

Table des matières

Introduction	1
Chapitre 1. Les fondements de l'analyse de données avec le langage SQL	11
1.1. Données au cœur du système d'information	11
1.1.1. Introduction	11
1.1.2. Notion de base de données	13
1.1.3. Système de gestion de base de données (SGBD)	14
1.1.4. Modèle relationnel	15
1.1.5. Langage SQL	16
1.1.5.1. Commandes SQL	16
1.1.5.2. Normes SQL	17
1.1.6. Fonctions analytiques du langage SQL	19
1.1.6.1. Fonctions d'agrégation	20
1.1.6.2. Fonctions de pagination	20
1.1.6.3. Fonctions de fenêtrage	20
1.1.6.4. Fonctions de classement	20
1.1.6.5. Fonctions de distribution et de décalage	20
1.1.6.6. Fonctions de traitement des valeurs nulles	21
1.1.6.7. Opérateurs de rotation	21
1.1.6.8. Fonctions hiérarchiques	21
1.1.7. Conclusion	21

1.2. SQL Server et l'analyse de données	21
1.2.1. Introduction	21
1.2.2. Architecture SQL Server	23
1.2.3. Versions SQL Server	24
1.2.4. Éditions de SQL Server	25
1.2.5. SQL Server dans le Cloud	26
1.2.6. Genèse des fonctions analytiques du langage SQL dans le SGBD SQL Server	26
1.2.7. Du langage SQL Analytique vers le <i>In-Database Analytics</i>	28
1.2.8. Environnement de test	29
1.2.8.1. Installation de SQL Server 2014	29
1.2.8.2. Base de données de test AdventureWorks2014	31
1.2.8.3. Projets de requêtes	34
1.2.9. Conclusion	35
Chapitre 2. Les requêtes	37
2.1. Filtrage des données	37
2.1.1. Introduction	37
2.1.2. Types de filtres	38
2.1.3. Recherche conditionnelle exacte	39
2.1.3.1. Recherche conditionnelle exacte avec un filtre par condition	39
2.1.3.2. Recherche conditionnelle exacte avec un filtre par motif	40
2.1.4. Recherche conditionnelle approchée (recherche en texte intégral).	43
2.1.4.1. Service de recherche en texte intégral	43
2.1.4.2. Catalogue de recherche en texte intégral	44
2.1.4.3. Index de type FULLTEXT	44
2.1.4.4. Listes d'arrêt STOPLIST	45
2.1.4.5. Recherche de type FULLTEXT	46
2.1.5. Conclusion	49
2.2. Tri des données	50
2.2.1. Introduction	50
2.2.2. Critères de tri	50

2.2.3. Sens de tri	51
2.2.4. Référencement des critères de tri	52
2.2.5. Tri et performance	53
2.2.6. Conclusion	54
2.3. Pagination des données.	54
2.3.1. Introduction	54
2.3.2. Option TOP	55
2.3.3. Clause OFFSET... FETCH.	57
2.3.4. Conclusion	58
2.4. Sous-requêtes.	58
2.4.1. Introduction	58
2.4.2. Sous-requêtes autonomes	59
2.4.2.1. Sous-requêtes autonomes scalaires.	60
2.4.2.2. Sous-requêtes autonomes à valeurs multiples	61
2.4.3. Sous-requêtes corrélées	62
2.4.4. Sous-requêtes optimisées	63
2.4.5. Conclusion	64
2.5. Options de la clause FROM.	64
2.5.1. Introduction	64
2.5.2. Tables	66
2.5.3. Vues	66
2.5.4. Tables dérivées	68
2.5.5. Fonctions retournant des tables (TVF).	69
2.5.6. Expressions de tables communes (CTE)	70
2.5.7. Conclusion	70

Chapitre 3. Les opérateurs 71

3.1. Jointure.	71
3.1.1. Introduction	71
3.1.2. Types de jointures	72
3.1.3. Jointure interne	74
3.1.3.1. Jointure interne entre deux tables.	74
3.1.3.2. Jointure interne entre plusieurs tables	74
3.1.3.3. Jointure interne avec un filtre	75
3.1.3.4. Jointure interne sur la même table (autojointure).	76

3.1.4. Jointure externe	77
3.1.4.1. Jointure externe gauche	77
3.1.4.2. Jointure externe droite	78
3.1.4.3. Jointure externe complète	78
3.1.5. Produit cartésien	79
3.1.6. Jointure latérale	80
3.1.7. Conclusion	81
3.2. Opérateurs ensemblistes	82
3.2.1. Introduction	82
3.2.2. Opérateur ensembliste UNION	83
3.2.3. Opérateur ensembliste INTERSECT	84
3.2.4. Opérateur ensembliste EXCEPT	85
3.2.5. Opérateur ensembliste de division	86
3.2.6. Conclusion	87
3.3. Opérateurs de rotation	87
3.3.1. Introduction	87
3.3.2. Opérateur PIVOT	88
3.3.3. Opérateur UNPIVOT	90
3.3.4. Conclusion	91

Chapitre 4. Les fonctions 93

4.1. Fonctions prédéfinies	93
4.1.1. Introduction	93
4.1.2. Fonctions prédéfinies scalaires	94
4.1.2.1. Fonctions scalaires numériques	95
4.1.2.2. Fonctions scalaires de traitement des chaînes de caractères	98
4.1.2.3. Fonctions scalaires de traitement des dates	100
4.1.2.4. Fonctions scalaires de conversion	102
4.1.2.5. Fonctions scalaires de traitement des valeurs nulles . .	103
4.1.2.6. Fonctions scalaires logiques	105
4.1.3. Fonctions utilisateur	107
4.1.3.1. Fonction utilisateur scalaire	107
4.1.3.2. Fonction utilisateur retournant une table	107
4.1.3.3. Fonction utilisateur d'agrégation	108
4.1.4. Conclusion	109

4.2. Fonctions d'agrégation	109
4.2.1. Introduction	109
4.2.2. Fonctions d'agrégation usuelles	110
4.2.3. Clause GROUP BY	112
4.2.4. Filtres WHERE et HAVING	113
4.2.5. Options de la clause GROUP BY	114
4.2.5.1. Option ROLLUP	115
4.2.5.2. Option CUBE	116
4.2.5.3. Option GROUPING SETS	117
4.2.5.4. Option GROUPING_ID	118
4.2.6. Conclusion	120
4.3. Fonctions de fenêtrage	121
4.3.1. Introduction	121
4.3.2. Mode de fonctionnement de la fonction de fenêtrage OVER()	122
4.3.3. Fonctions analytiques associées au mécanisme de fenêtrage	123
4.3.4. Jeu de données, partition et <i>frames</i>	124
4.3.5. Options de fenêtrage	124
4.3.5.1. Option PARTITION BY	125
4.3.5.2. Option ORDER BY	125
4.3.5.3. Options ROWS BETWEEN et RANGE	126
4.3.6. Conclusion	128
4.4. Fonctions analytiques	129
4.4.1. Introduction	129
4.4.2. Fonctions de classement	130
4.4.2.1. Introduction	130
4.4.2.2. Fonction RANK	130
4.4.2.3. Fonction DENSE_RANK	131
4.4.2.4. Fonction NTILE	131
4.4.2.5. Fonction ROW_NUMBER	132
4.4.2.6. Conclusion	132
4.4.3. Fonctions de distribution	132
4.4.3.1. Introduction	132
4.4.3.2. Écart type STDEV et STDEVP	133
4.4.3.3. Variance VAR et VARP	133

- 4.4.3.4. Distribution cumulative PERCENT_RANK
et CUME_DIST 135
- 4.4.3.5. Distribution inversée PERCENTILE 136
- 4.4.3.6. Conclusion 137
- 4.4.4. Fonctions de décalage 137
 - 4.4.4.1. Introduction 137
 - 4.4.4.2. Fonction LEAD. 137
 - 4.4.4.3. Fonction LAG. 138
 - 4.4.4.4. Fonction FIRST_VALUE 139
 - 4.4.4.5. Fonction LAST_VALUE. 140
 - 4.4.4.6. Conclusion. 141

Bibliographie 143

Index 145