

웹 개발자를 위한
**대규모 서비스를
지탱하는 기술**

“Web KAIHATSU-SHA NO TAMENO”

DAIKIBO-SWERVICE GIJUTSU-NYUMON

by Naoya Ito, Shinji Tanaka

Copyright © Naoya Ito, Shinji Tanaka 2010 All rights reserved.

Original Japanese edition published by Gijyutsu-Hyoron Co., Ltd., Tokyo

This Korean language edition published by arrangement with Gijutsu-Hyoron Co., Ltd., Tokyo in care of Tuttle-Mori Agency, Inc., Tokyo through Gaon Agency, Seoul.

Korean translation rights © 2011 by J-Pub

이 책의 한국어판 저작권은 터틀모리 에이전시와 가온 에이전시를 통한 Gijutsu-Hyoron Co., Ltd.와의 독점계약으로 제이펍에 있습니다. 신저작권법에 의하여 한국 내에서 보호를 받는 저작물이므로 무단전재와 무단복제를 금합니다.

웹 개발자를 위한 대규모 서비스를 지탱하는 기술

초판 1쇄 발행 2011년 2월 28일

지은이 이토 나오야, 다나카 신지

옮긴이 진명조 | 펴낸인 장성두 | 책임편집 안주연

본문디자인 북아이 | 표지디자인 미디어픽스

주소 경기도 파주시 교하읍 파주신도시 에이15-1블록 한빛마을 휴먼빌 201-502

전화 070-8201-9010 | 팩스 02-6280-0405

홈페이지 www.jpуб.kr | 펴낸곳 제이펍

출판신고 2009년 11월 10일 제406-2009-000087호

용지 신승지류유통 | 인쇄 해외정판사 | 제본 춘산제본

ISBN 978-89-94506-12-8 (13560)

값 25,000원

※ 이 책은 저작권법에 따라 보호를 받는 저작물이므로 무단전재와 무단복제를 금지하며, 이 책 내용의 전부 또는 일부를 이용하려면 반드시 저작권자와 제이펍의 서면동의를 받아야 합니다.

※ 잘못된 책은 구입하신 서점에서 바꾸어 드립니다.

제이펍은 독자 여러분의 책에 관한 아이디어와 원고 투고를 기다리고 있습니다. 책으로 펴내고자 하는 아이디어나 원고가 있으신 분은 책에 대한 간단한 개요와 차례, 구성과 저자 약력 등을 메일로 보내주세요. (보내실 곳: jeipub@gmail.com)

웹 개발자를 위한

대규모 서비스를 지탱하는 기술

데이터 구조, 메모리, OS, DB, 서버 / 인프라

1,000대의
시스템은
무엇이 다른가?

아마존 재팬
베스트셀러
★★★★★



이토 나오히, 다나카 신지 공저 / 진명조 옮김

Jpub
제이퍼블

드리는 말씀

- 이 책에 기재된 내용을 기반으로 한 운용결과에 대해 저자, 역자, 소프트웨어 개발자 및 제공자, 주식회사 기술평론사, 제이펍 출판사는 일체의 책임을 지지 않으므로 양해바랍니다.
- 이 책에 기재된 정보, 소프트웨어에 관한 언급 등은 이 책의 원고집필시점(2010년 4월) 및 하테나 여름 인턴 2009의 기술강의가 실시된 2009년 8월 시점의 내용을 게재하고 있습니다.
- 이 책에 등장하는 각 회사명, 제품명은 일반적으로 각 회사의 등록상표 또는 상표입니다. 본문 중에는 ™, ©, ® 마크 등이 표시되어 있지 않습니다.

차례

옮긴이 서문	XVI
지은이 서문	XVIII
이책의 구성	XX
하테나 소개	XXII

CHAPTER
01대규모 웹 서비스 개발 오리엔테이션
전체 그림 파악하기

강의 0 이 책의 근간 이 책에서 설명하는 것과 설명하지 않는 것	2
대규모 서비스 개발에 관련된 대학생 대상의 하테나 인턴십	2
이 책에서 설명하는 것	3
이 책에서 설명하지 않는 것	4
앞으로 대규모 서비스를 마주하게 될 여러분에게	5
강의 1 대규모 서비스와 소규모 서비스	6
하테나의 서비스 규모	6
하테나는 대규모, 구글 및 페이스북은 초대규모	8
소규모 서비스와 대규모 서비스의 차이	9
대규모 데이터량에 대한 대처	12
강의 2 계속 성장하는 서비스와 대규모화의 벽	14
웹 서비스의 어려움	14
하테나가 성장하기까지	15
시스템의 성장전략 — 미니멈 스타트, 변화를 내다본 관리와 설계	18

차 례

강의 3 서비스 개발의 현장	20
하테나의 기술팀 체제	20
하테나에서의 커뮤니케이션 방법	21
실제 서비스 개발	22
개발에 사용하는 툴	23
정리	26

CHAPTER
02

대규모 데이터 처리 입문

메모리와 디스크, 웹 애플리케이션과 부하

강의 4 하테나 북마크의 데이터 규모 데이터가 많을수록 처리에 시간이 걸린다	28
하테나 북마크를 예로 본 대규모 데이터	28
하테나 북마크의 데이터 규모	29
대규모 데이터로의 쿼리 — 대규모 데이터를 다루는 감각	30
강의 5 대규모 데이터 처리의 어려운 점 메모리와 디스크	32
대규모 데이터는 어떤 점이 어려운가? — 메모리 내에서 계산할 수 없다	32
메모리와 디스크의 속도차 — 메모리는 $10^8 \sim 10^9$ 배 이상 고속	33
디스크는 왜 늦을까? — 메모리와 디스크	34
OS 레벨에서의 연구	36
전송속도, 버스의 속도차	36
강의 6 규모조정 요소	41
규모조정, 확장성	41
규모조정 요소 — CPU 부하와 I/O 부하	42
웹 애플리케이션과 부하의 관계	42
DB 확장성 확보의 어려움	44

강의 7 대규모 데이터를 다루기 위한 기초지식	48
프로그래머를 위한 대규모 데이터 기초	48
대규모 데이터를 다루는 세 가지 급소 — 프로그램을 작성할 때의 요령	48
대규모 데이터를 다루기 전 3대 전제지식 — 프로그램 개발의 한층 아래 기초	50

CHAPTER
03

OS 캐시와 분산

대규모 데이터를 효율적으로 처리하는 원리

강의 8 OS의 캐시 구조	56
OS의 캐시 구조를 알고 애플리케이션 작성하기 — 페이지 캐시	56
가상 메모리 구조	58
Linux의 페이지 캐시 원리	60
VFS	62
Linux는 페이지 단위로 디스크를 캐싱한다	62
메모리가 비어 있으면 캐싱 — sar로 확인해보기	64
메모리를 늘려서 I/O 부하 줄이기	66
페이지 캐시는 투과적으로 작용한다	67
강의 9 I/O 부하를 줄이는 방법	72
캐시를 전제로 한 I/O 줄이는 방법	72
복수 서버로 확장시키기 — 캐시로 해결될 수 없는 규모일 경우	73
단순히 대수만 늘려서는 확장성을 확보할 수 없다	75
강의 10 국소성을 살리는 분산	79
국소성을 고려한 분산이란?	79
파티셔닝 — 국소성을 고려한 분산①	80
요청 패턴을 '섬'으로 분할 — 국소성을 고려한 분산②	83
페이지 캐시를 고려한 운용의 기본 규칙	85

CHAPTER
04

분산을 고려한 MySQL 운용

강의 11 인덱스를 올바르게 운용하기	분산을 고려한 MySQL 운용의 대전제	88
	분산을 고려한 MySQL 운용, 세 가지 포인트	88
	OS 캐시 활용	89
	인덱스의 중요성 — B트리	91
	인덱스의 효과	95
	인덱스가 작용하는지 확인하는 법 — explain 명령	98
강의 12 MySQL의 분산 확장을 전제로 한 시스템 설계		102
	MySQL의 레플리케이션 기능	102
	마스터/슬레이브의 특징 — 참조계열은 확장하고 갱신계열은 확장하지 않는다	104
강의 13 MySQL의 스케일아웃과 파티셔닝		107
	MySQL의 스케일아웃 전략	107
	파티셔닝(테이블 분할)에 관한 보충	107
	파티셔닝을 전제로 한 설계	108
	JOIN 배제 — where... in... 이용	110
	파티셔닝의 상반관계	112
	제2장~제4장 정리	116

CHAPTER
05

대규모 데이터 처리 실전 입문

애플리케이션 개발의 급소

강의 14 용도특화형 인덱싱	대규모 데이터를 능수능란하게 다루기	118
	인덱스와 시스템 구성 — RDBMS의 한계가 보일 때	118
	용도특화형 인덱싱 — 튜닝한 데이터 구조 사용하기	120
강의 15 이론과 실전 양쪽과의 싸움		125
	요구되는 기술적 요건 규명하기	125
	제2장~제5장 정리	127

CHAPTER
06**[과제] 압축 프로그래밍**

데이터 크기, I/O 고속화와의 관계 인식하기

강의 16 [과제] 정수 데이터를 컴팩트하게 가져가기	130
정수 데이터를 컴팩트하게 가져가기	130
출제의도 — 이 과제를 풀면 어떤 점이 좋을까?	130
과제에서 다루는 파일의 내부	132
강의 17 VB Code와 속도감각	133
VB Code — 정수 데이터를 컴팩트하게 저장하자	133
VB Code의 의사코드	135
정렬 완료된 정수를 'Gap'으로 가져가기	137
(보충❶) 압축의 기초	138
(보충❷) 대상이 정수인 경우 — 배경에 있는 이론	139
강의 18 과제에 대한 상세설명과 응답 사례	141
과제에 대한 상세설명	141
(참고❶) pack() 함수 — Perl 내부의 데이터 구조를 바이너리로 내보낸다	143
(참고❷) 바이너리의 read/write	146
(참고❸) 프로파일링	148
응답 사례와 사고방식	149

CHAPTER
07**알고리즘 실용화**

가까운 예로 보는 이론 · 연구의 실전 투입

강의 19 알고리즘과 평가	156
데이터 규모와 계산량 차이	156
알고리즘이란?	157
알고리즘을 배우는 의미 — 컴퓨터의 자원은 유한, 엔지니어의 공통언어	158

차 례

알고리즘의 평가 — Order 표기	159
티슈를 몇 번 접을 수 있을까? — $O(\log n)$ 과 $O(n)$ 의 차이	161
알고리즘과 데이터 구조 — 뿔레야 뿔 수 없는 관계?	163
알고리즘의 실제 활용 — 단순한 게 더 낫기도?	165
계산량과 상수항 — 역시 측정이 중요	164
써드파티 소스를 잘 활용하자 — CPAN 등	167
실제 사례를 보고 실전감각 익히기	169
강의 20 하테나 다이어리의 키워드 링크	170
키워드 링크란?	170
최초 구현방법	171
문제발생 — 키워드 사전의 대규모화	171
패턴매칭에 의한 키워드 링크의 문제점	172
정규표현 → Trie — 매칭 구현 변경	173
AC법 — Trie에 의한 매칭을 더욱 빠르게	175
Regexp::List로의 치환	177
키워드 링크 구현, 변이 및 고찰	178
강의 21 하테나 북마크의 기사 분류	179
기사 분류란?	179
기계학습과 대규모 데이터	180
대규모 데이터와 웹 서비스 — The google Way of Science	181
베이지안 필터의 원리	183
알고리즘이 실용화되기까지 — 하테나 북마크의 실제 사례	185
수비 자세, 공격 자세 — 기사 분류 구현으로부터의 고찰	187

CHAPTER
08**[과제] 하테나 키워드 링크 구현**
응용으로 가는 길 깨닫기

강의 22 [과제] 하테나 키워드 링크 만들기	192
AC법을 이용해서 하테나 키워드 링크 만들기	192
테스트 프로그램 작성	195
강의 23 응답 사례와 사고방식	198
응답 사례	198

CHAPTER
09**전문 검색기술 도전**
대규모 데이터 처리의 노하우

강의 24 전문 검색기술의 응용범위	202
하테나의 데이터로 검색엔진 만들기	202
하테나 다이어리의 전문 검색 — 검색 서비스 이외에 검색 시스템 이용	202
하테나 북마크의 전문 검색 — 세세한 요구를 만족시키는 시스템	205
강의 25 검색 시스템의 아키텍처	208
검색 시스템이 완성되기까지	208
다양한 검색엔진	210
전문 검색의 종류	212
강의 26 검색엔진의 내부구조	217
역 인덱스의 구조 — Dictionary+Postings	217
Dictionary 만드는 법 — 역 인덱스 작성법 #1	219
지금까지의 내용 정리	230
Postings 작성법 — 역 인덱스 작성법 #2	230
스코어링에 대한 보충	233
참고문헌	234

CHAPTER
10**[과제] 전문 검색엔진 작성**
기초, 상세부분 작성, 속도와 정확성 추구

강의 27 [과제] 하테나 북마크 전문 검색 만들기	238
전문 검색엔진 개발	238
과제내용	239
샘플 데이터 형식과 데이터 크기	240
사전의 구성 — Dictionary, Postings	241
인터페이스	241
기본적인 부분 + 심화 구현	242
속도와 정확성으로 승부	243
강의 28 응답 사례와 사고방식	244
응답 사례	244
indexer.pl 구현	245
searcher.pl 구현	247
개선할 수 있는 점은?	250

CHAPTER
11**대규모 데이터 처리를 지탱하는 서버/
인프라 입문 웹 서비스의 백엔드**

강의 29 엔터프라이즈 vs. 웹 서비스	254
엔터프라이즈 vs. 웹 서비스 — 응용범위로 보는 차이	254
웹 서비스의 인프라 — 중요시되는 세 가지 포인트	257
강의 30 클라우드 vs. 자체구축 인프라	259
클라우드 컴퓨팅	259
클라우드의 장단점	260
하테나에서의 클라우드 서비스 사용	261
자체구축 인프라의 장점	262

자체구축 인프라와 수직통합 모델	263
하테나의 서비스 규모	264
하테나 북마크의 시스템 구성도	265

CHAPTER
12

확장성 확보에 필요한 사고방식 규모 증대와 시스템 확장

강의 31 계층과 확장성	268
확장성에 대한 요구 — 서버 1대에서 처리할 수 있는 트래픽 한계	268
계층별 확장성	269
강의 32 부하 파악, 튜닝	271
부하 파악 — 가시화한 관리화면	271
부하를 측정하기 위한 항목 — Load Average, 메모리 관련, CPU 관련	273
용도에 맞는 튜닝 — 사용자용 서버, 봇용 서버	274
AP 서버/DB 서버의 튜닝 정책과 서버 대수	275
서비스 규모와 튜닝	276
확장성 확보	277

CHAPTER
13

다중성 확보, 시스템 안정화 100%에 근접한 가동률을 실현하는 원리

강의 33 다중성 확보	280
다중성 확보 — AP 서버	280
다중성 확보 — DB 서버	281
다중성 확보 — 스토리지 서버	284

차 례

강의 34 시스템 안정화	289
시스템 안정화를 위한 상반관계	289
시스템의 불안정 요인	290
강의 35 시스템 안정화 대책	296
실제 안정화 대책 — 적절한 버퍼 유지와 불안정 요인 제거	296

CHAPTER
14

효율향상전략

하드웨어의 리소스 사용률 높이기

강의 36 가상화 기술	300
가상화 기술의 도입	300
가상화 기술의 효용	301
가상화 서버 구축정책	302
가상화로 얻은 장점 정리	304
가상화와 운용 — 서버관리툴로 운용측면에서 가상화의 장점을 살리다 ...	305
가상화 도입 시 주의할 점	308
강의 37 하드웨어와 효율향상 저비용을 실현하는 요소기술	310
프로세서의 성능향상	310
메모리, HDD의 비용저하	311
저가 하드웨어의 유용한 이용법 — 가상화를 전제로 한 하드웨어 사용	313
SSD	315

CHAPTER
15

웹 서비스와 네트워크

서비스의 성장

강의 38 네트워크 분기점	320
서비스 성장과 네트워크 분기점	320

차 례

1Gbps의 한계 — PC 라우터의 한계	320
500호스트의 한계 — 1서브넷, ARP 테이블에서의 한계	321
네트워크 구조 계층화	323
글로벌화	324
강의 39 한층 높은 단계로	327
10Gbps 이상의 세계	327
하테나의 인프라 — 제11장~제15장 정리	328

APPENDIX

A

현대 웹 서비스 구축에 필요한 실전 기술 대규모 서비스에 대응하기 위해서

special 강의 1 작업큐(Job-Queue) 시스템 TheSchwartz, Gearman	332
special 강의 2 스토리지 선택 RDBMS와 key-value 스토어	337
special 강의 3 캐시 시스템 Squid, Varnish	350
special 강의 4 계산 클러스터 Hadoop	358

찾아보기 362

옮긴이 서문

“최근 안철수 교수님(카이스트 기술경영전문대학원)이 출연한 모 방송을 보면서 그가 말하는 ‘기업가(起業者) 정신’과 중소/벤처 기업의 역할에 대해 공감하고 다시 생각해볼 수 있는 시간이 있었다. 세상에 존재하지 않는 가치를 만들고 일자리를 창출하는 정신, 이를 실천하는 행동이 곧 기업가 정신이라는 것, 하지만 우리나라에서는 유독 기업가 정신이 사라지고 있고 성공한 중소/벤처 기업이 손꼽을 정도라는 등의 얘기를 들으면서, 문득 변변치 않은 서버 한 대에서 서비스를 시작해서 수백, 수천 대를 거쳐 이제는 수만 ~수백만 대의 서버에서 대규모 서비스를 운영하는 초대형 인터넷 기업이 된 Google과 Facebook에 대해 기업 환경적인 면에서 부러움을 느끼기도 하고, 한편으로는 기술적인 면에서 그들의 서비스 운용방식에 대해 호기심을 느끼기도 했다.”

이 책은 하테나 인턴십 강의 내용을 한 권의 책으로 정리한 것이다. ‘하테나’에 대해 생소하게 느끼는 독자들의 이해를 돕기 위해 본문 시작 전에 하테나 소개 페이지를 준비하였고, 강의 1에서는 하테나의 주요 서비스 및 서비스 규모에 대해서도 설명하고 있으니 참고하기 바란다.

하테나 인턴십의 전반부는 하테나의 엔지니어가 웹 기술 전반에 대한 내용을 강의하고, 후반부에는 실제 코드를 작성하면서 서비스 개발에 참여하게 되는 구성으로 되어 있다. 이 책은 그 중 전반부 강의 내용을 중심으로 해서 일부 실습 과제를 포함한 구성으로 되어 있다. 주요 강의 내용은 대규모 데이터를 다뤄야 하는 웹 서비스 개발자들이 알아두어야 할 메모리나 디스크, CPU와 I/O 부하, DB 분산 등의 서버 및 하드웨어에 대한 기초 지식과 대규모 데이터에 적합한 알고리즘, 데이터 구조, 인덱싱 등의 웹 서비스 개발 및 구현에 관련된 지식, 그리고 서비스 개발에만 초점이 가지 않도록 OS, 하드웨어, 네트워크

등 개발자가 의식하지 못할 수 있는 인프라 부분까지도 잘 담아내고 있다.

이러한 기술은 소프트웨어 엔지니어부터 인프라 엔지니어까지 웹 서비스와 관련된 모든 엔지니어가 알아두어야 하는 기술들이다. ‘자신이 작성한 코드가 어떤 시스템 상에서 동작하는지 또는 자신이 구성한 시스템 상에서 어떤 코드가 동작하는지’와 같은 문제의 기초를 이해해두는 것은 서비스 전체의 질을 향상시키기 위해 매우 중요하다. 이 책을 통해 독자 여러분은 하테나 정도의 중대규모의 서비스를 구축, 구현하기 위한 기초를 배울 수가 있다.

이와 같은 기본 지식 외에도 하테나의 서비스 개발 현장에서 엔지니어 간의 의사소통 방법과 활용도구에 대한 사용 이유, 목표도 소개하고 있어서 개발뿐 아니라 팀 운영면에서도 굉장히 도움이 될 만한 정보들이 가득 차 있다. 또한 앞으로 서비스 전체를 클라우드로의 이전 여부를 선택할 경우에 있을 수 있는 문제나 자체 서버를 사용하려는 입장에서의 고려사항 및 전망, 글로벌화를 위해 넘어야 할 장벽에 대한 설명도 중대규모 서비스를 운영하는 이들에게 도움이 될 것이다.

아무쪼록 이 책을 선택한 독자 여러분이 대규모 서비스에 대한 감을 익히고, 이를 지탱하는 각각의 요소 기술을 체득해서 안정적으로 중대규모 서비스를 개발 및 운용하고, 나아가 **Google**과 **Facebook** 같은 초대규모 서비스로 발전해가는 데 일조할 수 있는 기술참고서가 되었으면 하는 바람이다.

2011년 2월

진명조

지은이 서문

자신이 만든 웹 서비스의 규모가 앞으로 거대해지더라도 시스템은 괜찮을까? 이러한 불안감을 안고서 웹 서비스 개발에 종사하고 있는 분들도 있을 것이다. 또는 매일매일 시스템이 비명을 높여가고 있는데 어떻게 하면 이런 상황을 간파할 수 있을까? 성장한 웹 서비스를 앞에 두고 난감해하는 기술자도 있을 것이다. 필자도 동일한 경험을 해왔다.

월간 1,500만 명이 방문하는 하테나라고 하는 사이트. 이 대규모 시스템을 개발, 운영하는 일에 필자들은 몰두하고 있다. 1,000대의 호스트가 그 부하를 잘 처리해내고 있다. 100만 명 이상의 사용자가 블로그와 소셜 북마크에 계속해서 작성하고 있는 데이터는 매일 늘어가면서 서버 리소스를 핍박하고 있다. 기가바이트, 테라바이트 단위의 데이터량이 기술자들을 고민에 빠뜨린다. 그 래도 트래픽의 파도는 그칠 줄을 모른다.

일찍이 하테나가 아직 조직으로서도 미숙했던 시절, 대규모화하는 시스템을 앞에 두고 어쩔 줄을 몰라 했던 적도 있었다. 거대한 데이터로 대량의 트래픽이 밀어닥쳐 서버가 다운되고, 서비스는 중지, 황급히 심야에 달려가서 재기동. 어찌어찌 해서 안정되었다고 판단돼 새벽녘에 집으로 돌아가면 또다시 서버 다운. 이런 사태가 일상이었던 적도 있었다.

어떻게 하면 이 괴물 같은 대규모 서비스를 꿈쩍 못하게 할 수가 있을까? 수많은 시행착오를 거듭한 끝에 손에 넣은 기술과 노하우들이 이 책에 들어 있다.

이 책은 대규모 서비스를 개발, 운영하는 기술자를 위한 입문서다. 계속해서 성장하고 있는 웹 서비스가 간단히 처리할 수 없는 규모의 데이터를 떠맡게 되었을 때, 이것을 어떻게 요리하는 게 올바른 것일까? 자신이 작성한 코드가 시스템을 다운시키지 않기 위해서는 어떤 점에 주의해야 할까? 확장성을 고려해서 시스템을 설계하려면 어떤 것을 확보해두어야 할까? 이 책에서는 이러한 것

을 설명하고 있다.

하테나에서는 매년 여름에 학생을 대상으로 취업 체험을 목적으로 한 인턴십을 개최하고 있다. 이 인턴십에서는 학생들을 하테나의 실제 시스템 개발에 참여시키고 있다. 개발 경험이 적은 학생이더라도 사원과 동등하게 대우하면서 대규모 시스템 개발의 성공 체험을 가지고 돌아갈 수 있도록 하고 있다. 이것이 하테나 방식이다. 그 학생들이 개발에 앞서 알아두었으면 하는 것은 무엇일까? 우리가 먼 길을 돌고 돌아서 체득한 대규모 서비스 개발과 운용에 관한 지식, 바로 그것이였다.

인턴십 기획을 통해 하테나에서는 대규모 서비스 기술에 관한 교육 방법이 체계화되었다. 이 책에서는 이 인턴십에서 했던 강의를 기반으로 해서 대규모 서비스 기술에 대한 설명을 시도하고 있다. OS와 컴퓨터의 동작원리를 시작으로 DB 분산방법, 실전적 알고리즘을 시스템에 내장하는 원리, 대규모 데이터를 요리하는 검색엔진의 구조, 그리고 시스템 전체를 조망하기 위한 인프라 설계 지식 등 다방면에 걸쳐 있다.

실제로 1,500만 명에 달하는 사용자들이 이용하고 있는 하테나이기 때문에 전달할 수 있는 실전적이고 리얼한 기술과 현장감. 경험이 적은 학생을 불과 며칠 동안의 교육으로 대규모 서비스 개발현장으로 이끌어야 한다는 필요로부터 얻어진 지식의 체계화. 이것들을 섞고 융합함으로써 재미있고 질리지 않게 읽을 수 있고, 아울러 본질적인 지식을 얻을 수 있는 책이 될 수 있도록 힘썼다.

부디 이 책이 웹 서비스 개발에 종사하는 모든 엔지니어가 언제나 펼쳐 볼 수 있는 필수 도구가 되기를 바란다.

이토 나오야_(주)하테나 CTO

이 책의 구성

이 책은 하테나 인턴십 강의 내용을 기반으로 수록한 장과 새로 집필한 장으로 구성되어 있다. 상세한 구성과 집필담당자는 아래와 같다.

또한 이 책에서는 부하의 관점에서 OS 레벨의 기초 지식이 필요한 부분이 나온다. 이에 대한 보충 정보로, 『서버/인프라를 지탱하는 기술』(제이펍, 2009)의 4장 ‘성능향상, 튜닝’ 으로부터 일부 요약해서 이 책의 컬럼 코너에 수록했다(33쪽, 41쪽, 47쪽, 63쪽, 71쪽). 관심 있는 분은 꼭 참조하기 바란다.

장	내용	구분	집필담당
제1장	대규모 웹 서비스 개발 오리엔테이션 전체 그림 파악하기	새로 집필	
제2장	대규모 데이터 처리 입문 메모리와 디스크, 웹 애플리케이션과 부하	강의 기반	
제3장	OS 캐시와 분산 대규모 데이터를 효율적으로 처리하는 원리	강의 기반	
제4장	DB 스케일아웃 전략 분산을 고려한 MySQL 운용	강의 기반	
제5장	대규모 데이터 처리 실전 입문 애플리케이션 개발의 급소	강의 기반	(주)하테나 이토 나오야 (id:naoya)
제6장	[과제] 압축 프로그래밍 데이터 크기, I/O 고속화와의 관계 인식하기	강의 기반	
제7장	알고리즘 실용화 가까운 예로 보는 이론·연구의 실전 투입	새로 집필	
제8장	[과제] 하테나 키워드링크 구현 응용으로 가는 길 깨닫기	새로 집필	
제9장	전문 검색기술 도전 대규모 데이터 처리의 노하우	강의 기반	
제10장	[과제] 전문 검색엔진 작성 기초, 상세부분 작성, 속도와 정확성 추구	강의 기반	

이 책의 구성

제11장	대규모 데이터 처리를 지탱하는 서버/인프라 입문 웹 서비스의 백엔드	강의 기반	(주)하테나 다나카 신지 (id:stanaka)
제12장	확장성 확보에 필요한 사고방식 규모 증대와 시스템 확장	강의 기반	
제13장	다중성 확보, 시스템 안정화 100%에 근접한 가동률을 실현하는 원리	강의 기반	
제14장	효율향상전략 하드웨어의 리소스 사용을 높이기	강의 기반	
제15장	웹 서비스와 네트워크 서비스의 성장	강의 기반	
특별편	현대 웹 서비스 구축에 필요한 실전 기술 대규모 서비스에 대응하기 위해서	새로 집필	

● 도움(제9장, 제10장): (주)하테나 쿠라이 류타로(id:kurain)

※ 웹 지원 페이지 안내

이 책의 웹 지원 페이지에서는 샘플코드 다운로드를 제공하는 한편, 하테나 인턴십의 리포트 기사(최초공개 『WEB+DB PRESS』(Vol.53)의 'Special Report') 등 스페셜 콘텐츠를 게재하고 있다. 실제 강의 모습이나 커리큘럼, 엔지니어를 대상으로 한 강의 기간 중 하루 동안의 스케줄 등을 소개하고 있으므로 관심 있는 분은 꼭 참조하기 바란다.

- 스페셜 콘텐츠 : [URL http://gihyo.jp/magazine/wdpress/plus](http://gihyo.jp/magazine/wdpress/plus) (일어로 제공)
- 샘플 코드 다운로드 : [URL http://gihyo.jp/book/2010/978-4-7741-4307-1/support](http://gihyo.jp/book/2010/978-4-7741-4307-1/support)
- 이 책의 공식 태그 : hugedatabook(하테나 북마크, 트위터 등에서 사용 바랍니다.)
- Q & A : 제이펍 출판사(jeipub@gmail.com) 역자(chinmun@hanmail.net)

하테나 소개

하테나(hatena, はてな)는 ‘지식 커뮤니티’ 인력검색 서비스와 다이어리(블로그) 호스팅 서비스, 소셜 북마크 서비스 등을 개발해 운영하고 있는 일본의 인터넷 기업으로, 인터넷 서비스의 가장 매력적인 두 가지 요소인 ‘Fun’ 과 ‘Creativity’ 를 함께 추구하는 서비스, 생활 파트너로서 존재하는 서비스를 제공하고 있다.

● 주요 서비스

하테나 다이어리	광고 없는 심플한 블로그 서비스 웹 브라우저나 휴대전화로 자신만의 웹 다이어리(블로그) 작성 가능. 누군가에 의해서 등록된 키워드의 해설이 있기 내용 중에 있으면 자동적으로 링크되는 ‘키워드 링크 시스템’ 과 사용자가 CSS를 편집해 자신만의 디자인을 만들 수 있는 것이 특징이다.
하테나 하이쿠	키워드에 따라 언제든지 업데이트 가능한 미니 블로그
하테나 그룹	스케줄러, 게시판, 작업 관리, 파일 공유, 키워드 시스템 등의 기능을 갖춘 그룹웨어 서비스
우고메모 하테나	움직이는 메모를 통해 새로운 커뮤니케이션 체형 서비스
하테나 포토라이프	촬영한 사진을 공개할 수 있는 앨범 서비스
하테나 북마크	자신의 북마크를 다른 사용자와 공유하는 소셜 북마크 서비스
하테나 안테나	사용자가 등록된 웹사이트를 자동 방문하면서 사이트의 갱신 상황을 리스트로 표시해주는 서비스
하테나 키워드	단어(키워드)와 블로그를 연결하는 공유사전 서비스
인력검색 하테나	‘인력에 의한 검색’에 중점을 둔 지식 커뮤니티 서비스

● 하테나 서비스는 크게 두 가지 서비스로 나뉘어 있다.

대규모인 기본 서비스	중규모인 개별 서비스
<ul style="list-style-type: none"> • 알파벳 첫 글자가 서브도메인으로 덧붙여짐 • 많은 사용자들에게 기본 서비스가 되도록 계속 개발되고 있음 <p>☎ d.hatena.ne.jp h.hatena.ne.jp b.hatena.ne.jp a.hatena.ne.jp</p>	<ul style="list-style-type: none"> • 서브도메인이 복수 문자열로 되어 있음 • 기본 서비스의 보조적 역할, 용도가 한정된 서비스 <p>☎ graph.hatena.ne.jp search.hatena.ne.jp</p>

1.1 다중화의 기본