

Syllabus UE3 du Master ISIA D

Processus ETL

Vanderheyden Hervé

Business & Decision

9h00 cours + 24h00 TD

■ Objectifs

- Etude et pratique des outils d'alimentation d'entrepôts de données (Extraction, Transformation, Loading).
 - Illustration via le produit TALEND.
-

■ Contenu

- Historique et fonctionnalités des outils d'ETL
 - Problématiques de l'alimentation des entrepôts de données.
 - Mises-en œuvre des solutions via TALEND.
-

■ Bibliographie

- Ralph Kimball et Margy Ross: Guide pratique de modélisation décisionnelle
 - Ralph Kimball et Joe Caserta: The Data Warehouse ETL Toolkit
-

■ Pré-requis

- SQL
-

■ Contrôle des connaissances

- Examen
- Projet sous forme de ½ semaine bloquée

Base de données multidimensionnelles

J. Aligon

Université Toulouse 1 Capitole

6h00 cours + 21h00 TD

■ Objectifs

- Connaître les rudiments de l'optimisation de requêtes SQL, apprendre à utiliser les différents types d'index applicables aux bases de données multidimensionnelles pour optimiser les requêtes fréquentes et à vérifier l'efficacité de la stratégie d'indexation choisie
 - Être capable de concevoir des requêtes multidimensionnelles (par MDX)
-

■ Contenu

- Index B-Tree, Hash Table, Index BitMap, Clustered Index
 - Syntaxe et expression de requêtes MDX
 - Opérateurs de navigation pour MDX
 - Application à SQL Server
 - Application à MonetDB (stockage en colonne)
-

■ Bibliographie

- Database Systems : the complete book (Garcia-Molina et al.)
 - Database Systems : Design, Implementation and Management (Coronel et al.)
 - Physical Database Design (Lightstone et al.)
 - Fast Track to MDX (Whitehorn et al.)
-

■ Pré-requis

- Entrepôt de données & OLAP
 - Langage SQL
-

■ Contrôle des connaissances

- Contrôle TD (CC) + Projet sous la forme d'une ½ semaine bloquée
 - Examen final
-

Exploration et visualisation de données

Gil Nakache Viseo (P1) et **Emilie Legeay** Aosis (P2)

9h00 cours + 24h00 TD

■ Objectifs

■ P1

- Etude et pratique des fonctions utilisateurs et administrateur des outils de reporting et de visualisation de données.
- Illustration via le produit Business Objects, manipulé par les étudiants.

■ P2

- Etude et pratique des techniques de DataViz appliquées dans les outils de reporting et de visualisation de données.
 - Illustration via le produit Power BI, manipulé par les étudiants.
-

■ Contenu

- Les requêteurs et outils d'infocentre : positionnement et fonctionnalités
 - Le module Utilisateur Business Objects
 - Le module Designer Business Objects
 - Applications des techniques de Dataviz et de storytelling avec Power BI
-

■ Bibliographie

- SAP Business Objects BI 4.0 : The Complete Reference 3rd Edition (Cindi Howson)
-

■ Pré-requis

- Connaissances en SQL
-

■ Contrôle des connaissances

- Projet sous forme de semaine bloquée
-

Big Data Analytics

JM. Thévenin

Université Toulouse 1 Capitole

6h00 cours + 21h00 TD

■ Objectifs

- Rappeler les enjeux du Big Data et voir en quoi ils sont complémentaires avec la BI traditionnelle
 - Comprendre de rôle des composants entrant en jeu dans une architecture Big Data et ce que l'on entend par architecture 'scalable'
 - Comprendre le rôle des SGBD NoSQL dans ce contexte
 - Comprendre les principes des algorithmes de type Map-Reduce
 - (Découvrir le langage Python)
-

■ Contenu

• Cours

- Introduction : enjeux et cas d'usage des Big Data
- Scalability, Architecture Hadoop et architecture Spark
- Le modèle de programmation Map/Reduce
- L'éco-système Hadoop
- Objectifs et fonctionnement des différents SGBD NoSQL

• TD

- Découverte d'une machine virtuelle Cloudera simulant un cluster Hadoop
 - Utilisation des commandes Linux et HDFS
 - Découverte de l'environnement de développement Eclipse pour Java
 - Ecriture de programmes Map/Reduce en Java
 - Découverte du langage Scala, langage fonctionnel orienté objet dans lequel est implanté Spark
 - Utilisation de PIG, langage de haut niveau pour ceux ne maîtrisant la programmation Map/Reduce
-

■ Bibliographie

- "Hadoop The Definitive Guide", Tom White, O'Reilly editions
 - <http://bradhedlund.com/2011/09/10/understanding-hadoop-clusters-and-the-network/>
 - "Big Data Concepts et mise en œuvre de Hadoop", Laurent Jolia-Ferrier, ENI Editions
 - "Les Bases de Données NoSQL et le Big Data", Rudi Bruchez, Eyrolles
-

■ Pré-requis

- Gestion des données du SI
 - Algorithmique et JAVA
 - (Traitement statistique des données et datamining)
-

■ Contrôle des connaissances

- Projet sous la forme d'une semaine bloquée
- examen