

Technologies de l'information

Technologies de l'intelligence d'affaires - 6-701-12(Public)

A2016

Groupe J01

Enseignant(s)

Gregory Vial
Professeur(e) adjoint(e)
gregory.vial@hec.ca
Bureau : 4.368
Disponibilité : Sur rendez-vous
Inclure la mention "6-701" dans le titre du courriel.

Secrétaire(s)

Antonietta Florio
Secrétaire
antonietta.florio@hec.ca
514-340-6474
Bureau : 4.624

Présentation du cours

Objectifs



L'intelligence d'affaires vise à extraire de l'information à partir des données saisies par les systèmes opérationnels de l'entreprise. Pour être en mesure de déceler cette information cachée il faut premièrement enregistrer les données de bases dans des structures appropriées aux entrepôts de données. Ensuite, il faut analyser ces données à l'aide de techniques statistiques appropriées et de logiciels spécialisés qui aideront l'analyste dans sa démarche.

Ce cours porte sur les technologies qui supportent l'intégration et l'exploitation des données par les analystes en intelligence d'affaires, en se concentrant notamment sur les bases de données relationnelles. L'étudiant apprendra à bien comprendre, à manipuler et à extraire les données qui se trouvent dans les bases de données opérationnelles et dans les entrepôts de données. Pour cela, il devra comprendre les structures de données de ces environnements ainsi que les processus qui les gèrent. Il devra aussi comprendre les enjeux reliés à la gestion de ces technologies et processus.

Matériel pédagogique

Ressources bibliographiques



Kimball, Ralph et al. (2013). *The data warehouse toolkit the definitive guide to dimensional modeling, third edition*, Indianapolis, Ind., Wiley.
ISBN:9781118530771

 [Disponible à la bibliothèque](#)

Ressources bibliographiques complémentaires




IMPORTANT

William Yeoh (2010-04-01). « CRITICAL SUCCESS FACTORS FOR BUSINESS INTELLIGENCE SYSTEMS », *The Journal of Computer Information Systems*, vol.50, no.3.

 [Disponible à la bibliothèque](#)



Hughes, Ralph (2013). *Agile data warehousing project management business intelligence systems using Scrum*, Waltham, MA, Morgan Kaufmann.
ISBN:9780123965172

 [Disponible à la bibliothèque](#)



Chiang, Roger H. L (2012). « Business intelligence and analytics », *Management information systems*, vol.36, no.4.

 [Disponible à la bibliothèque](#)

Évaluations

Sommaire des évaluations

Travail individuel	40 %
Travail en équipe	20 %
Participation	5 %
Examen final	35 %

Voir *HEC en ligne* pour date

Travail individuel (40 %)

Individuel / À la maison - Hors classe / Écrit

Description

**IMPORTANT**

L'étudiant devra réaliser une série d'exercices à la maison afin d'appliquer les concepts vus en cours. Les exercices seront postés dans la rubrique Remise de Travaux et devront être remis via ZoneCours.

Le découpage des devoirs (tentatif) est le suivant (le nombre de devoirs ne changera pas mais la répartition des notes peut changer):

- 3 devoirs SQL (5%; 7.5%; 7.5%): 20%
- 1 devoir SSIS: 20%

Travail en équipe (20 %)

En équipe

Participation (5 %)

Individuel / En classe / Oral

Description**IMPORTANT**

La participation en classe représente un aspect important du cours permettant d'approfondir certains aspects de la matière.

Les étudiants sont invités à participer de façon active afin de s'assurer de leur bonne acquisition du matériel et de faire profiter leurs collègues de leurs expériences dans le domaine.

Examen final (35 %)

Voir *HEC en ligne* pour date

Individuel / En classe

Organisation du cours

1 - L'intelligence d'affaires dans les organisations et les entrepôts de données**Description**

**IMPORTANT**

Introduction au cours, présentation de l'intelligence d'affaires et de son utilisation dans les organisations. Présentation de définitions importantes. Mise en contexte de l'utilisation de l'intelligence d'affaires et des défis qui y sont associés. Distinction entre les systèmes transactionnels, bases de données opérationnelles et les systèmes d'intelligence d'affaires et les entrepôts de données. Dénormalisation des données.

Activités/Ressources avant la séance**IMPORTANT**

Kimball, Ralph et al. (2013). *The data warehouse toolkit the definitive guide to dimensional modeling, third edition*, Indianapolis, Ind., Wiley.
ISBN:9781118530771

Chapitre 1

 [Disponible à la bibliothèque](#)

**IMPORTANT**

Watson, H.J (2007). « The Current State of Business Intelligence », *Computer*, vol.40, no.9.

 [Disponible à la bibliothèque](#)

**IMPORTANT**

Chaudhuri, Surajit (2011-08-01). « An overview of business intelligence technology », *Communications of the ACM*, vol.54, no.8.

 [Disponible à la bibliothèque](#)

**IMPORTANT**

William Yeoh (2010-04-01). « CRITICAL SUCCESS FACTORS FOR BUSINESS INTELLIGENCE SYSTEMS », *The Journal of Computer Information Systems*, vol.50, no.3.

 [Disponible à la bibliothèque](#)

2 - Modélisation de systèmes transactionnels et d'intelligence d'affaires**Description**

**IMPORTANT**

Introduction à la modélisation conceptuelle et logique à des fins opérationnelle et dimensionnelle.
Distinction entre les besoins de modélisation à des fins transactionnelles et d'intelligence d'affaires (distinction entre les schémas en flocon et en étoile) et les formalismes qui leur sont associés.
Introduction au cas Adventureworks utilisé pour l'ensemble du cours.

Activités/Ressources avant la séanceEXERCICE

Après avoir fait les lectures, construisez deux modèles pour la dimension de produits d'AdventureWorks: Un modèle conceptuel et un modèle logique.

En cours, deux étudiants seront sélectionnés au hasard pour présenter et comparer leurs solutions.

**IMPORTANT**

Kimball, Ralph et al. (2013). *The data warehouse toolkit the definitive guide to dimensional modeling, third edition*, Indianapolis, Ind., Wiley.

ISBN:9781118530771

- Chapitres 2 & 3
- Lire le chapitre 5: pages 147 à 164 (Slowly Changing Dimensions)

 [Disponible à la bibliothèque](#)

**IMPORTANT**

Mundy, Joy et al. (2011). *The Microsoft data warehouse toolkit with SQL server 2008 R2 and the Microsoft Business Intelligence toolset*, Indianapolis, IN, Wiley.

ISBN:0470640383

Lire le chapitre 2

 [Disponible à la coop HEC](#)

 [Cliquer ici](#)

3 - Gestion et manipulation d'un modèle physique de bases de données**Description****IMPORTANT**

Présentation du langage SQL et inscription dans le contexte de la structuration de données. Création de requêtes simples pour faire des opérations de manipulation de données (DML) et de définition de structure de données (DDL) à travers différentes formes de requêtes et l'utilisation de jointures. Compréhension de la documentation du T-SQL et techniques de debugging de requêtes SQL.

Activités/Ressources avant la séance



IMPORTANT

[Query Fundamentals - SQL Server 2008 R2](http://msdn.microsoft.com/en-us/library/ms190659(v=sql.105).aspx) [Site web]

([http://msdn.microsoft.com/en-us/library/ms190659\(v=sql.105\).aspx](http://msdn.microsoft.com/en-us/library/ms190659(v=sql.105).aspx))

Lire la section Query Fundamentals et ses sous-sections: Parts of a SELECT statement; Using the select list; Using the FROM clause; Filtering rows by using WHERE and HAVING; Sorting rows with ORDER BY; Subquery fundamentals; Join fundamentals.



IMPORTANT

[Manipulation de données dans SQL Server 2008 R2](http://msdn.microsoft.com/en-us/library/ms181096(v=sql.105).aspx) [Site web]

([http://msdn.microsoft.com/en-us/library/ms181096\(v=sql.105\).aspx](http://msdn.microsoft.com/en-us/library/ms181096(v=sql.105).aspx))

Lire les sections suivantes: Inserting data into a table, Updating data in a table; Deleting data in a table; Inserting, updating, and deleting data with MERGE.

4 - Requêtes avancées pour manipuler un modèle physique de base de données

Description



IMPORTANT

Groupement, filtrage et ordonnancement avancé de résultats de requêtes. Création de vues logiques par-dessus un modèle physique. Mise en contexte avec le profilage de données et les problèmes associés à la création de requêtes complexes dans des systèmes transactionnels afin de mieux comprendre l'utilité des entrepôts de données utilisés dans les systèmes d'intelligence d'affaires.

5 - Utilisation de requêtes avancées à des fins analytiques

Description



IMPORTANT

Utilisation de fonctions de groupement avancées. Introduction aux Common Table Expressions (CTE) et utilisation dans le contexte du cas Adventureworks. Introduction aux opérateurs de pivot. Utilisation de ces fonctions dans des scénarios où les fonctions traditionnelles du langage SQL ne sont pas adaptées (opérations récursives etc.). Mise en contexte dans le cadre de problèmes associés à l'intelligence d'affaires.

6 - Programmation SQL et performance SQL

Description



IMPORTANT

Comprendre et implémenter les concepts et techniques de programmation dans un DBMS (stored procedures, fonctions) et introduction aux concepts de performance dans les bases de données relationnelles. Présentation et debugging de plans d'exécution de requêtes, optimisation de requêtes sur Adventureworks et mise en relation avec des modèles relationnels implémentés dans des entrepôts de données.

7 - Gouvernance, qualité et profilage de données

Description



IMPORTANT

Présentation des enjeux de gouvernance de données et mise en perspective dans le cadre de la conception de solutions d'intelligence d'affaires.

Présentation du profilage de données (techniques, enjeux, utilité dans le cadre d'un projet d'intelligence d'affaires).

Utilisation des connaissances acquises en SQL pour exécuter des opérations de profilage sur les données sources utilisées dans l'alimentation et la conception d'un entrepôt de données.

Activités/Ressources avant la séance



IMPORTANT

Tallon, P. P (2013). « Corporate Governance of Big Data: Perspectives on Value, Risk, and Cost », *Computer*, vol.46, no.6.

 [Disponible à la bibliothèque](#)

**IMPORTANT**

Khatri, Vijay (2010-01-01). « Designing data governance », *Communications of the ACM*, vol.53, no.1.

 [Disponible à la bibliothèque](#)

**IMPORTANT**

Abedjan, Ziawasch (2015-08-01). « Profiling relational data: a survey », *The VLDB Journal*, vol.24, no.4.

 [Disponible à la bibliothèque](#)

8 - Intégration de données: Conception de processus ETL #1

Description

**IMPORTANT**

Introduction à l'intégration de données. Présentation des enjeux et de l'ubiquité des enjeux d'intégration. Mise en contexte dans le cadre de la conception d'une solution d'intelligence d'affaires.

Introduction aux grandes étapes du processus ETL utilisé pour alimenter les entrepôts de données et intégrer différentes sources de données. Mise en contexte du processus dans le cadre de l'implantation d'un système d'intelligence d'affaires.

Première implémentation d'un processus ETL à l'aide de Microsoft SSIS.

Activités/Ressources avant la séance

**IMPORTANT**

Sur la VM, dans le répertoire C:\TIA\Videos SSIS, visionner et reproduire les exemples dans l'ordre suivant:

1. My First Integration Services Solution (<https://code.msdn.microsoft.com/windowsdesktop/My-First-Integration-fa41c0b1>)
2. SSIS Control Flow Basics (<https://code.msdn.microsoft.com/windowsdesktop/SSIS-Control-Flow-Basics-16bd36d0>)
3. SSIS Data Flow Basics (<https://code.msdn.microsoft.com/windowsdesktop/SSIS-Data-Flow-Basics-f429f5f1>)


Ces exercices vous permettront de vous familiariser avec SSIS. Prendre le temps de les faire. Au besoin les vidéos peuvent être visionnées séparément en suivant les liens ci-dessus.

**IMPORTANT**

Kimball, Ralph et al. (2013). *The data warehouse toolkit the definitive guide to dimensional modeling, third edition*, Indianapolis, Ind., Wiley.

ISBN:9781118530771

- Chapitre 5: description des types de dimensions (cf. le résumé p. 164)

 [Disponible à la bibliothèque](#)

**IMPORTANT**

Dayal, U., Castellanos, M., Simitsis, A., Wilkinson, K. (2009). « Data integration flows for business intelligence », *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*.

 [Version libre de droit en ligne](#)

**IMPORTANT**

Vassiliadis, P., Simitsis, A., Skiadopoulos, S. (2002). « Conceptual modeling for ETL processes », *Proceedings of the 5th ACM international workshop on Data Warehousing and OLAP*.

 [Version libre de droit en ligne](#)

9 - Intégration de données: Conception de processus ETL #2

Description

**IMPORTANT**

Continuation de l'implémentation du processus ETL centrée sur la population d'une table de fait.

Intégration de données externes.

Discussion sur les enjeux d'intégration de données dans contexte de collaboration entre partenaires et exploitation de données ouvertes ou non propriétaires.

10 - Intégration de données: Conception de processus ETL #3

Description

**IMPORTANT**

Continuation de l'implémentation du processus ETL centrée sur l'intégration de données hétérogènes dans un entrepôt de données.

11 - Visualisation des données: Théorie et pratique #1

Description



IMPORTANT

Discussion sur la visualisation des données, les enjeux et techniques couramment utilisées dans le domaine. Fondements cognitifs et théoriques justifiant l'utilisation de techniques de visualisation appropriées. Présentation des patterns les plus fréquemment utilisés ainsi que des écueils à éviter.

Expérimentation pratique avec Tableau Software.

Activités/Ressources avant la séance



Avant la séance, les étudiants doivent avoir installé Tableau Software. Les détails des licences seront fournis par l'enseignant avant la séance.



IMPORTANT

[Guides de démarrage Tableau](http://www.tableau.com/support/manuals/quickstart) [Site web]
(<http://www.tableau.com/support/manuals/quickstart>)



IMPORTANT

[Tutoriels Tableau](http://www.tableau.com/learn/training) [Site web]
(<http://www.tableau.com/learn/training>)
Liste de tutoriels en ligne pour Tableau

12 - Visualisation des données: Théorie et pratique #2 - Enjeux

Description



IMPORTANT

Discussion sur les différentes possibilités offertes par les outils de visualisation quant à l'exploitation et la transformation des données (création de hiérarchies, requêtage automatisé etc.).

Ouverture sur des enjeux et réflexions importants dans le domaine de l'intelligence d'affaires.

Activités/Ressources avant la séance



IMPORTANT

[Microsoft big data solutions](#) [Site web]

(<https://msdn.microsoft.com/en-us/library/dn749804.aspx>)

À lire pour comprendre certains enjeux courants qui créent des défis dans le monde du BI.



IMPORTANT

[Agile Best Practices for Data Warehouses](#) [Site web]

(<http://www.agiledata.org/essays/dataWarehousingBestPractices.html>)

Évolution de solutions BI d'après la philosophie Agile



IMPORTANT

Jeanne W Ross (2013-12-01). « You May Not Need Big Data After All », *Harvard Business Review*, vol.91, no.12.

 [Disponible à la bibliothèque](#)


Ressources générales



IMPORTANT

Hughes, Ralph (2013). *Agile data warehousing project management business intelligence systems using Scrum*, Waltham, MA, Morgan Kaufmann.

ISBN:9780123965172

 [Disponible à la bibliothèque](#)

13 - Présentation de travaux en équipe

Description



IMPORTANT

Présentation des travaux d'équipe (cf. la section Travail en Équipe).
Retour sur le cours et son contenu, révision dans le cadre de la préparation à l'examen final.

14 - Présentations de groupe et préparation à l'examen final

Règlements de HEC Montréal

Plagiat

Les étudiants sont priés de prendre connaissance des actes et des gestes qui sont considérés comme étant du plagiat ou une autre infraction de nature pédagogique, de la procédure et des sanctions, qui peuvent aller jusqu'à la suspension et même l'expulsion de HEC Montréal. Toute infraction sera analysée en fonction des faits et des circonstances, et une sanction sera appliquée en conséquence. [En savoir plus sur le plagiat...](#)

Calculatrices

Les étudiants sont priés de prendre connaissance de la politique d'utilisation de calculatrices lors d'examens lorsque celles-ci sont autorisées. [En savoir plus sur la politique d'usage de calculatrices...](#)