

SYSTEMES DISTRIBUES ET BASES DE DONNEES REPARTIES

Site web formations : www.hittechnologie.com

SGBD : Systèmes de Gestion des Bases de Données

NoSQL : Not only SQL

SD : Systèmes Distribués

COURS et EXERCICES CORRIGES

Pour concepteur et réalisateur des bases de données, d'applications, des Systèmes d'Information distribués

HIT-TECHNOLOGY SARL U

SYSTEMES DISTRIBUES ET BASES DE DONNEES REPARTIES COURS et EXERCICES CORRIGES

Formateur en développement applicatif
Enseignant à l'Université de Lomé et dans les Institutions privées (Togo)
Alassani AKANATE
Fondateur du Cabinet HIT-TECHNOLOGY SARL U
a.daviman@gmail.com
Site formations : www.hittechnologie.com
Site E-book : www.services.hittechnologie.com
Tel : (+228) 97 31 77 52 / 93 32 77 18

« L'avenir de l'Homme se trouve dans sa manière de penser »

L'équipe HIT-T

© Tous droits réservés 2026 HIT-T

PRIX : 4 000 F CFA

DANS LA MÊME COLLECTION

1	ARCHITECTURE ET INTERCONNEXION RESEAU
2	SGBD MySQL ET SQL
3	ANALYSE ET CONCEPTION DES SYSTEMES D'INFORMATION (ACSI) AVEC MERISE
4	SGBD MICROSOFT ACCESS 2010
5	SGBD MICROSOFT EXCEL 2007-2010 (Excel débutant, intermédiaire et avancé)
6	SQL POUR MySQL ET ORACLE
7	SE-OS, CONFIGURATION ET SECURITE
8	SECURITE INFORMATIQUE ET RESEAU
9	INFORMATIQUE ET CONFIGURATION
10	OFFICE 2007-2010 ET INITIATION (Word-Excel-PowerPoint-Publisher)
11	DEVELOPPEMENT DE SITES WEB AVEC LE CMS/SGC JOOMLA ET LES BASES DE DONNEES MySQL
12	P.O.O. : Java SE
13	P.O.O. : C++ et Qt
14	Ecrits Professionnels (EP)
15	Droit Social et de Travail (M. ALASSANI)
16	Economie Générale (M. AMANA)
17	Santé, Sécurité et Hygiène au travail (S. S. H. T.)
18	Tenue de caisse – Facturation
19	Electronique dans l'automobile
20	Climatisation dans l'automobile
21	MS SQL SERVER 2012-2014 et SQL / T-SQL
22	LES TECHNOLOGIES xDSL (Techniques d'accès à l'Internet)
23	MIEUX PARLER ANGLAIS
24	DROIT, NEGOCIATION ET CONTRATS INFORMATIQUES
25	SOCIAL MEDIA – NETWORKING (Marketing digital)
26	Patrons de conception (Design Patterns), UML, POO ET APPLICATIONS, etc.

COLLECTION HIT-T SARL U (E-book : www.services.hittechnologie.com)

Site web : www.hittechnologie.com

TABLE DES MATIERES

TABLE DES MATIERES 2
QUI S'ADRESSE CE LIVRE ? 10
POURQUOI CE LIVRE ? 10
OBJECTIFS SPECIFIQUES DE CE LIVRE 10
A PROPOS 10
INTRODUCTION GENERALE 11
CHAPITRE I : BIG DATA 12
INTRODUCTION 13
1. PAYSAGE DES SGBD 13
1.1. SGBD : un peu d'histoire 14
1.2. Les données aujourd'hui 14
1.3. Un point sur le relationnel 15
1.4. Un point sur le non-relationnel 15
1.4.1. Trois grandes catégories + SGBD colonnes 15
1.4.2. NoSQL : objectifs prioritaires 15
1.4.3. NoSQL : objectifs faibles 16
1.4.4. NoSQL : Les 4 catégories 16
1.4.5. Comment choisir un système NoSQL 22
1.5. Modèle de cohérence ? 22
1.6. ACID vs BASE 23
1.6.1. ACID 23
1.6.2. BASE 23
1.6.2.1. Signification de l'acronyme BASE 24
1.6.2.2. Comparatif : ACID vs BASE 24
1.6.2.3. Pourquoi utiliser BASE en Big Data ? 25
1.7. Pourquoi pas ACID+BASE ? 25
1.8. Théorème CAP [Brewer et Lynch] 26
1.8.1. Définition de CAP 26
1.8.2. Les 3 Piliers du Théorème CAP 27
1.8.3. Les Choix et Compromis (CP vs AP) 27
2. BIG DATA 27
2.1. Caractéristiques du Big Data 28
2.1.1. Volume 28

2.1.2. Vélocité 28
2.1.3. Variété 28
2.2. Autres caractéristiques 28
2.2.1. Veracity 29
2.2.2. Variabilité 29
2.2.3. Valeur 29
2.3. Comment fonctionne le big data ? 29
2.3.1. Intégration 29
2.3.2. Gestion 29
2.3.3. Analyse 30
2.4. Avantages du big data 30
2.4.1. Amélioration de la prise de décision 30
2.4.2. Agilité et innovation accrues 30
2.4.3. Meilleure expérience client 30
2.4.4. Intelligence continue 30
2.4.5. Des opérations plus efficaces 31
2.4.6. Amélioration de la gestion des risques 31
2.5. Défis liés à l'implémentation de l'analyse de big data 31
2.5.1. Manque de compétences et d'expertise en matière de données 31
2.5.2. Vitesse de croissance des données 31
2.5.3. Problèmes de qualité des données 31
2.5.4. Cas de non-conformité 32
2.5.5. Complexité de l'intégration 32
2.5.6. Problèmes de sécurité 32
2.6. Quelles sont les performances des entreprises basées sur les données ? 32
2.7. Stratégies et solutions de big data 33
2.7.1. Ouvert 33
2.7.2. Intelligent 34
2.7.3. Flexible 34
2.7.4. Confiance 34
2.8. Exemples d'utilisation 34
2.9. D'où vient le "Big Data" ? 35
2.9.1. Les données GPS : Le cas d'école du Big Data 36

2.9.2. Lien entre GPS et modèle BASE (Cohérence à terme) 37
2.9.3. Technologies Big Data pour le traitement GPS 37
2.10. Explosion de l'IoT 38
2.11. Dépenses de l'IoT 38
2.12. L'ampleur des réseaux sociaux 39
2.13. Types d'analyse 39
2.14. Cycle de vie de la donnée 40
2.15. Préparation des données 40
2.16. Qualité des données et sécurité 41
3. DATA CENTERS 42
3.1. Les plus grands data centers et campus mondiaux 42
3.2. Classement des pays ayant plus de Data Centers 43
3.3. Facteurs de domination 43
3.4. Durée de vie d'un Data Center 43
3.5. Combien de data centers y a-t-il dans le monde ? 44
3.5.1. Répartition par pays principaux 44
3.5.2. Les infrastructures géantes (Hyperscale) 44
3.5.3. Quels sont les inconvénients des data centers ? 44
3.5.4. Comment fonctionne un data center ? 45
4. OLTP 45
4.1. Caractéristiques de l'OLTP 46
4.2. Conformité ACID des systèmes OLTP 46
4.3. Les systèmes OLTP demandent une grande disponibilité 47
4.4. Exemples de logiciels OLTP 47
4.5. Domaines d'application 47
5. OLAP 47
5.1. Fonctionnement des systèmes OLAP 48
5.2. Cinq opérations d'analyse OLAP 49
5.3. Outils décisionnels et serveurs OLAP 50
5.4. Moteurs de bases de données OLAP modernes (orientés colonnes) 50
5.5. Les trois architectures : MOLAP, ROLAP et HOLAP 50
5.6. Comment choisir selon votre volume de données ? 51
6. DIFFERENCE ENTRE OLTP ET OLAP 51

6.1. Comparaison OLTP vs OLAP 51
6.2. Comparaison, quelques détails 52
6.3. Comment choisir pour votre projet ? 52
6.3.1. Bases de données OLTP (Transactionnel) : Outils 53
6.3.2. Solutions OLAP (Analytique / Data Warehouse) : Outils 53
7. EXEMPLE D'APPLICATIONS 53
7.1. Exemple OLTP : Gestion d'une vente en direct 54
7.1.1. Structure des tables (Normalisée) 54
7.1.2. Requête SQL OLTP typique 54
7.2. Exemple OLAP : Analyse des ventes annuelles 54
7.2.1. Structure de la table (Dénormalisée) 54
7.2.2. Requête SQL OLAP typique 54
8. TRANSFERT DES DONNEES DE OLTP VERS OLAP 55
8.1. Extraction (Extract) 55
8.2. Transformation (Transform) 55
8.3. Chargement (Load) 55
9. ETL CLASSIQUE VS ELT MODERNE 56
9.1. Outils ELT Modernes (Cloud & Big Data) 56
9.2. Plateformes ETL d'Entreprise (Visuelles & Robustes) 56
9.3. Orchestrateurs (Gestion des flux) 57
10. DIFFENCES DATA LAKE, DW, BDD 57
10.1. Data Lake : Définition 57
10.1.1. Data Lake : Qu'est-ce que c'est ? 57
10.1.2. L'architecture d'un Data Lake 57
10.1.2.1. Les Zones de Stockage Logiques 58
10.1.2.2. Les Composants Clés de l'Architecture 59
10.1.2.3. L'Évolution Moderne : Le Data Lakehouse 59
10.1.3. Big Data Analytics et les Data Lakes 59
10.1.4. Les enjeux du Data Lake 60
10.1.5. Conception d'un Data Lake 60
10.1.5.1. Écosystème Amazon Web Services (AWS) 60
10.1.5.2. Écosystème Microsoft Azure 60
10.1.5.3. Écosystème Google Cloud Platform (GCP) 61

10.1.5.4. L'Alternative Open-Source (Infrastructures Privées / On-Premise) 61
10.2. Data Lake, Data Warehouse, Base de données : Quelle est la différence ? 61
10.2.1. Les Bases de données 62
10.2.2. Le Data Warehouse 62
10.3. Le processus de DW 63
10.3.1. Le Data Lake 65
10.3.2. Schéma d'un Data Lake 65
10.3.3. Data warehouse vs Data Lake 65
11. AZURE DATA LAKE STORE GEN2 66
12. DE HADOOP A SPAR 68
12.1. Typologie des outils big data 68
12.1.1. Pile logicielle Big data 68
12.1.2. Machine Learning-Définition 72
12.1.3. Infrastructure matérielle pour le Big Data 74
12.2. Mapreduce et Hadoop 74
12.2.1. Architecture d'un cluster 74
12.2.2. Calcul à large échelle 75
13. INSIGHT 75
CONCLUSION 76
CHAPITRE II : SYSTEMES DISTRIBUES ET BASES DE DONNEES REPARTIES ... 77
INTRODUCTION 78
1. GENERALITES ET DEFINITIONS 78
1.1. Les soubassements technologiques de l'informatique distribuée 79
1.2. Rappel des protocoles de communication 80
1.2.1. Le mode non connecté 80
1.2.2. Le mode connecté 80
1.3. Les avantages de l'informatique distribuée 81
1.3.1. L'augmentation des ressources 81
1.3.2. La répartition des données et des services 81
1.3.3. Le pair à pair (peer-to-peer) 81
1.4. Les perspectives des architectures distribuées 82
2. CONCEPTS ET TECHNIQUES 82
2.1. Définition 82
2.2. Principe fondamental 82

2.2.1. Autonomie locale 83
2.2.2. Egalité entre sites 83
2.2.3. Fonctionnement continu 83
2.2.4. Localisation transparente 83
2.2.5. Fragmentation transparente 83
2.2.6. Indépendance à la réplication 83
2.2.7. Requêtes distribuées 84
2.2.8. Transactions réparties 84
2.2.9. Indépendance vis-à-vis du matériel 84
2.2.10. Indépendance vis-à-vis du SE 84
2.2.11. Indépendance vis-à-vis du réseau 84
2.2.12. Indépendance vis-à-vis du SGBD 84
2.3. Conception de BD répartie : Deux (2) approches 84
2.3.1. Conception ascendante 84
2.3.2. Conception descendante 85
2.4. Schémas d'une BD répartie 86
2.5. Fragmentation des données 86
2.5.1. Principaux types de fragmentation 87
2.5.2. Conséquences majeures 87
2.5.2.1. Fragmentation horizontale 87
2.5.2.2. Fragmentation verticale 88
2.5.3. Comparaison, conception et implémentation 89
2.5.3.1. Mode de découpage 89
2.5.3.2. Critère de sélection 89
2.5.3.3. Gestion des clés 89
2.5.3.4. Opération de reconstruction 90
2.5.3.5. Exemple concret 90
2.5.3.6. Exemples de requêtes 90
2.5.3.7. Exercice pratique 91
2.5.3.8. Pour aller plus loin : Requête de reconstruction totale 92
3. CDN (Content Delivery Network) 93
3.1. Qu'est-ce qu'un CDN ? 93
3.2. Les bénéfices d'un CDN 93

3.3. Le fonctionnement du CDN en 4 étapes 94
3.3.1. Les utilisateurs accèdent à vos contenus, partout dans le monde 94
3.3.2. Envoi des fichiers à l'utilisateur et communication entre le CDN et vos serveurs 94
3.3.3. L'utilisateur revient : Le site est plus rapide ! 95
3.3.4. Et sans CDN ? 96
4. PROGRAMMATION DISTRIBUEE 97
4.1. Concepts clés 97
4.2. Outils et technologies 97
4.3. Quelquess questions clés liées aux systèmes distribués 97
4.3.1. Comment les ordinateurs dans un système distribué communiquent-ils ? 97
4.3.2. Quels sont les modèles courants des systèmes distribués ? 98
4.3.3. Comment les données sont-elles stockées dans les systèmes distribués ? 98
4.3.4. Comment un système distribué traite-t-il les tâches plus rapidement ? 98
4.3.5. Qu'est-ce qu'un système distribué tolérant aux pannes ? 98
4.3.6. Quel est le rôle de l'équilibrage de la charge dans les systèmes distribués ? 99
4.3.7. Comment les logiciels intermédiaires simplifient-ils les systèmes distribués ? 99
4.3.8. Comment les risques de sécurité changent-ils dans les systèmes distribués ? 99
4.3.9. Quel est le théorème CAP dans les systèmes distribués ? 100
4.3.10. Comment les transactions sont-elles gérées dans les systèmes distribués ? 100
4.3.11. Puis-je utiliser n'importe quel langage de programmation pour les systèmes distribués ? 100
4.3.12. Quelles industries bénéficient le plus des systèmes distribués ? 100
4.4. Les technologies clés de distribution en Java 100
4.4.1. Exemple pratique : Serveur et Client HTTP simple (Java 11+) 101
4.4.2. Frameworks d'entreprise incontournables.....102
4.4.3. Exemple de programmation distribuée avec Spring Boot 103
4.5. Les Streams d'entrée – sortie : InputStream-OutputStream 107
4.5.1. InputStream (Flux d'entrée) 107
4.5.2. Méthodes clés.....107
4.5.3. OutputStream (Flux de sortie) 107
4.5.4. Méthodes clés.....107
4.5.5. Points importants.....108
CONCLUSION.....108
CONCLUSION GENERALE.....109

QUESTIONNAIRE AUX CHOIX MULTIPLES (QCM) 110
REPONSES AU QUESTIONNAIRE 114
RECUEIL DES EXERCICES 115
CORRIGES PROPOSES DES EXERCICES 124
GLOSSAIRE 125
BIBLIOGRAPHIE ET WEBOGRAPHIE 133

QUI S'ADRESSE CE LIVRE ?

Ce livre s'adresse aux débutants comme aux professionnels en activités et à toute personne désireuse de comprendre le fonctionnement, développer et déployer les systèmes distribués.

PREREQUIS

- Connaissance en informatique et Internet ;
- Connaissance des notions du client-serveur ;
- Connaissance d'un langage de programme orienté-objet ;
- Connaissance de base sur les bases de données reste un atout.

POURQUOI CE LIVRE ?

Ce livre répond aux questions et aux attentes des apprenants désireux de comprendre le fonctionnement, développer et déployer les systèmes distribués et les bases de données réparties. Ce livre discute également des concepts liés aux big data et à leur exploitation, concepts liés à la naissance de systèmes distribués.

OBJECTIFS SPECIFIQUES DE CE LIVRE

- Etudier et discuter les concepts de big data ;
- Présenter les outils de manipulation de big data ;
- Présenter les Data Centers ;
- Présenter et discuter des systèmes distribués ;
- Discuter des caractéristiques des bases de données réparties ;
- Discuter des modes de conception et de mise en œuvre des SD ;
- Présenter un recueil des exercices et corrigés ;
- Mettre en place un glossaire lié au domaine d'étude.

A PROPOS

HIT-TECHNOLOGY (HIT-T) SARL U est un cabinet spécialisé dans le développement des applications web et mobiles. Il offre également des services tels que : Les formations en Multimédia, Informatique et Télécommunications, l'installation des réseaux informatiques, la maintenance informatique et réseau, la conception et la réalisation de sites web, la conception et la réalisation des applications web et mobiles, le consulting dans le domaine de mise en place des bases de données, des systèmes d'information, la sécurité réseau, pour ne citer que ces points. Les formations sont offertes dans les filières tertiaires et technologiques. Le cabinet est enregistré sous le décret n° 2012-008 / PR du 07-03-2012, Arrêté n° 011 / MCPSP / CAB / DPSP du 13-04-2012.