

TP1 Entrepôts de données

Master 1 MIAGE

Exercice 1 : Requêtes de classement sous Oracle

Sur les tables Emp et Dept créées à partir du script « Exercice_1.sql »

- a) Quel est le classement des salaires des employés par département pour les départements 10 et 30 ?

RZ	DEPTNO	RZ	ENAME	RZ	SAL	RZ	RANG
	10		KING		5000		1
	10		CLARK		2450		2
	10		MILLER		1300		3
	30		BLAKE		2850		1
	30		ALLEN		1600		2
	30		TURNER		1500		3
	30		MARTIN		1250		4
	30		WARD		1250		4
	30		JAMES		950		6

Résultat souhaité

- c) Quel est le classement décroissant des salaires différents par département pour les départements 10 et 20 ?

RZ	DEPTNO	RZ	SAL	RZ	RANG
	10		5000		1
	10		2450		2
	10		1300		3
	20		3000		1
	20		2975		2
	20		1100		3
	20		800		4

Résultat souhaité

- b) Idem en ôtant les trous dans le classement.

RZ	DEPTNO	RZ	ENAME	RZ	SAL	RZ	RANG
	10		KING		5000		1
	10		CLARK		2450		2
	10		MILLER		1300		3
	30		BLAKE		2850		1
	30		ALLEN		1600		2
	30		TURNER		1500		3
	30		MARTIN		1250		4
	30		WARD		1250		4
	30		JAMES		950		5

Résultat souhaité

- d) Quel est le salaire total versé par profession (répondre avec deux méthodes possibles avec et sans « group by ») ?

RZ	JOB	RZ	TOT_SAL_JOB
	CLERK		4150
	SALESMAN		5600
	PRESIDENT		5000
	MANAGER		8275
	ANALYST		6000

Résultat souhaité

- e) Quelle est la différence entre un « group by » et un « partition by » dans une requête SQL ?

- f) Quel est le montant total des salaires versés, tous départements et jobs confondus, par département et par département et job ?

DEPTNO	JOB	SUM(SAL)
10	CLERK	1300
10	MANAGER	2450
10	PRESIDENT	5000
10	(null)	8750
20	CLERK	1900
20	ANALYST	6000
20	MANAGER	2975
20	(null)	10875
30	CLERK	950
30	MANAGER	2850
30	SALESMAN	5600
30	(null)	9400
(null)	(null)	29025

Résultat souhaité

- g) Idem en ôtant toute confusion avec les valeurs nulles (répondre avec deux méthodes possibles)

DEPARTEMENT	JOB	SUM(SAL)
10	TousEmployés	8750
10	PRESIDENT	5000
10	MANAGER	2450
10	CLERK	1300
20	TousEmployés	10875
20	ANALYST	6000
20	MANAGER	2975
20	CLERK	1900
30	TousEmployés	9400
30	SALESMAN	5600
30	MANAGER	2850
30	CLERK	950
TousDep	TousEmployés	29025

Résultat souhaité

Exercice 2 : OLAP/ROLAP sous Oracle

Dans le monde de l'informatique décisionnelle, le ROLAP (Relational On-Line Analytical Processing) est une solution OLAP où l'entrepôt de données est stocké sur un SGBD relationnel, sous la forme d'un schéma en étoile ou en flocon. Cela permet de l'interroger en utilisant des requêtes SQL étendu (souvent très complexes et très exigeantes en terme de ressources et de temps d'exécution).

Une requête ROLAP est en général exprimée comme suit.

- On calcule la jointure de la table des faits et des relations dimensionnelles.
- On sélectionne des tuples en fonctions des données dimensionnelles.
- On groupe ces données suivant certaines dimensions.
- On calcule une valeur agrégée (le plus souvent une somme).

Les tables correspondantes sont créées par le script "Exercice_rolap.sql".

Observez ce script, modifiez le si nécessaire et exécutez-le pour générer la base.

1. Donnez la moyenne des ventes par année, région, catégorie ; par année, région et par année, pour les années 2009 et 2010.

Extrait de la réponse,

671 lignes au total

	ANNEE	CL_R	CATEGORY	CA_MOYEN
654	2010	05454-...	Condiments	3789
655	2010	05454-...	Pâtes et céréales	1100
656	2010	05454-...	Poissons et fruits de mer	3800
657	2010	05454-...	(null)	3665,1875
658	2010	Isle o...	Boissons	1500
659	2010	Isle o...	Desserts	361,5
660	2010	Isle o...	Pâtes et céréales	1170
661	2010	Isle o...	Poissons et fruits de mer	2850
662	2010	Isle o...	(null)	1226,14285...
663	2010	Nueva...	Viandes	1987,5
664	2010	Nueva...	Boissons	2536
665	2010	Nueva...	Desserts	1690,66666...
666	2010	Nueva...	Condiments	1670,08333...
667	2010	Nueva...	Produits laitiers	2880
668	2010	Nueva...	Pâtes et céréales	10440
669	2010	Nueva...	(null)	2719,26785...
670	2010	(null)	(null)	3257,22376...
671	(null)	(null)	(null)	3158,31296...

2. Donnez la moyenne des ventes par année, selon les dimensions région et catégorie pour les années 2009 et 2010.

Extrait de la réponse,
1155 lignes au total

R	ANNEE	R	CL_R	R	CATEGORY	R	CA_MOYEN
422	(null)	OX15	4...		Desserts		3455,66666...
423	(null)	OX15	4...		Produits secs		3233
424	(null)	OX15	4...		Produits laitiers		3458,58333...
425	(null)	OX15	4...		Pâtes et céréales		5867,5
426	(null)	OX15	4...		Poissons et fruits de mer		560
427	(null)	S-958...		(null)			2899,75892...
428	(null)	S-958...			Viandes		1568,4
429	(null)	S-958...			Boissons		6586,75
430	(null)	S-958...			Desserts		5059,5
431	(null)	S-958...			Condiments		1321,33333...
432	(null)	S-958...			Produits secs		1860
433	(null)	S-958...			Produits laitiers		1112,125
434	(null)	S-958...			Pâtes et céréales		2585
435	(null)	S-958...			Poissons et fruits de mer		875
436	(null)	05432-...		(null)			1708
437	(null)	05432-...			Viandes		240
438	(null)	05432-...			Desserts		1200

3. Quel est le produit qui fait le plus grand CA par année et par catégorie (cf. RANK) ?

Réponse : 24 lignes

R	ANNEE	R	CATEGORY	R	PNAME
1	2009	Boissons			Côte de Blaye
2	2009	Condiments			Northwoods Cranberry Sauce
3	2009	Desserts			Tarte au sucre
4	2009	Pâtes et céréales			Raclette Courdavault
5	2009	Poissons et fruits de mer			Wimmers gute Semmelknodel
6	2009	Produits laitiers			Alice Mutton
7	2009	Produits secs			Manjimup Dried Apples
8	2009	Viandes			Carnarvon Tigers
9	2010	Boissons			Côte de Blaye
10	2010	Condiments			Sirop d érable
11	2010	Desserts			Tarte au sucre
12	2010	Pâtes et céréales			Raclette Courdavault
13	2010	Poissons et fruits de mer			Gnocchi di nonna Alice
14	2010	Produits laitiers			Alice Mutton
15	2010	Produits secs			Manjimup Dried Apples
16	2010	Viandes			Carnarvon Tigers
17	2011	Boissons			Côte de Blaye
18	2011	Condiments			Vegie-spread
19	2011	Desserts			Tarte au sucre
20	2011	Pâtes et céréales			Raclette Courdavault
21	2011	Poissons et fruits de mer			Wimmers gute Semmelknodel
22	2011	Produits laitiers			Alice Mutton
23	2011	Produits secs			Uncle Bob s Organic Dried Pears
24	2011	Viandes			Carnarvon Tigers

4. Pour chaque année, donner le total des ventes ainsi que le total des ventes par catégorie. On ne veut pas le résultat pour le total des années (utiliser la fonction GROUPING_ID).

Réponse : 27 lignes

	ANNEE	CATEGORY	CA_TOTAL
1	2009 (null)		883768
2	2009 Viandes		97455
3	2009 Boissons		242708
4	2009 Desserts		110631,5
5	2009 Condiments		92685,5
6	2009 Produits secs		40947
7	2009 Produits laitiers		75166
8	2009 Pâtes et céréales		183831
9	2009 Poissons et fruits de mer		40344
10	2010 (null)		2514576,75
11	2010 Viandes		300852
12	2010 Boissons		457464
13	2010 Desserts		349497,5
14	2010 Condiments		251353,5
15	2010 Produits secs		179707
16	2010 Produits laitiers		251224,25
17	2010 Pâtes et céréales		488837,5
18	2010 Poissons et fruits de mer		235641
19	2011 (null)		1688292,7
20	2011 Viandes		200525,2
21	2011 Boissons		480950
22	2011 Desserts		228732,5
23	2011 Condiments		131237,25
24	2011 Produits secs		94246,25
25	2011 Produits laitiers		91183,25
26	2011 Pâtes et céréales		342882
27	2011 Poissons et fruits de mer		118536,25

5. Quel est le meilleur mois de vente du produit "Sirop d érable" pour chacune des années ?

Réponse : 2 lignes

	ANNEE	MOIS	CA_TOTAL
1	2010	7	17100
2	2011	4	13822,5

6. Donnez tous les totaux de ventes par année selon la dimension catégorie, ainsi que selon la dimension cl_name (nom du client). Utilisez la commande GROUPING_SETS.

Extrait de la réponse,
229 lignes au total

	ANNEE	CL_NAME	CATEGORY	CA_TOTAL
192	2011	Wilman Kala	(null)	8351
193	2011	Magazzini Alimentari Riun...	(null)	8465
194	2011	Cactus Comidas para llevar	(null)	4535
195	2010	Maison Dewey	(null)	23860
196	2010	Franchi S.p.A.	(null)	809,5
197	2011	Que Delicia	(null)	6768
198	2011	LILA-Supermercado	(null)	28995,3
199	2011	Hungry Owl All-Night Groc...	(null)	27328,7
200	2010	Vins et alcools Chevalier	(null)	1899
201	2009	Seven Seas Imports	(null)	25512
202	2010	Old World Delicatessen	(null)	29480
203	2009	Old World Delicatessen	(null)	13082,5
204	2009	Bon app	(null)	21012,5
205	2010	Laughing Bacchus Wine Cel...	(null)	945
206	2010	(null)	Condiments	251353,5
207	2009	(null)	Poissons e...	40344
208	2011	(null)	Poissons e...	118536,25
209	2011	(null)	Boissons	480950
210	2010	(null)	Desserts	349497,5
211	2010	(null)	Boissons	457464
212	2011	(null)	Produits s...	94246,25
213	2010	(null)	Produits l...	251224,25
214	2010	(null)	Poissons e...	235641
215	2011	(null)	Desserts	228732,5
216	2011	(null)	Condiments	131237,25
217	2009	(null)	Condiments	92685,5
218	2009	(null)	Desserts	110631,5
219	2011	(null)	Pâtes et c...	342882

7. Quelle est la répartition par tiers des catégories selon leurs quantités totales vendues en 2010 ?

Réponse : 8 lignes

	CATEGORY	QTE_VENDUE_2010	TIERS
1	Pâtes et céréales	3431	1
2	Desserts	3355	1
3	Viandes	2751	1
4	Boissons	2256	2
5	Condiments	2233	2
6	Poissons et fruits de mer	1878	2
7	Produits laitiers	1820	3
8	Produits secs	1046	3

8. Quelle est la quantité de produits vendus pour chaque catégorie, les 5 premiers jours de chaque mois de 2010 ?

On peut imaginer plusieurs solutions, selon que l'on interprète « les 5 premiers jours de chaque mois » par « un jour compris entre 1 et 5 » ou « les 5 premiers jours de vente ».

La seconde solution semble plus logique, et on obtient comme résultat :

Extrait de la réponse,

72 lignes au total

QTE	CATEGORY	MOIS	JOUR1	JOUR5	QTE_5_JOURS
1	Boissons	1	153	171	179
2	Boissons	2	188	205	94
3	Boissons	3	218	238	143
4	Boissons	4	255	273	51
5	Boissons	5	276	294	134
6	Boissons	7	339	361	77
7	Boissons	8	375	384	125
8	Boissons	9	405	420	103
9	Boissons	10	441	449	216
10	Boissons	11	482	505	138
11	Boissons	12	513	530	205
12	Condiments	1	153	171	191
13	Condiments	2	194	204	313
14	Condiments	4	246	266	196
15	Condiments	5	280	303	240