발생한다. 사용자나 고객은 빠른 결과를 원하므로 성능을 무시한 모델은 좋은 모델이 아니다.

좋은 모델은 또한 비즈니스를 효율적으로 표현한 모델이다. 모델에 비즈니스를 표현하는 것은 필수적인 기본 조건이다. 하지만 비즈니스를 표현하지 못하는, 비즈니스를 수행하는 데 필요한 데이터 요소들이 많이 결여된 모델이 존재한다. 조금만 신경 쓰면 누락된 데이터 요소는 없도록 할 수 있다. 어렵지 않은 부분이므로 최소한 비즈니스 표현이 누락된 모델은 되지 않도록 신경 써야 한다. 누락되는 데이터 요소도 문제지만 필요치 않은 비즈니스를 모델에 표현하는 것도 문제가 된다. 필요로 하는 요건만을 데이터로 관리해야지 해당 기업에서는 발생하지 않을 요건이나 10년 후에나 발생할 수 있는 요건을 위해 고민하고 논쟁할 필요는 없다.

또한 업무의 변화에 유연하게 대처할 수 있는 모델이 좋은 모델이다. 신규로 업무가 추가되거나 바뀌더라도 모델 구조가 변하지 않도록 설계해야 한다. 확장성이 좋은 유연한 모델을 구축할 수 있는지는 모델러의 주요한 능력이며 모델러에 따라서 차이가 많이 나는 부분이다.

그리고 좋은 모델은 기본적으로 모델 표현이 정확해야 한다. ERD(Entity Relationship Diagram)의 표기법(Notation)에 맞게 정확하게 표현돼야 좋은 모델이다. 표기법이 틀린 모델은 당연히 좋은 모델이 아니다. 데이터 모델은 커뮤니케이션을 잘 수행할 수 있도록 지원하는 도구이기도 하다. 좋은 모델은 이론적으로도 완전해야 하지만 가독성도 좋아야 한다. 엔터티명이 애매하거나 관계가 불분명하면 가독성이 나빠지며 좋은 모델이 될 수 없다. 앞서 언급했지만 표현이 복잡한 모델은 좋은 모델이 아니다. 모델을 보면서 대략적인 업무를 이해할 수 있도록 표현이 명확하고 단순한 모델이 좋은 모델이다.

필자가 생각하는 좋은 모델은 ‘사용자가 원하는 데이터를 정확하고 빠르게 보여줄 수 있는 단순한 모델’이라고 표현하고 싶다.

1.6. 모델링이 왜 필요한가?

데이터 모델링이 왜 필요할까? 당연히 좋은 모델을 만들기 위해서 필요하다. 11장(모델링 방법론)에서 자세하게 설명하겠지만 데이터 모델링에도 방법론이 존재한다. 개념, 논리, 물리 등의 모델링 단계 없이 데이터베이스에 스키마를 바로 생성할 수는 없다. 모델링 과정 없이 생성된 모델은

조악한 모델이 될 가능성이 매우 크다.

데이터 모델링은 데이터를 분석하는 과정을 지원하여 관련자의 커뮤니케이션을 돕는다. 모델링을 수행하는 도중에 이루어지는 커뮤니케이션은 완성된 모델로 이루어지는 커뮤니케이션보다 더욱 중요하다. 모델링 과정 중의 모델은 업무와 데이터를 분석하는 데 중요한 역할을 하므로 모델링 과정이 필수적으로 필요하다.

데이터 모델링은 사용자의 요구 사항을 반영하는 작업이며 관련자가 모델을 이해할 수 있도록 하는 작업이다. 최종적으로 데이터베이스에 구현되기까지의 과정을 원활하게 지원하며, 애플리케이션 개발 중에 모델의 변경을 최소화할 수 있도록 하는 과정이 모델링이다. 모델링 과정을 거치지 않으면 좋은 품질의 모델이 나오기 어렵다.

또한 데이터를 정확히 관리하고 사용자에게 빠르게 제공하기 위해서 모델링은 필요하다. 데이터 모델링은 단지 애플리케이션과 관련된 것은 아니다. 흔히 모델러는 프로그래머가 개발을 잘하도록 지원하는 역할을 하며, 프로그램이 빈번하게 수정되지 않도록 견고한 모델을 만들어 빠른 개발을 지원하는 역할을 한다고 말하지만, 데이터 모델은 애플리케이션보다는 데이터와 더욱 관련된 것이다. 데이터가 잘 관리될 수 있도록 지원해 주는 것이 데이터 모델링이다. 오늘날 기업은 데이터를 효율적으로 관리하는 것에 성공이 달렸다.

내부, 외부 고객인 사용자의 불만은 틀린 데이터며 느리게 제공되는 정보(Information)이다. 치명적이지는 않지만 약간씩 틀린 데이터는 불신을 일으킨다. 나쁜 구조의 모델에서 특정 데이터를 요구하면 어렵게 가공해야 하므로 결과가 늦게 나온다. 조회가 늦어지는 것도 사용자의 주요한 불만이다. 업무가 수정되거나 추가될 때 구현이 느리다는 것 또한 불만이다.

데이터와 관련된 사용자의 불만은 의외로 심각하다. 정확한 정보를 빠르게 제공하고 업무를 신속하게 추진할 수 있도록 하려면 좋은 모델이 필요하며, 좋은 모델을 얻으려면 절차에 따른 데이터 모델링이 필요하다.

데이터를 정확히 관리하고 빠르게 제공하며 확장하기 쉬운 모델을 구축하면 같은 인력으로 많은 일을 할 수 있다. 이것이야말로 경영 혁신의 지름길이다.

현실적으로도 프로젝트에서 데이터 모델은 중요한 역할을 한다. 때로는 업무 분석/설계자에게 분석 데이터로 제공되기도 하며 관련자와 커뮤니케이션 도구로 사용되기도 한다. 때에 따라서 데이터 모델링이 프로젝트의 분석/설계 단계를 주도적으로 이끌고 가므로 모델링이 현실적으로 필수적인 부분이 된다. 프로젝트의 규모가 클수록 시스템의 터를 잡고 뼈대를 세울 수 있도록 모

RELATIONAL DATA MODELING PREMIUM GUIDE

좋은 모델은 사용자가 원하는 데이터를
정확하고 빠르게 보여줄 수 있는 단순한 모델이다.

업무요건 Business Requirements

모델링은 업무에서 필요로 하는 데이터를 설계하는 작업이다. 좋은 모델은 당연히 업무요건을 만족해야 한다. 업무가 기준이 되지 않은 모델은 좋은 모델이 아니다. 업무에서 요구하는 데이터를 분석해서 엔티티를 명확하게 정의하는 것이 정규화이므로 업무를 모델에 표현하려면 정규화 과정을 거쳐야 한다.

데이터 무결성 Data Integrity

데이터 무결성을 높이는 것이 모델링의 최우선 목표다. 데이터 무결성은 데이터에 결점이 없는 상태를 의미한다. 무결성이 훼손된 데이터는 아무리 업무 요구 사항을 충족한다 해도, 아무리 빠른 성능이 보장된다 해도 좋은 모델이 될 수 없다. 데이터 무결성을 보장하는 가장 기본적인 방법은 중복을 제거하고 관계를 관리하는 것이다. 더는 삭제할 것이 없는 모델이 좋은 모델이다. 데이터를 통합할수록 중복은 제거된다.

성능 Performance

데이터 무결성 다음으로 고려해야 하는 부분이 성능이다. 데이터 모델링은 단순히 업무를 개념적으로 표현하는 것에 그치는 게 아니라 DBMS에 구현됐을 때 문제가 발생하지 않도록 하는 것이다. 심각한 성능 이슈가 발생하지 않도록 하는 것 또한 모델링의 목표다. 데이터 무결성과 데이터베이스 성능을 높이는 것이 모델링의 양대 목표나 마찬가지다.

유연성 Flexibility

업무 변화에 유연하게 대처할 수 있는 모델이 좋은 모델이다. 신규로 업무가 추가되거나 바뀌더라도 모델 구조가 변하지 않도록 설계해야 한다. 확장성 높고 유연한 데이터 모델 도출은 모델러의 핵심 직무다. 그럼에도 모델러에 따라 모델링 결과물의 품질 차이가 심한 것이 현실이다. 데이터 자체를 기준으로 판단해 유사한 데이터라면 통합된 모델이 확장성이 좋은 유연한 모델이다. 정규화와 일반화를 할수록 확장하기가 좋은 유연한 모델이 된다.

정규화 Normalization 를 정확히 수행할 수 있고
데이터를 통합 Generalization 할 수 있는 능력만 있다면 좋은 모델을
만들 수 있을 것이다.



9 791196 319229
ISBN 979-11-963192-2-9

값 37,000 원
90560