

NoSQL 탄생 배경·
다양한 유형의 NoSQL 비교 분석·
효과적으로 설계하는 노하우까지
사례 연구로 실감나게 설명한다

PERFECT
INTRODUCTION
TO NoSQL

NoSQL 철저 입문

NoSQL 세계에 이해하고 첫걸음 NoSQL에 선택하자

댄 설리번 지음
홍형경 옮김

NOSQL FOR MERE MORTALS

Authorized translation from the English language edition, entitled NOSQL FOR MERE MORTALS, 1st Edition by DAN SULLIVAN, published by Pearson Education, Inc, publishing as Addison-Wesley Professional, Copyright © 2015 NOSQL FOR MERE MORTALS

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc.

KOREAN language edition published by GILBUT PUBLISHING CO., LTD., Copyright © 2015

이 책의 한국어판 저작권은 대니홍 에이전시를 통한 저작권사와의 독점 계약으로 (주)도서출판 길벗에 있습니다.

신저작권법에 의해 한국 내에서 보호를 받는 저작물이므로 무단전재와 복제를 금합니다.

NoSQL 철저 입문

초판 발행 · 2015년 12월 28일

지은이 · 댄 설리번

옮긴이 · 홍형경

발행인 · 이종원

발행처 · (주)도서출판 길벗

출판사 등록일 · 1990년 12월 24일

주소 · 서울시 마포구 월드컵로 10길 56(서교동)

대표 전화 · 02)332-0931 | 팩스 · 02)323-0586

홈페이지 · www.gilbut.co.kr | 이메일 · gilbut@gilbut.co.kr

기획 및 책임편집 · 서정민(jayseo@gilbut.co.kr) | 디자인 · 배진웅 | 제작 · 이준호, 손일수, 이진혁

영업마케팅 · 박성용, 전선하, 장봉석, 지운집 | 영업관리 · 김명자 | 독자지원 · 송혜란, 정은주

전산편집 · 박진희 | 출력 및 인쇄 · (주)이펙트앤피 | 제본 · 경문제책

- ▶ 잘못된 책은 구입한 서점에서 바꿔 드립니다.
- ▶ 이 책에 실린 모든 내용, 디자인, 이미지, 편집 구성의 저작권은 (주)도서출판 길벗과 지은이에게 있습니다.
허락 없이 복제하거나 다른 매체에 옮겨 실을 수 없습니다.

ISBN 979-11-86978-36-8 93560

(길벗 도서번호 006765)

정가 36,000원

독자의 1초를 아껴주는 정성 길벗출판사

(주)도서출판 길벗 | IT실용, IT전문서, IT/일반수험서, 경제경영, 취미실용, 인문교양(더퀘스트) www.gilbut.co.kr

길벗이지톡 | 어학단행본, 어학수험서 www.eztok.co.kr

길벗스쿨 | 국어학습, 수학학습, 어린이교양, 주니어 어학학습, 교과서 www.gilbutschool.co.kr

페이스북 · www.facebook.com/gilbutzigy | 트위터 · www.twitter.com/gilbutzigy

지은이의 말 003

옮긴이의 말 004

감사의 글 005

들어가기 006

1부 NoSQL 데이터베이스 개요

1장 요구 사항에 따른 데이터베이스의 종류 027

1.1 관계형 데이터베이스 설계 029

전자 상거래 애플리케이션 029

1.2 초기 데이터베이스 관리 시스템 030

플랫 파일 데이터 관리 시스템 030

계층형 데이터 모델 시스템 035

네트워크 데이터 관리 시스템 037

초기 데이터 관리 시스템 요약 039

1.3 관계형 데이터베이스의 혁명 040

관계형 데이터베이스 관리 시스템 041

1.4 NoSQL 데이터베이스의 출현 배경 048

확장성 049

비용 050

유연성 050

가용성 051

1.5 요약 052

1.6 사례 연구 053

1.7 참고자료 054

1.8 참고문헌 054

2장 다양한 NoSQL 데이터베이스 057

2.1 분산 데이터베이스에서의 데이터 관리 060

영구적인 데이터 저장 060

데이터 일관성 유지 061

데이터 가용성 확보 062

응답 시간·일관성·지속성 간의 균형 맞추기 066

CAP 이론: 일관성, 가용성, 그리고 파티셔닝 068

2.2 ACID와 BASE 071

ACID: 원자성, 일관성, 고립성, 지속성 071

BASE: 기본적인 가용성, 소프트 상태, 결과적 일관성 072

결과적 일관성의 유형 073

2.3 NoSQL 데이터베이스의 네 가지 유형 075

키-값 데이터베이스 075

문서 데이터베이스 080

컬럼 패밀리 데이터베이스 083

그래프 데이터베이스 085

2.4 요약 088

2.5 참고자료 089

2.6 참고문헌 090

연습문제 091

2부 키-값 데이터베이스

3장 키-값 데이터베이스 소개 093

3.1 배열에서 키-값 데이터베이스까지 095

배열: 보조 바퀴를 단 키-값 저장 구조 095

연관 배열: 보조 바퀴를 떼어버리자 097

캐시: 자전거에 기어를 달자 097

메모리와 디스크 키-값 데이터베이스: 자전거에서 오토바이로 100

3.2 키-값 데이터베이스의 핵심 특징 102

단순성: 복잡한 데이터 모델이 필요한 사람은 없다 102

속도: 빠를수록 좋다 103

확장성: 빠른 속도를 유지하라 105

3.3 키: 의미 없는 값을 가진 식별자, 그 이상인 존재 112

키 생성 방법 112

키를 사용해 값을 가져오기 114

3.4 값: 원하면 어떤 데이터든지 저장한다 118

값은 엄격한 데이터 타입을 요구하지 않는다 118

값 검색의 한계 119

3.5 요약 121

3.6 참고자료 122

3.7 참고문헌 123

연습문제 124

4장 키-값 데이터베이스 용어 정리 125

4.1 키-값 데이터베이스 데이터 모델링 용어 127

키 129

값 131

이름공간 131

파티션 133

파티션 키 135

스키마리스(스키마가 없는 데이터구조) 135

4.2 키-값 아키텍처 용어 137

클러스터 137

링 139

복제 140

4.3 키-값 데이터베이스 구현에 관한 용어 141

해시 함수 142

해시 충돌 143

압축 144

4.4 요약 145

4.5 참고자료 146

연습문제 147

5장 키-값 데이터베이스 설계 149

5.1 키 설계와 파티셔닝 151

키는 명명규칙을 따라야 한다 151

잘 설계된 키는 코드의 양을 줄여준다 152

값의 범위 처리 153

구현 상 제약을 고려해 키를 만들어야 한다 154

파티셔닝에서 키를 어떻게 사용할까? 155

5.2 구조화된 값 설계 156

구조화된 데이터 유형은 대기 시간을 줄이는 데 도움이 된다 157

값이 크면 읽기와 쓰기 연산이 비효율적일 수 있다 160

5.3 키-값 데이터베이스의 한계 163

키만으로 값 조회하기 163

범위 질의를 지원하지 않는 키-값 데이터베이스 164

관계형 데이터베이스의 SQL과 필적할 만한 표준 질의 언어의 부재 164

5.4 키-값 데이터베이스를 위한 설계 패턴 165

- TTL 키 166
- 테이블 모방 167
- 집계 169
- 원자적 집계 171
- 열거형 키 172
- 인덱스 172

5.5 요약 174

5.6 사례 연구: 모바일 애플리케이션을 위한 키-값 데이터베이스 구성 175

5.7 참고자료 178

연습문제 179

3부 문서 데이터베이스

6장 문서 데이터베이스 소개 181

6.1 문서란 무엇인가 183

- 문서란 그리 간단한 것이 아니다 183
- 문서와 키-값 쌍 186
- 컬렉션에서 여러 문서 관리하기 187

6.2 명시적으로 스키마를 정의하지 말자 196

6.3 문서 데이터베이스의 기본 연산 작업 198

- 컬렉션에 문서 삽입하기 199
- 컬렉션에 있는 문서 삭제하기 201
- 컬렉션에 있는 문서 갱신하기 203
- 컬렉션에서 문서 조회하기 204

6.4 요약 205

6.5 참고자료 206

연습문제 207

7장 문서 데이터베이스 용어 정리 209

7.1 문서와 컬렉션 용어 211

문서 211

컬렉션 213

중첩 문서 214

스키마리스 216

다형적 스키마 219

7.2 파티션 유형 220

수직 파티셔닝 221

수평 파티셔닝과 샤딩 222

7.3 데이터 모델링과 질의 처리 226

정규화 226

비정규화 228

질의 프로세서 229

7.4 요약 230

7.5 참고자료 230

연습문제 232

8장 문서 데이터베이스 설계 233

8.1 정규화와 비정규화, 둘 간에 균형을 맞춘 검색 235

일대다 관계 236

다대다 관계 236

조인의 필요성 237

조인: 관계형 데이터베이스가 짙어진 무거운 짐 238

문서 데이터베이스 모델러는 무엇을 해야 할까? 241

8.2 가변적인 문서 처리 계획 246

너무 커진 문서의 이동 회피 248

8.3 인덱스의 골디락스 존 249

읽기 부하가 큰 애플리케이션 249

쓰기 부하가 큰 애플리케이션 250

8.4 공통 관계 모델링 252

문서 데이터베이스에서 일대다 관계 252

문서 데이터베이스에서 다대다 관계 253

문서 데이터베이스의 계층 관계 모델링 255

8.5 요약 257

8.6 사례 연구: 고객 적하 목록 258

중첩할 것인가? 중첩하지 않을 것인가? 259

인덱스 선택 260

유형별로 컬렉션을 분리해야 할까? 261

8.6 참고자료 261

연습문제 263

4부 컬럼 패밀리 데이터베이스

9장 컬럼 패밀리 데이터베이스 소개 265

9.1 태초에 구글의 빅테이블이 있었다 267

- 컬럼을 동적으로 제어하기 268
- 로우, 컬럼 이름, 타임스탬프에 의한 인덱싱 269
- 데이터의 저장 위치 제어 269
- 원자적으로 처리되는 로우의 읽기와 쓰기 271
- 정렬 순서대로 로우 관리하기 272

9.2 키-값 데이터베이스와 문서 데이터베이스의 차이점과 유사점 273

- 컬럼 패밀리 데이터베이스의 특징 273
- 컬럼 패밀리 데이터베이스와 문서 데이터베이스 간의 유사점과 차이점 274
- 컬럼 패밀리 데이터베이스 대 관계형 데이터베이스 276

9.3 컬럼 패밀리 데이터베이스에서 사용되는 아키텍처 280

- HBase 아키텍처: 다양한 유형의 노드 280
- 카산드라 아키텍처: P2P(Peer-to-Peer) 281
- 용어 살펴보기: 가십 프로토콜 283
- 열역학과 분산 데이터베이스: 안티-엔트로피가 필요한 이유 285
- 이것 좀 들고 있어요: 힌트 핸드오프 286

9.4 어떤 경우에 컬럼 패밀리 데이터베이스를 사용하면 좋은가? 289

9.5 요약 290

9.6 참고자료 290

10장 컬럼 패밀리 데이터베이스 용어 정리 293

10.1 컬럼 패밀리 데이터베이스의 기본 구성 요소 295

키스페이스 295

로우 키 296

컬럼 296

컬럼 패밀리 298

10.2 구조와 프로세스: 컬럼 패밀리 데이터베이스 구현 299

컬럼 패밀리 데이터베이스의 내부 데이터 구조와 구성 매개변수 299

옛 친구: 클러스터와 파티션 300

내부 살펴보기: 또 다른 컬럼 패밀리 데이터베이스 구성 요소 302

10.3 프로세스와 프로토콜 307

복제 307

안티-엔트로피 307

가십 프로토콜 309

힌트 핸드오프 310

10.4 요약 311

10.5 참고자료 311

연습문제 312

11장 컬럼 패밀리 데이터베이스 설계 313

11.1 테이블 설계 지침 316

조인 대신 비정규화 316

값이 없는 컬럼 사용하기 318

컬럼 이름과 컬럼 값을 사용해 데이터 저장하기 318

단일 로우를 이용한 엔터티 모델 319

로우 키에서 핫스팟 회피 320

적당한 수의 컬럼 값 버전 유지 321
컬럼 값에서 복잡한 데이터 구조 회피 322

11.2 인덱스 구성 지침 323

컬럼 패밀리 데이터베이스가 관리하는 보조 인덱스는 언제 사용하면 좋을까? 324
테이블을 이용한 보조 인덱스 생성과 관리 327

11.3 빅데이터를 다루기 위한 도구 330

빅데이터의 추출, 가공, 적재 331
빅데이터 분석 332
빅데이터 모니터링 도구 336

11.4 요약 337

11.5 사례 연구: 고객 데이터 분석 338

사용자 요구 사항 파악 338

11.6 참고자료 340

연습문제 341

5부 그래프 데이터베이스

12장 그래프 데이터베이스 소개 343

12.1 그래프란 무엇인가? 345

12.2 그래프와 네트워크 모델링 346

지리적 위치 모델링 346
전염병 모델링 347
추상적이고 구체적인 엔터티 모델링 349
소셜 미디어 모델링 351

12.3 그래프 데이터베이스의 장점 353

조인을 피하면 더 빠른 질의가 가능하다 353

단순화한 모델링 355

엔터티 간의 여러 관계 356

12.4 요약 357

12.5 참고자료 357

연습문제 358

13장 그래프 데이터베이스 용어 정리 359

13.1 그래프의 구성 요소 361

정점 361

에지 362

경로 364

루프 365

13.2 그래프에서의 연산 작업 366

그래프의 합집합 366

그래프의 교집합 367

그래프의 순회 367

13.3 그래프와 정점의 속성 368

동형성 369

순서와 크기 370

차수 370

근접성 371

매개성 371

13.4 그래프의 유형 372

무방향 그래프와 방향 그래프 372

유동 네트워크 373

이분 그래프 374

다중 그래프 375

가중 그래프 375

13.5 요약 376

13.6 참고자료 377

연습문제 378

14장 그래프 데이터베이스 설계 379

14.1 그래프 설계 시작하기 381

소셜 네트워크 그래프 데이터베이스의 설계 382

질을 통한 설계 385

14.2 그래프 질의 387

사이파: 선언에 의한 질의 387

그렘린: 그래프 순회에 의한 질의 389

14.3 그래프 데이터베이스 설계 팁과 함정 393

검색 시간 향상을 위한 인덱스 사용하기 393

적절한 유형의 예지 사용하기 394

그래프 순회 시 사이클 감시하기 394

그래프 데이터베이스의 확장성 고려하기 396

14.4 요약 397

14.5 사례 연구: 운송 경로 최적화 398

사용자 요구 사항 파악 398

그래프 분석 솔루션 설계 398

14.6 참고자료 400

연습문제 401

6부 애플리케이션에 맞는 데이터베이스 선택하기

15장 데이터베이스 선택하기 403

15.1 NoSQL 데이터베이스 선택하기 405

키-값 데이터베이스 선택 기준 406

문서 데이터베이스 선택 기준과 사용 사례 406

컬럼 패밀리 데이터베이스 선택 기준과 사용 사례 407

그래프 데이터베이스 선택 기준과 사용 사례 409

15.2 NoSQL 데이터베이스와 관계형 데이터베이스 함께 사용하기 410

15.3 요약 412

15.4 참고자료 413

연습문제 415

7부 부록

부록A 연습문제 정답 417

부록B NoSQL 데이터베이스 목록 449

부록C 용어 해설 455

찾아보기 465

효율적인 빅데이터 처리 시스템, NoSQL

유형별로 분석한 철저 가이드!

빅데이터를 효과적으로 처리하자!

SNS의 등장과 스마트폰의 대중화로 새로운 형태의 데이터가 수억 건씩 매일 쏟아진다. 분산 처리, 스키마리스,고가
용성, 고확장성 등 NoSQL 유형별 특성과 관계형 데이터베이스의 장점을 활용하면 빅데이터를 효과적으로 처리할 수
있다.

네 가지 유형으로 분석했다!

지금까지 나온 NoSQL 데이터베이스를 분석해 키-값 데이터베이스, 문서 데이터베이스, 컬럼 패밀리 데이터베이스,
그래프 데이터베이스로 나누어 설명한다. 유형별로 기본 개념과 구조, 특성을 소개하고 설계 시 노하우도 담았다. 또
한, 연습문제를 통해 각 개념을 철저하게 학습한다.

최적의 NoSQL을 선택하라!

HBase, 카산드라, 몽고DB, Neo4j, 레디스 등 유명한 모델은 많다. 하지만 나에게 꼭 맞는 모델은 어느 것일까? 어떻
게 하면 효과적으로 구현할 수 있을까? 사례 연구를 통해 모델 선택 요령을 실감나게 설명하고, 부록에 데이터베이스
목록과 참조 링크를 따로 정리하여 더 자세히 살펴볼 수 있다.

CONTENTS

1부 ... NoSQL 데이터베이스 개요

- 1장 요구 사항에 따른 데이터베이스의 종류
- 2장 다양한 NoSQL 데이터베이스

2부 ... 키-값 데이터베이스

- 3장 키-값 데이터베이스 소개
- 4장 키-값 데이터베이스 용어 정리
- 5장 키-값 데이터베이스 설계

3부 ... 문서 데이터베이스

- 6장 문서 데이터베이스 소개
- 7장 문서 데이터베이스 용어 정리
- 8장 문서 데이터베이스 설계

4부 ... 컬럼 패밀리 데이터베이스

- 9장 컬럼 패밀리 데이터베이스 소개

- 10장 컬럼 패밀리 데이터베이스 용어 정리

- 11장 컬럼 패밀리 데이터베이스 설계

5부 ... 그래프 데이터베이스

- 12장 그래프 데이터베이스 소개
- 13장 그래프 데이터베이스 용어 정리
- 14장 그래프 데이터베이스 설계

6부 ... 애플리케이션에 맞는 데이터베이스 선택하기

- 15장 데이터베이스 선택하기

7부 ... 부록

- 부록A 연습문제 정답
- 부록B NoSQL 데이터베이스 목록
- 부록C 용어 해설



NoSQL 철저 입문
PERFECT INTRODUCTION
TO NOSQL

정가 36,000원



9 791186 978368
ISBN 979-11-86978-36-8