

Sigle : INF1473 Gr. 01

Titre : Entreposage et prospection de données

Session : Hiver 2021 Horaire et local

Professeur : Yapi, N'Dah Daniel

1. Description du cours paraissant à l'annuaire :

Objectifs

L'objectif du cours est de familiariser, par la pratique, l'étudiante, l'étudiant avec les techniques de recherche, traitement et diffusion de l'information et des connaissances au sein de l'entreprise en vue de la prise de décision.

Contenu

Entreposage de données : étapes de construction d'un entrepôt de données (acquisition, stockage, traitement et accès), modélisation multidimensionnelle des données et cubes de données, techniques OLAP, types d'architectures des entrepôts de données, optimisation des performances, produits et applications. Prospection de données : étapes de découverte de connaissances (prétraitement, prospection de données et interprétation des résultats), techniques de classification (arbres de décision, etc.), techniques de regroupement (treillis de concepts, classification hiérarchique), règles d'association et mesures de qualité, techniques statistiques d'analyse de données, produits et applications. Ce cours comporte des séances obligatoires de travaux dirigés (TD) de deux heures par semaine.

Descriptif – Annuaire

2. Objectifs spécifiques du cours :

- Familiariser l'étudiant(e) avec les techniques *Business Intelligence*.
- Présenter les principales techniques de prétraitement, prospection et entreposage de données.
- Illustrer ces techniques à l'aide du système SQL Server 2019 et particulièrement les modules SSIS (*SQL Server Integration Services*) et SSAS (*SQL Server Analysis Services*).
- Permettre à l'étudiant(e) de mettre en pratique toutes les connaissances et techniques acquises durant le cours via des exercices et un projet de session portant sur le prétraitement, la prospection et l'entreposage de données.

3. Stratégies pédagogiques :

- *Cours magistraux en non-présentiel avec ZOOM.*
- *Séances de travaux dirigés en non-présentiel.*

Liens et guides utiles :

- Guide d'utilisation de Zoom à l'intention des étudiants.
- Site pour soutien de réussite en mode non-présentiel : uqo.ca/etudier-non-presentiel.

Logistique du cours

- Le logiciel requis pour le cours et les travaux dirigés est le Client SQL Server 2019; il sera accessible à distance via le laboratoire virtuel de l'UQO : <https://uqo.ca/docs/39679>
- Accès à Moodle sur le Web pour la récupération des notes de cours, des énoncés de travaux, des consignes spécifiques et des résultats d'évaluation.

Outils : Les étudiant(e)s qui s'inscrivent à ce cours doivent s'assurer qu'ils ont un ordinateur; une connexion Internet; une webcam; un microphone; et un scanner pour numériser leurs examens.

Plan synthétisé du cours

Les thèmes suivants seront étudiés :

- Introduction à la veille économique (*Business intelligence*) et à la prospection de données (*Data mining*)
- Techniques de prétraitement de données : 1- *Épuration*. 2 - *Intégration et transformation*. 3 - *Sélection et réduction*
- Entreposage de données : raison d'être et concepts
- Étapes de construction d'un entrepôt de données
- Modélisation multidimensionnelle
- Stratégies de conception
- Techniques OLAP
- Architectures des entrepôts de données
- Optimisation des performances
- Applications et produits
- Découverte de connaissances : raison d'être et définitions
- Étapes du processus de découverte de connaissances
- Survol des techniques de prospection de données et des applications courantes
- Classification et prédiction : 1- *Définitions, principaux thèmes*. 2 - *Comparaison entre la classification et la prédiction*. 3 - *Arbres de décision et règles de classification*. 4 - *Réseaux bayésiens*. 5 - *Approches statistiques de prédiction (modèles de régression)*
- Regroupement : méthodes hiérarchiques comme *K-Means*, analyse formelle de concepts
- Règles d'association et mesures de qualité

NOTE : Toutes les parties du cours seront illustrées avec les outils Business Intelligence de SQL Server 2019.

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

Sur demande. L'étudiant(e) peut m'envoyer un courriel pour fixer un rendez-vous.

Courriel : ndahdaniel.yapi@uqo.ca

5. Plan détaillé du cours sur 15 semaines :

Semaine	Thèmes	Dates
1	Business Intelligence (Intelligence d'affaires) SQL Server 2019 : Survol des composantes	14 janv. 2021
2	Prétraitement de données (Épuration, intégration, transformation, sélection et réduction de données) Travaux dirigés 1 : Prétraitement et données (vendredi 22 janvier 2021)	21 janv. 2021
3	Construction d'un entrepôt de données (Data Warehousing) Travaux dirigés 2 : Prétraitement de données – SSIS (vendredi 29 janvier 2021)	28 janv. 2021

4	<p>Modélisation multidimensionnelle et stratégies de conception</p> <p>Travail à remettre : 1. <i>Prétraitement de données</i> – Pondération : 15 %</p> <p>Travaux dirigés 3 : Modélisation de données (vendredi 05 février 2021)</p>	04 févr. 2021
5	<p>Modélisation multidimensionnelle et stratégies de conception</p> <p>Travaux dirigés 4 : Modélisation de données (vendredi 12 février 2021)</p>	11 févr. 2021
6	<p>Techniques OLAP et optimisation</p> <p>Travaux dirigés 5 : Création des cubes de données – SSAS (vendredi 19 février 2021)</p>	18 févr. 2021
7	<p>Examen de mi-session – En non-présentiel. Pondération : 25 %</p> <p>Travail à remettre : 2.1. <i>Conception d'un entrepôt de données</i> – Pondération : 10 %</p> <p>Travaux dirigés 6 : Chargement et manipulation des entrepôts de données – SSAS (vendredi 26 février 2021)</p>	25 févr. 2021
8	<p>Semaine d'études</p>	04 mars 2021
9	<p>Produits et applications</p> <p>Travaux dirigés 7 : Création des cubes de données – SSAS (vendredi 12 mars 2021)</p>	11 mars 2021
10	<p>Découverte de connaissances (<i>Data Mining</i>)</p> <p>Travaux dirigés 8 : Manipulation des cubes de données – SSAS (vendredi 19 mars 2021)</p>	18 mars 2021
11	<p>Classification</p> <p>Travail à remettre : 2.2 <i>Manipulation d'un entrepôt de données</i> – Pondération : 5 %</p> <p>Travaux dirigés 9 : Fouille de données – arbres de décision SSAS (vendredi 26 mars 2021)</p>	25 mars 2021
12	<p>Regroupement</p> <p>Travaux dirigés 10 : Fouille de données – arbres de décision et regroupement – SSAS</p> <p>Vendredi 02 avril (jour férié – Vendredi de Pâques)</p> <p>Travaux dirigés déplacés au vendredi 09 avril en raison du Vendredi de Pâques.</p>	01 avril 2021

13	Règles d'association Travail à remettre : <i>3.1 Construction d'un modèle de classification</i> – Pondération : 7 % et de règles d'associations – Pondération : 8 %	08 avril 2021
14	Révision et discussion générale sur le cours et les perspectives de recherche en intelligence d'affaires Préparation pour l'examen final Travaux dirigés 11 : Fouille de données – Règles d'associations (vendredi 16 avril 2021)	15 avril 2021
15	Examen final – En non-présentiel. Pondération : 30 % Remise des derniers travaux	22 avril 2021

6. Évaluation du cours :

L'étudiant(e) dans ce cours sera évalué(e) par les examens de mi-session et final, ainsi que par des travaux pratiques. La pondération de la note finale sera comme suit :

- Examen de mi-session : 25 %
- Examen final : 30 %
- Travaux pratiques : 45 %

Les travaux pratiques comprendront les volets suivants : prétraitement, entreposage et prospection de données.

Une moyenne générale inférieure à **50 %** est éliminatoire et conduit automatiquement à l'échec de l'étudiant(e). Les travaux pratiques se feront par équipes de deux. La pénalité de retard pour la remise d'un travail est de **2 points** par jour (y compris les jours fériés et les fins de semaine).

Les présences aux séances de cours et de travaux dirigés seront considérées. Un(e) étudiant(e) qui s'absente souvent et de manière injustifiée aura une diminution de la note finale d'un maximum de **5 points**.

Des consignes sur l'échéancier et la réalisation des travaux pratiques seront précisées.

Des consultations de groupes seront organisées sur rendez-vous afin de guider et d'orienter les étudiant(e)s dans la réalisation de leurs 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
- Politique relative à la qualité de l'expression française écrite chez les étudiants et les étudiantes de premier cycle à l'UQQ
- 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 :

1. **Notes de cours disponibles sur Moodle (principale référence)**
2. Kimball, Ralph et al. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, Ind., Wiley, 2013.
3. Burquier, Bertrand. *Business intelligence avec SQL Server 2005*. Mise en œuvre d'un projet décisionnel, Dunod, 2007.
4. J. Han & M. Kamber. *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2nd edition, 2006.
5. D. J. Hand, H. Mannila and P. Smyth. *Principles of Data Mining*, MIT Press, 2001.
6. W.H. Inmon. *Building the Data Warehouse*, John Wiley, 3^e édition, 2002.
7. R. Kimball & M. Ross. *Guide pratique de modélisation dimensionnelle*, Vuibert informatique, Paris, 2002.
8. R. Kimball, L. Reeves, M. Ross et W. Thornthwaite. *Le Data warehouse*, collection Blanche, 2005.
9. R. Lefébure & G. Venturi. *Le Data Mining*. Eyrolles, 2001.
10. W. R. Stanek. *SQL Server 2005 : guide de l'administrateur*, Microsoft Press, 2006.
11. Z. Tang & J. MacLennan. *Data Mining with SQL Server 2005*, Wiley, 2005.
12. S. Tufféry. *Data mining et statistique décisionnelle – L'intelligence des données*, éditions TECHNIP, 2012.
13. I. H. Witten & E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd edition, Morgan Kaufmann, 2005.

9. Page Web du cours :

<https://moodle.uqo.ca>