



BIG DATA ET L'IA (INTELLIGENCE ARTIFICIELLE)

Site web formations : www.hittechnologie.com

SGBD : Systèmes de Gestion des Bases de Données

IA : Intelligence Artificielle

COURS et EXERCICES CORRIGES

*Pour concepteur et réalisateur des systèmes
d'IA et analyste des données et big data*

HIT-TECHNOLOGY SARL U

BIG DATA ET IA

COURS et EXERCICES CORRIGES

Alassani
AKANATE

Formateur en développement applicatif
Enseignant à l'Université de Lomé et dans les Institutions privées
(Togo)
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 : 4 400 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.

TABLE DES MATIERES

TABLE DES MATIERES.....	2
A QUI S'ADRESSE CE LIVRE ?	10
POURQUOI CE LIVRE ?	10
OBJECTIFS SPECIFIQUES DE CE LIVRE	10
A PROPOS.....	10
INTRODUCTION GENERALE	12
CHAPITRE I : BIG DATA ET LES TECHNOLOGIES ASSOCIEES.....	13
INTRODUCTION	14
1. PAYSAGE DES SGBD	14
1.1. SGBD : un peu d'histoire.....	15
1.2. Les données aujourd'hui	15
1.3. Un point sur le relationnel	16
1.4. Un point sur le non-relationnel	16
1.4.1. Trois grandes catégories + SGBD colonnes	16
1.4.2. NoSQL : objectifs prioritaires.....	16
1.4.3. NoSQL : objectifs faibles	17
1.4.4. NoSQL : Les 4 catégories.....	17
1.4.5. Comment choisir un système NoSQL.....	23
1.5. Modèle de cohérence ?.....	23
1.6. ACID vs BASE	24
1.6.1. ACID	24
1.6.2. BASE	24
1.7. Pourquoi pas ACID+BASE ?.....	25
1.8. Théorème CAP [Brewer et Lynch]	25
1.8.1. Définition de CAP	26
1.8.2. Les 3 Piliers du Théorème CAP	26
1.8.3. Les Choix et Compromis (CP vs AP)	26
2. BIG DATA	27
2.1. Caractéristiques du Big Data.....	27
2.1.1. Volume	27
2.1.2. Vitesse.....	27
2.1.3. Variété	27
2.2. Autres caractéristiques	28

2.2.1.	Veracity	28
2.2.2.	Variabilité.....	28
2.2.3.	Valeur.....	28
2.3.	Comment fonctionne le big data ?	28
2.3.1.	Intégration.....	29
2.3.2.	Gestion.....	29
2.3.3.	Analyse	29
2.4.	Avantages du big data	29
2.4.1.	Amélioration de la prise de décision.....	29
2.4.2.	Agilité et innovation accrues	29
2.4.3.	Meilleure expérience client	30
2.4.4.	Intelligence continue	30
2.4.5.	Des opérations plus efficaces	30
2.4.6.	Amélioration de la gestion des risques	30
2.5.	Défis liés à l'implémentation de l'analyse de big data.....	30
2.5.1.	Manque de compétences et d'expertise en matière de données.....	30
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	31
2.5.5.	Complexité de l'intégration	31
2.5.6.	Problèmes de sécurité	31
2.6.	Quelles sont les performances des entreprises basées sur les données ?	31
2.7.	Stratégies et solutions de big data	32
2.7.1.	Ouvert	33
2.7.2.	Intelligent.....	33
2.7.3.	Flexible	33
2.7.4.	Confiance.....	33
2.8.	Exemples d'utilisation.....	34
2.9.	D'où vient le "Big Data" ?	35
2.10.	Explosion de l'IoT.....	36
2.11.	Dépenses de l'IoT.....	37
2.12.	L'ampleur des réseaux sociaux	37
2.13.	Types d'analyse	38

2.14.	Cycle de vie de la donnée	38
2.15.	Préparation des données	39
2.16.	Qualité des données et sécurité.....	39
3.	DATA CENTERS	40
3.1.	Les plus grands data centers et campus mondiaux	40
3.2.	Classement des pays ayant plus de Data Centers	41
3.3.	Facteurs de domination	41
3.4.	Durée de vie d'un Data Center	42
3.5.	Combien de data centers y a-t-il dans le monde ?	42
3.5.1.	Répartition par pays principaux	42
3.5.2.	Les infrastructures géantes (Hyperscale)	42
3.5.3.	Quels sont les inconvénients des data centers ?	43
3.5.4.	Comment fonctionne un data center ?	43
4.	OLTP	43
4.1.	Caractéristiques de l'OLTP	44
4.2.	Conformité ACID des systèmes OLTP	44
4.3.	Les systèmes OLTP demandent une grande disponibilité	45
4.4.	Exemples de logiciels OLTP	45
4.5.	Domaines d'application	45
5.	OLAP	46
5.1.	Fonctionnement des systèmes OLAP	47
5.2.	Cinq opérations d'analyse OLAP	47
5.3.	Outils décisionnels et serveurs OLAP	48
5.4.	Moteurs de bases de données OLAP modernes (orientés colonnes).....	48
5.5.	Les trois architectures : MOLAP, ROLAP et HOLAP	48
5.6.	Comment choisir selon votre volume de données ?	49
6.	DIFFERENCE ENTRE OLTP ET OLAP	49
6.1.	Comparaison OLTP vs OLAP	49
6.2.	Comparaison, quelques détails	50
6.3.	Comment choisir pour votre projet ?	51
6.3.1.	Bases de données OLTP (Transactionnel) : Outils	51
6.3.2.	Solutions OLAP (Analytique / Data Warehouse) : Outils.....	51
7.	EXEMPLE D'APPLICATIONS	52
7.1.	Exemple OLTP : Gestion d'une vente en direct	52

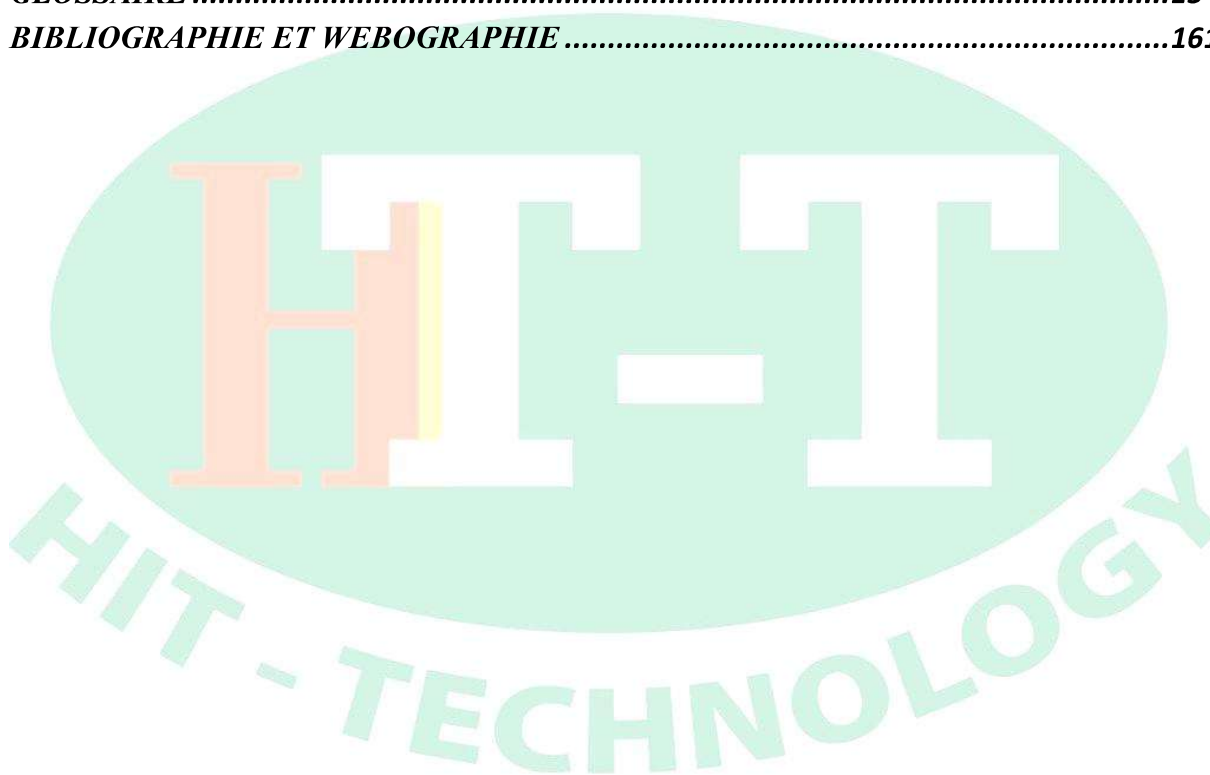
7.1.1.	Structure des tables (Normalisée)	52
7.1.2.	Requête SQL OLTP typique	52
7.2.	Exemple OLAP : Analyse des ventes annuelles	52
7.2.1.	Structure de la table (Dénormalisée)	52
7.2.2.	Requête SQL OLAP typique	52
8.	TRANSFERT DES DONNEES DE OLTP VERS OLAP	53
8.1.	Extraction (Extract)	53
8.2.	Transformation (Transform)	53
8.3.	Chargement (Load)	54
9.	ETL CLASSIQUE VS ELT MODERNE	54
9.1.	Outils ELT Modernes (Cloud & Big Data)	54
9.2.	Plateformes ETL d'Entreprise (Visuelles & Robustes)	54
9.3.	Orchestrations (Gestion des flux)	55
10.	DIFFERENCES DATA LAKE, DW, BDD	55
10.1.	Data Lake : Définition	55
10.1.1.	Data Lake : Qu'est-ce que c'est ?	55
10.1.2.	L'architecture d'un Data Lake	56
10.1.2.1.	Les Zones de Stockage Logiques	56
10.1.2.2.	Les Composants Clés de l'Architecture	57
10.1.2.3.	L'Évolution Moderne : Le Data Lakehouse	57
10.1.3.	Big Data Analytics et les Data Lakes	57
10.1.4.	Les enjeux du Data Lake	58
10.1.5.	Conception d'un Data Lake	58
10.1.5.1.	Écosystème Amazon Web Services (AWS)	58
10.1.5.2.	Écosystème Microsoft Azure	59
10.1.5.3.	Écosystème Google Cloud Platform (GCP)	59
10.1.5.4.	L'Alternative Open-Source (Infrastructures Privées / On-Premise)	59
10.2.	Data Lake, Data Warehouse, Base de données : Quelle est la différence ?	59
10.2.1.	Les Bases de données	60
10.2.2.	Le Data Warehouse	60
10.3.	Le processus de DW	61
10.3.1.	Le Data Lake	63
10.3.2.	Schéma d'un Data Lake	63

10.3.3. Data warehouse vs Data Lake	63
11. AZURE DATA LAKE STORE GEN2	64
12. DE HADOOP A SPAR	66
12.1. Typologie des outils big data	66
12.1.1. Pile logicielle Big data	66
12.1.2. Machine Learning-Définition.....	70
12.1.3. Infrastructure matérielle pour le Big Data.....	72
12.2. Mapreduce et Hadoop.....	72
12.2.1. Architecture d'un cluster	72
12.2.2. Calcul à large échelle	73
13. INSIGHT.....	73
CONCLUSION.....	74
CHAPITRE II : INTELLIGENCE ARTIFICIELLE (IA)	75
INTRODUCTION	76
1. DEFINITION.....	76
2. ORIGINE DE L'INTELLIGENCE ARTIFICIELLE (ChatGPT)	76
3. DOMAINES D'APPLICATION DE L'IA	76
3.1. La reconnaissance de formes	77
3.1.1. Composantes d'un système de RdF.....	79
3.1.2. Fonctionnement et méthodes.....	79
3.1.3. Processus de RdF	80
3.1.4. Etapes de traitement d'un processus de RdF	80
3.1.5. Applications concrètes au quotidien.....	81
3.1.6. Les principaux défis techniques	81
3.2. Le traitement automatique des langues ou NLP	82
3.2.1. Les niveaux d'analyse du langage	82
3.2.2. Les deux grands piliers du TAL.....	83
3.2.2.1. NLU	83
3.2.2.2. NLG	83
3.2.2.3. NLP.....	85
3.2.2.4. Différences entre NLP, NLU et NLG.....	88
3.2.2.5. Les meilleurs outils open-source pour le NLP/NLU/NLG.....	91
3.2.3. Domaines d'application du NLP en résumé	91
3.2.4. Fouille de Texte (Text Mining).....	91

3.2.4.1.	Fonctionnement	92
3.2.4.2.	Usages fréquents	92
3.2.4.3.	Processus de Text Mining: Vue simplifiée	92
3.2.4.4.	Calculs de fréquence des termes.....	93
3.2.4.5.	Domaines d'application.....	94
3.2.4.6.	Etapes de la fouille de textes.....	94
3.3.	Les systèmes experts.....	94
3.3.1.	Représentation schématique d'un système expert	94
3.3.2.	Fonctionnement et composition.....	95
3.3.2.1.	Les faits.....	95
3.3.2.2.	Règles.....	95
3.3.2.3.	Le moteur d'inférence	96
3.3.3.	Pourquoi un SE ?	96
3.3.4.	Les systèmes experts : Schéma de fonctionnement	96
3.3.5.	Mode de raisonnements	97
3.3.5.1.	Chaînage avant	97
3.3.5.2.	Chaînage arrière	98
4.	<i>DIFFERENCE ENTRE L'IA, MACHINE LEARNING ET DEEP LEARNING.....</i>	98
4.1.	Intelligence Artificielle (IA)	98
4.2.	Le Machine Learning	99
4.2.1.	Différents types d'apprentissage en Machine Learning (ML)	99
4.2.1.1.	Différences entre apprentissage supervisé et non supervisé.....	99
4.2.1.2.	Les objectifs et tâches principales.....	100
4.2.1.3.	Analogie simple pour comprendre	100
4.2.1.4.	Algorithmes couramment utilisés	100
4.2.1.5.	Apprentissage par renforcement	101
4.3.	Le Deep Learning	102
4.3.1.	Fonctionnement de base.....	102
4.3.2.	Domaines d'application.....	103
4.3.3.	Les langages et outils	103
4.3.3.1.	Le langage incontournable : Python	103
4.3.3.2.	Les frameworks de Deep Learning.....	103
4.3.3.3.	Un exemple de code minimal (en PyTorch ou Keras)	104

4.4. Différences entre DL et ML	105
4.4.1. L'extraction des caractéristiques (Feature Engineering).....	106
4.4.2. Exemples d'utilisation	106
5. COMMENT S'APPLIQUE L'INTELLIGENCE ARTIFICIELLE AU TRAITEMENT DES DOCUMENTS ?.....	106
6. INTELLIGENCE ARTIFICIELLE : DE QUOI PARLE-T-ON REELLEMENT ? ..	109
6.1. Intelligence Artificielle : de quoi parle-t-on réellement ?	109
6.1.1. Tentative de définition	109
6.1.2. D'où vient cette idée d'Intelligence Artificielle... ..	111
6.1.3. Qui sont les acteurs de cette révolution ?	112
6.1.4. La structuration du marché de l'IA	113
6.2. Un point de bascule de l'intelligence artificielle ?	113
6.2.1. Une puissance de calcul à des coûts abordables.....	113
6.2.2. La disponibilité de la donnée en grand nombre et de qualité	114
6.2.3. L'évolution des « réseaux de neurones »	114
6.2.4. Un effet de seuil sur la recherche	115
6.3. Les limites qui subsistent	116
7. INTELLIGENCE ARTIFICIELLE : QUELLES APPLICATIONS CONCRETES PEUT-ON EN ATTENDRE POUR LES COLLECTIVITES ?	116
7.1. Un contexte national et européen favorable.....	116
7.2. Les premiers grands constats d'utilisation de l'IA par les collectivités à ce jour	117
7.3. FOCUS - L'IA mise en œuvre dans le secteur public	119
7.4. Eléments de réflexion pour les collectivités avant de se lancer dans un projet d'IA...123	
8. MODELES IA	125
8.1. Types de modèles d'IA.....	125
8.2. Exemples populaires de modèles.....	125
9. AGENT IA	125
9.1. Différences clés : Modèle vs Agent.....	125
9.2. Composants principaux d'un agent	126
9.3. Exemples concrets d'utilisation.....	126
10. OCR.....	127
10.1. Fonctionnement de l'OCR moderne.....	127
10.2. OCR Traditionnel vs OCR propulsé par l'IA	127
10.3. Exemples d'outils populaires.....	128

10.4. Applications concrètes	128
11. LLM (Large Language Model)	128
11.1. Fonctionnement et entraînement	128
11.2. Exemples et usages.....	128
CONCLUSION.....	129
CONCLUSION GENERALE.....	130
QUESTIONNAIRE AUX CHOIX MULTIPLES (QCM)	131
REPONSES AU QUESTIONNAIRE	133
RECUEIL DES EXERCICES.....	134
CORRIGES PROPOSES DES EXERCICES	149
PROJETS IA (ETUDE DE CAS).....	153
GLOSSAIRE	154
BIBLIOGRAPHIE ET WEBOGRAPHIE	161



A QUI S'ADRESSE CE LIVRE ?

Ce livre s'adresse aux débutants et professionnels en activités, aux chefs d'entreprises dans les domaines du big data et de l'intelligence artificielle (IA).

PREREQUIS

- Connaissance en informatique et Internet ;
- 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 connaître le domaine de l'intelligence artificielle, un domaine en pleine expansion et le big data.

La demande de formation dans les domaines de l'intelligence artificielle et le big data étant sans cesse croissante, nous avons jugé utile de mettre à usage des apprenants ce livre afin de permettre aux débutants et aux professionnels du domaine de pouvoir répondre aux attentes du terrain.

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 les généralités sur l'IA ;
- Discuter les différents types de l'IA ;
- Présenter les différents modèles d'IA existants ;
- Présenter les opportunités offertes par l'IA ;
- Illustrer des exemples d'utilisation de l'IA ;
- 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.

