

Copyright © 2006, 2002 The McGraw-Hill Companies, S.r.l.  
Publishing Group Italia  
Via Ripamonti, 89 - 20139 Milano

**McGraw-Hill**



*A Division of The McGraw-Hill Companies*

I diritti di traduzione, riproduzione, memorizzazione elettronica e adattamento totale o parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche) sono riservati per tutti i Paesi.

Nomi e marchi citati nel testo sono generalmente depositati o registrati dalle rispettive case produttrici.

Editor: Chiara Tartara  
Produzione: Donatella Giuliani  
Realizzazione editoriale: Carmelo Giarratana  
Stampa: Rotolito Lombarda - Seggiano di Pioltello (MI)

Printed in Italy  
ISBN 88-386-6291-6  
1234567890GIMLIL09876

## Indice

### Prefazione *XI*

### Capitolo 1 Introduzione al data warehousing 1

- 1.1 I sistemi di supporto alle decisioni 2
- 1.2 Il data warehousing 3
- 1.3 Architetture per il data warehousing 6
  - 1.3.1 Architettura a un livello 7
  - 1.3.2 Architettura a due livelli 7
  - 1.3.3 Architettura a tre livelli 11
- 1.4 Gli strumenti ETL 13
  - 1.4.1 Estrazione 13
  - 1.4.2 Pulitura 15
  - 1.4.3 Trasformazione 15
  - 1.4.4 Caricamento 16
- 1.5 Il modello multidimensionale 17
  - 1.5.1 Restrizione 21
  - 1.5.2 Aggregazione 22
- 1.6 I meta-dati 22
- 1.7 Accedere al DW: reportistica e OLAP 25
  - 1.7.1 Reportistica 26
  - 1.7.2 OLAP 26
- 1.8 ROLAP e MOLAP 36
- 1.9 Altri aspetti 37
  - 1.9.1 La qualità 38
  - 1.9.2 La sicurezza 38
  - 1.9.3 L'evoluzione 39

**Capitolo 2 Il ciclo di vita dei sistemi di data warehousing 41**

- 2.1 Fattori di rischio 41
- 2.2 Top-down vs. bottom-up 42
  - 2.2.1 Il "Business Dimensional Lifecycle" 45
  - 2.2.2 La "Rapid Warehousing Methodology" 46
- 2.3 Le fasi di progettazione di un data mart 48
  - 2.3.1 Analisi e riconciliazione delle fonti dati 48
  - 2.3.2 Analisi dei requisiti 49
  - 2.3.3 Progettazione concettuale 51
  - 2.3.4 Raffinamento del carico di lavoro e validazione dello schema concettuale 51
  - 2.3.5 Progettazione logica 51
  - 2.3.6 Progettazione fisica 52
  - 2.3.7 Progettazione dell'alimentazione 52
- 2.4 Il quadro metodologico 52
  - 2.4.1 Scenario 1: approccio guidato dai dati 54
  - 2.4.2 Scenario 2: approccio guidato dai requisiti 55
  - 2.4.3 Scenario 3: approccio misto 55

**Capitolo 3 Analisi e riconciliazione delle fonti dati 57**

- 3.1 Ricognizione e normalizzazione degli schemi 60
- 3.2 Il problema dell'integrazione 61
  - 3.2.1 Diversità di prospettiva 63
  - 3.2.2 Equivalenza dei costrutti del modello 64
  - 3.2.3 Incompatibilità delle specifiche 64
  - 3.2.4 Concetti comuni 66
  - 3.2.5 Concetti correlati 67
- 3.3 Le fasi dell'integrazione 67
  - 3.3.1 Preintegrazione 68
  - 3.3.2 Comparazione degli schemi 70
  - 3.3.3 Allineamento degli schemi 73
  - 3.3.4 Fusione e ristrutturazione degli schemi 73
- 3.4 Definizione delle corrispondenze 75

**Capitolo 4 Analisi dei requisiti utente 77**

- 4.1 Le interviste 78
- 4.2 Analisi dei requisiti basata su glossari 81
  - 4.2.1 I fatti 82
  - 4.2.2 Il carico di lavoro preliminare 85
- 4.3 Analisi dei requisiti basata su obiettivi 88
  - 4.3.1 Introduzione a Tropos 88
  - 4.3.2 Modellazione dell'organizzazione 90
  - 4.3.3 Modellazione decisionale 93
- 4.4 Altri requisiti 96

**Capitolo 5 Modellazione concettuale 99**

- 5.1 Il Dimensional Fact Model: concetti di base 102
- 5.2 Modellazione avanzata 108
  - 5.2.1 Attributi descrittivi 108
  - 5.2.2 Attributi cross-dimensionali 112
  - 5.2.3 Convergenza 112
  - 5.2.4 Gerarchie condivise 114
  - 5.2.5 Archi multipli 115
  - 5.2.6 Archi opzionali 116
  - 5.2.7 Gerarchie incomplete 117
  - 5.2.8 Gerarchie ricorsive 119
  - 5.2.9 Dinamicità 120
  - 5.2.10 Additività 121
- 5.3 Aspetti intensionali: descrizione formale 123
  - 5.3.1 Il meta-modello 124
  - 5.3.2 Formalizzazione dei concetti di base del DFM 125
- 5.4 Sovrapposizione di schemi di fatto 127
- 5.5 Gli eventi 130
- 5.6 Aggregazione di eventi 134
  - 5.6.1 Aggregazione di misure additive 136
  - 5.6.2 Aggregazione di misure non-additive 138
  - 5.6.3 Aggregazione in presenza di convergenze e attributi cross-dimensionali 141
  - 5.6.4 Aggregazione in presenza di archi opzionali o multipli 141
  - 5.6.5 Aggregazione per schemi di fatto vuoti 146
  - 5.6.6 Aggregazione in presenza di dipendenze funzionali tra le dimensioni 147
  - 5.6.7 Aggregazione su gerarchie incomplete o ricorsive 148

**Capitolo 6 Progettazione concettuale 153**

- 6.1 Progettazione da schemi concettuali Entity/Relationship 154
  - 6.1.1 Definizione dei fatti 155
  - 6.1.2 Costruzione dell'albero degli attributi 157
  - 6.1.3 Potatura e innesto dell'albero degli attributi 163
  - 6.1.4 Le associazioni uno-a-uno 168
  - 6.1.5 Definizione delle dimensioni 169
  - 6.1.6 Il tempo 172
  - 6.1.7 Definizione delle misure 175
  - 6.1.8 Generazione dello schema di fatto 176
- 6.2 Progettazione da schemi logici relazionali 182
  - 6.2.1 Definizione dei fatti 183
  - 6.2.2 Costruzione dell'albero degli attributi 183
  - 6.2.3 Le altre fasi 186
- 6.3 Progettazione da schemi XML 190
  - 6.3.1 Modellazione delle associazioni in XML 190
  - 6.3.2 Fasi preliminari 193
  - 6.3.3 Scelta dei fatti e costruzione dell'albero degli attributi 193

- 6.4 Progettazione nell'approccio misto 196
  - 6.4.1 Mappatura dei requisiti 197
  - 6.4.2 Costruzione dello schema di fatto 198
  - 6.4.3 Raffinamento 199
- 6.5 Progettazione guidata dai requisiti 201

### Capitolo 7 Carico di lavoro e volume dati 203

- 7.1 Il carico di lavoro 204
  - 7.1.1 Espressioni dimensionali e interrogazioni sullo schema di fatto 204
  - 7.1.2 Interrogazioni di drill-across 209
  - 7.1.3 Interrogazioni composte 212
  - 7.1.4 Interrogazioni GPSJ annidate 213
  - 7.1.5 Validazione del carico di lavoro sullo schema concettuale 213
  - 7.1.6 Il carico di lavoro e gli utenti 214
- 7.2 Il volume dati 217

### Capitolo 8 Modellazione logica 221

- 8.1 I sistemi MOLAP 221
  - 8.1.1 Il problema della sparsità 222
- 8.2 I sistemi ROLAP 222
  - 8.2.1 Lo schema a stella 223
  - 8.2.2 Lo schema snowflake 226
- 8.3 Le viste 228
  - 8.3.1 Schemi relazionali in presenza di dati aggregati 232
- 8.4 Scenari temporali 235
  - 8.4.1 Gerarchie dinamiche: tipo 1 237
  - 8.4.2 Gerarchie dinamiche: tipo 2 238
  - 8.4.3 Gerarchie dinamiche: tipo 3 239
  - 8.4.4 Cancellazione di tuple 242

### Capitolo 9 Progettazione logica 243

- 9.1 Dagli schemi di fatto agli schemi a stella 244
  - 9.1.1 Attributi descrittivi 244
  - 9.1.2 Attributi cross-dimensionali 245
  - 9.1.3 Gerarchie condivise 246
  - 9.1.4 Archi multipli 247
  - 9.1.5 Archi opzionali 251
  - 9.1.6 Gerarchie incomplete 252
  - 9.1.7 Gerarchie ricorsive 253
  - 9.1.8 Dimensioni degeneri 255
  - 9.1.9 Problemi connessi all'additività 257
  - 9.1.10 Utilizzo di schemi snowflake 258
- 9.2 Materializzazione delle viste 260
  - 9.2.1 Risolvibilità delle interrogazioni sulle viste 266
  - 9.2.2 Formalizzazione del problema 267
  - 9.2.3 Un algoritmo di materializzazione 270

- 9.3 Frammentazione delle viste 272
  - 9.3.1 Frammentazione verticale delle viste 273
  - 9.3.2 Frammentazione orizzontale delle viste 276

### Capitolo 10 Progettazione dell'alimentazione 279

- 10.1 Alimentazione dello schema riconciliato 280
  - 10.1.1 L'estrazione dei dati 280
  - 10.1.2 La trasformazione dei dati 288
  - 10.1.3 Il caricamento dei dati 288
- 10.2 Pulizia dei dati 290
  - 10.2.1 Tecniche basate su dizionari 291
  - 10.2.2 Tecniche di fusione approssimata 292
  - 10.2.3 Tecniche ad hoc 295
- 10.3 Alimentazione delle dimension table 295
  - 10.3.1 Identificazione dei dati da caricare 296
  - 10.3.2 Sostituzione delle chiavi 296
- 10.4 Alimentazione delle fact table 297
- 10.5 Alimentazione delle viste materializzate 300

### Capitolo 11 Indici per il data warehouse 303

- 11.1 I B<sup>+</sup>-Tree 303
- 11.2 Gli indici bitmap 306
  - 11.2.1 Indici bitmap o B<sup>+</sup>-Tree? 309
  - 11.2.2 Indici bitmap evoluti 311
- 11.3 Gli indici di proiezione 314
- 11.4 Indici di join e indici a stella 317
  - 11.4.1 Indici Multi-Join 319
- 11.5 Indici spaziali 323
- 11.6 Algoritmi di join 325
  - 11.6.1 Nested loop 325
  - 11.6.2 Sort-merge 327
  - 11.6.3 Hash Join 327

### Capitolo 12 Progettazione fisica 331

- 12.1 L'ottimizzatore 331
  - 12.1.1 Gli ottimizzatori basati su regole 336
  - 12.1.2 Gli ottimizzatori basati sui costi 339
  - 12.1.3 Gli istogrammi 343
- 12.2 La scelta degli indici 346
  - 12.2.1 Indicizzazione delle dimension table 347
  - 12.2.2 Indicizzazione della fact table 349
- 12.3 Altri elementi di progettazione fisica 350
  - 12.3.1 Suddivisione in tablespace 351
  - 12.3.2 Allocazione dei datafile 352
  - 12.3.3 Dimensionamento dei blocchi di disco 356

**Capitolo 13 La documentazione di progetto 359**

- 13.1 Il livello del data warehouse 360
  - 13.1.1 Lo schema di data warehouse 360
  - 13.1.2 Lo schema di allocazione 362
- 13.2 Il livello dei data mart 364
  - 13.2.1 Lo schema di data mart 365
  - 13.2.2 Lo schema operativo 366
  - 13.2.3 Lo schema dell'alimentazione 367
  - 13.2.4 Il glossario dei domini 372
  - 13.2.5 Il carico di lavoro 372
  - 13.2.6 Lo schema logico e lo schema fisico 373
- 13.3 Il livello dei fatti 375
  - 13.3.1 Lo schema di fatto 375
  - 13.3.2 Glossario degli attributi e delle misure 376
- 13.4 Linee guida metodologiche 378

**Capitolo 14 Uno studio di caso 379**

- 14.1 Il dominio applicativo 379
- 14.2 Pianificazione del data warehouse di StraSport 380
- 14.3 Il data mart commerciale 381
  - 14.3.1 Analisi e riconciliazione delle fonti dati 381
  - 14.3.2 Analisi dei requisiti utente 391
  - 14.3.3 Progettazione concettuale 394
  - 14.3.4 Progettazione logica 400
  - 14.3.5 Progettazione dell'alimentazione 403
  - 14.3.6 Progettazione fisica 405
- 14.4 Il data mart del marketing 407

**Capitolo 15 Business intelligence: oltre il data warehouse 409**

- 15.1 Introduzione alla business intelligence 409
- 15.2 Analisi what-if 411
  - 15.2.1 Tecniche induttive 412
  - 15.2.2 Tecniche deduttive 414
- 15.3 Data mining 415
  - 15.3.1 Regole associative 417
  - 15.3.2 Clustering 419
  - 15.3.3 Alberi decisionali 420
  - 15.3.4 Serie temporali 421
- 15.4 Business Performance Management 421

**Glossario dei termini 427**

**Bibliografia 433**

**Indice analitico 445**