

Optimisation des bases de données

Mise en œuvre sous Oracle

Laurent Navarro

Avec la contribution technique
d'Emmanuel Lecoester

PEARSON

Pearson Education France a apporté le plus grand soin à la réalisation de ce livre afin de vous fournir une information complète et fiable. Cependant, Pearson Education France n'assume de responsabilités, ni pour son utilisation, ni pour les contrefaçons de brevets ou atteintes aux droits de tierces personnes qui pourraient résulter de cette utilisation.

Les exemples ou les programmes présents dans cet ouvrage sont fournis pour illustrer les descriptions théoriques. Ils ne sont en aucun cas destinés à une utilisation commerciale ou professionnelle.

Pearson Education France ne pourra en aucun cas être tenu pour responsable des préjudices ou dommages de quelque nature que ce soit pouvant résulter de l'utilisation de ces exemples ou programmes.

Tous les noms de produits ou marques cités dans ce livre sont des marques déposées par leurs propriétaires respectifs.

Publié par Pearson Education France
47 bis, rue des Vinaigriers
75010 PARIS
Tél. : 01 72 74 90 00
www.pearson.fr

Mise en pages : TyPAO

ISBN : 978-2-7440-2412-2
Copyright © 2010 Pearson Education France
Tous droits réservés

Aucune représentation ou reproduction, même partielle, autre que celles prévues à l'article L. 122-5 2° et 3° a) du code de la propriété intellectuelle ne peut être faite sans l'autorisation expresse de Pearson Education France ou, le cas échéant, sans le respect des modalités prévues à l'article L. 122-10 dudit code.

Table des matières

Préface	IX
À propos de l’auteur	XI
Introduction	1
1 Introduction aux SGBDR	5
1.1 Qu’est-ce qu’une base de données ?	5
1.1.1 Système de gestion des bases de données	6
1.2 Modèle de stockage des données	7
1.2.1 Organisation des données	7
1.2.2 Le RowID	10
1.2.3 Online Redo Log et Archived Redo Log	10
1.2.4 Organisation des tables	11
1.2.5 <i>Row Migration</i> et <i>Row Chaining</i>	12
1.2.6 Le cache mémoire	13
1.3 Intérêt des index dans les SGBDR	14
1.4 Analyse du comportement du SGBDR	15
1.4.1 Exécution d’une requête SQL	15
1.4.2 Optimiseur CBO (<i>Cost Based Optimizer</i>)	16

Axe 1

Étude et optimisation du modèle de données

2 Modèle relationnel	21
2.1 Présentation	21
2.2 Les bons réflexes sur le typage des données	25
2.2.1 Les types sous Oracle	28

2.2.2	Les types sous SQL Server.....	30
2.2.3	Les types sous MySQL.....	31
3	Normalisation, base du modèle relationnel	33
3.1	Normalisation	33
3.1.1	Première forme normale (1NF).....	34
3.1.2	Deuxième forme normale (2NF).....	35
3.1.3	Troisième forme normale (3NF)	36
3.1.4	Forme normale de Boyce Codd (BCNF)	37
3.1.5	Autres formes normales	38
3.2	Dénormalisation et ses cas de mise en œuvre	38
3.2.1	La dénormalisation pour historisation.....	38
3.2.2	La dénormalisation pour performance et simplification en environnement OLTP	40
3.2.3	La dénormalisation pour performance en environnement OLAP	42
3.3	Notre base de test.....	43

Axe 2

Étude et optimisation des requêtes

4	Méthodes et outils de diagnostic	49
4.1	Approche pour optimiser.....	49
4.1.1	Mesurer.....	50
4.1.2	Comprendre le plan d'exécution	55
4.1.3	Identifier les requêtes qui posent des problèmes.....	61
4.2	Outils complémentaires	64
4.2.1	Compteurs de performance Windows.....	64
4.2.2	SQL Tuning Advisor (Oracle)	65
4.2.3	SQL Access Advisor (Oracle).....	66
4.2.4	<i>SQL Trace</i> (Oracle).....	67
4.2.5	Outils SQL Server.....	72
4.2.6	Outils MySQL	73

5	Techniques d'optimisation standard au niveau base de données	77
5.1	Statistiques sur les données.....	77
5.1.1	Ancienne méthode de collecte.....	78
5.1.2	Nouvelle méthode de collecte.....	78
5.1.3	Sélectivité, cardinalité, densité.....	80
5.2	Utilisation des index.....	85
5.2.1	Index B*Tree.....	86
5.2.2	Index sur fonction.....	99
5.2.3	Reverse Index.....	100
5.2.4	Index bitmap.....	101
5.2.5	Bitmap Join Index.....	108
5.2.6	Full Text Index.....	111
5.3	Travail autour des tables.....	118
5.3.1	Paramètres de table.....	118
5.3.2	Index Organized Table	122
5.3.3	Cluster.....	132
5.3.4	Partitionnement des données.....	135
5.3.5	Les vues matérialisées	146
5.3.6	Reconstruction des index et des tables	148
6	Techniques d'optimisation standard des requêtes	153
6.1	Réécriture des requêtes	153
6.1.1	Transformation de requêtes.....	154
6.1.2	IN versus jointure.....	155
6.1.3	Sous-requêtes versus anti-jointures.....	156
6.1.4	Exists versus <i>Count</i>	158
6.1.5	<i>Exists versus IN</i>	159
6.1.6	Clause <i>Exists</i> * versus constante.....	160
6.1.7	Expressions sous requêtes	161
6.1.8	Agrégats : <i>Having versus Where</i>	162
6.2	Bonnes et mauvaises pratiques	163
6.2.1	Mélange des types	163
6.2.2	Fonctions et expressions sur index	163
6.2.3	Impact de l'opérateur <> sur les index.....	164
6.2.4	Réutilisation de vue.....	164

6.2.5	Utilisation de tables temporaires.....	165
6.2.6	Utilisation abusive de <i>SELECT</i> *.....	165
6.2.7	Suppression des tris inutiles.....	166
6.2.8	Utilisation raisonnée des opérations ensemblistes.....	166
6.2.9	Union versus <i>Union ALL</i>	169
6.2.10	<i>Count(*)</i> versus <i>count(colonne)</i>	170
6.2.11	Réduction du nombre de parcours des données.....	171
6.2.12	LMD et clés étrangères.....	172
6.2.13	Suppression temporaire des index et des CIR.....	173
6.2.14	<i>Truncate</i> versus <i>Delete</i>	173
6.2.15	Impacts des verrous et des transactions.....	173
6.2.16	Optimisation du <i>COMMIT</i>	175
6.2.17	DBLink et vues.....	176
7	Techniques d'optimisation des requêtes avancées.....	177
7.1	Utilisation des hints sous Oracle.....	177
7.1.1	Syntaxe générale.....	179
7.1.2	Les hints <i>Optimizer Goal</i>	179
7.1.3	Les hints <i>Access Path</i>	180
7.1.4	Les hints <i>Query Transformation</i>	181
7.1.5	Les hints de jointure.....	182
7.1.6	Autres hints.....	183
7.2	Exécution parallèle.....	183
7.2.1	Les hints de parallélisme.....	185
7.3	Utilisation du SQL avancé.....	186
7.3.1	Les <i>Grouping Sets</i>	186
7.3.2	<i>Rollup Group By</i>	188
7.3.3	<i>Cube Group By</i>	189
7.3.4	Utilisation de <i>WITH</i>	191
7.3.5	Les fonctions de classement (<i>ranking</i>).....	192
7.3.6	Autres fonctions analytiques.....	194
7.3.7	L'instruction <i>MERGE</i>	195
7.3.8	Optimisation des updates multitables.....	196
7.3.9	Insertion en mode <i>Direct Path</i>	198

7.4	PL/SQL.....	198
7.4.1	Impact des triggers	198
7.4.2	Optimisation des curseurs (<i>BULK COLLECT</i>).....	200
7.4.3	Optimisation du LMD (<i>FORALL</i>).....	203
7.4.4	SQL dynamique et <i>BULK</i>	208
7.4.5	Traitement des exceptions avec <i>FORALL</i>	209
7.4.6	Utilisation de cache de données	210
7.4.7	Utilisation du profiling	212
7.4.8	Compilation du code PL/SQL	212

Axe 3

Autres pistes d'optimisation

8	Optimisation applicative (hors SQL)	217
8.1	Impact du réseau sur le modèle client/serveur.....	217
8.2	Regroupement de certaines requêtes	217
8.3	Utilisation du binding.....	219
8.4	Utilisation de cache local à l'application.....	221
8.5	Utilisation du SQL procédural.....	221
8.6	Gare aux excès de modularité.....	221
9	Optimisation de l'infrastructure	223
9.1	Optimisation de l'exécution du SGBDR.....	223
9.1.1	Ajustement de la mémoire utilisable.....	223
9.1.2	Répartition des fichiers.....	224
9.2	Optimisation matérielle.....	224
9.2.1	Le CPU	224
9.2.2	La mémoire vive (RAM).....	225
9.2.3	Le sous-système disque.....	225
9.2.4	Le réseau	226
	Conclusion	227

Annexes

A	Gestion interne des enregistrements	231
A.1	Le RowID.....	231
A.2	Row Migration et Row Chaining	232
B	Statistiques sur les données plus en détail	235
B.1	Statistiques selon l'ancienne méthode de collecte	235
B.2	Statistiques selon la nouvelle méthode de collecte	236
B.3	Histogrammes.....	240
B.4	Facteur de foisonnement (<i>Clustering Factor</i>)	241
C	Scripts de création des tables de test BIGEMP et BIGDEPT	245
D	Glossaire	249
	Index	251