

Sigle : INF6143 Gr. 01**Titre : Bases de données avancées****Session : Automne 2021 Horaire et local****Professeur : Missaoui, Rokia****1. Description du cours paraissant à l'annuaire :****Objectifs**

Permettre aux étudiants de maîtriser les connaissances sur les fondements, concepts et problèmes reliés aux bases de données allant des bases de données conventionnelles (incluant les bases de données réparties) aux bases de données plus avancées comme les entrepôts de données et les bases multimédia (incluant les systèmes d'information géographique et les bases documentaires).

Contenu

Rappels sur les bases de données (BD). Contrôle et optimisation des performances dans un environnement centralisé. Bases de données réparties : principes, stratégies de conception, traitement des requêtes réparties, et gestion des transactions réparties. Veille économique (*business intelligence*) : fouille et entreposage de données. BD multimédia (particularités et exigences, stockage et exploitation, systèmes d'information géographique. BD documentaires). BD et Web (connexion à une BD via le Web, langage XML).

Descriptif – Annuaire

2. Objectifs spécifiques du cours :

Ce cours est destiné aux étudiant(e)s ayant suivi au préalable un ou des cours de bases de données au niveau baccalauréat et/ou qui ont une expérience pertinente en conception et utilisation des bases de données.

Ce cours vise à approfondir les connaissances de l'étudiant(e) sur des thèmes avancés en bases de données tels les bases de données réparties, la fouille et l'entreposage de données (*data mining et data warehousing*), les bases de données multimédia et les BD sur le Web.

Ce cours a comme objectifs plus particuliers :

- d'initier l'étudiant(e) aux fondements, concepts et problèmes reliés aux BD;
- de permettre à l'étudiant(e) d'approfondir, à travers les travaux de session, quelques aspects des BD avancées;
- d'initier l'étudiant(e) à un travail de recherche à travers la préparation de synthèses d'articles et de présentations orales, et la rédaction d'un rapport de session.

3. Stratégies pédagogiques :

Il s'agit d'un cours magistral en présentiel avec travaux individuels et travaux d'équipe.

4. Heures de disponibilité ou modalités pour rendez-vous :

Bureau : B-2032, téléphone : 819 595-3900, poste 1709

Courriel : rokia.missaoui@uqo.ca

Page Web : <http://w3.uqo.ca/missaoui>

Consultation : Vendredi, de 15 h 30 à 17 h 30

5. Plan détaillé du cours sur 15 semaines :

Semaine	Thèmes	Dates
1	Présentation du cours, entente d'évaluation, ...	10 sept. 2021

2	<p>Quelques rappels :</p> <ul style="list-style-type: none"> • fonctionnalités des systèmes de gestion de BD (SGBD); • conception des bases de données; • modèle relationnel. 	17 sept. 2021
3	<p>Principes des bases de données réparties (BDR) :</p> <ul style="list-style-type: none"> • objectifs; • nouveaux défis; • formes de compatibilité; • architecture ANSI/SPARC étendue. <p>Modèles architecturaux des BDR selon les trois dimensions :</p> <ul style="list-style-type: none"> • répartition; • hétérogénéité; • autonomie. <p>Architecture des systèmes répartis :</p> <ul style="list-style-type: none"> • modèle client-serveur; • modèle poste à poste (<i>peer-to-peer</i>). <p><i>Formation des équipes</i></p>	24 sept. 2021
4	<p>Conception et catégories des BDR :</p> <ul style="list-style-type: none"> • approche ascendante versus descendante; • fragmentation (horizontale, verticale et mixte); • allocation : principe et modèle. <p><i>Synthèse d'article : 10 points</i></p>	01 oct. 2021
5	<p>Veille économique (<i>Business Intelligence</i>) : raison d'être, techniques et défis :</p> <ul style="list-style-type: none"> • fouille de données (<i>data mining</i>); • entreposage de données (<i>data warehousing</i>). 	08 oct. 2021
6	Semaine d'études	15 oct. 2021
7	<p>Veille économique (<i>suite</i>) :</p> <ul style="list-style-type: none"> • fouille de données (<i>data mining</i>) <p><i>Synthèse d'article : 10 points</i></p>	22 oct. 2021
8	<p>Veille économique (<i>suite</i>) :</p> <ul style="list-style-type: none"> • entreposage de données (<i>data warehousing</i>) 	29 oct. 2021
9	<p>Optimisation des requêtes dans les BD centralisées :</p> <ul style="list-style-type: none"> • optimisation des sélections, projections et jointures; • optimisation dans SQL Server et Oracle. 	05 nov. 2021

10	<p>Optimisation des requêtes dans les BDR :</p> <ul style="list-style-type: none"> • décomposition des requêtes; • localisation des données; • traitement local versus global; • semi-jointure. <p><i>Ébauche du projet de session</i></p>	12 nov. 2021
11	<p>BD multimédia :</p> <ul style="list-style-type: none"> • texte, image, audio et vidéo; • particularités et exigences; • stockage et exploitation. 	19 nov. 2021
12	<p><i>BD et Web :</i></p> <ul style="list-style-type: none"> • BD XML : description et exploitation <p><i>Développement d'applications : 10 points</i></p>	26 nov. 2021
13	<p><i>Exposés</i></p> <p><i>Présentations orales : 15 points</i></p>	03 déc. 2021
14	<p><i>Exposés</i></p> <p><i>Présentations orales</i></p>	10 déc. 2021
15	<p>Examen : 30 points</p> <p>Rapport final : 25 points</p>	17 déc. 2021

6. Évaluation du cours :

Synthèse et critique de **deux articles** (*travail individuel*, 4^e et 7^e semaines) **2 * 10 points**

La synthèse ne doit pas dépasser un maximum de 1000 mots. De même, les normes de présentation de travaux (p. ex. : page de garde, marge d'un pouce, interligne à 1,5, taille des caractères de 12 points) doivent être **absolument** respectées. Tout travail non conforme aux consignes se verra attribuer une pénalité de 10 % de la note obtenue.

Développement d'applications (12^e semaine) **10 points**

Examen (15^e semaine) **30 points**

Projet de session (*en équipe de deux si le nombre d'étudiants inscrits dépasse 8*)

Le projet de session peut être soit théorique (synthèse d'articles), soit pratique (conception et implantation de programmes). Il comprend les deux composantes suivantes :

- **Ébauche du projet de session** (10^e semaine);
- **Présentations orales** (13^e et 14^e semaines) **15 points**;
- **Rapport final** (15^e semaine) **25 points**.

Voici quelques suggestions de sujets de projets de session :

- Gestion de données massives (*Big Data Management*)
- Fouille de données massives (*Big Data Analytics*)
- Web sémantique avec les graphes RDF et le langage SPARQL
- Infonuagique (*Cloud Computing*) avec Microsoft, Google ou Amazon
- Traitement parallèle (p. ex. MapReduce & Hadoop)

- Systèmes NoSQL
- Techniques avancées d'indexation et de stockage de données
- Analyse et fouille de réseaux sociaux (détection de communauté, prédiction de liens, déstabilisation de réseaux)
- Algorithmes d'ordonnancement du Web (p. ex. HITS, PageRank, ...)
- Optimisation des performances dans les entrepôts de données
- Fouille répartie des données (*Distributed Data Mining*)
- Interactions entre les BD et le Web
- BD XML (XQuery, Xpath, XSLT, etc.); BD actives
- Stockage et recherche de données spatiales
- Bases de données temporelles,
- Traitement de l'imprécision dans les bases de données, etc.

Les consignes de réalisation des divers travaux feront l'objet d'un document à part.

Tout retard dans la remise d'un travail entraîne une pénalité de **5 %** par jour sur la note attribuée à ce travail.

À la discrétion du professeur, une pénalité maximale de **10 %** de la note globale pourrait affecter la note de session si la professeure observe des absences fréquentes ou une passivité injustifiée et évidente au cours.

La qualité du français sera considérée lors de la correction des travaux.

7. Politiques départementales et institutionnelles :

- Politique du département d'informatique et d'ingénierie relative à la tenue des examens
- Note sur le plagiat et sur la fraude
- Absence aux examens : cadre de gestion, demande de reprise d'examen (formulaire)

À l'UQO, **les violences à caractère sexuel, c'est tolérance zéro!**

La communauté universitaire s'engage à lutter contre les inconduites, le harcèlement et les violences à caractère sexuel : parce que **le respect, c'est l'affaire de tout le monde!**

N'oubliez pas de faire la formation obligatoire :

uqo.ca/bimi/formation-obligatoire

Pour de plus amples renseignements :

bimi@uqo.ca



8. Principales références :

NOTES DE COURS

- Disponibles via MOODLE

LIVRE RECOMMANDÉ

- Elmasri, R. & Navathe S. B., *Fundamentals of Database Systems*, 7th edition, Pearson Education, 2016.

MONOGRAPHIE

- Atzeni, P., Ceri, S., Paraboschi, S., Torlone, R. *Database Systems – Concepts, Languages and Architectures*. McGraw-Hill, 2000.
- Abiteboul, S., Buneman, P. & Suci, D. *Data on the Web*, Morgan Kaufmann, 2000.
- Carlo Batini & Monica Scannapieco. *Data and Information Quality – Dimensions. Principles and Techniques*, Springer, 2016.
- Connolly, T. & Begg, C. *Database Systems: A Practical Approach to Design, Implementation and Management*, 6th edition, Pearson Education, 2015.
- Coulouris, G., Dollimore, J. & Kindberg, T. *Distributed Systems*, 4th edition, Addison-Wesley, 2005.
- Date, C.J. *An Introduction to Database Systems*, 8th edition, Pearson Education, 2004.
- Dunckley, L. *Multimedia Databases: An Object-Relational Approach*, 1st edition, Addison-Wesley, 2003.
- Garcia-Molina, H., Ullman, H.J. & Widom, J. *Database Systems – The Complete Book*, 2nd edition Prentice-Hall, 2009.
- Godin, R. *Systèmes de gestion de bases de données par l'exemple*, 3^e édition, Loze-Dion, 2015.
- Golfarelli, M. & Rizzi, S. *Data Warehouse Design: Modern Principles and Methodologies*, McGraw-Hill, 2009.
- Gray, J. and Reuter, A. *Transaction Processing: Concepts and Techniques*, Morgan Kaufmann Series in Data Management Systems, Jim Gray Series Editor, 1993.
- Han, J., Kamber, M. & Pei, J. *Data Mining: Concepts and Techniques*, Morgan Kaufmann. 3rd edition, 2011.
- Inmon, W.H. & Linstedt, D. *Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault*, Morgan Kaufmann, 2015.
- Kimball, R. & Ross, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 2013.
- Lewis, P.M., Bernstein, A. & Kifer, M. *Databases and Transaction Processing*, Addison-Wesley, 2002.
- Liu, L. and Özsu, M.T. Eds. *Encyclopedia of Database Systems*, Springer US, 2009.
- Özsu, M. T. & Valduriez, P. *Principles of Distributed Database Systems*, 4th Edition. Springer 2020, ISBN 978-3-030-26252-5, pp. 1-674.
- Vaisman, A. & Zimanyi, E. *Data Warehouse Systems – Design and Implementation*, Springer, 2014.
- Witten, H. & Frank, E., Hall, M.A. & Pal, C.J. *Data Mining: Practical Machine Learning Tools and Techniques*, 4th edition, Morgan Kaufmann, 2017.
- Zaniolo, C. Ceri, S., Faloutsos, C., Snodgrass, R.T., Subrahmanian & V.S., Zicari, R. *Advanced Database Systems*, The Morgan Kaufmann Series in Data Management Systems, 1997.
- Milos Radivojevic, M., Sarka, D., Durkin, W., Cote, C. & Lah, M. *Mastering SQL Server 2017: Build smart and efficient database applications for your organization with SQL Server*, Packt Publishing, 2019.

REVUES

- Communications of ACM, Journal of ACM
- ACM SIGMOD Record
- Transactions on Database Systems (TODS), ACM
- Transactions on Information Systems (TIS), ACM
- IEEE Database Engineering Bulletin, IEEE Computer
- IEEE Transactions on Knowledge and Data Engineering

ACTES DE CONFÉRENCE

- Proceedings of ACM International Conference on Management of Data (SIGMOD)
- Proceedings of ACM Symposium on Principles of Database Systems (PODS)
- Proceedings of Very Large Data Base (VLDB) Conference
- Proceedings of the IEEE International Conference on Data Engineering
- Proceedings of the International Conference on Extending Database Technology (EDBT)
- Proceeding of ACM International Conference on Knowledge Discovery from Databases (KDD)

QUELQUES SITES WEB UTILES

- <http://qdt.oqlf.gouv.qc.ca/index.aspx> (Grand dictionnaire terminologique de l'Office québécois de la langue française)
- <https://www.linguee.fr/> dictionnaire Linguee anglais-français
- <https://www.btb.termiumplus.gc.ca/tpv2alpha/alpha-fra.html?lang=fra> Termium Plus
- <http://www.sigmod.org/> (Groupe d'intérêt SIGMOD de l'ACM)

- <http://db.uwaterloo.ca/~tozsu/> (site de Prof. Tamer Özsu)
- <http://hanj.cs.illinois.edu/> (site de Prof. Jiawei Han)
- http://www.labunix.uqam.ca/~godin_r/SiteWebGodin2012/index.html (site de Prof. R. Godin)
- <http://scholar.google.ca/> (*Google scholar*)
- <https://dl.acm.org/> (*ACM Digital Library*)
- <http://www.informatik.uni-trier.de/~ley/db/> (*The DBLP Computer Science Bibliography*)

9. Page Web du cours :

<https://moodle.uqo.ca>